Inertial Mining: Equilibrium Implementation of the Bitcoin Protocol*

Manuel Mueller-Frank† Minghao Pan‡ Omer Tamuz§

Abstract

The value of proof-of-work cryptocurrencies critically depends on miners having incentives to follow the protocol. However, the Bitcoin mining protocol proposed by Nakamoto (2008) and implemented in practice is well known not to constitute an equilibrium: Eyal and Sirer (2018) construct a profitable deviation called "selfish mining" which relies on strategically delaying disclosure of newly mined blocks rather than publishing them immediately. We propose inertial mining, a novel mining protocol. When miners follow inertial mining, they produce the outcome intended by Nakamoto, i.e., a single longest chain. But unlike the Bitcoin mining protocol, inertial mining constitutes an equilibrium (assuming no miner controls more than half of the mining power). Indeed, neither selfish mining nor any other deviation is profitable. Furthermore, inertial mining only changes miners' behavior in the event of off-path forks, and can be implemented in Bitcoin without any changes to its consensus mechanism or blockchain architecture.

1 Introduction

Nakamoto (2008) introduced Bitcoin, a digital money that functions without a central entity controlling the monetary base and intermediating transactions. Although

*We thank Matt Weinberg for useful comments and suggestions.

†IESE Business School. Email: mmuellerfrank@iese.edu.

‡Caltech. Email: mpan2@caltech.edu.

§Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a National Science Foundation CAREER award (DMS-1944153) and a MURI grant.

Bitcoin so far failed to gain significant traction as a medium of exchange, it did so as an alternative financial asset and a digital store of value. Its market capitalization first surpassed $1 billion in 2013, $100 billion in 2017, and $1 trillion in 2021.[1] There is a growing literature that analyzes Bitcoin from an economic and game-theoretic perspective. See, for example, Biais et al. (2019), Schilling and Uhlig (2019), Leshno and Strack (2020), Huberman et al. (2021), and Pagnotta (2022). For a survey, see Halaburda et al. (2022).

The main feature that makes Bitcoin attractive as a reserve asset is its decentralized nature, which implies that no entity, country or otherwise, can censor transactions. Decentralization, as well as the absence of a central intermediary that selects and processes valid transactions, introduces two major challenges. First, it requires consensus among all participants about the allocation of Bitcoin ownership. Second, it requires that all entities that participate in the transaction selection and execution do not engage in malicious activity. Satoshi Nakamoto addressed these complications via the so-called *Nakamoto consensus* mechanism.

The initial understanding of Nakamoto consensus was that it is an equilibrium for consensus participants to act honestly and follow the standard *Bitcoin mining protocol* described by Nakamoto. The seminal paper by Eyal and Sirer (2018), however, establishes that the standard protocol is not an equilibrium, and constructs a profitable deviation strategy they call "selfish mining". As far as we know, it has since been an open question whether an equilibrium exists that replicates the outcome intended by Nakamoto. The contribution of this paper is to propose the *inertial mining protocol*. Inertial mining constitutes an equilibrium for the Nakamoto consensus mechanism, coincides with the Nakamoto protocol on-path, and thus produces the same outcome.

In order to enable decentralization, Nakamoto proposed a partitioning of the transaction history into discrete blocks that are assembled in regular intervals.[2] Each block is appended to one previously assembled block resulting in a set of directed paths of blocks that point to the genesis block. One such directed path (or chain of blocks) is called a *blockchain*. In Bitcoin, a transaction is considered part of the realized transaction history if and only if it is included in a block that belongs to the longest chain.[3] Blocks are added to the chain by so-called miners, i.e., agents that participate in the

[1] See www.coingecko.com/en/coins/bitcoin.

[2] In Bitcoin the expected time between blocks is approximately ten minutes.

[3] This is a simplification made for expositional convenience. Formally, the valid chain is the *most-worked* chain. Note that our main result carries forward for the "most-worked" chain selection

consensus mechanism. The power to assemble transactions into a block is chosen by proof-of-work, a contest that can be described as follows.

Each miner i controls α_i computational mining power (henceforth mining power)—normalized so that the sum of computational powers across agents equals one. The miners all use their mining power to solve a computational puzzle specific to the last block in the longest chain. When a miner succeeds (miner i succeeds with probability α_i) he publishes it in a new block which is propagated in the network. If the solution is valid, then all other miners add the new block to the end of the chain and start mining the next block upon it.

Agents are incentivized to follow the above-described standard Bitcoin mining protocol via so-called *block rewards*. In each valid mined block, a predetermined amount of Bitcoin is newly issued and assigned to the respective miner.[4] These block rewards incentivize honest mining as the property rights to the newly issued coins are only commonly accepted if the mined block ends up in the longest chain.

We assume that miners are long-lived and adopt the standard assumption that the payoff of a given miner is equal to the share of blocks he mined in the longest chain. Note that if all miners follow the standard mining protocol, then all blocks are consecutively mined on one chain and asymptotically each miner i's payoff is equal to his mining power α_i.

If a miner i has mining power α_i at least a half, equilibrium fails for a trivial reason: it can eventually overtake any public chain by extending his own private branch. Eyal and Sirer (2018) showed that there is a profitable deviation even if α_i is below one half (but not too low, see more below). The core idea is that a selfish miner withholds mined blocks, privately builds a parallel chain to the public longest chain, and selectively discloses it if his private chain generates a lead over the public one. This way the miners that follow the standard mining protocol waste resources and the selfish miner ends up with a payoff strictly larger than his computational power. They called this deviation *selfish mining*.

Our contribution is to construct an honest mining equilibrium such that no miner has an incentive to deviate, assuming $\max_i \alpha_i < 1/2$. The inertial mining protocol

rule.

[4]Total miner revenue is the sum of block rewards and transaction fees that senders pay for inclusion of their transaction in a block, and the transaction fee typically has a 1% to 10% share of miner revenue, which we abstract away from in our model.

we propose works as follows. Consider miners who are working on the last block in the current longest chain. If a new block gets appended to this chain, the miners will switch to working on the new last block, as in the standard mining protocol. However, if a new chain is published that does not extend the current longest chain, the miners will only switch to it if it is longer by at least $I > 0$ blocks. The number I is a parameter of the inertial mining protocol. To ensure that a specific symmetric strategy profile is an equilibrium, I needs to be chosen to be large enough, as a function of the miners' distribution of mining power. The closer the mining power of the largest miner is to half, the larger I needs to be. When there is a miner with power half or more, no choice of I constitutes an equilibrium.

Inertial mining differs from the standard mining protocol in its prescription of which chain to append to if there is more than one public chain, but results in one single chain as equilibrium outcome and thus achieves the intended outcome of the standard mining. Importantly, it does this robustly as an equilibrium, with miners having no incentive to deviate. It is straightforward to see that under inertial mining, selfish mining is no longer a profitable deviation. The main technical contribution of this paper is to show that no other possible deviation is profitable. This is challenging because the set of possible deviations is large. Our proof addresses this by assigning each honest block that is displaced from the equilibrium chain to a particular strategically mined "killer" block, and then showing that no such block can be expected to displace enough honest blocks to outweigh the miner's equilibrium share.

1.1 Related literature

This paper contributes to the growing literature on the incentive properties and equilibrium analysis of proof-of-work (PoW) blockchains. Our starting point is the vulnerability first identified by Eyal and Sirer (2018): Bitcoin's standard mining protocol does not constitute an equilibrium since a miner with less than majority hash power may profitably deviate by withholding blocks and releasing them strategically. This selfish mining strategy creates intentional and persistent forks. Sapirshtein et al. (2016) strengthen this critique by characterizing optimal withholding and disclosure strategies against honest miners following the standard protocol and showing that such strategies can strictly dominate the original selfish-mining deviation. The fact that the canonical Bitcoin mining protocol is strategically fragile motivates the cen-

tral question we study: whether there exists an equilibrium of a PoW blockchain that generates a single longest chain on the equilibrium path, as intended by Nakamoto (2008). Our main result answers this question affirmatively.

A closely related strand of work studies strategic mining in PoW blockchains but proposes to eliminate selfish mining by changing the consensus mechanism or the blockchain design itself. Heilman (2014) proposes the *Freshness Preferred* mining protocol, under which miners break ties by favoring blocks with more recent verifiable timestamps, thereby reducing the profitability of delayed block release and selfish mining. Because this approach relies on introducing unforgeable timestamps as a novel feature, it necessitates a redesign of blockchain. Solat and Potop-Butucaru (2017) propose changing Bitcoin's block-validation and chain-extension rules by requiring honest miners to create and mine on a special "ZeroBlock" whenever no valid block arrives within a predetermined interval based on expected block-finding and propagation times. Pass and Shi (2017) propose a redesigned blockchain architecture, FruitChains, that weakens the ability of strategic miners to benefit from fork manipulation and thereby improves robustness to selfish mining. While these contributions demonstrate that selfish-mining incentives can be mitigated, they all rely on changing the consensus mechanism and/or the architecture of the blockchain itself. By contrast, our approach does not modify Bitcoin's underlying architecture or mechanism; instead, we propose an alternative mining protocol which can be implemented in the same technical environment as the standard Bitcoin protocol.

Our paper is also related to the economics literature that models PoW consensus as a strategic interaction among rational miners. Relative to Eyal and Sirer (2018) and our paper, this literature typically restricts the strategy space by allowing miners to choose only which block to mine on, but not when to disclose newly found blocks. This rules out selfish mining by construction. Within such a framework, Biais et al. (2019) show that mining on the longest chain can be a Markov perfect equilibrium with convergence to a single chain, although other equilibria with persistent disagreement and forks also exist. Pagnotta (2022) embeds PoW mining incentives in a general-equilibrium environment in which the cryptocurrency's price and security are jointly determined. He shows that multiple equilibria can arise with different price-security combinations. Our paper differs from this literature by explicitly allowing the timing of block disclosure to be strategic.

More broadly, our analysis relates to the literature on attacks against Nakamoto

consensus and on the economic limits of PoW security. A related strand studies double-spend attacks; see, for example, Bonneau (2016) and Gans and Halaburda (2024). Budish (2025) argues that the permissionless security of Nakamoto consensus is intrinsically costly because the flow cost of security must scale with the value at risk from attack. Leshno, Shi, and Pass (2024) develop an alternative consensus design that preserves permissionless entry while delivering stronger economically meaningful security guarantees. These papers focus on different vulnerabilities or on alternative institutional designs, whereas our focus is the equilibrium resolution of selfish mining within PoW.

Finally, the motivation for our analysis is further strengthened by recent work emphasizing that selfish mining is not merely a theoretical possibility. Li, Campajola, and Tessone (2024) provide empirical evidence consistent with selfish-mining behavior in several PoW blockchains, with especially strong evidence for Monacoin and Bitcoin Cash. More conceptually, the absence of a known equilibrium in PoW that implements a single longest chain on the equilibrium path has been emphasized in discussions surrounding Ethereum's transition away from PoW; see Buterin (2017) and Hall, Shiaeles, and Li (2024). Our contribution to this debate is to show that such an equilibrium does exist: inertial mining sustains a single longest chain on the equilibrium path without requiring any modification of Bitcoin's consensus mechanism or blockchain architecture.

2 The Model

2.1 The mining game

We start by formally defining a game that captures the strategic interactions involved in Bitcoin mining. This model is widely used in the computer science literature (Eyal and Sirer, 2018; Sapirshtein et al., 2016; Bahrani and Weinberg, 2024).

A *block* is a pair (x, y), where x, y take values in a set of labels, which we take to be the unit interval $[0, 1]$. The label of the block (x, y) is x, and y is the label of the block's parent block. A chain is a finite or countable sequence of blocks $C = (x_1, y_1), (x_2, y_2), \ldots$ such that $x_i \neq x_j$ for $i \neq j$, $x_1 = y_1 = 0$, and for $i > 1$ it holds that $y_i = x_{i-1}$. We say that (x_i, y_i) is the *predecessor* block of (x_{i+1}, y_{i+1}), and that the latter is a *successor* of the former. For $i < j$ we say that (x_i, y_i) is an *ancestor* of

(x_j, y_j).

There is a finite set of players N. Each player i is exogenously assigned *mining power* $\alpha_i > 0$, with $\sum_i \alpha_i = 1$. There are discrete time periods $t \in \{1, 2, \ldots\}$.

Each player i has a finite set of *mined blocks* M_t^i at the end of period t. We denote by $M_t = \cup_i M_t^i$ the set of all mined blocks. There is a set of *public blocks* P_t at the end of period t, which is a subset of M_t. At the beginning of period t, player i can observe their own past blocks M_{t-1}^i as well as the past public blocks P_{t-1}, but not the other players' mined blocks. Thus, the history available to player i at the beginning of period t consists of their own actions up to that point, and in addition $P_0, P_1, \ldots, P_{t-1}, M_0^i, M_1^i, \ldots, M_{t-1}^i$. Beyond this, players do not observe the actions of other players.

We set $M_0^i = \varnothing$ for all $i \neq i_0$, and $M_0^{i_0} = \{(0,0)\}$. I.e., at the start of the game players have no mined blocks, except for some player i_0 who has the *genesis block* $(0,0)$. We also set $P_0 = \{(0,0)\}$, so that this block is public.

At each time period nature chooses one of the players, where the probability that i is chosen is i's mining power α_i. Nature's choices are independent across periods.

At the beginning of each time period each player i has to choose an action $b_t^i \in [0,1]$. As we shall see, this will be interpreted as the label of the block to which the player wants to add a block. If player i is chosen at period t, they mine a new block $m_t = (x, b_t^i)$, where x is chosen independently and uniformly at random from the set of labels $[0,1]$. That is, if they mine $m_t = (x, b_t^i)$ then $M_t^i = M_{t-1}^i \cup \{m_t\}$. Otherwise, $M_t^i = M_{t-1}^i$. Note that since block labels are chosen uniformly at random from $[0,1]$, each block will almost surely have a unique label.

Besides choosing b_t^i, players can, at the end of any time period, decide to *publish* any blocks in M_t^i. That is, each player chooses a subset $B_t^i \subset M_t^i$, and we set $P_t = P_{t-1} \cup (\cup_i B_t^i)$.

In addition, players observe a public randomization device: an i.i.d. process $(\Xi_t)_t$. This will be useful to coordinate on tie-breaking.[5] In summary, in period t, player i observes $\Xi_1, \ldots, \Xi_{t-1}$, $P_1, \ldots, P_{t-1}$, and $M_1^i, M_2^i, \ldots, M_{t-1}^i$; chooses b_t^i; then, with probability α_i, mines a block (x, b_t^i), which is added to M_{t-1}^i to form M_t^i; and finally can choose to publish any subset B_t^i of M_t^i.

Denote by $P = \cup_t P_t$ the set of all published blocks. Consider the set of all chains

[5] In practice, the chain itself can be used as such a device, by the use of standard secret-sharing techniques, applied to the block labels.

made out of blocks in P, and in particular the set of longest chains; these could be infinite.

If there is no unique longest chain in P, then all players' utility is 0. If there is a unique longest chain of published blocks we denote it by C^*. The utility of a player is the (asymptotic) fraction of blocks of C^* owned by the player:

$$u_i = \liminf_t \frac{|C^* \cap M_t^i|}{|C^* \cap M_t|}.$$

Utility depends on the fraction of blocks rather than the absolute number. This reflects a common assumption in the literature, which stems from the fact that the number of blocks (like the number of shares in a company) is arbitrary and can be scaled by a technical change in the protocol.

Our choice of utility reflects an assumption that miners are long-lived; indeed, due to immense fixed and variable costs, Bitcoin mining is conducted primarily by companies (and some governments) rather than individuals. The computer science literature likewise focuses on the asymptotic fraction rather than (say) the discounted fraction (see Eyal and Sirer, 2018; Sapirshtein et al., 2016; Bahrani and Weinberg, 2024), but we do expect that similar results hold for more standard discounted utilities. As our objective is resolve the selfish mining issue first uncovered in this literature, we naturally use the same utility function.[6]

2.2 The standard Bitcoin mining protocol and a profitable deviation

The standard Bitcoin mining protocol is simple, and is defined as follows: b_t^i is chosen to be the label of the last block in the longest chain in P_{t-1}. If there are multiple longest chains, then players choose uniformly at random among the last blocks of the longest chains, using the public randomization device. Whenever a player mines a block, they immediately publish it.

The seminal paper of Eyal and Sirer (2018) proves that the standard Bitcoin mining strategy profile is not an equilibrium. They introduce selfish mining, a profitable deviation strategy, which we describe now.

Suppose all agents follow the protocol, except i, who implements selfish mining. Consider the first time t in which i finds a block. Then up to time t all found blocks

[6]The only modification relative to Eyal and Sirer (2018) is our focus on the liminf of the fraction of own blocks in the longest chain as opposed to the limit. This is necessary as for general strategies the limit of the fraction might not exist.

form one chain. The selfish miner i appends his block to the last block of the longest public chain but does not communicate his block to other miners. He continues mining on his private chain and strategically discloses a subset of his private chain conditional on the number of blocks found by the other miners who always append to the longest public chain. Precisely, his disclosure strategy is as follows. In case of failing to generate a two-block lead over the public chain, the player publishes the block found in period t when the lead of his private chain drops to zero, i.e., when it has the same length of the public chain. In case of generating a private chain lead of two blocks or more, the selfish miner discloses the earlier blocks of the private chain so that from the moment of disclosure onward all other miners append to the last disclosed block of the chain mined by the selfish miner. This causes the miners who follow the Bitcoin mining protocol to waste their found blocks, increasing the share of blocks belonging to the selfish miner. While some blocks of the selfish miner are also lost, in some cases this is profitable.

The profitability of selfish mining crucially depends on the selfish miner's mining share α_i and the behavior of other miners in case of a tie in length between the public chain and the disclosed chain of the selfish miner. Following Eyal and Sirer (2018), let $\gamma \in [0,1]$ denote the probability of miners $j \neq i$ appending to the disclosed block of the selfish miner in case of a tie. Recall that our analysis concentrates on the case of $\gamma = \frac{1}{2}$. To see the intuition behind selfish mining, consider the extreme case of $\gamma = 1$. In this case, even initially withholding a block found in period t does not come with the risk of it not being included in the longest chain: if miner i fails to generate a two-block lead, he publishes the block i found at time t and all other miners subsequently append to his block, leaving stale the honest block found in period $t+1$. Thus, selfish mining results in a strictly higher payoff than honest mining for any mining share $\alpha_i > 0$, given the extreme case of $\gamma = 1$. Eyal and Sirer (2018) establish the following result.

Proposition 1. *For a given γ, a selfish miner i with computational power α_i achieves a payoff larger than α_i if $\alpha_i > \frac{1-\gamma}{3-2\gamma}$.*

This result implies that in our setting, the threshold above which selfish mining is a profitable deviation from the Bitcoin mining protocol is $\alpha_i > \frac{1}{4}$

2.3 The Inertial Mining Protocol

Suppose $\alpha_i < 1/2$ for all i. We propose the following symmetric strategy profile, and prove that it is an equilibrium. The protocol is parametrized by $I \geqslant 1$, and defined as follows.

If P_{t-1} consists of a single chain, choose b_t^i to be the label of the last block in this chain. Otherwise, let $C_1, \ldots, C_n$ be the chains in P_{t-1}, ordered by increasing lengths $\ell_1 \leqslant \ell_2 \leqslant \cdots \leqslant \ell_n$. If $\ell_n \geqslant \ell_{n-1} + I$, choose b_t^i to be the label of the last block in C_n. Otherwise, let $C_{i_1}, \ldots, C_{i_k}$ be the chains that contain b_{t-1}^i, and choose b_t^i uniformly at random among these chains, using the public randomization device Ξ_t, so that all players choose the same b_t^i. Finally, all blocks are published as soon as they are mined.

A few notes are in order:

1. On path, all blocks are published immediately, and so behavior is identical to the one resulting from all miners adopting the standard Bitcoin mining protocol.

2. The difference between inertial mining and standard mining occurs when more than one chain is longest. In Bitcoin miners switch to the longest chain (assuming no ties). In inertial mining they switch away from the chain they are currently mining only if the competing chain is longer by I.

3. The case $I = 1$ is similar to the standard Bitcoin protocol: if there is a unique longest chain then players mine that chain. However, tie-breaking works differently if there is no unique longest chain.

4. It is straightforward to see that if I is large enough then selfish mining is not profitable. The challenge is to prove that no other deviation is profitable; this is non-trivial, since the strategy space is rather rich. The analysis is further complicated by the asymptotic nature of the utility, which implies that no one-deviation principle holds.

5. It is important that when publishing at time t, Ξ_{t+1} is not yet known, so that tie-breaking is uniformly random, conditioned on the information available to players.

Claim 1. *Under inertial mining, the utility of each player i is almost surely α_i.*

Since under inertial mining there is a unique chain that contains every mined block, Claim 1 follows immediately from the strong law of large numbers.

3 Results and Analysis

Our main result is the following:

Theorem 1. *Given $(\alpha_i)_i$ with $\max_i \alpha_i < 1/2$, the inertial mining protocol is an equilibrium of the mining game for sufficiently large I.*

We leave for future work an explicit calculation of how large I needs to be, but conjecture that this is small enough to be practical, assuming $\max_i \alpha_i$ is not extremely close to one half. The remainder of this section is devoted to the proof of this theorem.

Suppose all players are playing according to the inertial mining protocol, except perhaps player i who is using some other strategy. We prove that player i's utility is at most α_i, which by Claim 1 shows that there is no profitable deviation. Without loss of generality we assume that player i's strategy is not randomized, since if there is a mixed profitable deviation then there is also a pure one. We assume that there is a longest published chain C^*, since otherwise player i has utility zero, and such their strategy is not profitable. We henceforth fix the strategies of all the players, including the deviating player i, and calculate probabilities and expectation with respect to the distribution over outcomes generated by these strategies and the randomness in the mining process.

Since all players other than i behave identically, we will assume that there is only one player other than i, and denote this player j. This is done for notational convenience only. The mining power of this player is $1 - \alpha_i$. Recall that b_t^j is the block to which player j is trying to mine a successor at time t. With more than two players, all players j different than i choose the same b_t^j, and so there is no loss of generality in assuming that there are only two players.

Denote by $H_t^i = (\Xi_1, \ldots, \Xi_{t-1}, P_1, \ldots, P_{t-1}, M_1^i, M_2^i, \ldots, M_{t-1}^i)$ the history observed by player i at the beginning of period t.

Recall that player i's utility is

$$u_i = \liminf_t \frac{|C^* \cap M_t^i|}{|C^* \cap M_t|}.$$

We need to show that $\mathbb{E}[u_i] \leqslant \alpha_i$. We claim that it suffices to show that $\mathbb{P}[u_i > \alpha_i] = 0$.

Claim 2. *Suppose that i has a strategy that $\mathbb{E}[u_i] > \alpha_i$. Then i has a (perhaps different) strategy such that $\mathbb{P}[u_i > \alpha_i] = 1$.*

Informally, this holds because if player i has a strategy that yields payoff greater than α_i with positive probability, then the player has a different strategy that yields payoff greater than α_i with probability one. The idea is that since the payoff does not depend on what fraction of blocks the player had at any finite time, the player can always "start from scratch" and draw a new u_i if they think it is unlikely to be above α_i.

Proof of Claim 2. Suppose that $\mathbb{P}[u_i > \alpha_i] = p > 0$, and let A be the event $\{u_i > \alpha_i\}$. Then $p_t = \mathbb{E}[A \,|\, H_t^i]$ is a bounded martingale, and hence converges almost surely to the indicator of the event A. Consider the following alternative strategy for player i: follow the original strategy, unless there is some time s such that $p_s < p/2$. If no such time exists then $\lim_t p_t = 1$, the event A occurs and the utility is strictly above α_i. If there is a time s such that $p_s < p/2$, the player "starts from scratch", implementing the original strategy but treating the block b_s^j as the genesis block. Since the payoff can also be written as

$$u_i = \liminf_t \frac{|C^* \cap (M_t^i \backslash M_{t'}^i)|}{|C^* \cap (M_t \backslash M_{t'})|}$$

for any time t', the player now again has chance p to have payoff strictly greater than α_i. Repeating this whenever $\mathbb{P}[u_i > \alpha_i \,|\, H_t^i]$ goes below $p/2$ yields, by the law of large numbers, a strategy such that $u_i > \alpha_i$ almost surely. □

Denote by $K^j = M^j \backslash C^*$ the set of *killed blocks*. These are the blocks of player j that are not in the longest chain. Since player j follows the inertial mining protocol, he will always add blocks to the tree originating in the genesis block $(0, 0)$. Hence each block $(x, y) \in K^j$ will have at least one ancestor in C^*. We denote by $a(x, y) \in C^*$ the closest ancestor of (x, y) that is in C^* (see Figure 1).

We say m_t, the block mined at time t, is *dishonestly mined* if either it is not published at time t, or its predecessor is not b_t^j. Otherwise, we say m_t is *honestly mined*. Note that all blocks mined by j are honestly mined. We say a dishonestly

mined block is an *initial dishonestly mined* (henceforth, initial) block if its predecessor is honestly mined.

The *depth* $\Delta(x,y)$ of a block (x,y) that is a descendant of the genesis block $(0,0)$ is the length of the chain starting at $(0,0)$ and ending at (x,y). Note that on C^* there is a unique block at each depth. It follows from the definition of inertial mining that $\Delta(b_1^j) \leqslant \Delta(b_2^j) \leqslant \cdots$, and that in periods in which j mines a block this inequality is strict. Of course, it is also strict if i mines honestly, and in some other cases. The following claim is an immediate consequence of the definitions.

Claim 3. *Suppose that the block m_t was honestly mined. Then $\Delta(b_{t+1}^j) > \Delta(b_t^j)$. It follows that blocks mined by j have distinct depths.*

Given a block (x,y) that is a descendant of the genesis block, denote by $\text{cousin}(x,y)$ the block (w,z) on C^* satisfying $\Delta(x,y) = \Delta(w,z)$ (see Figure 1).

Claim 4. *Let $(x,y) \in K^j$ be a killed block. The block* $\text{cousin}(x,y)$ *is dishonestly mined by i.*

Proof. Note that if m_t is honestly mined then $\Delta(m_t) = \Delta(b_t^j) + 1$, regardless if it was mined by i or j. It thus follows from Claim 3 that any two honestly mined blocks cannot have the same depth. Since every block in K^j is honestly mined, it must be that $\text{cousin}(x,y)$ is dishonestly mined. □

Fix a parameter I, and also fix $J \in \{1,\ldots,I-1\}$. We will show that conclusion of the theorem holds if J is large enough, and $I-J$ is also large enough.

Suppose $(x,y) \in K^j$ is a killed block. Let (w,z) be the ancestor block of $\text{cousin}(x,y)$ such that all the blocks between (w,z) and $\text{cousin}(x,y)$ (including (w,z) itself) are dishonestly mined, yet the immediate predecessor of (w,z) is honestly mined. Then (w,z) must be an initial dishonestly mined block. We define the *killer* of the killed block (x,y) as follows. If $\Delta(\text{cousin}(x,y)) \leqslant \Delta(w,z) + J$, then we set $\text{killer}(x,y) = \text{cousin}(x,y)$ to be the killer of (x,y). Otherwise, we set $\text{killer}(x,y) = (w,z)$. See Figure 1.

The definition of $\text{killer}(x,y)$ is the key to our proof: it sets up an accounting system that attributes every killed block to some (dishonestly mined) killer. The proof then shows that no block can be expected to kill enough blocks to make dishonest mining worthwhile.

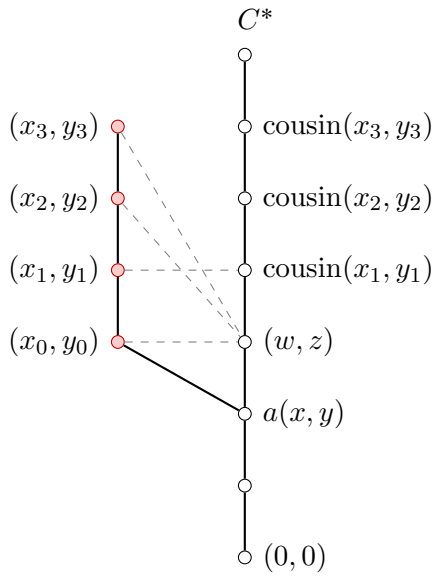

Figure 1: Illustration of the definitions of $a(x, y)$, cousin(x, y) and killer(x, y). The block $a(x, y)$ is the closest ancestor of (x, y) that lies on the longest chain C^*. The block cousin(x_i, y_i) is the unique block on C^* with the same depth as (x_i, y_i). Suppose $J = 1$. Then killer(x_1, y_1) = cousin(x_1, y_1) and (w, z) is the killer of (x_0, y_0), (x_2, y_2) and (x_3, y_3).

Let K_t be the number of blocks killed by m_t, the block mined at time t:

$$K_t = |\{(x, y) \in K^j \, : \, \text{killer}(x, y) = m_t\}|.$$

Let

$$\text{killer}^{-1}(m_t) := \{(x, y) : m_t = \text{killer}(x, y)\}.$$

Let $L_t = 1_{m_t \in M^i \cap C^*}$ the indicator of the event that the block mined at time t will belong to i and will be included in the longest chain.

Define the random variable

$$S_t = \alpha_i K_t + (1 - \alpha_i) L_t,$$

which we call the *payoff* of the block (potentially mined by i) at time t.

The next proposition is the key technical component in the proof of Theorem 1.

Proposition 2. *If J and $I - J$ are sufficiently large then*

$$\mathbb{E}\big[S_t \,\big|\, H_t^i\big] \leqslant \alpha_i(1-\alpha_i)$$

almost surely.

That is, given any history, the expected payoff is at most $\alpha_i(1-\alpha_i)$.

Proof. Since $S_t = 0$ if $m_t \notin M^i$, we need to show that $\mathbb{E}[S_t \,|\, H_t^i, m_t \in M^i] \leqslant 1-\alpha_i$.

We will consider the cases that the predecessor block of m_t is either honestly or dishonestly mined.

Case 1. Suppose that given $H_t^i = h_t^i$, the predecessor block of m_t is dishonest, so that m_t cannot be initial. Then, by definition, m_t can only possibly kill its cousin. Let (w, z) be the closest (youngest) initial dishonestly mined block among the ancestors of m_t. For example, this is the case for $m_t = \text{cousin}(x_i, y_i)$ in Figure 1.

Case 1.1. Suppose that $\Delta(w,z) + J < \Delta(m_t)$. For example, this is the case of $m_t = \text{cousin}(x_2, y_2)$ or $m_t = \text{cousin}(x_3, y_3)$ in Figure 1. Then by definition m_t cannot kill any block, so that

$$\mathbb{E}\big[S_t \,\big|\, H_t^i = h_t^i, m_t \in M^i\big] = (1-\alpha_i)\mathbb{E}\big[m_t \in C^* \,\big|\, H_t^i = h_t^i, m_t \in M^i\big] \leqslant 1-\alpha_i.$$

Case 1.2. Suppose that $\Delta(w,z)+J \geqslant \Delta(m_t)$, and assume that b_t^j is not an ancestor of m_t. Since the predecessor of (w,z) is honestly mined, we have

$$\Delta(b_t^j) \geqslant \Delta(w,z) \geqslant \Delta(m_t) - J.$$

If m_t is included in C^*, then eventually player j mines after m_t, and we let t' be the first time when player j ever mines at some descendant of m_t. By the definition of inertial mining, it must be that $\Delta(b_{t'}^j) \geqslant \Delta(b_{t'-1}^j) + I$.

Now, the chain starting at m_t and ending at $b_{t'}^j$ must be made of blocks mined by i; this follows from the minimality of t'. The length of this chain is $\Delta(b_{t'}^j) - \Delta(m_t)$, which by the argument of the previous paragraph is at least $\Delta(b_{t'-1}^j) + I - \Delta(m_t)$. Hence we have that the number of blocks $A_{t'}^i$ mined by i between time t and t' is at least

$$A_{t'}^i \geqslant \Delta(b_{t'-1}^j) + I - \Delta(m_t).$$

On the other hand, $b^j_{t'-1}$ must be deeper than b^i_t by at least $A^j_{t'}$, the number of blocks produced by j in this time interval. We thus have that

$$A^i_{t'} - A^j_{t'} \geqslant \big(\Delta(b^j_{t'-1}) + I - \Delta(m_t)\big) - \big(\Delta(b^j_{t'-1}) - \Delta(b^j_t)\big) \geqslant I - J.$$

That is, i produced at least $I - J$ blocks more than j in the time interval $[t, t']$.

The difference between the number of blocks mined by i and j behaves like a random walk that moves to the right with probability α_i, and to the left with probability $1-\alpha_i$. We define ε_k to be the probability that such an α_i-biased random walk starting from k ever hits I. A standard calculation yields that

$$\varepsilon_k = \begin{cases} \frac{\alpha_i^{I-k}}{(1-\alpha_i)^{I-k}} & \text{if } I > k \\ 1 & \text{otherwise.} \end{cases}$$

Then the probability that there exists some $t' > t$ such that i produced at least $I - J$ blocks more than j in the time interval $[t, t']$ is exactly ε_J. Since the existence of such t' is necessary for $m_t \in C^*$, we have

$$\mathbb{P}(m_t \in C^* | H^i_t = h^i_t, m_t \in M^i) \leqslant \varepsilon_J.$$

Since m_t can only possibly kill its cousin, and $m_t \in C^*$ is necessary for killing, we have

$$\mathbb{E}\big[S_t \,\big|\, H^i_t = h^i_t, m_t \in M^i\big] \leqslant \mathbb{P}(m_t \in C^* | H^i_t = h^i_t, m_t \in M^i)$$

When $I - J$ is large enough, we arrive at the desired bound

$$\mathbb{E}\big[S_t \,\big|\, H^i_t = h^i_t, m_t \in M^i\big] \leqslant 1 - \alpha_i$$

Case 1.3. Suppose now that $d\,(w, z) + J \geqslant \Delta(m_t)$, and that b^j_t is an ancestor of m_t. If miner i chooses to publish m_t at time t, we have $S_t \leqslant 1 - \alpha_i$, because m_t cannot kill any blocks. Our goal is then to show that

$$\mathbb{E}\big[S_t \,\big|\, H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ at time } t\big] \leqslant 1 - \alpha_i.$$

For $k \geqslant 0$, let E_k be the event that the k blocks mined starting at time $t+1$ are mined by player i and then player j gets the following one. Then $\mathbb{P}[E_k] = \alpha^k(1 - \alpha)$ and E_k is independent of $\{H^i_t, m_t \in M^i, \text{not publishing } m_t \text{ at time } t\}$. Under E_0, $m_t \in C^*$

only if m_t is published at $t+1$ and wins tie breaking, or descendants of m_t are being extended by I blocks longer than $b^j_{t'}$ at time t' for some $t' > t$. A union bound yields

$$\begin{aligned}
&\mathbb{E}\big[S_t \,\big|\, H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ at time } t, E_0\big] \\
\leqslant\ &\mathbb{E}\big[m_t \in C^* \,\big|\, H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ at time } t, E_0\big] \\
\leqslant\ &\frac{1}{2} + \varepsilon_J.
\end{aligned}$$

where we use the fact that m_t can only kill a block if it survives in the first inequality.

Under each E_k for $k \geqslant 1$, m_t may survive by publishing before time $t+k+1$, in which case $S_t = 1-\alpha_i$. If m_t is not published before time $t+k+1$, then again m_t can only survive if either m_t is published at $t+k+1$ and wins tie breaking, or extended by I blocks longer than $b^j_{t'}$ at time t' for some $t' > t+k$. Then

$$\begin{aligned}
&\mathbb{E}\big[S_t \,\big|\, H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ by time } t+k, E_k\big] \\
\leqslant\ &\mathbb{E}\big[m_t \in C^* \,\big|\, H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ by time } t+k, E_k\big] \\
\leqslant\ &\frac{1}{2} + \varepsilon_{k+J},
\end{aligned}$$

where the bound ε_{k+J} is because at time $t+k+1$, the descendants of m_t is at most $k+J$ ahead of b^j_{t+k+1}.

Summing up over E_k, we have

$$\begin{aligned}
&\mathbb{E}\big[S_t \,\big|\, H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ at time } t\big] \\
\leqslant\ &(1-\alpha_i)[\frac{1}{2} + \varepsilon_J] + \sum_{k=1}^{\infty} \alpha_i^k (1-\alpha_i) \max\{1-\alpha_i, \frac{1}{2} + \varepsilon_{k+J}\} \\
\leqslant\ &(1-\alpha_i)[\frac{1}{2} + \varepsilon_J] + \sum_{k=1}^{\infty} \alpha_i^k (1-\alpha_i)(1-\alpha_i + \varepsilon_{k+J}) \\
=\ &(1-\alpha_i)(\frac{1}{2} + \alpha_i) + (1-\alpha_i) \sum_{k=0}^{\infty} \alpha_i^k \varepsilon_{k+J}
\end{aligned}$$

Choosing $I - J$ large enough yields

$$\mathbb{E}[S_t | H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ at time } t] \leqslant 1-\alpha_i.$$

Case 2. Suppose that under history h^i_t the immediate predecessor of m_t is honest; this is the case of block (w, z) in Figure 1. If m_t is published at time t, then no block is killed by m_t and $S_t \leqslant 1-\alpha_i$. Next, we shall show that

$$\mathbb{E}\big[S_t \,\big|\, H^i_t = h^i_t, m_t \in M^i, \text{not publishing } m_t \text{ at time } t\big] \leqslant 1-\alpha_i.$$

We can write the payoff $S_t = S_t^1 + S_t^2$, where

$$S_t^1 := \alpha_i \sum_{k=J+1}^{\infty} 1_{\text{a block of depth } \Delta(m_t)+k \text{ is killed by } m_t}$$
$$S_t^2 := (1-\alpha_i) 1_{m_t \in C^*} + \alpha_i 1_{\text{a block of depth } \Delta(m_t) \text{ is killed by } m_t}$$

Here we use the fact that blocks of depth between $\Delta(m_t)+1$ and $\Delta(m_t)+J$ cannot be killed by m_t, and for each depth there is at most one honest block of that depth and can be possibly killed.

Suppose that a block (x,y) of depth of $\Delta(m_t)+k$ is killed by m_t, where $k \geqslant J+1$. Let $(x',y') = \text{cousin}(x,y) \in C^*$. Let t' be the first time when player j mines after descendants of (x',y'). Then $t' \geqslant k+t$. As all the blocks on the chain between (x,y) and $b_{t'}^j$ are mined by i, together with Claim 3, we know that between time t and t', player i mines weakly more blocks than player j. The probability that such t' exists conditional on $m_t \in M^i$ is at most ε_k^*, defined as the probability that there exists $k' > k$ such that the α_i-biased random walk starting from 1 hits 0 at time k'. Therefore,

$$\mathbb{E}\big[S_t^1 \,\big|\, H_t^i = h_t^i, m_t \in M^i, \text{not publishing } m_t \text{ at time } t\big] \leqslant \alpha_i \sum_{k=J+1}^{\infty} \varepsilon_k^*$$

Note that ε_k^* decays exponentially with k. Choosing J large enough, the conditional S_t^1 can be made arbitrarily small.

For the second part, we use the same computation as in Case 1.3 to get

$$\begin{aligned}
&\mathbb{E}\big[S_t^2 \,\big|\, H_t^i = h_t^i, m_t \in M^i, \text{not publishing } m_t \text{ at time } t\big] \\
\leqslant\ & (1-\alpha_i)[\frac{1}{2} + \varepsilon_1] + \sum_{k=1}^{\infty} \alpha_i^k (1-\alpha_i) \max\{1-\alpha_i, \frac{1}{2} + \varepsilon_{k+1}\}. \\
\leqslant\ & (1-\alpha_i)(\frac{1}{2} + \alpha_i) + (1-\alpha_i) \sum_{k=0}^{\infty} \alpha_i^k \varepsilon_{k+1}.
\end{aligned}$$

By choosing I large enough, the conditional S_t^2 can be made arbitrarily close to $(1-\alpha_i)(\frac{1}{2}+\alpha_i)$. Summing over the bounds on the conditional S_t^1 and S_t^2, and by choosing I and $I-J$ large enough, we have

$$\mathbb{E}\big[S_t \,\big|\, H_t^i = h_t^i, m_t \in M^i, \text{not publishing } m_t \text{ at time } t\big] \leqslant 1-\alpha_i.$$

□

Next, we prove a lemma that bounds the probability that a honest block is killed by another block created later.

Lemma 1. *For $s > t$,*

$$\mathbb{P}\big[m_t \text{ is killed by } m_s \,\big|\, H_t^i\big] \leqslant \varepsilon^*_{s-t},$$

where, as above, ε^*_k is the probability that there exists $k' > k$ such that the α_i-biased random walk starting from 1 hits 0 at time k'.

Before proving this lemma, we note that as ε^*_k decays exponentially in k, we get a universal upper bound on the number of blocks mined before time t and killed by blocks mined after t as a corollary.

Corollary 1. *There exists a constant C such that for any t:*

$$\sum_{k \geqslant t+1} \mathbb{E}\big[|M_t^j \cap \text{killer}^{-1}(m_k)|\big] < C$$

Proof of Corollary 1. We write the left hand side as a double summation:

$$\sum_{k \geqslant t+1} \mathbb{E}\big[|M_t^j \cap \text{killer}^{-1}(m_k)|\big] = \sum_{t' \leqslant t} \sum_{s \geqslant t+1} \mathbb{P}[m_{t'} \text{ is killed by } m_s]$$

. □

Proof of Lemma 1. Suppose that m_t is killed by m_s, where $s > t$. By definition of killing, $\Delta(m_t) \leqslant \Delta(m_s)$. As $s > t$, m_s cannot kill m_t by tie-breaking. Let $t' > s$ be the first time when player j mines after a descendant of m_s. Then descendants of m_s are extended by I blocks longer than $b^j_{t'}$ at time t'. This necessarily implies that the number of blocks player i mines between s and t' is at least I more than the number of blocks player j mines between t and t'. In particular, player i mines weakly more blocks than player j do between t and t'. The probability that there such $t' > s$ exists is at most ε^*_{s-t} defined in Case 2 of proof of Proposition 2.

□

We also need a bound on the number of killed blocks; this implies that the blockchain has a linear growth rate.

Lemma 2. *If J and $I - J$ are sufficiently large, then for every t:*

$$\mathbb{E}\big[K_t \,\big|\, H_t^i\big] \leqslant (1 - \alpha_i)\alpha_i$$

almost surely.

Proof. It suffices to show

$$\mathbb{E}\big[K_t \,\big|\, H_t^i, m_t \in M^i\big] \leqslant 1-\alpha$$

.

Suppose that m_t is not an initial dishonestly mined block. Then by definition, m_t can kill at most one block. Moreover, m_t can only kill a block when $m_t \in C^*$. Therefore, Proposition 2 implies that

$$\mathbb{E}\big[K_t | H_t^i, m_t \in M^i, m_t \text{ is not initial}\big] \leqslant 1-\alpha_i$$

Suppose now that m_t is initial. By definition, m_t can either kill its cousin, or kill blocks with depth at least $\Delta(m_t)+J$. Using notations from Case 2 of proof of Proposition 2, the number of blocks m_t kill that have depth at least $\Delta(m_t)+J$ is $\alpha^{-1}S_t^1$. Since m_t can only kill its cousin when $m_t \in C^*$, we have

$$1_{m_t \text{ kills its cousin}} \leqslant S_t^2.$$

Hence, $K_t \leqslant \alpha^{-1}S_t^1 + S_t^2$. For J and $I-J$ large enough, we have seen in the proof of Proposition 2 that the conditional S_t^1 can be made arbitrarily small and S_t^2 strictly bounded away from $1-\alpha$. Thus, we have

$$\mathbb{E}\big[K_t | H_t^i, m_t \in M^i, m_t \text{ is initial}\big] \leqslant \mathbb{E}\big[\alpha^{-1}S_t^1 + S_t^2 | H_t^i, m_t \in M^i, m_t \text{ is initial}\big] \leqslant 1-\alpha_i.$$

□

We will need to following general lemma regarding the limit of the ratio of two stochastic processes. Its proof is deferred to the appendix.

Lemma 3. *Suppose that $\{X_t\}$ and $\{Y_t\}$ are two random sequences and $a>0$ are constants such that*

1. $X_t, Y_t \in [0,1]$.
2. $\liminf_t \mathbb{E}[aY_t - X_t] \geqslant 0$.
3. $\liminf_t \mathbb{E}[Y_t] > 0$.

Then it is impossible that

$$\liminf_t \frac{X_t}{Y_t} > a \quad a.s.$$

Here we define the lim inf *to be* 0 *if* Y_t *is eventually* 0.

We are now ready to prove Theorem 1. We first choose J and $I - J$ large enough for the conclusions from Proposition 2 and Lemma 2 to be true.

We let

$$X_t := \frac{1}{t}|C^* \cap M_t^i|,\ Y_t := \frac{1}{t}|C^* \cap M_t|,\ \alpha := \alpha_i,$$

and check the conditions of Lemma 3 in the following. The first condition in the lemma is trivially satisfied as $t = |M_t| \geqslant |M_t^i|$. It remains to check the other two conditions.

Note that

$$\mathbb{E}\big[|C^* \cap M_t^i|\big] = \mathbb{E}[L_1 + \cdots L_t]$$

by the law of total expectation. Likewise,

$$\begin{aligned}
\mathbb{E}\big[|C^* \cap M_t^j|\big] &= \mathbb{E}\big[|M_t^j \setminus \cup_{k\geqslant 1} \text{killer}^{-1}(m_k)|\big] \\
&\geqslant \mathbb{E}\big[|M_t^j|\big] - \mathbb{E}[K_1 + \cdots + K_t] - \sum_{k\geqslant t+1} \mathbb{E}\big[|M_t^j \cap \text{killer}^{-1}(m_k)|\big] \qquad (1) \\
&\geqslant (1-\alpha_i)t - \mathbb{E}[K_1 + \cdots + K_t] - C.
\end{aligned}$$

The first inequality follows from the fact that the number of blocks killed by the blocks i mined up to time t upper bounds the number of blocks in M_t^j killed by these blocks. The second inequality follows from Corollary 1, where C is a constant independent of t. It follows that

$$\begin{aligned}
\mathbb{E}[aY_t - X_t] &\geqslant \alpha_i \left((1-\alpha_i) - \frac{1}{t}\mathbb{E}[K_1 + \cdots + K_t] + \frac{1}{t}\mathbb{E}[L_1 + \cdots + L_t] - \frac{C}{t}\right) \\
&\quad - \frac{1}{t}\mathbb{E}[L_1 + \cdots + L_t] \\
&= \frac{1}{t}\sum_{s=1}^{t}\Big[\alpha_i(1-\alpha_i) - \mathbb{E}[\alpha_i K_s + (1-\alpha_i)L_s]\Big] - \frac{C}{t}.
\end{aligned}$$

From Proposition 2 and law of iterated expectations, we know each summand is non-negative. Thus, $\liminf_t \mathbb{E}[aY_t - X_t] \geqslant 0$.

Lastly, continuing with (1) and using Lemma 2 and law of iterated expectations, we have

$$\begin{aligned}\mathbb{E}[Y_t] &\geqslant \mathbb{E}\big[|C^* \cap M_t^j|\big] \\ &\geqslant (1-\alpha_i) - \frac{1}{t}\mathbb{E}[K_1 + \cdots + K_t] - \frac{C}{t} \\ &\geqslant (1-\alpha_i) - (1-\alpha_i)\alpha_i - \frac{C}{t}.\end{aligned}$$

Taking lim inf then yields $\liminf_t \mathbb{E}[Y_t] \geqslant (1-\alpha_i)^2$.

The result from Lemma 3 states that it is impossible that

$$\liminf_t \frac{X_t}{Y_t} > \alpha,$$

which translates to it impossible that

$$u_i = \liminf_t \frac{|C^* \cap M_t^i|}{|C^* \cap M_t|} > \alpha.$$

By Claim 2, it is impossible that $\mathbb{E}[u_i] > \alpha_i$ as desired.

4 Conclusion

In this paper, we revisit a central vulnerability of Bitcoin's consensus mechanism: there exists a profitable deviation from the standard Bitcoin mining protocol, selfish mining, that results in time-persistent forks of the Bitcoin blockchain. This is not purely a theoretical concern: Li et al. (2024) provide empirical evidence consistent with selfish mining in several proof-of-work blockchains (most prominent for Monacoin and Bitcoin Cash). The lack of a known equilibrium for proof-of-work blockchains that generates a single longest chain on the equilibrium path has been highlighted by Ethereum's founder Buterin (2017) and by Hall et al. (2024) in the context of Ethereum's transition from proof-of-work to proof-of-stake.

A number of previous papers apply varying approaches to the issue of selfish mining (see,e.g., Heilman, 2014; Solat and Potop-Butucaru, 2017; Pass and Shi, 2017). To the best of our knowledge, all previously proposed solutions require changes to the consensus mechanism and/or the design of the blockchain itself. Our contribution is to propose an alternative mining protocol, inertial mining, that constitutes an equilibrium and results in one longest chain on the equilibrium path, as intended by Nakamoto.

As a direction for future research, we note that Proof-of Stake chains suffer from incentive problems that are similar to the selfish-mining deviations of Bitcoin (Brown-Cohen et al., 2019). It would be interesting to understand if a similar approach can be applied there to design equilibrium protocols.

References

M. Bahrani and S. M. Weinberg. Undetectable selfish mining. In *Proceedings of the 25th ACM Conference on Economics and Computation*, pages 1017–1044, 2024.

B. Biais, C. Bisière, M. Bouvard, and C. Casamatta. The Blockchain Folk Theorem. *The Review of Financial Studies*, 32(5):1662–1715, May 2019. doi: 10.1093/rfs/hhy095.

J. Bonneau. Why buy when you can rent? Bribery attacks on bitcoin-style consensus. In *Financial Cryptography and Data Security*, Lecture Notes in Computer Science, pages 19–26. Springer, 2016.

J. Brown-Cohen, A. Narayanan, A. Psomas, and S. M. Weinberg. Formal barriers to longest-chain proof-of-stake protocols. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 459–473, 2019.

E. Budish. Trust at scale: The economic limits of cryptocurrencies and blockchains. *The Quarterly Journal of Economics*, 140(1):1–62, 2025.

V. Buterin. Proof of stake faq. Blog post, 2017.

I. Eyal and E. G. Sirer. Majority is not enough: Bitcoin mining is vulnerable. *Communications of the ACM*, 61(7):95–102, 2018.

J. S. Gans and H. Halaburda. "zero cost" majority attacks on permissionless proof of work blockchains. *Management Science*, 70(6):4155–4165, 2024. doi: 10.1287/mnsc.2023.02426.

H. Halaburda, G. Haeringer, J. S. Gans, and N. Gandal. The microeconomics of cryptocurrencies. *Journal of Economic Literature*, 60(3):971–1013, Sept. 2022. doi: 10.1257/jel.20201593.

O. J. Hall, S. Shiaeles, and F. Li. A study of Ethereum's transition from proof-of-work to proof-of-stake in preventing smart contracts criminal activities. *Network*, 4(1): 33–47, 2024.

E. Heilman. One weird trick to stop selfish miners: Fresh bitcoins, a solution for the honest miner. In *Financial Cryptography and Data Security Workshops*. Springer, 2014. doi: 10.1007/978-3-662-44774-1_12.

G. Huberman, J. D. Leshno, and C. Moallemi. Monopoly without a monopolist: An economic analysis of the bitcoin payment system. *The Review of Economic Studies*, 88(6):3011–3040, 2021.

J. D. Leshno and P. Strack. Bitcoin: An axiomatic approach and an impossibility theorem. *American Economic Review: Insights*, 2(3):269–286, 2020.

J. D. Leshno, E. Shi, and R. Pass. On the viability of open-source financial rails: Economic security of permissionless consensus. *arXiv preprint arXiv:2409.08951*, 2024.

S.-N. Li, C. Campajola, and C. J. Tessone. Statistical detection of selfish mining in proof-of-work blockchain systems. *Scientific Reports*, 14:6251, 2024.

S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.

E. S. Pagnotta. Decentralizing money: Bitcoin prices and blockchain security. *The Review of Financial Studies*, 35(2):866–907, Feb. 2022. doi: 10.1093/rfs/hhaa149.

R. Pass and E. Shi. Fruitchains: A fair blockchain. In *Proceedings of the ACM Symposium on Principles of Distributed Computing*, pages 315–324, 2017. doi: 10.1145/3087801.3087809.

A. Sapirshtein, Y. Sompolinsky, and A. Zohar. Optimal selfish mining strategies in bitcoin. *International conference on financial cryptography and data security*, pages 515–532, 2016.

L. Schilling and H. Uhlig. Some simple bitcoin economics. *Journal of Monetary Economics*, 106:16–26, 2019.

S. Solat and M. Potop-Butucaru. Brief announcement: Zeroblock: Timestamp-free prevention of block-withholding attack in bitcoin. In *Networked Systems*. Springer, 2017. doi: 10.1007/978-3-319-69084-1_25.

A Missing proof

Proof of Lemma 3. Define $Z_t := X_t - aY_t$. Because $X_t \geqslant 0$ and $Y_t \leqslant 1$, we can establish a deterministic lower bound: $Z_t \geqslant -aY_t \geqslant -\max(0, a)$.

Assume for the sake of contradiction that

$$\mathbb{P}\left[\liminf_{t\to\infty} \frac{X_t}{Y_t} > a\right] = 1.$$

Let $L = \liminf_{t\to\infty} \frac{X_t}{Y_t}$ and define the random margin $\varepsilon := \frac{1}{2}\min(1, L - a)$. Then $\varepsilon > 0$ almost surely, and there is a random variable N taking values in $\mathbb{N}$ such that for $t \geqslant 0$ we have $\frac{X_{N+t}}{Y_{N+t}} \geqslant a + \varepsilon$. We can equivalently write this as

$$Z_{N+t} = X_{N+t} - aY_{N+t} \geqslant \varepsilon Y_{N+t}.$$

Combining this with the deterministic lower bound on Z_t, we have for all t that

$$Z_t \geqslant \varepsilon Y_t \mathbf{1}_{\{t\geqslant N\}} - a\mathbf{1}_{\{t<N\}}.$$

Taking the expectation and then lim sup of both sides and using (1) in the assumptions, we obtain:

$$0 \geqslant \limsup \mathbb{E}[Z_t] \geqslant \limsup \left(\mathbb{E}\left[\varepsilon Y_t \mathbf{1}_{\{t\geqslant N\}}\right] - a\mathbb{P}(N > t)\right) = \limsup \mathbb{E}\left[\varepsilon Y_t \mathbf{1}_{\{t\geqslant N\}}\right].$$

As $Y_t \geqslant 0$, this implies that $\lim_{t\to\infty} \mathbb{E}\left[\varepsilon Y_t \mathbf{1}_{\{t\geqslant N\}}\right] = 0$. This in particular implies that $\varepsilon Y_t \mathbf{1}_{\{t\geqslant N\}}$ converges to 0 in probability. It follows that Y_t converges to 0 in probability. Because the sequence Y_t is uniformly bounded, it guarantees convergence in expectation:

$$\lim_{t\to\infty} \mathbb{E}[Y_t] = 0.$$

However, this contradicts (3) in our initial hypothesis. Therefore, our assumption must be false, and it is impossible that $\liminf_{t\to\infty} \frac{x_t}{y_t} > a$ almost surely. □