THE BLOCK THEBLOCK.CO JANUARY 2024

BLOCKCHAIN DESIGN REPORTS SERIES

PROTOCOL PRISM
UNVEILING THE ARCHITECTURAL TRADEOFFS OF MODERN BLOCKCHAINS

COMMISSIONED BY SOLANA FOUNDATION

RESEARCHED BY THE BLOCK

OVERVIEW

Blockchains are poised to become a foundational element of digital infrastructure. This report focuses on fundamental differences in blockchain design that have been chosen by different protocols to pave the way for mass adoption. Despite all modern blockchains aiming to deliver public rails that are performant and secure, their approaches and timelines to achieve those goals vary.

To shed light on these differences, we outline the main functions blockchains perform and introduce a framework to categorize them. With case study-type analyses of Ethereum and Cosmos, along with a thorough review of Solana's design approach, we compare their strengths and weaknesses. Our study highlights the critical differences between horizontal and vertical scaling and the trade-offs between them. This comparison provides a detailed perspective on the technological choices and strategies influencing the evolution of smart contract platforms as the backbone of a growing on-chain financial economy.

The report is structured as follows: Part 1 defines the core functions of a blockchain system, and then classifies modern blockchains into several encompassing architectures according to the variations across their functions. Part 2 contextualizes the tradeoffs required for scaling blockchains across each function (such as data storage or execution), highlighting the way in which scalability optimization impacts important properties such as decentralization, security and interoperability. Part 3 takes a deep dive into Solana's design philosophy as an execution-focused, vertically integrated blockchain, as well as the many unique hurdles it has faced in the pursuit of achieving greater scalability. Finally, Part 4 compares the key design choices taken by protocols in the context of their varying scaling approaches, offering insights into their progress to date alongside critical commentary from the builders and thinkers shaping the future of blockchain design.

CONTACT

Email: research@theblock.co Twitter: @TheBlockPro

ACKNOWLEDGMENTS

We would like to thank the Solana Foundation for commissioning this research report. We would also like to thank their team for providing feedback and input for this report, in particular Austin Federa. We would also like to thank everyone at The Block who assisted with this report - design team: Zoe Ellyse Del Rosario; research team: Marcel Bluhm, George Calle, Greg Lim, Arnold Toh. We are also grateful to those that shared their valuable perspectives through interviews for this report:

Sunny Aggarwal (Osmosis Labs)

Lucas Bruder (Jito Labs)

Noam Cohen (Binary Builders)

Justin Drake (Ethereum Foundation)

Austin Federa (Solana Foundation)

Haseeb Qureshi (Dragonfly Capital)

Kyle Samani (Multicoin Capital)

Jeff Washington (Solana Labs)

Anatoly Yakovenko (Solana Labs)

The authors of this report may hold tokens mentioned in this report. Please refer to The Block's financial disclosures page for author token holdings.

AUTHOR

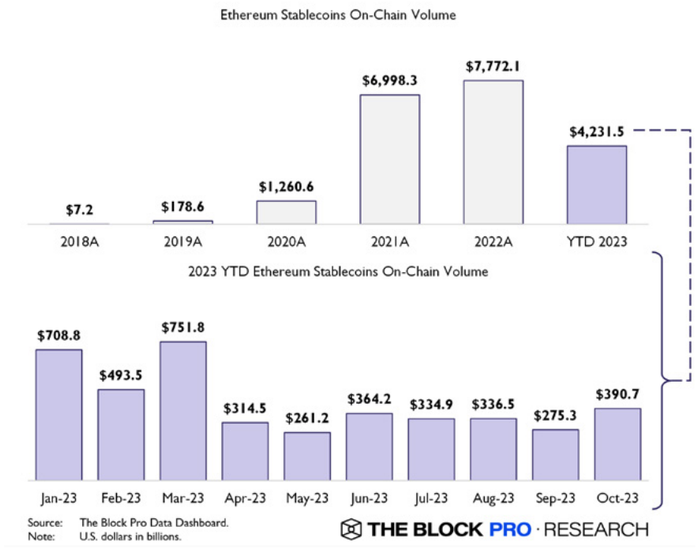
Kevin Peng, PhD
Research Analyst
@PengCapital
LinkedIn

TABLE OF CONTENTS

INTRODUCTION

Modern blockchains have evolved dramatically from their humble origins as simple peer-to-peer (P2P) networks for the secure transfer of money. Smart contracts, first introduced on Ethereum, have played an especially key role in enabling blockchains to expand their capabilities far beyond processing P2P transactions. With the rapid growth of decentralized finance (DeFi) protocols built upon smart contracts in recent years, including decentralized exchanges (DEXs), lending protocols, synthetic asset protocols, non-fungible token (NFT) marketplaces, and more, crypto users now have unprecedented access to a stunning array of [mostly] permissionless financial instruments.

Today, tens of billions of dollars worth of crypto assets are held in DeFi protocols alone. The smart contract-based blockchains that host these protocols are home to an even greater amount of capital overall, by an order of magnitude. As of November 2023, about $125 billion worth of fiat-backed stablecoins such as USDC and USDT reside on smart contract platforms. In October 2023, ~$390 billion in stablecoin volume was settled on the Ethereum blockchain alone, underscoring the immense trust that is now placed in smart contracts and their underlying blockchains.



This expanded optionality with respect to on-chain financial activity would not have been possible without a corresponding redesign of popular blockchain architecture in the early years following Bitcoin's rise to prominence. For instance, the introduction of the Ethereum Virtual Machine (EVM) and its Turing complete state machine laid the foundation for the smart contracts that underpin most crypto applications today. In the current blockchain landscape, core development teams have continued to design and build new functionalities for blockchains, primarily with the goal of enhancing scalability. After the sudden surge in both the volume and complexity of DeFi transactions on-chain beginning in late 2020, it has become especially clear to users and developers throughout the crypto ecosystem that blockchains still require significant improvements before they will be able to effectively meet the demands of a global digital economy.

However, one of the major challenges with evaluating the state of blockchain development today is the sheer variety of design goals and approaches that exist among various chains. Some chains exist solely to execute transactions while others are intended to provide security guarantees for other chains. In many cases, different development teams will use contrasting language to describe similar concepts, depending on their particular implementation or focus.

In this report, we aim to provide an in-depth understanding of the modern blockchain landscape in the context of the varying design philosophies that exist today. First, we provide an objective framework for classifying blockchains according to their key functions and architecture. Then, we explore some of the main components that underlie network functionality, as well as the various trade offs made within different initiatives to optimize these components. Finally, we conduct a deep dive into the Solana network's design and scaling approach, along with an assessment of its progress in comparison to other popular approaches today.

Let's begin with a look at different blockchain functions and how we can use them to categorize nearly all blockchains today.

PART 1
CLASSIFYING MODERN BLOCKCHAINS

At their core, fully-fledged smart contract platforms must all perform the same key responsibilities: transaction execution, settlement, consensus, and data availability (DA). These are often referred to as different “layers” in a hypothetical blockchain with fully modular components. Briefly, the execution layer is responsible for executing state transitions within a particular execution environment or virtual machine. In other words, execution defines nearly all aspects of user interactions with a specific blockchain on a fundamental level, from the way transactions are sent between addresses to the way contracts are written to perform certain actions.

The settlement layer is where transaction correctness and finality are established; transaction correctness is established through methods such as validity proof verification, zero-knowledge proofs, and cryptographic signatures, or via dispute resolution (a.k.a. social consensus). Finality is achieved through consensus mechanisms that include both social consensus and algorithmic consensus, such as Proof-of-Work (PoW), and Proof-of-Stake (PoS). Meanwhile, the consensus layer is responsible for - as the name suggests - achieving consensus between various nodes / validators on the order of transactions. The data availability layer is responsible for attesting to and providing transaction data availability - either to other layers or independent users querying the data. In many blockchains today, such as on Ethereum or Avalanche, the settlement layer is coupled with the consensus layer or the data availability layer, rather than existing as a standalone function.

Layer	Responsibilities
Execution	Execute transactions and produce new state commitments
Settlement	Establish transaction correctness and finality Facilitate cross-execution layer communication
Consensus	Reach agreement on transaction ordering
Data Availability	Attest to availability of transaction data Provide transaction data on-demand

THE BLOCK PRO · RESEARCH

These key blockchain functions – execution, settlement, consensus, and data availability – form the basis of our framework for classifying modern blockchains. This framework is sometimes referred to as a "modular" framework, which, along with the inclusion of the data availability layer, was largely popularized as a term by the Celestia team in recent years. It is worth noting the potential for some implicit bias with the use of these terms, given the fact that Celestia intends to be a leading data availability layer for blockchains with a rollup-centric roadmap. For instance, the terms "monolithic" and "modular," taken together as terms used to describe blockchains, can easily be misconstrued to imply that monolithic blockchains are non-modular and single-function, while modular chains are flexible and multi-functional. The truth is often far more complex. In fact, some would argue that the popular definition of a monolithic chain, i.e. one that performs all blockchain functions, might be more suitable for so-called modular layers like Celestia's data availability network, which only publishes transactions but does not execute them.

Nonetheless, we still find it useful to group and classify blockchains by the way they perform (or don't perform) key functions, and largely stick with the common definition for monolithic chains in this report. We also retain the data availability layer in our framework, given the growing acceptance of DA as a function for scaling rollups. In terms of our classification framework, we prefer to describe it as a blockchain function framework, as opposed to a modular framework, to more accurately reflect our focus on clearly defining blockchains by their responsibilities/functions, rather than suggesting that these responsibilities *must* be optimized individually.

1.1 MONOLITHIC BLOCKCHAINS

Most blockchains throughout history have been monolithic chains, performing all key responsibilities outlined above within a single architecture. The first blockchain, the Bitcoin blockchain, along with other early blockchains such as Litecoin and Dogecoin, are examples of monolithic blockchains. Notably, all three of these blockchains do not have smart contract functionality, and are thus limited to crypto transfers between peers on the network. Smart contracts, first introduced on Ethereum, enable a far wider range of transaction types and user behavior, but require robust Turing complete execution environments, which dictate execution logic.

Smart contract platforms like Ethereum, Avalanche, and Solana are commonly referred to as monolithic chains as well, despite the fact that they must execute more complex transaction logic and bear more computational load. In this report, we refer to monolithic chains solely in the context of smart contract platforms, which now serve as the backbone for the bulk of decentralized financial activity on blockchains today. Over time, smart contract functionality has given rise to increasingly sophisticated DeFi protocol design, providing blockchain users with new degrees of financial flexibility that have, in turn, created new channels of demand for block space. It is the confluence of these forces that has pushed early smart contract platforms like Ethereum to the limits of their natural capacity.

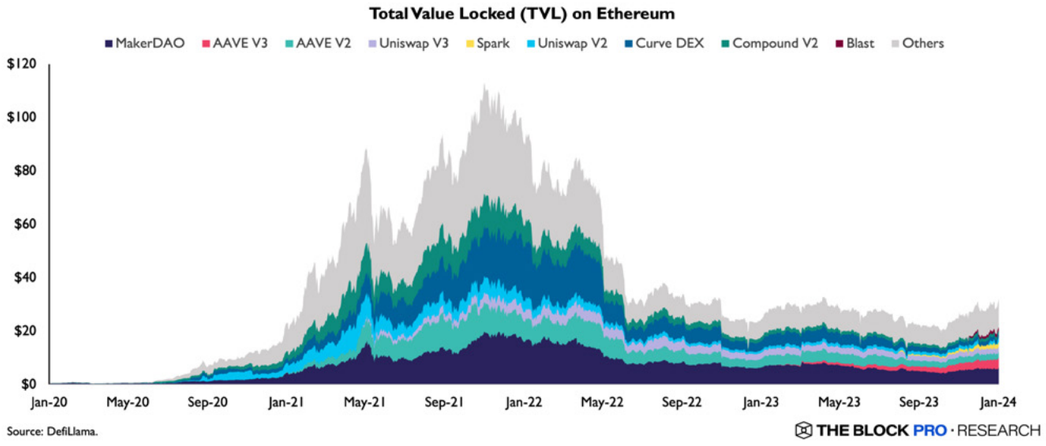
This brings us to the center of where a significant fraction of core blockchain innovation lies today: scaling. In general, monolithic blockchains – also referred to as Layer 1 (L1) networks - perform the full range of key blockchain functions within their own architecture, but that is roughly where their similarities end. In the pursuit of greater scalability, some L1s have begun to adopt a modular approach to development, embracing the idea of separating key functions into separate chains to be optimized on an individual level. Layer 2 (L2) networks, commonly referred to as rollups, embody this philosophy by focusing solely on execution to segregate computational load, leaving the remaining functions to an underlying L1, most commonly Ethereum. Other L1s, such as Solana, have taken a different approach, aiming instead to maximize the performance of their existing functions within a single, monolithic architecture. As it turns out, when it comes to execution, regardless of architecture, it is the underlying execution environments that largely give way to variations in overall performance.

1.1.1 RISE OF THE EVM

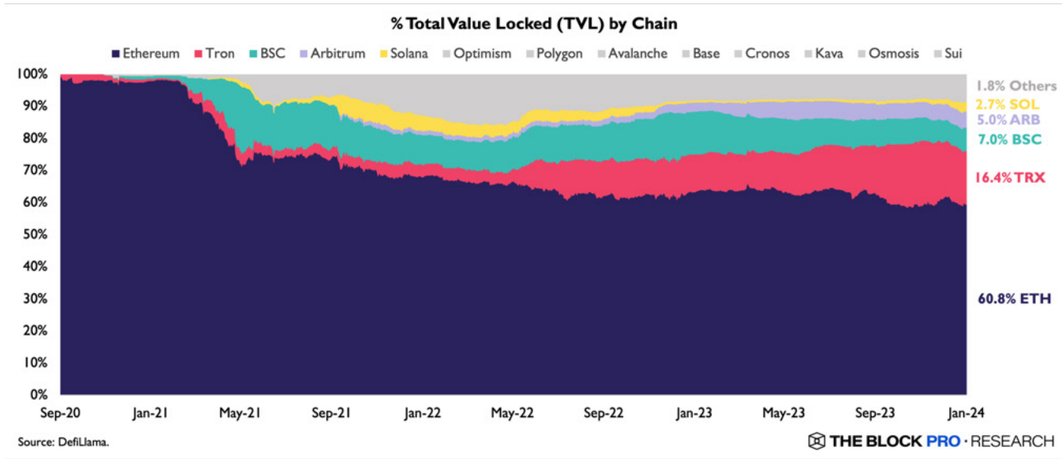
> *"First and foremost, Ethereum aspires to be the blockchain with highest (economic) security. And then, number two, we want to have enough scalability to provide a lot of utility to the world."*
> – Justin Drake (Ethereum Foundation)

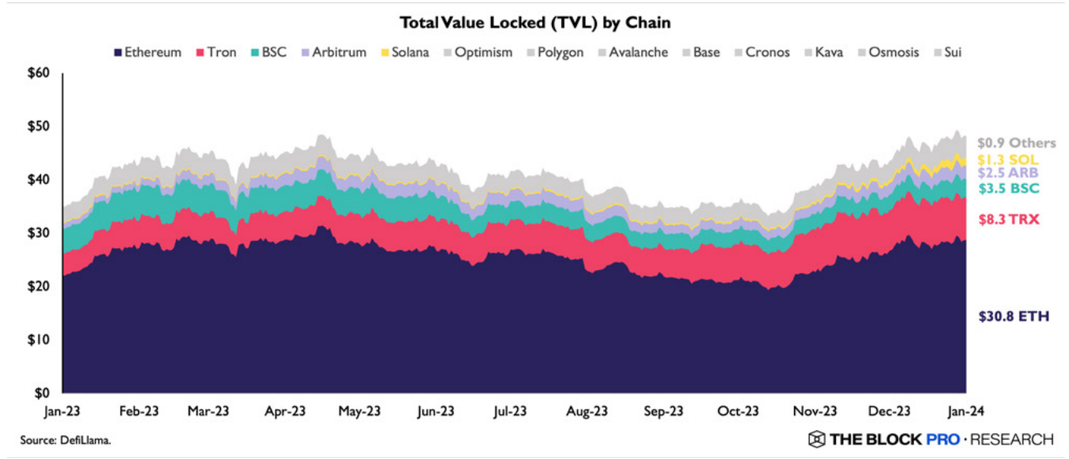
The EVM is by far the most dominant execution environment among both L1 and L2 smart contract platforms today; utilized by 9 of the 10 largest blockchains by total value locked (TVL), as of December 2023.

The explosive growth of EVM-compatible chains in recent years largely stems from Ethereum's pivotal role in fostering many of the core DeFi primitives still used today. In 2020, during what became known as "DeFi Summer," the combined market cap of DeFi protocols grew to a then-record high of ~$20.8 billion, capturing ~5.3% of crypto's total market cap at the time. By the end of 2020, nearly all the TVL and activity in DeFi remained on Ethereum, with a 96% market share firmly establishing the L1 network as the leader in the new DeFi space.

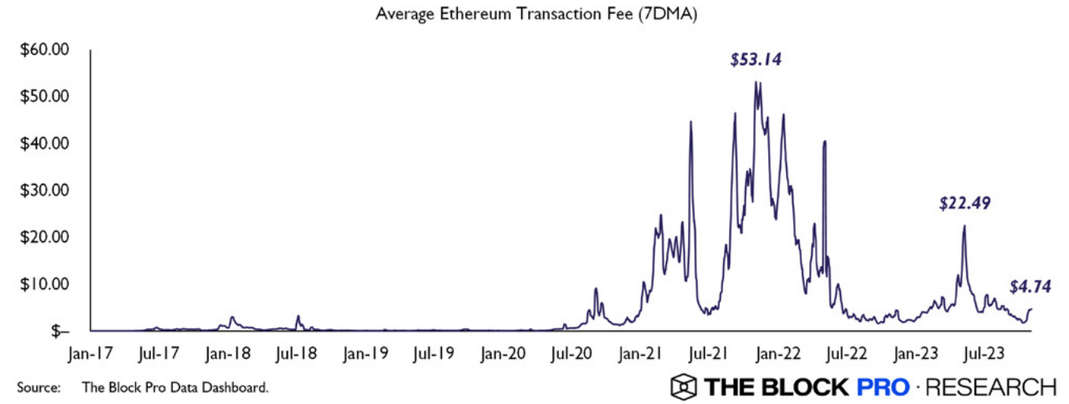


Over the ensuing years, Ethereum would go on to lose a significant portion of its DeFi market share to alternative L1s, even as DeFi's total market cap reached staggering new heights exceeding $170 billion at its peak. Buoyed by rampant speculation and record inflows to crypto markets, the events of 2020-2022 were ultimately highly instructive for identifying the key challenges in modern blockchain scaling, as well as the strengths and limitations of the EVM when bounded by a monolithic architecture.





One of the strongest drivers of the adoption of EVM-compatible L1s was the Ethereum network's pronounced struggles with surging user demand throughout 2021. Average transaction fees on Ethereum soared to record highs in the first half of 2021, coinciding with the rapid growth and initial adoption of Binance Smart Chain (now known as BNB Chain).



BNB Chain's out-of-the-box compatibility with the EVM meant that the growing community of Ethereum users and developers could seamlessly transition to the new L1 without significant changes to their typical workflow, such as accessing protocols via Metamask. With BNB Chain touting fast confirmation times and low transaction fees, the additional convenience of its EVM support proved to have a profound effect. By May 2021, the BNB Chain ecosystem comprised over 20% of total DeFi TVL. That same year, with a little help from liquidity mining incentives, several other EVM-compatible L1s would go on to see unprecedented growth while offering a similar value proposition as BNB Chain, with Avalanche, Fantom, and Polygon capturing ~5.8%, 2.4%, and 3.0%, respectively, of DeFi TVL by the end of 2021.

1.1.2 PUSHING THE LIMITS OF EVM-COMPATIBLE L1S

There are several reasons why, compared to Ethereum, transactions tend to be cheaper on L1s such as BNB Chain, Avalanche, or Fantom, all of which are EVM compatible and have a similar monolithic architecture. One of the main reasons is that demand for block space on these L1s is simply not as high relative to Ethereum. Even today, Ethereum remains dominant among smart contract platforms in terms of TVL, DEX volume, and NFT market share. As such, it becomes difficult to use nominal transaction fee amounts as reliable indicators of scalability in direct comparisons between L1s.

In the past few years, isolated spikes in demand for specific L1s have yielded spurts of useful insights into how EVM-based L1s realistically perform under duress. For example, in March 2022, the Fantom network saw its average gas prices temporarily skyrocket to record highs after one of its core devs abruptly announced his departure from the industry. Other notable spikes in gas prices include May 2022 during the collapse of Terra Luna, November 2022 during the collapse of FTX, and most recently, July 2023 during the shutdown of the Multichain bridge. All the while, average daily transactions have been on a steady decline since early 2022. These incidents highlight the fact that despite offering lower fees compared to Ethereum on average, Fantom's design does not make it immune to the negative effects of sudden increases in block space demand. Similar conclusions can be drawn with other monolithic chains that utilize the EVM, most of which exhibit the same gas profile as Fantom.

Sensitivity of gas prices to block space demand is one indicator of a network's overall scalability, though organic demand is generally tricky to measure or reproduce in a purely mathematical sense. One study conducted by Gengmo Qi and the Dragonfly Research team sought to quantify the throughput of various L1s in terms of DEX trades, factoring in the block gas limit, block time, and gas expenditure of a typical token swap on each chain in order to determine average swaps per second. Notably, the team found that EVM-compatible L1s were largely bound by the same scaling constraints due to using the same gas model and state transition function. For instance, BNB Chain is able to process far more trades per second relative to its peers, but it also makes significant tradeoffs in order to do so - as of this writing, the network still has only 29 active validators. The paper also suggests that users appear more interested in ecosystem strength, user experience, and low fees than in performance on non-Ethereum L1s. Performance isn't a competitive factor (yet) as these blockchains aren't consistently capacity constrained, barring occasional usage spikes. Further discussion on scaling tradeoffs in general can be found in Part 3 of this report.

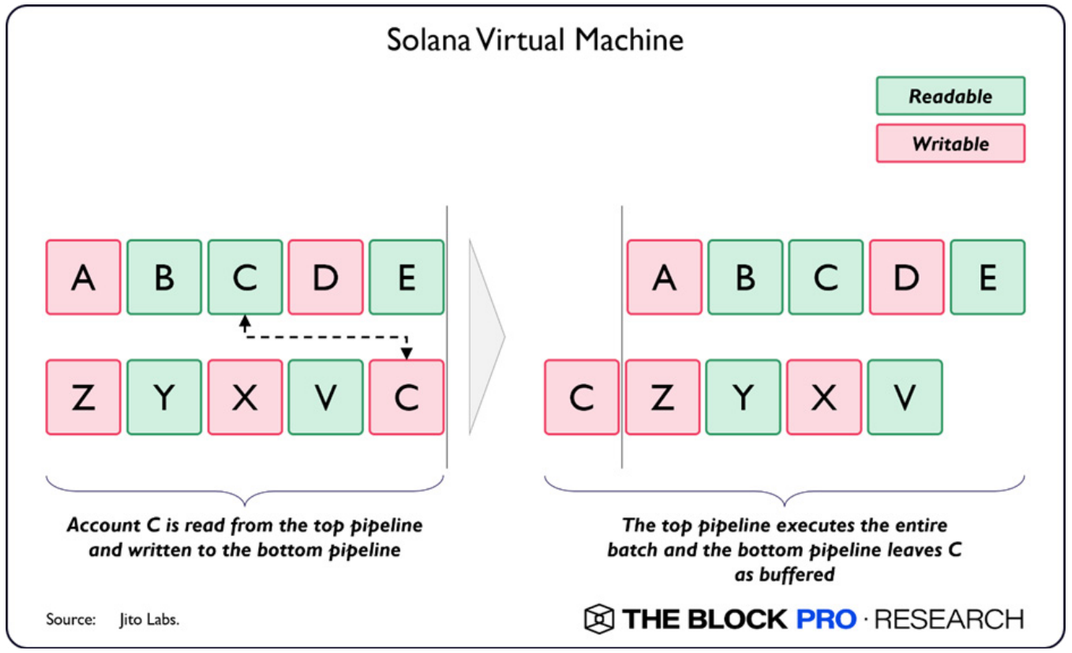
1.2 MONOLITHIC BLOCKCHAINS WITH PARALLEL PROCESSING

A common characteristic of most smart contract platforms today is that they utilize a single-threaded runtime, as in the EVM. In other words, transactions are processed sequentially, ensuring that conflicting transactions are not executed during state transitions. However, this behavior can also impose a limit on transaction throughput, especially when considering that transactions are often submitted simultaneously and much more frequently than they are being finalized on chain.

One of the main thrusts to directly address these throughput bottlenecks is the development of blockchains that employ parallel processing of transactions. If early non-smart contract blockchains such as Bitcoin are considered as "1st-gen" chains, and smart contract platforms such as Ethereum are considered "2nd-gen" chains, then blockchains with parallel processing capabilities can be reasonably categorized as "next-gen" for the theoretical benefits they provide.

> *"Because of [Solana's speed and low fees], you have more flexibility in the type of DApps that you can create."*
> – Lucas Bruder (Jito Labs)

The most prominent example of this design strategy in the market today is the Solana network, which grew at a meteoric pace alongside its EVM-based L1 competitors in 2021 before cooling down in the aftermath of the FTX collapse in 2022. Instead of the EVM, Solana utilizes a custom execution environment known as the Solana Virtual Machine (SVM).



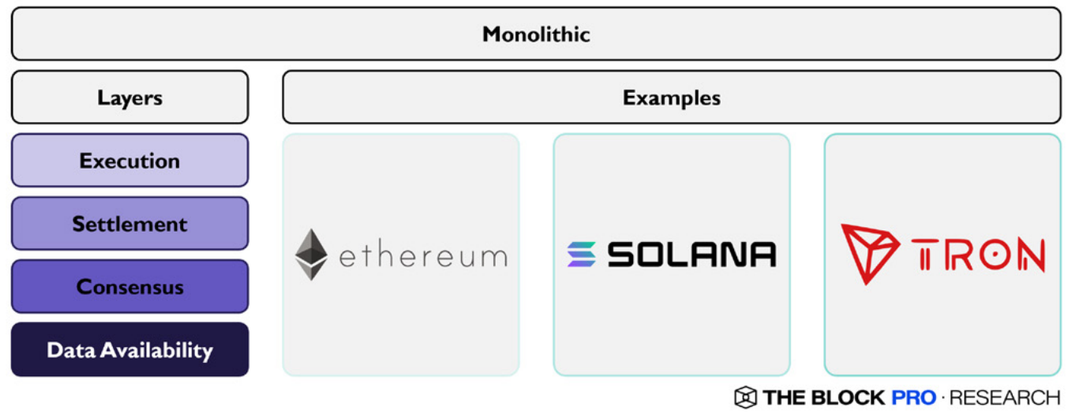
The key component of the SVM that enables parallel processing is the Sealevel runtime, which requires all transactions to explicitly define which accounts will be read from or written to. In practice, this adds an additional level of complexity for Solana developers, but it also enables an important step change in terms of transaction processing efficiency. Since transactions that do not introduce conflicting state changes are clearly marked prior to execution, they can be effectively grouped together and processed simultaneously, i.e. in parallel. For example, if account A sends 1 SOL to account B, this should have no impact on whether account C can send 1 SOL to account D; thus, processing these two transactions sequentially - as in the EVM - can be considered a source of inefficiency.

In order to fully leverage the efficiency gains imparted by the SVM, Solana relies on several additional features that are specific to its monolithic architecture, such as Pipelining transaction processing unit (TPU) and its Proof of History (PoH) synchronization system. At a high level, this approach can be summarized as an extreme version of vertical scaling with its own unique tradeoffs, which are discussed in further detail in Part 3 of this report.

In recent years, other core teams have begun to deploy blockchains that feature a similar focus on parallel

processing as a means of enhancing the scalability of monolithic systems. The most notable examples of this are Aptos and Sui, two L1s that were spun off from Meta's original Diem blockchain project. Aptos and Sui both use parallel processing mechanisms that bear similarities to Solana's. In Aptos's Block-STM engine, state-independent transactions are processed with a preset order as determined by its consensus engine; Sui's Narwhal/Tusk consensus engine performs a similar data ordering task that ultimately reduces computational load prior to transaction execution.

Perhaps the most notable trait shared between Aptos and Sui is their use of the Move virtual machine (Move VM), which requires contracts to be written in the Move programming language. Similar to the SVM's support for Rust and C/C++ over Solidity, the Move VM's unique programming requirements pose an additional barrier for existing EVM developers looking to transition to Aptos or Sui. This circumstance is largely derived from the fact that novel execution environments tend to require novel instructions as well. More importantly, it is clear from these examples that execution environments play a major role in dictating user experience, and are currently one of the key variables for enhancing transaction throughput as well.



1.3 MULTICHAIN PROTOCOLS

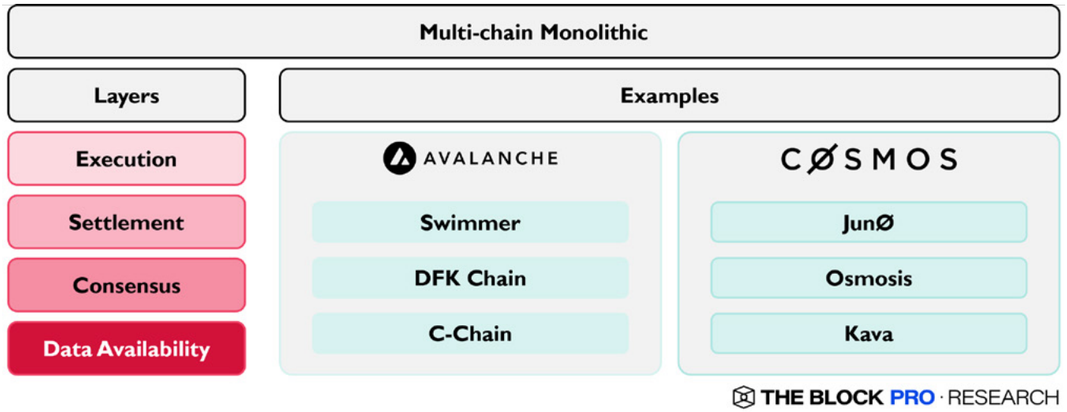
> *"The strength of Cosmos is that you can build your own application. If you want customizability on top of sovereignty, speed and decentralization, Cosmos is the way to go."*
> – Noam Cohen (Binary Builders)

In the previous section, we reviewed the current state of efforts to maximize throughput on individual smart contract platforms, where we found that execution environments are a central bottleneck for scaling. Whereas parallel processing is intended to improve the performance of monolithic blockchains under heavy demand, some protocols have taken an alternate approach to combatting congestion, namely, by separating user applications into individual blockchains. This strategy was largely popularized through the Cosmos team's original vision of an "internet of blockchains," which is essentially a collection of application-specific chains (app-chains) that are connected by a shared communication protocol. Aside from this dedicated communication standard, namely the Inter-Blockchain Communication (IBC) protocol, blockchains in the Cosmos ecosystems do not differ meaningfully from other monolithic chains in terms of their performed functions. With few exceptions, other "multichain" protocols are essentially the same; they perform their own execution, settlement, data availability sampling, and consensus. As such, they are most accurately described as "multichain monolithic" networks, and we focus primarily on the slight variations in their architecture in this section.

The general idea behind Cosmos is that application developers should be able to easily create their own blockchains that are tailored to their specific use cases. Cosmos chains are built via the Cosmos SDK and utilize the Tendermint BFT consensus engine, but they remain entirely sovereign networks. Sometimes referred to as "hubs," these networks typically feature their own validator sets, along with independently variable governance parameters and security guarantees. Meanwhile, the IBC protocol acts as a standardized means through which Cosmos networks can securely pass messages between one another, further enabling cross-chain asset transfers and general UX integration between applications.

It is important to note that transactions are settled on their respective chains, meaning that the confirmation of blocks on one chain has no direct impact or relation to other chains. One of the benefits of this architecture is that computational load is split amongst various chains depending on user demand at specific times. In addition, governance or consensus failures on one app-chain would be confined to that particular chain, as opposed to the broad impacts of a similar failure on a monolithic chain featuring many applications and protocols.

However, multichain protocols also face issues of fragmented resources; each chain is tasked with bootstrapping enough human and economic capital to establish a secure validator set, and liquidity is generally fragmented across multiple chains. Recent developments in the Cosmos ecosystem have aimed to address these challenges with features like Replicated Security, which would allow new "consumer chains" to essentially rent security from the Cosmos Hub chain by using its validator set. Under this design, Cosmos Hub validators perform all validation for consumer chains, collecting 100% of fees generated from the consumer chain. Of course, this also requires general social consensus on the Cosmos Hub serving as a settlement layer. It is important to note that settlement and consensus for consumer chains under the Replicated Security model remains separate from the Cosmos Hub; consumer chains receive periodic IBC packets to update their validator sets to match the Cosmos Hub, but activity on consumer chains still has no relation to activity on the Cosmos Hub, and vice versa. As of now, Hub validators are compelled to perform validation for consumer chains that have been onboarded to the Replicated Security model through governance, but there is a technical limit for how many chains Cosmos Hub validators can reasonably support. In the future, these validators would be able to "opt-in" to consumer chain validation, collecting fees only from those they choose to support.



Avalanche has employed a similar horizontal scaling approach with subnets, which are essentially independent chains that are constrained by the requirement of utilizing a subset of Avalanche C-Chain validators. This adds an additional layer of security to subnets, as subnet validators must already be honest participants on the C-Chain, but it is still limited by the size and economic security of the particular validator set. This is similar to the opt-in version of the Cosmos Hub's Replicated Security model, whereby validators

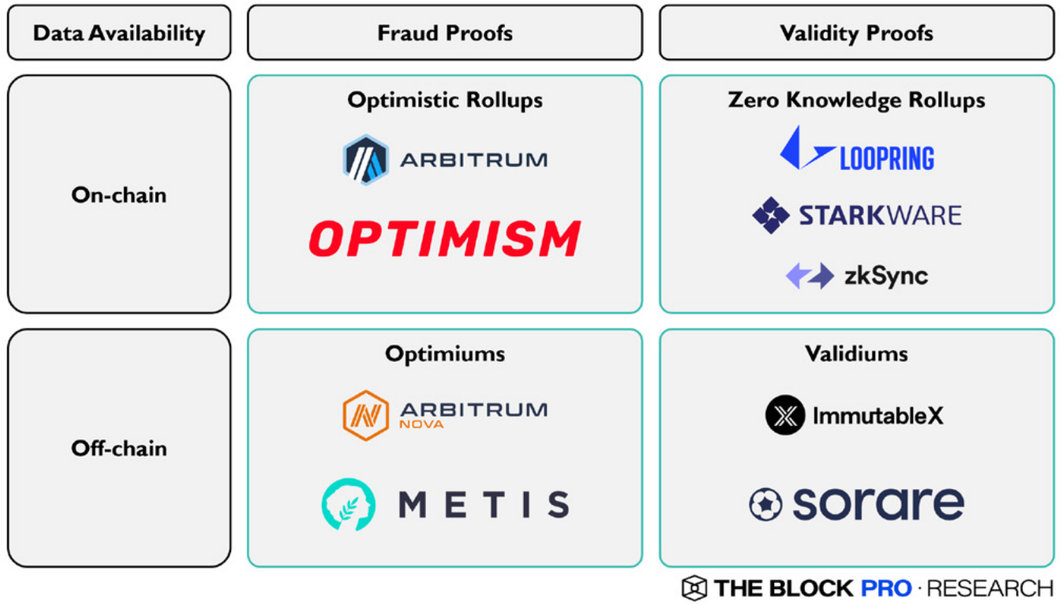
for the more popular C-Chain can perform validation for subnets, but consensus achieved on the C-Chain does not directly translate to consensus on subnets.

Polkadot is another multichain protocol that effectively delegates applications to individual blockchains, with the main differentiator being that all blockchains in its ecosystem, known as parachains, are secured by a single governing chain called the Polkadot Relay Chain. The overall Polkadot network is heavily dependent on the functioning of its Cross-Consensus Message (XCM) protocol, which facilitates both Relay Chain-parachain and parachain-parachain communication. As is the case with all multichain protocols, the cross-communication standard is ultimately an enabler but also the main limiting factor when it comes to overall throughput, security, and liquidity. Among multichain protocols, the Polkadot model is perhaps the closest to a true "shared security" model, where parachain transactions must also be confirmed on the primary Relay Chain. However, we have also discussed at length in previous reports how this arrangement has effectively bound the pace of development on parachains to the current state of development on the Relay Chain.

1.4 MODULAR ARCHITECTURES

In some ways, multichain protocols can be interpreted as an extension of modular architecture, wherein the design usually offers a choice between either siloed, application-specific security, or delegated security to a single consensus layer. The rollup-centric roadmap laid out by the Ethereum foundation has drawn considerable scaling efforts towards the idea of Ethereum being the "global settlement layer", leading to the proliferation of various L2 solutions. L2s, which can be further categorized by their validation mechanisms and data availability modes, are a fundamental extension of the basic premise that effective blockchain scaling can best be achieved by optimizing individual blockchain functions. As mentioned above, these are generally defined as execution, settlement, consensus, and data availability layers in a theoretical modular blockchain.

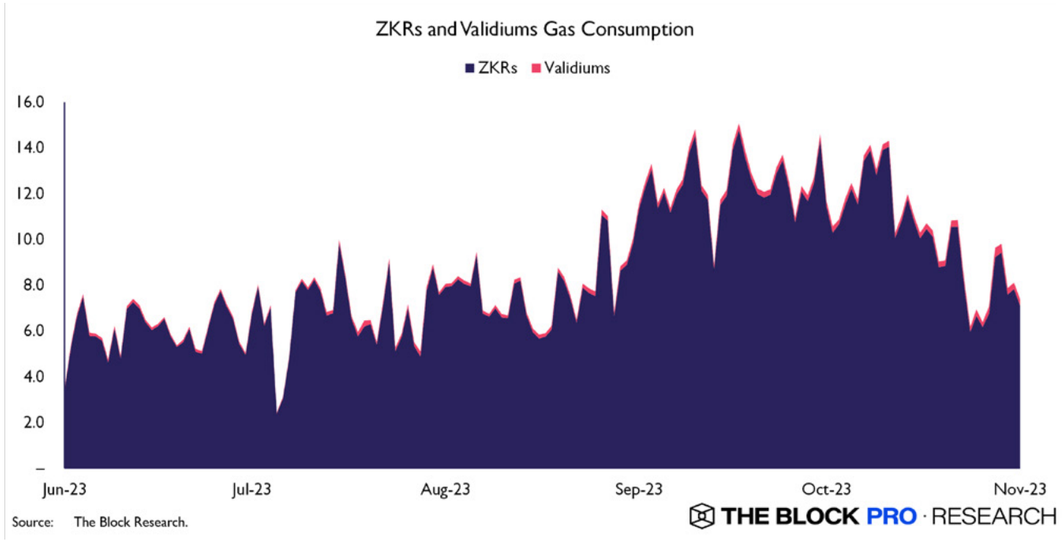
Since transaction processing and execution are computationally intensive tasks, rollups today generally exist for the purpose of separating execution from the settlement and consensus layers. In other words, rollups represent the execution layer (L2) for an underlying L1 such as Ethereum. The general mechanism for rollups is as follows: transactions are executed on a rollup chain, which are subsequently batched by a sequencer, producing a snapshot of the resultant VM state, which is then sent to the settlement layer for verification and consensus layer for finalization.



L2s that leverage fraud proofs differ greatly from those that leverage validity proofs in terms of their security assumptions. Examples of L2s using fraud proofs include Arbitrum, Optimism and Base and examples of L2s leveraging validity proofs include Starknet and Loopring. The former assumes that submitted batches are correct and waits for an arbitrary dispute period before finalization while the latter necessitates a validity proof that guarantees the correct execution of the proposed batch's transactions. Typically, validity proofs are to be submitted with every L2's state update, thereby consuming more computational resources on the underlying layer, Ethereum, than L2s using fraud proofs, since fraud proofs are only required in the event of a dispute. That said, validity proofs allow for state finality to be achieved much faster than fraud proofs.

There is also a growing opportunity in providing alternative data availability (DA) solutions for L2s, as rollups typically rely on the underlying L1, Ethereum, for data availability. This is further evidenced by the growth of Validiums like ImmutableX and Optimiums such as Mantle, which leverage off-chain data availability solutions. It should be noted that DA differs from data storage, as DA only requires that data can be accessed when queried while data storage refers to having the complete archive of past transaction data. Simply put, it is possible for one to make a DA layer much lighter than a data storage layer, in exchange for a marginal decrement in the guarantee of the data being accessible.

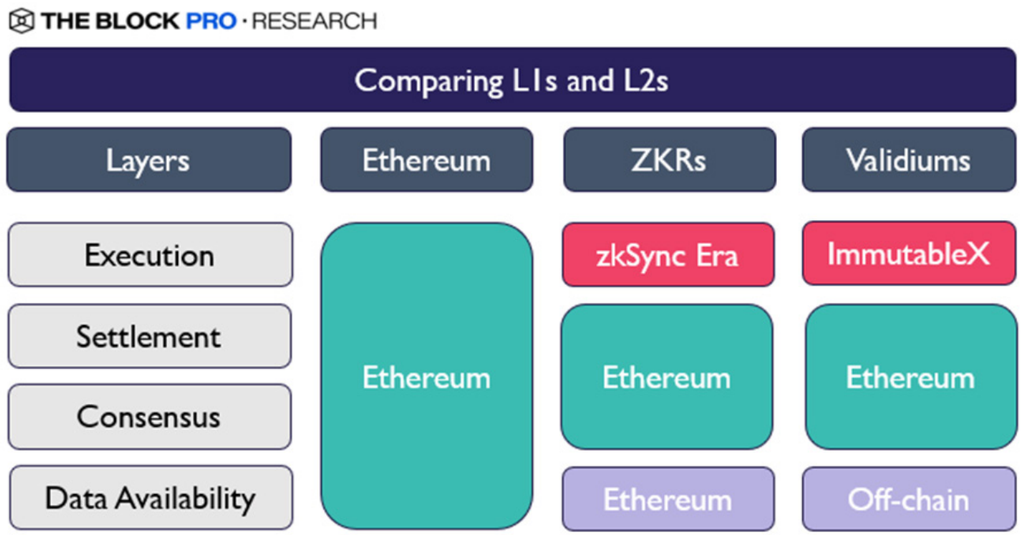
There are currently two types of DA layers: centralized DA committees, which are typically used by StarkEx protocols, and independent DA-optimized layers, such as Celestia and Avail (prev. Polygon Avail). The former comprises of multiple entities providing attestations that they possess a copy of proposed batches' transaction data while the latter submits random data samples on-chain to attest that the batched transaction data is available with high probability. Both approaches aim to alleviate the costs associated with storing all transaction data on-chain and it shows some extent of efficacy when comparing gas costs associated with rollups using on-chain DA against off-chain DA. The chart below compares the gas consumption of Ethereum Layer 2's leveraging validity proofs with on-chain DA (ZK rollups) against off-chain DA (Validiums).



However, the progress towards increased modularization and abstracting DA is not without costs. The debate between modularization and integration is not entirely novel; it had been present when operating systems tech stacks were being developed. Fundamentally, modularization does not enhance the scalability of any blockchain. Instead, it can be argued that it shifts the burden of computation to different entities and when demand for computation surges, the problem resurfaces. For example, both Arbitrum and Optimism saw sustained periods of high gas fees when Ethereum gas fees were consistently high. In other words, modularization can optimize how resources are utilized and potentially improve performance under certain conditions, it doesn't eliminate the fundamental scalability challenges facing blockchain networks. It's a step toward managing scalability issues, not a complete resolution.

Understanding data availability is fundamental to designing a blockchain that scales well because it is the main bottleneck for most, if not all, public blockchains today. Specifically, blockchain nodes must be certain that the data in a block is accessible if it is to accept the block as valid. As such, various approaches have been experimented with, from on-chain data availability, which are used by most Layer 1s, to off-chain data availability, employed by Validiums and Optimiums (previously known as Plasma), to data availability optimized layers, such as Celestia. As the need for increased throughput rises, it will be inevitable that we see more public blockchains today adopting some form of data availability optimization in order to keep up with the demand for block space.

A handful of Layer 1s aim to tackle this problem from the ground up, by integrating localized fee markets, such that users interacting with high-demand smart contracts pay more gas than users making a simple transfer. This provides a targeted incentive to lower demand for the frequent users. It remains to be seen how such novel incentive/pricing mechanisms play out in the market when demand is back and blockchain resources become scarce again.



To sum up, typical L1s have assumed the four responsibilities of execution, settlement, consensus and data availability. L2s have attempted to outsource both execution and data availability. While execution has proven to increase the throughput of the underlying L1 to some extent, evidenced by the growth in adoption of optimistic and ZK rollups, Validiums like ImmutableX and ApeX have shown more potential to scale blockchains. That said, Validiums make some trade-off in security by hosting data off-chain, which is why we see the rise of data availability optimized layers like Celestia. Based on the diagram above, we are about to reach the very edge of how modular a blockchain can be, which means the main focus will return to the fundamental design of a blockchain.

1.5 SHARDED BLOCKCHAINS AND THE FUTURE OF SCALING

Sharding is a technique used to improve the scalability and performance of blockchain networks by dividing the network's workload into smaller, manageable pieces, each called a shard. It can be seen as running a blockchain as parallel chains that merge every few blocks to establish a consensus on the global state, before dividing into multiple chains again for processing transactions. Sharding is an extremely complex approach to scaling that seeks to increase a blockchain's throughput without sacrificing security. The current state of development, however, reveals the difficulty and nuance involved in either building net new sharded blockchains or implementing sharding within existing blockchain architectures.

The complexity and scalability benefits vary depending on what layer of the blockchain stack is being sharded. Blockchains can either have a monolithic execution chain with shards used only to store data (data shards), or have nearly independent shards (execution shards), an approach that is marginally different from running separate Proof-of-Stake chains. It should be noted that data shards are different from data availability optimized layers, as data shards store transaction data in its entirety. As such, data shard nodes are expected to hold said data and ensure that the data remains accessible for all network participants.

While data shards are less challenging to implement, they do little to improve the chain's scalability unless accompanied by rollups, since L1 transaction data can be stored on the main chain. Thus, data shards are primarily meant as an alternative for rollups to post calldata, which then introduces the issue of the liveness of shards – whether shard nodes are able to provide transaction data upon request as readily as full nodes. This is crucial because rollups require transaction data to be accessible when verifying a validity proof or a fraud proof.

Execution shards, which can execute transactions on their own, with their own set of validators, are much more challenging to implement. This is expected considering how every shard's state must periodically align with a global consensus each time the shards merge. It is similar to a situation with multiple chains, each with its own consensus, arriving at a global consensus every few blocks. Additionally, execution shards require significant economic stake to secure, as they become capable of proposing fraudulent state transitions, giving malicious actors an incentive to compromise individual shards. However, neither splitting stakers across shards nor sharing a common stake across all shards would be optimal. Splitting stakers across shards would mean fragmenting the economic stake securing the whole chain, thereby reducing the economic security of a single shard, while sharing a common stake would mean that stakers are expected to validate the actions of all other stakers, essentially giving stakers the same computational work as a single unsharded chain.

Still, there have been numerous attempts at building shard chains, such as Near, Harmony and Elrond. These shard protocols all share one commonality: smart contracts are hosted on a single shard. This is because cross-shard smart contract interactions are significantly more complex, with more potential vulnerabilities.

To understand why, we can look at cross-chain smart contract interactions. Typical blockchain interactions are atomic; either the whole transaction happens, or not at all, while cross-chain interactions are not atomic. For most smart contract interactions, transactions need to be atomic. For example, a token swap on a decentralized exchange must either swap token A for token B, or make no swap. A break in atomicity, such as sending token A in the first transaction and receiving token B in the next transaction, increases the susceptibility for malicious attackers to revert the second transaction.

As a result, when performing a cross-chain swap, the user trusts the facilitator to behave honestly. In the case of cross-chain decentralized exchanges, there is typically a centralized actor or a centralized group of actors. However, a permissionless shard chain protocol like Near or Elrond would not be able to ensure that all nodes would honestly facilitate a cross-shard interaction, and the incentive design for such a responsibility is still not well-established, especially since a large cross-shard swap could be highly profitable for a malicious node to intercept. Thus, the design for execution shards is still incomplete and may not be ready in the foreseeable future.

In the case of Ethereum's Proto-Danksharding, data blobs are attached to Ethereum blocks to serve as cheaper alternative for on-chain data storage for L2s. These data blobs are also automatically discarded after some time, ideally within 3 months. This would be made even cheaper when Ethereum launches its shard chains, allowing for up to 64 different data blobs to be attached.

PART 2
SCALABILITY TRADEOFFS ACROSS BLOCKCHAIN FUNCTIONS

In the previous part, we looked at the general range of architectures for modern blockchains, tracing their evolution from simple p2p networks used primarily for the transfer of a native asset, to the complex, dynamic smart contract platforms that underlie on-chain financial ecosystems today. The rising computational demands of this emerging digital economy have driven most blockchain developmental efforts toward a single common goal: scaling. This landscape has largely informed our framework described above, which classifies blockchains according to their design with respect to key functions: execution, settlement, consensus, and data availability.

In recent years, growing efforts have been made toward individually optimizing these different functions of a blockchain, particularly in the Ethereum community, in the hopes of enhancing blockchain scalability overall - a so-called modular approach. Still, optimizing even a single function is a non-trivial task. In this section, we break down the main network components that underlie these functions, as well as the various tradeoffs made in the name of enhancing a particular functionality. As part of this discussion, it is useful to consider the impact of various blockchain functions through the lens of a popular mental model: the scalability trilemma.

The scalability trilemma describes the interdependent relationship between scalability, decentralization, and security in blockchain architecture. This dynamic serves as a general design constraint for scaling efforts, and essentially states that maximizing two of the three factors inevitably leads to decreased competency in the third. Each of these three properties is critical for blockchains to function at any point in time. Scalability generally refers to the ability of a network to process a high volume of transactions per second beyond the capabilities of a single consumer node. Decentralization is analogous to censorship resistance, which means that a properly decentralized network places little to no trust in a small group of nodes that could potentially fail or become compromised, impacting the entire network. Finally, security refers to a network's ability to resist an economic attack – either via collusion between participating nodes or a direct takeover of a controlling stake in the network.

Early efforts to scale blockchains have illustrated the difficulty in directly increasing throughput without having deleterious effects on decentralization or security. For instance, simply increasing block size can produce short-term benefits in terms of throughput, but ultimately imposes higher transaction costs on nodes, which has a net negative effect on decentralization by limiting the number of potential participating nodes. Similar challenges quickly arise when designing blockchains that maximize one or two properties without also solving for resultant weaknesses in the third. A highly scalable, decentralized blockchain faces

heightened security risks due to relatively lower costs of attacking the network. By the same token, highly decentralized, secure blockchains are often harder to scale, given that throughput generally becomes throttled by higher latency between a larger number of nodes.

Ultimately, the scalability trilemma should not be thought of as strict limiting factor for blockchain development, but more as a general mental framework to understand the tradeoffs that result from enhancing specific functions. Over time, this framework has become progressively less imposing on blockchain design as teams have continued to produce breakthroughs that are able to effectively mitigate the issues in past designs. Even so, the scalability trilemma still remains generally relevant as a model when considering blockchain design approaches as a whole, especially as teams continue to experiment with new ways to push their networks beyond their current scaling limits. Below, we review some of the underlying components that comprise the key blockchain functions in the context of the scalability trilemma, focusing on how these underlying components either directly or indirectly influence the delicate balance between scalability, decentralization, and security.

2.1 DATA STORAGE

Data storage is a critical part of blockchain functionality, and can be roughly divided into global state and historical data. Global state can be thought of as a live snapshot of all the data that can be accessed and modified for a blockchain at a given time. This data is typically stored in Merkle trie structures, which allows it to be quickly retrieved by validators for verification and execution of the next state change. On the other hand, historical data refers to the full range of data generated throughout the course of a blockchain's lifespan, including all confirmed blocks, all the transactions within those blocks, transaction signatures, etc.

Generally, historical data is needed to sync new validators to a blockchain, but does not need to be frequently accessed by active validators who rely on global state data to execute state transitions. However, over time, the constant buildup of raw data can lead to growing storage demands, progressively increasing the computational and economic load on validators. This is commonly referred to as state bloat, which can have a negative impact on transaction execution and throughput if validators begin to take longer to confirm transactions. At the same time, increasing hardware requirements for validators to combat state bloat can become a centralizing force, limiting the number of participants capable of running a validator.

One of the primary ways that historical data can be stored for EVM-based blockchains, which can have significant amounts of state bloat, is to use off-chain storage solutions. These solutions are typically blockchains themselves, offering a basic level of decentralization, but they also face the same issues that are relevant to smart contract platforms. For instance, they require some sort of incentive structure to ensure that nodes continue to write, store, and provide access to data, and they also require the ability to scale themselves. Some solutions, like Arweave, reward validators for adding new blocks, similar to Ethereum. Others, like Filecoin or Storj, rely on a contract-based mechanism for storing specific data over a given period of time, which enhances their scalability by reducing the complexity and persistence requirements of data storage. Still, it is important to note that most blockchains, including Ethereum, have not yet enshrined data storage solutions as a part of their protocol, and continue to rely on nodes being able to store progressively larger amounts of historical data over time.

Data storage is also important in the context of enabling rollups, which require the submission of either fraud proofs or validity proofs to the underlying L1 that can be accessed at any time. Note, however, that data storage typical refers to blockchain historical data in the context of state bloat, while data availability more specifically refers to the ability to access required data as needed, without which rollups could not be considered secure. As discussed in Part 1.3, rollups typically utilize one of two types of data availability layers, which guarantee data access on demand rather than full archival of historical data. These include centralized data availability committees and independent data availability layers, each with their own economic and security tradeoffs.

2.2 CONSENSUS AND SETTLEMENT

There are several key factors that ultimately contribute to a network's ability to consistently achieve consensus and settle transactions. One of the most important is governance, which is often overlooked but can have sweeping implications for a particular network's security and future development. At a high level, governance dictates the overall process for network upgrades and serves as a coordination mechanism between users, developers, validators, and stakeholders. In the event of potential network failures or malfunctions, governance can also be the ultimate adjudicator for consensus disputes.

Modern smart contract platforms feature a wide range of governance structures as part of their overall design. Ethereum, for example, employs an off-chain governance structure that is mostly conducted

through social consensus involving various stakeholders. Despite the fact that the network is the largest and likely most decentralized among smart contract platforms today, its governance has been relatively effective at enacting necessary changes to the network over time. Ethereum's successful transition from Proof of Work (PoW) to Proof of Stake (PoS) consensus in September 2022, albeit delayed, is a testament to this functionality, notably revamping its network architecture despite initial opposition from now-defunct Ethereum miners.

Decentralization of governance is important for maintaining a network's overall resistance to censorship, but it can sometimes be a barrier to timely protocol implementation as well. It is worth noting that despite Ethereum's eventual transition to PoS in 2022, many of its L1 competitors had already adopted the mechanism since their inception several years before. Some networks have established a more formalized governance structure by conducting the process on-chain. For instance, Cosmos chains delegate voting powers to validators, who have the ability to set custom governance parameters (voting period, quorum, veto threshold, etc.) depending on their individual needs.

In theory, this design has the effect of aligning interests between governance and stakeholders, as validators are often among the largest stakeholders in the network as well. Thus, one would expect them to vote with the intent of ensuring the long-term success of the network, which would enable a more streamlined development process as well. Indeed, a cursory glance at the proposal history of some of the largest Cosmos chains today reveals a prolific pace of protocol upgrades relative to L1s that mostly conduct governance off-chain. At the same time, the use of on-chain governance itself does not necessarily guarantee a productive or risk-free outcome. In past reports, we've discussed some of the unintended consequences of the Polkadot network's governance mechanism, which places a heavy emphasis on the Relay Chain as a minimum security threshold. Specifically, the network's reliance on its governance council for a wide range of parameters extending to its parachains has historically led to a situation where mostly-sovereign parachain teams have been bottlenecked by delays to technical upgrades and approvals from the core development team.

In practice, governance in most blockchains, aside from those in the Cosmos ecosystem, is primarily a process of social consensus that often involves coordination with core developers. Validators and large stakeholders typically have a voice in protocol-level decisions, but these decisions are ultimately carried out by the developers who maintain and release official client software. Such is the case for the Solana network, where validators in the past have played a key role in coordinating network restarts and re-establishing consensus after outages. More recently, the community has begun to establish a more formal governance process in anticipation of future growth.

One final note on the importance of governance in blockchains is its role in dictating economic incentives, which is in turn a major factor in economic security. Security in blockchains consist of both technical and economic components; whereas technical security is mostly addressed by the robustness of client codebases, economic risks often pose a more realistic risk to network stability. Networks that do not feature high economic value bear the risk of well-capitalized actors obtaining a controlling stake. In other words, they have a low cost of attack, and can thus be described as having weak economic security.

Many blockchain development teams are funded extensively by the distribution of native tokens from their treasuries, typically managed by a separate foundation entity. These foundations often have significant flexibility in the way they distribute funds, which can theoretically have an impact on both decentralization and security. For example, many blockchain foundations initiated incentive programs throughout 2021-2022 targeted at attracting users and developers, which effectively contributed to a wider distribution of tokens that can be used to secure the network. At the same time, these foundations have also used their treasuries to secure capital for future development, which has the opposite effect of concentrating large portions of token supply within a handful of entities.

Though it is not always intuitive, token economics often play a major role in the security and decentralization of blockchain networks. After all, blockchains fundamentally revolve around the secure, permissionless transfer of money. This dynamic is especially true in the case of blockchains secured by PoS consensus mechanisms, which encompasses nearly all smart contract platforms that exist in the market today. For PoS networks, the value of tokens that make up the network have a direct impact on the economic incentives afforded to validators, as well as the theoretical cost of attacking the network via consensus.

Examples of this interplay between economic value, consensus, security, and decentralization are abundant throughout crypto history, and they have often revealed the key tradeoffs between various blockchain designs. The Cosmos ecosystem is especially insightful in this regard, given the standard practice of combining consensus and governance among its networks. One notable case involves the Juno network, which, in March 2022, voted to forcibly confiscate the token holdings of a single holder after the core team determined that the holder in question had been able to accumulate far too much of the supply at low

cost. This incident highlights the fragility of PoS networks when the cost of obtaining a controlling stake is relatively low – as well as the overriding power of social consensus in extraordinary circumstances.

In fact, maintaining or achieving high economic value is a major factor for security and decentralization in most networks that are either new or that have relatively low market caps. Early periods of growth are the most risky for new blockchains, as they essentially face the constant risk of economic attack until it becomes economically unviable. In previous reports, we have commented on the fact that many Cosmos chains feature validator sets that are dominated by only a handful of validators, resulting in concentration of both governance and risk. One way to combat this risk is through the adoption of liquid staking, which essentially allows users to participate in securing the network while maintaining liquid control of their assets. The basic economic security case for liquid staking is that it reduces the opportunity cost of staking tokens to help secure the network. This was a major component of the Cosmos 2.0 proposal introduced in late 2022, which broadly sought to accrue more value and security to the Cosmos Hub, and is part of the reason why liquid staking derivatives (LSDs) have grown to become so popular on Ethereum in recent years as well.

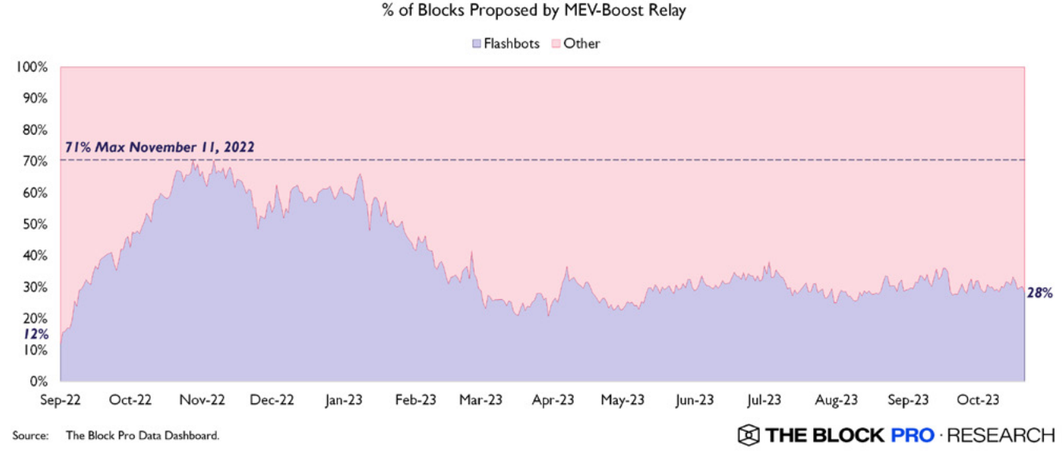
When it comes to economic security in multichain systems, one perspective worth noting is the negative impact of fragmented liquidity. As we suggested above, the ease of deploying new blockchains via the Cosmos SDK goes hand in hand with a period of low economic value and high security risk in the early days of an ecosystem. In theory, this issue would not be as problematic if Cosmos chains all shared the same liquidity and validator set, which would increase the network's overall cost of attack, as is the case for more integrated L1 chains. This same logic is a major part of the reason why Ethereum, as the second most valuable cryptocurrency by market cap, is often considered a global settlement layer in the broader blockchain ecosystem.

2.3 EXECUTION

In the previous part, we alluded to the fact that modularization and fragmentation of liquidity may pose notable risks in terms of economic security. Low economic value of a network lowers the barrier for a consensus-based attack, and it also makes it easier for the network to become quickly and inadvertently centralized. However, economic security alone is not always enough to prevent issues with decentralization. In this section, we explore some of the ways in which transaction execution and its variable underlying parts can grow to become a potentially problematic source of centralization as well.

From the standpoint of economic security, the Ethereum network is one of the most difficult to attack among all smart contract platforms, owing to the size of its market cap and the distribution of its nodes. At the same time, its validators and user base are not entirely immune to the allures of economic incentives either. This basic fact can have snowballing implications in terms of decentralization. Remember, validators are responsible not only for providing consensus, but also for transaction execution.

One of the main side effects of Ethereum having perhaps the most vibrant on-chain DeFi ecosystem today is that it is also a rich source of extractable value for opportunistic actors. In recent years, the pursuit of so-called maximal extractable value (MEV) has become one of the most competitive areas of the crypto industry. At a basic level, MEV searchers look for unique opportunities on-chain that allow them to siphon profits from organic user activity. This can come in the form of front-running transactions, arbitrage, and more. Often, these strategies are so profitable that it becomes economically viable for MEV searchers to coordinate with validators through bribes. In recent years, this system of competitive bidding for transaction execution priority has been largely automated through the use of tools like the Flashbots MEV-Boost Relay.



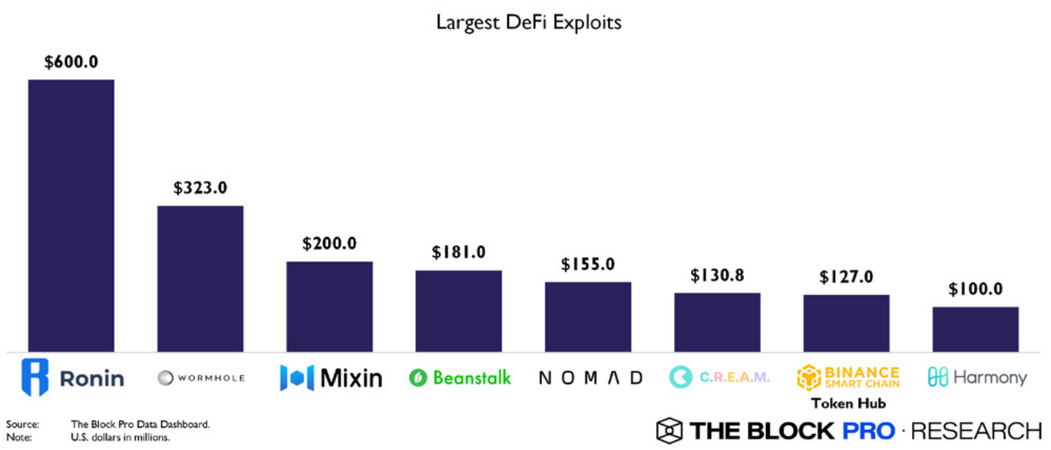
This tool has become so prevalent within the Ethereum ecosystem that as of this writing, roughly 33% of all blocks have been proposed by validators running MEV-Boost. From the perspective of users staking ETH for yield, the economically optimal choice often becomes simply choosing validators that capture

MEV. Validators have broadly accepted this strategy in order to remain competitive as well. Such tools may arguably introduce centralization vectors, for example if certain transactions or users are censored by them. However, it must be kept in mind that a larger share can only delay inclusion of censored transactions, but not prevent them (unless a share of 100% is achieved). For example, even if the share of MEV-Boost running validators is one third, it is highly likely that an MEV-Boost censored transaction becomes included in the blockchain in a short amount of time (the probability of inclusion approaches 1 within a minute).

The impact of optimizing transaction execution on decentralization extends to broader efforts to increase blockchain scalability as well, as approaches to interoperability introduce novel challenges. This is apparent when considering the role of communication protocols, which are ingrained in numerous aspects of the blockchain stack. The ability for nodes to communicate is critical in both vertical and horizontal scaling efforts. In the case of horizontal scaling, as with rollups, multichain protocols or sharding technologies, communication protocols also enable interoperability between often disparate systems. Yet no matter the case, enhancing communication essentially entails reducing latency, which often translates into greater hardware requirements.

Cross-chain communication protocols exist in various forms today, whether as the Cosmos ecosystem's IBC, the Polkadot ecosystem's XCM, or even Circle's Cross-Chain Transfer Protocol (CCTP) for USDC. In nearly all cases, one of the primary goals of implementing these protocols is to expand the available avenues for liquidity. In theory, the end result of this goal would be a wider distribution of assets across various networks (indirectly increasing decentralization), as well as increasing the overall throughput of blockchain transactions when viewed in aggregate.

Nonetheless, communication protocols today often suffer from the issue of being non-interoperable with each other. With each new standard for cross-chain communication comes another dimension for centralized security risk as well. Bridge exploits across both L1s and L2s have accounted for most of the largest DeFi exploits in history.



These risks become especially concerning when they can impact entire networks and vast swaths of capital. For instance, in late 2022, the exploit of the BSC Token Hub revealed a previously unknown bug related to the IBC protocol, highlighting the extent of damage that can arise from a single unexpected incident.

In recent years, new efforts have emerged to scale blockchains without introducing multiple new layers of complexity and risk, primarily through innovations in execution environments. These approaches typically require a full re-evaluation of blockchain architecture from the ground up - which can be challenging on its own - but they also reap the tried-and-true benefits of handling key blockchain functions within an integrated system. In the next section, we take a deep dive into the Solana network's uniquely iterative, execution-centric approach toward optimizing scalability and how it stacks up against the most popular smart contract platforms today.

PART 3
SOLANA: THE CASE FOR MONOLITHIC EXECUTION

Broadly speaking, current approaches to addressing the core issue of scalability in blockchains can be divided into two main categories: horizontal scaling and vertical scaling. Horizontal scaling generally involves splitting key blockchain functions across multiple systems to alleviate throughput bottlenecks. Multichain protocols such as Cosmos represent an extreme version of this approach, employing multiple app-specific blockchains that can independently perform all the key functions of typical blockchain to distribute network activity.

A more nuanced form of horizontal scaling is the use of modular architecture, which involves splitting key blockchain functions into separate layers - usually blockchains themselves - that can then be optimized individually to enhance overall throughput. This approach is exemplified by the modern rollup landscape, consisting mostly of various execution layers that settle transactions on Ethereum, as well as specialized data availability layers like Celestia. Notably, Ethereum development now follows a rollup-centric roadmap, which began in earnest with the creation of a new PoS consensus layer and subsequent transition from PoW during The Merge.

Vertical scaling represents a fundamentally different approach from horizontal scaling, focusing on maximizing blockchain performance within a monolithic architecture. Typically, this means increasing hardware requirements for validators, enabling greater computational capacity and higher transaction throughput while sacrificing decentralization to an extent. However, this is only one part of the story. As we saw in Part 2, every blockchain function has multiple underlying components, all of which can impact scalability, decentralization, or security during the process of optimization. In this section, we take a focused look at Solana's unique execution-focused architecture, its progressive evolution as a response to scaling challenges in recent years, and how its scaling approach compares to other major smart contract platforms today.

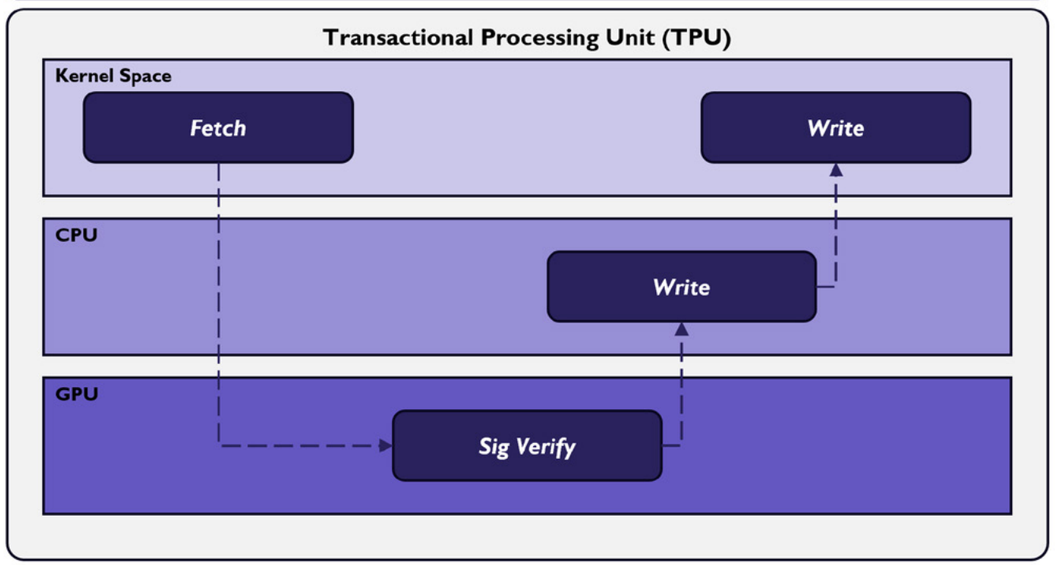
3.1 OPTIMIZING EXECUTION FROM THE GROUND UP

> *"Our premise is that blockchain becomes mainstream. That is only possible if accounts are cheap and useful in all kinds of things, right from finance to games to NFTs. We initially had a goal to be able to smoothly run a billion accounts on chain, to make that possible... we now can do 10 billion – at Solana's speed."*
> - Jeff Washington (Solana Labs)

Solana is a monolithic blockchain launched in 2020 with the primary aim of enabling high scalability and fast transaction finality while maintaining low transaction fees. At a high level, Solana's design approach can be described as an unwavering focus on optimizing blockchain execution by continually leveraging the state-of-the-art in hardware technology. Central to this approach is the network's use of a custom execution environment, known as the Solana Virtual Machine (SVM). As we touched upon in Part 1, the SVM's Sealevel runtime enables the parallelization of transaction processing, which results in an order of magnitude improvement in throughput compared to single-threaded runtimes like the EVM.

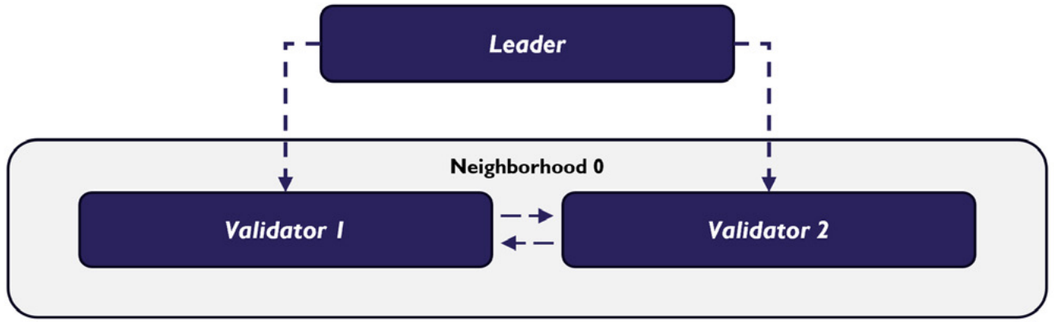
Solana's parallel processing capabilities are reliant on several other key innovations within its custom blockchain stack that operate in lockstep over the lifecycle of a transaction, ultimately putting significant strain on validator resources as well. For example, the Solana transaction processing unit (TPU) utilizes each validator's kernel space, CPU, and GPU at various stages in transaction processing, depending on whether the validator is running in "leader mode" or "validator mode." At any given moment, the entire network will have one validator serving as the leader, which is in charge of producing blocks, ordering transactions within the block, and propagating to all other validators.

Pipelining in Solana: The Transaction Processing Unit



Source: Solana Labs.

In validator mode, validators follow a logic scheme known as the transaction validation unit (TVU) - as opposed to the TPU used by the leader - in order to validate blocks and transmit data sent from the leader node. These processes rely on a proprietary block propagation protocol as well, known as Turbine, so as to not place excessive computational demand on the leader node. The Turbine mechanism requires leaders to break blocks into much smaller data packets according to the User Datagram Protocol (UDP) standard, which are then retransmitted via validator nodes to others operating in validator mode.

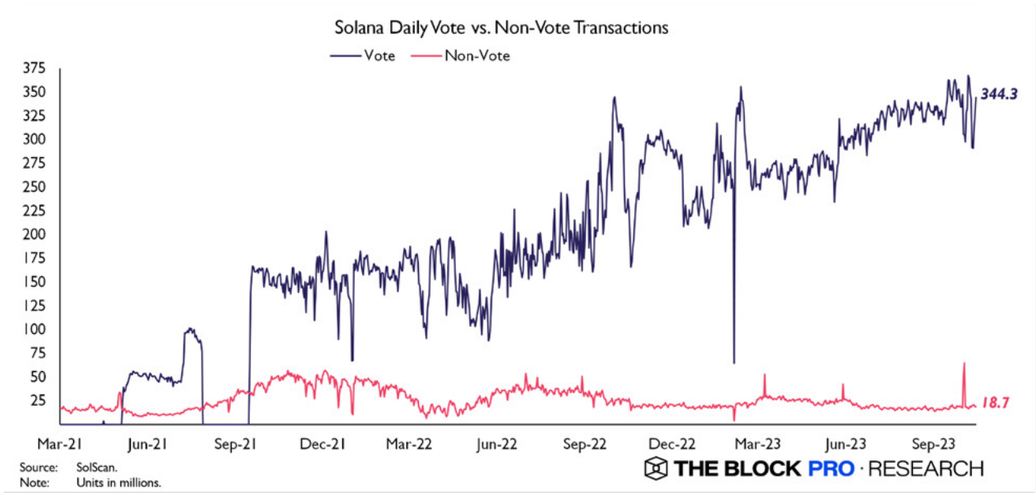


Source: Solana Labs.

In essence, this significantly reduces the bandwidth requirements for leader nodes, with the tradeoff being that UDP packets limit the size and complexity of transactions sent on the network as well. This, in turn, might mean that a single smart contract / program interaction in the SVM could require far more transactions compared to an equivalent one in the EVM. UDP packets are the base unit for both vote transactions, which refer to transactions sent between validators to achieve consensus, and non-vote transactions, which refer to transactions initiated by users.

The outcome of this simplified transaction processing scheme described above is that Solana validators must conduct frequent communication between one another for both execution and consensus. In fact, vote transactions on the network typically outnumber non-vote transactions by a significant margin, and have been rising steadily over the past year. This messaging-heavy design would naturally lead to excessive latency between Solana validators were it not for the network's unique Proof-of-History (PoH) algorithm, which essentially acts as an internal clock that allows validators to be constantly synchronized without additional network communication overhead. PoH is a verifiable delay function (VDF) run by every

validator, which means that validators can always quickly confirm that they are in sync with other validators with respect to this internal clock. This baseline level of additional information effectively replaces some of the typical communication required between validators with local computation. Paired with Solana's Tower BFT consensus mechanism, validators on the network are able to reach consensus on hundreds of non-vote transactions per second, representing a major increase in throughput relative to individual EVM chains.



3.2 IMPROVING VALIDATOR PERFORMANCE: EXPERIMENTATION IN PRODUCTION

From a development perspective, Solana's scaling approach largely entails fine-tuning validator clients (a.k.a. software) to fully leverage the performance of currently available hardware. In principle, this close relationship between standard validator hardware and blockchain performance would translate into increased throughput over time. However, if this logic holds, it would also imply that Solana's validator hardware requirements can be driven higher by both rising network demand and expansions of client responsibilities.

Solana Validator Requirements

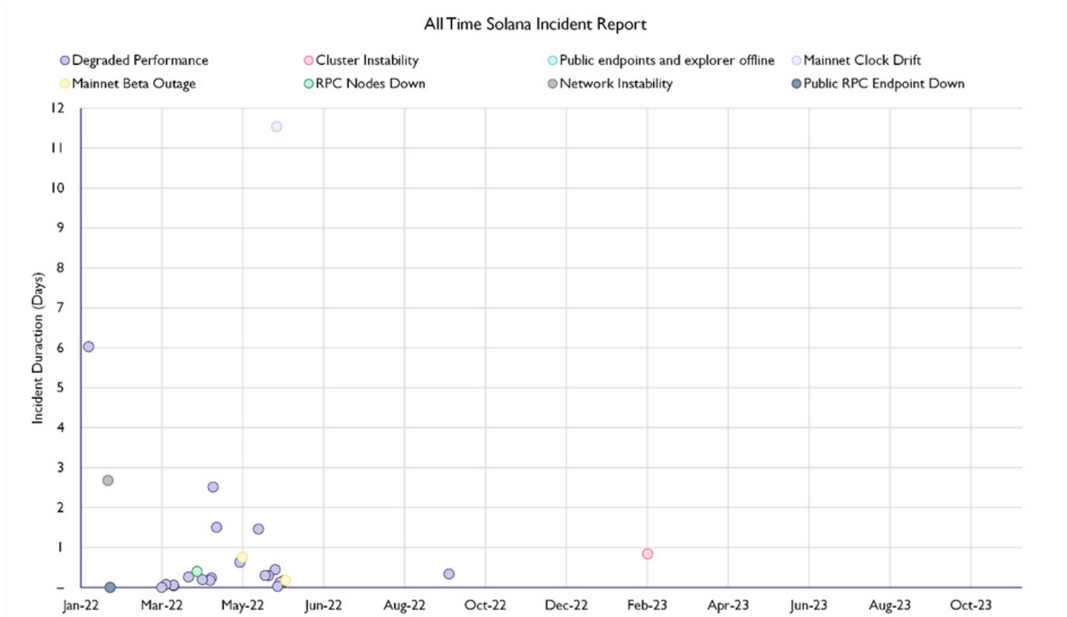
	Category	December 2021	December 2022	December 2023
CPU	Cores / Threads	12 cores / 24 threads	12 cores / 24 threads	12 cores / 24 threads
	Processor Speed	2.8 Ghz or faster	2.8 Ghz or faster	2.8 Ghz or faster
RAM	Memory	+128Gb	+128Gb	+256Gb
	Motherboard	+256Gb	+256Gb	+512Gb
GPU	General	Not Required	Not Required	Not Required
	NVIDIA CUDA	Required to use GPU	Required to use GPU	N/A
Networking		300 Mb/s Symmetric 1 Gb/s preferred	300 Mb/s Symmetric 1 Gb/s preferred	1 Gb/s Symmetric 1 Gb/s preferred
For RPC Nodes	CPU	16 cores / 32 threads	16 cores / 32 threads	16 cores / 32 threads
	GPU	+256Gb	+256Gb	+512Gb

Source: Solana Foundation.

THE BLOCK PRO · RESEARCH

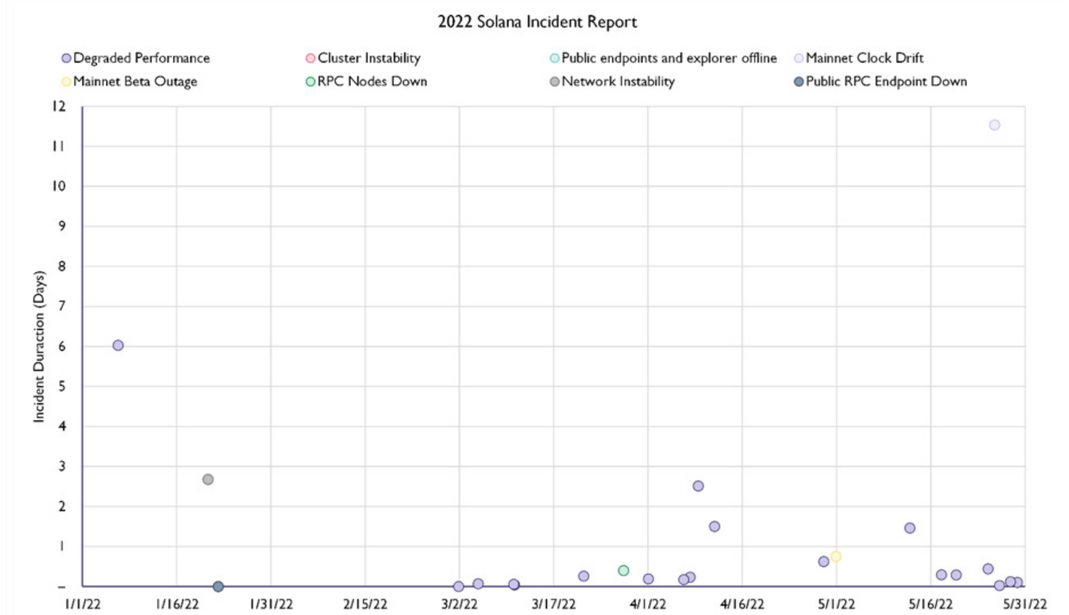
Indeed, taking a look at Solana validator documentation over the years reveals a notable jump in minimum spec requirements between December 2022 and 2023. CPU requirements have remained unchanged since late 2021, but RAM requirements doubled and networking requirements more than tripled over the same period. Interestingly, the NVIDIA CUDA toolkit was required to make use of validator GPUs as late as June 2023, but this requirement has since been removed, suggesting decreased reliance on GPUs going forward. Meanwhile, GPU requirements for RPC nodes specifically increased from 256 GB to 512 GB between December 2022 and 2023.

Perhaps the best way to understand the Solana network's overall scaling approach and current state of development is to trace its evolution over the past few years. As evident from our discussion of Solana's architecture and transaction lifecycle above, the network's liveness and performance rely on a delicate balance between multiple unique functions carried out by its validators. Any major disturbances to this balance can have a dramatic impact on the network's performance, or, even worse, its security and ability to reach consensus.



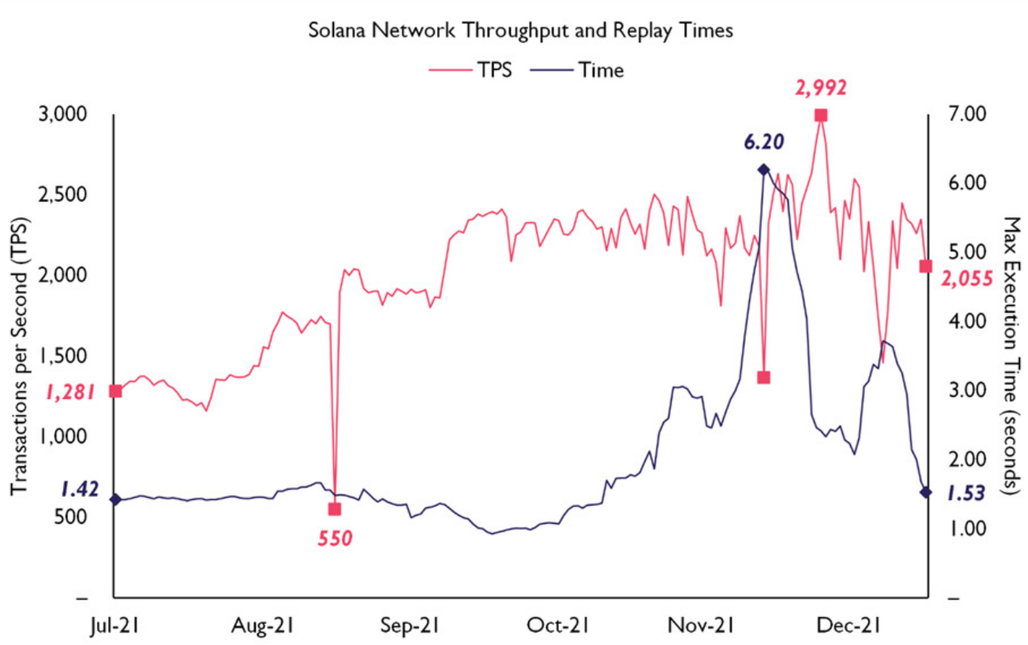
One of Solana's earliest major incidents came in September 2021, when the network experienced an outage that lasted about 17 hours and ultimately required a coordinated restart among validators. The cause of the outage was a sudden influx of resource-heavy bot transactions attempting to capitalize on a hotly-anticipated initial DEX offering (IDO) on the Raydium platform. As explained in the Solana Foundation's post-mortem, validators were essentially unable to process the unexpected spike in transactions, which led to an overload of the forwarder queue for unprocessed packets, followed by a rapid increase in proposed forks, which finally caused validator memory resources to become overwhelmed and crash.

The September 2021 incident was a sign of things to come, as a string of similar issues would go on to plague the network over the next 1.5 years. Not all incidents led to consensus failures, but a common theme that emerged was that validators were becoming overwhelmed by sudden spikes in transaction volumes, whether due to hype from token launches and NFT mints, or simply volatility in the markets, ultimately resulting in degraded performance and decreased throughput.



Sometimes, these incidents could be directly attributed to specific software bugs, highlighting the complex challenge of managing validator resources - which are crucial for the Solana network to function properly - under a variety of market conditions. For instance, the network experienced multiple days of partial outages throughout January 2022, ending the month with ~96% uptime overall. At the time, we noted how one particular bug caused Solana programs to be repeatedly evicted from the software cache, which in turn forced the SVM to recompile these programs, leading to dramatically longer transaction execution times, a.k.a. replay times.

Each of these incidents helped to identify areas where the Solana validator client was not yet fully optimized to combat large increases in network demand. In January 2022, this took the form of liquidator bots flooding the network amidst the early stages of an upcoming bear market. In late April 2022, bots targeting the Metaplex NFT minting program, Candy Machine, overwhelmed the network to the point of forcing yet another coordinated restart. By then, Solana validators and app developers had begun to implement temporary remedies to combat these issues, such as by penalizing repeated transaction submissions and restricting program access to whitelisted accounts during times of abnormally high demand.



Source: The Block Pro Research and RockawayX.

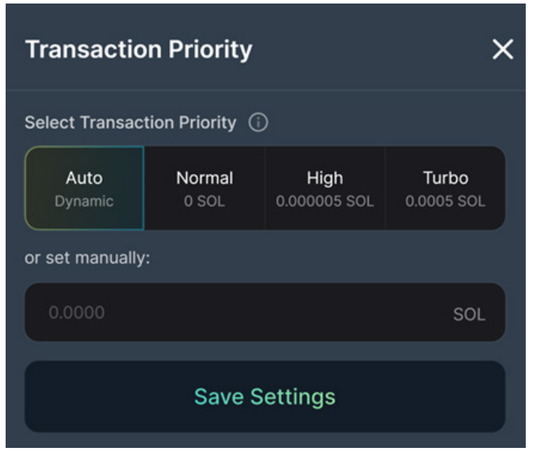
Still, one of the core problems of Solana's design remained largely unaddressed. That is, with trivially low transaction costs and fast confirmation times, there were almost no economic disincentives preventing bots and malicious actors from repeatedly spamming the network to capture potential market opportunities. These behaviors were further amplified by the fact that, as mentioned earlier, Solana validators use size and complexity-limited UDP packets to propagate blocks, which results in a high baseline number of transactions per second even under normal conditions.

After a tumultuous period of network instability throughout H1 2022, Solana developers were finally able to demonstrate notable improvements in validator client performance and resilience after releasing a flurry of key upgrades in June 2022 designed to directly address its historical issues. Chief among these upgrades was the network's initial rollout of the QUIC protocol, which, when integrated with validator TPUs, essentially allows validators to more efficiently filter and sort submitted transactions without significantly impacting throughput. This software implementation would continue to be refined over the next year, but it remains one of the most critical technical upgrades to the Solana network to this day.

Solana's issue with spam transactions has also been addressed with several key features that work more so through economic incentives than software / hardware optimizations. One of these is stake-weighted Quality of Service (QoS), which essentially scales the amount of transactions that can be transmitted from validators to leader nodes based on the size of their stake in the network. Another major feature that has been implemented is fee markets. Similar to those on Ethereum, Solana's fee markets introduce an economic component to regulate and prioritize transactions, allowing users to submit priority fees to increase the likelihood of their transactions being included in the next block.

One major difference is that Solana has implemented local fee markets, which effectively serve to isolate spikes in demand for specific programs without impacting fees and access to other programs on the network. For example, heightened activity surrounding an NFT mint on Solana should, to an extent, only increase fees for that particular program. This requires explicit declarations of access to parts of the network state, which would require a significant design overhaul to accomplish in the EVM, and thus represents a feature of transaction execution that is unique to Solana at the moment. For a more nuanced discussion regarding Solana's design advantages with respect to implementing multi-dimensional fee markets, and the importance of similar functionalities being added to the EVM, see this paper and recent talk by Gauntlet founder Tarun Chitra on the topic.

As illustrated in the chart above showing Solana's network outages, the combination of QUIC, stake-weighted QoS, and fee markets seem to have had a dramatic impact on maintaining network stability. Notably, priority fee usage has continued to rise steadily over the past year as well, suggesting further refinement and adoption of this feature in the future.



Source: Jupiter

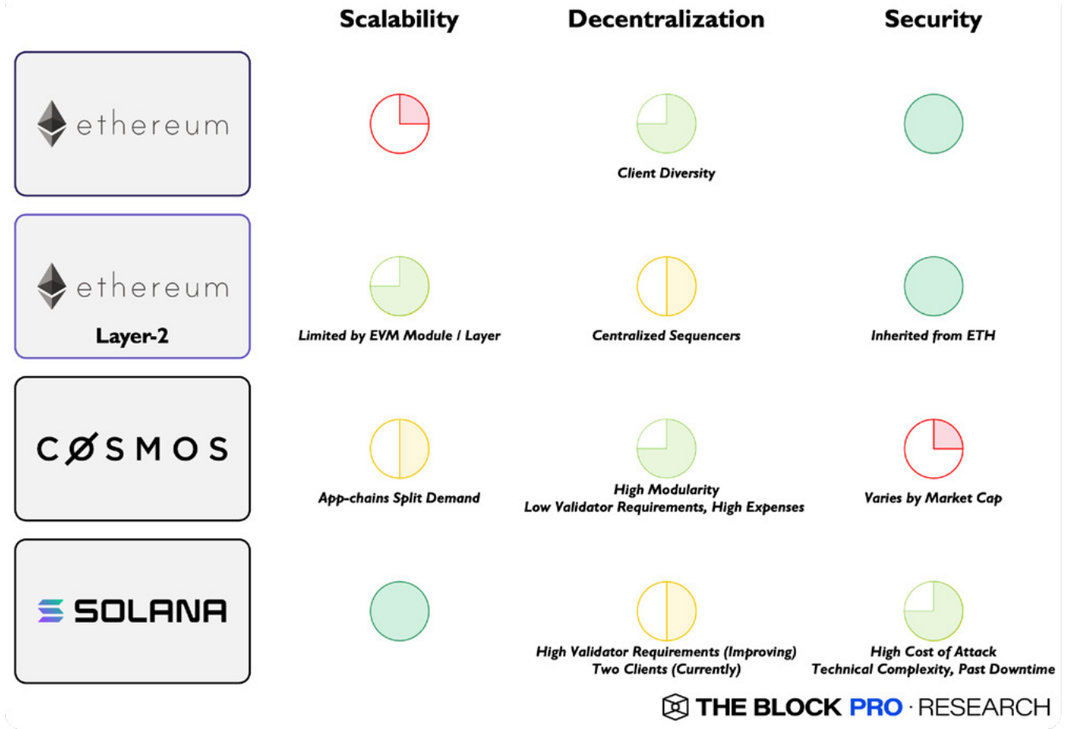
As of this writing, users of the DEX aggregator, Jupiter, can now set priority features through the web interface, simplifying access to Solana's fee markets for average users. As such, it can be expected that the impact of fee markets on network performance will become more pronounced over the next several months. Overall, the timeline of Solana's numerous upgrades over the past year

demonstrates the key challenges involved in scaling a monolithic blockchain, including the optimization of validator performance through computing and networking resource management, as well the balancing of economic incentives for network participants.

PART 4
COMPARING THE CURRENT STATE AND ROADMAPS OF VARIOUS BLOCKCHAIN ARCHITECTURES

> *"The reality is that when it comes to layer ones, there are so many metrics that immediately become absolute nonsense. Once upon a time, TVL was a useful metric and then it very quickly stopped being useful. Because immediately people started gaming it. There is no metric that you can look at that allows you to skip thinking... you have to think very clearly about where a number comes from and what it actually means."*
> – Haseeb Qureshi (Dragonfly Capital)

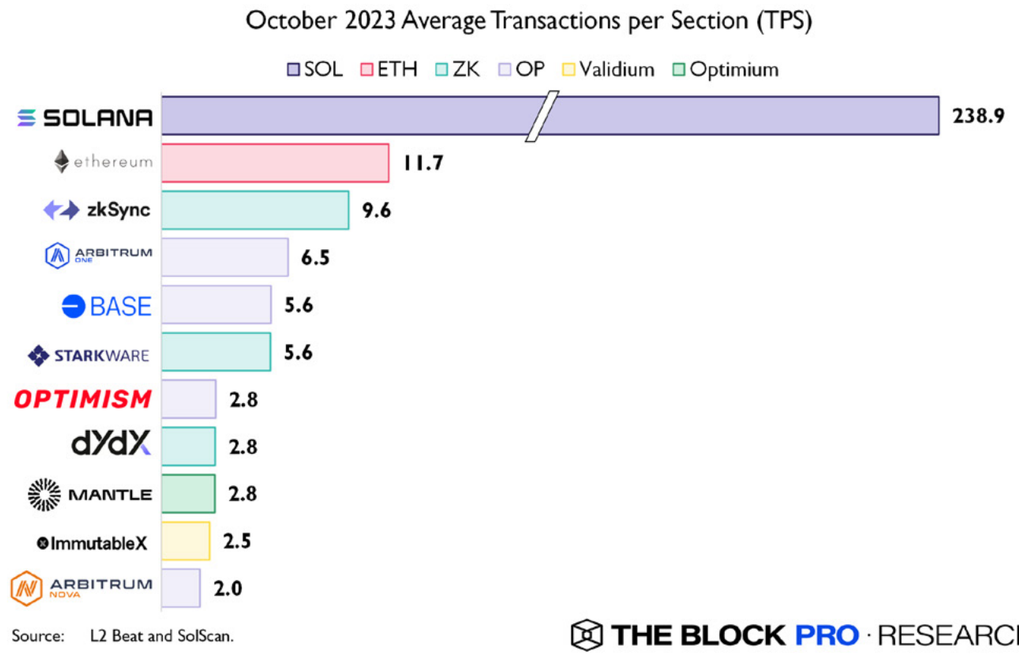
Now that we are familiar with the nuances of Solana's architecture and overall development strategy, let's dig into some of the key differences between Solana's monolithic scaling approach and modular approaches. We focus specifically on the Cosmos ecosystem's multichain scaling approach and Ethereum's rollup-centric approach, which represent two of the main manifestations of modular blockchain architecture. Finally, we take an objective look at the state of blockchain scalability today, and the road ahead from various design perspectives. The current state of each ecosystem along the dimensions of the blockchain trilemma is outlined below.



4.1 THROUGHPUT AND EXECUTION

First, let's take a look at execution, which we established earlier as one of the most important factors dictating blockchain scalability. The Solana team's focus on optimizing transaction processing through engineering has ostensibly given the network an advantage over other blockchains in terms of throughput; the question is by exactly how much? In terms of raw non-voting transactions per second (TPS), Solana processed a daily average of ~239 TPS in October 2023, with a YTD max of ~614 TPS reached on March 29th. Over the same period, Ethereum averaged only ~12 TPS, reaching a YTD max of ~19 TPS on September 13th.

Interestingly, most of the popular rollups today did not feature a higher TPS than Ethereum. In October, zkSync Era, Arbitrum One, and Starknet saw average daily TPS values of ~10, 7, and 6, respectively. Optimism and Base - which is built on the OP stack - had average daily TPS values of ~3 and 6, respectively.



However, it is important to note that TPS is not accurate as a standalone metric in direct comparisons of throughput between chains, though it may appear so on the surface. For one, TPS is directly affected by transaction volume since it is calculated by simply taking the number of transactions divided by the number of seconds over a certain period. Therefore, a chain with more activity than another at a given time will have a higher reported TPS, even if the two chains have the exact same theoretical throughput. The composition and size of transactions can vary between different chains as well. For example, Solana transactions are formatted as UDP packet bundles and have a data limit of 1,232 bytes, whereas Ethereum transactions have a calldata limit of ~1 MB as of EIP-4488.

A more empirical comparison of throughput between blockchains was conducted by Dragonfly Research (mentioned in Part 1), which evaluated throughput by assessing how many DEX trades can be fit into a single block. Notably, in this study, the Dragonfly team found that Solana has a theoretical limit of ~270 swaps per second on the Orca DEX, achieving similar numbers empirically on the Solana devnet as well. The study, conducted in late January 2022, assumes average slot times of ~590ms on mainnet, and a cost of ~74,400 compute units (CU) per Orca swap, given a 48 million CU limit per block and 12 million CU limit per account. As of this writing, the block and account CU limits remain unchanged at 48 million and 12 million, respectively, but slot times have decreased on average since the original study.

For this report, we sought to recreate the Dragonfly study given these updated metrics to gain a sense of how Solana stacks up against other L1s in the present day. Taking a look at epoch 274, roughly when the Dragonfly study was first conducted, we can calculate the average slot time with the following formula:

$$\textbf{Slot Time}(ms) = \frac{Last\ Block\ UNIX\ Timestamp - First\ Block\ Unix\ Timestamp}{Last\ Block\ \# - First\ Block\ \#}$$

$$\textbf{Slot Time}\ (ms) = \frac{1{,}643{,}652{,}202{,}000 - 1{,}643{,}431{,}535{,}000}{118{,}799{,}999 - 118{,}368{,}000} = \sim 510.8ms$$

Average slot times were ~511ms in epoch 274, roughly in line with the figure obtained by the Dragonfly team. We can use the same method to determine average slot times in late 2023, at the time of this writing. Taking epochs 520-524, spanning from October 19 to October 29, we find average epoch slot

times of ~414ms, 419ms, 430ms, 426ms, and 409ms, for an overall average slot time of ~420ms over the five epochs. This alone implies that Solana's throughput in terms of DEX swaps per second has increased substantially since early 2022.

Another change since the original Dragonfly study is that Orca swaps have become more efficient in terms of CU usage since the protocol implemented concentrated liquidity pools in March 2022. Analyzing a random swap transaction (here) from early 2022 that utilizes the older Orca V2 router reveals a cost of ~72,300 CU per swap. Meanwhile, a more recent swap (here) from October 2023 utilizing Orca's V3 "whirlpool" router cost only ~66,500 CUs, representing an efficiency improvement of ~8%. Note that swapping different tokens does not result in different CU for a given smart contract version. Therefore, we can base our analysis on a single random datapoint for each, the old and the new version. The same applies for our below analyses for Uniswap and Osmosis.

It is important to note that these compute optimizations do not constitute a direct gain in efficiency for the Solana network itself, but they do have an impact on empirical measurements of throughput when considering DEX swaps as a standard representation of typical network activity. Orca's reduction of CU costs for swaps can also be viewed as an example of the potential for further resource optimizations arising from Solana applications themselves that have not yet been fully explored to the same extent as on Ethereum. For instance, looking at a recent Uniswap V3 swap (here) reveals a gas cost of ~155,000, essentially unchanged since the Dragonfly study conducted on Uniswap V2 in 2022, which seems to reflect a tapering of significant protocol-based gas optimization on Ethereum in recent years.

Taking into account Solana's 12M per-account CU limits, as well as new values for both slot times (~420ms) and swap cost (~66,500 CUs), we can extrapolate an updated maximum throughput of ~430 swaps per second. At face value, this already represents an order of magnitude higher throughput than Ethereum (without associated rollups), but it also does not factor in the possibility of throughput gains from parallelizable swaps, which could, for example, originate from non-correlated pairs or liquidity pools on other DEXs aside from Orca.

As a whole, Solana's execution system has a clear advantage over competing networks – most of which use the EVM - in terms of transaction throughput. For the sake of completeness, let us also consider performance on Cosmos chains, which were omitted in the Dragonfly study. Osmosis, like most Cosmos chains, uses the CosmWasm execution environment, and is a suitable example for throughput evaluations given its position as the leader in Cosmos ecosystem DEX volumes. Given a block gas limit of ~12.67M, a cost of 430,310 gas per swap (example here), and a ~6.03s block time, Osmosis' throughput can be determined to be ~4.88 swaps per second. Ultimately, the inability of both the EVM and CosmWasm to stack up against the SVM's throughput performance is indicative of the latter's technical strengths from an architectural perspective. Rollups and Cosmos chains attempt to address scalability by essentially adding separate channels. While this approach has of course not yielded any major performance enhancements in terms of execution on a single chain, the hope is to increase throughput of the overall ecosystem via adding chains, or layers.

In some ways, this situation is the result of deliberate design and tradeoff choices, but some teams that have been primarily focused on offering modularity in the past are now starting to recognize the need for direct execution improvements as well. In a discussion with The Block Pro Research team on execution for Cosmos chains, Osmosis founder and original core developer for the Cosmos Hub and IBC, Sunny Agarwal, explained, "There's a lot of low-hanging fruit in the Cosmos SDK stack, in terms of performance engineering, that was never a top focus for us. We were kind of focused on everything else." He commented on a recent announcement from the co-founder of Maker - one of the largest DeFi protocols on Ethereum – that the protocol would be exploring Solana as one of the options for a template of a new blockchain:

> *That was a bit of a wake-up call, for everyone [in the Cosmos ecosystem]. It was like, alright, let's just get our act together on that. I think we can spend six months just fixing all these performance issues on the SDK and get it to be within an order of magnitude of the scalability of Solana. I think Solana's scalability comes from a number of sources – one, it's good engineering, I'll give it to them. They performance engineered it really well. Certain architectural design choices, some of which I agree with, some of which I disagree with. But the beauty of an app chain is you can decide if those design choices make sense for your app chain or not.*
>
> *For example, Solana does not merkelize state – it doesn't have Merkle trees at all. There's no such thing as a light client. And I personally think that's not OK for a public generalized L1 blockchain. But maybe that's OK for an app chain that doesn't care about it, right? So I think the nice thing about the Cosmos idea is like, hey, my chain cares about like client proofs. This chain doesn't. Well, we can do different things, right? So there's certain architectural things Solana does that, some of which we would adopt in Osmosis, some of which we wouldn't.*
>
> *And then there's the higher node requirements. You have to have pretty beefy servers, which, once again, I think that's an app chain level decision right? Like is that OK for a generalized L1? In my opinion, no. But is that OK for a perps exchange, maybe.*

Among teams in the Cosmos ecosystem, it is apparent that transaction execution will be an increasingly important priority in the coming years, especially for a community that has largely focused on enabling developer flexibility in the past. According to Cosmos Hub developer Noam Cohen, one of the key goals for the Hub will be releasing what is known as "Atomic IBC," which would essentially require merging binaries between some of the Cosmos chains that are being secured by the Hub's Interchain Security model and enable them to run in parallel. The idea is that Atomic IBC would create atomicity between chains, thus improving composability and helping to scale the current replicated security model. However, Atomic IBC is not expected to be released until late 2024 or early 2025. Aggarwal agrees that something like Atomic IBC is probably needed in the long run, but there are also easier targets to focus on in the next year or so, explaining, "It's easier for us to just focus on decreasing block times in Cosmos chains. Today the standard is like 5 to 6 seconds. Get this down to one second, sub-second blocks, and then even if you have to do async actions on another chain, it's fine. Half a second here, half a second there, I think that's actually a more productive and achievable goal." In the end, achieving this may require Cosmos validators to take on a greater computational load than they currently do. As we've seen with Solana, ensuring synchronization between validators through PoH is one of the defining characteristics of the network that enable it to consistently produce sub-second slot times.

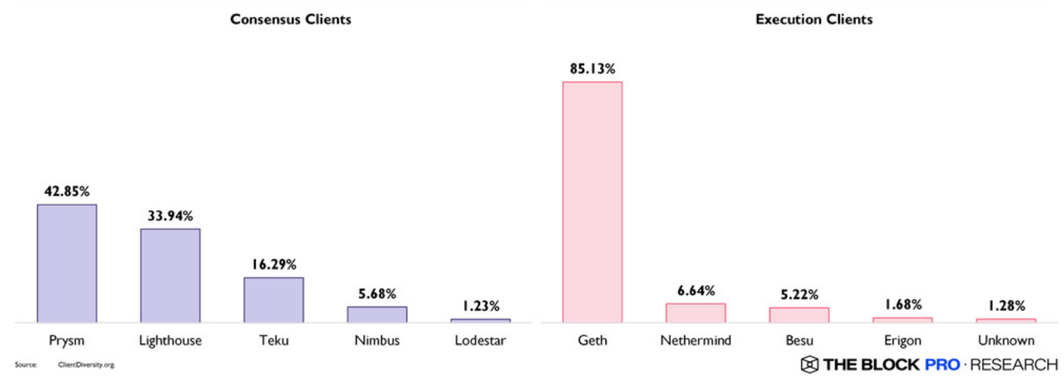
4.2 HARDWARE REQUIREMENTS AND DECENTRALIZATION

Part 3.1 breaks down some of the main challenges with vertical scaling through the lens of Solana. Perhaps the clearest takeaway in addition to our discussion in Part 3.2 thus far is that maximizing transaction throughput comes with a major cost: increased hardware requirements. As such, the barrier to entry becomes much higher for anyone looking to become a validator. This barrier exists in both an economic and technical sense; Solana's validator hardware requirements are not accessible to most people, nor are the skills and time commitment required to keep up with the pace of network developments.

On a purely technical level, the requirements to run an Ethereum - or Cosmos - node are clearly less stringent than those for a Solana node. As such, the total set of users that could potentially join the Ethereum or Cosmos networks as validators is higher compared to Solana. In theory, Solana's node requirements thus impact its overall decentralization, at least relative to Ethereum and Cosmos chains as baselines for comparison. However, decentralization itself is difficult to measure objectively. One could simply count the number of active nodes in a network, but this alone is not a useful metric without knowing how many nodes are controlled by a single entity.

Another, somewhat more insightful, metric is the number of nodes required to disrupt a network, commonly known as the Nakamoto Coefficient. The premise behind this metric is that the higher the coefficient, the harder it is to corrupt a critical mass of nodes, and thus the more "decentralized" a particular network is. Still, the Nakamato Coefficient does not fully account for key architectural variations between different networks, and thus fails to sufficiently capture the nuances involved in assessing decentralization. Even defining the nodes for comparison between blockchains can be tricky. For instance, on Ethereum, anyone can run a node - which enables user interaction with the network - for just the cost of the hardware to do so. However, only validators who stake significant capital (32 ETH)[1] are able to propose new blocks. This in turn creates a higher barrier to entry and effectively decreases the number of corrupted entities required to disrupt the network.

Another factor to consider is client diversity. Assuming a network uses the same client for all of its validators, this would mean that the cost of censoring the network would effectively boil down to a single, centralized point of failure. On Ethereum, this is addressed by having several clients for both consensus and execution.



Still, this diversity is best viewed as a spectrum. On the consensus side, Prysm and Lighthouse comprise ~77% of all clients, while ~85% of execution clients run Geth. Solana co-founder Anatoly Yakovenko has a specific view on decentralization. According to Yakovenko, the real scalability trilemma consists of three

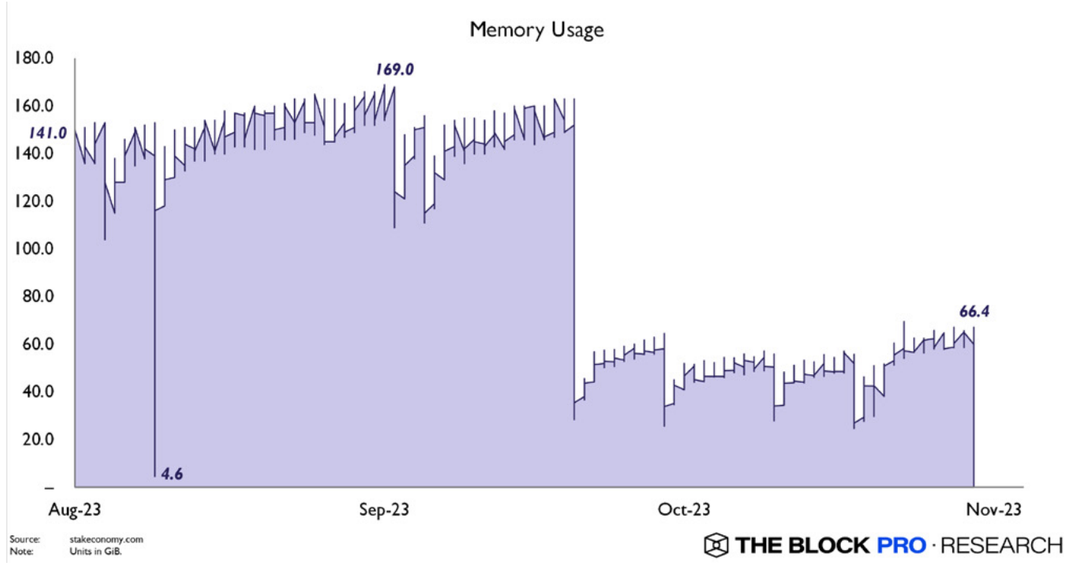
1 Ethereum node operators who use Rocketpool only need to put 8 ETH and 2.4 ETH worth of RPL at stake, theoretically lowering their individual cost. Nonetheless, these operators still need access to 32 ETH overall, with contributions from users, essentially spreading out the cost burden of staking through Rocketpool's coordination mechanism.

things: 1) cost to destroy all replicas, 2) cost to censor a message from arriving to all replicas in real time, and 3) cost of hardware. In his view, Solana has chosen to optimize on 1 and 2, while Ethereum has optimized for 1 and 3.

Client diversity is certainly a pending issue for the Solana network today. As of this writing, there are only two major validator clients live on Solana – the Solana Labs client, and Jito Labs' MEV-optimized fork of the Solana Labs client. This is to be expected in some ways – the pace of network upgrades originating through the Solana Labs team over the past year has likely made it practically difficult to develop an alternate client without significant coordination with Solana Labs. The most significant initiative currently underway to address this issue is the Firedancer validator client, an alternative client being developed by Jump Trading that aims to further maximize throughput on Solana. Yakovenko stressed the importance of Firedancer in a recent discussion with The Block Pro Research, stating, "I will finally consider Solana to be out of beta when Firedancer launches." This future may not be all too far off – at the Solana Breakpoint conference in late October, the Solana Foundation announced that Firedancer is now officially live on testnet.

There are additional existing efforts to increase the overall decentralization of the Solana network. One of the most significant is the Solana Foundation's Delegation Program, which delegates staked SOL controlled by the Foundation to smaller validators if they are able to meet the comprehensive list of requirements that would indicate reliability over the long term. The Foundation also runs the Server Program, which facilitates easier access to validator and RPC hardware through contracts with data centers around the world.

It is important to note that the issue of decentralization can be addressed from a technical standpoint as well. For instance, optimizing resource usage with existing specs is one of the main components of improving scalability. Over time this would lower the cost of running validators (Moore's law) and hence allow more people to run a node. This was clearly demonstrated with the release of Solana's v1.16 upgrade in September, which appears to have drastically reduced validator memory requirements in the process.



Using data collected from the Stakeconomy validator, we can see that average daily memory usage was ~132 GB in September, but has dropped to a daily average ~51 GB since the release of v1.16. In other words, increased scalability and throughput does not always translate to higher hardware requirements automatically, especially if optimizations for hardware resources are made on the software client side. For blockchains aside from Solana, there is already a non-trivial possibility that validator hardware requirements will eventually need to increase in order to support better execution. According to Osmosis's Aggarwal, it is likely that Cosmos validators will need to expand their responsibilities in the coming months:

> *I think over the next six months, you're going to start to see things that cause the SDK to become way more performant. One of the big things we're focusing on is having a lot of sidecars. What I mean by that is like – having the validators of a chain be able to run secondary things other than the nodes that, for example, basically offload computation. For those nodes, the computation basically has to happen on the state machine, which means every full node has to do it.*
>
> *But if you offload some of it to a validator sidecar and only have validators run that computation, then you can have that submitted as a proof – things like that, I think will also be a big focus of both - how do we increase performance, but also how do we decentralize the Cosmos stack? Today for our DEX, what we think a lot about is like, OK, our chain is decentralized, but the front end is centralized, the routing server is centralized, the data indexers are centralized, and it's like - how do we decentralize everything? Well, let's get our validators to run more and more of this.*

In this age of increasingly complex and demanding user activity on blockchain networks, it is not entirely far-fetched to expect that validators will soon need to evolve to meet these growing demands. The main question that remains is whether these validator requirements can be kept low enough over time to provide higher performance without drastically sacrificing decentralization or security in both the short and long term.

4.3 ONWARD AND UPWARD: THE STATE OF BLOCKCHAIN SCALING AND BEYOND

In an overarching sense, the contentious debate regarding the "best" approach to scaling blockchains today boils down to opposing viewpoints regarding the importance of various blockchain properties at different points in time. The key problem is that in reality, all of the properties that define a blockchain – scalability, decentralization, security, interoperability, usability, etc. – should be considered important at any given time. A network that focuses solely on scalability in the short term might quickly become so centralized that it becomes wholly unable to resist censorship in the future. Similarly, a network that maximizes security at all costs without developing short-term avenues for scaling bears the risk of becoming too expensive or unusable for a new generation of users, threatening its relevance in the long run as well. Meanwhile, any network that doesn't prioritize security at all times must accept the potential for devastating failure and the unrecoverable loss of user funds. One of the clearest examples of this risk is Terra Luna, whose fundamentally flawed economic model led the network's value to plummet to near zero in May 2022, resulting in the destruction of billions of dollars worth of value in the process.

One property of blockchains that has become increasingly relevant in recent years is interoperability, largely due to the rapid proliferation of a growing number of blockchains, whether it be monolithic L1s, multichain monolithic L1s, rollups, or any kind of function-specific chains for tasks like data storage or data availability. As such, this topic is also central to our discussion of the current and future blockchain scaling landscape. In some ways, the main paradox of interoperability is that most blockchains have a strong incentive to attract users, developers, and capital to their respective chains. This can be critical for maintaining economic security, and for ensuring continued growth and development of a particular platform. Therefore, from the perspective of some teams, there may be little benefit to maximizing interoperability with other chains if the long-term goal is to, say, become the de facto settlement layer in crypto.

At worst, tying the value of user capital on one chain to another chain or cross-chain communication protocol can potentially create a single point of failure with catastrophic consequences. This has been demonstrated multiple times throughout crypto history. In instances like the Wormhole exploit in 2022 or the Multichain collapse in 2023, the users that were primarily affected were those that sent their funds from Ethereum to use another network – in this case, Solana and Fantom, respectively. Solana users that held assets backed by the Sollet bridge had the additional misfortune of their assets quickly losing value during the collapse of FTX and Alameda in November 2022. To further exhaust this point, the Osmosis network itself was driven to the brink of liveness failure in May 2022 when staked liquidity pools comprising its native OSMO token and deteriorating Terra assets led to a near-uncontrolled downward spiral in the value of OSMO.

In light of all these factors, the fact remains nonetheless that establishing some level of interoperability with other chains is a basic requirement for attracting user activity and developer mindshare today. From this perspective, the Cosmos ecosystem today is perhaps one of the most focused in terms of leveraging interoperability as a baseline for development. This is evident from the ecosystem's embrace of IBC as a communication standard between chains, as well as the emergence of chains like Axelar and Celestia that primarily intend to service other blockchains (i.e. through bridging and data availability). As explained by the Cosmos Hub's Noam Cohen, "The goal is for IBC to become the communication standard between all blockchains."

At a high level, interoperability can be more precisely viewed as a key component of a horizontal scaling approach overall. The Cosmos ecosystem mostly utilizes interoperability via the application layer, relying on IBC to enable cross-chain capital flow and inter-application integrations. One could also loosely define this as composability. At the same time, modular design approaches embodied by Ethereum rollups and data availability layers utilize interoperability primarily from a security perspective, with a focus of extending security from the base layer to its rollups. This brings us back to the heart of our primary mental model for evaluating blockchain design choices and scaling approaches: horizontal vs. vertical scaling.

This concept has been described in detail by Multichain Capital's Kyle Samani, whose thesis on modular vs. integrated systems argues that modular systems suffer from a number of hidden, non-obvious costs to users and developers. Samani posits that increasing modularity in blockchain systems ultimately results in more complexity for application developers, who are effectively forced to contend with growing levels of fragmentation – both in terms of ensuring interoperability between modules and establishing social

consensus. At the same time, modular systems can create fragmented liquidity between chains (as in the Cosmos ecosystem) without addressing the problem of cross-app congestion. The example he cites is that surges in activity and transaction fees on an L1 like Ethereum ultimately result in higher fees for associated rollups as well, who must pay fees to the L1 in the process of posting fraud proofs. Speaking with The Block Pro Research, Samani clarified:

> *Modular systems give developers the illusion of more control, which is supposed to be a good thing, but in practice create far more problems than solutions. There are some instances in which flexibility and control are important, but those are few and far between—the exceptions, not the norm. In practice, the vast majority of developers only want to focus on the application layer. They don't want to maintain a dedicated chain, manage bridging or wrapped assets, fragmented liquidity, or any other cross-chain dependencies. Integrated blockchains give them just that: a platform to launch robust, scalable applications that centralize liquidity and abstract away the complexity of chain management.*

Osmosis's Aggarwal has a similar view on the concept of modularity, which he further breaks down into internal and external dimensions, as well as horizontal and vertical ones. Aggarwal argues that Celestia's definition of modularity is more accurately described as external modularity, wherein all blockchain functions are performed by different, distinct teams. This is also analogous to the structure of the Cosmos ecosystem as a whole, wherein various teams focus on developing applications on different chains, embodying a sort of horizontal integration approach overall. Aggarwal believes in more of a vertical integration approach, wherein teams aim to perform as many functions of its core product as possible in order to reduce costs and security vulnerabilities that can arise from outsourcing and communication barriers across organizations. He reasons that outsourcing should only be done in very select, strategic cases – as Osmosis does with bridge providers like Axelar. Importantly, this approach does not reject the concept of modularity as a whole, as components within a project (like Osmosis) or blockchain stack (like the Cosmos SDK) can still remain modular in order to support customizability and future optimization.

These definitions of vertical integration/scaling and internal modularity are perhaps the most accurate way to envision the Solana network's design as a whole. On the one hand, Solana is vertically integrated to a significant degree, relying on significant coordination between hardware and software components just to execute transactions and achieve consensus. On the other hand, individual parts of Solana's architecture can still be optimized to improve overall function, such as by updating software to reduce validator memory requirements.

One of the main challenges with a vertical scaling approach is a lack of interoperability with other ecosystems, and thus a reduced ability to accrue developer mindshare from a wide variety of sources. As we have seen with the EVM, the popularity of Ethereum's execution environment as a standard has enabled growth and DeFi innovation on many other chains that can eventually trickle back to Ethereum. Lucas Bruder, founder of Jito Labs, acknowledges that developer tooling is still lacking on Solana and remains more comprehensive on Ethereum. Popular tools such as the Etherscan block explorer and Debank on-chain portfolio tracker have functional counterparts that exist on Solana, such as Solscan and Step Finance, but the Solana versions are largely unable to benefit from the extensive usage across multiple EVM chains and resultant developer activity that naturally arises from the Ethereum-based tools.

Even so, the development of projects like Firedancer specifically for the Solana ecosystem by third-party teams shows that instances of external modularity are still possible for a mostly vertically integrated system like Solana. In the future, the deployment of cross-chain protocols such as the Eclipse SVM chain for Ethereum and the Neon EVM chain for Solana can also be expected to attract additional attention and developer mindshare to the Solana ecosystem. Forthcoming native asset integrations such as Circle's CCTP for USDC can be potentially viewed as a source of increased economic security for Solana as well. When it comes to composability overall, Solana's Yakovenko argues that composability among Solana DeFi protocols is even more robust than in other ecosystems that scale horizontally, due to the fact that access to state on Solana is more clearly defined.

Ultimately, as we mentioned at the beginning of this section, evaluating modern blockchains on the basis of scalability comes down to one's viewpoint on the importance of individual blockchain properties within different timeframes. Regarding the current and future state of rollups, Ethereum researcher Justin Drake explains:

> *If you're very short-sighted, then yes, the vast majority of these rollups are not going to be as secure as the L1. Basically, there's a period of transition when a lot of activity is moving to the L2s, and there's still a lot of work to make it as secure as the L1. But, as a researcher, my interest is in the end game. And I think with a little bit of vision, and a little bit of imagination, you can see that these downsides of the rollups will eventually go away. In terms of the exact timeline, it's hard to tell. My guess is that within a few years, maybe 1-2 years, we'll be in a position where the vast majority of activity is on extremely secure rollups.*

> *And not only will the security of the rollups improve, but the scalability of the rollups will dramatically improve. Partly through what's called Proto-Danksharding, which is an upgrade coming to Ethereum next year, as well as Danksharding, the total amount of data availability will increase. And so, things will be moving extremely fast in the next two years in terms of both security and scalability.*

Yakovenko also shares the viewpoint that timeframe is important when thinking about scalability in terms of blockchain design. However, his interpretation differs in that he believes that throughput gains from hardware improvements will naturally outpace those arising from innovations in blockchain-specific software implementations. He states:

> *Let's say you build some software today and you've got a really good complicated algorithm that makes it 80% faster - some huge improvement in performance and I don't do that. I just simply use hardware and I kind of get it out the door and I'm done. Your software implementation is faster, but every two years, the hardware gives me twice as much capacity and that marginal difference drops by 50% between your software implementation that is harder to scale. And with hardware that difference drops and drops and gets reduced and reduced, right? Eventually mine surpasses it pretty quickly.*
>
> *So, what you actually see is that if you take the complex route and try to squeeze out performance out of jumping through protocol hoops, versus let's just make sure every time bandwidth doubles, it gets faster. By the same amount you're trying to do this short-term improvement, you're losing out on the easy, long term improvements.*
>
> *In the very long term, we're talking about 20 year time span, you need to build in such a way that you just naturally take advantage of hardware. If your design does not naturally scale in every parameter including state size, bandwidth, compute, you've got to think about all those things because all those things get a thousand times bigger and cheaper and faster. You're massively going to be behind. This is why Solana has been built this way, because we actually want to do less work. We want to have a small engineering team that does these optimizations. And then all those scale out because validators can go and get a bigger box on Amazon over the weekend without talking to us. And that's all the work that they need to do to scale the network. And we just try to think six months ahead and make sure that the software can take advantage of their hardware.*

In essence, what we have uncovered throughout this report is that modern blockchain designs vary dramatically, particularly when it comes to scaling approach. The most popular blockchain on the market today in terms of user capital, Ethereum, has embodied a modular approach along the lines of horizontal scaling by focusing on ensuring security and interoperability with emerging rollup solutions, without increasing hardware requirements. Cosmos chains have historically taken a fully horizontal scaling approach and focused on enhancing cross-chain communication, but a general lack of improvements in single-chain throughput have started to drive more teams to consider scalability more so from an execution perspective in recent years as well. Finally, through a process of consistent optimization of software to meet existing hardware capabilities and user demand, the Solana network has grown to become an insightful example of the scaling benefits that can be derived from a vertically integrated approach to blockchain design. Over the coming years, with user adoption of blockchains expected to eventually return – and even exceed – previous all-time highs, it is expected that the long-term feasibility of these varying approaches will become dramatically clearer, and thus inform the next generation of blockchain development.

CONCLUSION

Modern blockchains have grown to become highly sophisticated, dynamic systems, featuring capabilities that far exceed those of early blockchains like Bitcoin. Smart contract platforms now dominate the field of blockchain innovation, providing the foundation for an ever-expanding range of on-chain financial activity. Just to maintain these systems in a steady state condition can be a challenging task, requiring extensive social and economic coordination between key stakeholders including validators, developers, and users. Meanwhile, as DeFi ecosystems mature and new avenues for block space demand continue to emerge, it has become clearer than ever that blockchains must continue to evolve in order to become scalable enough to meet the demands of the future.

In this pursuit of greater and greater scalability, the paths that various blockchains have taken to achieve this end continue to diverge. Some blockchain teams have begun to adopt newer, supplementary blockchain protocols to enhance their overall technology stack, while others have opted to maximize the capabilities of the core technologies within their existing architectures. Others have begun to converge between disparate philosophies in an attempt to reach a functional middle ground between scalability, decentralization, security, interoperability, and other important blockchain properties. All the while, the underlying challenge of preserving these critical properties remains ever-present, forcing teams to contend with the potential for existential risks in both the short and long term. The result of this environment is that it has become more difficult than ever to properly evaluate the state of blockchain development and the feasibility of different scaling approaches.

In this report, we broke down the key components that make up the leading smart contract platforms today in order to shed light on the current and future state of blockchain development. In Part 1, we identified the key functions for which blockchains are responsible, and provided an objective framework to classify blockchains based on these functions. In Part 2, we applied a popular mental model for evaluating blockchains – the scalability trilemma – and explored how the optimization of individual blockchain functions pulls on the various levers of the trilemma. We also discussed the various tradeoffs that begin to emerge when piecing together systems that are designed for specific functions, using case studies of two prominent blockchain architectures, modular and multichain, to highlight some of the existing advantages and challenges that have become more apparent in recent years.

In Part 3, we conducted a deep dive into a third architecture, monolithic, through the lens of Solana, following its developmental journey to overcome key technical hurdles in scalability over the past few years.

Finally, in Part 4, we conducted a broad comparison of the main scaling approaches today, applying a new mental model – horizontal vs. vertical scaling – to uncover their respective strengths and weaknesses. On one end of the spectrum, we have Cosmos and Ethereum following a horizontal scaling approach, while on the other end, we have Solana taking a vertical scaling approach. This has clearly led to different outcomes in terms of both performance and user experience. There is little doubt that today, Solana features the fastest execution environment, while the market considers Ethereum to be the most secure chain. Looking further into the future, both are taking different paths to converge upon the same goal: to provide a secure and fast smart contract platform for the masses. Ultimately, the market will decide which models have the best product market fit for the growing digital asset ecosystem. At the moment, it appears the pie is growing so fast that both approaches – and solutions in between – will have their opportunity to carve out a spot in a multichain future.

DISCLOSURES

This report is sponsored by the Solana Foundation. The content of this report contains views and opinions expressed by The Block's analysts which are solely their own opinions, and do not necessarily reflect the opinions of The Block or the organization that commissioned the report. The Block's analysts may have holdings in the assets discussed in this report and this statement is to disclose any perceived conflict of interest. Please refer to The Block's Financial Disclosures page for author holdings.

Beginning in 2021, Michael McCaffrey, the former CEO and majority owner of The Block, took a series of loans from founder and former FTX and Alameda CEO Sam Bankman-Fried. McCaffrey resigned from the company in December 2022 after failing to disclose those transactions.

This report is for informational purposes only and should not be relied upon as a basis for investment decisions, nor is it offered or intended to be used as legal, tax, investment, financial or other advice. You should conduct your own research and consult independent counsel on the matters discussed within this report. Past performance of any asset is not indicative of future results.