DELPHI PRO

The Year Ahead for Infra 2025

by Ceteris, Facundo Indabera, Muhammad Yusuf, Prasad Mahadik

INFRASTRUCTURE

The Year Ahead for Infra 2025

Dec 12th, 2024 • 128 min read

Written by: **Ceteris, Facundo Indabera, Muhammad Yusuf, Prasad Mahadik**

Key Takeaways

- 1. Solana leads the high-throughput blockchain narrative.
  Solana consistently outperforms competitors like Ethereum and Base in DEX volumes and transaction throughput.
  Median fees on Solana remain significantly lower than peers, highlighting its efficient architecture as avg fees rise.
  Validator income from MEV and fees surpassed issuance in mid-Nov, marking a pivotal shift in economic sustainability.
  2. High-throughput chains are redefining blockchain infrastructure.
  Chains like Monad, HyperLiquid, and MegaETH push technical limits with parallelization and custom databases.
  Rollup designs like MegaETH achieve unparalleled speed but trade off decentralization.
  General-purpose chains aim for both performance and user experience.
  3. Rollup ecosystems are consolidating around leading players.
  Base and Arbitrum dominate L2 TVL and user activity, accelerated by EIP-4844 implementation.
  Rollup fragmentation continues towards Base & Arbitrum.
  4. Ethereum's evolving roadmap lacks unified direction.
  Proposals like Native Rollups and Beam Chain aim to maintain Ethereum's decentralization ethos.
  Challenges include aligning on priorities while improving scalability, gas limits, and block-building mechanisms.
  5. Solana ecosystem innovation fosters a vibrant developer community.
  Projects like Firedancer and Anza enhance throughput and decentralization for Solana.

Temporal introduces fee-based models as a fix to swQoS
Solana fastest growing dev ecosystem.
6. MoveVM chains like Sui and Aptos emerge as Solana competitors.
Sui introduces innovative object-centric transaction models, allowing for consensus fast-path.
Aptos builds on original MoveVM principles, with growing ecosystems and liquidity.
Both chains highlight increasing adoption in Asia and beyond.
7. Modular and appchain ecosystems gain momentum.
Platforms like Initia and Movement integrate appchains and liquidity hubs for tailored experiences.
Initia targets Cosmos-like interoperability, while Movement has a larger focus on their general purpose chain.
Emerging appchains highlight the trade-off between ecosystem growth and operational complexity.
8. ZK rollups drive scalability with security.
Starknet and zkSync lead with distinct approaches; Starknet focuses on Cairo, zkSync on EVM & Elastic Chains.
Native rollups on Ethereum aim for deeper protocol-level ZK integration.
9. Modular ZK infrastructures expand.
Projects like Bullet and Nil use ZK technology to enhance performance and decentralization.
Polygon's AggLayer fosters liquidity and cross-chain interoperability via ZK-backed systems.
10. ZK faces centralization challenges.
Reliance on few zk proof vendors risks centralization, prompting Ethereum to diversify zk builders.
Beam Chain and zkSNARK integration aim to secure scalability within decentralized frameworks.
11. Industry-wide focus on high-throughput scalability intensifies.
Trends favor performant, general-purpose state machines like Solana over niche appchains.
Centralized sequencer models enable unparalleled speed but raise questions about long-term decentralization.

Performance competition remains a critical factor for dominance.

12. Infrastructure shifts signal broader implications for DeFi.

DeFi TVL trends show a steady rise for Solana and competitors like Sui, challenging Ethereum's dominance.

Projects integrating AI, stablecoins, and advanced UX highlight blockchain's broader use cases.

Stablecoin liquidity and cross-chain interoperability remain critical growth areas.

Table of Contents

Blockchain/acc – Time to go Fast

The industry is shifting to high throughput chains. This is not just about Solana, whose meteoric rise over the past year has led the thesis, but also alt L1s like Sui & L2s like Base who have surged in 2024, along with the numerous performant chains like Monad, HyperLiquid, Unichain, MegaETH and more all in the pipeline.

> unichain is not confirmation of the appchain thesis. it is confirmation of the performant, gp chain thesis.
>
> — ceteris (@ceterispar1bus) October 11, 2024

Users have shown that the main thing they've desired to use blockchains for is a good UX on a single general purpose shared state machine. There's really no better example of this than the recent AI meta. All of the activity is on Solana and Base. Not "AI focused" L1s like TAO, NEAR, or ICP. They are happening on two general purpose chains that have only aimed to create a shared state machine that can handle a lot of activity. Narratives will come to you if you can create this.

> yep.
>
> when you create a product ppl want to use, new narratives will just come to you without being forced.https://t.co/XdnWcD4jEA
>
> — ceteris (@ceterispar1bus) November 24, 2024

Even Ethereum has recently started to raise its gas limit with a re-focus on L1 execution and is seeing new protocols designed to speed everything up.

> // thread
> Introducing TOOL: Trustless Orderflow Operations Layer

> Making Ethereum 12x faster with 1-second execution confirmations.
> No protocol changes. No L2s.
>
> Launching MVP in Q1 2025 with support from industry leaders.
> pic.twitter.com/oI9OPUzjEX
>
> — 0xprincess (@0x9212ce55) November 25, 2024

While people will debate the merits of appchains, so far we have not seen demand in practice outside of a few perp chains (Hyperliquid, dYdX), and while there *are* some interesting new L1s like Initia, Delta, Pod, Hyle and more that we will discuss later in the report, the next year is all about performant chains.

> For some reason people believe that there can be only one high throughput blockchain, when the reality is that there will be none low throughput blockchains.
>
> — toly ■■ (@aeyakovenko) May 4, 2024

Lastly, you should note that the chains in this section are a mix of L1/L2 and monolithic/modular. These constructions have different tradeoffs but are all trying to make the same thing: **a fast shared state machine.** The next sections will go over such chains.

- **Solana**
- **High Throughput EVM:** *Hyperliquid, Monad, MegaETH, Unichain, Rise, Sonic, Ithaca, Nil, Berachain*
- **High Throughput SVM:** *Eclipse, Atlas, Soon, Fogo*
- **High Throughput MVM:** *Sui & Aptos*

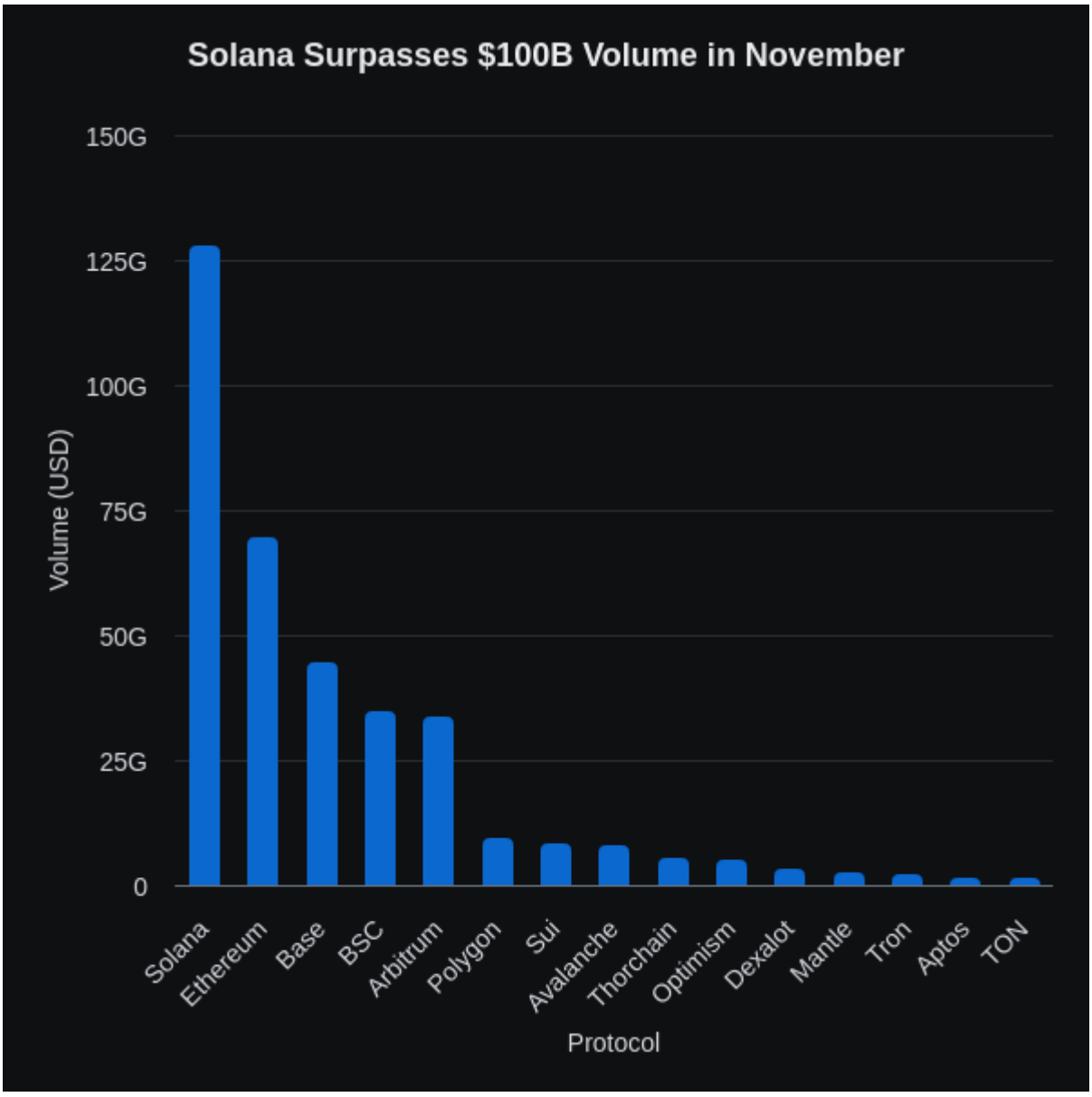
Solana – Increase Bandwidth, Reduce Latency

Solana Continues to Win

While we have grouped the other chains into EVM/SVM/MVM buckets, Solana has been the trend setter in the performant chain thesis and stands alone with a big

lead over the newer competitors (most of which haven't launched). All of Solana's metrics experienced rapid growth in 2024, and the introduction of so many new performant chains in the pipeline, is proof that Solana has been correct.

When we wrote about Solana in the 2024 Year Ahead report it was only one month into what turned into a banner year. DEX volumes had just cracked $5B weekly, spiked to ~50% of Ethereum L1 and priority fees were just starting to see usage. Today? Not only are Solana DEX volumes now more than Ethereum L1, in November they were comparable to Ethereum and all L2s combined, and it surpassed Ethereum's ATH DEX Volume of $118B from May 2021 (although below BSC's $138B).

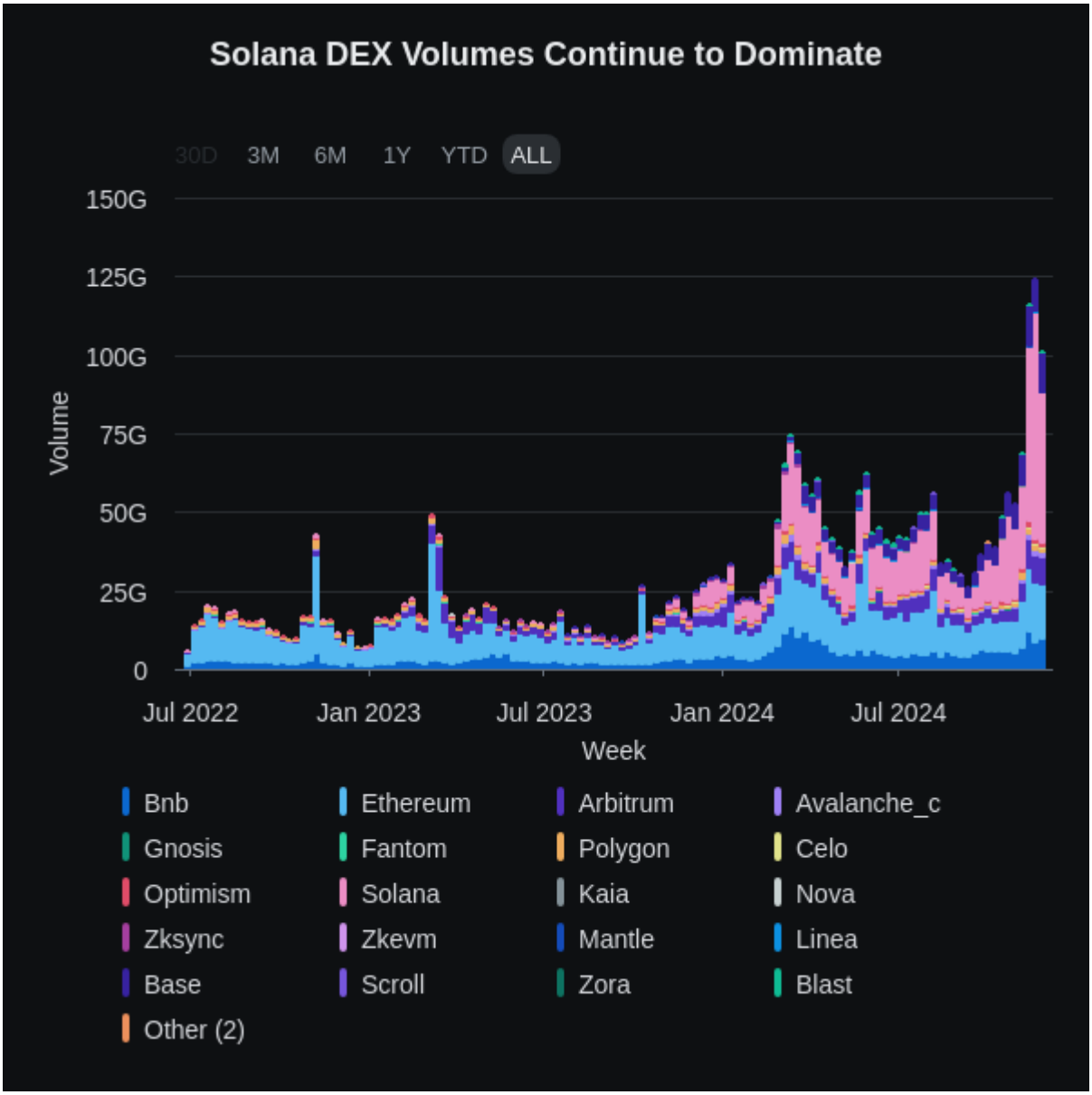


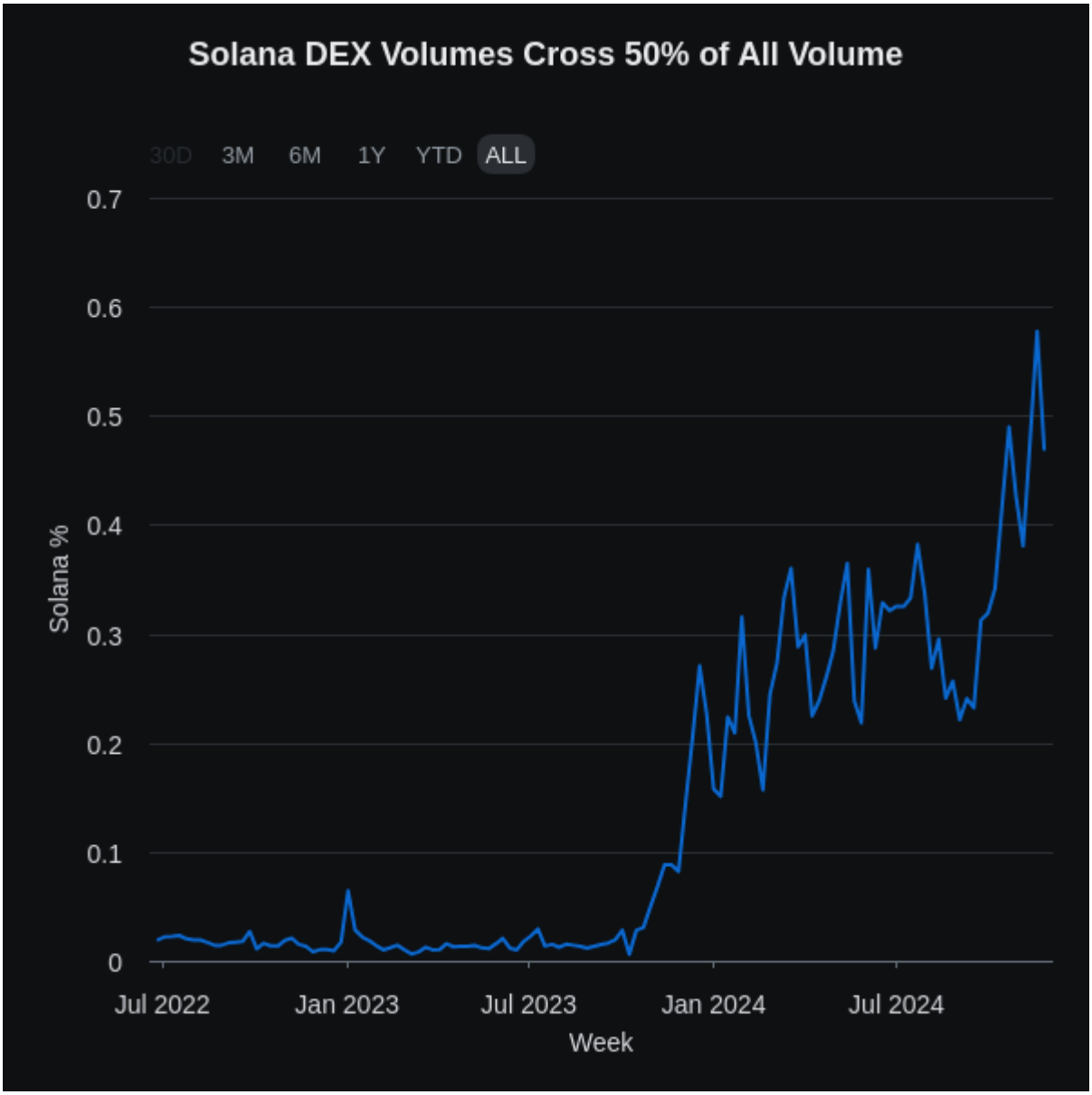
This is even more impressive when you realize every L2 runs a single centralized sequencer that doesn't need to participate in consensus (a process that

inherently degrades performance).

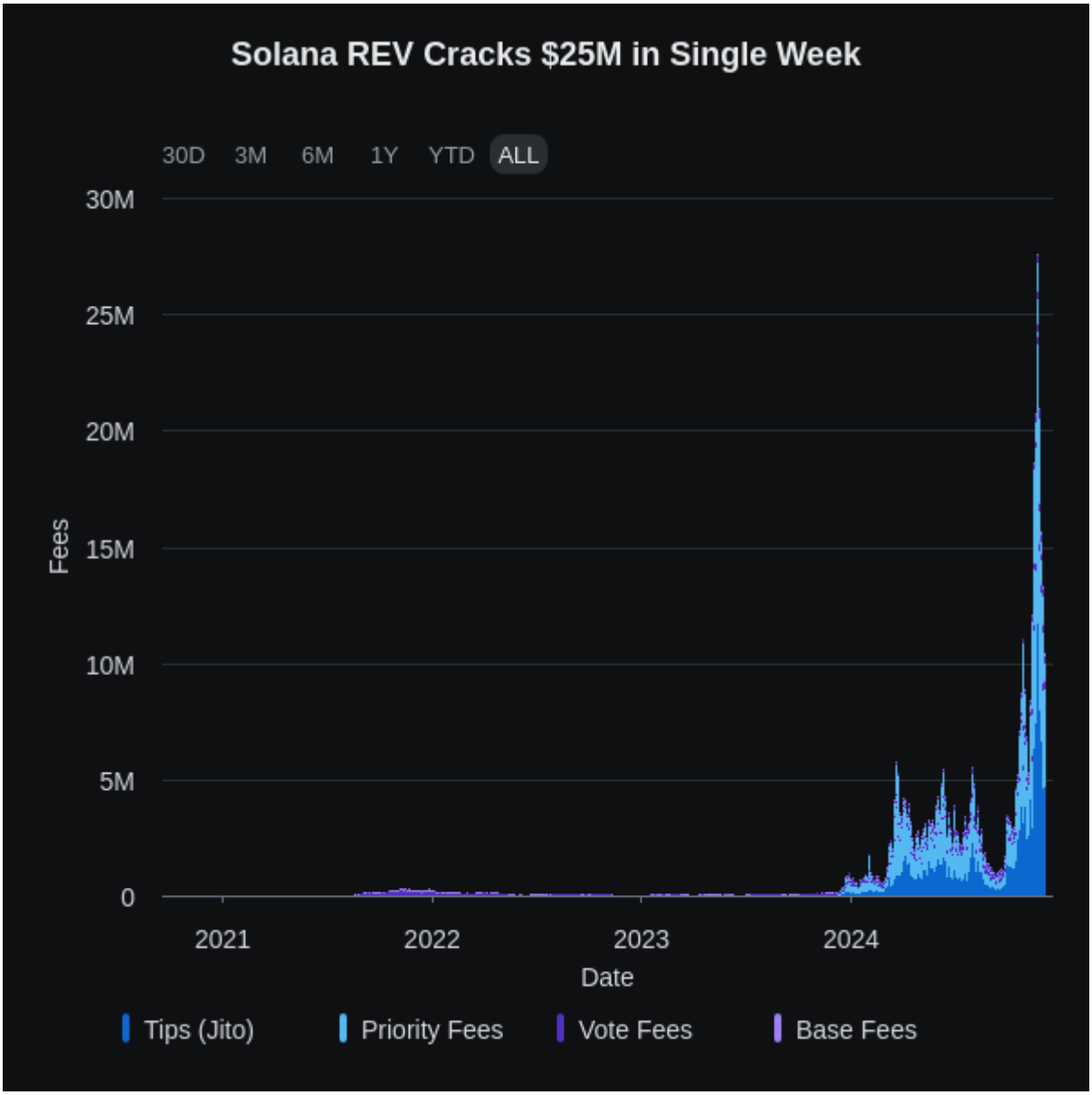
> a perfect demonstration of solana's engineering
>
> in the past day, for fees:
>
> Solana avg: 4 cents
> Base avg: 10 cents
>
> Solana median: 0.2 cents
> Base median: 2 cents
>
> the catch? Base did 6.5M txns, while Solana did 92M txns
>
> so — Solana has a 2.5x lower avg cost and a 10x lower...
> pic.twitter.com/RV7fUizbCM
>
> — mert | helius.dev (@0xMert_) December 1, 2024

When compared to Ethereum volumes over the past two years, it is now on par or surpasses them monthly, and its dominance as a percentage of total DEX volume continues to climb, now accounting for nearly 1/3 of all spot trading volume on blockchains.

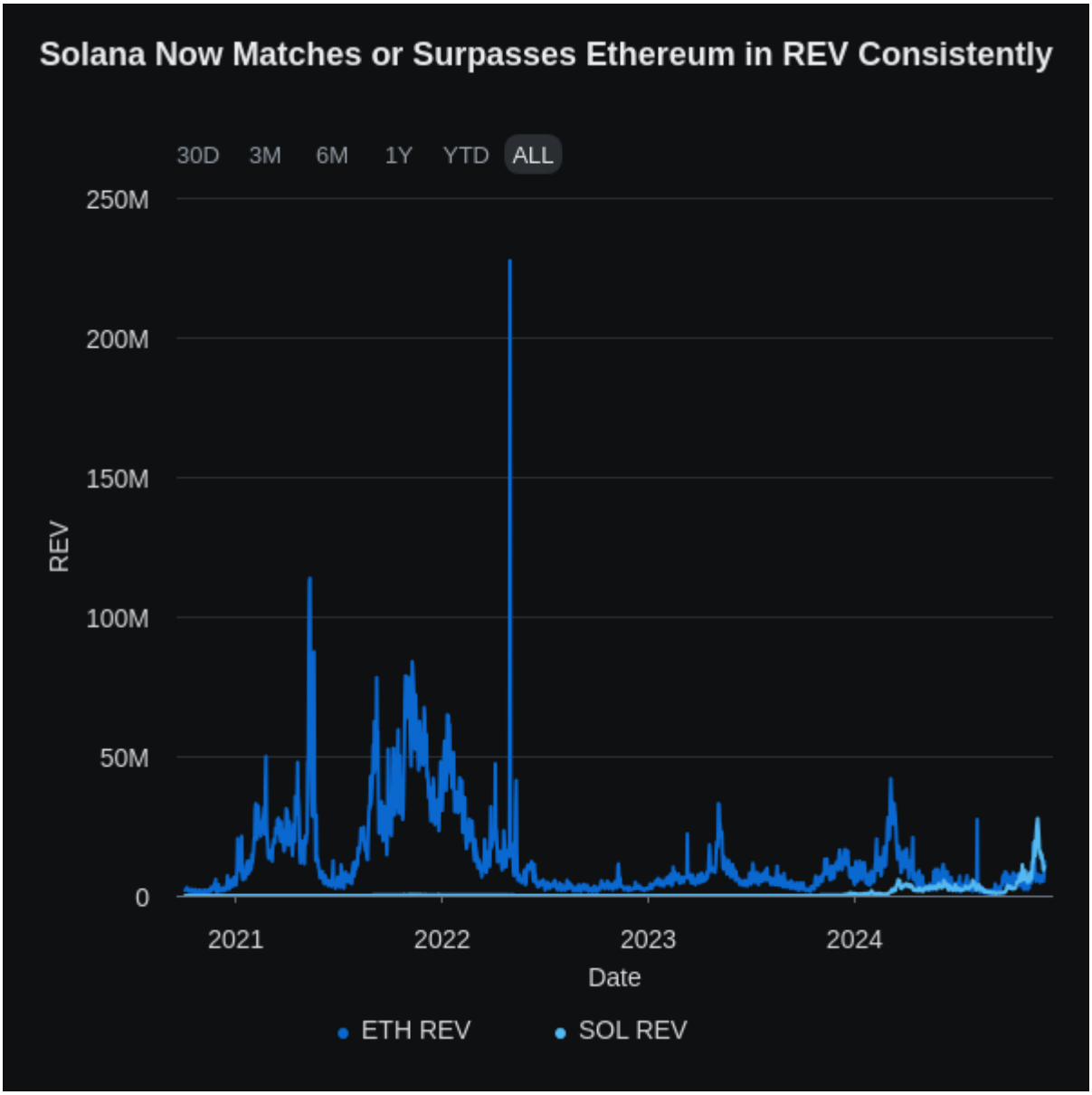




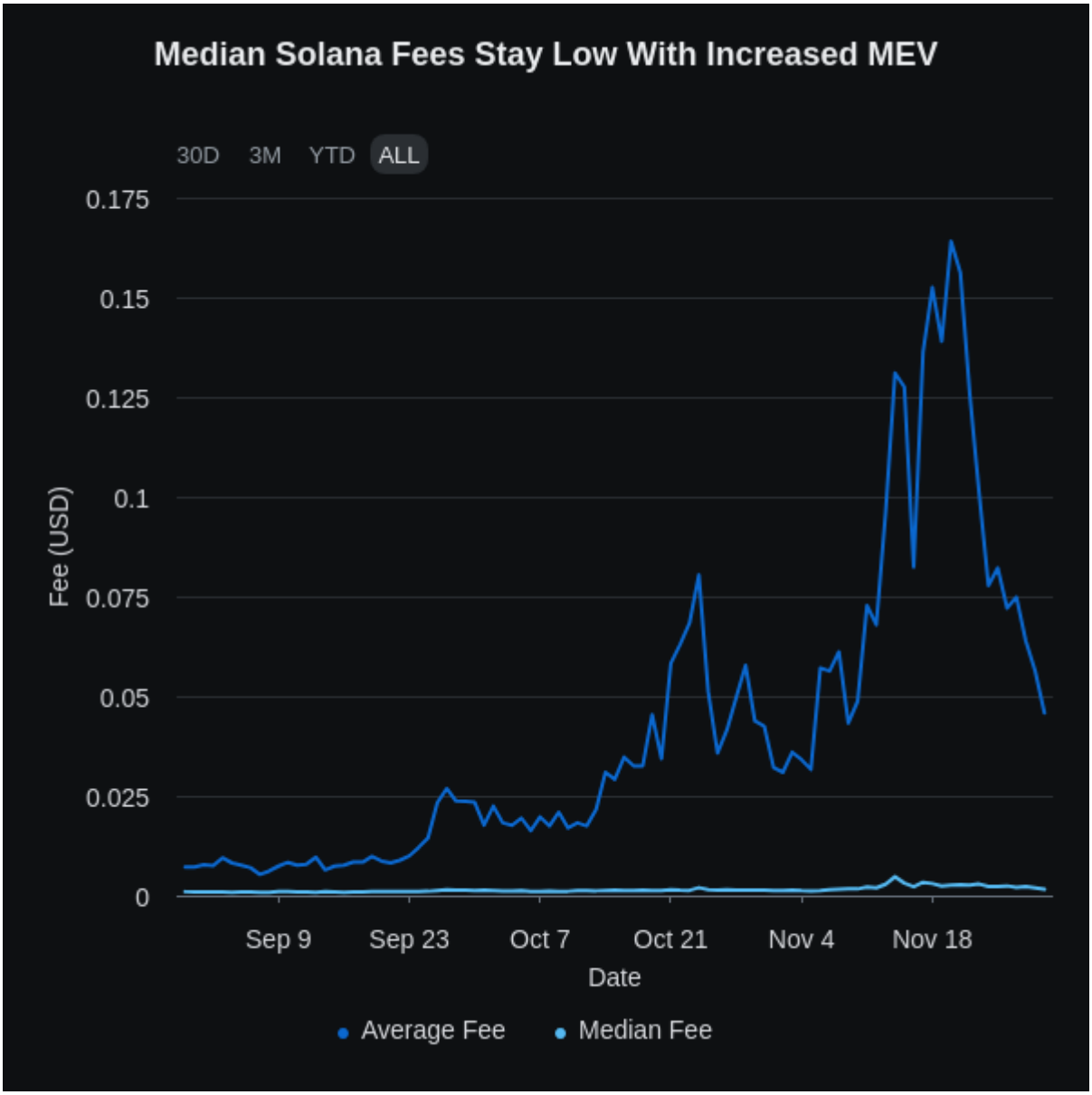
This has, of course, led to an explosion in protocol revenue. For years the main critique against Solana was that fees were too cheap to sustain the blockchain. This was faulty logic as it ignored the convexity that fees exhibit when you create valuable state. Solana's REV (total income from various fee sources) hit an ATH recently of >$25M in a week and is continuing to march up and to the right.



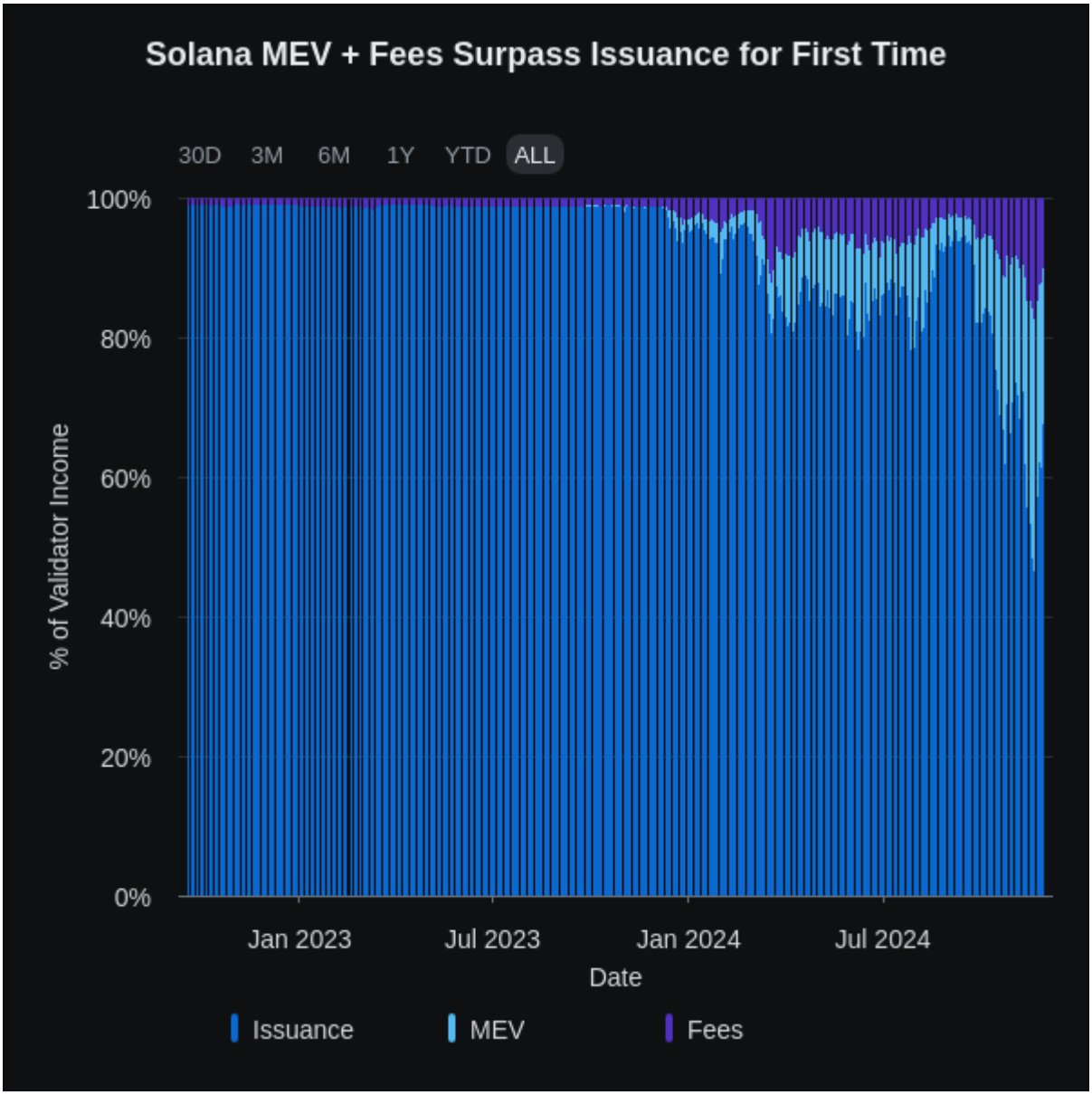
This momentum has propelled Solana to both tracking and surpassing Ethereum consistently, although recently Ethereum came back ahead (h/t to Blockworks for the data). Make sure to play around with the date ranges on these charts to get a picture of the all-time and recent change.



Also notable is that this hasn't affected *all* fee paying users as it would with a global fee market. The median fee on Solana still stayed sub-cent during this period (votes, transfers, etc).



Lastly, validator income from MEV and fees surpassed income from issuance for the first time ever in November. When we published "Solana the Monolith" in May 2023 this was nearly entirely issuance. There are calls for Solana to lower issuance further in light of this— keep an eye on this SIMD. With that being said, I don't think issuance rate matters too much (it's just a tax on non-stakers) and would refer people to this piece by Jon Charb if you disagree.



The critique against Solana now is that this activity isn't sustainable because it's all memecoin trading. Listen, I get it, but this is once again adding qualifiers to any Solana metric that proves another old thesis wrong. The reality is this: Solana is the place to trade. Yes, it is dominated by memecoins now, but there is no reason why it will stay that way in the long-run, and if this is your main critique then I believe you are missing the forest for the trees. Memecoins aren't any "less real" than other assets. **The market trades what it wants to trade, and the fees paid to trade them are very real.** I too hope we move past memecoins being the dominant assets traded but as of today that is just the reality, and they've become a good stress test for Solana.

People will also compare Solana to chains like BSC, Fantom and Polygon that gained users when Ethereum gas prices spiked last bull, and then lost them during the bear. Again, this is a flawed comparison; Solana went through an

explosion in activity all while Ethereum gas prices were 1-2 gwei. They were not pushed to Solana because of high fees elsewhere. Yes, low fees are an important part of Solana, but there's way more to it than low fees. And while you can hate memecoins, it's not like there was another meta out there that was gaining attention. If something new comes along and Solana doesn't capture the volume, then we can say Solana has lost ground.

If you need an example for why Solana is the place to trade, look at the recent AI meta. Agent AIXBT, an agent deployed on Base, had a significant amount of its volume traded on Solana. Solana users prefer it so much that they would rather trade a wrapped asset issued by a third party bridge (Wormhole) with much weaker security properties than use the chain where it was originally issued.

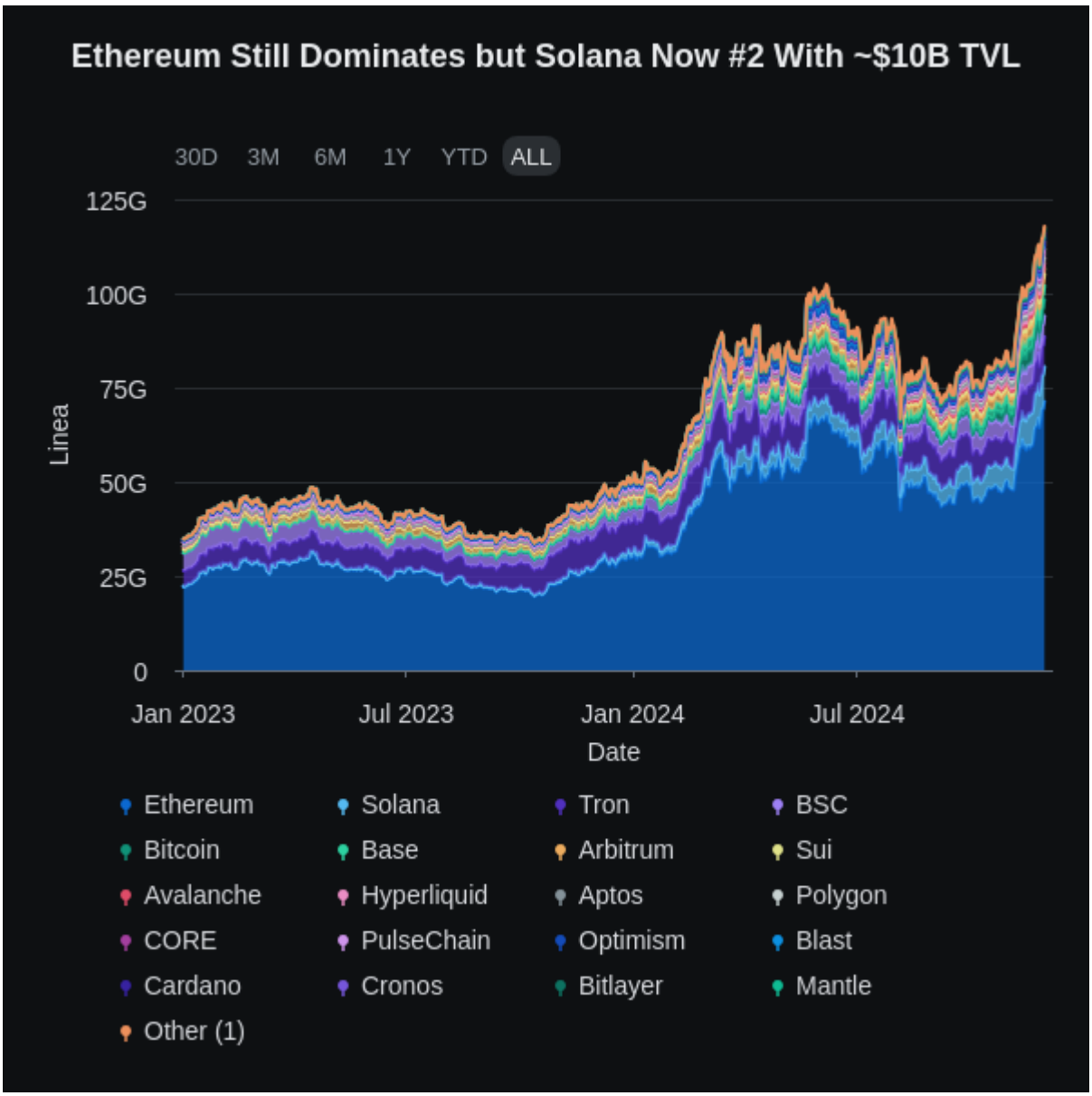
> rollup roadmap:
>
> theory – issue assets on eth l1, bridge to l2. can escape hatch back to L1 if needed
>
> practice – issue asset on l2, trade it on another l1 through a third party PoA bridge
>
> — ceteris (@ceterispar1bus) November 27, 2024

We saw a more extreme example of this happen when Runes on Bitcoin were going through their hype. The top traded Rune had the majority of its volume on Solana, not Bitcoin or any other chain. Bitcoin is a terrible trading chain so it makes sense price discovery would happen elsewhere, but it's notable that it's Solana. Solana is simply the preferred destination to trade today. Even large Ethereum NFT projects like Pudgy Penguins have chosen to launch their token on it. Solana is simply an L2 to every other chain that issues assets.

> everyone looking for bitcoin l2's don't realize we already have one and it's called solana pic.twitter.com/DrVxNXPhdy
>
> — ceteris (@ceterispar1bus) April 14, 2024

The main area where Solana still lacks is stablecoin TVL and by extension,

volume. This is a metric you would want to see continue to rise along with other non-meme (DePIN, RWA, DeFi, etc) volumes in the years ahead as FOREX is the largest market opportunity in the world. Also, while users prefer Solana to trade, a lot of capital still prefers to use stables and keep assets on Ethereum as it is home to most of the best DeFi applications today. While Jupiter is the best spot trading app in crypto, the rest of Solana DeFi is still well behind their Ethereum competitors. There is still a lot of room for Solana DeFi to improve, however TVL *has* risen steadily from a year ago and now sits #2 at ~$10B after being outside the top 10 just 1.5 years ago.



DeFi TVL also exhibits a sort of lindy effect. The longer people have assets in DeFi protocols the longer they will trust them, less likely they are to close out CDPs (and incur taxable events), and less likely to bear the operational overhead of moving assets. People continue to point to Ethereum vs Solana's TVL as a knock

against Solana. And it's true, Ethereum's is much larger, but this gap continues to move in one direction. If you used this argument a year ago when Ethereum had 40x Solana's TVL, you would have missed this moving down to 7.7x. I see no reason why this multiple doesn't continue to decrease over the next year.

The most important question now is, how does Solana maintain this lead? If crypto is truly entering its high throughput chain era, then how will Solana fare with so many competitors coming after the king? The answer is simple, but not easy: *increase bandwidth, reduce latency*.

How Solana Will Keep Winning

One commonly used argument against Solana's moat is that "there will always be a faster chain that comes along". This is a bad argument to me, for a few reasons.

1. SVM was a material improvement over EVM. Other VM's may provide some improvements over EVM but less against SVM (specifically a parallelized VM with no token approvals).
2. SVM is fastest growing alt-VM eco. Like Ethereum and EVM network effects, Solana is starting to exhibit the same. There are numerous new SVM chains we will discuss. This all creates a larger developer base.
3. You can't cheat physics. While centralized L2s with a single sequencer will be able to move faster than Solana, it is not clear if another globally distributed L1 can. With Firedancer, Anza improvements, and a ruthless dedication to IBRL, it's hard to see Solana be outcompeted. Toly and the team have always stayed laser focused on their mission to be a globally decentralized NASDAQ; the thesis hasn't changed.
4. The ZK argument: ZK can't speed up native consensus. Solana's goal is to be the most performant single state machine, something ZK cannot help with.

> The black pill is that if the only interop we need is sending USDC between giant state machines then ZK proving doesn't do anything.
>
> — Zaki ■■■ (@zmanian) November 24, 2024

With all that being said, success is by no means guaranteed. Solana is far from perfect, still has a lot of issues, and there are a lot of worthy competitors

approaching.

> MILLIONS of dollars in MEV are siphoned away DAILY by vpe (aka arsc), the top sandwich bot. But the worst part? This isn't just about MEV—it's about centralization of MEV. And it's the greatest threat to Solana's decentralization I've seen.
>
> Solana's volume is skyrocketing, and... pic.twitter.com/v6mlSQhtg1
>
> — Ben ▮ (@HypoNyms) December 10, 2024

Some key teams/initiatives for Solana are below.

Temporal

A research firm founded by ex-HFT, TradFi and MarginFi contributors (like Ben from above), Temporal looks to fix Solana's fee markets. Their main priority is to replace swQoS with a fee based model (tl;dr is swQoS gives priority to amount of stake whereas fee market is simply fee paid). Their goal is to create a new transport layer for Solana replacing QUIC and UDP.

They recently announced Nozomi, a new fee based way to land tx's on Solana (only pay tip if your tx is landed first). This product exploits Durable nonces to create a better priority fee mechanism. The team has a lot of experience and understands the infrastructure and nuances deeply – they are one of the main infra focused teams to watch.

Anza

While Firedancer gets the hype, the team behind Anza should not be written off or taken lightly. It's not like the Anza team is just going through the motions and maintaining Solana's validator client until Firedancer comes.

> If we are not faster than firedancer in a year I will quit my job
>
> — Alessandro Decina (@alessandrod) July 22, 2024

Anza recently released the Agave 2.0 client, with it some notable

improvements/changes:

- Full priority fee now given to validators instead of burned
- Central scheduler now on by default. The new scheduler went live a few months back and looks to improve network jitter (full breakdown)
- A new ZK ElGamal Proof program and new syscalls
- A break from old labs client and moves Solana towards multi-client world

And it's not like Anza isn't focused on IBRL.

> definitely nothing pic.twitter.com/SpXDllgYak
>
> — BW (@bw_solana) December 2, 2024

And they just got talent from Ethereum.

> Last week was my last week at Consensys. Today is my first day at @anza_xyz.
>
> I'm taking my talents to Solana.
>
> In my first 100 days, I plan on writing a spec for as much of the Solana protocol as I can get to, prioritizing fee markets and consensus implementations where I...
>
> — Max Resnick (@MaxResnick1) December 9, 2024

Firedancer

Frankendancer went live on mainnet in September. As of now they do not have Jito support so uptake has been slow. Once that is implemented expect the number of validators running it to increase significantly. The full client is running on testnet. You can watch a live feed of Superteam Germany's validator here: https://fd-mainnet.stakingfacilities.com/

Jito

The MEV backbone of Solana, Jito has become Solana's most profitable protocol. All of the fee/MEV charts from above are attributable to Jito's infrastructure. They recently released Jito Restaking which you can read more about in our Restaking report. As mentioned above, they are not yet integrated with Franken/Firedancer.

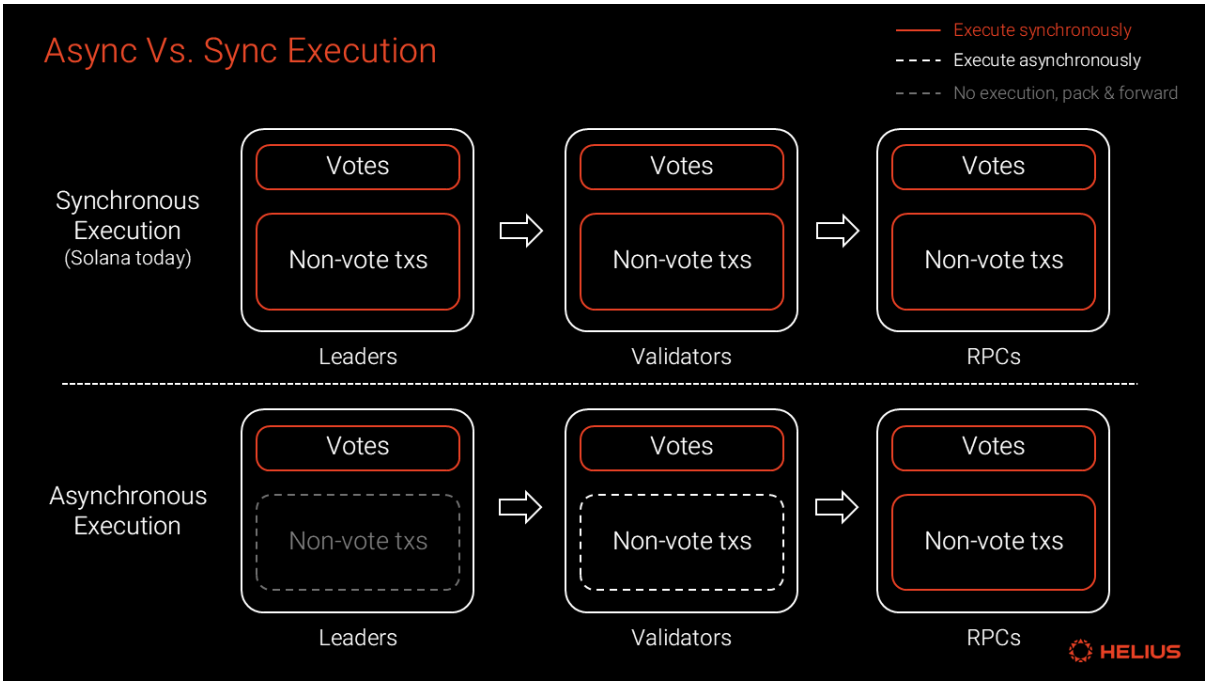
Wallet Infra

Solana's wallet infra continues to improve with teams like Squads leading the charge. They recently introduced virtual US banks accounts, allowing users to send directly between onchain and offchain without a CEX.

> Introducing virtual US bank accounts on Fuse.
>
> Powered by @Stablecoin, ▮your virtual bank account converts USD payments directly to USDC in your Fuse wallet.
>
> • Accept regular bank transfers
> • Skip multiple platforms & transfers
> • No expensive on-ramp fees pic.twitter.com/9Qvcban8Ke
>
> — Fuse (@fusewallet) December 4, 2024

Asynchronous Execution

An ambitious challenge, async execution allows validators to create blocks without executing them. There are numerous advantages to AE like shorter block times, lower validator requirements, faster finality, a pathway to MCP (multiple concurrent proposers) and more. A good breakdown can be found here by Helius.



Mithril

Created by the Overclock Validator team, Mithril is a Solana full node client that aims to make home verifiability plausible – one of Solana's largest critiques. They recently reached a new milestone.

New Ideas Replicated Elsewhere

One clear signal of an innovative ecosystem is when its products not only succeed but are also replicated elsewhere. The main ones that come to mind here are Pump.Fun, MetaDAO and AI Agents

- Pump.Fun gets a lot of hate for being memecoin focused but pump just gave users something they wanted with standardized regulation (stopping rug pulls). Competitors like clanker.fun and wow.xyz have launched on Base
- MetaDAO was first to launch Futarchy in prod and is now being replicated on Ethereum
- AI Agents & platforms took off on Solana & Base, not "AI focused" chains

DePIN & DeAI

More on this in the DePIN & DeAI Year Ahead report, but Solana is home to numerous DePIN networks like Helium, Hivemapper, Pipe, Dawn, and Teleport, and it's also leading the AI agent narrative, with other infrastructure projects such as grass

Numerous SVM Stack Chains

Eclipse, Soon, Atlas, Sonic and more are all building on the SVM stack, potentially increasing SVM network effects.

ZK & Modularity

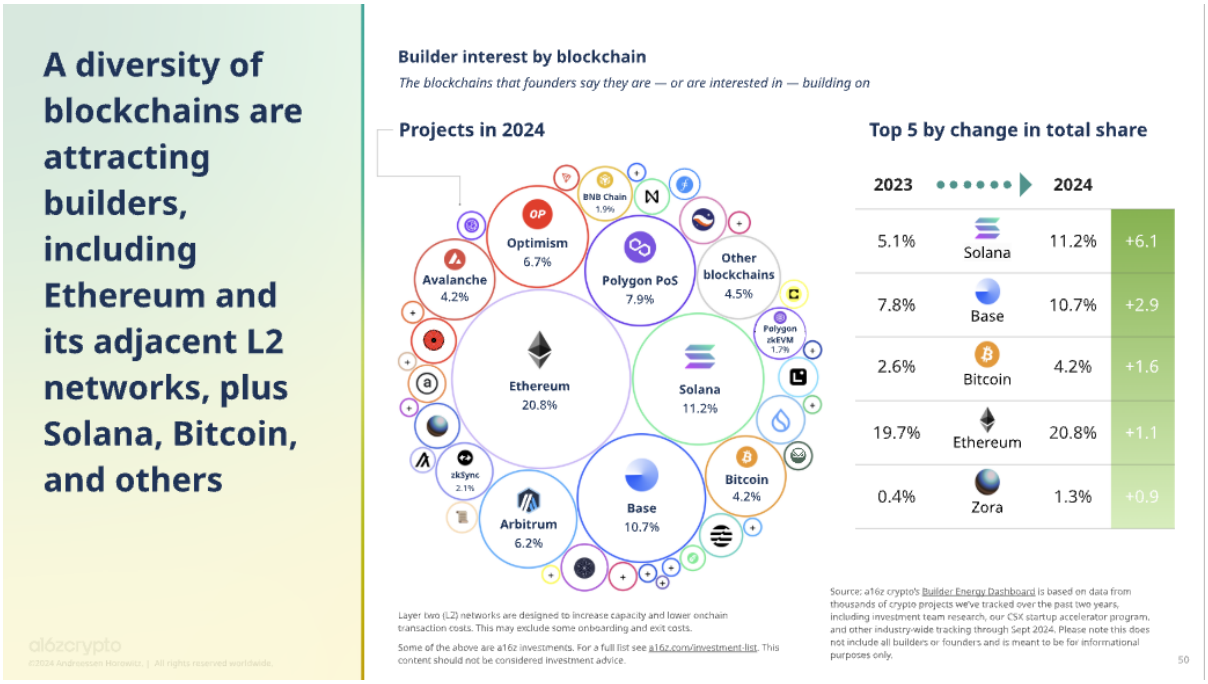
For the full breakdown just skip to the ZK deep dive part of this report and read "Solana the Modular" from August. The main project I wanted to highlight here is Bullet.

> Check out @Tristan0x 's article for an inside look at how Bullet works and how it stacks up against other high-throughput L2s launching this year https://t.co/N3jNV1N394
>
> — bullet (@bulletxyz_) November 25, 2024

The tl;dr is that Bullet is a perps appchain on Solana built using the Sovereign Labs SDK. They have a bridge to Solana and use it for DA, posting both ZK and fraud proofs to Solana. They have achieved sub 50ms latency with a target of 5ms – something you can only do with a centralized sequencer. As noted in the intro, perps dex's are the main proven use case for appchains and so it makes sense Zeta is going with this architecture. They will be the first to try this on Solana and will be able to seamlessly tap into Solana's native liquidity. After the success of Hyperliquid, it will be one to watch to see how they compete as Hyperliquid moves towards a general purpose L1; Bullet is VM-less, it's just Rust.

Growing Developer Interest

Per a16z's "State of Crypto 2024" report, Solana saw the largest jump in developer interest from 2023 to 2024 and sits at #2 in a sea of EVM chains. As we've said before, while Ethereum is in the lead, all of Solana's metrics continue to go up and to the right. It's also important to remember Ethereum launched 6 years (2014 vs 2020) before Solana.



DoubleZero

Announced on December 4th, DoubleZero is essentially trying to create a new internet for high performance blockchains (L1s and L2s). While not part of Solana itself, DZ is dubbed as "the new internet for modern distributed systems" and is built by ex-Solana Foundation employees with contributors from Firedancer and Malbec Labs. Basically, the current internet wasn't fast enough for blockchains and so they're trying to build a new internet. IBRL.

ETFs

While not related to network infrastructure, ETFs are a sign of institutional adoption and are a significant source of flows. There are 4 ETFs now on the table for Solana with the final decision date in August 2025. It is very possible with the new administration we will see a Solana ETF.

Job's Not Finished

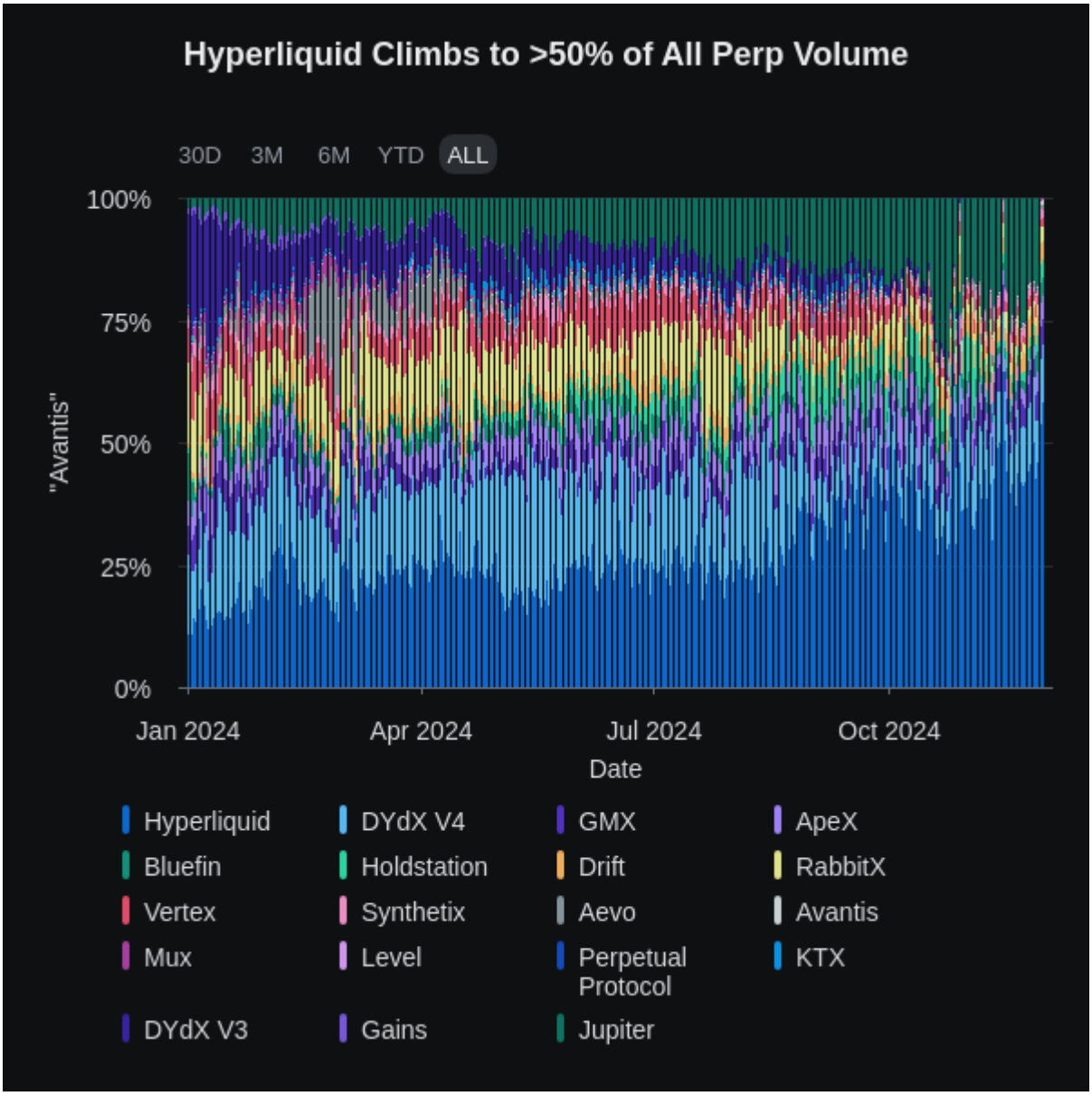
While Solana has had great momentum over the last year, and there are reasons to expect them to continue delivering, the competition is growing fast and Solana's continued success is by no means guaranteed. The next three sections are dedicated to these competitors. As mentioned above these are a mix of L1s and L2s, and while single sequencer chains don't compete towards being the dominant global shared state machine, they do compete for users and liquidity of various applications.

High Throughput EVM Chains

Hyperliquid, Monad, MegaETH, Unichain, Ithaca, Rise, Sonic, Nil, Berachain

Hyperliquid (L1)

Out of the chains listed here Hyperliquid is the only one live, so it makes sense to talk about them first. Hyperliquid in its current state is not a general purpose chain, it is a perps appchain that has seen remarkable success in a competitive sector and now accounts for >50% of all perp volume. The perp landscape has seen numerous protocols take top spot like Synthetix, Perp Protocol, GMX and dYdX over the years with none have been able to maintain the lead. The competition is fierce, but that's because the opportunity (Decentralized Binance) is so big.



Why are they in this section then if they're an appchain? Because they are moving to create a general purpose high throughout EVM L1. However, we should categorize it correctly today. It is not a decentralized perps exchange, it is a centralized, permissioned chain run by the team's 4 co-located validators in Tokyo. They are aiming to *transition* to a decentralized, open source L1, but that is not what they are today.

> BTW I have no issue with such products/services existing, I am even open to the idea that 'CeDeFi' or 'onchain CeFi' could be better/more sustainable than DeFi (and market seems to be telling us that), but things should be properly labeled. . .
>
> — _gabrielShapir0 (@lex_node) December 1, 2024

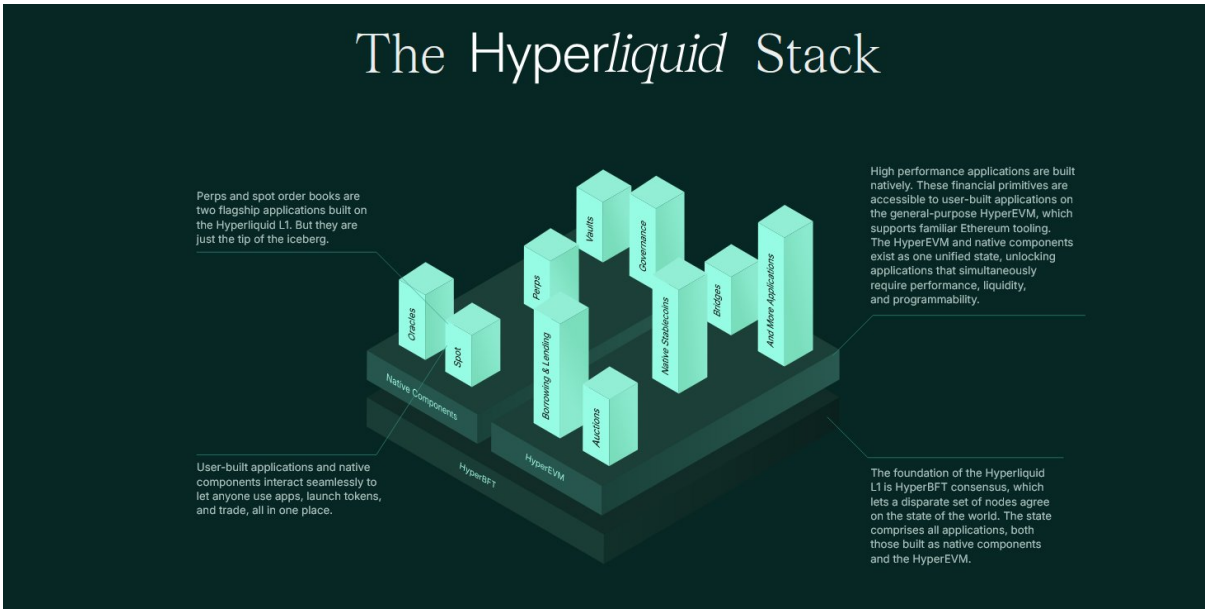
With that being said, they have done something rare in launching a new protocol without any funding and enriching a community of diehard supporters. They are kind of building backwards but now have a massive valuation to fund their vision. Hyperliquid is a good example of an appchain being successful and wanting more; wanting to become their own gp chain.

> many such cases https://t.co/tpS4LBJGR9 pic.twitter.com/NZm9wtV40M
>
> — Jon Charbonneau ■■ (@jon_charb) January 27, 2024

The challenge for Hyperliquid now will be how they transition to a decentralized L1 with a globally distributed validator set, all while maintaining the same UX that got them so many diehard users. If you talk to any power user of Hyperliquid they will tell you that it is the best trading experience in the market. Maintaining this experience will be harder as they decentralize and by extension network latency increases. They now have a testnet with more validators outside of the team run ones.

Hyperliquid created their own consensus mechanism HyperBFT after pivoting from Tendermint in earlier days. Internal testing with co-located validators shows an end-to-end latency median of 0.2 seconds and 99th percentile at 0.9 seconds. How this performs with a larger, globally distributed validator set is TBD.

Their EVM solution is unique and will be different than others on this list. Hyperliquid's perps and spot order books are all built in Rust and will continue to be isolated to their own environment. The EVM is being added as a generalized VM on top. The native VM is permissioned and tailored to their app, the EVM is permissionless and open to all. The challenge in having two distinct VMs is fragmenting liquidity between the two and getting them to work together seamlessly; the benefit is that EVM congestion shouldn't impact perps performance.



There's no question Hyperliquid has been a resounding success to date, and their go to market with a 30% community airdrop gives them a strong & dedicated user/holder base moving forward. L1s are never community owned, especially in the post-ICO era. There are significant technical hurdles to pass, but they now have a massive valuation and there are a lot of catalysts/narratives like leading perp dex, leading high throughput EVM chain, and best L1 distribution to help them get there.

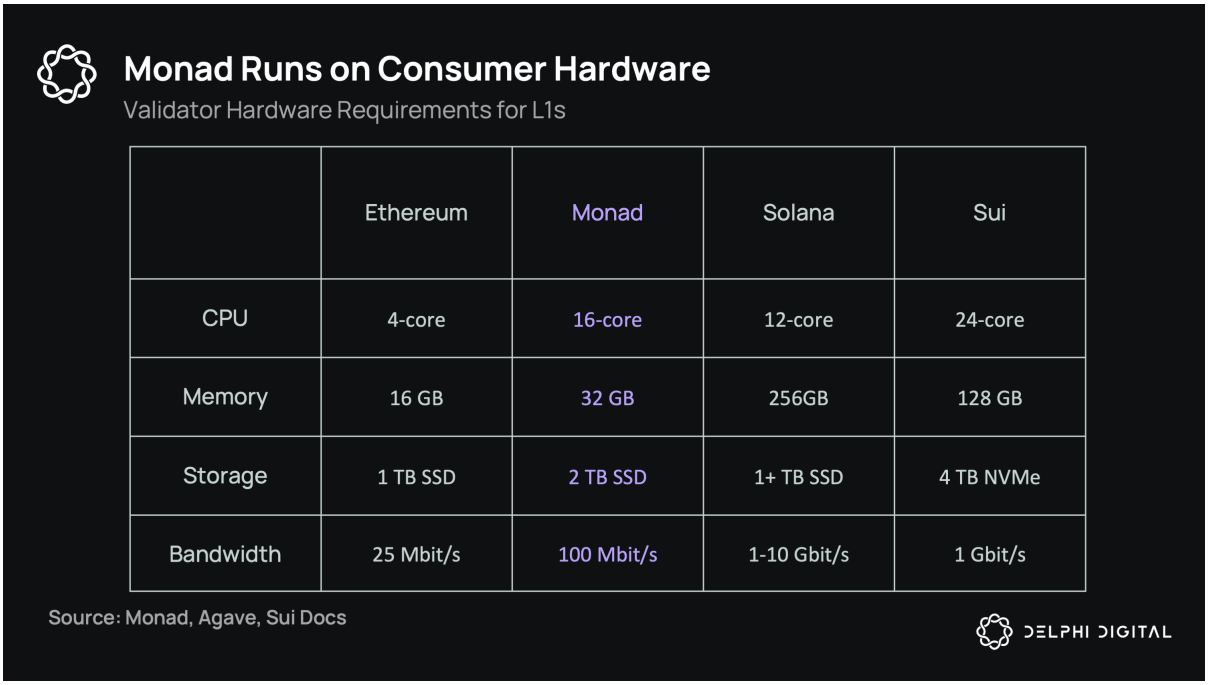
There is the other question as to if Hyperliquid should even *try* to decentralize meaningfully. Their product is good because it's centralized, and users don't seem to care, not only with Hyperliquid but also L2s governed by multisigs. It's not clear if the drawback of decentralization is actually worth the degradation in performance, especially as we transition to Trump's America.

A lot of people will point to Hyperliquid's airdrop as a successful way to launch a new protocol, but the reality is that they built something users just *really* wanted to use. You can't airdrop your way to success. Make sure to read the 2025 Markets Year Ahead for their take on the product.

Monad (L1)

While Monad frequently gets comped to Berachain, MegaETH, and now Hyperliquid, outside of them all being EVM they are quite different. Monad is building the most performant EVM chain they can with consumer hardware. What does this mean? When you decide to build a blockchain, first you must pick the hardware you will use to run the chain. On one end you have Ethereum with very low requirements, on the other Solana with significant requirements. Monad is

building for a more reasonable middle ground, being able to validate the chain on something like a MacBook Pro. They are then parallelizing the EVM. Monad is a possible long-term Ethereum L1 construction.



	Ethereum	Monad	Solana	Sui
CPU	4-core	16-core	12-core	24-core
Memory	16 GB	32 GB	256GB	128 GB
Storage	1 TB SSD	2 TB SSD	1+ TB SSD	4 TB NVMe
Bandwidth	25 Mbit/s	100 Mbit/s	1-10 Gbit/s	1 Gbit/s

In last year's year ahead report we wrote about Monad's 4 optimizations:

- *MonadBFT:* A derivative of Hotstuff, MonadBFT is a high performance consensus mechanism with 1 second slot times and single slot finality.
- *Deferred Execution:* In Monad, execution is decoupled from consensus. This allows for significant throughput increases as consensus can come to the order of transactions before executing.
- *Parallel Execution:* Uses optimistic execution. In contrast to Solana, where writes are required to be specified up-front, optimistic execution essentially executes all transactions "optimistically" and then re-tries the ones that conflict (later transaction is re-executed).
- *MonadDb*: A custom database for storing blockchain state.

The majority of the **secret sauce lies in MonadDB**. Given that Monad executes transactions in parallel, it would require a database structure that supports multiple read and write operations at once. Ethereum's LevelDB and RockDB do not natively support asynchronous I/O. If Ethereum would optimistically execute transactions in parallel, its synchronous database operations would be a bottleneck.

Monad uses MonadDB, which is purpose-built for parallelized execution. It implements the Patricia Trie data structure both on disk and in memory. This allows data to be updated and verified more efficiently than Ethereum's MPT. Linux's latest kernel support enables MonadDB to perform async I/O. Why does the latest kernel matter here? For example, whenever RockDB has to perform read and write operations, it opens up kernels to manage memory, threads, and synchronization. Executing transactions in parallel would open up even more kernels, which would lead to contention for CPU resources, increasing overhead. Linux's io_uring helps bypass this bottleneck and allows for multiple read and write operations to occur simultaneously.

Monad has teams building in numerous sectors and has run a few "Monad Madness" hackathons. A full list of teams interested in building on Monad can be found following Monad Eco on X.

Like Hyperliquid, they have done well to build a community before their chain launched. However, there is growing competition between the two camps, even if they aren't really that similar; Monad raised >$200M, Hyperliquid was bootstrapped without funding. Blockchains are not just about technology, they are about communities and tribes and there are people firmly in either camp (i.e., who made them rich?). Monad testnet should be launching shortly with mainnet next year.

Monad has created extremely high expectations over the past year and is now facing some social pressure (from other EVM chains nonetheless) with the

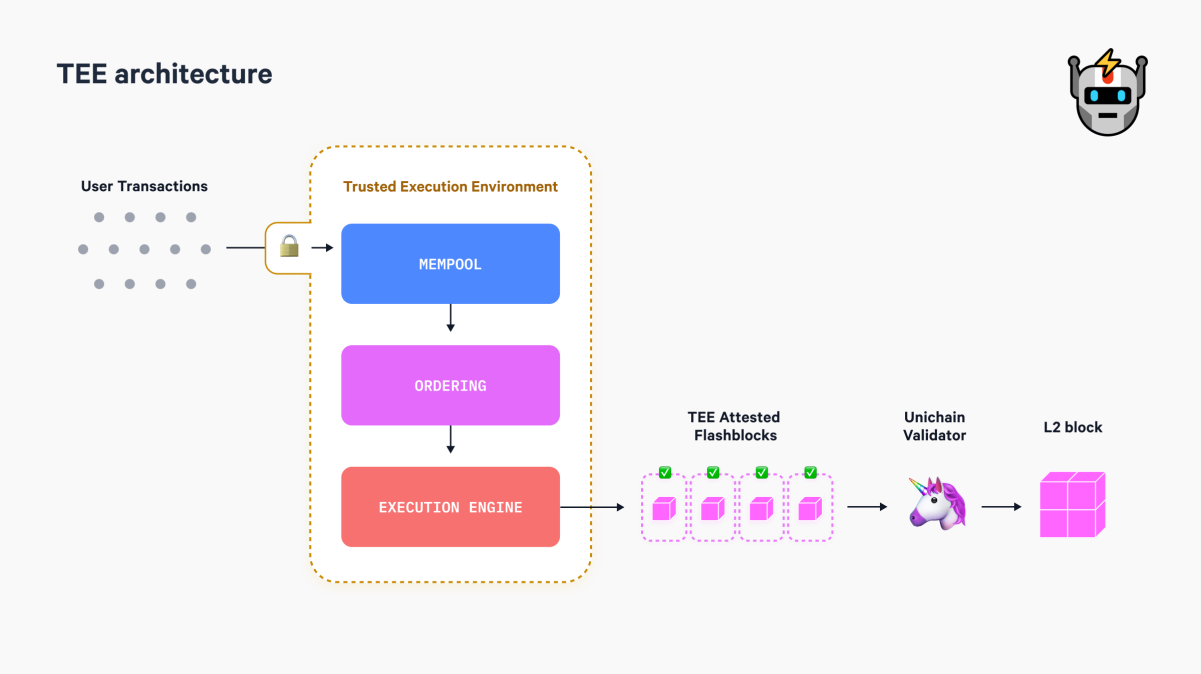
success of the less hyped HyperLiquid debut; I think this tweet by Keone sums up well how Monad is different. Regardless of whether HyperLiquid existed or not, Monad has a lot to prove in 2025. There is a big opportunity for a decentralized, performant EVM chain.

Unichain (L2)

Launched by Uniswap Labs, Unichain should not be confused with the Uniswap AMM product. Unichain is not like Osmosis in being a chain built around AMM infra, it's a general purpose (or "purpose built") chain made by Uniswap labs. Unichain aims to solve one of the main reasons why apps are enticed to launch their own appchain: capture their MEV. Unichain has the following properties:

- Single sequencer but decentralized validator network
- 1s block times with a longer-term 250ms target
- Native interop with Superchain
- Rollup-Boost: Near instant tx's & MEV/sequencing guarantees for apps

The one I want to hone in on is Rollup Boost. If you recall that tweet from Jon Charb above (app launches → takes up blockspace → becomes own appchain, repeat), Unichain aims to solve exactly this; the need for apps to become appchains to capture their own MEV. By utilizing TEEs & Flashbots, Unichain aims to give apps the ability to sequence their *own* transactions and retain their MEV. Instead of leaking MEV to validators/sequencers, it turns the economics of launching on a GP chain on its head.



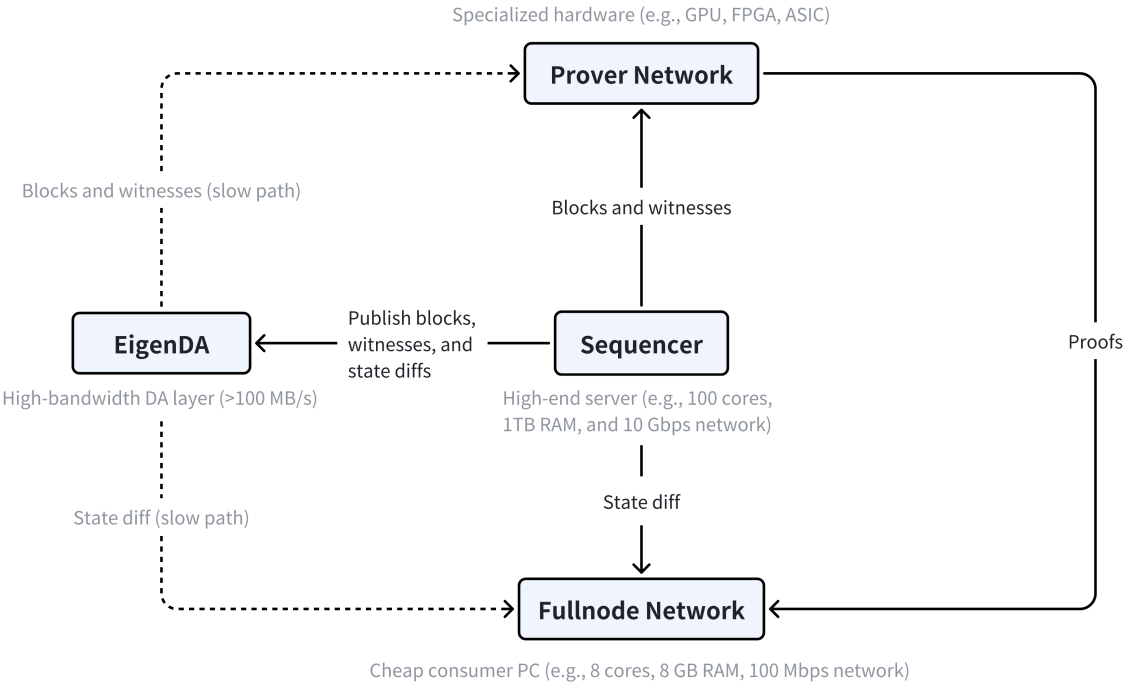
Unichain is essentially aiming for the best of both worlds: the shared liquidity and composability of a general purpose chain with the customizability and value accrual properties of an appchain. Unichain is another L2 in a long list of high profile OP stack chains led by Base.

MegaETH (L2)

The extreme end of rollup design, MegaETH is ruthlessly dedicated to performance. They will be running a centralized sequencer with insane requirements:

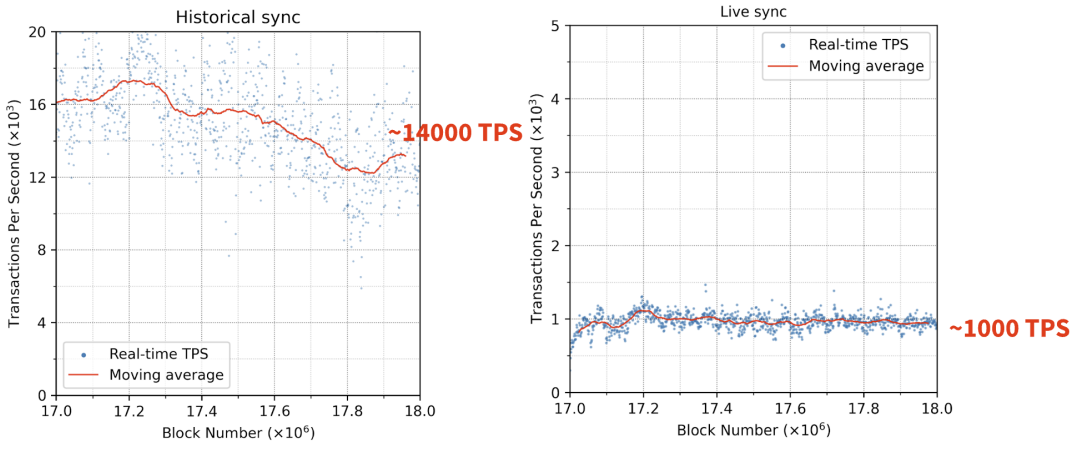
- 100 cores
- 1-4 TB RAM
- 10 Gbps network bandwidth

They will be using alt-DA in EigenDA which is the most abundant alt-DA in the market due to not having consensus.



Another main performance improvement is replacing the EVM's Merkle Patricia Trie (MPT) with their own state tree designed from scratch. In their internal testings, using the most performant execution client Reth (which we'll discuss with Ithaca), they were only able to get 1k TPS due mostly to merkleization (the MPT). Updating the MPT is ~10x more overhead than executing tx's, and so the main bottleneck doesn't lie in execution itself.

Historical vs. Live sync (512GB RAM)



Live sync is 14x slower even in a purely in-memory setting due to merkleization (9.3x) and more freq. disk flush (1.5x)

Slide 17

No other blockchain should be as fast as MegaETH given these parameters; if

they are it would be a pretty big fail for MegaETH. There is no consensus process and they are running with the most extreme hardware and network requirements. This extreme end of rollup construction will allow them to be the fastest chain on the market. The bottleneck for them is likely going to come from using EigenDA, as there are still network hops in using an external DA layer and it is possible MegaETH's DA needs are still above what Eigen will be able to provide.

I don't really consider MegaETH a competitor to networks like Solana, Monad, Sui and Aptos (and I think the team would agree), but more so to other rollups. All of these are already single sequencer chains and MegaETH *should be* most performant. Single sequencer L2s should always be faster than performant L1s and there are tradeoffs to both designs. With that being said, as mentioned before, every single blockchain still does compete for users and liquidity.

Ithaca (L2)

A project incubated by Paradigm and led by Georgios Konstantopoulos, Paradigm partner, CTO and developer of Reth. Ithaca is a high performant L2 built using Reth, the OP Stack, and Conduit. Reth is a high performance Ethereum node client with a stated goal of getting to 1 gigagas/s. For a comparison of how this compares to EVM chains today, see below. Note that Base is now at 16.5mg/s as they slowly look to scale up.

Select EVM Chains	Gas Per Second	Target Gas Per Block (Supply)	Block Time
opBNB	100.0 mg/s	100M	1.0s
BSC	46.6 mg/s	140M	3.0s
Polygon	7.5 mg/s	15M	2.0s
Avalanche C-Chain	7.5 mg/s	15M	2.0s
Arbitrum One	7.0 mg/s	1.75M	0.25s
Base	5.0 mg/s	15M	2.0s
Optimism Mainnet	2.5 mg/s	5M	2.0s
Conduit*	2.5 mg/s	5M	2.0s
Ethereum L1	1.25 mg/s	15M	12.0s

*Representative of most Conduit chains.

Ithica has an L2 on testnet called Odyssey. Odyssey currently has a 33 mg/s target and has some features such as:

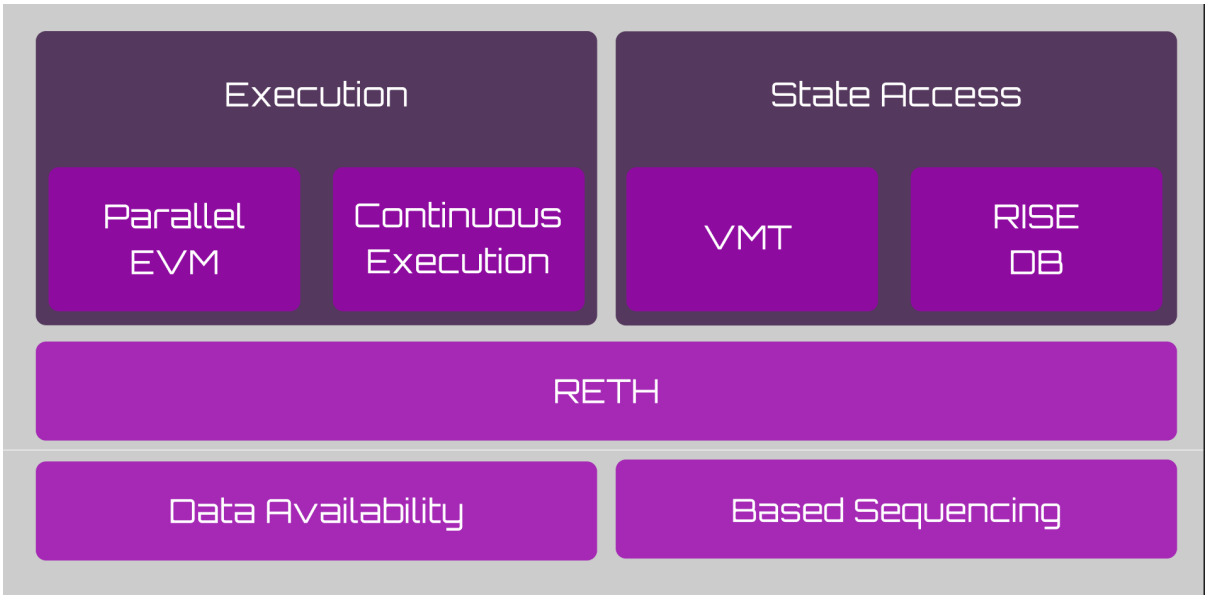
- EIP-7702: Allows EOA's to function like Smart Contract accounts
- EVM Object Format (EOF): Series of EVM improvements, making contracts cheaper
- EIP-2537: Useful for BLS signatures and ZK proofs
- RIP-7212: Allows signing with biometric authentication

Rise (L2)

Also in the EVM gigagas category we have Rise. Their targets are:

- 100k TPS
- 1+ Ggas/s
- <10ms latency

Like Ithaca, they're built on top of Reth with their own parallel EVM "PEVM", another contender in the "fast EVM L2" category. Like the other performant EVM chains they use a different merkle tree than Ethereum and have their own database. Different from other L2s, Rise aims to use based sequencing with preconfs instead of a centralized sequencer, allowing for better alignment and value accrual to the L1 and a higher degree of decentralization than something like MegaETH.



Rise is a fairly new project and I wouldn't expect mainnet in the near term.

Sonic (L1)

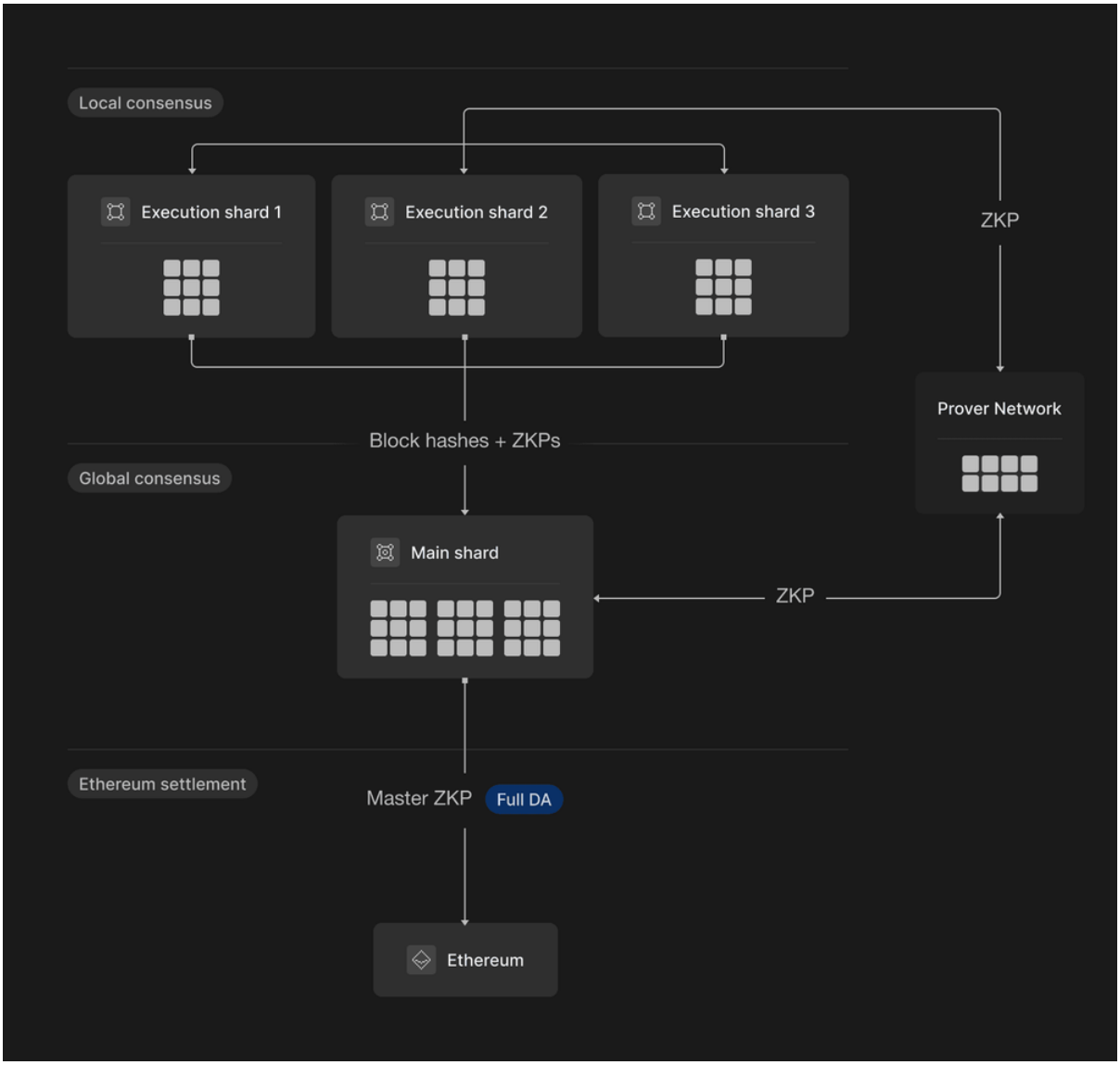
Rebranded from Fantom, Sonic is led by arguably the pioneer of DeFi summer, Andre Cronje. Their mainnet had their genesis block on December 2nd with public mainnet soon, and outside of a new ticker, brand, and chart, there are some improvements over Fantom.

- Consensus: Uses an Async BFT mechanism + DAG. With async BFT, nodes don't need to agree to blocks in sequential order before they're final. They can come to consensus individually and share blocks async. DAG's are seen in blockchains like Sui and Aptos and again allows validators to produce blocks async.
- Database Storage: Sonic's database prunes historical data automatically, reducing storage requirements for validators. The db is split into two parts: LiveDB and ArchiveDB. Validators use LiveDB which is only the current state and thus a lot lighter.
- Sonic Gateway: Bridge that has validators operating nodes on Ethereum.

There are a bunch of well-know Ethereum apps ready to deploy on Sonic in addition to newer ones, and it's possible Sonic takes some EVM activity like Fantom did last cycle.

Nil (L2)

An Ethereum L2 that goes back to Ethereum's original design choice before rollups: sharding. Nil splits validators into subcommittees to run individual local consensus' and execution shards. These executions shards run a zkEVM. Each committee then sends their proofs to the main shard for global consensus. Lastly, it is sent to Ethereum like any other ZK rollup.



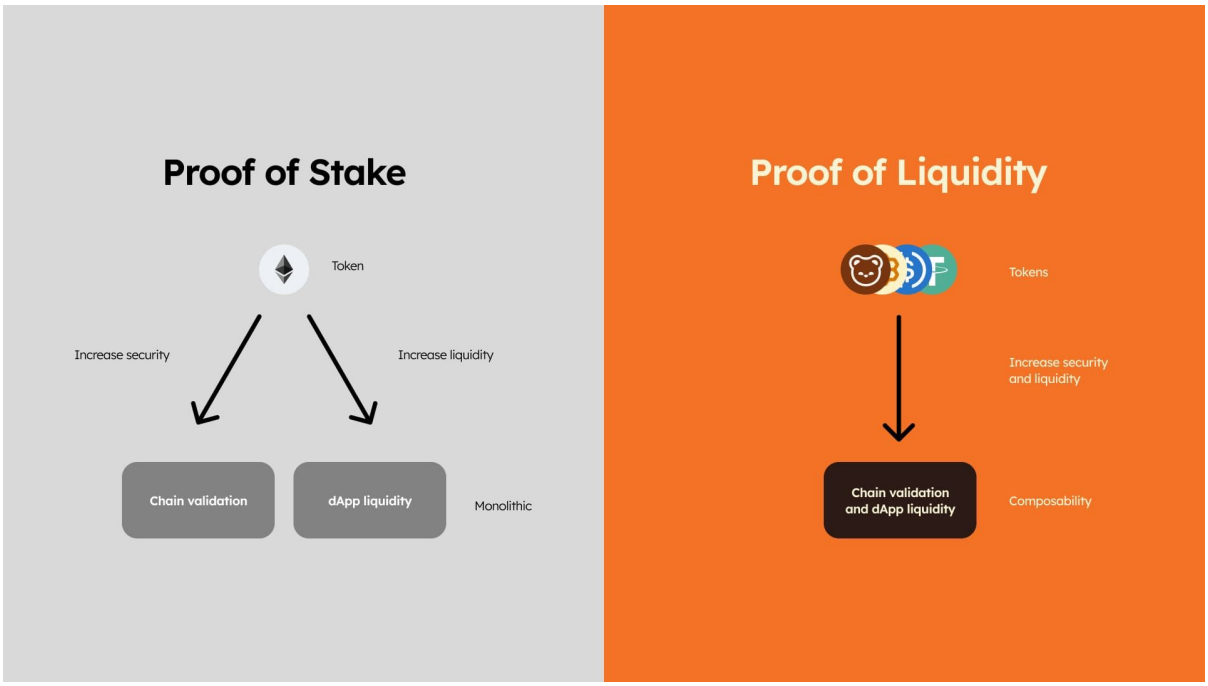
Nil launched their testnet in October.

Berachain (L1)

It is arguable if Berachain fits into this high throughput camp, but they are a well known, and well funded EVM L1 launching in 2025. Their differentiator and what they're known for (besides the memes) is their Proof of Liquidity consensus mechanism.

Berachain builds AMMs and liquidity directly into the consensus mechanism of the protocol instead of just relying on the native token. Protocols can create their own reward vaults with the objective of obtaining liquidity and then having validators direct emissions to them, aiming to create a sort of flywheel. This is a way for protocols and validators to create partnerships. The reward token for participating in consensus is a separate, soulbound (non-transferable) token than the native gas token, discussed below. You can read their deep-dive on PoL for

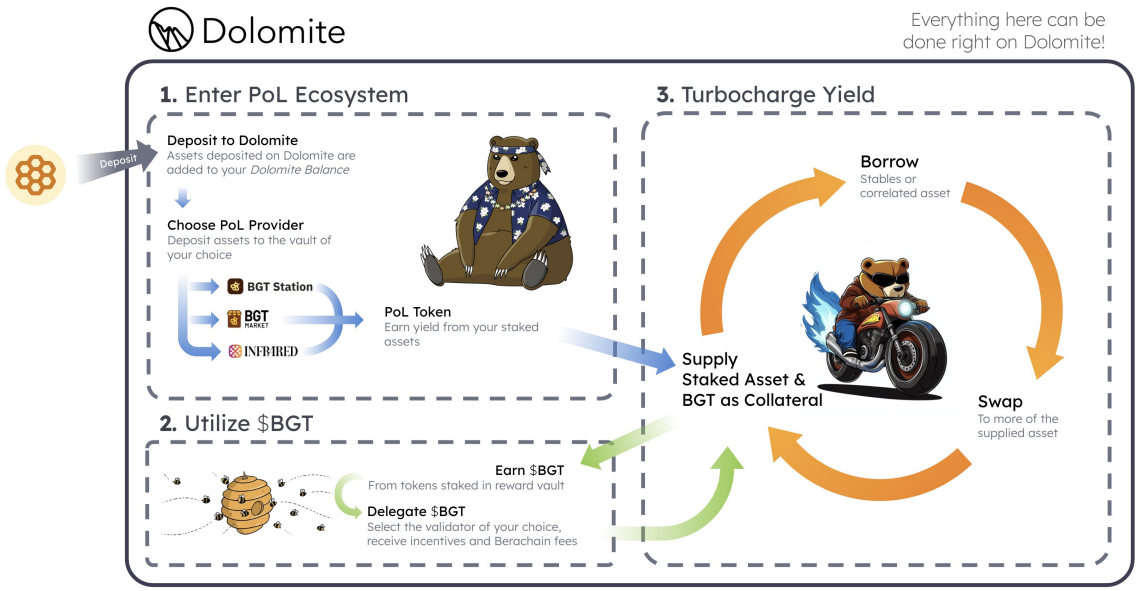
more.



Berachain's three tokens are as follows:

- **BERA:** Gas token and native staking token
- **BGT:** Non-transferable governance token earned through providing liquidity. BGT is used to vote on all governance proposals, is used to direct rewards, and can be burned for BERA (but not vice versa).
- **HONEY:** CDP stablecoin. Governance will decide which assets can be used to mint HONEY.

As an example of the flow between them all, see below chart from a DeFi protocol coming to Berachain.



Like Monad, Berachain has raised a lot of capital (>$100M) and is one of the most hyped projects with a large community coming in 2025. Their ecosystem can be viewed here. Berachain will probably be home to some of the wildest "flywheel" type apps, especially if they launch during the bull market. It did take inspiration from OHM, after all.

High Throughput SVM Chains

Eclipse, Atlas, Soon, Fogo

Eclipse (L2)

The first to bring SVM to a rollup, Eclipse is an L2 that uses ETH for gas & proofs and Celestia for DA. After a couple years of development, their mainnet went live on November 7th. Current apps on Eclipse are all Solana ports of Orca, Save, Invariant and Lifinity. Uptake has been slow a month in with ~$5m TVL.

The challenge for Eclipse (and all SVM rollups) is getting developers. If you're building an SVM app you can always deploy on Solana L1 and get access to Solana's deep and growing liquidity. You don't face the same fee restrictions that you do on Ethereum L1, so the natural extension of SVM rollups is not as clear. If you look at a bunch of the Solana rollups/extensions they are appchains and stuff like Magic Block's ephemeral rollups which are very different than gp rollups.

Atlas (L2)

Created by the team that built Phoenix, the CLOB on Solana, Atlas is an SVM L2 on

top of Ethereum (not clear what they will use for DA). They are not exactly a general purpose chain but more so in the "purpose built" category like Unichain.

Use case	General-purpose blockchain	Atlas
On-chain orderbooks	High jitter structurally favors toxic takers over market makers, resulting in worse liquidity for end users.	Low jitter and reliable transaction delivery allow market makers to provide competitive liquidity.
Margin systems	Oracles are unreliable during high volatility periods, when margining systems need them most.	Oracle updates are prioritized by the sequencer, ensuring they land regardless of market conditions.
High-frequency trading	High jitter and block-level confirmations increase uncertainty in state.	Low jitter and instant, transaction-level pre-confirmations allow traders to be confident in state transitions down to the millisecond.

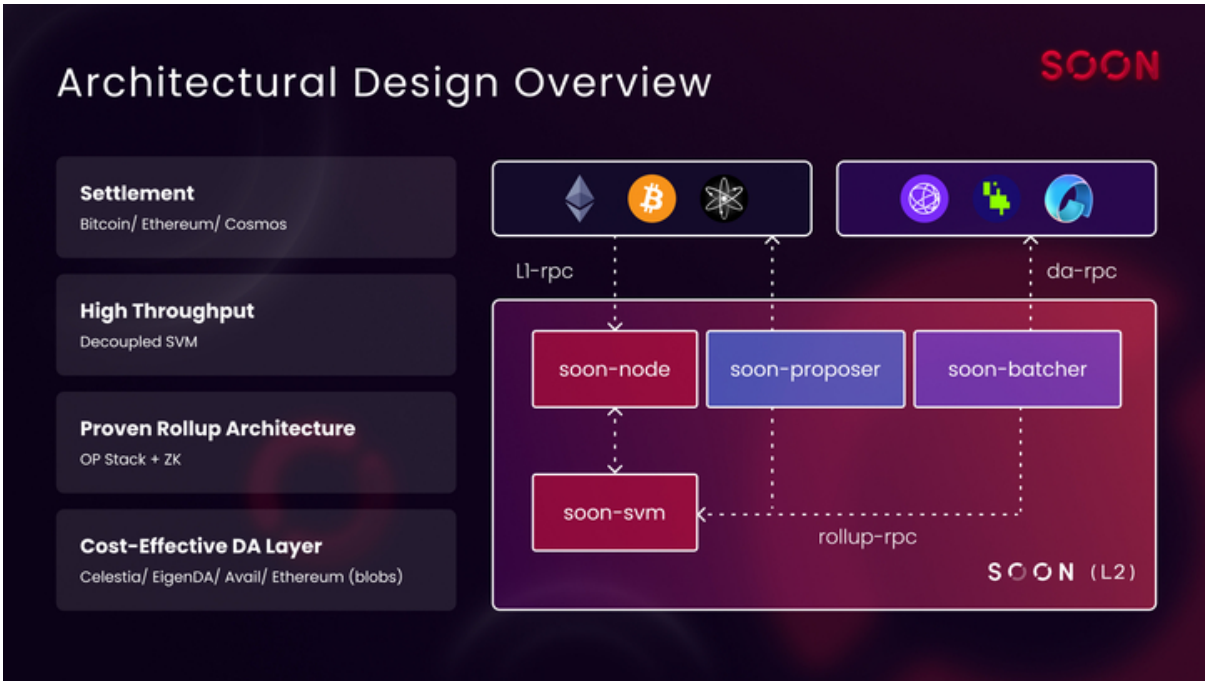
Applications they are targeting are orderbooks (which I assume will be built by them considering Phoenix is a port), margin systems and HFT. Solana developers can port applications easily, and the main differences from Solana are:

- 50ms vs 400ms block times
- Timestamps in milliseconds instead of seconds
- No confirmation levels due to no consensus (tx's are finalized or not)

A lot can still change because Atlas is new, but it's another "SVM on Ethereum" rollup.

Soon (L2)

Developed by ex-Optimism engineers, Soon is essentially OP stack but for SVM. Their mainnet will be deployed on top of Ethereum but can be used on top of any L1.



Soon is built on the OP Stack with a decoupled SVM (i.e. only the execution and removing consensus mechanism like voting). The Soon team believes SVM is the endgame over EVM and is bringing it to Ethereum and beyond. They have similar challenges to Eclipse highlighted above.

Fogo (L1)

Full disclosure I have no idea what this is but *apparently* they will be running an SVM fork with *only* the Firedancer client. Worth monitoring.

> Who we are: Fogo
>
> What we are: An experimental SVM layer 1 that deploys the Firedancer client in its purest form to date.
>
> What we do: Prioritize performance above all else, with the mission to achieve the optimal latency and bandwidth that Firedancer promises.
>
> Fogo: Faster than...
>
> — Fogo (@FogoChain) December 5, 2024

High Throughput MVM Chains

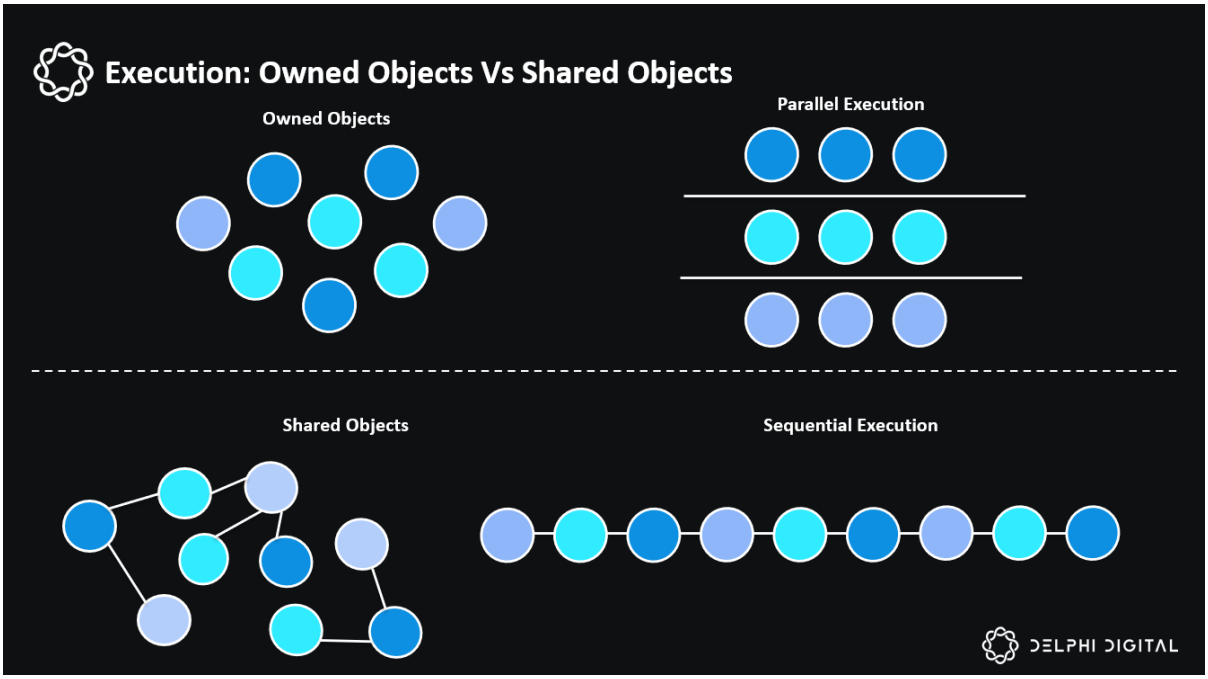
Sui & Aptos

Sui (L1)

In my opinion, Sui is the most viable threat to Solana out of all the chains in this report for a few reasons.

1. Like Solana, they are an L1 with globally distributed validators and consensus
2. The decision for an SVM app dev to launch on an SVM rollup vs Solana is not as clear as an EVM dev launching on an EVM rollup, thus I don't see an SVM rollup competing with Solana
3. MoveVM is performant like SVM and is more developer friendly
4. Sui engineers, like Solana's, are some of the smartest in the industry and come from Meta's former blockchain team at Project Diem (eg., Evan their CEO helped build numerous programming languages)
5. The market has (so far) chosen Sui > Aptos as the move chain, with Sui having an FDV 3x Aptos ($47B vs $15.5B) after launching around the same time and valuation
6. Multiple Concurrent Proposers through their DAG (still TBD how this plays out in Sui)
7. Sometimes I think of Sui as the "Academic Solana" (vs Solana's "we can engineer our way out of any problem" ethos). They have developed a lot of novel things like Narwhal, Mysticeti, native account abstraction, cryptographic extensions, zkSend, SuiLink, internetless transactions, and their new storage protocol Walrus

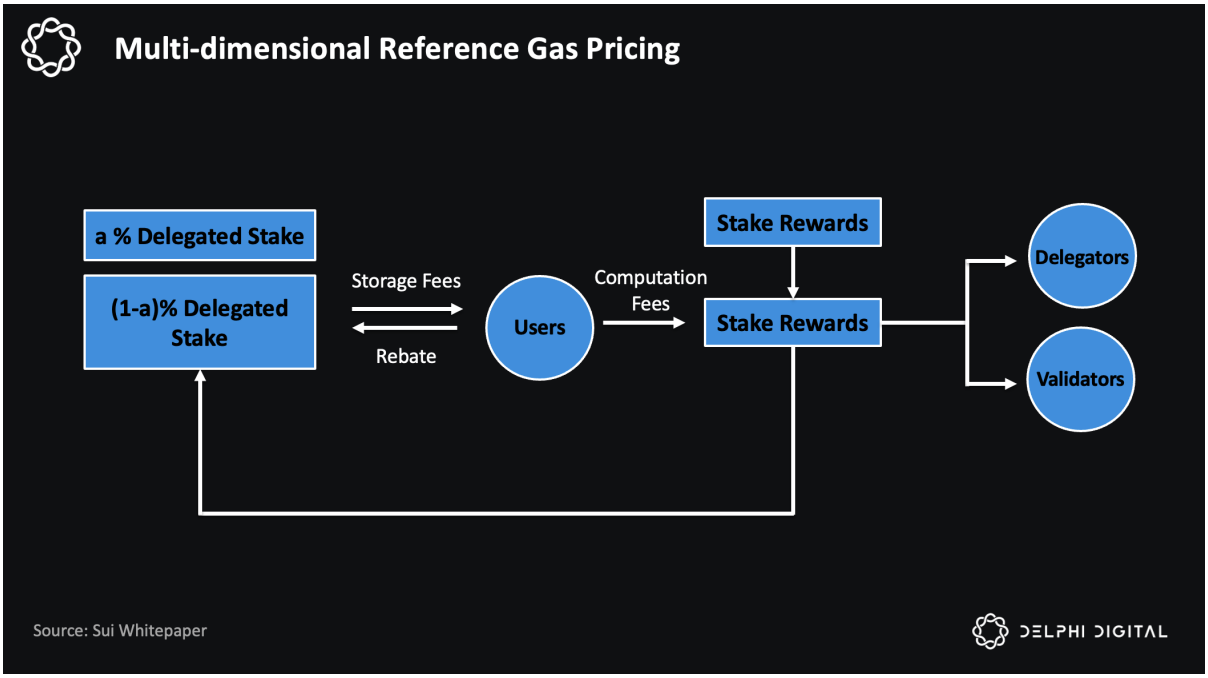
For the deep dive on Sui and explanation of these I will reference the report from Delphi Consulting. The biggest unlock is Sui's VM using the object centric model. Essentially, assets/tokens on Sui are identified as either shared (AMM pool) or owned (USDC in an account). By leveraging shared vs owned object, Sui can significantly speed up owned object tx's and is leveraged in their consensus fast path (i.e. p2p transfers like sending USDC from one person to another don't need to wait for consensus because ordering doesn't matter).



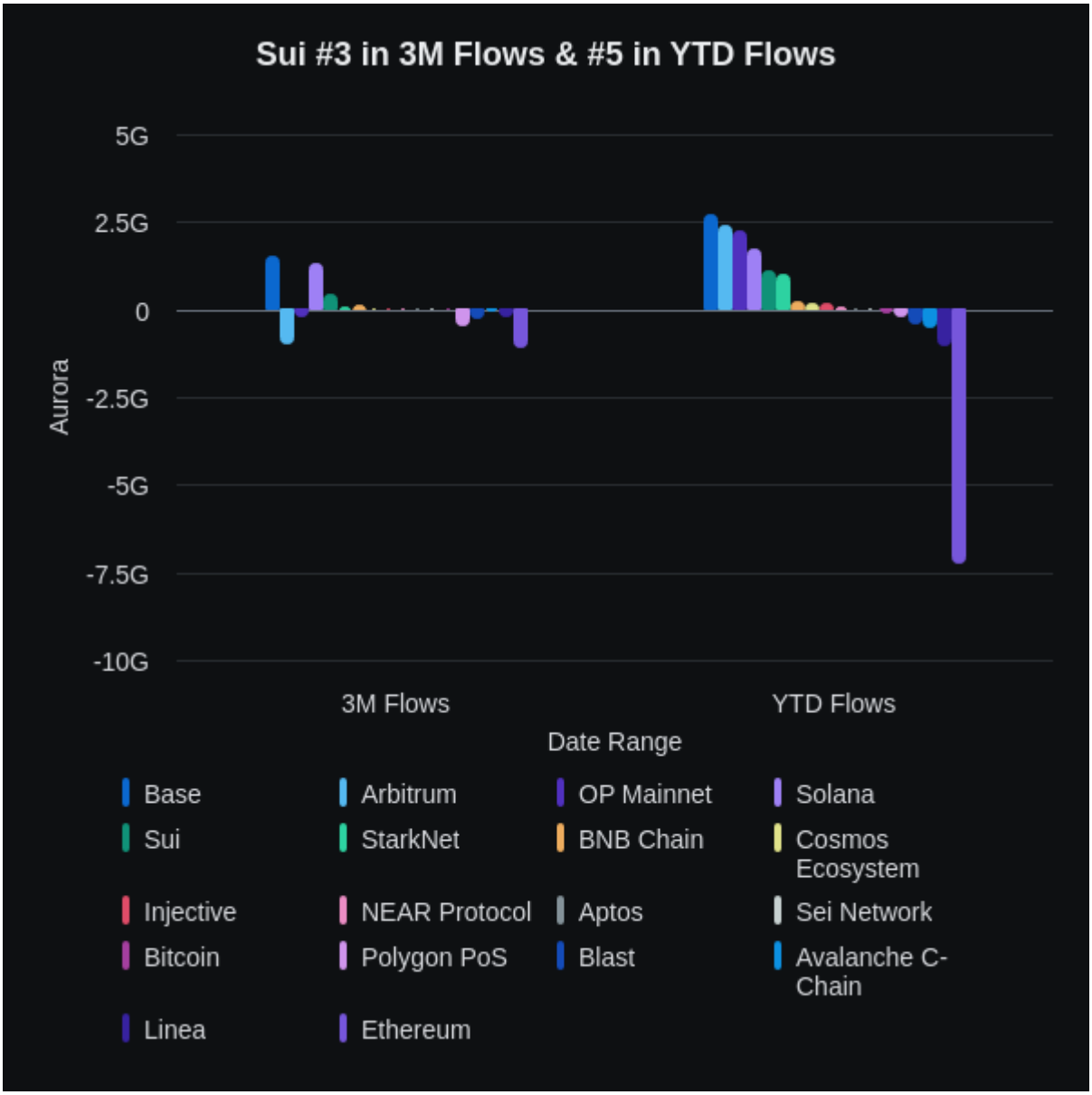
Sui also has a different gas model from other chains. Sui uses a multi-dimensional reference gas price mechanism. It separates fees into two main components: computation fees and storage fees. Computation fees cover the cost of processing transactions and consist of two parts. First is the reference price, determined at the beginning of each epoch (24-hour period) through a Gas Price Survey that determines the *reference gas price*.

The protocol surveys validators and then selects the 2/3 percentile of these prices, weighted by each validator's stake, as the reference price. This ensures that the reference price is acceptable to most validators. The reference gas price sets a floor on the price the network will accept, but users can specify higher prices for priority access to congested state.

Sui Objects live onchain and directly relate to accounts, tokens, NFTs, and smart contracts (packages). This is the second component of fees that users pay. Storage fees are calculated as Storage Fee = Storage Units × Storage Price. Storage Units represent the amount of storage a transaction requires, while the Storage Price is the cost (in SUI) of storing one data unit.



Looking at some data, Sui has been one of the biggest beneficiaries of flows from other chains this year, sitting at #3 over the past 3 months and #5 YTD (Note that a large part of Ethereum's outflows are to L2s).



Now, Sui was heavily incentivized with high yields for most of the year, so this is somewhat attributed to that. Still, they now sit at #8 in TVL overall, just behind Arbitrum and ahead of Avalanche with a growing ecosystem. They still don't have a killer app, but there's a lot of liquidity and all it really takes is one. Sui has more of an Asia focus, and while it has been leaning into gaming, it now boasts 6 DeFi protocols with >$100M TVL. They got native USDC support with Circle's CCTP protocol a few months ago and Phantom & Backpack wallet recently expanded there as well. Again, for the full deep dive I will reference the report from Delphi Consulting.

Aptos (L1)

The other move chain, Aptos shouldn't be counted out just because it is behind Sui. Aptos' TVL sits at #11 at $1.3B behind Sui's $1.7B and they too have a growing

DeFi ecosystem. Aptos and Sui's MoveVM's are actually different from one another. Aptos started from the original Diem Move whereas Sui built a new one (including their object model). Developers cannot port apps between the two and there are different nuances. Aptos' Move can be and is used by new modular ecosystems Initia and Movement whereas Sui's can only be used on Sui. The tradeoff here is that Sui developers are locked into Sui whereas Aptos developers can launch on other chains. It is TBD which long term strategy will win as there are pros/cons to each.

It's still early in the history of these Move chains, and even if one wins vs the other, it doesn't mean there's not a world where both are successful and adopted. They've both seen increasing development and adoption this year.

Slow is Not an Option

The high throughput era is just getting started. There are a lot of chains mentioned above and most will not survive. They are not just competing against Solana or with each other — they're also up against current L2 rollups and new stacks entering the space. We are becoming overloaded with blockspace, and as we've seen with token prices over the past year, consolidation is the name of the game.

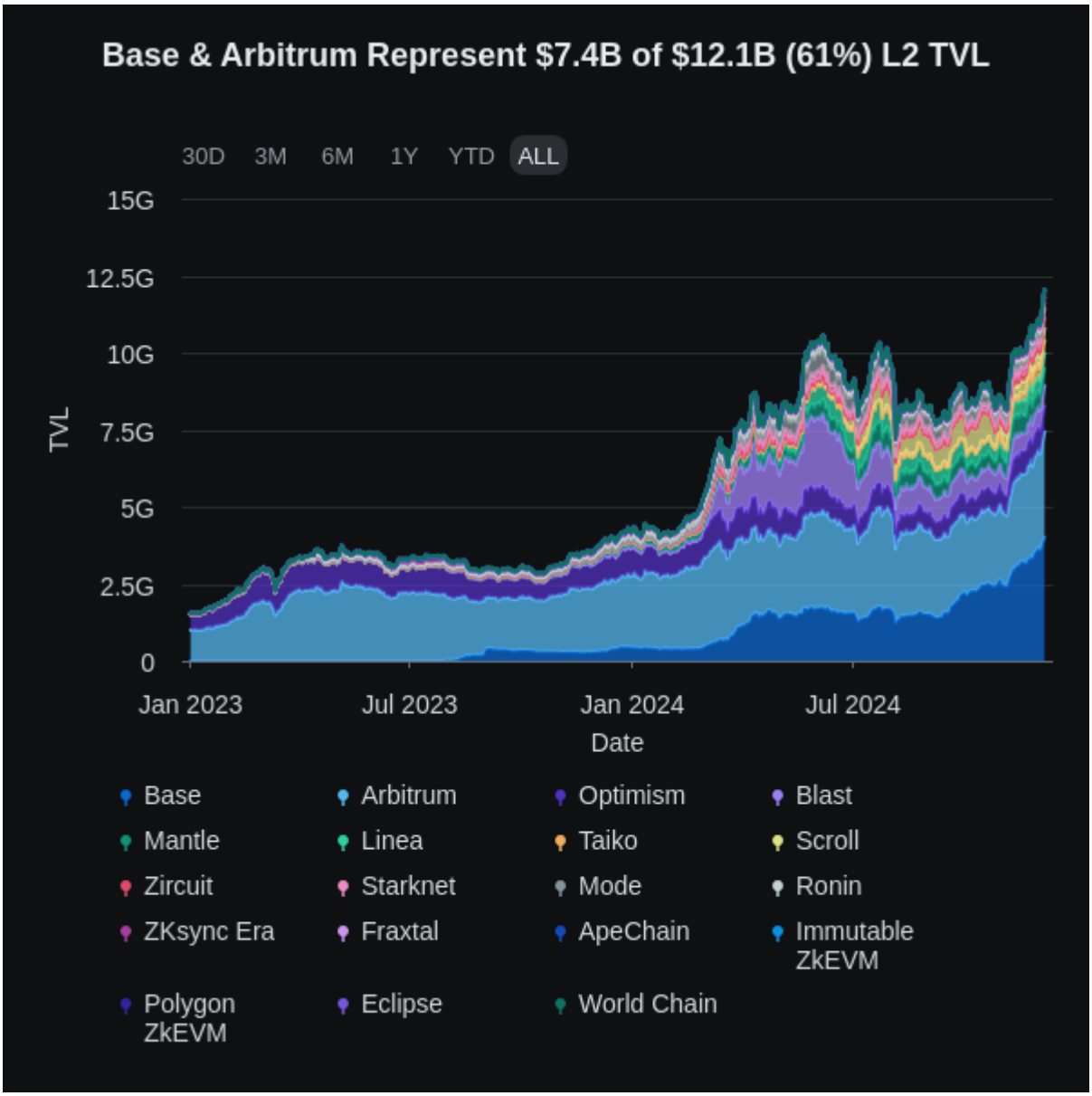
L2 Wars v2 – Base & Arbitrum Drive Consolidation

Ethereum Rollup Landscape

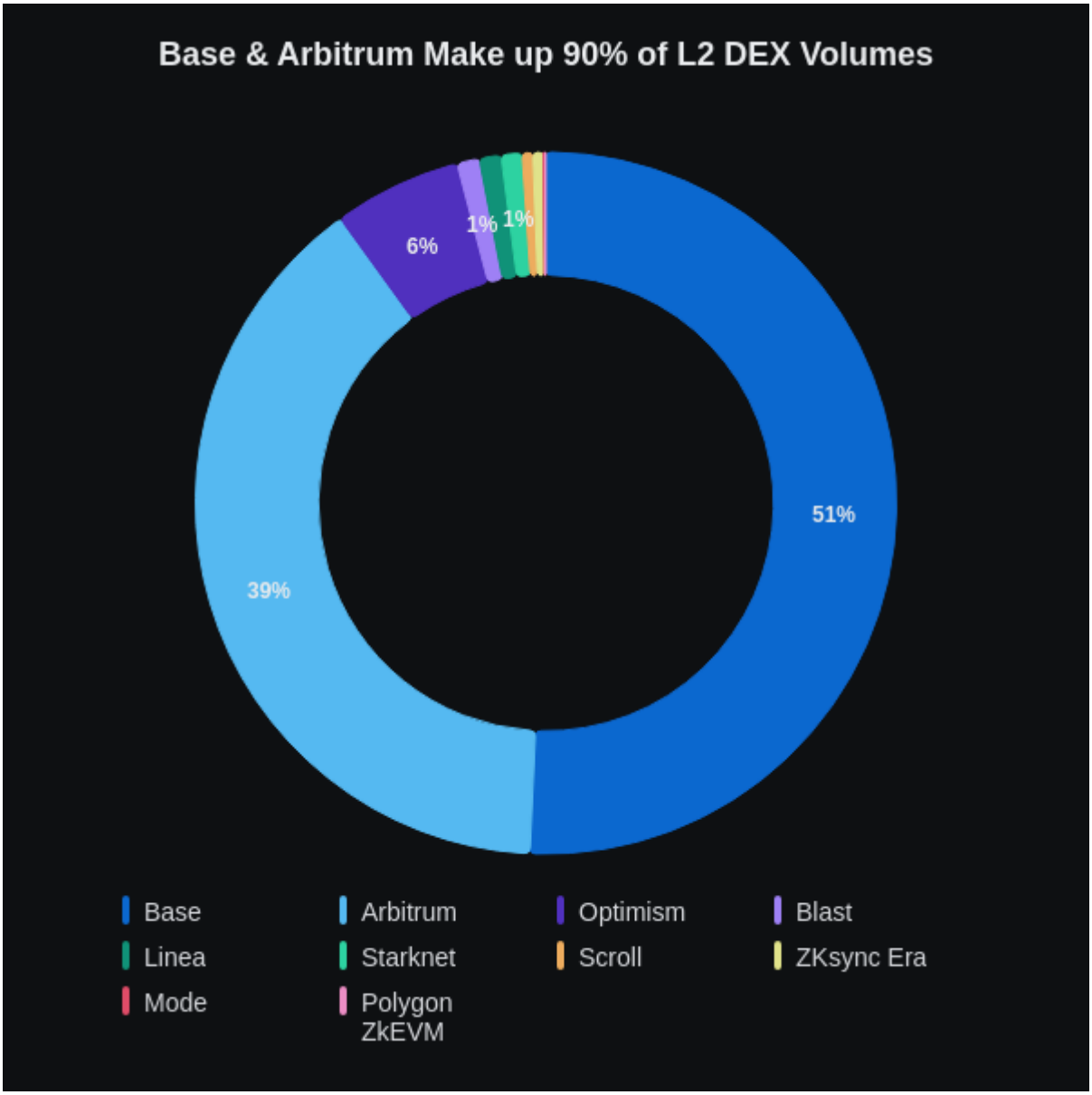
A year ago I wrote "The L2 Wars". In it I talked about what would be a year of numerous new L2s with fragmented liquidity, and the era of peace that would transition more into an era of war. This piece held up well, and these challenges are now properly recognized (plus the "war" vibes). As we've progressed further into the L2 wars, one thing is clear: they will consolidate. The majority of activity now takes place on either Base or Arbitrum, and this consolidation accelerated post EIP-4844.

Looking at DeFi TVL (not to be confused with L2Beat's "TVL" metric which is just assets minted on chain), Base and Arbitrum make up nearly 2/3 of all L2 TVL.

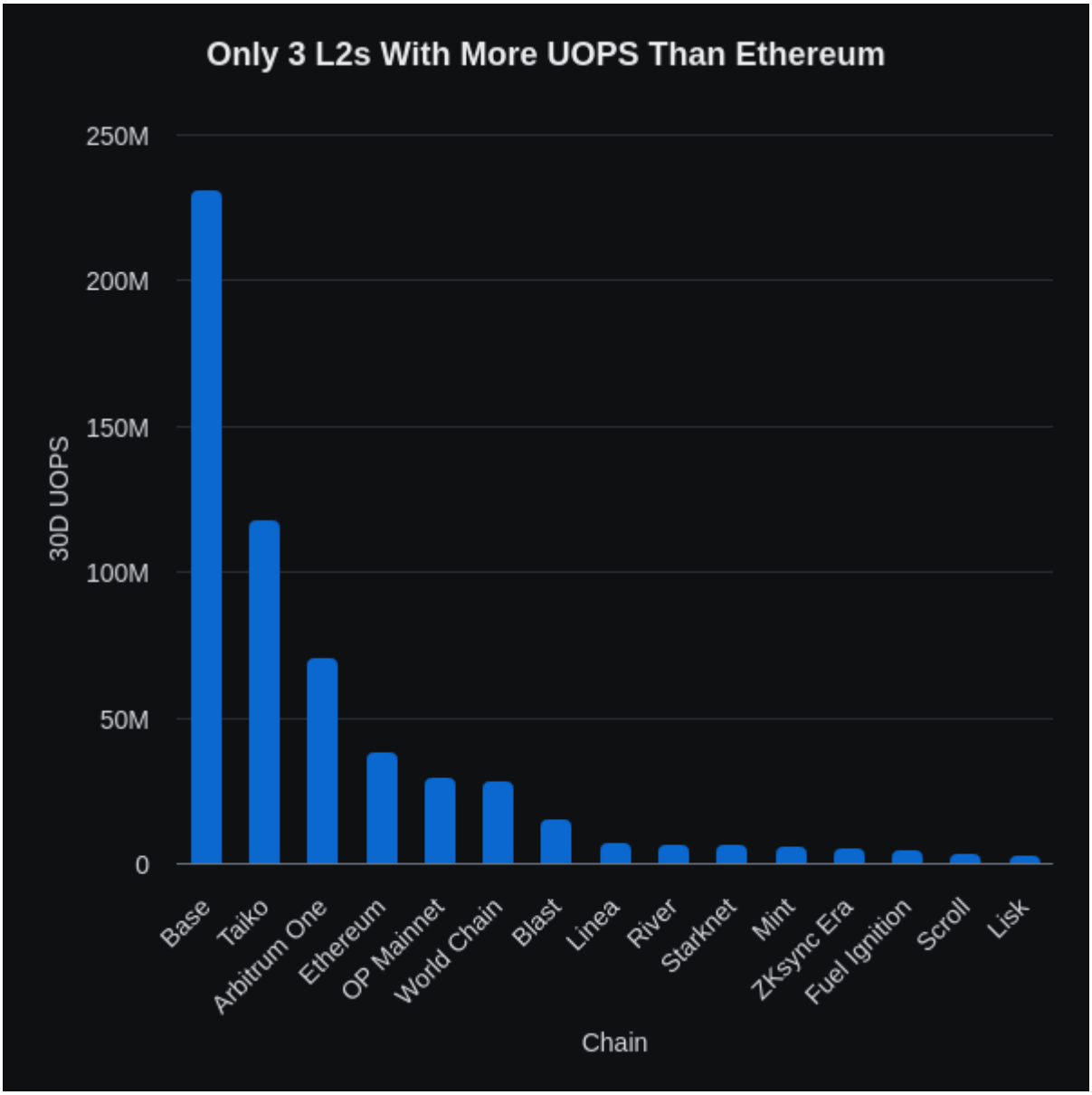
There are notable declines like Blast, zkSync Era and Optimism over the same period.



The same goes for DEX volumes. While TVL is 2/3 Base and Arbitrum, nearly *all* of L2 DEX volume is now happening on the two chains.



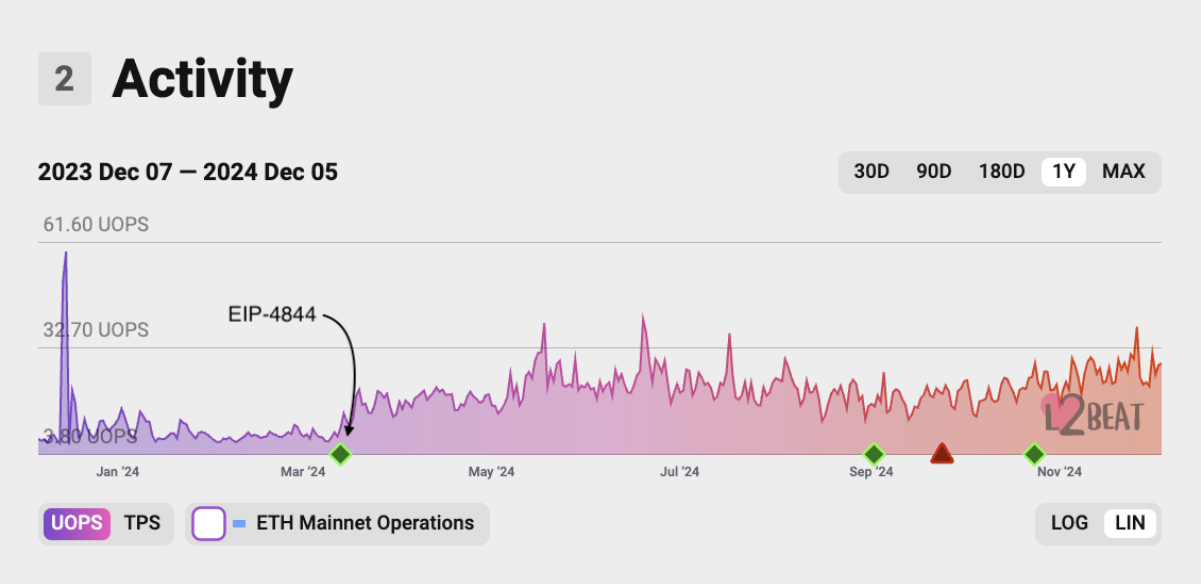
And lastly, User Operations per Second. Outside of Base and Arbitrum, only Taiko has more UOPS than Ethereum. Read that again. The low throughput, tx limited Ethereum L1 is doing more UOPS than the majority of L2s.



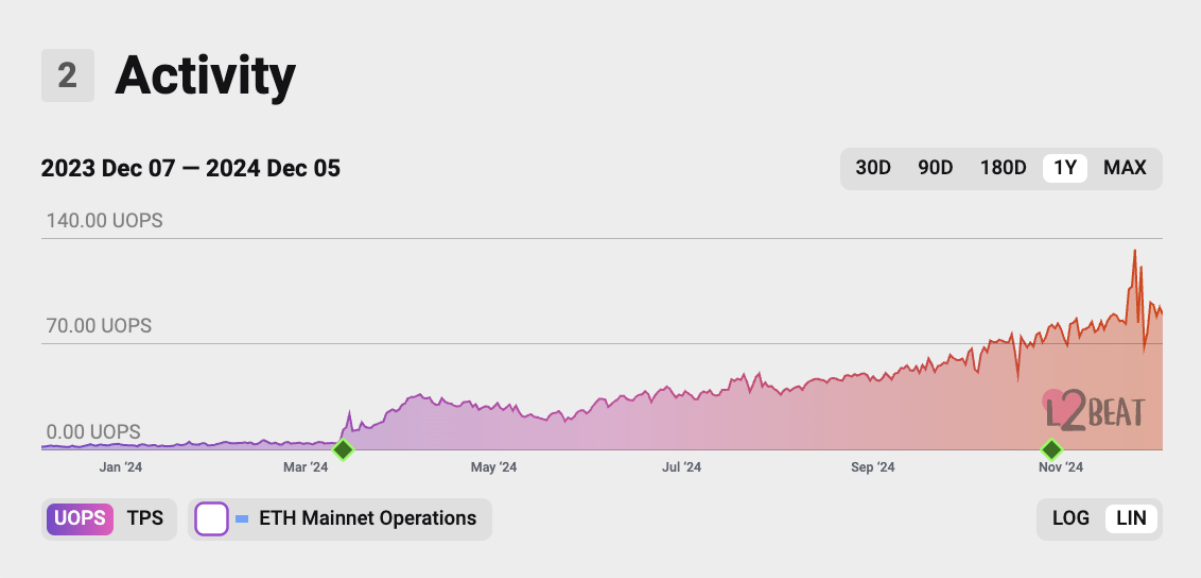
What does this all mean? Consolidation. All of these rollups are general purpose EVMs. Sure, some are optimistic, some are zk, and some are based; but they are all essentially the same product. There will be power laws here and we are seeing Base and Arbitrum start to reap those rewards. As I alluded to above, these divergences accelerated post EIP-4844. As a refresher, 4844 was the introduction of blob markets which made Ethereum DA significantly cheaper. But the benefits were not evenly distributed; now that every L2 can handle more tx's, the winners took more. We can look at Base & Arbitrum vs Blast & zkSync Era. That green diamond in March '24 is when 4844 went live.

Activity Spikes From EIP-4844

Arbitrum

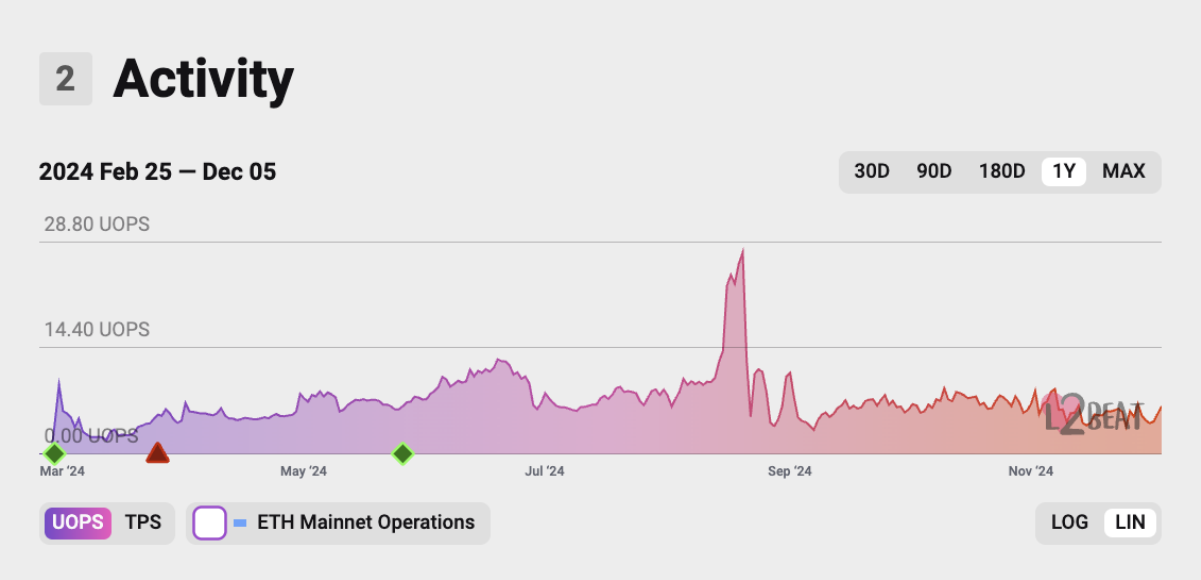


Base

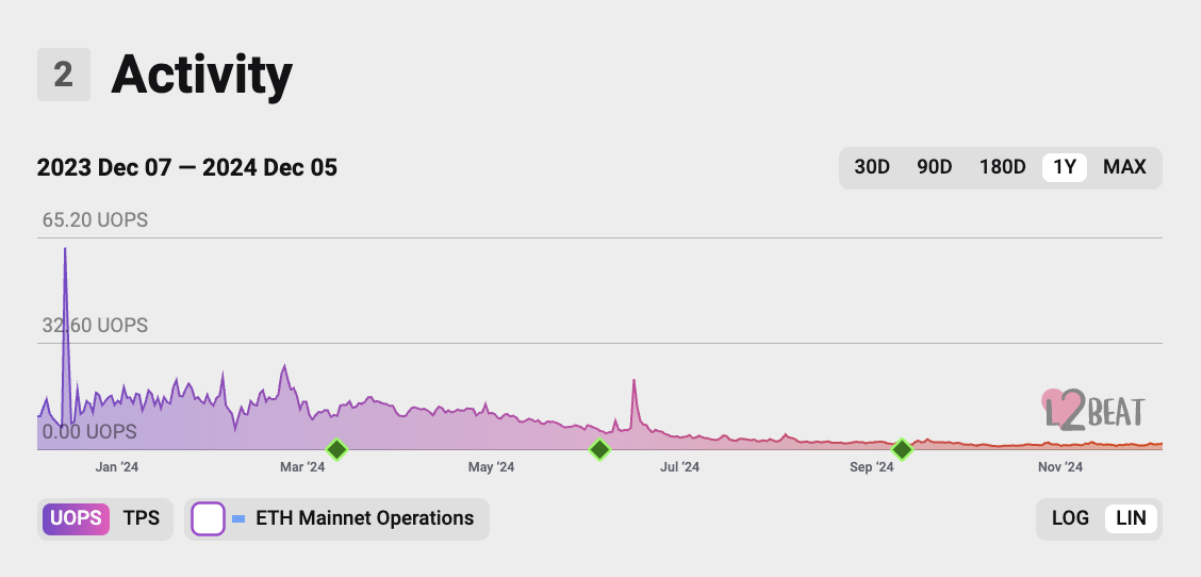


Activity Drops From EIP-4844

Blast



zkSync Era



As Base keeps raising their gas limit and Arbitrum continues to scale themselves, it seems likely that these chains will continue to move higher in adoption. Users need a *reason* to use another L2 and if the one they're using is handling their needs then they don't need to move; save it for an anticipated airdrop. Does the Pudgy Penguin airdrop signal anything here?

Pudgy Penguins Choosing Solana Over any L2

Pudgy Penguins, one of the most popular (and arguably most mainstream) Ethereum NFTs chose Solana as their chain of choice for their token over any Ethereum L2. This is an auspicious decision, and some have highlighted it is a "vampire" attack to eventually move users from Solana to their EVM L2 Abstract when it launches.

You may think it's a "stupid memecoin" and has no implications, but to me this highlights a failure of L2s to develop a product users want to use. As highlighted earlier, Solana has become the chain to launch tokens and trade assets. The fact that Pudgy chose Solana over any L2, especially as it's an Ethereum project, speaks volumes to this and the lost opportunity for L2s. I see it less a failure of Ethereum and more a failure of the L2s to develop a product users want.

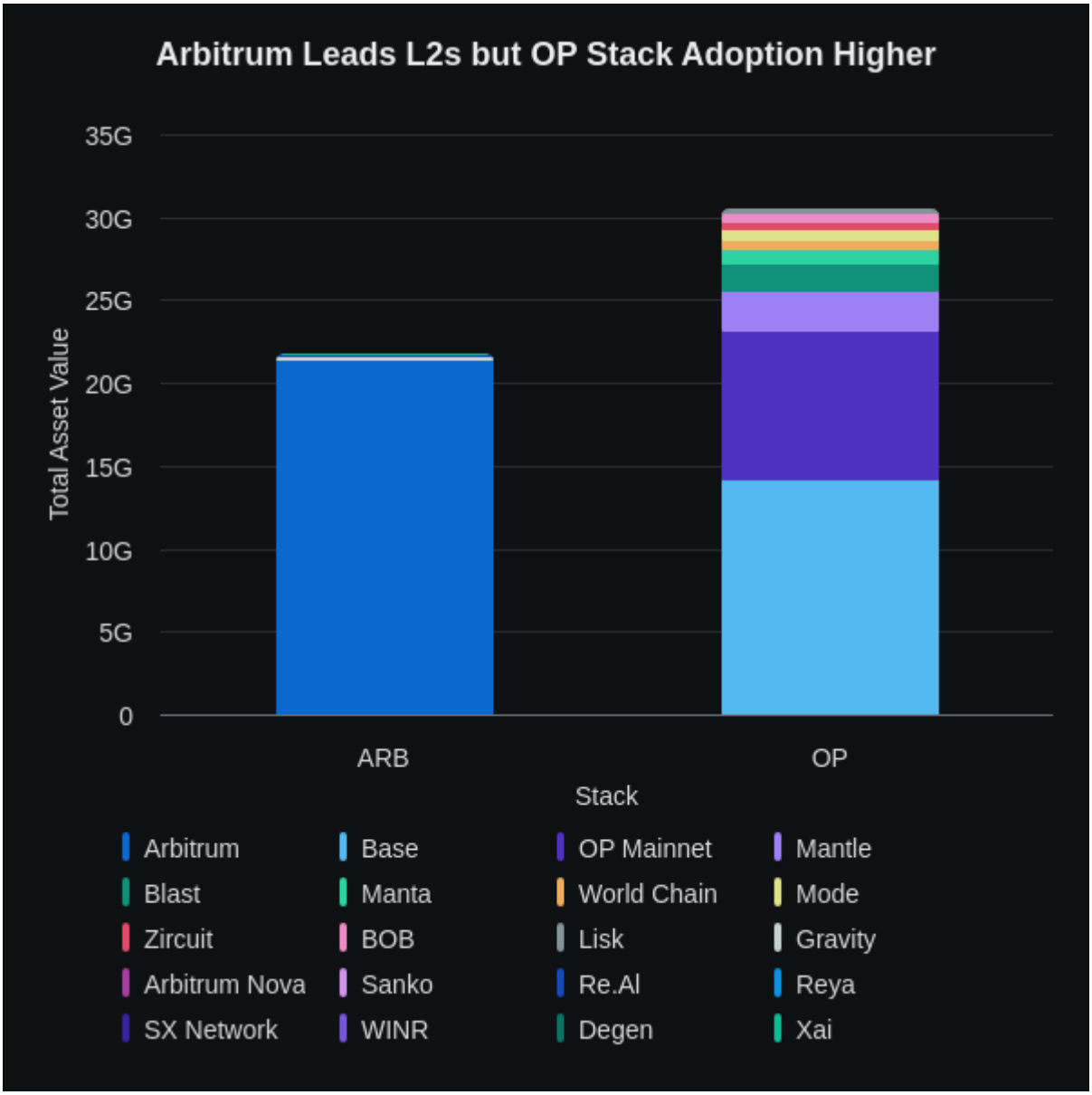
And it's not just NFTs, as we highlighted earlier it's also young talent leaving to the Solana ecosystem as well.

Leading Stacks: Optimism vs Arbitrum

It is no question that Optimism and Arbitrum are the leading stacks today, however they have taken very different approaches to one another. Optimism is leading the BD/OP Stack adoption charge, getting numerous rollups like Base, Blast, World Chain, Mode, Zircuit, Zora and upcoming Ink by Kraken, Unichain by Uni Labs, and even SVM stack Soon.

On the other side, Arbitrum's stack doesn't have nearly as much adoption but does target certain niches like Apechain, Sanko, Real RWA, gaming and a few others. Another notable difference is that Arbitrum main chain is the largest Arbitrum rollup (and largest in general) whereas Optimism main chain has seen activity continue to decline with Base and I would expect this trend to continue with some of the upcoming ones. A good way to visualize this is Velodrome ($137M mcap) vs Aerodrome ($1.5B mcap). Aerodrome is a fork of Velodrome by the same team that launched on Base; its valuation is >10x higher.

Looking at the top 10 rollups & validiums for each stack, it is clear that Arbitrum is the largest by total assets issued, while OP Stack has higher total adoption (note that $1.6B of Arbitrum's number is bridged to Hyperliquid).



Moving forward we will see which strategy is more effective: adoption of the OP stack at the cost of cannibalizing Optimism main chain vs Arbitrum main chain being the leader in their ecosystem.

Other Stacks: Starknet, Polygon, zkSync, Scroll

The other stacks are all of the zk ones, and maybe they need more time for the tech to improve to get adoption.

Starknet took a different path than other rollups, focusing on developing an entirely new ecosystem built around their language Cairo. The early days of Starknet, to put it lightly, were rough. Using the chain was like using a dial up modem and swaps would take minutes to confirm on average. However, that is no longer the case, and by forgoing EVM for Cairo they've been able to implement

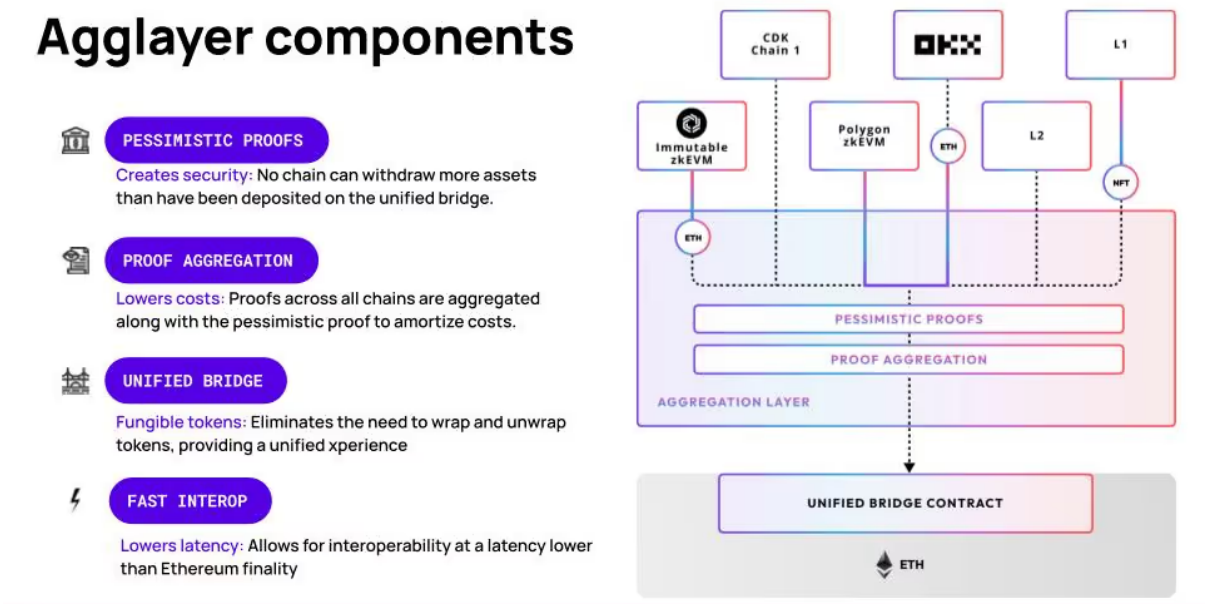
many major UX improvements. Every account is a contract account by default, which means you can do things like approve, swap, deposit, stake all in one tx. The Argent wallet works well and has nice web2 features like email 2FA. Lastly, Starknet is not just an Ethereum play as they have been deep in OP_CAT discussions and are expected to compete in the Bitcoin rollup eco as well. They took the long game by not going EVM which has taken them longer to develop an ecosystem (due to devs needing to learn a new language) but I believe this will eventually pay off and 2025 is the year we start to see these signs. For a dive into their DeFi eco be sure to read the DeFi year ahead next week.

> The first step toward a trustless L2 bridge connecting Starknet to Bitcoin has been achieved by @scryptplatform!
>
> A few months ago, we partnered with sCrypt to build a Bitcoin Signet (OP_CAT-enabled) PoC bridge, designed to lay the foundation for a production-grade bridge for... pic.twitter.com/xMlHlvMRjy
>
> — StarkWare ■■ (@StarkWareLtd) November 28, 2024

Polygon is interesting here because of their work on AggLayer whose launch is coming soon. The main issue that needs to be solved when it comes to rollups is fragmentation of liquidity. AggLayer, while developed by Polygon, is something *every* rollup can opt into for bridging and cross-chain interop needs. AggLayer has four components:

- **Pessimistic Proof:** ZK proof security mechanism that ensures no chain can withdraw more assets than they have deposited in the bridge (i.e. can't rug other chain's liquidity).
- **Proof Aggregation:** Aggregating proofs across many chains tapped in lowers costs for everyone. Has a cold start problem as it gets cheaper (i.e. further amortized) as more chains join.
- **Unified Bridge:** All chains share the assets in the same contract on Ethereum. Makes for better cross-chain interop.
- **Fast Interop:** A future feature, will allow cross-chain settlement to be faster than Ethereum.

zkSync has their own similar solution for this with their Elastic Chain.



As we mentioned last year, all of the rollup solutions have their own interop standards and bridges (Arbitrum bridge & orbit chains, Superchain, AggLayer, Elastic Chain). Competition is good because it will lead to the best tech getting developed, but it is still unclear if all Ethereum rollups will decide on a unified standard (seems unlikely, at least in the medium term, especially if new "SuperchainERC20" standards are being pushed).

Stage 2 or Bust

Lastly, it's time to see the training wheels come off. 2025 will show who has the tech to fulfill the promise of rollups.

> 2025 will be the year that reveals who's truly willing to move beyond stage 0 and who's just full of bullshit marketing. no more excuses
>
> — donnoh.eth ■ (@donnoh_eth) December 8, 2024

#	NAME	RISKS	TYPE	STAGE	TOTAL VALUE LOCKED
1	Arbitrum One		Optimistic Rollup	STAGE 1	$21.34B ▲9.62%
2	Base		Optimistic Rollup	STAGE 0	$14.08B ▲8.52%
3	OP Mainnet		Optimistic Rollup	STAGE 1	$8.96B ▲6.00%
4	Blast		Optimistic Rollup	STAGE 0	$1.65B ▲2.83%
5	ZKsync Era		ZK Rollup	STAGE 0	$1.44B ▲12.0%
6	Starknet		ZK Rollup	STAGE 0	$1.24B ▲12.5%
7	Linea		ZK Rollup	STAGE 0	$1.01B ▼8.43%
8	Scroll		ZK Rollup	STAGE 0	$920.73M ▲0.53%
9	World Chain		Optimistic Rollup	STAGE 0	$624.80M ▲16.4%
10	Mode		Optimistic Rollup	STAGE 0	$591.36M ▲10.0%
11	Zircuit		Optimistic Rollup	STAGE 0	$502.96M ▲45.2%
12	BOB		Optimistic Rollup	STAGE 0	$482.21M ▼11.6%
13	Fuel Ignition		Optimistic Rollup	STAGE 0	$385.65M ▲1.62%
14	Taiko		Optimistic Rollup	STAGE 0	$372.16M ▲2.12%
15	Lisk		Optimistic Rollup	STAGE 0	$232.19M ▲7.32%

New Stacks & L2s

Initia, Movement, Taiko, Spire, Fuel, Soon, Fluent, Rogue, Sigil, Last Network, Bitcoin Rollups

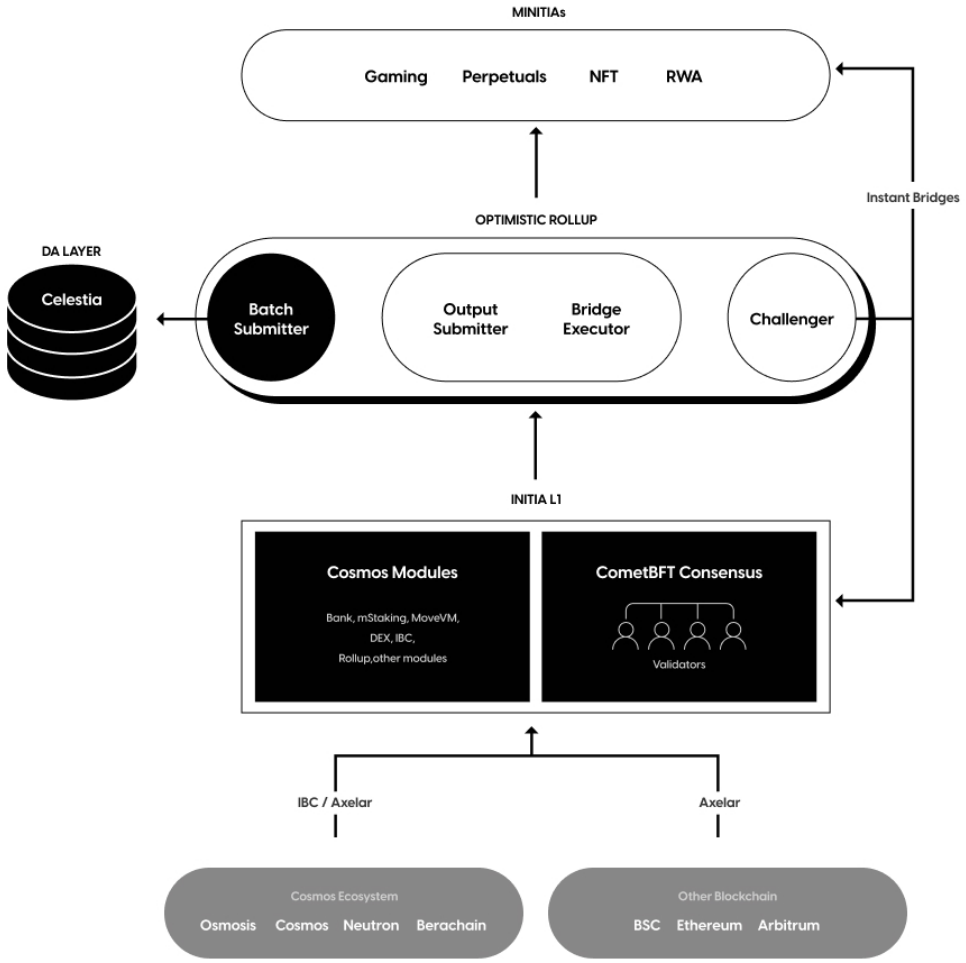
Not only do existing L2s have a lot of competition from the new high throughput chains, or the current crop of L2 chains, but there is a whole host of new stacks entering the ring as well. Competition is only getting fiercer. The main stacks & L2s to follow are below.

Initia

What Cosmos was meant to be. If there is a stack that will fulfill the appchain vision I believe it will be Initia or some of the newer, further out ones like Delta, Pod & InfinityVM (we'll discuss in detail later).

Initia is building an ecosystem dedicated to appchains from scratch. Initia's L1 serves as a coordination layer to all of the appchains (minitias) built on top. While

Initia's L1 uses the Aptos MoveVM, minitias can be a variety of alt-VMs like Move, EVM and WasmVM. The L1 also incorporates a liquidity hub with "enshrined liquidity". Think Balancer for the L1, this architecture diversifies the staking assets for the L1, is used for mintita to minitia routing, and gas fees.



Initia will have numerous integrations out of the box like support for Noble USDC through IBC, LayerZero integration for a coordinated standard for minitias, a kickback mechanism to minitias, oracles, programmable mempoools through Skip, and more. They want to go live ready to fully support an ecosystem of appchains. Some Minitia's in development are:

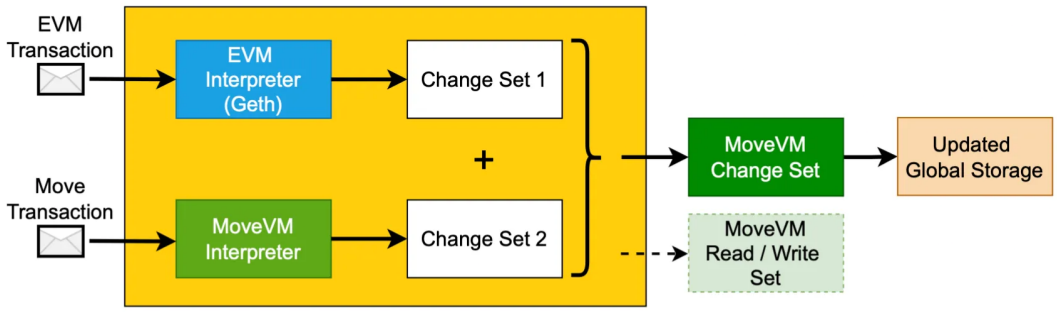
- Civitia: onchain finance board game
- Echelon: money market
- Contro: markets built around their new primitive Gradual Limit Order Books

(GLOB)

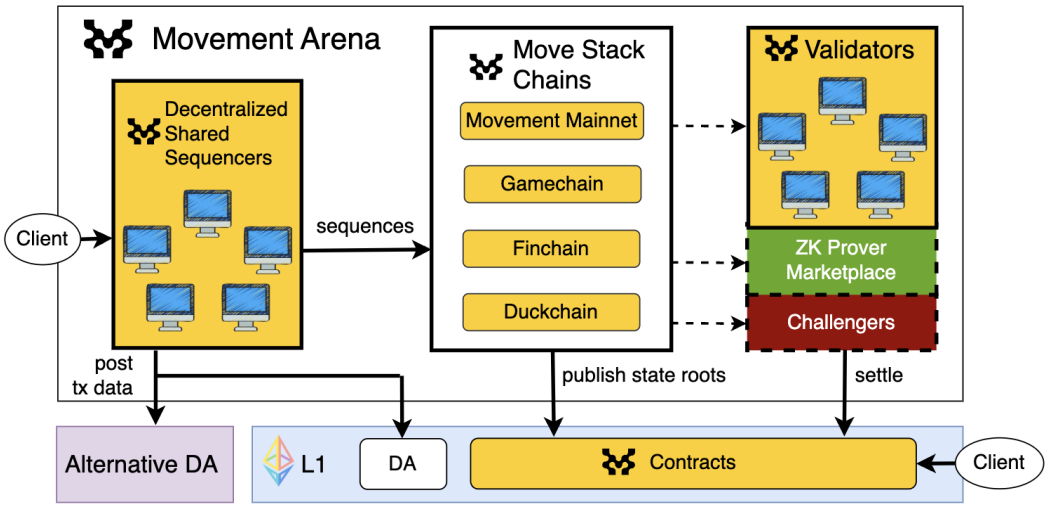
- Blackwing: liquidation free margin trading
- Inertia: LRT & lending protocol
- Infinity Ground: AI content creation platform
- Zaar: NFT Marketplace
- Milkway: liquid staking

Movement

Another rollup eco based off the MoveVM, Movement launched their token on Dec 9th and opened at an $8B FDV; the chain, however, is not yet live. Movement will have both a MoveVM and EVM as they aim to keep compatible with the large, existing EVM ecosystem while also supporting “next-gen” VMs like Move.



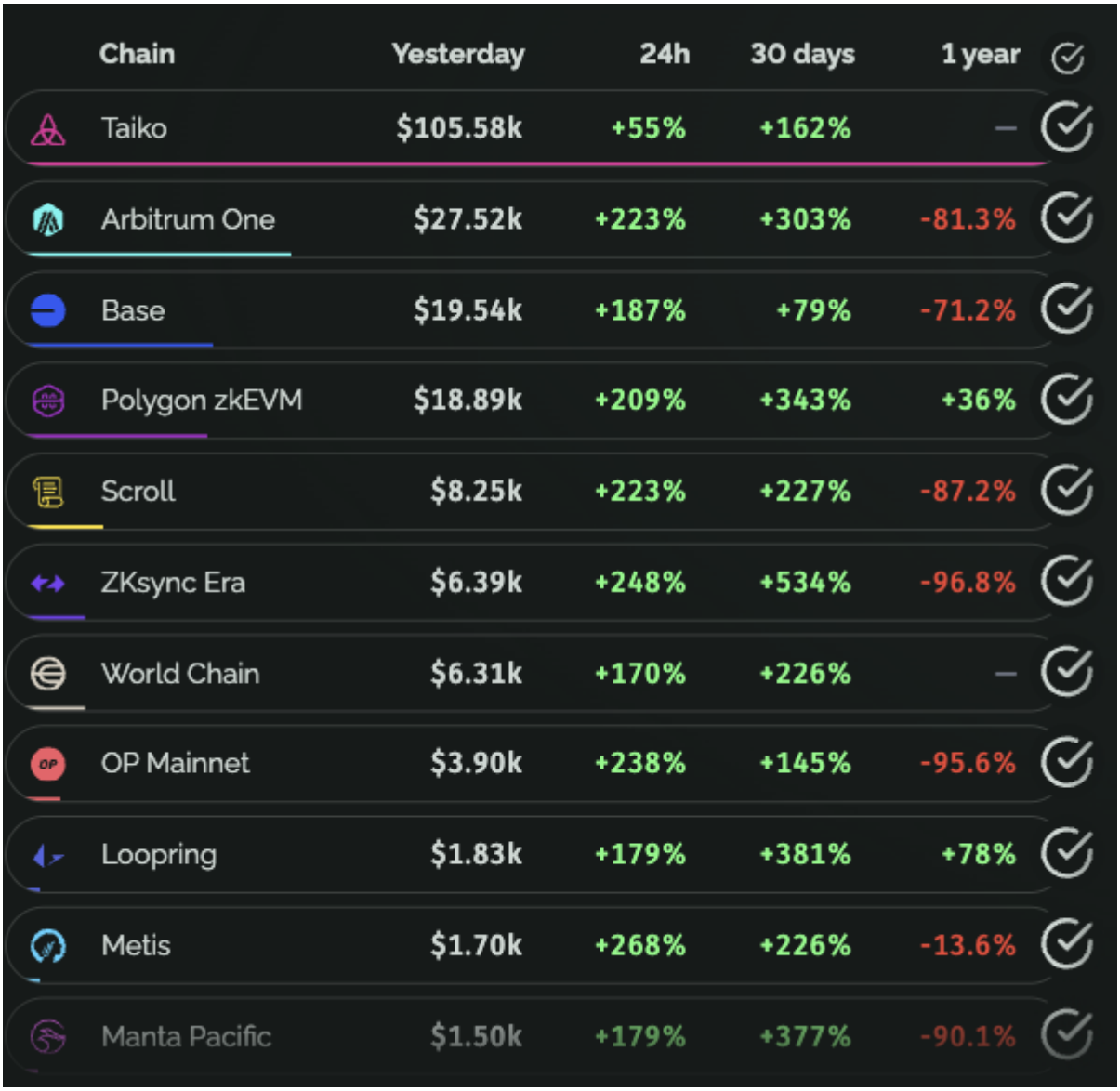
Their main chain is a general purpose L2 on top of Ethereum that uses Celestia for DA. Different from some other rollups that have a bridge to Ethereum, Movement will use their own MOVE token as the native gas paying token on their L2. Like other rollup stacks, they aim to support numerous constructions like ZK/ORU, shared sequencing, Ethereum & Alt-DA, and more.



Based Rollups: Taiko & Spire

We first wrote about Based Rollups in March and since then we have seen more teams developing them. These are not to be confused with *native rollups* which we'll talk about later.

Taiko is live today with $390M in TVL led by lending apps Avalon Labs and TakoTako. Due to the nature of being based (and using L1 for sequencing) their costs add up quite a bit, and are by far the top rent paying rollup on Ethereum. While not ideal for users on Taiko, it is aligned with the L1. Finding this balance for based and native rollups will be important and these costs should fall significantly with preconfs.



Chain	Yesterday	24h	30 days	1 year
Taiko	$105.58k	+55%	+162%	–
Arbitrum One	$27.52k	+223%	+303%	-81.3%
Base	$19.54k	+187%	+79%	-71.2%
Polygon zkEVM	$18.89k	+209%	+343%	+36%
Scroll	$8.25k	+223%	+227%	-87.2%
ZKsync Era	$6.39k	+248%	+534%	-96.8%
World Chain	$6.31k	+170%	+226%	–
OP Mainnet	$3.90k	+238%	+145%	-95.6%
Loopring	$1.83k	+179%	+381%	+78%
Metis	$1.70k	+268%	+226%	-13.6%
Manta Pacific	$1.50k	+179%	+377%	-90.1%

Spire isn't yet live but is the other based rollup stack, recently having raised $7M from some notable investors. One interesting decision is Spire's integration of Rome protocol, a protocol that uses Solana for based sequencing. We've spoken before about how the ideal L1 for based rollups is a fast with with short block times, and Solana is that. Based rollups are *provably better* for the economics of Ethereum L1, but they have challenges with respect to UX and user adoption. They are an area to watch (along with native rollups) as they may become the endgame for rollups.

Fuel

One of the most anticipated rollups from a few years ago (check the date of our deep dive), Fuel has had some internal issues that have delayed their launch. Their mainnet went live on October 16th.

> Next week is a big one for Fuel.
>
> For those who don't know, Fuel brings an entirely new high performance blockchain stack:
>
> New VM.
> New language.
> New toolchain.
> New development paradigms.
> New layer-2.
>
> And as you will see next week, so much more.
>
> We wouldn't be here without our...
>
> — nick.sway ■■ (@IAmNickDodson) December 6, 2024

Soon

Mentioned in high throughput section above but worth highlighting here again. OP Stack for SVM.

Fluent

The "blended execution network", Fluent supports all of EVM, SVM and Wasm all in one VM, allowing developers from every eco to build together. It is planning on extending support to CairoVM and MoveVM as well.

Rogue

A project out of LambdaClass (used to work with Starknet), Rogue is a no VC, no team allocation, fair launch ZK L2. Rogue will use based sequencing, proving and TEEs. The Lambda team are some of the smartest in the industry and the addition of the fair launch mechanics make this one to watch. A full description is in the tweet below.

> Some friends questioned me for doing this. But hey, life is about pursuing what you believe in and having fun in the process. This is the beginning of why and how I'm going rogue.

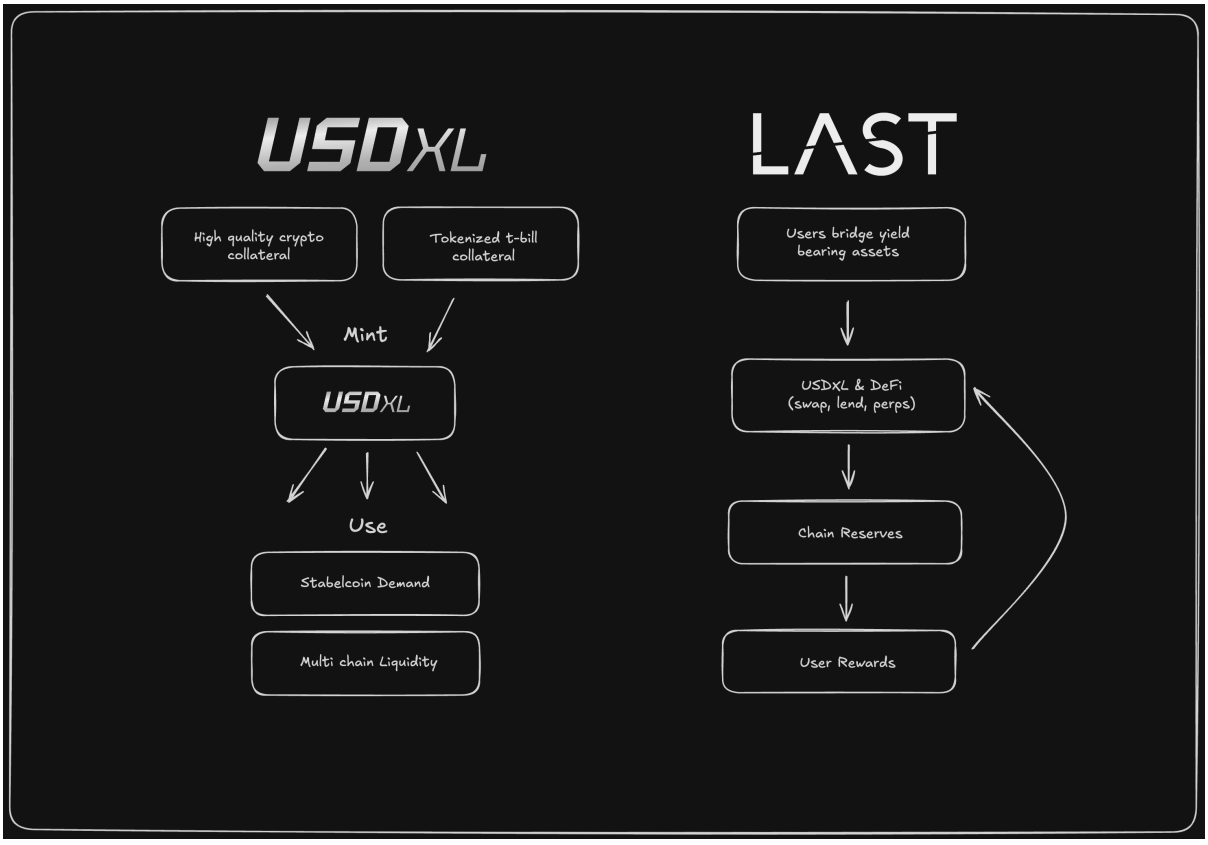
> I am going rogue
> The original promise of cryptocurrency was simple: decentralization, ownership by... pic.twitter.com/UmKLB7n5mx
>
> — Fede's intern ▮ (@fede_intern) October 15, 2024

Sigil

We talked about how rollups need to move to at least stage 1 in 2025 and the training wheels need to come off. Sigil is a rollup that is going to launch as a fully decentralized, Stage 2 rollup from day one (timeline TBD).

Last Network

Pitched as a "cashflow generating chain that reinvests back into the community", Last Network is a DeFi focused EVM rollup built around its stablecoin USDxl. This is a CDP stablecoin minted by crypto assets which directs chain profits into US Treasuries to distribute as USDxl profits to both USDxl holders and DeFi users.



Bitcoin Rollups

Lastly, we must mention that an eco of rollups are coming to Bitcoin powered by

BitVM and OP_CAT. We've spoken about Starknet and their intentions to bridge to Bitcoin (tbd if current rollup or a new one) but there are also many others like Citrea, Bitlayer, Yona, BOB and more. The Bitcoin rollup deep dive can be found here and we'll cover again later in the zk section of this report.

What About Ethereum L1

So far we have talked about Solana and other high throughput chains, the current rollup landscape, and the upcoming rollup landscape. But we have not yet talked about Ethereum. From the Beam chain, to single slot finality, to Native rollups, there are a lot of developments to track in 2025 with Ethereum.

Ethereum – The Near to Long Term Future

Previously, The Merge and the Beacon Chain upgrade served as Ethereum's North Star, rallying the community around a shared technical goal. Since then, it seems the Ethereum community lacks a clear, unified direction for prioritizing technical upgrades.

Ethereum has always focused on building a permissionless, transparent, and globally decentralized system—prioritizing decentralization over all other metrics, as Vitalik encapsulated with *DOPE*. This emphasis comes with trade-offs, such as sacrificing L1 performance. Most Devcon talks highlighted the importance of maintaining decentralization over and above performance and scalability.

Flashbots' co-founder, emphasized four pillars for *'Ethereum 3.0'*: permissionlessness, distribution, geo-economic decentralization, and neutral-builder efficiency.

The lack of consensus on Ethereum's priorities and roadmap has led some to describe this lack of coordination as Ethereum's "*Cosmosification*."

While Cosmos pioneered appchains, shared security, and async cross-chain messaging—trends now shaping the crypto ecosystem—it failed to gain traction. A key issue was the lack of coordination between appchains and the Cosmos Hub, leaving its purpose unclear. Similarly, Ethereum faces fragmentation: no

unified vision for its L1, L2s competing with each other, and L1 apps launching their own general-purpose L1s.

Ethereum lacks a unified "North Star" (or has multiple), as there's no consensus among stakeholders—researchers, app developers, ACD, clients, etc. While not ideal, this is the nature of a truly decentralized and distributed system. But this tradeoff comes at the expense of losing market share, with competitors outperforming Ethereum in key metrics.

Ethereum's situation is a byproduct of being an early pioneer. As other ecosystems mature and achieve mass adoption, they too will face similar issues, with diverse teams contributing to clients and upgrades taking different directions. For example, Solana contributors are increasing through client teams (firedancer, Jito, agave), Helius, zk teams, and even Anza splitting from Solana Labs.

Even though within the EF we find fragmentation and confrontations of roadmaps, Vitalik thinks that developers within the EF building different products "*don't interfere with each other*".

In reality, I don't know if this is true: some proposals are fundamentally opposed to each other. For instance, ePBS, which aims to remove dependency on relays, stands in contrast to preconfs, which would reintroduce relays or "gateways" for fast preconfirmations. So why enshrine ePBS when relays are going to stay anyway?

Scaling the L1

Increasing The Gas Limit

For context, the block gas limit is the amount of computation (which loosely means amount of transactions) that can be included in an Ethereum block, which is currently set at 30M gas. Ethereum validators can change the block gas limit if >50% of them are in favor of the increase. As of December 2024, 2.5% of them are signalling to increase it, which would be the first time since The Merge (!!) that Ethereum increased the gas limit. Back when Ethereum operated through PoW, miners increased the gas limit 6 different times.

Raising the gas block limit by 100% gives Ethereum L1 the ability to process 100% more transaction load in a day, help reduce network congestion, and lower txs fees (unless induced demand offsets it). Increasing the gas limit has been a hot

topic in Ethereum since, well, forever. Ethereum L1 hasn't increased block gas limits since 4 years ago. There were no agreements around this, basically there are 2 sides:

- Some argue that increasing it will lead to solo stakers not being able to continue validating, some are even in favor of further reduce gas limit,
- Some argue in favor of the increase, pointing to the natural improvements in hardware and client optimization over the past four years, which have made it more feasible for node operators to handle higher resource demands.

In January, Vitalik proposed a modest 33% increase. Nethermind developer Tomasz also stated in favor. Justin Drake raised his validator gas limit at 36M recently.

One downside argument is that increasing the gas limit also increases the resources required to run and maintain Ethereum nodes. A 100% increase in the gas limit could place a heavy burden on node operators, potentially causing some to drop off the network, negatively impacting the decentralization of the network, as fewer operators may be able to participate.

However, during Devcon, Marek Moraczyński presented data-driven research on why an increase is justified and safe for Ethereum. Also, three recent independent studies on node performance have been presented, suggesting nodes on Ethereum can handle larger blocks without major trade offs.

'*Decentralization and solo staking must be means to a clearer end, not the end goal themselves*'

It recently became a hot topic on CT again, with a push in favor of doubling it from 30M to 60M, and it seems this time the general sentiment is gaining some social consensus over raising it. A community-driven initiative led by Mariano Conti and Eric Conner launched a website to educate stakers on how they can signal their support for a gas limit increase called Pumpthegas.

Contrarian views, like Evan Van Ness, argued against any increase at all, saying that gas prices on Ethereum were already relatively low and an increase may negatively impact the health of the network.

As one of the largest pools running validators, with a ~28% market share, Lido community is also actively participating in the discussion, with voices in favor to increase it to at least 40M from Consensys, Everstake, and others.

SSF and Shorter Block Times

Scaling the L1 doesn't necessarily mean we need to aim for perfection in all aspects; improving L1 performance also makes it more suitable for rollups. L2s need an L1 with fast slots and SSF to achieve usable finality, interoperability, and, if they're based, sequencing.

With SSF and shorter block times, we could establish a more effective coordination layer to support cross-rollup composability and synchronization. Multiple protocols realized this and have already implemented these features to become an effective coordination layer (e.g. Celestia, AggLayer, Espresso). As a result, we may see increased competition among networks aiming to become the primary coordination layer for rollups to interoperate seamlessly, developing their own L2 interoperability standards (e.g. Superchain ERC20)

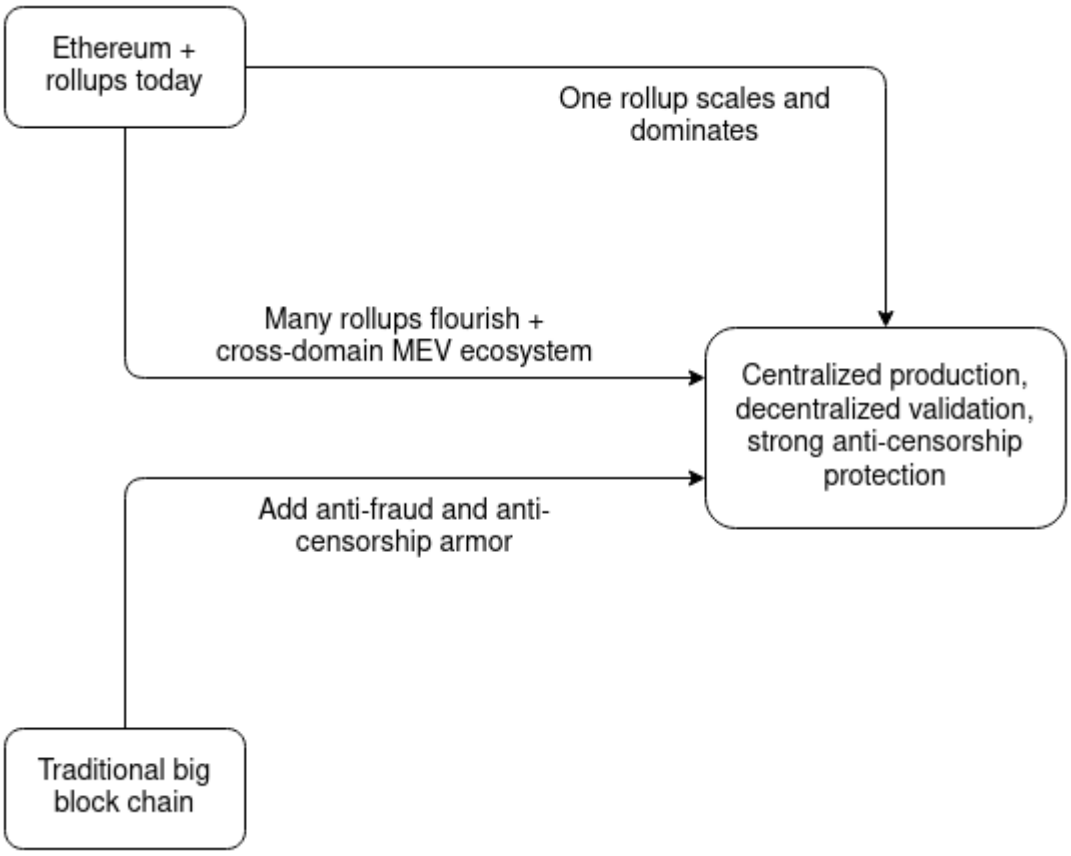
Currently, rollup interoperability is limited due to the lack of an efficient coordination layer. For rollups, waiting for two Ethereum L1 epochs (~13 minutes) to finalize blocks is impractical. To serve as a suitable coordination layer for facilitating cross-rollup commitments, a fast consensus mechanism and shorter block times are a must.

Decentralization and Censorship Resistance

Ethereum must safeguard its core values of global decentralization and censorship resistance, which will be improved by involving multiple parties in block construction and introducing a new consensus transaction policy (e.g., FOCIL and BuilderNet).

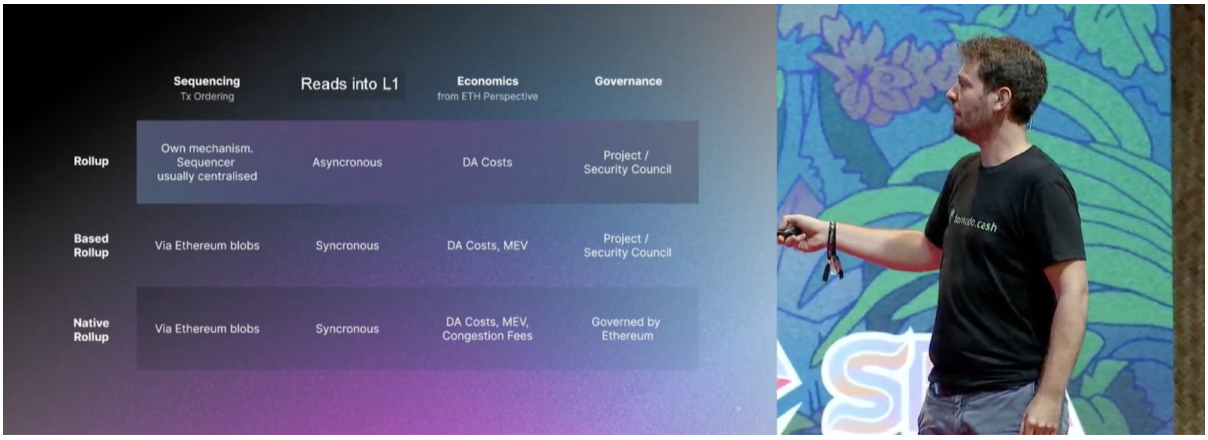
FOCIL works by having 16 validators in each slot who include the transactions they've seen into an aggregated master list, which serves as the starting point for builders to construct the block. Builders can modify the list in two ways: by reordering transactions or by inserting transactions to front-run or back-run, but they cannot exclude these txs from the block. FOCIL still ensures a healthy and decentralized block-building process from an inclusion standpoint, with only the final piece, handled by the builders, being centralized.

However, as of today, block building on Ethereum is dominated by just two builders currently building ~90% of Ethereum blocks, and three relays. Vitalik anticipated this scenario in the Endgame blog



Native Rollups

Martin Koeppelmann, Co-founder of Gnosis, introduced the idea of "Native Rollups" (NR) during Devcon. Native rollups are rollups built with Ethereum values such as decentralization and credible neutrality from scratch, translating into rollups fully controlled by the Ethereum protocol itself and not by multi-sigs. They will have multiple-client proof systems implementation, and rigorous testing of the rollup codebase, just like Ethereum.



It seems to just be an "*enshrined rollup*" rebranding. Native rollups are "*rollups validated on the opcode level, instead of solidity level*". They differ from based

rollups in that based rollup re-utilize the sequencing layer of the base layer, but execution still happens off-chain and then a proof of correctness is submitted to settle on Ethereum. But these proofs can be submitted arbitrarily by centralized operators whatever they want, whereas with NR proofs are built and verify in-protocol by Ethereum validators.

However, although in native rollups the execution also happens off-chain, the introduction of a zk proof (or various types of zk proofs) natively integrated into the opcode allows native rollups to be fully compatible with the EVM, meaning that the base layer directly supports and coordinates rollup operations.

Martin suggests we should create 128 highly interoperable instances of native rollups, which would act as a parallel execution environment while remaining interoperable with others and the Ethereum L1. This sounds a lot like the sharding implementation that Ethereum was seeking in the past. Main differences that I see are that,

1. NR are interoperable with each other as they share the base layer, but won't maintain perfect composability. Sharding would need Inter-shard communication, effectively losing composability between shards.
2. NR Inherits Ethereum L1 security through ZK proofs, while shards inherits Ethereum L1 security directly.

The Beam Chain

During Devcon, a new proposal self-named by the ETH community '*Ethereum 3.0*', surfaced on CT. While not quite what Doug forecasted, it's a proposal to introduce the Beam Chain, a bundled consensus layer upgrade, as a replacement for the Beacon Chain, and seeks to accelerate Ethereum's roadmap by leveraging zk technology. It integrates ZK proofs directly into block validation.

Justin Drake's motivation for introducing the Beam Chain proposal is "*to introduce a new meme*" that unites the Ethereum community and gives developers a common goal. He aims to streamline governance by categorizing upgrades into high-priority tasks and ambitious changes, bundling the latter into a single release to reduce governance complexity and drive collaboration. The estimated timeline? Five years.

Although the Beam Chain proposal is quite cool, it doesn't quite feel like a 'North Star' that everyone is ready to rally behind, especially when considering the

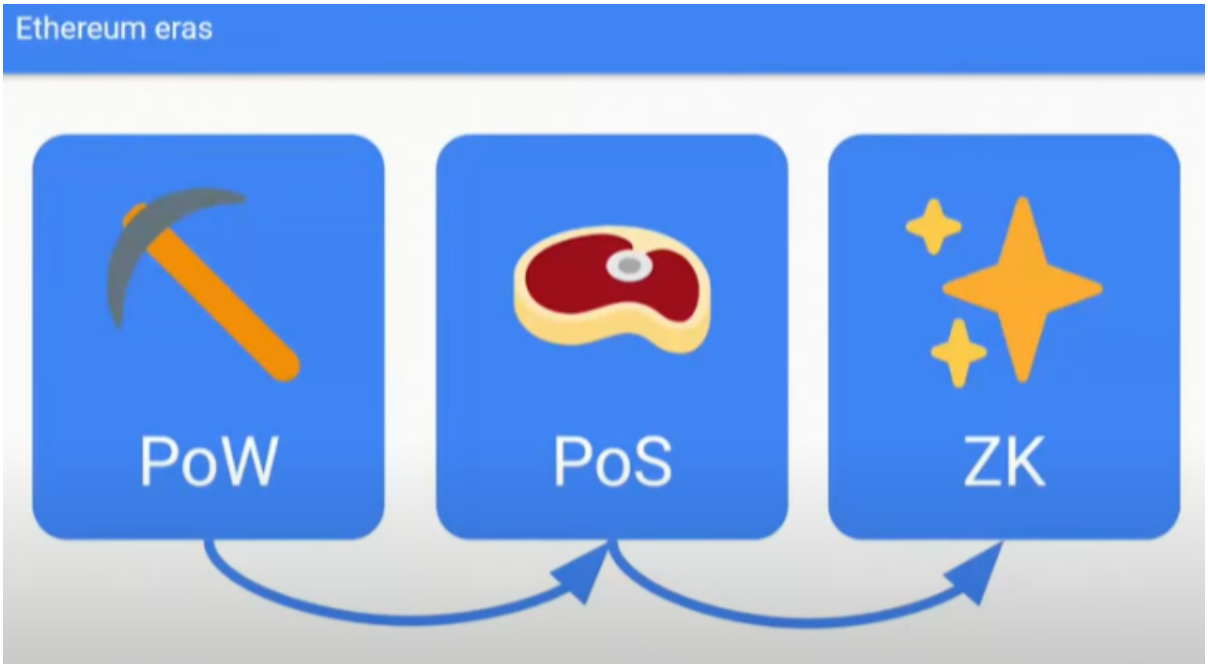
general reaction from CT to the 5y roadmap.

Let's break it down

The Merge transitioned Ethereum from PoW to PoS, replacing miners with validators staking 32 ETH each. This reduced energy consumption by ~99.9% and eliminated computational competition. The Beacon Chain Consensus Layer (CL) merged with Ethereum's execution layer (EL), preserving compatibility with existing smart contracts and dApps. This change set the stage for future improvements like sharding and lower gas costs via DA improvements.

We have all seen the complex and never-ending Vitalik image drawing Ethereum's roadmap, but to be honest I never fully understood it. Here's a simplified breakdown of the key phases:

- **The Surge (In Progress) –** The goal is to significantly scale Ethereum. Proto-Danksharding (EIP-4844) introduced *blobs*, a precursor to full danksharding. These blobs store large amounts of data off-chain, and are targeted for reducing L2 costs by replacing their use of calldata as DA. Once fully implemented, Ethereum will split up into 64 shards processing DA and txs in parallel.
- **The Verge (In Progress) –** focuses on simplifying and optimizing data verification using Verkle Trees, which are advanced data structures that allow nodes to store less information while verifying the entire chain. This will drastically reduce storage requirements and make it easier to run lightweight nodes by just having to verify a SNARK to validate a full block.
- **The Purge (Pending) –** Ethereum will get rid of unnecessary data, reducing chain complexity. Historical data will be removed; nodes will only need to store recent, relevant network data, reducing storage requirements. State optimization will improve sync times and ease the burden on new nodes. This will reduce operational costs, and increase node participation.
- **The Splurge (Pending) –** The '*everything else*' phase, including additional upgrades to improve Ethereum and resolve unforeseen issues.

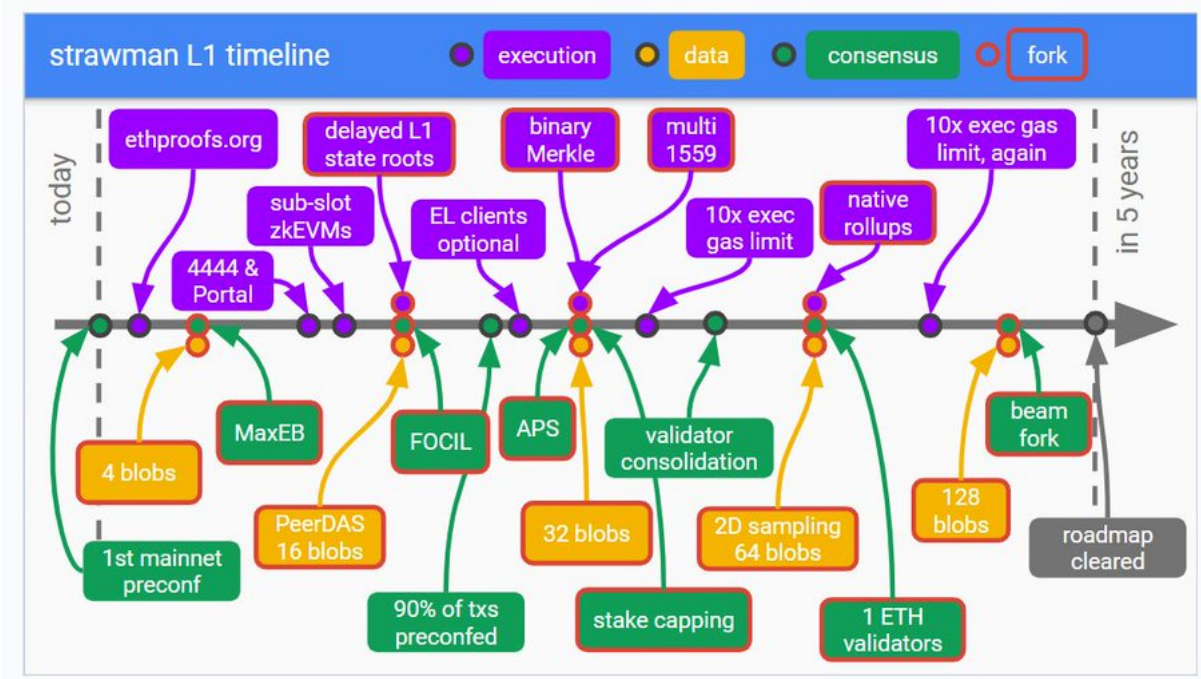


By integrating ZK proofs (SNARKs / STARKs) in-protocol, full blocks could be verified without requiring all nodes to process every transaction.

Community sentiment around the five year horizon was badly received, as many believed these five years would be a '*dark path*' until full implementation. However, I don't think this is going to be the case, as there will be numerous intermediate updates and incremental improvements across all three sublayers in the middle to be implemented in the Beacon Chain; the five year roadmap is going to be decoupled. I think what Justin was trying to achieve by bundling all upgrades into a single proposal is to create a shelling point for the Ethereum community.

For example, some upgrades across the three sublayers will be:****

- **EL** (*enshrined zkVMs, native rollups, delay the stateRoot to reduce L1 latency and increase throughput, increase gas limit by ~10x-100x*)
- **DA** (*increase max blobs per block, PeerDAS, full Danksharding*), and
- **CL** (*inclusion lists to improve censorship, MaxEB to consolidate ETH validator balances, reduce ETH staking requirements to 1 ETH, encrypted mempools to prevent toxic MEV, faster slots)*



ZK vendors (specialized nodes) will generate zk proofs validating sets of txs or even entire blocks. Nodes will need to verify only these proofs instead of processing all txs, reducing their computational burden.

However, I'm skeptical about this in the long-term as my concern lies in the centralization of these ZK vendors. If a small group of zk providers dominates the field, Ethereum could face a centralization risk similar to today's builder landscape, where only [two entities construct over 90% of blocks.]

In the Consensus Sublayer, some improvements include:

- Up to 90% of transactions could be preconfirmed by rollups
- By reducing the staking requirement to just 1 ETH, validators could consolidate their stake
- **Chain Snarkification** to reduce hardware dependency + combined with the 1 ETH staking would enable thousands of new validators to join the network in a much lighter way

But what does it mean to *'Snarkify the chain'*? From consensus to EVM, Beam Chain will try to add zk to every module of the Ethereum protocol. There will be a zk EVM pre-compile that will simplify launching zk rollups, and developers won't need to worry about bugs as the pre-compile will be enshrined in-protocol.

The proof verification process on Ethereum is costly for practically all of the zk

proofs in production today because it requires the execution of expensive on-chain operations to verify the proofs. Currently, the EVM does not have enshrined support (or precompiles) for zk proofs.

Precompiles are optimized and **pre**-built *'smart contracts'* in the EVM itself that can handle specific tasks more efficiently, e.g. the **ecrecover** (*Elliptic Curve Digital Signature Recovery*) precompile helps to confirm *'who'* signed a message being presented to a smart contract. The list of precompiles in the EVM is not capped, i.e. the network had a couple of upgrades to support multiple precompiles as the network needed them as the time went by, e.g. Pectra fork will include a precompile for BLS12-381 curve operations.

The lack of these precompiles for zk proofs means that the verification process is not optimized and computationally expensive to do it directly in Ethereum. Precompiles are designed for specific purposes, such as for a particular proving system like Groth16 or Plonk.

Ok, back to the Beam Chain: By *snarkifying* Ethereum, any entity which consumes the Ethereum chain can do it with extremely low resources, by verifying a single proof, and syncing to the tip of the chain. This would make sure we wouldn't have any further dependency on Infura, but does this imply we'll get rid of centralization forces? As stated earlier, with the introduction of ZK-vendors, I don't think so.

Other improvements discussed in the proposal include

- Reduce slot times from 12s to 4s
- Preconfirmations as fast as 100ms, which he stated that "*will soon give faster-than-Solana transactions*"
- Implement APS (Attester Proposer Separation). Similar to PBS, which decouples the tasks of proposers and builders, with APS the goal is to decouple attestation tasks from proposers to **I.** remove timing games as a concern to validators, **II.** remove MEV spikes due to volatility, **III.** remove worries about sophistication by having to deposit collateral with preconfirmations
    - However, validators are still responsible for the most critical part of block building: which is to include tx on the chain and be censorship resistant, something that could be addressed by FOCIL.

While I agree that current narratives aren't as compelling as they were in 2021, concepts like decentralization have lost traction, and users no longer seem to care about them. It's clear that what they want is fast confirmations and low fees.

In my opinion, builders are disconnected from users, focusing too much on infrastructure without aligning it with real-world user experiences or current needs. Ethereum still boasts the strongest network effects of any blockchain. It remains the most battle-tested blockchain (with no outages) for Web3 developers and the most thoroughly studied by researchers and developers tackling cutting-edge challenges like scaling, MEV, and censorship resistance.

Does this guarantee that Ethereum will always lead the pack? That's hard to say. Could Ethereum learn from its competitors? Absolutely. The Ethereum community needs to stop underestimating and dismissing its rivals and instead pay closer attention to the needs of the users that Solana has successfully attracted.

That being said, it's worth noting that Drake's Beam Chain is just a proposal, it still needs to gather feedback and community approval, something that's not really happening as of today, even within EF developers.

Pectra

Dencun was the last upgrade on Ethereum L1 deployed in March 2024. Pectra (Prague-Electra) is the next hard-fork upgrade expected to be activated as early as Q1 2025. It targets to improve UX for L1 users as well as protocol consensus upgrades.

- **Account Abstraction**

One of the potential upgrades included is EIP 7702 crafted by Vitalik, which targets EOA accounts to introduce a new tx type for end-users. Once implemented it will add additional capabilities to their accounts such as:

- batching transactions (so we wouldn't have to approve txs one by one),
- authorizing multiple on-chain actions from signing a single transaction,
- sponsorship, paying for a tx on behalf of another account, and
- introduces custom-logic on-chain to be able to e.g. customize the conditions of spending on the account balance.

Since most users interact with Ethereum through wallets, wallet developers will need to integrate the new tx type and make it easily accessible for users to improve their UX.

- **Staking Max Balance**

EIP 7251 increases the maximum effective balance of validators from 32 ETH to 2048 ETH and allows existing validators with a maximum effective balance of 32 ETH to consolidate their stake into a single validator instead of multiple ones, thereby reducing the number of validators on the network, which currently are >1 million.

- **Blobs**

A blob capacity increase in Pectra is likely to be included, as there is general consensus that it is necessary to accommodate rollup demand and prevent fee hikes. However, since EIP 7549 (PeerDAS)—an upgrade aimed at improving data availability (DA) efficiency through DA sampling instead of requiring full block downloads—is unlikely to be included in Pectra, an alternative proposal suggests a simpler adjustment to reduce DA costs. Currently, Ethereum processes up to six blobs per block, dynamically adjusting their costs to maintain an average of three blobs per block.

EIP 7742, *Uncouple Blob Count Between CL and EL*, introduces flexibility by allowing the CL to dynamically adjust the max and target blob limits. This change eliminates the need for hard forks on both the EL and CL to update blob capacity. Instead, future adjustments to DA capacity can be made exclusively through the CL, ensuring that the blob gas fee mechanism scales effectively with updates to target and maximum values.

EIP 7762 proposes increasing the `MIN_BASE_FEE_PER_BLOB_GAS`. When blob demand exceeds the target rate, the protocol automatically raises the base fee for blobs, following a mechanism similar to Ethereum transactions under EIP 1559. This adjustment aims to make blob fees more responsive to demand fluctuations, enabling faster price discovery in the blob fee market.

Lately, there has been criticism labeling rollups as parasitic to Ethereum and questioning the value of the rollup-centric roadmap. Introducing a small base fee could effectively address these concerns, and encourage rollups to flow a more substantial portion of their revenue back to Ethereum.

It's intriguing, to say the least, to see the Ethereum community align on increasing the blob count, which effectively increases block size, yet remain divided on raising the overall gas limit due to concerns that it could undermine decentralization by discouraging solo stakers. In the short term, these DA improvements may reduce Ethereum L1 protocol revenue, as rollups posting blobs will pay fewer fees to Ethereum, with the revenue instead going to L2 centralized sequencers. However, L2 users will greatly benefit from this, as it will lower tx fees for them.

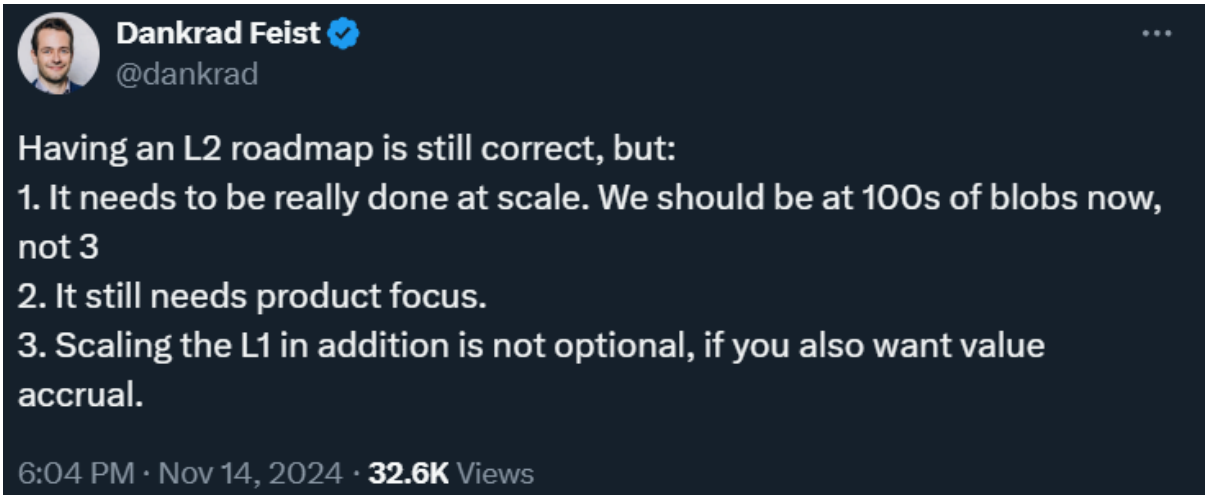
This has always been a common criticism of the rollup-centric roadmap: rollups posting DA pay minimal fees to Ethereum, raising concerns about the potential impact on ETH's value. On top of that, when rollup costs posting DA to Ethereum become too high and saturate the blob market, causing blob costs to rise for every rollup, they may switch to an alt-DA to reduce their costs.

That's why Ethereum needs to be prepared to increase the blob count for rollups, to support the potential demand they will face in the future.

Still, everyone agrees that the L2 roadmap is the correct approach. The million-dollar question that no one can give a clear answer to is, what should be kept on L1 and what should be outsourced to L2. It's, as Vitalik himself said, "*A big question that any L1 scaling roadmap needs to answer is: what is the ultimate vision for what belongs on L1 and what belongs on L2?"*.

Some claim that Ethereum should "*Maximize what you can do on L1, then overflow to L2*". Some others think that "*there is no future for the L1 directly. The L1 is a settlement layer"*. It's an open question that I believe will ultimately be decided by those who truly care about it: the users. So far, the "migration" from L1 to L2s has not been fully completed, as Ethereum L1 is still the preferred launchpad for some major apps (Ethena, Morpho, EigenLayer). It remains the dominant layer with the most network effects, liquidity, and composability between major apps.

I believe the rollup-centric roadmap is paving the way for Ethereum to support 100x-1000x more on-chain activity than the Ethereum L1 can handle. However, this needs to be considered in the long term, as protocol revenue will likely decrease in the short term if Ethereum improves DA for rollups. Over-optimizing for rollups might be counterproductive for ETH' value.



Yet, rollups are still highly centralized being operated by single sequencers, and Ethereum L1 core values of "*decentralization, permissionless, and censorship-resistance*" are no longer properties we have seen rollups prioritizing. Trading off these values for fast confirmations and low-fees was understandable a while ago as they were in early stages. However, as Vitalik said early this year, "*I take this seriously. Starting next year, I plan to only publicly mention (in blogs, talks, etc) L2s that are stage 1+, with maybe a short grace period for new genuinely interesting projects.*

It doesn't matter if I invested, or if you're my friend; stage 1 or bust."

> **The ecosystem's standards need to become stricter: so far, we have been lenient and accepted any project as long as it claims to be "on a path to decentralization". By the end of the year, I think our standards should increase and we should only treat a project as a rollup if it has actually reached at least stage 1.**
>
> After this, we can cautiously move toward stage 2: a world where rollups truly are backed by code, and a security council can only intervene if the code "provably disagrees with itself" (eg. accepts two incompatible state roots, or two different implementations give different answers). One path toward doing this safely is to use multiple prover implementations.

As a reminder, Stage 1 rollups are the ones that have a system which satisfies these properties: A complete and functional proof system is deployed, there are at least 5 external actors who can submit fraud proofs, users are able to exit without the help of the permissioned operators, and in case of an unwanted upgrade by actors more centralized than a Security Council, users have at least 7d to exit. As of December 2024, there are just three rollups at Stage 1 (Arbitrum, Optimism, and zkSync Lite)

These EIPs targeting Account Abstraction (AA) in L1 and improving blob capacity

seem to cater to different types of users, highlighting that Ethereum is working to improve both L1 and L2 UX. DA improvements are designed to help rollup users, which may shift traffic away from L1, while EIP 7702 (AA) focuses on UX improvements for L1 users.

However, it's important to remember that the upgrade lifecycle typically follows four phases: I. Devnet, II. Testnet, III. Mainnet, and after successful implementation in L1, they are generally adopted by rollups and side chains, so they will be benefiting from these improvements in the end.

Fusaka

The next upgrade after Pectra will be Fusaka (Fulu-Osaka), which aims to introduce a key single upgrade, *Verkle trees*, in order to minimize the technical risks associated with bundling multiple upgrades. This transition is expected to be major, as it will change how Ethereum nodes manage network state. Currently, Ethereum uses *Merkle Patricia Trees* to store network state (account balances, smart contracts, and storage data). The problem is that as Ethereum grows, the data stored in these trees becomes large and harder to manage. Nodes need to store all this data to fully validate the network, requiring more and more resources over time.

Verkle Trees are an advanced version of Merkle Trees that use vector commitments, enabling much smaller proofs of data inclusion. Nodes handling state will be relieved allowing them to be **stateless**, meaning that this will enable lightweight nodes to validate the network without storing all historical state data. These nodes rely on proofs provided by full nodes (**stateful clients**). With lower resource needs, more users would be able to join the network and run nodes.

Although these are not the main focus of the Pectra or Fusaka, Eth community is also actively discussing reducing ETH issuance and improving censorship resistance, which is likely to have a more significant impact on ETH price.

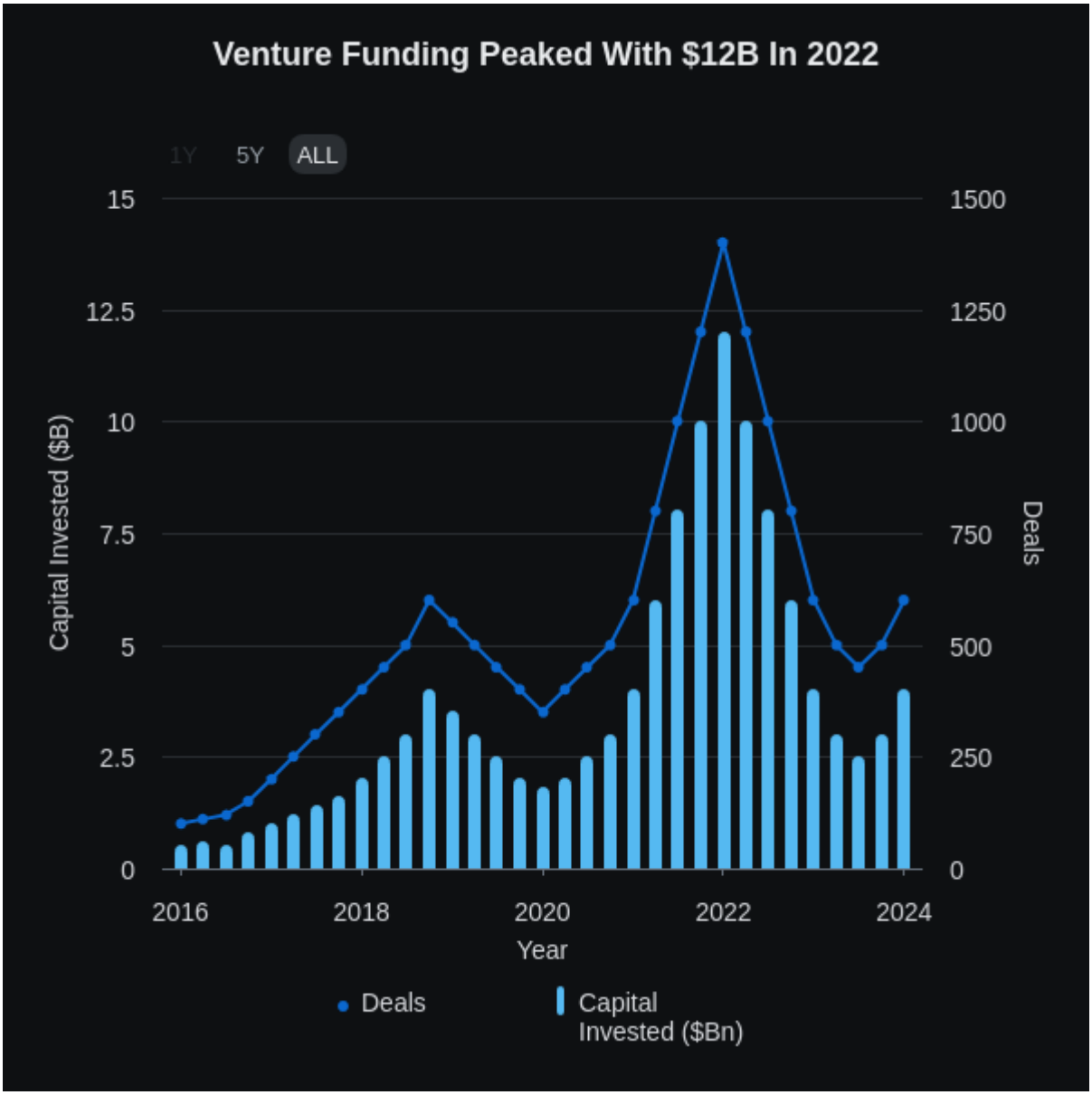
An Age of Cryptographic

Renaissance

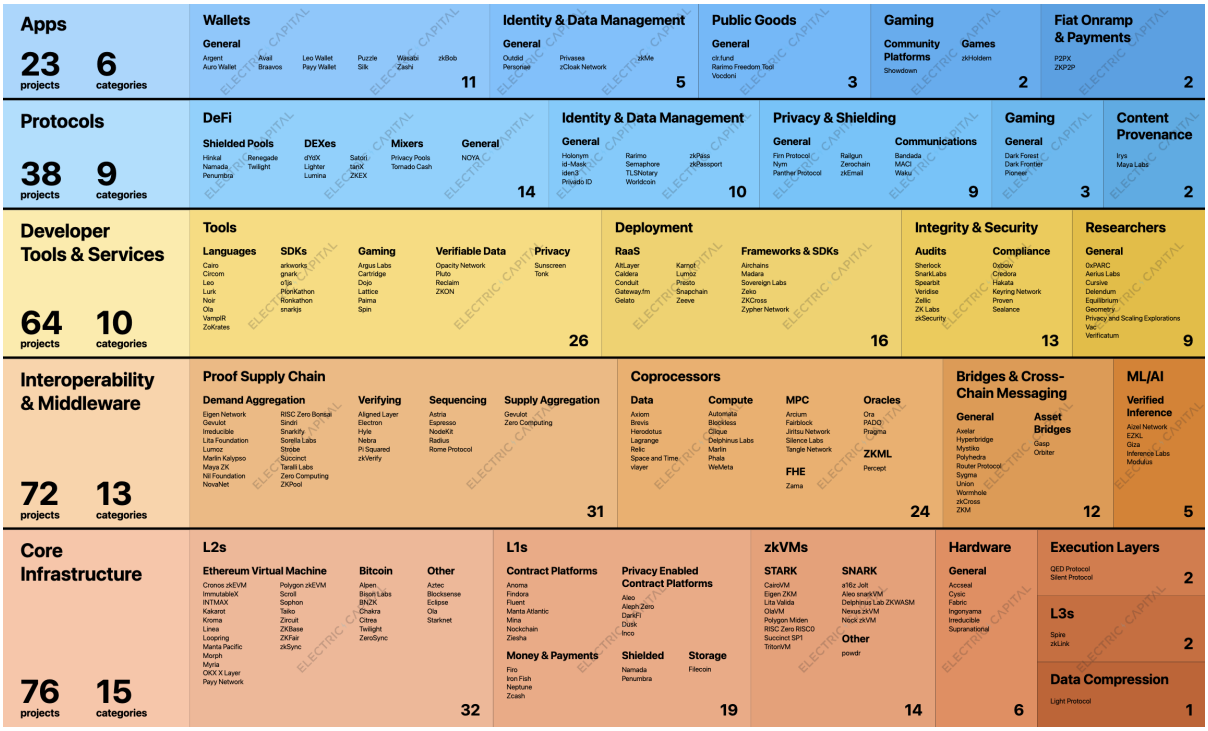
Over a hundred years ago, we used cryptography in the context of electronic warfare with specialized machines such as the Enigma, which was purpose-built to decrypt military communications in World War II. It shortened the war by several years and saved millions of lives, enabling the Allies to anticipate and counter German actions effectively.

Since then, with internet access raging and reaching further than ever, the rate at which humans innovate has drastically increased in the last few decades. We've been converging at generation-defining technologies at shorter intervals, time and again. As with all things in markets, we're in a race to identify and capitalize on these technologies.

> Now let me tell you how fundamental ZK is. ZK is as fundamental as information itself. Read on to see why:
>
> Think of information for a second. Wherever there is information, there is a marketplace: It gets created in one place and consumed in another. This naturally creates the...
>
> — Vanishree Rao (@vanishree_rao) December 11, 2024



In the last decade, crypto companies have raised an estimated $200B across 17K deals. A good chunk of this funding has flown to research and development on cryptographic methods across ZK, FHE, MPC, and TEEs.



Cryptography Market Map by Electric Capital

From scaling Ethereum with rollups, solving interoperability with bridges, and having distributed multi-sig wallets. **We are now at the cusp of entering the age of programmable cryptography with general-purpose systems that don't just cater to one-off use cases but rather act as a co-processor for n types of uses**. Similar to how we initially used specific devices to monitor oxygen levels, heart rate, and ECG to using wearable health monitors like smartwatches that integrate multiple health-tracking sensors (heart rate, blood oxygen levels, ECG, etc.) into a single, portable device.

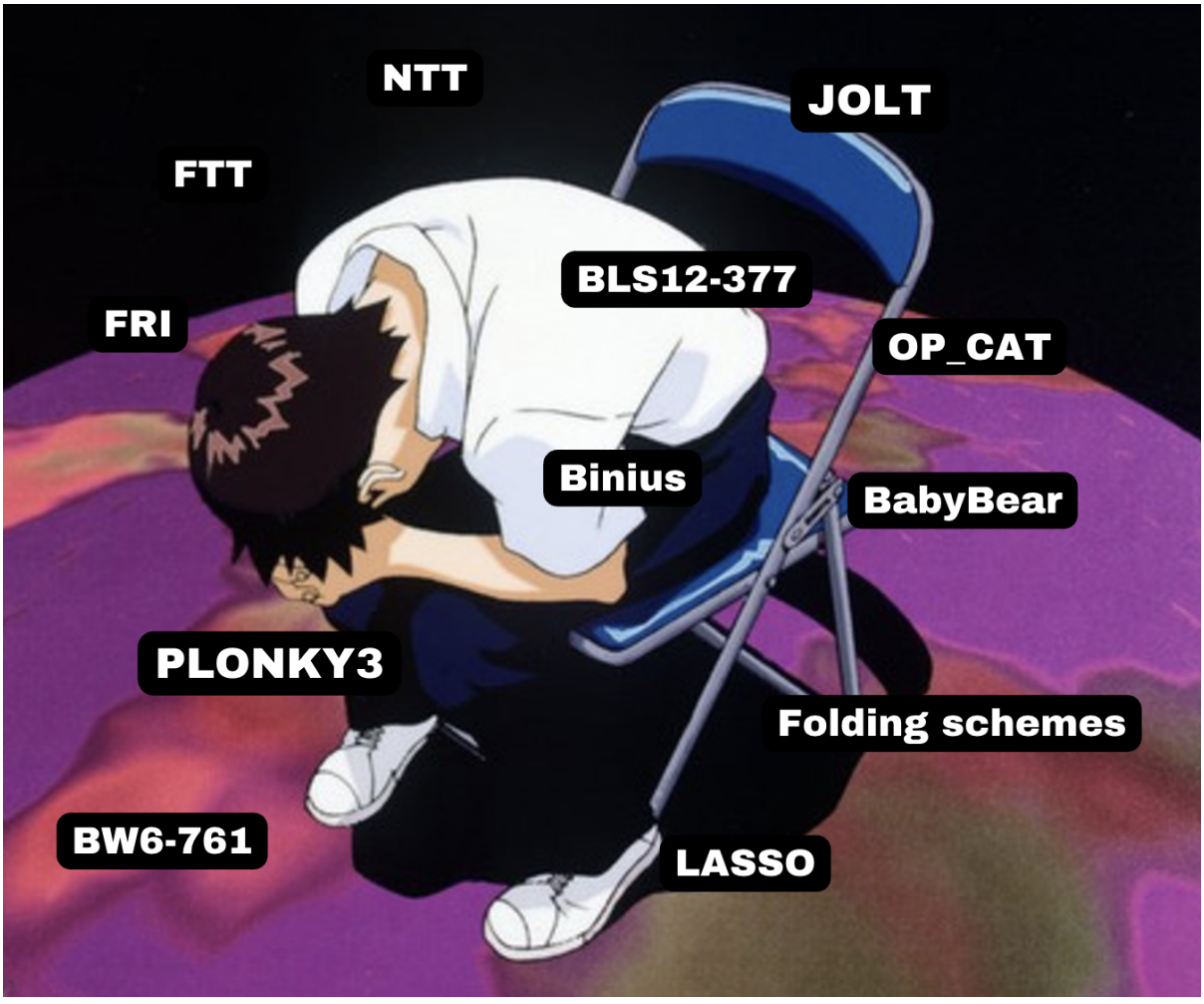
> Privacy Enabling Tech (ZK, MPC, FHE, TEE) may have more or less the same goal but each of them has a differentiated approach to privacy with its own set of ideal use cases and limitations.
>
> ZKPs enable single-user privacy without data disclosure. MPC relies on collective trust... pic.twitter.com/AT0YaAgyKp
>
> — Muhammad Yusuf (@yusufxzy) November 24, 2023

Over the last few years, we've had quite a few advancements across ZK, MPC, FHE, and TEEs that I believe are taking us closer to realizing the dream of a global private-shared state. The way programmable cryptography could play out

resonates with Vitalik's idea of glue and coprocessor architecture. We've seen this play out with other components.

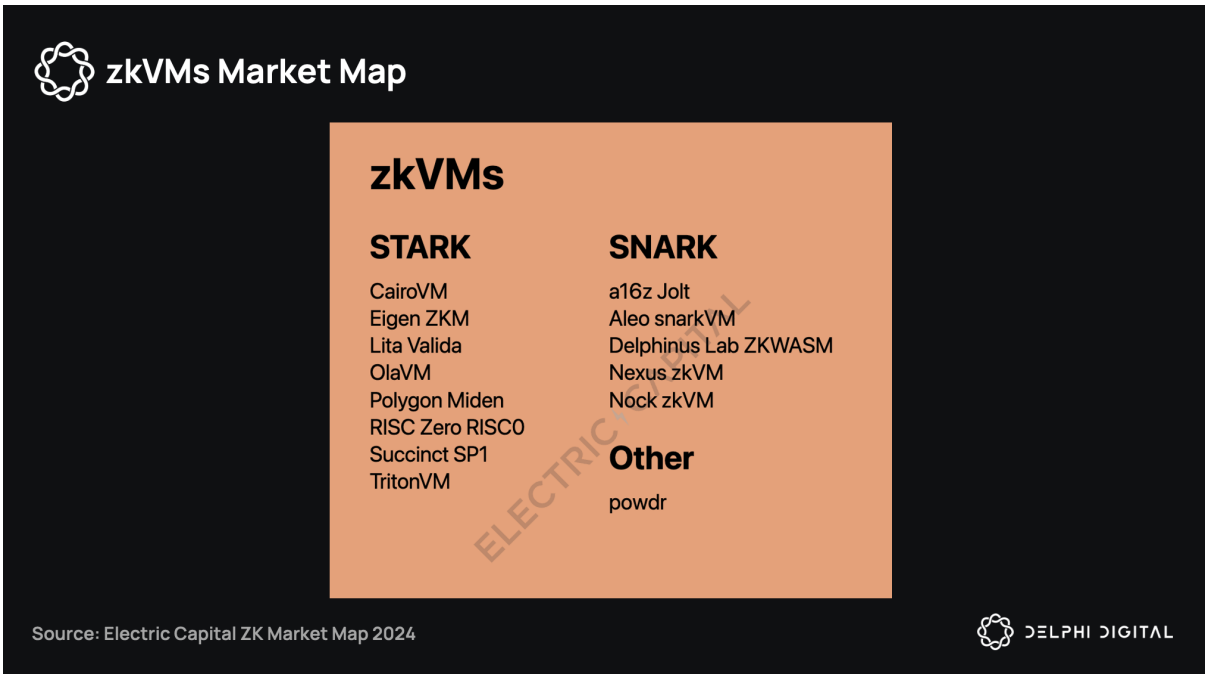
Blockchains optimize for throughput and offload computation requiring specific & external logic to off-chain components, as done with oracles and bridges. The same is going to happen with verifiability/confidentiality/privacy.

The "glue" acts as a flexible interface, such as the Ethereum Virtual Machine (EVM), enabling developers to build complex systems without being constrained by the performance limitations of a single component. "Co-processors," on the other hand, are designed to handle computationally intensive operations that have specialized designs such as generating zero-knowledge proofs, encrypting and decrypting data, or executing MPC or FHE algorithms. Developers can leverage zkVMs and fheVMs, among other coprocessors, without being cryptography experts themselves. **The next frontier for cryptography is with general-purpose cryptography, where we move from computing within silos to shared-private states, where we abstract cryptographic complexities to allow developers to get wild with their ideas.**



ZKfying Everything Under The Sun

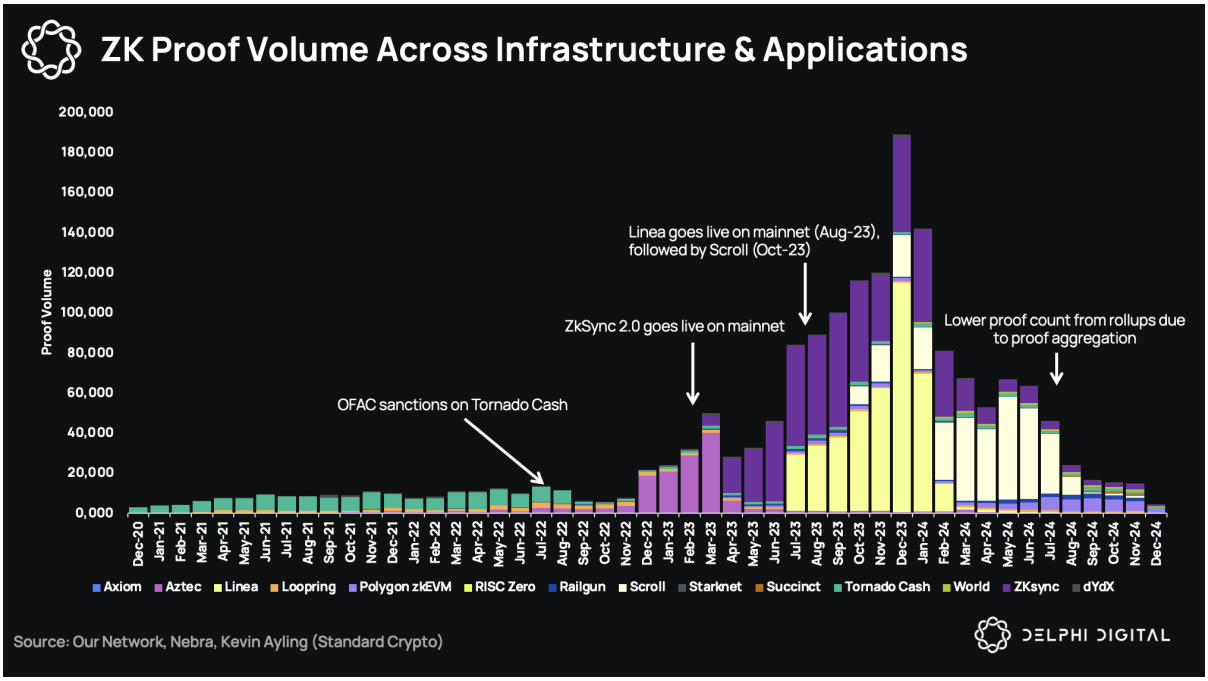
Out of all the cryptographic methods, research & development on ZK has received the biggest push with both capital and talent. From being used to create validity proofs across several zk-rollups to helping build the future of bridging with ZK light clients. Optimistic rollups such as Optimism, Arbitrum, Taiko, Fuel, and many others have been working on hybrid proving systems by integrating zkVMs such as Risc0 zkVM and SP1 by Succinct.



Beyond Ethereum, non-ZK/privacy L1s such as Solana and Sui have invested resources in ZK. One of the ways Solana is trying to manage state growth is with the use of ZK compression. Sui has built out ZK tooling with zkLogin that allows users to create and manage wallets with emails safely without compromising their data.

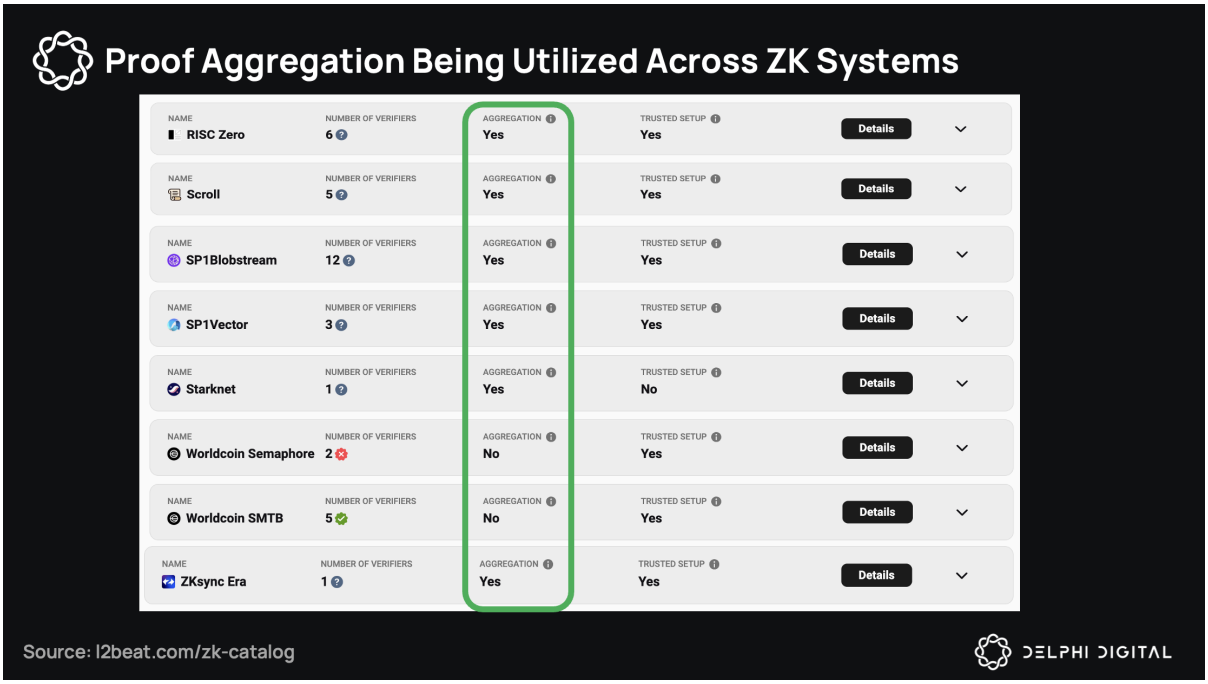
Tangible Impact of ZK Across Infrastructure & Applications

Based on data from Our Network and zkstats.io, over ~1.6M proofs have been submitted to Ethereum mainnet since 2021, with total settlement costs (TSF) crossing ~$60M. TSF peaked in Dec-23 with $15M, most of the fees being paid by Linea, ZkSync and Scroll.



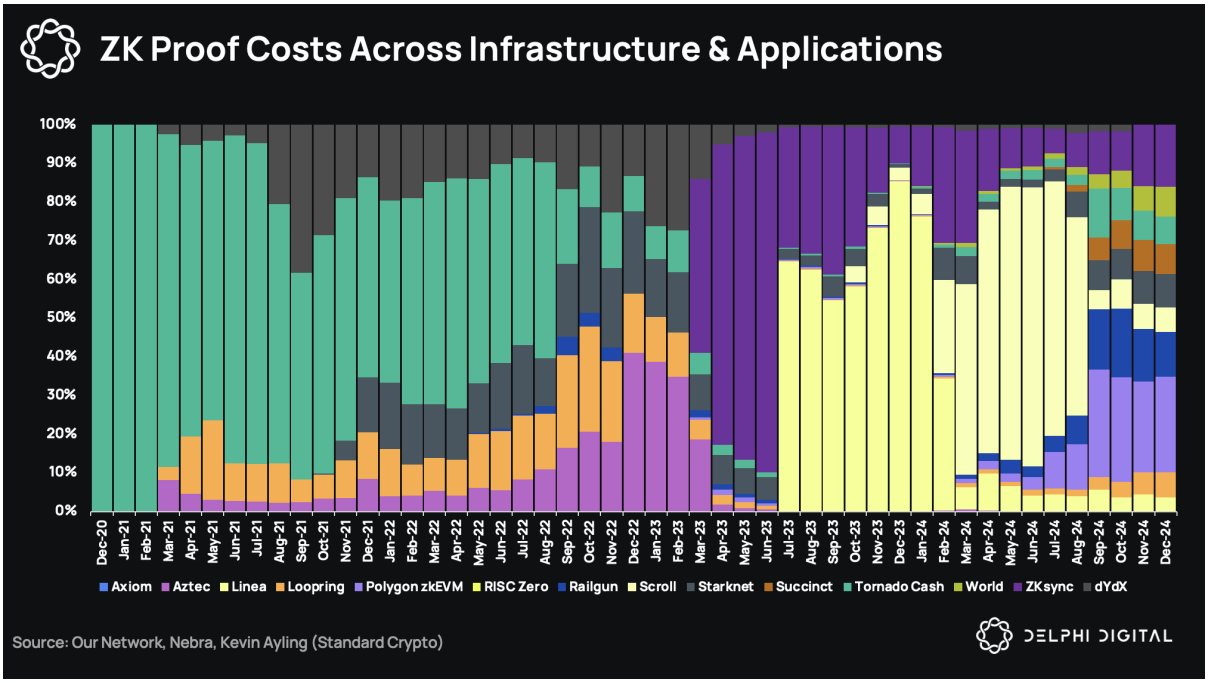
The numbers might seem underwhelming given the period of operations and the number of projects, but there are a few things to consider. Proof generation and verification costs have been getting more efficient.

Most, if not all, of the rollups, use recursive proof aggregation to cut down volume and costs. This is where one ZKP is used to validate a sequence of ZKPs, leading to only a single or few ZKPs being submitted for verification on Ethereum.

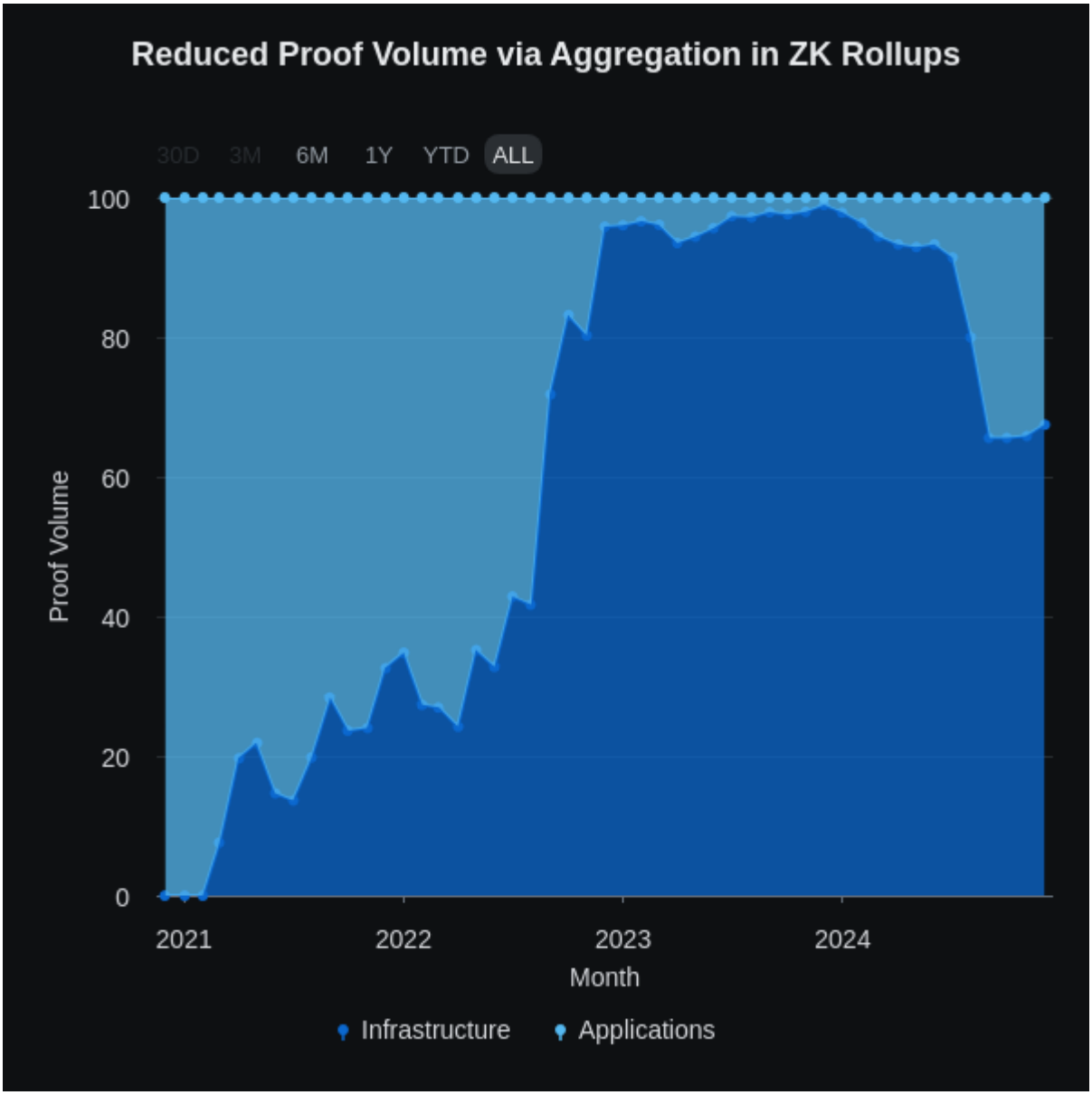


Most of the proof volume and TSF are seen coming from zk-rollups compared to applications. Applications such as Tornado Cash and Railgun not only operate on

Ethereum but also on L2s and networks such as BSC and Polygon. Given that the costs of privacy and transaction fees are magnitudes lower on L2s, a significant volume of their activity does not directly settle on Ethereum.



Proof volume from Tornado Cash started to decline after OFAC sanctions were placed on them in July 2022. In the chart below, you may see proof volume from infrastructure lowering in the last quarter, which is again due to proof aggregation practices from zk-rollups. I believe we may see ZKP activity rise from private DeFi applications. Dark pools such as Renegade may spring up in activity, but their settlement activity will be limited to L2s.

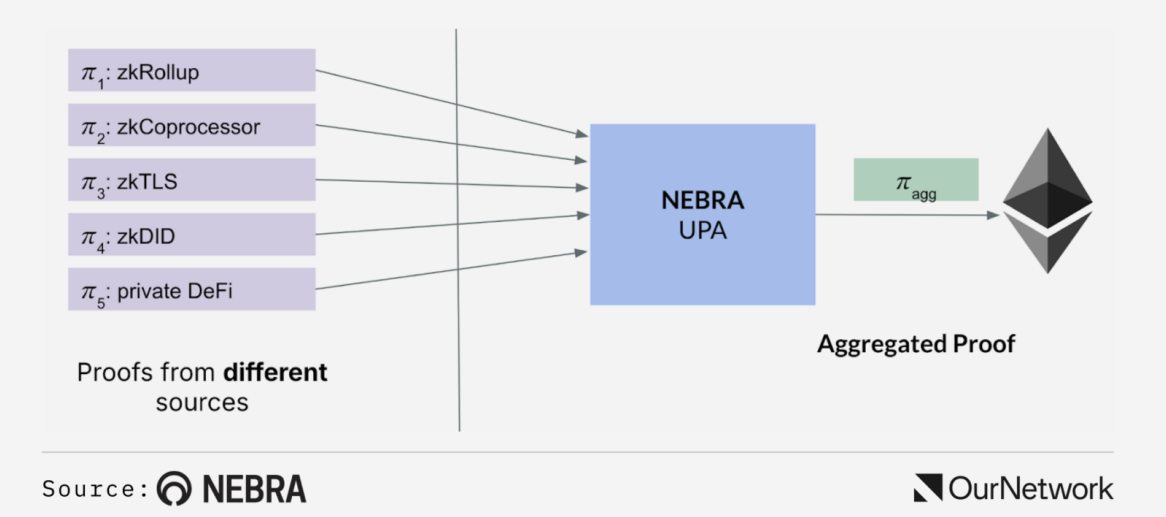


Economies of Proofs

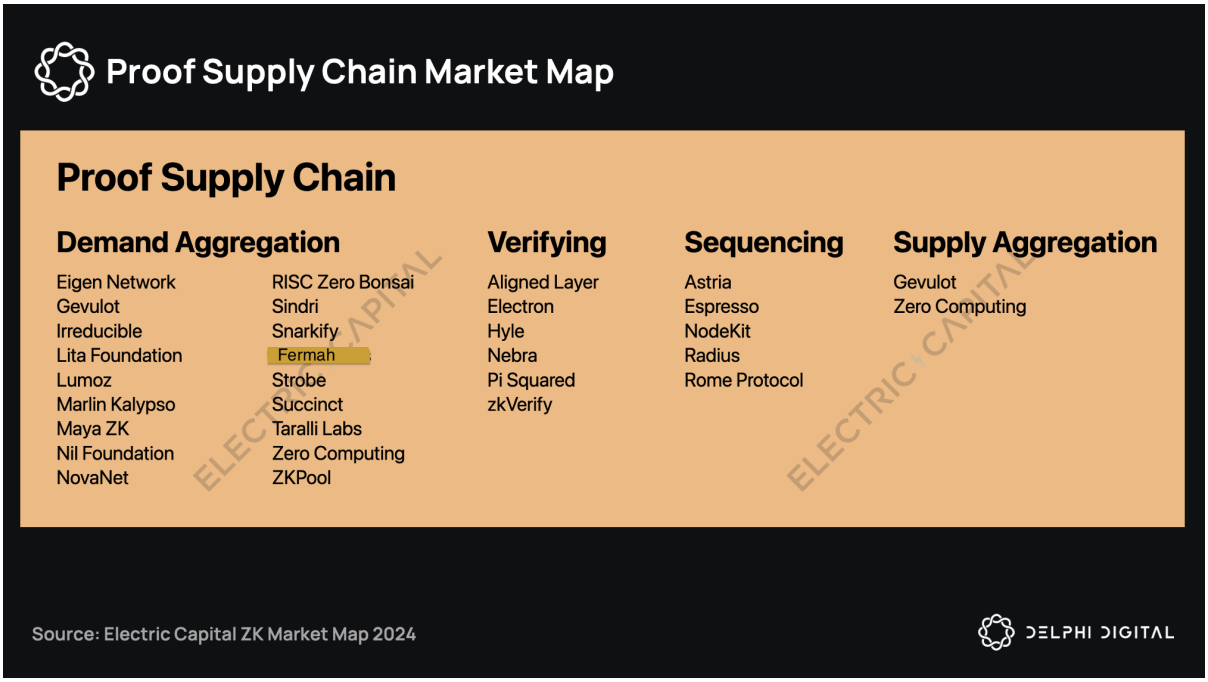
As more rollups and applications rely on zkVMs for verifiable computing and privacy, the demand for ZK proofs will rise. We're starting to see this uptick in demand, and this increase in appetite for verified computations will be served by next-gen zkVMs such as R0, SP1, and Nexus that run efficient and composable proving systems. ZkVMs tap into proof markets that pool multiple prover networks together to serve this appetite. These prover networks, composed of AISCs, GPUs, and other specialized hardware, compete to deliver proofs faster and cheaper.

With increased proof demand, proof markets have economies of scale that drive down the marginal cost for generating each proof. This could incentivize provers to invest more in dedicated hardware such as VPUs (Verifiable Processing Units),

ASICs, or GPUs. With costs being driven down, especially with recursive aggregation systems like Nebra, more developers, rollups, and applications may rely on zkVMs. This could lead to verifiable computing, confidential or otherwise, being closer to standard practice and not just a niche advantage, further increasing demand. This is already quite evident, as seen with the adoption of ZK above, and will only increase in the months and years to come.

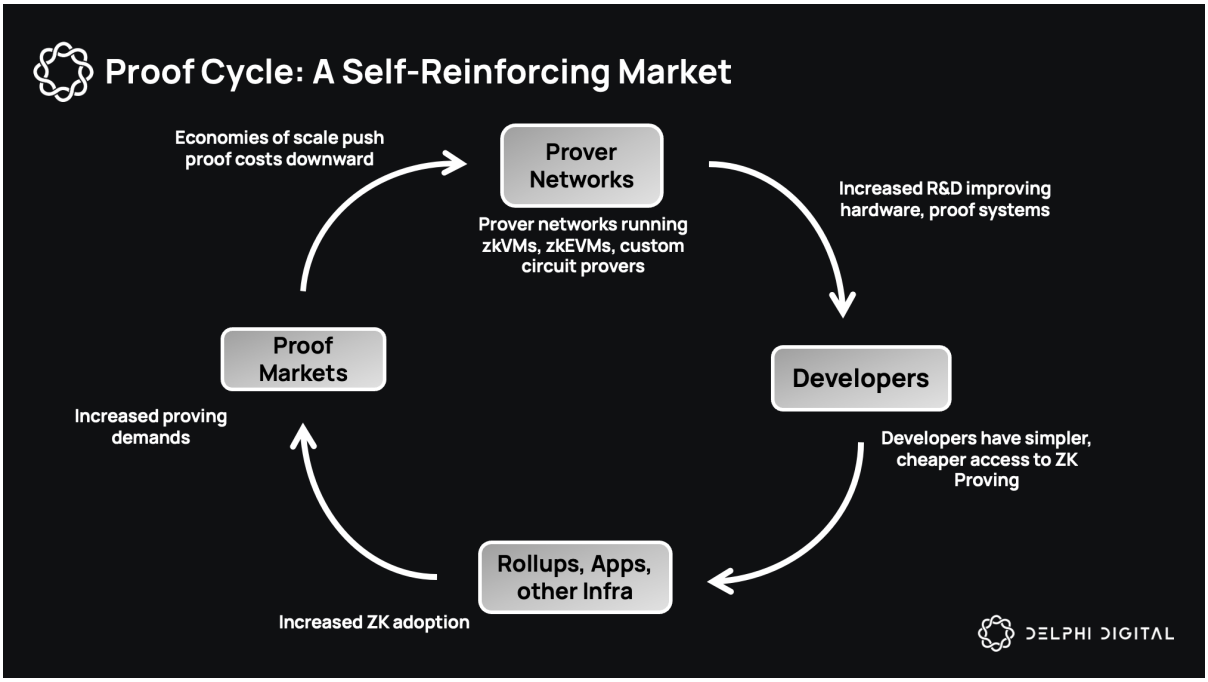


With more requests flowing into proof markets, individual zkVMs and prover networks will invest in improving their efficiency. We will see more experiments with cutting-edge proof systems, better compression techniques, and advanced cryptographic primitives that reduce the overhead of proof generation and verification. Improved zkVMs, in turn, lower barriers for developers, encouraging even broader usage of zero-knowledge proofs across a range of unique applications and rollups.



Proof markets such as Fermah, Risc Zero's Boundless, and Gevulot, among several others, play the role of orchestrating demand and supply between rollups and applications seeking proofs, and prover networks running zkVMs, zkEVMs, and custom circuits on provers. Their mechanism design will help price proofs and open doors for competition among provers.

With all this competition in place, proof markets can build moats to defend their market share. Better liveness for proof generation will lead to more demand flowing into that particular proof market. This also means that provers don't sit idle and can generate better ROI for the investments made on hardware, among other things. It also helps if the proof market can satisfy a diverse range of proof demands. This ensures increased utilization rates of machines. Provers will continue integrating with proof markets that help them get the best ROI.



Over time, I see this turning into a self-reinforcing cycle. More adoption of zero-knowledge proofs leads to larger proof markets and networks that have economies of scale and price reduction. Lower proof costs and improved performance encourage more dapps, rollups, and infrastructure layers to integrate with zkVMs. Proving systems continuously improve, reducing complexity and further expanding the pool of potential use cases. This cycle is already in motion, and we will soon be able to witness its indicators.

For specific applications, relying on proof markets may not be in their best interests. For example, Renegade, a dark pool on Arbitrum that has been live for over 3 months, processing ~$700K in volume, produces ZKPs attesting order matches within its orderbook that help settle user orders. If Renegade were to hand off proving to a proof market, prover nodes could view specific orders, defeating the purpose of using a dark pool in the first place. For most consumer-facing ZK applications, like Renegade, lightweight-proof systems are sufficient; they aren't computationally expensive.

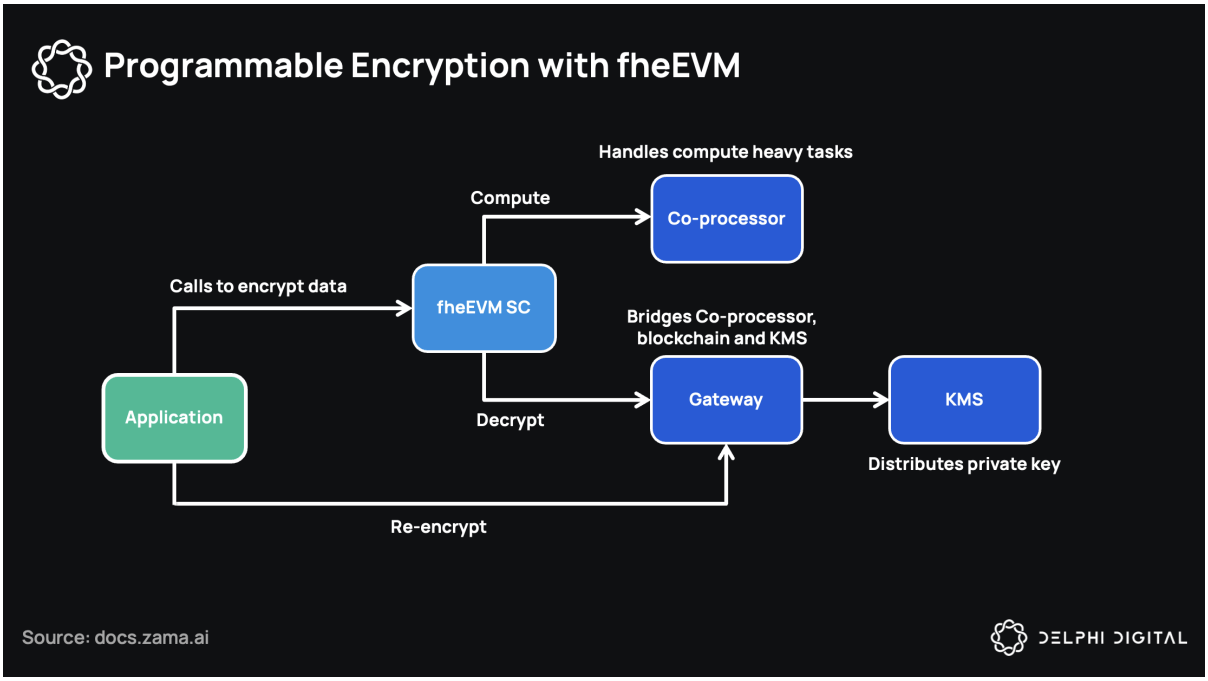
As we've seen with most markets like block building and sequencing, they tend to centralize because of the cost and latency benefits they provide. Until decentralized proof networks are fast enough, we may see similar levels of centralization even with prover networks as they may offer faster and cheaper proofs.

Practical FHE with Zama & Inco Network

FHE simply allows computations to be performed directly on encrypted data without the need to decrypt it first. In the last 2-4 years, FHE has come a long way from being theoretical to being capable of real-world implementations. However, running FHE programs on-chain is highly inefficient due to the computational intensity, the noise, and significant delays imposed by FHE operations.

It is more practical to leverage FHE through an off-chain coprocessor by offloading resource-intensive computations to be performed in specialized environments such as fheEVMs and even utilizing purpose-built hardware such as VPUs (Verifiable Processing Units) such as those being built by Fabric Cryptography.

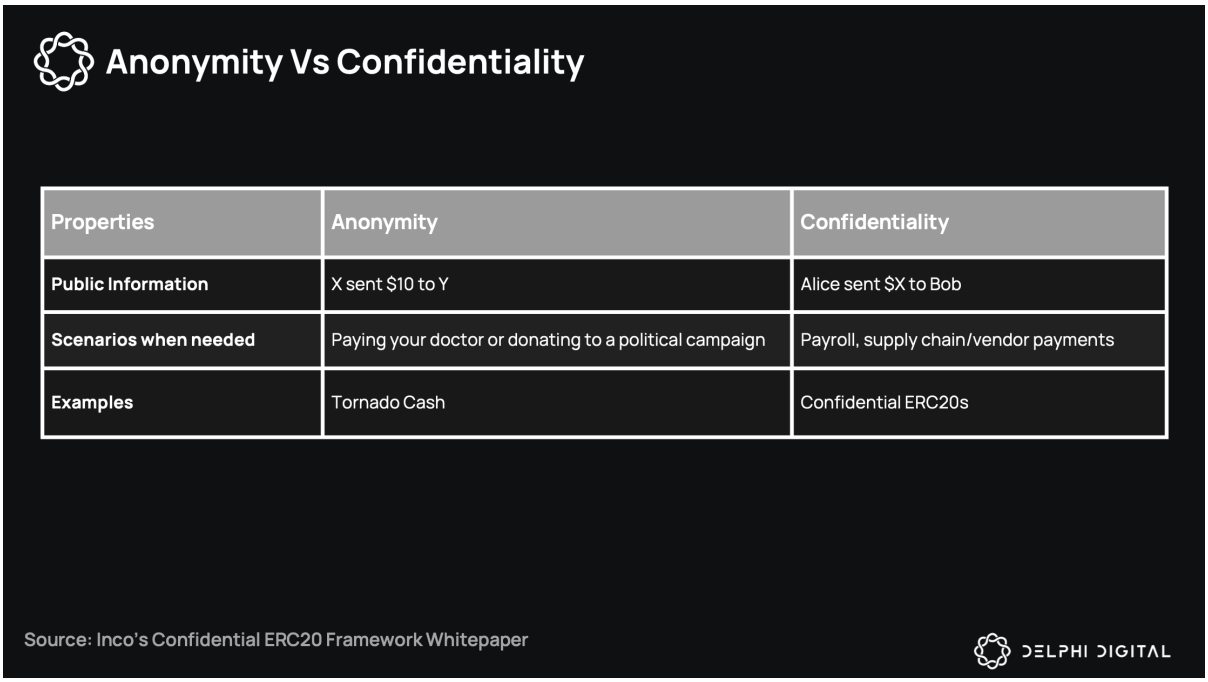
On this end, Zama and Inco Network have pushed the frontier by working their fheEVMs. They allow smart contracts to work on encrypted data while maintaining privacy of shared states. Data can be programmably encrypted, or decrypted as developers specify, enabling access control based on specific use cases across EVM networks.



When users encrypt data using their public key, their private key is not stored with a single validator. Instead, it is distributed across the entire validator network. The key management system relies on MPC, so to decrypt data, a threshold signature is required, meaning that x of n validators must agree to decrypt it.

Inco's Confidential ERC20 Framework

One can simply make global transfers and swaps in seconds on blockchains but the extent to which people rely on blockchains is limited in one capacity. As they are inherently transparent, any and all activity on-chain is traceable and attributable to specific entities or people. For example, on-chain payroll is not as convenient as it sounds. Employees can easily figure out how much their colleagues are being paid by searching their company's wallet address. The same can be applied to vendors or service provider payments. What if newer vendors or service providers figure out exactly how much you paid for your previous engagement? This would land the company in a pretty chaotic situation. That is why we draw the line between anonymity and confidentiality, as seen below.



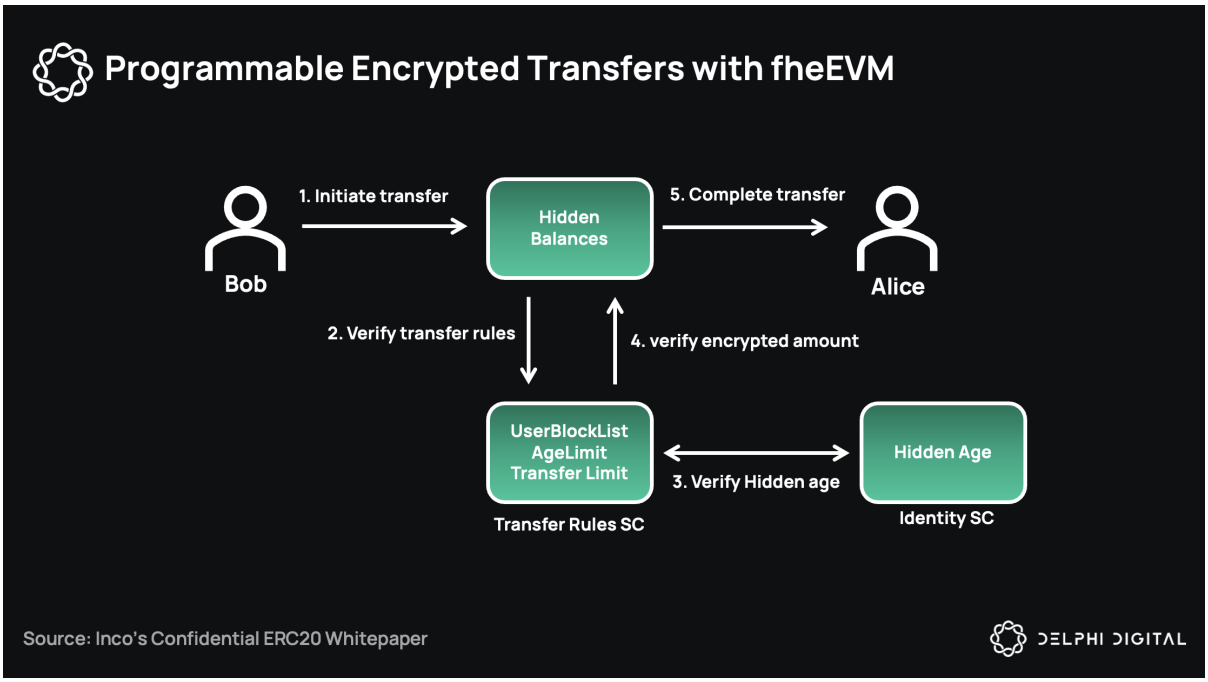
The confidential ERC20 framework builds on the existing ERC20 standard by adding confidentiality as a feature. It is now possible to

- Confidentially mint and burn new tokens
- Wrap existing tokens to be confidential
- Encrypt balances and transfers with access control

Let us take a look at an example of a confidential transfer of a wrapped token with transfer rules

1. Bob initiates a transfer of ERC20 tokens and wraps them into cERC20 with hidden balances using FHE

2. A smart contract initiates verification to check if the user's address is in a block list publicly
3. and then verifies if the sender/receiver's age and transfer limit are within bounds privately with the help of FHE algorithms
4. Once all transfer rules have been verified, verify the encrypted amount
5. the transfer is finalized



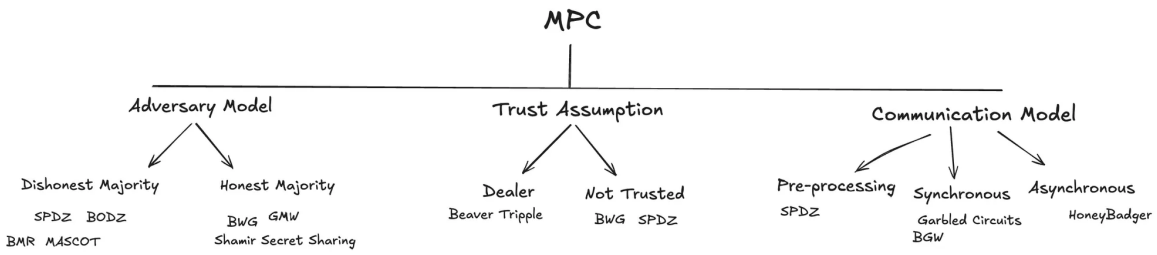
P2P confidential transfers are just one of many potential applications of programmable encryption. This approach could also extend to confidential token vesting, where transparency is required to show how many tokens are being vested over a specific duration to the team as a whole, while preserving privacy regarding the exact amounts allocated to individual employees.

This broadens the scope of how creative developers can get when creating consumer-facing applications. It is only a matter of time until we have Uniswap v4 hook for confidential swaps using cERC20s. Confidential AMMs, dark pools, lending, and blind auctions only scratch the surface.

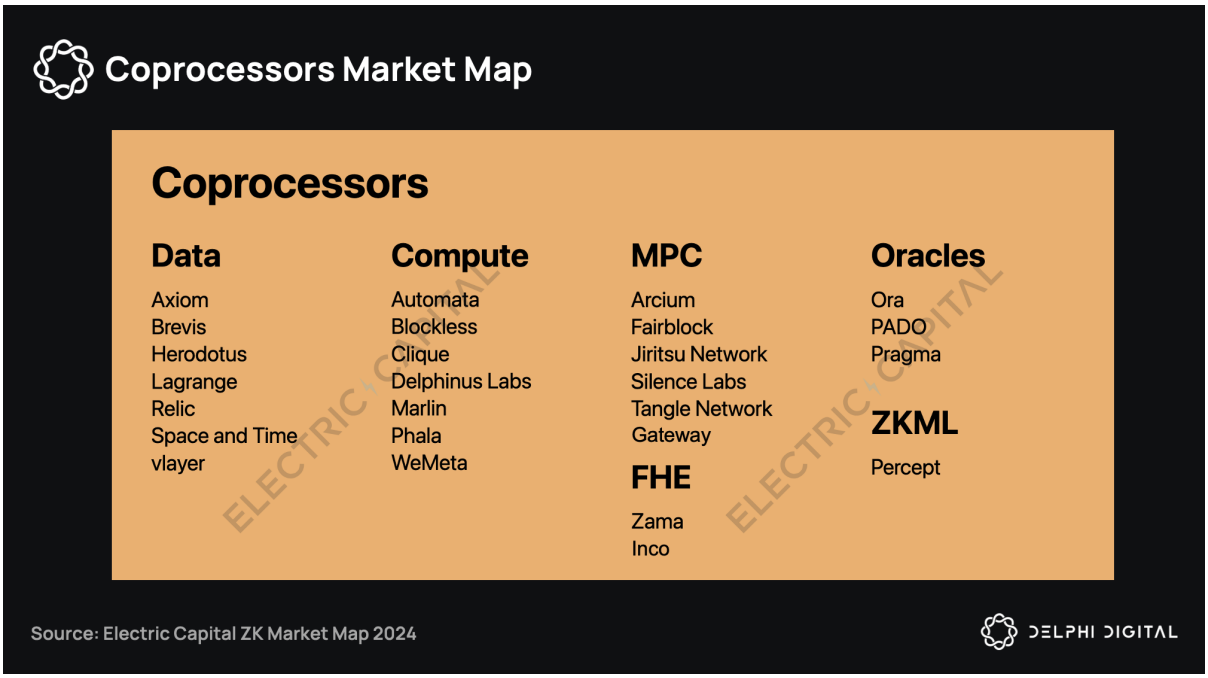
Dynamic MPC Becomes More Accessible

MPC allows multiple parties to compute over different inputs privately to arrive at an output. There are a range of different MPC protocols with their trade-offs across adversary models, trust assumptions, computational overhead,

communication models, and privacy guarantees.



Dishonest-majority protocols like SPDZ and MASCOT are more advanced MPC protocols that tolerate a higher fraction of compromised participants but have greater computational overheads. Honest-majority schemes such as BGW and Shamir Secret Sharing are relatively less compute-heavy but are less secure. Trust assumptions can shift from requiring a reliable dealer like Beaver triples to not trusting any single party with BWG or SPDZ. Communication models range from preprocessing-heavy frameworks (SPDZ) to synchronous (garbled circuits, BGW) and asynchronous models (HoneyBadger), each optimizing for different use cases with their own trade-offs.



As you can see, these are one too many complexities for teams just looking to ship and scale their applications. MPC co-processors such as Fairblock, Gateway, Stoffle, and Arcium handle a lot of heavy lifting under the hood. They abstract away the cryptographic complexity of dealing with multiple MPC protocols—no need to choose between SPDZ, MASCOT, or BGW yourself. They

also manage the underlying infrastructure that keeps these protocols running and coordinates a decentralized network of nodes to ensure threshold decryption works as intended. Co-processors can completely abstract complexities of privacy-preserving, verifiable computations, so developers don't have to spin up their own distributed networks, reinvent secure key management, or become overnight MPC experts.

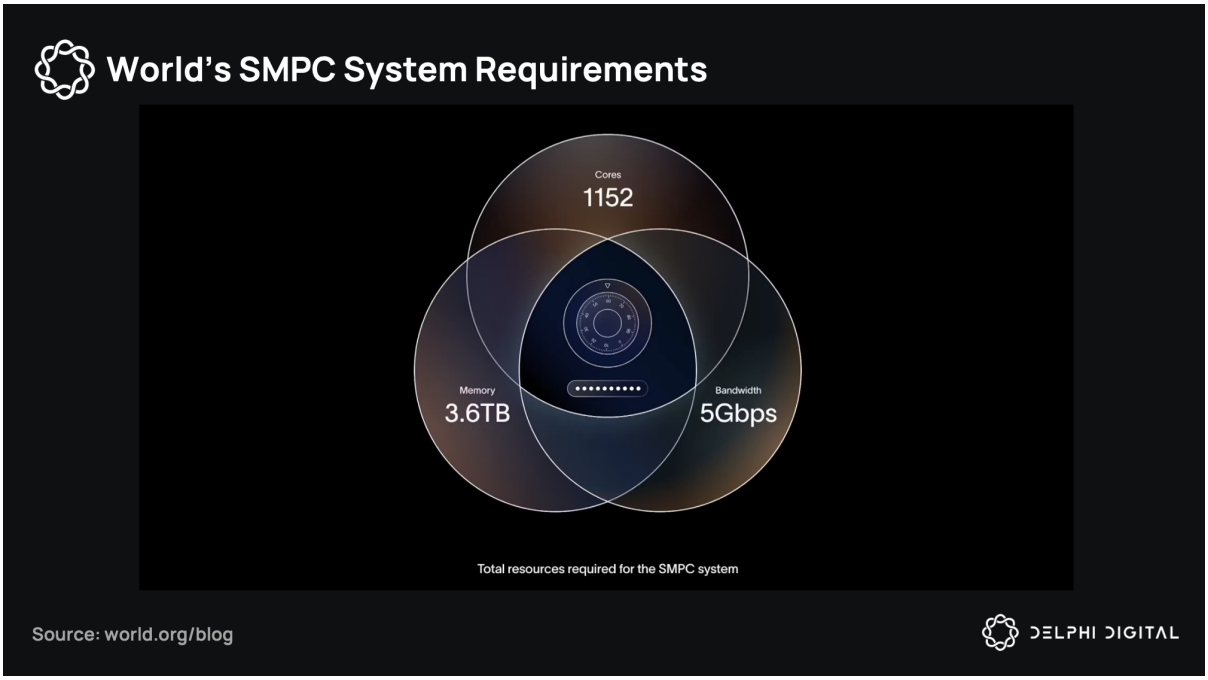
World Chain

World has been operating at the intersection of Crypto x AI. In July 2023, World had iris scanning stations across all major countries. Since then, over 8.7M people have scanned their iris, creating their own unique world IDs.

> No.
>
> 8,790,842 verifications so far.
> [■————————————————————————————]0.88%
>
> — Does World have a billion users yet? (@WorldIDCounter) December 11, 2024

The orb processes iris images on-device and converts them into an iris code which represents the iris texture. This code is stored in an encrypted form across multiple parties using secure multi-party computation (SMPC) built in collaboration with TACEO. In order to verify that each iris is, in fact, unique and not just the same person creating multiple IDs, every new iris code has to be verified against all the existing iris codes. If this were to be done using conventional methods, it would compromise user privacy.

With SMPC, each node holds only partial secret shares of every stored iris code. When a new code arrives, they collaboratively perform computations directly on these shares without decrypting anything. They collectively combine their partial results to produce a final match score, revealing only whether the new iris is unique, not any underlying iris data. This is by far the largest MPC network in terms of users.



A few things to note about how World operates

- With WorldIDs, World Chain will be differentiated between humans and bots. This is increasingly important as we're seeing AI agents become more active onchain and offchain.
- Verified users can periodically claim grants in WLD tokens. So far, over 370M WLD have been given out as grants to users.

While financial activity is yet to rise on World chain, I believe that its 8M WorldID user base is a dry powder for real-world consumer applications to leverage.

All ZK, FHE, MPC, TEE coprocessors will target DeFi use-cases because that is where they'll find users and liquidity. Fully autonomous, verifiable AI agents will have to rely on a mix of TEE, MPC, FHE solutions.

A couple of interesting developments to follow:

- Dark Pools, UniswapV4, CowSwap hooks – Renegade
- A range of different private DeFi uses, starting with private lending strategies and portfolio optimization – Stoffle, Fairblock
- Fully autonomous, verifiable AI agents such as TEE_HEE, an AI agent with its own twitter account. The scope of use-cases will rapidly expand – Nous Research, Flashbots, Teleport
- Biometrics and other personal data being ZK verified onchain to support

myriad of consumer facing applications – World, Reclaim

ZK's Exponential Era

2024 has been a big year for zk, and a number of developments emerged in all layers of the stack. Many of these are set to come into production and we'll explore how this convergence will affect different areas. We'll begin with zkVMs, which are democratizing ZK development by allowing any developer to build ZK applications without cryptographic expertise. Combined with rapidly falling costs, these zkVMs are positioned to challenge optimistic rollups' dominance, offering superior security guarantees and seamless interoperability.

We'll then see how zk is addressing fragmentation and enabling true unification across different rollups and creating a fluid user experience. We'll also see the impact of zk on Bitcoin, Ethereum and Solana. The rise of "ZK-first" L1s represents another shift – these are chains fundamentally take a different route, with zk verification as their core primitive, delivering horizontal scalability while maintaining validator accessibility and composability. Finally, we'll see how advances in hardware acceleration and proof systems are creating multiplicative performance gains across the ecosystem, from specialized chips and GPU optimization to revolutionary approaches with smaller field sizes.

> Zk feels like it's approaching an inflection point:
>
> 1. Transaction costs on ZK rollups are lower than ever
>
> 2. "Rollup Cluster" interop = more liquidity
>
> E.g., @zksync's Elastic Chain, @0xPolygon's AggLayer or @Optimism's Superchain
>
> 3. Rollups are prioritizing sustainably...
>
> — Ismael Hishon-Rezaizadeh ▮ (@Ismael_H_R) October 13, 2024

zkVMs

zkVMs have emerged as one of the biggest shifts and enablers of zk in the past year. Previously, teams who wanted benefits of zk, typically the rollups, needed cryptographers. Then the core primitive here was a circuit for generating proofs. This severely limited the use of zk to only those teams. The circuits were handwritten and had a larger surface area of bugs. The optimization was ad hoc.

zkVMs have changed the game.

A zkVM defines the instruction set and other things like memory, registers, etc. to run computations and generate proofs for it. It first runs the computation and records the changes in memory after applying the instructions. A proof of its correct execution is then generated. Continuations was a big breakthrough in zkVMs which is a standard practice now. Before this, the memory requirements for proof generation of entire computation made proof generation difficult for even a single transaction. But in continuations, large computation is divided into smaller parts and proofs for each is generated in parallel and recursively composed. This has made proof generation for entire Ethereum blocks possible.

The main unlock of zkVMs is how it expands the market for developers. Unlike before, they no longer have to learn a new language to develop zk applications. Now anyone who knows Rust or other familiar high level language can get proofs generated using a zkVM. This **opens up a huge developer base that can use zk for scaling or privacy. Transforming applications or programs to use zk capabilities has become significantly easier.**

We're seeing precisely this effect on the application space. All sorts of applications other than rollups are getting converted to its zk version – Bridges, Light Clients, OP Rollups, Ethereum Block proving, Individual Smart Contract acceleration, TEE attestations, etc. Applications can now focus on the business logic, while still getting the benefits of zk. A growing number of applications also means reductions in cost and latency, because the larger the proof demand, the better the costs and latency due to amortization across all the volume.

One way to categorize zkVMs is using the Instruction set that it proves

- **Hardware based ISA** like RiscV, MIPS, WASM (RiscZero, Succinct, zkWASM, zkm)
- **Custom ISA** (Valida, Binius) – These create new instruction set which is zk

friendly

- **Custom ISA + Custom Language** (Cairo)

Currently RiscV based zkVMs are leading in the market, but there are a number of others entering the market next year. We've already seen a **10x improvement in cost** in a year since their launch. The current costs for proving single transactions using zkVMs is around 0.1 cent. Next year we'll see another 10x improvement in cost or even more. This is a **combined effect of improvements through proof systems, hardware acceleration and zkVM optimizations.** There's a Moore's law like progression in zk proof efficiency in action here.

ISA choice has a big impact on zkVM costs and efficiency, especially in the long run. When looking at proof costs in zkVMs, we need to consider both how many execution cycles are generated and the cost per cycle. While RISC-V based zkVMs all generate the same number of cycles, Cairo's custom ISA helps Kakarot zkEVM generate about 5 times fewer cycles when proving Ethereum blocks. Combined with Cairo's optimization for proof generation, this makes Kakarot roughly 10 times cheaper. This difference is particularly significant because it matches the current difference between an OP rollup transaction and zkVM based OP rollup. This means that when it goes live, the OP rollups have a simple choice – become zk at the same cost with its benefits – a no brainer.

Performance for a zkVM can be considered across different dimensions: proof generation cost, proof size, verifier cost, verifier size, proof generation latency. Out of these, proof generation cost is the big one because that's the main hurdle for adoption from the customers. The reason for this is the demand upstream: If OP rollups want to shift to zk, they need the costs equal to or lower than their current costs. Users are generally willing to wait a little compared to paying higher costs.

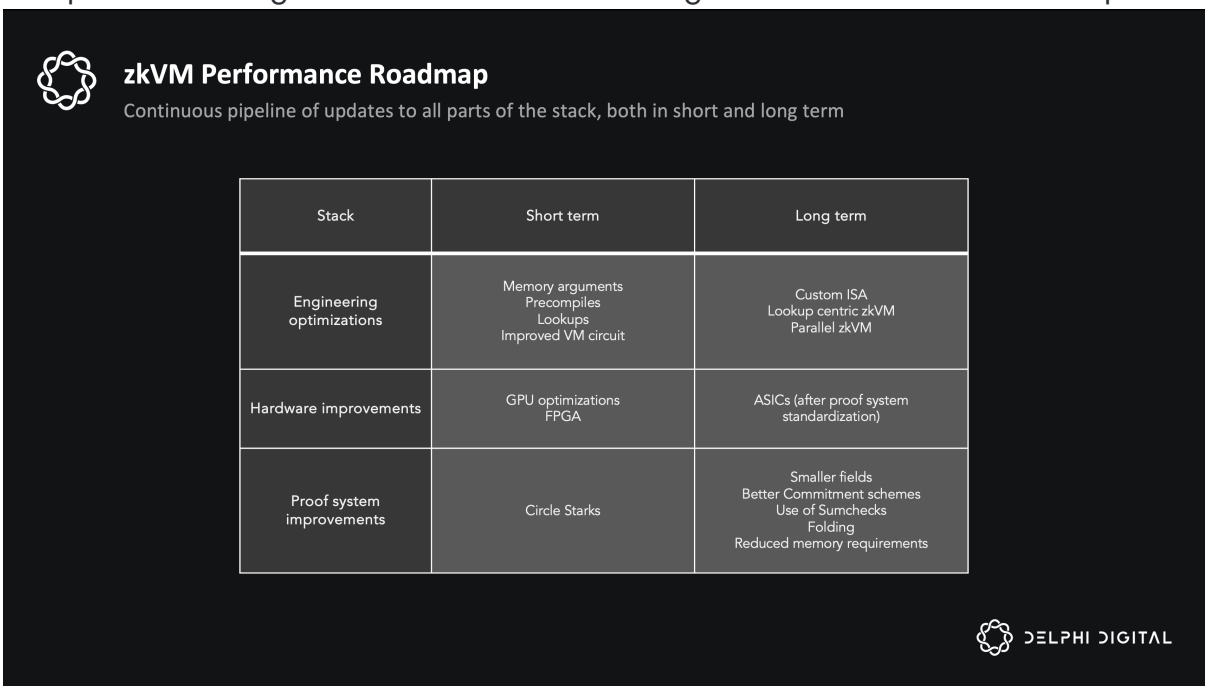
Current zkVMs are optimizing these by tweaking internal components. In zkVMs, we have a main CPU circuit and lookup circuits to carry out each kind of operation instead of one giant circuit. Between the parts of the computation, memory consistency has to be checked as well.

zkVM proving costs are reduced by

- Having an Improved VM circuit
- Using efficient proof systems

- Using smaller fields (operations are more efficient)
- Efficient use of Lookups (reduced columns and circuit size)
- Improved Memory arguments
- Use of accelerator circuits/ precompiles
- Optimized GPU acceleration
- Use of folding
- Engineering and implementation improvements (like implementing it in C instead of Rust)

The performance gains for zkVMs over the long term can be divided in two parts



In the short term (next year), zkVMs will be focusing on use of accelerator circuits, lookups, better memory argument, and GPU implementations rather than incorporating new proof systems, fields and folding, etc. This is because the engineering focused performance gains are easier and less risky than incorporating new proof systems, fields, using ASICs, etc. This is especially true because these things are still moving fast. These will contribute more in the longer term, once there is standardization and stability. This means that we have a long and continuous pipeline of improvements coming to the zkVMs – through zkVM engineering improvements as well as underlying stack improvements.

Precompiles

Another component which is considered to have a big impact on performance is Precompiles.

SP1 has a number of these precompiles live

- Secp256r1 precompile
- Poseidon2 BabyBear precompile
- BLS12-381 precompile
- Keccak precompile

But what are these precompiles? These are essentially specialized circuits for special computations that plug into the zkVM circuit. These circuits are run natively on hardware and have better efficiency. So instead of generating proofs for these computations through a VM and its instructions, precompiles do this without its overhead. Accelerators for common operations like keccak hashing, signature verification and curve operations will boost performance next year. These operations are intensive and precompiles will contribute to 6-10x improvement in these.

RiscZero also has accelerators for SHA256, ECDSA signature verification and Big integer arithmetic and are working on “Programmable Precompiles”, delivering 200x improvements in algebraic operations.

Execution and Witness generation before actual proving is another large bottleneck. Unlike proving, which can be parallelized using continuations, execution is a sequential process. Fabric’s VPU has made improvements on this part, which we’ll see in the hardware acceleration section later.

zkVMs are also not running on efficient GPU implementations, and there is a lot of improvement left on the table which will be bridged next year.

In the longer term, other zkVMs are going to enter the market

- **JOLT / LASSO** – zkVM based on lookups only and combining Binius
- **Irreducible / Polygon** – zkVM based on Binius proof system, efficient Polynomial Commitments, vertically integrated with their ASICs
- **Valida** – Custom ISA based zkVM
- **Polygon RiscV ZIVM**

- **Ceno** – zkVM by Scroll using proofs of parallel execution
- **Kakarot zkEVM** – EVM prover for all networks using CairoVM and Stwo prover

The zkVM market is primarily competing on cost and there is no direct moat. Price performance is going to be key, and making certain design decisions from the start might give approaches an advantage in the longer term. For instance, Custom ISAs definitely have edge in terms of performance and would win in longer term, but they do have challenges with compiler and tooling. There are few projects taking this route – Valida, Cairo etc and compared to RiscV zkVMs their performance is much better as we can see from the benchmarks.

Precompiles, though a special feature will become a standard industry practice and improve performance of all the zkVMs. As we slowly move towards Snarkification of Ethereum execution and consensus layer, it would be preferred to have different implementations of zkVMs similar to how we have client diversity in execution and consensus client.

Platform strategy

RiscV zkVMs are getting crowded, commoditized and there is a need for differentiation of switching costs. Similar to client diversity on Ethereum, we'd need diversity in zkVM implementations and being one of RiscV zkVMs would be tough. RiscZero is building a platform around its zkVM to build switching costs for the users. This is a good move.

> I believe RISC-V zkVMs will be the fastest technology to get commoditized. There won't be almost performance differences between the each implementation. One to two years max for this to happen.
>
> — Fede's intern ▮ (@fede_intern) October 24, 2024

RiscZero's platform is called Boundless and integrates a number of their offerings like –

- A Decentralized Prover market
- Steel – acceleration of individual contracts on Ethereum or any other chain

- zkVMs for EVM, SVM and other altVMs
- Light clients like Blobsteam for Celestia
- Proof Aggregator
- Settlement contracts

Network effects emerge from aggregating proving demand across diverse zkVM applications that they have access to. As more developers build on RiscZero's zkVM, this creates a larger pool of proving needs that can be routed to their decentralized proof market. Higher proving demand enables better capacity utilization and cost amortization, leading to lower prices. These improved economics attract more developers, creating a feedback loop.

The proof aggregation layer strengthens this effect – by combining proofs across applications, it reduces costs for all participants. The more applications in the ecosystem, the more opportunities for aggregation.

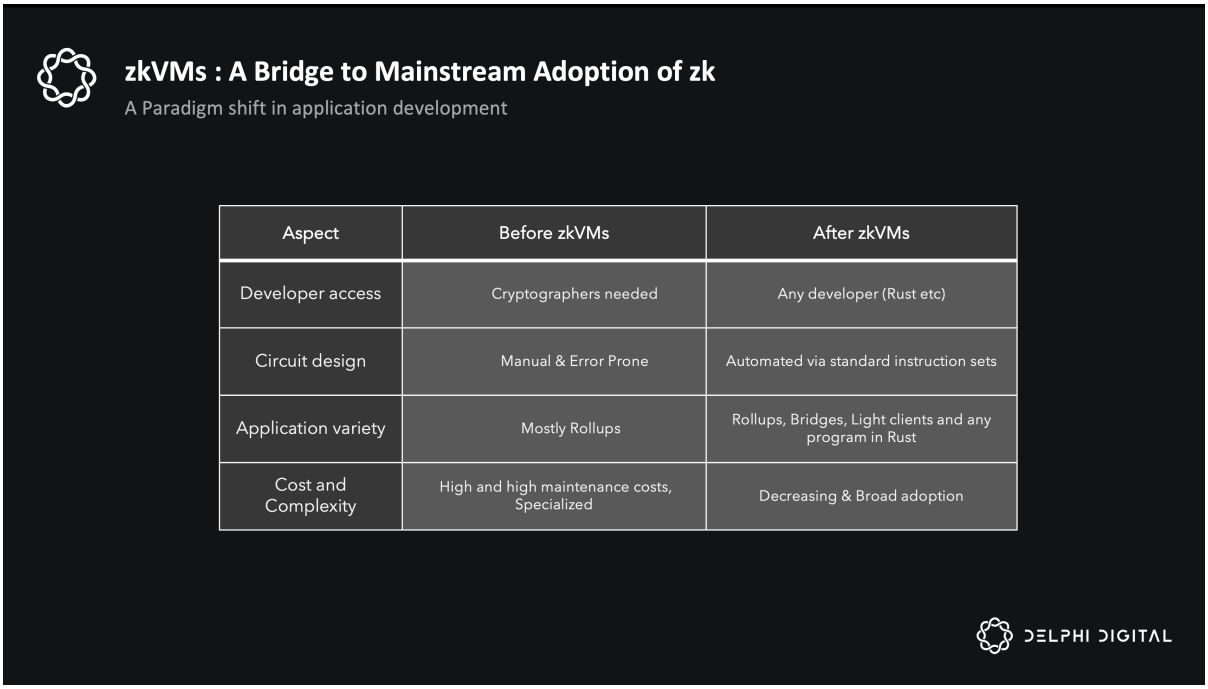
Rather than developers piecing together and managing multiple systems, they get a simplified interface to the full stack. This reduces development complexity while preserving flexibility in the unified platform, giving it a switching cost benefit. We like this strategy of building moats when proving costs could commoditize the market.

The pricing wars in zkVMs will ultimately benefit rollups, applications and finally take zk to mainstream next year.

Impact of zkVMs

zkVMs as an abstraction layer are leading to a number of changes in the way in which applications are developed, their limits and solving hardware acceleration of zk proofs.

Developer Access and Convergence of Software Paradigms



While zkVMs initially appear to drive accessibility through a feedback loop of mainstream language, better tooling and more developers, their true transformation runs deeper. They're creating a bridge between two previously isolated domains: the rich heritage of software engineering patterns and the privacy and scaling capacity of zk proofs. Circuit-based development made it impossible to apply decades of evolved software practices to ZK applications, while keeping ZK's powerful verification and privacy primitives locked away from traditional software. zkVMs dissolve this barrier, enabling a two-way exchange where software engineering wisdom flows into ZK development, while ZK's core principles reshape how we build conventional applications. This isn't merely about accessibility – it's about a new unified discipline that harnesses the combined power of both domains.

> The acceleration in ZK is real and measurable.
>
> I put together some data with @jtguibas on how much SP1 has sped up development timelines for production-grade codebases. It's pretty staggering! pic.twitter.com/oeQp4MSMoG
>
> — K Kulkarni (@ks_kulk) October 22, 2024

Implementation Compression

The shift from manual circuit design to automated compilation through zkVMs

represents a fundamental change in how ZK applications are built. This helps with composability of components because they're no longer tied down to circuits. Developers can focus on providing these legos, and others on assembling them for more complex applications without worrying about circuits or cryptography. A developer building a privacy-preserving voting system can use standard components for authentication, vote counting, and verification, focusing on the unique aspects of their application rather than reimplementing basic cryptographic primitives

Optimization Compression

Each different ZK circuit (whether it's for EVM execution, ML inference, or privacy-preserving computations) has its own unique pattern of computation. This creates a dilemma: Do you build specialized hardware for specific types of circuits (limiting your market) or try to build flexible hardware that can handle any circuit (sacrificing efficiency). zkVMs transform the problem. Instead of arbitrary circuits, we can design hardware that accelerates one type of computation – zkVM instruction set. This provides a fixed target for acceleration that stays stable over years, and optimization becomes predictable and simpler. This shifts the problem from "how do we accelerate unknown computation patterns?" to "how do we best execute this specific instruction set?"

This creates increasing returns to scale: As more developers use zkVMs, the tools get better, more components become available, and hardware optimization becomes more effective, making it even more attractive for new developers to enter the ecosystem. This positive feedback loop is similar to what happened with cloud computing – each layer of abstraction and standardization enabled new use cases and attracted more developers.

Applications Leveraging zkVMs

Transitioning OP Rollups to ZK Rollups

The shift from OP rollups to ZK rollups is one of the biggest drivers for zkVM adoption. The current OP rollup architecture has clear drawbacks – complex fraud proofs, interoperability that depends on bonds, and long withdrawal windows. These limitations make a strong case for moving to ZK technology.

Right now, the only real barrier stopping OP rollups from adopting zkVMs is the cost factor. Running transactions through zkVM-based systems is about ten times more expensive than traditional OP rollups. But here's what's changing: this

cost gap is exactly going to be improved next year – since we're expecting 10x improvement in zkVM proving cost next year. Once zkVM-based solutions reach price parity with OP rollups, the choice to switch becomes straightforward.

It always expected that OP rollups will eventually become ZK-based. With the economics finally aligning, this transition is set to kick off next year.

There are other benefits of zkVM based rollups compared to circuit based. zk rollups were maintained by cryptography heavy teams which wrote the circuits. Compared to that, zkVM based rollups don't have the problems of updatability and maintenance. The attack surface for bugs (etc) reduces compared to having circuits. In circuit based zkEVMs Keccak hashing, signature verification proofs are quite hard. Using precompiles in zkVMs makes this 6-10x more efficient.

In terms of migration, zkVM allows OP rollups to shift to being full zk rollups in less than 2k lines of code. It also takes almost the same on chain verification cost. The difference is the 2-3x cost of Ethereum transaction for generation of proofs – out of which, 80% of the cost is in execution. Compared to OP rollups, the withdrawal window will keep decreasing with decrease in proof generation times. Again here, the cost prohibits the frequent batching of proofs and their settlement, but this is on a downward trajectory. **This means the time to withdrawal will trend to zero – aggressively starting next year a step closer to seamless experience of a single chain**.

Though projects are experimenting with hybrid proofs, marginal improvement in terms of withdrawal window (because the window needs to be wider to allow dispute and zk proofs are a small fraction of actual window) and falling costs mean that full zk rollups would make more sense. Another difference would be in interop between the rollups. Though interop between OP stack rollups is one of the perks, they need nodes to be run for each rollup. zk stack rollups will have zk light clients, Agglayer like solutions which would be faster, less expensive and trustless.

But as zkVM makes zk rollups more easy to develop, just a good implementation would stop to be a differentiator. Added functionalities like privacy, client side proving, ecosystem support etc would matter more moving ahead.

We'll see the majority of rollups converted into zk by the end of next year, and this will be facilitated by the cost parity of running a zk rollup as cheap as an optimistic rollup.

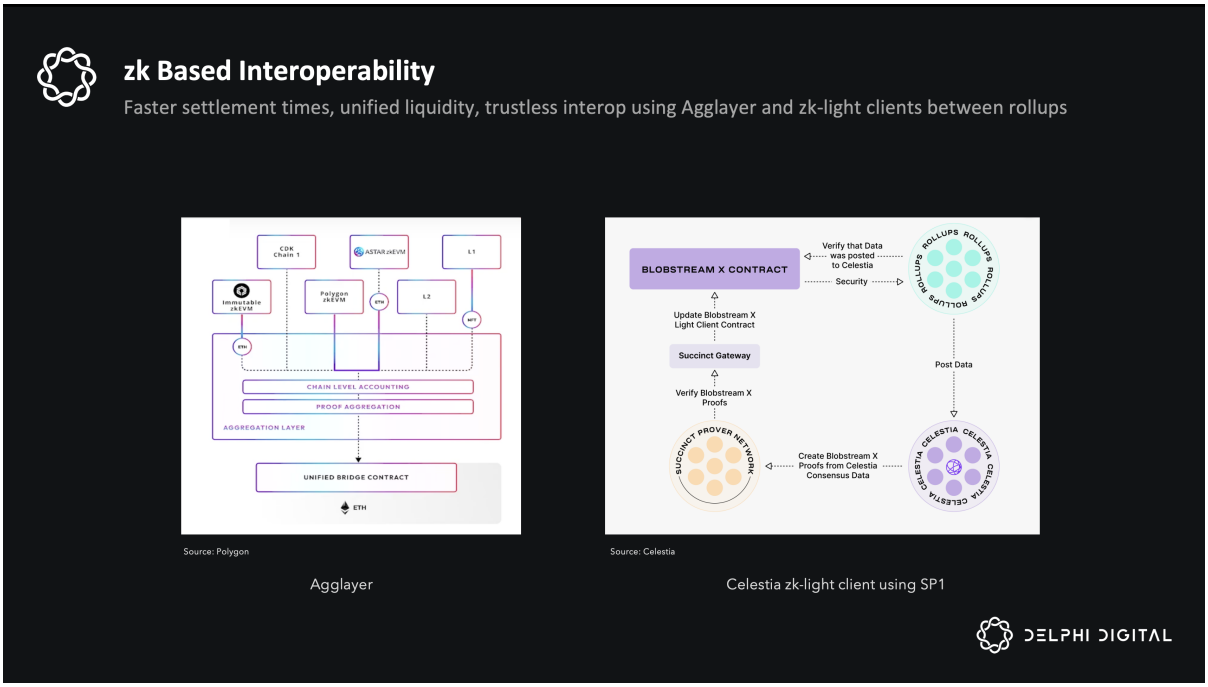
Interoperability

The rollup centric roadmap has created fragmentation and a single chain experience is needed asap. To make matters worse, each of the rollup clusters are building their own interoperability standards. Agglayer aims to be neutral and will lead to improved user experience. The first component of the AggLayer, specifically the unified bridge, is live.

> interesting bridging article commissioned by @Optimism, done by @norswap
>
> Conclusion: ZKPs enable trustless bridging and are poised to become the superior form of bridging in the medium-term, as proving time and cost further decrease.
>
> — elias tazartes ■■■ (@ETazou) September 25, 2024

Because each of the rollups has its own bridge, there is a lack of fungibility between tokens. This leads to capital inefficiency and bad user experience. Agglayer has a shared bridge, solving for this. Having a single shared bridge also introduces points of vulnerabilities which is handled using Pessimistic proofs and accounting checks. One of the main advantages of Agglayer is that it would allow rollups and chains to interact faster than Ethereum finality – so essentially being unaffected by the slow speed of base layer finality. It uses rollup state commitments and their enforcement in settlement via zk proofs for this. So, the unified liquidity, faster async and sync interop will bring a single chain experience. All of this will be accelerated using zkVMs and hardware acceleration next year. Fabric is working on releasing their VPUs next year, which will accelerate the Agglayer proof generation. More on Fabric and VPUs in later sections.

So we get this nice safety layer that solves the liquidity, sync and async interoperability problem between rollups via cryptography and it gets acceleration via hardware – both, next year.



Agglayer also works with Espresso, allowing the block builder to build rollup blocks simultaneously, and provide safety using its pessimistic proofs for safe preconfs. Behind the scenes, Agglayer has merkle trees to track withdrawals and balances for each rollup, Ethereum mainchain and a combined merkle tree for them all. They are called **LER**, **MER** and **GER**. The GER is passed on to all the rollups to check, verified against them, and aggregated to reduce costs. When rollups interact they create commitments to their state and while settled the verifier checks if this commitment is followed. So the rollup finality is dependent on them following their commitments.

So, all in all, Agglayer is accelerated on multiple levels:

- zkVM optimizations
- GPU proving and VPU proving
- Commitments and proofs for even faster guarantees than Ethereum L1 (rollups can start interop before settlement)

Another area where interop gets a boost using zkVMs is light clients between L2s. There are a number of implementations of zk light clients using zkVMs:

- Celestia Blobstream for DA (SP1)
- Celestia zkIBC, proofs for the zk rollups
- Gnosis light client (SP1)

- SP1-Helios (which is a mutli-chain light client focused on L2s) and a few others

Helios started out as a light client 2 years back and now has extended its support to OP stack chains and other rollups in its roadmap. It's a multi-chain light client to make interop simpler.

Currently, Helios runs on sync committee, like most of the light clients, introducing trust assumptions and lower security (sync committee has 512 validators only that are rotated). Helios' zkVM implementation allows rollups to just verify zk proofs of a light client running inside a zkVM, reducing the gas costs to run them as smart contracts. Polyhedra has done some interesting work on accelerating sumcheck proofs using GPUs. They're also working on proofs for the entire validator set instead of just the sync committee. This would turn it into a fully trustless light client. Hopefully these developments could be integrated with Helios for trustless bridging between the L1 and L2s. Again here, improvements in zkVMs, proof systems and hardware acceleration plays a major role to make this feasible and performant.

Espresso is live, and Agglayer's first component is live and next year, this combination will improve the user experience. This will be followed by reduced costs by aggregating proofs, zkVM optimizations and hardware acceleration using VPU. **The combination of Espresso, Agglayer and light clients will be the first steps towards a single chain experience next year.**

ZK by Ecosystem

Ethereum

There are longer term plans for snarkifying the execution and consensus layers. Snarkification of execution means that builders create blocks inside zkVMs and proofs of blocks are verified instead of re-running all the transactions. This reduces the load on validators as lighter machines can verify blocks and can have higher gas limits.

Snarkification of consensus layer focuses on improving the efficiency of signature aggregation. The current system requires collecting signatures from all validators before submission, creating a bottleneck. The proposed snarkified

approach would enable validators to sign with just their local peer group quickly, with these partial signature sets being aggregated through services. This would make the system more fault-tolerant and enable faster finality, potentially achieving it with just 60-70% participation.

These are longer term goals, still in specification phase. Main bottlenecks for these would be to generate proofs for blocks. Current latency is not much of a bottleneck compared to the cost. It is possible to have a large cluster of GPUs and get faster proofs, but the challenge is to make this economically viable – meaning faster proof generation on consumer hardware.

Kakarot zkEVM will be generating proofs for L2s and EVM chains and is working on generating sub 4 second proofs for entire blocks at a cost of a few cents per block by the end of 2025. It will use CairoVM, Stwo (new prover from Starkware), Circle Starks and other improvements. The key insight is that optimizing both cost and latency requires addressing the fundamental efficiency of the proving system itself, rather than just throwing more hardware at the problem. More on this in proof systems and hardware acceleration sections.

Hardware acceleration and proof system improvements will play a big role in reducing the validator load in coming year.

Bitcoin

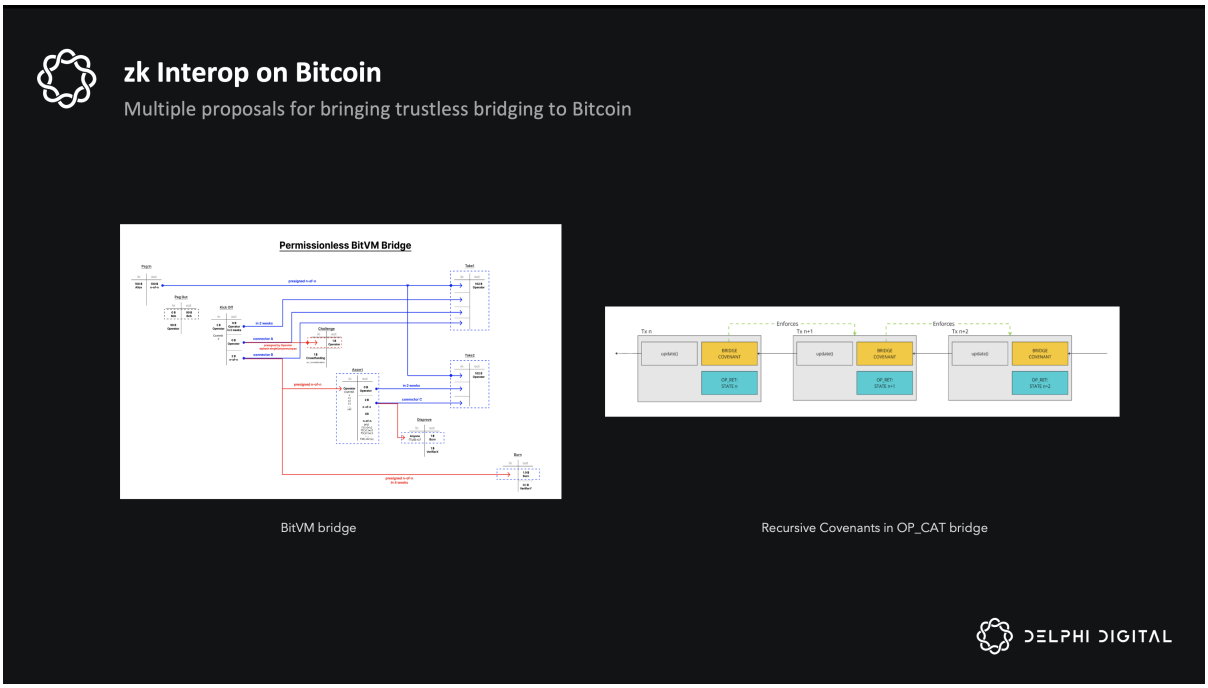
Bitcoin's evolution in 2024 marks a pivotal moment. While its fundamental value as a decentralized, censorship-resistant store of value has driven widespread adoption as a reserve asset, zk will dramatically expand its capabilities. The convergence of zkVMs, advanced bridging solutions, and protocol improvements suggests 2025 will be a transformative year for Bitcoin's technical capabilities and its role in DeFi.

Bitcoin L2s are evolving beyond simple payment channels to full zk-Rollups, offering orders of magnitude higher throughput while inheriting Bitcoin's security. This shift is enabled by major breakthroughs in proof systems:

- Circle Starks achieved 125x faster proving speeds on consumer hardware. This will be used by Starkware for their Bitcoin Rollup.
- Smaller field sizes, reducing computational overhead – Binius, M31.
- Development of BitVM, BitVM2, BitVMX and OP_CAT frameworks for verification of zk snarks on Bitcoin

- Hardware acceleration leading to 10x efficiency gains in 2025.

Bridging Innovations on Bitcoin



The critical challenge of trustless Bitcoin bridging saw three distinct approaches emerge in 2024, each pushing boundaries in different ways:

BitVM2 and Strata BitVM2 was proposed this year, and we have several implementations. Its used in bridging using validity proofs and optimization of Bitcoin's existing script capabilities. Alpen Labs' Strata implementation shows how this enables trustless 1:1 Bitcoin bridging without protocol changes. It combines validity proofs, emulated covenants, and economic incentives through operator collateral and challenge mechanisms. With active development and optimization underway, production-ready implementations are expected in early 2025, changing how assets move between Bitcoin, L2s and other chains.

OP_CAT Bridge Evolution The proposed OP_CAT opcode addition has shown potential using StarkWare's demonstration of STARK verification on Bitcoin's Signet. This allows direct covenant implementation and efficient transaction batching, offering a cleaner path to trustless bridging. While its activation would depend on Bitcoin community consensus, growing support suggests possible inclusion in 2025. The key advantage lies in its potential for near-instant bridging and settlement for rollups, but careful consideration of security implications remains crucial.

ColliderScript Exploration ColliderScript, developed by Blockstream researchers, proves that covenant-like functionality is possible on Bitcoin today without protocol modifications. Using sophisticated cryptographic techniques involving hash collisions, it shows untapped potential within Bitcoin's existing script capabilities. While current implementation costs (estimated at $50M per transaction) make it impractical for immediate use, this theoretical breakthrough is driving important research into optimization techniques. The shows how innovation can emerge from working within Bitcoin's constraints than changing them.

The parallel development of these approaches suggests 2025 will be a pivotal year for Bitcoin bridging. Each solution offers distinct advantages: BitVM2 provides immediate practicality, OP_CAT offers protocol-level improvements, and ColliderScript demonstrates theoretical possibilities. This diversity of approaches strengthens the ecosystem.

Bitcoin's Next Chapter

These solutions are coming together for something powerful. zkVMs based on RISC-V or CairoVM provide the computational engine for L2s. They work with the bridging solutions to ensure that Bitcoin can be securely transferred and utilized within these more expressive environments. New Proof systems enable efficient verification and hardware acceleration makes it performant.

This enables expressive applications while preserving Bitcoin's core properties. Users can now access decentralized lending and borrowing protocols, along with Bitcoin-native DeFi systems that eliminate the need for intermediaries. The Layer 2 solutions bring advanced smart contract capabilities to Bitcoin, while also enabling privacy-preserving financial applications that maintain the network's fundamental security and trustlessness.

The next year will see

- Multiple production-ready bridging solutions.
- Significant reduction in proof generation costs.
- Emergence of zkVM-based Bitcoin applications. Projects like Lava, which are building DeFi applications on Bitcoin L2s, will benefit from these advancements in bridging and zkVM technology.
- Potential protocol improvements enhancing these capabilities.

- Increased demand for Bitcoin, not just as a store of value, but as a productive asset within L2 ecosystem.

This suggests Bitcoin is entering a new phase where it not only serves as a store of value, but as a foundation for DeFi. The parallel development of multiple approaches provides redundancy and optionality, increasing the likelihood of successful implementation while maintaining Bitcoin's security-first ethos.

These developments don't just add features – they fundamentally expand what's possible with Bitcoin while preserving its core properties of decentralization and security.

Solana

zk helps Solana with its own, different challenge: managing explosive state growth while maintaining its high performance. With over 500 million accounts growing by ~1 million daily, Solana is taking a different approach to zk – not primarily for scaling throughput, but for state management and cost optimization.

> the big news is here
>
> today we're introducing ZK compression to Solana, directly on the L1 — without requiring L2s
>
> this changes everything you thought you knew about Solana and scaling L1s
>
> TL;DR — we compress onchain state to get 10,000x scale improvements and get 1 step closer... pic.twitter.com/7FtyLA3Jdp
>
> — mert | helius.dev (@0xMert_) June 21, 2024

This rapid growth has created significant challenges for state management. It now requires snapshot sizes of more than 70GB, which impacts node synchronization. Validators need more than 32GB of RAM only for account indexing, resulting in escalating costs for both validators and application developers.

These things point to need for managing the load as well as cost for both

validators as well as applications.

To address these challenges, Solana has implemented zk compression for state management. This approach differs from other chains that use zk primarily for throughput scaling. The compression achieves a 5000x reduction in state storage costs through several key innovations. The state is stored off-chain in Solana's ledger space while maintaining on-chain Merkle roots for verification. It is able to get 128-byte proof sizes through compression while preserving parallel transaction processing when states don't overlap. Importantly, it maintains full atomic composability, allowing transactions with both compressed and regular accounts to be combined atomically.

Solana has improved its infrastructure by adding zk syscalls this year for on-chain proof verification. These include Poseidon hash functions for efficient ZK operations, alt_bn128 syscalls for proof verification, and enhanced runtime support for elliptic curve operations. This infrastructure upgrade provides a foundation for ZK-based applications to manage state, enhance privacy, and enable scaling through network extensions.

This infrastructure upgrade creates a foundation for ZK-based applications, to manage state, add privacy and scaling via Network extensions.

The ecosystem is using these solutions in different ways. Wallet providers like Backpack are providing rent-free token storage, while Dark Protocol is developing privacy-preserving applications. Social applications such as Tribe are using compression for token distribution, and Helius is developing tooling and indexing solutions.

Apart from that, we're seeing zkSVMs being developed by running an SVM instance inside a zkVM (RiscZero and SP1) and allowing it for "network extensions" to quickly ship to production and settling via validity proofs.

These implementations show that zk is getting adopted in different ways on Solana, solving for different problems apart from scaling.

zk on Solana is new and looks promising. Next year, we would see

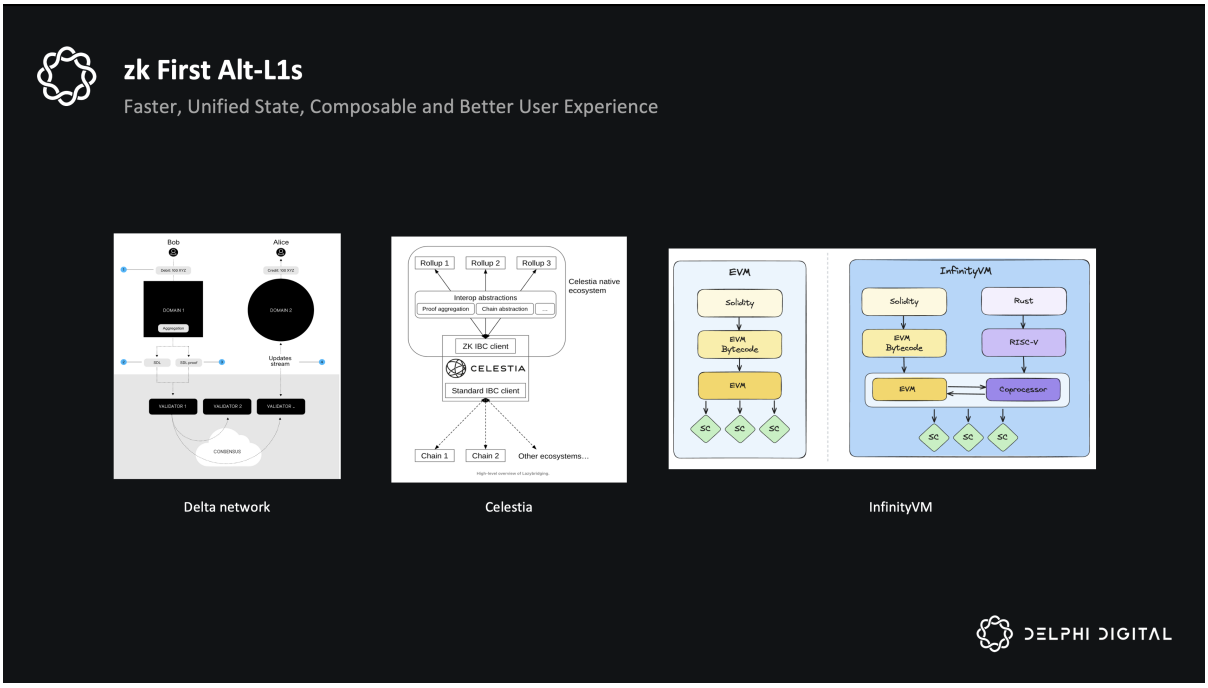
- Further optimization of state compression techniques, leading to further lower costs for applications, lower burden on validators
- Integration with cross-chain data availability solutions using things like Blobstream running in zkVMs like SP1

- Expansion of privacy-preserving applications
- Improved validator economics through compressed state

zk First L1s

Maturation of zk and its performance has led to a new generation of L1s which are fundamentally taking a different route, based on zk as a first class citizen.

All the L1s till now have been designed with re-execution of transactions as a core primitive. But with improved zk performance, easy development using zkVMs, we're seeing new L1s using verification of zk proofs as a core primitive.



These new L1s generally have shared accounts/state or completely separate state needing minimal co-ordination between them. The key insight here is, because the state is separate, it can manage its own ordering. This frees up the base layer to experiment with consensus and not rely on execution at all. These new consensus systems are much lighter, highly parallelizable and horizontally scalable. These consensus mechanisms don't manage total ordering like in other large chains, and instead only have partial ordering. They have parallel block building (different validators creating blocks at same time) and dissemination of blocks for parallel consensus on them.

The direct effects of this on the L1 are:

- No or minimal need for Execution on Base chain

- No or minimal need for Total Ordering on Base chain
- Local ordering by Rollups and value capture
- Fast & Light Consensus
- Horizontally scalable Execution
- Horizontally scalable Consensus
- New state models to account for unified but not totally ordered transactions

This translates into reduction in the amount of work a validator does on the base layer, but at the same time leading to better composability, better local control, and high performance. On Ethereum there is a split on whether to make the L1 faster or keeping the validators decentralized. In these cases, **they're able to do both**. The validator requirements remain light because it only needs to verify proofs and has a light and fast consensus. The L1 is fast because the reduced consensus is parallelizable and scalable, because weaker machines don't have to re-execute transactions. It just needs a quorum, which is very fast. **So the validator requirements is low, at the same time base layer performance and overall throughput is high.** In these designs, execution using rollups doesn't have fragmentation, and the network effects are shared with the rollups.

> delta's design is exceedingly simple. To get there, you can start with a rollup stack and just delete stuff:
>
> – Remove the need for bridges and 3rd party interoperability services by having all the "rollups" (domains) share state
>
> – Integrate the execution layer with the base… https://t.co/aEemPt66F8
>
> — Ole Hylland Spjeldnæs (@spjoleh) October 9, 2024

We cannot go into each of these designs in depth, but they are more or less similar:

- **Delta** – Shared assets, State and liquidity on the base layer with local execution, light consensus and zk verification

- **Celestia** – zkAccounts, lazybridging
- **Hyle** – Pipelined proofs, State commitments only onchain, light consensus
- **InfinityVM** – Enshrined zk coprocessors with all reorging or none
- **Pod** – light consensus, zk verification of rollups

Celestia has adopted zk for enabling network effects and interop between its rollups. Celestia doesn't want to add execution capabilities on base layer to reduce the state bloat and maximize its DA capacity. Instead, its introducing a zkIBC light client and zk accounts on the base layer. zk accounts corresponding to any program verifies if the program execution is valid and moves funds based on it. This is extended to the state transition functions of Rollup. These are run inside SP1 zkVM to generate proofs and get them verified using the zk IBC light client and zk account. This allows trustless bridging with fast finality (because Celestia has Single Slot finality). Here, the interop is bottlenecked only by proof generation times. And again, zkVMs and all related optimizations in zk make this continually improving.

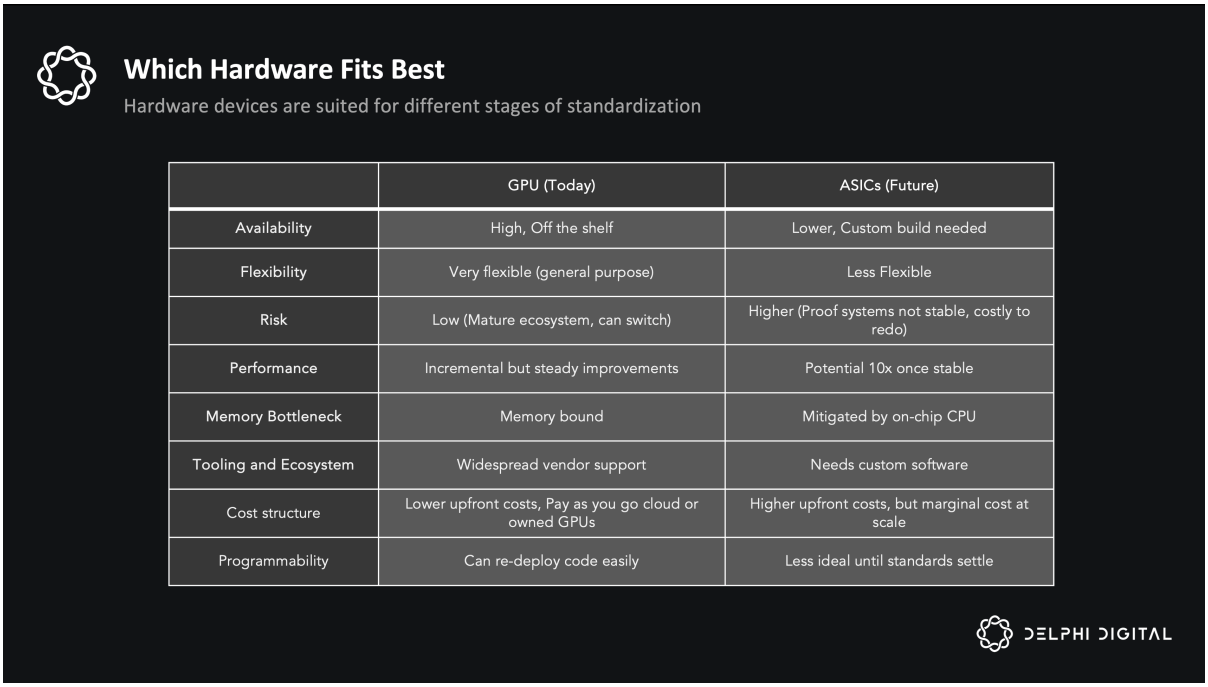
So looking ahead, we will see L1 approaches that have these benefits, challenging the tradeoff space for L1s

- They have low validator requirements
- They have fast base layer performance
- They allow network effects through sharing assets/ state and fast or no interop between them
- This is unlocked using zk, zkVMs and acceleration of proof generation

Further Advancements in ZK

Hardware acceleration

Because of Continuations, proof generation is parallelizable, but it is slow on CPUs. Hardware acceleration is crucial for zk proof generation, and understanding the technical constraints and market dynamics helps explain why certain solutions are gaining traction.



	GPU (Today)	ASICs (Future)
Availability	High, Off the shelf	Lower, Custom build needed
Flexibility	Very flexible (general purpose)	Less Flexible
Risk	Low (Mature ecosystem, can switch)	Higher (Proof systems not stable, costly to redo)
Performance	Incremental but steady improvements	Potential 10x once stable
Memory Bottleneck	Memory bound	Mitigated by on-chip CPU
Tooling and Ecosystem	Widespread vendor support	Needs custom software
Cost structure	Lower upfront costs, Pay as you go cloud or owned GPUs	Higher upfront costs, but marginal cost at scale
Programmability	Can re-deploy code easily	Less ideal until standards settle

The generation of zkp relies heavily on operations like MSMs (large integer arithmetic), NTTs (polynomial evaluations and FFTs) and hashing. CPUs are not particularly optimized for these operations. Modern GPUs and ASICs provide impressive raw computing power to handle these calculations quickly. However, they often hit a significant roadblock: memory bandwidth. This becomes particularly problematic for memory-bound operations that require constant shuffling of data between different memory levels. While pure computational tasks that minimize data movement can run at blazing speeds, memory operations are different. When a GPU or ASIC needs to access memory—especially global memory—it can slow down dramatically, sometimes running 50 times slower or even worse compared to pure computation. This speed difference creates a major bottleneck in generating zk-proofs, and finding optimization approaches that can strike the right balance between computational power and memory usage are important. It is also important to share that these bottlenecks are shared between GPUs and ASICs, unless they have some onbound CPU and high memory on chip.

zk proof acceleration is divided into few approaches:

- **GPUs** – Ingonyama, Snarkify
- **FPGAs**
- **ASICs** – Cysic, Ingonyama, Irreducible, Acceal, Fabric
- **Apple Silicon / Consumer devices** – Ingonyama

Continuation, which breaks large computations into smaller parts with intermediate checks, tells us that proof generation is highly parallelizable—making GPUs a perfect fit.

There is a general consensus that GPUs are going to win out in the short to medium term because of a number of factors

- They are cheaper compared to development of custom ASIC which needs a large demand to justify the fixed costs
- Widely available tooling and ecosystem
- Highly Performant and hard to compete with
- Far from being saturated in terms of performance
- Same bottlenecks for ASICs and GPUs – memory bound algorithms
- Harder memory design in ASICs
- Programmable and are flexible compared to ASICs

For these reasons, the zkVMs are prioritizing proof acceleration using GPUs. zkVMs have lot of demand sitting on top of it, and therefore are important to the downstream proof market. Infact the choices of proof systems, optimization for zkVMs are heavily guided by their ease of parallelization on GPUs. Both SP1 and RiscZero, as well as rollup stacks like Scroll, zkSync have a focus on GPU implementations.

But GPUs are a bit tricky, and having naive implementation keeps a lot of performance on the table. Deep hardware expertise is crucial for effective GPU acceleration. Most teams are converging with respect to algorithms and acceleration devices, proof systems and the key differentiator lies in optimization skill – specifically, the ability to implement solutions that fully understand and exploit hardware characteristics. This expertise in hardware-aware optimization can create performance gaps between otherwise similar implementations.

This is clear if we look at the ZPrize winners and their approaches. Snarkify brings Niall Emmart's background from Nvidia and years of GPU optimization experience. They were able to bring the proof generation time under a second. Both, the winning and the runner up implementations used the same approach, same algorithms. Yet there was a 50% difference in proof generation times for both – clearly indicating the role of experience and deep understanding. This

breakthrough came through implementation changes: completely rewriting in C instead of Rust, eliminating unnecessary columns, and leveraging large GPU memory to minimize data movement. These optimization techniques aren't just specific to their competition entry – they can be applied to proving systems in general, suggesting similar speedups are possible across the board. This demonstrates that we're far from reaching the limits of GPU optimizations.

Another factor that affects the cost of proof generation is the kind of GPUs that are used. Using cloud providers lead to higher costs, and having supply of GPUs leads to instant 10x cost difference. Snarkify has built a GPU capacity of 300, and generates proofs for Scroll, zkSync chains. But these are using larger primes, leading to higher costs. But, when shifted to smaller fields, would reap the benefits of lower costs. Similar observation confirms that building a GPU supply reduces costs by 4-10x.

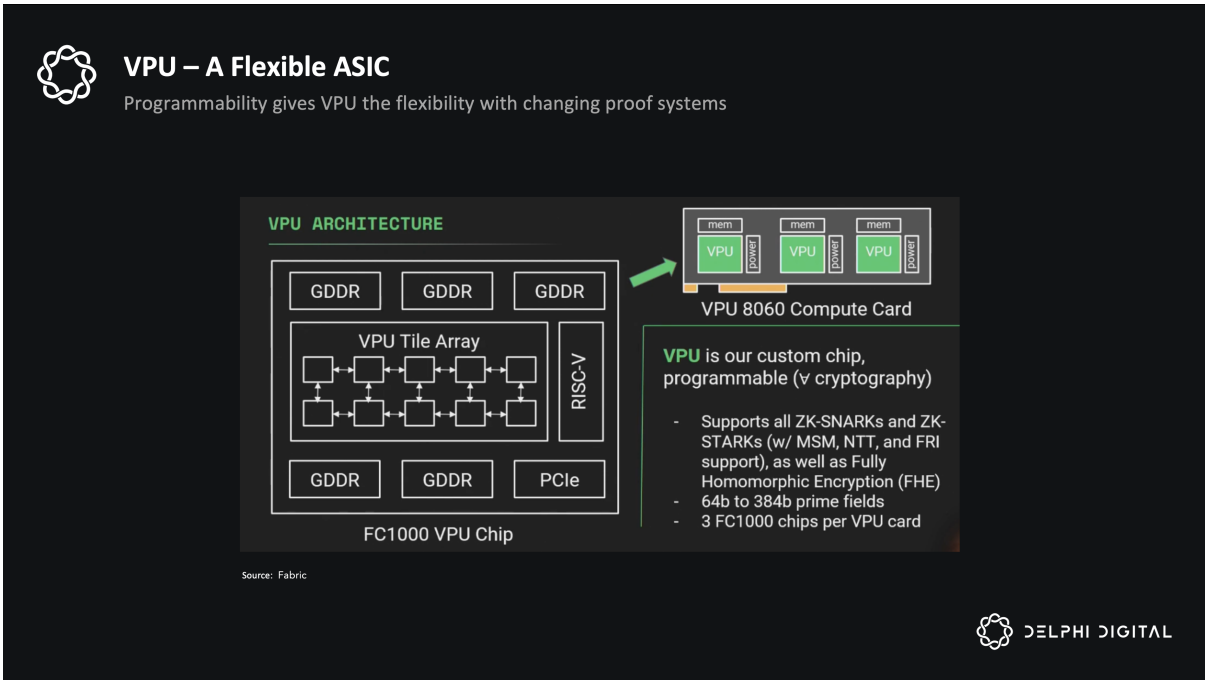
The key takeaway from this is that GPUs, with their underutilized potential and flexibility in programmability, emerge as the optimal choice. This is particularly true given that proof systems are still in flux and have yet to stabilize.

Now, looking at ASICs. Just like CPUs are not great for AI computations and GPU are something new, in the long term we would want custom accelerator chips. But currently it is too early with the proof systems not being fixed.

> Software-hardware co-design will unlock the next 10x for ZK performance.
>
> ZK proving is a two step process at the very high level:
> 1. Generate a large trace of witnesses (e.g. each risc-v instruction in RiscZero would equate to >1kB of witness data).
> 2. Heavy cryptographic... https://t.co/6mQxbaYlie
>
> — Wei Dai (@_weidai) September 27, 2024

In this context, Fabric has an interesting take where, the ASICs they will produce will not be just for accelerating zk, **but even for FHE and other programmable cryptographic computations**. The VPUs are programmable, have less hardened things. It is reusable for things like MSMs, NTTs, Poseiden hashing etc. Generally CPU and GPUs work with 32 or 64 bits and for working with large primes they have to use multiple registers, handle carry between them and manage pipelining –

leading to less efficiency and more cycles. The VPUs will have 384 bit width. By using 384 bit width each operation takes one cycle and is efficient. It is also compatible with smaller fields. So the same architecture handles variety of tasks like MSMs in larger fields, smaller fields and things like hashing efficiently. Its important to mention that **witness generation** is currently one of the largest bottlenecks, taking **more time than actual proof generation**. It is a sequential process needing back and forth between the CPU and GPU. And as we've seen before, data shuffling is much slower in GPU compared to actual computation. ASICs would face the same problems, since the memory technology is similar. To address this, the VPU has an on chip RiscV CPU for Witness generation. Having on chip CPU leads to 10x improvement in witness generation, which takes the processing capacity of VPUs to nearly 50Mhz (current state of the art is few Mhz) or more.



VPU also has lot of memory on chip, and high bandwidth. This helps a lot with the workloads, as these are memory bound. Fabric is also developing a Kernel Development Kit for writing low level Kernels to accelerate the primitives like NTTs, MSMs and a higher level Graph Development Kit to combine these Kernels on a higher level. So, they're working on the software support for their VPUs as well.

These chips are expected to be shipped next year, and design of V2 is already in works. This should provide 10x or more improvement in proof generation costs for all workloads.

So, for longer term, programmable cryptography (FHE and others) also gets benefits from acceleration using these VPUs.

We didn't cover other projects like Ingonyama and Cysic in depth, but they're also building great things. Ingonyama is building an unified API for accelerating proofs using on hardware. This includes CPUs, GPUs, ASICs accessed through their API. They're working on Apple Silicon support, **which would be great for client side proofs because of its unified memory**. They also have plans for their own ASIC based on a different ISA, but that is more in the long term.

Cysic is building 2 varieties of ASICs – ZK Air and ZK Pro. ZK Air is a lightweight ASIC for client-side proof generation, while ZK Pro is relatively high-performance ASIC for server-side proof generation. Recently Cysic's C1 chip was able to prove 1.3 Million keccak functions per second, which is about 100x faster compared to state of the art. This was using Sumchecks, GKR in collaboration with Polyhedra. **This points towards a promising future for client side proving in the coming year when the C1 chip is expected to be released.**

Apart from this Binius is also looking to build ASIC for accelerating their Binius proofs, which are highly anticipated and expected to improve the proving costs by large margin.

So on the hardware side,

- In the short term, GPUs dominate due to flexibility and market readiness. ASICs and VPUs may lead in the long term once proof standards stabilize, providing an efficiency leap. Competition and alliances may form around chip supply, engineering expertise, and integrated software toolchains.
- Cysic and Fabric are set to release their first batch of chips next year. So we should see an immediate bump in performance and on path towards incremental improvements going forward.
- Client side proofs would improve because of custom ASICs like Cysic's ZK Air, Apple Silicon. This would benefit privacy applications.
- Server side proofs would improve by 10x or more initially on GPUs and later on Fabric ASICs in next year via implementation improvements only. This is not accounting for the proof system improvements.
- Most of the algorithms for zkps are Memory Bound, and we'd see movement towards end to end proof generation on the same hardware.

Proof systems

Binius

Binius is a fundamental shift in SNARK design by addressing a core inefficiency in cryptographic protocols: the mismatch between computer hardware's natural data processing and cryptographic design. While modern computers work with binary data in fixed sizes (8-bit bytes, 32-bit words, 64-bit words), SNARKs typically operate over large prime fields requiring 256-bit arithmetic, creating significant performance gaps.

> Binius pairs exceptionally well with Jolt, and it's a top priority on our roadmap. The simplest, most performant zkVM will get much, much faster.
>
> If you're interested in helping implement Binius in Jolt, please reach out! https://t.co/JGqGFrazYp
>
> — Eddy Lazzarin ■■ (@eddylazzarin) April 29, 2024

This insight follows a broader industry trend toward smaller fields. The movement began with Plonky2 introducing 64-bit fields, followed by systems like Baby Bear and M31 adopting 32-bit fields. M31's practical implementation remained challenging until recent breakthroughs by the Polygon and Starkware teams. Binius takes this minimization to its logical conclusion, working with fields as small as 2 bits. The system's architecture is built on binary tower fields – a hierarchical structure where each field builds upon the previous one. This approach offers two key advantages: First, at the hardware level, binary field arithmetic reduces to simple logic gates (XOR, AND) and bit shifts, operations already optimized in modern processors. Second, it eliminates the "embedding overhead" found in traditional SNARKs, where single bits often require full 256-bit field elements for security. Instead, Binius allows data to be represented in its natural form – bits as bits, bytes as bytes – creating perfect alignment between data representation and computation. These efficiency gains are important for polynomial commitments, a major bottleneck in zk proofs. Irreducible has leveraged these improvements for GPU-based proof generation, getting an 80% cost reduction compared to CPU implementations. Their new Vision Mark-32 hash function operates ten times faster than existing alternatives, making client-side proof

generation increasingly feasible. Looking toward 2025, several developments are in progress. Irreducible is developing specialized CPU architecture for binary tower fields and ISA, though this remains a longer-term project. Memory usage is expected to decrease by 30-40%, and the technology is being integrated into various zkVMs, with both Irreducible and Jolt actively working on Binius-based implementations. As implementation costs continue to decrease, we can expect accelerated adoption across the zkVM ecosystem.

https://x.com/eddylazzarin/status/1784922354931397077https://x.com/gakonst/status/1813255778889048285

https://x.com/eddylazzarin/status/1784922354931397077https://x.com/gakonst/status/1813255778889048285

Circle Starks and Stwo

The second most important development that happened this year was Circle starks. As I mentioned, there is a move towards smaller fields and within it, the 32 bit fields are most interesting. This is because CPUs and GPUs use these lengths to compute. Using these, proof generation on consumer hardware becomes efficient. Currently, zkVMs use BabyBear field (another 32 bit field), but M31, which is used in Circle Starks is 1.3x faster than it. This is because working with numbers in this fields is much simpler, needing less cycles. Instead of calculation, often there is just rearrangement tricks in computations. The main technical challenge in implementing M31 was identifying a subgroup of order 2, essential for operations like polynomial evaluations, Fast Fourier Transforms (FFTs), and FRI protocol. The breakthrough came with the discovery of a new group structure (Circle Group) and corresponding mathematical operations that finally made these computations feasible within the M31 field.

This breakthrough is going to be used in Stwo prover by Starkware and integrated into Plonky3. Stwo is the replacement for Stone prover that Starknet used for the past 4 years. Going live in Q1/Q2 of next year, stwo includes few other improvements such as GKR lookups to reduce circuit size of the CairoVM (their zkVM – also the first one). zkVMs that use PLONKY3 will probably replace existing BabyBear with it, to take advantage of its efficiency. In terms of performance, Stwo is **125x faster** than the current Stone prover from Starkware! This means faster and cheaper proofs for Starknet next year. They already clocked in 800+ TPS this year. Along with that, Stwo is able to compute 22,000 hashes per second on Consumer hardware. Recursion needs about 24,000 hashes to verify. This means it can do recursive proofs in **under 3 seconds!** Since

this works so fast on 32 bit fields, this is going to be great, again for **Client side proving.** Using GPU acceleration, Stwo will be able to do more than a million hashes a second, which will be great for Ethereum (fast hashing is great for Ethereum)

Kakarot zkEVM is bringing EVM compatibility to CairoVM, powered by the new Stwo prover that uses Circle Starks. This combination is powerful for two reasons: CairoVM's naturally low cycle counts and the improved proof efficiency for proving per cycle. While Kakarot will first launch on Starknet, its plans extend beyond that – similar to existing zkVMs, it will be a prover for EVM blocks on rollups and EVM chains. This would be another important addition to the zkVM space, but with a key difference – instead of using hardware-based instruction sets like RiscV, it leverages CairoVM's specialized architecture along with Circle Starks' innovations in proof generation.

JOLT

JOLT looks at the problem of proving VMs differently. By using circuits for computations and their proof, there is an increase in inefficiency, because of its overhead. Lookups are already making the zkVMs efficient by reducing the VM circuit size and reduced number of columns. The first insight in JOLT is to look them up from a list of pre-computed table of inputs and their outputs – just like while multiplying, we often just remember instead of actually computing. For each instruction of CPU, we can map the inputs and the output in a table and all computations should be in somewhere in the table. But constructing these large tables and working with them becomes hard. And this is JOLT's second insight, not to use the entire table. This is how a real program runs – it only uses a small fraction of all possible instruction input combinations. JOLT introduces a way to efficiently prove lookups into these massive theoretical tables without actually storing them, only "paying" for the entries you actually use during execution.

This is huge because it means we can work directly with standard CPU instructions rather than translating everything into circuits.

JOLT is building a zkVM based, and has released a RV32I (RISC-V 32-bit base integer) instruction set zkVM which is faster than current RiscV zkVMs. Looking forward JOLT will be combining Binius with their Sumcheck for improved performance. It doesn't support recursion and continuations yet, and we should see improvement when these are integrated. JOLT is very early in its journey and there are a lot of optimizations that haven't been done.

Apart from Binius, Circle Stark, JOLT, there are number of improvements in proof systems like Sumcheck acceleration, Folding and many more. This is one of the fastest moving areas in the zk stack, and we're expecting a lot of improvement across different areas. Sumcheck, Smaller fields, SNARK friendly hashes, and Lookups are emerging some as the way to go, and we'll see more development in them.

There is so much in the proof systems development next year, but some things that we can say

- Proof systems are the fastest moving layer of the stack, there are multiple variants in the pipeline for next year. These will mature and will be integrated, leading to big performance gains for the zkVMs
- There are 2 fundamental shifts – move towards small fields, move towards lookups, driving performance. These would be great for client side proof generation, which would be great for privacy applications.
- Proof systems are a competitive frontier: everyone seeks faster, cheaper proofs. If one system achieves a drastic, stable performance lead, it could become the de facto standard. Until then, multiple contenders will coexist, each vying for ecosystem adoption and tooling support.

No More Excuses

A few years ago the main excuse for why we didn't have great applications was that the infra wasn't ready. Fees were high, UX was bad and transactions were slow. Today, that's no longer the case. With the numerous L1s/2s focused on performance, new stacks that enable better appchains, crypto primitives like FHE, TEEs and MPC, and the incredible advancements in ZK, the infra is here.

It's time to build some cool shit.