ALMOST ALL ORBITS OF THE COLLATZ MAP ATTAIN ALMOST BOUNDED VALUES

TERENCE TAO

Abstract. Define the *Collatz map* $\mathrm{Col}\colon \mathbb{N}+1 \to \mathbb{N}+1$ on the positive integers $\mathbb{N}+1 = \{1,2,3,\dots\}$ by setting $\mathrm{Col}(N)$ equal to $3N+1$ when N is odd and $N/2$ when N is even, and let $\mathrm{Col}_{\min}(N) := \inf_{n\in\mathbb{N}} \mathrm{Col}^n(N)$ denote the minimal element of the Collatz orbit $N, \mathrm{Col}(N), \mathrm{Col}^2(N), \dots$. The infamous *Collatz conjecture* asserts that $\mathrm{Col}_{\min}(N) = 1$ for all $N \in \mathbb{N}+1$. Previously, it was shown by Korec that for any $\theta > \frac{\log 3}{\log 4} \approx 0.7924$, one has $\mathrm{Col}_{\min}(N) \leq N^\theta$ for almost all $N \in \mathbb{N}+1$ (in the sense of natural density). In this paper we show that for *any* function $f\colon \mathbb{N}+1 \to \mathbb{R}$ with $\lim_{N\to\infty} f(N) = +\infty$, one has $\mathrm{Col}_{\min}(N) \leq f(N)$ for almost all $N \in \mathbb{N}+1$ (in the sense of logarithmic density). Our proof proceeds by establishing a stabilisation property for a certain first passage random variable associated with the Collatz iteration (or more precisely, the closely related Syracuse iteration), which in turn follows from estimation of the characteristic function of a certain skew random walk on a 3-adic cyclic group $\mathbb{Z}/3^n\mathbb{Z}$ at high frequencies. This estimation is achieved by studying how a certain two-dimensional renewal process interacts with a union of triangles associated to a given frequency.

1. Introduction

1.1. **Statement of main result.** Let $\mathbb{N} := \{0,1,2,\dots\}$ denote the natural numbers, so that $\mathbb{N}+1 = \{1,2,3,\dots\}$ are the positive integers. The *Collatz map* $\mathrm{Col}\colon \mathbb{N}+1 \to \mathbb{N}+1$ is defined by setting $\mathrm{Col}(N) := 3N+1$ when N is odd and $\mathrm{Col}(N) := N/2$ when N is even. For any $N \in \mathbb{N}+1$, let $\mathrm{Col}_{\min}(N) := \min \mathrm{Col}^{\mathbb{N}}(N) = \inf_{n\in\mathbb{N}} \mathrm{Col}^n(N)$ denote the minimal element of the Collatz orbit $\mathrm{Col}^{\mathbb{N}}(N) := \{N, \mathrm{Col}(N), \mathrm{Col}^2(N), \dots\}$. We have the infamous *Collatz conjecture* (also known as the $3x+1$ conjecture):

Conjecture 1.1 (Collatz conjecture)**.** *We have* $\mathrm{Col}_{\min}(N) = 1$ *for all* $N \in \mathbb{N}+1$.

We refer the reader to [9], [5] for extensive surveys and historical discussion of this conjecture.

While the full resolution of Conjecture 1.1 remains well beyond reach of current methods, some partial results are known. Numerical computation has verified $\mathrm{Col}_{\min}(N) = 1$ for all $N \leq 5.78 \times 10^{18}$ [12], and more recently for all $N \leq 10^{20}$ [13], while Krasikov and Lagarias [8] showed that

$$\#\{N \in \mathbb{N}+1 \cap [1,x] : \mathrm{Col}_{\min}(N) = 1\} \gg x^{0.84}$$

2010 *Mathematics Subject Classification.* 37P99.

for all sufficiently large x, where $\#E$ denotes the cardinality of a finite set E, and our conventions for asymptotic notation are set out in Section 2. In this paper we will focus on a different type of partial result, in which one establishes upper bounds on the minimal orbit value $\mathrm{Col}_{\min}(N)$ for "almost all" $N \in \mathbb{N}+1$. For technical reasons, the notion of "almost all" that we will use here is based on logarithmic density, which has better approximate multiplicative invariance properties than the more familiar notion of natural density (see [14] for a related phenomenon in a more number-theoretic context). Due to the highly probabilistic nature of the arguments in this paper, we will define logarithmic density using language of probability theory.

Definition 1.2 (Almost all). Given a finite non-empty subset R of $\mathbb{N}+1$, we define[1] $\mathbf{Log}(R)$ to be a random variable taking values in R with the logarithmically uniform distribution

$$\mathbb{P}(\mathbf{Log}(R) \in A) = \frac{\sum_{N \in A \cap R} \frac{1}{N}}{\sum_{N \in R} \frac{1}{N}}$$

for all $A \subset \mathbb{N}+1$. The *logarithmic density* of a set $A \subset \mathbb{N}+1$ is then defined to be $\lim_{x\to\infty} \mathbb{P}(\mathbf{Log}(\mathbb{N}+1 \cap [1,x]) \in A)$, provided that the limit exists. We say that a property $P(N)$ holds for *almost all* $N \in \mathbb{N}+1$ if $P(N)$ holds for N in a subset of $\mathbb{N}+1$ of logarithmic density 1, or equivalently if

$$\lim_{x\to\infty} \mathbb{P}(P(\mathbf{Log}(\mathbb{N}+1 \cap [1,x]))) = 1.$$

In Terras [15] it was shown that $\mathrm{Col}_{\min}(N) < N$ for almost all N. This was improved by Allouche [1] to $\mathrm{Col}_{\min}(N) < N^\theta$ for almost all N, and any fixed constant $\theta > 0.869$; the range of θ was later extended to $\theta > \frac{\log 3}{\log 4} \approx 0.7924$ by Korec [6]. (Indeed, in these results one can use natural density instead of logarithmic density to define "almost all'.) It is tempting to try to iterate these results to lower the value of θ further. However, one runs into the difficulty that the uniform (or logarithmic) measure does not enjoy any invariance properties with respect to the Collatz map: in particular, even if it is true that $\mathrm{Col}_{\min}(N) < x^\theta$ for almost all $N \in [1,x]$, and $\mathrm{Col}_{\min}(N') \leq x^{\theta^2}$ for almost all $N' \in [1,x^\theta]$, the two claims cannot be immediately concatenated to imply that $\mathrm{Col}_{\min}(N) \leq x^{\theta^2}$ for almost all $N \in [1,x]$, since the Collatz iteration may send almost all of $[1,x]$ into a very sparse subset of $[1,x^\theta]$, and in particular into the exceptional set of the latter claim $\mathrm{Col}_{\min}(N') \leq x^{\theta^2}$.

Nevertheless, in this paper we show that it is possible to locate an alternate probability measure (or more precisely, a family of probability measures) on the natural numbers with enough invariance properties that an iterative argument does become fruitful. More precisely, the main result of this paper is the following improvement of these "almost all" results.

[1]In this paper all random variables will be denoted by boldface symbols, to distinguish them from purely deterministic quantities that will be denoted by non-boldface symbols. When it is only the distribution of the random variable that is important, we will use multi-character boldface symbols such as **Log**, **Unif**, or **Geom** to denote the random variable, but when the dependence or independence properties of the random variable are also relevant, we shall usually use single-character boldface symbols such as **a** or **j** instead.

Theorem 1.3 (Almost all Collatz orbits attain almost bounded values)**.** *Let* $f\colon \mathbb{N}+1 \to \mathbb{R}$ *be any function with* $\lim_{N\to\infty} f(N) = +\infty$*. Then one has* $\mathrm{Col}_{\min}(N) < f(N)$ *for almost all* $N \in \mathbb{N}+1$*.*

Thus for instance one has $\mathrm{Col}_{\min}(N) < \log\log\log\log N$ for almost all N.

Remark 1.4. One could ask whether it is possible to sharpen the conclusion of Theorem 1.3 further, to assert that there is an absolute constant C_0 such that $\mathrm{Col}_{\min}(N) \leq C_0$ for almost all $N \in \mathbb{N}+1$. However this question is likely to be almost as hard to settle as the full Collatz conjecture, and out of reach of the methods of this paper. Indeed, suppose for any given C_0 that there existed an orbit $\mathrm{Col}^{\mathbb{N}}(N_0) = \{N_0, \mathrm{Col}(N_0), \mathrm{Col}^2(N_0), \dots\}$ that never dropped below C_0 (this is the case if there are infinitely many periodic orbits, or if there is at least one unbounded orbit). Then probabilistic heuristics (such as (1.16) below) suggest that for a positive density set of $N \in \mathbb{N}+1$, the orbit $\mathrm{Col}^{\mathbb{N}}(N) = \{N, \mathrm{Col}(N), \mathrm{Col}^2(N), \dots\}$ should encounter one of the elements $\mathrm{Col}^n(N_0)$ of the orbit of N_0 before going below C_0, and then the orbit of N will never dip below C_0. However, Theorem 1.3 is easily seen to be equivalent to the assertion that for any $\delta > 0$, there exists a constant C_δ such that $\mathrm{Col}_{\min}(N) \leq C_\delta$ for all N in a subset of $\mathbb{N}+1$ of logarithmic density at least $1-\delta$ (in fact our arguments give a constant of the form $C_\delta \ll \exp(\delta^{-O(1)})$). In particular[2], it is possible in principle that a sufficiently explicit version of the arguments here, when combined with numerical verification of the Collatz conjecture, can be used to show that the Collatz conjecture holds for a set of N of positive logarithmic density. Also, it is plausible that some refinement of the arguments below will allow one to replace logarithmic density by natural density in the definition of "almost all".

1.2. **Syracuse formulation.** We now discuss the methods of proof of Theorem 1.3. It is convenient to replace the Collatz map $\mathrm{Col}\colon \mathbb{N}+1 \to \mathbb{N}+1$ with a slightly more tractable acceleration $N \mapsto \mathrm{Col}^{f(N)}(N)$ of that map. One common instance of such an acceleration in the literature is the map $\mathrm{Col}_2\colon \mathbb{N}+1 \to \mathbb{N}+1$, defined by setting $\mathrm{Col}_2(N) := \mathrm{Col}^2(N) = \frac{3N+1}{2}$ when N is odd and $\mathrm{Col}_2(N) := \frac{N}{2}$ when N is even. Each iterate of the map Col_2 performs exactly one division by 2, and for this reason C_2 is a particularly convenient choice of map when performing "2-adic" analysis of the Collatz iteration. It is easy to see that $\mathrm{Col}_{\min}(N) = (\mathrm{Col}_2)_{\min}(N)$ for all $N \in \mathbb{N}+1$, so all the results in this paper concerning Col may be equivalently reformulated using Col_2. The triple iterate Col^3 was also recently proposed as an acceleration in [4]. However, the methods in this paper will rely instead on "3-adic" analysis, and it will be preferable to use an acceleration of the Collatz map which performs exactly one multiplication by 3 per iteration. More precisely, let $2\mathbb{N}+1 = \{1, 3, 5, \dots\}$ denote the odd natural numbers, and define the *Syracuse map* $\mathrm{Syr}\colon 2\mathbb{N}+1 \to 2\mathbb{N}+1$ (OEIS A075677) to be the largest odd number dividing $3N+1$; thus for instance

$$\mathrm{Syr}(1) = 1; \quad \mathrm{Syr}(3) = 5; \quad \mathrm{Syr}(5) = 1; \quad \mathrm{Syr}(7) = 11.$$

Equivalently, one can write

$$\mathrm{Syr}(N) = \mathrm{Col}^{\nu_2(3N+1)+1}(N) = \mathrm{Aff}_{\nu_2(3N+1)}(N) \tag{1.1}$$

[2]We thank Ben Green for this observation.

where for each positive integer $a \in \mathbb{N}+1$, $\mathrm{Aff}_a \colon \mathbb{R} \to \mathbb{R}$ denotes the affine map

$$\mathrm{Aff}_a(x) := \frac{3x+1}{2^a}$$

and for each integer M and each prime p, the *p-valuation* $\nu_p(M)$ of M is defined as the largest natural number a such that p^a divides M (with the convention $\nu_p(0) = +\infty$). (Note that $\nu_2(3N+1)$ is always a positive integer when N is odd.) For any $N \in 2\mathbb{N}+1$, let $\mathrm{Syr}_{\min}(N) := \min \mathrm{Syr}^{\mathbb{N}}(N)$ be the minimal element of the Syracuse orbit

$$\mathrm{Syr}^{\mathbb{N}}(N) := \{N, \mathrm{Syr}(N), \mathrm{Syr}^2(N), \dots\}.$$

This Syracuse orbit $\mathrm{Syr}^{\mathbb{N}}(N)$ is nothing more than the odd elements of the corresponding Collatz orbit $\mathrm{Col}^{\mathbb{N}}(N)$, and from this observation it is easy to verify the identity

$$\mathrm{Col}_{\min}(N) = \mathrm{Syr}_{\min}(N/2^{\nu_2(N)}) \tag{1.2}$$

for any $N \in \mathbb{N}+1$. Thus, the Collatz conjecture can be equivalently rephrased as

Conjecture 1.5 (Collatz conjecture, Syracuse formulation)**.** *We have* $\mathrm{Syr}_{\min}(N) = 1$ *for all* $N \in 2\mathbb{N}+1$.

We may similarly reformulate Theorem 1.3 in terms of the Syracuse map. We say that a property $P(N)$ holds for *almost all* $N \in 2\mathbb{N}+1$ if

$$\lim_{x \to \infty} \mathbb{P}(P(\mathbf{Log}(2\mathbb{N}+1 \cap [1,x])) = 1,$$

or equivalently if $P(N)$ holds for a set of odd natural numbers of logarithmic density $1/2$.

Theorem 1.6 (Almost all Syracuse orbits attain almost bounded values)**.** *Let* $f \colon 2\mathbb{N}+1 \to \mathbb{R}$ *be a function with* $\lim_{N\to\infty} f(N) = +\infty$*. Then one has* $\mathrm{Syr}_{\min}(N) < f(N)$ *for almost all* $N \in 2\mathbb{N}+1$.

Indeed, if Theorem 1.6 holds and $f \colon \mathbb{N}+1 \to \mathbb{R}$ is such that $\lim_{N\to\infty} f(N) = +\infty$, then from (1.2) we see that for any $a \in \mathbb{N}$, the set of $N \in \mathbb{N}+1$ with $\nu_2(N) = a$ and $\mathrm{Col}_{\min}(N) = \mathrm{Syr}_{\min}(N/2^a) < f(N)$ has logarithmic density 2^{-a}. Summing over any finite range $0 \le a \le a_0$ we obtain a set of logarithmic density $1 - 2^{-a_0}$ on which the claim $\mathrm{Col}_{\min}(N) < f(N)$ holds, and on sending a_0 to infinity one obtains Theorem 1.3. The converse implication (which we will not need) is also straightforward and left to the reader.

The iterates Syr^n of the Syracuse map can be described explicitly as follows. For any finite tuple $\vec{a} = (a_1, \dots, a_n) \in (\mathbb{N}+1)^n$ of positive integers, we define the composition $\mathrm{Aff}_{\vec{a}} = \mathrm{Aff}_{a_1,\dots,a_n} \colon \mathbb{R} \to \mathbb{R}$ to be the affine map

$$\mathrm{Aff}_{a_1,\dots,a_n}(x) := \mathrm{Aff}_{a_n}(\mathrm{Aff}_{a_{n-1}}(\dots(\mathrm{Aff}_{a_1}(x))\dots)).$$

A brief calculation shows that

$$\mathrm{Aff}_{a_1,\dots,a_n}(x) = 3^n 2^{-|\vec{a}|} x + F_n(\vec{a}) \tag{1.3}$$

where the *size* $|\vec{a}|$ of a tuple $\vec{a}$ is defined as

$$|\vec{a}| := a_1 + \dots + a_n, \tag{1.4}$$

and we define the n-*Syracuse offset map* $F_n\colon (\mathbb{N}+1)^n \to \mathbb{Z}[\frac{1}{2}]$ to be the function

$$\begin{aligned} F_n(\vec{a}) &:= \sum_{m=1}^{n} 3^{n-m} 2^{-a_{[m,n]}} \\ &= 3^{n-1}2^{-a_{[1,n]}} + 3^{n-2}2^{-a_{[2,n]}} + \cdots + 3^1 2^{-a_{[n-1,n]}} + 2^{-a_n}, \end{aligned} \tag{1.5}$$

where we adopt the summation notation

$$a_{[j,k]} := \sum_{i=j}^{k} a_i \tag{1.6}$$

for any $1 \leq j \leq k \leq n$, thus for instance $|\vec{a}| = a_{[1,n]}$. The n-Syracuse offset map F_n takes values in the ring $\mathbb{Z}[\frac{1}{2}] := \{\frac{M}{2^a} : M \in \mathbb{Z}, a \in \mathbb{N}\}$ formed by adjoining $\frac{1}{2}$ to the integers.

By iterating (1.1) and then using (1.3), we conclude that

$$\mathrm{Syr}^n(N) = \mathrm{Aff}_{\vec{a}^{(n)}(N)}(N) = 3^n 2^{-|\vec{a}^{(n)}(N)|} N + F_n(\vec{a}^{(n)}(N)) \tag{1.7}$$

for any $N \in 2\mathbb{N}+1$ and $n \in \mathbb{N}$, where we define n-*Syracuse valuation* $\vec{a}^{(n)}(N) \in (\mathbb{N}+1)^n$ of N to be the tuple

$$\vec{a}^{(n)}(N) := \left(\nu_2(3N+1), \nu_2(3\mathrm{Syr}(N)+1), \dots, \nu_2(3\mathrm{Syr}^{n-1}(N)+1)\right). \tag{1.8}$$

The identity (1.7) asserts that $\mathrm{Syr}^n(N)$ is the image of N under a certain affine map $\mathrm{Aff}_{\vec{a}^{(n)}(N)}$ that is determined by the n-Syracuse valuation $\vec{a}^{(n)}(N)$ of N. This suggests that in order to understand the behaviour of the iterates $\mathrm{Syr}^n(N)$ of a typical large number N, one needs to understand the behaviour of n-Syracuse valuation $\vec{a}^{(n)}(N)$, as well as the n-Syracuse offset map F_n. For the former, we can gain heuristic insight by observing that for a positive integer a, the set of odd natural numbers $N \in 2\mathbb{N}+1$ with $\nu_2(3N+1) = a$ has (logarithmic) relative density 2^{-a}. To model this probabilistically, we introduce the following probability distribution:

Definition 1.7 (Geometric random variable)**.** If $\mu > 1$, we use $\mathbf{Geom}(\mu)$ to denote a geometric random variable of mean μ, that is to say $\mathbf{Geom}(\mu)$ takes values in $\mathbb{N}+1$ with

$$\mathbb{P}(\mathbf{Geom}(\mu) = a) = \frac{1}{\mu}\left(\frac{\mu-1}{\mu}\right)^{a-1}$$

for all $a \in \mathbb{N}+1$. We use $\mathbf{Geom}(\mu)^n$ to denote a tuple of n independent, identically distributed (or *iid* for short) copies of $\mathbf{Geom}(\mu)$, and use $\mathbf{X} \equiv \mathbf{Y}$ to denote the assertion that two random variables $\mathbf{X}, \mathbf{Y}$ have the same distribution. Thus for instance

$$\mathbb{P}(\mathbf{a} = a) = 2^{-a}$$

whenever $\mathbf{a} \equiv \mathbf{Geom}(2)$ and $a \in \mathbb{N}+1$, and more generally

$$\mathbb{P}(\vec{\mathbf{a}} = \vec{a}) = 2^{-|\vec{a}|}$$

whenever $\vec{\mathbf{a}} \equiv \mathbf{Geom}(2)^n$ and $\vec{a} \in (\mathbb{N}+1)^n$ for some $n \in \mathbb{N}$.

In this paper we will only work with the geometric random variables $\mathbf{Geom}(2)$ and $\mathbf{Geom}(4)$.

We will then be guided by the following heuristic:

Heuristic 1.8 (Valuation heuristic)**.** *If N is a "typical" large odd natural number, and n is much smaller than $\log N$, then the n-Syracuse valuation $\vec{a}^{(n)}(N)$ behaves like $\mathbf{Geom}(2)^n$.*

We can make this heuristic precise as follows. Given two random variables $\mathbf{X}, \mathbf{Y}$ taking values in the same discrete space R, we define the *total variation* $d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y})$ between the two variables to be the total variation of the difference in the probability measures, thus

$$d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y}) := \sum_{r \in R} |\mathbb{P}(\mathbf{X} = r) - \mathbb{P}(\mathbf{Y} = r)|. \tag{1.9}$$

Note that

$$\sup_{E \subset R} |\mathbb{P}(\mathbf{X} \in E) - \mathbb{P}(\mathbf{Y} \in E)| \leq d_{\mathrm{TV}}(\mathbf{X}, \mathbf{Y}) \leq 2 \sup_{E \subset R} |\mathbb{P}(\mathbf{X} \in E) - \mathbb{P}(\mathbf{Y} \in E)|. \tag{1.10}$$

For any finite non-empty set R, let $\mathbf{Unif}(R)$ denote a uniformly distributed random variable on R. Then we have

Proposition 1.9 (Distribution of n-Syracuse valuation)**.** *Let $n \in \mathbb{N}$, and let $\mathbf{N}$ be a random variable taking values in $2\mathbb{N}+1$. Suppose there exist an absolute constant $c_0 > 0$ and some natural number $m \geq (2+c_0)n$ such that $\mathbf{N} \bmod 2^m$ is approximately uniformly distributed in $\mathbb{Z}/2^m\mathbb{Z}$, in the sense that*

$$d_{\mathrm{TV}}(\mathbf{N} \bmod 2^m, \mathbf{Unif}(\mathbb{Z}/2^m\mathbb{Z})) \ll 2^{-m}. \tag{1.11}$$

Then

$$d_{\mathrm{TV}}(\vec{a}^{(n)}(\mathbf{N}), \mathbf{Geom}(2)^n) \ll 2^{-c_1 n} \tag{1.12}$$

for some absolute constant $c_1 > 0$ (depending on c_0). The implied constants in the asymptotic notation are also permitted to depend on c_0.

Informally, this proposition asserts that Heuristic 1.8 is justified whenever N is expected to be uniformly distributed modulo 2^m for some m slightly larger than $2n$. The hypothesis (1.11) is somewhat stronger than what is actually needed for the conclusion (1.12) to hold, but this formulation of the implication will suffice for our applications.

Remark 1.10. Another standard way in the literature to justify Heuristic 1.8 is to consider the Syracuse dynamics on the 2-adic integers $\mathbb{Z}_2 := \varprojlim_m \mathbb{Z}/2^m\mathbb{Z}$, or more precisely on the odd 2-adics $2\mathbb{Z}_2 + 1$. As the 2-valuation ν_2 remains well defined on $\mathbb{Z}_2$, one can extend the Syracuse map Syr to a map on $2\mathbb{Z}_2 + 1$. As is well known (see e.g., [9]), the Haar probability measure on $2\mathbb{Z}_2 + 1$ is preserved by this map, and if $\mathbf{Haar}(2\mathbb{Z}_2 + 1)$ is a random element of $2\mathbb{N} + 1$ drawn using this measure, then it is not difficult (basically using the 2-adic analogue of Lemma 2.1 below) to show that the random variables $\nu_2(3\mathrm{Syr}^j(\mathbf{Haar}(2\mathbb{Z}_2 + 1)) + 1)$ for $j \in \mathbb{N}$ are iid copies of $\mathbf{Geom}(2)$. However, we will not use this 2-adic formalism in this paper.

In practice, the offset $F_n(\vec{a})$ is fairly small (in an Archimedean sense) when n is not too large; indeed, from (1.5) we have

$$0 \leq F_n(\vec{a}) \leq 3^n 2^{-a_n} \leq 3^n \tag{1.13}$$

for any $n \in \mathbb{N}$ and $\vec{a} \in (\mathbb{N}+1)^n$. For large N, we then conclude from (1.7) that we have the heuristic approximation

$$\mathrm{Syr}^n(N) \approx 3^n 2^{-|\vec{a}^{(n)}(N)|} N$$

and hence by Heuristic 1.8 we expect $\mathrm{Syr}^n(N)$ to behave statistically like

$$\mathrm{Syr}^n(N) \approx 3^n 2^{-|\mathbf{Geom}(2)^n|} N = N \exp(n \log 3 - |\mathbf{Geom}(2)^n| \log 2) \tag{1.14}$$

if n is much smaller than $\log N$. One can view the sequence $n \mapsto n \log 3 - |\mathbf{Geom}(2)^n| \log 2$ as a simple random walk on $\mathbb{R}$ with negative drift $\log 3 - 2\log 2 = \log \frac{3}{4}$. From the law of large numbers we expect to have

$$|\mathbf{Geom}(2)^n| \approx 2n \tag{1.15}$$

most of the time, thus we are led to the heuristic prediction

$$\mathrm{Syr}^n(N) \approx (3/4)^n N \tag{1.16}$$

for typical N; indeed, from the central limit theorem or the Chernoff bound we in fact expect the refinement

$$\mathrm{Syr}^n(N) = \exp(O(n^{1/2}))(3/4)^n N \tag{1.17}$$

for "typical" N. In particular, we expect the Syracuse orbit $N, \mathrm{Syr}(N), \mathrm{Syr}^2(N), \dots$ to decay geometrically in time for typical N, which underlies the usual heuristic argument supporting the truth of Conjecture 1.1; see [11], [7] for further discussion. We remark that the multiplicative inaccuracy of $\exp(O(n^{1/2}))$ in (1.17) is the main reason why we work with logarithmic density instead of natural density in this paper (see also [10] for a closely related "Benford's law" phenomenon).

1.3. **Reduction to a stablisation property for first passage locations.** Roughly speaking, Proposition 1.9 lets one obtain good control on the Syracuse iterates $\mathrm{Syr}^n(N)$ for almost all N and for times n up to $c \log N$ for a small absolute constant c; this already can be used in conjunction with a rigorous version of (1.16) or (1.17) to recover the previously mentioned result $\mathrm{Syr}_{\min}(N) \leq N^{1-c}$ for almost all N and some absolute constant $c > 0$; see Section 5 for details. In the language of evolutionary partial differential equations, these type of results can be viewed as analogous to "almost sure local wellposedness" results, in which one has good short-time control on the evolution for almost all choices of initial condition N.

In this analogy, Theorem 1.6 then corresponds to an "almost sure almost global wellposedness" result, where one needs to control the solution for times so large that the evolution gets arbitrary close to the bounded state $N = O(1)$. To bootstrap from almost sure local wellposedness to almost sure almost global wellposedness, we were inspired by the work of Bourgain [3], who demonstrated an almost sure global wellposedness result for a certain nonlinear Schrödinger equation by combining local wellposedness theory with a construction of an invariant probability measure for the dynamics. Roughly speaking, the point was that the invariance of the measure would almost surely keep the solution in a "bounded" region of the state space for arbitrarily long times, allowing one to iterate the local wellposedness theory indefinitely.

In our context, we do not expect to have any useful invariant probability measures for the dynamics due to the geometric decay (1.16) (and indeed Conjecture 1.5 would imply that the only invariant probability measure is the Dirac measure on $\{1\}$). Instead, we can construct a *family* of probability measures ν_x which are *approximately* transported to each other by certain iterations of the Syracuse map (by a variable amount of time). More precisely, given a threshold $x \geq 1$ and an odd natural number $N \in 2\mathbb{N}+1$, define the *first passage time*

$$T_x(N) \coloneqq \inf\{n \in \mathbb{N} : \mathrm{Syr}^n(N) \leq x\},$$

with the convention that $T_x(N) \coloneqq +\infty$ if $\mathrm{Syr}^n(N) > x$ for all n. (Of course, if Conjecture 1.5 were true, this latter possibility could not occur, but we will not be assuming this conjecture in our arguments.) We then define the *first passage location*

$$\mathrm{Pass}_x(N) \coloneqq \mathrm{Syr}^{T_x(N)}(N)$$

with the (somewhat arbitrary and artificial) convention that $\mathrm{Syr}^\infty(N) \coloneqq 1$; thus $\mathrm{Pass}_x(N)$ is the first location of the Syracuse orbit $\mathrm{Syr}^{\mathbb{N}}(N)$ that falls inside $[1,x]$, or 1 if no such location exists. Finally, let $\alpha > 1$ be an absolute constant sufficiently close to 1. The key proposition is then

Proposition 1.11 (Stabilisation of first passage). *Let $\alpha > 1$ be a constant sufficiently close to* 1*. For each sufficiently large y, let $\mathbf{N}_y$ be a random variable with distribution $\mathbf{N}_y \equiv \mathbf{Log}(2\mathbb{N}+1 \cap [y, y^\alpha])$. Then for sufficiently large x, we have the estimates*

$$\mathbb{P}(T_x(\mathbf{N}_y) = +\infty) \ll x^{-c} \tag{1.18}$$

for $y = x^\alpha, x^{\alpha^2}$, and also

$$d_{\mathrm{TV}}(\mathrm{Pass}_x(\mathbf{N}_{x^\alpha}), \mathrm{Pass}_x(\mathbf{N}_{x^{\alpha^2}})) \ll \log^{-c} x \tag{1.19}$$

for some absolute constant $c > 0$.

In Section 3 we shall see how Theorem 1.6 (and hence Theorem 1.3) follows from Proposition 1.11; basically the point is that (1.19), (1.18) imply that the first passage map Pass_x approximately maps the distribution ν_{x^α} of $\mathrm{Pass}_{x^\alpha}(\mathbf{N}_{x^{\alpha^2}})$ to the distribution ν_x of $\mathrm{Pass}_x(\mathbf{N}_{x^\alpha})$, and one can then iterate this to map almost all of the probabilistic mass of (say) $\mathbf{N}_x$ to be arbitrarily close to the bounded state $N = O(1)$. The implication is very general and does not use any particular properties of the Syracuse map beyond (1.18), (1.19). The estimate (1.18) is easy to establish; it is (1.19) that is the most important and difficult conclusion of Proposition 1.11. We remark that the bound of $O(\log^{-c} x)$ in (1.19) is stronger than is needed for this argument; any bound of the form $O((\log\log x)^{-1-c})$ would have sufficed. Conversely, it may be possible to improve the bound in (1.19) further, perhaps all the way to x^{-c}.

1.4. **Fine-scale mixing of Syracuse random variables.**

It remains to establish Proposition 1.11. For α close enough to 1, this proposition falls under the regime of a (refined) "local wellposedness" result, since from the heuristic (1.16) (or (1.17)) we expect the first passage time $T_x(\mathbf{N}_y)$ to be comparable to a small multiple of $\log \mathbf{N}_y$. Inspecting the iteration formula (1.7), the behaviour of the n-Syracuse valuation $\vec{a}^{(n)}(\mathbf{N}_y)$ for such times n is then well understood thanks to Proposition 1.9; the main remaining difficulty is to understand the behaviour of the n-Syracuse offset map $F_n\colon (\mathbb{N}+1)^n \to \mathbb{Z}[\frac{1}{2}]$, and more specifically to analyse the distribution of the random variable $F_n(\mathbf{Geom}(2)^n) \bmod 3^k$ for various n, k, where by abuse of notation we use $x \mapsto x \bmod 3^k$ to denote the unique ring homomorphism from $\mathbb{Z}[\frac{1}{2}]$ to $\mathbb{Z}/3^k\mathbb{Z}$ (which in particular maps $\frac{1}{2}$ to the inverse $\frac{3^k+1}{2} \bmod 3^k$ of $2 \bmod 3^k$). Indeed, from (1.7) one has

$$\mathrm{Syr}^n(N) = F_n(\vec{a}^{(n)}(N)) \bmod 3^k$$

whenever $0 \le k \le n$ and $N \in 2\mathbb{N}+1$. Thus, if $n, \mathbf{N}, m, c_0$ obey the hypotheses of Proposition 1.9, one has

$$d_{\mathrm{TV}}(\mathrm{Syr}^n(\mathbf{N}) \bmod 3^k, F_n(\mathbf{Geom}(2)^n) \bmod 3^k) \ll 2^{-c_1 n}$$

for all $0 \le k \le n$. If we now define the *Syracuse random variables* $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ for $n \in \mathbb{N}$ to be random variables on the cyclic group $\mathbb{Z}/3^n\mathbb{Z}$ with the distribution

$$\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) \equiv F_n(\mathbf{Geom}(2)^n) \bmod 3^n \tag{1.20}$$

then from (1.5) we see that

$$\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) \bmod 3^k \equiv \mathbf{Syrac}(\mathbb{Z}/3^k\mathbb{Z}) \tag{1.21}$$

whenever $k \le n$, and thus

$$d_{\mathrm{TV}}(\mathrm{Syr}^n(\mathbf{N}) \bmod 3^k, \mathbf{Syrac}(\mathbb{Z}/3^k\mathbb{Z})) \ll 2^{-c_1 n}.$$

We thus see that the 3-adic distribution of the Syracuse orbit $\mathrm{Syr}^{\mathbb{N}}(\mathbf{N})$ is controlled (initially, at least) by the random variables $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$. The distribution of these random variables can be computed explicitly for any given n via the following recursive formula:

Lemma 1.12 (Recursive formula for Syracuse random variables)**.** *For any $n \in \mathbb{N}$ and $x \in \mathbb{Z}/3^{n+1}\mathbb{Z}$, one has*

$$\mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^{n+1}\mathbb{Z}) = x) = \frac{\sum_{1 \le a \le 2\times 3^n : 2^a x = 1 \bmod 3} 2^{-a} \mathbb{P}\left(\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = \frac{2^a x - 1}{3}\right)}{1 - 2^{-2\times 3^n}},$$

where $\frac{2^a x-1}{3}$ is viewed as an element of $\mathbb{Z}/3^n\mathbb{Z}$.

Proof. Let $(\mathbf{a}_1, \dots, \mathbf{a}_{n+1}) \equiv \mathbf{Geom}(2)^{n+1}$ be $n+1$ iid copies of $\mathbf{Geom}(2)$. From (1.5) we have

$$F_{n+1}(\mathbf{a}_{n+1}, \dots, \mathbf{a}_1) = \frac{3F_n(\mathbf{a}_{n+1}, \dots, \mathbf{a}_2) + 1}{2^{\mathbf{a}_1}}$$

and thus we have

$$\mathbf{Syrac}(\mathbb{Z}/3^{n+1}\mathbb{Z}) \equiv \frac{3\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) + 1}{2^{\mathbf{Geom}(2)}},$$

where $3\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ is viewed as an element of $\mathbb{Z}/3^{n+1}\mathbb{Z}$, and the random variables $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}), \mathbf{Geom}(2)$ on the right hand side are understood to be independent. We therefore have

$$\mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^{n+1}\mathbb{Z}) = x) = \sum_{a \in \mathbb{N}+1} 2^{-a} \mathbb{P}\left(\frac{3\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) + 1}{2^a} = x \right)$$
$$= \sum_{a \in \mathbb{N}+1: 2^a x = 1 \bmod 3} 2^{-a} \mathbb{P}\left(\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = \frac{2^a x - 1}{3} \right).$$

By Euler's theorem, the quantity $\frac{2^a x-1}{3} \in \mathbb{Z}/3^n\mathbb{Z}$ is periodic in a with period 2×3^n. Splitting a into residue classes modulo 2×3^n and using the geometric series formula, we obtain the claim. ☐

Thus for instance, we trivially have $\mathbf{Syrac}(\mathbb{Z}/3^0\mathbb{Z})$ takes the value $0 \bmod 1$ with probability 1; then by the above lemma, $\mathbf{Syrac}(\mathbb{Z}/3\mathbb{Z})$ takes the values $0, 1, 2 \bmod 3$ with probabilities $0, 1/3, 2/3$ respectively; another application of the above lemma then reveals that $\mathbf{Syrac}(\mathbb{Z}/3^2\mathbb{Z})$ takes the values $0, 1, \dots, 8 \bmod 9$ with probabilities

$$0, \frac{8}{63}, \frac{16}{63}, 0, \frac{11}{63}, \frac{4}{63}, 0, \frac{2}{63}, \frac{22}{63}$$

respectively; and so forth. More generally, one can numerically compute the distribution of $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ exactly for small values of n, although the time and space required to do so increases exponentially with n.

Remark 1.13. One could view the Syracuse random variables $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ as projections

$$\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) \equiv \mathbf{Syrac}(\mathbb{Z}_3) \bmod 3^n$$

of a single random variable $\mathbf{Syrac}(\mathbb{Z}_3)$ taking values in the 3-adics $\mathbb{Z}_3 := \varprojlim_n \mathbb{Z}/3^n\mathbb{Z}$ (equipped with the usual metric $d(x,y) := 3^{-\nu_3(x-y)}$), which can for instance be defined as

$$\mathbf{Syrac}(\mathbb{Z}_3) \equiv \sum_{j=0}^\infty 3^j 2^{-\mathbf{a}_{[1,j+1]}}$$
$$= 2^{-\mathbf{a}_1} + 3^1 2^{-\mathbf{a}_{[1,2]}} + 3^2 2^{-\mathbf{a}_{[1,3]}} + \dots$$

where $\mathbf{a}_1, \mathbf{a}_2, \dots$ are iid copies of $\mathbf{Geom}(2)$; note that this series converges in $\mathbb{Z}_3$. One can view the distribution of $\mathbf{Syrac}(\mathbb{Z}_3)$ as the unique stationary measure for the discrete Markov process[3] on $\mathbb{Z}_3$ that maps each $x \in \mathbb{Z}_3$ to $\frac{3x+1}{2^a}$ for each $a \in \mathbb{N}+1$ with transition probability 2^{-a} (this fact is implicit in the proof of Lemma 1.12). However, we will not explicitly adopt the 3-adic perspective in this paper, preferring to work instead with the finite projections $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ of $\mathbf{Syrac}(\mathbb{Z}_3)$.

While the Syracuse random variables $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ fail to be uniformly distributed on $\mathbb{Z}/3^n\mathbb{Z}$, we can show that they do approach uniform distribution $n \to \infty$ at fine scales (as measured in a 3-adic sense), and this turns out to be the key ingredient needed to establish Proposition 1.11. More precisely, we will show

[3]This Markov process may possibly be related to the 3-adic Markov process for the *inverse* Collatz map studied in [18]. See also a recent investigation of 3-adic irregularities of the Collatz iteration in [17].

Proposition 1.14 (Fine scale mixing of n-Syracuse offsets)**.** *For all* $1 \leq m \leq n$ *one has*

$$\mathrm{Osc}_{m,n}\left(\mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = Y \bmod 3^n)\right)_{Y \in \mathbb{Z}/3^n\mathbb{Z}} \ll_A m^{-A} \tag{1.22}$$

for any fixed $A > 0$, *where the oscillation* $\mathrm{Osc}_{m,n}(c_Y)_{Y\in\mathbb{Z}/3^n\mathbb{Z}}$ *of a tuple of real numbers* $c_Y \in \mathbb{R}$ *indexed by* $\mathbb{Z}/3^n\mathbb{Z}$ *at* 3*-adic scale* 3^{-m} *is defined by*

$$\mathrm{Osc}_{m,n}(c_Y)_{Y\in\mathbb{Z}/3^n\mathbb{Z}} \coloneqq \sum_{Y\in\mathbb{Z}/3^n\mathbb{Z}} \left| c_Y - 3^{m-n} \sum_{Y'\in\mathbb{Z}/3^n\mathbb{Z}: Y'=Y \bmod 3^m} c_{Y'} \right|. \tag{1.23}$$

Informally, the above proposition asserts that the Syracuse random variable $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ is approximately uniformly distributed in "fine-scale" or "high-frequency" cosets $Y + 3^m\mathbb{Z}/3^n\mathbb{Z}$, after conditioning to the event $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = Y \bmod 3^m$. In Section 5, we show how Proposition 1.11 (and hence Theorem 1.3) follows from Proposition 1.14 and Proposition 1.9.

Remark 1.15. One can heuristically justify this mixing property as follows. The geometric random variable $\mathbf{Geom}(2)$ can be computed to have a Shannon entropy of $\log 4$; thus, by asymptotic equipartition, the random variable $\mathbf{Geom}(2)^n$ is expected to behave like a uniform distribution on $4^{n+o(n)}$ separate tuples in $(\mathbb{N}+1)^n$. On the other hand, the range $\mathbb{Z}/3^n\mathbb{Z}$ of the map $\vec{a} \mapsto F_n(\vec{a}) \bmod 3^n$ only has cardinality 3^n. While this map does have substantial irregularities at coarse 3-adic scales (for instance, it always avoids the multiples of 3), it is not expected to exhibit any such irregularity at fine scales, and so if one models this map by a random map from $4^{n+o(n)}$ elements to $\mathbb{Z}/3^n\mathbb{Z}$ one is led to the estimate (1.22) (in fact this argument predicts a stronger bound of $\exp(-cm)$ for some $c > 0$, which we do not attempt to establish here).

Remark 1.16. In order to upgrade logarithmic density to natural density in our results, it seems necessary to strengthen Proposition 1.14 by establishing a suitable fine scale mixing property of the entire random affine map $\mathrm{Aff}_{\mathbf{Geom}(2)^n}$, as opposed to just the offset $F_n(\mathbf{Geom}(2)^n)$. This looks plausibly attainable from the methods in this paper, but we do not pursue this question here.

To prove Proposition 1.14, we use a partial convolution structure present in the n-Syracuse offset map, together with Plancherel's theorem, to reduce matters to establishing a superpolynomial decay bound for the characteristic function (or Fourier coefficients) of a Syracuse random variable $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$. More precisely, in Section 6 we derive Proposition 1.14 from

Proposition 1.17 (Decay of characteristic function)**.** *Let* $n \geq 1$, *and let* $\xi \in \mathbb{Z}/3^n\mathbb{Z}$ *be not divisible by* 3*. Then*

$$\mathbb{E}e^{-2\pi i \xi \mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})/3^n} \ll_A n^{-A}$$

for any fixed $A > 0$.

Remark 1.18. In the converse direction, it is not difficult to use the triangle inequality to establish the inequality

$$|\mathbb{E}e^{-2\pi i \xi \mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})/3^n}| \leq \mathrm{Osc}_{n-1,n}\left(\mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = Y \bmod 3^n)\right)_{Y\in\mathbb{Z}/3^n\mathbb{Z}}$$

whenever ξ is not a multiple of 3 (so in particular the function $x \mapsto e^{-2\pi i \xi x/3^n}$ has mean zero on cosets of $3^{n-1}\mathbb{Z}/3^n\mathbb{Z}$). Thus Proposition 1.17 and Proposition 1.14 are in fact equivalent. One could also equivalently phrase Proposition 1.17 in terms of the decay properties of the characteristic function of $\mathbf{Syrac}(\mathbb{Z}_3)$ (which would be defined on the Pontryagin dual $\hat{\mathbb{Z}}_3 = \mathbb{Q}_3/\mathbb{Z}_3$ of $\mathbb{Z}_3$), but we will not do so here.

The remaining task is to establish Proposition 1.17. This turns out to be the most difficult step in the argument, and is carried out in Section 7. If the Syracuse random variable

$$\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) \equiv 2^{-\mathbf{a}_1} + 3^1 2^{-\mathbf{a}_{[1,2]}} + \cdots + 3^{n-1} 2^{-\mathbf{a}_{[1,n]}} \bmod 3^n \tag{1.24}$$

(with $(\mathbf{a}_1, \ldots, \mathbf{a}_n) \equiv \mathbf{Geom}(2)^n$) was the sum of independent random variables, then the characteristic function of $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ would factor as something like a Riesz product of cosines, and its estimation would be straightforward. Unfortunately, the expression (1.24) does not obviously resolve into such a sum of independent random variables; however, by grouping adjacent terms $3^{2j-2}2^{-\mathbf{a}_{[1,2j-1]}}, 3^{2j-1}2^{-\mathbf{a}_{[1,2j]}}$ in (1.24) into pairs, one can at least obtain a decomposition into the sum of independent expressions once one conditions on the sums $\mathbf{b}_j := \mathbf{a}_{2j-1} + \mathbf{a}_{2j}$ (which are iid copies of a Pascal distribution $\mathbf{Pascal}$). This lets one express the characteristic functions as an *average* of products of cosines (times a phase), where the average is over trajectories of a certain random walk $\mathbf{v}_1, \mathbf{v}_{[1,2]}, \mathbf{v}_{[1,3]}, \ldots$ in $\mathbb{Z}^2$ with increments in the first quadrant that we call a *two-dimensional renewal process*. If we color certain elements of $\mathbb{Z}^2$ "white" when the associated cosines are small, and "black" otherwise, then the problem boils down to ensuring that this renewal process encounters a reasonably large number of white points (see Figure 2 in Section 7).

From some elementary number theory, we will be able to describe the black regions of $\mathbb{Z}^2$ as a union of "triangles" Δ that are well separated from each other; again, see Figure 2. As a consequence, whenever the renewal process passes through a black triangle, it will very likely also pass through at least one white point after it exits the triangle. This argument is adequate so long as the triangles are not too large in size; however, for very large triangles it does not produce a sufficient number of white points along the renewal process. However, it turns out that large triangles tend to be fairly well separated from each other (at least in the neighbourhood of even larger triangles), and this geometric observation allows one to close the argument.

As with Proposition 1.14, it is possible that the bound in Proposition 1.17 could be improved, perhaps to as far as $O(\exp(-cn))$ for some $c > 0$. However, we will not need or pursue such a bound here.

The author is supported by NSF grant DMS-1764034 and by a Simons Investigator Award, and thanks Marek Biskup for useful discussions, and Ben Green, Lior Silberman, Sankeerth Rao, and several anonymous commenters on his blog for corrections and other comments.

2. Notation and preliminaries

We use the asymptotic notation $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote the bound $|X| \leq CY$ for an absolute constant C. We also write $X \asymp Y$ for $X \ll Y \ll X$. We also use $c > 0$ to denote various small constants that are allowed to vary from line to line, or even within the same line. If we need the implied constants to depend on other parameters, we will indicate this by subscripts unless explicitly stated otherwise, thus for instance $X \ll_A Y$ denotes the estimate $|X| \leq C_A Y$ for some C_A depending on A.

If E is a set, we use 1_E to denote its indicator, thus $1_E(n)$ equals 1 when $n \in E$ and 0 otherwise. Similarly, if S is a statement, we define the indicator 1_S to equal 1 when S is true and 0 otherwise, thus for instance $1_E(n) = 1_{n \in E}$.

The following alternate description of the n-Syracuse valuation $\vec{a}^{(n)}(N)$ (variants of which have frequently occurred in the literature on the Collatz conjecture) will be useful.

Lemma 2.1 (Description of n-Syracuse valuation)**.** *Let $N \in 2\mathbb{N}+1$ and $n \in \mathbb{N}$. Then $\vec{a}^{(n)}(N)$ is the unique tuple $\vec{a}$ in $(\mathbb{N}+1)^n$ for which $\mathrm{Aff}_{\vec{a}}(N) \in 2\mathbb{N}+1$.*

Proof. It is clear from (1.7) that $\mathrm{Aff}_{\vec{a}^{(n)}(N)} \in 2\mathbb{N}+1$. It remains to prove uniqueness. The claim is easy for $n = 0$, so suppose inductively that $n \geq 1$ and that uniqueness has already been established for $n-1$. Suppose that we have found a tuple $\vec{a} \in (\mathbb{N}+1)^n$ for which $\mathrm{Aff}_{\vec{a}}(N)$ is an odd integer. Then

$$\mathrm{Aff}_{\vec{a}}(N) = \mathrm{Aff}_{a_n}(\mathrm{Aff}_{a_1,\dots,a_{n-1}}(N)) = \frac{3\mathrm{Aff}_{a_1,\dots,a_{n-1}}(N)+1}{2^{a_n}}$$

and thus

$$2^{a_n}\mathrm{Aff}_{\vec{a}}(N) = 3\mathrm{Aff}_{a_1,\dots,a_{n-1}}(N) + 1. \tag{2.1}$$

This implies that $3\mathrm{Aff}_{a_1,\dots,a_{n-1}}(N)$ is an odd natural number. But from (1.3), $\mathrm{Aff}_{a_1,\dots,a_{n-1}}(N)$ also lies in $\mathbb{Z}[\frac{1}{2}]$. The only way these claims can both be true is if $\mathrm{Aff}_{a_1,\dots,a_{n-1}}(N)$ is also an odd natural number, and then by induction $(a_1, \dots, a_{n-1}) = \vec{a}^{(n-1)}(N)$, which by (1.7) implies that

$$\mathrm{Aff}_{a_1,\dots,a_{n-1}}(N) = \mathrm{Syr}^{n-1}(N).$$

Inserting this into (2.1) and using the fact that $\mathrm{Aff}_{\vec{a}}(N)$ is odd, we obtain

$$a_n = \nu_2(3\mathrm{Syr}^{N-1}(N) + 1)$$

and hence by (1.8) we have $\vec{a} = \vec{a}^{(n)}$ as required. □

We record the following concentration of measure bound of Chernoff type, which also bears some resemblance to a local limit theorem. We introduce the Gaussian-type weights

$$G_n(x) := \exp(-|x|^2/n) + \exp(-|x|)$$

for any $n \geq 0$ and $x \in \mathbb{R}^d$ for some $d \geq 1$, where we adopt the convention that $\exp(-\infty) = 0$ (so that $G_0(x) = \exp(-|x|)$). Thus $G_n(x)$ is comparable to 1 for

$x = O(n^{1/2})$, decays in a gaussian fashion in the regime $n^{1/2} \leq |x| \leq n$, and decays exponentially for $|x| \geq n$.

Lemma 2.2 (Chernoff type bound). *Let $d \in \mathbb{N}+1$, and let $\mathbf{v}$ be a random variable taking values in $\mathbb{Z}^d$ obeying the exponential tail condition*

$$\mathbb{P}(|\mathbf{v}| \geq \lambda) \ll \exp(-c_0\lambda) \tag{2.2}$$

for all $\lambda \geq 0$ and some $c_0 > 0$. Assume the non-degeneracy condition that $\mathbf{v}$ is not almost surely concentrated on any coset of any proper subgroup of $\mathbb{Z}^d$. Let $\vec{\mu} := \mathbb{E}\mathbf{v} \in \mathbb{R}^d$ and denote the mean of $\mathbf{v}$. In this lemma all implied constants, as well as the constant c, can depend on d, c_0, and the distribution of $\mathbf{v}$. Let $n \in \mathbb{N}$, and let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be n iid copies of $\mathbf{v}$. Following (1.6), *we write $\mathbf{v}_{[1,n]} := \mathbf{v}_1 + \cdots + \mathbf{v}_n$.*

(i) *For any $\vec{L} \in \mathbb{Z}^d$, one has*

$$\mathbb{P}(\mathbf{v}_{[1,n]} = \vec{L}) \ll \frac{1}{(n+1)^{d/2}} G_n(c(\vec{L} - n\vec{\mu})).$$

(ii) *For any $\lambda \geq 0$, one has*

$$\mathbb{P}(|\mathbf{v}_{[1,n]} - n\vec{\mu}| \geq \lambda) \ll G_n(c\lambda).$$

Thus, for instance for any $n \in \mathbb{N}$, we have

$$\mathbb{P}(|\mathbf{Geom}(2)^n| = L) \ll \frac{1}{\sqrt{n+1}} G_n(c(L-2n))$$

for every $L \in \mathbb{Z}$, and

$$\mathbb{P}\left(||\mathbf{Geom}(2)^n| - 2n| \geq \lambda\right) \ll G_n(c\lambda).$$

for any $\lambda \geq 0$.

Proof. We use the Fourier-analytic (and complex-analytic) method. We may assume that n is positive, since the claim is trivial for $n = 0$. We begin with (i). Let S denote the complex strip $S := \{z \in \mathbb{C} : |\mathrm{Re}(z)| < c_0\}$, then we can define the (complexified) moment generating function $M \colon S^d \to \mathbb{C}$ by the formula

$$M(z_1, \dots, z_d) := \mathbb{E}\exp((z_1, \dots, z_d) \cdot \mathbf{v}),$$

where $\cdot$ is the usual bilinear dot product. From (2.2) and Morera's theorem one verifies that this is a well-defined holomorphic function of d complex variables on S^d, which is periodic with respect to the lattice $(2\pi i\mathbb{Z})^d$. By Fourier inversion, we have

$$\mathbb{P}(\mathbf{v}_{[1,n]} = \vec{L}) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} M\left(i\vec{t}\right)^n \exp\left(-i\vec{t} \cdot \vec{L}\right)\, d\vec{t}.$$

By contour shifting, we then have

$$\mathbb{P}(\mathbf{v}_{[1,n]} = \vec{L}) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} M\left(i\vec{t} + \vec{\lambda}\right)^n \exp\left(-(i\vec{t} + \lambda) \cdot \vec{L}\right)\, d\vec{t}$$

whenever $\vec{\lambda} = (\lambda_1, \dots, \lambda_d) \in (-c_0, c_0)^d$. By the triangle inequality, we thus have

$$\mathbb{P}(\mathbf{v}_{[1,n]} = \vec{L}) \ll \int_{[-\pi,\pi]^d} \left|M\left(i\vec{t} + \vec{\lambda}\right)\right|^n \exp\left(-\vec{\lambda} \cdot \vec{L}\right)\, d\vec{t}.$$

From Taylor expansion and the non-degeneracy condition we have

$$M(\vec{z}) = \exp\left(\vec{z}\cdot\vec{\mu} + \frac{1}{2}\Sigma(\vec{z}) + O(|\vec{z}|^3)\right)$$

for all $\vec{z} \in S^d$ sufficiently close to 0, where Σ is a positive definite quadratic form (the covariance matrix of $\mathbf{v}$). From the non-degeneracy condition we also see that $|M(i\vec{t})| < 1$ whenever $\vec{t} \in [-\pi,\pi]^d$ is not identically zero, hence by continuity $|M(i\vec{t}+\vec{\lambda})| \leq 1-c$ whenever $\vec{t} \in [-\pi,\pi]^d$ is bounded away from zero and $\vec{\lambda}$ is sufficiently small. This implies the estimates

$$|M(i\vec{t}+\vec{\lambda})| \leq \exp\left(\vec{\lambda}\cdot\vec{\mu} - c|\vec{t}|^2 + O(|\vec{\lambda}|^2)\right)$$

for all $\vec{t} \in [-\pi,\pi]^d$ and all sufficiently small $\vec{\lambda} \in \mathbb{R}^d$. Thus we have

$$\begin{aligned}\mathbb{P}(\mathbf{v}_{[1,n]} = \vec{L}) &\ll \int_{[-\pi,\pi]^d} \exp\left(-\vec{\lambda}\cdot(\vec{L}-n\vec{\mu}) - cn|\vec{t}|^2 + O(n|\vec{\lambda}|^2)\right)\, d\vec{t} \\ &\ll n^{-1/2}\exp\left(-\vec{\lambda}\cdot(\vec{L}-n\vec{\mu}) + O(n|\vec{\lambda}|^2)\right).\end{aligned}$$

If $|\vec{L}-n\vec{\mu}| \leq n$, we can set $\vec{\lambda} := c(\vec{L}-n\vec{\mu})/n$ for a sufficiently small c and obtain the claim; otherwise if $|\vec{L}-n\vec{\mu}| > n$ we set $\vec{\lambda} := c(\vec{L}-n\vec{\mu})/|\vec{L}-n\vec{\mu}|$ for a sufficiently small c and again obtain the claim. This gives (i), and the claim (ii) then follows from summing in $\vec{L}$ and applying the integral test. □

Remark 2.3. Informally, the above lemma asserts that as a crude first approximation we have

$$\mathbf{v}_{[1,n]} \approx n\vec{\mu} + \mathbf{Unif}(\{k \in \mathbb{Z}^d : k = O(\sqrt{n})\}), \tag{2.3}$$

and in particular

$$|\mathbf{Geom}(2)^n| \approx \mathbf{Unif}(\mathbb{Z} \cap [2n - O(\sqrt{n}), 2n + O(\sqrt{n})]), \tag{2.4}$$

thus refining (1.15). The reader may wish to use this heuristic for subsequent arguments (for instance, in heuristically justifying (1.17)).

3. Reduction to stabilisation of first passage

In this section we show how Theorem 1.6 follows from Proposition 1.11.

For any threshold N_0, let $E_{N_0} \subset 2\mathbb{N}+1$ denote the set

$$E_{N_0} := \{N \in 2\mathbb{N}+1 : \mathrm{Syr}_{\min}(N) \leq N_0\}$$

of starting positions N of Syracuse orbits that reach N_0 or below. Let $\alpha > 1$ be a constant sufficiently close to one, let x be a sufficiently large quantity, and let $\mathbf{N}_y$ be the random variables from Proposition 1.11. Let B_x denote the event that $T_x(\mathbf{N}_{x^\alpha}) < +\infty$ and $\mathrm{Pass}_x(\mathbf{N}_{x^\alpha}) \in E_{N_0}$.

Observe that if $T_x(\mathbf{N}_{x^{\alpha^2}}) < +\infty$ and $\mathrm{Pass}_x(\mathbf{N}_{x^{\alpha^2}}) \in E_{N_0}$, then

$$T_{x^\alpha}(\mathbf{N}_{x^{\alpha^2}}) \leq T_x(\mathbf{N}_{x^{\alpha^2}}) < +\infty$$

and

$$\mathrm{Syr}^{\mathbb{N}}(\mathrm{Pass}_x(\mathbf{N}_{x^{\alpha^2}})) \subset \mathrm{Syr}^{\mathbb{N}}(\mathrm{Pass}_{x^\alpha}(\mathbf{N}_{x^{\alpha^2}}))$$

which implies that

$$\mathrm{Syr}_{\min}(\mathrm{Pass}_{x^\alpha}(\mathbf{N}_{x^{\alpha^2}})) \leq \mathrm{Syr}_{\min}(\mathrm{Pass}_x(\mathbf{N}_{x^{\alpha^2}})) \leq N_0.$$

In particular, the event B_{x^α} holds in this case. From this, (1.18), and (1.19), (1.10) we have

$$\begin{aligned}\mathbb{P}(B_{x^\alpha}) &\geq \mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_{x^{\alpha^2}}) \in E_{N_0} \wedge T_x(\mathbf{N}_{x^{\alpha^2}}) < +\infty) \\ &\geq \mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_{x^{\alpha^2}}) \in E_{N_0}) - O(x^{-c}) \\ &\geq \mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_{x^\alpha}) \in E_{N_0}) - O(\log^{-c} x) \\ &\geq \mathbb{P}(B_x) - O(\log^{-c} x)\end{aligned}$$

for all sufficiently large x.

Let J be the first natural number such that the quantity $y := x^{\alpha^{-J}}$ is less than $N_0^{1/\alpha}$. If N_0 is large enough, we then have (by replacing x with $y^{\alpha^{j-2}}$ in the preceding estimate) that

$$\mathbb{P}(B_{y^{\alpha^{j-1}}}) \geq \mathbb{P}(B_{y^{\alpha^{j-2}}}) - O((\alpha^j \log y)^{-c})$$

for all $j = 1, \dots, J$. The event $B_{y^{\alpha^{-1}}}$ occurs with probability $1 - O(y^{-c})$, thanks to (1.18) and the fact that $\mathbf{N}_y \leq y^\alpha \leq N_0$. Summing the telescoping series, we conclude that

$$\mathbb{P}(B_{y^{\alpha^{J-1}}}) \geq 1 - O(\log^{-c} y)$$

where we allow implied constants to depend on α (note that the $O(y^{-c})$ error can be absorbed into the $O(\log^{-c} y)$ factor. By construction, $y \geq N_0^{1/\alpha^2}$ and $y^{\alpha^J} = x$, so

$$\mathbb{P}(B_{x^{1/\alpha}}) \geq 1 - O(\log^{-c} N_0).$$

If $B_{x^{1/\alpha}}$ holds, then $\mathrm{Pass}_{x^{1/\alpha}}(\mathbf{N}_x)$ lies in the Syracuse orbit $\mathrm{Syr}^{\mathbb{N}}(\mathbf{N}_x)$, and thus $\mathrm{Syr}_{\min}(\mathbf{N}_x) \leq \mathrm{Syr}_{\min}(\mathrm{Pass}_{x^{1/\alpha}}(\mathbf{N}_x)) \leq N_0$. We conclude that for sufficiently large x, we have

$$\mathbb{P}(\mathrm{Syr}_{\min}(\mathbf{N}_x) > N_0) \ll \log^{-c} N_0.$$

By definition of $\mathbf{N}_x$ (and using the integral test to sum the harmonic series $\sum_{N \in 2\mathbb{N}+1 \cap [x, x^\alpha]} \frac{1}{N}$), we conclude that

$$\sum_{N \in 2\mathbb{N}+1 \cap [x, x^\alpha] : \mathrm{Syr}_{\min}(N) > N_0} \frac{1}{N} \ll \frac{1}{\log^c N_0} \log x$$

for all sufficiently large x.

Now let $f\colon 2\mathbb{N}+1 \to [0, +\infty)$ be such that $\lim_{N \to \infty} f(N) = +\infty$. Set $\tilde{f}(x) := \inf_{N \in 2\mathbb{N}+1 : N \geq x} f(N)$, then $\tilde{f}(x) \to \infty$ as $x \to \infty$. Applying the preceding estimate with $N_0 := \tilde{f}(x)$, we conclude that

$$\sum_{N \in 2\mathbb{N}+1 \cap [x, x^\alpha] : \mathrm{Syr}_{\min}(N) > f(N)} \frac{1}{N} \ll \frac{1}{\log^c \tilde{f}(x)} \log x$$

for all sufficiently large x. Since $\frac{1}{\log^c \tilde{f}(x)}$ goes to zero as $x \to \infty$, we conclude from telescoping series that the set $\{N \in 2\mathbb{N}+1 : \mathrm{Syr}_{\min}(N) > f(N)\}$ has zero logarithmic density, and Theorem 1.6 follows.

Remark 3.1. This argument also shows that for any large N_0, one has $\mathrm{Syr}_{\min}(N) \leq N_0$ for all N in a set of odd natural numbers of logarithmic density $\frac{1}{2} - O(\log^{-c} N_0)$; one can then show that one also has $\mathrm{Col}_{\min}(N) \leq N_0$ for all N in a set of positive natural numbers of logarithmic density $1 - O(\log^{-c} N_0)$. We leave the details to the interested reader.

4. 3-adic distribution of iterates

In this section we establish Proposition 1.9. Let $n, \mathbf{N}, c_0, m$ be as in that proposition. In this section we allow implied constants in the asymptotic notation, as well as the constants $c > 0$, to depend on c_0.

We first need a tail bound on the size of the n-Syracuse valuation $\vec{a}^{(n)}(\mathbf{N})$:

Lemma 4.1 (Tail bound)**.** *We have*

$$\mathbb{P}(|\vec{a}^{(n)}(\mathbf{N})| \geq m) \ll 2^{-cn}.$$

Proof. Write $\vec{a}^{(n)}(\mathbf{N}) = (\mathbf{a}_1, \dots, \mathbf{a}_n)$, then we may split

$$\mathbb{P}(|\vec{a}^{(n)}(\mathbf{N})| \geq m) = \sum_{k=0}^{n-1} \mathbb{P}(\mathbf{a}_{[1,k]} < m \leq \mathbf{a}_{[1,k+1]})$$

(using the summation convention (1.6)) and so it suffices to show that

$$\mathbb{P}(\mathbf{a}_{[1,k]} < m \leq \mathbf{a}_{[1,k+1]}) \ll 2^{-cn}$$

for each $0 \leq k \leq n-1$.

From Lemma 2.1 and (1.3) we see that

$$3^{k+1} 2^{-\mathbf{a}_{[1,k+1]}} \mathbf{N} + \sum_{i=1}^{k+1} 3^{k+1-i} 2^{-\mathbf{a}_{[i,k+1]}}$$

is an odd integer, and thus

$$3^{k+1} \mathbf{N} + \sum_{i=1}^{k+1} 3^{k+1-i} 2^{\mathbf{a}_{[1,i-1]}}$$

is a multiple of $2^{\mathbf{a}_{[1,k+1]}}$. In particular, when the event $\mathbf{a}_{[1,k]} < m \leq \mathbf{a}_{[1,k+1]}$ holds, one has

$$3^{k+1} \mathbf{N} + \sum_{i=1}^{k+1} 3^{k+1-i} 2^{\mathbf{a}_{[1,i-1]}} = 0 \bmod 2^m.$$

Thus, if one conditions to the event $\mathbf{a}_j = a_j, j = 1, \dots, k$ for some positive integers $a_1, \dots, a_k$, then $\mathbf{N}$ is constrained to a single residue class $b \bmod 2^m$ depending on $a_1, \dots, a_k$ (because 3^{k+1} is invertible in the ring $\mathbb{Z}/2^m\mathbb{Z}$). From (1.11), (1.9) we have the quite crude estimate

$$\mathbb{P}(\mathbf{N} = b \bmod 2^m) \ll 2^{-m}$$

and hence

$$\mathbb{P}(\mathbf{a}_{[1,k]} \leq m < \mathbf{a}_{[1,k+1]}) \ll \sum_{a_1,\dots,a_k \in \mathbb{N}+1: a_{[1,k]} < m} 2^{-m}.$$

The tuples $(a_1, \dots, a_k)$ in the above sum are in one-to-one correspondence with the k-element subsets $\{a_1, a_{[1,2]}, \dots, a_{[1,k]}\}$ of $\{1, \dots, m-1\}$, and hence have cardinality $\binom{m-1}{k}$, thus

$$\mathbb{P}(\mathbf{a}_{[1,k]} < m \leq \mathbf{a}_{[1,k+1]}) \ll 2^{-m}\binom{m-1}{k}.$$

Since $k \leq n-1$ and $m \geq (2+c_0)n$, the right-hand side is $O(2^{-cn})$ by Stirling's formula (one can also use the Chernoff inequality for the sum of $m-1$ Bernoulli random variables $\mathbf{Ber}(\frac{1}{2})$, or Lemma 2.2). The claim follows. $\square$

From Lemma 2.2 we also have

$$\mathbb{P}(|\mathbf{Geom}(2)^n| \geq m) \ll 2^{-cn}.$$

From (1.9) and the triangle inequality we therefore have

$$d_{\mathrm{TV}}(\vec{a}^{(n)}(\mathbf{N}), \mathbf{Geom}(2)^n) = \sum_{\vec{a} \in (\mathbb{N}+1)^n : |\vec{a}| < m} |\mathbb{P}(\vec{a}^{(n)}(\mathbf{N}) = \vec{a}) - \mathbb{P}(\mathbf{Geom}(2)^n = \vec{a})| + O(2^{-cn}).$$

From Definition 1.7 we have

$$\mathbb{P}(\mathbf{Geom}(2)^n = \vec{a}) = 2^{-|\vec{a}|}$$

so it remains to show that

$$\sum_{\vec{a} \in (\mathbb{N}+1)^n : |\vec{a}| < m} |\mathbb{P}(\vec{a}^{(n)}(\mathbf{N}) = \vec{a}) - 2^{-|\vec{a}|}| \ll 2^{-cn}. \tag{4.1}$$

By Lemma 2.1, the event $\vec{a}^{(n)}(\mathbf{N}) = \vec{a}$ occurs precisely when $\mathrm{Aff}_{\vec{a}}(\mathbf{N})$ is an odd integer, which by (1.3) we may write (for $\vec{a} = (a_1, \dots, a_n)$) as

$$3^n 2^{-a_{[1,n]}}\mathbf{N} + 3^{n-1}2^{-a_{[1,n]}} + 3^{n-2}2^{-a_{[2,n]}} + \dots + 2^{-a_n} \in 2\mathbb{N}+1.$$

Equivalently one has

$$3^n\mathbf{N} = -3^{n-1} - 3^{n-2}2^{a_1} - \dots - 2^{a_{[1,n-1]}} + 2^{|\vec{a}|} \bmod 2^{|\vec{a}|+1}.$$

This constrains $\mathbf{N}$ to a single odd residue class modulo $2^{|\vec{a}|+1}$. For $|\vec{a}| < m$, the probability of falling in this class can be computed using (1.11), (1.9) as $2^{-|\vec{a}|} + O(2^{-m})$. The left-hand side of (4.1) is then bounded by

$$\ll 2^{-m}\#\{\vec{a} \in (\mathbb{N}+1)^n : |\vec{a}| < m\} = 2^{-m}\binom{m-1}{n}.$$

The claim now follows from Stirling's formula (or Chernoff's inequality), as in the proof of Lemma 4.1. This completes the proof of Proposition 1.9.

5. Reduction to fine scale mixing of the n-Syracuse offset map

We are now ready to derive Proposition 1.11 (and thus Theorem 1.3) assuming Proposition 1.14. Let x be sufficiently large.

We begin with the proof of (1.18). Let y be one of x^α or x^{α^2}. Let n denote the natural number

$$n_0 := \left\lfloor \frac{\log x}{10 \log 2} \right\rfloor, \tag{5.1}$$

so that $2^{n_0} \asymp x^{0.1}$. Since $\mathbf{N}_y \equiv \mathbf{Log}(2\mathbb{N}+1 \cap [y, y^\alpha])$, a routine application of the integral test reveals that

$$d_{\mathrm{TV}}(\mathbf{N}_y \bmod 2^{3n_0}, \mathbf{Unif}(\mathbb{Z}/2^{3n_0}\mathbb{Z})) \ll 2^{-3n_0}$$

(with plenty of room to spare), hence by Proposition 1.9

$$d_{\mathrm{TV}}(\vec{a}^{(n_0)}(\mathbf{N}_y), \mathbf{Geom}(2)^{n_0}) \ll 2^{-cn_0}. \tag{5.2}$$

In particular, by (1.10) and Lemma 2.2 we have

$$\mathbb{P}(|\vec{a}^{(n_0)}(\mathbf{N}_y)| \leq 1.9n_0) \leq \mathbb{P}(|\mathbf{Geom}(2)^n| \leq 1.9n_0) + O(2^{-cn_0}) \ll 2^{-cn_0} \ll x^{-c} \tag{5.3}$$

(recall we allow c to vary even within the same line). On the other hand, from (1.7), (1.5) we have

$$\mathrm{Syr}^{n_0}(\mathbf{N}_y) \leq 3^{n_0} 2^{-|\vec{a}^{(n_0)}(\mathbf{N}_y)|} \mathbf{N}_y + O(3^{n_0}) \leq 3^{n_0} 2^{-|\vec{a}^{(n_0)}(\mathbf{N}_y)|} x^{\alpha^3} + O(3^{n_0})$$

and hence if $|\vec{a}^{(n_0)}(\mathbf{N}_y)| > 1.9n$ then

$$\mathrm{Syr}^{n_0}(\mathbf{N}_y) \ll 3^{n_0} 2^{-1.9n_0} x^{\alpha^3} + O(3^{n_0}).$$

From (5.1) and a brief calculation, we then conclude (for α sufficiently close to 1, and x large enough) that

$$\mathrm{Syr}^{n_0}(\mathbf{N}_y) \leq x,$$

and hence $T_x(\mathbf{N}_y) \leq n_0 < +\infty$ whenever $|\vec{a}^{(n_0)}(\mathbf{N}_y)| > 1.9n_0$. The claim (1.18) now follows from (5.3).

Remark 5.1. This argument already establishes that $\mathrm{Syr}_{\min}(N) \leq N^\theta$ for almost all N for any $\theta > 1/\alpha$. It also shows that most odd numbers do not lie in a periodic Syracuse orbit, or more precisely that

$$\mathbb{P}(\mathrm{Syr}^n(\mathbf{N}_y) = \mathbf{N}_y \text{ for some } n \in \mathbb{N}+1) \ll x^{-c}.$$

Indeed, the above arguments show that outside of an event of probabliity x^{-c}, one has $\mathrm{Syr}^m(\mathbf{N}_y) \leq x$ for some $m \leq n_0$, which we can assume to be minimal amongst all such m. If $\mathrm{Syr}^n(\mathbf{N}_y) = \mathbf{N}_y$ for some n, we then have $\mathbf{N}_y = \mathrm{Syr}^{n(M)-m}(M)$ for $M := \mathrm{Syr}^m(\mathbf{N}_y) \in [1, x]$ that generates a periodic Syracuse orbit with period $n(M)$. The number of possible pairs (M, m) obtained in this fashion is $O(xn_0)$, and hence this latter possibility only holds for at most $O(xn_0)$ possible values of $\mathbf{N}_y$; as this is much smaller than y, we thus see that this event is only attained with probability $O(x^{-c})$, giving the claim. It is then a routine matter to then deduce that almost all positive integers do not lie in a periodic Collatz orbit; we leave the details to the interested reader.

Now we establish (1.19). By (1.10), it suffices to show that for $y = x^\alpha, x^{\alpha^2}$ and $E \subset 2\mathbb{N}+1 \cap [1,x]$, that

$$\mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_y) \in E) = \left(1 + O(x^{-c})\right) Q + O(\log^{-c} x) \tag{5.4}$$

for some quantity Q that can depend on x, α, E but is independent of the choice of y (note that this bound automatically forces $Q = O(1)$ when x is large, so the first error term on the right-hand side may be absorbed into the second). The strategy is to manipulate the left-hand side of (5.4) into an expression that involves the Syracuse random variables $\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$ for various n (in a range depending on y) plus a small error, and then appeal to Proposition 1.14 to remove the dependence on n and hence on y in the main term.

We turn to the details. Fix E, y, and write $\vec{a}^{(n_0)}(\mathbf{N}_y) = (\mathbf{a}_1, \dots, \mathbf{a}_{n_0})$. From (5.2), (1.10), and Lemma 2.2 we see that for every $0 \le n \le n_0$, one has

$$\mathbb{P}(|\mathbf{a}_{[1,n]} - 2n| \geq \log^{0.6} x) \ll \exp(-c \log^{0.2} x).$$

Hence if we let $\mathcal{A}^{(n_0)} \subset (\mathbb{N}+1)^{n_0}$ denote the set of all tuples $(a_1, \dots, a_{n_0}) \in (\mathbb{N}+1)^{n_0}$ such that

$$|a_{[1,n]} - 2n| < \log^{0.6} x \tag{5.5}$$

for all $0 \le n \le n_0$, then we have from the union bound that

$$\mathbb{P}(\vec{a}^{(n_0)}(\mathbf{N}_y) \not\in \mathcal{A}^{(n_0)}) \ll \log^{-c} x \tag{5.6}$$

(with some room to spare); this can be viewed as a rigorous analogue of the heuristic (2.4). Hence

$$\mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_y) \in E) = \mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_y) \in E \wedge \vec{a}^{(n_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n_0)}) + O(\log^{-c} x).$$

Suppose that $\vec{a}^{(n_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n_0)}$. For any $0 \le n \le n_0$, we have from (1.7), (1.13) that

$$\mathrm{Syr}^n(\mathbf{N}_y) = 3^n 2^{-\mathbf{a}_{[1,n]}} \mathbf{N}_y + O(3^{n_0})$$

and hence by (5.5), (5.1) and some calculation

$$\mathrm{Syr}^n(\mathbf{N}_y) = (1 + O(x^{-0.1})) 3^n 2^{-\mathbf{a}_{[1,n]}} \mathbf{N}_y. \tag{5.7}$$

In particular, from (5.5) one has

$$\mathrm{Syr}^n(\mathbf{N}_y) = \exp(O(\log^{0.6} x)) (3/4)^n \mathbf{N}_y \tag{5.8}$$

which can be viewed as a rigorous version of the heuristic (1.17). As $T_x(\mathbf{N}_y)$ is the first time n for which $\mathrm{Syr}^n(\mathbf{N}_y) \le x$, this gives an approximation

$$T_x(\mathbf{N}_y) = \frac{\log(\mathbf{N}_y/x)}{\log\frac{4}{3}} + O(\log^{0.6} x) \tag{5.9}$$

(note that the right-hand side automatically lies between 0 and n_0 if α is close enough to 1 and x is large enough). In particular, if we introduce the interval

$$I_y := \left[\frac{\log(y/x)}{\log\frac{4}{3}} + \log^{0.8} x, \frac{\log(y^\alpha/x)}{\log\frac{4}{3}} - \log^{0.8} x \right] \tag{5.10}$$

then a straightforward calculation using the integral test shows that

$$\mathbb{P}(T_x(\mathbf{N}_y) \in I_y) = 1 - O(\log^{-c} x).$$

Conversely, suppose that $n \in I_y$. We introduce the natural number

$$m_0 := \left\lfloor \frac{\alpha - 1}{100} \log x \right\rfloor. \tag{5.11}$$

If $T_x(\mathbf{N}_y) = n$, then certainly $T_x(\mathrm{Syr}^{n-m_0}(\mathbf{N}_y)) = m_0$. Conversely, if $T_x(\mathrm{Syr}^{n-m_0}(\mathbf{N}_y)) = m_0$ and $\vec{a}^{(n_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n_0)}$, we have $\mathrm{Syr}^n(\mathbf{N}_y) \leq x < \mathrm{Syr}^{n-1}(\mathbf{N}_y)$, which by (5.8) forces

$$n = \frac{\log(\mathbf{N}_y/x)}{\log \frac{4}{3}} + O(\log^{0.6} x),$$

which by (5.9) implies that $T_x(\mathbf{N}_y) \geq n - m_0$, and hence

$$T_x(\mathbf{N}_y) = n - m_0 + T_x(\mathrm{Syr}^{n-m_0}(\mathbf{N}_y)) = n.$$

We conclude that the event

$$T_x(\mathbf{N}_y) = n \wedge \mathrm{Pass}_x(\mathbf{N}_y) \in E \wedge \vec{a}^{(n_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n_0)}$$

holds precisely when the event

$$B_n := T_x(\mathrm{Syr}^{n-m_0}(\mathbf{N}_y)) = m_0 \wedge \mathrm{Pass}_x(\mathrm{Syr}^{n-m_0}(\mathbf{N}_y)) \in E \wedge \vec{a}^{(n_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n_0)}$$

does. We therefore have

$$\mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_y) \in E) = \sum_{n \in I_y} \mathbb{P}(B_n) + O(\log^{-c} x).$$

Note from (5.8), (5.9) that when B_n occurs, one has

$$\mathrm{Syr}^{n-m_0}(\mathbf{N}_y) = \exp(O(\log^{0.6} x))(3/4)^{T_x(\mathbf{N}_y)-m_0}\mathbf{N}_y = \exp(O(\log^{0.6} x))(4/3)^{m_0}x.$$

We therefore have

$$\mathbb{P}(B_n) = \mathbb{P}(\mathrm{Syr}^{n-m_0}(\mathbf{N}_y) \in E' \wedge \vec{a}^{(n_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n_0)}),$$

for any $n \in I_y$, where E' is the set of odd natural numbers $M \in 2\mathbb{N}+1$ such that $T_x(M) = m_0$ and $\mathrm{Pass}_x(M) \in E$ with

$$\exp(-\log^{0.7} x)(4/3)^{m_0}x \leq M \leq \exp(\log^{0.7} x)(4/3)^{m_0}x. \tag{5.12}$$

The key point here is that E' does not depend on y. Since the event $\vec{a}^{(n_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n_0)}$ is contained in the event $\vec{a}^{(n-m_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n-m_0)}$, we conclude from (5.6) that

$$\mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_y) \in E) = \sum_{n \in I_y} \mathbb{P}(\mathrm{Syr}^{n-m_0}(\mathbf{N}_y) \in E' \wedge \vec{a}^{(n-m_0)}(\mathbf{N}_y) \in \mathcal{A}^{(n-m_0)}) + O(\log^{-c} x).$$

Suppose that $\vec{a} = (a_1, \dots, a_{n-m})$ is a tuple in $\mathcal{A}^{(n-m)}$, and $M \in E'$. From Lemma 2.1, we see that the event $\mathrm{Syr}^{n-m_0}(\mathbf{N}_y) = M \wedge \vec{a}^{(n-m_0)}(\mathbf{N}_y) = \vec{a}$ holds if and only if $\mathrm{Aff}_{\vec{a}}(\mathbf{N}_y) \in E'$, thus

$$\mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_y) \in E) = \sum_{n \in I_y} \sum_{\vec{a} \in \mathcal{A}^{(n-m_0)}} \sum_{M \in E'} \mathbb{P}(\mathrm{Aff}_{\vec{a}}(\mathbf{N}_y) = M) + O(\log^{-c} x).$$

Let $n, \vec{a}, M$ be as in the above sum. Then by (1.3), the event $\mathrm{Aff}_{\vec{a}}(\mathbf{N}_y) = M$ is only non-empty when

$$M = F_{n-m_0}(\vec{a}) \bmod 3^{n-m_0} \tag{5.13}$$

Conversely, if (5.13) holds, then $\mathrm{Aff}_{\vec{a}}(\mathbf{N}_y) = M$ holds precisely when

$$\mathbf{N}_y = 2^{|\vec{a}|} \frac{M - F_{n-m_0}(\vec{a})}{3^{n-m_0}}. \tag{5.14}$$

Note from (5.5), (1.13) that the right-hand side is equal to

$$2^{2(n-m_0)+O(\log^{0.6} x)} \frac{M + O(3^{n-m_0})}{3^{n-m_0}}$$

which by (5.12), (5.1) simplifies to

$$\exp(O(\log^{0.7} x))(4/3)^n.$$

Since $n \in I_y$, we conclude from (5.10) that the right-hand side of (5.14) lies in $[y, y^\alpha]$; from (5.13), (1.5) we also see that this right-hand side is a odd integer. Since $\mathbf{N}_y \equiv \mathbf{Log}(2\mathbb{N}+1 \cap [y, y^\alpha])$ and

$$\sum_{N \in 2\mathbb{N}+1 \cap [y,y^\alpha]} \frac{1}{N} = \left(1 + O\left(\frac{1}{x}\right)\right) \frac{\alpha - 1}{2} \log y,$$

we thus see that when (5.13) occurs, one has

$$\mathbb{P}(\mathrm{Aff}_{\vec{a}}(\mathbf{N}_y) = M) = \frac{1}{\left(1 + O(\frac{1}{x})\right) \frac{\alpha-1}{2} \log y} 2^{-|\vec{a}|} \frac{3^{n-m_0}}{M - F_{n-m_0}(\vec{a})}.$$

From (5.12), (5.1), (1.13) we can write

$$M - F_{n-m_0}(\vec{a}) = M - O(3^{n_0}) = (1 + O(x^{-c}))M$$

and thus

$$\mathbb{P}(\mathrm{Aff}_{\vec{a}}(\mathbf{N}_y) = M) = \frac{1 + O(x^{-c})}{\frac{\alpha-1}{2} \log y} \frac{2^{-|\vec{a}|} 3^{n-m_0}}{M}.$$

We conclude that

$$\mathbb{P}(\mathrm{Pass}_x(\mathbf{N}_y) \in E) = \frac{1 + O(x^{-c})}{\frac{\alpha-1}{2} \log y} \sum_{n \in I_y} 3^{n-m_0} \sum_{\vec{a} \in \mathcal{A}^{(n-m_0)}} 2^{-|\vec{a}|} \sum_{M \in E' : M = F_{n-m_0}(\vec{a}) \bmod 3^{n-m_0}} \frac{1}{M} + O(\log^{-c} x).$$

Since from (5.10) we have

$$\# I_y = (1 + O(\log^{-c} x)) \frac{\alpha - 1}{\log \frac{4}{3}} \log y,$$

it thus suffices to establish the estimate

$$3^{n-m_0} \sum_{\vec{a} \in \mathcal{A}^{(n-m_0)}} 2^{-|\vec{a}|} \sum_{M \in E' : M = F_{n-m_0}(\vec{a}) \bmod 3^{n-m_0}} \frac{1}{M} = Z + O(\log^{-c} x) \tag{5.15}$$

for all $n \in I_y$ and some quantity Z that does not depend on n or y.

Fix n. The left-hand side of (5.15) may be written as

$$\mathbb{E} 1_{(\mathbf{a}_1,\dots,\mathbf{a}_{n-m_0}) \in \mathcal{A}^{(n-m_0)}} c_n(F_{n-m_0}(\mathbf{a}_1, \dots, \mathbf{a}_{n-m_0}) \bmod 3^{n-m_0}) \tag{5.16}$$

where $(\mathbf{a}_1, \dots, \mathbf{a}_{n-m_0}) \equiv \mathbf{Geom}(2)^{n-m_0}$ and $c_n \colon \mathbb{Z}/3^{n-m_0}\mathbb{Z} \to \mathbb{R}^+$ is the function

$$c_n(X) := 3^{n-m_0} \sum_{M \in E' : M = X \bmod 3^{n-m_0}} \frac{1}{M}. \tag{5.17}$$

We have a basic estimate:

Lemma 5.2. *We have $c_n(X) \ll 1$ for all $n \in I_y$ and $X \in \mathbb{Z}/3^{n-m_0}\mathbb{Z}$.*

Proof. If $M \in E'$, then on setting $(a_1, \dots, a_{m_0}) := \vec{a}^{(m_0)}(M)$ we from (1.7) that

$$3^{m_0} 2^{-a_{[1,m_0]}} M + F_{m_0}(a_1, \dots, a_{m_0}) \leq x < 3^{m_0} 2^{-a_{[1,m_0-1]}} M + F_{m_0-1}(a_1, \dots, a_{m_0-1})$$

which by (5.11) and (1.13) implies that

$$3^{m_0} 2^{-a_{[1,m_0]}} M \leq x \ll 3^{m_0} 2^{-a_{[1,m_0-1]}} M$$

or equivalently

$$3^{-m_0} 2^{a_{[1,m_0-1]}} x \ll M \leq 3^{-m_0} 2^{a_{[1,m_0]}} x. \tag{5.18}$$

Also, from (1.7) we also have that

$$3^{m_0} M + 2^{a_{[1,m_0]}} F_{m_0}(a_1, \dots, a_{m_0}) = 2^{a_{[1,m_0]}} \bmod 2^{a_{[1,m_0]}+1}$$

and so M is constrained to a single residue class modulo $2^{a_{[1,m_0]}+1}$. In (5.17) we are also constraining M to a single residue class modulo 3^{n-m_0}; by the Chinese remainder theorem, these constraints can be combined into a single residue class modulo $2^{a_{[1,m_0]}+1} 3^{n-m_0}$. If $2^{a_{[1,m_0]}} \leq x^{0.5}$ (say), then the modulus here is much less than the lower bound on M in (5.18), and the contribution of the tuple $(a_1, \dots, a_{m_0})$ to $c_n(X)$ can then be bounded using the integral test by

$$\begin{aligned} &\ll 3^{n-m_0} (2^{a_{[1,m_0]}+1} 3^{n-m_0})^{-1} \log O\left(\frac{3^{-m_0} 2^{a_{[1,m_0]}} x}{3^{-m_0} 2^{a_{[1,m_0-1]}} x} \right) \\ &\ll 2^{-a_{[1,m_0]}} a_{m_0} \end{aligned}$$

which sums to $O(1)$. If instead $2^{a_{[1,m_0]}} \geq x^{0.5}$, we recall from (1.7) that

$$a_{m_0} = \nu_2\left(3(3^{m_0} 2^{-a_{[1,m_0-1]}} M + F_{m_0-1}(a_1, \dots, a_{m_0-1})) + 1\right)$$

so

$$2^{a_{m_0}} \ll 3^{m_0} 2^{-a_{[1,m_0-1]}} M + F_{m_0-1}(a_1, \dots, a_{m_0-1}) \ll 3^{m_0} 2^{-a_{[1,m_0-1]}} M$$

(using (1.13), (5.18) to handle the lower order term). Hence we we have the additional lower bound

$$M \gg 3^{-m_0} 2^{a_{[1,m_0]}}.$$

The integral test then gives an additional contribution in the $2^{a_{[1,m_0]}} \leq x^{0.5}$ case of

$$3^{n-m_0} 3^{m_0} 2^{-a_{[1,m_0]}}$$

to the contribution of the tuple $(a_1, \dots, a_{m_0})$ to $c_n(X)$; summing over $a_1, \dots, a_{m_0}$ using Lemma 2.2, (5.11), (5.1), we see that this contribution is $O(x^{-c})$ if α is sufficiently close to 1. The claim follows. $\square$

From the above lemma and (5.6), we may write (5.16) as

$$\mathbb{E} c_n(F_{n-m_0}(\mathbf{a}_1, \dots, \mathbf{a}_{n-m_0}) \bmod 3^{n-m_0}) + O(\log^{-c} x)$$

which by (1.20) is equal to

$$\sum_{X \in \mathbb{Z}/3^{n-m_0}\mathbb{Z}} c_n(X) \mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^{n-m_0}\mathbb{Z}) = X) + O(\log^{-c} x).$$

From (5.10), (5.11) we have $n - m_0 \geq m_0$. Applying Proposition 1.14, Lemma 5.2 and the triangle inequality, one can thus write the preceding expression as

$$\sum_{X \in \mathbb{Z}/3^{n-m_0}\mathbb{Z}} c_n(X) 3^{2m_0-n} \mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^{m_0}\mathbb{Z}) = X \bmod 3^{m_0}) + O(\log^{-c} x)$$

which by (5.17) can be rewritten as

$$Z + O(\log^{-c} x)$$

where Z is the quantity

$$Z := \sum_{M \in E'} \frac{\mathbb{P}(M = \mathbf{Syrac}(\mathbb{Z}/3^{m_0}\mathbb{Z}) \bmod 3^{m_0})}{M}.$$

Since Z does not depend on n or y, the claim follows.

6. Reduction to Fourier decay bound

In this section we derive Proposition 1.14 from Proposition 1.17. We first observe that to prove Proposition 1.14, it suffices to do so in the regime

$$0.9n \leq m \leq n. \tag{6.1}$$

(The main significance of the constant 0.9 here is that it lies between $\frac{\log 3}{2\log 2} \approx 0.7925$ and 1.) Indeed, once one has (1.22) in this regime, one also has from (1.21) that

$$\sum_{Y \in \mathbb{Z}/3^{n'}\mathbb{Z}} \left| 3^{n-n'}\mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = Y \bmod 3^n) - 3^{m-n'}\mathbb{P}(\mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z}) = Y \bmod 3^m) \right| \ll_A m^{-A}$$

whenever $0.9n \leq m \leq n \leq n'$, and the claim (1.22) for general $10 \leq m \leq n$ then follows from telescoping series, with the remaining cases $1 \leq m < 10$ following trivially from the triangle inequality.

Henceforth we assume (6.1). We also fix $A > 0$, and let C_A be a constant that is sufficiently large depending on A. We may assume that n (and hence m) are sufficiently large depending on A, C_A, since the claim is trivial otherwise.

Let $(\mathbf{a}_1, \dots, \mathbf{a}_n) \equiv \mathbf{Geom}(2)^n$, and define the random variable

$$\mathbf{X}_n := 2^{-\mathbf{a}_1} + 3^1 2^{-\mathbf{a}_{[1,2]}} + \cdots + 3^{n-1} 2^{-\mathbf{a}_{[1,n]}} \bmod 3^n,$$

thus $\mathbf{X}_n \equiv \mathbf{Syrac}(\mathbb{Z}/3^n\mathbb{Z})$. The strategy will be to split $\mathbf{X}_n$ (after some conditioning and removal of exceptional events) as the sum of two independent components, one of which has quite large entropy (or more precisely, Renyi 2-entropy) in $\mathbb{Z}/3^n\mathbb{Z}$ thanks to some elementary number theory, and the other having very small Fourier coefficients at high frequencies thanks to Proposition 1.17. The desired bound will then follow from some L^2-based Fourier analysis (i.e., Plancherel's theorem).

We turn to the details. Let E denote the event that one of the estimates

$$|\mathbf{a}_{[i+i,j]} - 2(j-i)| \leq C_A(\sqrt{(j-i)(\log n)} + \log n) \tag{6.2}$$

fails for some $1 \leq i < j \leq n$. By Lemma 2.2 and the union bound, we have

$$\begin{aligned}
\mathbb{P}(E) &\ll \sum_{1\leq i<j\leq n} G_{j-i}(cC_A(\sqrt{(j-i)(\log n)}+\log n)) \\
&\ll \sum_{1\leq i<j\leq n} \exp(-cC_A\log n)+\exp(-cC_A\log n) \\
&\ll n^2 n^{-cC_A} \\
&\ll n^{-A-1}
\end{aligned} \tag{6.3}$$

if C_A is large enough. By the triangle inequality, we may then bound the left-hand side of (1.22) by

$$\mathrm{Osc}_{m,n}\left(\mathbb{P}(\mathbf{X}_n = Y \wedge \overline{E})\right)_{Y\in\mathbb{Z}/3^n\mathbb{Z}} + O(n^{-A-1})$$

where $\overline{E}$ denotes the complement of E, so it now suffices to show that

$$\mathrm{Osc}_{m,n}\left(\mathbb{P}(\mathbf{X}_n = Y \wedge \overline{E})\right)_{Y\in\mathbb{Z}/3^n\mathbb{Z}} \ll_{A,C_A} n^{-A}.$$

Now suppose that E does not hold. From (6.2) we have

$$\mathbf{a}_{[1,n]} \geq 2n - C_A(\sqrt{n\log n}+\log n) > n\frac{\log 3}{\log 2}$$

since $\frac{\log 3}{\log 2} < 2$ and n is large. Thus, there is a well defined *stopping time* $0 \leq \mathbf{k} < n$, defined as the unique natural number for which

$$\mathbf{a}_{[1,\mathbf{k}]} \leq n\frac{\log 3}{\log 2} - C_A^2\log n < \mathbf{a}_{[1,\mathbf{k}+1]}.$$

From (6.2) we have

$$\mathbf{k} = n\frac{\log 3}{2\log 2} + O(C_A^2\log n).$$

It thus suffices by the union bound to show that

$$\mathrm{Osc}_{m,n}\left(\mathbb{P}(\mathbf{X}_n = Y \wedge \overline{E} \wedge B_k)\right)_{Y\in\mathbb{Z}/3^n\mathbb{Z}} \ll_{A,C_A} n^{-A-1} \tag{6.4}$$

for all

$$k = n\frac{\log 3}{2\log 2} + O(C_A^2\log n), \tag{6.5}$$

where B_k is the event that $\mathbf{k} = k$, or equivalently that

$$\mathbf{a}_{[1,k]} \leq n\frac{\log 3}{\log 2} - C_A^3\log n < \mathbf{a}_{[1,k+1]}. \tag{6.6}$$

Fix k. In order to decouple the events involved in (6.4) we need to shrink the exceptional event E slightly, so that it only depends on $\mathbf{a}_1,\dots,\mathbf{a}_{k+1}$ and not on $\mathbf{a}_{k+2},\dots,\mathbf{a}_n$. Let E_k denote the event that one of the estimates (6.2) fails for $1 \leq i < j \leq k+1$, thus E_k is contained in E. Then the difference between $\overline{E}$ and $\overline{E}_k$ has probability $O(n^{-A-1})$ by (6.3), so by the triangle inequality it now suffices to show that

$$\mathrm{Osc}_{m,n}\left(\mathbb{P}(\mathbf{X}_n = Y \wedge \overline{E}_k \wedge B_k)\right)_{Y\in\mathbb{Z}/3^n\mathbb{Z}} \ll_{A,C_A} n^{-A-1}.$$

From (6.6) and (6.2) we see that we have

$$n\frac{\log 3}{\log 2} - C_A^3\log n \leq \mathbf{a}_{[1,k+1]} \leq n\frac{\log 3}{\log 2} - \frac{1}{2}C_A^3\log n. \tag{6.7}$$

whenever one is in the event $\overline{E}_k \wedge B_k$. By a further application of the triangle inequality, it suffices to show that

$$\mathrm{Osc}_{m,n}\left(\mathbb{P}(\mathbf{X}_n = Y \wedge \overline{E}_k \wedge B_k \wedge C_{k,l})\right)_{Y \in \mathbb{Z}/3^n\mathbb{Z}} \ll_{A,C_A} n^{-A-2}$$

for all l in the range

$$n\frac{\log 3}{\log 2} - C_A^3 \log n \leq l \leq n\frac{\log 3}{\log 2} - \frac{1}{2}C_A^3 \log n, \tag{6.8}$$

where $C_{k,l}$ is the event that $\mathbf{a}_{[1,k+1]} = l$.

Fix l. If we let $g = g_{n,k,l} \colon \mathbb{Z}/3^n\mathbb{Z} \to \mathbb{R}$ denote the function

$$g_{n,k,l}(Y) := \mathbb{P}(\mathbf{X}_n = Y \wedge \overline{E}_k \wedge B_k \wedge C_{k,l})$$

then our task can be written as

$$\sum_{Y \in \mathbb{Z}/3^n\mathbb{Z}} \left| g(Y) - 3^{m-n} \sum_{Y' \in \mathbb{Z}/3^n\mathbb{Z}: Y' = Y \bmod 3^m} g(Y') \right| \ll_{A,C_A} n^{-A-2}.$$

By Cauchy-Schwarz, it suffices to show that

$$\sum_{Y \in \mathbb{Z}/3^n\mathbb{Z}} \left| g(Y) - 3^{m-n} \sum_{Y' \in \mathbb{Z}/3^n\mathbb{Z}: Y' = Y \bmod 3^m} g(Y') \right|^2 \ll_{A,C_A} n^{-2A-4}. \tag{6.9}$$

By Plancherel's theorem, the left-hand side of (6.9) may be written as

$$\sum_{\xi \in \mathbb{Z}/3^n\mathbb{Z}: 3^{n-m} \nmid \xi} \left| \sum_{Y \in \mathbb{Z}/3^n\mathbb{Z}} g(Y) e^{-2\pi i \xi Y/3^n} \right|^2.$$

We can write

$$\sum_{Y \in \mathbb{Z}/3^n\mathbb{Z}} g(Y) e^{-2\pi i \xi Y/3^n} = \mathbb{E} e^{-2\pi i \xi \mathbf{X}_n/3^n} 1_{\overline{E}_k \wedge B_k \wedge C_{k,l}}.$$

On the event $C_{k,l}$, one can use (1.5), (1.24) to write

$$\mathbf{X}_n = F_{k+1}(\mathbf{a}_{k+1}, \dots, \mathbf{a}_1) + 3^{k+1} 2^{-l} F_{n-k-1}(\mathbf{a}_n, \dots, \mathbf{a}_{k+2}) \bmod 3^n.$$

The key point here is that the expression $3^{k+1}2^{-l}F_{n-k-1}(\mathbf{a}_n, \dots, \mathbf{a}_{k+2})$ is independent of $\mathbf{a}_1, \dots, \mathbf{a}_{k+1}, \overline{E}_k, B_k, C_{k,l}$. Thus we may factor

$$\begin{aligned}\sum_{Y \in \mathbb{Z}/3^n\mathbb{Z}} g(Y) e^{-2\pi i \xi Y/3^n} &= \mathbb{E} e^{-2\pi i \xi (F_{k+1}(\mathbf{a}_{k+1}, \dots, \mathbf{a}_1) \bmod 3^n)/3^n} 1_{\overline{E}_k \wedge B_k \wedge C_{k,l}} \\ &\quad \times \mathbb{E} e^{-2\pi i \xi (2^{-l} F_{n-k-1}(\mathbf{a}_n, \dots, \mathbf{a}_{k+2}) \bmod 3^{n-k-1})/3^{n-k-1}}.\end{aligned}$$

We can write $\xi = 3^j 2^l \xi' \bmod 3^n$ where $0 \leq j < n - m \leq 0.1n$ and ξ' is not divisible by 3. In particular, from (6.5) one has

$$n - k - j - 1 \gg n.$$

Then by (1.21) we have

$$\mathbb{E} e^{-2\pi i \xi (2^{-l} F_{n-k-1}(\mathbf{a}_n, \dots, \mathbf{a}_{k+2}) \bmod 3^{n-k-1})/3^{n-k-1}} = \mathbb{E} e^{-2\pi i \xi' \mathbf{Syrac}(\mathbb{Z}/3^{n-k-j-1}\mathbb{Z})/3^{n-k-j-1}}$$

and hence by Proposition 1.17 this quantity is $O_{A'}(n^{-A'})$ for any A'. Thus we can bound the left-hand side of (6.9) by

$$\ll_{A'} n^{-2A'} \sum_{\xi \in \mathbb{Z}/3^n\mathbb{Z}} \left| \mathbb{E} e^{-2\pi i \xi (F_{k+1}(\mathbf{a}_{k+1}, \dots, \mathbf{a}_1) \bmod 3^n)/3^n} 1_{\overline{E}_k \wedge B_k \wedge C_{k,l}} \right|^2$$

which by Plancherel's theorem can be written as

$$\ll_{A'} n^{-2A'} 3^n \sum_{Y \in \mathbb{Z}/3^n\mathbb{Z}} \mathbb{P}(F_{k+1}(\mathbf{a}_{k+1}, \dots, \mathbf{a}_1) = Y \wedge \overline{E}_k \wedge B_k \wedge C_{k,l})^2.$$

Remark 6.1. If we ignore the technical events $\overline{E}_k, B_k, C_{k,l}$, this quantity is essentially the Renyi 2-entropy (also known as *collision entropy*) of $F_{k+1}(\mathbf{a}_{k+1}, \dots, \mathbf{a}_1) \bmod 3^n$.

Now we make a key elementary number theory observation:

Lemma 6.2 (Injectivity of offsets)**.** *For each natural number n, the n-Syracuse offset map $F_n \colon (\mathbb{N}+1)^n \to \mathbb{Z}[\frac{1}{2}]$ is injective.*

Proof. Suppose that $(a_1, \dots, a_n), (a'_1, \dots, a'_n) \in (\mathbb{N}+1)^n$ are such that $F_n(a_1, \dots, a_n) = F_n(a'_1, \dots, a'_n)$. Taking 2-valuations of both sides using (1.5), we conclude that

$$-a_{[1,n]} = -a'_{[1,n]}.$$

On the other hand, from (1.5) we have

$$F_n(a_1, \dots, a_n) = 3^n 2^{-a_{[1,n]}} + F_{n-1}(a_2, \dots, a_n)$$

and similarly for $a'_1, \dots, a'_n$, hence

$$F_{n-1}(a_2, \dots, a_n) = F_{n-1}(a'_2, \dots, a'_n).$$

The claim now follows from iteration (or an induction on n). □

We will need a more quantitative 3-adic version of this injectivity:

Corollary 6.3 (3-adic separation of offsets)**.** *The residue classes $F_{k+1}(a_{k+1}, \dots, a_1) \bmod 3^n$, as $(a_1, \dots, a_{k+1}) \in (\mathbb{N}+1)^{k+1}$ range over $k+1$-tuples of positive integers that obey the conditions*

$$|a_{[i+1,j]} - 2(j-i)| \leq C_A \left(\sqrt{(j-i)(\log n)} + \log n \right) \tag{6.10}$$

for $1 \leq i < j \leq k+1$ as well as

$$a_{[1,k+1]} = l, \tag{6.11}$$

are distinct.

Proof. Suppose that $(a_1, \dots, a_{k+1}), (a'_1, \dots, a'_{k+1})$ are two tuples of positive integers that both obey (6.10), (6.11), and such that

$$F_{k+1}(a_{k+1}, \dots, a_1) = F_{k+1}(a'_{k+1}, \dots, a'_1) \bmod 3^n.$$

Applying (1.5) and multiplying by 2^l, we conclude that

$$\sum_{j=1}^{k+1} 3^{j-1} 2^{l-a_{[1,j]}} = \sum_{j=1}^{k+1} 3^{j-1} 2^{l-a'_{[1,j]}} \mod 3^n. \tag{6.12}$$

From (6.11), the expressions on the left and right sides are natural numbers. Using (6.10), (6.8), and Young's inequality, the left-hand side may be bounded for C_A large enough by

$$\begin{aligned}\sum_{j=1}^{k+1} 3^{j-1} 2^{l-a_{[1,j]}} &\ll 2^l \sum_{j=1}^{k+1} 3^j 2^{-2j+C_A(\sqrt{j\log n}+\log n)} \\ &\ll \exp(-C_A^3 \log n) 3^n \sum_{j=1}^{k+1} \exp\left(-j\log\frac{4}{3} + O(C_A j^{1/2}\log^{1/2} n) + O(C_A \log n)\right) \\ &\ll \exp\left(-\frac{C_A^3}{2}\log n\right) 3^n \sum_{j=1}^{k+1} \exp(-cj) \\ &\ll n^{-\frac{C_A^3}{2}} 3^n;\end{aligned}$$

in particular, for n large enough, this expression is less than 3^n. Similarly for the right-hand side of (6.12). Thus these two sides are equal as natural numbers, not simply as residue classes modulo 3^n:

$$\sum_{j=1}^{k+1} 3^{j-1} 2^{l-a_{[1,j]}} = \sum_{j=1}^{k+1} 3^{j-1} 2^{l-a'_{[1,j]}}. \tag{6.13}$$

Dividing by 2^l, we conclude $F_{k+1}(a_{k+1},\dots,a_1) = F_{k+1}(a'_{k+1},\dots,a'_1)$. From Lemma 6.2 we conclude that $(a_1,\dots,a_{k+1}) = (a'_1,\dots,a'_{k+1})$, and the claim follows. □

In view of the above lemma, we see that for a given choice of $Y \in \mathbb{Z}/3^n\mathbb{Z}$, the event

$$F_{k+1}(\mathbf{a}_{k+1},\dots,\mathbf{a}_1) = Y \wedge \overline{E}_k \wedge B_k \wedge C_{k,l}$$

can only be non-empty for at most one value $(a_1,\dots,a_m)$ of the tuple $(\mathbf{a}_1,\dots,\mathbf{a}_m)$. By Definition 1.7, such a value is attained with probability $2^{-a_{[1,m]}} = 2^{-l}$, which by (6.8) is equal to $n^{O(C_A^3)}3^{-n}$. We can thus bound the left-hand side of (6.9) by

$$\ll_{A'} n^{-2A'+O(C_A^3)},$$

and the claim now follows by taking A' large enough. This concludes the proof of Proposition 1.14 assuming Proposition 1.17.

7. Decay of Fourier coefficients

In this section we establish Proposition 1.17, which when combined with all the implications established in preceding sections will yield Theorem 1.3.

Fix n, ξ, A; without loss of generality we may assume A to be larger than any fixed absolute constant. We let $\chi\colon \mathbb{Z}[\frac{1}{2}] \to \mathbb{C}$ denote the character

$$\chi(x) := e^{-2\pi i \xi (x \bmod 3^n)/3^n} \tag{7.1}$$

where $x \mapsto x \bmod 3^n$ is the ring homomorphism from $\mathbb{Z}[\frac{1}{2}]$ to $\mathbb{Z}/3^n\mathbb{Z}$. Note that χ is a group homomorphism from the additive group $\mathbb{Z}[\frac{1}{2}]$ to the multiplicative group $\mathbb{C}$. From (1.24), our task is now to show that

$$\mathbb{E}\chi(2^{-\mathbf{a}_1} + 3^1 2^{-\mathbf{a}_{[1,2]}} + \cdots + 3^{n-1} 2^{-\mathbf{a}_{[1,n]}}) \ll_A n^{-A}, \tag{7.2}$$

where $(\mathbf{a}_1, \dots, \mathbf{a}_n) \equiv \mathbf{Geom}(2)^n$.

To extract some usable cancellation on the left-hand side, we will group the sum on the left-hand side into pairs. For any real $x > 0$, let $[x]$ denote the discrete interval

$$[x] := \{j \in \mathbb{N}+1 : j \leq x\} = \{1, \dots, \lfloor x \rfloor\}.$$

For $j \in [n/2]$, set $\mathbf{b}_j := \mathbf{a}_{2j-1} + \mathbf{a}_{2j}$, so that

$$2^{-\mathbf{a}_1} + 3^1 2^{-\mathbf{a}_{[1,2]}} + \cdots + 3^{n-1} 2^{-\mathbf{a}_{[1,n]}} = \sum_{j \in [n/2]} 3^{2j-2} 2^{-\mathbf{b}_{[1,j]}} (2^{\mathbf{a}_{2j}} + 3) + 2^{-\mathbf{b}_{[1\lfloor n/2 \rfloor]} - \mathbf{a}_n}$$

when n is odd, where we extend the summation notation (1.6) to the $\mathbf{b}_j$. For n even, the formula is the same except that the final term $2^{-\mathbf{b}_{[1\lfloor n/2 \rfloor]} - \mathbf{a}_n}$ is omitted. Note that the $\mathbf{b}_1, \dots, \mathbf{b}_{\lfloor n/2 \rfloor}$ are jointly independent random variables taking values in $\mathbb{N}+2 = \{2, 3, 4, \dots\}$; they are iid copies of a Pascal (or negative binomial) random variable $\mathbf{Pascal} \equiv \mathbf{NB}(2, \frac{1}{2})$ on $\mathbb{N}+2$, defined by

$$\mathbb{P}(\mathbf{Pascal} = b) = \frac{b-1}{2^b}$$

for $b \in \mathbb{N}+2$.

For any $j \in [n]$, $\mathbf{a}_{2j}$ is independent of all of the $\mathbf{b}_1, \dots, \mathbf{b}_{\lfloor n/2 \rfloor}$ except for $\mathbf{b}_j$. For n odd, $\mathbf{a}_n$ is independent of all of the $\mathbf{b}_j$. We conclude that the left-hand side of (7.2) is equal to

$$\mathbb{E}\left(\prod_{j \in [n/2]} f(3^{2j-2} 2^{-\mathbf{b}_{[1,j]}}, \mathbf{b}_j)\right) g(2^{-\mathbf{b}_{[1,\lfloor n/2 \rfloor]}})$$

when n is odd, with the factor $g(2^{-\mathbf{b}_{[1,\lfloor n/2 \rfloor]}})$ omitted when n is even, where $f(x, b)$ is the conditional expectation

$$f(x, b) := \mathbb{E}\left(\chi(x(2^{\mathbf{a}_2} + 3)) | \mathbf{a}_1 + \mathbf{a}_2 = b\right) \tag{7.3}$$

(with $(\mathbf{a}_1, \mathbf{a}_2) \equiv \mathbf{Geom}(2)^2$) and

$$g(x) := \mathbb{E}\chi(x 2^{-\mathbf{Geom}(2)}).$$

Clearly $|g(x)| \leq 1$, so by the triangle inequality we can bound the left-hand side of (7.2) in magnitude by

$$\mathbb{E} \prod_{j \in [n/2]} |f(3^{2j-2} 2^{-\mathbf{b}_{[1,j]}}, \mathbf{b}_j)|$$

regardless of whether n is even or odd.

From (7.3) we certainly have

$$|f(x,b)| \leq 1. \tag{7.4}$$

We now perform an explicit computation to improve upon this estimate for many values of x in the case $b = 3$, which is the least value of $b \in \mathbb{N}+2$ for which the event $\mathbf{a}_1 + \mathbf{a}_2 = b$ does not completely determine $\mathbf{a}_1$ or $\mathbf{a}_2$. For any $(j,l) \in (\mathbb{N}+1) \times \mathbb{Z}$, we can write

$$\chi(3^{2j-2}2^{-l+1}) = e^{-2\pi i\theta(j,l)} \tag{7.5}$$

where $\theta(j,l) \in (-1/2, 1/2]$ denotes the argument

$$\theta(j,l) := \left\{ \frac{\xi 3^{2j-2}(2^{-l+1} \bmod 3^n)}{3^n} \right\} \tag{7.6}$$

and $\{\}\colon \mathbb{R}/\mathbb{Z} \to (-1/2, 1/2]$ is the signed fractional part function, thus $\{x\}$ denotes the unique element of the coset $x + \mathbb{Z}$ that lies in $(-1/2, 1/2]$. From (7.6) (or (7.5)) we observe the identity

$$3^{2(j_*-j)}2^{(l-l_*)}\theta(j,l) = \theta(j_*, l_*) \bmod \mathbb{Z} \tag{7.7}$$

whenever $j \leq j_*$ and $l \geq l_*$. Thus for instance $\theta(j+1,l) = 9\theta(j,l) \bmod \mathbb{Z}$ and $\theta(j,l-1) = 2\theta(j,l) \bmod \mathbb{Z}$.

Let $0 < \varepsilon < \frac{1}{100}$ be a small absolute constant to be chosen later. Call a point $(j,l) \in [n/2] \times \mathbb{Z}$ *black* if

$$|\theta(j,l)| \leq \varepsilon, \tag{7.8}$$

and *white* otherwise. We let B, W denote the black and white points of $[n/2] \times \mathbb{Z}$ respectively, thus we have the partition $[n/2] \times \mathbb{Z} = B \uplus W$.

Lemma 7.1 (Cancellation for white points)**.** *If (j,l) is white, then*

$$|f(3^{2j-2}2^{-l}, 3)| \leq \exp(-\varepsilon^3).$$

Proof. If $\mathbf{a}_1, \mathbf{a}_2$ are independent copies of $\mathbf{Geom}(2)$, then after conditioning to the event $\mathbf{a}_1 + \mathbf{a}_2 = 2$, the pair $(\mathbf{a}_1, \mathbf{a}_2)$ is equal to either $(1,2)$ or $(2,1)$, with each pair occuring with probability $1/2$. From (7.3) we thus have

$$f(x,3) = \frac{1}{2}\chi(5x) + \frac{1}{2}\chi(7x) = \frac{\chi(5x)}{2}(1 + \chi(2x))$$

for any x, so that

$$|f(x,3)| = \frac{|1 + \chi(2x)|}{2}.$$

We specialise to the case $x := 3^{2j-2}2^{-l}$. By (7.5) we have

$$\chi(2x) = e^{-2\pi i\theta(j,\mathbf{b}_{[1,j]})}$$

and hence by elementary trigonometry

$$|f(3^{2j-2}2^{-l}, 3)| = \cos(\pi\theta(j,l)).$$

By hypothesis we have

$$|\theta(j,l)| > \varepsilon$$

and the claim now follows by Taylor expansion (if ε is small enough). ☐

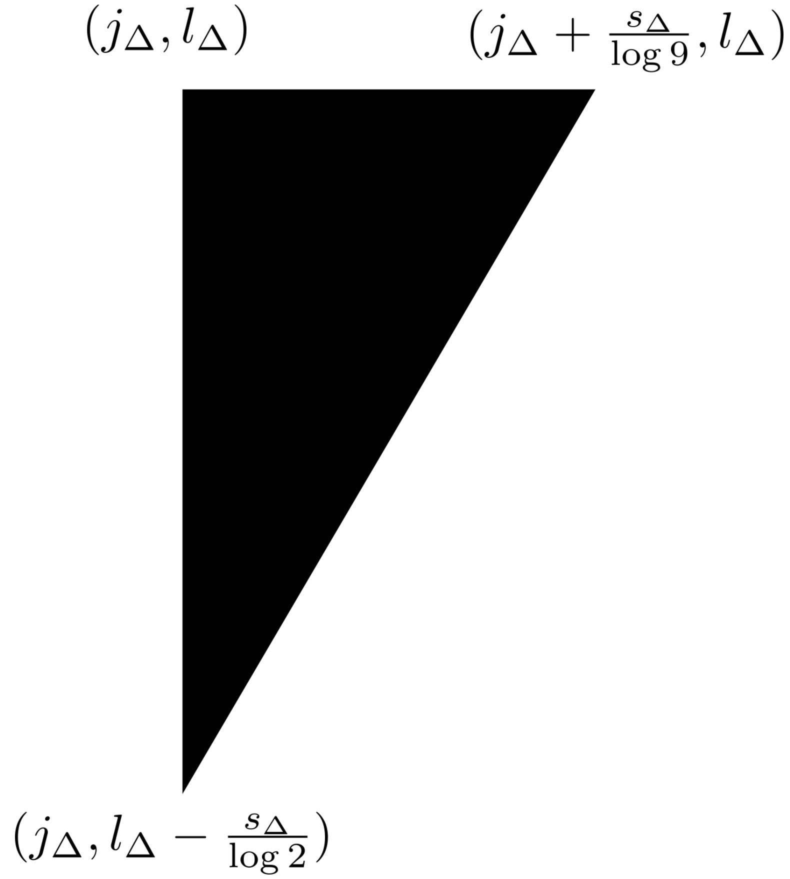

FIGURE 1. A triangle Δ, which we have drawn as a solid region rather than as a subset of the discrete lattice $\mathbb{Z}^2$.

From the above lemma and (7.4), we see that the left-hand side of (7.2) is bounded in magnitude by

$$\exp(-\varepsilon^3 \#\{j \in [n/2] : \mathbf{b}_j = 3, (j, \mathbf{b}_{[1,j]}) \in W\})$$

and so it suffices to show that

$$\mathbb{E} \exp(-\varepsilon^3 \#\{j \in [n/2] : \mathbf{b}_j = 3, (j, \mathbf{b}_{[1,j]}) \in W\}) \ll_A n^{-A}. \tag{7.9}$$

The proof of (7.9) consists of a "deterministic" part, in which we understand the description of the white set W (or the black set B), and a "probabilistic" part, in which we control the random walk $\mathbf{b}_{[1,j]}$ and the events $\mathbf{b}_j = 3$. We begin with the former task. Define a *triangle* to be a subset Δ of $(\mathbb{N}+1) \times \mathbb{Z}$ of the form

$$\Delta = \{(j, l) : j \geq j_\Delta; l \leq l_\Delta; (j - j_\Delta) \log 9 + (l_\Delta - l) \log 2 \leq s_\Delta\} \tag{7.10}$$

for some $(j_\Delta, l_\Delta) \in (\mathbb{N}+1) \times \mathbb{Z}$ (which we call the *top left corner* of Δ) and some $s_\Delta \geq 0$ (which we call the *size* of Δ); see Figure 1.

Lemma 7.2 (Description of black set)**.** *The black set can be expressed as a disjoint union*

$$B = \biguplus_{\Delta \in \mathcal{T}} \Delta$$

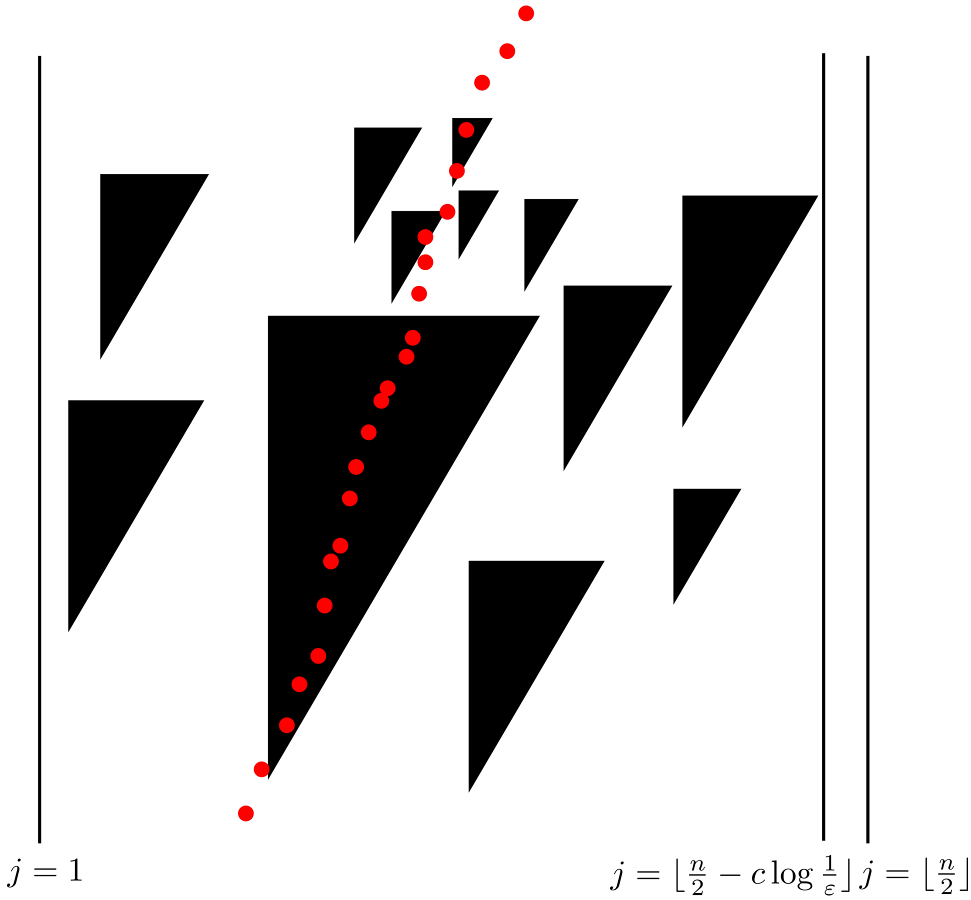

FIGURE 2. The black set is a union of triangles, in the strip $[\frac{n}{2} - c \log \frac{1}{\varepsilon}] \times \mathbb{Z}$, that are separated from each other by $\gg \log \frac{1}{\varepsilon}$. The red dots depict (a portion of) a renewal process $\mathbf{v}_1, \mathbf{v}_{[1,2]}, \mathbf{v}_{[1,3]}$ that we will encounter later in this section. We remark that the average slope $\frac{16}{4} = 4$ of this renewal process will exceed the slope $\frac{\log 9}{\log 2} \approx 3.17$ of the triangle diagonals, so that the process tends to exit a given triangle through its horizontal side.

of triangles Δ, each of which is contained in $[\frac{n}{2} - c \log \frac{1}{\varepsilon}] \times \mathbb{Z}$. Furthermore, any two triangles Δ, Δ' in $\mathcal{T}$ are separated by a distance $\geq c \log \frac{1}{\varepsilon}$ (using the Euclidean metric on $[n/2] \times \mathbb{Z} \subset \mathbb{R}^2$), where the constant $c > 0$ does not depend on ε. (See Figure 2.)

Proof. Suppose $(j, l) \in [n/2] \times \mathbb{Z}$ is black, then by (7.8), (7.6) we have

$$\frac{\xi 3^{2j-2} (2^{-l+1} \bmod 3^n)}{3^n} \in [-\varepsilon, \varepsilon] \bmod \mathbb{Z}$$

and hence

$$\frac{\xi 3^{n-1} (2^{-l+1} \bmod 3^n)}{3^n} \in [-3^{n+1-2j} \varepsilon, 3^{n+1-2j} \varepsilon] \bmod \mathbb{Z}.$$

On the other hand, since ξ is not a multiple of 3, the expression $\frac{\xi 3^{n-1}(2^{-l+1} \bmod 3^n)}{3^n}$ is either equal to 1/3 or 2/3 mod $\mathbb{Z}$. We conclude that

$$3^{n+1-2j}\varepsilon \geq \frac{1}{3}. \tag{7.11}$$

Hence all the white points in $[n/2] \times \mathbb{Z}$ actually lie in $[\frac{n}{2} - c\log\frac{1}{\varepsilon}] \times \mathbb{Z}$ for some $c > 0$ independent of ε.

Suppose that $(j,l) \in [n/2] \times \mathbb{Z}$ is such that (j,l') is black for all $l' \geq l$, thus

$$|\theta(j,l')| \leq \varepsilon$$

for all $l' \geq l$. But from (7.7) we have

$$\theta(j,l') = 2\theta(j,l'+1) \bmod \mathbb{Z}$$

for all l'. This implies that

$$\theta(j,l') = 2\theta(j,l'+1)$$

for all $l' \geq l$, hence

$$\theta(j,l') \leq 2^{l-l'}\varepsilon$$

for all $l' \geq l$. Repeating the proof of (7.11), one concludes that

$$3^{n+1-2j}2^{l-l'}\varepsilon \geq \frac{1}{3},$$

which is absurd for l' large enough. Thus it is not possible for (j,l') to be black for all $l' \geq l$.

Now let $(j,l) \in [n/2] \times \mathbb{Z}$ be black. By the preceding discussion, there exists a unique $l_* = l_*(j,l) \geq l$ such that (j,l') is black for all $l \leq l' \leq l_*$, but such that (j, l_*+1) is white. Now let $j_* = j_*(j,l) \leq j$ be the unique positive integer such that (j', l_*) is black for all $j_* \leq j' \leq j$, but such that either $j_* = 1$ or (j_*-1, l_*) is white. Informally, (j_*, l_*) is obtained from (j,l) by first moving upwards as far as one can go in B, then moving leftwards as far as one can go in B. As one should expect from glancing at Figure 2, (j_*, l_*) should be the top left corner of the triangle containing (j,l), and the arguments below are intended to support this claim.

By construction, (j_*, l_*) is black, thus by (7.8) we have

$$|\theta(j_*, l_*)| = \varepsilon \exp(-s_*) \tag{7.12}$$

for some $s_* \geq 0$. Let Δ_* denote the triangle with top left corner (j_*, l_*) and size s_*. If $(j,l) \in \Delta_*$, then by (7.10), (7.7) we have

$$|\theta(j,l)| \leq 3^{2(j-j_*)}2^{(l_*-l)}\varepsilon \exp(-s_*) \leq \varepsilon$$

and hence every element of Δ_* is black (and lies in $[\frac{n}{2} - c\log\frac{1}{\varepsilon}] \times \mathbb{Z}$).

Next, we claim that every point $(j',l') \in [n/2] \times \mathbb{Z}$ that lies outside of Δ_*, but is at a distance of at most $c_0 \log\frac{1}{\varepsilon}$ to Δ_*, is white if $c_0 > 0$ is a small enough absolute constant (not depending on ε). We divide into three cases:

Case 1: $j' \geq j_*, l' \leq l_*$. In this case we have from (7.10) that

$$s_* < (j' - j_*) \log 9 + (l_* - l') \log 2 \leq s_* + O\left(c_0 \log \frac{1}{\varepsilon}\right)$$

or equivalently

$$\exp(s_*) < 3^{2(j'-j_*)} 2^{(l_*-l')} \leq \exp(s_*) \varepsilon^{-O(c_0)}.$$

Combining this with (7.12), (7.7) we conclude (for c_0 small enough) that

$$\varepsilon < |\theta(j', l')| \leq \varepsilon^{1-O(c_0)} < \frac{1}{2}$$

and thus (j', l') is white as claimed.

Case 2: $j' \geq j_*, l' > l_*$. In this case we have from (7.10) that

$$l_* < l' \leq l_* + O\left(c_0 \log \frac{1}{\varepsilon}\right) \tag{7.13}$$

and

$$(j' - j_*) \log 9 \leq s_* + O\left(c_0 \log \frac{1}{\varepsilon}\right). \tag{7.14}$$

Suppose for contradiction that (j', l') was black, thus

$$|\theta(j', l')| \leq \varepsilon.$$

From (7.13), (7.7) this implies that

$$\theta(j', l_* + 1) \ll \varepsilon^{1-O(c_0)}.$$

On the other hand, from (7.12), (7.14) one has

$$\theta(j', l_*) \ll \varepsilon^{1-O(c_0)} \tag{7.15}$$

and hence by (7.7)

$$\theta(j', l_* + 1) = \frac{1}{2}\theta(j', l_*). \tag{7.16}$$

If $j' > j_*$, then from (7.12), (7.14) we also have

$$\theta(j' - 1, l_*) \ll \varepsilon^{1-O(c_0)}.$$

From (7.6), the identity

$$3^{2(j'-1)-2} 2^{-l_*} = 3^{2j'-2} 2^{-l_*} - 4 \times 3^{2(j'-1)-2} 2^{-l_*+1}$$

(which follows from the fact that $1 = 3^2 - 4 \times 2^1$) and the triangle inequality we have

$$|\theta(j' - 1, l_* + 1)| \ll |\theta(j', l_* + 1)| + |\theta(j' - 1, l_*)| \tag{7.17}$$

and thus from (7.15), (7.16) we have

$$\theta(j' - 1, l_* + 1) \ll \varepsilon^{1-O(c_0)}$$

and hence by (7.7)

$$\theta(j' - 1, l_* + 1) = \frac{1}{2}\theta(j' - 1, l_*).$$

One can iterate this argument to conclude that

$$\theta(j'', l_* + 1) = \frac{1}{2}\theta(j'', l_*)$$

for all $j_* \leq j'' \leq j'$. If $j \leq j'$, this implies in particular that

$$\theta(j, l_* + 1) = \frac{1}{2}\theta(j, l_*) \in [-\varepsilon/2, \varepsilon/2] \tag{7.18}$$

so that $(j, l_* + 1)$ is black, contradicting the construction of l_*. Now suppose that $j > j'$. Then from (7.15), (7.16) we have

$$\theta(j' + 1, l_* + 1) \ll \varepsilon^{1-O(c_0)}$$

which implies from (7.7) that

$$\theta(j' + 1, l_* + 1) = \frac{1}{2}\theta(j' + 1, l_*).$$

Iterating this we obtain (7.18), and obtain a contradiction as before.

Case 3: $j' < j_*$**.** Clearly this implies $j_* > 1$; also, from (7.10) we have

$$-O\left(c_0 \log \frac{1}{\varepsilon}\right) \leq (l_* - l') \log 2 \leq s_* + O\left(c_0 \log \frac{1}{\varepsilon}\right). \tag{7.19}$$

and

$$0 < j_* - j' \ll c_0 \log \frac{1}{\varepsilon}. \tag{7.20}$$

Suppose for contradiction that (j', l') was black, thus

$$|\theta(j', l')| \leq \varepsilon.$$

From (7.20), (7.7) we thus have

$$|\theta(j_* - 1, l')| \ll \varepsilon^{1-O(c_0)}. \tag{7.21}$$

If $l' \geq l_*$, then from (7.19), (7.7) we then have

$$|\theta(j_* - 1, l_*)| \ll \varepsilon^{1-O(c_0)}$$

and hence

$$\theta(j_* - 1, l_*) = \frac{1}{3}\theta(j_*, l_*) \in [-\varepsilon/3, \varepsilon/3] \tag{7.22}$$

so that $(j_* - 1, l_*)$ is black, contradicting the construction of j_*.

Now suppose that $l' < l_*$. Then from (7.21), (7.7) we have

$$\theta(j_* - 1, l') = \frac{1}{3}\theta(j_*, l').$$

Since (j_*, l_*) is black, we also have from (7.19), (7.7) that

$$\theta(j_*, l') \ll \varepsilon^{1-O(c_0)}$$

and

$$\theta(j_*, l' - 1) = \frac{1}{2}\theta(j_*, l').$$

From (7.17) (and some relabeling) we then have

$$\theta(j_* - 1, l' - 1) \ll \varepsilon^{1-O(c_0)}$$

and thus by (7.7)

$$\theta(j_* - 1, l' - 1) = \frac{1}{3}\theta(j_*, l' - 1).$$

Iterating this argument we eventually obtain (7.22), and again obtain the desired contradiction. This concludes the proof that all points outside Δ_* that lie within a distance $c_0 \log \frac{1}{\varepsilon}$ of Δ_* are white.

From the above claim and the construction of the functions $l_*(), j_*()$, we now see that for any $(j', l') \in \Delta_*$, that $l_*(j', l') = l_*$ and $j_*(j', l') = j_*$. In other words, we have

$$\Delta_* = \{(j', l') \in B : l_*(j', l') = l_*; j_*(j', l') = j_*\},$$

and so the triangles Δ_* form a partition of B. By preceding we see that these triangles lie in $[\frac{n}{2} - c \log \frac{1}{\varepsilon}] \times \mathbb{Z}$ and are separated from each other by at least $c_0 \log \frac{1}{\varepsilon}$. This proves the lemma. $\square$

Remark 7.3. One can say a little bit more about the structure of the black set B; for instance, from Euler's theorem we see that B is periodic with respect to the vertical shift $(0, 2 \times 3^{n-1})$ (cf. Lemma 1.12), and one could use Baker's theorem [2] that (among other things) establishes a Diophantine property of $\frac{\log 3}{\log 2}$ in order to obtain some further control on B. However, we will not exploit any further structure of the black set in our arguments beyond what is provided by Lemma 7.2.

We now return to the probabilistic portion of the proof of (7.9). Currently we have a finite sequence $\mathbf{b}_1, \dots, \mathbf{b}_{\lfloor n/2 \rfloor}$ of random variables that are iid copies of the sum $\mathbf{a}_1 + \mathbf{a}_2$ of two independent copies $\mathbf{a}_1, \mathbf{a}_2$ of $\mathbf{Geom}(2)$. We may extend this sequence to an infinite sequence $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \dots$ of iid copies of $\mathbf{a}_1 + \mathbf{a}_2$. The left-hand side of (7.9) can then be written as

$$\mathbb{E} \exp(-\varepsilon^3 \# \{j \in \mathbb{N} + 1 : \mathbf{b}_j = 3, (j, \mathbf{b}_{[1,j]}) \in W\})$$

since $W \subset [n/2] \times \mathbb{Z}$, so $(j, \mathbf{b}_{[1,j]})$ can only lie in W when $j \in [n/2]$.

We now describe the random set $\{(j, \mathbf{b}_{[1,j]}) : j \in \mathbb{N} + 1, \mathbf{b}_j = 3\}$ as[4] a two-dimensional renewal process. Since the events $\mathbf{b}_j = 3$ are independent and each occur with probability

$$\mathbb{P}(\mathbf{b}_j = 3) = \mathbb{P}(\mathbf{Pascal} = 3) = \frac{1}{4} > 0, \tag{7.23}$$

we see that almost surely one has $\mathbf{b}_j = 3$ for at least one $j \in \mathbb{N}$. Define the *two-dimensional holding time* $\mathbf{Hold} \in (\mathbb{N} + 1) \times (\mathbb{N} + 2)$ to be the random shift $(\mathbf{j}, \mathbf{b}_{[1,\mathbf{j}]})$, where $\mathbf{j}$ is the least positive integer for which $\mathbf{b}_\mathbf{j} = 3$; this random variable is almost surely well defined. Note from (7.23) that the first component $\mathbf{j}$ of $\mathbf{Hold}$ has the distribution $\mathbf{j} \equiv \mathbf{Geom}(4)$. A little thought then reveals that the random set

$$\{(j, \mathbf{b}_{[1,j]}) : j \in \mathbb{N} + 1, \mathbf{b}_j = 3\} \tag{7.24}$$

has the same distribution as the random set

$$\{\mathbf{v}_1, \mathbf{v}_{[1,2]}, \mathbf{v}_{[1,3]}, \dots\}, \tag{7.25}$$

where $\mathbf{v}_1, \mathbf{v}_2, \dots$ are iid copies of $\mathbf{Hold}$, and we extend the summation notation (1.6) to the $\mathbf{v}_j$. In particular, we have

$$\# \{j \in \mathbb{N} + 1 : \mathbf{b}_j = 3, (j, \mathbf{b}_{[1,j]}) \in W\} \equiv \# \{k \in \mathbb{N} + 1 : \mathbf{v}_{[1,k]} \in W\},$$

[4]We are indebted to Marek Biskup for this suggestion.

and so we can write the left-hand side of (7.9) as

$$\mathbb{E} \prod_{k \in \mathbb{N}+1} \exp(-\varepsilon^3 1_W(\mathbf{v}_{[1,k]}));$$

note that all but finitely many of the terms in this product are equal to 1.

For any $(j,l) \in \mathbb{N}+1 \times \mathbb{Z}$, let $Q(j,l)$ denote the quantity

$$Q(j,l) := \mathbb{E} \prod_{k \in \mathbb{N}} \exp(-\varepsilon^3 1_W((j,l) + \mathbf{v}_{[1,k]})) \tag{7.26}$$

then we have the recursive formula

$$Q(j,l) = \exp(-\varepsilon^3 1_W(j,l)) \mathbb{E} Q((j,l) + \mathbf{Hold}) \tag{7.27}$$

and the right-hand side of (7.9) is equal to

$$\mathbb{E} Q(\mathbf{Hold}).$$

One can think of $Q(j,l)$ as a quantity controlling how often one encounters white points when one walks along a two-dimensional renewal process $(j,l), (j,l)+\mathbf{v}_1, (j,l)+\mathbf{v}_{[1,2]}, \dots$ starting at (j,l) with holding times given by iid copies of **Hold**. The smaller this quantity is, the more white points one is likely to encounter. The main difficulty is thus to ensure that this renewal process is usually not trapped within the black triangles Δ from Lemma 7.2; as it turns out (and as may be evident from an inspection of Figure 2), the large triangles will be the most troublesome to handle (as they are so large compared to the narrow band of white points surrounding them that are provided by Lemma 7.2).

We now pause to record some basic properties about the distribution of **Hold**.

Lemma 7.4 (Basic properties of holding time)**.** *The random variable* **Hold** *has exponential tail (in the sense of* (2.2)*), is not supported in any coset of any proper subgroup of* $\mathbb{Z}^2$*, and has mean* $(4,16)$*. In particular, the conclusion of Lemma 2.2 holds for* **Hold** *with* $\vec{\mu} := (4,16)$*.*

Proof. From the definition of **Hold** and (7.23), we see that **Hold** is equal to $(1,3)$ with probability $1/4$, and on the remaining event of probability $3/4$, it has the distribution of $(1, \mathbf{Pascal}') + \mathbf{Hold}'$, where $\mathbf{Pascal}'$ is a copy of **Pascal** that is conditioned to the event $\mathbf{Pascal} \neq 3$, so that

$$\mathbb{P}(\mathbf{Pascal}' = b) = \frac{4}{3}\frac{b-1}{2^b} \tag{7.28}$$

for $b \in \mathbb{N}+2 \backslash \{3\}$, and $\mathbf{Hold}'$ is a copy of **Hold** that is independent of $\mathbf{Pascal}'$. Thus **Hold** has the distribution of $(1,3) + (1,\mathbf{b}'_1) + \dots + (1,\mathbf{b}'_{\mathbf{j}-1})$, where $\mathbf{b}'_1, \mathbf{b}'_2, \dots$ are iid copies of $\mathbf{Pascal}'$ and $\mathbf{j} \equiv \mathbf{Geom}(4)$ is independent of the $\mathbf{b}'_j$. In particular, for any $k = (k_1,k_2) \in \mathbb{R}^2$, one has from monotone convergence that

$$\mathbb{E}\exp(\mathbf{Hold} \cdot k) = \sum_{j=\mathbb{N}} \frac{1}{4}\left(\frac{3}{4}\right)^{j-1} \exp\left((1,3)\cdot k\right)\left(\mathbb{E}\exp((1,\mathbf{Pascal}')\cdot k)\right)^j. \tag{7.29}$$

From (7.28) and dominated convergence, we have $\mathbb{E}\exp((1,\mathbf{Pascal}')\cdot k) < \frac{4}{3}$ for k sufficiently close to 0, which by (7.29) implies that $\mathbb{E}\exp(\mathbf{Hold}\cdot k) < \infty$ for k sufficiently close to zero. This gives the exponential tail property by Markov's inequality.

Since $\mathbf{Hold}$ attains the value $(1,3)+(1,b)$ for any $b \in \mathbb{N}+2\backslash\{3\}$ with positive probability, we see that the support of $\mathbf{Hold}$ is not supported in any coset of any proper subgroup of $\mathbb{Z}^2$. Finally, from the description of $\mathbf{Hold}$ at the start of this proof we have

$$\mathbb{E}\mathbf{Hold} = \frac{1}{4}(1,3) + \frac{3}{4}\left((1,\mathbb{E}\mathbf{Pascal}') + \mathbb{E}\mathbf{Hold}\right);$$

also, from the definition of $\mathbf{Pascal}'$ we have

$$\mathbb{E}\mathbf{Pascal} = \frac{1}{4}3 + \frac{3}{4}\mathbb{E}\mathbf{Pascal}'$$

We conclude that

$$\mathbb{E}\mathbf{Hold} = (1,\mathbb{E}\mathbf{Pascal}) + \frac{3}{4}\mathbb{E}\mathbf{Hold};$$

since $\mathbb{E}\mathbf{Pascal} = 2\mathbb{E}\mathbf{Geom}(2) = 4$, we thus have $\mathbb{E}\mathbf{Hold} = (4,16)$ as required. □

This lemma allows us to control the distribution of first passage locations of renewal processes with holding times $\equiv \mathbf{Hold}$, which will be important for us as it lets us understand how such renewal processes exit a given triangle Δ:

Lemma 7.5 (Distribution of first passage location)**.** *Let $\mathbf{v}_1, \mathbf{v}_2, \dots$ be iid copies of $\mathbf{Hold}$, and write $\mathbf{v}_k = (\mathbf{j}_k, \mathbf{l}_k)$. Let $s \in \mathbb{N}$, and define the first passage time $\mathbf{k}$ to be the least positive integer such that $\mathbf{l}_{[1,k]} > s$. Then for any $j, l \in \mathbb{N}$ with $l > s$, one has*

$$\mathbb{P}(\mathbf{v}_{[1,\mathbf{k}]} = (j,l)) \ll \frac{e^{-c(l-s)}}{(1+s)^{1/2}} G_{1+s}\left(c(j-\frac{s}{4})\right).$$

Informally, this lemma asserts that as a rough first approximation one has

$$\mathbf{v}_{[1,\mathbf{k}]} \approx \mathbf{Unif}\left(\left\{(j,l) : j = \frac{s}{4} + O((1+s)^{1/2}); s < l \leq s + O(1)\right\}\right). \tag{7.30}$$

Proof. Note that by construction of $\mathbf{k}$ one has $\mathbf{l}_{[1,\mathbf{k}]} - \mathbf{l}_{\mathbf{k}} \leq s$, so that $\mathbf{l}_{\mathbf{k}} \geq \mathbf{l}_{[1,\mathbf{k}]} - s$. From the union bound, we therefore have

$$\mathbb{P}(\mathbf{v}_{[1,\mathbf{k}]} = (j,l)) \leq \sum_{k \in \mathbb{N}+1} \mathbb{P}(\mathbf{v}_{[1,k]} = (j,l) \wedge \mathbf{l}_k \geq l - s);$$

since $\mathbf{v}_k$ has the exponential tail and is independent of $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}$, we thus have

$$\mathbb{P}(\mathbf{v}_{[1,\mathbf{k}]} = (j,l)) \ll \sum_{k \in \mathbb{N}+1} \sum_{l_k \geq l-s} \sum_{j_k \in \mathbb{N}+1} e^{-c(j_k+l_k)} \mathbb{P}(\mathbf{v}_{[1,k-1]} = (j-j_k, l-l_k)).$$

Writing $l_k = l - s + l'_k$, we then have

$$\mathbb{P}(\mathbf{v}_{[1,\mathbf{k}]} = (j,l)) \ll e^{-c(l-s)} \sum_{k \in \mathbb{N}+1} \sum_{l'_k \in \mathbb{N}} \sum_{j_k \in \mathbb{N}+1} e^{-c(j_k+l'_k)} \mathbb{P}(\mathbf{v}_{[1,k-1]} = (j-j_k, s-l'_k)).$$

We can restrict to the region $l'_k \leq s$, since the summand vanishes otherwise. It now suffices to show that

$$\sum_{k\in\mathbb{N}+1}\sum_{0\leq l'_k\leq s}\sum_{j_k\in\mathbb{N}+1} e^{-c(j_k+l'_k)}\mathbb{P}\left(\mathbf{v}_{[1,k-1]}=(j-j_k,s-l'_k)\right) \ll (1+s)^{-1/2}G_{1+s}\left(c(j-\frac{s}{4})\right). \tag{7.31}$$

This is in turn implied by

$$\sum_{k\in\mathbb{N}+1}\sum_{0\leq l'_k\leq s} e^{-cl'_k}\mathbb{P}(\mathbf{v}_{[1,k-1]}=(j',s-l'_k)) \ll (1+s)^{-1/2}G_{1+s}\left(c(j'-\frac{s}{4})\right) \tag{7.32}$$

for all $j'\in\mathbb{Z}$, since (7.31) then follows by replacing j' by $j-j_k$, multiplying by $\exp(-cj_k)$, and summing in j_k (and adjusting the constants c appropriately). In a similar vein, it suffices to show that

$$\sum_{k\in\mathbb{N}+1}\mathbb{P}(\mathbf{v}_{[1,k-1]}=(j',s')) \ll (1+s')^{-1/2}G_{1+s'}\left(c(j'-\frac{s'}{4})\right)$$

for all $s'\in\mathbb{N}$, since (7.32) follows after setting $s'=s-l'_k$, multiplying by $\exp(-cl'_k)$, and summing in l'_k (splitting into the regions $l'_k\leq s/2$ and $l'_k>s/2$ if desired to simplify the calculations).

From Lemma 7.4 and Lemma 2.2 one has

$$\mathbb{P}(\mathbf{v}_{[1,k-1]}=(j',s')) \ll k^{-1}G_{k-1}\left(c((j',s')-(k-1)(4,16))\right),$$

and the claim now follows from summing in k and a routine calculation (splitting for instance into the regions $16(k-1)\in[s'/2,2s']$, $16(k-1)<s'/2$, and $16(k-1)>2s'$). □

For any $m\in[n/2]$, let Q_m denote the quantity

$$Q_m := \sup_{(j,l)\in\mathbb{N}+1\times\mathbb{Z}:j\geq\lfloor n/2\rfloor-m}\max(\lfloor n/2\rfloor-j,1)^A Q(j,l). \tag{7.33}$$

Clearly we have $Q_m\leq m^A$, since $Q(j,l)\leq 1$ for all j,l. We will shortly establish the monotonicity formula

$$Q_m\leq Q_{m-1} \tag{7.34}$$

whenever $C_{A,\varepsilon}\leq m\leq n/2$ for some sufficiently large $C_{A,\varepsilon}$ depending on A,ε. Assuming this, we conclude that $Q_m\ll_A 1$ for all $1\leq m\leq n/2$. In particular, we have

$$Q(\mathbf{Hold})\ll_A \max(\lfloor n/2\rfloor-\mathbf{j},1)^{-A}\ll_A n^{-A}\mathbf{j}^A$$

where $\mathbf{j}\equiv\mathbf{Geom}(4)$ is the first component of $\mathbf{Hold}$. As $\mathbf{Geom}(4)$ has exponential tail, we conclude that

$$\mathbb{E}Q(\mathbf{Hold})\ll_A n^{-A},$$

which gives (7.9).

It remains to establish (7.34). Let $C_A\leq m\leq n/2$ for some sufficiently large C_A. It suffices to show that

$$Q(j,l)\leq m^{-A}Q_{m-1} \tag{7.35}$$

whenever $j = \lfloor n/2 \rfloor - m$ and $l \in \mathbb{Z}$. We divide into three cases. Let $\mathcal{T}$ be the family of triangles from Lemma 7.2.

Case 1: $(j,l) \in W$**.** This is the easiest case, as one can immediately get a gain from the white point (j,l). From (7.27) we have

$$Q(j,l) = \exp(-\varepsilon^3)\mathbb{E}Q((j,l) + \mathbf{Hold})$$

and hence by (7.33) and the fact that the first component of **Hold** has the distribution of **Geom**(4), we have

$$Q(j,l) \le \exp(-\varepsilon^3)Q_{m-1}\mathbb{E}\max(m - \mathbf{Geom}(4), 1)^{-A}.$$

We can bound

$$\max(m-r,1)^{-1} \le m \exp\left(O\left(\frac{r\log m}{m}\right)\right) \tag{7.36}$$

for any $r \in \mathbb{N}+1$, and hence and thus

$$Q(j,l) \le \exp(-\varepsilon^3)m^{-A}Q_{m-1}\mathbb{E}\exp\left(O\left(\frac{A\log m}{m}\mathbf{Geom}(4)\right)\right).$$

For m large enough depending on A, ε, we then have

$$Q(j,l) \le \exp(-\varepsilon^3/2)m^{-A}Q_{m-1} \tag{7.37}$$

which gives (7.35) in this case (with some room to spare).

Case 2: $(j,l) \in \Delta$ **for some triangle** $\Delta \in \mathcal{T}$**, and** $l \ge l_\Delta - \frac{m}{\log^2 m}$**.** This case is slightly harder than the preceding one, as one has to walk randomly through the triangle Δ before one has a good chance to encounter a white point, but because this portion of the walk is relatively short, the degradation of the weight m^{-A} during this portion will be negligible.

Set $s := l_\Delta - l$, thus $0 \le s \le \frac{m}{\log^2 m}$. Let $\mathbf{v}_1, \mathbf{v}_2, \dots$ be iid copies of **Hold**, write $\mathbf{v}_k = (\mathbf{j}_k, \mathbf{l}_k)$ for each k with the usual summation notations (1.6), and define the first passage time $\mathbf{k} \in \mathbb{N}+1$ to be the least positive integer such that

$$\mathbf{l}_{[1,\mathbf{k}]} > s. \tag{7.38}$$

This is a finite random variable since the $\mathbf{l}_k$ are all positive integers. By iterating (7.27) appropriately (or using (7.26)), we have the identity

$$Q(j,l) = \mathbb{E}\exp\left(-\varepsilon^3\sum_{i=0}^{\mathbf{k}-1} 1_W((j,l) + \mathbf{v}_{[1,i]})\right)Q((j,l) + \mathbf{v}_{[1,\mathbf{k}]}) \tag{7.39}$$

and hence by (7.35), (7.37)

$$Q(j,l) \le Q_{m-1}\mathbb{E}\exp\left(-\frac{\varepsilon^3}{2}1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}]})\right)\max(m - \mathbf{j}_{[1,\mathbf{k}]}, 1)^{-A}$$

which by (7.36) gives

$$Q(j,l) \le m^{-A}Q_{m-1}\mathbb{E}\exp\left(-\frac{\varepsilon^3}{2}1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}]})\right)\exp\left(O\left(\frac{A\log m}{m}\mathbf{j}_{[1,\mathbf{k}]}\right)\right).$$

It thus suffices to show that

$$\mathbb{E}\exp\left(-\frac{\varepsilon^3}{2}1_W((j,l)+\mathbf{v}_{[1,\mathbf{k}]})\right)\exp\left(O\left(\frac{A\log m}{m}\mathbf{j}_{[1,\mathbf{k}]}\right)\right)\leq 1. \tag{7.40}$$

Since $\exp(-\varepsilon^3/2)\leq 1-\varepsilon^3/4$, we can bound the left-hand side by

$$\mathbb{E}\exp\left(O\left(\frac{A\log m}{m}\mathbf{j}_{[1,\mathbf{k}]}\right)\right)-\frac{\varepsilon^3}{4}\mathbb{P}((j,l)+\mathbf{v}_{[1,\mathbf{k}]}\in W). \tag{7.41}$$

By definition, the first passage location $(j,l)+\mathbf{v}_{[1,\mathbf{k}]}$ takes values in the region $\{(j',l')\in\mathbb{Z}^2 : j'>j, l'>l_\Delta\}$. From Lemma 7.5 we have

$$\mathbb{P}((j,l)+\mathbf{v}_{[1,\mathbf{k}]}=(j',l'))\ll\frac{e^{-c(l'-l_\Delta)}}{(1+s)^{1/2}}G_{1+s}\left(c(j'-j-\frac{s}{4})\right). \tag{7.42}$$

Summing in l', we conclude that

$$\mathbb{P}(\mathbf{j}_{[1,\mathbf{k}]}=j'-j)\ll(1+s)^{-1/2}G_{1+s}\left(c(j'-j-\frac{s}{4})\right)$$

for any j'; with $s\leq\frac{m}{\log^2 m}$, a routine calculation then gives the bound

$$\mathbb{E}\exp\left(O\left(\frac{A\log m}{m}\mathbf{j}_{[1,\mathbf{k}]}\right)\right)\leq 1+O\left(\frac{As\log m}{m}\right)=1+O\left(\frac{A}{\log m}\right). \tag{7.43}$$

Using (7.42) to handle all points (j',l') outside the region $l'=l_\Delta+O(1)$ and $j'=j+\frac{s}{4}+O((1+s)^{1/2})$, we have

$$\mathbb{P}\left((j,l)+\mathbf{v}_{[1,\mathbf{k}]}=\left(j+\frac{s}{4}+O((1+s)^{1/2}),l_\Delta+O(1)\right)\right)\gg 1 \tag{7.44}$$

(cf. (7.30)). On the other hand, since $(j,l)\in\Delta$ and $s=l_\Delta-l$, we have from (7.10) that

$$0\leq(j-j_\Delta)\log 9\leq s_\Delta-s\log 2$$

and thus (since $0<\frac{1}{4}\log 9<\log 2$) one has

$$-O(1)\leq(j'-j_\Delta)\log 9\leq s_\Delta+O(1)$$

whenever $j'=j+\frac{s}{4}+O((1+s)^{1/2})$. We conclude that with probability $\gg 1$, the first passage location $(j,l)+\mathbf{v}_{[1,\mathbf{k}]}$ lies outside of Δ, but at a distance $O(1)$ from Δ, hence is white by Lemma 7.2. We conclude that

$$\mathbb{P}((j,l)+\mathbf{v}_{[1,\mathbf{k}]}\in W)\gg 1 \tag{7.45}$$

and (7.40) now follows from (7.41), (7.43), (7.45).

Case 3: $(j,l)\in\Delta$ for some triangle $\Delta\in\mathcal{T}$, and $j<j_\Delta-\frac{m}{\log^2 m}$. This is the most difficult case, as one has to walk so far before exiting Δ that one needs to encounter multiple white points, not just a single white point, in order to counteract the degradation of the weight m^{-A}. Fortunately, the number of white points one needs to encounter is $O_{A,\varepsilon}(1)$, and we will be able to locate such a number of white points on average for m large enough.

We will need a large constant P (much larger than A or $1/\varepsilon$, but much smaller than m) depending on A,ε to be chosen later. As before, we set $s:=l_\Delta-l$, so now $s>\frac{m}{\log^2 m}$.

Since we have $j_\Delta \leq \lfloor n/2 \rfloor = j+m$, we also have $s \leq m$. We again let $\mathbf{v}_1, \mathbf{v}_2, \dots$ be iid copies of $\mathbf{Hold}$, write $\mathbf{v}_k = (\mathbf{j}_k, \mathbf{l}_k)$ for each k, and define the first passage time $\mathbf{k} \in \mathbb{N}+1$ to be the least positive integer such that (7.38) holds. From (7.39) we have

$$Q(j,l) \leq \mathbb{E} Q((j,l) + \mathbf{v}_{[1,\mathbf{k}]}).$$

Applying (7.27) we then have

$$Q(j,l) \leq \mathbb{E} \exp\left(-\varepsilon^3 \sum_{p=0}^{P-1} 1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}) \right) Q((j,l) + \mathbf{v}_{[1,\mathbf{k}+P]})$$

and hence by (7.33)

$$Q(j,l) \leq m^{-A} Q_{m-1} \mathbb{E} \exp\left(-\varepsilon^3 \sum_{p=0}^{P-1} 1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}) \right) \max\left(1 - \frac{\mathbf{j}_{[1,\mathbf{k}+P]}}{m}, \frac{1}{m}\right)^{-A}.$$

Thus it suffices to show that

$$\mathbb{E} \exp\left(-\varepsilon^3 \sum_{p=0}^{P-1} 1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}) \right) \max\left(1 - \frac{\mathbf{j}_{[1,\mathbf{k}+P]}}{m}, \frac{1}{m}\right)^{-A} \leq 1. \tag{7.46}$$

Let us first consider the event that $\mathbf{j}_{[1,\mathbf{k}+P]} \geq \frac{1}{2} m$. From Lemma 7.5 and the bound $s \leq m$, we have

$$\mathbb{P}(\mathbf{j}_{[1,\mathbf{k}]} \geq 0.4 m) \ll \exp(-cm)$$

(noting that $0.4 > \frac{1}{4}$) while from Lemma 2.2 (recalling that the $\mathbf{j}_k$ are iid copies of $\mathbf{Geom}(4)$) we have

$$\mathbb{P}(\mathbf{j}_{[\mathbf{k}+1,\mathbf{k}+P]} \geq 0.1 m) \ll_P \exp(-cm)$$

and thus by the triangle inequality

$$\mathbb{P}(\mathbf{j}_{[1,\mathbf{k}+P]} \geq \frac{1}{2} m) \ll_P \exp(-cm).$$

Thus the contribution of this case to (7.46) is $O_{P,A}(m^A \exp(-cm)) = O_{P,A}(\exp(-cm/2))$. If instead we have $\mathbf{j}_{[1,\mathbf{k}+P]} < \frac{1}{2} m$, then

$$\max\left(1 - \frac{\mathbf{j}_{[1,\mathbf{k}+P]}}{m}, \frac{1}{m}\right)^{-A} \leq 2^A.$$

Since m is large compared to A, P, it thus suffices to show that

$$\mathbb{E} \exp\left(-\varepsilon^3 \sum_{p=0}^{P-1} 1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}) \right) \leq 2^{-A-1}$$

which will in turn be implied by the bound

$$\mathbb{P}\left(\sum_{p=0}^{P-1} 1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}) \leq \frac{10A}{\varepsilon^3} \right) \leq 2^{-A-2} \tag{7.47}$$

Roughly speaking, the estimate (7.47) asserts that once one exits the large triangle Δ then one should almost always encounter at least $10A/\varepsilon^3$ white points by a certain time $P = O_{A,\varepsilon}(1)$. To prove this estimate we introduce another random statistic that measures the number of triangles that one encounters on an infinite two-dimensional renewal process $(j,l), (j,l)+\mathbf{v}_1, (j,l)+\mathbf{v}_{[1,2]}, \dots$, where $\mathbf{v}_1, \mathbf{v}_2, \dots$ are iid copies of $\mathbf{Hold}$. Given an initial point $(j,l) \in \mathbb{N}+1 \times \mathbb{Z}$, we recursively introduce the stopping times

$\mathbf{t}_1 = \mathbf{t}_1(j,l), \dots, \mathbf{t}_\mathbf{r} = \mathbf{t}_{\mathbf{r}(j,l)}(j,l)$ by defining $\mathbf{t}_1$ to be the first natural number (if it exists) for which $(j,l) + \mathbf{v}_{[1,\mathbf{t}_1]}$ lies in a triangle $\mathbf{\Delta}_1 \in \mathcal{T}$, then for each $i > 1$, defining $\mathbf{t}_i$ to be the first natural number (if it exists) with $l + \mathbf{l}_{[1,\mathbf{t}_i]} > l_{\mathbf{\Delta}_{i-1}}$ and $(j,l) + \mathbf{v}_{[1,\mathbf{t}_i]}$ lies in a triangle $\mathbf{\Delta}_i \in \mathcal{T}$. We set $\mathbf{r} = \mathbf{r}(j,l)$ to be the number of stopping times that can be constructed in this fashion (thus, there are no natural numbers k with $l + \mathbf{l}_{[1,k]} > l_{\mathbf{\Delta}_\mathbf{r}}$ and $(j,l) + \mathbf{v}_{[1,k]}$ black). Note that $\mathbf{r}$ is finite since the process $(j,l) + \mathbf{v}_{[1,k]}$ eventually exits the strip $[n/2] \times \mathbb{Z}$ when k is large enough, at which point it no longer encounters any black triangles.

The key estimate relating $\mathbf{r}$ with the expression in (7.47) is then

Lemma 7.6 (Many triangles usually implies many white points)**.** *Let* $\mathbf{v}_1, \mathbf{v}_2, \dots$ *be iid copies of* **Hold**. *Then for any* $(j,l) \in \mathbb{N}+1 \times \mathbb{Z}$ *and any positive integer* R*, we have*

$$\mathbb{E} \exp\left(- \sum_{p=1}^{\mathbf{t}_{\min(\mathbf{r},R)}} 1_W((j,l) + \mathbf{v}_{[1,p]}) + \varepsilon \min(\mathbf{r}, R) \right) \leq \exp(\varepsilon). \tag{7.48}$$

Informally the estimate (7.48) asserts that when $\mathbf{r}$ is large (so that the renewal process $(j,l), (j,l) + \mathbf{v}_1, (j,l) + \mathbf{v}_{[1,2]}, \dots$ passes through many different triangles), then the quantity $\sum_{p=1}^{\mathbf{t}_{\min(\mathbf{r},R)}} 1_W((j,l) + \mathbf{v}_{[1,p]}$ is usually also large, implying that the same renewal process also visits many white points. This is basically due to the separation between triangles that is given by Lemma 7.2.

Proof. Denote the quantity on the left-hand side of (7.48) by $Z((j,l),R)$. We induct on R. The case $R = 1$ is trivial, so suppose $R \geq 2$ and that we have already established that

$$Z((j',l'), R-1) \leq \exp(\varepsilon) \tag{7.49}$$

for all $(j',l') \in \mathbb{N}+1 \times \mathbb{Z}$. If $\mathbf{r} = 0$ then we can bound

$$\exp\left(- \sum_{p=1}^{\mathbf{t}_{\min(\mathbf{r},R)}} 1_W((j,l) + \mathbf{v}_{[1,p]}) + \varepsilon \min(\mathbf{r}, R) \right) \leq 1.$$

Suppose that $\mathbf{r} \neq 0$, so that the first stopping time $\mathbf{t}_1$ and triangle $\mathbf{\Delta}_1$ exists. Let $\mathbf{k}_1$ be the first natural number for which $l + \mathbf{l}_{[1,\mathbf{k}_1]} > l_{\mathbf{T}_1}$; then $\mathbf{k}_1$ is well-defined (since we have an infinite number of $\mathbf{l}_k$, all of which are at least 2) and $\mathbf{k}_1 > \mathbf{t}_1$. The conditional expectation of $\exp(-\sum_{p=1}^{\mathbf{t}_{\min(\mathbf{r},R)}} 1_W((j,l) + \mathbf{v}_{[1,p]}) + \varepsilon \min(\mathbf{r}, R))$ relative to the random variables $\mathbf{v}_1, \dots, \mathbf{v}_{\mathbf{k}_1}$ is equal to

$$\exp\left(- \sum_{p=1}^{\mathbf{k}_1} 1_W((j,l) + \mathbf{v}_{[1,p]}) + \varepsilon \right) Z(1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}_1]}, R-1)$$

which we can upper bound using the inductive hypothesis (7.49) as

$$\exp\left(-1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}_1]}) + 2\varepsilon \right).$$

We thus obtain the inequality

$$Z((j,l),M) \leq \mathbb{P}(\mathbf{r} = 0) + \exp(2\varepsilon) \mathbb{E} 1_{\mathbf{r} \neq 0} \exp(-1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}_1]}))$$

so to close the induction it suffices to show that

$$\mathbb{E}1_{\mathbf{r}\neq 0}\exp(-1_W((j,l)+\mathbf{v}_{[1,\mathbf{k}_1]}))\leq \exp(-\varepsilon)\mathbb{P}(\mathbf{r}\neq 0)$$

which will be implied by

$$\mathbb{P}(\mathbf{r}\neq 0\wedge (j,l)+\mathbf{v}_{[1,\mathbf{k}_1]}\in W)\gg \mathbb{P}(\mathbf{r}\neq 0).$$

For each $p\in\mathbb{N}+1$, triangle $\Delta_1\in\mathcal{T}$, and $(j',l')\in\Delta_1$, let $E_{p,\Delta_1,(j',l')}$ denote the event that $(j,l)+\mathbf{v}_{[1,p]}=(j',l')$, and $(j,l)+\mathbf{v}_{[1,p']}\in W$ for all $1\leq p'<p$. Observe that the event $\mathbf{r}\neq 0$ is the disjoint of the events $E_{p,T_1,(j',l')}$. It therefore suffices to show that

$$\mathbb{P}(E_{p,\Delta_1,(j',l')}\wedge (j,l)+\mathbf{v}_{[1,\mathbf{k}_1]}\in W)\gg \mathbb{P}(E_{p,\Delta_1,(j',l')}).$$

We may of course assume that the event $E_{p,\Delta_1,(j',l')}$ occurs with non-zero probability. Conditioning to this event, we see that $(j,l)+\mathbf{v}_{[1,\mathbf{k}_1]}$ has the same distribution as (the unconditioned random variable) $(j',l')+\mathbf{v}_{[1,\mathbf{k}']}$, where the first passage time $\mathbf{k}'$ is the first natural number for which $l'+\mathbf{l}_{[1,\mathbf{k}']}>l_{\Delta_1}$. By repeating the proof of (7.45), one has

$$\mathbb{P}((j',l')+\mathbf{v}_{[1,\mathbf{k}']}\in W)\gg 1$$

and the claim follows. □

To use this bound we need to show that the renewal process $(j,l),(j,l)+\mathbf{v}_1,(j,l)+\mathbf{v}_{[1,2]},\ldots$ either passes through many white points, or through many triangles. This will be established via a probabilistic upper bound on the size s_Δ of the triangles encountered. The key lemma in this regard is

Lemma 7.7 (Large triangles are rarely encountered shortly after a lengthy crossing)**.** *Let $p\in\mathbb{N}$ and $1\leq s'\leq m^{0.4}$. Let $E_{p,s'}$ denote the event that $(j,l)+\mathbf{v}_{[1,\mathbf{k}+p]}$ lies in a triangle $\Delta'\in\mathcal{T}$ of size $s_{\Delta'}\geq s'$. Then*

$$\mathbb{P}(E_{p,s'})\ll A^2\frac{1+p}{s'}+\exp(-cA^2(1+p)).$$

Proof. We can assume that

$$s'\geq CA^2(1+p) \tag{7.50}$$

for a large constant C, since the claim is trivial otherwise.

From Lemma 7.5 we have (7.42) as before, so on summing in j' we have

$$\mathbb{P}(l+\mathbf{l}_{[1,k]}=l')\ll \exp(-c(l'-l_\Delta))$$

and thus

$$\mathbb{P}(l+\mathbf{l}_{[1,k]}\geq l_\Delta+A^2(1+p))\ll \exp(-cA^2(1+p)).$$

Similarly, from Lemma 2.2 one has

$$\mathbb{P}(\mathbf{l}_{[\mathbf{k}+1,\mathbf{k}+p]}\geq A^2(1+p))\ll \exp(-cA^2(1+p))$$

and thus

$$\mathbb{P}(l+\mathbf{l}_{[1,\mathbf{k}+p]}\geq l_\Delta+2A^2(1+p))\ll \exp(-cA^2(1+p)).$$

In a similar spirit, from (7.42) and summing in l' (recalling that $\frac{m}{\log^2 m}\leq s\leq m$) one has

$$\mathbb{P}(j+\mathbf{j}_{[1,\mathbf{k}]}=j')\ll s^{-1/2}G_s\left(c(j'-j-\frac{s}{4})\right)$$

so in particular

$$\mathbb{P}\left(\left|\mathbf{j}_{[1,\mathbf{k}]} - \frac{s}{4}\right| \geq s^{0.6}\right) \ll \exp(-cs^{0.2}) \ll A^2 \frac{1+p}{s'}$$

from the upper bound on s'. From Lemma 2.2 we also have

$$\mathbb{P}(|\mathbf{j}_{[\mathbf{k}+1,\mathbf{k}+p]}| \geq s^{0.6}) \ll \exp(-cs^{0.6}) \ll A^2 \frac{1+p}{s'}$$

and hence

$$\mathbb{P}\left(\left|\mathbf{j}_{[1,\mathbf{k}+p]} - \frac{s}{4}\right| \geq 2s^{0.6}\right) \ll A^2 \frac{1+p}{s'}$$

Thus, if E' denotes the event that $l + \mathbf{l}_{[1,\mathbf{k}+p]} \geq l_\Delta + 2A^2(1+p)$ or $|\mathbf{j}_{[1,\mathbf{k}+p]} - \frac{s}{4}| \geq 2s^{0.6}$, then

$$\mathbb{P}(E') \ll A^2 \frac{1+p}{s'} + \exp(-cA^2(1+p)). \tag{7.51}$$

Suppose now that we are outside the event E', and that $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ lies in a triangle Δ', thus

$$l + \mathbf{l}_{[1,\mathbf{k}+p]} = l_\Delta + O(A^2(1+p)) \tag{7.52}$$

and

$$\mathbf{j}_{[1,\mathbf{k}+p]} = \frac{s}{4} + O(s^{0.6}) = \frac{s}{4} + O(m^{0.6}). \tag{7.53}$$

From (7.10) we then have

$$0 \leq j + \mathbf{j}_{[1,\mathbf{k}+p]} - j_{\Delta'} \leq \frac{1}{\log 9} s_{\Delta'} - \frac{\log 2}{\log 9}(l_{\Delta'} - l - \mathbf{l}_{[1,\mathbf{k}+p]}).$$

Suppose that the lower tip of Δ' lies well below the upper edge of Δ in the sense that

$$l_{\Delta'} - \frac{s_{\Delta'}}{\log 2} \leq l_\Delta - 10.$$

Then by (7.52) we can find an integer $j' = j + \mathbf{j}_{[1,\mathbf{k}+p]} + O(A^2(1+p))$ such that $j' \geq j_{\Delta'}$ and

$$0 \leq j' - j_{\Delta'} \leq \frac{1}{\log 9} s_{\Delta'} - \frac{\log 2}{\log 9}(l_{\Delta'} - l_\Delta).$$

In other words, $(j', l_\Delta) \in \Delta'$. But by (7.53) we have

$$j' = j + \frac{s}{4} + O(m^{0.6}) + O(A^2(1+p)) = j + \frac{s}{4} + O(m^{0.6}).$$

From (7.10) we have

$$0 \leq (j - j_\Delta)\log 9 \leq s_\Delta - s\log 2$$

and hence (since $s \geq \frac{m}{\log^2 m}$ and $\frac{1}{4}\log 9 < \log 2$)

$$0 \leq (j' - j_\Delta)\log 9 \leq s_\Delta$$

Thus $(j', l_\Delta) \in \Delta$. Thus Δ and Δ' intersect, which by Lemma 7.2 forces $\Delta = \Delta'$, which is absurd since $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ lies in Δ' but not Δ (the l coordinate is larger than l_Δ). We conclude that

$$l_{\Delta'} - \frac{s_{\Delta'}}{\log 2} > l_\Delta - 10.$$

On the other hand, from (7.10) we have

$$l_{\Delta'} - \frac{s_{\Delta'}}{\log 2} \leq l + \mathbf{l}_{[1,\mathbf{k}+p]}$$

hence by (7.52) we have

$$l_{\Delta'} - \frac{s_{\Delta'}}{\log 2} = l_\Delta + O(A^2(1+p)). \tag{7.54}$$

From (7.10), (7.52) we then have

$$\begin{aligned} 0 \leq j + \mathbf{j}_{[1,\mathbf{k}+p]} - j_{\Delta'} &\leq \frac{1}{\log 9} s_{\Delta'} - \frac{\log 2}{\log 9}(l_{\Delta'} - l - \mathbf{l}_{[1,\mathbf{k}+p]}) \\ &= O(A^2(1+p)). \end{aligned}$$

so that

$$j + \mathbf{j}_{[1,\mathbf{k}+p]} = j_{\Delta'} + O(A^2(1+p)).$$

Thus, outside the event E', the event that $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ lies in a triangle Δ' can only occur if $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ lies within a distance $O(A^2(1+p))$ of the point $(j_{\Delta'}, l_\Delta)$.

Now suppose we have two distinct triangles Δ', Δ'' in $\mathcal{T}$ obeying (7.54), with $s_{\Delta'}, s_{\Delta''} \geq s'$ with $j_{\Delta'} \leq j_{\Delta''}$. Set $l_* := l_\Delta + \lfloor s'/2 \rfloor$, and observe from (7.10) that $(j_*, l_*) \in \Delta'$ whenever j_* lies in the interval

$$j_{\Delta'} \leq j_* \leq j_{\Delta'} + \frac{1}{\log 9} s_{\Delta'} - \frac{\log 2}{\log 9}(l_{\Delta'} - l_*)$$

and similarly $(j_*, l_*) \in \Delta''$ whenever

$$j_{\Delta''} \leq j_* \leq j_{\Delta''} + \frac{1}{\log 9} s_{\Delta''} - \frac{\log 2}{\log 9}(l_{\Delta''} - l_*).$$

By Lemma 7.2, these two intervals cannot have any integer point in common, thus

$$j_{\Delta'} + \frac{1}{\log 9} s_{\Delta'} - \frac{\log 2}{\log 9}(l_{\Delta'} - l_*) \leq j_{\Delta''}.$$

Applying (7.54) and the definition of l_*, we conclude that

$$j_{\Delta'} + \frac{1}{2}\frac{\log 2}{\log 9} s' + O(A^2(1+p)) \leq j_{\Delta''}$$

and hence by (7.50)

$$j_{\Delta''} - j_{\Delta'} \gg s'.$$

We conclude that for the triangles Δ' in $\mathcal{T}$ obeying (7.54) with $s_{\Delta'} \geq s'$, the points $(j_{\Delta'}, l_\Delta)$ are $\gg s'$-separated. Let Σ denote the collection of such points, thus Σ is a $\gg s'$-separated set of points, and outside of the event E', $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ can only occur in a triangle Δ' with $s_{\Delta'} \geq s'$ if

$$\operatorname{dist}((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}, \Sigma) \ll A^2(1+p).$$

From (7.51) we conclude that

$$\mathbb{P}(E_{p,s'}) \ll A^2 \frac{1+p}{s'} + \exp(-cA^2(1+p)) + \mathbb{P}\left(\operatorname{dist}((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}, \Sigma) \ll A^2(1+p)\right).$$

From (7.42) we see that

$$\begin{aligned}\mathbb{P}\left((j,l)+\mathbf{v}_{[1,\mathbf{k}+p]}=(j_{\Delta'},l_{\Delta})+O(A^2(1+p))\right)\\ \ll \frac{A^2(1+p)}{s^{1/2}}G_s\left(c(j_{\Delta'}-j-\frac{s}{4})\right)\\ \ll \frac{A^2(1+p)}{s'}\sum_{j'=j_{\Delta'}+O(s')}\frac{1}{s^{1/2}}G_s\left(c(j'-j-\frac{s}{4})\right).\end{aligned}$$

Summing andusing the $\gg s'$-separated nature of Σ, we conclude that

$$\begin{aligned}\mathbb{P}(\mathrm{dist}((j,l)+\mathbf{v}_{[1,\mathbf{k}+p]},\Sigma)\ll A^2(1+p))&\ll \frac{A^2(1+p)}{s'}\sum_{j'\in\mathbb{Z}}\frac{1}{s^{1/2}}G_{1+s}(c(j'-j-\frac{s}{4}))\\ &\ll \frac{A^2(1+p)}{s'}\end{aligned}$$

and the claim follows. □

From the above lemma we have

$$\mathbb{P}(E_{p,4^A(1+p)^3})\ll A^2\frac{1}{4^A(1+p)^2}+\exp(-cA^2(1+p))$$

whenever $0\leq p\leq m^{0.1}$. Thus by the union bound, if E_* denotes the union of the $E_{p,4^A(1+p)^3}$ for $0\leq p\leq m^{0.1}$, then

$$\mathbb{P}(E_*)\ll A^2 4^{-A}.$$

Next, we apply Lemma 7.6 with (j,l) replaced by $(j,l)+\mathbf{v}_{[1,\mathbf{k}]}$ to conclude that

$$\mathbb{E}\exp\left(-\sum_{p=1}^{\mathbf{t}_{\min(\mathbf{r},R)}}1_W((j,l)+\mathbf{v}_{[1,\mathbf{k}+p]}+\varepsilon\min(\mathbf{r},R)\right)\leq\exp(\varepsilon),$$

where now $\mathbf{r}=\mathbf{r}((j,l)+\mathbf{v}_{[1,\mathbf{k}]})$ and $\mathbf{t}_i=\mathbf{t}_i((j,l)+\mathbf{v}_{[1,\mathbf{k}]})$. By Markov's inequality we thus see that outside of an event F_* of probability

$$\mathbb{P}(F_*)\leq 2^{-A-2},$$

one has

$$\exp\left(-\sum_{p=1}^{\mathbf{t}_{\min(\mathbf{r},R)}}1_W((j,l)+\mathbf{v}_{[1,\mathbf{k}+p]}+\varepsilon\min(\mathbf{r},R)\right)\ll 2^A$$

which implies that

$$\sum_{p=1}^{\mathbf{t}_{\min(\mathbf{r},R)}}1_W((j,l)+\mathbf{v}_{[1,\mathbf{k}+p]})\gg\varepsilon\min(\mathbf{r},R)-O(A).$$

In particular, if we set $R:=\lfloor A^2/\varepsilon^4\rfloor$, we have

$$\sum_{p=1}^{\mathbf{t}_R}1_W((j,l)+\mathbf{v}_{[1,\mathbf{k}+p]})\gg\frac{A^2}{\varepsilon^3}\tag{7.55}$$

whenever we lie outside of F_* and $\mathbf{r}\geq R$.

Now suppose we lie outside of both E_* and F_*. To prove (7.47), it will now suffice to show the deterministic claim

$$\sum_{p=0}^{P-1} 1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}) > \frac{10A}{\varepsilon^3}. \tag{7.56}$$

Suppose this is not the case, thus

$$\sum_{p=0}^{P-1} 1_W((j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}) \leq \frac{10A}{\varepsilon^3}.$$

Thus the point $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ is white for at most $10A/\varepsilon^3$ values of $0 \leq p \leq P-1$, so in particular for P large enough there is $0 \leq p \leq 10A/\varepsilon^3 + 1 = O_{A,\varepsilon}(1)$ such that $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ is black. By Lemma 7.2, this point lies in a triangle $\Delta' \in \mathcal{T}$. As we are outside E_*, we have

$$s_{\Delta'} < 4^A(1+p)^3.$$

Thus by (7.10), for p' in the range

$$p + 10 \times 4^A(1+p)^3 < p' \leq P - 1,$$

we must have $l + \mathbf{l}_{[1,\mathbf{k}+p']} > l_{\Delta'}$, hence we exit Δ' (and increment the random variable $\mathbf{r}$). In particular, if

$$p + 10 \times 4^A(1+p)^3 + 10A/\varepsilon^3 + 1 \leq P - 1,$$

then we can find

$$p' \leq p + 10 \times 4^A(1+p)^3 + 10A/\varepsilon^3 + 1 = O_{p,A,\varepsilon}(1)$$

such that $l + \mathbf{l}_{[1,\mathbf{k}+p']} > l_{\Delta'}$ and $(j,l) + \mathbf{v}_{[1,\mathbf{k}+p]}$ is black (and therefore lies in a new triangle Δ''). Iterating this R times, we conclude (if P is sufficiently large depending on A, ε) that $\mathbf{r} \geq R$ and that $\mathbf{t}_R \leq P$. The claim (7.56) now follows from (7.55), giving the required contradiction. This (finally!) concludes the proof of (7.35), and hence Proposition 1.17 and Theorem 1.3.

References

[1] J.-P. Allouche, *Sur la conjecture de "Syracuse-Kakutani-Collatz"*, Séminaire de Théorie des Nombres, 1978–1979, Exp. No. 9, 15 pp., CNRS, Talence, 1979.

[2] A. Baker, *Linear forms in the logarithms of algebraic numbers. I*, Mathematika. A Journal of Pure and Applied Mathematics, **13** (1966), 204–216.

[3] J. Bourgain, *Periodic nonlinear Schrödinger equation and invariant measures*, Comm. Math. Phys. **166** (1994), 1–26.

[4] T. Carletti, D. Fanelli, *Quantifying the degree of average contraction of Collatz orbits*, Boll. Unione Mat. Ital. **11** (2018), 445–468.

[5] M. Chamberland, *A $3x+1$ survey: number theory and dynamical systems*, The ultimate challenge: the $3x+1$ problem, 57–78, Amer. Math. Soc., Providence, RI, 2010.

[6] I. Korec, *A density estimate for the $3x+1$ problem*, Math. Slovaca **44** (1994), no. 1, 85–89.

[7] A. Kontorovich, J. Lagarias, *Stochastic models for the $3x+1$ and $5x+1$ problems and related problems*, The ultimate challenge: the $3x+1$ problem, 131–188, Amer. Math. Soc., Providence, RI, 2010.

[8] I. Krasikov, J. Lagarias, *Bounds for the $3x+1$ problem using difference inequalities*, Acta Arith. **109** (2003), 237–258.

[9] J. Lagarias, *The 3x+1 problem and its generalizations*, Amer. Math. Monthly **92** (1985), no. 1, 3–23.
[10] J. Lagarias, K. Soundararajan, *Benford's law for the* $3x + 1$ *function*, J. London Math. Soc. (2) **74** (2006), no. 2, 289–303.
[11] J. Lagarias, A. Weiss, *The* $3x + 1$ *problem: two stochastic models*, Ann. Appl. Probab. **2** (1992), no. 1, 229–261.
[12] T. Oliveira e Silva, *Empirical verification of the 3x+1 and related conjectures*, The ultimate challenge: the $3x + 1$ problem, 189207, Amer. Math. Soc., Providence, RI, 2010.
[13] E. Roosendaal, `www.ericr.nl/wondrous`
[14] T. Tao, *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations*, Forum Math. Pi **4** (2016), e8, 36 pp.
[15] R. Terras, *A stopping time problem on the positive integers*, Acta Arith. **30** (1976), 241–252.
[16] R. Terras, *On the existence of a density*, Acta Arith. **35** (1979), 101–102.
[17] A. Thomas, *A non-uniform distribution property of most orbits, in case the* $3x + 1$ *conjecture is true*, Acta Arith. **178** (2017), no. 2, 125–134.
[18] G. Wirsching, The Dynamical System Generated by the $3n + 1$ Function, Lecture Notes in Math. No. 1681, Springer-Verlag: Berlin 1998.

UCLA Department of Mathematics, Los Angeles, CA 90095-1555.

E-mail address: `tao@math.ucla.edu`