Galton Board: limit theorems and recurrence

N. Chernov[1] and D. Dolgopyat[2]

Abstract

We study a particle moving in $\mathbb{R}^2$ under a constant (external) force and bouncing off a periodic array of convex domains (scatterers); the latter must satisfy a standard 'finite horizon' condition to prevent 'ballistic' (collision-free) motion. This model is known to physicists as Galton board (it is also identical to a periodic Lorentz gas). Previous heuristic and experimental studies have suggested that the particle's speed $v(t)$ should grow as $t^{1/3}$ and its coordinate $x(t)$ as $t^{2/3}$. We prove these conjectures rigorously; we also find limit distributions for the rescaled velocity $t^{-1/3}v(t)$ and position $t^{-2/3}x(t)$. In addition, quite unexpectedly, we discover that the particle's motion is recurrent. That means that a ball dropped on an idealized Galton board will roll down but from time to time it should bounce all the way back up (with probability one).

1 Introduction

Galton board [21, Chapter V], also known as quincunx or bean machine, is one of the simplest mechanical devices exhibiting stochastic behavior. It consists of a vertical (or inclined) board with interleaved rows of pegs. A ball thrown into the Galton board moves under gravitation and bounces off the pegs on its way down. If many balls are thrown into the quincunx, then one can observe a normal distribution of balls coming to rest on the machine floor.

[1]Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294

[2]Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742

In this paper we deal with an idealized infinite Galton board: we consider a ball moving in a bean machine of infinite length under a constant external field. We neglect friction and the spin of the ball. Our pegs are convex obstacles (scatterers) positioned periodically on the board and satisfying the 'finite horizon' condition (the latter means that the ball cannot move in any direction indefinitely without meeting a scatterer).

This model is identical to a periodic Lorentz gas. Historically, Lorentz gas (in the 3D space and without necessarily periodic position of scatterers) was introduced in 1905, see [26], to illustrate the transport of electrons in metals in a spatially homogeneous electric field. Periodic Lorentz gases have been later studied mathematically [2, 3, 33]. Without external fields, the periodic Lorentz gas reduces to a billiard system on its fundamental domain (a torus minus scatterers). This is a dispersing billiard (Sinai billiard); it preserves a Liouville (equilibrium) measure and has strong ergodic and statistical properties, in particular it exhibits diffusive behavior, see [2, 3, 5, 33, 35].

Under a constant external field, the moving particle is likely to accelerate indefinitely, thus the system does not even have a stationary measure (physicists say that there is no steady state). Such a non-stationary behavior makes mathematical studies very difficult and may explain the lack of rigorous results (until now), despite persistent interest in physics community [4, 23, 24, 27, 28, 30, 31]. To make things more tractable, one can remove the excess of energy in various ways (deterministically or stochastically). One way to do that is to modify the equations of motion by introducing the so-called Gaussian thermostat [9, 10, 28]; this will keep the particle's speed constant. A rigorous investigation of the Gaussian thermostatted Lorentz particle under a small external field is done in [9]: it is proven that the dynamics has a stationary measure (steady state), the particle exhibits diffusive behavior and, in addition, it slowly drifts with an average velocity proportional to the field (Ohm's law), see [9, 10]. If the field is not small, computer experiments [12] show that the particle's trajectory converges either to a fractal set or to a stable periodic orbit.

Here we return to an 'unaltered' periodic Lorentz gas in a constant external field, i.e. we do not apply Gaussian (or any other) thermostat. To deal with indefinite acceleration and the lack of finite invariant measures we develop non-traditional approaches and arrive at unusual results, see below.

We note that due to the conservation of the total energy, the particle's speed $v(t)$ depends on its displacement in the direction of the field $x(t)$, see Fig. 1; precisely we have $v^2(t) = ax(t) + b$ for some constants $a, b > 0$. So the

2

farther the particle travels (in the x direction), the faster it moves. On the other hand, higher speed leads to a stronger scattering effect, thus increasing the chances that the particle bounces back and hence temporarily decelerates (this is similar to Fermi, or diffusive shock acceleration [22, 36]).
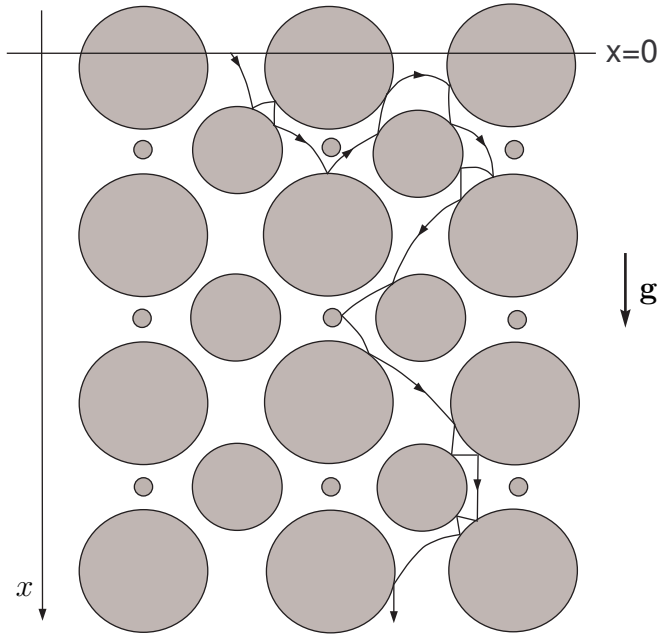

Figure 1: A trajectory of the Galton particle under an external field $\mathbf{g}$.

It turns out that the backscattering effect slows down the particle's drift in the direction of the field so much that its average displacement $\langle x(t) \rangle$ at time t will only grow as t^a with some $a < 1$. Physicists have estimated [24, 28, 30, 31], using heuristic and approximative arguments, as well as computer simulation, that the displacement of the particle typically grows as $t^{2/3}$. Due to the conservation of energy, the speed then grows as $t^{1/3}$.

We present here the first mathematically rigorous study of this model. We prove that, under certain conditions, the particle's displacement $x(t)$ indeed grows as $t^{2/3}$ and its velocity $v(t)$ grows as $t^{1/3}$. We also find limit distributions for the rescaled velocity $t^{-1/3}v(t)$ and rescaled position $t^{-2/3}x(t)$. In this respect our results agree with earlier heuristics.

On the other hand, we discover, quite surprisingly, that the particle's

3

motion is recurrent; i.e., with probability one

$$\liminf_{t\to\infty} v(t) \le v(0) \qquad \text{and} \qquad \liminf_{t\to\infty} x(t) \le x(0).$$

In other words, the particle slows down and effectively returns to a vicinity of its initial state infinitely many times! (Furthermore, we will show that our system satisfies the Poincaré Recurrence Theorem in the sense that for any set $\mathcal{Z}$ of positive Liouville measure, almost every point of $\mathcal{Z}$ visits $\mathcal{Z}$ infinitely often.)

Such recurrent behavior apparently contradicts common physical experience: we all know that a ball thrown into a real Galton board always rolls down and ends up on the floor. But on the idealized board, rather paradoxically, the ball will almost surely bounce all the way back up! (Naturally, this spectacular phenomenon is never observed in practice because physical collisions are non-elastic and the ball is subject to friction [4, 23, 27].)

The recurrence in our model leads to an additional complication: if the ball bounces back up too far, its speed may drop to almost zero, and its motion will not be chaotic enough for us to control it by our methods. Thus we need to prevent such returns, which we do in two different ways. First we can assume that our Galton board has 'open top' through which the ball simply escapes. Alternatively, we can close the top with a lid reflecting the ball back down every time it hits the lid on its way up. In the latter case the ball's speed will be bounded below $v \ge v_{\min} \gg 0$. We find that on a closed board the speed $v(t)$ will indeed grow as $t^{1/3}$ and the rescaled speed $t^{-1/3}v(t)$ will converge to a random variable that we describe in this paper.

Let us formulate our results precisely. To be consistent with other related papers ([24, 31] and especially [9]), we choose the coordinate frame so that the constant external field $\mathbf{g} = (g, 0)$ is directed along the x axis. We denote by $\mathbf{q} = (x, y)$ the particle's position and by $\mathbf{v}$ its velocity vector.

The particle has unit mass, so the equations of motion (between collisions with the scatterers) are

(1.1) $$\dot{\mathbf{q}} = \mathbf{v}, \qquad \dot{\mathbf{v}} = \mathbf{g}.$$

At every collision with scatterers, the velocity of the particle changes according to the law of elastic impact:

(1.2) $$\mathbf{v}^+ = \mathbf{v}^- - 2\langle \mathbf{v}, \mathbf{n_q}\rangle \mathbf{n_q},$$

4

where $\mathbf{v}^+$ and $\mathbf{v}^-$ refer to the postcollisional and precollisional velocities, respectively, $\mathbf{n_q}$ denotes the outward unit normal vector to the scatterer at the collision point $\mathbf{q}$, and $\langle \cdot, \cdot \rangle$ stands for the scalar product. The system preserves the total energy

(1.3) $$E = \tfrac{1}{2}[v(t)]^2 - gx(t),$$

where $v(t) = \|\mathbf{v}(t)\|$ is the particle's speed.

Since we are not interested in the displacement of the particle in the y direction, we can replace our infinite board with a strip $0 \leq y \leq L_y$ and impose periodic boundary conditions at $y = 0$ and $y = L_y$ (so y will be a cyclic coordinate). We assume a finite horizon, see above, to exclude collision-free ('ballistic') motion of the particle.

Our ball starts on the line $x = 0$ with its y coordinate uniformly distributed on $[0, L_y]$ (minus the scatterers that it crosses) and its initial velocity is uniformly distributed in a sector

(1.4) $$S_{V,\alpha} = \{\mathbf{v}(0)\colon c_1 V \leq \|\mathbf{v}(0)\| \leq c_2 V, \quad |\angle(\mathbf{v}(0), \mathbf{g})| \leq \alpha\}.$$

Here $0 < c_1 < c_2$ are two constants whose values are irrelevant, and we assume that V is large enough and α is small enough.

We distinguish between the 'open' machine where the ball coming back to the line $x = 0$ escapes and the 'closed' one where the line $x = 0$ acts as a mirror reflecting the ball back into $x > 0$.

Theorem 1. *In the open board the ball escapes through $x = 0$ with probability one.*

Theorem 2. *For the closed board there are constants c, V_0 such that if $V \geq V_0$, then $c\,t^{-1/3}v(t)$ converges, as $t \to \infty$, to a random variable with density*

$$\frac{3z}{\Gamma(2/3)} \exp\left[-z^3\right], \qquad z \geq 0.$$

Accordingly, $2gc^2 t^{-2/3}x(t)$ converges to a random variable with density

$$\frac{3}{2\Gamma(2/3)} \exp\left[-z^{3/2}\right], \qquad z \geq 0.$$

2 Preliminaries

Here we introduce basic notation and tools used in our analysis.

We denote by $\mathcal{D}$ the part of the Galton board available for the particle (i.e., the half-infinite cylinder $\{x \geq 0, \ \ 0 \leq y \leq L_y\}$ minus the scatterers). Thee phase space of a particle moving in $\mathcal{D}$ is $\mathcal{M} = \mathcal{D} \times \mathbb{R}^2$, where $\mathbb{R}^2$ accounts for the velocity vector. Equations of motion (1.1) define a continuous-time dynamical system (flow) $\Phi^t \colon \mathcal{M} \to \mathcal{M}$. It is a Hamiltonian flow, hence it preserves the Liouville measure, which is just the Lebesgue volume in $\mathcal{M}$. The 4D space $\mathcal{M}$ is foliated by the Φ^t-invariant 3D submanifolds of constant energy

$$\mathcal{M}_E = \{(\mathbf{q}, \mathbf{v}) \in \mathcal{M} \colon \tfrac{1}{2}v^2 - gx = E\}.$$

We also consider the 3D collision space Ω, which is the cross-section of $\mathcal{M}$ consisting of points $(\mathbf{q}, \mathbf{v}) \in \mathcal{M}$, where $\mathbf{q} \in \partial\mathcal{D}$ is the point of impact and $\mathbf{v}$ is the outgoing (postcollisional) velocity vector (pointing inward $\mathcal{D}$). Most of the time we deal with the closed board, so that the line $x = 0$ will be a part of the boundary (but the lines $y = 0$ and $y = L_y$ are never parts of $\partial\mathcal{D}$ because y is a cyclic coordinate).

We denote points of the collision space Ω by (q, u, K), where $q \in \partial\mathcal{D}$ belongs to the board's boundary, $u = \mathbf{v}/\|\mathbf{v}\|$ is a unit vector (pointing in the direction of the postcollisional velocity), and $K = \frac{1}{2}\|\mathbf{v}\|^2$ is the kinetic energy. Clearly the coordinates q, u, K are effectively one-dimensional (for this reason we use the lower case characters). We denote by $\mathcal{F} \colon \Omega \to \Omega$ the induced collision map. It preserves the 2D submanifolds $\Omega_E = \Omega \cap \mathcal{M}_E$ of constant total energy. The Liouville measure on $\mathcal{M}$ induces a smooth $\mathcal{F}$-invariant measure μ on Ω.

If we remove the external field (i.e. set $g = 0$), the particle will move with constant velocity between collisions and its speed will remain unchanged at all times, i.e. $K \equiv \text{const}$. Geometrically, its trajectory will not depend on K, and it will coincide with the trajectory of a billiard particle (moving at unit speed). Its collisions can be described by two variables, q and u, only. We denote by $\Omega_* = \{(q, u)\}$ the billiard collision space and by $\mathcal{F}_* \colon \Omega_* \to \Omega_*$ the billiard collision map. The map $\mathcal{F}_*$ preserves a smooth measure μ_* on Ω_*; its density is proportional to $\langle u, \mathbf{n}_q \rangle$, in the notation of (1.2). Observe that the spaces Ω and Ω_* are not compact and both measures μ and μ_* are unbounded (σ-finite).

Furthermore, due to the periodicity of our array of scatterers, we can

project the billiard dynamics onto a fundamental domain $\mathcal{D}_0$ of $\mathcal{D}$, after which the collision space will be compact, we denote it by Ω_0 (in fact, Ω_* is a countable union of replicas of Ω_0). The billiard map $\mathcal{F}_*$ commutes with the projection $\Omega_* \to \Omega_0$, thus it naturally generates a map $\mathcal{F}_0 \colon \Omega_0 \to \Omega_0$. The latter preserves a smooth *finite* measure μ_0 on Ω_0 (which is just the restriction of μ_* to Ω_0). We assume that μ_0 is normalized, i.e. it is a probability measure. The billiard map $\mathcal{F}_*$ is strongly hyperbolic, it has stable and unstable cones, as well as stable and unstable curves with uniform expansion and contraction rates, cf. [11, Chapters 4–5].

We will also use another dynamical system – the one generated by a particle moving in a small external field $\mathbf{e} = (e, 0)$ at unit speed and with Gaussian thermostat. Its motion is governed by equations

$$(2.1) \qquad\qquad d\mathbf{q}/dt = \mathbf{v}, \qquad d\mathbf{v}/dt = \mathbf{e} - \langle \mathbf{e}, \mathbf{v} \rangle \mathbf{v},$$

which ensure the unit speed throughout: $\|\mathbf{v}(t)\| \equiv 1$. The 'thermostatted' dynamics induces another collision map, we call it $\mathcal{F}_{*,e} \colon \Omega_* \to \Omega_*$. It also commutes with the above projection $\Omega_* \to \Omega_0$, thus it naturally induces a map $\mathcal{F}_e \colon \Omega_0 \to \Omega_0$. For small $e > 0$, the map $\mathcal{F}_e$ has been thoroughly investigated in [9, 6], it was proved to have a unique Sinai-Ruelle-Bowen (SRB) measure, μ_e, which has strong statistical properties.

Next, recall that our initial condition $(\mathbf{q}(0), \mathbf{v}(0))$ is chosen randomly according to a smooth probability measure, we denote it by μ_{ini}, which is concentrated on the set

$$(2.2) \qquad\qquad \Omega_{\mathrm{ini}} = \{x(0) = 0, \mathbf{v}(0) \in S_{V,\alpha}\} \subset \Omega,$$

cf. (1.4). This measure is not invariant under the collision map $\mathcal{F}$, and its images can be described (rather informally) as follows. Let us partition the sector (1.4) into small subdomains (cells) $D_i \subset \Omega_{\mathrm{ini}}$ and represent a smooth measure μ_{ini} on Ω_{ini} as a weighted sum of its restrictions to those cells (this is called 'coarse-graining' in physics).

Now the motion of our particle can be regarded as a small perturbation of the billiard dynamics in $\mathcal{D}$, as long as the kinetic energy K remains large enough (which is guaranteed by our assumption on the largeness of V in (1.4), see below). Thus the image of a small domain $D \subset \Omega$ under the map $\mathcal{F}^n$ gets strongly expanded in the unstable direction of the billiard map $\mathcal{F}_*$, strongly contracted in the stable direction of the billiard map $\mathcal{F}_*$, slightly deformed in the transversal direction (along the K coordinate axis), and possibly cut

by singularities into many pieces. Thus, $\mathcal{F}^n(D)$ will soon look like a union of one-dimensional curves, each of which resembles an unstable manifold of the billiard map $\mathcal{F}_*$. Henceforth the measure $\mathcal{F}^n(\mu_{\text{ini}})$ will evolve as a weighted sum of smooth measures on unstable curves. Furthermore, the total energy on such unstable curves will be virtually constant and its oscillations will decrease as $n \to \infty$.

Motivated by these observations we introduce a family of smooth one-dimensional measures on certain curves $\gamma \subset \Omega_E$ (we call them unstable curves) for large enough E. We follow a general scheme developed in [7, Section 3.2] and [11, Section 7.4].

A detailed construction of invariant (stable and unstable) cones for small perturbations of billiards, which have an integral of motion, is carried out in [6]. Thus it applies to our flow $\Phi^t \colon \mathcal{M}_E \to \mathcal{M}_E$ and our map $\mathcal{F} \colon \Omega_E \to \Omega_E$ as long as E is large enough, say $E > E_{\min} \gg 0$. We will assume that V in (1.4) is sufficiently large to ensure $E > E_{\min}$. Thus there are stable and unstable cones, $\mathcal{C}^s(X)$ and $\mathcal{C}^u(X)$, at every point $X \in \Omega_E$ for all $E > E_{\min}$. The derivative $D_X \mathcal{F}$ expands unstable vectors and contracts stable vectors:

$$\|D_X \mathcal{F}^n v\| \le c\vartheta^n \|v\| \quad \forall v \in \mathcal{C}^s(X), \quad \|D_X \mathcal{F}^{-n} v\| \le c\vartheta^n \|v\| \quad \forall v \in \mathcal{C}^u(X),$$

for all $n \ge 1$, here $\vartheta < 1$ and $c > 0$ are the *hyperbolicity constants* (there is an adapted metric in which $c = 1$, see [7, Section 4.2]).

Now a curve $\gamma \subset \Omega$ is stable (unstable) if its tangent vector lies in the stable (resp., unstable) cone at every point $X \in \gamma$. The image of a stable (unstable) curve under $\mathcal{F}^{-1}$ (resp., under $\mathcal{F}$) is a finite union of stable (resp., unstable) curves. (Note that the image of one curve may consist of several curves because the map $\mathcal{F}$ has singularities, which are caused by grazing collisions.) There is another technicality here – stable and unstable curves need be *homogeneous* in order for us to control distortions; but the details are irrelevant for us, they may be found in [6, page 216] or [11, Chapter 5].

Now a *standard pair* $\ell = (\gamma, \rho)$ is a C^2 smooth unstable curve $\gamma \subset \Omega_E$ with a probability density ρ whose logarithm is 'dynamically Hölder continuous' (this notion was introduced by Young [35]); the latter means that for some constants $C_{\mathcal{D}} > 0$ and $\theta_{\mathcal{D}} < 1$ and all $X, Y \in \gamma$

$$(2.3) \qquad |\ln \rho(X) - \ln \rho(Y)| \le C_{\mathcal{D}} \theta_{\mathcal{D}}^{s(X,Y)}$$

where $s(X, Y)$ is the *separation time* (the first iteration of $\mathcal{F}$ when the images of X and Y get separated by singularities); we refer to [7, Section 4.4] and

[11, Chapter 7] for precise definitions. We only note that $\rho(X)/\rho(Y) \leq e^{C_D}$, i.e. the density ρ is (uniformly) bounded away from zero and infinity.

For any standard pair $\ell = (\gamma, \rho)$ we denote by $\mathbb{P}_\ell$ the probability measure on γ with density ρ. For any function $A \colon \Omega \to \mathbb{R}$ we shall write

$$\mathbb{E}_\ell(A) = \int_\gamma A \, d\mathbb{P}_\ell = \int_\gamma A(X)\rho(X) \, dX.$$

The class of standard pairs remains invariant under $\mathcal{F}$ in the following sense: for any standard pair $\ell = (\gamma, \rho)$ and $k \geq 1$

$$(2.4) \qquad \mathbb{E}_\ell(A \circ \mathcal{F}^k) = \sum_{i=1}^\infty c_{i,k} \, \mathbb{E}_{\ell_{i,k}}(A)$$

where $\ell_{i,k} = (\gamma_{i,k}, \rho_{i,k})$ are standard pairs (here $\gamma_{i,k}$'s are the components of $\mathcal{F}^k(\gamma)$ and $\rho_{i,k}$ denotes the conditional density of the measure $\mathcal{F}^k(\mathbb{P}_\ell)$ on the component $\gamma_{i,k} \subset \mathcal{F}^k(\gamma)$); we naturally have $\sum_i c_{i,k} = 1$. Roughly speaking, a standard pair is transformed by $\mathcal{F}^k$ into a finite or countable family of standard pairs (with a factor measure defined by the sequence of the coefficients $\{c_{i,k}\}$).

This motivates the consideration of *standard families* – countable or even uncountable families $\mathcal{G} = \{\ell_\alpha\} = \{(\gamma_\alpha, \rho_\alpha)\}$, $\alpha \in \mathfrak{A}$, of standard pairs with a probability factor measure $\lambda_\mathcal{G}$ on the index set $\mathfrak{A}$. (The energy need not be the same on different standard pairs in $\mathcal{G}$.) Such a family induces a probability measure $\mu_\mathcal{G}$ on the union $\cup_\alpha \gamma_\alpha$ (and thus on Ω) defined by

$$\mu_\mathcal{G}(B) = \int_\mathfrak{A} \mathbb{P}_{\ell_\alpha}(B \cap \gamma_\alpha) \, d\lambda_\mathcal{G}(\alpha) \qquad \forall B \subset \Omega.$$

For any function $A \colon \Omega \to \mathbb{R}$ we shall write

$$\mathbb{E}_\mathcal{G}(A) = \int_\Omega A \, d\mathbb{P}_\mathcal{G} = \int_\mathfrak{A} \int_{\gamma_\alpha} A \, d\mathbb{P}_{\ell_\alpha} \, d\lambda_\mathcal{G}(\alpha).$$

Now it is clear that the image of a standard family under $\mathcal{F}^k$, $k \geq 1$, is always a standard family. It is important to control the size of curves γ_α in a standard family $\mathcal{G}$. Let

$$\mathcal{Z}_\mathcal{G} \colon = \sup_{\varepsilon > 0} \frac{\mu_\mathcal{G}(r_\mathcal{G} < \varepsilon)}{\varepsilon} = \sup_{\varepsilon > 0} \frac{\int \mathbb{P}_{\ell_\alpha}\big(X \in \gamma_\alpha \colon r_\mathcal{G}(X) < \varepsilon\big) \, d\lambda_\mathcal{G}(\alpha)}{\varepsilon}.$$

9

where $r_{\mathcal{G}}(X)$ denotes the distance from $X \in \gamma_\alpha$ to the closer endpoint of the curve γ_α. In other words, X divides γ_α into two subcurves, and $r_{\mathcal{G}}(X)$ denotes the length of the shorter one. Observe that if $\mathcal{G}$ consists of a single standard pair $\ell = (\gamma, \rho)$, then $\mathcal{Z}_{\mathcal{G}} \sim [\text{length}(\gamma)]^{-1}$.

A standard pair $\ell = (\gamma, \rho)$ is *proper* if $\text{length}(\gamma) \geq \delta_0 = \delta_0(\mathcal{D})$, where $\delta_0 > 0$ is a small constant that depends on the board $\mathcal{D}$, see [7, 11]. Similarly, a standard family $\mathcal{G}$ is said to be *proper* if $\mathcal{Z}_{\mathcal{G}} \leq C_0 = C_0(\mathcal{D})$, see [11]. We choose C_0 so large that any proper standard pair is a proper standard family.

It is not hard to see that the initial measure $\mu_{\text{ini}} = \mu_{\mathcal{G}}$ for a proper standard family $\mathcal{G}$ (one only needs to represent μ_{ini} by its conditional distributions on the fibers of a rather arbitrary smooth foliation of the sector (1.4) by unstable curves; such foliations are also used elsewhere, see e.g. [7, Section 3.2]).

Standard pairs and families can also be defined in Ω_0, so that they will be invariant (in the above sense) under the billiard map $\mathcal{F}_0$ and under the 'thermostatted' map $\mathcal{F}_e$ for small e; in fact such standard pairs are constructed in [7, Section 4.4] and [11, Section 7.4].

Lastly we state an important lemma that ensures that the images of any short standard pair grow fast, effectively at an exponential rate. For the proof we refer to [11, Chapter 5], [7, Lemma 4.10], and [6, Proposition 5.3] (though our situation is slightly different from those treated there, the two key components of the argument – a uniform expansion of unstable curves and complexity bounds – are clearly present here, too).

Let $\ell = (\gamma, \rho)$ be a standard pair and for $n \geq 1$ and $X \in \gamma$ let $r_n(X)$ denote the distance from the point $\mathcal{F}^n(X)$ to the closer endpoint of the component $\gamma_{i,n} \subset \mathcal{F}^n(\gamma)$ that contains the point $\mathcal{F}^n(X)$.

Lemma 2.1 ("Growth lemma"). *We have three properties:*
(a) *There are constants $\beta_1 \in (0, 1)$ and $\beta_2 > 0$, such that for any $\varepsilon > 0$*

$$(2.5) \qquad \mathbb{P}_\ell\big(X \colon \ r_n(X) < \varepsilon\big) \leq (\beta_1/\vartheta)^n \, \mathbb{P}_\ell\big(X \colon r_0 < \varepsilon \vartheta^n\big) + \beta_2 \varepsilon,$$

recall that $\vartheta < 1$ is the hyperbolicity constant;
(b) *There are constants $\beta_3, \beta_4 > 0$, such that if $n \geq \beta_3 \big| \ln \text{length}(\gamma) \big|$, then for any $\varepsilon > 0$ we have $\mathbb{P}_\ell\big(X \colon \ r_n(X) < \varepsilon\big) \leq \beta_4 \varepsilon$.*
(c) *There are constants $\beta_5, \beta_6 > 0$ and $\theta \in (0, 1)$ such that for any $n_2 > n_1 > \beta_5 \big| \ln \text{length}(\gamma) \big|$ we have*

$$\mathbb{P}_\ell\Big(X \colon \ \max_{n_1 < i < n_2} r_i(X) < \delta_0\Big) \leq \beta_6 \theta^{n_2 - n_1}$$

10

All these estimates are uniform in $\ell = (\gamma, \rho)$.

Images of standard families grow up in a similar way. If a standard family $\mathcal{G}$ is not proper (i.e. most of its standard pairs are 'too short'), then its image $\mathcal{G}_n$ under $\mathcal{F}^n$ satisfies

$$\mathcal{Z}_{\mathcal{G}_n} \leq c_1 \vartheta^n \mathcal{Z}_{\mathcal{G}} + c_2$$

for some constants $c_1, c_2 > 0$, see [11, page 171]. This effectively ensures an exponential growth of predominantly short standard families.

Many of our statements will be formulated for proper standard pairs, but they remain valid for proper standard families on which the energy E is constant; this is a common fact [7, 11].

3 Approximations

Here we construct two convenient approximations to our dynamics. First, if the speed $\|\mathbf{v}(t)\|$ is large, the velocity $\mathbf{v}(t)$ changes very little, only by $\mathcal{O}(1/\|\mathbf{v}(t)\|)$, between collisions, so the particle's path between collisions goes closely to a link of a billiard orbit. Billiard approximations are often used in the studies of non-billiard mechanical models [9, 7].

Our system can be even better approximated by a 'thermostatted' particle (2.1) in a small external field $\mathbf{e}$, so that we can use the results of [9, 6] to our advantage. Indeed, suppose our particle collides with a scatterer at a point $q_n \in \partial \mathcal{D}$ at time $t_n > 0$ (here n is the collision number), and its next collision occurs at a point $q_{n+1} \in \partial \mathcal{D}$. We change time variable $\hat{t} = (t - t_n)\|\mathbf{v}(t_n)\|$ so that $\|d\mathbf{q}/d\hat{t}|_{\hat{t}=0}\| = 1$ (the particle has unit speed at the moment of the nth collision). Then the equations of motion (1.1) transform to

$$(3.1) \qquad d\mathbf{q}/d\hat{t} = \hat{\mathbf{v}}, \qquad d\hat{\mathbf{v}}/d\hat{t} = \boldsymbol{\varepsilon} := \mathbf{g}/\|\mathbf{v}(t_n)\|^2.$$

This trajectory is $\mathcal{O}(\|\boldsymbol{\varepsilon}\|)$-close to the billiard trajectory moving with a constant (unit) velocity $\hat{\mathbf{v}}(0)$.

On the other hand, the motion with a Gaussian thermostat is governed by

$$(3.2) \qquad d\mathbf{q}/d\hat{t} = \hat{\mathbf{v}}_G, \qquad d\hat{\mathbf{v}}_G/d\hat{t} = \boldsymbol{\varepsilon} - \langle \boldsymbol{\varepsilon}, \hat{\mathbf{v}}_G \rangle \hat{\mathbf{v}}_G,$$

where the force $\boldsymbol{\varepsilon} - \langle \boldsymbol{\varepsilon}, \hat{\mathbf{v}}_G \rangle \hat{\mathbf{v}}_G$ is obtained by projecting the constant force $\boldsymbol{\varepsilon}$ of (3.1) onto the direction orthogonal to $\hat{\mathbf{v}}_G$. Thus our trajectory (3.1) will
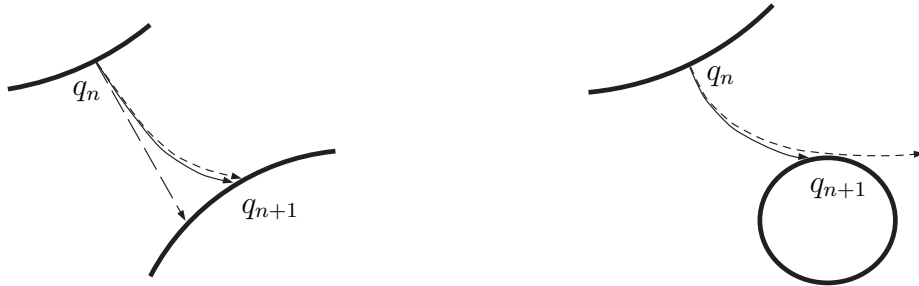
Figure 2: Our moving particle (solid line) is approximated by a billiard ball (straight dashed line) and by a Gaussian thermostatted particle (curved dashed line).

be $\mathcal{O}(\|\varepsilon\|^2)$-close to that of (3.2). In particular, the next collision point of the trajectory (3.2) will be $\mathcal{O}(\|\varepsilon\|^2)$-close to $\mathbf{q}_{n+1}$, unless the latter lies in the $\mathcal{O}(\|\varepsilon\|^2)$-vicinity of a singularity caused by a grazing collision, see Fig. 2 (right).

We now construct an approximation to our system on a longer time interval. Pick a moment $t_0 > 0$ and a small $\varepsilon > 0$. We will use a new (fast) time variable $s = (t - t_0)/\sqrt{\varepsilon}$, in which the particle's velocity is

$$(3.3) \qquad \tilde{\mathbf{v}}(s) = d\mathbf{q}/ds = \varepsilon^{1/2}\mathbf{v}(t(s))$$

and its kinetic energy is

$$(3.4) \qquad \tilde{K}(s) = \tfrac{1}{2}\|\tilde{\mathbf{v}}(s)\|^2 = \varepsilon K(t(s)),$$

where $K(t) = \tfrac{1}{2}\|\mathbf{v}(t)\|^2$. We will call this ε-*rescaled dynamics*.

Denote by s_n the moment of the nth collision (we start counting collisions after $s = 0$). At the nth collision we denote by $q_n \in \partial \mathcal{D}$ the collision point and by $u_n = \tilde{\mathbf{v}}_n/\|\tilde{\mathbf{v}}_n\|$ the normalized postcollisional velocity vector. Observe that

$$(3.5) \qquad d\tilde{\mathbf{v}}/ds = \varepsilon \mathbf{g},$$

i.e. the particle effectively moves in a small field $\varepsilon \mathbf{g}$. Now the collision map acts in the ε-rescaled coordinates as

$$(3.6) \qquad \mathcal{F}(q_n, u_n, \tilde{K}_n) = (q_{n+1}, u_{n+1}, \tilde{K}_{n+1})$$

12

where $\tilde{K}_n = \tilde{K}(s_n)$. Observe that the pair $X_n = (q_n, u_n)$ changes rapidly and chaotically, while $\tilde{K}_n$ changes slowly. Thus we are in a framework of 'slow-fast' systems, where X is a 'fast variable' and $\tilde{K}$ is a 'slow variable'. More precisely, since

$$\frac{d\tilde{K}}{ds} = \frac{d}{ds} \frac{1}{2} \langle \tilde{\mathbf{v}}, \tilde{\mathbf{v}} \rangle = \varepsilon \langle \tilde{\mathbf{v}}, \mathbf{g} \rangle,$$

we have

$$(3.7) \qquad \tilde{K}_{n+1} - \tilde{K}_n = \varepsilon \int_{s_n}^{s_{n+1}} \langle \tilde{\mathbf{v}}, \mathbf{g} \rangle \, ds = \varepsilon g L_n$$

where L_n denotes the displacement of the particle, in the x direction, between the nth and $(n+1)$st collisions. We denote this displacement by $L(q, u, \tilde{K})$, because it is a function of $(q, u, \tilde{K})$, and we write $L_n = L(q_n, u_n, \tilde{K}_n)$.

Next we approximate the motion between q_n and q_{n+1} by that of a thermostatted particle as described above. That particle starts at q_n with the unit speed u_n and then moves in the thermostatted constant field $\boldsymbol{\varepsilon}_n = \varepsilon \mathbf{g}/\|\tilde{\mathbf{v}}_n\|^2$ in accordance with (3.1) and (3.2). In the 'adjusted' time $\hat{s} = (s - s_n)\|\tilde{\mathbf{v}}_n\|$ we have

$$d\mathbf{q}/d\hat{s} = \hat{\mathbf{v}}, \qquad d\hat{\mathbf{v}}/d\hat{s} = \boldsymbol{\varepsilon}_n - \langle \boldsymbol{\varepsilon}_n, \hat{\mathbf{v}} \rangle \hat{\mathbf{v}},$$

which ensures a constant unit speed $\|\hat{\mathbf{v}}\| \equiv 1$. For the thermostatted particle, we record the next collision point $\hat{q}_{n+1}$ and the corresponding outgoing (unit) velocity vector $\hat{u}_{n+1}$. Then

$$(3.8) \qquad \hat{X}_{n+1} = (\hat{q}_{n+1}, \hat{u}_{n+1}) = \mathcal{F}_{*,e_n}(q_n, u_n)$$

where

$$(3.9) \qquad e_n = \|\boldsymbol{\varepsilon}_n\| = \frac{\varepsilon g}{2\tilde{K}_n}$$

and $\mathcal{F}_{*,e}$ again denotes the collision map of the particle moving in a constant external field $(e, 0)$ at unit speed with a Gaussian thermostat. Let $L_e(q, u)$ denote the corresponding displacement in the x direction of the thermostatted particle between collisions.

Recall that our original particle moves according to (3.5), i.e. its trajectory from X_n to X_{n+1} is identical to that of a particle starting at X_n at *unit speed* and moving toward X_{n+1} in a constant field $(e, 0) = (e_n, 0)$ (instead of

13

$\varepsilon\mathbf{g}$) without a thermostat. Now denote by $\mathcal{U}_e^1$ the set of points $(X, \tilde{K}) \in \Omega$ where these two trajectories land on different scatterers, i.e. where $L(X, \tilde{K}) - L_e(X) = \mathcal{O}(1)$. Clearly $\mathcal{U}_e^1$ is a strip of width $\mathcal{O}(e^2)$ between the singularity curves of the corresponding maps $\mathcal{F}$ and $\mathcal{F}_e$. Now we have

$$(3.10) \qquad L_n - L_{e_n}(X_n) = \begin{cases} \mathcal{O}(1) & \text{in } \mathcal{U}_{e_n}^1 \\ \mathcal{O}(e_n^2) & \text{elsewhere} \end{cases}$$

Now we can rewrite (3.7) as

$$(3.11) \qquad \tilde{K}_{n+1} - \tilde{K}_n = \varepsilon g L_{e_n}(X_n) + \mathcal{R}_1,$$

where

$$(3.12) \qquad \mathcal{R}_1 = \begin{cases} \mathcal{O}(\varepsilon) & \text{in } \mathcal{U}_{e_n}^1 \\ \mathcal{O}(\varepsilon^3 / \tilde{K}_n^2) & \text{elsewhere} \end{cases}$$

Let us also relate $L_{e_n}(X_n)$ to $L_0(X_n)$, which is the displacement in the x direction of the billiard (straight) trajectory starting at X_n and running till the next collision. Denote by $\mathcal{U}_e^2$ the set of points where the corresponding trajectories in these two dynamics land on different scatterers, i.e. where $L_e(X) - L_0(X) = \mathcal{O}(1)$. Clearly $\mathcal{U}_e^2$ is a strip of width $\mathcal{O}(e)$ between the singularity curves of the corresponding maps $\mathcal{F}_e$ and $\mathcal{F}_0$. Then we have

$$(3.13) \qquad L_e(X) - L_0(X) = \begin{cases} \mathcal{O}(1) & \text{if } X \in \mathcal{U}_e^2 \\ \mathcal{O}(e) & \text{otherwise} \end{cases}$$

4   Thermostatted dynamics

Here we recall and sharpen certain results of [9] on the 'thermostatted' collision map $\mathcal{F}_e$. For brevity, we use notation $\mu_e(A) = \int_{\Omega_0} A \, d\mu_e$ for any function $A \colon \Omega_0 \to \mathbb{R}$.

It is proved in [9, Proposition 3] that for any piecewise Hölder continuous function $A \colon \Omega_0 \to \mathbb{R}$

$$(4.1) \qquad \mu_e(A) = \mu_0(A) + \sum_{n=1}^{\infty} \mu_0\Big((A \circ \mathcal{F}_e^n) \big[1 - \exp(-eL_e)\big]\Big).$$

In physics, such expansions are known as Kawasaki formulas [34] (they describe a 'nonlinear response' of a system to perturbations). We note that

$\exp(-eL_e)$ is just the Jacobian of the map $\mathcal{F}_e$, see [9, page 584]. The terms in the series (4.1) decrease uniformly in e, and they are bounded by a stretched exponential function, see below.

One of the main results of [9] is a discrete-time version of Ohm's law:

$$(4.2) \qquad \mu_e(L_e) = \frac{e}{2} \sum_{n=-\infty}^{\infty} \mu_0\big((L_0 \circ \mathcal{F}_0^n)\, L_0\big) + o(e).$$

This follows from (4.1) via an approximation argument, see [9, page 585]. We note that the correlations in the series in (4.2) is symmetric about $n = 0$.

We need to slightly improve the error term in (4.2):

Proposition 4.1. *There is a $\delta > 0$ such that*

$$(4.3) \qquad \mu_e(L_e) = \frac{e}{2} \sum_{n=-\infty}^{\infty} \mu_0\big((L_0 \circ \mathcal{F}_0^n)\, L_0\big) + \mathcal{O}(e^{1+\delta}).$$

Proof. Let $A = \frac{1}{2}L_e + \frac{1}{2}L_e \circ \mathcal{F}_e^{-1}$. Due to the $\mathcal{F}_e$-invariance of the measure μ_e, and by (4.1), we have

$$\mu_e(L_e) = \mu_e(A) = \mu_0(A) + \sum_{n=1}^{\infty} \mu_0\Big((A \circ \mathcal{F}_e^n)\big[1 - \exp(-eL_e)\big]\Big)$$

$$= \mu_0(A) + {\sum_{n=0}^{\infty}}' \mu_0\Big((L_e \circ \mathcal{F}_e^n)\big[1 - \exp(-eL_e)\big]\Big),$$

where we use the following conventional summation sign:

$$\sum_{n=0}^{\infty}{}' a_n = \tfrac{1}{2}\, a_0 + a_1 + a_2 + a_3 + \cdots .$$

It is known [9, page 585] that $\mu_0(A) = 0$ (this follows from the invariance of the measure μ_0 under the time reversal). Now the error term in (4.3) can be expressed as

$$\mathcal{R}^* = e {\sum_{n=0}^{\infty}}' \mu_0\big((L_0 \circ \mathcal{F}_e^n - L_0 \circ \mathcal{F}_0^n)\, L_0\big)$$

$$+ e {\sum_{n=0}^{\infty}}' \mu_0\big((L_e \circ \mathcal{F}_e^n - L_0 \circ \mathcal{F}_e^n)\, L_0\big)$$

$$(4.4) \qquad + e {\sum_{n=0}^{\infty}}' \mu_0\big((L_e \circ \mathcal{F}_e^n)\, \Delta_e\big),$$

15

where
$$\Delta_e = e^{-1}\big[1 - \exp(-eL_e) - eL_0\big].$$

For brevity, we rewrite (4.4) as

$$\mathcal{R}^* = e \sum_{n=0}^{\infty}{}' I_n^{(1)} + e \sum_{n=0}^{\infty}{}' I_n^{(2)} + e \sum_{n=0}^{\infty}{}' I_n^{(3)}.$$

It is shown in [9, Theorem 17] that the correlation-type terms $I_n^{(k)}$, $k = 1, 2, 3$, decay uniformly in e, and they are bounded by a stretched exponential function: $|I_n^{(k)}| \leq C\theta^{\sqrt{n}}$, where $C > 0$ and $\theta \in (0, 1)$ are constants (this bound can be upgraded to an exponential one, in which $\sqrt{n}$ is replaced by n, by using the results of [6], but we will not need that).

We start analyzing the terms $I_n^{(1)}$, which are the hardest to deal with. First we will estimate $\Theta = A \circ \mathcal{F}_e^n - A \circ \mathcal{F}_0^n$ assuming that A is a smooth function on Ω_0. We use the following 'telescoping' sum:

$$\Theta = \sum_{i=0}^{n-1}\big(A \circ F_i - A \circ F_{i+1}\big),$$

where $F_i = \mathcal{F}_0^i \circ \mathcal{F}_e^{n-i}$. Due to the smoothness of A,

$$A(F_i(X)) - A(F_{i+1}(X)) = A(\mathcal{F}_0^i(Y')) - A(\mathcal{F}_0^i(Y'')),$$
(4.5)
$$\leq C \operatorname{dist}\big(\mathcal{F}_0^i(Y'), \mathcal{F}_0^i(Y'')\big),$$

where $Y' = \mathcal{F}_e^{n-i}(X)$ and $Y'' = \mathcal{F}_0(\mathcal{F}_e^{n-i-1}(X))$. Here and in the rest of the paper we denote by C various positive constants whose exact values are not essential. It is clear that

(4.6)
$$\operatorname{dist}(Y', Y'') = \mathcal{O}(e),$$

unless the point $\mathcal{F}_e^{n-i-1}(X)$ lies in the $\mathcal{O}(e)$-vicinity of singularities; that vicinity has measure $\mathcal{O}(e)$, which is too small to affect our estimates. (We will use our estimate for $n = \mathcal{O}(|\ln e|)$, so $\mathcal{O}(e)$ errors terms will give a negligible contribution to $\mathcal{R}^*$.)

Now assume that (4.6) holds and that there is a stable manifold $W^s(Y')$ for the billiard map $\mathcal{F}_0$, which passes through the point Y' and extends by the distance $2\hat{C}e$ in both directions where $\hat{C}$ is much larger than the implied constant in (4.6) and (if such a stable manifold does not exist, then Y' belongs

16

to a subset of points where stable manifolds are shorter than $2\hat{C}e$, and that subset has measure $\mathcal{O}(e)$, see [11, Section 4.12], so we can disregard that tiny set). Now if $\hat{C}$ is sufficiently large we can connect the point Y'' with a point $Y''' \in W^s(Y')$ by an unstable curve γ. Observe that $\text{dist}(Y'', Y''') \le 3\hat{C}e$. For any unstable curve $\gamma \subset \Omega_0$ and $X \in \gamma$ we denote by $J_\gamma \mathcal{F}_0^n(X)$ the Jacobian of the map $\mathcal{F}_0^n$ restricted to γ at X (i.e., $J_\gamma \mathcal{F}_0^n(X)$ is the factor of expansion of γ under $\mathcal{F}_0^n$). We have two cases:

(a) if the map $\mathcal{F}_0^i$ is smooth on γ, then

$$\text{(4.7)} \qquad \text{dist}(F_i(X), F_{i+1}(X)) \le Ce + C \int_\gamma \left| J_\gamma \mathcal{F}_0^i(Z) \right| dZ$$

(b) otherwise we have $\text{dist}(F_i(X), F_{i+1}(X)) = \mathcal{O}(1)$.

Though the factor of expansion of unstable curves under $\mathcal{F}_0$ is unbounded near singularities (thus the integral in (4.7) may be arbitrarily large for some X's), the situation is manageable because we only need to estimate the integral (4.4), i.e. we need to bound $\int_{\Omega_0} (A \circ F_i - A \circ F_{i+1}) L_0 \, d\mu_0$. To this end we foliate Ω_0 by smooth unstable curves on which the measure $L_0 \, d\mu_0$ has smooth and uniformly bounded conditional densities (with respect to the Lebesgue measure on those curves), thus we get $L_0 \, d\mu_0 = d\mu_{\mathcal{G}}$ for a proper standard family $\mathcal{G}$, see definition in Section 2. Then we take the curve γ in (4.7) to be a piece of the image of an original fiber of our foliation under $\mathcal{F}_0 \circ \mathcal{F}_e^{n-i-1}$ passing through Y''.

Lemma 4.2. *There are large constants $C, \Lambda > 1$ such that for any proper standard pair $\ell = (\tilde{\gamma}, \rho)$ and $i \ge 1$*

$$\int_{\tilde{\gamma}} \left| A \circ F_i - A \circ F_{i+1} \right| d\mathbb{P}_\ell \le Ce\Lambda^i.$$

Proof. We need to integrate the estimates (a) and (b) above. The contribution of the first term of (4.7) is negligible $\mathcal{O}(e)$. To estimate the second term of (4.7) we use the invariance of standard pairs (2.4) to write

$$\mathbb{E}_\ell \left(B \circ \mathcal{F}_0 \circ \mathcal{F}_e^{n-i-1} \right) = \sum_k c_{k,i} \mathbb{E}_{\ell_{k,i}}(B),$$

17

where the standard pairs $\ell_{k,i} = (\gamma_{k,i}, \rho_{k,i})$ are the (homogeneous) components of $\mathcal{F}_0(\mathcal{F}_e^{n-i-1}\tilde{\gamma})$ with the induced conditional densities, and B is any function. The contribution of $\mathbb{E}_{\ell_{k,i}}$ to (4.7) is bounded by

$$(4.8) \qquad C \int_{\gamma_{k,i}} \rho_{k,i}(Y) \left(\int_{\gamma_{k,i}} \mathbf{1}_{\text{dist}(Y,Z) < 3\hat{C}e} \big| J_{\gamma_{k,i}} \mathcal{F}_0^i(Z) \big| \, dZ \right) dY.$$

Since for any $Y, Z \in \gamma_{k,i}$ we have $\rho_{k,i}(Y) \le C\rho_{k,i}(Z)$, due to (2.3), we can bound (4.8) by

$$Ce \int_{\gamma_{k,i}} \rho_{k,i}(Z) \big| J_{\gamma_{k,i}} \mathcal{F}_0^i(Z) \big| \, dZ.$$

Summing over k we obtain

$$(4.9) \qquad Ce \int_{\tilde{\gamma}} J_{\mathcal{F}_0 \mathcal{F}_e^{n-i-1}(\tilde{\gamma})} \mathcal{F}_0^i(\mathcal{F}_0 \mathcal{F}_e^{n-i-1} X) \, d\mathbb{P}_\ell(X),$$

which is now an upper bound for all the terms coming from case (a). On the other hand, the contribution from (b) is estimated by the probability

$$(4.10) \qquad \mathbb{P}_\ell \big(\mathcal{F}_0 \mathcal{F}_e^{n-i-1} X \text{ is } Ce\text{-close to a singularity of } \mathcal{F}_0^i \big).$$

For $X \in \tilde{\gamma}$, let γ_X denote the component of $\mathcal{F}_0(\mathcal{F}_e^{n-i-1}\tilde{\gamma})$, i.e. one of the above curves $\gamma_{k,i}$, which contains the point $\mathcal{F}_0(\mathcal{F}_e^{n-i-1}(X))$. Let $K_{X,i}$ denote the number of smooth components of $\mathcal{F}_0^i(\gamma_X)$ (here we do not partition them into homogeneous curves); they are curves of uniformly bounded length. Then both (4.9) and (4.10) can be estimated by

$$Ce \int_{\tilde{\gamma}} K_{X,i} \, d\mathbb{P}_\ell(X)$$

(to bound (4.10) observe that the number of the singularities of $\mathcal{F}_0^i$ on γ_X is at most $2K_{X,i}$ and each singularity point only affects its $\mathcal{O}(e)$-neighborhood). Therefore

$$\int_{\tilde{\gamma}} |A \circ F_i - A \circ F_{i+1}| \, d\mathbb{P}_\ell \le Ce \left(1 + \int_{\tilde{\gamma}} K_{X,i} \, d\mathbb{P}_\ell(X) \right),$$

Lastly, $K_{X,i} \le C\Lambda^i$ for some constants $C, \Lambda > 1$, see [7, Section 4.4]. $\qquad \square$

The above argument is designed for smooth functions A, now we need to adapt it to $A = L_0$. This function has discontinuities which coincide with

those of the billiard map $\mathcal{F}_0$, besides, the derivatives of L_0 are unbounded. To handle this case, we make a simple geometric observation (see, e.g. (A1.5) in [3]) that for any unstable curve γ with endpoints X', X'' on which $\mathcal{F}_0$ is smooth

$$|L_0(X') - L_0(X'')| \leq C \left[\text{length}(\gamma) + \text{length}(\mathcal{F}_0(\gamma))\right].$$

Thus our analysis easily applies to the function $A = L_0$.

Now integrating over $\ell \in \mathcal{G}$ gives

$$\int_{\Omega_0} \left| L_0 \circ F_i - L_0 \circ F_{i+1} \right| L_0 \, d\mu_0 \leq C \, e \, \Lambda^i,$$

thus

$$|I_n^{(1)}| \leq \sum_{i=0}^{n} C e \Lambda^i \leq C_1 e \Lambda^n$$

for some constant $C_1 > 0$.

Next we estimate $I_n^{(2)}$. It follows from (3.13) that $|I_n^{(2)}| \leq C_2 e$ for some constant $C_2 > 0$, because $\mathcal{F}_e^n \mu_0(\mathcal{U}_e^2) = \mathcal{O}(e)$ (this last bound follows from Growth Lemma, see [6, Proposition 6.2] and [9, page 587]).

Next we estimate $I_n^{(3)}$. It is easily seen that

$$\Delta_e = \begin{cases} \mathcal{O}(1) & \text{in } \mathcal{U}_e^2 \\ \mathcal{O}(e) & \text{elsewhere} \end{cases}$$

hence $|I_n^{(3)}| \leq C_3 e$ for some $C_3 > 0$.

Lastly, combining all our estimates gives

$$\left| \sum_{n=0}^{\infty}{}' I_n^{(1)} \right| \leq \left| \sum_{n=0}^{c|\ln e|} C_1 e \Lambda_1^n \right| + \left| \sum_{n=c|\ln e|}^{\infty} C \theta^{\sqrt{n}} \right| = \mathcal{O}(e^\delta)$$

for some $\delta > 0$, provided $c > 0$ is small enough, and for $k = 2, 3$

$$\left| \sum_{n=0}^{\infty}{}' I_n^{(k)} \right| \leq \left| \sum_{n=0}^{c|\ln e|} C_k e \right| + \left| \sum_{n=c|\ln e|}^{\infty} C \theta^{\sqrt{n}} \right| = \mathcal{O}(e^\delta).$$

As a result, $\mathcal{R}^* = \mathcal{O}(e^{1+\delta})$. $\qquad\qquad\square$

5 Limiting process for the kinetic energy

In this and subsequent sections we use elements of Ito calculus. We first recall basic definitions and facts following [29, 32].

A (one-dimensional) *Ito diffusion process* is a stochastic process on $\mathbb{R}$ satisfying stochastic differential equation (SDE)

$$(5.1) \qquad dY = a(Y)\, dt + b(Y)\, dW_t, \qquad Y(0) = Y_0,$$

where W_t denotes the standard Wiener process on $\mathbb{R}$ and $a(y)$ and $b(y)$ are smooth functions; $a(Y)$ is called the *drift coefficient* and $b(Y)$ the *diffusion coefficient*. Equation (5.1) has a (unique) solution, which is a time-homogeneous Markov process and has continuous paths. It is a martingale if and only if $a = 0$.

Consider another process $Z = \phi(Y)$, where ϕ is a smooth function. The *Ito formula* asserts that

$$(5.2) \qquad dZ = \left[\phi'(Y)a(Y) + \tfrac{1}{2}\phi''(Y)b^2(Y)\right] dt + \phi'(Y)b(Y)\, dW_t,$$

thus Z is also an Ito diffusion process.

We now return to our system. Let $\bar{K} \geq 0$ and $\ell_\varepsilon = (\gamma_\varepsilon, \rho_\varepsilon)$ a one-parameter family of standard pairs such that $\mathrm{length}(\gamma_\varepsilon) \geq \varepsilon^{100}$ and for some (=for any) point $(X_\varepsilon, \tilde{K}_\varepsilon) \in \ell_\varepsilon$ we have $\tilde{K}_\varepsilon \to \bar{K}$ as $\varepsilon \to 0$. (We remind the reader that $\tilde{K} = \varepsilon K$ is the 'rescaled' kinetic energy, cf. (3.4).)

The following theorem will be proved in Sections 7 and 8.

Theorem 3. *Assume that $(X(0), \tilde{K}(0))$ is chosen according to the probability measure $\mathbb{P}_{\ell_\varepsilon}$, where ℓ_ε is as above. Then there is a constant $\sigma > 0$ such that in the ε-rescaled dynamics the random function $\tilde{K}(\xi\varepsilon^{-2})$, for $0 < \xi < 1$, weakly converges to an Ito diffusion process $\mathcal{K}(\xi)$ on $[0, \infty)$ satisfying SDE*

$$(5.3) \qquad d\mathcal{K} = \frac{\sigma^2}{2\sqrt{2\mathcal{K}}}\, d\xi + (2\mathcal{K})^{1/4}\sigma\, dW_\xi, \quad \mathcal{K}(0) = \bar{K}$$

which, additionally, reflects at the point $\mathcal{K} = 0$.

Observe that the drift coefficient in (5.3) has singularity at 0, so we need to be careful defining its solution at this point. We shall show that the process $\mathcal{Z} = \mathcal{K}^{3/2}$ satisfies equation (5.4), see below, whose coefficients are regular enough to guarantee the existence and uniqueness of its solution. By

20

a solution to (5.3) we shall mean $\mathcal{K}(\xi) = \mathcal{Z}^{2/3}(\xi)$, where $\mathcal{Z}$ is the solution to (5.4).

The foregoing discussion indicates that 0 is a special point for the limiting process (5.3). It also has a special character in the dynamical sense – when the particle's velocity is small (close to zero), its motion may differ significantly from that of a billiard particle. Accordingly, our proof of Theorem 3 consists of two parts. First we show in Section 7 that the result is true if $\bar{K} > 0$ and when $K(t)$ stays away from 0. Then in Section 8 we extend the result to the special case $\bar{K} = 0$.

Combining (5.3) and (5.2) shows that the process $\mathcal{Z} = \mathcal{K}^{3/2}$ satisfies

$$d\mathcal{Z} = \frac{3\sigma^2}{2\sqrt{2}}\,d\xi + \frac{3\sqrt[4]{2}\,\sigma}{2}\,\sqrt{\mathcal{Z}}\,dW_\xi$$

with reflections at 0. Changing the time variable $\xi = \frac{8\sqrt{2}}{9\sigma^2}\eta$ gives

$$(5.4) \qquad\qquad d\mathcal{Z} = \frac{4}{3}\,d\eta + 2\sqrt{\mathcal{Z}}\,dW_\eta.$$

Hence $\mathcal{Z}$ is a Bessel square process of dimension $4/3$ (its index $= -1/3$), see e.g. [32, Section XI.1]. Thus the properties of $\mathcal{K}$ can be derived from those of $\mathcal{Z}$.

First of all, the process solving (5.4) exists and is unique, hence so is the process $\mathcal{K}$ solving (5.3). Second, $\mathcal{Z}$ is recurrent, in fact it hits zero infinitely many times with probability one, hence our process $\mathcal{K}$ has the same properties. Third, if $\mathcal{K}(0) = 0$ then for any $\xi > 0$ the random variable $c\,\xi^{-2/3}\mathcal{K}(\xi)$ has density

$$(5.5) \qquad\qquad f(x) = C\,\exp(-x^{3/2}),$$

where $c = c(\sigma) > 0$.

This last fact readily implies Theorem 2, as we show next.

Proof of Theorem 2. We set $\bar{K} = 0$ and apply Theorem 3 to the proper standard family $\mathcal{G}$ such that $\mu_\mathcal{G} = \mu_{\text{ini}}$ (this is possible due to the remark in the end of Section 2 and because the kinetic energy for $X \in \Omega_{\text{ini}}$ is uniformly bounded). Remembering that $t = \sqrt{\varepsilon}s$ (we set $t_0 = 0$) and $K(t) = \varepsilon^{-1}\hat{K}(s)$, see (3.4), we conclude that

$$\varepsilon K(\xi\varepsilon^{-3/2}) = \tilde{K}(\xi\varepsilon^{-2}) \Rightarrow \mathcal{K}(\xi)$$

21

as $\varepsilon \to 0$. Now changing variable $\varepsilon = t^{-2/3}$, setting $\xi = 1$, and using (5.5) yield Theorem 2. $\qquad\square$

Next we derive other useful properties of the limit diffusion process $\mathcal{K}(t)$. Combining (5.3) and (5.2) shows that the process $\mathcal{L} = \sqrt{\mathcal{K}}$ satisfies

$$d\mathcal{L} = \frac{\sigma}{2^{3/4}\mathcal{L}^{1/2}}\, dW_\xi,$$

so the drift is missing, hence this process is a martingale (during intervals of time when it stays strictly positive).

Given $\xi_0 > 0$ with $\mathcal{K}(\xi_0) > 0$ we denote by $\xi_1 > \xi_0$ the first time the process $\mathcal{K}(\xi)$ goes up to $2\mathcal{K}(\xi_0)$ or down to $\frac{1}{2}\mathcal{K}(\xi_0)$. Denote

$$(5.6) \qquad p = \mathbb{P}\big(\mathcal{K}(\xi_1) = \tfrac{1}{2}\mathcal{K}(\xi_0)\big) \quad \text{and} \quad 1 - p = \mathbb{P}\big(\mathcal{K}(\xi_1) = 2\mathcal{K}(\xi_0)\big).$$

Since $\sqrt{\mathcal{K}}$ is a martingale, we have

$$(5.7) \qquad\qquad \sqrt{\mathcal{K}(\xi_0)} = p\sqrt{\mathcal{K}(\xi_0)/2} + (1 - p)\sqrt{2\mathcal{K}(\xi_0)},$$

whence $p = 2 - \sqrt{2} \approx 0.59$, i.e. the process $\mathcal{K}$ is more likely to decrease by half than to double its value.

Furthermore, let $\xi_0 < \xi_1 < \xi_2 < \cdots$ denote consecutive time moments such that ξ_k is the first time when $\mathcal{K}(\xi_k) = 2^n\mathcal{K}(\xi_{k-1})$ for some $n = \pm 1$. Then the process $Z_k = \log_2 \mathcal{K}(\xi_k) - \log_2 \mathcal{K}(\xi_0)$ is a simple random walk on $\mathbb{Z}$ defined by

$$(5.8) \qquad \mathbb{P}(Z_k = Z_{k-1} - 1) = p \quad \text{and} \quad \mathbb{P}(Z_k = Z_{k-1} + 1) = 1 - p.$$

Since $p > 0.5$, the random walk escapes to $-\infty$ with probability one. This actually confirms the recurrence of $\mathcal{K}(\xi)$.

Next, it is easy to see that the distribution of the rescaled process $c^{-2/3}\mathcal{K}(ct)$ is independent of c. Let τ_R denote the first time when $\mathcal{K}(\tau) = R$ (assuming that $\mathcal{K}(0) = 0$). Then the distribution of $\tau_R/R^{3/2}$ is independent of R.

Fix a $\delta > 0$. We define a δ-excursion of $\mathcal{K}$ to be a maximal interval of time $[a, b]$ such that $\mathcal{K}(a) = \delta$ and $\mathcal{K}(s) > \delta/2$ for $s \in [a, b]$. Let $\boldsymbol{\tau}_\delta = b - a$. There may be many δ-excursions $\boldsymbol{\tau}_{\delta,1}, \boldsymbol{\tau}_{\delta,2} \ldots$ on any trajectory of $\mathcal{K}$, and $\boldsymbol{\tau}_{\delta,1}, \boldsymbol{\tau}_{\delta,2} \ldots$ are i.i.d. random variables (because $\mathcal{K}$ is a time-homogeneous Markov process). We easily see that

$$(5.9) \qquad\qquad \text{the distribution of } \boldsymbol{\tau}_\delta/\delta^{3/2} \text{ is independent of } \delta.$$

Lemma 5.1. *We have the following 'tail estimate':*

(5.10) $$\mathbb{P}(\boldsymbol{\tau}_1 > A) \sim \frac{C}{A^{1/3}} \quad as \ \ A \to \infty.$$

The idea of the proof is the following. Observe that if $\boldsymbol{\tau}_1 > A$, then due to (5.5) we have

$$\max_{[a,b]} \mathcal{K}(t) \sim A^{2/3},$$

where $[a, b]$ is the interval corresponding to $\boldsymbol{\tau}_1$. On the other hand,

$$\mathbb{P}\big(\mathcal{K}(t) \text{ reaches } A^{2/3} \text{ before } 1/2\big) = \frac{C}{A^{1/3}},$$

since $\sqrt{\mathcal{K}}$ is a martingale. This explains (5.10). The precise proof is very similar to the proof of Lemma 8.1 given in Section 10, so we leave the proof of Lemma 5.1 to the reader.

Now let $T_n(\delta) = \sum_{i=1}^{n} \boldsymbol{\tau}_{\delta,i}$ denote the total time of the first n δ-excursions; it is the sum of n independent random variables each having the distribution of $\boldsymbol{\tau}_\delta$.

Corollary 5.2. *The distribution of $\frac{T_n(\delta)}{\delta^{3/2}n^3}$ does not depend on δ. It approaches a limit distribution as $n \to \infty$.*

Proof. The first claim follows from (5.9). The second follows from (5.10) and a standard limit theorem for sum of independent random variables with heavy tails (see, e.g., [20, Section 2.7]). □

Lastly we comment on the recurrent character of our limit process $\mathcal{K}(\xi)$. It means that the kinetic energy $K(t)$, and hence the coordinate $x(t)$, of the Galton particle must evolve in a similar manner – its excursions into the depth of the Galton board must alternate with retreats to a vicinity of the starting line $x = 0$. As time goes on, excursions tend to last longer and longer and extend farther and farther (because the coordinate $x(t)$ must grow as $t^{2/3}$), but every excursion is followed by a retreat of the particle back to the initial region $x \approx 0$. All this will be established in the subsequent sections.

6   Proof of Theorem 1

Since the limit stochastic process $\mathcal{K}(t)$ constructed in the previous section is recurrent, almost every trajectory of $\mathcal{K}(t)$ returns to zero infinitely many

times. In Section 10 we prove the following lemma, which effectively means recurrence for the Galton particle on a closed board.

Lemma 6.1. *There is a large constant $K_{\min} > 0$ such that for every standard pair $\ell = (\gamma, \rho)$ with $\gamma \subset \Omega_E$, $E > E_{\min}$, we have*

$$\mathbb{P}_\ell\Big(\big(X(0), K(0)\big) \in \gamma \colon \inf_{t>0} K(t) > \max\{K_{\min}, E\}\Big) = 0,$$

i.e. almost every trajectory originating from γ eventually slows down to the level $\max\{K_{\min}, E\}$.

We recall that $E_{\min}$ was introduced in Section 2 as the minimal energy for which the dynamics remains hyperbolic. Observe that $K(t) \geq E$ due to (1.3), hence the lemma tells us that if $E > K_{\min}$, then almost every trajectory starting at $\gamma \subset \Omega_E$ eventually hits the line $x = 0$. If E is not very large ($E < K_{\min}$), then almost every trajectory originating from $\gamma \subset \Omega_E$ eventually goes above the line $x = (K_{\min} - E)/g$. We also note that even if E is not very large, the initial kinetic energy $K(0)$ may be arbitrarily large, as it depends on the x coordinate of points on γ.

We now derive Theorem 1 from Lemma 6.1 using an idea of J. Littlewood [25]. Let
$$\Omega_{A,B} = \{(q, u, K) \in \Omega \colon \ x(q) \leq A, \ K \leq B\},$$
where $x(q)$ denotes the x coordinate of the point $q \in \partial \mathcal{D}$. Observe that $\mu(\Omega_{A,B}) < \infty$ for every A, B.

Let $E_{\max}$ denote the maximal total energy of trajectories originating from the set Ω_{ini} defined by (2.2) and pick any $A > (K_{\min} - E_{\min})/g$ and $B > E_{\max}$. Then the trajectories originating from (1.4) fill a subset of positive μ measure in $\Omega_{A,B}$, and almost all of them come back to $\Omega_{A,B}$ (after leaving that set), according to Lemma 6.1. Denote

$$\Omega^* = \Omega_{A,B} \cap \Big(\cup_{n=0}^{\infty} \mathcal{F}^n(\Omega_{\mathrm{ini}})\Big).$$

We see that $\mu(\Omega^*) < \infty$ and the return map

$$\mathcal{F}^*(x) = \mathcal{F}^k(x), \qquad k = \min\{n \geq 1 \colon \mathcal{F}^n(x) \in \Omega^*\}$$

is defined almost everywhere on Ω^*. It then easily follows from the Poincaré recurrence theorem that almost every initial state $(X, K) \in \Omega_{\mathrm{ini}}$ returns to

Ω_{ini}. But such a return obviously means that on the open Galton board the trajectory just escapes through the opening $x = 0$. $\qquad\square$

We remark that Theorem 1 remains true for any $V > 0$ used in the definition (1.4), i.e. V does not have to be large. Indeed, if V is small, then there are two types of trajectories originating from Ω_{ini}. Some never reach high kinetic energy which would ensure hyperbolic behavior; they simply return to the line $x = 0$ with probability one according to the Poincaré recurrence theorem. Others do reach high kinetic energy that ensures hyperbolicity, and to those we can apply the above argument again (observe that hyperbolicity is ensured whenever $K > E_{\text{min}}$).

Of course, in order to apply our argument to trajectories of the second type we need to foliate their union (at the moment when their kinetic energy exceeds E_{min}) by smooth unstable curves (which, together with conditional densities, become standard pairs). Such foliation is always possible because the trajectories of the second type, at every iteration of the map $\mathcal{F}$, occupy countably many domains in Ω with piecewise smooth boundaries, since our map $\mathcal{F}$ is piecewise smooth.

7 Proof of Theorem 3 away from $\tilde{K} = 0$

We now start proving Theorem 3. In this section we work away from the critical point $\tilde{K} = 0$, i.e. assume $\tilde{K} \geq \tilde{K}_{\text{min}}$ for some $\tilde{K}_{\text{min}} > 0$. The next section deals with a delicate issue of approaching zero.

Lemma 7.1 (Averaging). *Let $\bar{K} > 0$. Suppose $\ell_\varepsilon = (\gamma_\varepsilon, \rho_\varepsilon)$ is a family of standard pairs such that* length$(\gamma_\varepsilon) \geq \varepsilon^{100}$ *and chose a point $(X_\varepsilon, \tilde{K}_\varepsilon) \in \ell_\varepsilon$ we have $\tilde{K}_\varepsilon \to \bar{K}$ as $\varepsilon \to 0$. Then after $N = C|\ln \varepsilon|$ iterations we have*

$$\mathbb{E}_{\ell_\varepsilon}(L_{e_N} \circ \mathcal{F}^N) = \frac{\varepsilon g}{4\tilde{K}_\varepsilon} \sum_{n=-\infty}^{\infty} \mu_0\big((L_0 \circ \mathcal{F}_0^n)\, L_0\big) + \mathcal{O}(\varepsilon^{1+\delta}),$$

where $C \gg 1$ is a large constant, $\delta > 0$ is a small constant, and we use notation of Section 3.

Proof. This lemma is a close analogue of [7, Proposition 3.3] and [9, Proposition 4]. First observe that the kinetic energy $\tilde{K}$ changes by $\mathcal{O}(\varepsilon n)$ during n iterations, according to (3.11), hence it stays in a narrow interval $\bar{K} \pm C\varepsilon|\ln \varepsilon|$. Second, due to a standard growth lemma, see [7, Lemma 4.10],

it takes $\sim C_0 |\ln \varepsilon|$ iterations of $\mathcal{F}$ to transform ℓ_ε into a *proper* standard family, where $C_0 = C_0(\mathcal{D})$ is a constant. We choose $C \gg C_0$, hence effectively we can assume that ℓ_ε is a proper standard pair.

Next we put $e = \frac{\varepsilon g}{2K}$ and consider the 'thermostatted' collision map $\mathcal{F}_e \colon \Omega_0 \to \Omega_0$. Recall that the map $X_n \mapsto X_{n+1}$ is a $\mathcal{O}(\varepsilon^2)$-perturbation of the map $X_n \to \mathcal{F}_e(X_n)$, cf. Section 3. Now a standard shadowing-type argument used for hyperbolic systems with singularities yields

$$\mathbb{E}_{\ell_\varepsilon}(L_N \circ \mathcal{F}^N) = \mu_{e_0}(L_{e_0}) + \mathcal{O}(N\varepsilon^2)$$

and $\mathcal{O}(N\varepsilon^2) = \mathcal{O}(\varepsilon^2 |\ln \varepsilon|) \ll \varepsilon^{1+\delta}$. For a detailed exposition of the aforementioned shadowing-type argument we refer the reader to the proof of [7, Proposition 3.3]. Hence our result follows from Proposition 4.1. □

We introduce two convenient notation:

$$\bar{\sigma}^2 = g^2 \sum_{n=-\infty}^{\infty} \mu_0\big((L_0 \circ \mathcal{F}_0^n)\, L_0\big)$$

and

$$\bar{a}(K) = \frac{\bar{\sigma}^2}{4K}, \qquad K > 0.$$

In the next theorem we assume that $\mathcal{I} = (\tilde{K}_{\min}, \tilde{K}_{\max})$ is a finite interval such that $\tilde{K}_{\min} > 0$. Given a point $\bar{K} \in \mathcal{I}$, we suppose, just like in the previous lemma, that $\ell_\varepsilon = (\gamma_\varepsilon, \rho_\varepsilon)$ is a family of standard pairs such that $\mathrm{length}(\gamma_\varepsilon) \geq \varepsilon^{100}$ and for some (=for any) point $(X_\varepsilon, \tilde{K}_\varepsilon) \in \ell_\varepsilon$ we have $\tilde{K}_\varepsilon \to \bar{K}$ as $\varepsilon \to 0$.

Now we choose the initial state of the particle according to the probability measure $\mathbb{P}_{\ell_\varepsilon}$ and let the particle move under the iterations of the map $\mathcal{F}$ as long as its ε-rescaled kinetic energy $\tilde{K} = \varepsilon K$ stays within the interval $\mathcal{I}$. Once it exists $\mathcal{I}$, we stop the particle and 'freeze' its kinetic energy (forever). We denote by $\tilde{K}_n^*$ the so modified ε-rescaled kinetic energy of the particle at the nth collision and by $n_\mathcal{I} = \min\{n \colon \tilde{K}_n \notin \mathcal{I}\}$ be the first (discrete) time when $\tilde{K}_n$ exits $\mathcal{I}$ (i.e. when $\tilde{K}_n$ and $\tilde{K}_n^*$ start differ).

Theorem 4. *Under the above conditions*
(a) $\tilde{K}_{\bar{\xi}\varepsilon^{-2}}^*$ *weakly converges to a stochastic process* $\mathcal{K}^*(\bar{\xi})$ *satisfying SDE*

(7.1) $$d\mathcal{K}^* = \bar{a}(\mathcal{K}^*)\, d\bar{\xi} + \bar{\sigma}\, dW_{\bar{\xi}}, \quad \mathcal{K}(0) = \bar{K},$$

and stopped when it exits $\mathcal{I}$.

(b) *The process $\tilde{K}_n^*$ satisfies*

$$\mathbb{E}_{\ell_\varepsilon}\left([\tilde{K}_{n_\mathcal{I}}^*]^{1/2}\right) = \bar{K}^{1/2} + \mathcal{O}(\varepsilon^\delta).$$

We note that part (b) can be rewritten as

$$\mathbb{E}_{\ell_\varepsilon}\left(\Phi(\tilde{K}_n^*) - \Phi(\tilde{K}_0^*)\right) = \mathcal{O}(\varepsilon^{2+\delta}n)$$

for any function Φ that solves the differential equation $\frac{\bar{\sigma}^2}{2}\Phi'' + \bar{a}\Phi' = 0$; the general solution of the latter is $\Phi(x) = c_1\sqrt{x} + c_2$.

Observe that (7.1) and Ito's formula (5.1) imply that the process $Z = \sqrt{\mathcal{K}^*}$ satisfies the SDE $dZ = \frac{\bar{\sigma}}{2Z}\,dW_{\bar{\xi}}$, i.e. it has no drift, hence it is a martingale (during intervals of time when $\mathcal{K}^*$ stays within $\mathcal{I}$).

Theorem 4 can be proven by standard techniques based on moment estimation [7, 15, 19]. For completeness we outline its proof in Appendix.

Lastly we derive Theorem 3 (away from $\mathcal{K} = 0$) from Theorem 4. The latter describes the limit process $\mathcal{K}$ in the time variable $\bar{\xi} = \varepsilon^2 n$, which is proportional to discrete time n (the collision counter). All we need is to convert its result to continuous time $\xi = \varepsilon^2 s$ used in Theorem 3. This is done by a standard time change procedure which we sketch below.

As our particle moves at speed $\tilde{v} = \sqrt{2\mathcal{K}}$, it experiences

$$(7.2) \qquad\qquad n = s\tilde{v}/\bar{\tau} + o(s)$$

collisions during a time interval of length s, where

$$\bar{\tau} = \pi\,\mathrm{Area}(\mathcal{D}_0)/\mathrm{length}(\partial\mathcal{D}_0)$$

is the mean free path for the billiard dynamics ($\mathcal{D}_0$ denotes the fundamental domain of $\mathcal{D}$). The relation (7.2) is an analogue of the law of large numbers; and it is important that the remainder term $o(s)$ is uniform (in probability). It can be proved by fairly standard arguments, such as moment estimates [7, Section 6.8] or Kolmogorov-type inequalities [11, Theorem 7.68]. Thus the time variable is changed by the rule

$$d\bar{\xi} = \frac{\sqrt{2\mathcal{K}}}{\bar{\tau}}\,d\xi.$$

The discrete-time diffusion coefficient $\bar{\sigma}^2$ also needs to be replaced with that of the continuous time system according to a standard formula: $\sigma^2 = \bar{\sigma}^2/\bar{\tau}$, see [11, Theorem 7.68]. Now (7.1) reads

$$dK = \frac{\sigma^2}{2\sqrt{2K}}\, d\xi + (2K)^{1/4}\sigma\, dW_\xi,$$

as Theorem 3 claims.

8   Proof of Theorem 3 for small energies

Theorem 4 almost proves Theorem 3 except that it requires us to stop the process $\mathcal{K}^*(\bar{\xi})$ when its value becomes too small or too large. The later restriction is harmless since $\mathcal{K} = \infty$ is an 'inaccessible point' for our process; precisely this means

$$\mathbb{P}\Big(\max_{[0,1]} \mathcal{K}(\bar{\xi}) > R\Big) \to 0 \quad \text{as} \quad R \to \infty,$$

hence it can be easily ensured that the stopped process $\mathcal{K}^*$ differs from the actual one on a set of arbitrarily small probability.

The other restriction, which requires $\mathcal{K}$ to stay away from zero, is more serious. First of all, since our initial kinetic energy is $K = \mathcal{O}(1)$, the limit process starts at zero: $\mathcal{K}(0) = 0$. Second, even if the Galton particle starts with a high enough initial kinetic energy $K(0)$ to ensure that $\mathcal{K}(0) > 0$, we would still have

$$\liminf_{R \to 0} \mathbb{P}\Big(\min_{[0,1]} \mathcal{K}(\bar{\xi}) < R\Big) > 0.$$

In fact both limiting processes defined by (5.3) and (7.1) visit zero infinitely many times. Thus the vicinity of zero cannot be completely neglected.

To describe our idea we recall that Theorem 3 tells us that the actual kinetic energy of the Galton particle during the time interval $0 < t < T$ should be $K(t) = \mathcal{O}(T^{2/3})$. For every trajectory $K(t)$, $0 < t < T$, and a small $\delta > 0$ we call a δ-$excursion$ a maximal time interval (t_1, t_2) such that $K(t_1) = \delta T^{2/3}$ and $K(t) > \frac{1}{2}\delta T^{2/3}$ on $[t_1, t_2)$.

Observe that δ-excursions of $K(t)$, after ε-rescaling the dynamics and time (as done earlier), will correspond to the δ-excursions of the limit process $\mathcal{K}(\xi)$ defined in Section 5. Moreover, since we already have proved Theorem 3 away from $\mathcal{K} = 0$, it follows that for every $\delta > 0$ the δ-excursions

of the process $K(t)$ weakly converge to the δ-excursions of the process $\mathcal{K}(\xi)$. It precisely means that if we cut and remove intervals complementary to all the δ-excursions from both processes, we would have the desired weak convergence.

It remains to show that the complements of the δ-excursions become negligible as $\delta \to 0$. Precisely, let $\mathcal{J}_\delta \subset [0,1]$ be the union of all intervals complementary to the δ-excursions of the limit process $\mathcal{K}(\xi)$ and $|\mathcal{J}_\delta|$ denote its Lebesgue measure. Then for every $\varepsilon > 0$

$$\mathbb{P}\big(|\mathcal{J}_\delta| > \varepsilon\big) \to 0 \quad \text{as} \quad \delta \to 0$$

which follows from known properties of the Bessel square process $\mathcal{Z} = \mathcal{K}^{3/2}$, cf. Section 5.

Similarly, let $J_\delta \subset [0,T]$ be the union of all intervals complementary to the δ-excursions of a trajectory $K(t)$ and $|J_\delta|$ denote its Lebesgue measure. We need to show that for every $\varepsilon > 0$

$$(8.1) \qquad \mathbb{P}\big(|J_\delta| > \varepsilon T\big) \to 0 \quad \text{as} \quad \delta \to 0.$$

To this end, we will show that (i) there are not too many intervals complementary to the δ-excursions, and (ii) those intervals are not too large.

Let M_δ denote the number of the above intervals on a trajectory $K(t)$, $0 < t < T$; observe that the number of δ-excursions on that same trajectory must be $M_\delta \pm 1$. Since the δ-excursions of trajectories $K(t)$ weakly converge (with proper rescaling, see above) to those of the limit process $\mathcal{K}(\xi)$, an analogue of Corollary 5.2 holds for our process $K(t)$ as well; that is, the total length of M first δ-excursions divided by $(\delta T^{2/3})^{3/2} M^3 = \delta^{3/2} T M^3$ converges in distribution, as $M \to \infty$, to a finite random variable. Since the total length of all M_δ of our δ-excursions cannot exceed T, we conclude that $M_\delta = \mathcal{O}(\delta^{-1/2})$. Furthermore, for any $\varepsilon > 0$

$$(8.2) \qquad \mathbb{P}(M_\delta \geq C\delta^{-1/2-\varepsilon}) \to 0 \quad \text{as} \quad \delta \to 0.$$

Next we estimate the length of individual intervals complementary to the δ-excursions. Let $K > 0$ and assume that the Galton particle starts on a proper standard pair ℓ with a kinetic energy $\mathbb{E}_\ell(K(0)) < K$. Denote by τ_K the first time the particle's kinetic energy goes up to K.

Lemma 8.1. *There are constants $c, p > 0$ such that $\mathbb{P}_\ell\big(\tau_K \leq cK^{3/2}\big) \geq p$.*

29

Note that this bound holds uniformly for any $K(0) < K$, in particular for $K(0) = 0$. The proof of the lemma is quite long, it will be given in Section 10.

Now let Δ be an arbitrary interval between two neighboring δ-excursions on a trajectory $K(t)$; the previous lemma implies that

$$(8.3) \qquad \mathbb{P}\big(|\Delta| \geq c\delta^{3/2-\varepsilon}T\big) \leq (1-p)^{\delta^{-\varepsilon}}.$$

Combining (8.2) and (8.3) shows that for typical trajectories the total length of all the M_δ intervals complementary to the δ-excursions does not exceed $cC\delta^{1-2\varepsilon}T$. This proves (8.1) and completes the proof of Theorem 3.

9   Retreats of the Galton particle

In this section we prove Lemma 6.1. We need several elementary facts about simple biased random walks which can be found e.g. in [1].

Proposition 9.1. *Let* $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_n, \ldots$ *be i.i.d. random variables such that* $\tilde{\xi}_n \in \{-1, 1\}$ *and* $\mathbb{P}(\tilde{\xi}_n = -1) = p > 1/2$. *Denote* $\tilde{\mathcal{X}}_n = \tilde{\xi}_1 + \tilde{\xi}_2 + \cdots + \tilde{\xi}_n$. *Then*
(a) $\mathbb{P}\big(\tilde{\mathcal{X}}_n$ *hits* $k \geq 1$ *before* $-m \leq -1\big)$ *equals*

$$\frac{(q/p)^k - (q/p)^{m+k}}{1 - (q/p)^{m+k}}, \qquad \text{where } q = 1 - p;$$

(b) $\mathbb{P}\big(\tilde{\mathcal{X}}_n = k$ *for some* $n\big) = [(1-p)/p]^k$ *for every* $k \geq 1$;
(c) $\mathbb{P}\big(\tilde{\mathcal{X}}_n \leq 0$ *for all* $n\big) = 2 - 1/p > 0$;
(d) *For each* $c > \mathbb{E}(\tilde{\xi}_i) = 1 - 2p$ *there are constants* $C > 0$ *and* $\theta < 1$ *such that* $\mathbb{P}(\tilde{\mathcal{X}}_n > cn) \leq C\theta^n$.

Part (d) is just the standard law of large deviations. We shall also use the following comparison criterion (see e.g., [16]).

Proposition 9.2. *Let* $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_n, \ldots$ *be the same random walk as above and* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ *another sequence of random variables such that* $\xi_n \in \{-1, 1\}$ *and for every* $n \geq 1$ *we have the following bound on conditional probabilities:*

$$\mathbb{P}(\xi_n = -1 \mid \xi_1, \ldots, \xi_{n-1}) \geq p.$$

Denote $\mathcal{X}_n = \xi_1 + \xi_2 + \cdots + \xi_n$ *and* $\tilde{\mathcal{X}}_n = \tilde{\xi}_1 + \tilde{\xi}_2 + \cdots + \tilde{\xi}_n$. *Then for any* $n \geq 1$ *and* $m_1, m_2 \in \mathbb{Z}$ *we have*

$$\mathbb{P}\Big(\max_{k \leq n} \tilde{\mathcal{X}}_k \leq m_1 \ \& \ \min_{k \leq n} \tilde{\mathcal{X}}_k \leq m_2\Big) \leq \mathbb{P}\Big(\max_{k \leq n} \mathcal{X}_k \leq m_1 \ \& \ \min_{k \leq n} \mathcal{X}_k \leq m_2\Big).$$

We begin our proof of Lemma 6.1. We give a proof in case $E < K_{min}$, the case where $E \geq K_{min}$ is very similar. For an easy control over the kinetic energy we introduce some convenient notation. For any point $X = (q, u, K) \in \Omega$ we denote by $\hat{K}(X)$ the minimal kinetic energy of all points $X' = (q', u', K') \in \Omega$ such that q and q' belong to the same scatterer and X and X' have the same total energy. (Clearly, the minimum is achieved when q' has the smallest x coordinate on the scatterer.) For any standard pair $\ell = (\gamma, \rho)$ let $\hat{K}_\ell$ denote the common value of $\hat{K}(X)$ for $X \in \gamma$. For a given trajectory $\{X_n\}$ we will put $\hat{K}_n = \hat{K}(X_n)$.

Lemma 9.3. *If $K_{\min}$ is sufficiently large then there exists a small constant $p_0 > 0$ such that for every standard pair $\ell = (\gamma, \rho)$ there is $n_\ell \geq 0$ such that*

$$(9.1) \qquad \mathbb{P}_\ell\Big(\min_{n \leq n_\ell} \hat{K}_n \leq K_{\min}\Big) \geq p_0.$$

First we derive Lemma 6.1 from (9.1). Recall that the image $\mathcal{F}^n(\gamma)$ is a finite or countable union of unstable curves (which, along with the induced measures, make standard pairs, cf. Section 2) and $\hat{K}_n$ is constant on each of them. Thus we can apply the above lemma to the components of $\mathcal{F}^{n_\ell}(\gamma)$ on which $\hat{K}_{n_\ell}$ is still above $K_{\min}$, and then continue this procedure repeatedly; the total measure of the remaining components (where the kinetic energy exceeds $K_{\min}$), after k repetitions, will be $\leq (1-p_0)^k$. This proves Lemma 6.1. $\square$

Next we prove Lemma 9.3. Observe that small standard pairs expand under $\mathcal{F}^n$, $n \geq 1$, and their images consist, mostly, of proper standard pairs due to part (c) of the Growth Lemma. Thus we can (and will) assume that ℓ is a proper standard pair.

We put $R_n = 2^n K_\ell$ for $n \in \mathbb{Z}$. Let $\sigma_1 \colon \gamma \to \mathbb{N}$ denote the first time when either $\hat{K}_\sigma \geq R_1$ or $\hat{K}_\sigma \leq R_{-1}$. Thus $\hat{K}_{\sigma_1}$ is close to R_{ζ_1} where $\zeta_1 \in \{-1, +1\}$. Let $\sigma_2 \colon \gamma \to \mathbb{N}$, $\sigma_2 > \sigma_1$, denote the first time, after σ_1, when either $\hat{K}_\sigma \geq R_{\zeta_1+1}$ or $\hat{K}_\sigma \leq R_{\zeta_1-1}$. Then $\hat{K}_{\sigma_2}$ is close to R_{ζ_2} where $\zeta_2 \in \{-2, 0, 2\}$. Then we continue defining $\sigma_k \geq \sigma_{k-1}$ and $\zeta_k = \zeta_{k-1} \pm 1$ for $k > 2$ by the same procedure. We also set $\sigma_0 = \zeta_0 = 0$.

Recall that the limiting process $\mathcal{K}$ satisfies (5.6) with $p = 2 - \sqrt{2} \approx 0.59$. By Theorem 3 (note that we use it "away from zero", in which case it is already proved), if $\hat{K}_\ell \geq K_{\min}$ and $K_{\min} \gg 1$ is large enough, then

$$\mathbb{P}_\ell(\hat{K}_{\sigma_1} \leq R_{-1}) \geq 0.55 \quad \text{and} \quad \mathbb{P}_\ell(\hat{K}_{\sigma_1} \geq R_1) \leq 0.45$$

Thus on at least 55% of the images of ℓ at time σ_1 the kinetic energy will decrease by half, and on at most 45% it will double. To extent this result to σ_2 (and further to σ_n, $n \geq 3$), we need to apply Theorem 3 to the components $\gamma_{\sigma_n,i} \subset \mathcal{F}^{\sigma_n}(\gamma)$ for $n \geq 2$ (with induced measures on them). Recall, however, that Theorem 3 only applies to standard pairs which are not too small; it is easy to check that in our case it applies to components satisfying $\text{length}(\gamma_{\sigma_n,i}) \geq \hat{K}_{\ell_{n,i}}^{-100}$. Accordingly, we define a decreasing sequence of subsets $\gamma \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \cdots$ by

$$\mathcal{L}_n = \big\{ X \in \gamma \colon \text{length}\big(\gamma_{\sigma_k}(X)\big) \geq [\hat{K}_{\sigma_k}(X)]^{-100} \quad \text{for } k = 1 \ldots n \big\},$$

where $\gamma_j(X)$ denotes the component of $\mathcal{F}^j(\gamma)$ containing the point $\mathcal{F}^j(X)$. Also consider a decreasing sequence of subsets $\gamma \supset \mathcal{U}_1 \supset \mathcal{U}_2 \supset \ldots$ defined by

$$\mathcal{U}_n = \big\{ X \in \gamma \colon \min_{1 \leq k \leq n} \hat{K}_{\sigma_k}(X) \geq K_{\min} \big\}.$$

Now Theorem 3 applies and gives

$$\mathbb{P}_\ell \left(\{ \hat{K}_{\sigma_{n+1}} \leq R_{\zeta_n - 1} \} \bigcap \mathcal{L}_{n+1} \big| (\mathcal{L}_n \bigcap \mathcal{U}_n) \right) \geq 0.55.$$

Therefore the random variables ζ_n, $n \geq 2$, behave like the first variable ζ_1, but only after we restrict ζ_n to the set $\mathcal{L}_n \cap \mathcal{U}_n$. To discard that restriction, we define a new sequence of random variables $\{\hat{\zeta}_n\}$ on γ as follows: we put $\hat{\zeta}_n = \zeta_n$ on $\mathcal{L}_n \cap \mathcal{U}_n$; and on the complement $\gamma \setminus (\mathcal{L}_n \cap \mathcal{U}_n)$ we define $\hat{\zeta}_n$ arbitrarily in such a way that the increment $\hat{\zeta}_n - \hat{\zeta}_{n-1}$ is independent of the 'past history' $\{\hat{\zeta}_1 \ldots \hat{\zeta}_{n-1}\}$ and must take two values: -1 with conditional probability 0.55 and $+1$ with conditional probability 0.45. (Since $(\gamma, \mathbb{P}_\ell)$ is a Lebesgue space, such a construction is always possible.)

Thus, the sequence $\{\hat{\zeta}_n\}$ is in fact a combination of the real dynamics on the images of γ and some abstract random variables that 'take over' whenever the images of γ either become too short or hit the destination (i.e. their kinetic energy falls below $K_{\min}$). Observe that the process $\{\hat{\zeta}_n\}$ satisfies $\hat{\zeta}_{n+1} = \hat{\zeta}_n \pm 1$ and

$$\mathbb{P}(\hat{\zeta}_{n+1} = \hat{\zeta}_n - 1 | \hat{\zeta}_1, \ldots, \hat{\zeta}_n) \geq 0.55,$$

thus we can use Proposition 9.2 to compare $\{\hat{\zeta}_n\}$ to a simple random walk $\{\mathcal{X}_n\}$ with transition probabilities

(9.2) $\mathbb{P}(\mathcal{X}_{n+1} = \mathcal{X}_n - 1) = 0.55 \quad \text{and} \quad \mathbb{P}(\mathcal{X}_{n+1} = \mathcal{X}_n + 1) = 0.45.$

Note also that we need to make

$$L = [\log_2(\hat{K}_\ell/K_E)] + 1$$

steps down from $\hat{K}_\ell$ to reach the destination $K_{\min}$. Now Proposition 9.1 (d) ensures that there is a large constant $C > 0$ such that

$$\mathbb{P}_\ell(\hat{\zeta}_N \geq -L) < 0.01, \qquad N = [CL],$$

thus the random walk $\{\hat{\zeta}_n\}$ reaches the destination (the point $-L$) in $N = [CL]$ steps with probability > 0.99.

Now observe that if $\hat{\zeta}_n(X) \leq -L$, then either the particle's energy falls below $K_{\min}$ at some moment σ_j, $j \leq n$, or

(9.3) $\text{length}\big(\gamma_{\sigma_j}(X)\big) \leq [\hat{K}_{\sigma_j}(X)]^{-100}$ for some $1 \leq j \leq n$.

We claim that the probability of the second alternative can be made arbitrarily small by choosing a sufficiently large $K_{\min}$. First we shall show that there is a constant $\bar{\theta} < 1$ such that for any $k \geq 1$ and $m \geq 1$

(9.4) $\mathbb{P}_\ell\big(\text{Card}\{n\colon \hat{\zeta}_n = -L + m\} > k\big) \leq \bar{\theta}^k.$

and

(9.5) $\mathbb{P}_\ell\big(\text{Card}\{n\colon \hat{\zeta}_n = m\} > k\big) \leq \bar{\theta}^{m+k}.$

Indeed every time we have $\hat{\zeta}_n = -L + m$, the comparison with the 'model' random walk (9.2) and Proposition 9.1 (c) tell us that $\hat{\zeta}$ will never visit the point $-L + m$ again with a positive probability $p \geq 2/11$. This immediately proves (9.4) for $k = 1$ and $\bar{\theta} = 9/11$, and then we use induction on k. To prove (9.5), we note that by Proposition 9.1 (b) our sequence $\{\zeta_n\}$ reaches the point m with probability $\leq \bar{\theta}^m$, and then we repeat the proof of (9.4).

Given a constant $D > 0$, we shall say that a point $X \in \gamma$ and the corresponding sequence $\{\hat{\zeta}_n(X)\}$ are D-good if for every $m \geq 0$ we have

(9.6)
$$\begin{aligned}
&\text{Card}\{n\colon \zeta_n = -L + m\} \leq D(m + 1), \\
&\text{Card}\{n\colon \zeta_n = -m\} \leq D(m + 1), \\
&\text{Card}\{n\colon \zeta_n = m\} \leq D/2^m;
\end{aligned}$$

in other words, the trajectory of a D-good point does not visit any particular point $m > -L$ too many times (especially those near $-L$ and near 0). Denote by $\gamma_D \subset \gamma$ the set of all D-good points on γ. Then (9.4)–(9.5) imply

$$(9.7) \qquad \mathbb{P}_\ell(\gamma \setminus \gamma_D) \leq 2\sum_{m=1}^{\infty} \bar{\theta}^{Dm} + \sum_{m=1}^{\infty} \bar{\theta}^{m+D/2^m} =: \phi(D).$$

Observe that $\phi(D) \to 0$ as $D \to \infty$; hence the measure of the set of points that are not D-good can be made arbitrarily small by choosing D large enough. Next denote by $\gamma_{D,0}$ the set of D-good points $X \in \gamma_D$ where (9.3) holds, i.e. whose images accidentally fall into inadmissibly short standard pairs. Using part (b) of the Growth Lemma 2.1 gives

$$(9.8) \qquad \mathbb{P}_\ell(\gamma_{D,0}) \leq \sum_{m=1}^{\infty} \frac{\beta_4 Dm}{(2^m K_{\min})^{99}},$$

which can be made arbitrarily small by choosing a sufficiently large $K_{\min}$ (assuming that D has been chosen already). (Observe that we only used the first bound in (9.6), the other two will be used later.) Combining our estimates we see that the measure

$$\mathbb{P}_\ell\Big(X \in \gamma\colon \min_{n \leq \sigma_N} \hat{K}_n(X) \leq K_{\min}\Big)$$

can be made > 0.98 by first choosing large enough D to make (9.7) small, and then choosing large enough $K_{\min}$ to make (9.8) small. This completes the proof of Lemma 9.3. $\qquad \square$

Note that we have actually set $p_0 = 0.98$, and in fact p_0 could be made as close to one as we pleased.

For our future purposes we need to improve Lemma 9.3, and to this end we work with continuous time. Given a proper standard pair $\ell = (\gamma, \rho)$ we consider the family of trajectories $(X(t), K(t))$ such that $(X(0), K(0)) \in \gamma$ equipped with probability measure $\mathbb{P}_\ell$. We denote by K_ℓ the minimal initial kinetic energy $K(0)$ on γ. For any $R > 0$ we denote by $\tau_R = \min\{t > 0\colon K(t) = R\}$ the first time the kinetic energy hits R (we may have $R < K_\ell$ or $R > K_\ell$).

Now Lemma 9.3 can be restated as follows: for every standard pair $\ell = (\gamma, \rho)$ with $K_\ell > K_{\min}$ there is $t_\ell \geq 0$ such that $\mathbb{P}_\ell(\tau_{K_{\min}} \leq t_\ell) \geq p_0$. We need an explicit estimate for t_ℓ, at least for proper standard pairs.

34

Lemma 9.4. *There exist a small constant $p_1 > 0$ and a large constant $Q > 0$ such that for every proper standard pair $\ell = (\gamma, \rho)$ with initial kinetic energy $K_\ell > K_{\min}$ we have*

$$(9.9) \qquad \mathbb{P}_\ell\big(\tau_{K_{\min}} \leq QK_\ell^{3/2}\big) \geq p_1.$$

Proof. First we recall certain properties of the limiting process $\mathcal{K}(\xi)$, see Section 5. It is is scale-invariant; in particular, the distribution of $\tau_R/R^{3/2}$ is independent of R; here τ_R denotes the first time when $\mathcal{K}(\tau) = R$ (assuming $\mathcal{K}(0) = 0$). We restate this property a little differently. Let $R > 0$ and assume that $\mathcal{K}(0) = \bar{K} \in (R/2, 2R)$; we consider $\min\{\tau_{R/2}, \tau_{2R}\}$, which is the first time when the process hits either $R/2$ or $2R$. Now there is a constant $\bar{c} > 0$ such that for any $R > 0$

$$(9.10) \qquad \mathbb{P}\big(\min\{\tau_{R/2}, \tau_{2R}\} \leq \bar{c}R^{3/2}\big) \geq 1/2,$$

i.e. it takes $\bar{c}R^{3/2}$ units of time for 50% of the process' trajectories to escape from the interval $(R/2, 2R)$.

We now return to the Galton particle. By Theorem 3 (which we again use "away from zero", where it is already proved), a property similar to (9.10) holds for the kinetic energy $K(t)$ along trajectories starting from a standard pair. Precisely, there is a large constant $R_{\min} \gg 1$ such that for any $R \geq R_{\min}$ and any standard pair $\ell = (\gamma, \rho)$ such that $\mathrm{length}(\gamma) \geq R^{-100}$ and $R/2 < K(0) < 2R$ we have

$$(9.11) \qquad \mathbb{P}_\ell\big(\min\{\tau_{R/2}, \tau_{2R}\} \leq cR^{3/2}\big) \geq 1/2$$

(here $c > 0$ is a constant that should be somewhat larger than $\bar{c}$ in (9.10)).

Thus for at least 50% of points $X \in \gamma$ their kinetic energy will have left the interval $(R/2, 2R)$ by the time $cR^{3/2}$. The remaining points make standard pairs at their first collision that occurs right after $t = cR^{3/2}$. We can apply (9.11) again to those standard pairs whose length is $\geq R^{-100}$; and the union of shorter pairs will have relative measure $< \beta_4 R^{-100}$ due to part (c) of the Growth Lemma. It follows by induction on k that

$$(9.12) \qquad \mathbb{P}_\ell\big(\min\{\tau_{R/2}, \tau_{2R}\} > kcR^{3/2}\big) \leq \frac{k}{R^{99}} + \frac{1}{2^k}.$$

Next we need to recall the key moments σ_n's of our constructions in the proof of Lemma 9.3 and express them in continuous time. Recall that $\sigma_{n+1}(X)$ is

35

the first (discrete) moment of time, after $\sigma_n(X)$, when the point $\mathcal{F}^\sigma(X)$ belongs to a component $\gamma_{\sigma,i} \subset \mathcal{F}^\sigma(\gamma)$ on which $\hat{K}_\sigma$ moves above R_{ζ_n+1} or falls below R_{ζ_n-1}. Now we denote by $t_n(X)$ the corresponding continuous time when $\Phi^t(X) \in \gamma_{\sigma,i}$. Observe that $K(t_n)$ will be close to R_{ζ_n}. We also set $t_0 = 0$.

Now (9.12) implies that for any $n \geq 0$ and $k \geq 1$

$$(9.13) \qquad \mathbb{P}_\ell\Big(t_{n+1} - t_n > kcR_{\zeta_n} \,\big|\, (\mathcal{L}_n \bigcap \mathcal{U}_n), \zeta_n\Big) \leq \frac{k}{R_{\zeta_n}^{99}} + \frac{1}{2^k}.$$

Recall that in the proof of Lemma 9.3 we established that at least 98% of trajectories originating on γ have three key properties: (i) their kinetic energy falls below $K_{\min}$ within $N = [CL]$ steps, (ii) they avoid inadmissibly short standard pairs at every 'critical moment' σ_n, and (iii) they are D-good. Now we say that a point $X \in \gamma$ is D-nice if it has all the properties (i)–(iii) and, in addition, for each n

$$(9.14) \qquad \begin{aligned} \zeta_n = -L + m &\implies t_{n+1} - t_n \leq (m+1)DcR_{-L+m}^{3/2}, \\ \zeta_n = -m &\implies t_{n+1} - t_n \leq (m+1)DcR_{-m}^{3/2}, \\ \zeta_n = +m &\implies t_{n+1} - t_n \leq 2^{-m}DcR_{+m}^{3/2}; \end{aligned}$$

where $m \geq 0$. According to (9.13), the measure of the set of points that are D-good but not D-nice is bounded by

$$\sum_{m=0}^{\infty} \frac{D^2(m+1)^2}{2^{100m}K_{\min}^{100}} + 2\sum_{m=0}^{\infty} \frac{D(m+1)}{2^{D(m+1)}} + \sum_{m=0}^{\infty} \frac{\bar{\theta}^m D/2^m}{2^{D/2^m}};$$

in the last sum we used the fact that the sequence $\{\zeta_n\}$ reaches the point $m \geq 1$ with probability $\leq \bar{\theta}^m$, see the proof of Lemma 9.3. As in that proof, we can make the above sum less than 0.01 by first choosing large enough D to suppress the last two sums and then choosing large enough $K_{\min}$ to suppress the first one. Thus we can guarantee that 97% of points $X \in \gamma$ are D-nice.

On the other hand, for every D-nice point $X \in \gamma$ we can estimate the time $\tau_{K_{\min}}$ it takes for the kinetic energy $K(t)$ to fall below $K_{\min}$ by

$$\tau_{K_{\min}} \leq \sum_{m=0}^{\infty} D^2(m+1)^2 c[2^{-m}K_\ell]^{3/2} + \sum_{m=0}^{\infty} 4^{-m}D^2 c[2^m K_\ell]^{3/2} = QK_\ell^{3/2},$$

where $Q = Q(D) < \infty$. Lemma 9.4 is proved $\qquad\qquad\qquad\qquad$ □

Note that we have actually set $p_1 = 0.97$; and just like p_0 in the previous lemma, p_1 could be made as close to one as we pleased (at the expense of increasing Q).

10 Excursions of the Galton particle

In this section we prove Lemma 8.1. Our proof proceeds in several steps.

Step 1. We slightly modify (5.6) and the subsequent calculation of $p = 2 - \sqrt{2}$ by replacing the 'factor of 2' by a 'factor of 4'. More precisely, given $\xi_0 > 0$ with $\mathcal{K}(\xi_0) > 0$ we denote by $\xi_1 > \xi_0$ the first time the process $\mathcal{K}(\xi)$ goes up to $4\mathcal{K}(\xi_0)$ or down to $\frac{1}{4}\mathcal{K}(\xi_0)$. We denote $p^* = \mathbb{P}\big(\mathcal{K}(\xi_1) = \frac{1}{4}\mathcal{K}(\xi_0)\big)$ and, because $\sqrt{\mathcal{K}}$ is a martingale, we set up an equation similar to (5.7) and obtain $p^* = 2/3$.

Next we derive a similar fact for the kinetic energy $K(t)$ of trajectories originating from a standard pair $\ell = (\gamma, \rho)$. We use the notation of the previous section (in particular K_ℓ and τ_R). Note that $\tau_{K_\ell/4} \leq \tau_{4K_\ell}$ means that the kinetic energy $K(t)$ goes down to $\frac{1}{4}K_\ell$ earlier than it goes up to $4K_\ell$, i.e. $\mathbb{P}_\ell\big(\tau_{K_\ell/4} \leq \tau_{4K_\ell}\big)$ is the 'dynamical' counterpart of the above probability p^*. Of course, the process $\sqrt{K(t)}$ is not a martingale, but the part (b) of Theorem 4 allows us to treat the discrete time sequence $\{\sqrt{K_n}\}$ as a martingale approximately, with the error term being $\mathcal{O}(\varepsilon^{2+\delta}n)$; since $n = \mathcal{O}(\varepsilon^{-2})$, the error term is simply $\mathcal{O}(\varepsilon^\delta)$. Recall that Theorem 4 requires the standard pair satisfy $\text{length}(\gamma) \geq \varepsilon^{100}$, so we can assume that $\text{length}(\gamma) > K_\ell^{-100}$, identify ε with K_ℓ^{-1}, and then apply Theorem 4 (b); now the same calculation as above gives that

$$(10.1) \qquad \mathbb{P}_\ell\big(\tau_{\hat{K}_\ell/4} \leq \tau_{4\hat{K}_\ell}\big) = \frac{2}{3} + \mathcal{O}\big(\hat{K}_\ell^{-\delta_1}\big)$$

with some constant $\delta_1 > 0$.

Step 2. We construct a sequence of times t_n and a sequence of integers ζ_n, just like in the previous section, but replacing the 'factor of 2' by a 'factor of 4'; that is, we start with $R_n = 4^n K_\ell$, etc. In the previous step we established that the sequence $\{\zeta_n\}$ approximately behaves as a simple random walk $\{\xi_n\}$ on $\mathbb{Z}$ with transition probabilities

$$(10.2) \qquad \mathbb{P}(\xi_{n+1} = \xi_n - 1) = 2/3 \quad \text{and} \quad \mathbb{P}(\xi_{n+1} = \xi_n + 1) = 1/3.$$

By Proposition 9.1 (a) we have

(10.3) $\mathbb{P}(\xi_n \text{ hits } 1 \text{ earlier than } - m) = \dfrac{2^m - 1}{2^{m+1} - 1} = \dfrac{1}{2} + \mathcal{O}\Big(\dfrac{1}{2^m}\Big).$

Comparing our 'dynamical' sequence $\{\zeta_n\}$ with the random walk $\{\xi_n\}$ (we use Proposition 9.2 as in the previous section) gives

(10.4) $\mathbb{P}_\ell\big(\tau_{K_\ell^{1/2}} \leq \tau_{4K_\ell}\big) = \dfrac{1}{2} + \mathcal{O}\big(K_\ell^{-\delta_2}\big)$

with some constant $\delta_2 > 0$; note that we set $m = \log_4 K_\ell$ when making comparison with (10.3).

Step 3.   So far, like in the previous section, we were estimating the probability of our sequence $\{\zeta_n\}$ going down; but to prove Lemma 8.1 we need to estimate the probability of $K(t)$ going up! From now on we start looking in the right direction, where ζ_n increases.

We claim that there are constants $\bar{K} < \bar{\bar{K}}$ and $c_1 > 0$ such that for any standard pair $\ell = (\gamma, \rho)$ such that $K_\ell \geq \bar{\bar{K}}$ and length$(\gamma) \geq K_\ell^{-100}$ for all $K > K_\ell$ we have

(10.5) $\mathbb{P}_\ell(\tau_K \leq \tau_{\bar{K}}) \geq c_1 \sqrt{K_\ell / K},$

i.e. the probability that $K(t)$ goes up to K earlier than down to $\bar{K}$ is $\geq c_1 \sqrt{K_\ell / K}$.

To prove (10.5), we define a decreasing sequence of subsets $\gamma = \gamma_0 \supset \gamma_1 \supset \gamma_2 \supset \cdots$ as follows. Let γ_1 consist of points $X \in \gamma$ where $K(t)$ goes up to $4K_\ell$ before it comes down to $K_\ell^{1/2}$ and at the first collision after reaching $4K_\ell$ they belong to standard pairs of length $\geq (4K_\ell)^{-100}$. Then inductively, let $\gamma_{i+1} \subset \gamma_i$ consist of points whose kinetic energy $K(t)$, starting on the standard pairs of energy $\sim 4^i K_\ell$ formed in the construction of γ_i, will go up to $4^{i+1} K_\ell$ before it will come down to $(4^i K_\ell)^{1/2}$ and at the first collision after reaching $4^{i+1} K_\ell$ they belong to standard pairs of length $\geq (4^{i+1} K_\ell)^{-100}$.

We assume that $\bar{K}$ is large enough to ensure a good approximation for the transition probabilities in (10.1) (recall that $K_{\min}$ in the previous section served a similar purpose, so we may identify $\bar{K}$ with $K_{\min}$). We also assume that $\bar{\bar{K}} > \bar{K}^2$ and observe that

(10.6) $\gamma_m \subset \{X \in \gamma : \tau_K \leq \tau_{\bar{K}}\},$

38

where $m = [\log_4(K/K_\ell)] + 1$.

Due to (10.4) we have an estimate for conditional probability

$$\mathbb{P}_\ell(\gamma_{i+1} \mid \gamma_i) = \frac{1}{2} + \mathcal{O}((4^i K_\ell)^{-\delta_2})$$

and obtain

$$\mathbb{P}_\ell(\gamma_m) = \mathbb{P}_\ell(\gamma_m \mid \gamma_{m-1})\mathbb{P}_\ell(\gamma_{m-1} \mid \gamma_{m-2}) \cdots \mathbb{P}_\ell(\gamma_1 \mid \gamma_0)$$

(10.7)
$$= \left(\frac{1}{2}\right)^m \prod_{i=0}^{m-1} [1 + \mathcal{O}((4^i K_\ell)^{-\delta_2})] = \mathcal{O}(\sqrt{K_\ell/K}),$$

which readily implies (10.5).

Step 4. We claim that the estimate of Step 3 is sharp, namely that under the same conditions

(10.8)
$$\mathbb{P}_\ell(\tau_K \leq \tau_{\bar{K}}) \leq c_2\sqrt{K_\ell/K}$$

for some constant $c_2 > 0$. To this end we shall show, roughly speaking, that the optimal way for the kinetic energy $K(t)$ to reach a high level K (before dropping to $\bar{K}$) is exactly as described in the construction of γ_m, every other way to reach K (before dropping to $\bar{K}$) is less probable.

We begin by observing that the 'target' level $K(t) = K$ corresponds to $\zeta_n = m_+ := \log_4(K/K_\ell) > 0$ and the 'bottom' level $K(t) = \bar{K}$ corresponds to $\zeta_n = m_- := \log_4(K_\ell/\bar{K}) < 0$. The 'model' random walk $\{\xi_n\}$ defined by (10.2) reaches m_+ before m_- with probability

$$\frac{(1/2)^{m_+} - (1/2)^{m_++m_-}}{1 - (1/2)^{m_++m_-}} = \frac{1}{2^{m_+}}\left(1 + \mathcal{O}\left(\frac{1}{2^{m_-}}\right)\right)$$

$$= \sqrt{\frac{K_\ell}{K}}\left(1 + \mathcal{O}\left(\sqrt{\frac{\bar{K}}{K_\ell}}\right)\right),$$

cf. Proposition 9.1 (a). Our sequence $\{\zeta_n\}$ can be compared with another simple random walk $\{\xi_n^*\}$ whose transition probabilities are

$$\mathbb{P}(\xi_{n+1}^* = \xi_n^* - 1) = 2/3 - \varepsilon \quad \text{and} \quad \mathbb{P}(\xi_{n+1}^* = \xi_n^* + 1) = 1/3 + \varepsilon,$$

where $\varepsilon > 0$ can be made arbitrarily small by choosing sufficiently large $\bar{K}$. Then the comparison Proposition 9.2 shows that for any standard pair

$\ell = (\gamma, \rho)$ such that $\text{length}(\gamma) > K_\ell^{-100}$ and $\bar{K} < K_\ell < K$ we have

$$(10.9) \qquad \mathbb{P}_\ell\big(\tau_K \leq \tau_{\bar{K}}\big) \leq 2 \left(\frac{\frac{1}{3} + \varepsilon}{\frac{2}{3} - \varepsilon}\right)^{\log_4(K/K_\ell)} \leq 2 \left(\frac{K_\ell}{K}\right)^{0.49}.$$

This bound is weaker than the desired (10.8), but it will be helpful.

We now return to our sets $\gamma_0 \supset \gamma_1 \supset \cdots \supset \gamma_m$ constructed in Step 3 and examine which points of $\{X \in \gamma \colon \tau_K \leq \tau_{\bar{K}}\}$ we missed. First, these are points $X \in \gamma_i$, $0 \leq i \leq m - 1$, whose kinetic energy $K(t)$, starting on the standard pairs of energy $\sim 4^i K_\ell$ formed in the construction of γ_i, will come down to $(4^i K_\ell)^{1/2}$ before it goes up to $4^{i+1} K_\ell$, and then still it goes up to $4^{i+1} K_\ell$ before it comes further down to $\bar{K}$. Second, these are points $X \in \gamma_i$ whose images at the first collision after reaching $4^{i+1} K_\ell$ belong to standard pairs of length $\leq (4^{i+1} K_\ell)^{-100}$.

Consider the points of the first type whose images at the first collision after their kinetic energy drops to $(4^i K_\ell)^{1/2}$ belong to standard pairs of length $\geq (4^i K_\ell)^{-50}$. The relative measure of the set of such points is bounded, due to (10.9), by

$$\big[(4^i K_\ell)^{1/2}/4^{i+1} K_\ell\big]^{0.49} \leq (4^i K_\ell)^{-0.2},$$

which can be incorporated into the error term of (10.7). So it is unlikely that the kinetic energy deviates from the pattern described in Step 3.

It remains to estimate the measure of the set of points whose images fall into inadmissibly short standard pairs. This set cannot be ignored as even at the first step, $i = 1$, its measure is of order K_ℓ^{-100}, which may be much larger than our estimate $c_2\sqrt{K_\ell/K}$. So we need to wait until short standard pairs expand, by which time their kinetic energy may change significantly.

To account for the expansion of short standard pairs and the resulting extra change of $K(t)$, we define a sequence of 'stopping times' $0 = \tau_0^* < \tau_1^* < \cdots$ on γ as follows: $\tau_{i+1}^*(X)$ is the first time, after $\tau_i^*(X)$, such that $K(\tau_{i+1}^*) \geq 4K(\tau_i^*)$ and the image of X at the first collision after $\tau_{i+1}^*(X)$ belongs to a standard pair of length $\geq [K(\tau_i^*)]^{-100}$. In other words, not only the kinetic energy must quadruple between τ_i^* and τ_{i+1}^*, but we also wait until the point falls into an admissibly long standard pair (thus $K(\tau_{i+1}^*)$ may be actually much larger than $4K(\tau_i^*)$). As before, it should normally take $m = \log_4(K/K_\ell) + 1$ steps to reach the level K, but it may take fewer steps if standard pairs expand slowly (and the kinetic energy happens to grow fast).

Considering separately two cases, as before,

(10.10) $$\big\{X \colon \tau_{i+1}^* \le \tau_{K_{\tau_i^*}^{1/2}}\big\} \quad \text{and} \quad \big\{X \colon \tau_{i+1}^* > \tau_{K_{\tau_i^*}^{1/2}}\big\}$$

we obtain a conditional probability formula

$$\mathbb{P}_\ell\big(\tau_{i+1}^* \le \tau_{\bar{K}} \mid \tau_i^* \le \tau_{\bar{K}}\big) = \frac{1}{2} + \mathcal{O}\big((2^i K_\ell)^{-\delta_3}\big)$$

for some constant $\delta_3 > 0$, with the main contribution coming from the first alternative in (10.10). Therefore, as in (10.7), for every $i \ge 1$ we have

$$\mathbb{P}_\ell\big(\tau_i^* < \tau_{\bar{K}}\big) \le C'\Big(\frac{1}{2}\Big)^i,$$

where $C' > 0$ is a constant. Now we have

(10.11) $$\mathbb{P}_\ell\big(\tau_K \le \tau_{\bar{K}}\big) \le \sum_{i=0}^{m-1} \mathbb{P}_\ell\Big(\{\tau_{m-i}^* < \tau_{\bar{K}}\} \bigcap J_i\Big)$$

where

$$J_i = \Bigg\{\prod_{j=1}^{m-i}\Big(\frac{\hat{K}_{\tau_{j+1}^*}}{4K_{\tau_j^*}}\Big) \ge 4^i\Bigg\}$$

is the set of points $X \in \gamma$ where $K(\tau_{m-i}^*) \ge K$.

Observe that $J_0 = \gamma$, and actually the term $i = 0$ in (10.11) can be estimated as in (10.7); but for the other terms we need a good upper bound on the measure of J_i, $i \ge 1$. These sets consist of points whose images fall into short standard pairs for many consecutive iterations of $\mathcal{F}$; and by the Growth Lemma 2.1 (c) short standard pairs expand to proper standard pairs exponentially fast, which should make $\mathbb{P}_\ell(J_i)$ small.

Also recall that the kinetic energy is proportional to the x coordinate of the Galton particle, cf. (1.3), and the latter changes by $\le \max|L(X)| < \infty$ between collisions, due to our finite horizon assumption. Thus when the kinetic energy grows from $4^i K_\ell$ to $4^{i+1} K_\ell$, the Galton particle must experience at least $c(4^{i+1} K_\ell - 4^i K_\ell)$ collisions, where $c = c(\mathcal{D}) > 0$ is a constant, thus we have $> c4^i K_\ell$ iterations of $\mathcal{F}$ at our disposal. This allows us to prove that for some constants $C > 0$ and $\theta < 1$

$$\mathbb{P}_\ell\big(J_i \mid \tau_{m-i}^* \le \tau_{\bar{K}}\big) \le C\theta^{c4^i}$$

41

(see [18], the proof of Lemma 11.2 for more details). Lastly we obtain

$$\mathbb{P}_\ell\big(\tau_K \leq \tau_{\bar{K}}\big) \leq \sum_{i=0}^{m-1} \mathbb{P}_\ell\big(J_i \mid \tau^*_{m-i} \leq \tau_{\bar{K}}\big) \mathbb{P}_\ell\big(\tau^*_{m-i} < \tau_{\bar{K}}\big)$$

$$\leq \sum_{i=0}^{m-1} CC'\theta^{c4^i}(1/2)^{m-i} = \mathcal{O}\big((1/2)^m\big)$$

$$= \mathcal{O}\big(\sqrt{K_\ell/K}\big),$$

which completes the proof of (10.8).

Step 5. We need an extension of (10.8), namely

$$(10.12) \qquad \mathbb{P}_\ell\big(nK^{3/2} \leq \tau_K \leq \tau_{\bar{K}}\big) \leq c_3\theta^n \sqrt{K_\ell/K}$$

for some constants $c_3 > 0$ and $\theta < 1$ and all $n \geq 1$. To prove this we just amend our argument in Step 4 by using Lemma 9.4. The key estimate is

$$(10.13) \qquad \mathbb{P}_\ell\big(\tau^*_{i+1} - \tau^*_i \geq (4^{i+1}K_\ell)^{3/2}n \mid \tau^*_{i+1} \leq \tau_{\bar{K}}\big) \leq \theta^n$$

for every $i \geq 0$ and $n \geq 1$. Indeed, if $K(\tau^*_i) < 4^{i+1}K_\ell$ (the 'normal case'), we apply Lemma 9.4 consecutively n times (recall that we may identify $K_{\min}$ and $\bar{K}$). If $K(\tau^*_i) \geq 4^{i+1}K_\ell$ (the 'abnormal case'), then we combine Lemma 9.4 with the previous estimates on the measure of J_i. After proving (10.13) we continue arguing as in Step 4 to complete the proof of (10.12).

Step 6. For any $X \in \Omega$, we call an *excursion above* $\bar{K}$ a maximal time interval $[t_1, t_2]$ such that $K(t) > \bar{K}$ for $t_1 < t < t_2$ and at the first collision right after t_1 the image of X belongs to a proper standard pair. Let $[t_1^{(i)}, t_2^{(i)}]$ denote the ith excursion above $\bar{K}$ along the trajectory of X and $l^{(i)} = t_2^{(i)} - t_1^{(i)}$ its duration. The following 'tail bound' is a dynamical analogue of (5.10): for any proper standard pair ℓ

$$(10.14) \qquad \mathbb{P}_\ell(l^{(i)} \geq A) \leq CA^{-1/3}$$

for some constant $C > 0$ and all $A > 0$. Obviously, it is enough to prove (10.14) for $i = 1$. Let $M = \max_{[t_1^{(1)}, t_2^{(1)}]} K(t)$. Then

$$\mathbb{P}_\ell\big(l^{(1)} \geq A\big) \leq \mathbb{P}_\ell\big(M \geq A^{2/3}\big) + \sum_{m \geq 0} \mathbb{P}_\ell\Big(\frac{A^{2/3}}{2^m} \leq M \leq \frac{A^{2/3}}{2^{m-1}}, \ l^{(1)} \geq A\Big).$$

The first term here is $\mathcal{O}\big(1/\sqrt{A^{2/3}}\big)$ by Step 4, while the mth term of the sum is $\mathcal{O}\big(A^{-1/3}2^{m/2}\theta^{2^{3m/2}}\big)$ by Step 5. Summing up proves (10.14).

Step 7. According to Step 3, the probability that any given excursion above $\bar{K}$ reaches a high level K is $\geq cK^{-1/2}$ for some small constant $c > 0$. A standard decorrelation argument [7, Lemma 6.11] shows that different excursions are virtually independent. Hence there is a large $\bar{C} > 0$ such that at least one of the first $N = [\bar{C}K^{1/2}]$ excursions reaches the level K is > 0.99, i.e. $\mathbb{P}_\ell\big(\tau_K \leq t_2^{(N)}\big) \geq 0.99$.

Next let $l_N = \sum_{i=1}^N l^{(i)}$ denote the combined length of our N excursions above $\bar{K}$. In addition, denote by $d^{(i)} = t_1^{(i)} - t_2^{(i-1)}$ the length of the ith interval between consecutive excursions and put $d_N = \sum_{i=1}^N d^{(i)}$.

To complete the proof of Lemma 8.1, it is obviously enough to show that there are large constants $Q > 0$ and $D > 0$ such that

$$(10.15) \qquad\qquad \mathbb{P}_\ell\big(l_N \leq QK^{3/2}\big) \geq 0.99$$

and

$$(10.16) \qquad\qquad \mathbb{P}_\ell\big(d_N \leq DK^{1/2}\ln K\big) \geq 0.99.$$

To prove (10.15), we denote by $N_{m,Q}$ the number of excursions of length $\geq 4^{-m}QK^{3/2}$, i.e.

$$N_{m,Q} = \text{Card}\big\{1 \leq i \leq N\colon l^{(i)} > 4^{-m}QK^{3/2}\big\}.$$

According to (10.14), we have

$$\mathbb{E}_\ell(N_{m,Q}) = \sum_{i=1}^N \mathbb{P}_\ell\big(l^{(i)} > 4^{-m}QK^{3/2}\big) = \mathcal{O}\big(2^{2m/3}Q^{-1/3}\big).$$

Given a proper standard pair $\ell = (\gamma, \rho)$ let $\gamma_{m,Q}$ denote the set of points $X \in \gamma$ where $N_{m,Q} \geq 2^{m-1}$; now Chebyshev's inequality gives

$$\mathbb{P}_\ell\big(\gamma_{m,Q}\big) \leq 2^{-m+1}\mathbb{E}_\ell(N_{m,Q}) = \mathcal{O}\big(2^{-m/3}Q^{-1/3}\big).$$

Observe that for each $X \in \gamma \setminus \cup_{m\geq 1}\gamma_{m,Q}$ we have

$$l_N = \sum_{i=1}^N l^{(i)} \leq \sum_{m=1}^\infty 2^{-m}QK^{3/2} = QK^{3/2}.$$

43

On the other hand,

$$\mathbb{P}_\ell\big(\cup_{m\geq 1}\gamma_{m,Q}\big) \leq C \sum_{m\geq 1} 2^{-m/3} Q^{-1/3} \leq C Q^{-1/3}.$$

If Q is large enough, then $\mathbb{P}_\ell\big(\cup_{m\geq 1}\gamma_m\big) < 0.01$, which proves (10.15).

To prove (10.16), we first show that for some constants $\bar{c} > 0$ and $\bar{\theta} < 1$ and all $A > 0$

$$(10.17) \qquad\qquad \mathbb{P}_\ell\big(d^{(i)} > A\big) \leq \bar{c}\,\bar{\theta}^A.$$

Indeed, due to the conservation of energy (1.3) the condition $K(t) \leq \bar{K}$ corresponds to $x(t) \leq \bar{x}$ for some constant $\bar{x} > 0$, so (10.17) is equivalent to

$$(10.18) \qquad\qquad \mathbb{P}_\ell\Big(\max_{0<t<A} x(t) < \bar{x}\Big) \leq \bar{c}\,\bar{\theta}^A.$$

To prove this, consider a finite-size billiard table $\mathcal{D}_{\bar{x}}$ that is the union of the rectangle $\mathcal{D}_{\bar{x}}^+ = \{(x,y)\colon 0 < y < L_y,\ 0 < x < \bar{x}\}$ (minus the scatterers, of course) and its mirror image, $\mathcal{D}_{\bar{x}}^-$, across the line $x = 0$. Imposing periodic boundary conditions on $\mathcal{D}_{\bar{x}}$ we get a dispersing billiard on a torus of size $(2\bar{x}) \times L_y$ with finitely many convex scatters; it has finite horizon. The dynamics of the Galton particle, during time intervals when $x(t) \leq \bar{x}$, corresponds to the motion of the billiard particle on $\mathcal{D}_{\bar{x}}$ subjected to an external field that coincides with $\mathbf{g}$ on $\mathcal{D}_{\bar{x}}^+$ and with $-\mathbf{g}$ on $\mathcal{D}_{\bar{x}}^-$ (we note that adding the mirror image of $\mathcal{D}_{\bar{x}}^+$ across the line $x = 0$ is necessary to account for the reflections of the Galton particle off the 'lid' $x = 0$). The Galton particle crosses the line $x = \bar{x}$ whenever our billiard particle on the table $\mathcal{D}_{\bar{x}}$ crosses its left or right side.

The billiard system in $\mathcal{D}_{\bar{x}}$ has strong statistical properties, in particular diffusive behavior [6]; the latter means that on the universal cover the average displacement of the billiard particle in the x direction at time t is $\sim \sqrt{t}$. Hence for some constants $\bar{t} > 0$ and $\bar{p} > 0$ and any proper standard pair ℓ

$$\mathbb{P}_\ell\Big(\max_{0<t<\bar{t}} |x(t)| < \bar{x}\Big) \leq 1 - \bar{p}.$$

Now (10.18) can be obtained via Growth Lemma 2.1 and induction on $A/\bar{t}$.

Lastly, due to (10.17), we have

$$\mathbb{P}_\ell\Big(\max_{1\leq i\leq N} d^{(i)} > D \ln K\Big) \leq N\bar{c}\,\bar{\theta}^{D\ln K}$$

which is < 0.01 provided $D > 0$ large enough.

This completes the proof of (10.16) and that of Lemma 8.1.

A Proof of Theorem 4

The derivation of part (a) of Theorem 4 from Lemma 7.1 is very similar to the arguments of [7, 15, 19]. Here we present the main steps following the last reference since it gives a better bound on the error term (which is part (b) of Theorem 4).

Consider the stochastic process $K_\varepsilon^\dagger(\xi)$ which coincides with $\tilde{K}_{[\xi\varepsilon^{-2}]}$ as long as $\xi\varepsilon^{-2} \leq n_\mathcal{I}$, i.e. as long as $\tilde{K}_{[\xi\varepsilon^{-2}]}$ stays within the interval $\mathcal{I}$, and once it exists that interval (i.e. for $\xi\varepsilon^{-2} \geq n_\mathcal{I}$) we require that $K_\varepsilon^\dagger(\xi)$ continues as an abstract diffusion process satisfying (7.1) (in particular, its further increments will be independent of our dynamical system).

The proof of part (a) consists of two standard steps:

(I) showing that the family $\{K_\varepsilon^\dagger(\xi)\}$ is tight, hence it has at least one weak limit point (i.e. any sequence $\{K_{\varepsilon_k}^\dagger(\xi)\}$, $\varepsilon_k \to 0$, contains a subsequence weakly converging to a limit process);

(II) showing that finite-dimensional distributions of $\{K_\varepsilon^\dagger(\xi)\}$ converge, as $\varepsilon \to 0$, to those of the (unique) diffusion process $\mathcal{K}^*(\xi)$ satisfying (7.1) (but, unlike $\mathcal{K}^*$ in Theorem 4, it is *not* stopped when it exists $\mathcal{I}$).

Tightness means that typical trajectories of the family $K_\varepsilon^\dagger(\xi)$ do not oscillate too much; its verification is a fairly routine procedure: one uses equidistribution properties of the underlying dynamical system (in our case – the averaging lemma 7.1) to estimate higher moments of the given family and then deduces the tightness by a standard argument. Such proofs may be found in [15, Section 7, steps 1 and 2] and [7, Section 6.5], and we refer the reader to these works for more details.

The argument in Step (II) is specific to our system, so we sketch the proof here. To verify the convergence of finite-dimensional distributions it is enough to show that if $\mathcal{K}(\xi)$ is *any* weak limit point for the family $\{K_\varepsilon^\dagger(\xi)\}$, then for every finite ordered sequence of time moments $0 < s_1 < s_2 < \cdots < s_r < T_1 < T_2$ we have

$$\mathbb{E}\left(\left[\Phi(\mathcal{K}(T_2)) - \Phi(\mathcal{K}(T_1)) - \int_{T_1}^{T_2} \left(\bar{a}(\mathcal{K})\Phi'(\mathcal{K}) + \frac{\bar{\sigma}^2}{2}\Phi''(\mathcal{K})\right)(\bar{\xi})\,d\bar{\xi}\right] \right.$$
$$\left. \times \prod_{j=1}^{r} \Psi_j(\mathcal{K}(s_j))\right) = 0$$

45

for a sufficiently large class of functions Φ and Ψ_j. We will assume that Φ is C^3 smooth and Ψ_j are Hölder continuous.

In terms of the kinetic energy process $\tilde{K}_n$, the above relation means that given any fixed $T > 0$ we have

$$\mathbb{E}_{\ell_\varepsilon}\left(\left[\Phi(\tilde{K}_{n_1^*}) - \Phi(\tilde{K}_{n_2^*}) - \varepsilon^2 \sum_{n=n_1^*}^{n_2^*} \left(\bar{a}(\tilde{K}_n)\Phi'(\tilde{K}_n) + \frac{\bar{\sigma}^2}{2}\Phi''(\tilde{K}_n)\right)\right]\right.$$
$$\left. \times \prod_{j=1}^{r} \Psi_j(\tilde{K}_{m_j})\right) = o(1)$$

as $\varepsilon \to 0$, uniformly over $m_1 < m_2 < \cdots < m_r < n_1 < n_2 < T\varepsilon^{-2}$, where we denote $n_1^* = \min\{n_1, n_\mathcal{I}\}$ and $n_2^* = \min\{n_2, n_\mathcal{I}\}$.

A standard decorrelation estimate [7, Lemma 6.11] allows us to eliminate the $\Psi_j(\tilde{K}_{m_j})$ factors and reduce the above expression to

$$(\text{A.1}) \quad \mathbb{E}_\ell\left[\Phi(\tilde{K}_{\bar{n}^*}) - \Phi(\tilde{K}_0) - \varepsilon^2 \sum_{n=0}^{\bar{n}^*} \left(\bar{a}(\tilde{K}_n)\Phi'(\tilde{K}_n) + \frac{\bar{\sigma}^2}{2}\Phi''(\tilde{K}_n)\right)\right] = o(1)$$

as $\varepsilon \to 0$, uniformly over $\bar{n} < T\varepsilon^{-2}$ and standard pairs $\ell = (\gamma, \rho)$ of $\text{length}(\gamma) > \varepsilon^{100}$; here we denote $\bar{n}^* = \min\{\bar{n}, n_\mathcal{I}\}$.

In addition, to prove part (b) of Theorem 4 we need the right hand side of (A.1) be $\mathcal{O}(\varepsilon^\delta)$, so we will prove (A.1) with an explicit error estimate. Using (3.11) and Taylor expansion gives

$$\Phi(\tilde{K}_{n+1}) - \Phi(\tilde{K}_n) = \varepsilon\Phi'(\tilde{K}_n)gL_{e_n}(X_n) + \Phi'(\tilde{K}_n)\mathcal{R}_1^{(n)}$$
$$+ \tfrac{1}{2}\varepsilon^2\Phi''(\tilde{K}_n)g^2 L_{e_n}^2(X_n) + \varepsilon\Phi''(\tilde{K}_n)gL_{e_n}(X_n)\mathcal{R}_1^{(n)}$$
$$(\text{A.2}) \qquad\qquad + \tfrac{1}{2}\Phi''(\tilde{K}_n)(\mathcal{R}_1^{(n)})^2 + \mathcal{O}(\varepsilon^3).$$

First we observe that by (3.12), the transversality of unstable curves to the singularity manifolds, and Growth Lemma 2.1 (b), we have $\mathbb{E}_\ell(\mathcal{R}_1^{(n)}) = \mathcal{O}(\varepsilon^3)$ uniformly in $n \geq 0$. Thus the total contribution from every term in (A.2) which contains $\mathcal{R}_1^{(n)}$ will be $\mathcal{O}(\varepsilon^3\bar{n}) = \mathcal{O}(\varepsilon)$, so these terms can be neglected altogether. Thus (A.2) reduces to

$$(\text{A.3}) \qquad \Phi(\tilde{K}_{n+1}) - \Phi(\tilde{K}_n) = \varepsilon g\Phi'(\tilde{K}_n)L_{e_n}(X_n) + \tfrac{1}{2}\varepsilon^2 g^2\Phi''(\tilde{K}_n)L_{e_n}^2(X_n)$$

where negligible terms are dropped for brevity.

We start by estimating the sum of the second terms in (A.3). Our idea is that $\Phi''(\tilde{K}_n)$ changes with n very slowly, so we can accurately approximate it by a constant within $N \sim |\ln \varepsilon|$ consecutive iterations of $\mathcal{F}$, in the course of which the sum of $L_{e_n}^2(X_n)$ approaches its phase-space average, which, in turn, is close to the corresponding billiard phase-space average $\mu_0(L_0^2) = \int_{\Omega_0} L_0^2(X)\, d\mu_0(X)$.

Recall that the map $X_n \mapsto X_{n+1}$ is an $\mathcal{O}(\varepsilon)$-perturbation of the billiard map $X_n \to \mathcal{F}_0(X_n)$, cf. Section 3. We use again, as in Section 7, the shadowing-type argument (presented in detail in [7, Proposition 3.3]), which gives

$$\text{(A.4)} \qquad \mathbb{E}_\ell\big(L_{e_N}^2(X_N)\big) = \mu_0(L_0^2) + \mathcal{O}(\varepsilon N)$$

for any $N > 0$, uniformly over standard pairs of length $> \varepsilon^{100}$.

Next, we have $\tilde{K}_n = \tilde{K}_{n-2N} + \mathcal{O}(\varepsilon N)$, and since Φ'' is a Lipschitz continuous function we can write

$$\text{(A.5)} \quad \mathbb{E}_\ell\big[L_{e_n}^2(X_n)\Phi''(\tilde{K}_n)\mathbf{1}_{n<n_\mathcal{I}}\big] = \mathbb{E}_\ell\big[L_{e_n}^2(X_n)\Phi''(\tilde{K}_{n-2N})\mathbf{1}_{n<n_\mathcal{I}}\big] + \mathcal{O}(\varepsilon N),$$

where $\mathbf{1}_Z$ stands for the indicator of the set Z. Consider the decomposition

$$\mathbb{E}_\ell\big((A \circ \mathcal{F}^{n-N})\mathbf{1}_{n_\mathcal{I}>n-N}\big) = \sum_\alpha c_\alpha \mathbb{E}_{\ell_\alpha}(A),$$

where $\ell_\alpha = (\gamma_\alpha, \rho_\alpha)$ are the components of $\mathcal{F}^{n-N}(\gamma)$ on which $\tilde{K}$ has not exited the interval $\mathcal{I}$ yet, cf. (2.4). The crucial observation is that the function $\tilde{K}_{n-2N}$ is approximately constant on the $\mathcal{F}^{-N}$-preimage of every γ_α; more precisely, its oscillations do not exceed $\mathcal{O}(\vartheta^N)$, where $\vartheta < 1$ is the hyperbolicity constant. Replacing $\tilde{K}_{n-2N}$ with its average on the preimage of every γ_α allows us to apply (A.4) to every component γ_α of length $> \varepsilon^{100}$; the contribution of shorter components is $\mathcal{O}(\varepsilon^{100})$ due to the Growth Lemma 2.1 (b). Thus

$$\mathbb{E}_\ell\big[L_{e_n}^2(X_n)\Phi''(\tilde{K}_{n-2N})\mathbf{1}_{n_\mathcal{I}>n-N}\big] = \mu_0(L_0^2)\,\mathbb{E}_\ell\big[\Phi''(\tilde{K}_{n-2N}\mathbf{1}_{n_\mathcal{I}>n-N})\big]$$
$$\text{(A.6)} \qquad\qquad\qquad\qquad + \mathcal{O}(\varepsilon N) + \mathcal{O}(\vartheta^N).$$

Choosing $N = C|\ln \varepsilon|$ with a large $C > 0$ ensures that the combined error term in (A.6) is $\mathcal{O}(\varepsilon|\ln \varepsilon|)$.

However, (A.6) falls a little short of our goal because it contains $\mathbf{1}_{n_\mathcal{I}>n-N}$ rather than $\mathbf{1}_{n_\mathcal{I}>n-2N}$. The difference between those terms is

$$\mathcal{O}\big(\mathbb{E}_\ell(\mathbf{1}_{n_\mathcal{I}>n-N} \neq \mathbf{1}_{n_\mathcal{I}>n-2N})\big) = \mathcal{O}\big(\mathbb{P}_\ell(n_\mathcal{I} \in [n-2N, n-N])\big),$$

and we observe that $\sum_{n=0}^{\bar{n}^*} \mathbb{P}_\ell(n_\mathcal{I} \in [n-2N, n-N]) \leq N$, hence this difference will contribute to the final estimate a quantity of order $\varepsilon^2 N$, which is $\ll \varepsilon^\delta$. We record our final estimate for the sum of the second terms in (A.3)

$$(\text{A.7}) \quad \varepsilon^2 \, \mathbb{E}\left[\sum_{n=0}^{\bar{n}^*} \Phi''(\tilde{K}_n) L_{e_n}^2(X_n)\right] = \varepsilon^2 \, \mu_0(L_0^2) \, \mathbb{E}_\ell\left[\sum_{n=0}^{\bar{n}^*} \Phi''(\tilde{K}_n)\right] + \mathcal{O}(\varepsilon^\delta).$$

We proceed to estimating the sum of the first terms in (A.3). Again 'stepping back' $2N$ iterations (with $N = C|\ln \varepsilon|$), using Taylor expansion and (3.11) yield

$$\Phi'(\tilde{K}_n) = \Phi'(\tilde{K}_{n-2N}) + \varepsilon g \sum_{j=n-2N}^{n-1} \Phi''(\tilde{K}_{n-2N}) L_{e_j}(X_j)$$

$$(\text{A.8}) \qquad\qquad + \sum_{j=n-2N}^{n-1} \Phi''(\tilde{K}_{n-2N}) \mathcal{R}_1^{(j)} + \mathcal{O}(\varepsilon^2 N^2).$$

As before, it is easy to see that the contribution of the last two terms can be neglected, so we only have to analyze the first two.

An argument similar to the one used to derive (A.7) now gives

$$\varepsilon g \mathbb{E}_\ell\left[\sum_{n=0}^{\bar{n}^*} \Phi'(\tilde{K}_{n-2N}) L_{e_n}(X_n)\right] = \varepsilon^2 \mathbb{E}_\ell\left[\sum_{n=0}^{\bar{n}^*} \Phi'(\tilde{K}_n) \, \bar{a}(\tilde{K}_n)\right] + \mathcal{O}(\varepsilon^\delta),$$

except here we use Lemma 7.1 instead of (A.4).

It remains to estimate the sum of the second terms in (A.8), i.e.

$$(\text{A.9}) \qquad \varepsilon^2 g^2 \, \mathbb{E}_\ell\left[\sum_{n=0}^{\bar{n}^*} \sum_{j=n-2N}^{n-1} \Phi''(\tilde{K}_{n-2N}) L_{e_j}(X_j) L_{e_n}(X_n)\right]$$

Let $N_1 = N/100$. It is also convenient to change the time variable $k = n - j$. We have two cases:

Case 1: $k \geq 2N_1$. This is a relatively simple case, in which we can 'decorrelate' the factors $L_{e_{n-k}}(X_{n-k})$ and $L_{e_n}(X_n)$. Consider the decomposition

$$\mathbb{E}_\ell\left[\Phi''(\tilde{K}_{n-2N}) L_{e_{n-k}}(X_{n-k}) L_{e_n}(X_n) \mathbf{1}_{n<n_\mathcal{I}}\right]$$

$$= \sum_\alpha c_\alpha \mathbb{E}_{\ell_\alpha}\left[\Phi''(\tilde{K}_{N_1-2N}) L_{e_{N_1-k}}(X_{N_1-k}) L_{e_{N_1}}(X_{N_1})\right]$$

$$+ \mathcal{O}\left(\mathbb{P}_\ell(n_\mathcal{I} \in [n - N_1, n])\right),$$

48

where $\ell_\alpha = (\gamma_\alpha, \rho_\alpha)$ are the components of $\mathcal{F}^{n-N_1}(\gamma)$ on which $\tilde{K}$ has not exited the interval $\mathcal{I}$ yet. The first two factors, $\Phi''(\tilde{K}_{N_1-2N})$ and $L_{e_{N_1-k}}(X_{N_1-k})$, vary little on every component γ_α; more precisely, their oscillations do not exceed $\mathcal{O}(\vartheta^{N_1})$. So we can approximate them by constants (say, their averages) on every γ_α; now Lemma 7.1 gives $\mathbb{E}_{\ell_\alpha}(L_{e_{N_1}}(X_{N_1})) = \mathcal{O}(\varepsilon)$ for every component γ_α of length $> \varepsilon^{100}$, and the contribution of shorter components is $\mathcal{O}(\varepsilon^{100})$ due to the Growth Lemma 2.1 (b). Thus

$$\mathbb{E}_\ell\big[\Phi''\big(\tilde{K}_{n-2N}\big)L_{e_{n-k}}\big(X_{n-k}\big)L_{e_n}\big(X_n\big)\mathbf{1}_{n<n_\mathcal{I}}\big] = \mathcal{O}(\vartheta^{N_1}) + \mathcal{O}(\varepsilon^3)$$
$$+ \mathcal{O}\big(\mathbb{P}_\ell(n_\mathcal{I} \in [n-N_1, n])\big).$$

We assume that $N = C|\ln\varepsilon|$ with such a large $C > 0$ that $\vartheta^{N_1} = \mathcal{O}(\varepsilon^3)$, then the total contribution of the first two error terms to the final estimate will be negligible.

Case 2: $k < 2N_1$. In this case we no longer can 'decorrelate' the factors $L_{e_{n-k}}(X_{n-k})$ and $L_{e_n}(X_n)$, so their correlation will appear in the final estimate.

The transversality of unstable curves to singularities and the Growth Lemma 2.1 imply that $\mathbb{P}_\ell(\mathcal{F}^{-n}\mathcal{U}_\varepsilon^2) = \mathcal{O}(\varepsilon)$ uniformly in $n \geq C|\ln\varepsilon|$, thus (3.13) allows us to replace $L_{e_{n-k}}(X_{n-k})L_{e_n}(X_n)$ by $L_0(X_{n-k})L_0(X_n)$ at the cost of an extra error term $\mathcal{O}(\varepsilon^3\bar{n}^*N) = \mathcal{O}(\varepsilon N)$ in the final estimate. So (A.9) can be replaced with

$$(\text{A}.10) \qquad \varepsilon^2\mathbb{E}_\ell\Bigg[\sum_{n=0}^{\bar{n}^*}\sum_{k=1}^{2N}\Phi''\big(\tilde{K}_{n-2N}\big)L_0\big(X_{n-k}\big)L_0(X_n)\Bigg],$$

and in Case 2 we restrict the internal sum to $1 \leq k < 2N_1$.

We claim that

$$\mathbb{E}_\ell\big[\Phi''\big(\tilde{K}_{n-2N}\big)L_0(X_{n-k})L_0(X_n)\mathbf{1}_{n<n_\mathcal{I}}\big] = \mu_0\big((L_0 \circ \mathcal{F}_0^k)\,L_0\big)$$
$$(\text{A}.11) \quad \times \mathbb{E}_\ell\big[\Phi''\big(\tilde{K}_{n-2N}\big)\mathbf{1}_{n-2N<n_\mathcal{I}}\big] + \mathcal{O}(\varepsilon^\delta) + \mathcal{O}\big(\mathbb{P}_\ell(n_\mathcal{I} \in [n-2N, n])\big).$$

Indeed, as before, we can replace $\mathbf{1}_{n<n_\mathcal{I}}$ by $\mathbf{1}_{n-2N<n_\mathcal{I}}$ at the cost of an extra error term $\mathcal{O}(\mathbb{P}_\ell(n_\mathcal{I} \in [n-2N, n]))$ in the final estimate. Now the same shadowing argument as before allows us to replace $\mathcal{F}$ with $\mathcal{F}_0$ and obtain

$$\mathbb{E}_\ell\big[\Phi''\big(\tilde{K}(X_{n-2N})\big)L_0\big(\mathcal{F}^{2N-k}X_{n-2N}\big)L_0\big(\mathcal{F}^{2N}X_{n-2N}\mathbf{1}_{n-2N<n_\mathcal{I}}\big)\big]$$
$$= \mathbb{E}_\ell\big[\Phi''\big(\tilde{K}(X_{n-2N})\big)L_0\big(\mathcal{F}_0^{2N-k}X_{n-2N}\big)L_0\big(\mathcal{F}_0^{2N}X_{n-2N}\big)\mathbf{1}_{n-2N<n_\mathcal{I}}\big] + \mathcal{O}(N\varepsilon).$$

Next we consider a decomposition

$$
\mathbb{E}_\ell \big[\Phi'' \big(\tilde{K}(X_{n-2N}) \big) L_0 \big(\mathcal{F}_0^{2N-k} X_{n-2N} \big) L_0 \big(\mathcal{F}_0^{2N} X_{n-2N} \big) \mathbf{1}_{n-2N < n_\mathcal{I}} \big]
$$
$$
= \sum_\alpha c_\alpha \mathbb{E}_{\ell_\alpha} \big[\Phi'' \big(\mathcal{F}_0^{-N_1} \tilde{K}(Y) \big) L_0 \big(\mathcal{F}_0^{2N-k-N_1} Y \big) L_0 \big(\mathcal{F}_0^{2N-N_1} Y \big) \big] + \mathcal{O}(N\varepsilon),
$$

where $\ell_\alpha = (\gamma_\alpha, \rho_\alpha)$ denote the components of $\mathcal{F}_0^{N_1} \circ \mathcal{F}^{n-2N}(\gamma)$; more precisely, γ_α are obtained from the components of $\mathcal{F}^{n-2N}(\gamma)$ on which $\tilde{K}$ has not exited the interval $\mathcal{I}$ yet, by pushing them forward under the billiard map $\mathcal{F}_0^{N_1}$. On each component γ_α the first factor $\Phi''(\mathcal{F}_0^{-N_1} \tilde{K}(Y))$ varies little, namely, it equals $\phi_\alpha + \mathcal{O}(\vartheta^{N_1})$, where ϕ_α denotes its average on γ_α. So we need to estimate

$$
\sum_\alpha c_\alpha \phi_\alpha \, \mathbb{E}_{\ell_\alpha} \big[L_0 \big(\mathcal{F}_0^{2N-k-N_1} Y \big) L_0 \big(\mathcal{F}_0^{2N-N_1} Y \big) \big]
$$

where the sum can be restricted to indices with $\mathrm{length}(\gamma_\alpha) \geq \varepsilon^{100}$, as the contribution of shorter components is $\mathcal{O}(\varepsilon^{100})$. The equidistribution property of the billiard map for 'multiple observables' [11, Theorem 7.33] implies

$$
\left| \mathbb{E}_{\ell_\alpha} \big[(L_0 \circ \mathcal{F}_0^{2N-k-N_1})(L_0 \circ \mathcal{F}_0^{2N-N_1}) \big] - \mu_0 \big((L_0 \circ \mathcal{F}_0^k) L_0 \big) \right| = \mathcal{O}\big(\theta^{2N-N_1-k} \big)
$$

for some constant $\theta < 1$. So if $N = C|\ln \varepsilon|$ with a large enough $C > 0$, then the above estimate becomes $\mathcal{O}(\varepsilon^3)$.

Lastly we have

$$
\sum_\alpha c_\alpha \phi_\alpha = \mathbb{E}_\ell \big[\Phi'' \big(K_{n-2N} \big) \mathbf{1}_{n-2N < n_\mathcal{I}} \big] + \mathcal{O}(\theta^{N_1}) + \mathcal{O}(\varepsilon^{100}),
$$

which completes our proof of (A.11).

Now we perform summation over k in (A.11) and observe that

$$
\sum_{k=1}^{2N_1-1} \mu_0 \big((L_0 \circ \mathcal{F}_0^k) L_0 \big) = \sum_{k=1}^{\infty} \mu_0 \big((L_0 \circ \mathcal{F}_0^k) L_0 \big) + \mathcal{O}\big(\theta^{N_1} \big)
$$
$$
= \frac{1}{2g^2} \Big[\bar{\sigma}^2 - g^2 \mu_0(L_0^2) \Big] + \mathcal{O}(\theta^{N_1})
$$

for some constant $\theta < 1$, due to exponential decay of correlations in dispersing billiards.

Combining all our estimates we obtain (A.1) with error bound $\mathcal{O}\big(\varepsilon |\ln \varepsilon| \big)$, which is even better than $\mathcal{O}(\varepsilon^\delta)$ we need. This completes the proof of part

(a). It almost gives the part (b) as well, except it can happen that $n_{\mathcal{I}} > \bar{n}$. However by part (a) there exists $\theta < 1$ such that for any standard pair ℓ of length$> \varepsilon^{100}$ we have $\mathbb{P}_\ell(n_{\mathcal{I}} > \varepsilon^{-2}) < \theta$; hence by induction on k we get

$$(A.12) \qquad \mathbb{P}_\ell(n_{\mathcal{I}} > k\varepsilon^{-2}) < \theta^k + \mathcal{O}(k\varepsilon^{100}).$$

Next we decompose

$$\mathbb{E}_{\ell_\varepsilon}\big([\tilde{K}^*_{n_{\mathcal{I}}}]^{1/2}\big) = \mathbb{E}_{\ell_\varepsilon}\big([\tilde{K}^*_{n_{\mathcal{I}}}]^{1/2}\mathbf{1}_{n_{\mathcal{I}} \geq \varepsilon^{-2-\delta/2}}\big) + \mathbb{E}_{\ell_\varepsilon}\big([\tilde{K}^*_{n_{\mathcal{I}}}]^{1/2}\mathbf{1}_{n_{\mathcal{I}} < \varepsilon^{-2-\delta/2}}\big).$$

The first term is $\mathcal{O}(\varepsilon^{99})$ due to (A.12), whereas the second term is $\bar{K}^{1/2} + \mathcal{O}(\varepsilon^{\delta/2})$ by our main estimate. This completes the proof of Theorem 4. $\qquad\square$

References

[1] Billingsley P. *Probability and measure,* 3d edition. John Wiley & Sons, Inc., New York, 1995.

[2] Bunimovich L. A. and Sinai Ya. G. *Statistical properties of Lorentz gas with periodic configuration of scatterers,* Comm. Math. Phys. **78** (1980/81), 479–497.

[3] Bunimovich L. A., Sinai Ya. G., and Chernov N. I. *Statistical properties of two-dimensional hyperbolic billiards,* Russ. Math. Surv. **46** (1991), 47–106.

[4] Chepelianskii A. D. and Shepelyansky D. L. *Dynamical Turbulent Flow on the Galton Board with Friction,* Phys. Rev. Lett. **87** (2001) paper 034101 (4 pages).

[5] Chernov N. *Decay of correlations and dispersing billiards*, J. Stat. Phys. **94** (1999), 513–556.

[6] Chernov N. I. *Sinai billiards under small external forces,* Ann. Henri Poincare **2** (2001) 197–236.

[7] Chernov N. and Dolgopyat D. *Brownian Brownian Motion-1,* to appear in Memoirs AMS.

[8] Chernov N. and Dolgopyat D. *Hyperbolic billiards and statistical physics,* Proceedings ICM-2006, vol. II, 1679–1704.

[9] Chernov N. I., Eyink G. L., Lebowitz J. L. and Sinai Ya. G., *Steady-state electrical conduction in the periodic Lorentz gas*, Comm. Math. Phys. **154** (1993), 569–601.

[10] Chernov N. I., Eyink G. L., Lebowitz J. L. and Sinai Ya. G., *Derivation of Ohm's law in a deterministic mechanical model*, Phys. Rev. Lett. **70** (1993), 2209–2212.

[11] Chernov N. and Markarian R., *Chaotic Billiards*, Mathematical Surveys and Monographs, **127**, AMS, Providence, RI, 2006. (316 pp.)

[12] Dettmann C. P. and Morriss G. P., *Crisis in the periodic Lorentz gas*, Phys. Rev. E. **54** (1996), 4782–4790.

[13] Dolgopyat D. *On differentiability of SRB states for partially hyperbolic systems,* Invent. Math. **155** (2004) 389–449.

[14] Dolgopyat D. *Limit theorems for partially hyperbolic systems,* Trans. AMS **356** (2004) 1637–1689.

[15] Dolgopyat D. *Averaging and invariant measures,* Mosc. Math. J. **5** (2005), no. 3, 537–576.

[16] Dolgopyat D. *Bouncing balls in non-linear potentials,* Discrete Cont. Dynam. Syst.–A **22** (2008) 165–182.

[17] Dolgopyat D. *Fermi acceleration,* Cont. Math. **469** (2008) 149–166.

[18] Dolgopyat D., Szasz D. and Varju T. *Recurrence properties of Lorentz process,* Duke Math. J. **142** (2008) 241-281.

[19] Dolgopyat D., Szasz D. and Varju T. *Limit Theorems for perturbed planar Lorentz process,* preprint.

[20] Durrett R. *Probability: theory and examples,* 2d edition. Duxbury Press, Belmont, CA, 1996.

[21] Galton F., *Natural Inheritance,* MacMillan, 1989 (facsimile available at www.galton.org).

[22] Fermi E. *On the origin of cosmic radiation,* Phys. Rev. **75** (1949), 1169–1174.

[23] Kozlov V. V. and Mitrofanova M. Yu. *Galton Board*, Reg. Chaot. Dynam. **8** (2003) 431–439.

[24] Krapivsky P. and Redner S., *Slowly divergent drift in the field-driven Lorentz gas*, Phys. Rev. E **56** (1997) 3822.

[25] Littlewood J. E., *On the problem of n bodies*, Comm. Sem. Math. Unvi. Lund Tome Supplementaire (1952), 143–151.

[26] Lorentz H. A., *The motion of electrons in metallic bodies*, Proc. Amst. Acad. **7** (1905), 438–453.

[27] Lue A. and Brenner H. *Phase flow and statistical structure of Galton-board systems,* Phys. Rev. E **47** (1993) 3128–3144.

[28] Moran B. and Hoover W., *Diffusion in a periodic Lorentz gas*, J. Stat. Phys. **48** (1987), 709–726.

[29] Oksendal B., *Stochastic Differential Equations*, Springer, Berlin, 2003.

[30] Piasecki J. and Wajnryb E., *Long-time behavior of the Lorentz electron gas in a constant, uniform electric field*, J. Stat. Phys. **21** (1979), 549–559.

[31] Ravishankar K. and Triolo L., *Diffusive limit of Lorentz model with uniform field from the Markov approximation*, Markov Proc. Rel. Fields **5** (1999), 385–421.

[32] Revuz D. and Yor M. *Continuous martingales and Brownian motion,* 3d edition. Grundlehren der Mathematischen Wissenschaften, **293** (1999) Springer-Verlag, Berlin.

[33] Sinai Ya. G., *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.

[34] Yamada T., Kawasaki K. *Nonlinear effects in the shear viscosity of a critical mixture*, Prog. Theor. Phys. **38** (1967), 1031–1051.

[35] Young L.-S., *Statistical properties of dynamical systems with some hyperbolicity*, Ann. Math. **147** (1998) 585–650.

[36] Zaslavsky, G. M. *Chaos in dynamic systems,* Translated from the Russian by V. I. Kisin. Harwood Academic Publishers, Chur, 1985.