

Ephemeral Rollups are All you Need

Gabriele Picco, Andrea Fortugno

www.magicblock.gg

Abstract

In the realm of open and composable gaming, we envision platforms where users actively expand, create, engage, and immerse themselves in a rich world of entertainment. One promising avenue for achieving this vision is through fully on-chain (FOC) games, where both game state and logic reside on the blockchain, maximizing composability. However, we must grapple with inherent limitations and trade-offs, particularly in terms of costs and scalability. This paper proposes BOLT, a framework that leverages the Solana Virtual Machine (SVM) to scale FOC games without state fragmentation or compromised trust assumptions. The framework introduces a systematic approach for discovering, utilizing, and publishing modular pieces of logic as components deeply rooted in the Entity-Component-System (ECS) pattern. To enhance scalability and resource optimization, we introduce the concept of Ephemeral Rollups (ERs) that overcome the tradeoffs of L2s horizontal scaling. These dedicated runtimes can be customized to provide higher operational speed, configurable ticking mechanisms, provable sessions and gasless transactions without composability-scalability tradeoffs.

1 Introduction

Fully on-chain (FOC) games intertwine digital experiences with decentralized immutable ledgers. Any participant of a FOC game can integrate additional content, logic or modifications without explicit consent from the original creator ("permissionless modding") or entertain with the creation of "Autonomous Worlds", unstoppable digital Worlds [19]. Thanks to smart contracts - which enforce the substrate of these virtual realities - it is impossible to unplug them like traditional game servers. Once deployed, a FOC game will run forever. This degree of creativity and persistency introduces a whole new level of significance to games, which become avenues to play, socialize, work, and conduct commerce in an even more pervasive and profound fashion.

In accordance with the definitions given in [19], [2], [14], our perspective holds that for a game to be classified as fully-on-chain, it should inherently possess these crucial attributes:

- **Blockchain as single source of truth:** both game state and logic must reside on the blockchain. It serves as a transparent data store, guaranteeing the permanence of all meaningful data and allowing trustless execution of logic. The game is client-agnostic.
- **Permissionless Composability:** the game can accommodate new extensions, logic and/or components without the need for explicit permission from the original creator. Trustless computation enabled by smart contracts ensures unbounded access for any enhancement ("everything is a public API"). Even though not every plug-in, extension or third-party client will necessarily be used, everyone has a chance to contribute and add their modifications to it.
- **Persistency:** The game should demonstrate resilience, with no single point of failure. Even though servers or centralized components may serve to augment performances or playability, the game/world should hold the potential to progress without their presence, thereby ensuring the continuous existence of said digital reality.

Our conviction is that FOC games will become the key drivers for the inception of innovative platforms and self-reliant ecosystems. These digital worlds will evolve according to the will of consumers and online communities aligned in this creative endeavour. This concept holds the potential to transcend gaming and blur the line between the virtual and the physical world, seamlessly blending the two.

1.1 Game Logic, State, and Transactions

In a fully on-chain game, the game logic resides within smart contracts deployed on the blockchain, ensuring transparency and immutability. It comprises the rules, player interactions, and other essential behaviours that define the gaming experience.

The game state, on the other hand, refers to the real-time status and attributes of the game elements, which are stored on-chain, providing a consistent and synchronized environment for all players. It reflects the game's current situation, from player positions to assets and scores.

Transactions in this context denote the interactions and exchanges occurring within the game. Players' actions, asset transfers, and other in-game activities are state transitions recorded on-chain, ensuring their integrity and permanent record. The blockchain's decentralization makes these interaction transparent, verifiable, and immutable.

1.2 Outline

In the upcoming sections, 2 and 4, we introduce BOLT, a performant and scalable SVM-based framework for FOC games and Autonomous Worlds. We dive into some of key features, notably its scalability and the efficient re-use of resources and on-chain components. For an overview of the emerging gaming frameworks and their limitations refer to Appendix A. For an overview of the commonly adopted scalability solutions and their limitations and tradeoffs refer to Appendix B.

2 Framework Architecture

Developing multiplayer games is a complex endeavour in conventional settings. FOC games further intensify these challenges as we must adhere to registering every interaction on-chain, a limiting factor both in terms of scalability and costs. These limitations primarily arise from its decentralized nature and the consensus mechanisms employed.

Considering a straightforward scenario, a game with 1000 players, with each player’s position updating every 100 ms, results in 10,000 transactions per second. Scaling this to 100,000 players, the rate is leaping to 1,000,000 transactions per second — a volume no blockchain, at the time of writing, can support. Even for a high-speed blockchain like Solana, which might manage such traffic in the future, the demands quickly become insurmountable when considering multiple games and higher numbers of players. From a cost perspective, Solana’s localized fee market does alleviate some concerns by preventing a single game from affecting fee costs for other games (“noisy neighbor” problem). However, even if the individual transaction fees are low, the cumulative cost of such a vast number of transactions can become untenable for a single fee payer (i.e., the game developer).

2.1 SVM runtime

The Scalability in Solana is intrinsically linked to its unique method of state storage and management, distinguishing it from other virtual machines like the EVM. In Solana, everything is an account. Accounts can hold any data, as well as native SOL tokens.

Broadly, accounts fall into two categories:

- Executable (program accounts): These are akin to smart contracts, housing code and often referred to as “programs”
- Non-executable (data accounts): These can store tokens or data but lack the capability to execute code.

Solana’s account model inherently grants programs the privilege to create and oversee specific accounts. This allows developers to create custom rules and logic for their governance. These special accounts are called Program Derived

Addresses (PDAs) [27]. This mechanism empowers programs to endorse various on-chain actions on a PDA’s behalf, ensuring the Solana network recognizes and validates these actions without needing a private key [28].

Another core feature of the Solana Virtual Machine is that it parallelizes transaction processing. It identifies which parts of a transaction are non-overlapping and can be processed in parallel, maximizing hardware utilization. Unlike other blockchains that use a merkle tree to keep track of the global state and process transactions sequentially, Solana’s Sealevel executes them concurrently, significantly boosting the network’s throughput and scalability.

2.2 Ephemeral Rollups

The core intuition is that by harnessing the SVM’s account structure and its capacity for parallelization, we can split the app/game state into clusters. Users can lock one or multiple accounts to temporarily transfer the state to an auxiliary layer, which we define as the “ephemeral rollup”, a configurable dedicated runtime. This process temporarily entrusts the sequencer to modify the accounts in the ephemeral rollup, where the state can be forcefully reverted and unlocked on the L1 if constraints are not satisfied (see section 2.3). Despite this delegation, operations and transactions can still use the delegate accounts as readable on the base layer. Non-delegated accounts remain unaffected and modifiable. The ephemeral rollup operates as a specialized SVM runtime to facilitate transaction processing at an elevated throughput. Additionally, this specialized runtime can be customized to include configurations like gasless transactions, quicker block-time, and the inclusion of a ticking mechanism (i.e., an integrated transaction scheduling system like clockwork [3] operated without fees). The entire process is transparent to the end user - a specialized RPC provider can route the transactions to the base layer and the ephemeral rollup(s) in parallel during the game session.

Figure 1 provides an overview of the system. The game program defines the game’s logic and mechanisms. Its state consists of various accounts, such as PDAs representing the positions of P1, P2, and a Chest designated for distributing rewards. The game program leverages BOLT logic and tools, for instance, an off-the-shelf movement or an energy system.

The steps involved in provisioning and utilizing the ER are:

1. The game program communicates with the BOLT ER program, initiating a transaction that requests an ephemeral rollup provision. This request specifies the ER’s configuration details, such as its lifetime, base layer update frequency, targeted transactions per second, and block time.
2. Monitoring the BOLT ER program, the provisioner detects the provisioning request and promptly launches the corresponding runtime based on the configuration.

3. The game program delegates accounts to the BOLT ER, granting it the authority for state updating and settlement. From now on, the delegated accounts can only be updated on the ER. This third step can even occur in parallel with, or prior to, the second step.
4. Clients wanting to execute transactions or retrieve data via RPC send their transactions to an RPC router. Depending on the accounts involved in the transactions, this router then forwards the transactions to the appropriate chain, either on the base layer or one of the ERs. A comprehensive discussion of the routing mechanism is available in 2.2.2.
5. Periodically, or upon ER termination, the sequencer updates the state on the base layer.
6. Once the ER concludes (or if the game program forces its closure), account control reverts to the owner program and the provisioner terminates the ER.

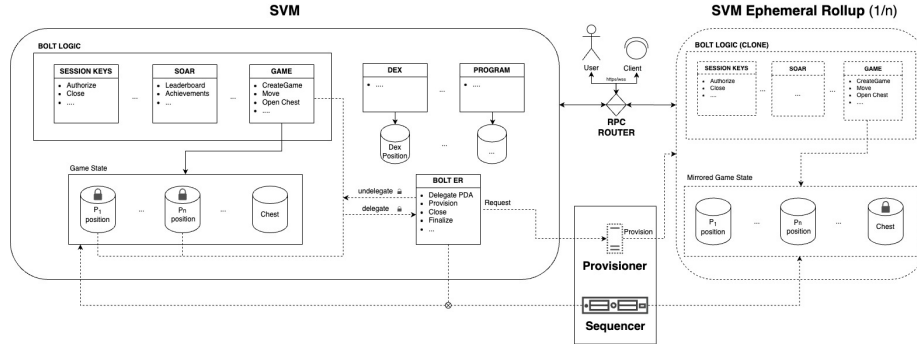


Figure 1: The figure showcases the architecture of the BOLT scalability solution, focusing on how a game program manages its state across multiple accounts. The example presents two PDAs, each responsible for recording the position of Player 1 and Player 2 respectively. These PDAs are then delegated to the BOLT ER (Ephemeral Rollup) program. In the background, the Provisioner constantly observes the BOLT ER program, watching for provisioning requests. Upon detection, it dynamically launches an SVM runtime, adapting to specific configuration parameters such as block time and ticking, among others. Within the Ephemeral Rollup (ER) framework, transactions using these delegated accounts are accelerated. The sequencer is the component responsible for updating and finalizing the states of these accounts on the L1.

To illustrate with a practical example, consider Figure 1. A game program might reward players with an NFT reward upon reaching a specified area on a map. PDAs, tracking players' positions, can be frequently updated within the ER, facilitating a low-latency multiplayer session among players in the ER. When a player reaches the designated map location that triggers a reward, according to the game logic, the client can submit the transaction in the usual manner. If the sequencer doesn't update rapidly enough, one might first submit an optional state settlement request. Behind the scenes, the Remote Procedure Call (RPC)

router relays this transaction to the base layer. Operating on the base layer, the game program can accurately process this transaction and dispense the reward. This is because the position account, which is read-only, reflects the updated location, and modifications can be made to the Chest PDA which wasn't delegated to the BOLT ER program.

2.2.1 Avoiding Fragmentation and Benefits

The benefit of ER is that programs and assets reside directly on the base layer. Transactions can be accelerated through ERs, which are fully compatible with the Solana Virtual Machine (SVM) down to the bytecode level. Any improvements or advancements at the base layer are immediately available, without the need to modify or re-deploy programs.

Ephemeral Rollups has the following benefits:

- Developers deploy programs to the base layer (e.g., Solana), rather than on a separate chain as it would normally happen with rollups. Programs live on the base layer and can interact with any existing protocol and assets. ERs don't fragment the existing ecosystem and allow the speed-up of targeted operations without creating an isolated environment.
- Users, developers and programs using ERs can take advantage of Solana's infrastructure. This includes programming languages, code libraries, testing tools, client software, deployment infrastructure etc.
- The specialized runtime can accommodate game-specific customizations (e.g., ticking or passive events, typical in games, as opposed to the event-driven runtime of blockchains) without the need to pay gas for fees on those specific accounts.
- This approach enables a highly scalable system capable of launching rollups on-demand, auto-scaling horizontally to accommodate users performing millions of transactions without the tradeoffs of traditional L2s.

BOLT allows FOC games to scale without compromising composability on the base layer.

2.2.2 Reading and Modifying the state

As highlighted in the preceding sections, the state is maintained in accounts. These accounts can concurrently exist on the base layer and within an ephemeral rollup (ER). While accounts are universally readable from both layers, they can be modified either from the base layer or the ER. It's important to note that if a user or client seeks to access data from an account, the most recent or updated data might reside in the ER, if it hasn't been settled on the base layer yet. Consider a scenario where a game client employs a WebSockets connection to display other players on a map, and the position PDAs are delegated to the ER. In such a case, a WebSockets connection to the ER would allow observation

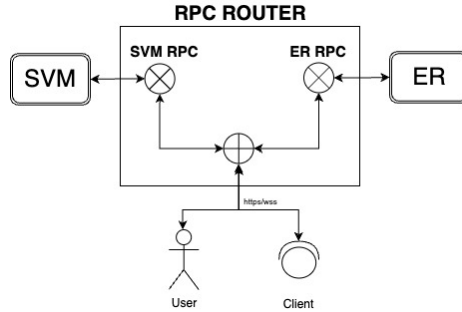


Figure 2: The RPC router module routes both HTTPS and WebSocket requests either to the base layer RPC or to the Ephemeral Rollup RPC, based on the accounts being accessed or utilized.

and streaming of transactions with a latency and block time comparable to a traditional multiplayer game server, typically between 10 to 100 milliseconds. The added complexity of determining where to submit transactions and where to retrieve data from can be abstracted using an RPC router, as depicted in 2. Clients can subscribe and send transactions to the RPC router as they would to a traditional RPC. The RPC Router then directs transactions and connections either to the base layer or to the ER RPC, depending on the accounts involved.

Considering the following scenarios:

For Reading the state:

- If the request accesses accounts delegated to an ER, the RPC router relays the requests to the ER RPC.
- Otherwise, the base layer RPC processes the request.

For Sending transactions:

- If the transaction contains readable accounts from the base/ER layer and all writable accounts are delegated to the ER, the transaction is relayed to the ER RPC.
- If the transaction includes readable accounts from the base/ER layer and no writable accounts are delegated, the transaction is relayed to the base layer RPC.
- A challenging scenario arises when a transaction aims to modify accounts present in both the ER and the base layer. There are two options to resolve this situation. The first is to force an intermediate settlement on the base layer, undelegate the accounts and send the transaction through the base layer RPC. The second is to discard the transaction as invalid. The preferred solution depends on the use case.

Additionally, the RPC Router provides the correct blockhash for transactions, a fundamental component in how Solana ensures security and prevents double spending [25].

2.2.3 Account/programs lazy loading and Data Availability

An interesting feature to note about ephemeral rollups is the ability to perform lazy loading of the accounts/programs used in transactions. If a transaction uses readable accounts for state or programs (logic) that are part of the main layer, the ER can clone the account state as needed, without having to initialize a complete clone of the main chain. Transactions are available for a specific time window. They can also be indefinitely stored in a compressed account on Solana [26] or within an external DA layer such as Celestia [18]. Transactions persist after the ephemeral rollup has been closed, making it easier to verify the correctness of the executed computation.

2.3 Security and Computation Trustworthiness

The two critical security aspects of the ER are as follows:

- Preventing an indefinite lock of delegated accounts or compromised state.
- Consensus and verification of ER’s computation.

The mechanism to avoid the first point hinges on the nature of Ephemeral Rollups; as the name suggests, they are ephemeral. Anyone can request the undelegation of accounts when the life of the ER ends by interacting with the program in the base layer. Furthermore, the program that carried out the delegation can revoke it or request settlement at any time, thereby forcing the closure of the ER.

The second point is a common challenge in all rollup architectures, typically addressed through optimistic rollups or ZK rollups. For a detailed description of zk-Rollups, see [33]; for Optimistic Rollups, see [22]; and for a comparison, see [11]. [24] describes ZK channels, a conceptually similar idea for the gaming use-case. As mentioned, transactions history remain available natively in a compressed account on Solana or an external DA layer like Celestia [18]). The proposed approach for BOLT combines these two techniques: it initially performs an optimistic computation for speed, and then produces a ZK proof (e.g., through RISC0 [31]) for verification. This proof can be efficiently verified on the base layer in a single transaction. Anyone can become a Provisioner for the Ephemeral Rollups sessions by staking a predetermined amount of tokens to the ER program, essentially creating a DePIN infrastructure [21]. Furthermore, ER participants themselves can act as SVM client clients [15], thereby enhancing the trustworthiness of the network.

3 SVM as a Multiplayer Game Server

The architecture presented addresses the challenge of processing millions of transactions for an arbitrary number of games. By horizontally scaling the traffic across multiple SVM ERs, it ensures high throughput and low latency, without compromising composability on the base layer and making it suitable for multiplayer gaming scenarios.

Estimating the theoretical throughput of this system is complex due to its dependency on the number of ERs and the sharding strategy. However, a comparative analysis with a multiplayer game server like Nakama [17] can provide a performance benchmark. Nakama is renowned for its low-latency performance, often achieving latency close to 50ms, and high-throughput capabilities, supporting thousands of concurrent players in real-time multiplayer games (see [16] for a more detailed Nakama benchmark).

As of the time of writing, Solana Sealevel can achieve a throughput ranging from 50,000 to 65,000 TPS [4]. The advent of newer runtimes, such as Firedancer [10], is anticipated to further boost the throughput to more than 100k TPS. This level of throughput is already substantial, yet the ER can further enhance the throughput with an arbitrarily large number of dedicated runtimes. In terms of latency, Solana produces blocks approximately every 400 milliseconds. However, the customizable ER runtime can decrease this time to between 10 and 50ms. It's worth noting that the reported performance figures for Solana are based on a network comprising around 2,000 nodes.

Solana's technology, particularly the SVM, could foster a new era of decentralized gaming. The Solana Virtual Machine, with its efficient transaction processing and support for decentralized applications, emerges as a promising infrastructure for hosting multiplayer game servers.

4 ECS, Mapping and Registry

The framework includes a standardized way to model the game logic, using the ECS (Entity, Components and Systems), which is commonly used in the gaming industry and also adopted in most on-chain engines, such as [14] and [20].

This pattern decouples logic from state, enabling optimizations in terms of performance and scalability. Additionally, its modular nature facilitates code reusability and extensibility, which are two essential properties for fully on-chain games and autonomous worlds.

The Solana Virtual Machine makes use of a paradigm similar to an ECS. The state (the accounts) and the logic (the programs) are natively separated. This separation is the key mechanism for Solana's parallelism and the scalability solution presented in Section 2.

When drawing a comparison with Solana's architecture, we can see how an ECS can be readily implemented:

- **Entities:** An entity is a general-purpose object typically represented by a unique identifier. It doesn't directly contain any data or behaviour but

serves as an identifier for a collection of components. Entities can be registered in a world instance account on Solana and could themselves be accounts.

- **Components:** Raw data structure that represents some aspect or attribute of an entity. For instance, a `PositionComponent` might just contain `x`, `y`, and `z` coordinates. This concept is essentially equivalent to accounts on Solana.
- **Systems:** Performs the game logic or behaviour by acting upon entities that have specific components. Similar to programs on Solana, which only specify the logic and the accounts they operate on.

Solana’s Program Derived Addresses (PDAs) facilitate the efficient storage and retrieval of components, mimicking a hash table-like structure. [12] provides an example implementation of an ECS on Solana.

4.1 Mapping and State listening

The standardized pattern allows for easy integration of game components with a rendering engine to display the game interface. The presented framework can be viewed as an open and permissionless alternative to a backend intended as a multiplayer game server. For visualization and rendering, existing engines such as Unity [30], Unreal Engine [32], Phaser [23], and others can be used. The standardized structure of the components allows for automatic mapping of components and entity properties (abstracting serialization and deserialization, akin to the mechanism of the Anchor framework on Solana [5]). State updates can be easily performed and intercepted, providing a mechanism similar to the Observer pattern [9] to listen to state changes and update the rendering.

4.2 Public Components Registry

A registry is a complementary infrastructure that works effectively in conjunction with an architecture that leverages the ECS pattern or the general paradigm of Solana. A public registry allows for publishing, discovering, and reusing pieces of logic (such as systems and data structures) shared across apps and games. Every program deployed on Solana can be perceived as a system that delineates the logic and receives accounts as input for computation. Specifically for apps and games, we envision a registry housing typical developmental logic - a collision system, an energy system, a movement system etc. — commonly found in traditional engines. Thus, the registry serves as a single hub where developers can discover, publish, and utilize systems and components. This facilitates the swift assembly of games by developers without the need to rewrite the entire underlying logic from scratch.

4.3 Game Loops, Ticking and FRP

Typically, game engines integrate a Game Loop, which allows for the continuous updating of the state (sometimes referred as ticking) in a loop that performs the following steps:

- Process input.
- Update the game state (e.g., physics, AI, object positions).

A client-side engine would also include the rendering step. As mentioned in section 2.2, continuous state updates can be achieved using a transaction scheduling mechanism like Clockwork. While it is a viable solution, it presents some issues such as the cost of fees and the uncertainty of the exact block/timestamp of the transaction execution. These issues are much less prominent in Ephemeral Rollup, where the cost of state update transactions can be lifted, and the block time can be configured to be in the range of 10 to 50ms.

While having game loops constitutes the most adopted solution for most game engines, there are also less explored alternatives. There are Haskell engines that instead leverage the functional programming paradigm, such as Yampa [1] and implementations of the functional and asynchronous ECS pattern, such as Apecs [6]. The underlying idea is that values are derived through a function instead of evolving the state based on update loops (which can depend on input, time, or other sources). A simple example is an energy system that depends on time, e.g., a function that increases energy by one point every 30 seconds. This paradigm aligns perfectly with the realm of blockchain, where transactions are inherently asynchronous and where the game instance potentially lives forever on the blockchain, evolving over time. The idea is conceptually identical to Functional Reactive Programming (FRP) and Functional Reactive Animation [13], which is a programming paradigm that deals with asynchronous data streams and changes in data over time.

Although it is theoretically possible to write an engine that fully utilizes this paradigm, avoiding an update loop, in practice, a combination of the two techniques allows for greater configurability and the ability to use the suitable paradigm considering the game mechanics being developed. For example, it is possible to update the position of one non-performing character (NPC) with a loop and derive the position of n other NPCs. The framework will evolve supporting both paradigms.

5 Conclusion

In this article, we have presented the design and architecture of BOLT, a performant and scalable framework for fully on-chain games and Autonomous Worlds. The framework adopts an Entity Component System (ECS) architecture to facilitate the creation and reuse of on-chain components and logic. By leveraging the unique features of the Solana Virtual Machine — its account structure, the

delegation of account modifications, and parallelization capabilities - the framework proposes Ephemeral Rollups as a promising path to scale FOCG. This design enables fast, specialized runtimes that auto-scale horizontally to meet demand without creating isolated environments. As a result, the system maintains composability with existing protocols, tooling and runtime advancements at the base layer without any state fragmentation. It's worth noting that while the solution is designed for fully on-chain games, this mechanism can be broadly applied to various other use cases.

References

- [1] Mun Hon Cheong, Manuel M. T. Chakravarty, and Ken Robinson. "Functional Programming and 3D Games". In: 2005. URL: <https://www.semanticscholar.org/paper/Functional-Programming-and-3D-Games-Cheong-Chakravarty/0308575c4253507c50ff63e5ab383f5bb0c84b9b/>.
- [2] Terry Chung. *Autonomous Worlds: The Case for Fully On-chain Games*. Accessed: 2023-10-07. URL: <https://medium.com/1kxnetwork/autonomous-worlds-the-case-for-fully-on-chain-games-3066db695a5a>.
- [3] Clockwork. *Solana automation engine*. Accessed: 2023-10-07. URL: <https://github.com/clockwork-xyz/clockwork>.
- [4] Solana Compass. *Solana Performance Statistics: Live TPS, Compute + Fee Tracker*. Accessed: 2023-10-07. URL: <https://solanacompass.com/>.
- [5] Anchor Contributors. *Anchor Framework for Solana*. <https://github.com/coral-xyz/anchor>. Accessed: 2023-10-07. 2023.
- [6] Apecs Contributors. *A fast, extensible, type driven Haskell ECS framework for games*. <https://github.com/jonascarpay/apecs>. Accessed: 2023-10-07. 2023.
- [7] Loot Chain Contributors. *What is Loot Chain?* Accessed: 2023-10-07. URL: <https://loot-chain.gitbook.io/loot-chain-documentation/getting-started/overview>.
- [8] Paima Contributors. *What is Paima Engine?* Accessed: 2023-10-07. URL: <https://docs.paimastudios.com/>.
- [9] Wikipedia contributors. *Observer pattern*. Accessed: 2023-10-07. URL: https://en.wikipedia.org/wiki/Observer_pattern.
- [10] Jump Crypto. *FireDancer: A New Validator Client for Solana*. Accessed: 2023-10-07. URL: <https://github.com/firedancer-io/firedancer>, <https://solana.com/firedancer>, <https://jumpcrypto.com/firedancer>, <https://www.alchemy.com/what-is-firedancer>, <https://thewealthmastery.io/solana-firedancer>.

- [11] Ivan Cryptoslav. *Optimistic Rollups vs. ZK-Rollups: The Ultimate Comparison*. Accessed: 2023-10-07. 2021. URL: <https://coinmarketcap.com/alexandria/article/optimistic-rollups-vs-zk-rollups-the-ultimate-comparison>.
- [12] Shanav K Mehta Dev Bharel. *Gaming Infrastructure Part 2: Introduction to ARC*. Accessed: 2023-10-07. URL: <https://jumpcrypto.com/writing/introduction-to-arc/>.
- [13] Conal Elliott and Paul Hudak. "Functional Reactive Animation". In: *ACM SIGPLAN Notices* 32.5 (1997), pp. 263–273. DOI: 10.1145/258948.258967. URL: <http://conal.net/papers/icfp97/icfp97.pdf>.
- [14] dojo engine. *Dojo: The Provable Game Engine*. Accessed: 2023-10-07. URL: <https://book.dojoengine.org/theory/autonomous-worlds.html>.
- [15] Anoushk Kharangate Harsh Patel. *Tinydancer: Diet Client v0 for the Solana Blockchain*. Accessed: 2023-10-07. URL: <https://www.tinydancer.io/whitepaper.pdf>.
- [16] Heroic Labs. *Benchmarks - Nakama Documentation*. Accessed: 2023-10-07. URL: <https://heroiclabs.com/docs/nakama/getting-started/benchmarks/>.
- [17] Heroic Labs. *Nakama: Open Source Online and Multiplayer Framework*. Accessed: 2023-10-07. URL: <https://heroiclabs.com/docs/>.
- [18] *LazyLedger: A Distributed Data Availability Ledger With Client-Side Smart Contracts*. Accessed: 2023-10-07. URL: <https://arxiv.org/abs/1905.09274>.
- [19] ludens. <https://0xparc.org/blog/autonomous-worlds>. Accessed: 2023-10-07. URL: <https://0xparc.org/blog/autonomous-worlds>.
- [20] Alvarius Ludens. *MUD: An engine for Autonomous Worlds*. Accessed: 2023-10-07. URL: <https://lattice.xyz/blog/mud-an-engine-for-autonomous-worlds>.
- [21] Kuleen Nimkar et al. *Solana DePIN Projects: A New Horizon for the Gig Economy*. Accessed: 2023-10-07. 2023. URL: <https://www.theblock.co>, <https://en.coinotag.com>, <https://solanafloor.com>, <https://coinmarketcap.com>.
- [22] *Optimistic Rollups — ethereum.org*. Accessed: 2023-10-07. 2023. URL: <https://ethereum.org/en/developers/docs/scaling/optimistic-rollups/>.
- [23] *Phaser*. <https://phaser.io/>. Accessed: 2023-10-07.
- [24] Will Robinson. *Unblocking On-Chain Games Pt. 5: Bat-Channels*. Accessed: 2023-10-07. URL: <https://medium.com/alliancedao/unblocking-on-chain-games-pt-5-bat-channels-f24a3845e465>.
- [25] Solana. *A new architecture for a high performance blockchain v0.8.13*. Accessed: 2023-10-07. URL: <https://solana.com/solana-whitepaper.pdf>.

- [26] Solana. *Account Compression Program*. Accessed: 2023-10-07. URL: <https://spl.solana.com/account-compression/>.
- [27] Solana. *Program Derived Addresses (PDAs)*. Accessed: 2023-10-07. URL: <https://solanacookbook.com/core-concepts/pdas.html>.
- [28] Squad. *An Introduction to Account Abstraction*. Accessed: 2023-10-07. URL: <https://squads.so/blog/what-is-account-abstraction-ethereum-vs-solana>.
- [29] Scott Sunarto. *Introducing world engine by Argus*. Accessed: 2023-10-07. URL: <https://blog.argus.gg/world-engine>.
- [30] Unity. <https://unity.com/>. Accessed: 2023-10-07.
- [31] Unknown. *RISC Zero's zk protocol*. Accessed: 2023-10-07. Unknown. URL: <https://risczero.com/proof-system-in-detail.pdf>.
- [32] Unreal Engine. <https://www.unrealengine.com/>. Accessed: 2023-10-07.
- [33] *Zero-Knowledge rollups* — *ethereum.org*. Accessed: 2023-10-07. 2023. URL: <https://ethereum.org/en/developers/docs/scaling/zk-rollups/>.

A Emerging gaming frameworks

Recently, several frameworks have emerged in an attempt to simplify the development of this new domain of games. Each of these "on-chain game engines" possesses distinct features and iterates along different verticals to facilitate the creation and management of FOC games, with various degrees of trade-offs. This section delineates the technical nuances and provides a comparative analysis of these frameworks.

A.1 EVM frameworks

One of the earliest on-chain gaming frameworks is MUD [20]. Lattice, the company behind the MUD framework, was the first to introduce the ECS (Entity-Component-System) paradigm on-chain. While currently on Optimism, MUD aims to resolve the EVM limitation by introducing an opinionated database that indexes the game state from the chain and can be easily queried like any normal relational DB. This enables quicker access to the table of components that constitute an Autonomous World instance compared to retrieving the state directly from the chain. The focus of MUD is on modularity, featuring a system of libraries and plugin-ins. The biggest limitation is the ability to scale or add customization at the runtime level. Assuming the projected path to scale games built on MUD is with L3, this would fragment the state of the game deployed across new rollups. A similar OP-stack-derived engine is Keystone, developed by Curio, a rollup framework improving the performance of the EVM. Curio is focusing on the more responsive types of games, thanks to the introduction of "ticks", atomic units of time, that ultimately makes the game feel and look more responsive. Similarly to MUD, every game built on Keystone will have to deploy their own independent rollup, with the same fragmentation problem of an isolated state.

LootChain [7] is yet another low block-time, customizable layer 2 deployed with Caldera. It is developed by the community of Loot and uses ADGL as gas. Aside from a lower block time it doesn't solve the above-mentioned problem of fragmentation across games, nor gas congestion due to the "noisy neighbour" problem.

A.2 Non-EVM frameworks

Dojo [14] is an open-source framework developed by Cartridge and other contributors on Starknet. The main advantage of Dojo is relying on Starknet prover to create provable-game sessions leveraging zero-knowledge computation and being able to post state diffs rather than entire state updates. The focus on Dojo is on shifting intense computation (i.e., physics) off-chain and proving the integrity of said computation on-chain. Dojo, similarly to other engines, proposes an ECS for the components, tools for scaffolding, indexing and the setup of local nodes to facilitate development. With the release of Madara, a fast L3, Dojo is making explicit that the path to scaling presents the same tradeoffs as other solutions.

Paima engine [8] is a stack developed by Paima Studio supporting FOC on Cardano, Polkadot, Algorand and EVM chains. It places emphasis on cross-chain development and at the core it makes use of a state machine that replays the smart contract logic and pushes the update on a sovereign rollup. Paima is explicitly trading off composability in order to achieve higher scalability and parallelization in its state machine. Every game is deployed on its own sovereign rollups. Even a trivial global leaderboard across games built on Paima becomes complicated, if even possible, with this approach.

A.3 Other frameworks

World engine [29], developed by Argus Labs, is an attempt to solve performance issues and introduce a ticking mechanic while also keeping composability in mind. Argus acknowledges that atomic composability is not necessarily a prerequisite for games (i.e., no flash-loans like dynamics) and proposes an EVM L2 with sharded execution environments and customizable runtimes. The insight is that gaming runtimes are inherently different from blockchains as they can have passive events (i.e., day and night cycles, energy regeneration...). Passive events are not generated from user input. This architecture is referred to as "loop-driven" runtime as opposed to "event-driven" runtime. While on paper, this looks like a promising approach, Argus is effectively bootstrapping a new blockchain from scratch. None of the existing infrastructure, tooling or smart contracts works out-of-the-box on non-EVM shards and the Argus ecosystem is not directly befitting from the broader advancements and activity of the base layer. We believe Autonomous Worlds promise to bring a new level of significance to play, which makes them more likely to succeed in an ecosystem where global payments are settled, userbase and liquidity are abundant, protocols are innovating on new primitives etc. Additionally, atomic composability can be leveraged to improve UX or to unleash novel use cases in non-trivial manners (i.e., in the context of eSports, an on-chain Bot could borrow a powerful item from an NFT marketplace right before a critical hit to complete a high-stakes fight and repay the loan with the prize won or a user could onramp with FIAT, swap to SOL, create an account abstraction wallet and commit to a prize pool in 1 atomic transaction).

A.4 Considerations

In their own unique ways, every framework is trying to increase the performance in terms of blocktime and throughput but introducing composability and network effects tradeoffs. Isolated app chains and horizontal scaling in its current form feel extremely short-sighted as it fragments the state and reduce the synergies with other protocols at the base layer. If it's true that Autonomous Worlds introduce a new level of significance in games and enable virtual experiences that act as avenues for commerce, play and work, we believe blockchains settling large volume of real-world economic activity, RWO, consumer-facing

applications, decentralized Physical infrastructure, and synchronizing state at low latency globally are a better fit for these kind of new experiences.

B Scalability Approaches

A prevalent approach to address scalability challenges involves layers 2, protocols built on top of the base blockchain (Layer 1) to increase transaction speed and reduce costs, and partitioning traffic into isolated segments, known as sharding. Both methods share similarities and scales by segregating the state and computation into independent layers/shards. A similar technique is often employed in multiplayer games, where the multiple server instances manage traffic, e.g., sharding by geographic zones, whether virtual or real-world regions. Layer and Shards must be interoperable; for instance, a player might transition from one map area to another, effectively moving between shards. This method of scalability is incorporated in most game engine described in section ??(directly or indirectly). This solution, taken to the extreme, can lead to app-specific L3s, where state and traffic are sharded for each app/game, an emerging scaling strategy for games built on the dojo and mud engines.

This technique inherently leads to state fragmentation, complicating the composability of games, applications, or other protocols on the blockchain. While bridges can solve this issue, they’ve historically been unreliable, added user interface complexity, and introduced potential sources of bugs. Additionally, it’s impossible to execute atomic transactions, which are transactions that are fully executed or not at all. Such atomicity can be essential for specific game mechanics, particularly if a game has economic implications or interfaces with atomic-requiring protocols.

Both solutions can use an improved runtime that can be faster and more performant. In this category, we can conceptually identify two typologies:

- The first encompasses various runtime implementations that, while diverse, remain compliant with the parent virtual machine’s interface. Speed and performance can be superior either due to a more efficient implementation (for example, an optimized implementation written in Rust vs. Python for the EVM). Alternatively, it might result from different underlying assumptions about security and decentralization. Examples include employing a novel or more centralized consensus mechanism or using an EVM runtime capable of creating a zero-knowledge proof of the execution, decoupling computation and consensus/verification.
- In contrast, the second category adopts a distinct state machine, free from the limitations of the base layer’s virtual machine. This offers an avenue to leverage optimized runtimes—like Paima [8] and the World Engine [29] architecture (with Go and Nakama as a multiplayer game server). The state machine define it’s own state transition mechanism, and eventually also the consensus and settlement on the base layer.

Both categories fall under the umbrella of rollups, further classified into distinct categories like ZK, optimistic, smart contract rollup, and sovereign, contingent on the mechanisms of execution, settlement, consensus, and data availability they employ. For a more detailed description, refer to section ??.

B.1 Scalability vs Composability

The performance and throughput of the execution layer are constrained by the features of the virtual/state machine and its runtime (the implementation). One clear advantage of using a runtime compliant with the specifications of the virtual machine is maintaining compatibility with the base layer. For instance, with the EVM, protocols and tech stacks are typically reusable or deployable without modifications across any EVM-compatible blockchain and L2/L3 solutions. Such compatibility fosters the growth and rapid expansion of an ecosystem. It also facilitates the integration of technological and application innovations from diverse ecosystems, leading to a general improvement in virtual machines. Moreover, users and products can easily be acquired and shared between compatible chains. However, the downside lies in architectural limitations, mainly arising from the blockchain’s emphasis on security, decentralization, immutability, and consensus.

Conversely, using a custom or pre-existing state machine not tailored for blockchain allows for performance, latency, and throughput optimization, but complicates composability and typically reduces trust and security assumptions. Additionally, protocols and products that rely on this state machine cannot be transferred between ecosystems, leading to increased fragmentation.

Our framework achieves both benefits, minimizing the scalability/composability trade-off. By leveraging the unique features of the Solana Virtual Machine, described in subsequent sections, it can accommodate any runtime for the SVM and scale it horizontally without compromising the ability to interact with protocols/components on the base layer, even atomically.