

# Tight Consistency Bounds for Bitcoin

Peter Gaži<sup>1</sup>, Aggelos Kiayias<sup>2,1</sup>, and Alexander Russell<sup>3,1</sup>

<sup>1</sup>IOHK Research

<sup>2</sup>University of Edinburgh

<sup>3</sup>University of Connecticut

June 2, 2020

## Abstract

We establish the optimal security threshold for the Bitcoin protocol in terms of adversarial hashing power, honest hashing power, and network delays. Specifically, we prove that the protocol is secure if

$$r_a < \frac{1}{\Delta + 1/r_h},$$

where  $r_h$  is the expected number of honest proof-of-work successes in unit time,  $r_a$  is the expected number of adversarial successes, and no message is delayed by more than  $\Delta$  time units. In this regime, the protocol guarantees consistency and liveness with exponentially decaying failure probabilities. Outside this region, the simple private chain attack prevents consensus.

Our analysis immediately applies to any Nakamoto-style proof-of-work protocol; we also present the adaptations needed to apply it in the proof-of-stake setting, establishing a similar threshold there.

## 1 Introduction

The Bitcoin protocol, proposed in 2008 by Satoshi Nakamoto [13], has received extraordinary attention from both the applied and theoretical communities. The protocol’s survival in the permissionless setting—where parties may freely join and depart—and the promise of digital currencies and contracts that can thrive in such a hostile environment have led to widespread experimentation and numerous implementation projects. Likewise, the algorithmic core has proven to be a successful framework for designing and analyzing consensus algorithms.

Despite over a decade of study, the fundamental guarantees of the protocol are not well understood. Roughly, the essential ledger properties—consistency and liveness—are determined by three interacting features: the hashing power of the adversary, the hashing power of the honest parties, and networking delays. Ideally, one would like to establish the precise relationship between these parameters, exactly characterizing the parametrizations that guarantee the Bitcoin ledger properties.

We establish this relationship, proving that Bitcoin is secure if

$$r_a < \frac{1}{\Delta_0 + 1/r_h}, \tag{1}$$

where  $r_a$  is the expected number of adversarial proof-of-work successes in unit time,  $r_h$  is the expected number of honest successes, and no message is delayed by more than  $\Delta_0$  time units. Here, adversarial and honest proof-of-work successes are modeled as independent Poisson processes, with parameters  $r_h$  and  $r_a$ . In this region, consistency accrues exponentially quickly in the sense that blocks appearing at depth  $k$  in a longest chain can only be later abandoned with probability  $\exp(-\Omega(k))$ . Additionally, liveness in this region follows from a simple chain growth argument that provides similar exponential guarantees. This result is tight: if  $r_a$  exceeds this threshold, the simple private-chain attack prevents consensus. The threshold, as a function of  $r_h$ , is indicated by the solid black curve in Figure 1.

**Our results in more detail.** We work with the standard discrete approximation to the Poisson distribution to simplify bookkeeping. Specifically, we treat time as divided into small slots of length  $s$  and let  $p_a = s \cdot r_a$  denote the probability of an adversarial hashing success in a single slot;  $p_h = s \cdot r_h$  is likewise defined for the honest parties. This distribution limits to the Poisson distribution as  $s \rightarrow 0$ ; though classical results provide explicit upper bounds for the distance between these distributions (e.g., [2]) we do not need such high precision estimates because we model a system with finite lifetime. We remark that if slots are short enough, there is no loss in assuming that no more than a single success appears per slot. This is discussed formally in Remark 1 below.

We reflect network delays with a single parameter  $\Delta = \lceil \Delta_0/s \rceil$ : while any message sent by honest parties is always delivered, the adversary may delay its arrival by up to  $\Delta$  slots. Delivery is assumed to take place “at the beginning” of the slot, which is to say that the minimum value  $\Delta = 1$  corresponds to the case where messages transmitted in slot  $t$  are available for other parties’ full consideration in slot  $t + 1$ . In this setting, we prove that Bitcoin is secure if

$$p_a < \frac{1}{\Delta - 1 + 1/p_h}. \quad (2)$$

As mentioned above, if  $p_a$  exceeds the bound there is an attack that prevents any Bitcoin block from settling and succeeds with probability tending to 1. This natural attack goes back to the original Bitcoin whitepaper: it calls for the adversary to mine on a private chain with the intention to double spend if this private chain catches up to honestly held chains. The attack naturally generalizes to the setting with delays by calling for maximum possible delay of all honest messages. A notable, and perhaps unexpected, conclusion of our work is that the viability of this straightforward attack precisely captures the security regime of Bitcoin; in particular, when the adversarial hashing power exceeds the optimal security threshold this very attack prevents the protocol from reaching consensus and thus represents the best one can do to subvert consistency.

Finally, we point out that although we mention Bitcoin for concreteness, our results are obtained in a model sufficiently general to immediately cover any Nakamoto-style proof-of-work protocol. Additionally, an adaptation of our techniques can be used to establish similar results also for Nakamoto-style proof-of-stake protocols; see Section 5 for a detailed discussion of this case.

**Related work.** Analyzing the security of Bitcoin has a long history. The first rigorous results, due to Garay et al. [8], were obtained in the lock-step synchronous model. Pass et al. [14] gave a new treatment that established results in the  $\Delta$ -synchronous model, subsequently adopted by Garay et al. [7]. Kiffer et al. [12] tightened the consistency bound of [14] by associating security with the behavior of a Markov chain. Ren [15] simplified and condensed these results, adopting the continuous-time Poisson model.

This line of research culminated in identifying the “ $\Delta$ -isolated bound,” establishing security if

$$p_a < p_h(1 - p_h)^{2\Delta - 1}.$$

As  $\lim_{s \rightarrow 0} (1 - sr_h)^{2\Delta_0/s - 1} = \exp(-2r_h\Delta_0)$ , this corresponds to the Poisson model bound

$$r_a < r_h \exp(-2r_h\Delta_0).$$

The  $\Delta$ -isolated bound is compared side-by-side with the optimal bound in Figure 1. Roughly, the  $\Delta$ -isolated bound can only leverage honest hashing successes when surrounded by a  $\Delta$  region with no competing honest successes. While the slopes of the two bounds coincide at zero, the “ $\Delta$ -isolation” criterion penalizes larger values of  $r_h$ . It is natural to parameterize blockchain algorithms in the “sweet spot” where  $r_h + r_a \approx 1/\Delta_0$ , as this intuitively maximizes block throughput; the graph of Figure 1 illustrates this region.

The relevance of  $\Delta$ -isolated (honest) hashing victories to longest chain rule analysis was recognized at least as early as [14], and also plays a prominent role in our analysis: they arise in the treatment in the “critical zone,” where the adversary has roughly “caught up with” the honest players. (A more precise discussion appears below.)

Finally, an independent preprint by Dembo *et al.* [5] appeared online a few days prior to this paper and seems to investigate the same questions and obtain similar results. We defer a detailed comparison of the results and techniques to a future update.

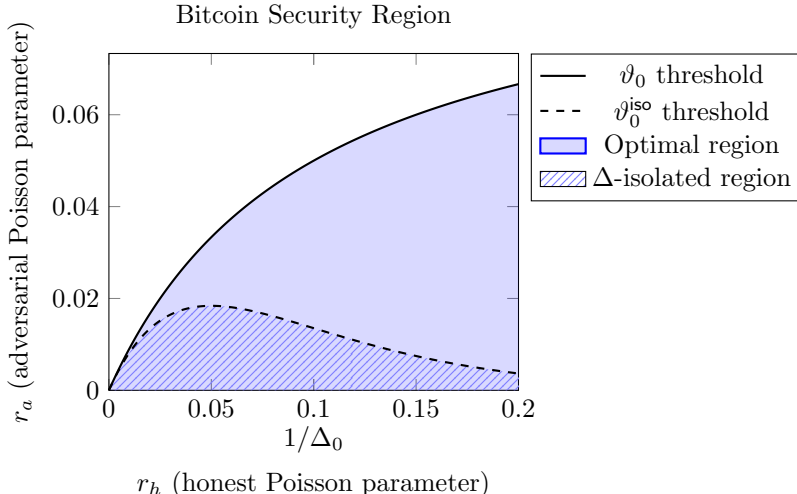


Figure 1: The region of pairs  $(r_a, r_h)$  for which Bitcoin is secure with  $\Delta_0 = 10$ . The optimal bound established in this paper  $\vartheta_0 = 1/(\Delta_0 + 1/r_h)$  is shown in solid black; the region beneath this line—filled in light blue—is the region of pairs  $(r_a, r_h)$  for which we establish security. The best prior bound  $\vartheta_0^{\text{iso}} = r_h \exp(-2\Delta_0 r_h)$  is shown in dashed black; its corresponding region of security is shown with blue hatching.

**A technical survey of the proof.** To motivate the optimal threshold itself, consider the baseline blockchain height achieved by the honest parties if the adversary contributes no blocks and subjects every honest message to a maximum  $\Delta$  delay. With honest hashing victories given by a sequence of i.i.d. indicator random variables corresponding to the time slots  $w_1, w_2, \dots$ , the height  $h_i$  achieved at slot  $i$  satisfies

$$h_i = \begin{cases} h_{i-\Delta} + 1, & \text{if } w_i = 1, \text{ and} \\ h_{i-1}, & \text{if } w_i = 0. \end{cases} \quad (3)$$

It is not difficult to show that the expectation  $\mathbb{E}[h_n] = n/\alpha + O(1)$  where  $\alpha = (\Delta - 1) + 1/p_h$  (as above,  $p_h$  is the probability of an honest hashing success). It is then clear that if  $p_a$  exceeds  $1/\alpha$  (exactly the optimal threshold discussed earlier) an adversary can dominate Bitcoin with the private-chain attack: in particular, the adversary may pick any undesirable block in the system, begin building a private chain prior to that block, and eventually overtake the honest chains which grow at a rate of  $1/\alpha$ .

As for demonstrating security below this threshold, we develop a set of new techniques for reasoning about the longest chain rule in the  $\Delta$ -synchronous setting. We begin by borrowing the notion of a “fork,” the bookkeeping tool originating in [11] and adapted to  $\Delta$ -delays in [4], and the technique of “relative margin” from [1]. In the context of a history of hashing successes—which indicates the prior slots in time during which honest and adversarial hashing victories occurred—the notion of relative margin provides a precise metric for “how far ahead” of the honest chains an adversarial chain could possibly be. (In fact, one has to specify a particular point in time before which the adversary’s chain must diverge to make sense of this notion, but we ignore such details in this summary.) Previous work analyzed the behavior of relative margin in the synchronous setting, first showing that it satisfies a relatively simple recurrence relation, and then analyzing the long term behavior of the process that emerges by applying this to i.i.d. random variables, as above. Existing analyses break down entirely in the  $\Delta$ -synchronous case—to sidestep this difficulty, one can use a pessimistic “ $\Delta$ -synchronous to synchronous reduction mapping,” [4] but this route leads to precisely the  $\Delta$ -isolated bound described above.

Our principal technical contribution is an analysis of relative margin in the  $\Delta$ -synchronous setting. We mention a few of the technical curiosities that arise; the full details are in Section 3. Intuitively, one would like to show that each adversarial success increases relative margin by one, and that each time an honest success

gives rise to a height increase, according to the rule (3), the relative margin should decrease. Unfortunately, this intuition fails: there are circumstances—occurring when the “race is close” and relative margin is close to zero—where the appearance of an honest victory actually works in the adversary’s favor. However, we show that this favorable intuition can indeed be established when relative margin is bounded away from zero. Our results rely heavily on a “compression transformation” that places forks in a semi-normal form; this guarantees that among all honest blocks of a particular depth, there is at least one that is “tight,” in the sense that it is placed at the minimum depth history would allow. In the critical zone around zero, we show looser bounds that rely on  $\Delta$ -isolated successes.

With these recurrence relations in place, we analyze the resulting stochastic process obtained by the appropriate i.i.d. distribution of hashing successes in Section 4. This yields a random walk with three regions, which we analyze separately: when relative margin is bounded away from zero, it is stochastically dominated by a negatively biased random walk; the bias is determined by the gap between  $p_a$  and the optimal threshold. When in a particular region near zero, it follows a positively biased random walk, but one which descends with constant probability. Fortunately, the critical zone around zero has only constant thickness, so the global random walk still has the desired features: in particular, after  $k$  steps the probability it will ever again rise to zero (or any other constant value) is  $\exp(-\Omega(k))$ .

**Remarks and future directions.** We work with a very strong adversary, one who is apprised of all future adversarial and honest mining successes and their exact times. It is an interesting fact that the security of the protocol is independent of such adversarial future knowledge. In particular, such an adversary never has to contend with regret for building on the wrong chain. On the other hand, we analyze the “static setting”:  $p_a$  and  $p_h$  are constant. It is reasonable to expect that the analysis can be extended to a setting where these are variable (but always satisfy, say  $p_a < (1 - \delta)p_h$ ); but we do not explore these issues. Our results focus on the “cryptographic” setting where mining power is split between honest parties following the protocol and adversarial parties deviating arbitrarily; hence we cannot capture rational attacks by honest parties, such as “selfish mining” [6]—of course the effect of such attacks can be reflected in our model if the selfish miners are treated as adversarial. Finally, we have made no particular attempt to control the constants; a more precise understanding of the critical region could presumably result in constants that directly inform practice.

## 2 Preliminaries

Throughout the paper, let  $\mathbb{N} = \{0, 1, 2, \dots\}$  denote the set of natural numbers (including zero). For  $n \in \mathbb{N}$ ,  $[n]$  denotes the set  $\{1, \dots, n\}$  (hence  $[0] = \emptyset$ ). For a word  $w = w_1 \dots w_n \in \Sigma^n$  we denote by  $w_{i:j}$  its subword  $w_i w_{i+1} \dots w_j$ , and by  $\#_a(w)$  we denote the number of occurrences of the symbol  $a \in \Sigma$  in  $w$ . We extend this notation also to multiple symbols, for example  $\#_{a,b}(w) \triangleq \#_a(w) + \#_b(w)$ . We denote by  $\|$  the concatenation of languages.

### 2.1 Our Model and the Bitcoin Protocol

We begin with an informal, abstract description of the Bitcoin protocol that suffices to describe our model. We delay formal definitions of the consistency and liveness events to later in this section.

The Bitcoin protocol is carried out by a family of parties of two types: *honest* parties follow the letter of law, carrying out the specified protocol, while *adversarial* parties may diverge arbitrarily from the specifications. All parties actively engage in searching for “proofs-of-work,” which afford them the right to contribute to the ledger. For the purposes of analysis we treat time as divided into small slots and use a *characteristic string* to indicate whether a proof-of-work was discovered in a particular time slot, and whether the successful party was honest or adversarial. In particular, the characteristic string  $w = w_1 w_2 \dots \in \{0, \mathfrak{h}, \mathfrak{a}\}^*$  associated with an

execution of the protocol is defined so that

$$w_s = \begin{cases} 0 & \text{if no proof-of-work was discovered in slot } s, \\ \mathbf{h} & \text{if an honest party discovered a proof-of-work in slot } s, \text{ and} \\ \mathbf{a} & \text{if an adversarial party discovered a proof-of-work in slot } s. \end{cases}$$

It is occasionally convenient to treat infinite characteristic strings in  $\{0, \mathbf{h}, \mathbf{a}\}^{\mathbb{N}}$  for which we use the same conventions. We study a probability distribution  $\mathcal{B}(p_a, p_h)$  of characteristic strings that reflects different rates of adversarial and honest success.

**Definition 1.** Let  $p_a, p_h > 0$  satisfy  $p_a + p_h \leq 1$ . Let  $\mathcal{B}(p_a, p_h)$  denote the distribution on characteristic strings  $w_1 w_2 \dots \in \{\mathbf{a}, \mathbf{h}, 0\}^{\mathbb{N}}$  given by independent selection of each  $w_i$  so that

$$w_i = \begin{cases} \mathbf{a} & \text{with probability } p_a, \\ \mathbf{h} & \text{with probability } p_h, \\ 0 & \text{with probability } 1 - p_a - p_h. \end{cases}$$

**Remark 1** (The discrete approximation to the Poisson process). The most natural mathematical model for the distribution of proof-of-work successes is a Poisson process, which reflects both the memoryless aspect of the proof-of-work challenge and the fact that it takes place in (effectively) continuous time. We work in the standard discrete approximation to the Poisson process since it simplifies the accounting in Section 3; however, the proof could as well have been presented in the Poisson setting. To clarify the relationship between these models, consider the Poisson process with parameter  $\lambda$  occurring on  $[0, L) \subset \mathbb{R}$ . Dividing the interval into  $L/s$  subintervals of length  $s$ , let  $X_t$  be the indicator random variable for the event that at least one success appears in the  $t$ th subinterval. Then  $\mathbb{E}[X_t] = 1 - \exp(-\lambda s) \approx \lambda s$  and the probability that two Poisson successes appear in any of the subintervals is  $L/s \cdot [1 - \exp(-\lambda s)(1 + \lambda s)] = O(L\lambda^2 s)$  by the union bound, which limits to zero linearly in  $s$ . It follows that, except with probability  $O(L\lambda^2 s)$ , the results of the independent random variables  $X_t$  are sufficient to determine the position of every success in  $[0, L)$  with accuracy  $\pm s/2$  and to determine their relative order exactly. Selecting a sufficiently small  $s$  then suffices to bound the probability of all the events of interest for our analysis. This also explains the assumption that no more than one proof-of-work success can arise in a particular time slot—this does not change the limiting model. A final remark: scaling the discrete  $\Delta$ -synchronous models of Pass et al. [14] and Garay et al. [7]—which do reflect multiple hashing successes—likewise leads to this very same Poisson model (for the same reason). Ren [15] adopts precisely the Poisson model.

The Bitcoin protocol calls for parties to exchange *blockchains*, each of which is an ordered sequence of blocks beginning with a distinguished “genesis block,” known to all parties. Each proof-of-work success confers on that party the right to add exactly one block to an existing blockchain. (In fact, the party must identify the previous chain on which she wishes to build ahead of time, but this will not affect our analysis.) Honest parties follow the *longest-chain rule* which dictates that they always choose to add to the longest blockchain they have yet observed and broadcast the result to all other parties. The basic dynamics of the system, with a particular characteristic string  $w$  and an adversary, can be described as follows.

We shall let  $\mathcal{C}_t$  denote the collection of all blockchains created by time  $t$  and let  $H(\mathcal{C}_t)$  denote the subset of all chains in  $\mathcal{C}_t$  whose last block was created by an honest party. Set  $\mathcal{C}_0 = \{G\}$ , where  $G$  denotes the unique chain consisting solely of the genesis block. The genesis block is “honest”; thus  $H(\mathcal{C}_0) = \mathcal{C}_0$ . It is convenient to adopt the convention that  $\mathcal{C}_{-t} = H(\mathcal{C}_{-t}) = \{G\}$  for any negative integer  $-t < 0$ . Then the protocol execution proceeds as follows. For each timestep  $t = 1, \dots$ :

- If  $w_t = 0$ , define  $\mathcal{C}_t = \mathcal{C}_{t-1}$  and  $H(\mathcal{C}_t) = H(\mathcal{C}_{t-1})$ .
- If  $w_t = \mathbf{a}$ , the adversary may select a single blockchain  $C$  from  $\mathcal{C}_{t-1}$  and add a block to create a new chain  $C'$ . Then  $\mathcal{C}_t = \mathcal{C}_{t-1} \cup \{C'\}$  and  $H(\mathcal{C}_t) = H(\mathcal{C}_{t-1})$ .

- If  $w_t = \mathbf{h}$ , the adversary may select any collection of chains  $\mathcal{V}$  for which  $H(\mathcal{C}_{t-\Delta}) \subseteq \mathcal{V} \subseteq \mathcal{C}_{t-1}$ . This is the “view” of the honest player, who applies the longest chain rule to  $\mathcal{V}$ , selects the longest chain  $L \in \mathcal{V}$  where ties are broken by the adversary, and adds a new block to create a new chain  $L'$ . Then  $\mathcal{C}_t = \mathcal{C}_{t-1} \cup \{L'\}$  and  $H(\mathcal{C}_t) = H(\mathcal{C}_{t-1}) \cup \{L'\}$ .

In each time step  $t$  we also maintain the set of  $\Delta$ -dominant chains  $\mathcal{D}_t \subseteq \mathcal{C}_t$ , determined entirely by  $\mathcal{C}_t$  and  $H(\mathcal{C}_{t-\Delta})$ : namely,  $\mathcal{D}_t$  is the set of all chains in  $\mathcal{C}_t$  that are at least as long as the longest chain in  $H(\mathcal{C}_{t-\Delta})$ . The intuition behind the definition of  $\Delta$ -dominant chains is that, in a time slot  $t$ , it is in principle possible to manipulate an honest party into adopting any  $\Delta$ -dominant chain, as the adversary is only obligated to deliver those chains in  $H(\mathcal{C}_{t-\Delta})$  and the chains in  $\mathcal{D}_t$  are at least as long as those in  $H(\mathcal{C}_{t-\Delta})$ .

This description implicitly places several constraints on the adversary; most notably, the only means of producing new chains is to append a block (associated with a proof-of-work success) to an existing chain. In practice, these constraints are guaranteed with cryptographic hash functions. Note that the  $\Delta$ -synchrony assumption is reflected in the rule for the case  $w_t = \mathbf{h}$ : the adversary is obligated to deliver all chains produced by honest players that are  $\Delta$  slots old. Finally, we permit the adversary to have full view of the characteristic string during this process. Of course, in practice a Bitcoin adversary must make decisions “on line.”

While expressed as a game between the adversary and the honest players, considering that the adversary selects both the view  $\mathcal{V}$  of each honest player and is empowered to break ties, the resulting sequence of chains is determined entirely by the adversary and the characteristic string.

In this context, we are interested in preserving two properties, consistency and liveness, which can be expressed in terms of the variables introduced above. In more detail:

- **Consistency; with parameter  $k$ .** For any  $t_1 \leq t_2$ , any  $C_1 \in \mathcal{D}_{t_1}$  and  $C_2 \in \mathcal{D}_{t_2}$  satisfy  $C_1^{\lceil k} \preceq C_2$ , where  $\preceq$  indicates the prefix relation and  $C_1^{\lceil k}$  is  $C_1$  without the blocks originating from the slots  $\{t_1 - k + 1, \dots, t_1\}$ .
- **Liveness; with parameter  $u$ .** There is a parameter  $u$  such that for any two slots  $t_1 < t_2$  with  $t_1 + u \leq t_2$ , and any chain  $C \in \mathcal{D}_{t_2}$ , there is a time  $t' \in \{t_1, \dots, t_1 + u\}$  and a chain  $C' \in H(\mathcal{C}_{t'})$  such that  $C' \preceq C$ .

Intuitively, consistency mandates that any blockchain possibly held by an honest party at time  $t_2$  extends a blockchain that was held by an honest party at time  $t_1$ , except perhaps for a  $k$ -slot suffix which could have been abandoned. Liveness, on the other hand, mandates that the blockchain held by an honest party incorporates at least one fresh honest block over a period of  $u$  slots.

We next proceed to introduce the formal notion of a *fork* for a characteristic string, which is one of the key tools for our analysis.

## 2.2 Characteristic Strings and Forks

We let  $\Sigma = \{0, \mathbf{h}, \mathbf{a}\}$  and consider characteristic strings  $w = w_1 \dots w_L$  drawn from the set  $\Sigma^L$ . Recall that intuitively, the  $i$ -th symbol  $w_i$  of  $w$  describes the outcome of the  $i$ -th slot in an  $L$ -slot execution of the Bitcoin protocol: the values 0,  $\mathbf{h}$ ,  $\mathbf{a}$  indicate no mining success, an honest success, and an adversarial success, respectively.

We next introduce the concept of a fork which will be the core analytical tool for establishing the security properties of the protocol. In particular a  $\Delta$ -fork abstracts a protocol execution with a simple but sufficiently descriptive discrete structure.

**Definition 2** (PoW  $\Delta$ -fork). Let  $\Delta$  be a positive integer and  $L \in \mathbb{N}$ . A *PoW  $\Delta$ -fork* for the string  $w = w_1 \dots w_L \in \Sigma^L$  is a directed, rooted tree  $F = (V, E)$  with a labeling function

$$\text{lb} : V \rightarrow \{0\} \cup \{i \in [L] : w_i \neq 0\}$$

satisfying the axioms (A1)–(A4) below. Edges are directed “away from” the root so that there is a unique directed path from the root to any vertex. The value  $\text{lb}(v)$  is referred to as the *label* of  $v$ . A non-root vertex  $v$  is called *honest* when  $w_{\text{lb}(v)} = \mathbf{h}$ ; otherwise it is *adversarial*.

- (A1) The root  $r \in V$  has label  $\text{lb}(r) = 0$  and is considered honest.
- (A2) The sequence of labels  $\text{lb}(\cdot)$  along any directed path is increasing.
- (A3) If  $w_i = \mathbf{h}$  then there is exactly one vertex with the label  $i$ , if  $w_i = \mathbf{a}$  then there is at most one vertex with the label  $i$ .
- (A4) For any pair of honest vertices  $v, w$  for which  $\text{lb}(v) + \Delta \leq \text{lb}(w)$ ,  $\text{len}(v) < \text{len}(w)$ , where  $\text{len}(\cdot)$  denotes the depth of the vertex.

It is easy to see the correspondence between the above axioms and the constraints imposed in the protocol execution. In particular, (A1) corresponds to the trusted nature of the genesis block; (A2) reflects that the blocks' ordering in a chain must be consistent with slot order; (A3) reflects that honest players produce exactly one block per PoW success, while the adversary might forgo a block-creation opportunity; finally (A4) reflects the fact that given sufficient time, as needed for block propagation in the network, an honest party will take into account the blocks produced by previously activated honest parties.

**Definition 3** (Fork notation). We write  $F \vdash_{\Delta} w$  to indicate that  $F$  is a  $\Delta$ -fork for the string  $w$ . If  $F' \vdash_{\Delta} w'$  for a prefix  $w'$  of  $w$ , we say that  $F'$  is a *subfork* of  $F$ , if  $F$  contains  $F'$  as a consistently-labeled subgraph. A fork  $F \vdash_{\Delta} w$  is *closed* if all its leaves are honest. By convention the trivial fork, consisting solely of a root vertex, is closed. The *closure* of a fork  $F$ , denoted  $\overline{F} \vdash_{\Delta} w$ , is the maximal closed subfork of  $F$ .

**Definition 4** (Tines). A path in a fork  $F$  originating at the root is called a *tine* (note that tines do not necessarily terminate at a leaf). As there is a one-to-one correspondence between directed paths from the root and vertices of a fork, we routinely overload notation so that it applies to both tines and vertices.

Specifically, we let  $\text{len}(T)$  denote the *length* of the tine  $T$ , equal to the number of edges on the path (see axiom (A4)). In the unusual cases where we wish to emphasize the fork from which  $v$  is drawn, we write  $\text{len}_F(v)$ . We further overload this notation by letting  $\text{len}(F)$  denote the length of the longest tine in a fork  $F$ .

Likewise, we let  $\text{lb}(\cdot)$  apply to tines by defining  $\text{lb}(T) \triangleq \text{lb}(v)$ , where  $v$  is the terminal vertex on the tine  $T$ . For a vertex  $v$  in a fork  $F$ , we denote by  $F(v)$  the tine in  $F$  terminating in  $v$ . We say that a tine is *honest* if the last vertex of the tine is honest, otherwise it is adversarial.

**Definition 5** (Branches). For an integer  $\ell \geq 1$  and for two tines  $T, T'$  of a fork  $F$ , we write  $T \sim_{\ell} T'$  if the two tines share a vertex with a label greater or equal to  $\ell$ . The set of all tines  $T' \in F$  such that  $T \sim_{\ell} T'$  is called the *branch* of  $T$  in  $F$  and denoted  $\mathbf{B}_F(T)$ .

**Definition 6** (Fork trimming; dominance). For a string  $w = w_1 \dots w_n$  and a positive integer  $k$ , we let  $w_{\lceil k} = w_1 \dots w_{n-k+1}$  denote the string obtained by removing the last  $k - 1$  symbols. For a fork  $F \vdash_{\Delta} w_1 \dots w_n$  we let  $F_{\lceil k} \vdash_{\Delta} w_{\lceil k}$  denote the fork obtained by retaining only those vertices labeled from the set  $\{0, \dots, n - k + 1\}$ . In the singular case  $k > n$  we postulate  $w_{\lceil k}$  to be the empty string and  $F_{\lceil k}$  to be the trivial fork containing only the root. For convenience, we sometimes prefer to emphasize the *remaining* length of the string (resp. fork), and denote by  $w_{\lceil m}$  and  $F_{\lceil m}$  the  $m$ -symbol prefix of  $w$  and the corresponding fork, formally  $w_{\lceil m} \triangleq w_{\lceil n-m+1}$  and  $F_{\lceil m} \triangleq F_{\lceil n-m+1}$ .

For an integer  $\delta > 0$ , we say that a tine  $T$  in  $F$  is  $\delta$ -*dominant* if

$$\text{len}(T) \geq \text{len}(\overline{F}_{\lceil \delta})$$

and simply call it *dominant* in the case  $\delta = 1$  (i.e., when  $\text{len}(T) \geq \text{len}(\overline{F})$ ).

Observe that honest tines appearing in  $F_{\lceil \Delta}$  are those that are necessarily visible to honest players at a timeslot just beyond the last one described by the characteristic string. Correspondingly, in the special case  $\delta = \Delta$ , the notion of a  $\Delta$ -dominant tine corresponds to  $\Delta$ -dominant chains as defined in the experiment described in Section 2.1. More broadly, here and below we will always only be interested in two possible values of the parameter  $\delta$ : either  $\delta = \Delta$  or  $\delta = 1$ ; and whenever we suppress  $\delta$  in the notation, it indicates the case  $\delta = 1$ .

## 2.3 Advantage and Margin

**Definition 7** (Advantage, margin). For a  $\Delta$ -fork  $F \vdash_{\Delta} w$  and  $\delta > 0$ , we define the  $\delta$ -*advantage* of a tine  $T \in F$  as

$$\alpha_F^{\delta}(T) = \text{len}(T) - \text{len}(\overline{F}_{\lceil \delta}).$$

Observe that  $\alpha_F^{\delta}(T) \geq 0$  if and only if  $T$  is  $\delta$ -dominant in  $F$ . We often suppress the subscript if  $F$  is clear from the context. For  $\ell \geq 1$ , we define the  $\delta$ -*margin* of a fork  $F$  as

$$\beta_{\ell}^{\delta}(F) = \max_{\substack{T_h \not\sim_{\ell} T_a \\ T_h \text{ is } \delta\text{-dominant}}} \alpha_F^{\delta}(T_a),$$

this maximum extended over all pairs of tines  $(T_h, T_a)$  where  $T_h$  is  $\delta$ -dominant and  $T_h \not\sim_{\ell} T_a$ . We call the pair  $(T_h, T_a)$  the  $\delta$ -*witness tines* for  $F$  if the above conditions are satisfied; i.e.,  $T_h$  is  $\delta$ -dominant,  $T_h \not\sim_{\ell} T_a$ , and  $\beta_{\ell}^{\delta}(F) = \alpha_F^{\delta}(T_a)$ . Note that there might exist multiple such pairs in  $F$ , but under the condition  $\ell \geq 1$  there will always exist at least one such pair, as the trivial tine  $T_0$  containing only the root vertex satisfies  $T_0 \not\sim_{\ell} T$  for any  $T$  and  $\ell \geq 1$ . For this reason, we will always consider  $\beta_{\ell}^{\delta}$  only for  $\ell \geq 1$ .

We overload the notation and define the  $\delta$ -margin of  $w$  as

$$\beta_{\ell}^{\delta}(w) = \max_{F \vdash_{\Delta} w} \beta_{\ell}^{\delta}(F).$$

We call a fork  $F \vdash_{\Delta} w$  a  $\delta$ -*witness fork* for  $w$  if  $\beta_{\ell}^{\delta}(w) = \beta_{\ell}^{\delta}(F)$ , again multiple  $\delta$ -witness forks may exist for a single string  $w$ .

We often write  $\alpha_F$  and  $\beta_{\ell}$  as shorthands for  $\alpha_F^1$  and  $\beta_{\ell}^1$ , respectively; for brevity we also refer to 1-witness tines and 1-witness forks as witness tines and witness forks, respectively.

**Remark 2.** The bulk of the analysis focuses on the quantity  $\beta_{\ell}(w)$ . This quantity, without the special considerations on tine dominance, appears to be somewhat more tractable than  $\beta_{\ell}^{\Delta}(w)$ . However, the direct relationship between settlement failures and margin is most easily expressed using  $\beta_{\ell}^{\Delta}(w)$ . The two notions have a simple relationship which justifies the choice to study  $\beta_{\ell}(\cdot)$ : if  $w, x \in \Sigma^*$  and  $|x| \geq \Delta$ , then  $\beta_{\ell}^{\Delta}(wx) \leq \beta_{\ell}(wy)$ , where  $y \in \Sigma^*$  is the string obtained by replacing every  $\mathfrak{h}$  in  $x$  with the symbol  $\mathfrak{a}$ . (See Lemma 19.)

**Remark 3.** In the special case  $|w| < \ell$ , we can observe that any fork  $F \vdash_{\Delta} w$  and any tines  $T, T' \in F$  satisfy  $T \not\sim_{\ell} T'$  (in particular,  $T \not\sim_{\ell} T$ ). Hence, in this case the quantity  $\beta_{\ell}^{\delta}(w)$  simplifies to

$$\beta_{\ell}^{\delta}(w) = \max_{F \vdash_{\Delta} w} \beta_{\ell}^{\delta}(F) = \max_{\substack{F \vdash_{\Delta} w \\ T \in F}} \alpha_F^{\delta}(T) = \max_{\substack{F \vdash_{\Delta} w \\ T \in F}} \text{len}_F(T) - \text{len}(\overline{F}_{\lceil \delta})$$

and so in this case we always have  $\beta_{\ell}^{\delta}(w) \geq 0$ .

It is easy to see that if a fork  $F \vdash_{\Delta} w$  has  $\beta_{\ell}^{\delta}(F) < 0$  then all tines of length at least  $\text{len}(\overline{F}_{\lceil \delta})$  belong to the same branch. This justifies the following definition.

**Definition 8** (Main branch). Let  $w \in \Sigma^n$ ,  $\ell \geq 1$ , and  $F \vdash_{\Delta} w$  such that  $\beta_{\ell}^{\delta}(F) < 0$ . The unique branch of  $F$  that contains all tines of length at least  $\text{len}(\overline{F}_{\lceil \delta})$  (and possibly other tines) is called the  $\delta$ -*main* branch of  $F$  and denoted  $M_{\delta}(F)$ ; we again omit  $\delta$  in the notation to indicate that  $\delta = 1$ .

## 2.4 Margin and Consistency

We now show how consistency can be established based on the margin quantity defined above.

Consider an execution of the Bitcoin protocol over a lifetime of  $L$  slots, let  $w = w_1 \dots w_L$  be the corresponding characteristic string. Let  $F \vdash_{\Delta} w$  be the fork consisting of vertices corresponding to all blocks created during the execution, connected via the natural ‘‘child-block’’ relation and labeled by their creation slot. For brevity, for each  $t \in [L]$  let  $F_t, w_t$  be the shorthands for  $F_{t\lceil}$ ,  $w_{t\lceil}$ , respectively.



**Lemma 1.** *Consider the Bitcoin execution described above. If for every  $\ell \in [L-k]$  and every  $t \in \{\ell+k, \dots, L\}$  we have  $\beta_\ell^\Delta(w_t) < 0$  then  $k$ -consistency was maintained during that execution.*

*Proof.* Let  $1 \leq t_1 \leq t_2 \leq L$  be slots and let  $C_i \in \mathcal{D}_{t_i}$  be the  $\Delta$ -dominant chains from the definition of the consistency property. If  $t_1 \leq k$  then there is nothing to prove, hence assume  $t_1 > k$  and consider  $\ell := t_1 - k$ .

Fix any  $t \in \{t_1, \dots, L\}$ . Since  $\beta_\ell^\Delta(F_t) \leq \beta_\ell^\Delta(w_t)$  is negative by assumption, there is a  $\Delta$ -main branch  $M_\Delta(F_t)$  in  $F_t$ , and tines in this branch share a vertex in or after slot  $\ell$ , hence the corresponding blockchains “agree” on their view of the content of the blockchain up to slot  $\ell$ . Moreover, any  $T \notin M_\Delta(F_t)$  has  $\alpha_{F_t}^\Delta(T) < 0$  and therefore  $\text{len}(T) < \text{len}(\overline{(F_t)_{\uparrow\Delta}})$ , hence  $T$  is not  $\Delta$ -dominant in  $F_t$ . Therefore, for each fixed  $t \in \{t_1, \dots, L\}$ , all  $\Delta$ -dominant blockchains  $\mathcal{D}_t$  in slot  $t$  agree up to slot  $\ell$ .

It remains to show that for  $t \in \{t_1, \dots, L-1\}$ , tines in  $M_\Delta(F_t)$  share their prefix up to slot  $\ell$  with tines in  $M_\Delta(F_{t+1})$ . If  $\text{len}(\overline{(F_t)_{\uparrow\Delta}}) = \text{len}(\overline{(F_{t+1})_{\uparrow\Delta}})$  then this is clear as  $M_\Delta(F_t) \subseteq M_\Delta(F_{t+1})$  and as argued above, all tines in  $M_\Delta(F_{t+1})$  agree up to  $\ell$ . On the other hand, if  $\text{len}(\overline{(F_t)_{\uparrow\Delta}}) < \text{len}(\overline{(F_{t+1})_{\uparrow\Delta}})$  then no extension of a tine  $T \in F_t$ ,  $T \notin M_\Delta(F_t)$  can belong to  $M_\Delta(F_{t+1})$ , as we had  $\text{len}(T) < \text{len}(\overline{(F_t)_{\uparrow\Delta}})$ , and  $T$  could be extended by at most one vertex in  $F_{t+1}$ , hence the extended tine is still shorter than  $\text{len}(\overline{(F_{t+1})_{\uparrow\Delta}})$ . Therefore, by an induction argument, all chains in  $\mathcal{D}_{t_1} \cup \mathcal{D}_{t_2}$  agree on their prefix up to  $\ell$  and so this is also true for  $C_1$  and  $C_2$ , establishing consistency.  $\square$

### 3 The Margin Recurrence

Our goal in this section is to establish upper bounds on  $\beta_\ell(w)$  for characteristic strings  $w \in \Sigma^*$ . Our bounds are expressed inductively, having the form  $\beta_\ell(wx) \leq \beta_\ell(w) + f(x)$  where  $w, x \in \Sigma^*$  and  $f$  is an appropriate function of the suffix  $x$ . Roughly, the bounds show that when  $\beta_\ell(w)$  is “suitably large” or “suitably small,” it satisfies the *ideal recurrence*: for  $w, x \in \Sigma^*$  and  $|x| \geq \Delta - 1$ ,

$$\beta_\ell(w\mathbf{a}x\mathbf{h}) \leq \beta_\ell(w) + \#_{\mathbf{a}}(x) - 1. \quad (4)$$

Note that  $\beta_\ell(\cdot)$  increases by 1 for each ‘a’-symbol; and it decreases by 1—intuitively accounting for the last ‘h’-symbol which is at least  $\Delta$  slots ahead of any of the slots associated with  $w$ .

The region around zero is more problematic; in this case we only show that  $\beta_\ell$  cannot move too quickly, and that there are certain suffixes (like  $0^{\Delta-1}\mathbf{h}$ ) which indeed force  $\beta_\ell(\cdot)$  to decrease. Because this difficult region will have only constant width, we will see that it does not adversely affect the final probabilistic results.

**The step decomposition.** The decomposition of  $w$  appearing in the ideal recurrence (4) above plays a special role in the analysis. We lay down some notation to reflect this.

**Definition 9** (The step decomposition). Let  $w = w_1w_2\dots \in \{\mathbf{a}, \mathbf{h}, 0\}^{\mathbb{N}}$ . For such a string, we consider the decomposition

$$w = \sigma_1\sigma_2\dots \quad \text{where each} \quad \sigma_i \in \Sigma_{\mathcal{S}} \triangleq \{\mathbf{a}, \mathbf{h}, 0\}^{\Delta-1} \parallel \{\mathbf{a}, 0\}^* \parallel \{\mathbf{h}\}.$$

We reserve the word *symbol* to refer to elements of  $\Sigma$ , and the word *step* to refer to elements of  $\Sigma_{\mathcal{S}}$ . We write  $\mathcal{S}(w) \triangleq \sigma_1\sigma_2\dots$  to indicate the resulting sequence of elements of  $\Sigma_{\mathcal{S}}$ . Throughout, we let  $|\gamma|$  denote the number of symbols in the step  $\gamma \in \Sigma_{\mathcal{S}}$ .

We remark that this decomposition is unique and has the following direct interpretation: (1) write  $w = xhw'$  where  $x$  is the shortest prefix of length at least  $\Delta - 1$  that is followed by the symbol  $\mathbf{h}$ . (2) emit the symbol  $\sigma = x\mathbf{h}$ ; (3) repeat the process on  $w'$ . The sequence of symbols produced by this process corresponds to the  $\sigma_i$  above.

**Organization.** We start by introducing some key technical tools below in Section 3.1. As a warm-up, in Section 3.2 we establish a variant of the ideal recurrence (4) in the considerably simpler setting “before the slot  $\ell$ ,” i.e., for  $w$  such that  $|w| < \ell$ . Then we turn to the more interesting case of general  $|w|$ , considering

three separate regions: the “critical” region where  $\beta_\ell(w)$  is close to zero (Section 3.3); and the “cold” resp. “hot” regions where it is sufficiently far from zero on the negative, resp. positive side (Sections 3.4 and 3.5, respectively).

### 3.1 Compressed Forks and the Restructuring Lemma

In our arguments we make use of special honest vertices called *tight* that are, informally speaking, at the minimal depth that the preceding part of the fork allows without violating the axiom (A4). Here we define these vertices formally and summarize several useful properties they have: in particular, in Lemma 4 we show how a fork that has a tight vertex at each possible depth (we call such forks *compressed*) allows for a complex restructuring operation that leads to a lower-bound on the margin of the underlying characteristic string.

**Definition 10.** Let  $F \vdash_\Delta w \in \Sigma^n$ . An honest vertex  $v$  of  $F$  is called *tight* if  $\text{len}(v) = \text{len}(\overline{F_{\text{lb}(v)-\Delta}}) + 1$ . The fork  $F$  is said to be *compressed* if, for every depth  $0 \leq d \leq \text{len}(\overline{F})$ , there is a tight honest vertex  $v$  of depth  $d$ .

**Lemma 2.** Let  $F \vdash_\Delta w \in \Sigma^n$ . Let  $v$  be a tight vertex and let  $v'$  be an honest vertex with  $\text{lb}(v) \leq \text{lb}(v')$ ; then  $\text{len}(v) \leq \text{len}(v')$ .

*Proof of Lemma 2.* As  $\text{lb}(v) \leq \text{lb}(v')$  and  $v$  is tight,

$$\text{len}(v) = \text{len}(\overline{F_{\text{lb}(v)-\Delta}}) + 1 \leq \text{len}(\overline{F_{\text{lb}(v')-\Delta}}) + 1 \leq \text{len}(v'). \quad \square$$

In the proofs below, we often use the contrapositive of Lemma 2: if  $\text{len}(v') < \text{len}(v)$  then  $\text{lb}(v') < \text{lb}(v)$ .

**Lemma 3.** Let  $w \in \Sigma^*$ , there exists a compressed witness fork  $F \vdash_\Delta w$  for  $w$ .

*Proof.* Let  $F \vdash_\Delta w$  be a witness fork for  $w$ . We describe a transformation, which we naturally call “compression,” that converts  $F$  into a compressed fork  $F^c \vdash_\Delta w$  for which  $\beta_\ell(F^c) = \beta_\ell(F)$ . If  $F$  is compressed, the transformation makes no change. Otherwise, the transformation is given as a sequence of “compression steps,” each of which reduces the total depth of the fork and locally improves tightness violations.

In particular, if  $F$  is not compressed, there is a smallest depth  $d \leq \text{len}(\overline{F})$  for which there is no tight honest vertex of depth  $d$ . Let  $F'$  denote the labeled rooted tree obtained from  $F$  by carrying out the following alterations:

- If  $d = 1$ , for every vertex  $v$  of depth  $d = 1$ , replace any edge  $(v, u)$  by an edge  $(r, u)$  where  $r$  is the root.
- If  $d > 1$ , raise every vertex  $v$  of depth  $d$  one level in the tree by replacing the unique edge of the form  $(u, v)$  with the edge  $(p, v)$ , where  $p$  is the parent of  $u$ .

The labels of all vertices in  $F'$  remain the same as those of the corresponding vertices in  $F$ . As indicated above, we refer to the procedure carrying  $F \mapsto F'$  as a compression step.

We verify that  $F' \vdash_\Delta w$ : Axiom (A1) is trivially satisfied. Axiom (A2) holds for  $F'$  as all directed edges added to  $F'$  respect the label ordering. Axiom (A3) holds as all labels are preserved. Finally, we consider axiom (A4). Note the effect that the process has on the depth of honest vertices in the general case  $d > 1$ : The depths of all honest vertices with (initial) depth less than  $d$  are preserved, while the depths of all honest vertices with depths at least  $d$  are decreased by exactly one. Thus the only possible violations of axiom (A4) could occur among those honest vertices at depth exactly  $d$ ; however, as all such vertices are non-tight by assumption, reducing their depth by one guarantees axiom (A4). Finally, observe that if  $d = 1$ , all vertices of depth  $d$  are adversarial, as any honest vertex of depth 1 is tight by definition, hence the above reasoning applies as well despite a different alteration rule.

In light of the comments above, we note that  $\text{len}(F') = \text{len}(F) - 1$  and  $\text{len}(\overline{F'}) = \text{len}(\overline{F}) - 1$ . It follows that a finite number of compression steps results in a compressed fork,  $F^c$ , as desired.

In general, we show below that  $\beta_\ell(F') \geq \beta_\ell(F)$ ; thus, if  $\beta_\ell(F) = \beta_\ell(w)$  then  $\beta_\ell(F') = \beta_\ell(w)$  and  $F'$  is likewise an optimal fork. Consider a tine  $T$  of  $F$ ; we may naturally associate this with the tine  $T'$  of  $F'$  that terminates with the same vertex. If  $\alpha_F(T) \geq 0$  it follows from the discussion above that  $\alpha_{F'}(T') = \alpha_F(T)$ , as

$\text{len}(T') = \text{len}(T) - 1$  and  $\text{len}(\overline{F}') = \text{len}(\overline{F}) - 1$ . If  $\alpha_F(T) < 0$  it follows that  $\alpha_F(T) \leq \alpha_{F'}(T') \leq \alpha_F(T) + 1$ , depending on whether the alterations involve any vertices of the  $T$ . It follows immediately that  $\beta_\ell(F') \geq \beta_\ell(F)$ . Specifically, let  $(T_h, T_a)$  be two witness tines for  $F$  so that  $\alpha_F(T_h) \geq 0$ ,  $\alpha_F(T_a) = \beta_\ell(F)$ , and  $T_h \not\prec_\ell T_a$ . Let  $T'_a$  and  $T'_h$  be the two tines corresponding to  $T_a$  and  $T_h$  in  $F'$ , respectively; clearly  $T'_h \not\prec_\ell T'_a$  and note that this does not depend on  $\ell$ . Then  $\alpha_{F'}(T'_h) = \alpha_F(T_h)$  and  $\alpha_{F'}(T'_a) \geq \alpha_F(T_a)$ ; therefore  $T'_h$  is dominant in  $F'$  and we have  $\beta_\ell(F') \geq \beta_\ell(F)$ , as desired.  $\square$

**Lemma 4** (Restructuring lemma). *Let  $w \in \Sigma^*$  be a characteristic string and  $F \vdash_\Delta w$  be a compressed fork for  $w$ , let  $T_1 \not\prec_\ell T_2$  be arbitrary tines in  $F$ . For  $i \in \{1, 2\}$ , let  $v_i$  be an honest vertex on  $T_i$  and let  $A_i$  denote the set of all adversarial vertices on  $T_i$  deeper than  $v_i$ . If  $\text{lb}(v_1) \leq \text{lb}(v_2)$  then*

$$\beta_\ell(w) \geq \alpha_F(v_1) + |A_1 \cup A_2| .$$

*Proof.* On a high level, we restructure the fork  $F$  to obtain a valid fork  $\tilde{F} \vdash_\Delta w$  that satisfies  $\beta_\ell(\tilde{F}) \geq \alpha_F(v_1) + |A_1 \cup A_2|$ , establishing the claim. This restructuring consists of two main modifications: (i) use (at least) all adversarial vertices in  $A_1 \cup A_2$  to build a tine  $\tilde{T}_a$  on top of  $v_1$  with  $\text{len}(\tilde{T}_a) \geq \text{len}(v_1) + |A_1 \cup A_2|$ ; and (ii) use tight vertices of depths  $\text{len}(v_2) + 1, \text{len}(v_2) + 2, \dots, \text{len}(\overline{F})$  to build an honest tine  $\tilde{T}_h$  on top of  $v_2$  with  $\text{len}(\tilde{T}_h) = \text{len}(\overline{F})$ . Additional care is needed to ensure the validity of the resulting fork.

Towards a more detailed description, we identify sets of vertices in  $F$  that will be modified in the same way. Let  $y$  denote the last common vertex of  $T_1$  and  $T_2$ , and let  $z$  denote the deeper one of the vertices  $\{v_1, y\}$ . First, we define  $A'_i \triangleq \{a \in A_i : \text{lb}(a) > \text{lb}(z)\}$ , we write  $A \triangleq A'_1 \cup A'_2$  and refer to the individual vertices in  $A$  as  $a_1, \dots, a_{|A|}$  so that  $i < j$  implies  $\text{lb}(a_i) < \text{lb}(a_j)$ .<sup>1</sup> Next, as  $F$  is compressed, it contains a tight honest vertex for each depth in  $\{\text{len}(v_2) + 1, \dots, \text{len}(\overline{F})\}$ . We label these vertices  $h_1, \dots, h_g$ , where  $g = \text{len}(\overline{F}) - \text{len}(v_2)$  and  $h_i$  has  $\text{len}(h_i) = \text{len}(v_2) + i$ , and denote  $H \triangleq \{h_1, \dots, h_g\}$ . Note that there might be several tight vertices of a particular depth in  $F$ , the choice of vertices for  $H$  is arbitrary, we just ensure that it contains one vertex for each of the relevant depths. Finally, we denote by  $D = \{d_1, \dots, d_{|D|}\}$  the set of all vertices that at the same time (a) are honest, (b) are (possibly indirect) descendants of either  $v_1$  or  $v_2$ , (c) are not (possible indirect) predecessors of either  $z$  or  $v_2$ , and (d) are not in  $H$ . We again index the vertices in  $D$  in an increasing order of labels.

We first modify  $F$  as follows:

**Set A:** The unique edge of the form  $(u, a_1)$  is replaced with the edge  $(z, a_1)$  and for each  $i \in \{2, \dots, |A|\}$ , the unique edge of the form  $(u, a_i)$  is replaced with the edge  $(a_{i-1}, a_i)$ .

**Set H:** The unique edge of the form  $(u, h_1)$  is replaced with the edge  $(v_2, h_1)$  and for each  $i \in \{2, \dots, g\}$ , the unique edge of the form  $(u, h_i)$  is replaced with the edge  $(h_{i-1}, h_i)$ .

We denote the resulting labeled tree  $F_0$ , note that  $F_0$  is not necessarily a valid fork. To reestablish validity, we proceed with the following sequence of modifications:

**Set D:** For each  $i \in \{1, \dots, |D|\}$  the unique edge of the form  $(u, d_i)$  in  $F_{i-1}$  is replaced by  $(\tilde{u}_i, d_i)$ , where  $\tilde{u}_i$  is the vertex in  $F_{i-1}$  with maximum depth out of all honest vertices with label at most  $\text{lb}(d_i) - \Delta$ ; note that this in particular excludes  $d_j$  for  $j > i$ . Formally,

$$\tilde{u}_i \triangleq \underset{\substack{u \in F_{i-1}; w_{\text{lb}(u)} = \text{h} \\ \text{lb}(u) \leq \text{lb}(d_i) - \Delta}}{\arg \max} \text{len}_{F_{i-1}}(u) , \quad (5)$$

where ties in max can be broken arbitrarily. The labeled tree resulting from the  $i$ -th iteration is called  $F_i$ .

Finally, we let  $\tilde{F} \triangleq F_{|D|}$ .

We now show that  $\tilde{F}$  is a valid fork for  $w$ . The axioms (A1) and (A3) are clearly maintained by the above modifications and hence inherited from  $F$ . Axiom (A2) is satisfied in  $\tilde{F}$  as each newly added edge  $(u, v)$  has

<sup>1</sup>We recommend the reader to first consider the simplest situation where  $\text{len}(y) < \text{len}(v_i)$  for both  $i \in \{1, 2\}$  and hence  $z = v_1$  and  $A'_i = A_i$ .

$\text{lb}(u) < \text{lb}(v)$ . For new edges  $\{(\cdot, d_i) : d_i \in D\}$  this directly follows from (5), for new edges  $\{(\cdot, a_i) : a_i \in A\}$  this is a consequence of the definition of  $A'_i$  and the ordering within  $A$ . Finally, in  $H$  we have by construction

$$\text{len}(v_2) < \text{len}(h_1) < \dots < \text{len}(h_g) = \text{len}(\overline{F}),$$

$v_2$  is honest, and each  $h_i$  is tight (and hence honest). Applying Lemma 2 to each  $h_i$  implies that

$$\text{lb}(v_2) < \text{lb}(h_1) < \dots < \text{lb}(h_g)$$

as required.

To verify axiom (A4), note that when moving from  $F$  to  $\tilde{F}$ , the depths of all honest vertices outside of  $D$  remained unchanged. The depth of a vertex  $d_i \in D$  might have changed, but it has not increased; this can be shown by simple induction on  $i$ : by induction hypothesis, also the depths of all honest vertices with labels up to  $\text{lb}(d_i) - \Delta$  have not increased from  $F$  to  $\tilde{F}$ , and hence

$$\text{len}_F(d_i) \geq \max_{\substack{u \in F, w_{\text{lb}(u)} = \text{h} \\ \text{lb}(u) \leq \text{lb}(d_i) - \Delta}} \text{len}_F(u) + 1 \geq \max_{\substack{u \in \tilde{F}, w_{\text{lb}(u)} = \text{h} \\ \text{lb}(u) \leq \text{lb}(d_i) - \Delta}} \text{len}_{\tilde{F}}(u) + 1 = \text{len}_{\tilde{F}}(d_i),$$

where the first inequality follows from axiom (A4) in  $F$  and the last equality is a consequence of (5). Given the above, the only possible violation of axiom (A4) in  $\tilde{F}$  could occur for a pair  $(v, w)$  with  $w = d_i \in D$ , but this is exactly prevented by the rule (5). This concludes the argument that  $\tilde{F} \vdash_{\Delta} w$ .

To finish the proof, denote by  $\tilde{T}_a$  and  $\tilde{T}_h$  the tines in  $\tilde{F}$  terminating in  $a_{|A|}$  and  $h_{|H|}$ , respectively. Given  $T_1 \not\sim_{\ell} T_2$  we have  $\text{lb}(y) < \ell$ , and note that the last common vertex of  $\tilde{T}_a$  and  $\tilde{T}_h$  has label at most  $\text{lb}(y)$ , hence we have  $\tilde{T}_a \not\sim_{\ell} \tilde{T}_h$ . Furthermore,  $\text{len}(\tilde{T}_h) = \text{len}(\tilde{F})$  by construction. Hence we have

$$\beta_{\ell}(w) \geq \beta_{\ell}(\tilde{F}) \geq \alpha_{\tilde{F}}(\tilde{T}_a) = \alpha_F(z) + |A|.$$

Finally,  $\alpha_F(z) + |A| \geq \alpha_F(v_1) + |A_1 \cup A_2|$ : if  $z = v_1$  then each  $A_i = A'_i$  and hence  $A = A_1 \cup A_2$ ; otherwise  $z = y$  and  $\alpha_F(z) \geq \alpha_F(v_1) + |(A_1 \cup A_2) \setminus A|$ . This concludes the proof.  $\square$

### 3.2 Warm-up: Margin Prior To $\ell$

We start by describing the behavior of  $\beta_{\ell}(w)$  for  $|w| < \ell$ . Note that this significantly simplifies the notion as discussed in Remark 3, in particular  $|w| < \ell$  implies  $\beta_{\ell}(w) \geq 0$ .

**Lemma 5.** *Let  $\ell \geq 1$ ,  $w \in \{0, \text{h}, \text{a}\}^{<\ell}$  and  $x \in \{0, \text{h}, \text{a}\}^{\geq \Delta - 1}$ . Then*

$$\beta_{\ell}(wxh) \leq \begin{cases} \beta_{\ell}(w) + \#_{\text{a}}(x) - 1 & \text{if } \beta_{\ell}(w) \geq 1, \\ \#_{\text{a}}(x) & \text{if } \beta_{\ell}(w) = 0. \end{cases}$$

*Proof.* First consider the case  $\beta_{\ell}(w) \geq 1$ . If we have  $\beta_{\ell}(wxh) \leq \#_{\text{a}}(x)$  then the lemma follows immediately, hence assume  $\beta_{\ell}(wxh) > \#_{\text{a}}(x)$ . Let  $F' \vdash_{\Delta} wxh$  be a witness fork for  $w' \triangleq wxh$  and let  $(T'_h, T'_a)$  be a witness pair in  $F'$ . Let  $F \triangleq F'_{|w|} \vdash_{\Delta} w$  and define  $T \triangleq (T'_a)_{|w|}$  as the restriction of  $T'_a$  to vertices with labels at most  $|w|$ ; we have  $T \in F$ . By our assumption on  $\beta_{\ell}(wxh)$ , more than  $\#_{\text{a}}(x)$  deepest vertices of  $T'_a$  are adversarial and hence we get  $\text{len}(T'_a) - \text{len}(T) \leq \#_{\text{a}}(x)$ . The pair  $(T, T)$  can serve as a witness pair in  $F$ , and we have  $\text{len}(\overline{F}') > \text{len}(\overline{F})$  since  $|x| \geq \Delta - 1$ . We can hence conclude

$$\beta_{\ell}(w) \geq \beta_{\ell}(F) \geq \alpha_F(T) = \text{len}(T) - \text{len}(\overline{F}) \geq [\text{len}(T'_a) - \#_{\text{a}}(x)] - [\text{len}(\overline{F}') - 1] = \beta_{\ell}(wxh) - \#_{\text{a}}(x) + 1,$$

as desired.

In the case  $\beta_{\ell}(w) = 0$ , the situation  $\beta_{\ell}(wxh) > \#_{\text{a}}(x)$  cannot occur, as the same reasoning as above would give us  $0 = \beta_{\ell}(w) \geq \beta_{\ell}(wxh) - \#_{\text{a}}(x) + 1 \geq 1$ , a contradiction. Hence in this case we must have  $\beta_{\ell}(wxh) \leq \#_{\text{a}}(x)$  as desired.  $\square$

### 3.3 The Critical Region

In the “critical region” (near zero) we will rely on rather loose information about the behavior of  $\beta_\ell$ . The first bound (Lemma 6) establishes that  $|\beta_\ell(wx) - \beta_\ell(w)| \leq \#_{\mathbf{h}, \mathbf{a}}(x)$ —each symbol of  $x$  can change  $\beta_\ell()$  by at most one. The second bound (Lemma 7) shows that for  $|w| \geq \ell$ ,  $\beta_\ell(w0^t\mathbf{h}) < \beta_\ell(w)$  when  $t \geq \Delta - 1$ . Note that a similar statement (with a singular special case) for  $|w| < \ell$  follows from Lemma 5.

**Lemma 6** ( $\beta_\ell$  is 1-Lipschitz). *Let  $w \in \{0, \mathbf{h}, \mathbf{a}\}^*$  be a characteristic string. Then  $\beta_\ell(w0) = \beta_\ell(w)$  and for  $x \in \{\mathbf{h}, \mathbf{a}\}$  we have*

$$|\beta_\ell(wx) - \beta_\ell(w)| \leq 1.$$

*Proof.* The lower bound  $\beta_\ell(wx) \geq \beta_\ell(w) - 1$  is straightforward; in fact one can establish higher precision bounds

$$\begin{aligned} \beta_\ell(w0) &= \beta_\ell(w), \\ \beta_\ell(w\mathbf{a}) &\geq \beta_\ell(w) + 1 \\ \text{and } \beta_\ell(w\mathbf{h}) &\geq \beta_\ell(w) - 1. \end{aligned}$$

These follow by considering an optimal fork  $F \vdash_\Delta w$  with witness tines  $(T_h, T_a)$ : if  $x = \mathbf{a}$ , an adversarial vertex can be added to the end of  $T_a$ ; if  $x = \mathbf{h}$ , this honest vertex can be added to the end of  $T_h$ . The resulting forks clearly achieve the statistics above.

We turn our attention to the upper bound  $\beta_\ell(wx) \leq \beta_\ell(w) + 1$ . Let  $F' \vdash_\Delta wx$  be a compressed optimal fork with witness tines  $(T'_h, T'_a)$ . Let  $F \vdash_\Delta w$  denote the fork that results by removing the vertex  $v$  associated with the symbol  $x$ . If  $v$  does not appear on either of the witness tines, the same tines establish that  $\beta_\ell(F) \geq \beta_\ell(F')$  and we conclude that  $\beta_\ell(w) \geq \beta_\ell(wx)$ , as desired. If  $v$  appeared on  $T'_a$  (and possibly also on  $T'_h$  if  $T'_h = T'_a$ ), we let  $(T_h, T_a)$  denote the restrictions of  $(T'_h, T'_a)$  to  $F$  and note that the witness tines  $(T_h, T_a)$  establish that

$$\beta_\ell(w) \geq \beta_\ell(F) \geq \alpha_F(T_a) \geq \alpha_{F'}(T'_a) - 1 = \beta_\ell(wx) - 1,$$

as desired. It remains to consider the case that  $v$  appears on  $T'_h$  and not on  $T'_a$ . As above, let  $T_h$  denote the tine in  $F$  resulting from removing  $v$  from  $T'_h$ , and observe that  $F$  is compressed. If  $\alpha_{F'}(T'_a) = \beta_\ell(wx) \geq 0$ , we invoke Lemma 4. Let  $v_h$  and  $v'_a$  denote the deepest honest vertices on  $T_h$  and  $T'_a$  respectively; let  $A_h$  (resp.  $A'_a$ ) be the set of adversarial vertices on  $T_h$  (resp.  $T'_a$ ) deeper than  $v_h$  (resp.  $v'_a$ ). If  $\text{lb}(v_h) \leq \text{lb}(v'_a)$  then Lemma 4 gives us

$$\beta_\ell(w) \geq \alpha_F(v_h) + |A'_a \cup A_h| \geq (\alpha_F(v_h) + |A_h|) + |A'_a \setminus A_h| \geq -1 + |A'_a \setminus A_h| \geq \beta_\ell(wx) - 1$$

as desired. On the other hand, if  $\text{lb}(v'_a) \leq \text{lb}(v_h)$  we similarly have

$$\beta_\ell(w) \geq \alpha_F(v'_a) + |A'_a \cup A_h| \geq \alpha_F(v'_a) + |A'_a| \geq \alpha_{F'}(T'_a) = \beta_\ell(wx).$$

Finally, we consider the case that  $\alpha_{F'}(T'_a) = \beta_\ell(wx) < 0$ . Letting  $T_H$  denote a maximum length honest tine in  $F$  we consider two cases: if  $T_H \not\sim T'_a$ , these two tines witness  $\beta_\ell(w) \geq \alpha_F(T'_a) \geq \alpha_{F'}(T'_a) = \beta_\ell(wx)$ , as desired. Otherwise,  $T_H \sim T_h$  and these two tines witness  $\beta_\ell(w) \geq \alpha_F(T_h) \geq \alpha_{F'}(T'_h) - 1 \geq \beta_\ell(wx) - 1$ , as desired.  $\square$

**Lemma 7.** *Let  $\ell \geq 1$  and  $w \in \{0, \mathbf{h}, \mathbf{a}\}^{\geq \ell}$  be a characteristic string. Then*

$$\beta_\ell(w0^{\Delta-1}\mathbf{h}) \leq \beta_\ell(w) - 1.$$

*Proof.* Let  $F' \vdash_\Delta w0^{\Delta-1}\mathbf{h}$  be a witness fork for  $w0^{\Delta-1}\mathbf{h}$  and let  $T'_a$  and  $T'_h$  denote a pair of witness tines in  $F'$  so that  $\alpha(T'_h) \geq 0$  and  $\alpha(T'_a) = \beta_\ell(w0^{\Delta-1}\mathbf{h})$ . Let  $v$  denote the vertex in  $F'$  corresponding to the final  $\mathbf{h}$  symbol and let  $F \vdash_\Delta w$  denote the fork obtained by removing the vertex  $v$ . Note that  $\text{len}(\overline{F}) \leq \text{len}(\overline{F'}) - 1$  by axiom (A4).

Note that as  $|w| \geq \ell$  and  $T'_h \not\sim_\ell T'_a$ ,  $v$  cannot appear on both these tines. If  $v$  appears on  $T'_a$ , let  $T_a$  denote the tine in  $F$  resulting from removal of  $v$ . As  $T'_a$  terminated with an honest vertex and, by definition,  $\beta_\ell(w0^{\Delta-1}\mathbf{h}) = \alpha_{F'}(T'_a)$ , we conclude that  $\beta_\ell(w0^{\Delta-1}\mathbf{h}) = 0$ . In this special case, then, we wish to show that  $\beta_\ell(w) \geq 1$ . Observe that  $T_a$  is dominant in  $F$ , as  $\text{len}(T_a) = \text{len}(\overline{F'}) - 1 = \text{len}(\overline{F})$ . On the other hand,  $\alpha_F(T'_h) = \alpha_{F'}(T'_h) + 1$  so the two tines (now playing reverse roles) witness  $\beta_\ell(w) \geq 1$ , as desired. Otherwise,  $v$  does not appear on  $T'_a$ . In this case, we let  $T_h$  denote the tine corresponding to  $T'_h$  in  $F$ : specifically, if  $v$  does not appear in  $T'_h$  then define  $T_h = T'_h$ ; otherwise, define  $T_h$  to be the result of removing  $v$  from  $T'_h$ . In either case, however,  $T_h$  is dominant in  $F$  (as  $\text{len}(\overline{F}) = \text{len}(\overline{F'}) - 1$ ). Thus the tines  $T'_a$  and  $T_h$  (in  $F$ ) witness  $\beta_\ell(w) \geq \beta_\ell(w0^{\Delta-1}\mathbf{h}) + 1$ , as desired.  $\square$

### 3.4 The Cold Region

We now study the setting when  $\beta_\ell$  is sufficiently small. Specifically, consider a string of steps  $\sigma = \sigma_1 \dots \sigma_n \in \Sigma_S^n$ , where each  $\sigma_i \in \Sigma_S$ . We identify the set

$$\text{Cold} = \{ \sigma = \sigma_1 \dots \sigma_n \in \Sigma_S^+ : \beta_\ell(\sigma_1 \dots \sigma_{n-1}) + \#_{\mathbf{h}, \mathbf{a}}(\sigma_n) + \#_{\mathbf{h}}(\sigma_{n-1}) < 0 \}$$

where naturally  $\#_{\mathbf{h}}(\sigma_0) = 0$ . We show that in the region defined by **Cold**,  $\beta_\ell$  satisfies the ideal recurrence (4).

Note that the following lemma does not require any relationship between  $|w|$  and  $\ell$ , the value  $\ell$  can be an arbitrary positive constant. Nonetheless, the lemma is only useful to control margin after slot  $\ell$ , as we know from Lemma 5 that before that slot, margin cannot be negative.

**Lemma 8.** *Let  $\ell \geq 1$ ; let  $w \in \{0, \mathbf{h}, \mathbf{a}\}^*$ ,  $x \in \{0, \mathbf{h}, \mathbf{a}\}^{\geq \Delta-1}$ , and let  $z \in \{0, \mathbf{h}, \mathbf{a}\}^{\leq \Delta}$  be the  $\Delta$ -long suffix of  $w$  (if  $|w| < \Delta$  then  $z = w$ ). If  $\beta_\ell(w) < -\#_{\mathbf{h}, \mathbf{a}}(x) - \#_{\mathbf{h}}(z)$  then*

$$\beta_\ell(w\mathbf{xh}) \leq \beta_\ell(w) + \#_{\mathbf{a}}(x) - 1.$$

*In particular, for any  $\sigma \in \Sigma_S^*$  and any step  $\gamma \in \Sigma_S$ , if  $\sigma\gamma \in \text{Cold}$  then  $\beta_\ell(\sigma\gamma) \leq \beta_\ell(\sigma) + \#_{\mathbf{a}}(\gamma) - 1$ .*

*Proof.* Let  $w' \triangleq w\mathbf{xh}$  and let  $F'$  be a compressed witness  $\Delta$ -fork  $F' \vdash_\Delta w'$ ; let  $(T'_h, T'_a)$  be a pair of witness tines in  $F'$  such that  $\text{len}(T'_h) = \text{len}(\overline{F'})$ . Furthermore, let  $F \triangleq F'_{|w|} \vdash_\Delta w$  and define  $T_h \triangleq (T'_h)_{|w|}$  and  $T_a \triangleq (T'_a)_{|w|}$ , i.e.,  $T_h$  and  $T_a$  are the restrictions of  $T'_h$  and  $T'_a$  to vertices with labels at most  $|w|$ ; we have  $T_h, T_a \in F$  by definition of  $F$ .

Consider any honest tine  $T_H$  of maximum length in  $F$ , we have  $\text{lb}(T_H) \leq |w|$ . Now let  $T'_H$  be the unique honest tine in  $F'$  that satisfies  $\text{lb}(T_H) = |w'|$  as it terminates with the unique honest vertex corresponding to the trailing  $\mathbf{h}$ -symbol of  $w'$  according to axiom (A3) of Definition 2. As  $|\mathbf{xh}| \geq \Delta$ , axiom (A4) gives us  $\text{len}(T_H) < \text{len}(T'_H)$  and hence

$$\text{len}(\overline{F}) < \text{len}(\overline{F'}). \quad (6)$$

By our assumption of negative  $\beta_\ell(w)$ , there is a well-defined main branch  $\mathbf{M}(F)$ . We first establish the following claim.

**Claim 9.** *Consider any tine  $T \in F$  such that  $T \notin \mathbf{M}(F)$  and any  $T' \in F'$  that extends  $T$  in  $F'$  so that  $T = T'_{|w|}$ . Then the set of vertices  $T' \setminus T$  contains no honest vertices.*

To see this, observe that any honest vertex in  $F'$  with label greater than  $|w|$  must have depth at least  $\text{len}(\overline{F}_{\lceil \Delta}) + 1$  by axiom (A4), hence all vertices in  $T' \setminus T$  with depth at most  $\text{len}(\overline{F}_{\lceil \Delta})$  must be adversarial. Furthermore,  $\text{len}(\overline{F}) - \text{len}(\overline{F}_{\lceil \Delta}) \leq \#_{\mathbf{h}}(z) + \#_{\mathbf{h}}(x)$ : this is because  $F'$  is compressed and contains an honest vertex for each depth  $d \in \{\text{len}(\overline{F}_{\lceil \Delta}) + 1, \dots, \text{len}(\overline{F'})\}$ ; but at most  $\#_{\mathbf{h}}(z)$  of these honest vertices can have labels from  $[|w|]$  (by definition of  $\overline{F}_{\lceil \Delta}$ ), similarly at most  $\#_{\mathbf{h}}(x)$  of these honest vertices can have labels greater than  $|w|$  (by Axiom (A4)). This gives us  $\text{len}(T) + \#_{\mathbf{a}}(x) < \text{len}(\overline{F}_{\lceil \Delta})$ , as we have  $\alpha_F(T) \leq \beta_\ell(w) < -\#_{\mathbf{h}, \mathbf{a}}(x) - \#_{\mathbf{h}}(z)$  by our assumption on  $\beta_\ell(w)$ , and hence  $\text{len}(T') < \text{len}(\overline{F}_{\lceil \Delta})$ . This already implies that there are no honest vertices in  $T' \setminus T$  and establishes Claim 9.

We now argue that  $T_h \in \mathbf{M}(F)$ . Towards contradiction, assume that  $T_h \notin \mathbf{M}(F)$ . Then Claim 9 applies to  $T_h$  and  $T'_h \setminus T_h$  contains no honest vertices, hence

$$\text{len}(T'_h) \leq \text{len}(T_h) + \#_a(x). \quad (7)$$

However, by assumption  $\text{len}(T_h) - \text{len}(\overline{F}) = \alpha_F(T_h) \leq \beta_\ell(w) < -\#_a(x)$  and hence  $\text{len}(T_h) < \text{len}(\overline{F}) - \#_a(x)$ , and using equations (7) and (6) gives us  $\text{len}(T'_h) < \text{len}(\overline{F}) < \text{len}(\overline{F}')$ , a contradiction with the definition of  $T'_h$ . Therefore,  $T_h \in \mathbf{M}(F)$ .

Since  $T'_h \not\sim_\ell T'_a$ , it also follows that  $T_h \not\sim_\ell T_a$ , and at most one of these tines belongs to  $\mathbf{M}(F)$ , hence we have  $T_a \notin \mathbf{M}(F)$ . By Claim 9,  $T'_a \setminus T_a$  contains no honest vertices. Hence we have

$$\text{len}(T'_a) \leq \text{len}(T_a) + \#_a(x) \quad (8)$$

and we can combine equations (6) and (8) to get

$$\begin{aligned} \beta_\ell(w) &\geq \alpha_F(T_a) = \text{len}(T_a) - \text{len}(\overline{F}) \geq \text{len}(T'_a) - \#_a(x) - \text{len}(\overline{F}') + 1 = \alpha_{F'}(T'_a) - \#_a(x) + 1 \\ &= \beta_\ell(w') - \#_a(x) + 1, \end{aligned}$$

finishing the proof of Lemma 8. □

### 3.5 The Hot Region

We shift attention to the setting when  $\beta_\ell$  is sufficiently large. Specifically, consider a string of steps  $\sigma = \sigma_1 \dots \sigma_n \in \Sigma_S^n$  where each  $\sigma_i \in \Sigma_S$  and  $n \geq \#_h(\sigma_n) + 3$ . We write  $\sigma = \tilde{\sigma}\tau\sigma_n$ , where  $\tau$  consists of  $\#_h(\sigma_n) + 2$  steps, and identify the set

$$\text{Hot} = \{\sigma = \tilde{\sigma}\tau\sigma_n \mid \beta_\ell(\tilde{\sigma}\tau) \geq \#_a(\tau) + 2\}.$$

We show that in the region defined by  $\text{Hot}$ ,  $\beta_\ell$  satisfies the ideal recurrence (4).

We first need to formally define the minimal honest height  $h_\Delta(\cdot)$ .

**Definition 11.** Let  $x \in \{0, 1\}^*$  and recall that  $x_{\uparrow\Delta}$  denotes the string obtained by removing the last  $\Delta - 1$  symbols from  $x$ , with the understanding that the result is  $\epsilon$  if  $|x| < \Delta$ . We define  $h_\Delta(x)$  inductively so that  $h_\Delta(\epsilon) = 0$ ,  $h_\Delta(x0) = h_\Delta(x)$ , and  $h_\Delta(x1) = h_\Delta(x_{\uparrow\Delta}) + 1$ . We often overload  $h_\Delta$  to apply to strings from  $\{0, \mathbf{h}, \mathbf{a}\}^*$ , in that case only the honest symbols  $\mathbf{h}$  are counted as 1s, while symbols 0 and  $\mathbf{a}$  are treated as 0s.

Now we can state the result describing  $\beta_\ell$  in the  $\text{Hot}$  region.

**Lemma 10.** *Let  $\ell \geq 1$ , let  $x \in \{0, \mathbf{h}, \mathbf{a}\}^{\Delta-1} \parallel \{0, \mathbf{a}\}^*$ , and let  $w \in \{0, \mathbf{h}, \mathbf{a}\}^*$  with  $h_\Delta(w) > \#_h(x) + 3$ . Let  $z$  be the shortest suffix of the string  $w$  with the property that  $h_\Delta(z) \geq \#_h(x) + 3$ . If  $\beta_\ell(w) > \#_a(z) + 2$  then we have*

$$\beta_\ell(w\mathbf{x}\mathbf{h}) \leq \beta_\ell(w) + \#_a(x) - 1.$$

*In particular, for any  $\sigma \in \Sigma_S^*$  and any step  $\gamma \in \Sigma_S$ , if  $\sigma\gamma \in \text{Hot}$  then  $\beta_\ell(\sigma\gamma) \leq \beta_\ell(\sigma) + \#_a(\gamma) - 1$ .*

*Proof.* Let  $F' \vdash w\mathbf{x}\mathbf{h}$  be an optimal compressed fork for  $w' \triangleq w\mathbf{x}\mathbf{h}$  and  $F \vdash w$  the restriction to  $w$ ; if  $T'$  is a tine in  $F'$ , we let  $T$  denote the associated tine for  $F$ . Let  $T'_h$  and  $T'_a$  be a pair of witness tines for  $F'$ . Observe that (6) can be again established in the same way.

We first prove a lower bound on  $\beta_\ell(w')$ . Towards that, consider a witness fork  $G \vdash_\Delta w$  for  $w$ , and let  $(U_h, U_a)$  be witness tines for  $G$  such that  $\text{len}(U_h) = \text{len}(\overline{G})$ . For  $s \in \{\mathbf{h}, \mathbf{a}\}$ , let

$$I_s \triangleq \{i \in \{|w| + 1, \dots, |w\mathbf{x}\mathbf{h}|\} : w'_i = s\}.$$

Construct a labeled rooted tree  $G'$  from  $G$  by (i) adding  $\#_h(x)$  honest vertices labelled by indices from  $I_h$ , all of them as direct descendants of the terminal vertex of  $U_h$ ; (ii) adding a single honest vertex with label  $|w'|$  as a direct descendant of any of the above-added honest vertices; and finally (iii) extending the tine  $U_a$  by a

path consisting of  $\#_a(x)$  adversarial vertices labelled by the increasing sequence of indices from  $I_a$ . Let  $U'_h$  denote the tine terminating in the vertex labelled  $|w'|$  and let  $U'_a$  be this newly-constructed tine extending  $U_a$  in  $G'$ . Observe that  $G'$  is a valid  $\Delta$ -fork for  $w'$ : the axioms (A1)–(A3) are trivially satisfied, and the axiom (A4) also holds as all newly added honest vertices only share depth with honest vertices labelled closer than  $\Delta$  to their own label. Clearly  $\text{len}(U'_h) = \text{len}(\overline{G'})$  and moreover,  $U_h \not\sim_\ell U_a$  implies  $U'_h \not\sim_\ell U'_a$ ; hence we have

$$\beta_\ell(w') \geq \text{len}(U'_a) - \text{len}(U'_h) = (\text{len}(U_a) + \#_a(x)) - (\text{len}(U_h) + 2) = \beta_\ell(w) + \#_a(x) - 2 > \#_a(zx), \quad (9)$$

where the last inequality follows by our assumption on  $\beta_\ell(w)$ .

We now establish that also in this setting there are no honest vertices on  $T'_a$  with a label greater than  $|w|$ , in other words, there are no honest vertices in  $T'_a \setminus T_a$ . Towards a contradiction, assume that there is an honest vertex in  $T'_a \setminus T_a$  and let  $v'_a$  be the honest vertex on  $T'_a$  with maximum label (and hence also maximum depth). Since  $\text{lb}(v'_a) > |w|$ , all vertices  $u$  on  $T'_a$  with  $\text{len}(u) > \text{len}(v'_a)$  also have  $\text{lb}(u) > \text{lb}(v'_a) > |w|$ , and by maximality of  $v'_a$  all these vertices are adversarial, hence there are at most  $\#_a(x)$  such vertices by axiom (A3). However, we also have  $\text{len}(v'_a) \leq \text{len}(\overline{F'})$  as  $v'_a$  is honest. Put together, we have  $\beta_\ell(w') = \text{len}(T'_a) - \text{len}(\overline{F'}) \leq \text{len}(T'_a) - \text{len}(v'_a) \leq \#_a(x)$ . This contradicts (9), concluding the proof that there are no honest vertices on  $T'_a \setminus T_a$ . Hence we have  $\text{len}(T'_a) - \text{len}(T_a) \leq \#_a(x)$ .

Let  $v_a$  be the last honest vertex on  $T_a$  (we now know that it is also the last honest vertex on  $T'_a$ ). Likewise, let  $v_h$  be the last honest vertex on  $T_h$ . We consider two cases depending on  $\text{lb}(v_a)$  and  $\text{lb}(v_h)$ .

**The case  $\text{lb}(v_h) < \text{lb}(v_a)$ .** For a tine  $T$  and a portion  $y$  of the characteristic string, we let  $\#_h(y; T)$  denote the number of honest vertices in  $T$  labeled with symbols from  $y$ ; we similarly overload also the notation  $\#_a$  and  $\#_{h,a}$ .

We first establish that

$$\text{len}(v_a) < \text{len}(T_h) + \#_a(xh; T'_h). \quad (10)$$

Note, first of all, that  $\text{len}(T_h) \geq \text{len}(\overline{F'}) - \#_{h,a}(xh; T'_h)$ . Now consider  $\text{len}(v_a)$ . Observe that  $v_a$  cannot be labeled from the string  $z$ : if it were, then  $\beta_\ell(F') \leq \#_a(zx)$  which contradicts (9). Hence  $v_a$  is labeled prior to  $z$  and it follows that  $\text{len}(F') \geq \text{len}(v_a) + [h_\Delta(z) - 1] \geq \text{len}(v_a) + \#_h(x; T'_h) + 2$  by definition of  $z$ . Hence

$$\text{len}(T_h) \geq \text{len}(\overline{F'}) - \#_{h,a}(xh; T'_h) \geq (\text{len}(v_a) + \#_h(x; T'_h) + 2) - \#_{h,a}(xh; T'_h) > \text{len}(v_a) - \#_a(xh; T'_h),$$

proving (10).

Now we invoke Lemma 4 with tines  $T'_h, T'_a$  and vertices  $v_h, v_a$  in  $F'$ . By assumption  $\text{lb}(v_h) < \text{lb}(v_a)$  and hence we obtain  $\beta_\ell(w') \geq \alpha_{F'}(v_h) + |A_h \cup A_a|$ , where  $A_h$  (resp.  $A_a$ ) is the set of adversarial vertices in  $T'_h$  (resp. in  $T'_a$  after  $v_a$ ). Note that  $\text{lb}(v_h) < \text{lb}(v_a)$  means  $v_h \neq v_a$  and together with the definition of  $v_h, v_a$  this means that  $A_h \cap A_a = \emptyset$  and  $|A_h \cup A_a| = |A_h| + |A_a|$ .

Recall that  $T_h$  (resp.  $T_a$ ) contains only adversarial vertices after  $v_h$  (resp.,  $v_a$ ) by definition of  $v_h$  (resp.  $v_a$ ). Moreover,  $T'_a \setminus T_a$  also only contains adversarial vertices. Hence we get

$$\begin{aligned} \beta_\ell(w') &\geq \alpha_{F'}(v_h) + (\text{len}(T_h) - \text{len}(v_h)) + (\text{len}(T'_a) - \text{len}(v_a)) + \#_a(xh; T'_h) \\ &\geq \alpha_{F'}(T_h) + (\text{len}(T'_a) - \text{len}(v_a)) + \#_a(xh; T'_h) \\ &> \alpha_{F'}(v_a) + (\text{len}(T'_a) - \text{len}(v_a)) \geq \alpha_{F'}(T'_a), \end{aligned}$$

where the third inequality follows from (10). This contradicts the optimality of  $F'$  and  $T'_a$ , and shows that this case cannot occur.

**The case  $\text{lb}(v_h) \geq \text{lb}(v_a)$ .** Let  $T_H$  denote a maximal length honest tine in  $F$ . If  $T_H \not\sim_\ell T_a$ , these two tines witness

$$\beta_\ell(w) \geq \alpha_F(T_a) \geq (\text{len}(T'_a) - \#_a(x)) - \text{len}(\overline{F}) \geq (\text{len}(T'_a) - \#_a(x)) - (\text{len}(\overline{F'}) - 1) = \beta_\ell(w_xh) - \#_a(x) + 1$$

as desired. Otherwise, we assume that  $T_H \sim_\ell T_a$  and hence  $T_H \not\sim_\ell T_h$ .

In this case, we begin by compressing the fork  $F$ : Let  $c(F) \vdash_\Delta w$  denote the compression of  $F$ . If  $v$  is a vertex of  $F$  it appears in  $c(F)$ ; in order for context to be clear we let  $c(v)$  denote the vertex  $v$  as it appears in



$c(F)$ . If  $T$  is a tine of  $F$ , we let  $c(T)$  denote the associated tine in the compression (that is, the tine that terminates with the vertex that terminates  $T$ ). As  $\alpha_F(T_a) \geq 0$ , recall that  $\alpha_{c(F)}(c(T_a)) = \alpha_F(T_a)$ . Note that the last honest vertex on  $c(T_a)$  may not be  $c(v_a)$  (due to a compression step); let  $w_a$  be the vertex for which  $c(w_a)$  is the last honest vertex on  $c(T_a)$ . As  $\text{lb}(w_a) \leq \text{lb}(v_a)$ , we still have the inequality  $\text{lb}(w_a) \leq \text{lb}(v_h)$  (and hence, of course,  $\text{lb}(c(w_a)) < \text{lb}(c(v_h))$ ).

We again invoke Lemma 4, this time for tines  $c(T_a)$ ,  $c(T_h)$  and vertices  $c(w_a)$ ,  $c(v_h)$  in  $c(F)$ . Since  $\text{lb}(c(w_a)) < \text{lb}(c(v_h))$ , we obtain the following, where  $|A|$  is the set of adversarial vertices on  $c(T_a)$  after  $c(w_a)$  in  $c(F)$ :

$$\begin{aligned} \beta_\ell(w) &\geq \alpha_{c(F)}(c(w_a)) + |A| = \alpha_{c(F)}(c(T_a)) = \alpha_F(T_a) = \text{len}_F(T_a) - \text{len}(\overline{F}) \\ &\geq (\text{len}_{F'}(T'_a) - \#_a(x)) - (\text{len}(\overline{F}') - 1) = \alpha_{F'}(T'_a) - \#_a(x) + 1 = \beta_\ell(w') - \#_a(x) + 1. \end{aligned}$$

This concludes the proof for the second case. □

## 4 Analysis of the Stochastic Process

Recall from the introduction the critical security threshold.

**Definition 12.** For  $p_h > 0$  and  $\Delta \in \mathbb{N}$ , we define the discrete critical threshold

$$\vartheta(p_h, \Delta) := \frac{1}{(\Delta - 1) + 1/p_h}.$$

For  $r_h > 0$  and  $\Delta_0 > 0$ , we likewise define the Poisson critical threshold

$$\overline{\vartheta}(r_h, \Delta_0) := \frac{1}{\Delta_0 + 1/r_h}.$$

While  $\vartheta$  is a function of  $p_h$  and  $\Delta$ , we simply write  $\vartheta$  when these parameters can be inferred from context;  $\overline{\vartheta}$  is treated similarly.

To relate the security threshold  $\overline{\vartheta}$  in the Poisson setting to discrete threshold  $\vartheta$ , recall that the discrete approximation is given by taking  $p_h = sr_h$ ,  $p_a = sr_a$  and  $\Delta = \lceil \Delta_0/s \rceil$  for a (small ‘‘slot length’’) parameter  $s$ . If  $r_a, r_h$ , and  $\Delta_0$  satisfy  $r_a < \overline{\vartheta}$ , which is to say  $1/r_a > \Delta_0 + 1/r_h$  then, by scaling this inequality by  $1/s$ , we find that

$$\frac{1}{p_a} = \frac{1}{sr_a} > \Delta_0/s + \frac{1}{sr_h} > (\lceil \Delta_0/s \rceil - 1) + \frac{1}{p_h} = \Delta - 1 + \frac{1}{p_h}.$$

This proves the following.

**Fact 11.** For all  $s > 0$ ,  $s\overline{\vartheta}(r_h, \Delta_0) \leq \vartheta(sr_h, \lceil \Delta_0/s \rceil)$ . Hence, if  $r_a < \overline{\vartheta}(r_h, \Delta_0)$  then  $s \cdot r_a < \vartheta(s \cdot p_h, \lceil \Delta_0/s \rceil)$ .

We remark that  $\vartheta$  satisfies the equality

$$\frac{1}{\vartheta} = (\Delta - 1) + \frac{1}{p_h},$$

which gives an immediate and intuitive interpretation: note that if  $w_i = x$  for a symbol  $x \in \Sigma$  that occurs with probability  $p$  then  $1/p$  is the expected waiting time before the next occurrence of  $x$ . Thus the threshold corresponds to the setting where the average waiting time for a symbols is larger, by an additive factor of  $\Delta - 1$ , than the average waiting time for  $h$  symbols.

**Definition 13.** When  $w$  has the distribution  $\mathcal{B}(p_a, p_h)$  of Definition 1, the steps  $\sigma_i$  arising from  $w$  are independent and identically distributed random variables, taking values in  $\{\mathbf{a}, \mathbf{h}, 0\}^*$ . We denote this distribution  $\mathcal{S}(p_a, p_h; \Delta)$ .

Observe that when  $\gamma$  is drawn from  $\mathcal{S}(p_a, p_h; \Delta)$ , its length follows a translated geometric distribution: for each  $k \geq 0$ ,

$$\Pr[|\gamma| = \Delta + k] = p_h(1 - p_h)^k. \quad (11)$$

This immediately yields the following tail bound.

**Fact 12.** *Let  $\gamma$  be drawn according to  $\mathcal{S}(p_a, p_h; \Delta)$ . Then  $\Pr[\#_a(\gamma) \geq k] \leq \Pr[|\gamma| \geq k] = \exp(-\Omega(k))$ .*

The estimates of Lemma 5, Lemma 8, and Lemma 10 show that, with particular exceptions,  $\beta$  adheres to the ideal recurrence

$$\beta(\sigma\gamma) \mapsto \beta(\sigma) + \#_a(\gamma) - 1.$$

The claim below confirms that when  $p_a < \vartheta$ , this transition has negative bias (for  $\gamma$  drawn from  $\mathcal{S}(p_a, p_h; \Delta)$ ), its proof appears in Appendix A.

**Claim 13.** *Let  $p_a < \vartheta(p_h, \Delta)$  and let  $\gamma$  have the distribution  $\mathcal{S}(p_a, p_h; \Delta)$ . Then  $\mathbb{E}[\#_a(\sigma) - 1] < p_a/\vartheta - 1 < 0$ . We remark, additionally, that if  $p_a = sr_a$ ,  $p_h = sr_h$ , and  $\Delta = \lceil \Delta_0/s \rceil$ , then  $p_a/\vartheta < r_a/\bar{\vartheta}$ .*

Thus, the random walk described by  $\beta_\ell(\sigma_1 \dots)$  initially observes a negative bias with a barrier at zero (arising from the rules of Lemma 5); once the length of the string exceeds  $\ell$ , the walk is more complicated: it is negatively biased when sufficiently far from zero and positively biased near zero.

In the following (Sections 4.1–4.3) we work towards establishing the estimate for the behavior of the full walk formulated as Theorem 14 below. Finally, in Section 4.4 we apply it to control Bitcoin consistency failures.

**Theorem 14.** *Let  $p_a$ ,  $p_h$  and  $\Delta$  satisfy  $p_a < \vartheta(p_h, \Delta)$ . Fix  $m \geq 0$ . Let  $\sigma = \sigma_1, \dots$  denote a sequence of steps, each identically distributed according to  $\mathcal{S}(p_a, p_h; \Delta)$ . Let  $\ell$  denote the random variable  $|\sigma_1 \dots \sigma_m|$ , i.e., the length of  $\sigma_1 \dots \sigma_m$  in symbols. Then for any  $T \geq 0$  and  $M > 0$ ,*

$$\Pr[\exists t \geq m + T, \beta_\ell(\sigma_1 \dots \sigma_t) \geq -M] = \exp(-\Omega(T) + O(M)),$$

where the constants hidden by the asymptotic notation are universal aside from dependence on  $|p_a/\vartheta - 1|$ .

## 4.1 Minimal Honest Height

Recall the notion of  $h_\Delta(\cdot)$  from Definition 11. We record the McDiarmid inequality, which immediately implies a large deviation bound on  $h_\Delta$ .

**Theorem 15** (McDiarmid's inequality). *Let  $X = X_1, \dots, X_n$  be a sequence of independent random variables taking values in  $\{0, 1\}$ . Let  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  have the property that for any  $x = (x_1, \dots, x_n) \in \{0, 1\}^n$  and  $y = (y_1, \dots, y_n) \in \{0, 1\}^n$  that differ only in a single coordinate  $|f(x) - f(y)| < C$ . Then*

$$\Pr_X[|f(X) - \mathbb{E}[f(Y)]| \geq \lambda\sqrt{n}] \leq 2 \exp(-2\lambda^2/C),$$

where  $Y$  has the same distribution as  $X$ .

**Lemma 16.** *Let  $p \in (0, 1)$  and  $\Delta \in \{1, 2, \dots\}$ . Let  $X_1, \dots, X_n$  be independent Bernoulli random variables, each taking the value 1 with probability  $p$ . Let  $\alpha = \Delta + (1 - p)/p$  and  $X = (X_1, \dots, X_n)$ ; then*

$$|n/\alpha - \mathbb{E}[h_\Delta(X)]| \leq 1. \quad (12)$$

Furthermore,

$$\Pr[|h_\Delta(X) - n/\alpha| \geq \lambda\sqrt{n} + 1] \leq 2 \exp(-2\lambda^2).$$

The straightforward induction proof appears in Appendix A.

## 4.2 The Stationary Distribution Prior to $\ell$

We consider the distribution of  $\beta_\ell(w)$  where  $|w| \leq \ell$ . Focusing on the step decomposition  $w = \sigma_1 \dots$  and the recurrence relation established in Lemma 5, we observe that for any  $\sigma = \sigma_1 \dots \sigma_k \in \Sigma_{\mathcal{S}}^k$  for which  $|\sigma| \leq \ell$ ,  $\beta_\ell(\sigma) \leq B(\sigma)$ , where  $B(\cdot)$  is given by the upper bounds of Lemma 5. That is,  $B(\epsilon) = 0$ , and

$$B(\sigma\gamma) = \begin{cases} B(\sigma) + \#_a(\gamma) - 1 & \text{if } B(\sigma) > 0 \\ \#_a(\gamma) & \text{if } B(\sigma) = 0, \end{cases}$$

where  $\sigma \in \Sigma_{\mathcal{S}}^*$  and  $\gamma$  is a single step,  $\gamma \in \Sigma_{\mathcal{S}}$ . In light of Claim 13, when the symbols are drawn from  $\mathcal{S}(p_a, p_h; \Delta)$  the quantity  $B(\cdot)$  follows a negatively biased random walk on  $\mathbb{N}$  with a barrier at zero; in this setting where the upper tails of the walk are sub-geometric (that is, there is an upper bound on the one-step tails of the form  $Ca^{-k}$  for some  $a > 1$ ), it follows immediately that the random variables  $B(\sigma)$  converge to a stationary distribution.

To articulate the result formally, we recall the notion of stochastic dominance. For two random variables  $X$  and  $Y$  taking values in  $\mathbb{R}$ , we say that  $Y$  *stochastically dominates*  $X$ , written  $X \prec Y$ , if for all  $\lambda \in \mathbb{R}$ ,  $\Pr[X \geq \lambda] \leq \Pr[Y \geq \lambda]$ . Note that if  $X \prec Y$  we can transfer tail bounds on  $Y$  to tail bounds on  $X$ : if  $\Pr[Y \geq \lambda] \leq f(\lambda)$  then  $\Pr[X \geq \lambda] \leq f(\lambda)$ . The discussion above implies that for any random variable  $\sigma$ ,  $\beta_\ell(\sigma) \prec B(\sigma)$  (so long as  $|\sigma| \leq \ell$ ).

**Lemma 17.** *Let  $X_1, X_2, \dots$  be a sequence of i.i.d. random variables taking values in  $\mathbb{N}$  for which (i.)  $\mathbb{E}[X_i] < 1$  and (ii.) there are constants  $a > 1$  and  $A > 0$  so that  $\Pr[X_i = k] \leq A \cdot a^{-k}$ . Let  $W_t$  denote the random walk on  $\mathbb{N}$  given by the rule  $W_0 = 0$ ,*

$$W_t = \begin{cases} W_{t-1} + X_t - 1 & \text{if } W_{t-1} > 0, \text{ and} \\ W_{t-1} + X_t & \text{if } W_{t-1} = 0. \end{cases}$$

*Then there is a random variable  $S$ , taking values in  $\mathbb{N}$ , for which  $W_t \prec S$  for all  $t$  and, moreover, there are constants  $a_* > 1$  and  $A_* > 0$  so that  $\Pr[S = k] \leq A_* \cdot a_*^{-k}$ .*

The proof appears in Appendix A. Applying Lemma 17 to the random variables  $X_i = \#_a(\sigma_i)$  (with the  $\sigma_i$  drawn as above), yields the following bound on  $\beta_\ell(\cdot)$ .

**Corollary 18.** *Let  $\sigma = \sigma_1 \dots, \sigma_m \in \Sigma_{\mathcal{S}}^m$  be independently generated according to  $\mathcal{S}(p_a, p_h; \Delta)$ . Let  $\ell = \ell(\sigma)$  denote the random variable  $|\sigma|$ . Then  $\Pr[\beta_\ell(\sigma) \geq k] \leq \exp(-\Omega(k))$ .*

## 4.3 The Descent to $-\infty$ After $\ell$

This section proves Theorem 14. In general, the proof proceeds by considering several coupled stochastic processes:

$$\begin{array}{lll} \mathcal{B}(p_a, p_h) & : & \underbrace{w_1 \dots w_{t_1}}_{\sigma_1} \quad \underbrace{w_{t_1+1} \dots w_{t_2}}_{\sigma_2} \quad \dots \\ \mathcal{S} = \mathcal{S}(w) & : & \sigma_1 \quad \sigma_2 \quad \dots \\ \mathcal{M} & : & m_1 = \beta_\ell(\sigma_1) \quad m_2 = \beta_\ell(\sigma_1\sigma_2) \quad \dots \\ \mathcal{P} & : & \pi_1 = \pi(\sigma_1) \quad \pi_2 = \pi(\sigma_1\sigma_2) \quad \dots \\ \mathcal{I} & : & i_1 = \iota(\sigma_1) \quad i_2 = \iota(\sigma_1\sigma_2) \quad \dots \\ \mathcal{I}^D & : & i_1^D = \iota^D(\sigma_1) \quad i_2^D = \iota^D(\sigma_1\sigma_2) \quad \dots \end{array}$$

The random variables  $w_i$  of  $\mathcal{B} = \mathcal{B}(p_a, p_h)$  are described in Definition 1. The process  $\mathcal{S}$  is given by the rule  $\sigma(w)$  described in Definition 9 above. The ‘‘margin process’’  $\mathcal{M}$  is determined by application of  $\beta(\cdot)$  to  $\mathcal{S}$ . This is the principal process of interest, and the subject of Theorem 14.

The final processes, which are introduced solely for the purposes of analysis, are  $\mathcal{I}$ , the *ideal process*, which carries out the ideal recurrence, and  $\mathcal{P}$ , the *pessimistic process*, which only relies on the generally applicable results from Lemma 6 and Lemma 7. We study, additionally, a ‘‘deformation’’ of the ideal process denoted  $\mathcal{I}^D$ . These are defined by the following recurrences.

**Definition 14** (The ideal process; the pessimistic process). Fix  $\Delta > 0$ . Define  $\iota, \pi : \Sigma_{\mathcal{S}}^* \rightarrow \mathbb{Z}$  by the rule  $\iota(\epsilon) = \pi(\epsilon) = 0$ , for the empty string  $\epsilon$ , and, for  $\sigma \in \Sigma_{\mathcal{S}}^*$  and a single step  $\gamma \in \Sigma_{\mathcal{S}}$ ,

$$\iota(\sigma\gamma) = \iota(\sigma) + \#_{\mathbf{a}}(\gamma) - 1,$$

and

$$\pi(\sigma\gamma) = \begin{cases} \pi(\sigma) - 1 & \text{if } \gamma = 0^t \mathbf{h} \text{ with } t \geq \Delta - 1, \\ \pi(\sigma) + \#_{\mathbf{a}, \mathbf{h}}(\gamma) & \text{otherwise.} \end{cases}$$

In general, we apply these functions to suffixes of a sequence  $\sigma \in \Sigma_{\mathcal{S}}^*$ . To reflect this usage, we generalize these functions to prescribed *basepoints* for convenience. Specifically, we define

$$\iota(\sigma; \tau) = \beta_{\ell}(\sigma) + \iota(\tau) \quad \text{and} \quad \pi(\sigma; \tau) = \beta_{\ell}(\sigma) + \pi(\tau),$$

for  $\sigma, \tau \in \Sigma^*$ . (Note that  $\ell$  and  $\Delta$  are implicit in this notation, but they can be inferred from context.)

Note that  $\iota$  and  $\pi$  are motivated by the margin bounds of Section 3. In particular, in light of Lemma 6 and Lemma 7,

$$\beta_{\ell}(\sigma\tau) \leq \pi(\sigma; \tau)$$

for any  $|\sigma| \geq \ell$ . (In fact, for  $\pi$  this is true even without the length restriction on  $\sigma$ .) We note here an important relationship with  $\beta_{\ell}^{\Delta}(\cdot)$ , which will be relevant for the proof of Theorem 14.

**Lemma 19.** *Let  $w \in \Sigma^*$  and  $x \in \Sigma^*$  satisfy  $|x| \geq \Delta$ . Let  $y$  denote the string obtained from  $x$  by replacing all  $\mathbf{h}$  with  $\mathbf{a}$ . Then*

$$\beta_{\ell}^{\Delta}(wx) \leq \beta_{\ell}(wy) \leq \beta_{\ell}(w) + \#_{\mathbf{h}, \mathbf{a}}(x).$$

*Similarly, let  $w \in \Sigma^*$  and write  $w = \sigma_1 \dots \sigma_m w'$ , where each  $\sigma_i \in \Sigma_{\mathcal{S}}$  and  $w' \in \Sigma^*$  so that the expression  $\sigma_1 \dots \sigma_m w'$  denotes a partial rendering of  $w$  into steps followed by an arbitrary suffix. Then*

$$\beta_{\ell}^{\Delta}(w) \leq \beta_{\ell}(\sigma_1 \dots \sigma_{m-1}) + \#_{\mathbf{h}, \mathbf{a}}(\sigma_m w').$$

*Proof.* Let  $F \vdash_{\Delta} wx$  where, as in the statement of the lemma,  $|x| \geq \Delta$ . Then,  $F$  may be viewed as a fork of  $wy$ , as any vertex of  $F$  associated with an honest symbol of  $x$  can be associated with an adversarial symbol. We let  $F' \vdash_{\Delta} wy$  denote this fork  $F$  viewed as a fork of  $wy$ ; note that while  $F$  and  $F'$  have identical structure, there may be honest vertices of  $F$  that are adversarial in  $F'$ . Note, also, that  $\text{len}(\overline{F}) = \text{len}(\overline{F'}_{\Gamma_{\Delta}})$ . Thus a pair of  $\Delta$ -witness tines  $F$  are a pair of witness tines in  $F'$ ; this proves the statement.  $\square$

Likewise, roughly speaking, if  $\sigma\tau'$  satisfies the Hot or Cold conditions for each prefix  $\tau'$  of  $\tau$ , then

$$\beta_{\ell}(\sigma\tau) \leq \iota(\sigma; \tau).$$

Finally, to manage some technical issues in the proof, we introduce a related process with a less ready interpretation.

**Definition 15** (The deformed ideal process;  $D$ -typicality). For each  $D \geq 4$ , define the function  $\iota^D : \Sigma_{\mathcal{S}}^* \rightarrow \mathbb{Z}$  by the following rules. In general, for a string  $\sigma \in \Sigma_{\mathcal{S}}^*$ , we write  $\sigma = \sigma_{\text{base}}\sigma_{\text{tail}}$ , with the convention that  $\sigma_{\text{tail}}$  consists of the last  $D$  steps of  $\sigma$ ; if  $\sigma$  consists of fewer than  $D$  steps, we define  $\sigma_{\text{tail}} = \sigma$  and  $\sigma_{\text{base}} = \epsilon$ . Then for  $\sigma \in \Sigma_{\mathcal{S}}^*$  and a single step  $\gamma \in \Sigma_{\mathcal{S}}$ , we say that  $\sigma\gamma$  is  $D$ -typical if

$$\#_{\mathbf{h}, \mathbf{a}}(\gamma) \leq D - 2, \quad \text{and} \quad \#_{\mathbf{h}, \mathbf{a}}(v) \leq D \quad \text{for each step } v \text{ in } \sigma_{\text{tail}}.$$

Note that typicality is determined only by the last  $D$  steps of  $\sigma$  and the step  $\gamma$ . Then we define  $\iota^D(\epsilon) = 0$ , and in general

$$\iota^D(\sigma\gamma) = \begin{cases} \iota^D(\sigma) + \#_{\mathbf{a}}(\gamma) - 1 & \text{if } \sigma\gamma \text{ is } D\text{-typical,} \\ \iota^D(\sigma) + \#_{\mathbf{h}, \mathbf{a}}(\gamma) & \text{otherwise.} \end{cases}$$

We similarly define a notion with a basepoint:

$$\iota^D(\sigma; \tau) = \beta_{\ell}(\sigma) + [\iota^D(\sigma\tau) - \iota^D(\sigma)],$$

for  $\sigma, \tau \in \Sigma^*$ .

Observe that  $\iota^\infty(\cdot) = \iota(\cdot)$ , which explains the name.  $D$ -typicality is a convenient “local” criterion for membership in **Hot** or **Cold** (as it depends only on the most recent  $D + 1$  steps); in particular, the details of the definition are meant to appropriately correspond to the definitions of **Hot** and **Cold** as described by the claim below.

**Claim 20.** *Let  $\sigma\gamma \in \Sigma_{\mathcal{S}}^* \times \Sigma_{\mathcal{S}}$  be  $D$ -typical. Then*

$$\begin{aligned}\beta_\ell(\sigma) > D^2 &\implies \sigma\gamma \in \mathbf{Hot}, \\ \beta_\ell(\sigma) < -D^2 &\implies \sigma\gamma \in \mathbf{Cold}.\end{aligned}$$

*As a result, if  $\sigma\gamma$  is  $D$ -typical and  $|\beta_\ell(\sigma)| > D^2$  then  $\iota^D(\sigma; \gamma) = \iota(\sigma; \gamma) = \beta_\ell(\sigma\gamma)$ . More generally, if  $|\beta_\ell(\sigma)| > D^2$  then  $\beta_\ell(\sigma\gamma) \leq \iota^D(\sigma; \gamma)$ .*

*Proof.* This follows immediately from Lemma 8, Lemma 10, and Lemma 6.  $\square$

We first develop a standard tail bound for the random variables that arise naturally in the ideal process  $\iota$  (that is, the  $\#_{\mathfrak{a}}(\sigma_i) - 1$ ). Recall that for a real-valued random variable  $X$ , the *moment-generating function*  $m_X$  is defined by the rule  $z \mapsto \mathbb{E}[e^{zX}]$ . The proofs of the following five lemmas appear in Appendix A.

**Lemma 21.** *Let  $a > 1$  and  $C > 0$ . Let  $A$  be a random variable on  $\{-1, 0, 1, \dots\}$  satisfying  $\mathbb{E}[A] < 0$  and  $\Pr[A = k] \leq Ca^{-k}$ . Then*

$$m_A(\lambda) \leq 1 + \mathbb{E}[A] \lambda / 2 \tag{13}$$

*for sufficiently small  $\lambda$ .*

**Lemma 22.** *Let  $a > 1$ ,  $C > 0$ , and  $\gamma > 0$ . Consider a sequence of i.i.d. integer-valued random variables  $Z_1, Z_2, \dots$  satisfying  $\mathbb{E}[Z_i] = -\gamma < 0$  and  $\Pr[Z_i = k] \leq Ca^{-k}$ . Let  $S_n = \sum_{i=1}^n Z_i$ . Then there is a constant  $\alpha > 0$  so that*

$$\forall \Lambda \geq -\gamma n / 2, \quad \Pr[S_n \geq \Lambda] \leq e^{-\alpha(\Lambda + \gamma n / 2)}. \tag{14}$$

*It follows that for any  $N > 0$ ,*

$$\Pr[\exists n \geq N, S_n \geq -\gamma n / 4] = e^{-\Theta(N)}. \tag{15}$$

**Lemma 23** (Gambler’s ruin). *Let  $a > 1$  and  $C > 0$ . Let  $Z_1, \dots$  be a sequence of i.i.d. random variables taking values in  $\{-1, 0, 1, \dots\}$  satisfying  $\mathbb{E}[Z_i] < 0$  and  $\Pr[Z_i = k] \leq Ca^{-k}$ . Let  $S_n = \sum_{i=1}^n Z_i$ . Then*

1. *for any  $D > 0$ ,  $\Pr[\exists t > 0, S_t \geq D] = \exp(-\Theta(D))$ , and*
2.  $\Pr[\forall t > 0, S_t < 0] > 0$ .

**Lemma 24.** *Let  $H_1, H_2, \dots$  be a sequence of i.i.d. random variables taking values in  $\mathbb{N}$  for which  $\Pr[H_1 = k] = \exp(-\Omega(k))$ . Likewise, let  $G$  be a random variable taking values in  $\mathbb{N}$ , independent from the  $H_i$ , for which  $\Pr[G = k] \leq \exp(-\Omega(k))$ . Then*

1.  $\Pr[G + H_1 \geq k] = \exp(-\Omega(k))$ , and
2.  $\Pr[H_1 + \dots + H_G \geq k] = \exp(-\Omega(k))$ .

*The constants hidden by these instances of asymptotic notation may be different.*

**Lemma 25.** *Let  $X_1, X_2, \dots$  be a sequence of independent geometrically distributed random variables, so that each  $X_i$  has the distribution  $\Pr[X_i = k] = p(1 - p)^k$  for a parameter  $p \in (0, 1]$ . Then  $\mathbb{E}[X_i] = (1 - p)/p$  and, for any  $\lambda \geq 1$ ,*

$$\Pr\left[\sum_{i=1}^n X_i > \lambda n / p\right] \leq e^{-n(1-\lambda)}.$$

*Let  $Y_1, Y_2, \dots$  be a sequence of independent exponentially distributed random variables, so that each  $Y_i$  has the probability density function  $pe^{-px}$  ( $x \geq 0$ ). Then  $\mathbb{E}[Y_i] = 1/p$  and, for any  $\lambda \geq 2$ ,*

$$\Pr\left[\sum_{i=1}^n Y_i > \lambda n / p\right] \leq e^{-n \frac{1}{4(p+1)}(\lambda-1)}.$$

Finally we return to the proof of Theorem 14.

*Proof of Theorem 14.* Denote by  $\text{bias} = \mathbf{E}[\iota(\gamma)] = \mathbf{E}[\#\mathbf{a}(\gamma) - 1] < 0$  the (negative) bias of the “ideal walk” (with  $\gamma$  drawn from  $\mathcal{S}(p_a, p_h; \Delta)$ ). We organize the proof around three “zones,” corresponding roughly to the Hot, Cold and critical cases studied in Section 3.

Specifically, we select a “typicality limit”  $D \geq 4$  and define the following subsets of the integers:

$$R^- = \{z \in \mathbb{Z} \mid z < -D^2\}, \quad R^0 = \{z \in \mathbb{Z} \mid -D^2 \leq z \leq D^2\}, \quad \text{and} \quad R^+ = \{z \in \mathbb{Z} \mid D^2 < z\}.$$

To determine the limit  $D$  defining these regions, consider independent selection of  $C \geq 4$  steps  $\gamma_1 \dots \gamma_C$  and a final step  $\gamma'$  (each independently according to  $\mathcal{S}(p_a, p_h; \Delta)$ ). Then examine the random variable

$$P_C = \iota^C(\gamma_1 \dots \gamma_C \gamma') - \iota^C(\gamma_1 \dots \gamma_C) = \begin{cases} \#\mathbf{a}(\gamma') - 1 & \text{if } \gamma_1 \dots \gamma_C \gamma' \text{ is } C\text{-typical,} \\ \#\mathbf{a},\mathbf{h}(\gamma') & \text{otherwise.} \end{cases}$$

As  $C \rightarrow \infty$ , note that  $\Pr[\#\mathbf{h},\mathbf{a}(\gamma_i) > C] = \exp(-\Omega(C))$  and hence  $\Pr[\max(\#\mathbf{h},\mathbf{a}(\gamma_1), \dots, \#\mathbf{h},\mathbf{a}(\gamma_C), \#\mathbf{h},\mathbf{a}(\gamma')) \geq C] = C \exp(-\Omega(C))$ . It follows that  $\lim_{C \rightarrow \infty} \Pr[\gamma_1 \dots \gamma_C \gamma' \text{ is } C\text{-typical}] = 1$  and hence that  $\lim_{C \rightarrow \infty} \mathbf{E}[P_C] \rightarrow \text{bias}$ . Define  $D$  to be the smallest value of  $C$  for which  $\mathbf{E}[P_C] < \text{bias}/2$ . We explain the relevance of this rule for selecting  $D$  below.

Throughout the proof, we often write the sequence  $\sigma_1 \sigma_2 \dots$  as  $\sigma_1 \dots \sigma_s \tau_1 \dots$  (so that  $\tau_i = \sigma_{s+i}$ ) with the implicit understanding that  $s \geq m$ . With this convention, the initial steps always determine the “pre- $\ell$ ” dynamics  $\beta_\ell(\sigma_1 \dots \sigma_m)$ , whose statistics are controlled by Corollary 18.

We associate with every prefix of steps  $\sigma_1 \dots \sigma_m \tau_1 \dots \tau_t$  a *state* in the set  $\{R^-, R^0, R^+\}$ , depending on which of these contains the integer  $\beta_\ell(\sigma\tau)$ . We now consider the transitions between these states.

**Dynamics in the region  $R^+$ .** Consider entry to  $R^+$  at “time”  $s$  (i.e., after  $s$  steps), and let  $\sigma = \sigma_1 \dots \sigma_s$ . We examine the random walk  $\beta_\ell(\sigma), \beta_\ell(\sigma\tau_1), \dots$  by comparing it to  $\iota^D(\sigma; \epsilon), \iota^D(\sigma; \tau_1), \dots$ . Returning to the definition of  $\iota^D(\sigma; \tau)$ , observe that while  $\sigma\tau_1 \dots \tau_t \in R^+$ ,  $\beta_\ell(\sigma\tau_1 \dots \tau_t) > D^2$  and, by the results of Section 3.5 and Claim 20,

$$\beta_\ell(\sigma\tau_1 \dots \tau_t) \leq \iota^D(\sigma; \tau_1 \dots \tau_t).$$

We wish to show that this  $\iota^D(\sigma; \tau_1 \dots)$  walk will descend to  $D^2$  (so that  $\beta_\ell(\sigma\tau)$  returns to  $R^0$ ) with certainty and, moreover, that the descent will occur quickly: that is, the probability that the descent will take  $k$  steps is  $\exp(-\Omega(k))$ .

A typical entry into  $R^+$  arrives with an initial value  $Z_{\text{init}}$  for which  $\Pr[Z_{\text{init}} = k] = \exp(-\Omega(k))$ . This can either occur at time  $m$ , when the distribution of  $\beta_\ell(\sigma_1 \dots \sigma_m)$  is given by Corollary 18, or as a result of a transition from  $R_0$  (or  $R^-$ ) in which case the value is bounded above by  $D + \#\mathbf{a}(\gamma)$  for the last step prior to the transition to  $R^+$ , in which case the height is bounded by Fact 12.

Fixing  $\sigma$ , we consider the random variables

$$W_t = \iota^D(\sigma; \tau_1 \dots \tau_t) - \iota^D(\sigma; \tau_1 \dots \tau_{t-1}).$$

By the definition of  $D$ ,  $\mathbf{E}[W_t] < -\text{bias}/2$  for any  $t > D$  (so long as the walk remains in  $R^+$ ); for  $t \leq D$ , the exact behavior of  $W_t$  may depend on  $\sigma$  but in any case  $W_t \leq \#\mathbf{a},\mathbf{h}(\tau_t)$ . To account for this, we define

$$Z_{\text{warm}} = \sum_{t=1}^D W_t \quad \text{and} \quad Z_t = W_{t+D} \quad (\text{for } t \geq 1).$$

In light of Fact 12 and Lemma 24,  $\Pr[Z_{\text{warm}} \geq k] = \exp(-\Omega(k))$  as it is a sum of a constant number of variables with exponential tails. We wish to show that

$$\Pr \left[ Z_{\text{init}} + Z_{\text{warm}} + \sum_{t=1}^k Z_t \geq D^2 \right] = \exp(-\Omega(k)),$$

and hence that  $\beta()$  returns to  $R_0$  quickly.

Though the random variables  $Z_i$  satisfy the conditions of Lemma 21, we cannot directly apply Lemma 22 to this sequence of random variables as they are not independent. However,  $Z_{t_1}$  and  $Z_{t_2}$  are independent if  $|t_1 - t_2| > D$  since the conditioning arising from  $\iota^D()$  only involves the previous  $D$  steps. Thus we may partition the  $k$  random variables into  $D$  subsets indexed by arithmetic progressions with multiple  $D$ ; each subset then contains random variables that are never closer than  $D$  from each other. The tail bounds of Lemma 22 apply to each subset. By the union bound, it then follows that there is a constant  $C > 0$  so that  $\Pr[Z_1 + \dots + Z_k > -Ck] = \exp(-\Omega(k))$ . Combining this with the bounds on  $Z_{\text{init}}$  and  $Z_{\text{warm}}$ , we conclude that the probability that the walk remains in  $R^+$  for  $k$  steps is  $\exp(-\Omega(k))$ , as desired.

**Dynamics in the region  $R^0$ .** Consider entry to  $R^0$  at time  $s \geq m$  and let  $\sigma = \sigma_1 \dots \sigma_s$ . We examine the random walk  $\beta_\ell(\sigma), \beta_\ell(\sigma; \tau_1), \dots$  by comparing it to  $\pi(\sigma; \epsilon), \pi(\sigma; \tau_1), \dots$ . Considering that the “width” of the region,  $2D^2 + 1$ , is a fixed constant the next distinct observed state will be  $R^-$  with nonzero constant probability; at worst, this is the probability of observing a sequence of back-to-back  $0^{\Delta-1}h$  steps that carry  $\pi()$ , and hence  $\beta()$ , into  $R^-$  (Lemma 7). Moreover, considering that each non-overlapping block of  $2D^2 + 1$  steps independently escapes from  $R^0$  to  $R^-$  with positive constant probability, it follows that the probability that the walk remains in  $R^0$  for more than  $k$  steps is  $\exp(-\Omega(k))$ .

**Dynamics in the region  $R^-$ .** The analysis is nearly identical to the case for  $R^+$ , though in this setting we must specially handle the event that the walk never returns to  $R^0$  (or  $R^+$ ). Consider entry to  $R^-$  at a time  $s \geq m$  and let  $\sigma = \sigma_1 \dots \sigma_s$ . We examine the random walk  $\beta_\ell(\sigma), \beta_\ell(\sigma; \tau_1), \dots$  by comparing it to  $\iota^D(\sigma; \epsilon), \iota^D(\sigma; \tau_1), \dots$ . As in the analysis of the  $R^+$  case, we condition on an arbitrary history  $\sigma$  and note that so long as  $\beta_\ell(\sigma; \tau_1 \dots \tau_t)$  remains in  $R^-$

$$\beta_\ell(\sigma; \tau_1 \dots \tau_t) \leq \iota^D(\sigma; \tau_1 \dots \tau_t);$$

again this follows from Claim 20. This yields the random variables

$$W_t = \iota^D(\sigma; \tau_1 \dots \tau_t) - \iota^D(\sigma; \tau_1 \dots \tau_{t-1}).$$

As in the case of  $R^+$ , it is convenient to decompose the steps of the walk into an initial “warm” region consisting of at least  $D$  steps—which may suffer from some conditioning from  $\sigma$ —and the remaining steps. By the definition of  $D$ ,  $\mathbb{E}[W_t] < -\text{bias}/2$  for any  $t > D$  (so long as the walk remains in  $R^-$ ); for  $t \leq D$ , the exact behavior of  $W_t$  may depend on  $\sigma$  but in any case  $W_t \leq \#_{a,h}(\tau_t)$ . To account for this, we will select a constant  $C_{\text{warm}} > D$  and define

$$Z_{\text{warm}} = \sum_{t=1}^{C_{\text{warm}}} W_t \quad \text{and} \quad Z_t = W_{t+C_{\text{warm}}} \quad (\text{for } t \geq 1).$$

As pointed out above, the random variables  $Z_t$  are independent of  $\sigma$ . We set the exact value of  $C_{\text{warm}}$  in the argument below.

First we establish that with constant probability, this walk never returns to  $R^0$  (or  $R^+$ ). Adopting the approach from the  $R^+$  case above, we partition the sequence of random variables  $Z_t$  into  $D$  families of i.i.d. random variables; specifically, let  $Z_i^{(s)} = Z_{Di+s}$  (for  $s \in \{1, \dots, D\}$ ). Then for each fixed  $s$  the sequence  $Z_i^{(s)}$  are independent random variables that satisfy the assumptions of Lemma 22 and hence Lemma 23. In particular, there is a constant  $C > 0$  so that

$$\Pr \left[ \exists t > 0, \sum_{i=0}^t Z_i^{(s)} \geq C \right] < \frac{1}{2D}$$

for each fixed  $s$ . Hence

$$\Pr \left[ \exists t > 0, \sum_{i=0}^t Z_i \geq C \cdot D \right] \leq \Pr \left[ \exists s \exists t > 0, \sum_{i=0}^t Z_i^{(s)} \geq C \right] < \frac{1}{2}. \quad (16)$$

We now assign  $C_{\text{warm}} = C \cdot D$ . To complete the argument, note that with constant probability the first  $C_{\text{warm}}$  of the random variables  $W_t$ —precisely those comprising  $Z_{\text{warm}}$ —all take the value  $-1$  as this is guaranteed by the possibility that each of these steps is  $0^{\Delta-1}h$ . The variables  $Z_i^{(1)}$  may depend on this conditioning, as  $D$ -typicality depends on the prior  $D$  steps; however, the conditioning assigns the “warm” random variables values that contain no adversarial symbols: this can only increase the probability that a particular  $Z_s^{(1)}$  is  $D$ -typical and hence reduce its expected value. We conclude that the  $Z_i^{(s)}$  satisfy equation (16), even under conditioning. It follows that with constant probability  $\sum_i^t W_i$  never rises above  $-D^2$  and hence that  $\beta()$  never departs  $R^-$ .

Conditioned on the event that  $\beta_\ell()$  never departs  $R^-$ , the value  $\beta_\ell(\sigma\tau_1 \dots \tau_D)$  is bounded above by  $-D^2$ ; hence  $\beta_\ell(\sigma\tau_1 \dots \tau_t)$  is bounded above by  $-D^2 + \sum_{i=1}^{i=t-D} Z_i$ . Again applying Lemma 22 to this sum (after decomposing it into  $\sum_i Z_i^{(s)}$  as above), we find that, for any  $M, T > 0$ ,

$$\Pr[\exists t \geq T, \beta(\sigma\tau_1 \dots \tau_t) \geq -M \mid \beta_\ell(\sigma\tau_1 \dots) \text{ never escapes } R^-] \leq \exp(-\Omega(T) + O(M)).$$

Finally, we wish to show that if  $\beta()$  returns to  $R^0$  (or  $R^+$ ), it does so quickly. This proceeds exactly as in the case of  $R^+$ : in light of Lemma 22, for a constant  $E > 0$ , the probability that any one of the sums  $\sum_{i=1}^t Z_i^{(s)}$  exceeds  $-Et$  is  $\exp(-\Omega(t))$  and hence that  $\beta()$  departs  $R^-$  quickly, if it does so at all.

Finally, consider the transitions among the states  $R^-$ ,  $R^0$ , and  $R^+$ . Any arrival into the state  $R^-$  results in the desired permanent descent beyond  $-M$  with constant probability. Otherwise, the waiting time to leave any of the states after entry has a worst-case exponential tail: specifically, there are constants  $A_{\text{trans}} > 0$  and  $a_{\text{trans}} > 1$  so that for any of the three states the probability that the waiting time between arrival and departure in that state exceeds  $k$  is no more than  $A_{\text{trans}} a_{\text{trans}}^{-k}$ . Furthermore, from either  $R^0$  or  $R^+$ , the probability of transitioning to  $R^-$  in the next two state transitions is a nonzero constant. It follows that  $T$ , the number of transitions that occur before observing the permanent descent, has an exponential tail. As the convolution of  $T$  waiting time distributions has an exponential tail by Lemma 24, it follows that the walk permanently descends past  $-M$  with certainty and, moreover, the number of steps before this event takes place has an exponential tail.  $\square$

#### 4.4 Bitcoin Security Threshold

**Theorem 26.** *If  $p_a < \vartheta(p_h, \Delta)$  then a Bitcoin execution over a lifetime of  $L$  slots achieves  $k$ -consistency and  $u$ -liveness except with error probabilities  $L \cdot \exp(-\Omega(k))$  and  $L \cdot \exp(-\Omega(u))$ , respectively.*

*If  $p_a > \vartheta(p_h, \Delta)$ , then the private chain attack is successful (with probability tending to 1 exponentially quickly), and Bitcoin is insecure.*

*Proof sketch.* The main claim is the positive statement for consistency. Consider an  $L$ -slot execution of the protocol and let  $\sigma_1, \sigma_2, \dots$  be the resulting sequence of steps as per Definition 9 (for convenience, we see this as an infinite sequence, while only being interested in its prefix covering the first  $L$  slots). First, recall that by Fact 12 and a union bound, we have

$$\Pr[\exists i \in [L]: |\sigma_i| \geq z] \leq L \cdot \exp(-\Omega(z)). \quad (17)$$

Moreover, given (11) and Lemma 25, for  $T \triangleq (p_h/4\Delta) \cdot k$  and for any  $m \in [L]$  we have  $|\sigma_m \sigma_{m+1} \dots \sigma_{m+T-1}| \leq k$  except with error  $\exp(-\Omega(k))$ .

We invoke Theorem 14 for each  $m \in [L]$ , using  $T$  as above and  $M \triangleq ck$  for a suitable constant  $c > 0$  depending on the constants hidden in the asymptotic notation of Theorem 14, so that the error term remains  $\exp(-\Omega(k))$  in each invocation. Applying union bound over all  $m$ , this gives us

$$\Pr[\exists m \in [L], \exists \tilde{t} \geq m + T: \beta_\ell(\sigma_1 \dots \sigma_{\tilde{t}}) \geq -ck] = L \cdot \exp(-\Omega(k)).$$

Recalling the notation of Section 2.4, this means that for any two slots  $\ell, t$  that are step boundaries separated by at least  $T$  steps, and hence also if they are separated by at least  $k$  slots, we have  $\beta_\ell(F_t) < -ck$  except with error  $\exp(-\Omega(k))$ .



We extend this to arbitrary slots  $\ell, t$  that are not necessarily step boundaries. Thanks to the Lipschitz property of  $\beta_\ell$  (Lemma 6), we know that in any slot  $t'$  belonging to the step following immediately after the slot  $t$ ,  $\beta_\ell(F_{t'})$  may differ from  $\beta_\ell(F_t)$  by at most  $t' - t$ , which can be upper-bounded by a sufficiently small multiple of  $k$  using (17), with error  $\exp(-\Omega(k))$ . By a similar argument, for any slot  $\ell'$  from the step immediately before the slot  $\ell$ ,  $\beta_{\ell'}(F_t)$  may differ from  $\beta_\ell(F_t)$  by at most  $\ell - \ell'$ , which can be similarly bounded using (17). To conclude, by a proper choice of  $c'$ , we have that for any  $\ell \in [L - k]$  and  $t \in \{\ell + k, \dots, L\}$ ,  $\beta_\ell(F_t) < -c'k$  except with probability  $L \cdot \exp(-\Omega(k))$ .

Finally, we transition from  $\beta_\ell$  to  $\beta_\ell^\Delta$ : for any  $w = \sigma_1 \dots \sigma_r w'$  where  $w'$  is an incomplete step,

$$\beta_\ell^\Delta(w) - \beta_\ell(w) = [\beta_\ell^\Delta(w) - \beta_\ell(\sigma_1 \dots \sigma_{r-1})] + [\beta_\ell(\sigma_1 \dots \sigma_{r-1}) - \beta_\ell(w)] \leq 2\#\_{h,a}(\sigma_r w'),$$

where the second inequality uses both Lemma 19 and Lemma 6. The resulting quantity can be again bounded using (17), we can hence conclude that for any  $\ell \in [L - k]$  and  $t \in \{\ell + k, \dots, L\}$ ,  $\beta_\ell^\Delta(F_t) < 0$  except with an overall error probability  $L \cdot \exp(-\Omega(k))$ , and invoke Lemma 1 to establish consistency.

For the positive result on liveness, the argument follows exactly the same path as in previous work (e.g., [14, 15]), with the single exception that the honest chain growth is lower-bounded by  $h_\Delta(\cdot)$  rather than the number of so-called “left-isolated slots” or “non-tailgaters.”

Finally, the negative result is straightforward: if  $p_a > \vartheta(p_h, \Delta)$ , then the expected growth rate of a private chain dominates the expected growth rate of an honest chain with maximally delayed blocks (with strong tail bounds), and so the private-chain attack succeeds with overwhelming probability.  $\square$

**Corollary 27.** *If  $r_a < \bar{\vartheta}(r_h, \Delta_0)$  then a Bitcoin execution over a lifetime of  $L$ , where (1) honest and adversarial PoW successes are modeled by Poisson processes with parameters  $r_h$  and  $r_a$ , respectively, and (2) honest messages are delayed by no more than  $\Delta_0$  time, achieves  $k$ -consistency and  $u$ -liveness except with error probabilities  $L \cdot \exp(-\Omega(k))$  and  $L \cdot \exp(-\Omega(u))$ , respectively.*

*If  $r_a > \bar{\vartheta}(r_h, \Delta_0)$ , then the private chain attack is successful (with probability tending to 1 exponentially quickly), and Bitcoin is insecure.*

*Proof.* This follows directly from the proof of Theorem 26. In particular, for a parameter  $s > 0$ —the length in time of a discrete slots—and define  $p_a = sr_a$ ,  $p_h = sr_h$ , and  $\Delta = \lceil \Delta_0 \rceil$ . In the limit, as  $s \rightarrow 0$ , this yields the Poisson model. As noted in Fact 11, it follows that for any  $s$ ,  $p_a < \vartheta(p_h, \Delta)$ , and the proof of Theorem 26 applies. A critical feature of the proof is that the conclusions are independent of  $s$ . In particular, the error rates and constants (e.g., “D”) selected in the proof are independent of  $s$ —indeed, the dynamics of walk are given by the bias, which is bounded for any  $s$  by  $-(1 - p_a/\vartheta) < -(1 - r_a/\bar{\vartheta})$ .

We remark that in the Poisson setting, we naturally wish to parameterize consistency and soundness in terms of absolute time (rather than an integer number of discrete slots, which would scale with  $1/s$ ). Note that the time  $t_{\text{step}}$  associated with a single step drawn from  $\mathcal{S}(p_a, p_h; \Delta)$  converges (as  $s \rightarrow 0$ ) to the shifted exponential distribution  $\Delta_0 + X$ , where  $X$  is exponentially distributed with parameter  $p_h$  (so that the density function of  $X$  is given by  $p_h \exp(-xp_h)$ ). Applying the tail bounds of Lemma 25 (for exponential random variables), we find that in time  $t$  one must observe  $\Omega(t/\Delta_0)$  of these “Poisson steps” except with probability  $\exp(-\Omega(t/\Delta_0))$ . Thus the error bounds of Theorem 26 scale in  $t/\Delta_0$ , as desired.  $\square$

## 5 Tight Security Threshold for Proof-of-Stake Blockchains

Our results apply, with small adaptations, also to Nakamoto-style protocols in the proof-of-stake (PoS) setting where they likewise yield a tight threshold. The PoS setting has a few notable differences which must be appropriately reflected in the proof.

**Protocol modeling.** PoS blockchain security can be analyzed in an abstract framework analogous to that described above for PoW blockchains. The fact of the matter is that PoS blockchains rely on fundamentally more sophisticated protocols because they are faced with the challenge of generating randomness for leader election and appropriately managing stake distribution snapshots. Despite this, typical PoS protocol security

proofs proceed by showing that randomness generation can be carried out with high fidelity and, hence, that the protocol can be analyzed in an idealized setting where slot leadership is determined by an i.i.d. distribution. (The possibility of adaptive corruption can lead to a distortion of this i.i.d. condition, but this can be generically managed by stochastic dominance arguments [1].) This yields the comparatively simple setting described below, which can be directly applied to the security proofs of such protocols as Snow White [3] and Ouroboros Praos [4].

**Multiple slot leaders; the PoS characteristic string alphabet.** In the PoS setting, one adopts a discrete time model to reflect the round-based nature of the protocol itself. This means that events involving multiple leaders per slot must be handled explicitly by the analysis. Specifically, the fundamental parameters of the protocol naturally determine three probabilities:

- $p_A$  – the probability of at least one adversarial leader,
- $p_H$  – the probability of at least one honest leader,
- $p_h$  – the probability of exactly one honest leader.

These probabilities determine a distribution on a richer alphabet of characteristic string symbols. That is, we now consider the alphabet  $\Sigma_{\text{PoS}} = \{0, \text{h}, \text{H}\} \times \{0, \text{A}\}$  and place the distribution on  $w = (x, y) \in \Sigma_{\text{PoS}}$  which independently assigns values to  $x \in \{0, \text{h}, \text{H}\}$  and  $y \in \{0, \text{A}\}$  so that:

- $x = \text{H}$  with probability  $p_H - p_h$ ;  $x = \text{h}$  with probability  $p_h$ .
- $y = \text{A}$  with probability  $p_A$ .

The interpretation of the symbol  $w = (x, y)$  mirrors the definitions of the probabilities above. The  $x$  symbol corresponds to honest participation in the slot in question: 0 indicates no honest leaders, h indicates exactly one honest leader, and H indicates a nonzero number of honest leaders. (For analytic purposes, it is convenient to permit  $x = \text{H}$  to be a valid assignment even if the slot has a unique leader.) Likewise,  $y = 0$  indicates that there are no adversarial leaders and  $y = \text{A}$  if there is at least one. As adversarial slot leaders can issue as many blocks as they please in a PoS slot, there is no need to distinguish the case of a unique adversarial leader. Observe that  $p_H$  is the probability of *any* honest slot leader in a particular slot (and that, in general,  $p_h \leq p_H$ ).

The security threshold in the PoS case is analogous to the PoW case:

$$p_A < \frac{1}{\Delta - 1/p_H}$$

under the added assumption that  $p_h$  is nonzero. Note that no particular relationship is assumed between  $p_h$  and  $p_H$ ; it suffices for  $p_h$  to be bounded above zero. This additional constraint (on  $p_h$ ) is necessary in general circumstances where the adversary may break ties in the longest-chain rule: even without any adversarial blocks, if all honest leaders are paired with at least one simultaneous leader a simple attack can force honest maintenance of two forking chains of equal depths. It is an interesting fact that with the stronger assumption of *deterministic tie breaking*, the positivity demand on  $p_h$  can be removed (see [10] for a detailed discussion and the full analysis of the synchronous case), and indeed this would also suffice in our setting. (We return to this point below.)

**The analysis: PoS  $\Delta$ -forks and PoS margin.** The definition of PoS fork is analogous; we only give a brief overview as our treatment directly follows the definitions of PoS  $\Delta$ -forks, relative margin, and settlement in [4, 1, 10]. The notion of a PoS  $\Delta$ -fork is adapted to the richer PoS characteristic strings in the natural way: the only alteration is axiom (A3), which is updated so that for a symbol  $w_i = (x_i, y_i)$ , we have

- if  $x_i = 0$  there are no honest vertices associated with slot  $i$ ; if  $x_i = \text{h}$  there is exactly one; if  $x_i = \text{H}$  there may be an arbitrary positive number; and

- if  $y_i = 0$  there are no adversarial vertices associated with slot  $i$ ; if  $y_i = A$  there may be an arbitrary number.

Observe that in case  $x_i = H$ , we place no particular constraints on the (positive) number of honest vertices that label the slot; as we effectively permit the adversary to “choose” the fork for a given characteristic string, this gives the adversary the power to determine the number of honest leaders if the characteristic string permits the possibility that there is more than one. As it happens, giving the adversary this extra latitude slightly simplifies the analysis without changing the (optimal) final results.

The most substantial change involves the elementary metric for measuring the ability of an adversary to produce settlement violations for a characteristic string. In the PoS setting this is exactly captured by the notion of relative margin; we reuse the name *margin* in this setting, since no confusion can arise. For a fork  $F \vdash_{\Delta} w$ , we define the relative  $\delta$ -margin as

$$\mu_{\ell}^{\delta}(F) \triangleq \max_{T_1 \neq T_2} \left( \min\{\alpha_F^{\delta}(T_1), \alpha_F^{\delta}(T_2)\} \right),$$

this maximum extended over all pairs of tines satisfying the criterion. As usual, we again overload the notation by defining

$$\mu_{\ell}^{\delta}(w) \triangleq \max_{F \vdash_{\Delta} w} \mu_{\ell}^{\delta}(F).$$

As in the PoW case, we will be considering the two variants  $\delta = \Delta$  and  $\delta = 1$ , which are again closely related. The variant with  $\delta = \Delta$  is directly relevant for settlement violations; on the other hand, the bulk of the analysis focuses on the simpler variant  $\mu_{\ell}^1(w)$ , which we simply denote  $\mu_{\ell}(w)$ .

It is convenient to define the *reach*  $\rho$  as the simpler quantity  $\rho(F) = \max_T \alpha_F(T)$  and, likewise,  $\rho(w) = \max_{F \vdash_{\Delta} w} \rho(F)$ . Margin plays a role directly analogous to the PoW setting: that is, a settlement violation can occur in slot  $\ell$  at time  $\ell + t$  exactly if  $\mu_{\ell}^{\Delta}(w) \geq 0$ , where  $w$  is the characteristic string through time  $\ell + t$ . While  $\mu$  is a structurally distinct quantity from  $\beta$ , it can be analyzed using exactly the machinery we have already developed. Specifically,

- When  $\mu_{\ell}(w) < 0$ ,  $\mu_{\ell}(w) = \beta_{\ell}(w)$  and thus adheres precisely to the recurrence for the Cold region, described in Section 3.4.
- When  $\mu_{\ell}(w) \geq 0$ , its exact behavior is unimportant because  $\mu_{\ell}(w) \leq \rho(w)$  and  $\mu_{\ell}()$  cannot descend below zero until  $\rho(w)$  returns to zero. In particular, once  $\mu_{\ell}()$  rises to zero, the quantity of interest becomes  $\rho(w)$ .  $\rho()$  simply follows directly the recurrence and dynamics for  $\beta_{\ell}()$  in the pre- $\ell$  regime, described in Section 3.2. As in the analysis of Ouroboros Praos [1], this somewhat complicates the analysis of the resulting random walk because  $\rho()$  will typically be nonzero when  $\mu_{\ell}()$  climbs to zero, so the resulting return time of  $\rho()$  to zero depends on its initial height. One can develop a highly precise accounting for this initial height, but for our purposes it suffices to simply bound it with the length of the previous excursion of  $\mu_{\ell}()$  below zero (which is already controlled by the analysis).
- The final point of interest is the critical region around zero. Here the same argument works, though one must adopt a stronger notion of “magic” sequence to guarantee descending across 0. In particular, one identifies the sequence of  $(\Delta - 1)$ -isolated occurrences of *uniquely* honest slots (with no adversarial presence) to guarantee descent. It is an interesting fact that this is the only place in the analysis that requires unique honest leaders. Of course, this changes the constant probability associated with descent compared to the PoW case.

## References

- [1] E. Blum, A. Kiayias, C. Moore, S. Quader, and A. Russell. The combinatorics of the longest-chain rule: Linear consistency for proof-of-stake blockchains. In S. Chawla, editor, *31st SODA*, pages 1135–1154. ACM-SIAM, Jan. 2020.

- [2] L. L. Cam. On the distribution of sums of independent random variables. In *Bernoulli, Bayes, Laplace: Anniversary Volume*. Springer-Verlag, 1965.
- [3] P. Daian, R. Pass, and E. Shi. Snow white: Robustly reconfigurable consensus and applications to provably secure proof of stake. In I. Goldberg and T. Moore, editors, *FC 2019*, volume 11598 of *LNCS*, pages 23–41. Springer, Heidelberg, Feb. 2019.
- [4] B. David, P. Gazi, A. Kiayias, and A. Russell. Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake blockchain. In J. B. Nielsen and V. Rijmen, editors, *EUROCRYPT 2018, Part II*, volume 10821 of *LNCS*, pages 66–98. Springer, Heidelberg, Apr. / May 2018.
- [5] A. Dembo, S. Kannan, E. N. Tas, D. Tse, P. Viswanath, X. Wang, and O. Zeitouni. Everything is a Race and Nakamoto Always Wins. <https://arxiv.org/abs/2005.10484>, 2020.
- [6] I. Eyal and E. G. Sirer. Majority is not enough: Bitcoin mining is vulnerable. In N. Christin and R. Safavi-Naini, editors, *FC 2014*, volume 8437 of *LNCS*, pages 436–454. Springer, Heidelberg, Mar. 2014.
- [7] J. Garay, A. Kiayias, and N. Leonardos. The bitcoin backbone protocol: Analysis and applications. Cryptology ePrint Archive, Report 2014/765, revision Feb 2017, 2017. <https://eprint.iacr.org/2014/765>.
- [8] J. A. Garay, A. Kiayias, and N. Leonardos. The bitcoin backbone protocol: Analysis and applications. In E. Oswald and M. Fischlin, editors, *EUROCRYPT 2015, Part II*, volume 9057 of *LNCS*, pages 281–310. Springer, Heidelberg, Apr. 2015.
- [9] Y. Jin, Y. Li, Y. Wang, and Y. Zhou. On asymptotically tight tail bounds for sums of geometric and exponential random variables, 2019.
- [10] A. Kiayias, S. Quader, and A. Russell. Consistency in proof-of-stake blockchains with concurrent honest slot leaders. *IACR Cryptology ePrint Archive*, 2020:41, 2020.
- [11] A. Kiayias, A. Russell, B. David, and R. Oliynykov. Ouroboros: A provably secure proof-of-stake blockchain protocol. In J. Katz and H. Shacham, editors, *CRYPTO 2017, Part I*, volume 10401 of *LNCS*, pages 357–388. Springer, Heidelberg, Aug. 2017.
- [12] L. Kiffer, R. Rajaraman, and a. shelat. A better method to analyze blockchain consistency. In D. Lie, M. Mannan, M. Backes, and X. Wang, editors, *ACM CCS 2018*, pages 729–744. ACM Press, Oct. 2018.
- [13] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [14] R. Pass, L. Seeman, and a. shelat. Analysis of the blockchain protocol in asynchronous networks. In J. Coron and J. B. Nielsen, editors, *EUROCRYPT 2017, Part II*, volume 10211 of *LNCS*, pages 643–673. Springer, Heidelberg, Apr. / May 2017.
- [15] L. Ren. Analysis of Nakamoto consensus. Cryptology ePrint Archive, Report 2019/943, 2019. <https://eprint.iacr.org/2019/943>.

## A Omitted Proofs

### A.1 Proof of Claim 13

*Proof.* We have

$$\begin{aligned}
\mathbb{E}[\#\alpha(\sigma)] &= \sum_{\ell=\Delta}^{\infty} \Pr[|\sigma| = \ell] \cdot \mathbb{E}[\#\alpha(\sigma) \mid |\sigma| = \ell] \\
&= \sum_{\ell=\Delta}^{\infty} \Pr[|\sigma| = \ell] \cdot (\ell - 1)p_\alpha \\
&\leq p_\alpha \sum_{\ell=\Delta}^{\infty} \Pr[|\sigma| = \ell] \cdot \ell = p_\alpha \mathbb{E}[|\sigma|] \\
&= p_\alpha((\Delta - 1) + \mathbb{E}[G_{p_\alpha}]) \\
&= p_\alpha((\Delta - 1) + 1/p_h) = p_\alpha/\vartheta < 1,
\end{aligned}$$

where  $G_p$  denotes a geometrically distributed random variable with distribution  $\Pr[G_p = t] = (1-p)^{t-1}p$  (for  $t \geq 1$ ).  $\square$

### A.2 Proof of Lemma 16

*Proof.* Fix  $p \in [0, 1]$  and define

$$h_n = \mathbb{E}[h_\Delta(X_1, \dots, X_n)]$$

where, as in the statement of the lemma, the  $X_i$  are independent Bernoulli random variables with  $\mathbb{E}[X_i] = p$ . To match the definition of  $h_\Delta$ , we define  $h_0 = 0$ . Expanding the expectation around the outcome of the last random variable  $X_n$ , we find that

$$h_n = (1-p)h_{n-1} + p(1 + h_{n-\Delta}) \quad \text{for all } n > \Delta. \quad (18)$$

Likewise, we find that  $h_n = (1-p)h_{n-1} + p$  for all  $0 < n \leq \Delta$ . In this regime (where  $n \leq \Delta$ ) we can directly solve for  $h_n$ :

$$h_n = 1 - (1-p)^n \quad \text{for } 0 < n \leq \Delta.$$

Define  $f(n) = n/\alpha$ . A calculation confirms that  $f(\cdot)$  satisfies the recurrence relation (18) for all  $n \in \mathbb{Z}$ :

$$f(n) = (1-p) \cdot f(n-1) + p[1 + f(n-\Delta)].$$

To complete the proof, we proceed by induction. For any  $0 \leq n \leq \Delta$ , both  $h_n$  and  $f(n)$  lie in the interval  $[0, 1]$  which establishes (12). Assuming (12) for all  $k < n$  (where  $n > \Delta$ ), we observe that

$$f(n) - h_n = (1-p)[f(n-1) - h_{n-1}] + p[f(n-\Delta) - h_{n-\Delta}]$$

as both  $f$  and  $h$  satisfy (18). Thus

$$\begin{aligned}
|f(n) - h_n| &\leq (1-p)|f(n-1) - h_{n-1}| + p|f(n-\Delta) - h_{n-\Delta}| \\
&\leq (1-p) + p = 1,
\end{aligned}$$

as desired.

Finally, note that  $|h_\Delta(x_1, \dots, x_n) - h_\Delta(y_1, \dots, y_n)| \leq 1$  if  $(x_1, \dots, x_n)$  and  $(y_1, \dots, y_n)$  differ in only one coordinate. The large deviation bound follows directly, then, from the McDiarmid inequality (Theorem 15).  $\square$

### A.3 Proof of Lemma 17

*Proof.* Let  $g_k = \Pr[X_i = k]$  and let  $G(x) = \sum_{k \geq 0} g_k x^k$  denote the corresponding ordinary generating function associated with the random variables  $X_i$ . By assumption, for sufficiently small  $a > 1$ ,  $g_k \leq A \cdot a^{-k}$ ; it follows that  $G(x)$  is well-defined and differentiable around 1. As  $\mathbb{E}[X] < 1$ , expanding the derivative we find that  $G'(1) = \mathbb{E}[X_i] < 1$ ; in particular, for sufficiently small  $b > 1$ ,  $G(b) < b$ . Thus, we may adopt an  $a_* = \min(a, b)$  satisfying both of these inequalities. Let

$$B = Aa_*/(a_* - G(a_*)).$$

Then we show by induction that (for all  $t$  and  $k$ )  $\Pr[W_t = k] \leq Ba_*^{-k}$ . The base case  $t = 0$  is immediate.

$$\begin{aligned} \Pr[W_t = k] &= \Pr[W_{t-1} = 0] \cdot g_k + \sum_{j=0}^k g_j \Pr[W_{t-1} = k - j + 1] \\ &\leq Aa_*^{-k} + \sum_{j=0}^{\infty} g_j \Pr[W_{t-1} = k - j + 1] \\ &\leq Aa_*^{-k} + \sum_{j=0}^{\infty} g_j Ba_*^{-k+j-1} \\ &= Aa_*^{-k} + Ba_*^{-k-1} \sum_{j=0}^{\infty} g_j a_*^j \\ &= Aa_*^{-k} + Ba_*^{-k-1} G(a_*) \\ &= a_*^{-k} (A + BG(a_*)/a_*) = Ba_*^{-k}, \end{aligned}$$

as desired. □

### A.4 Proof of Lemma 21

*Proof.* Let  $\alpha = 1/a < 1$ . Assume that  $\lambda < 1/e$  is small enough to satisfy the following additional inequalities:

$$\lambda \ln^2(\lambda^{-1}) \leq \frac{|\mathbb{E}[A]|}{4 \cdot 3^2} (1 - \sqrt{\alpha})^2 \leq^* \frac{|\mathbb{E}[A]|}{4 \cdot 6^2} \ln^2(1/\alpha), \quad \text{and} \quad (19)$$

$$\lambda \leq 1/C. \quad (20)$$

The inequality  $\leq^*$  above follows directly from the fact that  $1 - \sqrt{\alpha} < (1/2) \ln(1/\alpha)$  for all  $\alpha \in (0, 1)$ . Decompose  $m_A(\lambda)$  into two sums:

$$m_A(\lambda) = \underbrace{\sum_{-1 \leq k < S} e^{\lambda k} \Pr[A = k]}_{(\dagger)} + \underbrace{\sum_{S \leq k} e^{\lambda k} \Pr[A = k]}_{(\ddagger)},$$

where  $S = 2 \cdot \ln(\lambda^2/C)/\ln(\alpha)$  is a threshold chosen to balance error terms determined below. We record the fact that under constraint (20) we have the simpler upper bound

$$S \leq 6 \ln(\lambda)/\ln(\alpha). \quad (21)$$

Consider the sum  $(\dagger)$ . We first record two estimates applied in the bound. Observe that

$$\lambda S^2 \leq \lambda \left( \frac{6 \ln(\lambda^{-1})}{\ln(\alpha^{-1})} \right)^2 \leq \lambda \ln^2(\lambda^{-1}) \left( \frac{6}{\ln(\alpha^{-1})} \right)^2 \leq \frac{|\mathbb{E}[A]|}{4} \quad (22)$$

and hence  $\lambda S \leq \lambda S^2 < 1/2$ . We additionally remark that for  $|\delta| < 1$ ,  $|\exp(\delta) - (1 + \delta)| \leq \delta^2/[2(1 - |\delta|)]$ . Then we find that

$$\begin{aligned} \sum_{-R \leq k < S} e^{\lambda k} \Pr[A = k] &\leq \sum_{-R \leq k < S} \left(1 + \lambda k + \frac{\lambda^2 k^2}{1 - |\lambda k|}\right) \Pr[A = k] \\ &\leq 1 + \lambda \mathbf{E}[A] + \frac{\lambda^2 S^2}{2(1 - \lambda S)} \\ &\leq 1 + \lambda \mathbf{E}[A] + \lambda^2 S^2. \end{aligned} \tag{23}$$

$$\leq 1 + \lambda \mathbf{E}[A] + \frac{\lambda}{4} |\mathbf{E}[A]|, \tag{24}$$

where line 23 follows because  $\lambda S \leq 1/2$ , as noted above, and line 24 follows from (22).

As for the sum (‡), in light of constraint 19 we find that

$$\begin{aligned} \sum_{S \leq k} e^{\lambda k} \Pr[A = k] &\leq \sum_{S \leq k} e^{\lambda k} C \alpha^k = \frac{C(\alpha e^\lambda)^S}{1 - \alpha e^\lambda} \\ &\leq \frac{C \sqrt{\alpha}^S}{1 - \sqrt{\alpha}} \leq \frac{\lambda^2}{1 - \sqrt{\alpha}} \leq \lambda \frac{|\mathbf{E}[A]|}{4}. \end{aligned} \tag{25}$$

where we used the fact that  $\lambda < \ln(1/\sqrt{\alpha})$  and hence  $\alpha e^\lambda < \sqrt{\alpha}$  in the first inequality of the last line. Thus

$$m_A(\lambda) \leq 1 + \lambda \mathbf{E}[A] + 2 \frac{|\mathbf{E}[A]|}{4} \leq 1 + \lambda \frac{\mathbf{E}[A]}{2},$$

as desired.  $\square$

## A.5 Proof of Lemma 22

*Proof.* Applying Lemma 21 to the random variables  $Z_i$ , there is a constant  $\lambda^*$  for which  $m_{Z_i}(\lambda^*) \leq 1 - \lambda^* \gamma/2 \leq \exp(-\lambda^* \gamma/2)$ . As the  $Z_i$  are independent,

$$m_{S_n}(\lambda^*) = \mathbf{E} \left[ e^{\lambda^* \sum_i^n Z_i} \right] = \prod_i^n \mathbf{E} \left[ e^{\lambda^* Z_i} \right] \leq \exp(-n \lambda^* \gamma/2).$$

Thus

$$\begin{aligned} \Pr[S_n \geq T] &= \Pr[e^{\lambda^* S_n} \geq e^{\lambda^* T}] \leq \frac{\mathbf{E} [e^{\lambda^* S_n}]}{e^{\lambda^* T}} \\ &\leq e^{-n \lambda^* \gamma/2} e^{-\lambda^* T} = \exp(-\lambda^* (T + n \gamma/2)). \end{aligned} \tag{26}$$

## A.6 Proof of Lemma 23

*Proof of Lemma 23.* For a constant  $D$ ,

$$\begin{aligned} \Pr[\exists n > 0, S_n \geq D] &\leq \sum_{n=1}^{\infty} \Pr[S_n \geq D] \leq \sum_{n=1}^{\infty} e^{-\alpha [D + \gamma n/2]} \\ &= e^{-\alpha D} \sum_{n=1}^{\infty} e^{-\gamma n/2} = \frac{e^{-\alpha D}}{e^{\gamma/2} - 1}, \end{aligned} \tag{26}$$

where  $\alpha$  is the constant promised by Lemma 22. Let  $D^*$  be a constant for which (26) is less than 1. Then, with non-zero probability  $Z_1 = Z_2 = \dots = Z_{D^*} = -1$ , so that  $S_n = -D^*$  and no future  $S_n$  is zero.  $\square$

## A.7 Proof of Lemma 24

*Proof.* Let  $G(X) = \sum_k a_k X^k$  and  $H(X) = \sum_k b_k X^k$  be the ordinary generating functions for the random variables  $G$  and  $H_i$ , respectively. By assumption there are constants  $a > 1$  and  $A > 0$  for which  $a_k = \Pr[G = k] \leq Aa^{-k}$  and constants  $b > 1$  and  $B > 0$  for which  $b_k = \Pr[H_i = k] \leq Bb^{-k}$ . Thus  $G(X)$  converges inside  $[0, a)$ ; likewise  $H(X)$  converges inside  $[0, b)$ .

Recall that  $G(X) \times H(X) = \sum c_n X^n$  is the generating function for  $G + H_1$ . As  $G(X) \cdot H(X)$  converges in  $[0, \min(a, b))$  (and,  $\min(a, b) > 1$ ),  $\Pr[G + H_1 = k] = c_k = \exp(-\Omega(k))$ . The tail bound in the statement of the theorem follows immediately.

Recall that  $G(H(X))$  is the generating function associated with the convolution of  $G$  copies of  $H$  (the random variable  $\sum_{i=1}^G H_i$ ). As  $H$  converges in a neighborhood around 1 and  $\lim_{z \rightarrow 1^+} H(z) = 1$ , there is a value  $z^* > 1$  for which  $H(z^*)$  converges to a value less than  $a$  (which is  $> 1$ ). Then  $G(H(z^*))$  converges; writing  $G(H(Z)) = \sum_k c_k X^k$ , we conclude that  $c_k < \exp(-\Omega(k))$ . The tail bound in the statement of the theorem follows immediately.  $\square$

## A.8 Proof of Lemma 25

This is a result of [9, Thm. 1].