Probab. Theory Relat. Fields (2011) 150:613–633
DOI 10.1007/s00440-010-0285-8

Monotone loop models and rational resonance

Alan Hammond · Richard Kenyon

Received: 28 March 2008 / Revised: 16 March 2009 / Published online: 7 May 2010

Abstract Let $T_{n,m} = \mathbb{Z}_n \times \mathbb{Z}_m$, and define a random mapping $\phi\colon T_{n,m} \to T_{n,m}$ by $\phi(x, y) = (x + 1, y)$ or $(x, y + 1)$ independently over x and y and with equal probability. We study the orbit structure of such "quenched random walks" ϕ in the limit $m, n \to \infty$, and show how it depends sensitively on the ratio m/n. For m/n near a rational p/q, we show that there are likely to be on the order of $\sqrt{n}$ cycles, each of length $O(n)$, whereas for m/n far from any rational with small denominator, there are a bounded number of cycles, and for typical m/n each cycle has length on the order of $n^{4/3}$.

Mathematics Subject Classification (2000) 60D05

1 Introduction

We study a model of monotone non-intersecting lattice paths in $\mathbb{Z}^2$. While this is a classically studied model in statistical mechanics, related to Dyson's Brownian motion and random matrices, there are few studies concerned with the influence of the boundary conditions at the critical point of the model. The authors of [3]

A. Hammond
Mathematics Department, New York University,
251 Mercer St., New York, NY 10012, USA

Present Address:
A. Hammond (✉)
Department of Statistics, Oxford University, 1 South Parks Road,
Oxford OX1 3TG, UK
e-mail: alanmhammond@gmail.com

R. Kenyon
Mathematics Department, Brown University,
151 Thayer Street, Providence, RI 02912, USA

studied a model in a similar setting, finding some surprising "resonance" phenomena, which showed how the shape of the domain—in particular the rationality of the aspect ratio—had a strong influence on the partition function and other observables in the model. Here we study a model very closely related to that in [3], on which we can obtain more accurate and complete information using simple probabilistic methods. One of our goals is to explain some of the conjectured behavior in [3]. However we feel that our model is of primary interest as an example of a quenched random dynamical system for which a fairly complete analysis can be obtained.

For positive integers m, n let $\Gamma_{n,m}$ be the sublattice of $\mathbb{Z}^2$ generated by $(n, 0)$ and $(0, m)$. Let $T_{n,m} = \mathbb{Z}^2/\Gamma_{n,m} \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$, the n, m-torus. We will consider configurations consisting of collections of vertex-disjoint monotone lattice cycles on $T_{n,m}$. The law on such configurations that we consider has two simple definitions.

Firstly, we define $\phi : T_{n,m} \to T_{n,m}$ by setting $\phi(x, y)$ to be equal, with equal probability, either to $(x + 1, y)$ or $(x, y + 1)$, these choices being made independently over the nm vertices of $T_{n,m}$. We call ϕ a *quenched random walk*. Each ϕ represents a dynamical system, with at least one periodic orbit. The law on disjoint unions of cycles that we consider is given by the collection of distinct orbits of this randomly selected ϕ.

The dynamical system ϕ is equivalent to another model, the *cycle-rooted spanning forest*. We may make $T_{n,m}$ into a directed graph, with each vertex (x, y) having two outgoing edges, that point to $(x + 1, y)$ and $(x, y + 1)$. A cycle-rooted spanning forest is a subgraph in which each vertex has one outgoing edge. The components in these subgraphs may be several, but each component contains a single cycle (directed northeast), which is topologically nontrivial. The remaining edges of the component form in-directed trees, attached to this cycle. The uniform probability measure on cycle-rooted spanning forests is called the CRSF measure, μ_{CRSF}. Each component is referred to as a cycle-rooted spanning tree. See Fig. 1.

It is not hard to see that the cycles of the quenched random walk ϕ are precisely the cycles in the CRSF model. We simply define $\phi(x, y)$ to be the vertex pointed to by the random outgoing edge from (x, y) in the CRSF.

The model is closely related to the monotone non-intersecting lattice path (MNLP) model, which was studied in [3]. The state space of the measure again consists of collections of disjoint, monotone northeast-going lattice cycles. A given configuration is chosen according to a Boltzmann measure μ at temperature T, which is the probability measure assigning to a configuration a probability proportional to $e^{-\frac{E_b N_b + E_c N_c}{T}}$. Here, E_b and E_c are positive constants, and N_b and N_c denote the total number of eastgoing or northgoing steps in the configuration.

In [3], the behaviour of the MNLP model was examined near its critical temperature T, which is the temperature at which $e^{-E_b/T} + e^{-E_c/T} = 1$. In the critical case, a configuration is being chosen with probability proportional to $b^{N_b} c^{N_c}$, where b, c satisfy $b + c = 1$.

The two models, CRSF and MNLP, have similarities in their definition, and we will shortly explain their connection more precisely. As [3] showed, MNLP is amenable to an exact solution analysis using Kasteleyn theory. The CRSF model, on the

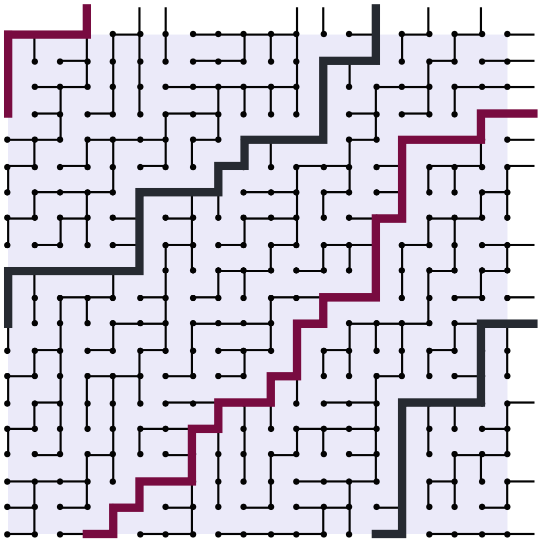

Fig. 1 A quenched random walk and its orbit structure, with the cycles highlighted

other hand, has a dynamical definition that permits a more geometric discussion of its behaviour.

The authors of [3] determine that the behaviour of MNLP depends sensitively, for large tori, on the aspect ratio m/n of the torus, with radical changes in behaviour occurring near rational values for this ratio. In Theorem 1 of [3], for example, an asymptotic expression for the value of partition function (whose definition we will shortly give), and for the mean and variance of the number of northgoing edges present in an MNLP configuration is computed. Figure 3 of [3] illustrates a conjecture of [3]: the number of edges typically present in an MNLP configuration appears to be highly peaked when the aspect ratio is precisely equal to a given small rational, it experiences a rapid decay if the aspect ratio is slightly increased or decreased, and, strikingly, it increases again if the change in aspect ratio is further accentuated, while still remaining "far" from other rationals.

Consider the critical case of MNLP where $b = c = 1/2$ (we might consider other values of b, c such that $b + c = 1$, but this is essentially indistinguishable, for asymptotic behaviour, from changing the aspect ratio of the torus). In this case, MNLP assigns a weight of $2^{-|\mathcal{C}|}$ to any configuration $\mathcal{C}$, a disjoint union of north- and east-going cycles, where $|\mathcal{C}|$ is the total number of edges of $\mathcal{C}$. We define $Z_{MNLP} = \sum_{\mathcal{C}} 2^{-|\mathcal{C}|}$, where the sum is over all configurations, so that the probability of a configuration is $2^{-|\mathcal{C}|}/Z_{MNLP}$. The quantity Z_{MNLP} is the *partition function* of the MNLP model.

We now discuss the relation between the two models, MNLP and CRSF. Consider a variant of CRSF, called oriented CRSF, which is given by the uniform measure on cycle-rooted spanning forests, each of whose cycles is given an orientation (either to the northeast, or the southwest). We can exhibit a measure-preserving correspondence between MNLP and oriented CRSF. Indeed, if we take a sample of MNLP, and orient each of its cycles in the southwesterly direction, and then assign to each remaining vertex in the torus an edge pointing to the east or to the north, independently and with

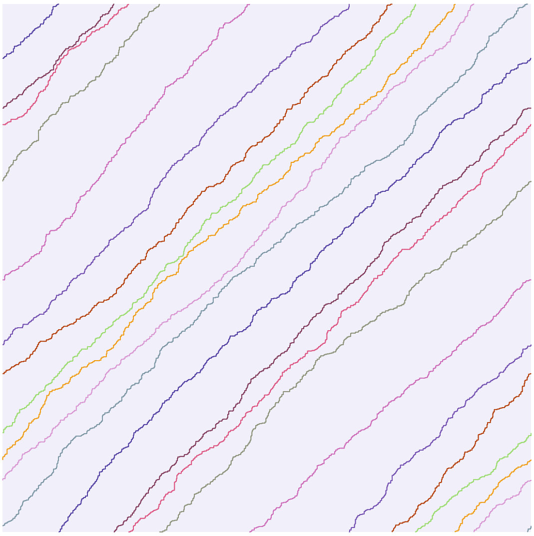

Fig. 2 Cycles in a sample when $n = m = 1{,}000$

equal probability, declaring that the newly formed cycles, if any, are to be oriented in the northeasterly direction, we claim that the resulting law is the uniform measure on oriented cycle-rooted spanning forests. We see this as follows: for a given cycle-rooted spanning forest whose cycles are oriented, the procedure will alight on this configuration if, firstly, the sample of MNLP happens to pick its set $\mathcal{C}$ of southwesterly oriented cycles and, secondly, the correct choice of north or east outgoing edges is made in each of the remaining vertices. The probability of the first event is $2^{-|\mathcal{C}|}/Z_{MNLP}$, while the second then occurs with probability $2^{-(mn-|\mathcal{C}|)}$. So the probability of obtaining the given oriented cycle-rooted spanning forest, which is $2^{-mn}Z_{MNLP}^{-1}$, does indeed not depend on the configuration of the oriented CRSF.

This argument also demonstrates that the number of oriented CRSFs is given by $2^{mn}Z_{MNLP}$. We also know that it is equal to the sum over CRSFs of $2^{\sharp \text{ cycles}}$, or to $2^{mn}\mathbb{E}_{CRSF}(2^{\sharp \text{ cycles}})$, since there a total of 2^{mn} CRSFs.

We summarise these deductions:

Theorem 1.1 *The MNLP model is in measure-preserving correspondence with oriented CRSF, and*

$$\mathbb{E}_{CRSF}\left(2^{\sharp\text{cycles}}\right) = Z_{MNLP}, \tag{1}$$

That is, the partition function for the MNLP model is the expected value of 2 *to the number of components in the CRSF model.*

In [3] it was conjectured that $Z_{MNLP} \geq 2$. This follows trivially from (1).

In this paper we study the number and the length of cycles in CRSF, as well as their homology class. Our results, which we shortly outline, yield a geometric understanding of CRSF: the sharp changes in behaviour that occur near rational choices for the aspect ratio for a large torus, and the remaining generic case in which the aspect ratio is not close to any small rational (Figs. 2, 3).

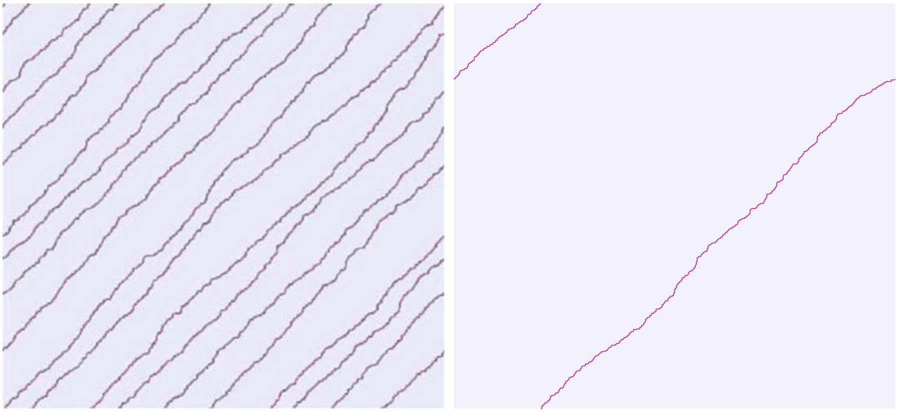

Fig. 3 Cycles in two samples when $n = 1{,}090$, $m = 1{,}000$. In the first there is one cycle with homology class (9, 10)

1.1 Outline

In Sect. 2, we prove two propositions, the first detailing the behaviour of the number of cycles when $m-n = O(n^{1/2})$, and the second when $n^{1/2} << |m-n| << \sqrt{2n\log n}$. We show that, in the first case, $\Theta(n^{1/2})$ cycles are likely to be present in the CRSF, while this number experiences a rapid decay as we enter the second case. In Sect. 3, we turn to the behaviour of the model when $m = n + C\sqrt{n\log n}$, where $C > \sqrt{2}$ is a fixed constant, showing that precisely one loop is likely to exist, and that this loop is global in nature, having length $n^{3/2+o(1)}$. We see, then, that the mean number of edges present in the CRSF configuration, as the aspect ratio is perturbed from $m/n = 1$ by varying the value of m, is peaked at an order of $n^{3/2}$ when $|m/n-1| = O(n^{-1/2})$, falls below n^{ϵ} when $m = n + (\sqrt{2}+c)\sqrt{n\log n}$ for $c = c(\epsilon)$ a small negative constant, and then rises again to $n^{3/2+o(1)}$ as this constant is taken to be positive. These phases appear to correspond to the primary spike, the valley, and the secondary spike observed for the mean total length of cycles in MNLP near a given rational aspect ratio that is described in Section 6 of [3].

In Sect. 4, we extend these results to the case m/n close to a rational p/q. In Sect. 5, we deal with the case that m/n is not close to a rational with small denominator. In this case, typically, a constant number of cycles form, each having a length of order $n^{4/3}$. These cycles cross the torus about $n^{1/3}$ times and divide it into pieces whose widths are about $n^{2/3}$.

1.2 Notation

We identify $T_{n,m}$ with the rectangle $\{0,\dots,n-1\}\times\{0,\dots,m-1\}$ in $\mathbb{Z}^2$. A closed orbit (or *cycle*) has *homology class* (p,q) if it crosses any horizontal line q times and any vertical line p times. We refer to such an orbit as a (p,q)-*cycle*. Two disjoint closed orbits necessarily have the same homology class, and p, q are necessarily relatively prime, since the orbits are simple closed curves. The length of a (p, q)-cycle is $np + mq$ edges; it necessarily has np horizontal edges and mq vertical edges.

A *strand* of a cycle $C = \{c_0, c_1, \ldots, c_n\}$, with $c_n = c_0$, is defined to be a subpath $\{c_i, c_{i+1}, \ldots, c_{j-1}, c_j\}$ between two consecutive passes of the line $x = \frac{1}{2}$, that is, such that (c_{i-1}, c_i) and (c_{j-1}, c_j) are successive pairs whose terms have x-coordinate equal to zero and one respectively. A cycle is partitioned by the set of its strands.

For a particular realization of ϕ, and for $(p, q) \in \mathbb{N}^2$, we write $N_{(p,q)}$ for the number of cycles of homology class (p, q), and $N = \sum_{(p,q)\in\mathbb{N}^2} N_{(p,q)}$ for the total number (note that exactly one term in the sum is nonzero).

For each $(i, j) \in \{0, \ldots, n-1\} \times \{0, \ldots, m-1\}$, associate a random walk $W_{i,j} : \mathbb{N} \to T_{n,m}$, starting at (i, j) and whose steps are independently up and to the right with equal probability. We call such a random walk an *up-right random walk*.

The CRSF can be obtained by iteratively running the random walks $\{W_{ij}\}_{0\le i\le n-1, 0\le j\le m-1}$, stopping each when it intersects its trace or the trace of those run earlier. We will form the configuration in this way, or by some variation of this approach. Although in constructing the CRSF, we have no cause to examine the walks after they intersect themselves, the independence properties of the walks without stopping will be useful in the proofs of the propositions.

Similarly to the definition of a strand of a cycle, we say that a walk $W_{x,y}$ performs a traversal on the interval $\{t_1, \ldots, t_2 - 1\}$ if the x-coordinate of $W_{x,y}$ is $m-1, 0, m-1, 0$ at times $t_1 - 1, t_1, t_2 - 1, t_2$ respectively (the first condition we omit if $t_1 = 0$), and t_2 is the first return to the line $x = 0$ after t_1. That is, a traversal is a horizontal crossing of the torus by the walk.

2 The primary spike and the valley when $m \approx n$

We discuss in this section the case when $m - n = O(\sqrt{n})$. In this case there are many $(1, 1)$-cycles. This case is generalized to $m/n \approx p/q$ in Sect. 4.

Proposition 1 *Let $\rho \in (0, \infty)$ be fixed and $m = n + \rho\sqrt{n}(1 + o(1))$. There exists $c = c(\rho) > 0$ with the following property. The probability that each closed orbit has homology class $(1, 1)$ exceeds $1 - \exp\{-cn^{1/2}\}$. The number $N_{(1,1)}$ of such orbits satisfies*

$$\mathbb{P}\left(N_{(1,1)} > cn^{1/2}\right) \geq 1 - \exp\left\{-cn^{1/2}\right\}$$

for all sufficiently large n. Moreover there exists a (large) $C > 0$ so that we have

$$\mathbb{P}\left(N_{(1,1)} > Cn^{1/2}\right) \leq \exp\left\{-C^{-1}n^{1/2}\right\}$$

for all sufficiently large n.

Proof Let $K > 0$ be large but fixed as $n \to \infty$. We partition the torus into strips of width $2K\sqrt{n}$, parallel to the closed path $y = (m/n)x$. That is, let $A_i, i \in \{0, \ldots, \lfloor \frac{m}{2K\sqrt{n}} \rfloor - 1\}$ denote the set of vertices in the ith strip:

$$A_i = \left\{ (x, y) \in T_{n,m} : 2Kn^{1/2} i \le |y - (m/n)x| < 2Kn^{1/2}(i+1) \right\}.$$

The last strip, $A_{\lfloor \frac{m}{2K\sqrt{n}} \rfloor}$, may be thinner than the others, but this does not matter for our purpose.

We will form the CRSF configuration in the following way. Let $z_i = \lfloor Kn^{1/2}(2i+1) \rfloor$ be the y-coordinate of the point on the line $x = 0$ in the center of the ith strip. We will run the walks W_{0,z_i}, $i = 1, 2, \ldots$ in increasing order, stopping any such walk at the stopping time σ_i, where

$$\sigma_i = \min \left\{ j \ge 0 : W_{0,z_i}(j) \notin A_i, \text{ or } W_{0,z_i}(j) = W_{0,z_i}(t) \text{ for some } t < j \right\}$$

denotes the first time at which W_{0,z_i} either leaves ∂A_i or hits its own trace.

After these segments of random walks have been run, we choose an arbitrary order of successive sites as the initial locations of further random walks, until a cycle-rooted spanning forest has been determined.

Let $E_i, i \in \{0, \ldots, \lfloor \frac{m}{2K\sqrt{n}} \rfloor\}$ denote the event that the walk W_{0,z_i} remains in A_i during its first two traversals, and that its first return to the line $\{x = 0\}$ occurs at a y-coordinate exceeding z_i, while its second return has a y-coordinate at most z_i. We claim that $\mathbb{P}(E_i) > c = c(\rho) > 0$. Indeed, let $\{X_j : j \in \mathbb{N}\}$ denote the number of upward displacements made by W_{0,z_i} between its $j-1$-st and jth rightward displacements. The event that W_{0,z_i} remains in A_i during its first two traversals occurs precisely when

$$\left| \sum_{k=1}^{j} X_k - \frac{mj}{n} \right| \le Kn^{1/2} \tag{2}$$

for each $j \in \{1, \ldots, 2n\}$. The condition on m implies that $\left| \frac{mj}{n} - j \right| \le 2\sqrt{n}\rho(1+o(1))$ for such values of j, from which we learn that (2) is implied by the inequalities

$$\left| \sum_{k=1}^{j} X_k - j \right| \le (K - 3\rho)\, n^{1/2} \tag{3}$$

each being satisfied, for $j \in \{1, \ldots, 2n\}$. For the occurrence of E_i, we require in addition that

$$\sum_{k=1}^{n} X_k - m \in \left[0, Kn^{1/2} \right] \tag{4}$$

and that

$$\sum_{k=1}^{2n} X_k - 2m \in \left[-Kn^{1/2}, 0 \right]. \tag{5}$$

Equivalent to (3) is the assertion that the random walk $\sum_{i=1}^{j} 2^{-1/2}(X_i - 1)$, that has a step distribution with mean zero and variance one, has a maximum absolute value of at most $(K - 3\rho)(n/2)^{1/2}$, among $j \in \{1, \dots, 2n\}$. As (3) is sufficient for (2), so are (4) and (5) implied by

$$\sum_{k=1}^{n} 2^{-1/2} (X_k - 1) \in \left[\frac{3\rho}{\sqrt{2}} n^{1/2}, \frac{K - 3\rho}{\sqrt{2}} n^{1/2} \right] \tag{6}$$

and

$$\sum_{k=1}^{2n} 2^{-1/2} (X_k - 1) \in \left[-\frac{K - 3\rho}{\sqrt{2}} n^{1/2}, -\frac{3\rho}{\sqrt{2}} n^{1/2} \right]. \tag{7}$$

We have then that $\mathbb{P}(E_i) \geq p_n$, where p_n is the probability that each of the conditions (3), (6) and (7) is satisfied. By Donsker's theorem [1, page 365], we have that $p_n \to p$ as $n \to \infty$, where

$$p = \mathbb{P}\left(\left\{ \sup_{t \in [0,2]} \left| B(t) \right| \leq \frac{K - 3\rho}{\sqrt{2}} \right\} \cap \left\{ B(1) \in \left[\frac{3\rho}{\sqrt{2}}, \frac{K - 3\rho}{\sqrt{2}} \right] \right\} \cap \left\{ B(2) \in \left[-\frac{K - 3\rho}{\sqrt{2}}, \frac{-3\rho}{\sqrt{2}} \right] \right\} \right), \tag{8}$$

with $B : [0, \infty) \to \mathbb{R}$ denoting a standard Brownian motion. Note that $p > 0$: the event

$$\left\{ B(1) \in \left[\frac{3\rho}{\sqrt{2}}, \frac{K - 3\rho}{\sqrt{2}} \right] \right\} \cap \left\{ B(2) \in \left[-\frac{K - 3\rho}{\sqrt{2}}, \frac{-3\rho}{\sqrt{2}} \right] \right\}$$

occurs with positive probability, because $B(1)$ and $B(2) - B(1)$ are independent normal random variables. Conditionally on the pair $(B(1), B(2))$ taking a given value in the set $[\frac{3\rho}{\sqrt{2}}, \frac{K-3\rho}{\sqrt{2}}] \times [-\frac{K-3\rho}{\sqrt{2}}, \frac{-3\rho}{\sqrt{2}}]$, there is a uniformly positive probability that the first condition in the event on the right-hand-side of (8) is satisfied, as we see from the law of the maximum of a Brownian bridge [1, exercise 8.2, page 391].

Thus, indeed, each event E_i has a positive probability $> c(\rho)$, bounded below independently of n and i.

If E_i occurs, then, at the stopping time σ_i, the walk W_{0,z_i} hits its own trace. The choice of the order of the walks in the formation of the CRSF configuration ensures that this event produces a $(1, 1)$-cycle in A_i. Moreover, the events E_i are pairwise independent. Thus, the number $N_{(1,1)}$ of $(1, 1)$-cycles is bounded below by a binomial random variable with parameters $\lfloor \frac{m}{2K\sqrt{n}} \rfloor$ and $c(\rho)$. If the configuration contains one $(1, 1)$-cycle, then all the other cycles are also of this type, so the absence of a $(1, 1)$-cycle implies that E_i does not occur for any $i \in \{0, \dots, \lfloor m/(2Kn^{1/2}) \rfloor - 1\}$. We infer the first two statements of the proposition.

To treat the last assertion, we form the CRSF by running the random walk W_{00} and adding an edge $(x, y)(x', y')$ traversed by W_{00} to the configuration on each occasion for which the site (x, y) is visited by W_{00} for the first time, until a cycle-rooted spanning forest is formed. We may assume that at least one $(1, 1)$-cycle is formed, or, equivalently, that every cycle in the configuration is a $(1, 1)$-cycle.

Define the wraparound time of the walk W_{00} to be the earliest time t such that the set $W_{00}[0, t]$ of vertices visited by the walk up to time t has the property that every $(1, 1)$-cycle in $T_{n,m}$ intersects $W_{00}[0, t]$.

We record the successive maxima and minima of the y-coordinate of the intersection of the walk W_{00} with the line $\{x = \frac{1}{2}\}$ (i.e. the first horizontal step after each visit to the line $x = 0$). Let $u_0 = X_1$ be the y-coordinate of the walk on the first occasion that it crosses the line $x = \frac{1}{2}$. When the walk next crosses the line $\{x = \frac{1}{2}\}$, its y-value, which, with the natural choice of shift by a multiple of m, we take to be $\sum_{i=1}^{n+1} X_i - m$, may or may not exceed u_0. If it is greater than u_0, we record its value as u_1, and, if it is smaller than u_0, we record it as v_1. We do not set the value of either u_1 or v_1 if $\sum_{i=1}^{n+1} X_i - m$ equals u_0. The y-value of the walk on the occasion of the kth return to the line $\{x = 1/2\}$ is given by $\sum_{i=1}^{kn+1} X_i - km$. We iteratively record the successive maxima of these statistics as $u_2, u_3, \ldots$ and the successive minima as $v_2, v_3, \ldots$.

We no longer record either maxima or minima on a return to $\{x = 1/2\}$ if this return occurs after the wraparound time. Let $\{u_0, \ldots, u_{J_1}\}$ and $\{v_1, \ldots, v_{J_2}\}$ denote the final record. Let Q denote the set of horizontal edges crossing $\{x = 1/2\}$ that are traversed by the walk at one of the recorded times.

By assumption, in each tree T of the CRSF configuration lies a unique $(1, 1)$-cycle, and in this cycle lies a unique horizontal edge $e = e(T)$ that crosses the line $\{x = 1/2\}$. Let E denote the set of edges of the form $e(T)$ for some tree T in the configuration. Let $e_0 \in E$ be the element in E lying in the cycle in the configuration which is the last to be formed.

We claim that

$$E \setminus \{e_0\} \subseteq Q. \tag{9}$$

To see this, note that if a $(1, 1)$-cycle C lies in the configuration, there exists a vertex $c \in C$ and $t, s \in \mathbb{N}$, $t < s$, such that the walk W_{00} makes its first and second visits to c at times t and s, the set $W_{00}[t + 1, s]$ is the vertex set of C, and

$$W_{00}\,[t, s] \cap W_{00}\,[0, t - 1] = \emptyset. \tag{10}$$

We call t the start-time of the cycle C, and s, the end-time.

Note that the start-time of C necessarily occurs before the wraparound time. The intervals of time during which distinct cycles of the CRSF configuration form being disjoint, we learn that every cycle except possibly that which forms last has an end-time that occurs before the wraparound time.

We claim that the cycle C, whose vertex set is $W_{00}[t + 1, s]$, either lies above or below the configuration $W_{00}[0, t-1]$ present prior to its formation. More precisely, the

y-coordinate of every vertex $W_{00}(n)$, $n \in \{t, \ldots, s\}$, exceeds the maximum y-coordinate of any $W_{00}(m)$, $m \in \{0, \ldots, t-1\}$, sharing its x-coordinate, or the y-coordinate of every such vertex is less than the corresponding minimum. Indeed, this statement is readily verified from (10), with the aid, for example, of the intermediate value theorem.

We have shown that every edge $e \in E$, except possibly e_0, is traversed before the wraparound time, at a time which is recorded on either the $\{u_i\}$ or $\{v_i\}$ list. That is, we have obtained (9).

Note then that

$$N = |E| \leq |Q| + 1 = J_1 + J_2 + 2, \tag{11}$$

the inequality by (9).

We now show that, if $\rho = 0$, then for $i \in \{1, 2\}$, and if $\rho > 0$, for $i = 1$,

$$\mathbb{P}\left(J_i > Cn^{1/2}\right) \leq \exp\left\{-C^{-1}n^{1/2}\right\}, \tag{12}$$

for C sufficiently large. Indeed, the sequence of increments $\{u_{i+1} - u_i : i \in \mathbb{N}\}$ consists of independent random variables, each of which has, for small enough c, probability at least c of exceeding $n^{1/2}$. To see this, note that the value u_{j+1} will be recorded on the first return to $\{x = 1/2\}$ after that at which u_j is recorded, provided that this return occurs at a higher value of y, which occurs with probability at least $2^{-1}(1 + o(1))$, in which case, the difference $u_{j+1} - u_j$ will exceed $n^{1/2}$ with positive probability, by the central limit theorem. If a proposed entry u_j exceeds m, then the wraparound time has already occurred, and the entry is not recorded. We see that (12) follows.

Suppose now that $\rho > 0$. Then for any k,

$$\mathbb{P}\left(J_2 > k\right) \leq \gamma^k, \tag{13}$$

for some $\gamma = \gamma(\rho) \in (0, 1)$. Indeed, it is readily seen that a new term is added to the v-sequence, independently of its history, with a probability that is bounded away from one. So J_2 satisfies (12) in this case also.

This completes the proof of the last assertion of the proposition. □

The number of cycles experiences a rapid decline as the value of m is increased beyond that treated in Proposition 1.

Proposition 2 *We set $m = n + \lfloor C\sqrt{n \log n} \rfloor$, for $C \in (0, \infty)$ a fixed constant. Then*

$$\mathbb{E}\left(N_{(1,1)}\right) = \frac{1}{2\sqrt{\pi}} n^{\frac{1}{2} - \frac{C^2}{4}} \left(1 + O\left(\frac{(\log n)^3}{n^{1/2}}\right)\right).$$

Remark The principal interest of this result is for values $C \in (0, \sqrt{2})$, in which case, all cycles are of homology class $(1, 1)$ with high probability. Indeed, the proof of the first part of Proposition 1 may be adapted to show that every cycle in the configuration is a (1, 1)-cycle with probability exceeding

$$1 - \exp\left\{ - c\frac{n^{1/2-C^2/4}}{\sqrt{\log n}} \right\},$$

where $c > 0$ is a small constant.

Proof We estimate the expected number $\mathbb{E}N_{(1,1)}$ of (1, 1) cycles as follows. Every (1, 1) cycle contains exactly one of the edges $(0, y)(1, y)$, for $y = 0, \ldots, m-1$. Hence, the number of (1, 1) cycles is equal to the number of edges $(0, y)(1, y)$ with $0 \le y < m$ that are present in the CRSF configuration and are such that the trajectory begun at $(1, y)$ first visits the line $\{x = 1/2\}$ at $(1/2, y)$. Thus,

$$\mathbb{E}\left(N_{(1,1)}\right) = m\mathbb{P}\left(\{Z = m\} \cap \{(0,0)(1,0) \in \mathcal{C}\}\right),$$

where $\mathcal{C}$ denotes the CRSF configuration, and Z denotes the y-coordinate of the trajectory starting at $(1, 0)$ on its first return to the line $x = \frac{1}{2}$. Noting that $Z = m$ implies that $(0, 0)(1, 0) \in \mathcal{C}$, we obtain

$$\mathbb{E}\left(N_{(1,1)}\right) = m\mathbb{P}\left(Z = m\right). \tag{14}$$

Note further that $Z = \sum_{i=1}^{n} Y_i$, where $\{Y_i : i \in \{1, \ldots, n\}\}$ is an independent sequence of geometric random variables of mean one and variance two. We find that

$$\mathbb{P}\left(Z = m\right) = \mathbb{P}\left(\sum_{i=1}^{n} (Y_i - 1) = \lfloor C\sqrt{n \log n} \rfloor\right).$$

We require a local limit theorem for a sum of independent identically distributed random variables in a regime of moderate deviations. Theorem 3 of [6] yields

$$\mathbb{P}\left(\sum_{i=1}^{n} (Y_i - 1) = \lfloor C\sqrt{n \log n} \rfloor\right) = \frac{1}{2\sqrt{\pi}} n^{-\frac{1}{2}-\frac{C^2}{4}} \left(1 + O\left(\frac{(\log n)^3}{n^{1/2}}\right)\right),$$

so that the result follows from (14), since $m = n + O(\sqrt{n \log n})$. □

3 The secondary spike when $m \approx n$

We prove two propositions regarding the behavior of the model in a regime where $m = n + C\sqrt{n \log n}$, with $|C| \in (\sqrt{2}, \infty)$ a fixed constant. In Proposition 3, we show that it is likely that there is a cycle of length $n^{3/2+o(1)}$, and, in Proposition 4, we establish that it is likely to be the only cycle.

Proposition 3 *For* $|C| > \sqrt{2}$, *set* $m = n + C\sqrt{n \log n}(1 + o(1))$. *Then, for any* $\epsilon > 0$ *and for* n *sufficiently large,*

$$\mathbb{P}\left(\textit{any cycle has length at least } \frac{n^{3/2}}{3|C|\sqrt{\log n}} \right) \geq 1 - n^{\frac{1}{2} - \frac{C^2}{4} + \epsilon}.$$

Proof We treat the case that $C < 0$, the other being similar. Let $\phi^* : \{0, \ldots, m-1\} \to \mathbb{N}$ be the y-coordinate of the return map of the line $x = 0$ to itself. That is, let $(0, \phi^*(y)) = \phi^\tau(0, y)$ where $\tau = \tau_y > 0$ is the first time that $\phi^\tau(0, y)$ has x-coordinate zero after the first positive time at which it has a strictly positive x-coordinate. Note that $y < y'$ implies $\phi^*(y) \leq \phi^*(y')$, in other words ϕ^* is non-decreasing.

For $i \in \mathbb{N}$, let D_i denote the event that

$$\phi^* \left(\lfloor i \epsilon C \sqrt{n \log n} \rfloor \right) > (i + 1)\epsilon C \sqrt{n \log n}.$$

Note that, for any given y, we may write $\phi^*(y) = y - m + \sum_{i=1}^{n} Y_i$, where $\{Y_i : i \in \mathbb{N}\}$ is an independent sequence of geometric random variables of mean one and variance two. By Theorem 5.23 of [5], we have the bound

$$\mathbb{P}(D_i) \geq 1 - C_0 n^{-\frac{C^2}{4}(1-\epsilon)^2}, \tag{15}$$

for some large constant C_0. Set $D = \bigcap_{i=0}^{\lfloor \frac{m}{C\epsilon\sqrt{n \log n}} \rfloor} D_i$. We claim that, if D occurs, then $\phi^*(k) \geq k$ for all $k \in \{0, \ldots, m-1\}$. Suppose on the contrary $\phi^*(k) < k$. Let $j \in \mathbb{N}$ be maximal such that $\lfloor j \epsilon C \sqrt{n \log n} \rfloor \leq k$. Then

$$\phi^* \left(\lfloor j \epsilon C \sqrt{n \log n} \rfloor \right) \leq \phi^*(k) < k < \lfloor (j+1)\epsilon C \sqrt{n \log n} \rfloor < \phi^* \left(\lfloor j \epsilon C \sqrt{n \log n} \right),$$

a contradiction.

For $i \in \mathbb{N}$, let E_i denote the event that

$$\phi^* \left(\lfloor i C \sqrt{n \log n} \rfloor \right) \leq \lfloor (i+2) C \sqrt{n \log n} \rfloor.$$

Arguing similarly to (15), we note that

$$\mathbb{P}(E_i) \geq 1 - C_0 n^{-\frac{C^2}{4}}, \tag{16}$$

where C_0 again denotes a large constant.

Set $E = \bigcap_{i=0}^{\lfloor \frac{m}{C\sqrt{n \log n}} \rfloor} E_i$.

Define $y_0 = 0$ and for $i > 0$ define $y_i = \phi^*(y_{i-1})$. We will show that, if E occurs,

$$y_{i+1} - y_i \leq 3C\sqrt{n \log n} \tag{17}$$

for each $i \in \mathbb{N}$. Let $j \in \{1, \ldots, \lfloor \frac{m}{C\sqrt{n \log n}} \rfloor + 1\}$ satisfy

$$\lfloor (j-1)C\sqrt{n \log n} \rfloor \leq y_i < \lfloor jC\sqrt{n \log n} \rfloor.$$

We have that

$$\begin{aligned} y_{i+1} = \phi^* (y_i) &\leq \phi^* \left(\lfloor jC\sqrt{n \log n} \rfloor \right) \\ &\leq \lfloor (j+2)C\sqrt{n \log n} \rfloor \leq y_i + 3C\sqrt{n \log n}, \end{aligned}$$

the first inequality since ϕ^* is non-decreasing, the second due to the occurrence of E_j. We have obtained (17).

Let $K \in \mathbb{N}$ be maximal subject to $y_K < m$ (note that K is finite if D occurs). We claim that, on the event D, any cycle has at least K strands, and that, on the event E, $K \geq \frac{m}{3C\sqrt{n \log n}}$.

Indeed, setting $C_i = \{y_i, \ldots, y_{i+1}\}$ for $i \in \{0, \ldots, K-1\}$, we have that $\phi^*(C_i) \subseteq C_{i+1}$ for such i, if D occurs. The monotonicity of ϕ^* implies that any cycle contains a point $(0, a_1)$ with $a_1 \in C_1$ and, by the sequence of inclusions, distinct points $(0, a_i)$ with $a_i \in C_i$ for each value of i. Hence, the cycle has at least K strands.

The lower bound on K follows by noting that, from (17),

$$m - 3C\sqrt{n \log n} \leq y_K = \sum_{i=1}^{K-1} (y_{i+1} - y_i) \leq 3C(K-1)\sqrt{n \log n}.$$

The proof is completed by noting the following bounds on $\mathbb{P}(D)$ and $\mathbb{P}(E)$, which follow from (15) and (16):

$$\mathbb{P}(D) \geq 1 - \frac{2C_0 n^{1/2 - \frac{C^2}{4}(1-\epsilon)^2}}{C\epsilon\sqrt{\log n}}$$

and

$$\mathbb{P}(E) \geq 1 - \frac{2C_0 n^{1/2 - \frac{C^2}{4}}}{C\sqrt{\log n}}.$$

□

Proposition 4 *Set $m = n + C\sqrt{n \log n}\,(1 + o(1))$ with $|C| > \sqrt{2}$. Then, in the CRSF, for each $\epsilon > 0$,*

$$\mathbb{P}\,(\textit{there exist at least two disjoint cycles}) \leq n^{\frac{1}{2} - \frac{C^2}{4} + \epsilon}$$

for n sufficiently large.

Remark As the proof will show, in the presence of one cycle of the length given in the statement of Proposition 3, the conditional probability of another cycle decays as $\exp\{-n^{1/2+o(1)}\}$. In the regime that Propositions 3 and 4 treat, then, the most probable means by which two cycles form is by local fluctuations in the generating random walks that create two (1, 1)-cycles. This occurs with a probability that decays polynomially in n.

Proof Recall the events D and E, the quantity K and the intervals C_i from the proof of Proposition 3. Suppose that the event $D \cap E$ occurs. Let a_i be the intersection of a cycle with the interval C_i. Then

$$a_{i+1} - a_i \leq y_{i+2} - y_i \leq 6C\sqrt{n \log n}.$$

Suppose that the CRSF configuration is formed by firstly running the random walk $W_{0,0}$ until it meets its own trace, and then running the walks W_{0,w_i} until existing trees or the current trace is hit, where $w_i \in \{0, \ldots, m-1\}$ are selected in an arbitrary manner from the subset of the line $\{x = 0\}$ not yet belonging to any tree.

If two cycles are to be present in the configuration, then, for some z_1 with $a_1 < z_1 < a_2$, W_{0,z_1} must not meet the first cycle before it hits its own trace. Set $z_0 = w_1$, and let z_i denote the y-coordinate of the ith return of the walk W_{0,w_1} to the line $\{x = 0\}$.

In the event $D \cap E$, if W_{0,w_1} does not meet the first cycle before visiting its own trace, then

$$a_i < z_i < a_{i+1} \tag{18}$$

for each $i \in \{1, \ldots, K-2\}$.

We sample the sequence z_i when i is a multiple of $L = \lfloor n^\epsilon \rfloor$. Note that (18) implies that

$$|a_{(j+1)L} - a_{jL+1}| \leq |z_{(j+1)L} - z_{jL}| \leq |a_{(j+1)L+1} - a_{jL}| \tag{19}$$

for each j. So $z_{(j+1)L} - z_{jL}$ is restricted to an interval of length at most $12C\sqrt{n \log n}$.

The quantity $z_{(j+1)L} - z_{jL}$ has the distribution of $\sum_{i=1}^{nL} X_i - mL$, where $\{X_i\}$ is a sequence of independent geometric random variables of mean one and variance two (these are the vertical displacements of the walk W_{0z} in between successive rightward movements).

Now (19) at the given value of j requires that this sum $\sum_{i=1}^{nL} X_i$ of independent random variables lie in a fixed interval of length at most $12C\sqrt{n \log n}$. It follows readily from Theorem 3 of [6] that the probability of this event is maximized by choosing the interval to be centred at nL, and, thus, to be bounded above by

$$C_0 \frac{12C\sqrt{n \log n}}{\sqrt{nL}} \leq n^{-\frac{\epsilon}{2}}.$$

Thus, on the event $D \cap E$, the probability that each of the inequalities (19) is satisfied is at most

$$\left(n^{-\epsilon/2}\right)^{n^{1-\epsilon}} \leq \exp\{-n^{1-\epsilon}\}.$$

The bounds on the probabilities of D and E presented in the proof of Proposition 3 complete the proof. □

4 Near $m/n = p/q$

We extend the previous results to the case m/n is near a rational p/q with small denominator.

Proposition 5 *Let p, q be fixed and relatively prime. Let $\rho \in (0, \infty)$ be fixed and $m = (p/q)n + \rho\sqrt{n}(1 + o(1))$. For $c > 0$ small enough, each closed orbit has homology class (p, q) with probability at least $1 - \exp\{-cn^{1/2}\}$, while*

$$\mathbb{P}\left(N_{(p,q)} > cn^{1/2}\right) \geq 1 - \exp\left\{-cn^{1/2}\right\}$$

for sufficiently large n. For $C > 0$ sufficiently large, we have

$$\mathbb{P}\left(N_{(p,q)} > Cn^{1/2}\right) \leq \exp\left\{-C^{-1}n^{1/2}\right\}$$

for sufficiently large n.

Proof The first part of the proof is essentially the same as the proof of Proposition 1, with the following changes. We again partition the torus into strips, but in this case the strips have horizontal length pn instead of n. Thus each strip winds p times around horizontally, and q times vertically, before closing up. The direction of the strip is now parallel to the closed curve of homology class (p, q) on the torus, and the width of the strips is still $2K\sqrt{n}$ for some large K.

For the second half of the proof, we require some variations on the sequence of maxima and minima that we record. We divide returns to the line $\{x = 1/2\}$ into p classes, according to the value of the index of the return reduced mod p. We then form p separate lists $\{u_i^j\}$, $\{v_i^j\}$ of maxima and minima, where the y-coordinate of the kth return to the line $\{x = 1/2\}$ is entered as a maximum u_i^j or as a minimum v_i^j on the list j, $j = k \bmod p$, if this y value exceeds, or is less than, any y-coordinate for an l-return to the line $\{x = 1/2\}$ with $l \bmod p$ equal to j.

We define the wraparound time to be that moment at which there no longer exists a cycle of homology (p, q) that is disjoint from the existing trace of W_{00}. We no longer record the y-coordinate of a return to the line $\{x = 1/2\}$ after the wraparound time.

Let $\{u_i^j : 0 \leq i \leq J_1^j\}$ and $\{v_i^j : 1 \leq i \leq J_2^j\}$ denote the maxima and minima recorded on the jth list. Similarly to the case treated in Proposition 1, the y-coordinate of each horizontal edge crossing $\{x = 1/2\}$ is recorded on one of the lists, for each cycle in the configuration, except possibly the last one. We learn that

$$N \leq \sum_{j=1}^{p} \left(J_1^j + J_2^j + 1 \right) + p.$$

The proof is completed by estimating the tail of the random variables J_1^j and J_2^j as in the previous proof. □

Proposition 6 *We set* $m = (p/q)n + C\sqrt{n}\sqrt{\log n}$*, for* $C \in (0, \sqrt{2p})$*. Then*

$$\mathbb{E}N_{(p,q)} = \frac{\sqrt{p}}{2q\sqrt{\pi}} n^{\frac{1}{2} - \frac{C^2}{4p}} \left(1 + O\left(\frac{(\log n)^3}{\sqrt{n}} \right) \right).$$

Proof We replace the (1, 1)-cycles considered in the proof of Proposition 2 by (p, q)-cycles, and note the following variation: we have

$$\mathbb{E}\left(N_{(p,q)}\right) = m\mathbb{P}\left(\{Z = pm\} \cap \{(0,0)(1,0) \in C\} \cap A\right),$$

where, in this instance, Z denotes the y-coordinate of the trajectory starting at $(1, 0)$ on its pth return to the line $\{x = 1/2\}$. The event A is that the walk $W_{0,0}$, after visiting $(0, 1)$, does not meet itself before its p-th return to the line $\{x = 1/2\}$. Noting that $\mathbb{P}(A \big| (0,0)(0,1) \in C) \geq 1 - \exp\{-cn\}$, and that $Z = pm$ implies that $(0,0)(0,1) \in C$, we see that

$$\mathbb{E}\left(N_{(p,q)}\right) = m\mathbb{P}\left(Z = pm\right) \Big(1 + O\left(\exp\{-cn\}\right) \Big).$$

Noting that $Z = pm$ if and only if $\sum_{i=1}^{pn} X_i = pn + p\lfloor C\sqrt{n\log n} \rfloor$, and applying Theorem 3 of [6], we find that

$$\mathbb{P}\left(Z = pm\right) = \frac{1}{2\sqrt{\pi p}} n^{-\frac{1}{2} - \frac{C^2}{4p}} \left(1 + O\left(\frac{(\log n)^3}{\sqrt{n}} \right) \right).$$

We thus have

$$\mathbb{E}\left(N_{(p,q)}\right) = \frac{\sqrt{p}}{2q\sqrt{\pi}} n^{\frac{1}{2} - \frac{C^2}{4p}} \left(1 + O\left(\frac{(\log n)^3}{\sqrt{n}} \right) \right).$$

□

The next two propositions, whose proofs mimic those of Propositions 3 and 4, treat the secondary spike for a torus with aspect ratio close to a general rational.

Proposition 7 *Let $p, q \in \mathbb{N}$ be relatively prime. Set $m = (p/q)n + C\sqrt{n \log n}$, for $|C| > \sqrt{2p}$. Then, for $K = K(p, q)$,*

$$\mathbb{P}\left(\text{there exists a loop of length at least } \frac{n^{3/2}}{\sqrt{\log n}}\right) \geq 1 - n^{\frac{1}{2} - \frac{C^2}{4p} + \epsilon},$$

as $n \to \infty$.

Proposition 8 *Set $m = (p/q)n + C\sqrt{n \log n}\,(1 + o(1))$ with $|C| > \sqrt{2p}$. Then, in the CRSF, for each $\epsilon > 0$,*

$$\mathbb{P}\,(\text{there exist at least two disjoint cycles}) \leq n^{\frac{1}{2} - \frac{C^2}{4p} + \epsilon},$$

for n sufficiently large.

5 The irrational regime

Let $m, n \in \mathbb{N}$ with $Cn > m > n$ for a constant $C > 1$.

We begin by collecting some elementary facts about continued fractions. These can be found in, for example, [2]. Let

$$\frac{m}{n} = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_l}}}$$

be the continued fraction decomposition of m/n. Define $p_0/q_0 = a_0$ and for $0 \leq j \leq l$, define

$$\frac{p_j}{q_j} = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_j}}}$$

to be the rational approximants to m/n.

We have $\frac{p_k}{q_k} < \frac{m}{n} \leq \frac{p_{k+1}}{q_{k+1}}$ for $k \leq l$ even and $\frac{p_{k+1}}{q_{k+1}} \leq \frac{m}{n} < \frac{p_k}{q_k}$ for $k \leq l$ odd. In each case, $\frac{p_{k+1}}{q_{k+1}}$ is the closer endpoint to m/n. Also $|\frac{p_k}{q_k} - \frac{p_{k+1}}{q_{k+1}}| = \frac{1}{q_k q_{k+1}}$. Hence,

$$\frac{n}{2q_{j+1}} < |np_j - mq_j| < \frac{n}{q_{j+1}} \tag{20}$$

We also have

$$\begin{aligned} p_k &= a_k p_{k-1} + p_{k-2} \\ q_k &= a_k q_{k-1} + q_{k-2}. \end{aligned}$$

Choose j_0 so that

$$q_{j_0} \leq n^{1/3} < q_{j_0+1}. \tag{21}$$

We will invoke upper bounds on certain coefficients a_k as hypotheses. In this regard, note that, for typical m, n, the maximum of the coefficients a_k behaves as a multiple of $\log n$. Indeed, writing $D(m, n)$ for this maximum, the main result of [4] shows that, for each $\epsilon > 0$, and uniformly in $\alpha > \epsilon$ as $x \to \infty$.

$$\Big| \{(a, b) : 0 \le a < b \le x, \gcd(a, b) = 1, D(a, b) > \alpha \log x\} \Big| \\ \sim 3\pi^{-2} x^2 \left(1 - \exp\left\{-12\alpha^{-1}\pi^{-2}\right\}\right).$$

Theorem 5.1 *Define j_0 as in (21). For each $k \in \mathbb{N}$, there exists $c = c(k) > 0$ independent of n and m, such that the probability that there at least k cycles, each of homology class (p_{j_0}, q_{j_0}), is at least c.*

Remark The length of each such cycle is equal to $np_{j_0} + mq_{j_0} \ge nq_{j_0} \ge nq_{j_0+1}/(a_{j_0} + 1) > n^{4/3}/(a_{j_0} + 1)$. Since a_{j_0} is typically $O(1)$, we see that, for n large and most choices of m, cycles of length $n^{4/3}$ form with positive probability in the CRSF configuration.

Proof We argue the case $k = 2$, the general one being no harder. Note that

$$\left| \frac{p_{j_0}}{q_{j_0}} - \frac{m}{n} \right| < \frac{1}{q_{j_0} q_{j_0+1}},$$

so that

$$\left| \frac{p_{j_0}}{q_{j_0}} - \frac{m}{n} \right| < \frac{1}{q_{j_0} n^{1/3}}, \tag{22}$$

which implies, by $1 \le \frac{m}{n} \le C$ and (21), that

$$\frac{q_{j_0}}{2} \le p_{j_0} < (C + 1) q_{j_0}.$$

Using (21), then,

$$\frac{p_{j_0}}{C + 1} \le n^{1/3}. \tag{23}$$

Note that, by (22), $|np_{j_0} - mq_{j_0}| < n^{2/3}$. Let R_1 be the closed line/loop through the origin of slope mq_{j_0}/np_{j_0} on the torus. The vertical distance between strands of R_1 is $m/p_{j_0} \ge \frac{n^{2/3}}{C+1}$. Let S_1 be a strip centered on R_1 and of width $\frac{n^{2/3}}{3(C+1)}$. Let S_2 be a translate of S_1 which is disjoint from S_1.

The probability that $W_{0,0}$ does not exit S_1 before making a cycle can be bounded as follows. It suffices that:

1. The walk stays in the strip for two circuits, i.e. for $2p_{j_0}$ returns to the line $x = 0$.
2. At the end of the first circuit (at the p_{j_0}-th return) the walk is in the upper half of the strip.
3. At the end of the second circuit the walk is in the lower half of the strip.

In the notation used to argue that $\mathbb{P}(E_i) \geq c$ in the proof of Proposition 1, the first event is

$$\left\{ \sup_{\ell \in \{1,2,\ldots,2p_{j_0}n\}} \left| \sum_{i=1}^{\ell} X_i - \ell \frac{mq_{j_0}}{np_{j_0}} \right| \leq \frac{n^{2/3}}{6(C+1)} \right\}.$$

Using $|1 - \frac{mq_{j_0}}{np_{j_0}}| \leq \frac{1}{p_{j_0}n^{1/3}}$, we may now argue similarly to the deduction of $\mathbb{P}(E_i) \geq c$ that the three events listed above occur simultaneously with a probability that is positive, uniformly in $n \geq n_0$ and $n < m < Cn$. □

Theorem 5.2 *Fix* $\epsilon > 0$*, and choose* j *so that* $q_j \leq n^{1/3-\epsilon/2} < q_{j+1}$*. Suppose that* $a_j < n^{\epsilon/2}$ *and* $a_{j+1} \leq n^{\epsilon/4}$*. Then*

$$\mathbb{P}\left(\textit{there is a cycle of length at most } O\left(n^{4/3-\epsilon}\right)\right) \leq \exp\left\{-cn^{\epsilon/2}\right\}.$$

Proof From (20) we have $|np_j - mq_j| > \frac{n}{2q_{j+1}}$. This implies that after p_j traversals, the strands of the ray R of slope 1 starting at the origin do not come within $\frac{n}{2q_{j+1}}$ of each other. In particular, there is an embedded strip $U = U_j$, centered on the ray R starting at the origin, of width $\frac{n}{2q_{j+1}}$ and horizontal length np_j.

By assumption, $a_j < n^{\epsilon/2}$ so that $q_{j+1} < (a_j+1)q_j < (n^{\epsilon/2}+1)n^{1/3-\epsilon/2} < 2n^{1/3}$. This implies that the strip U_j has width at least $n^{2/3}/4$. Arguing similarly to (23),

$$\frac{p_j}{C+1} \leq n^{1/3-\epsilon/2}. \tag{24}$$

Suppose that the CRSF configuration is formed by firstly running the walk $W_{0,0}$ until it hits its trace. If the walk $W_{0,0}$ remains in the strip U_j until it hits itself, forming a cycle, then the cycle to which $(0,0)$ belongs has length at least $mq_j \geq nq_{j+1}/(a_{j+1}+1) \geq 2^{-1}n^{4/3-(3/4)\epsilon}$, since $a_{j+1} \leq n^{\epsilon/4}$. Each cycle has the same length. Hence, the event that there is a cycle of length at most $n^{4/3-\epsilon}$ implies that the walk $W_{0,0}$ leaves the strip U_j before reaching its end. The strip U_j having horizontal length $np_j \leq (C+1)n^{4/3-\epsilon/2}$ by (24), we see that, with $\{X_i : i \in \mathbb{N}\}$ being defined in the proof of Proposition 1, if $\left|\sum_{i=1}^{j} X_i - j\right| \leq n^{2/3}/8$ for each $j \in \{1,\ldots,\lfloor (C+1)n^{4/3-\epsilon/2}\rfloor\}$, then the walk $W_{0,0}$ remains in U_j until reaching its end.

A brief argument using Theorem 5.23 of [5] yields

$$\mathbb{P}\left(\max_{0 \leq j \leq (C+1)n^{4/3-\epsilon/2}} \left|\sum_{i=1}^{j} X_i - j\right| < n^{2/3}/8\right) > 1 - e^{-cn^{\epsilon/2}}$$

which completes the proof. □

Theorem 5.3 *Fix $\epsilon > 0$ and choose $k < \ell$ so that $p_k \leq n^{1/3-\frac{7}{48}\epsilon} < p_{k+1}$ and $p_\ell \leq n^{1/3+\frac{47}{48}\epsilon} < p_{\ell+1}$. Suppose that* $\max\{a_{k+1}, a_{l+1}\} < n^{\frac{5}{48}\epsilon}$. *Then*

$$\mathbb{P}\Big(\textit{there is a cycle of length at least } n^{4/3+\epsilon})\Big) \leq \exp\left\{-n^{\frac{\epsilon}{7}}\right\}.$$

Proof We in fact prove a modified statement: choose $k \leq \ell$ so that $p_k \leq n^{1/3-\beta\epsilon} < p_{k+1}$ and $p_\ell \leq n^{1/3+\gamma\epsilon} < p_{\ell+1}$, and suppose that $\max\{a_{k+1}, a_{l+1}\} < n^{\alpha\epsilon}$. Let $\delta > \alpha + \beta$ with $\beta > \alpha$. Then

$$\mathbb{P}\Big(\text{there is a cycle of length at least } 3n^{4/3+\gamma\epsilon}\Big) \leq e^{-n^{\beta\epsilon+o(1)}} + e^{-n^{(\gamma-\alpha-2\delta)\epsilon+o(1)}}. \tag{25}$$

The statement of the proposition then follows by the choices $\alpha = 5/48$, $\beta = 7/48$, $\delta = 13/48$ and $\gamma = 47/48$.

Let E be the event

$$E = \left\{ \max_{j\in\{1,\ldots,np_k\}} \left| \sum_{i=1}^{j} X_i - j \right| \leq n^{2/3} \right\}$$

and F be the event

$$F = \left\{ \max_{j\in\{1,\ldots,np_\ell\}} \left| \sum_{i=1}^{j} X_i - j \right| \geq n^{2/3+\delta\epsilon} \right\}.$$

Suppose that the CRSF configuration is formed by firstly running the walk $W_{0,0}$ until it hits its trace. We claim that, on the event $E \cap F$, the walk $W_{0,0}$ completes a cycle before its q_l-th return to the line $\{x = 0\}$. To see this, note that after p_k returns to $x = 0$, the ray R splits the x-axis into intervals of lengths between m/p_{k+1} and m/p_k. We have

$$\frac{n^{1-\alpha\epsilon}}{2p_k} < \frac{m}{p_{k+1}} < \frac{m}{n^{1/3}}n^{\beta\epsilon} < \frac{m}{p_k} < \frac{2Cn^{1+\alpha\epsilon}}{p_{k+1}},$$

so these intervals are of order at least $m/p_{k+1} \geq n/p_{k+1} \geq \frac{m}{2Cn^{1/3}}n^{(\beta-\alpha)\epsilon} \geq (2C)^{-1}n^{2/3+(\beta-\alpha)\epsilon}$ and at most $m/p_k \leq Cn/p_k \leq 2C\frac{m}{n^{1/3}}n^{\beta\epsilon}n^{\alpha\epsilon} \leq 2C^2n^{2/3+(\alpha+\beta)\epsilon}$. Due to the occurrence of E, the path $W_{0,0}$ lies within $n^{2/3}$ of these points, and, due to $\beta > \alpha$, the path of $W_{0,0}$ does not intersect itself before time np_k. For the event F, at some point before the p_ℓ-th return there is a displacement of at least $n^{2/3+\delta\epsilon}$ from R. Since $\delta > \alpha + \beta$ the path must intersect itself.

The number of horizontal steps in the cycle to which $(0, 0)$ is rooted is at most $np_l \leq n^{4/3+\gamma\epsilon}$. We have demonstrated that $W_{0,0}$ hits its trace before the first moment j at which $\left|\sum_{i=1}^{j} X_i - j\right| \geq n^{2/3+\delta\epsilon}$. We learn that the number of vertical steps in the cycle is at most $np_l + n^{2/3+\delta\epsilon} \leq 2n^{4/3+\gamma\epsilon}$. Hence, on the event $E \cap F$, the cycle to which $(0, 0)$ is rooted has length at most $3n^{4/3+\gamma\epsilon}$.

By invariance under vertical translation, the probability that there exists a cycle whose length exceeds $3n^{4/3+\gamma\epsilon}$ is at most $m\mathbb{P}(E \cap F)$.

With the aid of Theorem 5.23 of [5],

$$\mathbb{P}(E) \geq 1 - e^{-n^{\beta\epsilon+o(1)}}. \tag{26}$$

To obtain an analogous bound on $\mathbb{P}(F)$, we split the time interval $[1, \ldots, np_l]$ into a succession of intervals, each but the last of length $n^{4/3+2\delta\epsilon}$. (We discard the final, shorter interval.) If F^c is to occur, the mean-zero random walk $X_j - j$ must make a displacement of at most $n^{2/3+\delta\epsilon}$ in each of these intervals. By the central limit theorem, the probability of such a displacement is at most a constant strictly less than one, uniformly in high values of n. These displacements being independent for the disjoint intervals under consideration, we find that $\mathbb{P}(F^c) \leq c^L$, where L is the number of such intervals, and c is a constant less than one. Note that

$$L \geq \frac{np_l}{n^{4/3+2\delta\epsilon}} - 1.$$

Using $p_l \geq \frac{p_{l+1}}{a_{l+1}+1} > \frac{p_{l+1}}{n^{\alpha\epsilon}+1} \geq 2^{-1}n^{(\gamma-\alpha)\epsilon}$, we conclude that

$$\mathbb{P}(F) \geq 1 - e^{-c'n^{(\gamma-\alpha-2\delta)\epsilon}}, \tag{27}$$

for some constant $c' > 0$. From (26) and (27), we obtain (25). □

Acknowledgments We thank Yuval Peres for comments and references. The work of R. Kenyon was supported by NSERC and NSF. This work was started while both authors were at the University of British Columbia.

References

1. Durrett, R.: Probability: Theory and Examples. The Wadsworth and Brooks/Cole Statistics/Probability Series, Belmont (1991)
2. Hardy, G.H., Wright, E.M.: An Introduction to the Theory of Numbers, 5th edn. Oxford Science Publications, Oxford (1979)
3. Kenyon, R., Wilson, D.: Resonance in the non-intersecting lattice path model. Probab. Theory Related Fields **130**(3), 289–318 (2004)
4. Hensley, D.: The largest digit in the continued fraction expansion of a rational number. Pac. J. Math. **151**(2), 237–255 (1991)
5. Petrov, V.V.: Limit Theorems of Probability Theory. Oxford University Press, Oxford (1995)
6. Richter, W.: Local limit theorems for large deviations. Theory Probab. Appl. **2**(2), 206–220 (1957)