©Copyright 2022

Peter Francis Rudzis

Rough Collisions

Peter Francis Rudzis

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2022

Reading committee:
Krzysztof Burdzy, Chair
Robin Graham
Soumik Pal

Program authorized to offer degree:
Mathematics

University of Washington

Abstract

Rough Collisions

Peter Francis Rudzis

Chair of the Supervisory Committee:
Krzysztof Burdzy
Department of Mathematics

A rough collision law describes the limiting contact dynamics of a pair of rough rigid bodies, as the scale of the rough features (asperities) on the surface of each body goes to zero. The class of rough collision laws is quite large and includes random elements. The main results of this work characterize the rough collision laws for a freely moving rough disk and a fixed rough wall in dimension 2. Any collision law which (i) is symmetric with respect to a certain well-known invariant measure from billiards theory, and (ii) conserves the projection of the phase space velocity onto the "rolling velocity" is a rough collision law. We provide a method for explicitly constructing rough collision laws for a broad range of choices of microstructure on the disk and wall. Having established the collision dynamics, we also investigate the ergodic properties of the system consisting of a rough disk bouncing between two parallel rough walls. For such a system, we describe necessary and sufficient conditions for when Lambertian measure is ergodic. In our introduction, we review

past work in billiards, including characterizations of other rough billiard systems due to Plakhov [31, 35] and Angel, Burdzy, and Sheffield [1], which our results build upon.

Contents

1 Introduction

1.1 Motivation and main results

1.1.1 Frictional collisions

Frictional forces between colliding physical bodies arise from a combination of electrical forces and asperities (microscopic rough features) on the surface of each body. Most mathematical models for friction are phenomenological, in the sense that they do not reduce to more fundamental physical principles and typically contain basic parameters (e.g. the coefficient of friction) depending on the physical materials in play, which must be determined through empirical measurement. Models for frictional collisions can lead to paradoxical results, and there is no single model which describes friction well in all scenarios (see the review [5]). The statistical mechanical point of view has been taken much more rarely, and the relationship between the microscopic surface features on each body and the macroscopic contact dynamics is only vaguely understood.

This monograph is a mathematical work concerned with an idealized statistical model for frictional collisions. We derive dynamics under the following assumptions: (1) the frictional forces arise only from rigid asperities on the surfaces of each body (and not from electrical forces); and (2) the kinetic energy of the colliding bodies is conserved.

These postulates allow us to frame our objective in the language of mathematical billiards. Consider two rigid bodies whose surfaces are endowed with small geometric

features – bumps, crevices, etc. Associated with the two bodies is a "collision law" which governs the dynamics when the two bodies collide. The physical assumptions of our model imply that a collision may be represented by a point particle undergoing specular (mirror) reflection from the boundary of the configuration space. A *rough collision law* will be defined as a limit of a sequence of collision laws as the scale of the asperities on each body goes to zero. The limiting collision law may in general have a random "noise" component, and thus an appropriate sense of convergence must be defined to capture the full breadth of possible limiting behavior. Our goal is to describe the kinds of collision laws which may arise from such a limiting procedure.

1.1.2 Rigid body collisions

The mathematical literature on rigid body interactions falls into two categories. On the one hand, we find extensive literature on hard sphere models, where the particle-to-particle interactions are simple to describe. On the other hand, the literature on colliding rigid bodies of more general shape is much more restricted in scope, being mainly concerned with foundational issues (well-posedness) and describing the local (in space and time) contact dynamics.

Problems about interacting rigid bodies become an order of magnitude harder when one passes from spherical to non-spherical bodies. In the latter case, the configuration space can contain complicated singularities, and the dynamical evolution may not be well-defined for a small set of initial conditions, even for smooth bodies. This leads to paradoxes. The authors of [28], for example, construct convex non-spherical rigid bodies which, for certain initial conditions, must either interpenetrate

upon collision or violate the classical balance laws of rigid body mechanics. Cox, Feres, and Ward have developed a theory of rigid body collisions from a differential geometric point of view [10]. To avoid issues with singularities, these authors assume that the difference in the shape operators on the boundaries of the two bodies, expressed in a certain common frame, are non-singular. One can also consider weak solutions to the dynamical equations governing rigid body interactions. Ballard has developed an existence theory along these lines [3]. Wilkinson shows that typically such systems are underdetermined in the weak sense [41]. A rare case in which a well-known hard sphere model has been extended to the non-spherical setting is Saint-Raymond and Wilkinson's study of the Boltzmann equation [36]. The challenges described by these authors in their introduction exemplify the general difficulty of working outside of the hard sphere paradigm.

Rough collisions have the potential to provide a kind of mean between well-understood questions in the hard sphere setting, and their corresponding generalizations to rigid bodies of more arbitrary shape. In the rough collisions setting, one can choose the microscopic features to be quite complicated, even fractal-like, while keeping the macroscopic shape of each body relatively simple (e.g. a sphere). In the limit as the scale of the rough features goes to zero, the complicated singularities in the configuration space become invisible, but some information about the rough features is still preserved in the limiting rough collision law.

1.1.3 Model and main results: informal description

Our main results characterize collisions between a freely moving rough disk and a fixed rough wall. This characterization provides a way to explicitly construct the collision dynamics for various choices of microstructure on the disk and the wall. We give a mathematically rigorous description of our model in §1.3, stating our main results in §1.3.7. Here we limit ourselves to an informal description of our model and results.

Consider a disk D with unit radius, moving freely in two-dimensional space and colliding with a fixed wall W lying in the lower half-plane $\{(x_1, x_2) \in \mathbb{R}^2 : x_2 \leq 0\}$. The surfaces of the disk and the wall are covered in small asperities. We allow the asperities on the wall to be fairly arbitrary in shape, while requiring the asperities of the disk to be of a quite specific type, namely "geostationary satellites," as illustrated in Figure 1. The satellites should be spaced far enough apart that the event that multiple satellites interact with the wall during a single collision is rare. During a collision event, a single satellite may hit the wall multiple times however. The limiting (possibly random) collision dynamics, obtained as the scale of the roughness on D and W goes to zero, are governed by a *rough collision law*, which specifies the post-collision linear and angular velocities of the disk after it leaves the wall.

The somewhat unrealistic surface structure on D is necessary to avoid some of the difficulties encountered in rigid body mechanics, described above. If the satellites are too close together, then the boundary of the configuration space will be too singular to derive the kinds of estimates needed to prove our main results. The model is nonetheless "universal," in a sense to be described shortly.

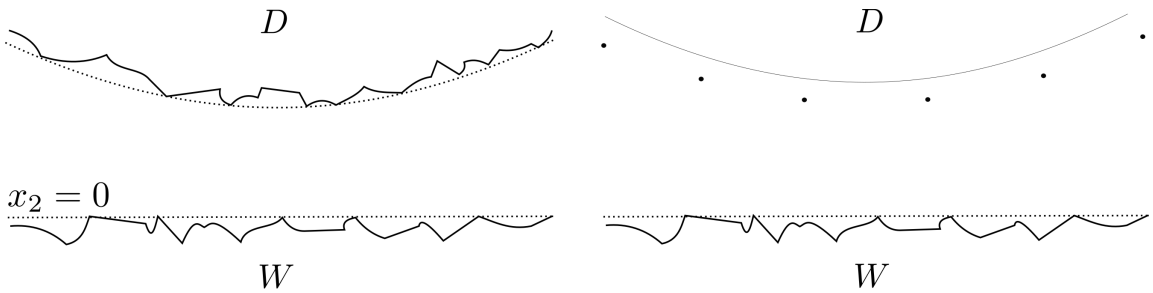

Figure 1: The collision dynamics are too complicated to describe when we put arbitrarily shaped asperities on both D and W (left). The problem becomes more tractable if we assume the asperities on D are well-spaced "geostationary satellites" (right).

We represent the state of the system by a sextuple $(x_1, x_2, \alpha, v_1, v_2, \omega)$, where (x_1, x_2) is the center of mass of D in $\mathbb{R}^2$ and α its angular orientation, and (v_1, v_2) is the linear velocity of the disk and ω its angular velocity. We assume in our analysis that the mass density of the disk is rotationally symmetric, and that the kinetic energy of the system is conserved.

There are three properties which we expect the disk and wall system with rough collision dynamics to satisfy.

(I) Liouville measure on the phase space $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}\alpha\mathrm{d}v_1\mathrm{d}v_2\mathrm{d}\omega$ should be preserved.

(II) The collision dynamics should be "time-reversible," in the sense that the evolution of the system will look the same from a statistical point of view, whether time is run forward or backward.

(III) The quantity

$$-mv_1 + J\omega, \tag{1.1}$$

where m is the mass of the disk and J is the moment of inertia of the disk

about its center of mass, should be conserved.

The basis for properties (I) and (II) comes from billiards theory. It is well-known that the dynamics of classical billiard systems preserve Liouville measure and are time-reversible. Consequently, the rough collision dynamics, obtained in the weak limit, should also preserve Liouville measure and be time-reversible.

The intuition behind (III) is that the quantity (1.1) is the projection (with respect to an inner product coming from kinetic energy) of the phase space velocity onto the "rolling velocity" $\chi = (-1, 0, 1)$. If the disk comes into contact with the wall with velocity $(v_1, v_2, \omega) \approx (-1, 0, 1)$, then the disk will "roll" along the wall. The relative velocity of the wall and the point of contact on the disk will be negligible. Consequently the impact will be negligible, and the disk will continue rolling indefinitely with approximately the same velocity as before. In other words, translation in the direction χ should be a "symmetry" of the system.

Modulo some technical assumptions, our main results (Theorems 1.27, 1.28, and 1.31) say that a rough collision law not only satisfies properties (I)-(III), but these properties characterize the class of rough collision laws for the disk and wall system described above. That is, any collision law which produces dynamics satisfying (I)-(III) may be approximated by a deterministic collision law obtained by equipping W with small, appropriately shaped asperities.

Note that the heuristic justification for properties (I)-(III) does not depend on the special surface structure imposed on D. Thus the range of dynamics manifested in our model is much broader than the setup suggests.

Contained in our results is a way to construct rough collision laws. A rough

collision law is described by a Markov kernel $\mathbb{K}(y_1, y_3, \theta, \psi; \mathrm{d}y_1' \mathrm{d}y_3' \mathrm{d}\theta' \mathrm{d}\psi')$, where (y_1, y_3) are a certain choice of coordinates on the $x_1\alpha$-plane in the configuration space, and (θ, ψ) are spherical coordinates on the velocity space. We will see that rough collision laws always take the form

$$\mathbb{K}(y_1, y_3, \theta, \psi; \mathrm{d}y_1' \mathrm{d}y_3' \mathrm{d}\theta' \mathrm{d}\psi') = \delta_{(y_1, y_3)}(\mathrm{d}y_1' \mathrm{d}y_3') \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') \delta_{\pi - \psi}(\mathrm{d}\psi'). \tag{1.2}$$

The single non-trivial factor $\widetilde{\mathbb{P}}$ describes the way a point particle reflects from a rough wall $\widetilde{W}$, obtained by foreshortening the original wall W in one direction by a factor of $(1 + m/J)^{1/2}$. In many cases, the Markov kernel $\widetilde{\mathbb{P}}$ can be computed explicitly.

The proof of our main results depends on a characterization of rough reflection laws, discovered independently by Plakhov and by Angel, Burdzy, and Sheffield (see §1.2.5. and §1.2.7). The main novelty in this work – as well as the main technical challenge – is to prove rigorously that the quantity (1.1) is conserved.

The results of this book relate to the work of R. Feres and collaborators on two separate fronts – first, in relation to *rough reflections* (see §1.2.6), and second, in relation to *no-slip collisions*, a type of idealized, deterministic frictional collision (see §2.3.4). Our results imply that no-slip collisions belong to the class of rough collisions; thus the dynamics of a freely moving disk and fixed wall undergoing no-slip collisions can be approximated by a pair of bodies undergoing classical non-frictional collisions.

1.1.4 Ergodicity

Having described the collision dynamics, we can next consider the long-term evolution of more complex billiard systems with rough collisions. For example, suppose that a rough unit disk D moves freely in the strip $\{(x_1, x_2) : -1 \leq x_2 \leq 1 + r\}$, and undergoes a rough collision when it hits either of the two walls. In this setting, locally, collisions with each wall conserve a quantity of type (1.1). But the conserved quantity associated with each wall is not the same, and thus no quantity of type (1.1) is conserved globally.

Under relatively weak conditions on the Markov kernel $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ appearing in the product decomposition (1.2), we will see that the discrete velocity evolution of the system tends to explore the entire constant-energy level set in the phase space of the system. The conditions under which this occurs are that (1) the angle $\gamma := \arccos\left(\frac{1}{1+J/m}\right)$ is not a rational multiple of π, and (2) the Markov kernel $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is ergodic and is not singular with respect to Lebesgue measure. Ergodicity of $\widetilde{\mathbb{P}}$ is also a necessary condition for this result to hold. The precise statement of these ergodicity results is given in Theorems 7.1 and 7.2.

1.1.5 Organization

Rough collisions belong to a subbranch of stochastic billiards which we refer to as *rough billiards*. An introduction to past work in this subject area may be found in §1.2. A more rigorous description of our model and main results is given in §1.3.

The purpose of §2 is to apply our main results to construct a number of examples of rough reflection laws and rough collision laws, for various choices of microstructure

on the wall W.

A collision between two rigid bodies may be represented by a point mass reflecting specularly from the boundary of the configuration space. This fact allows us to apply techniques from billiards theory to analyze our model. In §3 we derive from physical principles in rigid body mechanics the specular reflection law for our model.

§4 is devoted to preliminaries for the proof of our main results. First comes a careful description of the elementary properties of the configuration space of the disk and wall system. Subsequent sections provide a rigorous definition of the collision law associated with the system, and introduce two auxiliary collision laws which play a role in our proofs.

Our main results characterizing rough collision laws are proved in §5. For a high-level summary of our arguments, see also §1.3.8.

§6 is concerned with the "abstract theory" of rough billiards. Both the rough reflections described in §1.2 and the rough collision laws introduced in §1.3 are special cases of the rough reflections defined in §6. We will refer to results proved in this section a number of times throughout the book.

§7 investigates the ergodicity of a billiard system consisting of a rough disk bouncing between two parallel rough walls. The main result of this section characterizes when the Lambertian measure is an ergodic measure for the discrete velocity evolution of the system.

1.1.6 Acknowledgments

I would like to sincerely thank my advisor Krzysztof Burdzy, who has been an invaluable source of help and insight from start to finish. This work owes much to his patience and unabating encouragement. I would also like to thank my committee members Robin Graham, Soumik Pal, and Matt Lorig for their helpful comments on the draft.

In addition, I am grateful to my parents and siblings who nourished me in my pursuits from the beginning and who provided much appreciated moral support during my graduate studies. Especially, I wish to remember my father Michael Rudzis who passed two years ago but whose wisdom and kindness I will always hold dear in my heart.

1.2 Rough billiards

The following section serves two purposes: first, to introduce results concerning rough billiards in the upper half-plane, upon which the main results of this work depend; and second, to provide a general survey of past work in rough billiards. Consequently, we are careful about giving technically accurate statements in §§1.2.2-1.2.5, whereas the style of §§1.2.6-1.2.7 is a bit more informal.

1.2.1 Notation and other conventions

We will use the following notation throughout the book. If X is a topological space and $Y \subset X$, then $\operatorname{Int} Y$ and $\overline{Y}$ denote, respectively, the topological interior and closure

of Y relative to X. The notation $\partial Y := \overline{Y} \cap \overline{Y^c}$ always refers to the topological boundary of Y relative to X, and should not be confused with the boundary of a manifold. Context will be sufficient to distinguish the ambient space X in most cases (usually $X = \mathbb{R}^d$ for some $d \geq 1$).

We let $C_c(X)$ denote the space of compactly supported functions on X. If X has the structure of a differentiable manifold and $k \geq 0$, then $C^k(X)$ denotes the space of k-times continuously differentiable functions on X, and $C^\infty(X)$ denotes the space of infinitely differentiable functions on X. We let $C_c^k(X) = C_c(X) \cap C^k(X)$ and $C_c^\infty(X) = C_c(X) \cap C^\infty(X)$.

If (X, μ) is a measure space and $1 \leq p \leq \infty$, then $L^p_\mu(X)$ denotes the space of p-integrable functions on (X, μ). We denote the L^p-norm on this space by $\|\cdot\|_{L^p_\mu(X)}$. We suppress X and μ from our notation if they are clear from the context.

If $f, g : \mathbb{R} \to \mathbb{R}_+$ are real functions, we write $f = O(g)$ if $\limsup_{t\to 0} f(t)/g(t) < \infty$, and we write $f = o(g)$ if $\lim_{t\to 0} f(t)/g(t) = 0$.

If B is a subset of $\mathbb{R}^d$, and v is a vector in $\mathbb{R}^d$, then we denote the translate of B by v as follows: $B + v := \{p \in \mathbb{R}^d : p - v \in B\}$.

Here and throughout the book, the term *billiard* refers to any dynamical system in which a point particle moves linearly in the complement of a closed subset $W \subset \mathbb{R}^d$ and reflects from the boundary of W (specularly or according to some other rule). The subset W is called the *wall*, and is usually assumed to have a piecewise smooth boundary (where the meaning of "piecewise smooth" is made precise in more specific contexts). The complement of the interior of W is referred to as either the *billiard table* or *billiard domain*. The piecewise linear curve traced out by the point

particle for some choice of initial conditions is called the *billiard trajectory*. For more background on mathematical billiards, we refer the reader to [6] and [39].

1.2.2 Rough billiards in the upper half-plane

We begin by describing the construction of rough billiards in the upper half-plane. Our approach is essentially the same as that of [1], and similar to that of [13].

The billiard table we initially consider is the complement of a closed set $W \subset \mathbb{R}^2$ satisfying the following assumptions:

A1. W is the closure of its interior in $\mathbb{R}^2$.

A2. $\mathbb{R}^2 \setminus W$ is path-connected.

A3. The following inclusions hold: $\mathbb{R} \times (-\infty, -1] \subset W \subset \mathbb{R} \times (-\infty, 0]$; thus $\partial W \subset \mathbb{R} \times [-1, 0]$.

A4. $\partial W = \bigcup_{i=1}^{\infty} \Gamma_i$, where $\{\Gamma_i\}_{i \geq 1}$ is some collection of compact curve segments satisfying the following conditions:

(i) The collection $\{\Gamma_i\}_{i \geq 1}$ is locally finite, in the sense that any bounded set intersects only finitely many of the curve segments Γ_i;

(ii) each Γ_i is the image of an injective C^2 map $\gamma_i : [0, 1] \to \mathbb{R}^2$ with nonvanishing left and right-hand derivatives (where C^2 means that γ_i has a C^2 extension to an open interval containing $[0, 1]$);

(iii) the curves Γ_i are allowed to intersect each other only at their endpoints; and

(iv) for each i, if Γ_i intersects the line $\{(x_1, x_2) : x_2 = 0\}$ at a point other than one of its two endpoints, then $\Gamma_i \subset \{(x_1, x_2) : x_2 = 0\}$.

In condition A4, the decomposition of ∂W into curve segments Γ_i is not unique. We shall refer more generally to a curve Γ with a decomposition $\Gamma = \bigcup_{i \geq 1} \Gamma_i$ such that conditions (i)-(iii) are satisfied as a *piecewise C^2 curve*. Condition (iv) lets us avoid certain pathological situations when defining the macro-reflection law below (we would like to avoid the situation where some Γ_i intersects the line $x_2 = 0$ in a "fat Cantor set" for example). In most typical situations, it will be easy to choose a decomposition of ∂W such that (iv) holds.

Let $\mathbb{R}^2_\pm = \{(x_1, x_2) \in \mathbb{R}^2 : \pm x_2 > 0\}$. Consider a point particle moving freely in W^c and reflecting *specularly* (angle of incidence equals angle of reflection) from ∂W. When the point particle leaves the upper half-plane $\mathbb{R}^2_+$, the particle may hit ∂W multiple times before returning to the upper half-plane, as illustrated in Figure 2. The limiting behavior of this interaction as $\epsilon \to 0$ will be described by a rough reflection law.

The *kinetic energy* of a point particle with velocity $v = (v_1, v_2)$ is the quantity $\frac{1}{2}v_1^2 + \frac{1}{2}v_2^2$. We assume that kinetic energy is conserved for all time. Without loss generality we take the velocity of the point particle to be restricted to the Euclidean unit circle $\mathbb{S}^1 \subset \mathbb{R}^2$ for all time. We identify points in $\mathbb{S}^1$ with angles θ in the interval $[0, 2\pi)$, and we let $\mathbb{S}^1_+ = \mathbb{S}^1 \cap \mathbb{R}^2_+ = (0, \pi)$ and $\mathbb{S}^1_- = \mathbb{S}^1 \cap \mathbb{R}^2_- = (\pi, 2\pi)$.

The *macro-reflection law associated with W* is the map $P^W : \mathbb{R} \times \mathbb{S}^1_+ \to \mathbb{R} \times \mathbb{S}^1_+$ defined as follows. As shown in Figure 2, if initially the point particle starts on the x_1-axis with velocity pointing into the lower half-plane $\mathbb{R}^2_-$, its state may be

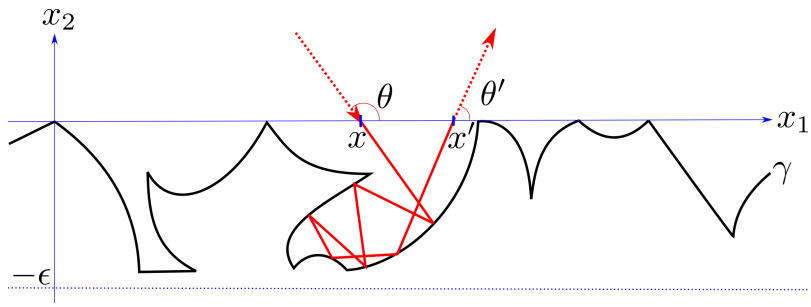

Figure 2: The macro-reflection law maps the initial state (x, θ) to the state (x', θ').

represented by a pair $(x, \theta) \in \mathbb{R} \times \mathbb{S}^1_+$, where x is first coordinate of the particle on the x_1-axis, and $\pi + \theta \in \mathbb{S}^1_-$ is its velocity. After reflecting from the boundary ∂W a certain number of times, the particle returns to the x_1-axis at a position $x' \in \mathbb{R}$ and with velocity $\theta' \in \mathbb{S}^1_+$. We define

$$P^W(x, \theta) = (x', \theta'). \tag{1.3}$$

The term "macro-reflection law" should be understood in contradistinction to the specular reflection law which describes the "micro" reflection of the trajectory from ∂W at a single instant in time.

The map P^W may fail to be defined at pairs (x, θ) such that the billiard trajectory hits the boundary ∂W tangentially or at a "corner" or never returns to the x_1-axis. Thus we impose the following additional assumption.

A5. For almost every $(x, \theta) \in \mathbb{R} \times \mathbb{S}^1_+$, the billiard trajectory starting from (x, θ) is well-defined for all time, and returns to the x_1-axis after only finitely many collisions with W.

This condition is implied by conditions A1-A4 together with either one of the fol-

lowing conditions:

A5a. There exists a countable collection $\{R_j\}_{j\geq 1}$ of disjoint bounded open subsets of $\mathbb{R}^2$ such that $\mathbb{R}^2_- \setminus W = \bigcup_{j=1}^{\infty} R_j$.

A5b. The wall W is ϵ-periodic in the x_1-coordinate, in the sense that

$$W = \{(x_1, x_2) : (x_1 - \epsilon, x_2) \in W\} = W + \epsilon e_1, \qquad \text{where } e_1 = (1, 0). \tag{1.4}$$

The first condition means that the point particle will always become trapped in some bounded region when it interacts with the wall. Both conditions allow us to apply the Poincaré Recurrence Theorem to obtain A5. For more details, we refer the reader to §6, where we define macro-reflection laws in a more general setting. See specifically the discussion of upper half-space billiards in §6.2.4.

Define a measure Λ^1 on $\mathbb{R} \times \mathbb{S}^1_+$ by

$$\Lambda^1(\mathrm{d}x\mathrm{d}\theta) = \sin\theta \mathrm{d}x\mathrm{d}\theta. \tag{1.5}$$

The most important elementary properties of P^W are summed up in the following proposition.

Proposition 1.1. *(i) The map P^W is involutive in the sense that $P^W \circ P^W(x, \theta) = (x, \theta)$ whenever the left-hand side is defined.*

(ii) The map P^W preserves the measure Λ^1 in the sense that, for any measurable set $A \subset \mathbb{R} \times \mathbb{S}^1_+$,

$$\Lambda^1((P^W)^{-1}(A)) = \Lambda^1(A). \tag{1.6}$$

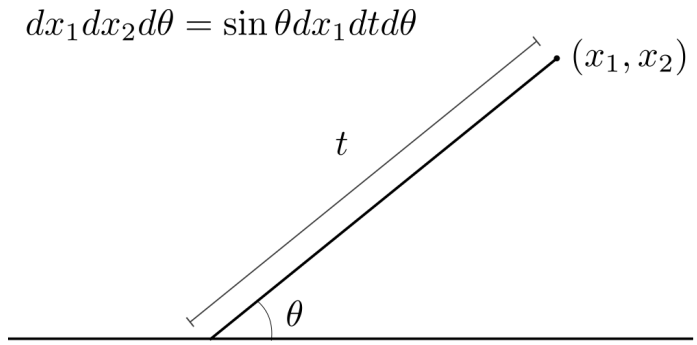

Figure 3: If (x_1, x_2, θ) is the state of the point particle, then $x_2 = t\sin\theta$, where t is the time to hit the x_1-axis. Liouville measure in these new coordinates is $\sin\theta \mathrm{d}x_1\mathrm{d}t\mathrm{d}\theta$.

To understand (i), note that specular reflection is involutive; so "running the evolution backward" from $(x', \theta') := P^W(x, \theta)$, the trajectory is guaranteed to return to the x_1-axis in state (x, θ). Part (ii) is a corollary of a well-known theorem in billiards theory (see Lemma 6.2). If we accept that the continuous billiard evolution should preserve Liouville measure $\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}\theta$ on the phase space, then Figure 3 should make property (ii) quite believable. A more general version of Proposition 1.1 is proved in §6 (see Proposition 6.5).

The macro-reflection law P^W is naturally associated with a deterministic Markov kernel on $\mathbb{R} \times \mathbb{S}^1_+$, defined by

$$\mathbb{P}^W(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \delta_{P^W(x,\theta)}(\mathrm{d}x'\mathrm{d}\theta'). \tag{1.7}$$

In what follows, by "wall" we mean a subset $W \subset \mathbb{R}^2$ satisfying conditions A1-A5.

Definition 1.2. We call a Markov kernel $\mathbb{P}$ on $\mathbb{R} \times \mathbb{S}^1_+$ a *rough reflection law* in the upper half-plane $\mathbb{R}^2_+$ if there exists a sequence of positive numbers $\epsilon_n \to 0$ and a

sequence of walls W_n such that $\partial W_n \subset \{(x_1, x_2) : -\epsilon_n \le x_2 \le 0\}$ and

$$\mathbb{P}^{W_n}(x, \theta; \mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \to \mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \tag{1.8}$$

weakly in the space of measures on $\mathbb{R} \times \mathbb{S}^2_+$.

The two properties of macro-reflection laws P^W described in Proposition 1.1 carry over to rough reflection laws in the following sense.

Proposition 1.3. *Let $\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta')$ be a rough reflection law. The Markov kernel $\mathbb{P}$ is symmetric with respect to the measure Λ^1, in the sense that, for any $f \in C_c((\mathbb{R} \times \mathbb{S}^1_+)^2)$,*

$$\begin{aligned} &\int_{(\mathbb{R}\times\mathbb{S}^1_+)^2} f(x, \theta, x', \theta')\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \\ &\quad = \int_{(\mathbb{R}\times\mathbb{S}^1_+)^2} f(x', \theta', x, \theta)\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta). \end{aligned} \tag{1.9}$$

Symmetry generalizes time-reversibility in the sense that the left-hand side of (1.9) is transformed into the right-hand side by interchanging the pre- and post-reflection variables (x, θ) and (x', θ'). Symmetry also implies that $\mathbb{P}$ preserves Λ^1. Indeed, by letting $f(x, \theta, x', \theta') \uparrow g(x', \theta') \in C_c(\mathbb{R} \times \mathbb{S}^2_+)$ in (1.9), we obtain

$$\int_{(\mathbb{R}\times\mathbb{S}^1_+)^2} g(x', \theta')\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) = \int_{\mathbb{R}\times\mathbb{S}^1_+} g(x, \theta)\Lambda^1(\mathrm{d}x\mathrm{d}\theta). \tag{1.10}$$

Proposition 1.3 is a special case of Proposition 6.8, proved in §6.

Remark 1.4. In Definition 1.2, it is equivalent to replace Λ^1 with any measure which is mutually absolutely continuous with respect to Lebesgue measure on $\mathbb{R} \times \mathbb{S}^1_+$. The

measure Λ^1 happens to be a convenient choice, due to Proposition 1.3.

Remark 1.5. The convergence (1.8) means that, for any $h(x,\theta,x',\theta') \in C_c((\mathbb{R} \times \mathbb{S}^1_+)^2)$,

$$\begin{aligned} \lim_{n\to\infty} \int_{(\mathbb{R}\times\mathbb{S}^1_+)^2} h(x,\theta,x',\theta')\mathbb{P}^{W_n}(y,w;\mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}x\mathrm{d}\theta) \\ = \int_{(\mathbb{R}\times\mathbb{S}^1_+)^2} h(x,\theta,x',\theta')\mathbb{P}(y,w;\mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}x\mathrm{d}\theta). \end{aligned} \tag{1.11}$$

Since the tensor product $C_c^\infty(\mathbb{R} \times \mathbb{S}^1_+) \otimes C_c^\infty(\mathbb{R} \times \mathbb{S}^1_+)$ is dense in $C_c((\mathbb{R} \times \mathbb{S}^1_+)^2)$, it is sufficient to verify (1.11) for functions for form $h(y,w,y',w') = f(y,w)g(y',w')$, where $f,g \in C_c^\infty(\mathbb{R} \times \mathbb{S}^1_+)$.

Remark 1.6. There is a sense in which $\mathbb{P}^{W_n}$ converges to $\mathbb{P}$ as a limit with respect to a pseudometric topology on the space of Markov kernels on $\mathbb{R} \times \mathbb{S}^1_+$. This topology is described in §6.2.3.

Some care must be taken when working with this sense of convergence, because limits may not be unique. With respect to the pseudometric, the distance between two Markov kernels $\mathbb{P}$ and $\mathbb{P}'$ is zero if and only if $\mathbb{P}(\cdot;\mathrm{d}x\mathrm{d}\theta)$ and $\mathbb{P}'(\cdot;\mathrm{d}x\mathrm{d}\theta)$ agree on a Λ^1-full measure subset of $\mathbb{R} \times \mathbb{S}^1_+$. If we identify Markov kernels which agree on a Λ^1-full measure subset of $\mathbb{R} \times \mathbb{S}^1_+$, then limits will be unique and the pseudometric will be a metric.

From this point on, we will write $\lim_{i\to\infty} \mathbb{P}^{W_n} = \mathbb{P}$ to indicate that (1.8) holds.

1.2.3 Simple example: the rectangular teeth microstructure

A simple example of a rough reflection law may be obtained by considering a sequence of walls W_n with periodic boundary structure consisting of "rectangular teeth." That is, we first define real functions

$$t_n(x) = \begin{cases} 0 & \text{if } 2k\epsilon_n \le x \le (2k+1)\epsilon_n, \\ -r\epsilon_n & \text{if } (2k+1)\epsilon_n < x < (2k+2)\epsilon_n \end{cases} \qquad \text{for } k \in \mathbb{Z}. \tag{1.12}$$

See Figure 6 in §2. The quantity $r > 0$ is a fixed parameter representing the ratio of the height of the teeth to the width. We define

$$W_n = \{(x_1, x_2) : x_2 \le t_n(x_1)\}. \tag{1.13}$$

If $v^- = (v_1^-, v_2^-)$ is the incoming velocity of a point particle, then after hitting W_n a certain number of times, the particle will return to the upper half-plane with velocity either $v^+ = (v_1^-, -v_2^-)$ or $v^+ = (-v_1^-, -v_2^-) = -v^-$. The first of these velocities corresponds to a specular reflection, while the second corresponds to a *retroreflection* – i.e. a reflection in which the outgoing trajectory of the point particle goes in the opposite direction as the incoming trajectory. Thus, as $\epsilon_n \to 0$, we expect the limiting rough reflection law to randomly select between specular reflection and retroreflection.

In §2, we derive explicit formulas for rough reflections from several different types of microstructures, including the rectangular teeth microstructure described above.

1.2.4 Periodic case

We now comment on the special case where the wall W satisfies the periodicity condition A5b. In this setting, it is useful to abstract the *shape* of the wall from the *scale*. In the limit, as the scale of the wall goes to zero, we expect at least some information about the shape of the wall to be preserved, and we would like to be able to talk about the shape of the wall independently of the scale.

To accomplish this, we observe that a periodic wall is determined uniquely by a pair (Σ, ϵ), where Σ is a subset of $\mathbb{S}^1 \times (-\infty, 0]$ with $\partial\Sigma \subset \mathbb{S}^1 \times [-1, 0]$, and $\epsilon > 0$. In particular, W is the unique wall satisfying periodicity condition (1.4) such that the image of W under the covering map

$$(x_1, x_2) \mapsto (e^{2\pi i x_1/\epsilon}, \epsilon^{-1} x_2) : \mathbb{R} \times [-\epsilon, 0] \to \mathbb{S}^1 \times [-1, 0] \tag{1.14}$$

is Σ. We denote the wall so determined by $W(\Sigma, \epsilon)$. Drawing on terminology from Feres [13], we refer to Σ as the *cell* and we refer to $\epsilon > 0$ as the *roughness scale*.

If the wall $W = W(\Sigma, \epsilon)$ arises from a cell and roughness scale, as described above, then we will denote the corresponding macro-reflection law by $\mathbb{P}^{\Sigma,\epsilon}$.

To illustrate the use of this concept, consider rough reflections from a "fractal microstructure." In general, fractals do not have well-defined normal vectors at most boundary points, so we cannot define specular reflection on such a surface directly. But sense can be made of this in the case of rough reflections. For example, we might take Σ_n to be a sequence of sets generating a fractal whose boundary is a Koch curve – see Figure 4. A rough reflection "from a Koch curve microstructure" can then be

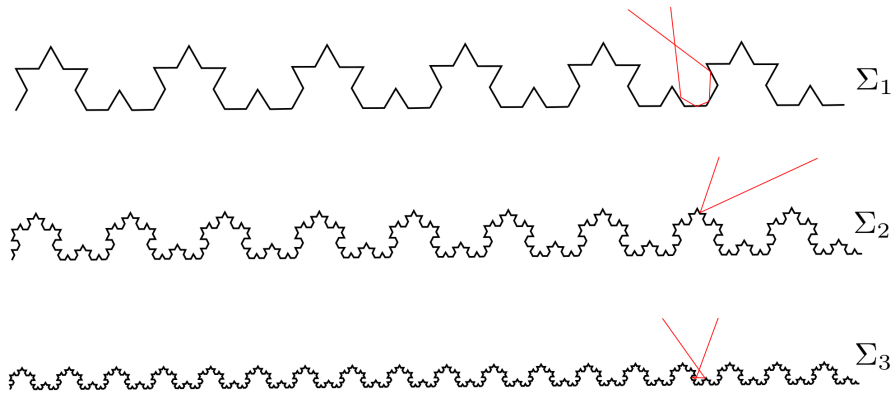

Figure 4: Reflections from a Koch curve.

defined as a rough reflection obtained as the limit (in the sense of (1.8)) of a sequence of deterministic Markov kernels $\mathbb{P}^{\Sigma_n,\epsilon_n}$, where ϵ_n is a sequence of positive numbers converging to zero. The paper [22] carries out numerical experiments for a related model.

Conditions on Σ which are sufficient to guarantee that the wall $W = W(\Sigma, \epsilon)$ satisfies conditions A1-A5 are the following:

B1. Σ is the closure of its interior in $\mathbb{S}^1 \times \mathbb{R}$.

B2. $\mathbb{S}^1 \times \mathbb{R} \setminus \Sigma$ is connected.

B3. The following inclusions hold: $\mathbb{S}^1 \times (-\infty, -1] \subset \Sigma \subset \mathbb{S}^1 \times (-\infty, 0]$; thus $\partial\Sigma \subset \mathbb{S}^1 \times [-1, 0]$.

B4. There exists a finite collection of compact C^2 curve segments $\{\widetilde{\Gamma}_i\}_{i=1}^m$ such that $\partial\Sigma = \bigcup_{i=1}^m \widetilde{\Gamma}_i$. The curve segments $\widetilde{\Gamma}_i$ satisfy conditions A4(i)-(iv), with the obvious modifications.

We of course get the periodicity condition A5b for free.

One condition which is sufficient to guarantee that A5a holds is:

B5. There exists a non-trivial loop $\gamma \subset \mathbb{S}^1 \times \mathbb{R}$, starting and ending at the point $(1, 0)$, which lies entirely in Σ.

Here "non-trivial" means that γ cannot be contracted in $\mathbb{S}^1 \times \mathbb{R}$ to a point. This condition implies the points $(\epsilon k, 0) \in \mathbb{R}^2$, where $k \in \mathbb{Z}$, all lie in a single connected component of $W = W(\Sigma, \epsilon)$. Condition B5 will always be satisfied if Σ is connected, satisfies conditions B1-B4, and contains the point $(1, 0) \in \mathbb{S}^1 \times \mathbb{R}$.

The main reason we would want to impose the condition B5 is the following. Under generic circumstances, we expect a rough collision law obtained from a sequence of periodic walls to take the form

$$\mathbb{P}(x, \theta; \mathrm{d}x' \mathrm{d}\theta') = \delta_x(\mathrm{d}x') \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta'). \tag{1.15}$$

The intuition behind this is that the point particle should leave a rough wall at approximately the same spatial position that it hits. Moreover, if the microstructure on the wall is periodic, then the distribution of the angle of reflection θ' should only depend on the angle of incidence θ, and not on the position where the particle hits the wall. Lemma 2.1 from §2 implies that if $\mathbb{P} = \lim_{n \to \infty} \mathbb{P}^{\Sigma_n, \epsilon_n}$, where the cells Σ_n satisfy B1-B5, then $\mathbb{P}$ takes the form (1.15). The assumption B5 guarantees that the billiard trajectory will get trapped in small "hollow" within a single period of the wall W, and consequently the distance between the points where the trajectory first hits and returns to the x_1-axis will be of order ϵ apart.

Although we do not know of a specific counter-example, it is most likely not possible to obtain (1.15) if we just assume conditions B1-B4. One can imagine constructing a periodic wall with a large "asteroid field" of connected components, such that the point particle will be forced to travel a great distance, reflecting from the various components many times, before eventually leaving the wall. If each wall W_n is constructed in this way, the limiting reflection law might not satisfy (1.15).

Remark 1.7. We can, however, weaken condition B5 as follows, and (1.15) will still hold:

B5'. There exists $a \in [-1, 0]$ and $u \in \mathbb{S}^1$ such that

$$\mathbb{S}^1 \times (-\infty, -1] \subset \Sigma \subset \mathbb{S}^1 \times (-\infty, a], \tag{1.16}$$

and there exists a nontrivial loop $\gamma \subset \mathbb{S}^1 \times \mathbb{R}$ starting and ending at the point (u, a), which lies entirely in Σ.

Under this assumption, the proof of Lemma 2.1 in §2 goes through with only minor modifications.

Remark 1.8. The factor $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is a Markov kernel on $\mathbb{S}^1_+$. When $\mathbb{P}$ takes the form (1.15), the symmetry property (1.9) reduces to the following: for any $f \in C_c(\mathbb{S}^1_+ \times \mathbb{S}^1_+)$,

$$\int_{\mathbb{S}^1_+ \times \mathbb{S}^1_+} f(\theta, \theta') \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') \sin\theta \mathrm{d}\theta = \int_{\mathbb{S}^1_+ \times \mathbb{S}^1_+} f(\theta', \theta) \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') \sin\theta \mathrm{d}\theta. \tag{1.17}$$

For more information on the situation when W is periodic, see §2.1.

1.2.5 Characterization of rough reflections laws in upper half-plane

Rough reflections were originally characterized by Plakhov in the context of optimization problems in aerodynamics. This author's setting is quite general and at least superficially different from the one above, considering the *scattering law* on a bounded convex body in $\mathbb{R}^d$, instead of the rough reflection law in the upper half-plane. Angel, Burdzy, and Sheffield independently obtained a characterization of the rough reflection laws as we have defined them above. We state this characterization first, since it is the one most closely related to the main results obtained in this work. In §1.2.7 we discuss Plakhov's ideas and how they are related to the result which we state presently.

Theorem 1.9 (1, Theorem 2.3)**.** *Suppose* $\mathbb{P}(x,\theta;\mathrm{d}x'\mathrm{d}\theta') = \delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(x,\theta;\mathrm{d}\theta')$ *and is a symmetric with respect to the measure* Λ^1 *in the sense of (1.9). Then there exists a sequence of walls* W_n *with piecewise analytic boundaries* $\partial W_n \subset \{(x_1,x_2) : -1/n < x_2 \leq 0\}$ *such that*

$$\mathbb{P}^{W_n}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \to \mathbb{P}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta), \tag{1.18}$$

weakly on the space of measures on $\mathbb{R}\times\mathbb{S}^1_+$.

Remark 1.10. The actual walls W_n constructed in [1, Theorem 2.3] have boundaries ∂W_n which satisfy the following conditions: (1) the boundary is composed of a locally finite collection of compact, analytic curve segments (where analytic on a compact interval means having an analytic extension to an open interval); (2) the curve segments intersect only at their endpoints and do not form cusps at their

intersection points (i.e. the angle between two intersecting curve segments at their endpoints is not zero); (3) each curve segment either has non-vanishing curvature of one sign or is a line segment; and (4) the condition A5a (see above) is satisfied. Thus the walls satisfy the same hypotheses as those of [6, Chapter 2] for example.

To appreciate the significance of Theorem 1.9, consider the following two Markov kernels which are easily shown to be symmetric with respect to the measure Λ^1.

- retroreflection: $\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \delta_{(x,\theta)}(\mathrm{d}x'\mathrm{d}\theta')$.
- Lambertian reflection: $\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \frac{1}{2}\delta_x(\mathrm{d}x')\sin\theta'\mathrm{d}\theta'$.

The first of these reflection laws is deterministic. Examples of approximate retroreflectors in real life include cat's eyes and street signs with reflective paint [43]. The second reflection law is random. It was introduced by Lambert in 1760 to model light reflecting from a matte surface [19]. There are many other examples of Markov kernels which preserve the measure Λ^1 and have a trivial spatial factor $\delta_x(\mathrm{d}x')$ (the collection of deterministic reflection laws alone is isomorphic to the space of measure preserving transformations of $(0, 1)$ with Lebesgue measure – see Remark 2.2). Theorem 1.9 says that each of these is a rough reflection law; that is, each may be approximated by a deterministic reflection from a surface with a geometric microstructure.

In the case where $\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$, the sequence of approximating reflectors may be taken to be periodic.

Corollary 1.11. *Suppose* $\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$*, where* $\widetilde{\mathbb{P}}$ *is a Markov kernel which is symmetric with respect to the measure* $\sin\theta\mathrm{d}\theta$ *in the sense of (1.17).*

Then there exists a sequence of cells $\{\Sigma_n\}_{n\geq 1}$ (satisfying conditions B1-B5) and positive numbers $\epsilon_n \to 0$ such that

$$\mathbb{P}^{\Sigma_n,\epsilon_n}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \to \mathbb{P}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta), \tag{1.19}$$

weakly on the space of measures on $\mathbb{R}\times\mathbb{S}^1_+$.

Remark 1.12. We call this a "corollary" because it follows from the proof of Theorem 2.3 in [1]. In this proof, the sequence of approximating reflectors W_n is obtained by locally constructing reflectors beneath the intervals $[k/n,(k+1)/n]$ in the x_1-axis for $k\in\mathbb{Z}$, and then piecing the reflectors together. When $\widetilde{\mathbb{P}}$ does not depend on x, these local reflectors can be taken to be of the same type on each interval, and consequently W_n is periodic.

Remark 1.13. It will follow from Lemma 2.1 that the limit of $\mathbb{P}^{\Sigma_n,\epsilon_n}$ does not depend on the choice of positive numbers $\epsilon_n \to 0$, but only on the sequence of cells $\{\Sigma_n\}$.

1.2.6 Operators on a Hilbert space

What information about the microgeometry of the rough surface can be recovered from the rough reflection law $\mathbb{P}$? In a series of papers, Feres and collaborators have sought to address this and related questions.

These authors are motivated in part by a problem in gas kinematics. Suppose that an inert gas at low pressure is released from a long but finite cylindrical chamber with rough interior walls. How is the microgeometry of the interior walls related to the time of escape of the gas particles? Questions of this nature go back to

Knudsen's studies of gas kinematics in 1907 (see [21], [12]). Knudsen assumed, based on physical heuristics, that the angle of reflection of a gas particle from the interior wall is independent of the angle of incidence, and hence the distribution of the angle of reflection is $\frac{1}{2}\sin\theta \mathrm{d}\theta$ (the same distribution introduced by Lambert in optical studies, as noted above). This is of course directly related to the fact that $\widetilde{\mathbb{P}}$ preserves the measure $\sin\theta \mathrm{d}\theta$.

Consider the case of a periodic wall where the sequence of cells $\Sigma_n = \Sigma$ is constant. Recall that in this case, rough reflection laws take the form (1.15). It is proved in [13] that $\widetilde{\mathbb{P}}$ is a bounded self-adjoint operator on the Hilbert space $L^2(\mathbb{S}^1_+, \sin\theta \mathrm{d}\theta)$. Self-adjointness is a direct consequence of the time-reversibility property mentioned above. Moreover, under additional assumptions (more or less, the sides of Σ should be dispersing or Sinai), $\widetilde{\mathbb{P}}$ becomes a Hilbert-Schmidt operator.

The time for a gas particle to exit an open-ended cylindrical chamber of length L and radius r is related in [13] to the Markov kernel $\widetilde{\mathbb{P}}$ as follows. If the gas particle repeatedly hits the interior boundary of the cylinder with angle of incidence θ_n and angle of reflection θ'_n, then we must have

$$\theta'_n \sim \widetilde{\mathbb{P}}(\theta_n, \mathrm{d}\theta') \qquad \text{and} \qquad \theta_{n+1} = \theta'_n, \qquad n = 0, 1, 2, \ldots. \tag{1.20}$$

The sequence of random angles θ'_n together with the constant speed v of the gas particle determine a continuous, piecewise linear process X_t on the real line which is the projection of the position of the gas particle at time t onto the cylindrical axis.

The exit time for the gas particle from the cylinder is then

$$\tau_L = \inf\{t \geq 0 : X_t > L\}. \tag{1.21}$$

By a central limit theorem argument, under restrictive assumptions on the stationary distribution of the angle process θ'_n, the scaling limit of the process $\{\xi X_{t/\xi}, t \geq 0\}$ as $\xi \to 0$ is Brownian motion with variance depending on the spectrum of $\widetilde{\mathbb{P}}$. The assumptions in [13] under which this is proved exclude the case where the process θ'_n is ergodic (i.e. the unique stationary distribution is $\sin\theta \mathrm{d}\theta$). In the ergodic case, the linear increments of the process X_t have infinite variance, and to obtain a Brownian motion scaling limit, one must instead use the scaling $\{\xi X_{t|\log\xi|/\xi}, t \geq 0\}$, $\xi \to 0$. The analysis of the latter case is carried out in [13] for one example.

With this motivation, the spectrum of $\widetilde{\mathbb{P}}$ is further analyzed in [14] and [15]. The first of these papers examines $\widetilde{\mathbb{P}}$ for a special class of cells Σ whose sides consist of dispersive circular arcs. By analyzing the moments of $\widetilde{\mathbb{P}}$, a relationship to spherical harmonics emerges. Namely, for smooth functions u on $\mathbb{S}^2$ which are rotationally invariant about the vertical axis,

$$\widetilde{\mathbb{P}}u - u = \frac{k_2}{6}\Delta_{\mathbb{S}^2} u + O(K^3), \tag{1.22}$$

where K is a scale invariant curvature parameter depending on Σ, and $\Delta_{\mathbb{S}^2}$ is the spherical Laplacian on $\mathbb{S}^2$. (The meaning of $\widetilde{\mathbb{P}}u$ is as follows: If we give $\mathbb{S}^2$ spherical coordinates (ϕ, ψ) where $\psi \in \mathbb{S}^1_+$ is the angle from vertical axis, then $u(\phi, \psi) = \widetilde{u}(\psi)$ for some $\widetilde{u}$ by rotational invariance, and $\widetilde{\mathbb{P}}u := \widetilde{\mathbb{P}}\widetilde{u}$.)

The paper [15] examines the spectrum of $\widetilde{\mathbb{P}}$ for more general types of cells, and the authors obtain explicit bounds on the spectral gap of $\widetilde{\mathbb{P}}$. These bounds depend on the curvature of the "most exposed" parts of the boundary of the cell Σ. These results are applied to estimate the rate of convergence of the Markov process determined by $\widetilde{\mathbb{P}}$ to stationary distribution.

The paper [7] considers similar questions but in an even more general setting, where the wall W is allowed to have moving parts and energy can be exchanged between the wall and the point particle.

1.2.7 Rough scattering laws on bounded convex bodies

In the context of optimization problems in aerodynamics, Plakhov has considered rough reflections on general bounded convex bodies. Here the main object of interest is the scattering law, which describes the equilibrium distribution of the incoming and outgoing particle velocities and the normal direction at the point of contact. The scattering law does not contain precisely the same information as the rough reflection law, but the important ideas are similar. The concepts and results summarized here originally appeared in [31, 32, 33, 34] and were later assembled in a book [35, Chapter 4].

Given a bounded convex body $C \subset \mathbb{R}^d$ and a body $B \subset C$ with piecewise smooth boundary, define subspaces

$$(\partial C \times \mathbb{S}^{d-1})_{\pm} = \{(x, u) \in \partial C \times \mathbb{S}^{d-1} : \pm u \cdot n(x) > 0\}, \tag{1.23}$$

where $n(x)$ is the outward-pointing unit normal vector at $x \in \partial C$, and $\cdot$ is the Eu-

clidean dot product. Define a measure Λ_C on $(\partial C \times \mathbb{S}^{d-1})_+$ by

$$\Lambda_C(\mathrm{d}x\mathrm{d}u) = (u \cdot n(x))_+ \mathrm{d}x\mathrm{d}u, \tag{1.24}$$

where $\mathrm{d}x$ and $\mathrm{d}u$ are Lebesgue measure on ∂C and $\mathbb{S}^{d-1}$ respectively. The macro-reflection law $P^{B,C}$ is defined in the same way as before. That is, if $(x, -u) \in (\partial C \times \mathbb{S}^{d-1})_+$ is the initial state of a point particle, then the particle will enter the C and hit the body B some number (possibly zero) times before returning to the boundary ∂C in state $(x', u') = (x'(x, u), u'(x, u)) \in (\partial C \times \mathbb{S}^{d-1})_+$. The *macro-reflection law determined by* B is the map $P^{B,C} : (\partial C \times \mathbb{S}^{d-1})_+ \to (\partial C \times \mathbb{S}^{d-1})_+$ such that

$$P^{B,C}(x, u) = (x', u'). \tag{1.25}$$

The map $P^{B,C}$ may not be defined on a measure zero subset of $(\partial C \times \mathbb{S}^{d-1})_+$. Like in the case of the upper half-plane, $P^{B,C}$ is an involution and preserves the measure Λ_C.

Let $\nu_{B,C}$ be the measure on $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} = (\mathbb{S}^{d-1})^3$ giving the equilibrium joint distribution of the incoming velocity $-u$, the outgoing velocity u', and the unit normal vector $n(x)$ at the point of contact $x \in \partial C$. That is, $\nu_{B,C}$ is defined by

$$\int_{(\mathbb{S}^{d-1})^3} f(u, u', n)\nu_{B,C}(\mathrm{d}u\mathrm{d}u'\mathrm{d}n) = \int_{(\partial C \times \mathbb{S}^{d-1})_+} f(u, u'(x, u), n(x))\Lambda_C(\mathrm{d}x\mathrm{d}u). \tag{1.26}$$

A measure ν on $(\mathbb{S}^{d-1})^3$ is called a *rough scattering law* on C if there exists a sequence of bodies $B_n \subset C$ such that

(i) $\mathrm{Vol}(C \setminus B_n) \to 0$ as $n \to \infty$, and

(ii) the sequence of measures $\nu_{B_n,C}$ converges weakly to the measure ν.

Remark 1.14. This terminology departs slightly from the terminology in [35, Chapter 4]. Here one considers equivalence classes $\mathcal{B}$ of sequences of bodies B_n satisfying (i) such that the sequences of measures $\{\nu_{B_n,C}\}$ have the same weak limit $\nu_{\mathcal{B}}$. One says that $\mathcal{B}$ is a *rough body* obtained by *grooving* C, and the measure $\nu_{\mathcal{B}}$ is given no special name.

Remark 1.15. Condition (i) looks different from the corresponding condition in the definition of a rough reflection law, where the wall boundaries Γ_n approach the line $x_2 = 0$ uniformly. Nonetheless, an equivalent definition of a rough scattering law is obtained by replacing (i) with the more restrictive condition:

(i)' $\sup\{\mathrm{dist}(x, \partial C) : x \in B_n\} \to 0$ as $n \to \infty$.

This is not immediate, but it follows from the characterization of rough scattering laws (Theorem 1.17 – below) and its proof. The sequence of bodies B_n constructed in the proof of the characterization theorem [35, Thm 4.5] can in fact be taken to satisfy (i)'.

Define a measure τ_C on $(\mathbb{S}^{d-1})^2$ by

$$\int_{(\mathbb{S}^{d-1})^2} g(u,n)\tau_C(\mathrm{d}u\mathrm{d}n) = \int_{\partial C \times \mathbb{S}^{d-1}} g(u, n(x))\Lambda_C(\mathrm{d}x\mathrm{d}u). \tag{1.27}$$

Let $\pi_{u,n} : (u, u', n) \mapsto (u, n)$ and $\pi_{u',n} : (u, u', n) \mapsto (u', n)$ be the natural projections. Let $\mathrm{Adj}(u, u', n) = (u', u, n)$. The most important elementary properties of a rough

scattering law are summarized in the following proposition.

Proposition 1.16. *A rough scattering law ν on C has the following properties:*

(i) $\pi^{\#}_{u,n}\nu = \tau_C = \pi^{\#}_{u',n}\nu$.

(ii) $Adj^{\#}\nu = \nu$.

(Here if $g : X \to Y$ is a map between measurable spaces, and μ is a measure on X, then $g^{\#}\mu$ is the *pushforward measure* on Y defined by $g^{\#}\mu(A) = \mu(g^{-1}(A))$.)

The proposition above follows from the fact that any measure $\nu_{B,C}$ must satisfy (i) and (ii), and these properties are preserved in weak limits. For $\nu_{B,C}$, property (i) is a consequence of the fact that $P^{B,C}$ preserves the measure Λ_C, while property (ii) follows from involutivity of $P^{B,C}$.

In fact, the properties (i) and (ii) completely characterize the rough scattering laws.

Theorem 1.17 (35, Theorem 4.5)**.** *A measure ν on $\mathbb{S}^3$ is a rough scattering law on C if and only if ν satisfies properties (i) and (ii).*

The original motivation for considering the scattering law and its characterization is its relationship to certain *resistance functionals* of the form

$$R_\chi[P^{B,C}] := \int_{(\partial C \times \mathbb{S}^{d-1})_+} c(u, u'(x,u))(u \cdot n(x))\mathrm{d}x\widetilde{\chi}(\mathrm{d}u), \tag{1.28}$$

where c is a "cost function" on $(\mathbb{S}^{d-1})^2$, and χ is a Borel measure on $\mathbb{S}^{d-1}$. When χ

is Lebesgue measure, the above functional may be expressed as

$$\begin{aligned} R[P^{B,C}] &= \int_{(\partial C \times \mathbb{S}^{d-1})_+} c(u, u'(x,u)) \Lambda_C(\mathrm{d}x\mathrm{d}u) \\ &= \int_{(\mathbb{S}^{d-1})^3} c(u,u') u \cdot n \ \nu_{B,C}(\mathrm{d}u\mathrm{d}u'\mathrm{d}n). \end{aligned} \tag{1.29}$$

By taking weak limits, we can extend the definition of resistance functionals to rough scattering laws:

$$R[\nu] := \int_{(\mathbb{S}^{d-1})^3} c(u,u') u \cdot n \ \nu(\mathrm{d}u\mathrm{d}u'\mathrm{d}n). \tag{1.30}$$

With the characterization given by Theorem 1.17, one can thus reduce problems of minimizing air resistance of rough convex bodies to problems in mass optimal transport. This leads to some counterintuitive results. It is possible, for example, to actually *decrease* the air resistance of a convex body by appropriately roughening its surface. In fact, one can construct nonconvex bodies which have arbitrarily small air resistance in one direction, as well as bodies which are invisible in one direction. By contrast, the air resistance of a smooth convex body has long been known to have a strictly positive lower bound [35, Chapters 5 and 8].

Let us now describe the connection between rough reflections and rough scattering laws. Just as in the upper half-plane case, we can define deterministic Markov kernels on $(\partial C \times \mathbb{S}^{d-1})_+$ by

$$\mathbb{P}^{B,C}(x,u;\mathrm{d}x'\mathrm{d}u') = \delta_{P^{B,C}(x,u)} \mathrm{d}x'\mathrm{d}u'. \tag{1.31}$$

We say that Markov kernel $\mathbb{P}$ is a *rough reflection law on C* if there exists a sequence

of bodies $B_n \subset C$ such that $\mathrm{Vol}(C \setminus B_n) \to 0$, and

$$\mathbb{P}^{B_n,C}(x,u;\mathrm{d}x'\mathrm{d}u')\Lambda_C(\mathrm{d}x\mathrm{d}u) \to \mathbb{P}(x,u;\mathrm{d}x'\mathrm{d}u')\Lambda_C(\mathrm{d}x\mathrm{d}u) \tag{1.32}$$

weakly in the space of measures on $(\partial C \times \mathbb{S}^{d-1})_+ \times (\partial C \times \mathbb{S}^{d-1})_+$. A measure ν is a rough scattering law if and only if there exists a rough reflection law $\mathbb{P}$ such that

$$\begin{aligned}&\int_{(\mathbb{S}^{d-1})^3} f(u,u',n)\nu(\mathrm{d}u\mathrm{d}u'\mathrm{d}n)\\ &\quad = \int_{((\partial C\times\mathbb{S}^{d-1})_+)^2} f(u,u',n(x))\mathbb{P}(x,u;\mathrm{d}x'\mathrm{d}u')\Lambda_C(\mathrm{d}x\mathrm{d}u).\end{aligned} \tag{1.33}$$

Rough scattering laws and rough reflection laws are not in one-to-one correspondence. One way to see this is to consider a convex body C with a flat side $V \subset \partial C$. The rough scattering law will not distinguish pointwise variation in reflection from V, because the unit normal vector at each point in V is the same.

As noted above, the characterization of rough scattering laws predates the characterization of rough reflection laws. Although Theorem 1.17 does not imply Theorem 1.9, the proof of the former can almost certainly be modified to obtain the latter. In fact, higher dimensional versions of Theorem 1.9 can probably be proved by an appropriate modification of the arguments in [35, Chapter 4]. The proofs of both theorems begin with a local construction of reflectors which redirect a point particle hitting the boundary in a prescribed way. It is then a matter of appropriately assembling the local reflectors to produce the desired rough scattering law or rough reflection law.

1.3 Disk and wall model

1.3.1 Rigid body system

The model we study consists of a fixed wall in the lower half-plane of $\mathbb{R}^2$, together with a disk which is given some rotationally symmetric mass density and allowed to move freely in the complement of the wall. Each body is furnished with asperities (microscopic structures) on its surface. Aside from a periodicity requirement, the asperities on the surface of the wall are allowed to be fairly arbitrary in shape. On the other hand, the asperities on the disk are of a specific type, namely "geostationary satellites" spaced in such a way as to guarantee that non-local interactions between the two bodies are rare.

The wall W will be built from a cell Σ and a roughness scale $\epsilon > 0$, in the same way as we have done in §1.2.2. The cell Σ is assumed to satisfy conditions B1-B5, stated in §1.2.4. The definition of the wall $W = W(\Sigma, \epsilon)$ is slightly modified as follows: $W(\Sigma, \epsilon)$ is the unique subset of $\mathbb{R}^2$ which is ϵ-periodic in the x_1-direction, and whose image under the covering map

$$(x_1, x_2) \mapsto (e^{2\pi i x_1/\epsilon}, \epsilon^{-1}(x_2 + 1)) \tag{1.34}$$

is Σ. Such a wall will satisfy conditions A1-A4 and A5a and A5b, stated in §1.2.2, except that condition A3 must be modified as follows:

A3'. The following inclusions hold: $\mathbb{R} \times (-\infty, -1 - \epsilon] \subset W \subset \mathbb{R} \times (-\infty, -1]$; thus $\partial W \subset [-1 - \epsilon, -1] \times \mathbb{R}$.

The advantage of having W located below the line $x_2 = -1$, instead of $x_2 = 0$, will become apparent later.

By assumption, the cell boundary $\partial\Sigma$ decomposes into a finite collection of closed C^2 curve segments $\{\widetilde{\Gamma}_i\}_{i=0}^{m-1}$. Such a decomposition determines a decomposition of ∂W into a countable, locally finite collection $\{\Gamma_i\}_{i\in\mathbb{Z}}$ of C^2 curve segments such that distinct pairs Γ_i and Γ_j can intersect only at their endpoints, and Γ_i maps to $\widetilde{\Gamma}_{i \bmod m}$ via the covering map (1.34) defined above.

Remark 1.18. Recall what it means for a closed curve segment Γ_i to be C^2: There an open interval $I \supset [0,1]$ and a C^2 map $\gamma_i : I \to \mathbb{R}^2$ such that $\gamma_i([0,1]) = \Gamma_i$. Consequently, by compactness, each of the curve segments Γ_i has bounded curvature.

The decomposition of $\partial\Sigma$ into C^2 curve segments $\widetilde{\Gamma}_i$ is not unique, and correspondingly the decomposition of ∂W into C^2 curve segments Γ_i is not unique. To ensure that terms introduced below are well-defined, we assume from this point on that some decomposition $\{\widetilde{\Gamma}_i\}_{0\le i\le m-1}$ of $\partial\Sigma$ and correspondingly $\{\Gamma_i\}_{i\in\mathbb{Z}}$ of ∂W has been fixed, and we shall refer to these as the *given decompositions* of $\partial\Sigma$ and ∂W respectively.

We denote the relative interior of a curve segment Γ_i by $\operatorname{Int}\Gamma_i$. We call a point $p \in \partial\Sigma$ *regular* if there is some C^2 curve segment $\Gamma \subset \partial\Sigma$ (not necessarily coming from the given decomposition) such that $p \in \operatorname{Int}\Gamma$. We denote the set of regular points in $\partial\Sigma$ by $\partial_{\text{reg}}\Sigma$, and we denote the set of regular points in ∂W by $\partial_{\text{reg}}W$. We let $\partial_s\Sigma = \partial\Sigma \setminus \partial_{\text{reg}}\Sigma$, and we let $\partial_s W = \partial W \setminus \partial_{\text{reg}}W$. We refer to points in $\partial_s\Sigma$ and $\partial_s W$ as *singular points* of $\partial\Sigma$ and ∂W respectively. The sets $\partial_s W$ and $\partial_s\Sigma$ are measure zero subsets of $\partial\Sigma$ and ∂W respectively.

For $p \in \partial_{\text{reg}} W$, we let $\kappa(p)$ denote the unsigned curvature of ∂W at p. We define

$$\kappa_{\max} = \sup\{\kappa(p) : p \in \partial_{\text{reg}} W(\Sigma, 1)\}. \tag{1.35}$$

The quantity above depends only on the cell Σ. Per Remark 1.18 and periodicity of ∂W, $\kappa_{\max}$ is finite.

We take $\rho(\epsilon)$ to be some positive, non-decreasing function of ϵ such that

$$\rho(\epsilon) \to 0 \quad \text{and} \quad \frac{\epsilon^{1/2}}{\rho(\epsilon)} \to 0 \quad \text{as } \epsilon \to 0. \tag{1.36}$$

In addition, we assume that $2\pi/\rho(\epsilon)$ is an integer.

The freely moving body in our system is a "disk with satellites." Namely, we first define a *reference body*

$$D = D(\epsilon) = \left(\bigcup_{k=0}^{N-1} S_k \right) \cup D_0 \subset \mathbb{R}^2, \tag{1.37}$$

where

$$N = \frac{2\pi}{\rho}, \qquad S_k = (\sin k\rho(\epsilon), -\cos k\rho(\epsilon)), \tag{1.38}$$

and

$$D_0 \subset \{(x_1, x_2) : x_1^2 + x_2^2 \leq (1 - 2\rho(\epsilon)^2)^2\}. \tag{1.39}$$

See Figure 5. The quantity N is the number of satellites, and S_k is the position of the k'th satellite in the reference body.

For $x = (x_1, x_2) \in \mathbb{R}^2$ and $\alpha \in \mathbb{R}$, we let $D(x, \alpha)$ denote the subset of the plane

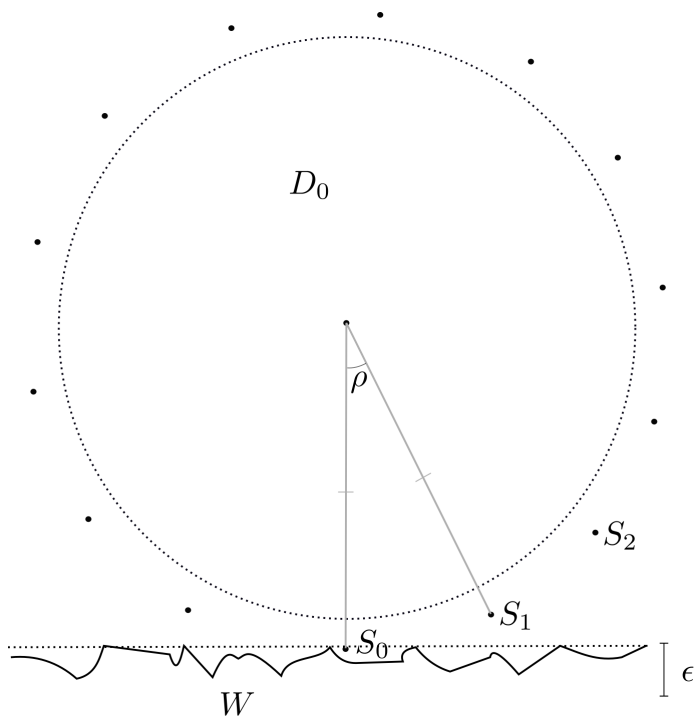

Figure 5: A freely moving disk with satellites, and a fixed wall.

occupied by the reference body after rotating it counterclockwise about the origin by an angle α and translating it by x from the reference configuration (1.37). We will sometimes abuse terminology by using D to denote both the reference body (1.37) (a fixed subset of the plane) and the freely moving physical body which it represents.

We assume that D has some mass density λ which is rotationally symmetric about the origin when D is in reference configuration (1.37), and we introduce parameters:

$$m = \int_D \lambda(\mathrm{d}x), \qquad J = \int_D |x|^2 \lambda(\mathrm{d}x) \tag{1.40}$$

– the *mass* and *moment of inertia*, respectively, of the disk. We assume that λ does not depend on the parameter $\epsilon > 0$.

In the analysis of our model, we will see that there is no loss of generality in assuming that $m = J = 1$; however, we will not introduce this simplification until later.

Remark 1.19. Additional motivation for our choice of the body D can be gained from the following remarks.

- The first member of the union (1.37) is a discrete set of points S_k (the "satellites") which are disconnected from the rest of the body and evenly spaced at angles of $\rho(\epsilon)$ along the unit circle. We have stipulated that $\rho(\epsilon)$ should converge to zero, but at a rate more slowly than $\epsilon^{1/2}$. This will guarantee that interactions between multiple satellites during a single collision event are rare as the roughness scale goes to zero.

- The subset D_0 is of no mathematical significance in the analysis of the disk and

wall model, but serves only to persuade the reader of the physical generality of the model. One may imagine that D_0 carries the mass of the body D. The inclusion (1.39) guarantees that D_0 can never come into contact with the wall W. Only the satellites S_k can come into contact with W.

- Choosing the "roughness" on D, as we have above, so that only isolated points can interact with the wall greatly simplifies the description of the configuration space of the system. Unfortunately, the choice of a disconnected body is physically unrealistic. One way to reconcile this with our intuition is to think of D as a "hockey puck" with additional features in the third (vertical) dimension. In the same spirit as the analysis of polygonal chains in [11, §5.3], we obtain a physical system which is equivalent to the one above by joining each satellite to the inner body D_0 by a curved rod which extends into the third dimension.

- Whether theorems analogous to the ones we state in §1.3.7 can be proved for a connected body D is an open question. On the other hand, we expect the range of possible dynamics for a system consisting of a freely moving rough disk and fixed rough wall to be exhausted by our model. See the discussion of our main results in §1.1.3.

1.3.2 Configuration space

The *configuration space* of the disk and wall system is the topological closure in $\mathbb{R}^3$ of the set of configurations $y = (x_1, x_2, \alpha)$ of the disk such that the disk and the wall are disjoint, i.e.

$$\mathcal{M} = \overline{\{(x_1, x_2, \alpha) : D(x_1, x_2, \alpha) \cap W = \varnothing\}}. \tag{1.41}$$

Note that points in $\mathcal{M}$ which differ only in their angular coordinates by multiples of 2π represent the same physical configuration. It is in fact easy to see that $\mathcal{M}$ is doubly periodic: Let

$$e_1 = (1,0,0), \quad e_2 = (0,1,0), \quad e_3 = (0,0,1) \tag{1.42}$$

be the standard basis for $\mathbb{R}^3$. Then

$$\begin{aligned} \mathcal{M} + \epsilon e_1 &= \{(x_1, x_2, \alpha) : (x_1 - \epsilon, x_2, \alpha) \in \mathcal{M}\} = \mathcal{M}, \qquad \text{and} \\ \mathcal{M} + \rho e_3 &= \{(x_1, x_2, \alpha) : (x_1, x_2, \alpha - \rho) \in \mathcal{M}\} = \mathcal{M}. \end{aligned} \tag{1.43}$$

Indeed, $\mathcal{M}$ is ϵ-periodic in the x_1-coordinate because W is ϵ-periodic in the x_1-coordinate, and $\mathcal{M}$ is ρ-periodic in the α-coordinate because, as noted above, only the satellites of D can come into contact with the wall, and rotating the disk about its center of mass by an angle of $\rho(\epsilon)$ maps the set of satellites onto itself.

We can alternatively study the configuration space $\mathcal{M}/\sim$, where $\sim$ is the equivalence relation which identifies points y and y' such that $y' = y + i\epsilon e_1 + j\rho e_3$, for some $i, j \in \mathbb{Z}$. The space $\mathcal{M}/\sim$ may be regarded as a subspace of $\mathbb{S}^1 \times \mathbb{R} \times \mathbb{S}^1$. This configuration space will be useful for applying dynamical results which require compactness, but otherwise we will mainly just work with $\mathcal{M}$.

The topological boundary $\partial\mathcal{M}$ of the configuration space corresponds to the set of *collision configurations* of the system, i.e. the set of configurations in which the

rough disk and the wall are in contact. The boundary of $\mathcal{M}$ lies just below the plane

$$\mathbf{P} := \{(x_1, x_2, \alpha) \in \mathbb{R}^3 : x_2 = 0\}. \tag{1.44}$$

The space $\mathcal{M}$ does not fit neatly into a well-studied class of manifolds. It is probably not even a manifold with corners for many choices of the wall W. The space $\mathcal{M}$ may loosely be described as a C^2 manifold with boundary and "singularities." We will see that there exists a closed subset $\mathcal{S} \subset \partial\mathcal{M}$ such that the 2-dimensional Hausdorff measure of $\mathcal{S}$ is zero, and $\mathcal{M}_{\text{reg}} := \mathcal{M} \setminus \mathcal{S}$ is an embedded C^2 submanifold of $\mathbb{R}^3$ with boundary. In particular, $\partial_{\text{reg}}\mathcal{M} := \partial\mathcal{M} \setminus \mathcal{S}$ is a full-measure subset of the boundary on which there exists a C^1 field of inward-pointing unit normal vectors $n(q), q \in \partial_{\text{reg}}\mathcal{M}$.

A more detailed description of the configuration space $\mathcal{M}$ is given in §4.1. In §6, we reprove a number of standard results from billiards for a very general class of manifolds to which $\mathcal{M}$ may be shown to belong.

1.3.3 Cylindrical approximation

Some additional understanding of the structure of $\mathcal{M}$ can be gained by imagining the following physical situation. Suppose the configuration of the disk D is initially $y = (x_1, x_2, 0)$. Then at most one satellite of D can be in contact with ∂W, and this satellite is S_0 (this follows from Proposition 4.2 for example). In this configuration, the coordinates of S_0 are $(x_1, x_2{-}1)$. All other satellites must lie in W^c. Consequently, we may rotate the disk counterclockwise about the satellite S_0 by a small angle $\Delta\alpha$, and the resulting configuration y' will still lie in $\mathcal{M}$. Moreover, if $y \in \partial\mathcal{M}$ then

$y' \in \partial\mathcal{M}$. We compute explicitly:

$$y' = (x_1 - \sin\Delta\alpha, x_2 - 1 + \cos\Delta\alpha, \Delta\alpha) \approx (x_1 - \Delta\alpha, x_2, \Delta\alpha). \tag{1.45}$$

Let

$$\mathbf{Q}_0 = \{(x_1, x_2, \alpha) : \alpha = 0\}. \tag{1.46}$$

We will see that $\mathcal{M} \cap \mathbf{Q}_0 = \overline{W^c} + e_2$. Thus (1.45) suggests that, in a neighborhood of the plane $\mathbf{Q}_0$, the space $\mathcal{M}$ is well modeled by the cylinder

$$\mathcal{M}_{\text{cyl}} := \{(x_1, x_2, \alpha) : (x_1 + \alpha, \alpha) \in \overline{W^c} + e_2\}. \tag{1.47}$$

This is the cylinder with base

$$\widehat{B} := \overline{W^c} + e_2 \tag{1.48}$$

and axis

$$\widetilde{\chi} := (-1, 0, 1). \tag{1.49}$$

The vector $\widetilde{\chi}$ is the "rolling velocity" described in the introduction.

More generally, we will see in §4.1 that $\mathcal{M}$ may be decomposed into large "cylindrical regions" and small "gap regions." The "gap regions" are difficult to describe explicitly and correspond to configurations where multiple satellites can come into contact with ∂W. Near any point in one of the "cylindrical regions" $\mathcal{M}$ is well modeled by a cylinder whose base is an order ϵ translate of $\widehat{B}$ and whose axis is $\widetilde{\chi}$.

The main results of this work were originally derived informally by assuming that we can replace $\mathcal{M}$ with $\mathcal{M}_{\text{cyl}}$ in our arguments.

1.3.4 Phase space and dynamics

The dynamical state of the system is completely specified by a pair (y, w), where $y = (x, \alpha) = (x_1, x_2, \alpha) \in \mathcal{M}$ and $w = (v, \omega) = (v_1, v_2, \omega) \in \mathbb{R}^3$. Here $x = (x_1, x_2)$ is the center of mass of the disk, and α is the angular orientation of the disk, as above; $v = (v_1, v_2)$ is the linear velocity of the center of mass of the disk, and ω is the angular velocity.

The *phase space* is the set of all possible states (y, w) of the system:

$$T\mathcal{M} \cong \mathcal{M} \times \mathbb{R}^3. \tag{1.50}$$

The notation $T\mathcal{M}$ refers to the fact that geometrically we regard the phase space as the tangent space of the configuration space. For technical purposes, the phase space can be taken simply as the Cartesian product of the configuration space $\mathcal{M}$ and the "velocity space" $\mathbb{R}^3$; however it will occasionally be convenient to use the above geometric language, for example when referring to the tangent space $T_y\mathcal{M} \subset T\mathcal{M}$ at a point y in $\mathcal{M}$, which in this context is just the subset $\{y\} \times \mathbb{R}^3$.

We equip $\mathbb{R}^3$ with the inner product

$$\langle w, w' \rangle = mv_1v_1' + mv_2v_2' + J\omega\omega', \quad \text{for} \quad w = (v_1, v_2, \omega), \quad w' = (v_1', v_2', \omega'), \tag{1.51}$$

and we denote the corresponding norm by $\|\cdot\|$. If the disk moves with velocity

$w = (v_1, v_2, \omega)$, then the kinetic energy of the disk is

$$\frac{1}{2}mv_1^2 + \frac{1}{2}mv_2^2 + \frac{1}{2}J\omega^2 = \frac{1}{2}||w||^2. \tag{1.52}$$

The sum of the first two terms is the kinetic energy arising from the linear motion of the disk, while the third term is the kinetic energy arising from rotational motion. For this reason we will often refer to the above inner product as the *kinetic energy inner product.*

Unless specified otherwise, *distances, angles, etc. in $\mathcal{M}$ are assumed to be with respect to the inner product $\langle \cdot, \cdot \rangle$ defined above.*

The main physical assumptions which govern the dynamics of the model are:

P1. The bodies D and W do not interpenetrate.

P2. The system is subject to Euler's laws of rigid body motion (see §3).

P3. When not in contact, the net force applied to each body is zero, and upon contact, a single impulsive force is applied to the disk at the point of contact and directed parallel to the unit normal vector on the wall.

P4. The kinetic energy of the disk is conserved for all time.

These assumptions are discussed in greater detail and given a more precise mathematical form in §3.

Assumptions P1-P3 are standard for rigid body interactions. Assumption P4 is the limiting case for a system of bodies of finite mass in which the total kinetic energy, linear momentum, and angular momentum are conserved. After letting the mass and

moment of inertia of one body diverge to infinity, the kinetic energy of the other body is conserved in the limit. (Note that the wall W may be regarded as having infinite mass and infinite moment of inertia.) For more details, see Proposition 3.3.

The dynamics of the disk and wall system may be represented by a moving point mass in $\mathcal{M}$, where the position and velocity of the point mass is given by the state (y, w) of the disk and wall system. It follows from the assumptions P3 and P4 that the point mass moves linearly with constant velocity in the interior of $\mathcal{M}$. When the point mass comes into contact with the boundary at a point $q \in \partial\mathcal{M}$, the velocity should be redirected back into $\mathcal{M}$ by a mapping $R_q : T_q\mathcal{M} \to T_q\mathcal{M}$. The dynamical description of our system is completed by the following proposition.

Proposition 1.20. *Let $q \in \partial_{\text{reg}}\mathcal{M}$ and let $n(q)$ denote the inward-pointing unit normal vector at q. The unique map $R_q : T_q\mathcal{M} \to T_q\mathcal{M}$ for which assumptions P1-P4 are satisfied is specular reflection with respect to the kinetic energy inner product:*

$$R_q(w) = w - 2\langle w, n(q)\rangle n(q), \qquad w \in T_q\mathcal{M}. \tag{1.53}$$

Proposition 1.20 is proved in §3. The main takeaway is that the system under consideration has the same dynamical evolution as a classical billiard: free motion in the interior of the billiard domain and specular reflection on the boundary.

Remark 1.21. The inner product $\langle\cdot,\cdot\rangle$ makes $\mathbb{R}^3$ into a geodesically complete Riemannian manifold, and thus results proved in §6 for billiards in general Riemannian manifolds will apply in this case. To avoid confusion, we will refer to $\langle\cdot,\cdot\rangle$ as an "inner product" in §§1-5, reserving the word "metric" to refer to either (a) the dis-

tance functional on a metric space, or (b), in §6, a Riemannian metric on a general geodesically complete Riemannian manifold. Context will be sufficient to distinguish senses (a) and (b).

1.3.5 Collision laws

The configuration space $\mathcal{M}$ contains the plane $\mathbf{P} := \{(x_1, x_2, \alpha) : x_2 = 0\} \subset \mathbb{R}^3$. The boundary $\partial\mathcal{M}$ lies just below $\mathbf{P}$ within distance $O(\rho(\epsilon)^2)$ from the plane (this is proved in Proposition 4.1).

Let $\mathbb{R}^3_\pm := \{(x_1, x_2, \alpha) : x_2 \geq 0\}$. Let $\mathbb{S}^2$ denote the unit sphere in $\mathbb{R}^3$ with respect to the kinetic energy norm $\|\cdot\|$ defined above. By conservation of energy, the velocity of a point particle representing the system may be taken to lie in the unit sphere $\mathbb{S}^2$ for all time. Let $\mathbb{S}^2_\pm = \mathbb{S}^2 \cap \mathbb{R}^3_\pm = \{(v_1, v_2, \omega) \in \mathbb{S}^2 : \pm v_2 > 0\}$.

Let $(y, w) \in \mathbf{P} \times \mathbb{S}^2_+$, and suppose the initial state of the point particle is $(y, -w)$. The velocity $-w$ is directed toward the boundary $\partial\mathcal{M}$. Therefore, the point particle will reflect from $\partial\mathcal{M}$ specularly a certain number of times before eventually returning to a point y' in the plane $\mathbf{P}$ with some velocity $w' \in \mathbb{S}^2_+$. The *collision law* is the map $\mathbf{P} \times \mathbb{S}^2_+ \to \mathbf{P} \times \mathbb{S}^2_+$, defined by

$$K^{\Sigma,\epsilon}(y, w) = (y', w'). \tag{1.54}$$

The set of inputs (y, w) such that the billiard trajectory starting from $(y, -w)$ cannot be continued for all time (say because it hits a singular point of $\partial\mathcal{M}$) or never returns to $\mathbf{P}$ constitutes a measure zero subset of $\mathbf{P} \times \mathbb{S}^2_+$, and thus $K^{\Sigma,\epsilon}$ is well-defined up to null sets. This fact is rigorously proved in §4.3.

Remark 1.22. The reader should be careful not to confuse the terms "collision law" and "reflection law." In our usage, the latter expression is a generic term for any rule whereby the particle trajectory is redirected into the billiard domain at the instant in time when it hits the wall (e.g. the specular reflection law). The collision law, by contrast, is an analogue of the macro-reflection law defined in §1.2.2. In fact, we will see that the collision law is a special case of the general macro-reflection laws defined in §6.2.1.

Just as in the random reflections case, the map $K^{\Sigma,\epsilon}$ is naturally associated with the deterministic Markov kernel on $\mathbf{P} \times \mathbb{S}^2_+$,

$$\mathbb{K}^{\Sigma,\epsilon}(y, w; \mathrm{d}y'\mathrm{d}w') := \delta_{K^{\Sigma,\epsilon}(y,w)}(y', w')\mathrm{d}y'\mathrm{d}w', \tag{1.55}$$

We equip $\mathbf{P} \times \mathbb{S}^2_+$ with the well-known measure from billiards theory,

$$\Lambda^2(\mathrm{d}y\mathrm{d}w) := \langle w, e_2\rangle \mathrm{d}y\sigma(\mathrm{d}w), \tag{1.56}$$

where $\mathrm{d}y$ denotes Lebesgue measure on $\mathbf{P}$ and $\sigma(\mathrm{d}w)$ denotes surface measure on $\mathbb{S}^2$. This is the 3-dimensional analogue of the Lambertian measure Λ^1, defined by (1.5).

Definition 1.23. We call a Markov kernel $\mathbb{K}$ on $\mathbf{P} \times \mathbb{S}^2_+$ a *rough collision law* if there exists a sequence of cells Σ_i (satisfying conditions B1-B5 of §1.2.4) and a sequence of positive numbers $\epsilon_i \to 0$ such that,

$$\mathbb{K}^{\Sigma_i,\epsilon_i}(y, w; \mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w) \to \mathbb{K}(y, w; \mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w) \tag{1.57}$$

weakly in the space of measures on $\mathbf{P} \times \mathbb{S}^2_+$.

Remark 1.24. Similar comments to Remarks 1.4 and 1.5 apply. In the above definition, it is equivalent to replace Λ^2 with any measure which is mutually absolutely continuous with respect to surface measure on $\mathbf{P} \times \mathbb{S}^2_+$. The convergence (1.57) is in duality to functions $h(y, w, y', w') \in C_c((\mathbf{P} \times \mathbb{S}^2_+)^2)$, but by a density argument it is sufficient to verify the convergence in duality to functions of form $h(y, w, y', w') = f(y, w)g(y', w')$ where $f, g \in C_c^\infty(\mathbf{P} \times \mathbb{S}^2_+)$.

Remark 1.25. In (1.57), there is a sense in which $\mathbb{K}^{\Sigma_i, \epsilon_i}$ converges to $\mathbb{K}$ as a limit with respect to a pseudometric topology on the space of Markov kernels on $\mathbf{P} \times \mathbb{S}^2_+$. This topology is described in §6.2.3.

From this point on, we will write $\lim_{i \to \infty} \mathbb{K}^{\Sigma_i, \epsilon_i} = \mathbb{K}$ to indicate that (1.57) holds.

The most important elementary properties of collision laws are summarized in the following proposition. (Compare with Propositions 1.1 and 1.3.)

Proposition 1.26. *(i) There exists a full measure open set $\mathcal{F} \subset \mathbf{P} \times \mathbb{S}^2_+$ such that $K^{\Sigma,\epsilon} : \mathcal{F} \to \mathcal{F}$ is a C^2 diffeomorphism and $K^{\Sigma,\epsilon} \circ K^{\Sigma,\epsilon} = Id_{\mathcal{F}}$ on $\mathcal{F}$.*

Moreover, $K^{\Sigma,\epsilon}$ preserves the measure Λ^2, in the sense that for any set $A \subset \mathbf{P} \times \mathbb{S}^2_+$, $\Lambda^2((K^{\Sigma,\epsilon})^{-1}(A)) = \Lambda^2(A)$.

(ii) A rough collision law $\mathbb{K}$ is symmetric with respect to Λ^2, in the sense that for any $f, g \in C_c(\mathbf{P} \times \mathbb{S}^2_+)$,

$$\begin{aligned} &\int_{(\mathbf{P} \times \mathbb{S}^2_+)^2} g(y', w') \mathbb{K}(y, w; \mathrm{d}y' \mathrm{d}w') f(y, w) \Lambda^2(\mathrm{d}y \mathrm{d}w) \\ &\qquad = \int_{(\mathbf{P} \times \mathbb{S}^2_+)^2} f(y', w') \mathbb{K}(y, w; \mathrm{d}y' \mathrm{d}w') g(y, w) \Lambda^2(\mathrm{d}y \mathrm{d}w). \end{aligned} \tag{1.58}$$

Consequently, $\mathbb{K}$ *preserves* Λ^2*, in the sense that*

$$\int_{(\mathbf{P}\times\mathbb{S}^2_+)^2} g(y',w')\mathbb{K}(y,w;\mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w) = \int_{\mathbf{P}\times\mathbb{S}^2_+} g(y,w)\Lambda^2(\mathrm{d}y\mathrm{d}w). \tag{1.59}$$

This result is proved in §4.3.

1.3.6 Coordinates on $\mathbf{P}\times\mathbb{S}^2_+$

Recall the cylindrical axis $\widetilde{\chi}$, defined by (1.49). We equip $\mathbb{R}^3\times\mathbb{S}^2$ with coordinates $(y_1,y_2,y_3,\theta,\psi)$ defined as follows. Let

$$\chi = \|\widetilde{\chi}\|^{-1}\widetilde{\chi} = (m+J)^{-1/2}(1,0,-1), \qquad \chi^\perp = (m^{-1}+J^{-1})^{-1/2}(m^{-1},0,J^{-1}). \tag{1.60}$$

Also let

$$\widehat{e}_2 = m^{-1/2}e_2 = (0,m^{-1/2},0). \tag{1.61}$$

Then $(\chi^\perp,\widehat{e}_2,\chi)$ is an orthonormal basis for $\mathbb{R}^3$ (with respect to the inner product $\langle\cdot,\cdot\rangle$). The two vectors χ and $\chi^\perp$ span $\mathbf{P}$. We define a coordinate map $G:\mathbb{R}^3\times(0,\pi)^2\to\mathbb{R}^3\times\mathbb{S}^2$ by

$$G:(y_1,y_2,y_3,\theta,\psi)\mapsto(y,w), \tag{1.62}$$

where

$$\begin{aligned} y &= y_1\chi^\perp + y_2\widehat{e}_2 + y_3\chi, \\ w &= (\cos\theta\sin\psi)\chi^\perp + (\sin\theta\sin\psi)\widehat{e}_2 + (\cos\psi)\chi. \end{aligned} \tag{1.63}$$

In other words, the spatial coordinates (y_1, y_2, y_3) are obtained by rotating (with respect to the kinetic energy inner product) the original coordinates (x_1, x_2, α) so that the vertical axis coincides with χ. The velocity coordinates (θ, ψ) are just spherical coordinates with the "north pole" at χ.

In these coordinates, $\mathbf{P} = \{(y_1, y_2, y_3) : y_2 = 0\}$. The invariant measure Λ^2 has the following coordinate representation:

$$\Lambda^2(\mathrm{d}y_1 \mathrm{d}y_3 \mathrm{d}\theta \mathrm{d}\psi) = \sin\theta \sin^2\psi \mathrm{d}y_1 \mathrm{d}y_3 \mathrm{d}\theta \mathrm{d}\psi. \tag{1.64}$$

To see this, note that $\langle w, e_2 \rangle = \sin\theta \sin\psi$ and the spherical volume form is $\sin\psi \mathrm{d}\theta \mathrm{d}\psi$.

1.3.7 Main results

The first of our main results concerns the case in which the rough collision law is obtained through *pure scaling*. This means that the sequence of cells $\Sigma_i = \Sigma$ is constant.

Theorem 1.27. *Consider a constant sequence of cells $\Sigma_i = \Sigma$ for $i \geq 1$. There exists a Markov kernel $\mathbb{K}$ such that for any sequence of positive numbers $\epsilon_i \to 0$, the limit $\lim_{i\to\infty} \mathbb{K}^{\Sigma_i, \epsilon_i}$ exists and is equal to $\mathbb{K}$. Moreover, $\mathbb{K}$ takes the form*

$$\mathbb{K}(y_1, y_3, \theta, \psi; \mathrm{d}y_1' \mathrm{d}y_3' \mathrm{d}\theta' \mathrm{d}\psi') = \delta_{(y_1, y_3)}(\mathrm{d}y_1' \mathrm{d}y_3') \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') \delta_{\pi-\psi}(\mathrm{d}\psi'), \tag{1.65}$$

where $\widetilde{\mathbb{P}}$ is a Markov kernel on $\mathbb{S}^1_+$ satisfying the following properties:

i. $\widetilde{\mathbb{P}}$ is symmetric with respect to the measure $\sin\theta \mathrm{d}\theta$, in the sense of (1.17)

ii. Let

$$\begin{aligned}
\widetilde{\Sigma} &= \{(x_1, x_2) \in \mathbb{S}^1 \times \mathbb{R} : (x_1, (1 + mJ^{-1})^{-1/2} x_2) \in \Sigma\}, \\
\widetilde{\epsilon}_i &= (1 + mJ^{-1})^{-1/2} \epsilon_i, \\
\mathbb{P} &= \delta_y(\mathrm{d}y') \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta').
\end{aligned} \tag{1.66}$$

Then

$$\mathbb{P} = \lim_{i \to \infty} \mathbb{P}^{\widetilde{\Sigma}, \widetilde{\epsilon}_i}. \tag{1.67}$$

Consequently, $\mathbb{K}$ *is symmetric with respect to the measure* Λ^2*.*

Our next two results concern more general rough collision laws. For an arbitrary sequence of cells $\{\Sigma_i\}_{i \geq 1}$, there is no guarantee that the limit of $\mathbb{K}^{\Sigma_i, \epsilon_i}$ exists or is uniquely determined by the sequence of cells Σ_i. Nonetheless, if $\epsilon_i \to 0$ sufficiently fast (where the rate depends on the sequence of cells Σ_i), a strict dichotomy holds.

Theorem 1.28. *For any sequence of cells* $\{\Sigma_i\}_{i \geq 1}$*, there exist numbers* $b_1 \geq b_2 \geq b_3 \geq \cdots > 0$ *such that exactly one of the following is true:*

(A) There exists a unique Markov kernel $\mathbb{K}$ *such that for any sequence of positive numbers* $\epsilon_i \leq b_i$ *with* $\epsilon_i \to 0$*,*

$$\lim_{i \to \infty} \mathbb{K}^{\Sigma_i, \epsilon_i} = \mathbb{K}. \tag{1.68}$$

(B) For any sequence of positive numbers $\epsilon_i \leq b_i$ *with* $\epsilon_i \to 0$*,* $\lim_{i \to \infty} \mathbb{K}^{\Sigma_i, \epsilon_i}$ *does not exist.*

Remark 1.29. Both possibilities in the dichotomy are realized. In §2 we will construct a few different examples of rough collision laws. Denote two of these by

$\mathbb{K} = \lim \mathbb{K}^{\Sigma,\epsilon_i}$ and $\mathbb{K}' = \lim \mathbb{K}^{\Sigma',\epsilon_i}$ where Σ and Σ' are fixed cells and $\mathbb{K} \neq \mathbb{K}'$ (where inequality means that $\mathbb{K}(\cdot; \mathrm{d}y'\mathrm{d}w')$ and $\mathbb{K}'(\cdot; \mathrm{d}y'\mathrm{d}w')$ disagree on a non-null set). Note that by Theorem 1.27, these limits always exist and do not depend on the sequence $\epsilon_i \to 0$, and thus we may take $b_i = \infty$ for example. Hence we see that (A) is realized. Define a sequence of cells by $\Sigma_{2i-1} = \Sigma$ and $\Sigma_{2i} = \Sigma'$ for $i \geq 1$. Then the sequence $\mathbb{K}^{\Sigma_i,\epsilon_i}$ has two distinct limit points $\mathbb{K}$ and $\mathbb{K}'$ for any sequence $\epsilon_i \to 0$, and thus (B) is realized.

Remark 1.30. The proof of this theorem depends on the Poincaré Recurrence Theorem. As such, it does not yield quantitative estimates for the numbers b_i.

We let $\mathcal{A}_0$ denote the set of all Markov kernels $\mathbb{K}$ obtained as a limit of form $\lim_{i\to\infty} \mathbb{K}^{\Sigma_i,\epsilon_i}$, where $\epsilon_i \leq b_i$ for all i, and the b_i are chosen as in Theorem 1.28. The following theorem says that the collision laws in $\mathcal{A}_0$ have essentially the same properties as those stated in Theorem 1.27. Moreover, these properties completely characterize the members of $\mathcal{A}_0$.

Theorem 1.31. *Let $\mathbb{K} \in \mathcal{A}_0$. In the coordinates (y_1, y_3, θ, ψ), $\mathbb{K}$ takes the form*

$$\mathbb{K}(y_1, y_3, \theta, \psi; \mathrm{d}y_1'\mathrm{d}y_3'\mathrm{d}\theta'\mathrm{d}\psi') = \delta_{(y_1,y_3)}(\mathrm{d}y_1'\mathrm{d}y_3')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')\delta_{\pi-\psi}(\mathrm{d}\psi'), \tag{1.69}$$

where $\widetilde{\mathbb{P}}$ is a Markov kernel on $\mathbb{S}^1_+$ satisfying the following properties:

i. $\widetilde{\mathbb{P}}$ is symmetric with respect to the measure $\sin\theta\mathrm{d}\theta$ on $\mathbb{S}^1_+$, in the sense of (1.17).

ii. Suppose $\{\Sigma_i\}$ is a sequence of cells such that for any $\epsilon_i \leq b_i$ with $\epsilon_i \to 0$,

$\mathbb{K}^{\Sigma_i,\epsilon_i} \to \mathbb{K}$ *in the sense of Theorem 1.28(A). Let*

$$\begin{aligned}
\widetilde{\Sigma}_i &= \{(x_1, x_2) \in \mathbb{S}^1 \times \mathbb{R} : (x_1, (1 + mJ^{-1})^{-1/2} x_2) \in \Sigma_i\}, \\
\widetilde{\epsilon}_i &= (1 + mJ^{-1})^{-1/2} \epsilon_i, \\
\mathbb{P} &= \delta_y(\mathrm{d}y')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta').
\end{aligned} \tag{1.70}$$

Then

$$\mathbb{P} = \lim_{i\to\infty} \mathbb{P}^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}. \tag{1.71}$$

Consequently, $\mathbb{K}$ *is symmetric with respect to the measure* Λ^2*.*

Conversely, if $\mathbb{K}$ *is a Markov kernel on* $\mathbf{P} \times \mathbb{S}^2_+$ *of form (1.69) such that* $\widetilde{\mathbb{P}}$ *is symmetric with respect to the measure* $\sin\theta \mathrm{d}\theta$ *on* $\mathbb{S}^1_+$*, then* $\mathbb{K} \in \mathcal{A}_0$*.*

Theorems 1.27 and 1.31 establish a one-to-one correspondence between rough collision laws and rough reflection laws in the upper half-plane $\{(x_1, x_2) : x_2 \geq 0\}$, indicated schematically as follows:

$$\begin{array}{ccc}
\{(\Sigma_i, \epsilon_i)\}_{i\geq 1} & \longleftrightarrow & \{(\widetilde{\Sigma}_i, \widetilde{\epsilon}_i)\}_{i\geq 1} \\
\downarrow & & \downarrow \\
\mathbb{K} & \longleftrightarrow & \mathbb{P}
\end{array} \tag{1.72}$$

The correspondence at the top is between the sequences of cell-roughness scale pairs giving rise to the walls $W_i = W(\Sigma_i, \epsilon_i)$ and foreshortened walls $\widetilde{W}_i = W(\widetilde{\Sigma}_i, \widetilde{\epsilon}_i)$ respectively. The correspondence on the bottom is given by (1.69). The downward arrows map the sequence $\{(\Sigma_i, \epsilon_i)\}$ (resp. $\{(\widetilde{\Sigma}_i, \widetilde{\epsilon}_i)\}$) to $\lim_{i\to\infty} \mathbb{K}^{\Sigma_i,\epsilon_i}$ (resp. $\lim_{i\to\infty} \mathbb{P}^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}$). Provided $\epsilon_i \to 0$ sufficiently fast, the limit on the left exists if and only if the limit

on the right exists. We will take advantage of this correspondence in §2 for building examples of rough collision laws.

The reader may wish to compare the statement of Theorem 1.31, above, to the informal description of our main results in §1.1.3. The fact that $\mathbb{K}$ preserves the measure Λ^2 is equivalent to the fact that the rough collision dynamics preserve the Liouville measure on the phase space (one approach for proving this is suggested by Figure 3), while symmetry with respect to Λ^2 makes precise the notion of "time-reversibility." The product decomposition (1.69) implies that projection of the phase space velocity onto χ is a conserved quantity.

1.3.8 Outline of proof of main results

The proofs of our main results are given in §5. Here we provide a high level synopsis of our arguments.

Step 1. The first step will be to show that versions of Theorems 1.27, 1.28, 1.31 hold if we replace the configuration space $\mathcal{M}$ with its cylindrical approximation $\mathcal{M}_{\text{cyl}}$. (More precisely, see Theorem 5.1.)

Consider a point particle moving freely in $\mathcal{M}_{\text{cyl}}$ and reflecting specularly from $\partial\mathcal{M}_{\text{cyl}}$. The boundary of $\mathcal{M}_{\text{cyl}}$ lies just below the plane $\mathbf{P}$, within ϵ distance from the plane. In analogy to the collision law defined for the space $\mathcal{M}$, we define the *cylindrical collision law* $K_{\text{cyl}}^{\Sigma,\epsilon} : \mathbf{P} \times \mathbb{S}_+^2 \to \mathbf{P} \times \mathbb{S}_+^2$ by

$$K_{\text{cyl}}^{\Sigma,\epsilon} : (y, w) \mapsto (y', w,'), \tag{1.73}$$

where (y', w') is the state of the freely moving point particle upon its first return

to $\mathbf{P}$, after reflecting from $\partial\mathcal{M}_{\text{cyl}}$ some number of times. We then consider Markov kernels $\mathbb{K}$ on $\mathbf{P}\times\mathbb{S}^2_+$ such that for some sequence of cells $\{\Sigma_i\}$ and positive numbers $\epsilon_i\to 0$,

$$\delta_{K^{\Sigma_i,\epsilon_i}_{\text{cyl}}(y,w)}(\mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w)\to\mathbb{K}(y,w;\mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w), \tag{1.74}$$

weakly in the space of measures on $(\mathbf{P}\times\mathbb{S}^2_+)^2$.

To prove that such a Markov kernel $\mathbb{K}$ takes the form (1.69), the key observation is that the billiard trajectory $y(t)$ in $\mathcal{M}_{\text{cyl}}$ decouples into two independent evolutions: $y(t)=(y_{1,2}(t),y_3(t))$, where the evolution $y_3(t)$ is the projection of $y(t)$ onto the cylindrical axis χ, and the evolution $y_{1,2}(t)$ is the projection of $y(t)$ onto the orthogonal complement of χ, $\mathbf{Q}_1:=\{y\in\mathbb{R}^3:\langle y,\chi\rangle=0\}$. By virtue of the cylindrical structure of $\mathcal{M}_{\text{cyl}}$, $y_3(t)$ evolves linearly, with constant velocity for all time, while $y_{1,2}(t)$ follows the trajectory of a point particle moving freely in $\mathbf{Q}_1\setminus\mathcal{M}_{\text{cyl}}$ and reflecting specularly from the boundary $\mathbf{Q}_1\cap\partial\mathcal{M}_{\text{cyl}}$. After appropriately identifying $\mathbf{Q}_1$ with $\mathbb{R}^2$, this billiard domain in $\mathbf{Q}_1$ may be shown to coincide with the complement of the foreshortened wall $\widetilde{W}_i=W(\widetilde{\Sigma}_i,\widetilde{\epsilon}_i)$. This accounts for the non-trivial factor $\widetilde{\mathbb{P}}$ in (1.69).

On the other hand, if $\mathbb{K}$ is a Markov kernel on $\mathbf{P}\times\mathbb{S}^2_+$ of form (1.69), then one may show that for some sequence of cells $\{\Sigma_i\}$ and positive numbers $\epsilon_i\to 0$, a limit of form (1.74) holds. The idea is to apply the characterization of rough reflection laws in the upper half-plane given by Theorem 1.9 to show that there exist Σ_i and ϵ_i such that the limiting rough reflection law on $\widetilde{W}_i=W(\widetilde{\Sigma}_i,\widetilde{\epsilon}_i)$ is $\widetilde{\mathbb{P}}$ in (1.69).

For more details on the cylindrical configuration space and the cylindrical collision law, see §4.2 and §4.3.2. The argument in this step is presented in §5.1.

Step 2. Most of the work involved with proving our main results is concerned with the case of pure scaling – that is, the case where the sequence of cells $\Sigma_i = \Sigma$ is constant. We will prove that, for any $f, g \in C_c^\infty(\mathbf{P} \times \mathbb{S}_+^2)$,

$$\int_{\mathbf{P}\times\mathbb{S}_+^2} g(y,w)[f \circ K^{\Sigma,\epsilon}(y,w) - f \circ K_{\text{cyl}}^{\Sigma,\epsilon}(y,w)]\mathrm{d}\Lambda^2(\mathrm{d}y\mathrm{d}w) \to 0 \quad \text{as } \epsilon \to 0. \tag{1.75}$$

This is the statement of Lemma 5.2. The above limit together with the previous step can be used to prove Theorem 1.27.

To obtain (1.75), we introduce a *modified collision law* $\widetilde{K}^{\Sigma,\epsilon}$ to which we may compare both $K^{\Sigma,\epsilon}$ and $K_{\text{cyl}}^{\Sigma,\epsilon}$. The modified collision law is defined on a large (but not full measure) open subset $\widetilde{\mathcal{F}} \subset \mathbf{P} \times \mathbb{S}_+^2$, and satisfies

$$\widetilde{K}^{\Sigma,\epsilon} = \eta^{-1} \circ K^{\Sigma,\epsilon} \circ \eta, \tag{1.76}$$

where $\eta : \widetilde{\mathcal{F}} \to \widetilde{\mathcal{F}}$ is a smooth perturbation which "corrects" for the large-scale spherical shape of the body D. The map η is defined in §4.3.3. We will see that η converges to Id in $C^1(\widetilde{\mathcal{F}})$ as $\epsilon \to 0$. The comparison between $\widetilde{K}^{\Sigma,\epsilon}$ and $K^{\Sigma,\epsilon}$ will be carried out in §5.2 by making estimates on the differential of η.

The comparison between $\widetilde{K}^{\Sigma,\epsilon}$ and $K_{\text{cyl}}^{\Sigma,\epsilon}$ will be carried out in §5.3 via a "zooming argument." By double-periodicity of the configuration spaces $\mathcal{M}$ and $\mathcal{M}_{\text{cyl}}$, it is enough to compare the behavior of $\widetilde{K}^{\Sigma,\epsilon}$ and $K_{\text{cyl}}^{\Sigma,\epsilon}$ on a single parallelogram

$$R_\epsilon = \{(x_1, \alpha) \in \mathbf{P} : 0 \le x_1 + \alpha \le \epsilon, -\rho/2 \le x_2 \le \rho/2\}. \tag{1.77}$$

We will see that there is a large subset $\Omega \subset \mathbf{P} \times \mathbb{S}^2_+$ such that

$$\sup\left\{\|\widetilde{K}^{\Sigma,\epsilon}(y,w) - K^{\Sigma,\epsilon}_{\mathrm{cyl}}(y,w)\| : (y,w) \in \Omega \cap (R_\epsilon \times \mathbb{S}^2_+)\right\} \to 0, \text{ as } \epsilon \to 0. \tag{1.78}$$

This is the content of Lemma 5.3. The comparison is best carried out in "zoomed" coordinates. That is, we let

$$\sigma_{\epsilon^{-1}}(y,w) = (\epsilon^{-1}y, w), \qquad (y,w) \in \mathbf{P} \times \mathbb{S}^2_+, \tag{1.79}$$

and we compare the two maps

$$\widetilde{K}^* := \sigma_{\epsilon^{-1}} \circ \widetilde{K}^{\Sigma,\epsilon} \circ \sigma_\epsilon, \qquad K^*_{\mathrm{cyl}} := \sigma_{\epsilon^{-1}} \circ K^{\Sigma,\epsilon}_{\mathrm{cyl}} \circ \sigma_\epsilon. \tag{1.80}$$

The advantage of this point of view can be seen by observing that the scaled cylindrical configuration space $\mathcal{M}^*_{\mathrm{cyl}} := \epsilon^{-1}\mathcal{M}_{\mathrm{cyl}}$ is simply the cylinder with base $\overline{W(\Sigma,1)^c} + e_2$ and axis χ. Thus $\mathcal{M}^*_{\mathrm{cyl}}$ does not depend on ϵ. This will allow us to control the billiard trajectories in the zoomed spaces $\mathcal{M}^* := \epsilon^{-1}\mathcal{M}$ and $\mathcal{M}^*_{\mathrm{cyl}}$ in terms of properties of their "projections" onto the fixed cylindrical base $\overline{W(\Sigma,1)^c} + e_2$. This idea is fleshed out in Lemmas 5.7 and 5.8.

Step 3. Finally, to prove Theorems 1.28 and 1.31, we will apply the previous two steps and take advantage of the fact that the convergence (1.57) comes from a pseudometric $d^{\Lambda^2}_{\mathcal{G}}$ on the space of Markov kernels on $\mathbf{P} \times \mathbb{S}^2_+$. (For the definition of

this pseudometric, see §6.2.3.) For any Σ_i and ϵ_i, the triangle inequality gives us

$$d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}, \mathbb{K}^{\Sigma_i,\epsilon_i}) \leq d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}, \mathbb{K}_{\mathrm{cyl}}^{\Sigma_i,\epsilon_i}) + d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}_{\mathrm{cyl}}^{\Sigma_i,\epsilon_i}, \mathbb{K}^{\Sigma_i,\epsilon_i}), \tag{1.81}$$

where $\mathbb{K}_{\mathrm{cyl}}^{\Sigma_i,\epsilon_i}(y, w; \mathrm{d}y'\mathrm{d}w') := \delta_{K_{\mathrm{cyl}}^{\Sigma_i,\epsilon_i}(y,w)}(\mathrm{d}y'\mathrm{d}w')$. The convergence (1.75) implies that there exist positive numbers $b_i^{(0)}$ such that if $\epsilon_i \leq b_i^{(0)}$ for all i, then the second term in the right-hand side of (1.81) is negligible. Consequently, Theorems 1.28 and 1.31 will follow by applying the corresponding results obtained for the cylindrical configuration space in Step 1.

2 Examples

Here we construct examples of rough reflection laws and rough collision laws. Most of the work goes into building rough reflection laws. For each rough reflection law we construct, the correspondence (1.72) gives us a rough collision law "for free."

2.1 Lemma for constructing rough reflections

Throughout this section, we assume that the cells Σ satisfy conditions B1-B5 of §1.2.4.

Fix a cell Σ and a positive number $\epsilon > 0$. Suppose that $(x, \theta) \in \mathbb{R} \times \mathbb{S}^1_+$, and let (x', θ') be the random variable in $\mathbb{R} \times \mathbb{S}^1_+$ whose law is given by $\mathbb{P}^{\Sigma,\epsilon}(x, \theta, \mathrm{d}x'\mathrm{d}\theta')$. We define $\widetilde{\mathbb{P}}^{\Sigma,\epsilon}(\theta, \mathrm{d}\theta')$ to be the law of θ' as x varies uniformly in the period $[0, \epsilon]$ and θ stays fixed. In other words, for any $f \in C_c(\mathbb{S}^1_+)$,

$$\int_{\mathbb{S}^1_+} f(\theta')\widetilde{\mathbb{P}}^{\Sigma,\epsilon}(\theta, \mathrm{d}\theta') := \frac{1}{\epsilon}\int_0^\epsilon \int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma,\epsilon}(x, \theta, \mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x. \tag{2.1}$$

The following lemma gives us a way to construct rough reflection laws from a periodic microstructure.

Lemma 2.1. *(i) If the limit*

$$\lim_{i\to\infty} \mathbb{P}^{\Sigma_i,\epsilon_i}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') \tag{2.2}$$

exists, then the limit

$$\lim_{i\to\infty} \widetilde{\mathbb{P}}^{\Sigma_i,\epsilon_i}(\theta, \mathrm{d}\theta') \tag{2.3}$$

exists.

(ii) If the limit

$$\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') := \lim_{i\to\infty} \widetilde{\mathbb{P}}^{\Sigma_i,\epsilon_i}(\theta, \mathrm{d}\theta') \tag{2.4}$$

exists, then the limit

$$\lim_{i\to\infty} \mathbb{P}^{\Sigma_i,\epsilon_i}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') \tag{2.5}$$

exists and is equal to $\delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$*, where "equal" means the two Markov kernels belong to the same equivalence class (see Remark 1.6).*

(iii) Suppose that $\Sigma_i = \Sigma$ *is constant. Then for any sequence* $\epsilon_i \to 0$*, the limit* $\lim_{i\to\infty} \mathbb{P}^{\Sigma,\epsilon_i}(x, \theta, \mathrm{d}x'\mathrm{d}\theta')$ *exists and is equal to* $\delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$*, where*

$$\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') = \widetilde{\mathbb{P}}^{\Sigma,1}(\theta, \mathrm{d}\theta'). \tag{2.6}$$

The lemma is proved in §2.4. The benefit of the lemma is that the law $\widetilde{\mathbb{P}}^{\Sigma_i,\epsilon_i}(\theta, \mathrm{d}\theta')$ is usually much easier to compute than $\mathbb{P}^{\Sigma_i,\epsilon_i}(x, \theta; \mathrm{d}x'\mathrm{d}\theta')$, owing to the fact that in the former case we do not have to worry about the spatial variable x'.

2.2 Examples of rough reflection laws

2.2.1 Rectangular teeth

First we consider a microstructure of "rectangular teeth." That is, we define real functions

$$t_n(x) = \begin{cases} 0 & \text{if } 2k\epsilon_n \le x \le (2k+1)\epsilon_n, \\ -r\epsilon_n & \text{if } (2k+1)\epsilon_n < x < (2k+2)\epsilon_n \end{cases} \qquad \text{for } k \in \mathbb{Z}. \tag{2.7}$$

The quantity $r > 0$ is a fixed parameter representing the ratio of the height of the teeth to the width. Define ϵ_n-periodic walls

$$W_n = W(\Sigma, \epsilon_n) = \{(x_1, x_2) : x_2 \le t_n(x_1)\}, \tag{2.8}$$

To find the macro-reflection law, let X be uniform in $[0,1]$, and let $\theta \in (0,\pi)$ be fixed. Let $(X'_\theta, \Theta'_\theta) = P^{\Sigma,1}(X,\theta)$. By Lemma 2.1(iii), the limiting reflection law $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law of the random variable Θ'_θ.

As illustrated in Figure 6, Θ'_θ is equal to either $\pi - \theta$ (specular reflection) or θ (retroreflection). With probability 1/2 the starting point X of the point particle is on top of a tooth, in which case $\Theta'_\theta = \pi - \theta$. Otherwise, X will be in the interval above the crevice between two teeth.

Conditioned on the latter event, the probability that $\Theta'_\theta = \pi - \theta$ may be determined by "unfolding" the rectangular billiard between the two teeth, as shown in Figure 7. By inspecting this figure, we see that if $\lfloor 2r|\cot\theta| \rfloor$ is even, then the probability

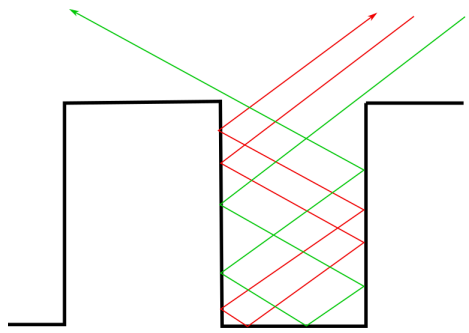

Figure 6: Reflection from rectangular teeth. The post-reflection angle will be either $\theta' = \pi - \theta$ (specular reflection – green) or $\theta' = \theta$ (retroreflection – red).

of specular reflection is $\{2r|\cot\theta|\}$. If on the other hand $\lfloor 2r|\cot\theta|\rfloor$ is odd, then the probability of specular reflection is $1 - \{2r|\cot\theta|\}$.

Putting these observations together, we conclude that

$$\Theta'_\theta = \begin{cases} \pi - \theta & \text{w.p. } p_r(\theta), \\ \theta & \text{w.p. } 1 - p_r(\theta), \end{cases} \tag{2.9}$$

where

$$p_r(\theta) = \begin{cases} \frac{1}{2} + \frac{1}{2}\{2r|\cot\theta|\} & \text{if } \lfloor 2r|\cot\theta|\rfloor \text{ is even}, \\ 1 - \frac{1}{2}\{2r|\cot\theta|\} & \text{if } \lfloor 2r|\cot\theta|\rfloor \text{ is odd}. \end{cases} \tag{2.10}$$

Thus the limiting rough reflection law is $\mathbb{P}(x, \theta, \mathrm{d}x'\mathrm{d}\theta') = \delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$, where $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law of Θ'_θ.

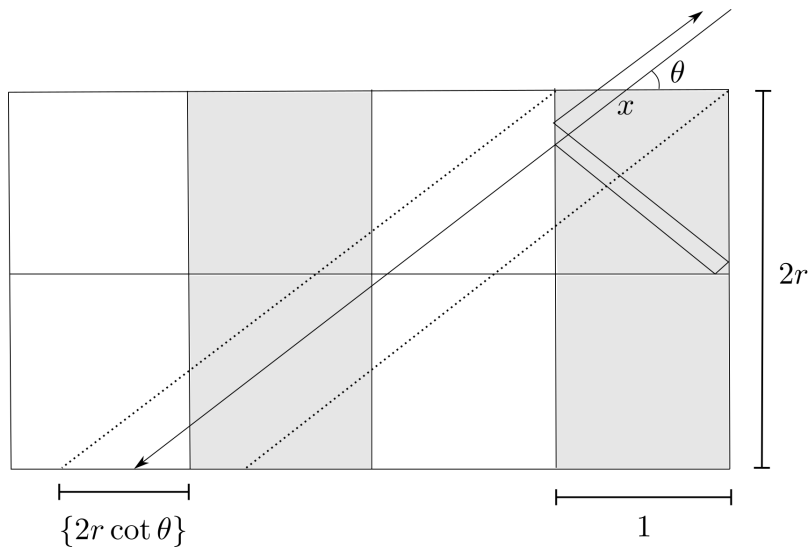

Figure 7: If the end of the trajectory of the point particle in the unfolded rectangular billiard falls in a gray region, then the reflection will be specular. Otherwise it will be a retroreflection.

2.2.2 Triangular teeth

We can similarly construct reflections from triangular teeth. Let $\psi \in (0, \pi)$. Define a "tooth function"

$$t_n(x) = \begin{cases} -\cot\left(\frac{\psi}{2}\right)x & \epsilon_n k \le x < \epsilon_n\left(\frac{2k+1}{2}\right), \\ \cot\left(\frac{\psi}{2}\right)(x-1) & \epsilon_n\left(\frac{2k+1}{2}\right) \le x < \epsilon_n(k+1) \end{cases} \qquad \text{for } k \in \mathbb{Z}. \tag{2.11}$$

Then define ϵ_n-periodic walls

$$W_n = W(\Sigma_n, \epsilon_n) = \{(x_1, x_2) : x_2 \le t_n(x_1)\}. \tag{2.12}$$

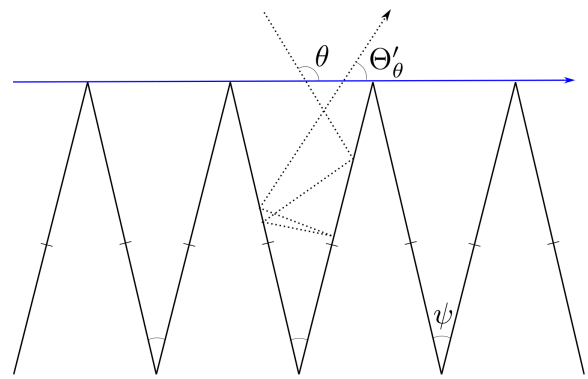

Figure 8: Triangular teeth

Each period of the wall is an isosceles triangle such that each peak and valley spans an angle of ψ. See Figure 8.

As before, let X be uniform in $[0,1]$, let $\theta \in (0,\pi)$ be fixed, and let $(X'_\theta, \Theta'_\theta) = P^{\Sigma,1}(X,\theta)$. The limiting reflection law $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law of Θ'_θ.

Similarly to the case of rectangular teeth, we can deduce the distribution of Θ'_θ by considering the unfolding of the region between two triangular teeth. See Figure 9. In the figure, the number of triangles crossed by the top of the horizontal strip is

$$N = \left\lceil \frac{2\theta}{\psi} \right\rceil. \tag{2.13}$$

The angle of the outgoing trajectory depends on whether N is even or odd and whether the unfolded trajectory exits the unfolded billiard in a white region or a gray region. The trajectory leaves in a white region with probability $p_\psi(\theta)$ and leaves in gray region with probability $1 - p_\psi(\theta)$.

It is tedious but elementary to compute the angle of exit in each case, and the

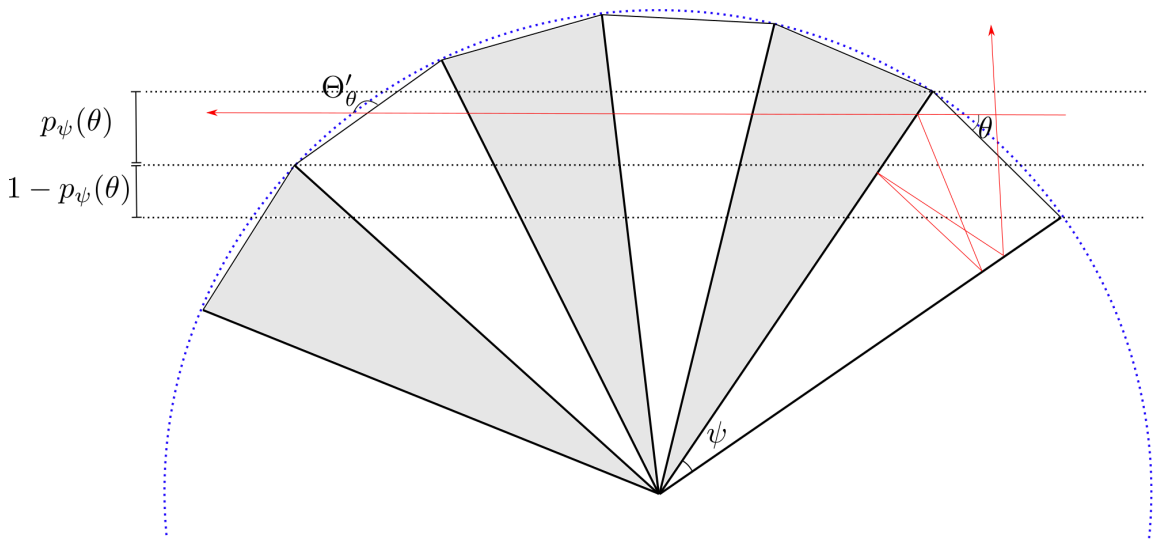

Figure 9: Unfolding of the billiard between two teeth.

probability $p_\psi(\theta)$. One obtains that

$$\begin{aligned} &\text{if } \left\lceil \frac{2\theta}{\psi} \right\rceil \text{ is even,} \quad \Theta'_\theta = \begin{cases} \pi + \theta - \psi\lceil \frac{2\theta}{\psi} \rceil & \text{w.p. } p_\psi(\theta), \\ \psi\left(\lceil \frac{2\theta}{\psi} \rceil + 1\right) - \theta & \text{w.p. } 1 - p_\psi(\theta), \end{cases} \\ &\text{and if } \left\lceil \frac{2\theta}{\psi} \right\rceil \text{ is odd,} \quad \Theta'_\theta = \begin{cases} \psi\lceil \frac{2\theta}{\psi} \rceil - \theta & \text{w.p. } p_\psi(\theta), \\ \pi + \theta - \psi\left(\lceil \frac{2\theta}{\psi} \rceil + 1\right) - \theta & \text{w.p. } 1 - p_\psi(\theta), \end{cases} \end{aligned} \tag{2.14}$$

where

$$p_\psi(\theta) = \frac{\left[\cos\left(\theta - \frac{\psi}{2}\right) - \cos\left(\frac{\psi}{2}\lceil \frac{2\theta}{\psi} \rceil\right)\right]_+}{\cos\left(\theta - \frac{\psi}{2}\right) - \cos\left(\theta + \frac{\psi}{2}\right)}. \tag{2.15}$$

Here $[x]_+ = x$ if $x \geq 0$ and zero otherwise. The limiting reflection law is $\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$, where $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law of Θ'_θ.

2.2.3 Focusing circular arcs

The discrete distribution of both of the previous reflection laws is an artifact of the polygonal boundary of the walls. When the boundary of the wall contains curve segments with non-zero curvature, we expect the distribution to be non-singular in general.

As an example of this, consider the wall whose periods consist of focusing (i.e. concave-up) circular arcs. Let $0 < \xi \leq \pi/2$. We define

$$t_n(x) = \frac{\epsilon_n}{2}\cot\xi - \sqrt{\frac{\epsilon_n^2}{4}\csc^2\xi - \left(x - \frac{\epsilon_n}{2}\right)^2} \qquad \text{for } x \in [0, \epsilon_n], \tag{2.16}$$

and extend t_n to be ϵ_n-periodic on the line $\mathbb{R}$. The wall is

$$W_n = W(\Sigma, \epsilon_n) = \{(x_1, x_2) : x_2 \leq t_n(x_1)\}. \tag{2.17}$$

The arc forming one period of the wall spans an angle of 2ξ.

Consider the wall $W(\Sigma, 1)$, depicted in Figure 10. Let X be uniform in $[0, 1]$, and let $\theta \in (0, \pi)$ be fixed. Let $(X'_\theta, \Theta'_\theta) = P^{\Sigma,1}(X, \theta)$. Let C denote the center of the circle whose arc forms the period of the wall below $[0, 1]$, and let R denote the radius. In coordinates,

$$C = \left(\frac{1}{2}, \frac{1}{2}\cot\xi\right), \qquad R = \frac{1}{2}\csc\xi. \tag{2.18}$$

Let P be the first point in the circular arc hit by the billiard trajectory starting from (X, θ), and let γ be the signed angle measured counterclockwise from the vector $-e_2$

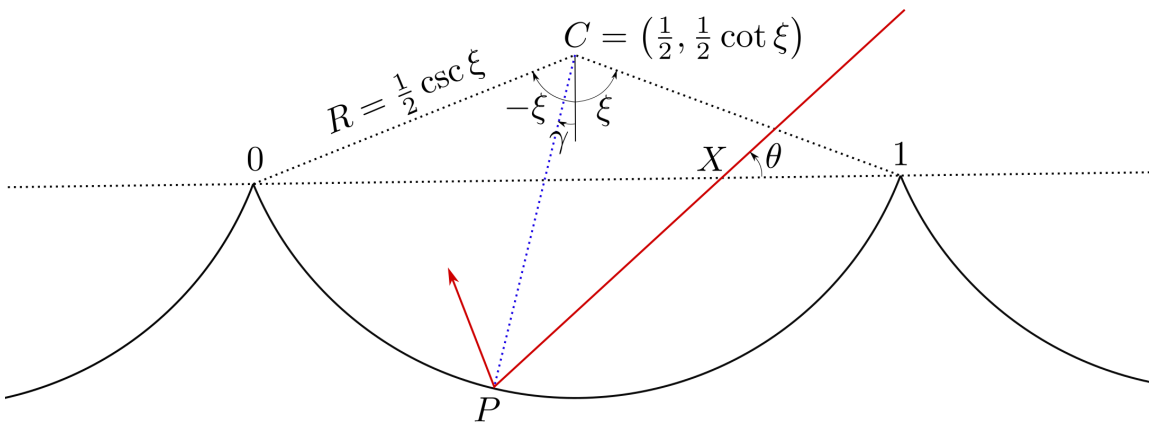

Figure 10: A wall formed from circular arcs.

to the ray $\overrightarrow{CP}$. (Thus if $X = 0$ then $\gamma = -\xi$ and if $X = 1$ then $\gamma = \xi$.)

It turns out to be more convenient to express Θ'_θ in terms of γ, rather than X.

Let N denote the number of times that the point particle hits the circular arc before leaving the wall. Referring to Figure 11, let $P_0 = P$, and suppose that the point particle subsequently hits the wall at points $P_1, P_2, \ldots, P_{N-1}$. The sequence of points P_i will progress along the arc in the counterclockwise direction if $\theta \geq \frac{\pi}{2} + \gamma$, and in the clockwise direction if $\theta < \frac{\pi}{2} + \gamma$ (strictly speaking, the case where $\theta = \frac{\pi}{2} + \gamma$ is vacuous since the trajectory will hit the boundary only once). The signed angle from P_{i-1} to P_i is

$$\Delta\gamma = \begin{cases} 2(\pi + \gamma - \theta) & \text{if } \theta \geq \frac{\pi}{2} + \gamma, \\ 2(\gamma - \theta) & \text{if } \theta < \frac{\pi}{2} + \gamma. \end{cases} \tag{2.19}$$

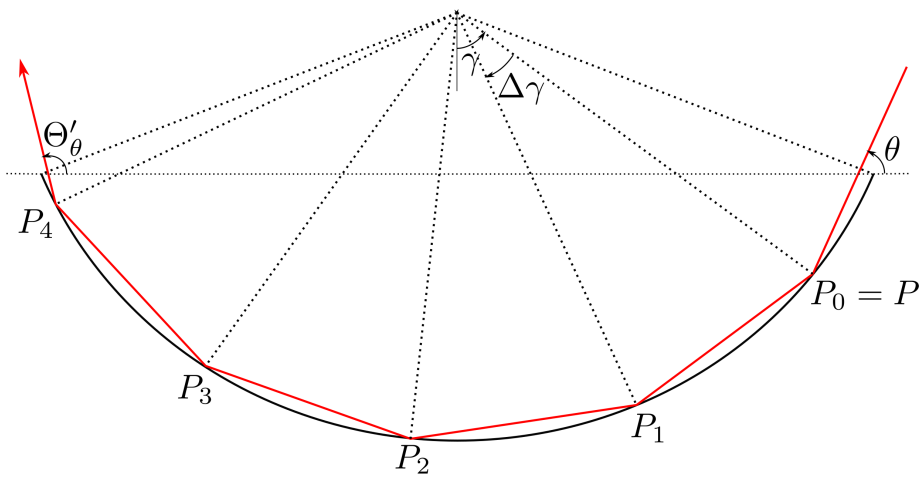

Figure 11: Multiple reflections from a circular arc.

The number of times which the billiard trajectory hits the arc is

$$N = \begin{cases} \lceil \frac{\xi-\gamma}{|\Delta\gamma|} \rceil & \text{if } \theta \geq \frac{\pi}{2} + \gamma, \\ \lceil \frac{\xi+\gamma}{|\Delta\gamma|} \rceil & \text{if } \theta < \frac{\pi}{2} + \gamma. \end{cases} \tag{2.20}$$

Each time the point particle hits the circular arc, the angle which the billiard trajectory makes with the $+x_1$-axis increments by $\Delta\gamma$. Thus,

$$\Theta'_\theta = \theta + N\Delta\gamma = \begin{cases} \theta + \lceil \frac{\xi-\gamma}{2|\pi+\gamma-\theta|} \rceil 2(\pi + \gamma - \theta) & \text{if } \theta \geq \frac{\pi}{2} + \gamma, \\ \theta + \lceil \frac{\xi+\gamma}{2|\gamma-\theta|} \rceil 2(\gamma - \theta) & \text{if } \theta < \frac{\pi}{2} + \gamma. \end{cases} \tag{2.21}$$

We will now find an explicit formula for γ in terms of X and θ. This, together with (2.21), will give us the distribution of Θ'_θ.

An elementary calculation shows that

$$\begin{aligned} X &= \frac{\cot\theta\cos\gamma}{2\sin\xi} + \frac{\sin\gamma}{2\sin\xi} + \frac{1}{2} - \frac{1}{2}\cot\theta\cot\xi \\ &= \frac{\cos(\theta+\gamma) - \cos(\theta+\xi)}{\cos(\theta-\xi) - \cos(\theta+\xi)}. \end{aligned} \tag{2.22}$$

Geometrically, it is clear that X is uniquely determined by γ and θ and that X should increase with γ. Note that $-\pi/2 \le -\xi \le \theta + \gamma < \pi + \xi \le 3\pi/2$. On the interval $[-\pi/2, 3\pi/2)$, the cosine function increases on the intervals $[-\pi/2, 0]$ and $[\pi, 3\pi/2)$. Thus $\theta + \gamma \in [-\pi/2, 0] \cup [\pi, 3\pi/2)$. Thus if we define

$$\operatorname{Arccos}(x) = \begin{cases} 2\pi - \arccos(x) & \text{if } -1 \le x < 0, \\ -\arccos(x) & \text{if } 0 \le x \le 1, \end{cases} \tag{2.23}$$

then $\operatorname{Arccos}(\cos(\theta+\gamma)) = \theta + \gamma$. Hence, solving (2.22) gives us

$$\gamma = -\theta + \operatorname{Arccos}\left(X\cos(\theta-\xi) + (1-X)\cos(\theta+\xi)\right). \tag{2.24}$$

2.2.4 Retroreflection

Consider the retroreflection law

$$\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \delta_{(x,\theta)}(\mathrm{d}x'\mathrm{d}\theta'). \tag{2.25}$$

It follows from Theorem 1.9 that there is a sequence of walls W_i such that $\mathbb{P}^{W_i} \to \mathbb{P}$ as $i \to \infty$.

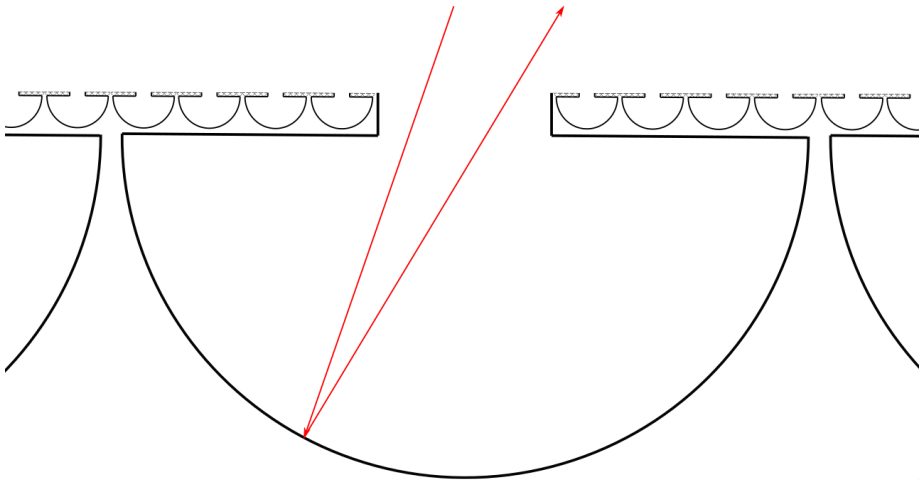

Figure 12: Retroreflection from a fractal of mushroom billiards.

We obtain a sequence of walls which generates retroreflection by considering the fractal of mushroom billiards depicted in Figure 12. The "cap" of each mushroom is a semicircular arc. The walls consist of partial iterations of the mushroom fractal. To each W_i is added an additional row of "mushroom hallows," while the scale of the wall converges to zero as $u \to \infty$. The ratio of the width of the stem to the width of the cap of each mushroom is taken to converge to zero as $i \to \infty$, but at not too fast a rate. The horizontal segments at the top the boundary of W_i form a partial iteration of a Cantor set. Provided the width of the stems does not converge to zero too quickly, this set will become negligible as $i \to \infty$.

For additional examples and discussion of walls which generate retroreflection, see Chapter 9 of [35].

2.3 Examples of rough collision laws

Given a sequence of periodic walls, $W_i = W(\Sigma, \epsilon_i)$, let $\widetilde{W}_i = W(\widetilde{\Sigma}, \widetilde{\epsilon}_i)$ denote the foreshortened wall (where $\widetilde{\Sigma}$ and $\widetilde{\epsilon}_i$ are defined by (1.70)). Let X be uniform in $[0,1]$, and let

$$(X'_\theta, \Theta'_\theta) = P^{\widetilde{\Sigma},1}(X, \theta), \tag{2.26}$$

The angle Θ'_θ is the angle of exit of the trajectory hitting the foreshortened wall, as X varies uniformly over one period. Theorem 1.31 tells us that the limiting collision law for the disk and wall system $\mathbb{K} = \lim_{i\to\infty} \mathbb{K}^{W_i}$ is given by

$$\mathbb{K}(y_1, y_3, \theta, \psi; \mathrm{d}y'_1 \mathrm{d}y'_3 \mathrm{d}\theta' \mathrm{d}\xi') = \delta_{(y_1,y_3)}(\mathrm{d}y'_1 \mathrm{d}y'_3)\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')\delta_{\pi-\psi}(\mathrm{d}\psi'), \tag{2.27}$$

where $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law of the random angle of exit Θ'_θ.

We obtain examples of rough collision laws by choosing walls W_i, such that the limiting reflection law for the sequence of foreshortened walls $\widetilde{W}_i$ is known.

2.3.1 Rectangular teeth

Consider walls $W_i = W(\Sigma, \epsilon_i)$ consisting of rectangular teeth with parameter r (the ratio of the width of the teeth to the height). For this choice of walls, we denote the angle of exit by $\Theta'_\theta(r)$ to indicate the dependence of Θ'_θ on r in the formula (2.9).

The class of walls with rectangular teeth is invariant under foreshortening. After foreshortening by a factor $(1 + m/J)^{1/2}$, the new wall $\widetilde{W}_i$ is composed of rectangular

teeth with parameter

$$\widetilde{r} = (1 + m/J)^{-1/2} r. \tag{2.28}$$

Thus the rough collision law $\mathbb{K} = \lim_{i\to\infty} \mathbb{K}^{W_i}$ is given by (2.27), where $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law of $\Theta'_\theta(\widetilde{r})$.

Using (1.63) to convert back to the original coordinates, we obtain that the collision law $\mathbb{K}$ maps the incoming velocity $w = (v_1, v_2, \omega)$ to $A_{\text{smooth}} w$ with probability $p_{\widetilde{r}}(\theta)$ and to $A_{\text{no-slip}} w$ with probability $1 - p_{\widetilde{r}}(\theta)$, where A_{smooth} and $A_{\text{no-slip}}$ are the matrices defined as follows:

$$A_{\text{smooth}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad A_{\text{no-slip}} = \begin{bmatrix} \frac{m-J}{m+J} & 0 & \frac{-2J}{m+J} \\ 0 & -1 & 0 \\ \frac{-2m}{m+J} & 0 & \frac{-m+J}{m+J} \end{bmatrix}. \tag{2.29}$$

The significance of the notation A_{smooth} and $A_{\text{no-slip}}$ is explained in §2.3.4.

2.3.2 Triangular teeth

Consider walls $W_i = W(\Sigma, \epsilon_i)$, consisting of triangular teeth with parameter ψ. We denote the random angle or reflection by $\Theta'_\theta(\psi)$, to indicate the dependence of Θ'_θ on ψ in the formula (2.14)

For each i, the foreshortened wall $\widetilde{W_i}$ consists of triangular teeth with parameter

$$\widetilde{\psi} = 2 \arctan\left(\left(1 + \frac{m}{J}\right)^{-1/2} \tan\left(\frac{\psi}{2}\right)\right). \tag{2.30}$$

The rough collision law $\mathbb{K} = \lim_{i\to\infty} \mathbb{K}^{W_i}$ is given by (2.27), where $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law

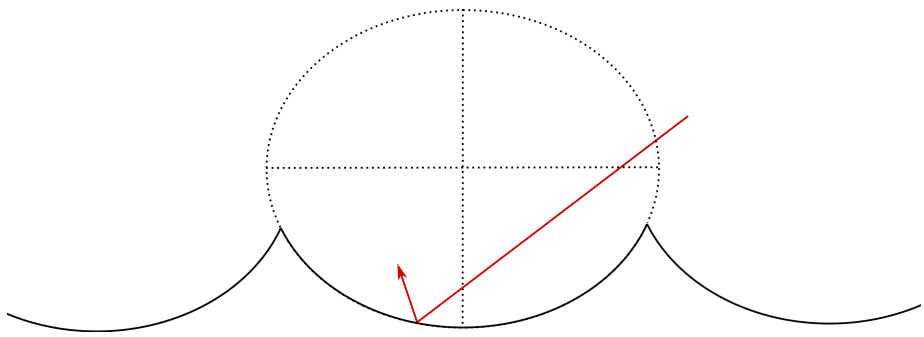

Figure 13: Focusing elliptical arcs

of $\Theta'_\theta(\widetilde{\psi})$.

In principle, one can use (1.63) to express $\mathbb{K}$ in the original coordinates $(x_1, x_2, \alpha, v_1, v_2, \theta)$, although the formula is not as simple as in the previous case.

2.3.3 Focusing elliptical arcs

Consider walls $W_i = W(\Sigma, \epsilon_i)$ constructed of focusing (concave-up) elliptical arcs, where the ratio of the horizontal to vertical axes is $(1 + m/J)^{1/2}$. In other words, we define

$$t_n(x) = \frac{\epsilon_n}{2}\cot\xi - (1+m/J)^{-1/2}\sqrt{\frac{\epsilon_n^2}{4}\csc^2\xi - \left(x - \frac{\epsilon_n}{2}\right)^2} \qquad \text{for } x \in [0, \epsilon_n], \tag{2.31}$$

and extend t_n to be periodic on $\mathbb{R}$. We define

$$W_n = W(\Sigma, \epsilon_n) = \{(x_1, x_2) : x_2 \le t_n(x_1)\}. \tag{2.32}$$

See Figure 13

The foreshortened wall $\widetilde{W}_i$ is precisely the wall of circular arcs with parameter

ξ, discussed in §2.2.3. Therefore, the rough collision law $\mathbb{K} = \lim_{i\to\infty} \mathbb{K}^{W_i}$ is given by (2.27), where $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is the law of $\Theta'_\theta = \Theta'_\theta(\xi)$, given by the formula (2.21).

2.3.4 Smooth and no-slip collisions

As we have seen, there are two basic examples of deterministic rough reflection laws: specular reflection

$$\mathbb{P}_{\text{spec}}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') := \delta_{(x,\pi-\theta)}(\mathrm{d}x'\mathrm{d}\theta') \tag{2.33}$$

and retroreflection

$$\mathbb{P}_{\text{retro}}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') := \delta_{(x,\theta)}(\mathrm{d}x'\mathrm{d}\theta'). \tag{2.34}$$

Trivially, $\mathbb{P}_{\text{spec}}$ is the limit of the reflection laws on the constant sequence of flat walls $W_i = \{(x_1, x_2) : x_2 \le 0\}$. Under the correspondence (1.72), $\mathbb{P}_{\text{spec}}$ corresponds to

$$\mathbb{K}_{\text{smooth}}(y_1, y_3, \theta, \psi; \mathrm{d}y_1'\mathrm{d}y_3'\mathrm{d}\theta'\mathrm{d}\psi') = \delta_{(y_1,y_3,\pi-\theta,\pi-\psi)}(\mathrm{d}y_1', \mathrm{d}y_3'\mathrm{d}\theta'\mathrm{d}\psi'). \tag{2.35}$$

This is just the classical collision law describing a frictionless collision between a hard disk and fixed wall in the plane.

Under the correspondence (1.72), $\mathbb{P}_{\text{retro}}$ corresponds to

$$\mathbb{K}_{\text{no-slip}}(y_1, y_3, \theta, \psi; \mathrm{d}y_1'\mathrm{d}y_3'\mathrm{d}\theta'\mathrm{d}\psi') = \delta_{(y_1,y_3,\theta,\pi-\psi)}(\mathrm{d}y_1', \mathrm{d}y_3'\mathrm{d}\theta'\mathrm{d}\psi'). \tag{2.36}$$

This type of deterministic collision is known as a *no-slip collision*.

Using (1.63) to convert back to our original coordinates $(x_1, x_2, \alpha, v_1, v_2, \beta)$, one may show that the collision law $\mathbb{K}_{\text{smooth}}$ maps the velocity $w = (v_1, v_2, \omega)$ to $A_{\text{smooth}}w$,

and the collision law $\mathbb{K}_{\text{no-slip}}$ maps w to $A_{\text{no-slip}}w$, where A_{smooth} and $A_{\text{no-slip}}$ are the matrices defined by (2.29).

Let W_i be a sequence of walls such that $\mathbb{P}_{\text{spec}} = \lim_{i\to\infty} \mathbb{P}^{W_i}$ (for example, the walls of §2.2.4). Let $\widehat{W_i} = \{(x_1, x_2) : (x_1, (1 + m/J)^{-1/2}x_2) \in W_i\}$. That is, we choose walls $\widehat{W_i}$ whose foreshortening is W_i. Theorem 1.31 tells us that $\mathbb{K}_{\text{no-slip}} = \lim_{i\to\infty} \mathbb{K}^{\widehat{W_i}}$.

No-slip collisions were first introduced by Broomhead and Gutkin in [4] and have been further investigated Cox, Feres, and Ward [10, 9].

Remark 2.2. It turns out that $\mathbb{P}_{\text{spec}}$ and $\mathbb{P}_{\text{retro}}$ are the only two deterministic rough reflection laws which act continuously on $\mathbb{R}\times\mathbb{S}^1_+$. Thus, by the correspondence (1.72), $\mathbb{K}_{\text{smooth}}$ and $\mathbb{K}_{\text{no-slip}}$ are the only two rough collision laws which act continuously on $\mathbf{P} \times \mathbb{S}^2_+$. The latter statement is also closely related to Corollary 2.2 in [10], which classifies deterministic collision laws for more general rigid bodies.

To see why the first statement is true, let $\mu(\mathrm{d}\theta) = \sin\theta \mathrm{d}\theta$, and let m denote Lebesgue measure on the line. Recall that by Proposition 6.5, if $\mathbb{P}(x, \theta; \mathrm{d}x'\mathrm{d}\theta') = \delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ is a rough reflection law, then $\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$ preserves the measure μ. The measure space $((0, \pi), \mu)$ is isomorphic to $((-1, 1), m)$, via the mapping $f(\theta) = -\cos\theta$. The only continuous transformations $(-1, 1) \to (-1, 1)$ which preserve Lebesgue measure are $T(x) = x$ and $T(x) = -x$, and these transformations pull back via f to retroreflection and specular reflection respectively.

By considering transformations which are not continuous, one can obtain a much larger class of deterministic reflection laws and collision laws. For example, any involutive interval exchange transformation of $(-1, 1)$ corresponds to some deterministic reflection law and to some deterministic collision law.

2.4 Proof of the lemma for constructing random reflections

Proof of Lemma 2.1. For $(x,\theta) \in \mathbb{R} \times \mathbb{S}^1_+$, let $(x',\theta') = P^{\Sigma_i,\epsilon_i}(x,\theta)$. By condition B5 on the cells Σ_i (see §1.2.4), the boundary of the wall $W(\Sigma_i,\epsilon_i)$ is periodic and has maximums at integer multiples of ϵ_i along the line $\{(x_1,x_2) : x_2 = 0\}$. Moreover, if $x \in [k\epsilon_i,(k+1)\epsilon_i]$ for $k \in \mathbb{Z}$, then the billiard trajectory will remain in the hollow bounded between the two maximums and will return to line $x_2 = 0$ at $x' \in [k\epsilon_i,(k+1)\epsilon_i]$. Therefore,

$$|x - x'| \le \epsilon_i. \tag{2.37}$$

We will use this estimate several times in the argument below.

Proof of (i) Suppose that $\mathbb{P}(x,\theta;\mathrm{d}x'\mathrm{d}\theta') = \lim_{i\to\infty} \mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')$ exists. Let $f,g \in C_c(\mathbb{S}^1_+)$. We have

$$\begin{aligned}
&\int_{\mathbb{S}^1_+} g(\theta) \int_{\mathbb{S}^1_+} f(\theta')\widetilde{\mathbb{P}}^{\Sigma_i,\epsilon_i}(\theta,\mathrm{d}\theta')\sin\theta\mathrm{d}\theta \\
&= \int_{\mathbb{S}^1_+} g(\theta)\frac{1}{\epsilon_i}\int_0^{\epsilon_i}\int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x\sin\theta\mathrm{d}\theta \\
&= \int_{\mathbb{S}^1_+} g(\theta)\frac{1}{\epsilon_i\lceil 1/\epsilon_i\rceil}\int_0^{\epsilon_i\lceil 1/\epsilon_i\rceil}\int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x\sin\theta\mathrm{d}\theta,
\end{aligned} \tag{2.38}$$

where the last line uses the fact that $x \mapsto \int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')$ is ϵ_i-periodic. Noting that $\mathbb{P}^{\Sigma_i,\epsilon_i}$ is a probability measure, we also have the following bound:

$$\int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta') \le \|f\|_{L^\infty}. \tag{2.39}$$

Using this bound, we see that the last line in (2.38) is equal to

$$\int_{\mathbb{S}^1_+} g(\theta) \int_0^1 \int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x \sin\theta \mathrm{d}\theta + O(\epsilon_i), \tag{2.40}$$

where the error term may depend on f and g. Let $\phi_\delta \in C_c(\mathbb{R})$ be chosen so that $0 \le \phi_\delta \le 1$ and $\operatorname{supp}\phi_\delta \subset [0,1]$ and $\int |\phi_\delta - \mathbf{1}_{[0,1]}|\mathrm{d}x < \delta$. Let $\eta_\delta \in C_c(\mathbb{R})$ be chosen so that $0 \le \eta_\delta \le 1$ and $\eta_\delta = 1$ on the interval $[-\delta^{-1}, 1+\delta^{-1}]$. We see that (2.40) is equal to

$$\begin{aligned}
&\int_{\mathbb{R}\times\mathbb{S}^1_+} \phi_\delta(x) g(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) + O(\delta+\epsilon_i) \\
&= \int_{\mathbb{R}\times\mathbb{S}^1_+} \phi_\delta(x) g(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} \eta_\delta(x') f(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) + O(\delta+\epsilon_i),
\end{aligned} \tag{2.41}$$

noting that by the bound (2.37), $x' \in [-\delta^{-1}, 1+\delta^{-1}]$ (for i sufficiently large) whenever $x \in [0,1]$; thus $\eta_\delta(x') = 1$ whenever $\phi_\delta(x) \neq 0$. Taking the limit as $i \to \infty$, (2.41) converges to

$$\int_{\mathbb{R}\times\mathbb{S}^1_+} \phi_\delta(x) g(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} \eta_\delta(x') f(\theta')\mathbb{P}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) + O(\delta). \tag{2.42}$$

Letting $\delta \to 0$, by boundedness and pointwise convergence of ϕ_δ and η_δ, dominated convergence implies (2.42) converges to

$$\begin{aligned}
&\int_{[0,1]\times\mathbb{S}^1_+} g(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x \sin\theta\mathrm{d}\theta \\
&\quad = \int_{\mathbb{S}^1_+} g(\theta) \int_{\mathbb{S}^1_+} f(\theta')\widetilde{\mathbb{P}}_0(\theta,\mathrm{d}\theta')\sin\theta\mathrm{d}\theta,
\end{aligned} \tag{2.43}$$

where $\widetilde{\mathbb{P}}_0$ is the Markov kernel on $\mathbb{S}^1_+$ defined by

$$\int_{\mathbb{S}^1_+} h(\theta')\widetilde{\mathbb{P}}_0(\theta, \mathrm{d}\theta') = \int_{[0,1]} \int_{\mathbb{R}\times\mathbb{S}^1_+} h(\theta')\mathbb{P}(x,\theta;\mathrm{d}x'\mathrm{d}\theta'). \tag{2.44}$$

Putting all this together, we see that (2.38) converges to (2.43) as $i \to \infty$, as desired.

Proof of (ii). Conversely, suppose that $\widetilde{\mathbb{P}} = \lim_{i\to\infty} \widetilde{\mathbb{P}}^{\Sigma_i,\epsilon_i}$ exists. Given $h(x,\theta,x',\theta') = f(x,\theta)g(x',\theta')$, where $f, g \in C_c^\infty(\mathbb{R}\times\mathbb{S}^1_+)$, we want to show that a limit of form (1.11) holds. Since the tensor product $C_c^\infty(\mathbb{R}) \otimes C_c^\infty(\mathbb{S}^1_+)$ is dense in $C_c^\infty(\mathbb{R}\times\mathbb{S}^1_+)$, we may assume without loss of generality that $f(x,\theta) = f_1(x)f_2(\theta)$ and $g(x,\theta) = g_1(x)g_2(\theta)$ for some $f_1, g_1 \in C_c^\infty(\mathbb{R})$ and $f_2, g_2 \in C_c^\infty(\mathbb{S}^1_+)$. We have

$$\begin{aligned}
&\int_{\mathbb{R}\times\mathbb{S}^1_+} g_1(x)g_2(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} f_1(x')f_2(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \\
&= \int_{\mathbb{R}\times\mathbb{S}^1_+} g_1(x)g_2(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} f_1(x)f_2(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \\
&\quad + \int_{\mathbb{R}\times\mathbb{S}^2_+} g_1(x)g_2(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} [f_1(x') - f_1(x)]f_2(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \\
&\qquad\qquad + O(\delta^2) \\
&= \int_{\mathbb{R}\times\mathbb{S}^1_+} g_1(x)f_1(x)g_2(\theta) \int_{\mathbb{R}\times\mathbb{S}^1_+} f_2(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) + O(E_i),
\end{aligned} \tag{2.45}$$

where

$$E_i := \sup\{|f_1(x') - f_1(x)| : x \in \mathbb{R}\}. \tag{2.46}$$

By the bound (2.37) and the fact that f_1 is continuous and compactly supported, we see that E_i converges to zero as $i \to \infty$. By approximation by simple functions, we

may write

$$g_1(x)f_1(x) = \sum_{k\in\mathbb{Z}} g_1(\epsilon_i k)f_1(\epsilon_i k)\mathbf{1}_{[\epsilon_i k,\epsilon_i(k+1)]}(x) + h_i(x), \tag{2.47}$$

where $h_i \to 0$ in $L^1(\mathbb{R})$. Also, observing that the function

$$x \mapsto \int_{\mathbb{R}\times\mathbb{S}^1_+} f_2(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta') \tag{2.48}$$

is ϵ_i-periodic, we may write the last quantity in (2.45) as

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} g_1(\epsilon_i k)f_1(\epsilon_i k)\int_k^{k+\epsilon_i}\int_{\mathbb{S}^1_+} g_2(\theta)\int_{\mathbb{R}\times\mathbb{S}^1_+} f_2(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\sin\theta\mathrm{d}\theta\mathrm{d}x \\
&\qquad\qquad + O(\|h_i\|_{L^1} + E_i) \\
&= \left(\sum_{k\in\mathbb{Z}} g_1(\epsilon_i k)f_1(\epsilon_i k)\epsilon_i\right)\left(\int_{\mathbb{S}^1_+} g_2(\theta)\frac{1}{\epsilon_i}\int_0^{\epsilon_i}\int_{\mathbb{R}\times\mathbb{S}^1_+} f_2(\theta')\mathbb{P}^{\Sigma_i,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x\sin\theta\mathrm{d}\theta\right) \\
&\qquad\qquad + O(\|h_i\|_{L^1} + E_i). \\
&= \left(\sum_{k\in\mathbb{Z}} g_1(\epsilon_i k)f_1(\epsilon_i k)\epsilon_i\right)\left(\int_{\mathbb{S}^1_+} g_2(\theta)\int_{\mathbb{S}^1_+} f_2(\theta')\widetilde{\mathbb{P}}^{\Sigma_i,\epsilon_i}(\theta;\mathrm{d}\theta')\sin\theta\mathrm{d}\theta\right) \\
&\qquad\qquad + O(\|h_i\|_{L^1} + E_i).
\end{aligned} \tag{2.49}$$

Taking the limit first as $i \to \infty$, and noting that the first factor is a Riemann sum, the quantity above converges to

$$\begin{aligned}
&\left(\int_{\mathbb{R}} f_1(x)g_1(x)\mathrm{d}x\right)\left(\int_{\mathbb{S}^1_+} g_2(\theta)\int_{\mathbb{S}^1_+} f_2(\theta')\widetilde{\mathbb{P}}(\theta;\mathrm{d}\theta')\sin\theta\mathrm{d}\theta\right) \\
&= \int_{\mathbb{R}\times\mathbb{S}^1_+} g_1(x)g_2(\theta)\int_{\mathbb{S}^1_+} f_1(x)f_2(\theta')\widetilde{\mathbb{P}}(\theta;\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta) \\
&= \int_{\mathbb{R}\times\mathbb{S}^1_+} g_1(x)g_2(\theta)\int_{\mathbb{R}\times\mathbb{S}^1_+} f_1(x')f_2(\theta')\delta_x(\mathrm{d}x')\widetilde{\mathbb{P}}(\theta;\mathrm{d}\theta')\Lambda^1(\mathrm{d}x\mathrm{d}\theta).
\end{aligned} \tag{2.50}$$

This proves that $\lim_{i\to\infty} \mathbb{P}^{\Sigma_i,\epsilon_i}$ exists and is equal to $\delta_x \times \widetilde{\mathbb{P}}$.

Proof of (iii). Take $\Sigma_i = \Sigma$ to be constant. For all ϵ_i, the macro-reflection map P^{Σ,ϵ_i} is just the macro-reflection map $P^{\Sigma,1}$ re-expressed after scaling the spatial coordinates by ϵ_i, that is

$$P^{\Sigma,\epsilon_i} = \sigma_{\epsilon_i} \circ P^{\Sigma,1} \circ \sigma_{\epsilon_i}^{-1}, \qquad \text{where } \sigma_{\epsilon_i}(x,\theta) = (\epsilon_i x, \theta). \tag{2.51}$$

Consequently,

$$\mathbb{P}^{\Sigma,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta') = (\sigma_{\epsilon_i})_{\#}\mathbb{P}^{\Sigma,1}(\sigma_{\epsilon_i}^{-1}(x,\theta);\mathrm{d}x'\mathrm{d}\theta'). \tag{2.52}$$

Thus, for $f, g \in C_c(\mathbb{S}^1_+)$,

$$\begin{aligned}
&\int_{\mathbb{S}^1_+} g(\theta) \int_{\mathbb{S}^1_+} f(\theta')\widetilde{\mathbb{P}}^{\Sigma,\epsilon_i}(\theta,\mathrm{d}\theta')\sin\theta\mathrm{d}\theta \\
&= \int_{\mathbb{S}^1_+} g(\theta)\frac{1}{\epsilon_i}\int_0^{\epsilon_i}\int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma,\epsilon_i}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x\sin\theta\mathrm{d}\theta \\
&= \int_{\mathbb{S}^1_+} g(\theta)\frac{1}{\epsilon_i}\int_0^{\epsilon_i}\int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')(\sigma_{\epsilon_i})_{\#}\mathbb{P}^{\Sigma,1}(\sigma_{\epsilon_i}^{-1}(x,\theta);\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x\sin\theta\mathrm{d}\theta \\
&= \int_{\mathbb{S}^1_+} g(\theta)\int_0^{1}\int_{\mathbb{R}\times\mathbb{S}^1_+} f(\theta')\mathbb{P}^{\Sigma,1}(x,\theta;\mathrm{d}x'\mathrm{d}\theta')\mathrm{d}x\sin\theta\mathrm{d}\theta \\
&= \int_{\mathbb{S}^1_+} g(\theta)\int_{\mathbb{S}^1_+} f(\theta')\widetilde{\mathbb{P}}^{\Sigma,1}(\theta,\mathrm{d}\theta')\sin\theta\mathrm{d}\theta.
\end{aligned} \tag{2.53}$$

Here the second equality follows from (2.52) and the third equality follows by making the changes of variables $(x',\theta') \mapsto \sigma_{\epsilon_i}(x',\theta')$ and $(x,\theta) \mapsto \sigma_{\epsilon_i}(x,\theta)$. This shows that the sequence $\widetilde{\mathbb{P}}^{\Sigma,\epsilon_i}$ is constant and equal to $\widetilde{\mathbb{P}}^{\Sigma,1}$. Thus (iii) follows from (i) and (ii). $\square$

3 Specular Reflection Law

The dynamics of the disk and wall system are described by an evolution $t \mapsto (y(t), w(t))$ in the phase space $T\mathcal{M} \cong \mathcal{M} \times \mathbb{R}^3$. This evolution should satisfy, and ideally be uniquely determined by, accepted physical laws governing rigid body interactions. We assume that:

P1. The bodies D and W do not interpenetrate.

P2. The system is subject to Euler's laws of rigid body motion (see equations (3.1) and (3.2)).

P3. When not in contact, the net force applied to each body is zero, and upon contact, a single impulsive force is applied to the disk at the point of contact and directed parallel to the unit normal vector on the wall.

P4. The kinetic energy of the disk is conserved for all time.

The first three assumptions are standard for rigid body interactions in which no friction is present. The fourth is the limiting case for a system of bodies of finite mass in which the total kinetic energy, linear momentum, and angular momentum are conserved. After letting the mass and moment of inertia of one body diverge to infinity, the kinetic energy of the other body is conserved in the limit (see in Proposition 3.3).

The material in this section does not depend on results from the other sections, with the exception of two results proved in §4, namely (i) that there is a full measure

open subset $\partial_{\text{reg}}\mathcal{M} \subset \partial\mathcal{M}$ on which a C^1 field of unit normal vectors is defined, and (ii) a formula for the unit normal vector to the surface. See Proposition 4.2 and Remark 4.6. The material on the geometric properties of the configuration space does not depend on this section, so there is no circularity.

Let us give a more precise mathematical formulation to the physical assumptions P1-P4.

The assumption that the bodies do not interpenetrate means that the configuration of the system y is confined to the set $\mathcal{M} \subset \mathbb{R}^3$ for all time. The boundary $\partial\mathcal{M}$ corresponds to collision configurations for the system. The collision dynamics cannot be defined when multiple satellites simultaneously make contact with the wall W, or when a satellite makes contact with a singular point of W. We therefore only derive the collision law for when y lies in $\partial_{\text{reg}}\mathcal{M}$, which is a full measure subset of $\partial\mathcal{M}$ (see Proposition 4.2).

We use the following notation: for $y \in \partial_{\text{reg}}\mathcal{M}$, we let $\sigma(y) \in \partial W$ denote the position in $\mathbb{R}^2$ of the unique satellite of $D(y)$ which lies in ∂W. As usual, we write $y = (x, \alpha)$ and $w = (v, \omega)$ where x is the center of mass of the disk, α is the angular configuration, v is the linear velocity of the center of mass, and ω is the angular velocity of the disk.

When $y \in \operatorname{Int}\mathcal{M}$, the net force on D is zero. Thus Euler's equations of motion say that

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = 0, \quad \text{and} \quad J\frac{\mathrm{d}\omega}{\mathrm{d}t} = 0 \quad \text{whenever } y \in \operatorname{Int}\mathcal{M}, \tag{3.1}$$

here recalling that the total mass of the disk is m and the moment of inertia of the disk about its center of mass is J. In other words, D moves freely in the complement

of W.

Euler's laws must be given an impulsive interpretation when the disk makes contact with the wall. For $p \in \partial_{\text{reg}} W$, let $k(p)$ denote the outward pointing unit normal vector to the wall at the point p. Our assumptions say that the impulse when the disk makes contact with the wall in configuration $y \in \partial_{\text{reg}} \mathcal{M}$ is $\lambda k(\sigma(y))$ for some $\lambda \in \mathbb{R}$. Since each satellite lies at unit distance from the center of mass, the induced impulsive torque at the point of contact is then $\lambda \sin \beta$ where $\beta = \beta(y) \in (-\pi/2, \pi/2)$ is the signed angle measured counterclockwise from $k(\sigma(y))$ to $x - \sigma(y)$. In impulsive form, Euler's laws say that

$$m(v^+ - v^-) = \lambda k(\sigma(y)), \qquad J(\omega^+ - \omega^-) = \lambda \sin \beta(y) \quad \text{whenever } y \in \partial_{\text{reg}} \mathcal{M}, \tag{3.2}$$

where

$$v^{\pm}(t) := \lim_{s \to t\pm} v(t), \quad \text{and} \quad \omega^{\pm} := \lim_{s \to t\pm} \omega(t). \tag{3.3}$$

The kinetic energy of the disk in state (y, w) is

$$\frac{1}{2} m v_1^2 + \frac{1}{2} m v_2^2 + \frac{1}{2} J \omega^2 = \frac{1}{2} \|w\|^2. \tag{3.4}$$

By conservation of the kinetic energy of the disk, up to a change of units, we may assume that $\|w\|^2 = 1$ for all time.

Summarizing, we wish to solve the following initial value problem: for $(y^0, w^0) \in$

$\operatorname{Int}\mathcal{M}$, to find an evolution $t \mapsto (y(t), w(t)) \in \mathbb{R}^3 \times \mathbb{R}^3$ which satisfies

$$\begin{cases} y \in \mathcal{M}, \\ y \in \operatorname{Int}\mathcal{M} \quad \Rightarrow \quad \frac{\mathrm{d}w}{\mathrm{d}t} = 0, \\ y \in \partial_{\mathrm{reg}}\mathcal{M} \quad \Rightarrow \quad (\exists \lambda \in \mathbb{R}) : w^+ - w^- = \begin{pmatrix} m^{-1}\lambda k(\sigma(y)) \\ J^{-1}\lambda \sin\beta(y) \end{pmatrix}, \\ ||w|| = 1, \\ (y(0), w(0)) = (y^0, w^0). \end{cases} \tag{3.5}$$

Contained in these conditions is the assumption that the derivative $\frac{\mathrm{d}w}{\mathrm{d}t}$ exists when $y \in \operatorname{Int}\mathcal{M}$ and that the limits w^- and w^+ exist when $y \in \partial_{\mathrm{reg}}\mathcal{M}$.

Remark 3.1. Our approach to formulating the above problem aims to obtain a precise mathematical statement quickly while keeping the amount of technical machinery to a minimum. One drawback to our approach is that it requires taking the distinct forms of Euler's laws for impulsive and non-impulsive interactions as basic. The reader may wonder whether there is a unified framework in which to handle these different types of interactions. In addition, while Proposition 3.2 shows that the problem (3.5) is well-posed, it is not so clear how well the approach would generalize to other rigid body settings.

Much work has been done in the last 40 years to put physical systems with impulsive interactions within a rigorous and unified mathematical framework. One such approach is based on *differential inclusions*. The idea is to solve a differential

relation of form

$$F(q,t) - \ddot{q} \in \partial I_{V(q)}, \tag{3.6}$$

where q is the position in the configuration space $\mathcal{M}$, F is the external force acting on the system, $V(q)$ is the tangent cone in $\mathcal{M}$ at the point q, $I_{V(q)}(r) = 0$ if $r \in V(q)$ and $I_{V(q)}(r) = \infty$ otherwise, and $\partial I_{V(q)}$ is the subdifferential of $I_{V(q)}$ in the convex analysis sense. Some early work to formulate and solve rigid body problems within this setting may be found in [24, 26, 27, 29, 30, 25, 23].

Another approach is to formulate impulse problems as variational problems in Lagrangian mechanics. This leads to geometric integration algorithms for systems with impacts, as detailed in [16].

We now solve the problem (3.5).

Recall that n denotes the field of inward-pointing unit normal vectors on $\partial_{\text{reg}}\mathcal{M}$. Define the subbundles of the tangent space

$$\begin{aligned} V_{\text{in}} &= \{(y,w) \in T\mathcal{M} : y \in \partial_{\text{reg}}\mathcal{M} \text{ and } \langle w, n(y)\rangle > 0\}, \\ V_{\text{out}} &= -V_{\text{in}} = \{(y,w) \in T\mathcal{M} : y \in \partial_{\text{reg}}\mathcal{M} \text{ and } \langle w, n(y)\rangle < 0\}. \end{aligned} \tag{3.7}$$

Given $(y,w) \in \mathcal{M} \times \mathbb{R}^3$, let

$$\bar{t}(y,w) = \inf\{t > 0 : y + tw \in \partial\mathcal{M}\}. \tag{3.8}$$

If $y \in \operatorname{Int}\mathcal{M}$, or $(y,w) \in V_{\text{in}}$, then by compactness and continuity, if $\bar{t}(y,w) < \infty$ then it is a minimum.

Proposition 3.2. *Let $(y^0, w^0) \in \operatorname{Int} \mathcal{M} \times \mathbb{R}^3$, and assume that $\bar{t}_1 := \bar{t}(y^0, w^0) < \infty$. Let $y^1 = y^0 + \bar{t}_1 w^0$, and assume that $(y^1, w^0) \in V_{out}$. There exists $\delta > 0$ and a unique evolution $t \mapsto (y(t), w(t))$, $t \in [0, \bar{t}_1 + \delta)$ such that y is continuous, w is right-continuous, and the conditions (3.5) hold. Explicitly, the evolution is given by linear motion in the interior of $\mathcal{M}$ and specular reflection on the boundary:*

$$(y(t), w(t)) = \begin{cases} (y^0 + tw^0, w^0) & \text{if } 0 \le t < \bar{t}_1, \\ (y^1, w^0 - 2\langle w^0, n(y^1)\rangle n(y^1)) & \text{if } t = \bar{t}_1, \\ (y^1 + (t - \bar{t}_1) w(\bar{t}_1), w(\bar{t}_1)) & \text{if } \bar{t}_1 < t < \bar{t}_1 + \delta. \end{cases} \tag{3.9}$$

Proof of Proposition 3.2. Let $w^1 = w^0 - 2\langle w^0, n(y^1)\rangle n(y^1)$. It follows from $(y^1, w^0) \in V_{\text{out}}$ that $(y^1, w^1) \in V_{\text{in}}$, and consequently $\bar{t}_2 := \bar{t}(y^1, w^1) > 0$. We choose $\delta = \bar{t}_2$.

Assume that $t \mapsto (y(t), w(t))$ is right-continuous and satisfies the conditions (3.5) for $0 \le t \le \bar{t}_1 + \delta$. To prove uniqueness, it is enough to show that (3.9) holds. Since the motion is free in the interior of $\mathcal{M}$, it is clear that $(y(t), w(t)) = (y^0 + tw^0, w^0)$ for $0 \le t < \bar{t}_1$, and in particular $y^1 = y(\bar{t}_1)$. Write $y^1 = (x^1, \alpha^1) \in \partial_{\text{reg}}\mathcal{M}$. We have $w^- = (v^-, \omega^-) = \lim_{t \to \bar{t}_1 -} w(t) = w^0$, and $w^+ = (v^-, \omega^-) = \lim_{t \to \bar{t}_1 +} w(t) = w(\bar{t}_1)$, where the last equality follows by right-continuity. By assumption, we have

$$w^+ - w^- = w(\bar{t}_1) - w^0 = \begin{pmatrix} m^{-1}\lambda k(\sigma(y^1)) \\ J^{-1}\lambda \sin\beta(y^1) \end{pmatrix}. \tag{3.10}$$

Write $k = k(\sigma(y^1)) = (k_1, k_2) \in \mathbb{R}^2$, and write $y^1 = (x^1, \alpha^1)$. The vector $x^1 - \sigma(y^1)$ is equal to $(-\sin\alpha^1, \cos\alpha^1)$, and consequently $\frac{\pi}{2} - \beta(y^1)$ is the angle measured coun-

terclockwise from the vector $(\cos\alpha^1, \sin\alpha^1)$ to the normal $k(y^1)$. We therefore have

$$\sin\beta = \cos\left(\frac{\pi}{2} - \beta\right) = k \cdot (\cos\alpha^1, \sin\alpha^1) = k_1 \cos\alpha^1 + k_2 \sin\alpha^1. \tag{3.11}$$

Thus

$$w(\bar{t}_1) - w^0 = \lambda \begin{pmatrix} m^{-1}k_1 \\ m^{-1}k_2 \\ J^{-1}(k_1 \cos\alpha + k_2 \sin\alpha) \end{pmatrix} = R\lambda n, \tag{3.12}$$

where $R = (m^{-1} + J^{-1}(k_1 \cos\alpha^1 + k_2 \sin\alpha^1)^2)^{1/2}$, here using the formula (4.39) for the unit normal n in the configuration space given in Remark 4.6. By conservation of energy and the above, we have

$$\begin{aligned} \|w^0\|^2 = \|w(\bar{t}_1)\|^2 &= \|w^0 + R\lambda n\|^2 \\ &= \|w^0\|^2 + 2R\lambda\langle w^0, n\rangle + (R\lambda)^2 \\ \Rightarrow \quad 0 &= 2R\lambda\langle w^0, n\rangle + (R\lambda)^2 \\ \Rightarrow \quad R\lambda &= -2\langle w^0, n\rangle \quad \text{or} \quad R\lambda = 0. \end{aligned} \tag{3.13}$$

As $(y^1, w^0) \in V_{\text{out}}$, we see that $\langle w^0, n\rangle \neq 0$. By (3.12) and (3.13) either $w(\bar{t}_1) = w^0$ or $w(\bar{t}_1) = w^0 - 2\langle w^0, n\rangle n$. But we cannot have $w(\bar{t}_1) = w^0$ since this would imply by right-continuity that the point mass enters $\mathcal{M}^c$.

It remains to show that the motion is linear for $\bar{t}_1 < t < \delta$. From the above, $(y^1, w(\bar{t}_1)) \in V_{\text{in}}$, and thus by continuity there exists a maximal (possibly infinite) $\eta > 0$ such that $y(t) \in \operatorname{Int}\mathcal{M}$ for $\bar{t}_1 < t < \bar{t}_1 + \eta$. By free motion in the interior of $\mathcal{M}$,

it follows that

$$(y(t), w(t)) = (y^1 + (t - \bar{t}_1)w(\bar{t}_1), w(\bar{t}_1)) \quad \text{for} \quad \bar{t}_1 < t < \bar{t}_1 + \eta. \tag{3.14}$$

It is easy to see that $\eta = \delta$. For (3.14) and the definition of η imply that $\eta \leq \bar{t}_2 = \delta$. On the other hand, if $\eta < \bar{t}_2$, then $y(\bar{t}_1 + \eta) \in \operatorname{Int} \mathcal{M}$ by continuity of the trajectory, and thus there exists $\eta' > \eta$ such that $y(t) \in \operatorname{Int} \mathcal{M}$ for $\bar{t}_1 < t < \bar{t}_1 + \eta'$, contradicting maximality of η. We conclude uniqueness.

To prove existence, we take $(y(t), w(t))$ to be defined by (3.9) for $0 \leq t < \bar{t}_1 + \delta$. Then $y(t)$ is continuous and $w(t)$ is right-continuous. Further, the only condition in (3.5) which is not trivial to verify is the third. For this, note that we may take $\lambda = -2R^{-1}\langle w^0, n\rangle$, where R is defined as above. It follows from the expression (4.39) for the normal vector n that the third condition in (3.5) holds. □

We close by providing some support for the assumption that the energy of the disk is conserved for all time.

Proposition 3.3. *Consider a physical system consisting of two bodies B_1 and B_2 in the plane, of mass M_1 and M_2 respectively and with moments of inertia J_1 and J_2 respectively about their centers of mass. Assume that the kinetic energy, linear momentum, and angular momentum of the system are conserved for all time. Let $t < t'$, let v^i, $v^{i'}$ be the linear velocity of the center of mass of B_i at times t, t' respectively, let $\omega^i, \omega^{i'}$ be the angular velocity of B_i at times t, t' respectively, and assume that $v^2 = \omega^2 = 0$. Then, keeping M_1, J_1, v^1, and ω^1 fixed, the following limit*

holds:

$$\lim_{M_2,J_2\to\infty} M_1\|v^{1\prime}\|^2 + J_1|\omega^{1\prime}|^2 = M_1\|v^1\|^2 + J_1|\omega^1|^2. \tag{3.15}$$

Proof of Proposition 3.3. Since we assume $v^2 = 0$ and $\omega^2 = 0$, conservation of linear and angular momentum give us

$$M_1 v^{1\prime} + M_2 v^{2\prime} = M_1 v^1, \tag{3.16}$$

and

$$J_1\omega^{1\prime} + J_2\omega^{2\prime} = J_1\omega^1. \tag{3.17}$$

We obtain from these,

$$M_2^2\|v^{2\prime}\|^2 \le 2M_1^2\|v^{1\prime}\|^2 + 2M_1^2\|v^1\|^2, \qquad J_2^2|\omega^{2\prime}|^2 \le 2J_1^2|\omega^{1\prime}|^2 + 2J_1^2|\omega^1|^2, \tag{3.18}$$

and hence

$$\begin{aligned} M_2\|v^{2\prime}\|^2 + J_2|\omega^{2\prime}|^2 &\le \frac{2M_1^2}{M_2}(\|v^{1\prime}\|^2 + \|v^1\|^2) + \frac{2J_1^2}{J_2}(|\omega^{1\prime}|^2 + |\omega^1|^2) \\ &\le \max\left\{\frac{2M_1}{M_2}, \frac{2J_1}{J_2}\right\}(M_1\|v^{1\prime}\|^2 + J_1|\omega^{1\prime}|^2 + M_1\|v^1\|^2 + J_1|\omega^1|^2). \end{aligned} \tag{3.19}$$

By conservation of energy, we also have

$$M_1\|v^{1\prime}\|^2 + J_1|\omega^{1\prime}|^2 + M_2\|v^{2\prime}\|^2 + J_2|\omega^{2\prime}|^2 = M_1\|v^1\|^2 + J_1|\omega^1|^2. \tag{3.20}$$

Therefore,

$$M_1\|v^{1\prime}\|^2 + J_1|\omega^{1\prime}|^2 \leq M_1\|v^1\|^2 + J_1|\omega^1|^2. \tag{3.21}$$

From (3.21) and (3.19) we obtain that

$$M_2\|v^{2\prime}\|^2 + J_2|\omega^{2\prime}|^2 \leq \max\left\{\frac{4M_1}{M_2}, \frac{4J_1}{J_2}\right\}(M_1\|v^1\|^2 + J_1|\omega^1|^2). \tag{3.22}$$

From this we see that

$$M_2\|v^{2\prime}\|^2 + J_2|\omega^{2\prime}|^2 \to 0 \quad \text{as } M_2, J_2 \to \infty. \tag{3.23}$$

Hence, by conservation of energy (3.20), the limit (3.15) holds. □

4 Elementary Properties of the Billiard System

The goals for this section are first to describe the basic properties of the configuration space $\mathcal{M}$ and its cylindrical approximation $\mathcal{M}_{\text{cyl}}$, and second to provide a rigorous definition for the collision law described in the introduction. §4.1 and 4.2 are concerned exclusively with the geometry of $\mathcal{M}$ and $\mathcal{M}_{\text{cyl}}$, while §4.3 introduces dynamics.

The collision law turns out to be a special case of the general macro-reflection laws, described in §6. Results proved in §6 can be applied to show, for example, that the collision law is symmetric with respect to the measure Λ^2.

In addition to describing the collision law in $\mathcal{M}$, we will also define two auxiliary collision laws, the *cylindrical collision law* and the *modified collision law*, which will play important roles in the proofs of our main results.

4.1 Properties of the configuration space

Throughout this subsection, the cell Σ – assumed to satisfy conditions B1-B5 of §1.2.4 – and the roughness scale $\epsilon > 0$ are fixed. As described in §1.3, the cell and roughness scale give rise to a fixed wall $W = W(\Sigma, \epsilon)$ and a freely moving disk with satellites $D = D(\epsilon)$, where the satellites are spaced at angles $\rho(\epsilon)$ apart. This physical system has a configuration space $\mathcal{M} = \mathcal{M}(\Sigma, \epsilon) \subset \mathbb{R}^3$, defined by the expression (1.41).

We first describe some elementary properties of the configuration space $\mathcal{M}$. In what follows, $\mathcal{H}^2$ denotes 2-dimensional Hausdorff measure on $\mathbb{R}^3$.

Proposition 4.1. *i. For all $j, k \in \mathbb{Z}$, $\mathcal{M} + j\epsilon e_1 + k\rho e_3 = \mathcal{M}$.*

ii. $\{(x_1, x_2, \alpha) : x_2 \geq 0\} \subset \mathcal{M} \subset \{y \in \mathbb{R}^3 : D(y) \cap \operatorname{Int} W = \varnothing\}$.

iii. $\operatorname{Int} \mathcal{M} = \{y \in \mathbb{R}^3 : D(y) \cap W = \varnothing\}$*; consequently,* $\mathcal{M} = \overline{\operatorname{Int} \mathcal{M}}$.

iv. $\partial \mathcal{M} \subset \{y \in \mathbb{R}^3 : D(y) \cap \operatorname{Int} W = \varnothing,$ *and for some* $0 \leq k \leq N,\ S_k(y) \in \partial W\}$

v. $\partial \mathcal{M} \subset \{(x_1, x_2, \alpha) : -\epsilon - \frac{1}{8}\rho^2 \leq x_2 \leq 0\}$.

vi. Assume ϵ *is sufficiently small. For any* $y \in \mathcal{M}$*, there are at most two satellites of* $D(y)$ *which lie on* ∂W*. If two satellites of* $D(y)$ *lie in* ∂W*, then the satellites are adjacent to each other on the disk.*

vii. $\mathcal{H}^2(\{(x_1, x_2, \alpha) : x_2 \geq 0\} \cap \overline{\mathcal{M}^c}) = 0$.

Next, we will describe the regularity properties of $\mathcal{M}$, and introduce a large subset of $\mathcal{M}$ which has a simple parametrization. This requires some additional notation. We define a disjoint union of open intervals in the line:

$$\mathcal{Z} = \mathcal{Z}(\epsilon) = \bigcup_{k \in \mathbb{Z}} \left(\frac{2k-1}{2}\rho(\epsilon) + \delta_0(\epsilon), \frac{2k+1}{2}\rho(\epsilon) - \delta_0(\epsilon) \right) \tag{4.1}$$

where

$$\delta_0 = \delta_0(\epsilon) = \arcsin\left(\frac{\epsilon}{2\sin(\rho(\epsilon)/2)} \right). \tag{4.2}$$

We also let

$$\widehat{\mathcal{Z}} = \mathbb{R}^2 \times \mathcal{Z}. \tag{4.3}$$

One may easily check that, for any $R > 0$, the Lebesgue measure of $[-R, R] \setminus \mathcal{Z}$ is of order $O(\delta_0) = O(\epsilon/\rho) = o(\epsilon^{1/2})$. Thus $\widehat{\mathcal{Z}}$ is a large subset of $\mathbb{R}^3$ in the sense that

for any bounded subset $B \subset \mathbb{R}^3$, the Lebesgue measure of $B \setminus \widehat{\mathcal{Z}}$ converges to zero as $\epsilon \to 0$. We define

$$\mathcal{M}_{\text{roll}} = \mathcal{M} \cap \widehat{\mathcal{Z}}, \tag{4.4}$$

$$\Gamma_{\text{roll}} = \partial\mathcal{M} \cap \widehat{\mathcal{Z}}. \tag{4.5}$$

We have

Proposition 4.2. *Assume ϵ is sufficiently small.*

i. There exists a closed subset $\mathcal{S} \subset \partial\mathcal{M}$ such that

i.1. $\mathcal{H}^2(\mathcal{S}) = 0$; and

i.2. for every $q \in \partial\mathcal{M} \setminus \mathcal{S}$, there exists a neighborhood $U \subset \mathbb{R}^3$ of q and a C^2 diffeomorphism $\phi : U \to \mathbb{R}^3$ such that $\phi(q) = 0$ and $\phi(U \cap \mathcal{M}) = \mathbb{H}^3 := \{(x_1, x_2, x_3) : x_3 \geq 0\}$.

ii. For any $y \in \mathcal{M}_{\text{roll}}$, $D(y) \cap W$ has cardinality at most 1.

For any $y \in \Gamma_{\text{roll}}$, the set $D(y) \cap W$ has cardinality 1, and consists of a single satellite of D.

iii. $\mathcal{M}_{\text{roll}}$ is parametrized by the function $F : \overline{W^c} \times \mathcal{Z} \to \mathbb{R}^3$ defined by

$$F(x_1, x_2, \alpha) = (x_1 - \sin\overline{\alpha}, x_2 + \cos\overline{\alpha}, \alpha), \tag{4.6}$$

where $\overline{\alpha} := \alpha - \overline{k}\rho$ and $\overline{k} := \operatorname{argmin}\{|\alpha - k\rho| : k \in \mathbb{Z}\}$.

Moreover, the restriction of F to $\partial W \times \mathcal{Z}$ parametrizes Γ_{roll}.

Remark 4.3. The quantity $\overline{k}$ is uniquely defined because $\mathcal{Z}$ does not contain points of form $(\frac{2j+1}{2})\rho$, $j \in \mathbb{Z}$.

Remark 4.4. The parametrization (4.6) explains the notation $\mathcal{M}_{\text{roll}}$ and Γ_{roll}. Suppose that the disk initially has the configuration $(x_1, x_2, k\rho)$ where $k \in \mathbb{Z}$, and imagine rotating the center of the disk counterclockwise about the satellite S_k by a small angle $\Delta\alpha = \alpha - k\rho$, keeping the position of S_k fixed. After "rolling" the disk in this way, the final configuration is given by $F(x_1, x_2, \alpha)$.

We call a point in $\mathcal{M}$ a *regular point* if it belongs to the set

$$\mathcal{M}_{\text{reg}} := \mathcal{M} \setminus \mathcal{S}. \tag{4.7}$$

To avoid ambiguity, we will assume that $\mathcal{S}$ is the minimal subset of $\partial\mathcal{M}$ (with respect to inclusion) such that $i.1$ and $i.2$ in Proposition 4.2 hold.

In §6.1.1, we define the class of billiard domains $\text{CES}_0^2(\mathbb{R}^3)$. The class is sufficiently large to encompass the kinds of spaces dealt with in this work, while narrow enough that standard results from billiards theory still apply. The acronym CES stands for "closed and embedded, with singularities." The subscript 0 means that the codimension in $\mathbb{R}^3$ of a submanifold in this class is zero, and the superscript 2 indicates that the submanifold is twice differentiable. An immediate consequence of Proposition 4.2(i) is the following fact.

Corollary 4.5. *$\mathcal{M}$ belongs to the class* $\text{CES}_0^2(\mathbb{R}^3)$.

To prove the propositions, we introduce the following notation. Recall that N denotes the number of satellites of the disk. For $y = (x, \alpha) \in \mathbb{R}^3$ and $0 \le k \le N-1$,

let $S_k(y)$ denote the position of the satellite S_k after rotating the disk in reference configuration (1.37) about its center of mass counterclockwise by an angle of α and translating by x. Explicitly,

$$S_k(x_1, x_2, \alpha) = (x_1 + \sin(\alpha + k\rho), x_2 - \cos(\alpha + k\rho)). \tag{4.8}$$

The map $y \mapsto S_k(y) : \mathbb{R}^3 \to \mathbb{R}^2$ is smooth, and for each fixed α, the map $x \mapsto S_k(x, \alpha)$ is a translation in the plane $\mathbb{R}^2$.

As noted in Remark 1.19, the inner body D_0 cannot come into contact with the wall W. Thus, for the purposes of describing the geometry of $\mathcal{M}$, there is no loss of generality if we assume $D_0 = \varnothing$. We use this fact without comment in the proofs below.

Proof of Proposition 4.1. (i) We must show that $\mathcal{M}$ is invariant under translation in $\mathbb{R}^3$ by ϵe_1 and by ρe_3. The first of these follows because the wall $W = W(\Sigma, \epsilon)$ is invariant under translation by ϵe_1. The second follows by noting that a rotation of the disk through an angle of ρ about its center of mass maps the disk onto itself.

(ii) Since the satellites of D lie at unit distance from the center, and $W \subset \{(x_1, x_2) : x_2 \leq -1\}$, it follows that $\{(x_1, x_2, \alpha) : x_2 > 0\} \subset \mathcal{M}$. Since $\mathcal{M}$ is closed, the first inclusion follows.

By construction, $D \cap \operatorname{Int} W = \varnothing$ if and only if none of the satellites of D lie in $\operatorname{Int} W$. That is,

$$\{y : D(y) \cap \operatorname{Int} W = \varnothing\} = \{y : \text{for } 0 \leq k \leq N-1, S_k(y) \notin \operatorname{Int} W\}. \tag{4.9}$$

By continuity of the S_k's, this set is closed. The second inclusion then follows by the definition (1.41) of $\mathcal{M}$.

(iii) Let $\mathcal{M}_0 = \{y : D(y) \cap W = \varnothing\}$. Equivalently,

$$\mathcal{M}_0 = \{y : \text{for } 0 \le k \le N-1, S_k(y) \notin W\}. \tag{4.10}$$

By continuity of the S_k's, this is an open set, and it is also evident that $\mathcal{M}_0 \subset \mathcal{M}$. Thus $\mathcal{M}_0 \subset \operatorname{Int} \mathcal{M}$. For the reverse inclusion, suppose that $y \notin \mathcal{M}_0$. Then for some k, $S_k(y) \in W$. Let $V \subset \mathbb{R}^3$ be any neighborhood of y. Since W is the closure of its interior by condition A1 in §1.2.2, it follows that $V \cap \operatorname{Int} W \neq \varnothing$. Therefore, by (ii) V intersects $\mathcal{M}^c$. Hence $y \notin \operatorname{Int} \mathcal{M}$, and this proves $\mathcal{M}_0 \supset \operatorname{Int} \mathcal{M}$.

(iv) It follows from (ii) and (iii) that

$$\partial \mathcal{M} \subset \{y \in \mathbb{R}^3 : D(y) \cap \operatorname{Int} W = \varnothing \text{ and } D(y) \cap \partial W \neq \varnothing\}. \tag{4.11}$$

If D intersects W, then one of its satellites lies in W, so the result follows.

(v) By definition of $W = W(\Sigma, \epsilon)$,

$$\{(x_1, x_2) : x_2 \le -1 - \epsilon\} \subset W \subset \{(x_1, x_2) : x_2 \le -1\}. \tag{4.12}$$

Consequently, since the satellites lie at unit distance from the center of mass of D, if $y = (x_1, x_2, \alpha) \in \partial \mathcal{M}$, then $x_2 \le 0$; otherwise $D(y)$ could not intersect W. On the other hand, a strict lower bound is obtained by putting D in a configuration such that two adjacent satellites lie on the line $x_2 = -1 - \epsilon$. See Figure 14. Trigonometry

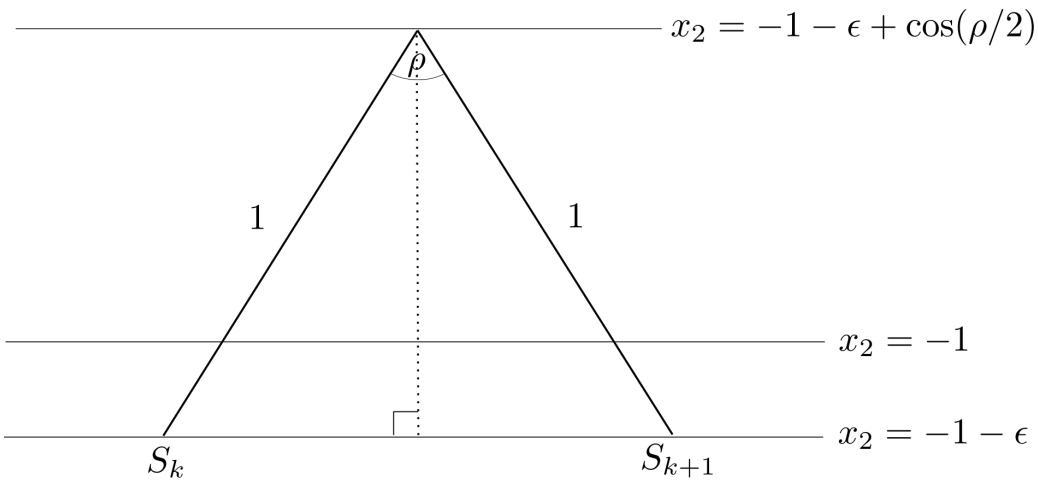

Figure 14: Two satellites lying on the line $x_2 = -1 - \epsilon$.

yields that the x_2-coordinate of the center of the disk in this configuration is

$$-1 - \epsilon + \cos(\rho/2) \geq -\epsilon - \frac{1}{8}\rho^2. \tag{4.13}$$

(vi) Let $y = (x_1, x_2, \alpha) \in \mathcal{M}$, and without loss of generality, suppose that $S_0(y)$, $S_1(y)$, and $S_2(y)$ all lie in W. An upper bound for x_2 is attained in the case where $S_0(y)$ and $S_2(y)$ both lie on the line $x_2 = -1$. In this case, $x_2 = -1 + \cos\rho \leq -\frac{1}{4}\rho^2$ for ρ sufficiently small. Since $\rho(\epsilon) \to 0$ and $\epsilon/\rho(\epsilon)^2 \to 0$, this contradicts the bound (4.13) if ϵ is sufficiently small.

(vii) Suppose $y = (x_1, x_2, \alpha) \in \overline{\mathcal{M}^c} = (\operatorname{Int}\mathcal{M}^c)$, and $x_2 \geq 0$. Then at least one satellite of $D(y)$ lies in W. By (4.12) and the fact that each satellite lies at unit distance from the center of mass of D, this implies that $x_2 = 0$, and $\alpha = k\rho$ for some

$k \in \mathbb{Z}$. That is,

$$\{(x_1, x_2, \alpha) : x_2 \geq 0\} \cap \overline{\mathcal{M}^c} \subset \{(x_1, x_2, \alpha) : x_2 = 0 \text{ and } \alpha = k\rho \text{ for some } k \in \mathbb{Z}\}, \quad (4.14)$$

and the set on the right is $\mathcal{H}^2$-null. $\square$

Proof of Proposition 4.2. (i) We begin by introducing the mapping $g : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^3$ defined by

$$g(p_1, p_2, \alpha) = (p_1 - \sin\alpha, p_2 + \cos\alpha, \alpha). \quad (4.15)$$

This is the mapping which takes (p_1, p_2, α) to the unique configuration y such that $S_0(y) = (p_1, p_2)$ and the angular orientation of the disk is α. One may easily check that g is a diffeomorphism. Consequently, g takes $\mathcal{H}^2$-null sets to $\mathcal{H}^2$-null sets.

Consider the following subsets of $\mathcal{M}$:

$$A = \{y \in \mathcal{M} : D(y) \cap W \text{ contains a singular point of the boundary } \partial W\}, \quad (4.16)$$

$$B = \{y \in \mathcal{M} : D(y) \cap W \text{ contains more than one satellite of } D\}. \quad (4.17)$$

We define

$$\mathcal{S} = A \cup B. \quad (4.18)$$

By continuity and the fact that W is closed, it is easy to see that A and B are closed; therefore $\mathcal{S}$ is closed.

To prove (i.1) holds, we must show that A and B are $\mathcal{H}^2$-null. To show that $\mathcal{H}^2(A) = 0$, by symmetry, it is enough to show that $\mathcal{H}^2(\{y : S_0(y) \in \partial_s W\}) = 0$. To

prove this, it is enough to show that $g^{-1}\{y : S_0(y) \in \partial_s W\}$ is an $\mathcal{H}^2$-null set. But this set is equal to $\partial_s W \times \mathbb{R}$, and the result follows since $\partial_s W$ is a discrete set of points.

It takes more effort to show that $\mathcal{H}^2(B) = 0$. By Proposition 4.1(ii) and (vi) and symmetry, it is enough to show that

$$\{y \in \mathbb{R}^3 : S_0(y) \in \partial W \text{ and } S_1(y) \in \partial W\} \tag{4.19}$$

is $\mathcal{H}^2$-null. By pulling back with respect to g, it is equivalent to show that

$$K := \{(p, \alpha) \in \partial W \times \mathbb{R} : S_1(g(p, \alpha)) \in \partial W\} \tag{4.20}$$

is $\mathcal{H}^2$-null. Furthermore, by periodicity, it is enough to show that

$$K_1 := \{(p, \alpha) \in \partial W_1 \times [0, 2\pi) : S_1(g(p, \alpha)) \in \partial W\} \tag{4.21}$$

is $\mathcal{H}^2$-null, where $\partial W_1 := \partial W \cap [0, \epsilon) \times \mathbb{R}$ is a single period of the boundary of the wall W.

Note that the restriction of $\mathcal{H}^2$ to $\partial W_1 \times [0, 2\pi)$ is just the product measure induced by Lebesgue measure on each factor. We denote Lebesgue measure on both factors by m. For each $p \in \partial W_1$, let

$$F_p = \{\alpha \in [0, 2\pi) : S_1(g(p, \alpha)) \in \partial W\}, \tag{4.22}$$

and let

$$E = \{p \in \partial W_1 : m(F_p) > 0\}. \tag{4.23}$$

For $r > 0$ and $p \in \mathbb{R}^2$, let $C(r,p)$ denote the circle of radius r centered at p. Trigonometry shows that the distance between any two satellites is

$$2\sin(\rho/2). \tag{4.24}$$

Therefore for each $p \in \partial W_1$, the mapping $\alpha \mapsto S_1(g(p,\alpha))$ takes α to the point on the circle $C(2\sin(\rho/2),p)$ making an angle of α counterclockwise from the point $S_1(g(p,0))$ in the same circle. Therefore, for any $U \subset [0,2\pi)$,

$$S_1(g(p,U)) \subset C(2\sin(\rho/2),p), \tag{4.25}$$

and

$$m(S_1(g(p,U))) = 2\sin(\rho/2)m(U). \tag{4.26}$$

For each $p \in \partial W_1$, let $G_p = S_1(g(p,F_p))$. If $p \in E$, then

$$G_p \subset \partial W_2 \tag{4.27}$$

where $\partial W_2 := \partial W \cap [-1,\epsilon+1]$, a bounded set. Moreover, since each G_p is a subset of a circle centered at p, if $p \neq p'$, then G_p and $G_{p'}$ can intersect in at most two points. In particular $m(G_p \cap G_{p'}) = 0$. Applying this and (4.26) and (4.27), we obtain

$$\sum_{p\in E} 2\sin(\rho/2)m(F_p) = \sum_{p\in E} m(G_p) = m\left(\bigcup_{p\in E} G_p\right) \leq m(\partial W_2) < \infty. \tag{4.28}$$

Since $m(F_p) > 0$ for each $p \in E$, it follows that E is countable. Therefore,

$$\mathcal{H}^2(K_1) = \int_E m(F_p) m(\mathrm{d}p) = 0, \tag{4.29}$$

and this concludes the proof that $\mathcal{H}^2(B) = 0$.

It remains to show that (i.2) holds. Let $q = (p_1, p_2, \beta) \in \partial\mathcal{M} \setminus \mathcal{S}$. Then exactly one satellite of $D(q)$ lies in $\partial_{\mathrm{reg}} W$, and all the other satellites lie in W^c. Without loss of generality, we may suppose $S_0(q) \in \partial_{\mathrm{reg}} W$. Let $V \subset \mathbb{R}^2$ be a neighborhood of $S_0(q)$ chosen small enough that there is some diffeomorphism $\psi : V \to \mathbb{R}^3$ such that $\psi(S_0(q)) = 0$, and $\psi(V \cap \overline{W^c}) = \{(x_1, x_2) : x_2 \geq 0\}$. Let

$$\widehat{U} = S_0^{-1}(V) = g(V \times \mathbb{R}), \tag{4.30}$$

and let $\phi : \widehat{U} \to \mathbb{R}^3$ be defined by

$$\phi(x_1, x_2, \alpha) = (\psi(S_0(x_1, x_2, \alpha)), \alpha - \beta). \tag{4.31}$$

This is easily seen to be a diffeomorphism, with inverse given by $\phi^{-1}(x_1, x_2, \alpha) = g(\psi^{-1}(x_1, x_2), \alpha + \beta)$ and with $\phi(q) = 0$. We may find a neighborhood $U \subset \widehat{U}$ of q

such that $S_k(y) \in W^c$ for all $y \in U$ and $1 \le k \le N-1$. Then

$$\begin{aligned}
\phi(U \cap \mathcal{M}) &= \phi(U \cap \overline{\{y \in \mathbb{R}^3 : S_0(y) \in W^c\}}) \\
&= \phi(U \cap \{y \in \mathbb{R}^3 : S_0(y) \in \overline{W^c}\}) \\
&= \phi(U) \cap \psi(V \cap \overline{W^c}) \times \mathbb{R} \\
&= \phi(U) \cap \{(x_1, x_2, x_3) : x_2 \ge 0\}.
\end{aligned} \tag{4.32}$$

Thus $\phi : U \to \phi(U)$ gives us the desired map.

(ii) Since $\Gamma_{\text{roll}} \subset \partial\mathcal{M}$, it is clear that for any $y \in \Gamma_{\text{roll}}$, at least one satellite of $D(y)$ lies in W. Since $\Gamma_{\text{roll}} \subset \mathcal{M}_{\text{roll}} \subset \mathbb{R} \times [-\epsilon, \infty) \times \mathcal{Z}$, (ii) will follow from:

Claim 4.2.1. *Suppose $y = (x_1, x_2, \alpha) \in \mathbb{R} \times [-\epsilon, \infty) \times \mathcal{Z}$. Then at most one satellite of $D(y)$ lies in W.*

To prove this, note that if $y \in \mathbb{R} \times (0, \infty) \times \mathcal{Z}$, then none of the satellites can lie in W. Thus, to prove the claim, we may suppose that $y \in \mathbb{R} \times [-\epsilon, 0] \times \mathcal{Z}$. It is then enough to show that, for such a y, at most one satellite of $D(y)$ lies in the strip $\{(x_1, x_2) : -1 - \epsilon < x_2 \le -1\}$. Suppose for a contradiction that two satellites lie in the strip. By (vi) the two satellites are adjacent. By symmetry, we may without loss of generality take the adjacent pair to be S_0 and S_1. Up to symmetry, the situation is as depicted in Figure 15. We let $\epsilon_1 \ge 0$ be the magnitude of the difference in the x_2-coordinates of S_0 and S_1, and we let δ denote the unsigned angle between the ray $\overrightarrow{S_0S_1}$ and the horizontal line through S_0. The center of D and the two satellites S_0

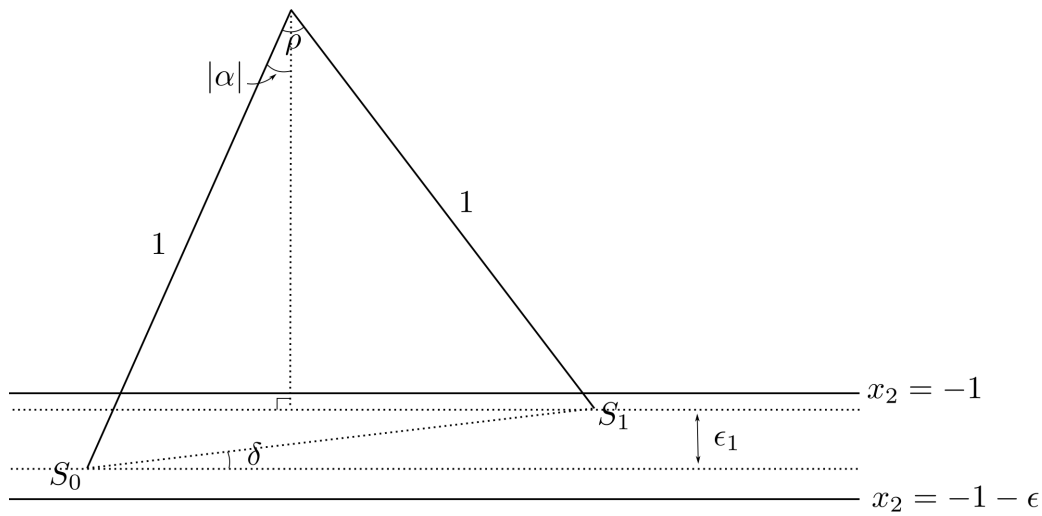

Figure 15: Isosceles triangle formed by the center of the disk and two satellites.

and S_1 form an isosceles triangle. Elementary triangle geometry shows that

$$\delta = \frac{\rho}{2} - |\alpha|, \qquad 2\sin\left(\frac{\rho}{2}\right)\sin\delta = \epsilon_1. \tag{4.33}$$

Therefore,

$$\begin{aligned} \frac{\rho}{2} \geq |\alpha| = \frac{\rho}{2} - \delta &= \frac{\rho}{2} - \arcsin\left(\frac{\epsilon_1}{2\sin(\rho/2)}\right) \\ &\geq \frac{\rho}{2} - \arcsin\left(\frac{\epsilon}{2\sin(\rho/2)}\right) = \frac{\rho}{2} - \delta_0, \end{aligned} \tag{4.34}$$

here using the fact that $\epsilon_1 \leq \epsilon$. Hence α is not in $\mathcal{Z}$, giving us the desired contradiction.

(iii) Let

$$\mathcal{Z}^0 = \left(-\frac{1}{2}\rho + \delta_0, \frac{1}{2}\rho - \delta_0\right), \qquad \widehat{\mathcal{Z}}^0 = \mathbb{R}^2 \times \mathcal{Z}^0. \tag{4.35}$$

The set $\mathcal{Z}^0$ is the member of the union (4.1) corresponding to $k = 0$. Let F^0 denote the restriction of F to $\overline{W^c} \times \mathcal{Z}^0$; then

$$F^0(x_1, x_2, \alpha) = (x_1 - \sin\alpha, x_2 + \cos\alpha, \alpha), \qquad (x_1, x_2, \alpha) \in \overline{W^c} \times \mathcal{Z}^0. \tag{4.36}$$

Let $\mathcal{M}^0_{\text{roll}} = \mathcal{M} \cap \widehat{\mathcal{Z}}^0$, and $\Gamma^0_{\text{roll}} = \partial\mathcal{M} \cap \widehat{\mathcal{Z}}^0 \subset \Gamma_{\text{roll}}$. As $\mathcal{M}$ is invariant under translation by ρe_3, so is $\mathcal{M}_{\text{roll}}$ and Γ_{roll}. Consequently, it is enough to show that $F^0(\overline{W^c} \times \mathcal{Z}^0) = \mathcal{M}^0_{\text{roll}}$ and $F^0(\partial W \times \mathcal{Z}^0) = \Gamma^0_{\text{roll}}$.

We observe that F^0 is just the restriction of g to $\overline{W^c} \times \mathcal{Z}^0$. In particular, F^0 maps a point $(p_1, p_2, \alpha) \in \overline{W^c} \times \mathcal{Z}^0$ to a configuration $y = (x_1, x_2, \alpha)$ such that $S_0(y) = (p_1, p_2)$. It follows from Claim 4.2.1 in the proof of (ii) that at most one satellite of $D(y)$ can lie in the strip $\mathbb{R} \times [-1-\epsilon, -1]$ and all other satellites lie in $\mathbb{R} \times (0, \infty)$. If one satellite lies in the strip, then as $S_0(y)$ has the minimal x_2-coordinate of all the satellites of $D(y)$, it follows that this satellite must be $S_0(y)$. As $S_0(y) = (p_1, p_2) \in \overline{W^c}$, it follows that $y \in \mathcal{M}$. This proves $F^0(\overline{W^c} \times \mathcal{Z}^0) \subset \mathcal{M}^0_{\text{roll}}$.

Moreover, if we suppose that $(p_1, p_2, \alpha) \in \partial W \times \mathcal{Z}^0$, then the same reasoning goes through, but $S_0(y) \in \partial W$. Hence $y \in \Gamma^0_{\text{roll}}$, and this proves $F^0(\partial W \times \mathcal{Z}^0) \subset \Gamma^0_{\text{roll}}$.

For the reverse inclusions, note that if $y = (x_1, x_2, \alpha) \in \mathcal{M}^0_{\text{roll}}$, then $(p_1, p_2) := S_0(y) \in \overline{W^c}$, so $F^0(p_1, p_2, \alpha) = y$, and this proves $F^0(\overline{W^c} \times \mathcal{Z}^0) \supset \mathcal{M}^0_{\text{roll}}$.

Moreover, if we suppose that $y \in \Gamma^0_{\text{roll}}$, then exactly one satellite must lie in ∂W, by (ii). Since $\alpha \in \mathcal{Z}^0$, this satellite is S_0. Thus if we take $(p_1, p_2) := S_0(x_1, x_2, \alpha)$, then $F^0(p_1, p_2, \alpha) = (x_1, x_2, \alpha)$, and this proves $F^0(\partial W \times \mathcal{Z}^0) \supset \Gamma^0_{\text{roll}}$. □

Proof of Corollary 4.5. We verify the definition of the class $\text{CES}^2_0(\mathbb{R}^3)$, given in

§6.1.1. Taking $\mathcal{S} = \mathcal{S}$ as in Proposition 4.1(i), it is immediate that condition C1 of the definition holds. To see that condition C2 holds, let $y \in \mathcal{M} \setminus \mathcal{S}$. If $y \in \operatorname{Int} \mathcal{M}$, then trivially there exists a neighborhood of y which is C^2-diffeomorphic to $\mathbb{R}^3$. On the other hand, if $y \in \partial\mathcal{M} \setminus \mathcal{S}$, then we apply Proposition 4.1(i). □

Remark 4.6. One can extract from the proof of Proposition 4.2 a formula for the unit normal vector to the configuration space boundary at a point $q = (p_1, p_2, \alpha) \in \partial\mathcal{M} \setminus \mathcal{S}$. Indeed, for such a q, there exists a unique satellite of $D(q)$ which lies in ∂W. Suppose that $S_0(q) \in \partial W$. The proof of (i) tells us that in a neighborhood of q, for some $\delta > 0$, the boundary $\partial\mathcal{M}$ is parametrized by

$$\widetilde{g}(t, \beta) = g(p(t), \beta), \qquad (t, \beta) \in (-\delta, \delta) \times (\alpha - \delta, \alpha + \delta), \tag{4.37}$$

where $p : (-\delta, \delta) \to \partial_{\mathrm{reg}} W$ is a parametrization of an open subset of the regular part of ∂W, chosen so that $p(0) = p = (p_1, p_2)$. Let $k = (k_1, k_2)$ denote the unit normal to ∂W at the point p, with respect to the Euclidean inner product on $\mathbb{R}^2$. Let

$$n(q) = \left(m^{-1} + J^{-1}(k_1 \cos\alpha + k_2 \sin\alpha)^2\right)^{-1/2} \begin{pmatrix} m^{-1} k_1 \\ m^{-1} k_2 \\ J^{-1}(k_1 \cos\alpha + k_2 \sin\alpha) \end{pmatrix}. \tag{4.38}$$

One may easily check that $\|n\| = 1$, and that $\langle \partial_t \widetilde{g}(0, \alpha), n \rangle = \langle \partial_\alpha \widetilde{g}(0, \alpha), n \rangle = 0$. Consequently, n is the unit normal to $\partial\mathcal{M}$ at the point q.

By symmetry, if more generally $q = (p_1, p_2, \alpha)$ is a point in $\partial\mathcal{M} \setminus \mathcal{S}$ such that $S_j(q) \in \partial W$ and $k = (k_1, k_2)$ is the unit normal to ∂W at the point $S_j(q)$, then the

unit normal to $\partial\mathcal{M}$ at q is given by the formula

$$n(q) = R^{-1}\begin{pmatrix} m^{-1}k_1 \\ m^{-1}k_2 \\ J^{-1}\left(k_1\cos(\alpha - j\rho) + k_2\sin(\alpha - j\rho)\right) \end{pmatrix}, \tag{4.39}$$

where $R := \left(m^{-1} + J^{-1}(k_1\cos(\alpha - j\rho) + k_2\sin(\alpha - j\rho)^2\right)^{1/2}$.

4.2 Cylindrical configuration space

For a fixed wall $W = W(\Sigma, \epsilon)$, the *cylindrical configuration space* is defined by

$$\mathcal{M}_{\text{cyl}} = \{(x_1, x_2, \alpha) : (x_1 + \alpha, x_2 - 1) \in \overline{W^c}\}. \tag{4.40}$$

This is the cylinder with base

$$\widehat{B} := \overline{W^c} + e_2 = \mathcal{M} \cap \{(x_1, x_2, \alpha) : \alpha = 0\} \tag{4.41}$$

and axis

$$\chi := (m + J)^{-1/2}(1, 0, -1) \in \mathbb{S}^2. \tag{4.42}$$

The space $\mathcal{M}_{\text{cyl}}$ is much simpler than $\mathcal{M}$. It is parametrized by

$$f_{\text{lin}}\begin{pmatrix} x_1 \\ x_2 \\ \alpha \end{pmatrix} = \begin{pmatrix} x_1 - \alpha \\ x_2 + 1 \\ \alpha \end{pmatrix}, \qquad (x_1, x_2, \alpha) \in \overline{W^c} \times \mathbb{R}. \tag{4.43}$$

The restriction of F_{lin} to $\overline{W^c} \times \mathbb{R}$ is just the linearization about $\alpha = 0$ of the parametrization F for $\mathcal{M}_{\text{roll}}$, defined by (4.6).

Recall the definition of $\mathcal{Z}^0$ and $\widehat{\mathcal{Z}}^0$ (4.35). The subset of the configuration space $\mathcal{M} \cap \widehat{\mathcal{Z}}^0$ can be expressed as a smooth perturbation of $\mathcal{M}_{\text{cyl}} \cap \widehat{\mathcal{Z}}^0$. To make this statement precise, let $H_1 : \widehat{\mathcal{Z}} \to \widehat{\mathcal{Z}}$ be defined by

$$H_1 \begin{pmatrix} x_1 \\ x_2 \\ \alpha \end{pmatrix} = \begin{pmatrix} x_1 + \overline{\alpha} - \sin(\overline{\alpha}) \\ x_2 - 1 + \cos(\overline{\alpha}) \\ \alpha \end{pmatrix}, \tag{4.44}$$

where $\overline{\alpha} := \alpha - \overline{k}\rho$ and $\overline{k} := \operatorname{argmin}\{|\alpha - k\rho| : k \in \mathbb{Z}\}$. (Note that $\overline{k}$ is uniquely determined since $\mathcal{Z}$ does not contain points of the form $\rho/2 + k\rho$.) The map H_1 is a diffeomorphism with inverse given by

$$H_1^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \alpha \end{pmatrix} = \begin{pmatrix} x_1 - \overline{\alpha} + \sin(\overline{\alpha}) \\ x_2 + 1 - \cos(\overline{\alpha}) \\ \alpha \end{pmatrix}. \tag{4.45}$$

The following identity holds:

$$H_1 \circ f_{\text{lin}} = f \quad \text{on } \overline{W^c} \times \mathcal{Z}^0. \tag{4.46}$$

Consequently, by Proposition 4.2(iii),

$$\begin{aligned} H_1(\mathcal{M}_{\text{cyl}} \cap \widehat{\mathcal{Z}}^0) &= \mathcal{M} \cap \widehat{\mathcal{Z}}^0, \quad \text{and} \\ H_1(\partial\mathcal{M}_{\text{cyl}} \cap \widehat{\mathcal{Z}}^0) &= \partial\mathcal{M} \cap \widehat{\mathcal{Z}}^0. \end{aligned} \tag{4.47}$$

Notice that $|\alpha| \le \rho(\epsilon)/2$ in $\mathcal{Z}^0$. Consequently

$$\begin{aligned} \sup_{\widehat{\mathcal{Z}}^0} \|H_1 - \mathrm{Id}_{\widehat{\mathcal{Z}}^0}\| &\le |\sin(\rho/2) - \rho/2| + |-1 + \cos(\rho/2)| \\ &= O(\rho^2) = o(1). \end{aligned} \tag{4.48}$$

To define the cylindrical collision law in the space $\mathcal{M}_{\text{cyl}}$, we will need the following result.

Lemma 4.7. *$\mathcal{M}_{\text{cyl}}$ belongs to the class* $\mathrm{CES}_0^2(\mathbb{R}^3)$.

Proof. Recall the given decomposition of ∂W into countably many compact C^2 curve segments: $\partial W = \bigcup_{i\in\mathbb{Z}} \Gamma_i$. For each $i \in \mathbb{Z}$, let $\widehat{\Gamma}_i = \Gamma_i + e_2$. For each i, let p_i and p_i' denote the two endpoints of $\widehat{\Gamma}_i$ and let $\operatorname{Int}\widehat{\Gamma}_i = \widehat{\Gamma}_i \setminus \{p_i, p_i'\}$. Let

$$\mathcal{V}_i = \{(x_1, x_2, \alpha) : (x_1 + \alpha, x_2 - 1) \in \operatorname{Int}\widehat{\Gamma}_i\}, \tag{4.49}$$

$$\mathcal{S} = \left\{(x_1, x_2, \alpha) : (x_1 + \alpha, x_2 - 1) \in \bigcup_{i\in\mathbb{Z}} \{p_i, p_i'\}\right\}. \tag{4.50}$$

Then the boundary of $\mathcal{M}_{\text{cyl}}$ may be written as the disjoint union

$$\partial\mathcal{M}_{\text{cyl}} = \mathcal{S} \cup \bigcup_{i\in\mathbb{Z}} \mathcal{V}_i \tag{4.51}$$

The set $\mathcal{S}$ is a countable, locally finite union of parallel lines. Thus $\mathcal{S}$ is closed, and the 2-dimensional Hausdorff measure of $\mathcal{S}$ is zero.

Suppose $y = (x_1, x_2, \alpha) \in \mathcal{M}_{\text{cyl}} \setminus \mathcal{S}$. If $y \in \operatorname{Int} \mathcal{M}_{\text{cyl}}$, then trivially there exists a neighborhood U of y which is entirely contained in $\mathcal{M}_{\text{cyl}}$ and which is C^2-diffeomorphic to $\mathbb{R}^3$. If $y \in \mathcal{V}_i$ for some i, then

$$p := (x_1 + \alpha, x_2 - 1) \in \operatorname{Int} \widehat{\Gamma}_i. \tag{4.52}$$

It follows from the assumptions on the curve segments Γ_i that there exists a neighborhood $\widetilde{U} \subset \mathbb{R}^2$ of p and a diffeomorphism $\widetilde{\phi} : \widetilde{U} \to \mathbb{R}^2$ such that $\widetilde{\phi}(p) = 0$ and $\widetilde{\phi}(\widehat{B} \cap \widetilde{U}) = \mathbb{H}^2 := \mathbb{R} \times [0, \infty)$. Let

$$U = \{(x_1, x_2, \alpha) : (x_1 + \alpha, x_2 - 1) \in \widetilde{U}\}, \tag{4.53}$$

and define $\phi : U \to \mathbb{R}^3$ by

$$\phi(x_1, x_2, \alpha) = (\widetilde{\phi}(x_1 + \alpha, x_2 - 1), \alpha). \tag{4.54}$$

This is a diffeomorphism, and moreover $\phi(\mathcal{M}_{\text{cyl}} \cap U) = \mathbb{R} \times [0, \infty) \times \mathbb{R} \cong \mathbb{H}^3$. This completes the proof. $\square$

4.3 Collision laws

4.3.1 Definition and basic properties

As before, let Σ be a fixed cell, and let $\epsilon > 0$. Recall the plane $\mathbf{P} = \{(x_1, x_2, \alpha) : x_2 = 0\}$, and the hemispheres $\mathbb{S}^2_\pm = \mathbb{S}^2 \cap \{(v_1, v_2, \omega) : \pm v_2 > 0\}$.

In the introduction, we defined the collision law $K^{\Sigma,\epsilon}$ as follows: Let $(y, w) \in \mathbf{P} \times \mathbb{S}^2_+$, and consider the billiard trajectory starting in initial state $(y, -w)$. Suppose the trajectory hits the boundary $\partial\mathcal{M}$ and reflects specularly a certain number of times before returning to the plane $\mathbf{P}$ in a state $(y', w') \in \mathbf{P} \times \mathbb{S}^2_+$. The *collision law (associated with the cell Σ and scale ϵ)* is the mapping

$$K^{\Sigma,\epsilon}(y, w) = (y', w'). \tag{4.55}$$

The domain of this mapping consists of pairs (y, w) such that the billiard trajectory starting from $(y, -w)$ is well defined for all time and returns to the plane $\mathbf{P}$ after only finitely many reflections at regular points of $\partial\mathcal{M}$. For certain initial conditions, the trajectory may not be well-defined for all time. In particular, the trajectory may not in general be continued beyond a point where it hits a singular point of $\partial\mathcal{M}$ (i.e. the set $\mathcal{S}$ defined above) or where it hits the boundary tangentially.

Useful facts about the collision are obtained by observing that the mapping $K^{\Sigma,\epsilon}$ is a special case of the general macro-reflection laws defined in §6, and applying the results derived in that section.

Recall the measure on $\mathbf{P} \times \mathbb{S}^2_+$,

$$\Lambda^2(\mathrm{d}y\mathrm{d}w) = \langle w, e_2 \rangle \mathrm{d}y\sigma(\mathrm{d}w), \tag{4.56}$$

where $\mathrm{d}y$ is Lebesgue measure on $\mathbf{P}$ and $\sigma(\mathrm{d}w)$ is surface measure on $\mathbb{S}^2$. This is just a special case of the measure Λ defined in §6.2.1.

Proposition 4.8. *There exists a full measure open set $\mathcal{F} \subset \mathbf{P} \times \mathbb{S}^2$ such that:*

(i) $K^{\Sigma,\epsilon} : \mathcal{F} \to \mathbf{P} \times \mathbb{S}^2_+$ is a well-defined C^1 mapping.

(ii) $K^{\Sigma,\epsilon}$ maps $\mathcal{F}$ into $\mathcal{F}$ and is an involution in the sense that $K^{\Sigma,\epsilon} \circ K^{\Sigma,\epsilon} = Id_{\mathcal{F}}$. Consequently, $K^{\Sigma,\epsilon} : \mathcal{F} \to \mathcal{F}$ is a C^1 diffeomorphism.

(iii) $K^{\Sigma,\epsilon}$ preserves the measure Λ^2.

Proof. We apply the results of §6.2.1 with $\mathcal{R} = \mathbb{R}^3$, $\mathcal{M}_1 = \mathcal{M}$, $\mathcal{M}_0 = \{(x_1, x_2, \alpha) : x_2 \geq 0\}$, and $\mathcal{N} = \overline{\mathcal{M}_1 \setminus \mathcal{M}_0}$. For these results to hold, we need to check that conditions D1, D2, D3', D4, and D5 from §6.2.1 hold (for the statement of D3', see Remark 6.6). If they do, then in the notation of §6.2.1

$$K^{\Sigma,\epsilon} = P^{\mathcal{M},\mathcal{M}_0}, \tag{4.57}$$

and Proposition 4.8 is just a special case of Proposition 6.5.

Condition D1 is obvious, and condition D2 follows from Corollary 4.5.

By Proposition 4.1(ii), $\mathcal{M}_0 \subset \mathcal{M}$. In the notation of Remark 6.6,

$$A_2 := \mathbf{P} \cap (\overline{\mathcal{M} \setminus \mathcal{M}_0}) \setminus \operatorname{Int} \mathcal{M} \subset \{(x_1, x_2, \alpha) : x_2 \geq 0\} \cap \mathcal{M}, \tag{4.58}$$

and by Proposition 4.1(vii), the two-dimensional Hausdorff measure of the set on the right is zero. Therefore, $\mathcal{H}^2(A_2) = 0$, and condition D3' holds.

To verify condition D4, note that the boundary of $\mathcal{N}$ is contained in $\partial\mathcal{M} \cup \mathbf{P}$. It follows from Proposition 4.1(v) that $\mathcal{N}$ is bounded between the planes $x_2 = 0$ and $x_2 = -\epsilon - \frac{1}{2}\rho^2$. Thus the only way for a billiard trajectory in $\mathcal{N}$ starting from a state (y, w) to not return to the boundary is if the trajectory is parallel to the two planes. Hence $\mathcal{U}^+ \setminus \mathcal{U}_{\text{fin}} \subset \{(y, w) \in \mathcal{U}^+ : \langle w, e_2 \rangle = 0\}$, and this is a measure zero subset of $\mathcal{U}^+$.

To verify condition D5, we use the double periodicity of $\mathcal{M}$. Let $\widetilde{\mathcal{N}}$ be the space obtained by identifying points in $\mathcal{N}$ which are translates of each other by $i\epsilon e_1 + j\rho e_3$, $i, j \in \mathbb{Z}$. Since $\mathcal{N}$ is bounded between the planes $x_2 = 0$ and $x_2 = -\epsilon - \frac{1}{2}\rho^2$, the reduced billiard domain $\widetilde{\mathcal{N}}$ is compact, and the induced invariant measure on $\widetilde{\mathcal{N}}$ is finite. Thus the Poincaré Recurrence Theorem implies that, except on a null set of initial conditions, the billiard trajectory will return to the plane $\mathbf{P}$ after only finitely many collisions with $\partial\mathcal{M}$. □

For $(j, k) \in \mathbb{Z}^2$, define translations

$$\tau_{jk}(y) = y + j\epsilon e_1 + k\rho e_3, \qquad y \in \mathbf{P}, \tag{4.59}$$

and

$$\overline{\tau}_{jk}(y, w) = (\tau_{jk}(y), w), \qquad (y, w) \in \mathbf{P} \times \mathbb{S}^2_+. \tag{4.60}$$

An elementary but important fact is that $K^{\Sigma,\epsilon}$ commutes with these translations.

Proposition 4.9. *For all $(j, k) \in \mathbb{Z}^2$, $\overline{\tau}_{jk} \circ K^{\Sigma,\epsilon} = K^{\Sigma,\epsilon} \circ \overline{\tau}_{jk}$.*

Proof. This is immediate from Proposition 4.1(i), which tells us that the configuration space $\mathcal{M}$ is invariant under the translations $\overline{\tau}_{jk}$. $\square$

The collision law $K^{\Sigma,\epsilon}$ is a associated with a Markov kernel on $\mathbf{P} \times \mathbb{S}^2_+$:

$$\mathbb{K}^{\Sigma,\epsilon}(y, w; \mathrm{d}y'\mathrm{d}w') := \delta_{K^{\Sigma,\epsilon}(y,w)}(\mathrm{d}y'\mathrm{d}w'). \tag{4.61}$$

We say that a Markov kernel $\mathbb{K}(y, w; \mathrm{d}y'\mathrm{d}w')$ is a *rough collision law* if there exist a sequence of cells Σ_i (satisfying conditions B1-B5 of §1.2.4) and positive numbers $\epsilon_i \to 0$ such that

$$\mathbb{K}^{\Sigma_i,\epsilon_i}(y, w; \mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w) \to \mathbb{K}(y, w; \mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w) \tag{4.62}$$

weakly in the space of measures on $\mathbf{P} \times \mathbb{S}^2_+$.

The most important basic property of the rough collision laws is the following:

Proposition 4.10. *If $\mathbb{K}(y, w; \mathrm{d}y'\mathrm{d}w')$ is a rough collision law, then it is symmetric with respect to the measure Λ^2 in the sense that for any $h \in C_c((\mathbf{P} \times \mathbb{S}^2_+)^2)$,*

$$\begin{aligned} &\int_{(\mathbf{P}\times\mathbb{S}^2_+)^2} h(y, w, y', w')\mathbb{K}(y, w; \mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w) \\ &\qquad \int_{(\mathbf{P}\times\mathbb{S}^2_+)^2} h(y', w', y, w)\mathbb{K}(y, w; \mathrm{d}y'\mathrm{d}w')\Lambda^2(\mathrm{d}y\mathrm{d}w). \end{aligned} \tag{4.63}$$

Proof. It follows from the proof of Proposition 4.8 that the collision law is a special case of the macro-reflection law defined in §6.2.1. Therefore, the rough collision law is a special case of the rough reflection law defined in §6.2.2. Thus the proposition follows from Proposition 6.8. $\square$

4.3.2 Cylindrical collision law

Recall the definition of the cylindrical configuration space. By condition A3' of §1.3.1, $\mathbb{R} \times (-\infty, -1-\epsilon] \subset W \subset \mathbb{R} \times (-\infty, -1]$. It thus follows from the definition of $\mathcal{M}_{\text{cyl}}$ that

$$\mathbb{R} \times [0, \infty) \times \mathbb{R} \subset \mathcal{M}_{\text{cyl}} \subset \mathbb{R} \times [-\epsilon, \infty) \times \mathbb{R}. \tag{4.64}$$

Let $(y, w) \in \mathbf{P} \times \mathbb{S}^2_+$. Suppose a point particle starting from initial state $(y, -w)$ hits and reflects specularly from $\partial\mathcal{M}_{\text{cyl}}$ a certain number of times before eventually returning to $\mathbf{P}$ in a state $(y', w') \in \mathbf{P} \times \mathbb{S}^2_+$. By definition, the *cylindrical collision law (associated with cell Σ and scale ϵ)* is the mapping

$$K^{\Sigma,\epsilon}_{\text{cyl}}(y, w) = (y', w'). \tag{4.65}$$

In the notation of §6.2.1 with $\mathcal{M}_0 = \{(x_1, x_2, \alpha) : x_2 \geq 0\}$,

$$K^{\Sigma,\epsilon}_{\text{cyl}} := P^{\mathcal{M}_{\text{cyl}}, \mathcal{M}_0}. \tag{4.66}$$

As in the previous case, $K^{\Sigma,\epsilon}_{\text{cyl}}$ is well-defined on a full-measure subset of $\mathbf{P} \times \mathbb{S}^2_+$, and the following proposition holds.

Proposition 4.11. *There exists a full measure open set $\mathcal{F}_{\text{cyl}} \subset \mathbf{P} \times \mathbb{S}^2$ such that:*

(i) $K^{\Sigma,\epsilon}_{\text{cyl}} : \mathcal{F}_{\text{cyl}} \to \mathbf{P} \times \mathbb{S}^2_+$ is a well-defined C^1 mapping.

(ii) $K^{\Sigma,\epsilon}_{\text{cyl}}$ maps $\mathcal{F}_{\text{cyl}}$ into $\mathcal{F}_{\text{cyl}}$ and is an involution in the sense that $K^{\Sigma,\epsilon} \circ K^{\Sigma,\epsilon} = Id_{\mathcal{F}_{\text{cyl}}}$. Consequently, $K^{\Sigma,\epsilon} : \mathcal{F}_{\text{cyl}} \to \mathcal{F}_{\text{cyl}}$ is a C^1 diffeomorphism.

(iii) $K^{\Sigma,\epsilon}_{\text{cyl}}$ preserves the measure Λ^2.

Proof. The situation is similar to that of Proposition 4.8. Taking $\mathcal{M}_1 = \mathcal{M}_{\text{cyl}}$, $\mathcal{M}_0 = \{(x_1, x_2, \alpha) : x_2 \geq 0\}$, and $\mathcal{N} = \overline{\mathcal{M}_1 \setminus \mathcal{M}_0}$, it is sufficient to check that conditions D1, D2, D3', D4, and D5 from §6.2.1 hold. The proposition then follows as a special case of Proposition 6.5.

The verification of conditions D1, D2, D4, and D5 is almost identical to the verification of the same conditions in the proof of Proposition 4.8, so we omit it.

To verify condition D3', we must show that the 2-dimensional Hausdorff measure of

$$A_2 := \mathbf{P} \cap (\overline{\mathcal{M}_{\text{cyl}} \setminus \mathcal{M}_0}) \setminus \operatorname{Int} \mathcal{M}_{\text{cyl}} \tag{4.67}$$

is zero. Let $G : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$G(x_1, x_2, \alpha) = (x_1 + \alpha, x_2). \tag{4.68}$$

Since $\mathcal{M}_{\text{cyl}}$ is the cylinder with base $\widehat{B}$ and axis χ, we have $\mathcal{M}_{\text{cyl}} = G^{-1}(\widehat{B})$. Let $\mathbf{L}$ be the line $\{(x_1, x_2) : x_2 = 0\} \subset \mathbb{R}^2$. As the set A_2 is cylindrical with axis χ, the projection of A_2 under G onto $\mathbb{R}^2$ is

$$\widetilde{A}_2 := \mathbf{L} \cap \overline{\{(x_1, x_2) \in \widehat{B} : x_2 < 0\}} \setminus \operatorname{Int} \widehat{B}. \tag{4.69}$$

Since $\mathbf{L}$ is a subset of $\widehat{B}$, we have $\widetilde{A}_2 \subset \partial \widehat{B}$.

Fix any $q \in \widetilde{A}_2$, and let $\{\widehat{\Gamma}_i, i \in \mathbb{Z}\}$, be the collection of curve segments constituting the boundary of $\widehat{B}$, as in the proof of Lemma 4.7. Evidently, $q \in \widehat{\Gamma}_i$ for some i. But by our assumptions on the Γ_i, if the interior of $\widehat{\Gamma}_i$ intersects $\mathbf{L}$, then $\widehat{\Gamma}_i \subset \mathbf{L}$. Thus

either q is an endpoint of $\widehat{\Gamma}_i$, or $q \in \operatorname{Int}\widehat{\Gamma}_i \subset \mathbf{L}$. In the second case we may choose a neighborhood U of q small enough that $U \cap \widehat{B} = U \cap \{(x_1, x_2) : x_2 \geq 0\}$. But then q is not in the closure of $\{(x_1, x_2) \in \widehat{B} : x_2 < 0\}$, contrary to our assumption. We conclude that q is an endpoint of $\widehat{\Gamma}_i$.

Let E be the set of endpoints of the Γ_i's. By the argument above, $\widetilde{A}_2 \subset E$, and thus $A_2 \subset G^{-1}(E)$ is a discrete collection of lines. Therefore $\mathcal{H}^2(A_2) = 0$. □

For $(j,k) \in \mathbb{Z}^2$ and $s \in \mathbb{R}$, define translations

$$\tau_{jk}^{(s)}(y) = y + j\epsilon e_1 + ks\chi, \qquad y \in \mathbf{P}, \tag{4.70}$$

and

$$\overline{\tau}_{jk}^{(s)}(y,w) = (\tau_{jk}^{(s)}(y), w), \qquad (y,w) \in \mathbf{P} \times \mathbb{S}^2_+. \tag{4.71}$$

The cylindrical collision law $K_{\text{cyl}}^{\Sigma,\epsilon}$ commutes with these translations.

Proposition 4.12. *For all* $(j,k) \in \mathbb{Z}^2$, $\overline{\tau}_{jk}^{(s)} \circ K_{\text{cyl}}^{\Sigma,\epsilon} = K_{\text{cyl}}^{\Sigma,\epsilon} \circ \overline{\tau}_{jk}^{(s)}$.

Proof. Since $\mathcal{M}_{\text{cyl}}$ is a cylinder with base $\widehat{B}$ and axis χ, and the base is ϵ-periodic in the e_1 direction, it follows that $\mathcal{M}_{\text{cyl}}$ is invariant under the translations $\tau_{jk}^{(s)}$. This implies that $K_{\text{cyl}}^{\Sigma,\epsilon}$ commutes with the translations $\overline{\tau}_{jk}^{(s)}$. □

4.3.3 Modified collision law

When we compare the mappings $K^{\Sigma,\epsilon}$ and $K_{\text{cyl}}^{\Sigma,\epsilon}$, the modified collision law will serve as a kind of intermediate mapping to which we may compare both. As the definition of the modified collision law is somewhat technical, the reader may wish to refer to

§5.2 before reading this section for better motivation.

The plane $\mathbf{P}$ forms the boundary of the two half-spaces $\mathbb{R}^3_+ := \{(x_1, x_2, \alpha) : x_2 > 0\}$ and $\mathbb{R}^3_- := \{(x_1, x_2, \alpha) : x_2 < 0\}$. Recall the mapping $H_1 : \widehat{\mathcal{Z}} \to \widehat{\mathcal{Z}}$ defined by (4.44). We define subsets of $\widehat{\mathcal{Z}}$:

$$\widetilde{\mathbf{P}} = H_1(\mathbf{P} \cap \widehat{\mathcal{Z}}), \tag{4.72}$$

$$\mathcal{O}_+ = H_1(\mathbb{R}^3_+ \cap \widehat{\mathcal{Z}}), \tag{4.73}$$

$$\mathcal{O}_- = H_1(\mathbb{R}^3_- \cap \widehat{\mathcal{Z}}). \tag{4.74}$$

Then $\mathcal{O}_+$ and $\mathcal{O}_-$ are disconnected open subsets of $\widehat{\mathcal{Z}}$ satisfying

$$\overline{\mathcal{O}_+} \cup \overline{\mathcal{O}_-} = \widehat{\mathcal{Z}} \quad \text{and} \quad \overline{\mathcal{O}_+} \cap \overline{\mathcal{O}_-} = \widetilde{\mathbf{P}}, \tag{4.75}$$

where closure is taken in $\widehat{\mathcal{Z}}$.

As H_1 fixes the α-coordinate, H_1 maps $\mathbb{R}^3_- \cap \widehat{\mathcal{Z}}$ into itself. Since $\partial\mathcal{M}_{\text{cyl}} \subset \overline{\mathbb{R}^3_-}$, and $\mathbf{P} \subset \overline{\mathbb{R}^3_+}$, it follows from (4.47) that

$$\partial\mathcal{M} \cap \widehat{\mathcal{Z}} \subset \overline{\mathcal{O}_-} \text{ and } \mathbf{P} \cap \widehat{\mathcal{Z}} \subset \overline{\mathcal{O}_+}. \tag{4.76}$$

The set $\widetilde{\mathbf{P}}$ is a smoothly embedded surface in $\widehat{\mathcal{Z}}$. For $(x_1, \alpha) \in \mathbf{P} \cap \widehat{\mathcal{Z}}$, we define

$$u(x_1, \alpha) = \cos\overline{\alpha} - 1, \tag{4.77}$$

where $\overline{\alpha} = \alpha - \overline{k}\rho$ and $\overline{k} = \operatorname{argmin}\{|\alpha - k\rho| : k \in \mathbb{Z}\}$. One may verify directly from the

definition of H_1 that $\widetilde{\mathbf{P}}$ is the graph of u, i.e.

$$\widetilde{\mathbf{P}} = \{(x_1, x_2, \alpha) \in \widehat{\mathcal{Z}} : x_2 = u(x_1, \alpha)\}. \tag{4.78}$$

Two useful observations about u are the following: First, for any $k \in \mathbb{Z}$, the restriction of u to $\mathbf{P} \cap \mathbb{R}^2 \times (k\rho - \rho/2 + \delta_0, k\rho + \rho/2 - \delta_0)$ is concave. Second, we have the estimate on the gradient:

$$\|\nabla u(x_1, \alpha)\| \leq |\sin \overline{\alpha}| \leq \sin(\rho/2) = o(1), \qquad \text{as } \epsilon \to 0. \tag{4.79}$$

Thus $\widetilde{\mathbf{P}}$ is almost "flat" and almost parallel to $\mathbf{P}$.

Let $\mathbb{S}^2_{+\rho/2} = \{w \in \mathbb{S}^2 : \langle w, e_2 \rangle > \sin(\rho/2)\}$, and define $\Psi : \widetilde{\mathbf{P}} \times \mathbb{S}^2_{+\rho/2} \to \mathbf{P} \times \mathbb{S}^2_{+\rho/2}$ by

$$\Psi(y, w) = \left(y - \frac{\langle y, e_2 \rangle}{\langle w, e_2 \rangle} w, w \right). \tag{4.80}$$

That is, Ψ maps (y, w) to (y', w), where y' is the point of intersection of the ray $\{y + tw : t \geq 0\}$ with $\mathbf{P}$. The fact that, for $(y, w) \in \widetilde{\mathbf{P}} \times \mathbb{S}^2_{+\rho/2}$, the ray makes an angle of at most $\pi/2 - \rho/2$ with e_2 guarantees that the ray cannot intersect $\widetilde{\mathbf{P}}$ at any point other than its initial point and thus Ψ is injective. In fact, the following is true:

Lemma 4.13. *Ψ is a diffeomorphism onto an open subset of $\mathbf{P} \times \mathbb{S}^2_{+\rho/2}$.*

Proof. By injectivity, it is enough to show that Ψ is a local diffeomorphism. Let $\overline{\Psi}$ denote a function given by the formula (4.80) and defined on an open subset of

$\mathbb{R}^3 \times \mathbb{R}^3$ containing $\widetilde{\mathbf{P}} \times \mathbb{S}^2_{+\rho/2}$. The differential of $\overline{\Psi}$ takes the form

$$\mathrm{d}\overline{\Psi}(x_1, x_2, \alpha, v_1, v_2, \omega) = \begin{bmatrix} 1 & -\frac{v_1}{v_2} & 0 & \\ 0 & 0 & 0 & * \\ 0 & -\frac{\omega}{v_2} & 1 & \\ & & & \\ & 0_{3\times 3} & & I_{3\times 3} \\ & & & \end{bmatrix}, \tag{4.81}$$

where $0_{3\times 3}$ and $I_{3\times 3}$ are the 3×3 zero matrix and identity matrix respectively. The null space of this matrix is

$$\mathrm{Span}(\overline{w}), \quad \text{where } \overline{w} = (v_1, v_2, \omega, 0, 0, 0) \in \mathbb{R}^6. \tag{4.82}$$

A consequence of the estimate (4.79) is that if $w = (v_1, v_2, \omega) \in \mathbb{S}^2_{+\rho/2}$, then w cannot be tangent to $\widetilde{\mathbf{P}}$ at (x_1, x_2, α); otherwise $v_2 \leq \nabla g(x_1, \alpha) \leq \sin(\rho/2)$. Thus $\overline{w}$ is not tangent to $\widetilde{\mathbf{P}} \times \mathbb{S}^2_+ \subset \mathbb{R}^3 \times \mathbb{R}^3$. Since Ψ is the restriction of $\overline{\Psi}$ to $\widetilde{\mathbf{P}} \times \mathbb{S}^2_{+\rho/2}$, we conclude that the differential of Ψ has full rank, and the result follows by the inverse function theorem. □

Set

$$\overline{H}_1(y, w) = (H_1(y, w), w) \qquad \text{for } (y, w) \in \widehat{\mathcal{Z}} \times \mathbb{S}^2. \tag{4.83}$$

This is a diffeomorphism of $\widehat{\mathcal{Z}} \times \mathbb{S}^2$. Define $\eta : \mathbf{P} \cap \widehat{\mathcal{Z}} \times \mathbb{S}^2_{+\rho/2} \to \mathbf{P} \cap \widehat{\mathcal{Z}} \times \mathbb{S}^2_{+\rho/2}$ by

$$\eta(y, w) = \Psi \circ \overline{H}_1(y, w). \tag{4.84}$$

This is a diffeomorphism onto its image because $\overline{H}_1$ and Ψ are.

Recall the full-measure subset open subset $\mathcal{F} \subset \mathbf{P} \times \mathbb{S}^2_+$ on which the collision law $K^{\Sigma,\epsilon} : \mathcal{F} \to \mathcal{F}$ is a well-defined involutive C^1 diffeomorphism, and let

$$\widetilde{\mathcal{F}} = \eta^{-1}(\mathcal{F} \cap K^{\Sigma,\epsilon}(\operatorname{Im}\eta))) \subset \mathbf{P} \times \mathbb{S}^2_{+\rho/2}. \tag{4.85}$$

We define the *modified collision law* $\widetilde{K}^{\Sigma,\epsilon} : \widetilde{\mathcal{F}} \to \widetilde{\mathcal{F}}$ by

$$\widetilde{K}^{\Sigma,\epsilon} = \eta^{-1} \circ K^{\Sigma,\epsilon} \circ \eta. \tag{4.86}$$

This is an involutive C^1 diffeomorphism of $\widetilde{\mathcal{F}}$ because $K^{\Sigma,\epsilon}$ is an involutive C^1 diffeomorphism of $\mathcal{F}$.

Note that $\widetilde{\mathcal{F}}$ is not a full-measure subset of $\mathbf{P} \times \mathbb{S}^2$. Nonetheless, we will see below that $\Lambda^2(B \setminus \widetilde{\mathcal{F}}) \to 0$ as $\epsilon \to 0$ for any set B of finite Λ^2-measure (see Lemma 5.3 and Remark 5.4).

Recall the translation maps τ_{jk} and $\overline{\tau}_{jk}$ defined by (4.59) and (4.60). It is easy to show directly that η commutes with $\overline{\tau}_{jk}$. This observation and Proposition 4.9 give us

Proposition 4.14. *For all* $(j,k) \in \mathbb{Z}^2$, $\overline{\tau}_{jk} \circ \widetilde{K}^{\Sigma,\epsilon} = \widetilde{K}^{\Sigma,\epsilon} \circ \overline{\tau}_{jk}$.

We think of $\widetilde{K}^{\Sigma,\epsilon}$ as a collision law obtained by "modifying" $K^{\Sigma,\epsilon}$ in the spatial coordinates by η. Let us consider the perturbation η in more detail. We denote its domain more succinctly by

$$\mathcal{G} := (\mathbf{P} \cap \widehat{\mathcal{Z}}) \times \mathbb{S}^2_{+\rho/2}. \tag{4.87}$$

Let $(y, w) \in \mathcal{G}$ and let $(y', w') = \eta(y, w) \in \mathbf{P} \times \mathbb{S}^2_+$. Then $w' = w$, and

$$\begin{aligned} \|y' - y\| &\leq \|y - H_1(y)\| + \left| \frac{\langle H_1(y), e_2 \rangle}{\langle w, e_2 \rangle} \right| \\ &\leq |\alpha - \sin \overline{\alpha}| + |-1 + \cos \overline{\alpha}| + \frac{|-1 + \cos \overline{\alpha}|}{\sin(\rho/2)} \leq C\rho \end{aligned} \tag{4.88}$$

for some constant C, here using $|\overline{\alpha}| < \rho/2$. This shows that

$$\|\eta - \mathrm{Id}_{\mathcal{G}}\|_{L^\infty(\mathcal{G})} \leq C\rho = o(1) \quad \text{as } \epsilon \to 0. \tag{4.89}$$

Since $\eta : \mathcal{G} \to \eta(\mathcal{G})$ is a diffeomorphism, we also have the following estimate:

$$\|\mathrm{Id}_{\eta(\mathcal{G})} - \eta^{-1}\|_{L^\infty(\eta(\mathcal{G}))} \leq C\rho = o(1) \quad \text{as } \epsilon \to 0. \tag{4.90}$$

5 Proofs of Main Results

In this section we prove Theorems 1.27, 1.28, and 1.31. For a short summary of our arguments, see §1.1.3.

5.1 Cylindrical configuration space

Our first task is to show that a version of Theorem 1.31 holds if we replace $\mathcal{M}$ with $\mathcal{M}_{\text{cyl}}$.

Theorem 5.1. *Given a sequence of cells Σ_i, there exists decreasing sequence of positive numbers $\{b_i\}$ such that exactly one of the following is true:*

(A) There exists a Markov kernel $\mathbb{K}$ such that, for any sequence $\epsilon_i \leq b_i$ with $\epsilon_i \to 0$, the limit $\lim_{i\to\infty} \mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i}$ exists and is equal to $\mathbb{K}$.

(B) For any sequence of positive numbers $\epsilon_i \leq b_i$ with $\epsilon_i \to 0$, the limit $\lim_{i\to\infty} \mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i}$ does not exist.

If (A) holds, then $\mathbb{K}$ takes the following form

$$\mathbb{K}(y_1, y_2, \theta, \psi; \mathrm{d}y_1'\mathrm{d}y_2'\mathrm{d}\theta'\mathrm{d}\psi') = \delta_{(y_1,y_2)}(y_1', y_2')\mathrm{d}y_1'\mathrm{d}y_2' \times \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') \times \delta_{\pi-\psi}(\psi')\mathrm{d}\psi', \quad (5.1)$$

where $\widetilde{\mathbb{P}}$ is a Markov kernel on $\mathbb{S}^1_+$ satisfying the following properties:

i. $\widetilde{\mathbb{P}}$ is symmetric with respect to the measure $\sin\theta\mathrm{d}\theta$ on $\mathbb{S}^1_+$.

ii. Let

$$\begin{aligned}
\widetilde{\Sigma}_i &= \{(y_1, y_2) : (y_1, (1 + mJ^{-1})^{-1/2} y_2) \in \Sigma_i\}, \\
\widetilde{\epsilon}_i &= (1 + mJ^{-1})^{-1/2} \epsilon_i, \\
\mathbb{P} &= \delta_{y_1}(y_1') \mathrm{d}y_1' \times \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta').
\end{aligned} \tag{5.2}$$

Then

$$\mathbb{P} = \lim_{i \to \infty} \mathbb{P}^{\widetilde{\Sigma}_i, \widetilde{\epsilon}_i}. \tag{5.3}$$

Consequently, $\mathbb{K}$ *is symmetric with respect to the measure* Λ^2*.*

Moreover, if the sequence of cells $\Sigma_i = \Sigma$ *is constant, then we may take* $b_i = \infty$ *and (A) always holds.*

Conversely, if $\mathbb{K}$ *is a Markov kernel on* $\mathbf{P} \times \mathbb{S}^2_+$ *of form (5.1) such that the measure* $\sin\theta \mathrm{d}\theta$ *on* $\mathbb{S}^1_+$ *is invariant with respect to* $\widetilde{\mathbb{P}}$*, then there exist a sequence of cells* $\{\Sigma_i\}$ *and a decreasing sequence of positive numbers* $\{b_i\}$ *such that* $\mathbb{K} = \lim_{i \to \infty} \mathbb{K}^{\Sigma_i, \epsilon_i}_{\mathrm{cyl}}$ *whenever* $\epsilon_i \to 0$ *and* $0 < \epsilon_i \le b_i$*.*

Proof. First, we will define the sequence b_i. Define the parallelogram

$$R_{1\times 1} = \{(x_1, \alpha) \in \mathbf{P} : 0 \le x_1 + \alpha \le 1, 0 \le \alpha \le 1\}. \tag{5.4}$$

By Proposition 4.11, $K^{\Sigma_i, 1}_{\mathrm{cyl}}$ is well-defined and finite Λ^2-almost everywhere. Since $\Lambda^2(R_{1\times 1} \times \mathbb{S}^2_+) < \infty$, for each i there exists a constant C_i such that

$$\Lambda^2(\{(y, w) \in R_{1\times 1} \times \mathbb{S}^2_+ : \|K^{\Sigma_i, 1}_{\mathrm{cyl}}(y, w) - (y, w)\| \ge C_i\}) \le i^{-1}. \tag{5.5}$$

By replacing each C_i with $\max\{C_1, \ldots, C_i\}$, we may suppose that the constants C_i are increasing. We define

$$b_i = C_i^{-1} i^{-1}. \tag{5.6}$$

We assume possibility (B) does not hold, i.e. we assume that $\mathbb{K} = \lim_{i\to\infty} \mathbb{K}_{\mathrm{cyl}}^{\Sigma_i, \epsilon_i^{(0)}}$ does exist for some sequence of positive numbers $\epsilon_i^{(0)} \le b_i$ such that $\epsilon_i^{(0)} \to 0$. We will argue that (A) holds and that $\mathbb{K}$ takes the form described above.

Fix any sequence of positive numbers $\epsilon_i \le b_i$ with $\epsilon_i \to 0$, and consider the billiard trajectory in $\mathcal{M}_{\mathrm{cyl}} = \mathcal{M}_{\mathrm{cyl}}(\epsilon_i)$.

Step 1a. The key fact is that the billiard evolution decouples into two independent evolutions. To describe this decoupling, first recall that a point particle in $\mathcal{M}_{\mathrm{cyl}}$ moves linearly in $\operatorname{Int} \mathcal{M}_{\mathrm{cyl}}$ and reflects specularly from the boundary $\partial\mathcal{M}_{\mathrm{cyl}}$. Note that, for every $p \in \partial\mathcal{M}_{\mathrm{cyl}}$, χ is tangent to the boundary $\partial\mathcal{M}_{\mathrm{cyl}}$ at p. Therefore, specular reflection preserves the angle between the velocity of the point particle and χ. Hence, the angle between the velocity and χ is conserved for all time.

In the coordinates (y_1, y_2, y_3), the y_3-axis is parallel to the cylindrical axis χ. Consequently, $\mathcal{M}_{\mathrm{cyl}}$ may be identified with the product space $\mathcal{M}_{\mathrm{cyl}}^{1,2} \times \mathbb{R}$, where $\mathcal{M}_{\mathrm{cyl}}^{1,2} = \chi^\perp \cap \mathcal{M}_{\mathrm{cyl}} = \{(y_1, y_2) : (y_1, y_2, 0) \in \mathcal{M}_{\mathrm{cyl}}\}$. The trajectory $y(t)$ of the point particle in $\mathcal{M}_{\mathrm{cyl}}$ decouples into a pair of independent trajectories $(y_{1,2}(t), y_3(t)) \in \mathcal{M}_{\mathrm{cyl}}^{1,2} \times \mathbb{R}$. The trajectory $y_{1,2}(t)$ moves linearly in the interior of $\mathcal{M}_{\mathrm{cyl}}^{1,2}$ with velocity $\dot{y}_{1,2}(t) = \dot{y}(t) - \langle \dot{y}(t), \chi\rangle\chi$, and reflects specularly from $\partial\mathcal{M}_{\mathrm{cyl}}^{1,2}$. The trajectory $y_3(t)$ moves freely in $\mathbb{R}$ with constant velocity $\dot{y}_3 = \langle \dot{y}(t), \chi\rangle = \langle \dot{y}(0), \chi\rangle$ for all time.

Step 1b. Let us consider the two-dimensional billiard in $\mathcal{M}_{\mathrm{cyl}}^{1,2}$ in more detail. Define planes $\mathbf{Q}_0 = \{(x_1, x_2, \alpha) : \alpha = 0\}$ and $\mathbf{Q}_1 = \chi^\perp = \{(y_1, y_2, y_3) : y_3 = 0\}$. The

angle between these two planes (with respect to the kinetic energy inner product) is $\gamma = \arccos(\langle e_3, \chi\rangle) = \arccos((1 + mJ^{-1})^{-1/2})$. Let $\pi_0 : \mathbf{Q}_0 \to \mathbf{Q}_1$ denote orthogonal projection from $\mathbf{Q}_0$ onto $\mathbf{Q}_1$. The base of the cylinder $\mathcal{M}_{\text{cyl}}$ is $\widehat{B} \subset \mathbf{Q}_0$, and therefore $\mathcal{M}^{1,2}_{\text{cyl}} = \pi_0(\widehat{B})$. Identifying $\mathbf{Q}_0$ with $\mathbb{R}^2$ with coordinates (x_1, x_2) and $\mathbf{Q}_1$ with $\mathbb{R}^2$ with coordinates (y_1, y_2), π_0 is just the "foreshortening map"

$$\pi_0 : (x_1, x_2) \mapsto (y_1, y_2) = ((1 + mJ^{-1})^{-1/2} x_1, x_2). \tag{5.7}$$

Thus, the particle with position $y_{1,2}(t)$ moves freely in the complement of the "foreshortened" wall

$$\widetilde{W} := \{(y_1, y_2) \in \mathbf{Q}_1 : ((1 + mJ^{-1})^{1/2} y_1, y_2) \in W + e_2\}, \tag{5.8}$$

and reflects specularly from $\partial\widetilde{W}$. The boundary $\partial\widetilde{W}$ is piecewise C^2 and bounded between the lines $y_2 = -\epsilon$ and $y_2 = 0$, because ∂W is piecewise C^2 and bounded between the lines $y_2 = -1 - \epsilon$ and $y_2 = -1$. Letting $\mathbf{L} = \{(y_1, y_2) : y_2 = 0\}$, we may therefore define a macro-reflection law $P^{\widetilde{\Sigma}_i, \widetilde{\epsilon}_i} : \mathbf{L} \times (0, \pi) \to \mathbf{L} \times (0, \pi)$, in the way described in §1.2.2 (with $\mathbf{L}$ playing the role of $\mathbb{R}$).

Step 1c. We now describe the marginals of the collision law. The relationship between the macroscopic reflection law $P^{\widetilde{\Sigma}_i, \widetilde{\epsilon}_i}$ and the collision law $K^{\Sigma_i, \epsilon_i}_{\text{cyl}}$ is as follows. Note that $\mathbf{L} = \mathbf{P} \cap \mathbf{Q}_1$. The macro-reflection law $P^{\widetilde{\Sigma}_i, \widetilde{\epsilon}_i}$ describes the orthogonal projection onto $\mathbf{Q}_1$ of the state of the point particle after returning to the plane $\mathbf{P}$. In more detail, letting $\pi_{1,3} : \mathbf{P} \times (0, \pi)^2 \to \mathbb{R} \times (0, \pi)$ be the mapping $(y_1, y_3, \theta, \psi) \mapsto$

(y_1, ψ), we have

$$\pi_{1,3} \circ K_{\text{cyl}}^{\Sigma_i,\epsilon_i}(y_1, y_3, \theta, \psi) = P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1, \theta). \tag{5.9}$$

Note that the right-hand side only depends on y_1 and θ.

We also consider the projections $\pi_2 : \mathbf{P} \times (0,\pi)^2 \to \mathbb{R}$ mapping $(y_1, y_3, \theta, \psi) \mapsto y_3$, and $\pi_4 : \mathbf{P} \times (0,\pi)^2 \to (0,\pi)$ mapping $(y_1, y_2, \theta, \psi) \mapsto \psi$. We define

$$Q_2(y_1, y_3, \theta, \psi) = \pi_2 \circ K_{\text{cyl}}^{\Sigma_i,\epsilon_i}(y_1, y_3, \theta, \psi), \tag{5.10}$$

$$Q_4(y_1, y_3, \theta, \psi) = \pi_4 \circ K_{\text{cyl}}^{\Sigma_i,\epsilon_i}(y_1, y_3, \theta, \psi). \tag{5.11}$$

Suppose $(y, w) \in \mathbf{P} \times \mathbb{S}^2_+$ with coordinates (y_1, y_2, θ, ψ), and let $(y', w') = K^{\Sigma_i,\epsilon_i}(y, w)$ with coordinates $(y_1', y_2', \theta', \psi')$. Then, since the angle between the velocity and χ is conserved for all time, $\cos\psi' = \langle w', \chi\rangle = \langle -w, \chi\rangle = \cos(\pi - \psi)$. Hence, $\psi' = \pi - \psi$. Therefore, Q_4 only depends on ψ, and

$$Q_4(\psi) = \pi - \psi. \tag{5.12}$$

To describe Q_2, let us write

$$Q_2(y_1, y_3, \theta, \psi) = y_3 + E(y_1, y_3, \theta, \psi). \tag{5.13}$$

The definition of E makes sense for any choice of cell Σ and roughness scale ϵ. In the argument below, we will indicate explicitly the dependence of E on Σ and ϵ by writing $E = E_\Sigma^\epsilon$. We will prove

Claim 5.1.1. *For any bounded set $B \subset \mathbf{P} \times \mathbb{S}^2_+$, there exists a constant C_B depending only on B such that*

$$\Lambda^2(B \cap \{(y,w) : |E^{\epsilon_i}_{\Sigma_i}(y,w)| \geq i^{-1}\}) \leq C_B i^{-1}. \tag{5.14}$$

Proof of Claim 5.1.1. The key observation is that the change of spatial coordinates $y \mapsto \epsilon_i^{-1} y$ maps $\mathcal{M}_{\mathrm{cyl}}(\epsilon_i)$ to $\mathcal{M}_{\mathrm{cyl}}(1)$. Consequently $K^{\Sigma_i,1}_{\mathrm{cyl}}$ and $E^1_{\Sigma_i}$ may be viewed respectively as $K^{\Sigma_i,\epsilon_i}_{\mathrm{cyl}}$ and $E^{\epsilon_i}_{\Sigma_i}$ re-expressed in "zoomed" coordinates. Let

$$R_{\epsilon_i \times \epsilon_i} = \epsilon_i R_{1\times 1} = \{(x_1, \alpha) \in \mathbf{P} : 0 \leq x_1 \leq \epsilon_i, 0 \leq x_1 + \alpha \leq \epsilon_i\}. \tag{5.15}$$

Making the change of coordinates $y \mapsto \epsilon_i y$, recalling that the spatial factor of Λ^2 is just Lebesgue measure on $\mathbf{P}$, and using the observation above, we have

$$\begin{aligned}
\Lambda^2(R_{\epsilon_i\times\epsilon_i} \times \mathbb{S}^2_+ \cap \{|E^{\epsilon_i}_{\Sigma_i}| \geq i^{-1}\}) &= \epsilon_i^2 \Lambda^2(R_{1\times 1} \times \mathbb{S}^2_+ \cap \{|E^1_{\Sigma_i}| \geq \epsilon_i^{-1} i^{-1}\}) \\
&\leq \epsilon_i^2 \Lambda^2(R_{1\times 1} \times \mathbb{S}^2_+ \cap \{|E^1_{\Sigma_i}| \geq C_i\}) \\
&\leq \epsilon_i^2 \Lambda^2(R_{1\times 1} \times \mathbb{S}^2_+ \cap \{\|K^{\Sigma_i,1}_{\mathrm{cyl}} - \mathrm{Id}_{\mathbf{P}\times\mathbb{S}^2_+}\| \geq C_i\}) \\
&\leq \epsilon_i^2 i^{-1},
\end{aligned} \tag{5.16}$$

where the second line follows from (5.6) and $\epsilon_i \leq b_i$, the third line follows because $\|K^{\Sigma_i,1}_{\mathrm{cyl}} - \mathrm{Id}_{\mathbf{P}\times\mathbb{S}^2_+}\| \geq |E^{\epsilon_i}_{\Sigma_i}|$, and the fourth line follows from (5.5). For $(j,k) \in \mathbb{Z}^2$, define translations,

$$\tau_{jk}(y) = y + j\epsilon_i e_1 + (1 + mJ^{-1})^{1/2} k \epsilon_i \chi, \qquad y \in \mathbf{P}. \tag{5.17}$$

The plane $\mathbf{P}$ is tessellated by the translates $\tau_{jk}R_{\epsilon_i\times\epsilon_i}$ in the sense that

$$\mathbf{P} = \bigcup_{(i,k)\in\mathbb{Z}^2} \tau_{jk}R_{\epsilon_i\times\epsilon_i}, \tag{5.18}$$

and for $(j,k) \neq (j',k')$, the set $\tau_{jk}R_{\epsilon_i\times\epsilon_i} \cap \tau_{j'k'}R_{\epsilon_i\times\epsilon_i}$ has measure zero. We also note that the cylindrical set $\mathcal{M}_{\mathrm{cyl}}(\epsilon_i)$ is invariant under the translates τ_{jk}. Consequently, $E^{\epsilon_i}_{\Sigma_i}$ is invariant under the τ_{jk} in the sense that $E^{\epsilon_i}_{\Sigma_i} \circ \tau_{jk} = E^{\epsilon_i}_{\Sigma_i}$. Therefore, for any $(j,k) \in \mathbb{Z}^2$,

$$\Lambda^2(\tau_{jk}R_{\epsilon_i\times\epsilon_i} \times \mathbb{S}^2_+ \cap \{|E^{\epsilon_i}_{\Sigma_i}| \geq i^{-1}\}) = \Lambda^2(R_{\epsilon_i\times\epsilon_i} \times \mathbb{S}^2_+ \cap \{|E^{\epsilon_i}_{\Sigma_i}| \geq i^{-1}\}) \leq \epsilon_i^2 i^{-1}. \tag{5.19}$$

Fix a bounded set $B \subset \mathbf{P} \times \mathbb{S}^2_+$. Note that there exists a constant C_B depending only on B such that B may be covered by $C_B\epsilon_i^{-2}$ sets of form $\tau_{jk}P_{\epsilon_i\times\epsilon_i} \times \mathbb{S}^2_+$. Consequently, by (5.19),

$$\Lambda^2(B \cap \{|E^{\epsilon_i}_{\Sigma_i}| \geq i^{-1}\}) \leq C_B i^{-1}. \tag{5.20}$$

This implies the claim. ■

Step 1d. We now conclude the proof that $\lim_{i\to\infty} \mathbb{K}^{\Sigma_i,\epsilon_i}_{\mathrm{cyl}}$ exists and is of form (5.1).

We write

$$
\begin{aligned}
&\int_{(\mathbf{P}\times\mathbb{S}^2_+)^2} g(y',w')\mathbb{K}^{\Sigma_i,\epsilon_i}_{\mathrm{cyl}}(y,w;\mathrm{d}y'\mathrm{d}w')f(y,w)\Lambda^2(\mathrm{d}y\mathrm{d}w)\\
&= \int_{\mathbf{P}\times\mathbb{S}^2_+} g(K^{\Sigma_i,\epsilon}_{\mathrm{cyl}}(y,w))f(y,w)\Lambda^2(\mathrm{d}y\mathrm{d}w)\\
&= \int_{\{E<i^{-1}\}} g(K^{\Sigma_i,\epsilon_i}_{\mathrm{cyl}}(y,w))f(y,w)\Lambda^2(\mathrm{d}y\mathrm{d}w)\\
&\quad + \int_{\{E\geq i^{-1}\}} g(K^{\Sigma_i,\epsilon_i}_{\mathrm{cyl}}(y,w))f(y,w)\Lambda^2(\mathrm{d}y\mathrm{d}w).
\end{aligned}
\tag{5.21}
$$

Noting that g is bounded and f has compact support, we see that by Claim 5.1.1, with $B = \operatorname{supp} f$, the second term above converges to zero as $i \to \infty$.

To handle the first term in (5.21), we introduce the following notation: For $h \in C_c(\mathbf{P}\times\mathbb{S}^2_+)$, let $\widehat{h}(y_1,\theta,y_3,\psi) = h(y_1,y_3,\theta,\psi)$. Using (5.9), (5.12), and (5.13), the first term in (5.21) is equal to

$$
\begin{aligned}
&\int_{\{E<i^{-1}\}} g(K^{\Sigma_i,\epsilon_i}_{\mathrm{cyl}}(y,w))f(y,w)\Lambda^2(\mathrm{d}y\mathrm{d}w)\\
&= \int_{\{E<i^{-1}\}} \widehat{g}(P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta),y_3+E,\pi-\psi)f(y_1,y_3,\theta,\psi)\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi\\
&= \int_{\{E<i^{-1}\}} \widehat{g}(P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta),y_3,\pi-\psi)f(y_1,y_3,\theta,\psi)\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi\\
&\quad + \int_{\{E<i^{-1}\}} (\tau_{(0,0,E,0)}\widehat{g}-\widehat{g})(P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta),y_3,\pi-\psi)f(y_1,y_3,\theta,\psi)\\
&\qquad\qquad\qquad\qquad \times\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi\\
&= \int_{\{E<i^{-1}\}} \widehat{g}(P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta),y_3,\pi-\psi)f(y_1,y_3,\theta,\psi)\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi\\
&\quad + \int_{P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(\{E<i^{-1}\})} (\tau_{(0,0,E,0)}\widehat{g}-\widehat{g})(y_1,\theta,y_3,\pi-\psi)\widehat{f}(P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta),y_3,\psi)\\
&\qquad\qquad\qquad\qquad \times\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi,
\end{aligned}
\tag{5.22}
$$

using in the last line the fact that $P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta)$ is an involution which preserves the measure $\sin\theta \mathrm{d}y_1\mathrm{d}\theta$ by Proposition 1.1 in §1.2. The second term above is bounded in absolute value by

$$\begin{aligned}
&\sup_{\|h\|\le i^{-1}} \|\tau_h\widehat{g}-\widehat{g}\|_{L^\infty}\int_{\mathbf{P}\times\mathbb{S}^2_+}|\widehat{f}(P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta),y_3,\psi)|\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi \\
&= \sup_{\|h\|\le i^{-1}} \|\tau_h\widehat{g}-\widehat{g}\|_{L^\infty}\|f\|_{L^1}\to 0 \quad \text{ as } i\to\infty,
\end{aligned} \tag{5.23}$$

again using invariance of $P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}$ with respect to $\sin\theta\mathrm{d}\theta\mathrm{d}y_1$ to obtain the equality. On the other hand, by Claim 5.1.1 and the fact that f and g are bounded and f has compact support, the first term in (5.22) is equal to

$$\begin{aligned}
&\int_{\mathbf{P}\times\mathbb{S}^2_+}\widehat{g}(P^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta),y_3,\pi-\psi)f(y_1,y_3,\theta,\psi)\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi+o_{i\to\infty}(1) \\
&=\int_{\mathbf{P}\times\mathbb{S}^2_+}\left(\int_{\mathbf{P}\times\mathbb{S}^2_+}g(y_1',y_3',\theta',\psi')\mathbb{P}^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta;\mathrm{d}y_1'\mathrm{d}\theta')\delta_{y_3}(\mathrm{d}y_3')\delta_{\pi-\psi}(\mathrm{d}\psi')\right) \\
&\qquad\qquad\times f(y_1,y_3,\theta,\psi)\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi+o_{i\to\infty}(1).
\end{aligned} \tag{5.24}$$

To summarize, we have shown that

$$\begin{aligned}
&\int_{\mathbf{P}\times\mathbb{S}^2_+}\left(\int_{\mathbf{P}\times\mathbb{S}^2_+}g(y',w')\mathbb{K}^{\Sigma_i,\epsilon_i}_{\mathrm{cyl}}(y,w;\mathrm{d}y'\mathrm{d}w')\right)f(y,w)\Lambda^2(\mathrm{d}y\mathrm{d}w) \\
&=\int_{\mathbf{P}\times\mathbb{S}^2_+}\left(\int_{\mathbf{P}\times\mathbb{S}^2_+}g(y_1',y_3',\theta',\psi')\mathbb{P}^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i}(y_1,\theta;\mathrm{d}y_1'\mathrm{d}\theta')\delta_{y_3}(\mathrm{d}y_3')\delta_{\pi-\psi}(\mathrm{d}\psi')\right) \\
&\qquad\qquad\times f(y_1,y_3,\theta,\psi)\sin\theta\sin^2\psi\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}\theta\mathrm{d}\psi+o_{i\to\infty}(1).
\end{aligned} \tag{5.25}$$

Since by assumption $\mathbb{K}=\lim_{i\to\infty}\mathbb{K}^{\Sigma_i,\epsilon_i^{(0)}}_{\mathrm{cyl}}$ exists, the equality above implies that

the following limit exists:

$$\mathbb{P} = \lim_{i\to\infty} \mathbb{P}^{W(\widetilde{\Sigma}_i, \widetilde{\epsilon}_i^{(0)})}, \tag{5.26}$$

and $\mathbb{K} = \mathbb{P} \times \delta_{y_3} \times \delta_{\pi-\psi}$. By Corollary 1.11 and Remark 1.13, $\mathbb{P} = \lim \mathbb{P}^{\gamma(\widetilde{\Sigma}_i, \widetilde{\epsilon}_i)}$ does not depend on the sequence $\epsilon_i \to 0$, and moreover $\mathbb{P}(y_1, \theta; \mathrm{d}y_1' \mathrm{d}\theta') = \delta_{y_1}(\mathrm{d}y_1')\widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta')$, for some Markov kernel $\widetilde{\mathbb{P}}$ on $\mathbb{S}^1_+$ which is symmetric with respect to the measure $\sin\theta \mathrm{d}\theta$. Thus by (5.25) we see that $\lim_{i\to\infty} \mathbb{K}^{\Sigma_i, \epsilon_i}$ exists for any $\epsilon_i \le b_i$, and this limit is equal to $\mathbb{K} = \delta_{y_1} \times \widetilde{\mathbb{P}} \times \delta_{y_3} \times \delta_{\pi-\psi}$. This proves the first part of the theorem.

Step 2. Now let us assume that $\Sigma_i = \Sigma$ is constant, let $\epsilon_i \to 0$ be arbitrary (not necessarily bounded by b_i) and consider how the proof above goes through in this case. In steps 1a-1d, to obtain the equality (5.25) we did not use the assumption that $\mathbb{K}^{\Sigma_i, \epsilon_i}_{\mathrm{cyl}}$ converges. Thus (5.25) always holds. On the other hand, $\widetilde{\Sigma}_i = \widetilde{\Sigma}$ is constant, it follows from Lemma 2.1 that the limit $\mathbb{P} = \lim_{i\to\infty} \mathbb{P}^{\widetilde{\Sigma}, \widetilde{\epsilon}_i}$ exists. This implies that the limit $\lim_{i\to\infty} \mathbb{K}^{\Sigma, \epsilon_i}_{\mathrm{cyl}}$ exists. This proves the second statement in the theorem.

Step 3. Finally, the converse may be proved as follows. Suppose $\mathbb{K}$ takes the form (5.1) where $\widetilde{\mathbb{P}}$ is a Markov kernel on $\mathbb{S}^1_+$ such that $\sin\theta \mathrm{d}\theta$ is invariant with respect to $\widetilde{\mathbb{P}}$. By Corollary 1.11 in §1.2 there is a sequence $\widetilde{\Sigma}_i$ of cells such that for any sequence $\widetilde{\epsilon}_i \to 0$, $\delta_{y_1} \times \widetilde{\mathbb{P}} = \lim_{i\to\infty} \mathbb{P}^{\widetilde{\Sigma}_i, \widetilde{\epsilon}_i}$. Let

$$\Sigma_i = \{(x_1, x_2) : (x_1, (1 + mJ^{-1})^{1/2} x_2) \in \widetilde{\Sigma}_i\}, \tag{5.27}$$

and let b_i be chosen with respect to the Σ_i above as in the beginning of the proof. Take a sequence of positive numbers $\epsilon_i \to 0$ with $\epsilon_i \le b_i$, and let $\widetilde{\epsilon}_i = (1 + mJ^{-1})^{1/2}\epsilon_i$. Let $W_i = W(\Sigma_i, \epsilon_i)$ and $\widetilde{W}_i = W(\widetilde{\Sigma}_i, \widetilde{\epsilon}_i)$, and observe that $\widetilde{W}_i$ is just foreshortened

version of W_i in the sense of (5.8). Consider the cylindrical configuration space $\mathcal{M}_{\text{cyl}}(\Sigma_i, \epsilon_i)$ and the cylindrical collision law $K_{\text{cyl}}^{\Sigma_i,\epsilon_i}$ which the wall W_i gives rise to. By Steps 1a-1d, the equality (5.25) holds. Since $\mathbb{P}^{\widetilde{\Sigma}_i,\widetilde{\epsilon}_i} \to \mathbb{P} = \delta_{y_1} \times \widetilde{\mathbb{P}}$, it follows that $\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i} \to \delta_{y_1} \times \widetilde{\mathbb{P}} \times \delta_{y_3} \times \delta_{\pi-\psi} = \mathbb{K}$. □

5.2 Cylindrical approximation in the pure scaling case

In this section we prove Theorem 1.27, delegating the proof of a key lemma to a subsequent subsection.

Since the sequence of cells is constant, we take Σ to be fixed throughout this section, and we denote the collision law in $\mathcal{M} = \mathcal{M}(\Sigma, \epsilon)$ by K^ϵ, dropping the explicit dependence on Σ from our notation. Similarly, we denote the collision law in the cylindrical configuration space $\mathcal{M}_{\text{cyl}} = \mathcal{M}_{\text{cyl}}(\Sigma, \epsilon)$ by K_{cyl}^ϵ, and we denote the modified collision law by $\widetilde{K}^\epsilon$.

The following lemma gives us the sense in which the collision laws K^ϵ and K_{cyl}^ϵ approximate each other.

Lemma 5.2. *For any $f, g \in C_c^\infty(\mathbf{P} \times \mathbb{S}_+^2)$,*

$$\int_{\mathbf{P}\times\mathbb{S}_+^2} g[f \circ K^\epsilon - f \circ K_{\text{cyl}}^\epsilon]\mathrm{d}\Lambda^2 \to 0 \quad \textit{as } \epsilon \to 0. \tag{5.28}$$

The proof of the lemma involves first throwing away a small set of "bad" inputs, and then making two comparisons: (i) a comparison between the true collision law K^ϵ and the modified collision law $\widetilde{K}^\epsilon$, and (ii) a comparison between the modified collision law $\widetilde{K}^\epsilon$ and the cylindrical collision law K_{cyl}^ϵ.

Recall the definition (4.86) of $\widetilde{K}^\epsilon$. The main idea for making the comparison (i) is to use estimates for η and its differential to argue that $\widetilde{K}^\epsilon = \eta^{-1} \circ K^\epsilon \circ \eta$ approximates K^ϵ as $\epsilon \to 0$ in the sense of (5.28).

Most of the work in this section is concerned with making the comparison (ii).

Consider the parallelogram

$$R_\epsilon = \{(x_1, \alpha) \in \mathbf{P} : 0 \le x_1 + \alpha \le \epsilon, -\rho(\epsilon)/2 \le \alpha \le \rho(\epsilon)/2\}. \tag{5.29}$$

For $(j,k) \in \mathbb{Z}^2$, define translations

$$\tau_{jk}(y) = y + j\epsilon e_1 + k\rho e_3, \qquad y \in \mathbf{P}. \tag{5.30}$$

We also let

$$\overline{\tau}_{jk}(y,w) = (\tau_{jk}(y), w), \qquad (y,w) \in \mathbf{P} \times \mathbb{S}^2_+. \tag{5.31}$$

The plane $\mathbf{P}$ may be tessellated by the parallelograms $\tau_{jk}R_\epsilon$ in the sense that $\mathbf{P} = \bigcup_{(j,k)\in\mathbb{Z}^2} \tau_{jk}R_\epsilon$ and $\tau_{jk}R_\epsilon \cap \tau_{j'k'}R_\epsilon$ has Lebesgue measure zero whenever $(j,k) \ne (j',k')$.

Recall that $\widetilde{K}^\epsilon$ is defined on an open subset $\widetilde{\mathcal{F}} \subset \mathbf{P} \times \mathbb{S}^2_+$.

Lemma 5.3. *For $\epsilon > 0$ sufficiently small, there exist sets $\Omega(\epsilon) \subset \widetilde{\mathcal{F}}$ such that*

i. $\Omega(\epsilon)$ is invariant under the translations $\overline{\tau}_{jk}$;

ii. $\frac{1}{\epsilon\rho(\epsilon)}\Lambda^2(R_\epsilon \times \mathbb{S}^2_+ \setminus \Omega(\epsilon)) \to 0$ as $\epsilon \to 0$; and

iii. the following limit holds:

$$\lim_{\epsilon\to 0} \sup_{(y,w)\in R_\epsilon\cap\Omega(\epsilon)} \|\widetilde{K}^\epsilon(y,w) - K^\epsilon_{\mathrm{cyl}}(y,w)\| = 0. \tag{5.32}$$

Remark 5.4. Lemma 5.3(i) and (ii), together with Claim 5.2.1 from Step 1 of its proof, below, imply that for any bounded set $B \subset \mathbf{P}\times\mathbb{S}^2_+$, $\Lambda^2(B\setminus\Omega(\epsilon)) \to 0$ as $\epsilon\to 0$. Consequently, $\Lambda^2(B\setminus\widetilde{\mathcal{F}}) \to 0$ as $\epsilon \to 0$.

The proof of this lemma is given in §5.3. We will now prove Lemma 5.2 using the lemma above.

Proof of Lemma 5.2. Let $\Omega(\epsilon) \subset \widetilde{\mathcal{F}}$ be as in Lemma 5.3. We split up the integral $\int g[f\circ K^\epsilon - f\circ K^\epsilon_{\mathrm{cyl}}]\mathrm{d}\Lambda^2$ as follows:

$$\begin{aligned}
&\int_{\mathbf{P}\times\mathbb{S}^2_+} g[f\circ K^\epsilon - f\circ K^\epsilon_{\mathrm{cyl}}]\mathrm{d}\Lambda^2 \\
&= \underbrace{\int_{\mathbf{P}\times\mathbb{S}^2_+\setminus\Omega} g[f\circ K^\epsilon - f\circ K^\epsilon_{\mathrm{cyl}}]\mathrm{d}\Lambda^2}_{=:\, I_1} + \underbrace{\int_\Omega g[f\circ K^\epsilon - f\circ \widetilde{K}^\epsilon]\mathrm{d}\Lambda^2}_{=:\, I_2} \\
&\qquad + \underbrace{\int_\Omega g[f\circ \widetilde{K}^\epsilon - f\circ K^\epsilon_{\mathrm{cyl}}]\mathrm{d}\Lambda^2}_{=:\, I_3}.
\end{aligned} \tag{5.33}$$

Note that the modified collision law $\widetilde{K}^\epsilon$ is defined on Ω because $\Omega\subset\widetilde{\mathcal{F}}$. We will show separately that each of I_1, I_2, and I_3 converge to zero as $\epsilon\to 0$.

Step 1. To show $I_1 \to 0$, first we write

$$\begin{aligned} |I_1| &\le \int_{\mathbf{P}\times\mathbb{S}^2_+\setminus\Omega} |g|\cdot|f\circ K^\epsilon|\mathrm{d}\Lambda + \int_{\mathbf{P}\times\mathbb{S}^2_+\setminus\Omega} |g|\cdot|f\circ K^\epsilon_{\mathrm{cyl}}|\mathrm{d}\Lambda \\ &\le 2\|f\|_{L^\infty}\Lambda^2(\operatorname{supp} g\setminus\Omega). \end{aligned} \tag{5.34}$$

Claim 5.2.1. *Let B be a compact set. There exists a constant $C = C_B < \infty$ depending only on B such that, for ϵ sufficiently small, the set B can be covered by $\frac{C}{\epsilon\rho(\epsilon)}$ sets of form $\tau_{jk}R_\epsilon \times \mathbb{S}^2_+$.*

Proof of Claim 5.2.1. Since B is compact, there exist minimal $a, b < \infty$ such that

$$B \subset \{(x_1,\alpha)\in\mathbf{P} : -a \le x_1+\alpha \le a, -b\le\alpha\le b\}\times\mathbb{S}^2_+. \tag{5.35}$$

Note that parallelogram R_ϵ has length ϵ and height ρ, so the parallelogram appearing on the right-hand side above can be covered by $\frac{2a+1}{\epsilon}\cdot\frac{2b+1}{\rho(\epsilon)}$ translates $\tau_{jk}R_\epsilon$. The claim follows by taking $C = (2a+1)(2b+1)$. ■

Since Ω is invariant with respect to the translations τ_{jk} and $\operatorname{supp} g$ is compact, it follows that

$$\Lambda^2(\operatorname{supp} g\setminus\Omega) \le \frac{C}{\epsilon\rho(\epsilon)}\Lambda^2(R_\epsilon\setminus\Omega), \tag{5.36}$$

where $C = C_{\operatorname{supp} g}$. The right-hand side converges to zero by Lemma 5.3.

Step 2. Next we show $I_2 \to 0$. Recall the definition (4.84) of the diffeomorphism $\eta : \mathcal{G} \to \eta(\mathcal{G}) \subset \mathbf{P}\times\mathbb{S}^2_{+\rho/2}$. We may write $\eta(y,w) = (\varphi_w(y), w)$, where for each

$w = (v_1, v_2, \omega) \in \mathbb{S}^2_{+\rho/2}$, $\varphi_w : \mathbf{P} \cap \widehat{\mathcal{Z}} \to \mathbf{P}$ is given by the formula

$$\varphi_w(x_1, \alpha) = \begin{pmatrix} x_1 - \overline{\alpha} + \sin\overline{\alpha} - (1 - \cos\overline{\alpha})\frac{v_1}{v_2} \\ \alpha - (1 - \cos\overline{\alpha})\frac{\omega}{v_2} \end{pmatrix}, \tag{5.37}$$

where $\overline{\alpha} = \alpha - \overline{k}\rho$ and $\overline{k} = \operatorname{argmin}\{|\alpha - k\rho| : k \in \mathbb{Z}\}$. The differential of φ_w is thus given by the formula:

$$\mathrm{d}(\varphi_w)\,(x_1, \alpha) = \begin{bmatrix} 1 & -1 + \cos\overline{\alpha} - \sin(\overline{\alpha})\frac{v_1}{v_2} \\ 0 & 1 - \sin(\overline{\alpha})\frac{\omega}{v_2} \end{bmatrix}. \tag{5.38}$$

For $(y, w) \in \mathcal{G}$, we define

$$U(y, w) = |\det \mathrm{d}(\varphi_w)\,(x_1, \alpha)| = \left|1 - \sin(\overline{\alpha})\frac{\omega}{v_2}\right|. \tag{5.39}$$

Then for $(y, w) \in \eta(\mathcal{G})$,

$$|\det \mathrm{d}\big(\varphi_w^{-1}\big)\,(x_1, \alpha)| = \frac{1}{U \circ \eta^{-1}(y, w)}. \tag{5.40}$$

Claim 5.2.2. *If B is a compact subset of $\mathbf{P} \times \mathbb{S}^2_+$, then*

$$\sup_{(y,w)\in B\cap\mathcal{G}} |U(y, w) - 1| \to 0 \qquad \textit{as } \epsilon \to 0. \tag{5.41}$$

Proof of Claim 5.2.2. By compactness, for all $(y, w) \in B$, $v_2 = \langle w, e_2\rangle > c > 0$ for some

fixed constant c. Also note that $|\overline{\alpha}| \leq \rho(\epsilon)/2 \to 0$ as $\epsilon \to 0$. Therefore,

$$|U(y,w) - 1| \leq \sin(\rho/2) \cdot c^{-1} \to 0 \quad \text{as } \epsilon \to 0, \tag{5.42}$$

and the claim is proved. ■

Recall also the definition (4.86) of the modified collision law $\widetilde{K}^\epsilon$. Making the change of variables $(y,w) \mapsto \eta^{-1}(y,w) = (\varphi_w^{-1}(y), w)$ and noting that the first factor of Λ^2 is just Lebesgue measure on $\mathbf{P}$, we have

$$\int_\Omega g[f \circ \widetilde{K}^\epsilon] \mathrm{d}\Lambda^2 = \int_{\eta(\Omega)} [g \circ \eta^{-1}][f \circ \eta^{-1} \circ K^\epsilon] \frac{1}{U \circ \eta^{-1}} \mathrm{d}\Lambda^2. \tag{5.43}$$

Using this, we have

$$\begin{aligned}
&\left| \int_\Omega g[f \circ K^\epsilon - f \circ \widetilde{K}^\epsilon] \mathrm{d}\Lambda^2 \right| \\
&= \left| \int_\Omega g[f \circ K^\epsilon] \mathrm{d}\Lambda^2 - \int_{\eta(\Omega)} [g \circ \eta^{-1}][f \circ \eta^{-1} \circ K^\epsilon] \frac{1}{U \circ \eta^{-1}} \mathrm{d}\Lambda^2 \right| \\
&\leq \left| \int_{\Omega \cap \eta(\Omega)} g[f \circ K^\epsilon - f \circ \eta^{-1} \circ K^\epsilon] \mathrm{d}\Lambda^2 \right| \\
&\quad + \left| \int_{\Omega \cap \eta(\Omega)} [g - g \circ \eta^{-1}][f \circ \eta^{-1} \circ K^\epsilon] \mathrm{d}\Lambda^2 \right| \\
&\quad + \left| \int_{\Omega \cap \eta(\Omega)} [g \circ \eta^{-1}][f \circ \eta^{-1} \circ K^\epsilon] \left[1 - \frac{1}{U \circ \eta^{-1}} \right] \mathrm{d}\Lambda^2 \right| \\
&\quad + \left| \int_{\Omega \setminus \eta(\Omega)} g[f \circ K^\epsilon] \mathrm{d}\Lambda^2 \right| + \left| \int_{\eta(\Omega) \setminus \Omega} [g \circ \eta^{-1}][f \circ K^\epsilon] \frac{1}{U \circ \eta^{-1}} \mathrm{d}\Lambda^2 \right|.
\end{aligned} \tag{5.44}$$

We will estimate separately each term appearing above. First, recall (4.89) and (4.90) which say that η and η^{-1} converges uniformly to the identity on their respective domains $\mathcal{G}$ and $\eta(\mathcal{G})$ as $\epsilon \to 0$. This has two important consequences: First, since f

and g are continuous and compactly supported,

$$\|f - f \circ \eta\|_{L^\infty(\mathcal{G})} \to 0 \quad \text{and} \quad \|g - g \circ \eta\|_{L^\infty(\mathcal{G})} \to 0 \quad \text{as} \quad \epsilon \to 0, \tag{5.45}$$

and

$$\|f - f \circ \eta^{-1}\|_{L^\infty(\eta(\mathcal{G}))} \to 0 \quad \text{and} \quad \|g - g \circ \eta^{-1}\|_{L^\infty(\eta(\mathcal{G}))} \to 0 \quad \text{as} \quad \epsilon \to 0. \tag{5.46}$$

Second, there exists a fixed compact set $B \subset \mathbf{P} \times \mathbb{S}^2_+$ such that, for ϵ sufficiently small,

$$\begin{aligned} &\mathcal{G} \cap \left\{ \operatorname{supp} f \cup \operatorname{supp} g \cup \operatorname{supp}(f \circ \eta) \cup \operatorname{supp}(g \circ \eta) \right\} \subset B, \quad \text{and} \\ &\eta(\mathcal{G}) \cap \left\{ \operatorname{supp} f \cup \operatorname{supp} g \cup \operatorname{supp}(f \circ \eta^{-1}) \cup \operatorname{supp}(g \circ \eta^{-1}) \right\} \subset B. \end{aligned} \tag{5.47}$$

We recall that by Lemma 5.3, $\Omega \subset \widetilde{\mathcal{F}} \subset \mathcal{G}$, and hence

$$\widetilde{K}^\epsilon(\Omega) \subset \widetilde{K}^\epsilon(\widetilde{\mathcal{F}}) = \widetilde{\mathcal{F}} \subset \mathcal{G} \quad \Rightarrow \quad K^\epsilon(\eta(\Omega)) \subset \eta(\mathcal{G}). \tag{5.48}$$

By invariance of Λ^2 with respect to K^ϵ and (5.46),

$$\begin{aligned} &\left| \int_{\Omega \cap \eta(\Omega)} g[f \circ K^\epsilon - f \circ \eta^{-1} \circ K^\epsilon] \mathrm{d}\Lambda^2 \right| \\ &\quad = \int_{K^\epsilon(\Omega \cap \eta(\Omega))} |g \circ K^\epsilon| \cdot |f - f \circ \eta^{-1}| \mathrm{d}\Lambda^2 \\ &\quad \le \|g \circ K^\epsilon\|_{L^1_{\Lambda^2}(\mathbf{P} \times \mathbb{S}^2_+)} \|f - f \circ \eta^{-1}\|_{L^\infty(\mathcal{G})} \\ &\quad = \|g\|_{L^1_{\Lambda^2}(\mathbf{P} \times \mathbb{S}^2_+)} \|f - f \circ \eta^{-1}\|_{L^\infty(\mathcal{G})} \to 0 \quad \text{as } \epsilon \to 0. \end{aligned} \tag{5.49}$$

Also, by (5.46) and (5.47), we have

$$\begin{aligned}&\left|\int_{\Omega\cap\eta(\Omega)}[g-g\circ\eta^{-1}][f\circ\eta^{-1}\circ K^{\epsilon}]\mathrm{d}\Lambda^2\right|\\&\leq\|f\|_{L^\infty}\left|\int_{B\cap\Omega\cap\eta(\Omega)}[g-g\circ\eta^{-1}]\mathrm{d}\Lambda^2\right|\to 0\quad\text{ as }\epsilon\to 0.\end{aligned}\tag{5.50}$$

In addition, note that $\operatorname{supp} g\subset\eta(B)$, and by Claim 5.2.2, $\frac{1}{U}\to 1$ uniformly on $B\cap\mathcal{G}$; thus

$$\begin{aligned}&\left|\int_{\Omega\cap\eta(\Omega)}[g\circ\eta^{-1}][f\circ\eta^{-1}\circ K^{\epsilon}]\left[1-\frac{1}{U\circ\eta^{-1}}\right]\mathrm{d}\Lambda^2\right|\\&\leq\|f\|_{L^\infty}\|g\|_{L^\infty}\int_{\Omega\cap\eta(B\cap\Omega)}\left|1-\frac{1}{U\circ\eta^{-1}}\right|\mathrm{d}\Lambda^2\to 0\quad\text{ as }\epsilon\to 0.\end{aligned}\tag{5.51}$$

By Claim 5.2.2, there exists a constant $d_1<\infty$ such that $U\leq d_1$ on $B\cap\mathcal{G}$, for ϵ sufficiently small. Thus

$$\begin{aligned}\left|\int_{\Omega\setminus\eta(\Omega)}g[f\circ K^{\epsilon}]\mathrm{d}\Lambda^2\right|&=\left|\int_{\eta^{-1}(\Omega)\setminus\Omega}(g\circ\eta)[f\circ K^{\epsilon}\circ\eta]U\mathrm{d}\Lambda^2\right|\\&=d_1\|f\|_{L^\infty}\int_{\eta^{-1}(\Omega)\setminus\Omega}|g\circ\eta|\mathrm{d}\Lambda^2\\&\leq d_1\|f\|_{L^\infty}\|g\|_{L^\infty}\Lambda^2(B\setminus\Omega)\\&\leq d_1\|f\|_{L^\infty}\|g\|_{L^\infty}\frac{C}{\epsilon\rho(\epsilon)}\Lambda^2(R_\epsilon\setminus\Omega)\to 0\text{ as }\epsilon\to 0,\end{aligned}\tag{5.52}$$

where $C=C_B$ as in Claim 5.2.1 from Step 1, and the convergence to zero follows by Lemma 5.3.

Also by Claim 5.2.2 there exists $d_2<\infty$ such that $\frac{1}{U}\leq d_2$ on $B\cap\mathcal{G}$. Noting that

$\operatorname{supp}(g \circ \eta^{-1}) \subset B \cap \eta(B)$ by (5.47), we similarly have

$$\begin{aligned}
&\left| \int_{\eta(\Omega) \setminus \Omega} [g \circ \eta^{-1}][f \circ K^\epsilon] \frac{1}{U \circ \eta^{-1}} \mathrm{d}\Lambda^2 \right| \\
&= \left| \int_{B \cap \eta(\Omega \cap B) \setminus \Omega} [g \circ \eta^{-1}][f \circ K^\epsilon] \frac{1}{U \circ \eta^{-1}} \mathrm{d}\Lambda^2 \right| \\
&\leq d_2 \|f\|_{L^\infty} \|g\|_{L^\infty} \Lambda^2(B \setminus \Omega) \\
&\leq d_2 \|f\|_{L^\infty} \|g\|_{L^\infty} \frac{C}{\epsilon \rho(\epsilon)} \Lambda^2(R_\epsilon \setminus \Omega) \to 0 \text{ as } \epsilon \to 0.
\end{aligned} \tag{5.53}$$

From (5.49), (5.50), (5.51), (5.52), and (5.53), we see that each term in the bound (5.44) converges to zero. Thus $I_2 \to 0$.

Step 3. Finally, we show that $I_3 \to 0$. We must deal with the fact that $\widetilde{K}^\epsilon$ and $K^\epsilon_{\mathrm{cyl}}$ have different periodicity (see Propositions 4.12 and 4.14). Let τ_{jk} be defined as above, and for $(j,k) \in \mathbb{Z}^2$, let

$$\tau'_{jk}(y) = y + \left(j + \left\lceil \frac{j(1+mJ^{-1})^{1/2}\rho}{\epsilon} \right\rceil \right) \epsilon e_1 + k(1+mJ^{-1})^{1/2} \rho \chi, \qquad y \in \mathbf{P}. \tag{5.54}$$

In the lattice in $\mathbf{P}$ spanned by integer combinations of e_1 and $(1+mJ^{-1})^{1/2}\rho\chi$, the point $\tau'_{jk}(0)$ is the closest point in the lattice to the right of the point $\tau_{jk}(0) = j\epsilon e_1 + k\rho e_3$. We also let

$$\overline{\tau}'_{jk}(y,w) = (\tau'_{jk}(y), w), \qquad (y,w) \in \mathbf{P} \times \mathbb{S}^2_+. \tag{5.55}$$

Then

$$K^\epsilon \circ \overline{\tau}_{jk} = \overline{\tau}_{jk} \circ K^\epsilon, \quad \text{and} \quad K^\epsilon_{\mathrm{cyl}} \circ \overline{\tau}'_{jk} = \overline{\tau}'_{jk} \circ K^\epsilon_{\mathrm{cyl}}. \tag{5.56}$$

Also, for $k \in \mathbb{Z}$, define

$$\mathbf{P}_k = \{(x_1, \alpha) \in \mathbf{P} : k\rho - \rho/2 \le \alpha < k\rho + \rho/2\}, \tag{5.57}$$

and define $\theta : \mathbf{P} \times \mathbb{S}^2_+ \to \mathbf{P} \times \mathbb{S}^2_+$ by

$$\theta(y, w) = \left(y - \epsilon \left\{ \frac{k(1 + mJ^{-1})^{1/2}\rho}{\epsilon} \right\} e_1, w \right), \text{ if } (y, w) \in \mathbf{P}_k \times \mathbb{S}^2_+. \tag{5.58}$$

Here $\left\{ \frac{k(1+mJ^{-1})^{1/2}\rho}{\epsilon} \right\}$ is the fractional part of $\frac{k(1+mJ^{-1})^{1/2}\rho}{\epsilon}$. We see that θ is an order ϵ shift which "corrects" for the difference of periods, in the sense that

$$\theta \circ \overline{\tau}'_{jk} = \overline{\tau}_{jk} \quad \text{ on } \mathbf{P}_k \times \mathbb{S}^2_+. \tag{5.59}$$

We may write

$$\begin{aligned} \int_\Omega g[f \circ \widetilde{K}^\epsilon - f \circ K^\epsilon_{\text{cyl}}] = & \int_\Omega g[f \circ \widetilde{K}^\epsilon - f \circ \theta \circ K^\epsilon_{\text{cyl}} \circ \theta^{-1}] \mathrm{d}\Lambda^2 \\ & + \int_\Omega g[f \circ \theta \circ K^\epsilon_{\text{cyl}} \circ \theta^{-1} - f \circ K^\epsilon_{\text{cyl}} \circ \theta^{-1}] \mathrm{d}\Lambda^2 \\ & + \int_\Omega g[f \circ K^\epsilon_{\text{cyl}} \circ \theta^{-1} - f \circ K^\epsilon_{\text{cyl}}] \mathrm{d}\Lambda^2. \end{aligned} \tag{5.60}$$

We show separately that each of the above terms converges to zero. If u is any Lipschitz function on $\mathbf{P} \times \mathbb{S}^2_+$, we let

$$\text{Lip}(u) = \sup_{(y,w) \neq (y',w')} \frac{|u(y, w) - u(y', w')|}{\|(y, w) - (y', w')\|}. \tag{5.61}$$

By Claim 5.2.1 from Step 1, there is an index set $I \subset \mathbb{Z}^2$ with $|I| \le C/\epsilon\rho$ such that

$\operatorname{supp} g$ is covered by the parallelograms $\tau_{jk} R_\epsilon$, $(j,k) \in I$. Using the fact that g is bounded and f is Lipschitz (since it is in C_c^∞), we have

$$
\begin{aligned}
&\left| \int_\Omega g[f \circ \widetilde{K}^\epsilon - f \circ \theta \circ K^\epsilon_{\mathrm{cyl}} \circ \theta^{-1}] \mathrm{d}\Lambda^2 \right| \\
&\leq \|g\|_{L^\infty} \operatorname{Lip}(f) \int_{\Omega \cap \operatorname{supp} g} \|\widetilde{K}^\epsilon - \theta \circ K^\epsilon_{\mathrm{cyl}} \circ \theta^{-1}\| \mathrm{d}\Lambda^2 \\
&\leq \|g\|_{L^\infty} \operatorname{Lip}(f) \sum_{(j,k) \in I} \int_{\overline{\tau}_{jk}(R_\epsilon \times \mathbb{S}^2_+) \cap \Omega} \|\widetilde{K}^\epsilon - \theta \circ K^\epsilon_{\mathrm{cyl}} \circ \theta^{-1}\| \mathrm{d}\Lambda^2 \\
&= \|g\|_{L^\infty} \operatorname{Lip}(f) \sum_{(j,k) \in I} \int_{(R_\epsilon \times \mathbb{S}^2_+) \cap \Omega} \|\widetilde{K}^\epsilon \circ \overline{\tau}_{jk} - \theta \circ K^\epsilon_{\mathrm{cyl}} \circ \overline{\tau}'_{jk}\| \mathrm{d}\Lambda^2 \\
&= \|g\|_{L^\infty} \operatorname{Lip}(f) \sum_{(j,k) \in I} \int_{(R_\epsilon \times \mathbb{S}^2_+) \cap \Omega} \|\overline{\tau}_{jk} \circ \widetilde{K}^\epsilon - \overline{\tau}_{jk} \circ K^\epsilon_{\mathrm{cyl}}\| \mathrm{d}\Lambda^2 \\
&= \|g\|_{L^\infty} \operatorname{Lip}(f) |I| \int_{(R_\epsilon \times \mathbb{S}^2_+) \cap \Omega} \|\widetilde{K}^\epsilon - K^\epsilon_{\mathrm{cyl}}\| \mathrm{d}\Lambda^2 \\
&\leq \|g\|_{L^\infty} \operatorname{Lip}(f) |I| \Lambda^2(R_\epsilon \times \mathbb{S}^2_+) \sup_{(R_\epsilon \times \mathbb{S}^2_+) \cap \Omega} \|\widetilde{K}^\epsilon - K^\epsilon\| \\
&= 2\pi C \|g\|_{L^\infty} \operatorname{Lip}(f) \sup_{(R_\epsilon \times \mathbb{S}^2_+) \cap \Omega} \|\widetilde{K}^\epsilon - K^\epsilon\|,
\end{aligned}
\tag{5.62}
$$

using in the last line the fact that $|I| = C/\epsilon\rho$ and $\Lambda^2(R_\epsilon \times \mathbb{S}^2_+) = 2\pi\rho\epsilon$. The last quantity converges to zero by Lemma 5.3. To show that the other two terms in (5.60) converge to zero, observe that

$$
\begin{aligned}
&\left| \int_\Omega g[f \circ \theta \circ K^\epsilon_{\mathrm{cyl}} \circ \theta^{-1} - f \circ K^\epsilon_{\mathrm{cyl}} \circ \theta^{-1}] \mathrm{d}\Lambda^2 \right| \\
&\leq \int_{\mathbf{P} \times \mathbb{S}^2_+} |g| |f \circ \theta \circ K^\epsilon_{\mathrm{cyl}} \circ \theta^{-1} - f \circ K^\epsilon_{\mathrm{cyl}} \circ \theta^{-1}| \mathrm{d}\Lambda^2 \\
&= \int_{\mathbf{P} \times \mathbb{S}^2_+} |g| |f \circ \theta - f| \mathrm{d}\Lambda^2,
\end{aligned}
\tag{5.63}
$$

where the last equality follows by invariance of Λ^2 under translation and $K^\epsilon_{\mathrm{cyl}}$. The

quantity above converges to zero because $\theta \to 0$ uniformly as $\epsilon \to 0$. To handle the last term in (5.60), we make the change of variables $(y, w) \mapsto \theta(y, w)$ to obtain

$$\begin{aligned}
&\left| \int_\Omega g[f \circ K^\epsilon_{\text{cyl}} \circ \theta^{-1} - f \circ K^\epsilon_{\text{cyl}}] \mathrm{d}\Lambda^2 \right| \\
&= \left| \int_{\theta^{-1}(\Omega)} [g \circ \theta][f \circ K^\epsilon_{\text{cyl}}] \mathrm{d}\Lambda^2 - \int_\Omega g[f \circ K^\epsilon_{\text{cyl}}] \mathrm{d}\Lambda^2 \right| \\
&\leq \int_{\mathbf{P} \times \mathbb{S}^2_+} |g \circ \theta - g| |f \circ K^\epsilon_{\text{cyl}}| \mathrm{d}\Lambda^2 \\
&\qquad + \|g\|_{L^\infty} \|f\|_{L^\infty} \left\{ \Lambda^2(\theta^{-1}(\operatorname{supp} g) \setminus \theta^{-1}(\Omega)) + \Lambda^2(\operatorname{supp} g \setminus \Omega) \right\} \\
&\leq \|g \circ \theta - g\|_{L^\infty} \|f\|_{L^1_{\Lambda^2}} + 2\|g\|_{L^\infty} \|f\|_{L^\infty} \Lambda^2(\operatorname{supp} g \setminus \Omega),
\end{aligned} \tag{5.64}$$

using invariance of Λ^2 with respect to K^ϵ_{cyl} and θ to obtain the last line. The first term in the line above converges to zero because $\theta \to 0$ uniformly, and the second term converges to zero by the same argument as in Step 1. Thus all three terms on the right-hand side of (5.60) converge to zero, so $I_3 \to 0$. □

We now apply Lemma 5.2 and Theorem 5.1 to prove our main results.

Proof of Theorem 1.27. Fix a constant sequence of cells $\Sigma_i = \Sigma$ and a sequence of positive numbers $\epsilon_i \to 0$. By Theorem 5.1, there exists a Markov kernel $\mathbb{K}$ not depending on the sequence ϵ_i such that the limit $\lim_{i \to \infty} \mathbb{K}^{\Sigma_i, \epsilon_i}_{\text{cyl}}$ exists and is equal to $\mathbb{K}$. Moreover, $\mathbb{K}$ takes the form 1.65. We may write

$$\begin{aligned}
&\int_{\mathbf{P} \times \mathbb{S}^2_+} g(y, w) \left(\int_{\mathbf{P} \times \mathbb{S}^2_+} f(y, w) \mathbb{K}^{\Sigma_i, \epsilon_i}(y, w; \mathrm{d}y' \mathrm{d}w') \right) \mathrm{d}\Lambda^2(\mathrm{d}y \mathrm{d}w) \\
&= \int_{\mathbf{P} \times \mathbb{S}^2_+} g(y, w) \left(\int_{\mathbf{P} \times \mathbb{S}^2_+} f(y, w) \mathbb{K}^{\Sigma_i, \epsilon_i}_{\text{cyl}}(y, w; \mathrm{d}y' \mathrm{d}w') \right) \mathrm{d}\Lambda^2(\mathrm{d}y \mathrm{d}w) \\
&\qquad + \int_{\mathbf{P} \times \mathbb{S}^2_+} g(y, w) [f(K^{\Sigma_i, \epsilon_i}(y, w)) - f(K^{\Sigma_i, \epsilon_i}_{\text{cyl}}(y, w))] \mathrm{d}\Lambda^2(\mathrm{d}y \mathrm{d}w).
\end{aligned} \tag{5.65}$$

The second term above converges to zero as $\epsilon_i \to 0$ by Lemma 5.2. It follows that $\mathbb{K}^{\Sigma_i,\epsilon_i} \to \mathbb{K}$, and the theorem is proved. $\square$

Proof of Theorem 1.28. Fix a sequence of cells $\{\Sigma_i\}_{i\geq 1}$. Lemma 5.2 implies that, for each fixed $i \geq 1$,

$$d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon}, \mathbb{K}^{\Sigma_i,\epsilon}) \to 0 \quad \text{as } \epsilon \to 0, \tag{5.66}$$

where $d_{\mathcal{G}}^{\Lambda^2}$ is the pseudometric introduced in 6.2.3. Thus we may choose a decreasing sequence $\{b_i^{(1)}\}_{i\geq 1}$ such that, for each $i \geq 1$, if $\epsilon \leq b_i^{(1)}$, then

$$d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon}, \mathbb{K}^{\Sigma_i,\epsilon}) \leq i^{-1}. \tag{5.67}$$

Let another decreasing sequence $\{b_i^{(2)}\}_{i\geq 1}$ be chosen as in Theorem 5.1, and let $b_i = \min\{b_i^{(1)}, b_i^{(2)}\}$.

To prove Theorem 1.28, we will show that $\neg(B) \Rightarrow (A)$. Assume that there exists a sequence $\epsilon_i^{(0)} \to 0$, $\epsilon_i^{(0)} \leq b_i$ such that $\mathbb{K} := \lim_{i\to\infty} \mathbb{K}^{\Sigma_i,\epsilon_i^{(0)}}$ exists. Then by the triangle inequality,

$$\begin{aligned} d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i^{(0)}}, \mathbb{K}) &\leq d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i^{(0)}}, \mathbb{K}^{\Sigma_i,\epsilon_i^{(0)}}) + d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}^{\Sigma_i,\epsilon_i^{(0)}}, \mathbb{K}) \\ &\to 0 \quad \text{as } i \to \infty. \end{aligned} \tag{5.68}$$

By Theorem 5.1, it follows that, for any sequence $\epsilon_i \to 0$, $\epsilon_i \leq b_i$, $\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i} \to \mathbb{K}$ as $i \to \infty$; so by the triangle inequality again

$$\begin{aligned} d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}^{\Sigma_i,\epsilon_i}, \mathbb{K}) &\leq d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i}, \mathbb{K}^{\Sigma_i,\epsilon_i}) + d_{\mathcal{G}}^{\Lambda^2}(\mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i}, \mathbb{K}) \\ &\to 0 \quad \text{as } i \to \infty. \end{aligned} \tag{5.69}$$

This proves the theorem. $\square$

Remark 5.5. The pseudometric $d_{\mathcal{G}}^{\Lambda^2}$ is an important ingredient in the proof. A direct application of (5.28) to handle convergence would not allow for a choice of $\{b_i\}_{i\geq 1}$ which is independent of f and g.

Proof of Theorem 1.31. Suppose that $\mathbb{K} \in \mathcal{A}_0$. Then there exists a sequence of cells $\{\Sigma_i\}_{i\geq 1}$ and a decreasing sequence of positive numbers $\{b_i\}_{i\geq 1}$ such that (A) holds, i.e. for any sequence $\epsilon_i \to 0$ with $0 < \epsilon_i \leq b_i$, $\lim \mathbb{K}^{\Sigma_i,\epsilon_i} = \mathbb{K}$. By the triangle inequality argument giving us (5.68) in the proof of Theorem 1.28, if $\epsilon_i \to 0$ sufficiently fast then we also have $\lim \mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i}$ exists and is equal to $\mathbb{K}$, and it follows from Theorem 5.1 that $\mathbb{K}$ must take the form (1.69). The proves the forward direction.

To prove the converse statement, suppose $\mathbb{K}$ is a Markov kernel on $\mathbf{P} \times \mathbb{S}_+^2$ which takes the form (1.69) where $\widetilde{\mathbb{P}}$ preserves the measure $\sin\theta \mathrm{d}\theta$. By Theorem 5.1, there is a sequence of cells Σ_i and a decreasing sequence of positive numbers $b_i^{(1)}$ such that $\mathbb{K} = \lim_{i\to\infty} \mathbb{K}_{\text{cyl}}^{\Sigma_i,\epsilon_i}$ whenever $0 < \epsilon_i \leq b_i^{(1)}$ and $\epsilon_i \to 0$. By the triangle inequality argument giving us (5.69) in the proof of Theorem 1.28, we have that $\mathbb{K}^{\Sigma_i,\epsilon_i} \to \mathbb{K}$ provided that $\epsilon_i \to 0$ sufficiently fast. This implies that $\mathbb{K} \in \mathcal{A}_0$. $\square$

5.3 Zooming argument

In this subsection, we will define the set Ω and prove Lemma 5.3. The important idea from this section is to compare the billiard evolutions in the respective configuration spaces $\mathcal{M}$ and $\mathcal{M}_{\text{cyl}}$ after "zooming in" by a factor of ϵ^{-1}.

Throughout this subsection, the cell Σ is fixed. The scale $\epsilon > 0$ is also fixed, except at the very end.

5.3.1 Notation and elementary observations

For $r > 0$, we define

$$\sigma_r(y, w) = (ry, w), \qquad (y, w) \in \mathbb{R}^3 \times \mathbb{R}^3. \tag{5.70}$$

Throughout this subsection, if A is a subset of $\mathbb{R}^3$ and B is a subset of $\mathbb{R}^3 \times \mathbb{R}^3$ we will denote

$$\begin{aligned} A^* &= \epsilon^{-1} A = \{y \in \mathbb{R} : \epsilon y \in A\}, \\ B^* &= \sigma_{\epsilon^{-1}}(B) = \{(y, w) \in \mathbb{R}^3 \times \mathbb{R}^3 : (\epsilon y, w) \in B\}. \end{aligned} \tag{5.71}$$

Thus, for example, we will consider $\mathcal{M}^*$, $\mathcal{M}^*_{\text{cyl}}$, $\mathcal{Z}^*$, $\widehat{\mathcal{Z}}^*$, $\Gamma^*_{\text{roll}} \subset \partial\mathcal{M}^*$, and so forth.

As before, we let $\mathbf{Q}_0 = \{(x_1, x_2, \alpha) : \alpha = 0\}$. We will often identify this plane with $\mathbb{R}^2$.

One motivation for "zooming" is that the zoomed cylindrical configuration space $\mathcal{M}^*_{\text{cyl}}$ does not depend on ϵ. In particular, $\mathcal{M}^*_{\text{cyl}}$ is the cylinder with base

$$\widehat{B}^* := \epsilon^{-1}\widehat{B}(\epsilon) = \overline{W(\Sigma, 1)^c} + e_2 \subset \mathbf{Q}_0 \tag{5.72}$$

and axis χ. Neither the base nor the axis depend on ϵ. Another important observation is that

$$\mathcal{M}^*_{\text{cyl}} \cap \mathbf{Q}_0 = \widehat{B}^* = \mathcal{M}^* \cap \mathbf{Q}_0. \tag{5.73}$$

(The second equality follows from the parametrization of the configuration space given by Proposition 4.2(iii).) Thus, while $\mathcal{M}^*$ does depend on ϵ, its intersection

with $\mathbf{Q}_0$ does not.

By Proposition 4.2(iii), the subset of the zoomed configuration space $\mathcal{M}^* \cap \widehat{\mathcal{Z}}^*$ is parametrized by $f^* : \overline{W(\Sigma,1)^c} \times \mathcal{Z}^* \to \mathbb{R}^3$, defined by

$$f^*(x_1, x_2, \alpha) = \epsilon^{-1} f(\epsilon x_1, \epsilon x_2, \epsilon\alpha) = \begin{pmatrix} x_1 - \epsilon^{-1}\sin(\epsilon\overline{\alpha}^*) \\ x_2 + 1 + \epsilon^{-1}(-1 + \cos(\epsilon\overline{\alpha}^*)) \\ \alpha \end{pmatrix}, \tag{5.74}$$

where $\overline{\alpha}^* = \alpha - \overline{k}^*\rho$ and $\overline{k}^* = \operatorname{argmin}\{|\alpha - k\epsilon^{-1}\rho| : k \in \mathbb{Z}\}$. In particular, note that the restriction of f^* to $\partial W(\Sigma,1) \times \mathcal{Z}^* \to \mathbb{R}^3$ parametrizes $\Gamma^*_{\text{roll}} = \partial\mathcal{M}^* \cap \widehat{\mathcal{Z}}^*$.

Recall the diffeomorphism $H_1 : \widehat{\mathcal{Z}} \to \widehat{\mathcal{Z}}$ defined by (4.44). We define $H_\epsilon : \widehat{\mathcal{Z}}^* \to \widehat{\mathcal{Z}}^*$ by

$$H_\epsilon(x_1, x_2, \alpha) = \epsilon^{-1} H_1(\epsilon x_1, \epsilon x_2, \epsilon\alpha) = \begin{pmatrix} x_1 + \overline{\alpha}^* - \epsilon^{-1}\sin(\epsilon\overline{\alpha}^*) \\ x_2 + \epsilon^{-1}(-1 + \cos(\epsilon\overline{\alpha}^*)) \\ \alpha \end{pmatrix}. \tag{5.75}$$

(Arguably, we should denote this map instead by H^*. However, the parameter ϵ will play an important role in our arguments.)

Corresponding to (4.47), we have

$$\begin{aligned} &H_\epsilon(\mathcal{M}^*_{\text{cyl}} \cap \widehat{\mathcal{Z}}^*) = \mathcal{M}^* \cap \widehat{\mathcal{Z}}^*, \text{ and} \\ &H_\epsilon(\partial\mathcal{M}^*_{\text{cyl}} \cap \widehat{\mathcal{Z}}^*) = \partial\mathcal{M}^* \cap \widehat{\mathcal{Z}}^*. \end{aligned} \tag{5.76}$$

The motivation behind "zooming" becomes clearer by observing that

$$H_\epsilon(y) = y + \text{Error}(y, \epsilon), \tag{5.77}$$

where the error term and its first derivatives converge uniformly to zero as $\epsilon \to 0$. Thus H_ϵ maps straight chords in $\mathcal{M}^*_{\text{cyl}}$ to "almost straight" chords in $\mathcal{M}^*$, and this suggests that a billiard trajectory in $\mathcal{M}^*$ will in some sense be closely tracked by a billiard trajectory in $\mathcal{M}^*_{\text{cyl}}$ as $\epsilon \to 0$.

5.3.2 Projection onto the cylindrical base

We will analyze separately the projection of the billiard trajectory onto the cylindrical base and the cylindrical axis. Define a projection map $G_0 : \mathbb{R}^3 \to \mathbf{Q}_0$ by

$$G_0(x_1, x_2, \alpha) = (x_1 + \alpha, x_2). \tag{5.78}$$

Note that $G_0(\chi) = 0$, and consequently, G_0 projects the cylinder $\mathcal{M}^*_{\text{cyl}}$ onto its base $\widehat{B}^*$ and the cylindrical boundary $\partial\mathcal{M}^*_{\text{cyl}}$ onto $\partial\widehat{B}^*$.

Therefore, by (5.76), the "perturbed" projection map

$$G_\epsilon := G_0 \circ H_\epsilon^{-1} : \widehat{\mathcal{Z}}^* \to \mathbf{Q}_0 \tag{5.79}$$

maps $\mathcal{M}^* \cap \widehat{\mathcal{Z}}^*$ onto $\widehat{B}^*$ and $\partial\mathcal{M}^* \cap \widehat{\mathcal{Z}}^*$ onto $\partial\widehat{B}^*$. Explicitly, G_ϵ is given by the formula

$$G_\epsilon(x_1, x_2, \alpha) = \begin{pmatrix} x_1 + \epsilon^{-1}\sin(\epsilon\overline{\alpha}^*) \\ x_2 - \epsilon^{-1}(-1 + \cos(\epsilon\overline{\alpha}^*)) \end{pmatrix}. \tag{5.80}$$

The goal is to compare the projection under G_0 of the billiard trajectory in $\mathcal{M}^*_{\text{cyl}}$ to the projection under G_ϵ of the billiard trajectory in $\mathcal{M}^*$. The advantage of considering the projection is that the image of G_ϵ is $\widehat{B}^*$ which does not depend on ϵ.

Towards our goal, we first provide estimates on the differential of G_ϵ.

Lemma 5.6. *Let $\epsilon \geq 0$, $y = (x_1, x_2, \alpha) \in \mathbb{R}^3$ and $w = (v_1, v_2, \omega) \in \mathbb{S}^2$. Set $\varphi = \angle(w, \chi)$. Then*

$$\begin{aligned}[1 - ((mJ)^{-1/2} + 1)|\epsilon\overline{\alpha}^*|] \sin\varphi \leq \| \mathrm{d}(G_\epsilon)_y (w)\| \\ \leq (1 + mJ^{-1})^{1/2}[1 + ((mJ)^{-1/2} + 1)|\epsilon\overline{\alpha}^*|] \sin\varphi\end{aligned} \tag{5.81}$$

Proof. One may easily prove the result if $\epsilon = 0$, because then G_0 is just the linear map which projects vectors in the direction χ onto the plane $\mathbf{Q}_0$, and

$$\mathrm{d}(G_0)_y (w) = \begin{pmatrix} v_1 + \omega \\ v_2 \end{pmatrix}. \tag{5.82}$$

Indeed, if θ is the angle between χ and the plane $\mathbf{Q}_0$, then $\mathrm{d}(G_0)_y (w)$ must lie on the boundary of the ellipse with major and minor axes $\sin\varphi$ and $\frac{\sin\varphi}{\sin\theta}$ respectively; hence

$$\sin\varphi \leq \mathrm{d}(G_0)_y (w) \leq \frac{\sin\varphi}{\sin\theta}. \tag{5.83}$$

Noting that $\widehat{e}_3 := (0, 0, J^{-1/2})$ is the unit normal to the plane $\mathbf{Q}_0$, we have

$$\sin\theta = \langle \chi, n_0 \rangle = \left(\frac{J}{m + J}\right)^{1/2} \tag{5.84}$$

Substituting this into the previous equality yields

$$\sin\varphi \leq \mathrm{d}(G_0)_y (w) \leq (1 + mJ^{-1})^{1/2} \sin\varphi. \tag{5.85}$$

Now assume $\epsilon > 0$. The differential of G_ϵ is

$$\mathrm{d}(G_\epsilon)_y = \begin{bmatrix} 1 & 0 & \cos(\epsilon\overline{\alpha}^*) \\ 0 & 1 & \sin(\epsilon\overline{\alpha}^*) \end{bmatrix}. \tag{5.86}$$

Thus

$$\|\mathrm{d}(G_\epsilon)_y(w)\|^2 = (v_1+\omega)^2 + (v_2)^2 + E(\epsilon, y, w) = \|\mathrm{d}(G_0)_y(w)\|^2 + E(\epsilon, y, w), \tag{5.87}$$

where

$$E(\epsilon, y, w) = 2\omega[-v_1(1-\cos(\epsilon\overline{\alpha}^*)) + v_2\sin(\epsilon\overline{\alpha}^*)] + \omega^2\sin^2(\epsilon\overline{\alpha}^*). \tag{5.88}$$

We estimate the middle term

$$\begin{aligned}
&2\omega[-v_1(1-\cos(\epsilon\overline{\alpha}^*)) + v_2\sin(\epsilon\overline{\alpha}^*)] \\
&\le 2\omega\sqrt{(v_1)^2+(v_2)^2}\sqrt{(1-\cos(\epsilon\overline{\alpha}^*))^2+\sin^2(\epsilon\overline{\alpha}^*)} \qquad \text{(Cauchy-Schwarz)} \\
&= 2\omega\sqrt{(v_1)^2+(v_2)^2}\sqrt{2-2\cos(\epsilon\overline{\alpha}^*)} \\
&\le 2\omega\sqrt{(v_1)^2+(v_2)^2}|\epsilon\overline{\alpha}^*| \qquad \text{(using } 1-\cos(x) \le x^2/2) \\
&= 2(mJ)^{-1/2}(J^{1/2}\omega m^{1/2}\sqrt{(v_1)^2+(v_2)^2}|\epsilon\overline{\alpha}^*| \\
&\le (mJ)^{-1/2}[m(v_1)^2 + m(v_2)^2 + J\omega^2]|\epsilon\overline{\alpha}^*| \\
&= (mJ)^{-1/2}\|w\|^2|\epsilon\overline{\alpha}^*| = (mJ)^{-1/2}|\epsilon\overline{\alpha}^*|,
\end{aligned} \tag{5.89}$$

where the second-to-last line above follows from the Arithmetic-Geometric Mean

Inequality. Since $\omega^2 \le \|w\|^2 = 1$, we have

$$|E(\epsilon, y, w)| \le ((mJ)^{-1/2} + 1)|\epsilon\overline{\alpha}^*|. \tag{5.90}$$

By (5.87), we obtain

$$\begin{aligned} \|\,\mathrm{d}(G_0)_y\,(w)\|^2 - ((mJ)^{-1/2} + 1)|\epsilon\overline{\alpha}^*| &\le \|\,\mathrm{d}(G_\epsilon)_y\,(w)\|^2 \\ &\le \|\,\mathrm{d}(G_\epsilon)_y\,(w)\|^2 + ((mJ)^{-1/2} + 1)|\epsilon\overline{\alpha}^*|. \end{aligned} \tag{5.91}$$

Taking the square root of each side, and using the fact that $\sqrt{x+h} \le \sqrt{x}(1+h)$ for $h > 0$ and $\sqrt{x+h} \ge \sqrt{x}(1+h)$ for $h < 0$, we conclude that

$$\begin{aligned} \|\,\mathrm{d}(G_0)_y\,(w)\|(1 - ((mJ)^{-1/2} + 1)|\epsilon\overline{\alpha}^*|) &\le \|\,\mathrm{d}(G_\epsilon)_y\,(w)\| \\ &\le \|\,\mathrm{d}(G_0)_y\,(w)\|(1 + ((mJ)^{-1/2} + 1)|\epsilon\overline{\alpha}^*|). \end{aligned} \tag{5.92}$$

By (5.85) and the above, we obtain (5.81). □

5.3.3 Radius of transversality

To state our next result, we must introduce some additional terminology. Consider a closed set $\mathcal{D} \subset \mathbb{R}^2$. Assume that $\mathcal{D}$ is the closure of its interior and has a piecewise C^2 boundary (in the sense of §1.2.2). Assume that some decomposition $\partial\mathcal{D} = \bigcup_{i=1}^{\infty} I_i$ into closed C^2 curve segments has been fixed.

We will call a line segment between two points p and p' in $\partial\mathcal{D}$ such that the interior of the line segment lies in $\operatorname{Int}\mathcal{D}$ a *chord* of $\mathcal{D}$.

Consider a chord C of the domain $\mathcal{D}$ with endpoints $p, p' \in \partial\mathcal{D}$, and let $v = \frac{p'-p}{\|p'-p\|}$.

(For example, C might be the segment of the billiard trajectory starting from the initial state (p, v).)

Let $I, I' \subset \partial\mathcal{D}$ denote C^2 curve segments from the fixed decomposition $\{I_i\}$ containing p, p' respectively, and let $L(p, p')$ denote the line through points p and p'.

Let $Z \subset I$ be the set of all points $q \in \operatorname{Int} I$ such that the ray starting from q with direction v first returns to the boundary at a point $q' \in \operatorname{Int} I'$ and such that the line $L(q, q')$ is not tangent to the curve segments I, I' at the points q, q' respectively.

We define the *radius of transversality associated with* C to be the number

$$r_{\mathcal{D}}(C) = r_{\mathcal{D}}(p, v) = \sup\{r \geq 0 : (\forall q \in I)\ \operatorname{dist}(q, L(p, p')) \leq r \Rightarrow q \in Z\}. \tag{5.93}$$

This quantity is well-defined up to the choice of decomposition of $\partial\mathcal{D}$ into C^2 curve segments $\{I_i\}$. Indeed, if either p or p' lies in at the endpoint of a curve segment in the fixed decomposition of $\partial\mathcal{D}$, then $r_{\mathcal{D}}(p, v) = 0$ for any choice of I and I'. On the other hand, if p and p' both lie in the interiors of curve segments, then I and I' are uniquely determined.

The concept is illustrated in Figure 16. The term "transversality" refers to the fact that the lines $L(q, q')$ must cross the segments I, I' transversally. We will also sometimes refer to this as the radius of transversality associated with state (p, v) if C arises as a segment of a billiard trajectory.

5.3.4 Modulus of continuity lemmas

To state our next results, we need to refer to some aspects of the construction of the cylindrical collision law $K^{\epsilon}_{\text{cyl}}$. According to the definition of $K^{\epsilon}_{\text{cyl}}$ given in §4.3.2, the

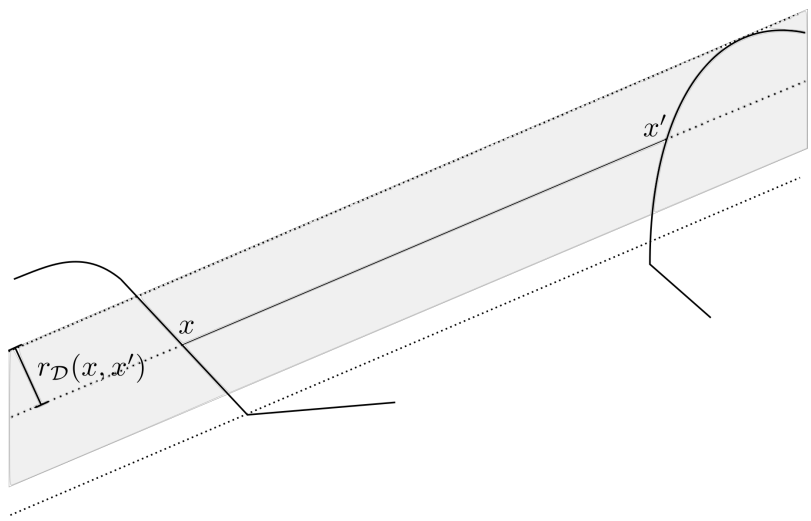

Figure 16: Radius of transversality

cylindrical collision law is a special case of the general macro-reflection laws defined in §6.2.1 (taking $\mathcal{M}_1 = \mathcal{M}_{\text{cyl}}$ and $\mathcal{M}_0 = \{(x_1, x_2, \alpha) : x_2 \geq 0\}$). Adapting notation from Remark 6.6 to this setting, we have

$$\mathcal{N} = \overline{\{(x_1, x_2, \alpha) \in \mathcal{M}_{\text{cyl}} : x_2 < 0\}}, \tag{5.94}$$

and the plane $\mathbf{P}$ may be expressed as the disjoint union

$$\mathbf{P} = A_1 \cup A_2 \cup A_3, \tag{5.95}$$

where

$$A_1 = \mathbf{P} \cap \mathcal{N} \cap \operatorname{Int} \mathcal{M}_{\text{cyl}}, \qquad A_2 = \mathbf{P} \cap \mathcal{N} \setminus \operatorname{Int} \mathcal{M}_{\text{cyl}}, \qquad A_3 = \mathbf{P} \setminus \mathcal{N}. \tag{5.96}$$

The set A_2 has Lebesgue measure zero, and consequently $\mathcal{N}$ belongs to the class $\mathrm{CES}_0^2(\mathbb{R}^3)$. (See the proof of Proposition 4.11.)

The billiard domain $\mathcal{N}$ plays a key role in the definition of $K_{\mathrm{cyl}}^{\epsilon}$. If $(y,w) \in A_3 \times \mathbb{S}_+^2$, then the collision law maps (y,w) to $R(y,-w)$, where R denotes specular reflection. And if $(y,w) \in A_1 \times \mathbb{S}_+^2$, then (except on a null set) the billiard trajectory starting from $(y,-w)$ will enter $\mathcal{N}$ and reflect from $\partial\mathcal{M}_{\mathrm{cyl}}$ a finite number of times before eventually returning to $\mathbf{P}$ in a state (y',w'); in this case $K_{\mathrm{cyl}}^{\epsilon}(y,w) := (y',w')$. The details of the more general case are covered in §6.2.1 and Remark 6.6.

We would like to study the billiard trajectory in $\mathcal{N}$ after zooming. That is, we introduce "zoomed" versions of the sets defined above:

$$\mathcal{N}^* = \epsilon^{-1}\mathcal{N} = \overline{\{(x_1,x_2,\alpha) \in \mathcal{M}_{\mathrm{cyl}}^* : x_2 < 0\}}, \qquad A_i^* = \epsilon^{-1}A_i, \quad i = 1,2,3. \tag{5.97}$$

The set $\mathcal{N}^*$ is cylindrical, with axis χ and base

$$\mathcal{D}^* := G_0(\mathcal{N}^*) = \overline{\{(x_1,x_2) \in \widehat{B}^* : x_2 < 0\}} \subset \mathbf{Q}_0, \tag{5.98}$$

where we recall that $\widehat{B}^*$ is the base of the cylinder $\mathcal{M}_{\mathrm{cyl}}^*$ in the plane $\mathbf{Q}_0$; see (5.72). By our assumptions on the wall, the boundary $\partial\widehat{B}^*$ is piecewise C^2; it decomposes as a union of compact C^2 curves $\partial\widehat{B}^* = \bigcup_{i\in\mathbb{Z}} \widehat{\Gamma}_i^*$; for each i, if the interior of $\widehat{\Gamma}_i^*$ intersects the line $\mathbf{L} := \{(x_1,x_2) : x_2 = 0\}$, then $\widehat{\Gamma}_i^* \subset \mathbf{L}$. Consequently, $\mathcal{D}^*$ also has a piecewise C^2 boundary, consisting of a union of some subcollection of the curve segments $\{\widehat{\Gamma}_i^*, i \in \mathbb{Z}\}$ and a collection of line segments in $\mathbf{L}$. Fixing such a decomposition of $\partial\mathcal{D}^*$, we can unambiguously refer to the radius of transversality of a chord of $\mathcal{D}^*$.

We will now very carefully study a single segment of the billiard trajectory in $\mathcal{N}^*$ and its projected image in $\mathcal{D}^*$.

Let $y^0 \in \partial\mathcal{N}^*$ and $w^0 \in \mathbb{S}^2$. We assume that the billiard trajectory starting from initial state (y^0, w^0) immediately enters the interior of $\mathcal{N}^*$ and first returns to $\partial\mathcal{N}^*$ at a point y^1. We also assume that if $y^i \in \partial\mathcal{M}^*_{\text{cyl}}$ then the unit normal vector at y^i exists, for $i = 0, 1$. These assumptions will be satisfied for example if (y^0, w^0) is in the domain of the billiard map for the billiard domain $\mathcal{N}^*$.

If $y^1 \in \partial\mathcal{M}^*_{\text{cyl}}$, let w^1 denote the velocity of the point particle after reflecting specularly at y^1. Otherwise (if $y^1 \in \mathbf{P}$) let $w^1 = w^0$.

Recall the surface $\widetilde{\mathbf{P}} = H_1(\mathbf{P} \cap \widehat{\mathcal{Z}})$, which is the boundary of the open sets $\mathcal{O}_\pm = H_1(\mathbb{R}^3_\pm \cap \widehat{\mathcal{Z}}^*)$ (see §4.3.3). We have

$$\widetilde{\mathbf{P}}^* := \epsilon^{-1}\widetilde{\mathbf{P}} = H_\epsilon(\mathbf{P} \cap \widehat{\mathcal{Z}}^*), \qquad \mathcal{O}^*_\pm := \epsilon^{-1}\mathcal{O}_\pm = H_\epsilon(\mathbb{R}^3_\pm \cap \widehat{\mathcal{Z}}^*). \tag{5.99}$$

We want to compare the segment of the billiard trajectory running from (y^0, w^0) to (y^1, w^1) to a segment of a billiard trajectory in $\mathcal{M}^*$. However, rather than taking as given that the segment of the trajectory in $\mathcal{M}^*$ has nice properties, as we have for the segment in $\mathcal{M}^*_{\text{cyl}}$, we will state conditions in the lemmas below which guarantee that this is the case. Our goal in Lemmas 5.7 and 5.8 is to be able to control the billiard trajectory in $\mathcal{M}^*$ in terms of properties of the billiard trajectory in $\mathcal{M}^*_{\text{cyl}}$. See Figure 17

Let $\widetilde{y}^0 \in (\partial\mathcal{M}^* \cap \widehat{\mathcal{Z}}^*) \cup \widetilde{\mathbf{P}}^*$, and let $\widetilde{w}^0 \in \mathbb{S}^2$. We say that the pair $(\widetilde{y}^0, \widetilde{w}^0)$ is *licit* if (i) the billiard trajectory starting from initial state $(\widetilde{y}^0, \widetilde{w}^0)$ immediately enters $\operatorname{Int}\mathcal{M}^* \cap \mathcal{O}_-$ and first returns to $(\partial\mathcal{M}^* \cap \widehat{\mathcal{Z}}^*) \cup \widetilde{\mathbf{P}}^*$ at a point $\widetilde{y}^1$; and (ii) if $\widetilde{y}^i \in \partial\mathcal{M}^*$

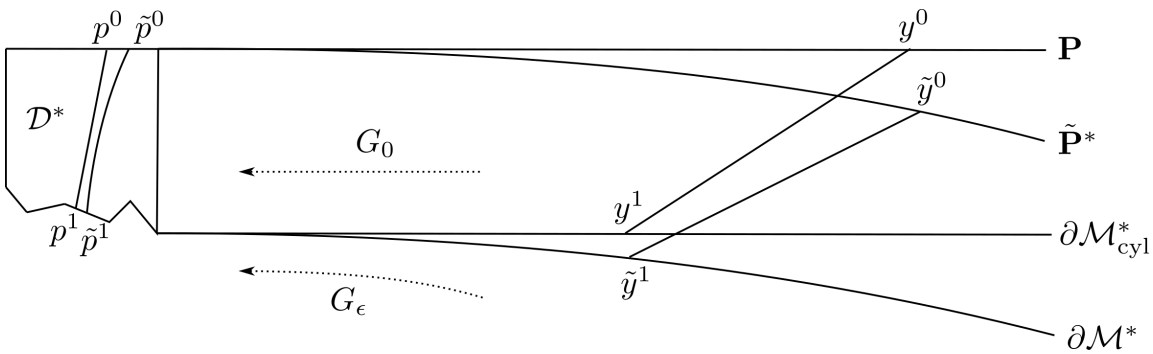

Figure 17: The line segment from y^0 to y^1 projects under G_0 to the chord of the cylindrical base $\mathcal{D}^*$ from p^0 to p^1. The line segment from $\widetilde{y}^0$ to $\widetilde{y}^1$ projects under G_ϵ to a (slightly) curved segment from $\widetilde{p}^0$ to $\widetilde{p}^1$ whose interior lies in $\operatorname{Int}\mathcal{D}^*$.

then there is a well-defined inward-pointing unit normal vector at $\widetilde{y}^i$, $i = 0, 1$. If $\widetilde{y}^1 \in \partial\mathcal{M}^*$, let $\widetilde{w}^1$ denote the velocity of the point particle after reflecting at y^1. Otherwise (if $\widetilde{y}^1 \in \widetilde{\mathbf{P}}^*$) let $\widetilde{w}^1 = \widetilde{w}^0$.

For $i = 0, 1$, we use the following notation:

- α^i denotes the angular coordinate of y^i, and $\widetilde{\alpha}^i$ denotes the angular coordinate of $\widetilde{y}^i$.
- ω^i denotes the angular coordinate of w^i, and $\widetilde{\omega}^i$ denotes the angular coordinate of $\widetilde{w}^i$.
- $(p^i, u^i) = \mathrm{d}G_0(y^i, w^i)$ and $(\widetilde{p}^i, \widetilde{u}^i) = \mathrm{d}G_\epsilon(\widetilde{y}^i, \widetilde{w}^i)$.

The quantities above whose definition depends on $(\widetilde{y}^0, \widetilde{w}^0)$ being licit are: $\widetilde{y}^1$, $\widetilde{w}^1$, $\widetilde{\alpha}^1$, $\widetilde{\omega}^1$, $\widetilde{p}^1$, and $\widetilde{u}^1$.

The projection G_0 maps $\mathcal{N}^* = \overline{\mathbb{R}_- \cap \mathcal{M}_{\mathrm{cyl}}}$ onto $\mathcal{D}^*$, mapping the boundary onto the boundary and the interior onto the interior. Consequently, G_ϵ maps $H_\epsilon(\overline{\mathbb{R}_- \cap \mathcal{M}_{\mathrm{cyl}}}) =$

$\overline{\mathcal{M}^* \cap \mathcal{O}_-^*}$ onto $\mathcal{D}^*$, mapping the boundary onto the boundary and the interior onto the interior. The line segment from p^0 to p^1 is therefore a chord of $\mathcal{D}^*$. Similarly, the image under G_ϵ of the line segment from y^0 to y^1 is a closed (not necessarily straight) curve segment whose interior lies in $\operatorname{Int}\mathcal{D}^*$.

We also use the following notation:

- $L = \|p^1 - p^0\|$.
- θ_0 is the angle between u^0 and the inward pointing normal vector at $p^0 \in \partial\mathcal{D}^*$, and θ_1 is the angle between $-u^0$ and the inward pointing normal at $p^1 \in \partial\mathcal{D}^*$.
- ν is the angle between w and the line spanned by χ.
- $\widehat{u}^0 = u^0/\|u^0\|$.
- ξ is the unique unit vector such that $(\xi, \widehat{u}^0)$ is a positively oriented orthonormal basis.
- $r_0 = r_{\mathcal{D}^*}(p, \widehat{u}^0)$ is the radius of transversality associated with the chord from p^0 to p^1 in $\mathcal{D}^*$.

All of these quantities depend only on the segment of the billiard trajectory in the cylindrical configuration space. See Figure 18.

By the definition of the radius of transversality and the implicit function theorem, there exist C^2 functions $f_0, f_1 : (-r_0, r_0) \to \mathbb{R}$ such that for all $h \in (-r_0, r_0)$, $p^0 + h\xi + f_0(h)\widehat{u}^0$ lies in the C^2 curve segment containing p^0 and $p^0 + h\xi + f_1(h)\widehat{u}^0$ lies in the

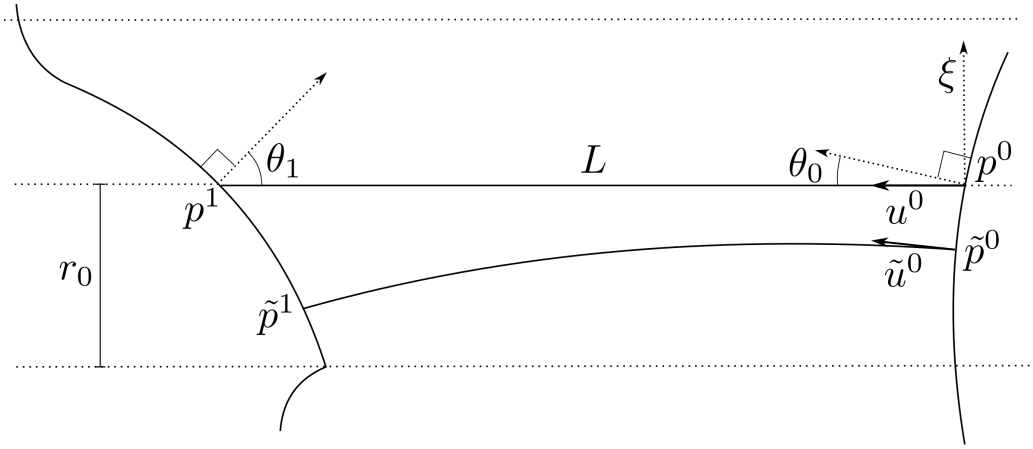

Figure 18: Projected segments of the billiard trajectories.

C^2 curve segment containing p^1. Moreover, the strip

$$\{h\xi + h'\widehat{u}^0 : -r_0 < h < r_0, f_0(h) < h' < f_1(h)\} \subset \mathbb{R}^2 \tag{5.100}$$

does not intersect the wall $W(\Sigma, 1)$.

Recall that the maximum curvature $\kappa_{\max}$ of the C^2 curve segments constituting $\partial W(\Sigma, 1)$ is finite (see Remark 1.18). Let $L_{\max}$ denote the supremum of the lengths of all chords of $\partial W(\Sigma, 1)$. Recall that by construction, $W(\Sigma, 1)$ satisfies conditions A5a and A5b from §1.2.2. These conditions imply that $L_{\max}$ is finite. Let

$$\overline{\kappa} = \max\{\kappa_{\max}, 1\}, \qquad \overline{L} = \max\{L_{\max}, 1\}. \tag{5.101}$$

In our formulas below, we would like to consider the inverse of the function ρ, but this is not possible since ρ is not strictly increasing (by assumption, it is increasing

but its values are always some fraction of 2π). Therefore, we instead consider

$$\breve{\rho}^{-1}(s) := \sup\{t \in (0,\infty) : \rho(t) < s\}, \qquad s \in (0,\infty). \tag{5.102}$$

It follows from the assumptions on the growth of $\rho(\epsilon)$, stated in 1.3.1 that $\frac{\breve{\rho}^{-1}(s)}{s^2} \to 0$ as $s \to 0$.

Lemma 5.7. *Let* $0 < s_2, s_3 \le 1$ *and*

$$0 < s_1 \le \min\left\{r_0, \frac{5\cos\theta_1}{2\overline{\kappa}}, \frac{5s_2\sin\nu}{12}, \frac{s_3}{24\sqrt{10}\overline{\kappa}}\right\}, \tag{5.103}$$

and assume that

$$\epsilon \le \min\left\{\breve{\rho}^{-1}\left(\frac{s_1\cos\theta_1\sin\nu}{10(L+1)}\right), \frac{s_1\cos\theta_1\sin^2\nu}{10(L+1)^2}, \frac{5s_2\sin\nu}{12}, \breve{\rho}^{-1}\left(\frac{s_3}{12\sqrt{2}}\right),\right\}, \tag{5.104}$$

and

$$|\alpha^0| < \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) - \left(\frac{L}{\sin\nu} + 1\right). \tag{5.105}$$

Assume in addition that

$$\begin{aligned} &\|\widetilde{p}^0 - p^0\| \le \frac{s_1\cos\theta_1}{20}, \\ &|\widetilde{\alpha}^0 - \alpha^0| \le \frac{s_2}{3}, \qquad \textit{and} \\ &\|\widetilde{w}^0 - w^0\| \le \min\left\{\frac{s_1\cos\theta_1\sin\nu}{20(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(L+1)}, \frac{s_2\sin\nu}{3L}, \frac{s_3}{12}\right\}. \end{aligned} \tag{5.106}$$

Then the pair $(\widetilde{y}^0, \widetilde{w}^0)$ *is licit, in the sense defined above. Moreover, the point* $\widetilde{p}^1$ *lies on the same curve segment of* $\partial W(\Sigma, 1)$ *as* p^1 *(with respect to the given decomposi-*

tion), and

$$\|\widetilde{p}^1 - p^1\| \le s_1, \quad |\widetilde{\alpha}^1 - \alpha^1| \le s_2, \qquad \|\widetilde{w}^1 - w^1\| \le s_3. \tag{5.107}$$

A weaker but more convenient form of the lemma is as follows:

Lemma 5.8. *Assume that*

$$s \le 79 \min\left\{r_0, (\overline{\kappa})^{-1} \cos\theta_1\right\}, \tag{5.108}$$

and

$$\epsilon \le \min\left\{\breve{\rho}^{-1}\left(\frac{s\cos\theta_1 \sin^2\nu}{800\overline{\kappa}(\overline{L}+1)}\right), \frac{s\cos\theta_1\sin^3\nu}{800\overline{\kappa}(\overline{L}+1)^2}\right\}, \tag{5.109}$$

and

$$|\alpha^0| \le \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) - \left(\frac{\overline{L}}{\sin\nu} + 1\right). \tag{5.110}$$

Assume in addition that

$$\|\widetilde{p}^0 - p^0\| + |\widetilde{\alpha}^0 - \alpha^0| + \|\widetilde{w}^0 - w^0\| \le \frac{s\cos\theta_1\sin^2\nu}{1600(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})\overline{\kappa}(\overline{L}+1)}. \tag{5.111}$$

Then the pair $(\widetilde{y}^0, \widetilde{w}^0)$ *is licit. Moreover,* $\widetilde{p}^1$ *lies on the same curve segment of* $\partial W(\Sigma, 1)$ *as* p^1*, and*

$$\|\widetilde{p}^1 - p^1\| + |\widetilde{\alpha}^1 - \alpha^1| + \|\widetilde{w}^1 - w^1\| \le s. \tag{5.112}$$

We now prove the two lemmas.

Proof of Lemma 5.7. Step 1. Recall that if I' denotes the curve segment containing

p^1, and I_1 denotes the connected component of the intersection of I' with the strip $\{x \in \mathbb{R}^2 : |\langle x - p^0, \xi\rangle| < r_0\}$, then the transversality radius I_1 may be represented as the graph of the function $f_1 : (-r_0, r_0) \to \mathbb{R}$, in the sense that

$$I_1 = \{h\xi + f_1(h)\widetilde{u}^0 : h \in (-r_0, r_0)\}. \tag{5.113}$$

The proof of the lemma depends on the following two claims.

Claim 5.7.1. *Let $p : [a, b] \to \mathbb{R}^2$ be a unit speed parametrization of the curve segment I_1 containing p^1, with $a < 0 < b$ and $p(0) = p^1$. Let $0 < c < \min\{|a|, |b|\}$ be such that*

$$|\langle p(c) - p(0), \xi\rangle| < \min\left\{r_0, \frac{(\overline{\kappa})^{-1}}{2}\right\}. \tag{5.114}$$

Then the functional $\sigma \mapsto \|p(\sigma) - p(0)\|$ is monotone increasing on the interval $[0, c]$ and monotone decreasing on the interval $[-c, 0]$.

Claim 5.7.2. *Consider the projected billiard trajectories $\gamma(t) := G_0(y^0 + tw^0)$, and $\widetilde{\gamma}(t) := G_\epsilon(\widetilde{y}^0 + t\widetilde{w}^0)$, where $t \geq 0$. Let*

$$t_1 = \inf\{t > 0 : \gamma(t) \in \partial\mathcal{D}^*\}, \quad \widetilde{t}_1 = \inf\{t > 0 : \widetilde{\gamma}(t) \in \partial\mathcal{D}^*\} \tag{5.115}$$

(defined equal to ∞ if the set over which we take the infimum is empty). Then $t_1 < \infty$. Moreover, if we assume that

$$\|\widetilde{\gamma}(t) - \gamma(t)\| \leq s_1 \cos\theta_1/5 \quad \textit{for all } 0 \leq t \leq t_1 + \frac{4s_1}{5\|u^0\|}, \tag{5.116}$$

then

(i) $t_1 - \frac{4s_1}{5\|u^0\|} \le \widetilde{t}_1 \le t_1 + \frac{4s_1}{5\|u^0\|}$, *and*

(ii) $\|\widetilde{p}^1 - p^1\| \le s_1$.

Proof of Claim 5.7.1. By symmetry it is enough to prove that the functional is increasing on $[0, c]$. Let $g_1(\sigma) = \langle p(\sigma) - p(0), \xi\rangle$ and $g_2(\sigma) = \langle p(\sigma) - p(0), \widehat{u}^0\rangle$ for $\sigma \in [0, c]$. Since $|\langle p(c) - p(0), \xi\rangle| < r_0$, it follows that the part of the curve segment between $p(0)$ and $p(c)$ lies on the graph of the function f_1. Consequently, g_1 is nondecreasing on $[0, c]$. We may write

$$\|p(\sigma) - p(0)\|^2 = g_1(\sigma)^2 + g_2(\sigma)^2.$$

If g_2 does not attain an extreme value in $(0, c)$, then the above functional is nondecreasing, and we are done. Suppose g_2 does attain an extreme value at some point $\sigma^* \in (0, c)$. Then $g_2'(\sigma^*) = \langle p'(\sigma^*), \widehat{u}^0\rangle = 0$ and $\langle p'(\sigma^*), \xi\rangle = \pm 1$. Taylor expanding about σ^* gives us

$$p(\sigma) - p(\sigma^*) = p'(\sigma^*)(\sigma - \sigma^*) + E(\sigma), \quad \|E(\sigma)\| \le \frac{\overline{\kappa}}{2}|\sigma - \sigma^*|^2. \tag{5.117}$$

Since the curve segment between $\sigma = 0$ and $\sigma = c$ lies on the graph of the function f_1,

for all $\sigma \in [0, c]$, $|\langle p(c) - p(0), \xi\rangle| \geq |\langle p(\sigma) - p(\sigma^*), \xi\rangle|$, and hence using the hypothesis

$$\begin{aligned}
\frac{(\overline{\kappa})^{-1}}{2} &> |\langle p(c) - p(0), \xi\rangle| \\
&\geq |\langle p(\sigma) - p(\sigma^*), \xi\rangle| \\
&= |\langle p'(\sigma^*), \xi\rangle(\sigma - \sigma^*) - \langle E(\sigma), \xi\rangle| \\
&\geq |\sigma - \sigma^*| - \frac{\overline{\kappa}}{2}|\sigma - \sigma^*|^2.
\end{aligned} \tag{5.118}$$

The function $f(s) = s - \frac{\overline{\kappa}}{2}s^2$ achieves a maximum value of $\frac{(\overline{\kappa})^{-1}}{2}$ at $s = (\overline{\kappa})^{-1}$. It follows that $|\sigma - \sigma^*| < (\overline{\kappa})^{-1}$ for all $\sigma \in [0, c]$. Therefore, $c < 2(\overline{\kappa})^{-1}$.

Set $h(\sigma) = \|p(\sigma) - p(0)\|^2$, $\sigma \in [0, c]$. Then

$$\begin{aligned}
h'(\sigma) &= 2\langle p(\sigma) - p(0), p'(\sigma)\rangle \\
&\geq \sigma - \frac{\overline{\kappa}}{2}\sigma^2,
\end{aligned} \tag{5.119}$$

where we Taylor expand about σ to obtain the second line. Since $0 \leq \sigma \leq c < 2(\overline{\kappa})^{-1}$, we see that the last quantity above is nonnegative, and $h'(\sigma) \geq 0$ for $\sigma \in [0, c]$. This proves the claim. ■

Proof of Claim 5.7.2. Since G_0 is linear, we have

$$\gamma(t) = p^0 + tu^0, \tag{5.120}$$

and $t_1 = L/\|u^0\| < \infty$. We also have for $0 \le t \le t_1 + \frac{4s_1}{5\|u^0\|}$,

$$\begin{aligned} |\langle \widetilde{\gamma}(t) - p^0, \xi\rangle| &= |\langle \widetilde{\gamma}(t) - \gamma(t), \xi\rangle + \langle \gamma(t) - p^0, \xi\rangle| \\ &= |\langle \widetilde{\gamma}(t) - \gamma(t), \xi\rangle| \\ &\le \|\langle \widetilde{\gamma}(t) - \gamma(t)\| \le \frac{s_1 \cos\theta_1}{5} < r_0, \end{aligned} \tag{5.121}$$

using (5.103) in the last line. Thus $\widetilde{\gamma}(t)$ lies in the strip $\{x : |\langle x - p^0\rangle| < r_0\}$, for all times $0 \le t \le t_1 + \frac{4s_1}{5\|u^0\|}$.

As before, let $p : [a, b] \to \mathbb{R}^2$ be a unit speed parametrization of I_1 with $p(0) = p^1$. Since the segment I_1 coincides with the graph of the function $f_1 : (-r_0, r_0) \to \mathbb{R}$, as described above, and $s_1 \cos\theta_1/5 < r_0$ by (5.103), there is a maximal interval $[c_1, c_2]$ with $c_1 < 0 < c_2$ such that $p(\sigma)$ lies in the strip $\{x : |\langle x - p^0\rangle| \le s_1 \cos\theta_1/5\}$ for all $\sigma \in [c_1, c_2]$. Reversing the direction of the parametrization if necessary, we may assume that $\sigma \mapsto \langle p(\sigma) - p(0), \xi\rangle$ is increasing on $[c_1, c_2]$. By Taylor expansion, we have

$$p(\sigma) = p(0) + p'(0)\sigma + E(\sigma), \quad \|E(\sigma)\| \le \frac{\overline{\kappa}}{2}\sigma^2. \tag{5.122}$$

Thus for all $\sigma \in [c_1, c_2]$,

$$\begin{aligned} \frac{s_1 \cos\theta_1}{5} &\ge |\langle p(\sigma) - p(0), \xi\rangle| \\ &= |(\cos\theta_1)\sigma + \langle E(\sigma), \xi\rangle| \\ &\ge \cos\theta_1 |\sigma| - \frac{\overline{\kappa}}{2}|\sigma|^2. \end{aligned} \tag{5.123}$$

The function $f(s) = (\cos\theta_1)s - \frac{\overline{\kappa}}{2}s^2$ is increasing on the interval $[0, (\overline{\kappa})^{-1}\cos\theta_1]$ and

achieves a maximum of $\frac{\cos^2\theta_1}{2\overline{\kappa}}$ at $s = (\overline{\kappa})^{-1}\cos\theta_1$. By (5.103) $\frac{s_1\cos\theta_1}{5} < \frac{\cos^2\theta_1}{2\overline{\kappa}}$, and therefore $|\sigma| < (\overline{\kappa})^{-1}\cos\theta_1$. Thus by (5.123), for all $\sigma \in [c_1, c_2]$

$$\begin{aligned} \frac{s_1\cos\theta_1}{5} &\geq \cos\theta_1|\sigma| - \frac{\overline{\kappa}}{2}|\sigma|[(\overline{\kappa})^{-1}\cos\theta_1] \geq \frac{\cos\theta_1|\sigma|}{2} \\ \Rightarrow \quad & \frac{2s_1}{5} \geq |\sigma|. \end{aligned} \tag{5.124}$$

We conclude that

$$\max\{|c_1|, |c_2|\} \leq \frac{2s_1}{5}. \tag{5.125}$$

Let $t^- = t_1 - \frac{4s_1}{5\|u^0\|}$, and let $t^+ = t_1 + \frac{4s_1}{5\|u^0\|}$. Fix $t_0 < t^-$ and $\sigma \in [c_1, c_2]$. Let $\widetilde{x}$ be a point in the disk of radius $s_1\cos\theta_1/5$ centered at $\gamma(t_0)$. We have

$$\|p(\sigma) - \widetilde{x}\| \geq \|\gamma(t_0) - p(0)\| - \|\gamma(t_0) - \widetilde{x}\| - \|p(\sigma) - p(0)\|. \tag{5.126}$$

Note that

$$\|\gamma(t_0) - p(0)\| = \|\gamma(t_0) - \gamma(t_1)\| = |t_0 - t_1| \cdot \|u^0\| \geq |t^- - t_1| \cdot \|u^0\| = \frac{4s_1}{5}. \tag{5.127}$$

Also, by Taylor expansion,

$$\|p(\sigma) - p(0)\| \leq |\sigma| + \frac{\overline{\kappa}}{2}|\sigma|^2 \leq \frac{3}{2}|\sigma| \leq \frac{3}{5}s_1, \tag{5.128}$$

here using the fact proved above that $|\sigma| \leq (\overline{\kappa})^{-1}$ and $|\sigma| \leq 2s_1/5$. Substituting (5.127)

and (5.128) as well as the hypothesized bound on $\|\gamma(t) - \widetilde{\gamma}(t)\|$ into (5.126) gives us

$$\|p(\sigma) - \widetilde{x}\| \geq \frac{4s_1}{5} - \frac{s_1 \cos\theta_1}{5} - \frac{3s_1}{5} > 0. \tag{5.129}$$

Since $\gamma(t_0)$ lies below the graph of f_1 in the center of the strip $\{x : |\langle x - p^0, \xi\rangle| \leq s_1 \cos\theta_1/5\}$, by connectedness this proves that the entire disk of radius $s_1 \cos\theta_1/5$ centered at $\gamma(t_0)$ lies below the graph of f_1 in the strip $\{x : |\langle x - p^0, \xi\rangle| \leq s_1 \cos\theta_1/5\}$. In particular, by the hypothesized bound on $\|\gamma(t) - \widetilde{\gamma}(t)\|$, we see that $\gamma(t_0)$ lies below the graph of f_1 in the strip $\{x : |\langle x - p^0, \xi\rangle| \leq s_1 \cos\theta_1/5\}$. This proves that $\widetilde{t}_1 > t^-$.

A symmetric argument shows that the disk of radius $s_1 \cos\theta_1/5$ and centered at $\gamma(t^+)$ lies above the graph of f_1 in the strip $\{x : |\langle x - p^0, \xi\rangle| \leq s_1 \cos\theta_1/5\}$. Consequently, $\widetilde{\gamma}(t^+)$ lies above the graph of f_1 in the strip $\{x : |\langle x - p^0, \xi\rangle| \leq s_1 \cos\theta_1/5\}$. By the intermediate value theorem, we conclude that $t^- < \widetilde{t}_1 < t^+$, and this proves part (i) of the claim.

To prove part (ii), we make the following estimate using part (i):

$$\begin{aligned}
\|\widetilde{p}^1 - p^1\| &= \|\widetilde{\gamma}(\widetilde{t}_1) - \gamma(t_1)\| \\
&\leq \|\widetilde{\gamma}(\widetilde{t}_1) - \gamma(\widetilde{t}_1)\| + \|\gamma(\widetilde{t}_1) - \gamma(t_1)\| \\
&\leq \frac{s_1 \cos\theta_1}{5} + |\widetilde{t}_1 - t_1| \cdot \|u^0\| \\
&\leq \frac{s_1 \cos\theta_1}{5} + \frac{4s_1}{5} \leq s_1.
\end{aligned} \tag{5.130}$$

The claim is proved. ■

Step 2. We will now show that, under the hypotheses of the lemma, $\|\widetilde{p}^1 - p^1\| \leq s_1$.

Let γ and $\widetilde{\gamma}$ be as in Claim 5.7.2 above. By Taylor expansion,

$$\widetilde{\gamma}(t) = \widetilde{p}^0 + t\widetilde{u}^0 + E(t), \qquad \text{where } \|E(t)\| \leq \frac{t^2}{2} \sup_{u\in[0,t]} \|\gamma''(u)\|. \tag{5.131}$$

To compute $\widetilde{\gamma}''(t)$, write $\widetilde{y}^0 = (\widetilde{x}_1^0, \widetilde{x}_2^0, \widetilde{\alpha}^0)$ and $\widetilde{w}^0 = (\widetilde{v}_1^0, \widetilde{v}_2^0, \widetilde{\omega}^0)$, and note that by (5.86)

$$\widetilde{\gamma}'(t) = \mathrm{d}(G_\epsilon)_{\widetilde{y}^0 + t\widetilde{w}^0}(\widetilde{w}^0) = \begin{pmatrix} \widetilde{v}_1^0 + \widetilde{\omega}^0 \cos(\epsilon(\widetilde{\alpha}^0 + t\widetilde{\omega}^0)) \\ \widetilde{v}_2^0 + \widetilde{\omega}^0 \sin(\epsilon(\widetilde{\alpha}^0 + t\widetilde{\omega}^0)) \end{pmatrix}, \tag{5.132}$$

and therefore

$$\widetilde{\gamma}''(t) = \begin{pmatrix} -\epsilon(\widetilde{\omega}^0)^2 \sin(\epsilon(\widetilde{\alpha}^0 + t\widetilde{\omega}^0)) \\ \epsilon(\widetilde{\omega}^0)^2 \cos(\epsilon(\widetilde{\alpha}^0 + t\widetilde{\omega}^0)) \end{pmatrix}. \tag{5.133}$$

Thus $\|\widetilde{\gamma}''(t)\| = \epsilon(\widetilde{\omega}^0)^2$ and $\|E(t)\| \leq \epsilon t^2(\widetilde{\omega}^0)^2/2$. This, together with (5.120) and (5.131), gives us

$$\|\widetilde{\gamma}(t) - \gamma(t)\| \leq \|\widetilde{p}^0 - p^0\| + \|\widetilde{u}^0 - u^0\|t + \frac{\epsilon t^2(\widetilde{\omega}^0)^2}{2}. \tag{5.134}$$

Using (5.86) and Lemma 5.6, we compute

$$\begin{aligned} \|\widetilde{u}^0 - u^0\| &= \|(\mathrm{d}G_0)_{y^0}(\widetilde{w}^0) - (\mathrm{d}G_\epsilon)_{\widetilde{y}^0}(w^0)\| \\ &\leq \|(\mathrm{d}G_0)_{y^0}(w^0) - (\mathrm{d}G_\epsilon)_{\widetilde{y}^0}(w^0)\| + \|(\mathrm{d}G_\epsilon)_{\widetilde{y}^0}(\widetilde{w}^0 - w^0)\| \\ &\leq |\omega^0|\sqrt{(1-\cos(\epsilon\widetilde{\alpha}^0))^2 + \sin^2(\epsilon\widetilde{\alpha}^0)} \\ &+ (1+mJ^{-1})^{1/2}(1 + ((mJ)^{-1/2}+1)|\epsilon\widetilde{\alpha}^0|)\|\widetilde{w}^0 - w^0\| \\ &\leq |\omega^0||\epsilon\widetilde{\alpha}^0| + (1+mJ^{-1})^{1/2}(1 + ((mJ)^{-1/2}+1)|\epsilon\widetilde{\alpha}^0|)\|\widetilde{w}^0 - w^0\|. \end{aligned} \tag{5.135}$$

Using the hypotheses, we have

$$|\epsilon\widetilde{\alpha}^0| \le \epsilon(|\alpha^0| + |\widetilde{\alpha}^0 - \alpha^0|) \le \frac{\rho(\epsilon)}{2} - \delta_0 - \epsilon(\frac{L}{\sin\nu} + 1) + \frac{\epsilon s_2}{3} \le \frac{\rho(\epsilon)}{2} \le 1. \tag{5.136}$$

and $|\omega^0| \le \|w^0\| = 1$. Hence from (5.135) we obtain

$$\|\widetilde{u}^0 - u^0\| \le \frac{\rho(\epsilon)}{2} + (1 + mJ^{-1})^{1/2}(2 + (mJ)^{-1/2})\|\widetilde{w}^0 - w^0\|. \tag{5.137}$$

Using the same notation as in the proof of Claim 5.7.2, we let

$$t^+ = t_1 + \frac{4s_1}{5\|u^0\|} = \frac{L + 4s_1/5}{\|u^0\|}. \tag{5.138}$$

By Lemma 5.6,

$$\|u^0\| = \|\mathrm{d}G_0(w^0)\| \ge \sin\nu, \tag{5.139}$$

and hence

$$t^+ \le \frac{L + 4s_1/5}{\sin\nu} \le \frac{L+1}{\sin\nu} \tag{5.140}$$

(note that $s_1 \le 5/4$ by the given bounds (5.103)). Substituting (5.137) into (5.134) gives us, for $0 \le t \le t^+$,

$$\begin{aligned} \|\widetilde{\gamma}(t) - \gamma(t)\| \le \|\widetilde{p}^0 - p^0\| + \frac{\rho(\epsilon)}{2}t^+ \\ + (1 + mJ^{-1})^{1/2}(2 + (mJ)^{-1/2})\|\widetilde{w}^0 - w^0\|t^+ + \frac{\epsilon(t^+)^2}{2}. \end{aligned} \tag{5.141}$$

By the bounds in the hypothesis (5.106) and the bound (5.140) on t^+, we obtain

$$\begin{aligned}
\|\widetilde{\gamma}(t) - \gamma(t)\| \leq \frac{s_1 \cos\theta_1}{20} &+ \frac{1}{2}\left(\frac{s_1 \cos\theta_1 \sin\nu}{10(L+1)}\right)\left(\frac{L+1}{\sin\nu}\right) \\
&+ (1 + mJ^{-1})^{1/2}(2 + (mJ)^{-1/2}) \\
&\times \left(\frac{s_1 \cos\theta_1 \sin\nu}{20(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(L+1)}\right)\left(\frac{L+1}{\sin\nu}\right) \\
&+ \frac{1}{2}\left(\frac{s_1 \cos\theta_1 \sin^2\nu}{10(L+1)^2}\right)\left(\frac{L+1}{\sin\nu}\right)^2 \\
&= \frac{s_1 \cos\theta_1}{5}.
\end{aligned} \tag{5.142}$$

By Claim 5.7.2, this implies $\|\widetilde{p}^1 - p^1\| \leq s_1$.

Step 3. To bound the difference in the angle coordinates $|\widetilde{\alpha}^1 - \alpha^1|$, we proceed as follows. Let $t_1, \widetilde{t}_1 > 0$ be the times defined above at which the trajectory hits the boundary in the cylindrical configuration space and the true configuration space respectively. As we argued above $\|\widetilde{\gamma}(t) - \gamma(t)\| \leq s_1 \cos\theta_1/5$ for all $0 \leq t \leq (L + 4s_1/5)/\|u^0\|$. By Claim 5.7.2, this implies that $t_1 - 4s_1/5\|u^0\| \leq \widetilde{t}_1 \leq t_1 + 4s_1/5\|u^0\|$. We want to bound $|\widetilde{\alpha}(\widetilde{t}_1) - \alpha(t_1)|$. Since $\widetilde{\alpha}(t) = \widetilde{\alpha}^0 + t\widetilde{\omega}^0$ is linear, the function $t \mapsto \widetilde{\alpha}(t) - \alpha(t_1)$ is either increasing or decreasing. Consequently, by the bounds on $\widetilde{t}_1$ here described,

$$|\widetilde{\alpha}(\widetilde{t}_1) - \alpha(t_1)| \leq \max\left\{\left|\widetilde{\alpha}\left(t_1 + \frac{4s_1}{5\|u^0\|}\right) - \alpha(t_1)\right|, \left|\widetilde{\alpha}\left(t_1 - \frac{4s_1}{5\|u^0\|}\right) - \alpha(t_1)\right|\right\}. \tag{5.143}$$

Recall that $t_1 = L/\|u^0\|$. Substituting $\alpha(t) = \alpha^0 + t\omega^0$ and $\widetilde{\alpha}(t) = \widetilde{\alpha}^0 + t\widetilde{\omega^0}$ into the

above and applying the triangle inequality, we obtain

$$\begin{aligned} |\widetilde{\alpha}(\widetilde{t_1}) - \alpha(t_1)| &\le |\widetilde{\alpha}^0 - \alpha^0| + |\widetilde{\omega}^0 - \omega^0| \left(\frac{L}{\|u^0\|} \right) + |\widetilde{\omega}^0| \left(\frac{4s_1}{5\|u^0\|} \right) \\ &\le |\widetilde{\alpha}^0 - \alpha^0| + \|\widetilde{w}^0 - w^0\| \left(\frac{L}{\sin \nu} \right) + \left(\frac{4s_1}{5 \sin \nu} \right), \end{aligned} \tag{5.144}$$

using (5.139) in the last line. Substituting the hypothesized bounds (5.106) for the quantities above, we obtain

$$|\widetilde{\alpha}^1 - \alpha^1| \le \frac{s_2}{3} + \frac{s_2 \sin \nu}{3L} \left(\frac{L}{\sin \nu} \right) + \frac{4}{5 \sin \nu} \left(\frac{5 s_2 \sin \nu}{12} \right) = s_2. \tag{5.145}$$

Step 4. Finally, we bound the difference in velocities $\|\widetilde{w}^1 - w^1\|$. Let $\widetilde{n}^1$ be the inward-pointing unit normal at $\widetilde{y}^1 \in \partial \mathcal{M}^*$ and let n^1 be the unit normal at $y^1 \in \partial \mathcal{M}^*_{\text{cyl}}$. Then by specular reflection

$$\widetilde{w}^1 = \widetilde{w}^0 - 2\langle \widetilde{w}^0, \widetilde{n}^1 \rangle \widetilde{n}^1 \quad \text{and} \quad w^1 = w^0 - 2\langle w^0, n^1 \rangle n^1. \tag{5.146}$$

Therefore, writing

$$\langle \widetilde{w}^0, \widetilde{n}^1 \rangle \widetilde{n}^1 - \langle w^0, n^1 \rangle n^1 = \langle \widetilde{w}^0 - w^0, \widetilde{n}^1 \rangle n^1 + \langle w^0, \widetilde{n^1} - n^1 \rangle \widetilde{n}^1 + \langle w^0, n^1 \rangle (\widetilde{n}^1 - n^1), \tag{5.147}$$

we obtain by (5.146) and Cauchy-Schwarz

$$\|\widetilde{w}^1 - w^1\| \le 3\|\widetilde{w}^0 - w^0\| + 4\|\widetilde{n}^1 - n^1\|. \tag{5.148}$$

To bound the difference in the unit normal vectors, first we obtain an expression

for $\widetilde{n}^1$ in terms of n^1. Suppose once again that $p(\sigma)$, $\sigma \in (a,b)$ is a unit speed parametrization of the curve segment I_1 with $p(0) = p^1$. Note that by Step 2, $\widetilde{p}^1 \in I_1$, and we may assume without loss of generality that $p(c) = \widetilde{p}^1$ for some $c \in (0,b)$.

We obtain a parametrization of a neighborhood of $\widetilde{y}^1$ in $\partial\mathcal{M}^*$ as follows. First, from the hypothesis and the conclusion of Step 3, we have

$$\begin{aligned}
|\widetilde{\alpha}^1| &\leq |\alpha^0| + |\alpha^1 - \alpha^0| + |\widetilde{\alpha}^1 - \alpha^1| \\
&< \frac{\epsilon^{-1}\rho}{2} - \epsilon^{-1}\delta_0 - (\frac{L}{\sin\nu} + 1) + |t_1||\omega^0| + s_2 \\
&\leq \frac{\epsilon^{-1}\rho}{2} - \epsilon^{-1}\delta_0 - (\frac{L}{\sin\nu} + 1) + \frac{L}{||u^0||} + s_2 \\
&\leq \frac{\epsilon^{-1}\rho}{2} - \epsilon^{-1}\delta_0,
\end{aligned} \tag{5.149}$$

using $t_1 = L/||u^0||$ and $|\omega^0| \leq 1$ to obtain the second to last line, and (5.139) for the last line. Consequently, $\widetilde{y}^1 \in \Gamma^*_{\text{roll}}$. Recall the map f^* defined by (5.74) which parametrizes the surface $\partial\mathcal{M}^*_{\text{roll}}$, and observe that $f^*(\widetilde{p}^1, \widetilde{\alpha}^1) = \widetilde{y}^1$. Thus, the map

$$Q(\sigma,\alpha) = f(p(\sigma),\alpha), \quad (\sigma,\alpha) \in (a,b) \times (-\frac{\epsilon^{-1}\rho}{2} + \epsilon^{-1}\delta_0, \frac{\epsilon^{-1}\rho}{2} - \epsilon^{-1}\delta_0), \tag{5.150}$$

parametrizes a neighborhood of $\widetilde{y}^1$ in $\partial\mathcal{M}^*$. Explicitly,

$$Q(\sigma,\alpha) = p(\sigma) + \begin{pmatrix} \epsilon^{-1}\sin(\epsilon\alpha) \\ 1 + \epsilon^{-1}(-1 + \cos(\epsilon\alpha)) \\ \alpha \end{pmatrix}, \tag{5.151}$$

Observe that $\widetilde{y}^1 = Q(c, \widetilde{\alpha}^1)$. The tangent space of $\partial\mathcal{M}^*$ at $\widetilde{y}^1$ is therefore spanned by

the following unit vectors

$$\partial_\sigma Q(c,\widetilde{\alpha}^1) = p'(c), \quad (m+J)^{-1/2}\partial_\alpha Q(c,\widetilde{\alpha}^1) = (m+J)^{-1/2}\begin{pmatrix}\cos(\epsilon\widetilde{\alpha}^1)\\ \sin(\epsilon\widetilde{\alpha}^1)\\ 1\end{pmatrix} =: \widetilde{\chi}. \tag{5.152}$$

(These are linearly independent since the α-component of $\widetilde{\chi}$ is nonzero.) Let

$$\widehat{\chi} = \frac{\widetilde{\chi} - \langle p'(c), \widetilde{\chi}\rangle p'(c)}{\|\widetilde{\chi} - \langle p'(c), \widetilde{\chi}\rangle p'(c)\|}. \tag{5.153}$$

Then $(p'(c), \widehat{\chi})$ is an orthonormal basis for the tangent space to Γ_{roll} at $\widetilde{y}^1$. Consequently, the unit normal at $\widetilde{y}^1$ is given by the formula

$$\begin{aligned}\widetilde{n}^1 &= \frac{n^1 - \langle n^1, p'(c)\rangle p'(c) - \langle n^1, \widehat{\chi}\rangle\widehat{\chi}}{\|n^1 - \langle n^1, p'(c)\rangle p'(c) - \langle n^1, \widehat{\chi}\rangle\widehat{\chi}\|}\\ &= \frac{n^1 - \langle n^1, p'(c)\rangle p'(c) - \langle n^1, \widehat{\chi}\rangle\widehat{\chi}}{\sqrt{1 - \langle n^1, p'(c)\rangle^2 - \langle n^1, \widehat{\chi}\rangle^2}}.\end{aligned} \tag{5.154}$$

Therefore,

$$\langle n^1, \widetilde{n}^1\rangle = \frac{1 - \langle n^1, p'(c)\rangle^2 - \langle n^1, \widehat{\chi}\rangle^2}{\sqrt{1 - \langle n^1, p'(c)\rangle^2 - \langle n^1, \widehat{\chi}\rangle^2}} = \sqrt{1 - \langle n^1, p'(c)\rangle^2 - \langle n^1, \widehat{\chi}\rangle^2}. \tag{5.155}$$

It is even easier to specify a frame at $y^1 \in \partial\mathcal{M}^*_{\text{cyl}}$. Because of the cylindrical structure of $\partial\mathcal{M}^*_{\text{cyl}}$, both $p'(0)$ and $\chi = (m+J)^{-1/2}(-1,0,1)$ are tangent to $\partial\mathcal{M}^*_{\text{cyl}}$ at y^1. Hence $\langle p'(0), n^1\rangle = \langle \chi, n^1\rangle = 0$, and

$$|\langle n^1, p'(c)\rangle| = |\langle n^1, p'(c) - p'(0)\rangle| \le \|p'(c) - p'(0)\|. \tag{5.156}$$

To bound the quantity on the right, first note that

$$|\langle p(c) - p(0), \xi\rangle| \leq \|p(c) - p(0)\| = \|\widetilde{p}^1 - p^1\| \leq s_1 < \min\{r_0, \frac{1}{2}(\overline{\kappa})^{-1}\}, \tag{5.157}$$

by (5.103). Thus, by Claim 5.7.1, $\sigma \mapsto \|p(\sigma) - p(0)\|$ is monotone increasing on $[0, c]$. Consequently, for all $\sigma \in [0, c]$,

$$s_1 \geq \|p(\sigma) - p(0)\| \geq |\sigma| - \frac{\overline{\kappa}}{2}|\sigma|^2, \tag{5.158}$$

using Taylor's theorem for the last estimate. Since the function $f(s) = s^2 - \frac{\overline{\kappa}}{2}s^2$ is monotone increasing on $[0, (\overline{\kappa})^{-1}]$ and achieves a maximum of $\frac{(\overline{\kappa})^{-1}}{2} > s_1$ at $s = (\overline{\kappa})^{-1}$, it follows that $c < (\overline{\kappa})^{-1}$. Thus by (5.158)

$$s_1 \geq |c| - \frac{\overline{\kappa}}{2}|c|(\overline{\kappa})^{-1} = \frac{1}{2}|c|. \tag{5.159}$$

It follows that

$$\|p'(c) - p(0)\| \leq \int_0^c \|p''(u)\| \mathrm{d}u \leq \overline{\kappa}|c| \leq 2\overline{\kappa}s_1. \tag{5.160}$$

Substitution into (5.156) yields

$$|\langle n^1, p'(c)\rangle| \leq 2\overline{\kappa}s_1. \tag{5.161}$$

In addition, we have

$$
\begin{aligned}
|\langle n^1, \widetilde{\chi}\rangle| &= |\langle n^1, \widetilde{\chi} - \chi\rangle| \leq \|\widetilde{\chi} - \chi\| \\
&= \sqrt{(1 - \cos(\epsilon\widetilde{\alpha}^1))^2 + \sin^2(\epsilon\widetilde{\alpha}^1)} \\
&= \sqrt{2 - 2\cos(\epsilon\widetilde{\alpha}^1)} \leq |\epsilon\widetilde{\alpha}^1|.
\end{aligned}
\tag{5.162}
$$

Therefore, by (5.153)

$$
\begin{aligned}
|\langle n^1, \widetilde{\chi}\rangle| &= \frac{|\langle n^1, \widetilde{\chi}\rangle - \langle p'(c), \widetilde{\chi}\rangle\langle n^1, p'(c)\rangle|}{\sqrt{1 - \langle p'(c), \widetilde{\chi}\rangle^2}} \\
&\leq \frac{|\epsilon\widetilde{\alpha}^1| + 2\overline{\kappa}s_1}{\sqrt{1 - 4(\overline{\kappa})^2 s_1^2}} \leq \sqrt{2}(|\epsilon\widetilde{\alpha}^1| + 2\overline{\kappa}s_1),
\end{aligned}
\tag{5.163}
$$

where the last line follows by the assumption that $s_1 \leq 1/(2\sqrt{2}\overline{\kappa})$ by (5.103). By (5.155), (5.161), and (5.163), we conclude that

$$
\begin{aligned}
\langle \widetilde{n}^1, n^1\rangle^2 &\geq 1 - 4(\overline{\kappa})^2 s_1^2 - 2(|\epsilon\widetilde{\alpha}^1| + 2\overline{\kappa}s_1)^2 \\
&\geq 1 - 20(\overline{\kappa})^2 s_1^2 - 4\epsilon^2|\widetilde{\alpha}^1|^2 \\
&\geq 1 - 20(\overline{\kappa})^2 s_1^2 - \rho(\epsilon)^2,
\end{aligned}
\tag{5.164}
$$

noting that $|\widetilde{\alpha}^1| \leq \epsilon^{-1}\rho/2$ by (5.149). From this we obtain the estimate

$$
\begin{aligned}
\|\widetilde{n}^1 - n^1\| &= \sqrt{2 - 2\langle \widetilde{n}^1, n^1\rangle} \\
&\leq \sqrt{40(\overline{\kappa})^2 s_1^2 + 2\rho(\epsilon)^2} \\
&\leq 2\sqrt{10}\overline{\kappa}s_1 + \sqrt{2}\rho(\epsilon).
\end{aligned}
\tag{5.165}
$$

From (5.148) we obtain the following bound on the difference of velocities

$$\|\widetilde{w}^1 - w^1\| \le 3\|\widetilde{w}^0 - w^0\| + 8\sqrt{10}\overline{\kappa}s_1 + 4\sqrt{2}\rho(\epsilon). \tag{5.166}$$

Substituting the bounds (5.103), (5.104), and (5.106) for the quantities above, we obtain

$$\|\widetilde{w}^1 - w^1\| \le 3(\frac{s_3}{12}) + 8\sqrt{10}\overline{\kappa}(\frac{s_3}{24\sqrt{10}\overline{\kappa}}) + 4\sqrt{2}(\frac{s_3}{12\sqrt{2}}) = s_3. \tag{5.167}$$

This completes the proof. □

Proof of Lemma 5.8. This lemma is corollary of the previous one. Take

$$\begin{aligned} s_1 &= \left(1 + \frac{12}{5\sin\nu} + 24\sqrt{10}\overline{\kappa}\right)^{-1} s, \\ s_2 &= \frac{12s_1}{5\sin\nu}, \\ s_3 &= 24\sqrt{10}\overline{\kappa}s_1. \end{aligned} \tag{5.168}$$

Since by the above and (5.120) and the assumption $s \le 1$, we have $s_2 \le 1$ and $s_3 \le 1$. Since $0 < \sin\nu \le 1$ and $\overline{\kappa} \ge 1$, we have

$$79 \le 1 + \frac{12}{5\sin\nu} + 24\sqrt{10}\overline{\kappa} \le \frac{80\overline{\kappa}}{\sin\nu}. \tag{5.169}$$

Thus by (5.108),

$$s_1 \le 79^{-1}s \le \min\{r_0, (\overline{\kappa})^{-1}\cos\theta_1\} \le \min\left\{r_0, \frac{5\cos\theta_1}{2\overline{\kappa}}\right\}. \tag{5.170}$$

Also, by definition of s_2 and s_3 above, we have

$$s_1 = \frac{5s_2 \sin\nu}{12} = \frac{s_3}{24\sqrt{10}\overline{\kappa}} \tag{5.171}$$

Therefore the bound (5.103) is satisfied.

To see that (5.104) is satisfied, first observe that

$$\begin{aligned} \frac{s\cos\theta_1 \sin^2\nu}{800\overline{\kappa}(\overline{L}+1)} &= \left(1 + \frac{12}{5\sin\nu} + 24\sqrt{10}\overline{\kappa}\right)\frac{s_1\cos\theta_1\sin^2\nu}{800\overline{\kappa}(\overline{L}+1)} \\ &\leq \frac{s_1\cos\theta_1\sin\nu}{10(\overline{L}+1)} \leq \frac{s_1\cos\theta_1\sin\nu}{10(L+1)}, \end{aligned} \tag{5.172}$$

where the first inequality above follows from (5.169). Consequently, since $\breve{\rho}^{-1}$ is increasing, by (5.109) we have

$$\epsilon \leq \breve{\rho}^{-1}\left(\frac{s\cos\theta_1\sin^2\nu}{800\overline{\kappa}(\overline{L}+1)}\right) \leq \breve{\rho}^{-1}\left(\frac{s_1\cos\theta_1\sin\nu}{10(L+1)}\right). \tag{5.173}$$

By (5.109) and the estimate (5.172), we also have

$$\epsilon \leq \frac{s\cos\theta_1\sin^3\nu}{800\overline{\kappa}(\overline{L}+1)^2} \leq \frac{s_1\cos\theta_1\sin^2\nu}{10(\overline{L}+1)^2} \leq \frac{s_1\cos\theta_1\sin^2\nu}{10(L+1)^2}. \tag{5.174}$$

It also follows from (5.174) and the definition of s_2 that

$$\epsilon \leq s_1 = \frac{5s_2\sin\nu}{12}, \tag{5.175}$$

and from (5.173) and the definition of s_3 that

$$\epsilon \le \breve{\rho}^{-1}(s_1) \le \breve{\rho}^{-1}\left(2\sqrt{5}s_1\right) = \breve{\rho}^{-1}\left(\frac{s_3}{12\sqrt{2}}\right). \tag{5.176}$$

Combining (5.173), (5.174), (5.175), and (5.176), we obtain the desired bound (5.104) on ϵ.

To see that (5.105) is satisfied, observe that by (5.110),

$$|\alpha^0| \le \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) - \left(\frac{\overline{L}}{\sin\nu} + 1\right) \le \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) - \left(\frac{L}{\sin\nu} + 1\right). \tag{5.177}$$

From (5.111) we get the following bounds on $\|\widetilde{p}^0 - p^0\|$, $|\widetilde{\alpha}^0 - \alpha^0|$, and $\|\widetilde{w}^0 - w^0\|$. First, by definition of s_1 and (5.111), we have

$$\begin{aligned}
&\|\widetilde{p}^0 - p^0\| + |\widetilde{\alpha}^0 - \alpha^0| + \|\widetilde{w}^0 - w^0\| \\
&\qquad \le \left(1 + \frac{12}{5\sin\nu} + 24\sqrt{10}\overline{\kappa}\right) \\
&\qquad\quad \times \frac{s_1\cos\theta_1\sin^2\nu}{1600(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})\overline{\kappa}(\overline{L}+1)} \\
&\qquad \le \left(\frac{80\overline{\kappa}}{\sin\nu}\right)\frac{s_1\cos\theta_1\sin^2\nu}{1600(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})\overline{\kappa}(\overline{L}+1)} \\
&\qquad = \frac{s_1\cos\theta_1\sin\nu}{20(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(\overline{L}+1)}.
\end{aligned} \tag{5.178}$$

It follows from the above that

$$\|\widetilde{p}^0 - p^0\| \le \frac{s_1\cos\theta_1}{20}. \tag{5.179}$$

Since $s_1 = \frac{5s_2 \sin\nu}{12}$, we also obtain from (5.178)

$$|\widetilde{\alpha}^0 - \alpha^0| \leq \frac{s_2 \cos\theta_1 \sin^2\nu}{48(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(\overline{L}+1)} \leq \frac{s_2}{3}. \tag{5.180}$$

In addition, (5.178) implies

$$\|\widetilde{w}^0 - w^0\| \leq \frac{s_1 \cos\theta_1 \sin\nu}{20(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(L+1)}. \tag{5.181}$$

Substituting $s_1 = \frac{5s_2 \sin\nu}{12}$ into (5.178) gives us

$$\|\widetilde{w}^0 - w^0\| \leq \frac{s_2 \cos\theta_1 \sin^2\nu}{48(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(\overline{L}+1)} \leq \frac{s_2 \sin\nu}{3L}, \tag{5.182}$$

and substituting $s_1 = \frac{s_3}{24\sqrt{10}\overline{\kappa}}$ gives us

$$\|\widetilde{w}^0 - w^0\| \leq \frac{s_3 \cos\theta_1 \sin\nu}{48\sqrt{10}\overline{\kappa}(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(\overline{L}+1)} \leq \frac{s_3}{12}. \tag{5.183}$$

We conclude from the last five displays that the bounds (5.106) are satisfied. Therefore, the hypotheses of Lemma 5.7 hold, and we conclude that

$$\begin{aligned} \|\widetilde{p}^1 - p^1\| + |\widetilde{\alpha}^1 - \alpha^1| + \|\widetilde{w}^1 - w^1\| &\leq s_1 + s_2 + s_3 \\ &= \left(1 + \frac{12}{5\sin\nu} + 24\sqrt{10}\overline{\kappa}\right) s_1 = s, \end{aligned} \tag{5.184}$$

here using the definitions (5.168). □

5.3.5 Definition of Ω and proof of Lemma 5.3

Recall the sets $\mathcal{N}^*$ and A_1^*, A_2^*, A_3^* defined by (5.97). The set A_2^* has Lebesgue measure zero. (See the proof of Proposition 4.11.)

Recall that $K_{\text{cyl}}^{\Sigma,\epsilon}$ is a well-defined C^1 involutive diffeomorphism on a Λ^2-full measure open subset $\mathcal{F}_{\text{cyl}} \subset \mathbf{P} \times \mathbb{S}_+^2$. For any $(y, w) \in \mathcal{F}_{\text{cyl}}$, the billiard trajectory in $\mathcal{M}_{\text{cyl}}$ starting from $(y, -w)$ is well-defined for all time and hits $\partial\mathcal{M}_{\text{cyl}}$ only finitely many times before returning to $\mathbf{P}$. (See Propositions 4.11 and 6.5.)

Let $\mathcal{F}_{\text{cyl}}^* = \sigma_{\epsilon^{-1}}(\mathcal{F}_{\text{cyl}}) = \{(y, w) \in \mathbf{P} \times \mathbb{S}_+^2 : (\epsilon y, w) \in \mathcal{F}_{\text{cyl}}\}$. For any $(y, w) \in \mathcal{F}_{\text{cyl}}^*$, the billiard trajectory in $\mathcal{M}_{\text{cyl}}^*$ starting from initial state $(y, -w)$ is defined for all time and returns to $\mathbf{P}$ after only finitely many collisions with $\partial\mathcal{M}_{\text{cyl}}^*$.

Fix $(y, w) \in (A_1^* \times \mathbb{S}^2) \cap \mathcal{F}_{\text{cyl}}^*$, and let $(y^0, w^0) = (y, -w)$. Let N denote the number of times which the point particle hits $\partial\mathcal{M}_{\text{cyl}}^*$ before returning to the plane $\mathbf{P}$. For $1 \le j \le N$, let y^j denote the point in $\partial\mathcal{M}_{\text{cyl}}^*$ where the billiard trajectory starting from (y^{j-1}, w^{j-1}) first returns to the boundary, and let w^j denote the velocity of the point particle immediately after reflecting from the boundary at y^j. Let (y^{N+1}, w^{N+1}) denote the state of the point particle upon returning to the plane $\mathbf{P}$ (thus $w^{N+1} = w^N$).

We will also use the following notation: For $0 \le j \le N + 1$

- $(p^j, u^j) = \mathrm{d}G_0(y^j, w^j)$.
- k^j is the unit normal vector to $\partial\widehat{B}^*$ at p^j if $1 \le j \le N$. We also let $k^0 = k^{N+1} = -e_2$. (Thus each k^j is the inward-pointing unit normal at $p^j \in \partial\mathcal{D}^*$, where $\mathcal{D}^*$ is the region defined by (5.98)).
- θ_j is the angle between k^j and $-u^{j-1}$ (i.e. $\cos\theta_j = \langle k^j, -u^{j-1}\rangle$) if $1 \le j \le N + 1$.

We let θ_0 be the angle between w^0 and $k^0 = -e_2$.

- $r_j = r_{\mathcal{D}^*}(p_j, \widehat{u}_j)$ is the radius of transversality associated with the chord from p_j to p_{j+1} if $0 \le j \le N$.

- ν is the angle between w^0 and the line spanned by χ.

Recall that in the cylindrical configuration space the projection of the point particle velocity onto χ is preserved. It follows that ν is the angle between w^j and the line spanned by χ for $0 \le j \le N+1$.

In the notation above, we let $F = F(\epsilon)$ denote the set of all states states $(y, w) \in (A_1 \times \mathbb{S}^2) \cap \mathcal{F}^*_{\text{cyl}}$ such that

F1. $N \le \frac{\log(1/\rho)}{4 \log\log(1/\rho)}$,

F2. $r_j \ge \frac{1}{\log(1/\rho)}$ for $0 \le j \le N$,

F3. $\cos\theta_j \ge \frac{\overline{\kappa}}{\log(1/\rho)}$ for $0 \le j \le N+1$, and

F4. $\sin\nu \ge \frac{1}{\log(1/\rho)}$.

We also define $\Xi = \Xi(\epsilon)$ to be the set of all states $(y, w) \in (A_1 \times \mathbb{S}^2) \cap \mathcal{F}^*_{\text{cyl}}$ such that if α is the angular coordinate of y, then the following holds:

$$\begin{aligned} &\text{for all } j \in \mathbb{Z}, \text{ if } \alpha \in \left(\epsilon^{-1}\Big(-\frac{\rho}{2} + j\rho\Big), \epsilon^{-1}\Big(\frac{\rho}{2} + j\rho\Big)\right], \\ &\qquad \text{then } |\alpha - j\epsilon^{-1}\rho| \le \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0 - \frac{12(\overline{L}+1)}{5\log(1/\rho)^2}\right) - \frac{\log(1/\rho)^2 \overline{L}}{4\log\log(1/\rho)}. \end{aligned} \tag{5.185}$$

Recall the parallelogram R_ϵ, defined by (5.29). Let

$$R_\epsilon^* = \epsilon^{-1} R_\epsilon = \{(x_1, \alpha) \in \mathbf{P} : 0 \le x_1 + \alpha \le 1 \text{ and } -\epsilon^{-1}\rho(\epsilon)/2 \le \alpha \le \epsilon^{-1}\rho(\epsilon)/2\}. \quad (5.186)$$

Also define translation maps

$$\tau_{jk}^*(y) = \epsilon^{-1}\tau_{jk}(\epsilon y) = y + je_1 + k\epsilon^{-1}\rho e_3, \qquad y \in \mathbf{P}, \quad (5.187)$$

$$\overline{\tau}_{jk}^*(y, w) = (\tau_{jk}^*(y), w), \qquad (y, w) \in \mathbf{P} \times \mathbb{S}_+^2. \quad (5.188)$$

The translates $\tau_{jk}^* R_\epsilon^*$ tessellate the plane $\mathbf{P}$, in the sense that $\mathbf{P} = \bigcup_{(j,k)\in\mathbb{Z}^2} \tau_{jk}^* R_\epsilon^*$, and $\tau_{jk}^* R_\epsilon^* \cap \tau_{j'k'}^* R_\epsilon^*$ has Lebesgue measure zero whenever $(j, k) \ne (j', k')$.

We let $\widehat{R}_\epsilon^*$ denote the parallelogram R_ϵ^* minus its upper and right boundary segments, i.e.

$$\widehat{R}_\epsilon^* = \{(x_1, \alpha) \in \mathbf{P} : 0 \le x_1 + \alpha < 1 \text{ and } -\epsilon^{-1}\rho(\epsilon) \le \alpha < \epsilon^{-1}\rho(\epsilon)/2\}. \quad (5.189)$$

Then the translates $\tau_{jk}^* \widehat{R}_\epsilon^*$ form a collection of disjoint sets which tessellate the plane $\mathbf{P}$.

We define Ω^* to be the subset of $\mathbf{P} \times \mathbb{S}_+^2$ which is invariant under the translations $\overline{\tau}_{jk}^*$, $(j, k) \in \mathbb{Z}^2$ and such that

$$\Omega^* \cap (\widehat{R}_\epsilon^* \times \mathbb{S}_+^2) = \big(F \cup \Xi \cup (A_3^* \times \mathbb{S}_+^2)\big) \cap (\widehat{R}_\epsilon^* \times \mathbb{S}_+^2). \quad (5.190)$$

We then define $\Omega \subset \mathbf{P} \times \mathbb{S}^2$ by

$$\Omega = \sigma_\epsilon(\Omega^*) = \{(y, w) \in \mathbf{P} \times \mathbb{S}^2_+ : (\epsilon^{-1}y, w) \in \Omega^*\}. \tag{5.191}$$

Proof of Lemma 5.3. (i) follows immediately from the definition of Ω.

(ii) Changing to zoomed coordinates, it is enough to show that $\epsilon\rho(\epsilon)^{-1}\Lambda^2(\mathbf{P} \times \mathbb{S}^2 \setminus \Omega^*) \to 0$ as $\epsilon \to 0$. Note that $A_1^* \cup A_3^*$ is a full measure subset of $\mathbf{P}$. Consequently, it is enough to show that $\epsilon\rho(\epsilon)^{-1}\Lambda^2((R_\epsilon^* \cap A_1^*) \times \mathbb{S}^2 \setminus F) \to 0$ and $\epsilon\rho(\epsilon)^{-1}\Lambda^2((R_\epsilon^* \cap A_1^*) \times \mathbb{S}^2 \setminus \Xi) \to 0$ as $\epsilon \to 0$.

First we deal with F. We let $k = \lceil \epsilon^{-1}\rho(\epsilon)/2 \rceil$, and we chop R_ϵ into $2k$ parts

$$R_\epsilon^{*j} = R_\epsilon^* \cap \left\{(x_1, \alpha) : \frac{(j-1)\epsilon^{-1}\rho}{2k} \le \alpha \le \frac{j\epsilon^{-1}\rho}{2k}\right\}, \qquad -k+1 \le j \le k. \tag{5.192}$$

By definition of the parallelogram R_ϵ^*, we see that

$$R_\epsilon^{*j} = R_\epsilon^{*0} + \frac{(1 + mJ^{-1})^{-1/2}\epsilon^{-1}\rho j}{k}\chi. \tag{5.193}$$

By observing that the sets F and A_1^* are invariant under any translation in the direction χ, it follows that

$$\Lambda^2((R_\epsilon^{*j} \cap A_1^*) \times \mathbb{S}^2_+ \setminus F) = \Lambda^2((R_\epsilon^{*0} \cap A_1^*) \times \mathbb{S}^2_+ \setminus F), \quad \text{for } -k+1 \le j \le k. \tag{5.194}$$

Therefore,

$$\Lambda^2((R_\epsilon^{*0} \cap A_1^*) \times \mathbb{S}^2_+ \setminus F) = (2k)^{-1}\Lambda^2((R_\epsilon^* \cap A_1^*) \times \mathbb{S}^2_+ \setminus F). \tag{5.195}$$

Noting that $(2k)^{-1} \sim \frac{\epsilon}{\rho}$, we see it is enough to show that $\Lambda^2((R_\epsilon^{*0} \cap A_1^*) \times \mathbb{S}_+^2 \setminus F) \to 0$ as $\epsilon \to 0$. To this end, note that $R_\epsilon^{*0} \times \mathbb{S}_+^2$ lies in the fixed bounded set $\{(x_1, \alpha) : 0 \le x_1 \le 1, -1 \le x_1 + \alpha \le 0\} \times \mathbb{S}_+^2$ which has finite Λ^2-measure. Also notice that $F(\epsilon) \supset F(\epsilon')$ whenever $\epsilon < \epsilon'$. Consequently, by continuity from above, it is enough to show that $\bigcup_{\epsilon>0} F(\epsilon)$ has full Λ^2 measure in $A_1^* \times \mathbb{S}_+^2$. Now $\bigcup_{\epsilon>0} F(\epsilon)$ is the set of all $(y, w) \in A_1^* \times \mathbb{S}_+^2$ such that

1. $N < \infty$,
2. $r_j > 0$ for $1 \le j \le N - 1$,
3. $\theta_j < \frac{\pi}{2}$ for $0 \le j \le N + 1$, and
4. $\nu > 0$.

By construction of the cylindrical collision law, the set of initial states (y, w) such that $N = \infty$ is a measure zero subset of $A_1^* \times \mathbb{S}_+^2$ (see §4.3.2 and §6.2.1). If $r_j = 0$, this means that the billiard trajectory hits $\partial\mathcal{M}_{\text{cyl}}$ at y^j either tangentially or at a singularity of $\partial\mathcal{M}_{\text{cyl}}$, and the set of initial states such that this happens has measure zero (see Lemma 6.3). If $\theta_j = \pi/2$, this means the billiard trajectory hits $\partial\mathcal{M}_{\text{cyl}}$ tangentially at y^j, and again the set of initial states such that this happens has measure zero. Finally, since ν is a conserved quantity, we see that $\nu = 0$ if and only if $w = \pm\chi$, and $\Lambda^2(A_1^* \times \{\pm\chi\}) = 0$. This proves $\bigcup_{\epsilon>0} F(\epsilon)$ is a full-measure subset of $A_1^* \times \mathbb{S}_+^2$, as desired.

To show that $\frac{\epsilon}{\rho(\epsilon)}\Lambda^2(R_\epsilon^* \times \mathbb{S}_+^2 \setminus \Xi(\epsilon)) \to 0$, observe that $R_\epsilon^* \times \mathbb{S}_+^2 \setminus \Xi(\epsilon)) = (R_\epsilon^* \setminus$

$R'_\epsilon) \times \mathbb{S}^2_+$, where R'_ϵ is the parallelogram defined by

$$R'_\epsilon = \Bigg\{(x_1, \alpha) \in \mathbf{P} : 0 \le \alpha + x_1 \le 1, \\ \text{and } |\alpha| \le \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) - \left(\frac{\overline{L}}{\sin\nu} + 1\right) - \frac{\log(1/\rho)^2\overline{L}}{4\log\log(1/\rho)}\Bigg\}. \tag{5.196}$$

It is enough to show that $\frac{\epsilon}{\rho(\epsilon)} m(R_\epsilon \setminus R'_\epsilon) \to 0$, where m denotes Lebesgue measure on $\mathbf{P}$. Note that $R'_\epsilon \subset R_\epsilon$. Computing the area of each rectangle, we have

$$\begin{aligned} \frac{\epsilon}{\rho(\epsilon)} m(R_\epsilon \setminus R'_\epsilon) &= \frac{\epsilon}{\rho(\epsilon)}\left[\epsilon^{-1}\rho - 2\epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) + 2\left(\frac{\overline{L}}{\sin\nu} + 1\right) + \frac{2\log(1/\rho)^2\overline{L}}{4\log\log(1/\rho)}\right] \\ &= \frac{2\delta_0}{\rho} + \frac{2\epsilon}{\rho}\left(\frac{\overline{L}}{\sin\nu} + 1\right) + \frac{\epsilon\log(1/\rho)^2\overline{L}}{2\rho\log\log(1/\rho)}. \end{aligned} \tag{5.197}$$

Recalling that $\frac{\epsilon^{1/2}}{\rho} = o(1)$, and $\delta_0 = O(\frac{\epsilon}{\rho}) = o(\epsilon^{1/2})$, we see that the above quantity converges to zero as $\epsilon \to 0$.

Part (iii) of Lemma 5.3 is proved by iterating the result of Lemma 5.8. Fix $(y, w) \in \Omega$ and let N, (y^j, w^j), θ_j, and r_j be defined as in the definition of Ω above. Let $\widetilde{y}^0 = H_\epsilon(y^0)$ and let $\widetilde{w}^0 = w^0$. Let Φ^* denote the billiard map in the zoomed configuration space $\mathcal{M}^*$. We will use Lemma 5.8 to prove that

$$(\widetilde{y}^j, \widetilde{w}^j) := \Phi^*(\widetilde{y}^{j-1}, \widetilde{w}_{j-1}), \quad 1 \le j \le N. \tag{5.198}$$

are well-defined, and that the trajectory starting from $(\widetilde{y}^0, \widetilde{w}^0)$ will return to the surface $\widetilde{\mathbf{P}}^*$ in some state $(\widetilde{y}^{N+1}, \widetilde{w}^{N+1})$ after N collisions with the boundary. The same lemma will also give us bounds on $\|(\widetilde{y}^{N+1}, \widetilde{w}^{N+1}) - (y^{N+1}, w^{N+1})\|$. Indeed, we

define by backwards induction

$$
\begin{aligned}
s_{N+1} &= 79\log(1/\rho)^{-1}, \\
s_j &= \frac{s_{j+1}\cos\theta_{j+1}\sin^2\nu}{1600(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})\overline{\kappa}(\overline{L}+1)}, \quad 0 \le j \le N.
\end{aligned}
\tag{5.199}
$$

Then by definition of Ω^*, for $0 \le j \le N$,

$$
s_j \le s_{N+1} = 79\log(1/\rho)^{-1} \le 79\min\{r_j, (\overline{\kappa})^{-1}\cos\theta_{j+1}\}. \tag{5.200}
$$

Also observe that for $1 \le j \le N+1$,

$$
s_j \ge s_1 = s_{N+1}\prod_{j=1}^{N}\frac{\cos\theta_{j+1}\sin^2\nu}{1600(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})\overline{\kappa}(\overline{L}+1)}. \tag{5.201}
$$

Hence, using the bounds in the definition of Ω^*,

$$
\begin{aligned}
&\frac{s_j\cos\theta_j\sin^2\nu}{800\overline{\kappa}(\overline{L}+1)} \\
&\ge \left(s_{N+1}\prod_{j=1}^{N}\frac{\cos\theta_{j+1}\sin^2\nu}{1600(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})\overline{\kappa}(\overline{L}+1)}\right)\frac{\cos\theta_j\sin^2\nu}{800\overline{\kappa}(\overline{L}+1)} \\
&\ge \frac{79}{\log(1/\rho)}\left(\frac{\log(1/\rho)^{-3}}{1600(1+mJ^{-1})^{1/2}(2+(mJ)^{-1/2})(\overline{L}+1)}\right)^N\frac{\log(1/\rho)^{-3}}{800(\overline{L}+1)} \\
&= \log(1/\rho)^{-4N}\cdot\frac{79\log(1/\rho)^{N-4}}{[1600(1+J^{-1})^{1/2}(2+J^{-1/2})(\overline{L}+1)]^N 800(\overline{L}+1)}.
\end{aligned}
\tag{5.202}
$$

Substituting in the given upper bound for N in the definition of Ω^*, and noting the that the right factor above converges to infinity as $\rho = \rho(\epsilon) \to 0$, we obtain that for ϵ

sufficiently small, the last line above is bounded below by

$$\log(1/\rho)^{-\frac{\log(1/\rho)}{\log\log(1/\rho)}} = \rho = \rho(\epsilon). \tag{5.203}$$

Thus

$$\breve{\rho}^{-1}\left(\frac{s_j \cos\theta_j \sin^2\nu}{800\overline{\kappa}(\overline{L}+1)}\right) \geq \epsilon. \tag{5.204}$$

An almost identical calculation gives us, for ϵ sufficiently small,

$$\frac{s_j \cos\theta_j \sin^3\nu}{800\overline{\kappa}(\overline{L}+1)^2} \geq \log(1/\rho)^{-5N} \geq \rho(\epsilon)^{\frac{5}{4}} \geq \epsilon, \tag{5.205}$$

here recalling that $\epsilon^{1/2}/\rho = o(1)$. Assume $\alpha^0 \in [-\epsilon^{-1}\rho/2 + k\epsilon^{-1}\rho, \epsilon^{-1}\rho/2 + k\epsilon^{-1}\rho)$. For $0 \leq j \leq N-1$,

$$|\alpha^{j+1} - \alpha^j| \leq \|w^j\| t_j = t_j, \tag{5.206}$$

where t_j is the time for the billiard trajectory to go from y^j to y^{j+1}. We have $t_j = \frac{\|p^{j+1}-p^j\|}{\|u^j\|} \leq \frac{\overline{L}}{\sin\nu}$ by Lemma 5.6, recalling that by definition $u^j = \mathrm{d}(G_0)_{y^j}(w^j)$. Hence, $|\alpha^{j+1} - \alpha^j| \leq \frac{\overline{L}}{\sin\nu}$. Thus, for $0 \leq j \leq N$, applying the bound on $|\alpha^0 - k\epsilon^{-1}\rho|$ coming from the definition of Ω, we obtain

$$\begin{aligned}
|\alpha^j - k\epsilon^{-1}\rho| &\leq |\alpha^0 - k\epsilon^{-1}\rho| + \sum_{j=0}^{N-1} |\alpha^{j+1} - \alpha^j| \\
&\leq \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) - \left(\frac{\overline{L}}{\sin\nu} + 1\right) - \frac{\log(1/\rho)^2\overline{L}}{4\log\log(1/\rho)} + \frac{N\overline{L}}{\sin\nu} \\
&\leq \epsilon^{-1}\left(\frac{\rho}{2} - \delta_0\right) - \left(\frac{\overline{L}}{\sin\nu} + 1\right).
\end{aligned} \tag{5.207}$$

By definition of y^0 and $\widetilde{y}^0$, and the fact that H_ϵ fixes the α-coordinate, we have

$$p^0 = G_0(y^0) = G_0 \circ H_\epsilon^{-1}(H_\epsilon(y^0)) = G_\epsilon(\widetilde{y}^0) = \widetilde{p}^0. \tag{5.208}$$

Also $\widetilde{w}^0 = w^0$ by definition. Thus trivially

$$||\widetilde{p}^0 - p^0|| + |\widetilde{\alpha}^0 - \alpha^0| + ||\widetilde{w}^0 - w^0|| \leq s_0. \tag{5.209}$$

From (5.200), (5.204), (5.205), and (5.207) we see that the hypotheses of Lemma 5.8 are satisfied. Thus, by induction, for $1 \leq j \leq N+1$, $\widetilde{p}^j$, the pair $(\widetilde{y}^{j-1}, \widetilde{w}^{j-1})$ is licit, in the sense of §5.3.4. Hence, $\widetilde{\alpha}^j$, and $\widetilde{w}^j$ are well-defined, p^j and $\widetilde{p}^j$ lie on the same smooth curve segment in $\partial\mathcal{D}^*$, and

$$||\widetilde{p}^j - p^j|| + |\widetilde{\alpha}^j - \alpha^j| + ||\widetilde{w}^j - w^j|| \leq s_j. \tag{5.210}$$

In particular, p^j and $\widetilde{p}^j$ lie on $\partial\widehat{B}^*$ for $1 \leq j \leq N$; p^0 and $\widetilde{p}^{N+1}$ both lie on the line $\{(x_1, x_2) : x_2 = 0\}$, and

$$||\widetilde{p}^{N+1} - p^{N+1}|| + |\widetilde{\alpha}^{N+1} - \alpha^{N+1}| + ||\widetilde{w}^{N+1} - w^{N+1}|| \leq s_{N+1} = \frac{79}{\log(1/\rho)}. \tag{5.211}$$

We will use this estimate shortly.

Recall the diffeomorphism $H_\epsilon : \widehat{\mathcal{Z}}^* \to \widehat{\mathcal{Z}}^*$, and define a diffeomorphism $\overline{H}_\epsilon : \widehat{\mathcal{Z}}^* \times \mathbb{S}^2 \to \widehat{\mathcal{Z}}^* \times \mathbb{S}^2$ by

$$\overline{H}_\epsilon(y, w) = (H_\epsilon(y), w). \tag{5.212}$$

In other words, $\overline{H}_\epsilon = \sigma_{\epsilon^{-1}} \circ \overline{H}_1 \circ \sigma_\epsilon$, where $\overline{H}_1$ is defined as in (4.83). We also define "zoomed versions" of Ψ, $K^{\Sigma,\epsilon}$, $K^{\Sigma,\epsilon}_{\text{cyl}}$, and $\widetilde{K}^{\Sigma,\epsilon}$ as follows:

$$\begin{aligned}
\Psi^* &= \sigma_{\epsilon^{-1}} \circ \Psi \circ \sigma_\epsilon : \widetilde{\mathbf{P}}^* \times \mathbb{S}^2_{+\rho/2} \to \mathbf{P}^* \times \mathbb{S}^2_{+\rho/2}, \\
K^* &= \sigma_{\epsilon^{-1}} \circ K^\epsilon \circ \sigma_\epsilon : \sigma_{\epsilon^{-1}}(\mathcal{F}) \to \sigma_{\epsilon^{-1}}(\mathcal{F}), \\
K^*_{\text{cyl}} &= \sigma_{\epsilon^{-1}} \circ K^\epsilon_{\text{cyl}} \circ \sigma_\epsilon : \sigma_{\epsilon^{-1}}(\mathcal{F}_{\text{cyl}}) \to \sigma_{\epsilon^{-1}}(\mathcal{F}_{\text{cyl}}), \\
\widetilde{K}^* &= \sigma_{\epsilon^{-1}} \circ \widetilde{K}^\epsilon \circ \sigma_\epsilon : \sigma_{\epsilon^{-1}}(\widetilde{\mathcal{F}}) \to \sigma_{\epsilon^{-1}}(\widetilde{\mathcal{F}}).
\end{aligned} \tag{5.213}$$

Then from (4.86) we see that

$$\widetilde{K}^* = (\overline{H}_\epsilon)^{-1} \circ (\Psi^*)^{-1} \circ K^* \circ \Psi^* \circ \overline{H}_\epsilon. \tag{5.214}$$

For the rest of our argument to be valid, we need $(\widetilde{y}^0, -\widetilde{w}^0)$ and $(\widetilde{y}^{N+1}, \widetilde{w}^{N+1})$ to lie in the domain of Ψ^*. This follows from:

Claim 5.3.1. *For ϵ sufficiently small, $(\widetilde{y}^0, -\widetilde{w}^0) \in \widetilde{\mathbf{P}}^* \times \mathbb{S}^2_{+\rho/2}$ and $(\widetilde{y}^{N+1}, \widetilde{w}^{N+1}) \in \widetilde{\mathbf{P}}^* \times \mathbb{S}^2_{+\rho/2}$.*

Proof of Claim 5.3.1. As explained above $\widetilde{y}^0$ and $\widetilde{y}^{N+1}$ lie in $\widetilde{\mathbf{P}}^*$. Let $\psi_0 = \angle(-\widetilde{w}^0, e_2)$, and let $\psi_{N+1} = \angle(\widetilde{w}^{N+1}, e_2)$. We must show that $\psi_0 \le \frac{\pi}{2} - \rho$ and $\psi_{N+1} \le \frac{\pi}{2} - \rho$. We

have

$$\begin{aligned}\cos\psi_0 = \langle -\widetilde{w}^0, e_2\rangle &= \langle \mathrm{d}(G_\epsilon)_{\widetilde{y}^0}(-\widetilde{w}^0), e_2\rangle + \langle \mathrm{d}(G_0)_{\widetilde{y}^0}(-\widetilde{w}^0) - \mathrm{d}(G_\epsilon)_{\widetilde{y}^0}(-\widetilde{w}^0), e_2\rangle \\ &\quad + \langle -\widetilde{w}^0 - \mathrm{d}(G_0)_{\widetilde{y}^0}(-\widetilde{w}^0), e_2\rangle \\ &\geq \|\mathrm{d}(G_\epsilon)(-\widetilde{w}^0)\|\cos\theta_0 - \|\mathrm{d}(G_0)_{\widetilde{y}^0}(-\widetilde{w}^0) - \mathrm{d}(G_\epsilon)_{\widetilde{y}^0}(-\widetilde{w}^0)\|,\end{aligned} \tag{5.215}$$

here using that $\langle -\widetilde{w}^0 - \mathrm{d}(G_0)_{\widetilde{y}^0}(-\widetilde{w}^0), e_2\rangle = 0$. Let $\widetilde{\alpha}^0$ denote the angular coordinate of $\widetilde{y}^0$. By Lemma 5.6,

$$\begin{aligned}\|\mathrm{d}(G_\epsilon)(-\widetilde{w}^0)\| &\geq [1 - ((mJ)^{-1/2} + 1)|\epsilon\widetilde{\alpha}_0|]\sin\varphi \\ &\geq [1 - ((mJ)^{-1/2} + 1)\epsilon\rho/2]\sin\varphi.\end{aligned} \tag{5.216}$$

Also, by (5.86) we deduce that

$$\|\mathrm{d}(G_0)_{\widetilde{y}^0}(-\widetilde{w}^0) - \mathrm{d}(G_\epsilon)_{\widetilde{y}^0}(-\widetilde{w}^0)\| \leq |\epsilon\widetilde{\alpha}^0| \leq \frac{\epsilon\rho}{2}. \tag{5.217}$$

Substituting these bounds into (5.215), we obtain

$$\begin{aligned}\cos\psi_0 &\geq [1 - ((mJ)^{-1/2} + 1)\epsilon\rho/2]\sin\varphi\cos\theta_0 - \frac{\epsilon\rho}{2} \\ &\geq [1 - ((mJ)^{-1/2} + 1)\epsilon\rho/2](\sin\varphi)\log(1/\rho)^{-1} - \frac{\epsilon\rho}{2} \geq \sin(\rho),\end{aligned} \tag{5.218}$$

for ϵ sufficiently small. Therefore, $\psi_0 \leq \frac{\pi}{2} - \rho(\epsilon)$ for ϵ sufficiently small. We omit the argument showing that $\psi_{N+1} \leq \frac{\pi}{2} - \rho(\epsilon)$ since it is almost identical. ■

By Claim 5.3.1, we may define $(\widehat{y}^0, \widehat{w}^0) = \Psi^*(\widetilde{y}^0, -\widetilde{w}^0)$, and $(\widehat{y}^{N+1}, \widehat{w}^{N+1}) =$

$\Psi^*(\widetilde{y}^{N+1}, \widetilde{w}^{N+1})$. By definition of Ψ and Ψ^*, $\widehat{y}^0$ and $\widehat{y}^{N+1}$ both lie in the plane $\mathbf{P}$; the interior of the line segment from $\widehat{y}^0$ to $\widetilde{y}^1$ lies in $\operatorname{Int}\mathcal{M}^*$ and contains the point $\widehat{y}^0$; and the interior of the line segment from $\widetilde{y}^N$ to $\widehat{y}^{N+1}$ lies in the interior of $\mathcal{M}^*$ and contains the point $\widetilde{y}^{N+1}$ in its interior. Thus the billiard trajectory in $\mathcal{M}^*$ starting in state $(\widehat{y}^0, -\widehat{w}^0)$ hits the boundary N times before returning to the plane $\mathbf{P}$ in state (y^{N+1}, w^{N+1}). Therefore,

$$K^*(\widehat{y}^0, \widehat{w}^0) = (\widehat{y}^{N+1}, \widehat{w}^{N+1}), \qquad \text{and} \qquad K^*_{\text{cyl}}(y^0, -w^0) = (y^{N+1}, w^{N+1}). \tag{5.219}$$

Consequently,

$$\begin{aligned}
\widetilde{K}^*(y^0, -w^0) &= (\overline{H}_\epsilon)^{-1} \circ (\Psi^*)^{-1} \circ K^* \circ \Psi^* \circ \overline{H}_\epsilon(y^0, -w^0) \\
&= (\overline{H}_\epsilon)^{-1} \circ (\Psi^*)^{-1} \circ K^* \circ \Psi^*(\widehat{y}^0, -\widetilde{w}^0) \\
&= (\overline{H}_\epsilon)^{-1} \circ (\Psi^*)^{-1} \circ K^*(\widehat{y}^0, \widehat{w}^0) \\
&= (\overline{H}_\epsilon)^{-1} \circ (\Psi^*)^{-1}(\widehat{y}^{N+1}, \widehat{w}^{N+1}) \\
&= (\overline{H}_\epsilon)^{-1}(\widetilde{y}^{N+1}, \widetilde{w}^{N+1}) = (H_\epsilon^{-1}(\widetilde{y}^{N+1}), w^{N+1}).
\end{aligned} \tag{5.220}$$

Therefore,

$$\begin{aligned}
\|\widetilde{K}^*(y^0, -w^0) - K^*_{\text{cyl}}(y^0, -w^0)\| &\le \|H_\epsilon^{-1}(\widetilde{y}^{N+1}) - y^{N+1}\| + \|\widetilde{w}^{N+1} - w^{N+1}\| \\
&\le \|G_0(H_\epsilon^{-1}(\widetilde{y}^{N+1}) - y^{N+1})\| + |\widetilde{\alpha}^{N+1} - \alpha^{N+1}| + \|\widetilde{w}^{N+1} - w^{N+1}\|,
\end{aligned} \tag{5.221}$$

here using the definition of G_0 and the fact that H_ϵ fixes the angular coordinate.

But by definition of G_ϵ,

$$G_0(H_\epsilon^{-1}(\widetilde{y}^{N+1})) = G_\epsilon(\widetilde{y}^{N+1}) = \widetilde{p}^{N+1}. \tag{5.222}$$

Thus

$$\|G_0(H_\epsilon^{-1}(\widetilde{y}^{N+1}) - y^{N+1})\| = \|\widetilde{p}^{N+1} - p^{N+1}\|. \tag{5.223}$$

Substituting the above into (5.221), and applying the bound (5.155), we obtain

$$\begin{aligned} &\|\widetilde{K}^*(y^0, -w^0) - K^*_{\mathrm{cyl}}(y^0, -w^0)\| \\ &\qquad \le \|\widetilde{p}^{N+1} - p^{N+1}\| + |\widetilde{\alpha}^{N+1} - \alpha^{N+1}| + \|\widetilde{w}^{N+1} - w^{N+1}\| \\ &\qquad\qquad \le \frac{80}{\log(1/\rho)^3}. \end{aligned} \tag{5.224}$$

Note that this bound does not depend on the choice of $(y, w) = (y^0, -w^0) \in \Omega^*$. Thus

$$\begin{aligned} \sup_\Omega \|\widetilde{K}^{\Sigma,\epsilon} - K^{\Sigma,\epsilon}\| &\le \sup_\Omega \|\sigma_{\epsilon^{-1}} \circ \widetilde{K}^{\Sigma,\epsilon} - \sigma_{\epsilon^{-1}} \circ K^{\Sigma,\epsilon}\| \\ &= \sup_{\Omega^*} \|\widetilde{K}^* - K^*_{\mathrm{cyl}}\| \le \frac{80}{\log(1/\rho)^3}, \end{aligned} \tag{5.225}$$

and the right-hand side converges to zero as $\epsilon \to 0$, as desired. □

6 Rough Reflections in General Billiard Domains

6.1 The billiard map and the invariant measure

The results stated in this section are well-known, but proofs under the weak regularity conditions adopted in this work may not be found in standard references. We therefore provide careful proofs here. More traditional treatments of billiards may be found in [6] and [39]. For billiards in Riemannian manifolds, see [8, Chapter 5].

6.1.1 A billiard domain with singularities

Let $\mathcal{R}$ be a d-dimensional C^2 Riemannian manifold with metric $\langle \cdot, \cdot \rangle$. Assume that $\mathcal{R}$ is geodesically complete.

Our billiard domain will be taken from a certain class of closed C^2 submanifolds of $\mathcal{R}$ with singularities. Namely, let $\mathcal{N}$ be a subset of $\mathcal{R}$, and let $\partial\mathcal{N}$ and $\operatorname{Int}\mathcal{N}$ denote the topological boundary and interior of $\mathcal{N}$. We say that $\mathcal{N}$ belongs to the class $\mathrm{CES}_0^2(\mathcal{R})$ if $\mathcal{N}$ is a closed subset of $\mathcal{R}$, and there exists a closed subset $\mathcal{S} \subset \partial\mathcal{N}$ such that the following two conditions hold:

C1. The $(d-1)$-dimensional Hausdorff measure of $\mathcal{S}$ is zero.

C2. For every point $q \in \mathcal{N} \setminus \mathcal{S}$, there exists a neighborhood $U \subset \mathcal{R}$ of q and a C^2 diffeomorphism $\phi : U \to \mathbb{R}^d$ such that $\phi(q) = 0$ and the image of $\mathcal{N} \cap U$ is either $\mathbb{R}^d$ or the upper half-space $\mathbb{H}^d := \{(x_1, \ldots, x_d) \ : x_d \geq 0\}$.

The second condition means that $\mathcal{N} \setminus \mathcal{S}$ is a codimension 0, embedded, C^2 submanifold of $\mathcal{R}$ with boundary. We let $\mathcal{N}_{\mathrm{reg}} = \mathcal{N} \setminus \mathcal{S}$.

We will refer to points in $\mathcal{S}$ as *singular points* of $\mathcal{N}$, and to points in $\mathcal{N} \setminus \mathcal{S}$ as *regular points* of $\mathcal{N}$. To avoid ambiguities, we will always assume that $\mathcal{S}$ is chosen to be minimal. That is, $\mathcal{S}$ is a smallest subset of $\partial\mathcal{N}$, with respect to inclusion, such that condition C2 holds.

In condition C2, we will refer to points q with a neighborhood mapping to $\mathbb{R}^d$ as *interior points* of $\mathcal{N}$, and we will refer to points q with a neighborhood mapping to $\mathbb{H}^d$ as *regular boundary points* of $\mathcal{N}$. These notions do not depend on ϕ. The set of regular boundary points coincides with $\partial\mathcal{N} \setminus \mathcal{S}$, and the set of interior points coincides with $\operatorname{Int}\mathcal{N}$.

6.1.2 Billiard map

Let $T\mathcal{R}$ denote the tangent space over $\mathcal{R}$, and let $T^1\mathcal{R} = \{(q,p) \in T\mathcal{R} : \langle p,p\rangle_q = 1\}$ denote the unit tangent bundle. We denote the *geodesic flow* on $T^1\mathcal{R}$ by G^t. That is, for each (q,p), if $\gamma(t)$ is the unique (unit speed) geodesic starting from state (q,p) (defined for all time by completeness), then by definition $G^t(q,p) = (\gamma(t), \dot\gamma(t)) \in T^1\mathcal{R}$. In local coordinates (q^i, p^i) on the tangent space, the geodesics satisfy:

$$\dot q^i = p^i, \qquad \dot p^i = -\sum_{i,k=1}^{d} \Gamma^i_{jk} p^j p^k, \tag{6.1}$$

where Γ^i_{jk} are the Christoffel symbols associated with the Riemannian metric, and "dot" denotes the time derivative. The flow G^t is generated by the vector field on

$T^1\mathcal{R}$, expressed in local coordinates as

$$X = \sum_{i=1}^{d} p^i \frac{\partial}{\partial q^i} - \sum_{j,k=1}^{d} \Gamma^i_{jk} p^j p^k \frac{\partial}{\partial p^i}. \tag{6.2}$$

It follows from this expression that

$$\mathrm{d}\pi(X(q,p)) = \sum_{i=1}^{d} p^i \frac{\partial}{\partial q_i} = (q,p), \tag{6.3}$$

where

$$\pi : T\mathcal{R} \to \mathcal{R} \tag{6.4}$$

is natural projection.

For each $q \in \partial\mathcal{N}_{\mathrm{reg}}$, let $n(q) \in T^1_p\mathcal{R}$ denote the unit normal vector pointing into $\mathcal{N}$. We denote the restriction of the unit tangent bundle on $\mathcal{R}$ to the regular boundary of $\mathcal{N}$ by

$$\mathcal{U} = \{(q,p) \in T^1\mathcal{R} : q \in \partial\mathcal{N}_{\mathrm{reg}}\}. \tag{6.5}$$

We also denote the inward (+) and outward (−)-pointing unit tangent bundles on the regular boundary by

$$\mathcal{U}^{\pm} = \{(q,p) \in T^1\mathcal{R} : q \in \partial\mathcal{N}_{\mathrm{reg}} \text{ and } \pm\langle p, n(q)\rangle_q > 0\}. \tag{6.6}$$

An important observation is that the vector field X is nowhere tangent to $\mathcal{U}^+$. Indeed, if $f(q)$ is a local defining function for $\partial\mathcal{N}$, then $\nabla f(q) = cn(q)$ for some $c \neq 0$,

and $F(q,p) = f \circ \pi(q,p)$ is a local defining function for $\mathcal{U}$. Thus, for $(q,p) \in \mathcal{U}^+$,

$$\mathrm{d}F \circ X(q,p) = \mathrm{d}f \circ \mathrm{d}\pi \circ X(q,p) = \mathrm{d}f(q,p) = \mathrm{d}f_q(p) = c\langle p, n(q)\rangle_q \neq 0, \tag{6.7}$$

where the second equality follows by (6.3). This implies $X(q,p)$ does not lie in $\ker \mathrm{d}F_{(q,p)} = T_{(q,p)}\mathcal{U}$. A consequence of this observation is that the flowout

$$G : \mathcal{U}^+ \times \mathbb{R} \to T^1\mathcal{R}, \qquad G(q,p,t) := G^t(q,p), \tag{6.8}$$

is a local diffeomorphism.

A point particle starting from $(q,p) \in \mathcal{U}^+$ first returns to ∂M at time

$$\bar{t}(q,p) := \inf\{t > 0 : \pi \circ G^t(q,p) \notin \operatorname{Int}\mathcal{N}\}. \tag{6.9}$$

(If $\pi \circ G^t(q,p) \in \operatorname{Int}\mathcal{N}$ for all $t > 0$, then by convention the above quantity is infinite.) Since the geodesic starting from a point $(q,p) \in \mathcal{U}^+$ initially points into $\mathcal{N}$, the time $\bar{t}$ is strictly positive. In fact, if $\bar{t}(q,p)$ is finite, then since the complement of $\operatorname{Int}\mathcal{N}$ is closed, the infimum in (6.9) must be a minimum.

We introduce the following subbundles of $\mathcal{U}^+$:

$$\begin{aligned} \mathcal{U}^+_{\mathrm{fin}} &= \{(q,p) \in \mathcal{U}^+ : \bar{t}(q,p) < \infty\}, \\ \mathcal{U}^+_1 &= \{(q,p) \in \mathcal{U}^+_{\mathrm{fin}} : G^{\bar{t}(q,p)}(q,p) \in \mathcal{U}^-\}. \end{aligned} \tag{6.10}$$

In applications, $\mathcal{U}^+_{\mathrm{fin}}$ will be a full-measure subset of $\mathcal{U}^+$. The set $\mathcal{U}^+_1$ is an open subset of $\mathcal{U}$ (not just of $\mathcal{U}^+_{\mathrm{fin}}$). This may be proved by showing that $\bar{t}$ is finite and continuous

on a neighborhood in $\mathcal{U}$ of a point $(q,p) \in \mathcal{U}_1^+$. (See the proof of Lemma 6.1).

The specular reflection map $R : \mathcal{U} \to \mathcal{U}$ is defined by

$$R(q,p) = (q, p - 2\langle p, n(q)\rangle_q n(q)). \tag{6.11}$$

Note that R maps $\mathcal{U}^\pm$ onto $\mathcal{U}^\mp$ and is involutive (that is, $R \circ R = \mathrm{Id}$).

The *billiard map* $\Phi : \mathcal{U}_1^+ \to \mathcal{U}^+$ is defined by

$$\Phi(q,p) = R \circ G^{\bar{t}(q,p)}(q,p). \tag{6.12}$$

Lemma 6.1. *(i) Φ is injective, and its left-inverse is given by $-R \circ \Phi \circ -R$, where $-R(q,p) := R(q,-p)$.*

(ii) Suppose $(q_1,p_1) := \Phi(q_0,p_0)$ lies in $\mathcal{U}_1^+$. Then Φ restricts to a C^1 diffeomorphism from a neighborhood of (q_0,p_0) to a neighborhood of (q_1,p_1).

Proof. (i) Consider the unit speed geodesic running from $(q,p) \in \mathcal{U}_1^+$ to $(q',p') := G^{\bar{t}(q,p)}(q,p) \in \mathcal{U}^+$. Noting that R is involutive, we have $-R \circ \Phi(q,p) = (q',-p')$. The geodesic starting from $(q',-p')$ retraces the original geodesic and returns to the boundary in the state $(q,-p)$. Thus by symmetry $-R \circ \Phi(q',-p') = (q,p)$. Hence $-R \circ \Phi \circ -R \circ \Phi(q,p) = (q,p)$.

(ii) Write $\Phi = R \circ \Phi_-$, where $\Phi_-(q,p) = G^{\bar{t}(q,p)}(q,p)$. Let $(q_1,p_1^-) = \Phi_-(q_0,p_0)$. Since $\partial\mathcal{N}_{\mathrm{reg}}$ is an C^2 submanifold of $\mathcal{R}$ of codimension 1, it has a C^2 local defining function f defined in a neighborhood of q_1. Then $F(q,p) := f(q) = f \circ \pi(q,p)$ is a

local defining function for $\mathcal{U}$. The function

$$K(q,p,t) := F(G^t(q,p)) \tag{6.13}$$

is defined in a neighborhood of $(q_0, p_0, \bar{t}(q_0,p_0))$, and

$$\begin{aligned} \frac{\partial K}{\partial t}(q_0,p_0,\bar{t}(q_0,p_0)) &= \mathrm{d}F(q_1,p_1^-)X(q_1,p_1^-) = \mathrm{d}f \circ \mathrm{d}\pi \circ X(q_1,p_1^-) \\ &= \mathrm{d}f_{q_1}(p_1^-) = c\langle n(q_1), p_1^- \rangle_{q_1} \neq 0, \end{aligned} \tag{6.14}$$

for some $c \neq 0$. The third equality above follows from (6.3), and the last equality follows because the gradient of f is parallel to $n(q_1)$ at q_1. By the Implicit Function Theorem, there is a C^2 function $\widehat{t}$ defined on a neighborhood $U \subset \mathcal{U}_1^+$ of (q_0,p_0), and a finite open interval $I = (a,b)$ containing $\bar{t}(q_0,p_0)$ such that solutions to

$$K(q,p,t) = 0, \quad (q,p,t) \in U \times I \tag{6.15}$$

take the form $(q,p,\widehat{t}(q,p))$.

We claim $\bar{t} = \widehat{t}$ in a neighborhood of (q_0,p_0). To see this, let $V \subset\subset U$ be a precompact open set containing (q_0,p_0), and let

$$A = \{(q,p,t) \in \overline{V} \times [0,a] : \pi \circ G^t(q,p) \notin \operatorname{Int} M\}. \tag{6.16}$$

For $(q,p) \in \overline{V}$, it is evident that if $\bar{t}(q,p) \neq \widehat{t}(q,p)$, then $\bar{t}(q,p) \leq a$, since otherwise $(q,p,\bar{t}(q,p))$ would solve (6.15). Therefore $\{(q,p) \in V : \bar{t}(q,p) \neq \bar{t}(q,p)\} \subset \pi_{\mathcal{U}}(A)$, where $\pi_{\mathcal{U}} : \mathcal{U} \times \mathbb{R} \to \mathcal{U}$ is projection. Note that $\pi_{\mathcal{U}}(A)$ is compact. Since $\bar{t}(q_0,p_0) > a$,

we have $(q_0, p_0) \notin \pi_{\mathcal{U}}(A)$, and $\bar{t} = \widehat{t}$ on $V \setminus \pi_{\mathcal{U}}(A)$.

Consequently, Φ_- is C^2 in a neighborhood of (q_0, p_0). The unit normal field n and the specular reflection map R are C^1. Therefore $\Phi = R \circ \Phi_-$ is C^1 in a neighborhood of (q_0, p_0).

By the same reasoning, Φ is C^1 in a neighborhood of $(q_1, -p_1^-)$. By (i), Φ has a left inverse $-R \circ \Phi \circ -R$ which is C^1 in a neighborhood of (q_1, p_1). Thus Φ is a local C^1 diffeomorphism. $\square$

6.1.3 Invariant measure

At each point $q \in \mathcal{R}$, the tangent space $T_q\mathcal{R}$ inherits a Euclidean metric from the metric on the manifold. Let σ_q denote surface measure on the unit sphere $T_q^1\mathcal{R} \subset T_q\mathcal{R}$ with respect to this metric, and let τ denote the measure on $\mathcal{R}$ generated by the Riemannian metric on the manifold. Define a measure μ on the unit tangent bundle by

$$\mu(\mathrm{d}p\mathrm{d}q) = \sigma_q(\mathrm{d}p)\tau(\mathrm{d}q). \tag{6.17}$$

In other words,

$$\int_{T^1\mathcal{R}} f \mathrm{d}\mu = \int_{\mathcal{R}} \int_{T_q^1\mathcal{R}} f(q,p)\sigma_q(\mathrm{d}p)\tau(\mathrm{d}q), \qquad \text{for all } f \in C_c(T^1\mathcal{R}). \tag{6.18}$$

It is well-known that the geodesic flow preserves the measure μ, in the sense that $\int_{T^1\mathcal{R}} f \circ G^t \mathrm{d}\mu = \int_{T^1\mathcal{R}} f \mathrm{d}\mu$ for any f and $t \in \mathbb{R}$. This is Liouville's Theorem, stated for a Hamiltonian system with energy function $H(q,p) = \frac{1}{2}\langle p, p\rangle_q$. See [2, §16], or [8, Ch. 5].

Let τ_1 denote the surface measure on $\partial\mathcal{N}_{\text{reg}}$ induced by the metric on the ambient space $\mathcal{R}$. We define a measure μ_1 on $\mathcal{U}_+$ as follows:

$$\mu_1(\mathrm{d}p\mathrm{d}q) = \langle q, n(p)\rangle_p \sigma_p(\mathrm{d}q)\tau_1(\mathrm{d}p). \tag{6.19}$$

Lemma 6.2. *The billiard map Φ preserves the measure μ_1, in the sense that, for any $A \subset \mathcal{U}_1^+$, $\mu_1(\Phi(A)) = \mu_1(A)$.*

Proof. We equip $T\mathcal{R}$ with the *Sasaki metric* $\langle\cdot,\cdot\rangle_S$. This is the metric, first introduced in [37], which restricts on $\mathcal{R}$ to the given Riemannian metric and on the fibers $T_q\mathcal{R}$ to the inherited Euclidean metric, and which makes $\mathcal{R}$ and $T_q\mathcal{R}$ orthogonal. Given $\xi, \zeta \in T_{(q,p)}T\mathcal{R}$ and curves $\gamma = (\alpha, U)$ and $\eta = (\beta, V)$ with $\dot\gamma(0) = \xi$ and $\dot\eta(0) = \zeta$, the metric is defined by

$$\langle \xi, \zeta\rangle_S = \langle \dot\alpha(0), \dot\beta(0)\rangle + \langle D_{\dot\alpha}U(0), D_{\dot\beta}V(0)\rangle, \tag{6.20}$$

where D denotes covariant differentiation.

The metric $\langle\cdot,\cdot\rangle_S$ induces a metric on the unit tangent bundle $T^1\mathcal{R}$. The volume measure on $T^1\mathcal{R}$ coming from this metric is just the measure μ defined above.

The following simple argument comes from [8, Chapter 6]. With respect to the Sasaki metric, the unit normal vector field on the codimension 1 submanifold $\mathcal{U}_1^+ \subset T^1\mathcal{R}$ is given by $\iota^* n$, where $\iota : \mathcal{R} \to T\mathcal{R}$ is inclusion. Consequently, the flux of the vector field X through any $A \subset \mathcal{U}_1^+$ is

$$\int_A \langle X(q,p), \iota^* n(q,p)\rangle_S \mu(\mathrm{d}q\mathrm{d}p) = \int_A \langle p, n(q)\rangle \mu(\mathrm{d}q\mathrm{d}p) = \mu_1(A). \tag{6.21}$$

As above, write $\Phi = R \circ \Phi_-$. Let $A' = \Phi_-(A)$. Since the flow G^t preserves μ, this implies that the flux is preserved:

$$\mu_1(A) = \mu_1(A'). \tag{6.22}$$

Thus Φ_- preserves the measure μ_1. Since specular reflection R is an isometry and preserves the quantity $\langle p, n(q)\rangle_q$, it too preserves the measure μ_1. □

Lemma 6.3. *(i)* $\mu_1(\mathcal{U}^+_{\text{fin}} \setminus \mathcal{U}^+_1) = 0$.

(ii) For each $m \geq 1$, *let* $\mathcal{U}^+_m = \Phi^{-1}(\mathcal{U}^+_{m-1})$. *Then* $\mu_1(\mathcal{U}^+_{\text{fin}} \setminus \mathcal{U}^+_m) = 0$, *and the iterate* $\Phi^m = \Phi \circ \Phi \circ \cdots \circ \Phi$ *(m times)* $: \mathcal{U}^+_m \to \mathcal{U}^+$ *is a* C^1 *diffeomorphism onto its image.*

Proof. (i) For $(q,p) \in \mathcal{U}^+_{\text{fin}} \setminus \mathcal{U}^+_1$, set $(q',p') = \Phi_-(q,p)$. Let A be the set of all such points (q,p) such that q' lies in the singular part of the boundary $\mathcal{S}$. Let B be the set of all such points (q,p) such that $q' \in \partial\mathcal{N}_{\text{reg}}$ and p' is tangent to $\partial\mathcal{N}_{\text{reg}}$ at q'. We see that $\mathcal{U}^+_{\text{fin}} \setminus \mathcal{U}^+_1$ is the disjoint union of A and B. Thus it enough to show that A and B have measure zero in $\mathcal{U}^+$.

Recall that the flowout G defined by (6.8) is a local diffeomorphism. Let $\widetilde{\mathcal{S}} \subset \mathcal{U}^+ \times \mathbb{R}$ denote the preimage of $\mathcal{S}$ under the composition

$$\mathcal{U}^+ \times \mathbb{R} \xrightarrow{G} T^1\mathcal{R} \xrightarrow{\pi} \mathcal{R}, \tag{6.23}$$

where π is natural projection. Then $A \subset \pi_{\mathcal{U}^+}(\widetilde{\mathcal{S}})$, where $\pi_{\mathcal{U}^+}$ is projection onto the first factor of $\mathcal{U}^+ \times \mathbb{R}$.

Let s denote the Hausdorff dimension of $\mathcal{S}$. By assumption, $s < d-1$. Since the first map in (6.23) is a local C^2 diffeomorphism, and the second is a C^2 submersion

from a $(2d-1)$-dimensional space to a d-dimensional space, we see that the Hausdorff dimension of $\widetilde{\mathcal{S}}$ is $s+d-1$. Thus the Hausdorff dimension of $\pi_{\mathcal{U}^+}(\widetilde{\mathcal{S}})$ is at most $s+d-1<2d-2$. Since $\mathcal{U}^+$ is a $(2d-2)$-dimensional submanifold of $\mathcal{R}$, we conclude that $A \subset \pi_{\mathcal{U}^+}(\widetilde{\mathcal{S}})$ has measure zero in $\mathcal{U}^+$.

To see that B also has measure zero in $\mathcal{U}^+$, let $\mathcal{T} \subset \mathcal{U}^+ \times \mathbb{R}$ be the preimage under G of $T^1\partial\mathcal{N}_{\text{reg}} \subset T^1\mathcal{R}$. Observe that $B \subset \pi_{\mathcal{U}^+}(\mathcal{T})$. The dimension of $T^1\partial\mathcal{N}_{\text{reg}}$ is a $2d-3$, and therefore the dimension of $\mathcal{T}$ is $2d-3$, and the dimension of $\pi_{\mathcal{U}^+}(\mathcal{T})$ is at most $2d-3$. It follows that $B \subset \pi_{\mathcal{U}^+}(\mathcal{T})$ has measure zero in $\mathcal{U}^+$.

(ii) Recall that $\mathcal{U}_1^+$ is an open subset of $\mathcal{U}^+$ and $\Phi : \mathcal{U}_1^+ \to \Phi(\mathcal{U}_1^+)$ is a C^1 diffeomorphism. It clearly follows that each $\mathcal{U}_m^+$ is an open subset of $\mathcal{U}^+$ and each iterate Φ^m is a C^1 diffeomorphism from $\mathcal{U}_m^+$ onto its image. To see that $\mu_1(\mathcal{U}_{\text{fin}}^+ \setminus \mathcal{U}_1^+) = 0$, let $F \subset \mathcal{U}_1^+$ be any measurable set with $\mu_1(F) < \infty$. Using (i) and the fact that Φ preserves μ_1, we have

$$\begin{aligned}
\mu_1(F \setminus \mathcal{U}_2^+) &= \mu_1(F \setminus \Phi^{-1}(\Phi(F) \cap \mathcal{U}_1^+)) \\
&= \mu_1(F) - \mu_1(\Phi^{-1}(\Phi(F) \cap \mathcal{U}_1^+)) \\
&= \mu_1(F) - \mu_1(\Phi(F) \cap \mathcal{U}_1^+) \\
&= \mu_1(F) - \mu_1(\Phi(F)) = 0.
\end{aligned} \tag{6.24}$$

It follows that $\mu_1(\mathcal{U}_1^+ \setminus \mathcal{U}_2^+) = 0$. Inductively, we similarly obtain that $\mu_1(\mathcal{U}_{m-1} \setminus \mathcal{U}_m) = 0$ for all $m \geq 1$. As $\mathcal{U}_1^+$ is a full measure subset of $\mathcal{U}_{\text{fin}}^+$, the result follows. $\square$

6.2 Rough Reflection Laws

The construction of rough reflections described here generalizes a construction appearing in [1]. It is also similar to constructions appearing in [13] and [35].

6.2.1 Macro-reflection laws

Consider two nested billiard domains $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \mathcal{R}$. The boundary of $\mathcal{M}_0$ provides an interface from which we can observe reflections from the boundary of $\mathcal{M}_1$. The goal for this subsection is to describe precisely what it means to "observe" the reflections from $\partial\mathcal{M}_1$ from the interface $\partial\mathcal{M}_0$. In the next subsection, we will describe what happens when $\mathcal{M}_1$ approaches $\mathcal{M}_0$ uniformly.

We suppose that $\mathcal{M}_1, \mathcal{M}_0$ are subsets of $\mathcal{R}$ which satisfy the following conditions:

D1. $\mathcal{M}_0$ is closed, embedded C^2 submanifold of $\mathcal{R}$ of codimension 0, with boundary (but no corners or singularities),

D2. $\mathcal{M}_1$ belongs to the class $\mathrm{CES}_0^2(\mathcal{R})$, and

D3. $\operatorname{Int}\mathcal{M}_1 \supset \mathcal{M}_0$

We would also like to consider the case where $\mathcal{M}_0$ is allowed to intersect the boundary of $\mathcal{M}_1$, but this introduces some technicalities. We describe how to modify our arguments to accommodate this situation in Remark 6.6.

Let

$$\mathcal{N} = \overline{\mathcal{M}_1 \setminus \mathcal{M}_0}. \tag{6.25}$$

Here the closure is taken in the ambient space $\mathcal{R}$. The fundamental fact in our analysis is that:

Lemma 6.4. *$\mathcal{N}$ belongs to the class* $\mathrm{CES}_0^2(\mathcal{R})$.

Proof. Let $\mathcal{S}$ be the set of singular points of $\mathcal{M}_1$, and let $\mathcal{S}_1 = \mathcal{S} \cap \mathcal{N}$. It is enough to show that $\mathcal{N} \setminus \mathcal{S}_1$ is a codimension 0, embedded C^2 submanifold of $\mathcal{R}$ with boundary. Since $\mathcal{M}_0 \subset \operatorname{Int} \mathcal{M}_1$, we may write

$$\mathcal{N} \setminus \mathcal{S}_1 = (\operatorname{Int} \mathcal{M}_1 \setminus \mathcal{M}_0) \cup \partial_{\mathrm{reg}} \mathcal{M} \cup \partial \mathcal{M}_0, \tag{6.26}$$

and the above union is disjoint.

If $q \in \operatorname{Int} \mathcal{M}_1 \setminus \mathcal{M}_0$, then there exists a neighborhood $U_1 \subset \mathcal{R}$ of q which lies entirely in $\operatorname{Int} \mathcal{M}_1 \setminus \mathcal{M}_0$ and which is C^2-diffeomorphic to $\mathbb{R}^d$.

If $q \in \partial_{\mathrm{reg}} \mathcal{M}$, then there exists a neighborhood $U_2 \subset \mathcal{R}$ of q and a C^2-diffeomorphism $\phi_2 : U_2 \to \mathbb{R}^d$ such that $\phi_2(0) = 0$ and $\phi_2(\mathcal{M}_1 \cap U_2) = \mathbb{H}^d$. Shrinking U if necessary, we may suppose that $U_2 \cap \mathcal{M}_0 = \varnothing$, in which case $\phi_2(\mathcal{N} \cap U_2) = \mathbb{H}^d$.

Finally, if $q \in \partial \mathcal{M}_0$, then there exists a neighborhood $U_3 \subset \mathcal{R}$ of q and a C^2-diffeomorphism $\phi_3 : U_3 \to \mathbb{R}^d$ such that $\phi_3(0) = 0$ and $\phi_3(\mathcal{M}_0 \cap U_2) = \mathbb{H}^d$. Shrinking U_3 if necessary, we may suppose that $U_3 \subset \operatorname{Int} \mathcal{M}_1$, in which case

$$\phi_3(\mathcal{N} \cap U_3) = \Phi_3(U \setminus \operatorname{Int} \mathcal{M}_0) = \mathbb{H}^d_- := \{(x_1, ..., x_d) : x_d \leq 0\} \cong \mathbb{H}^d, \tag{6.27}$$

and we are done. □

Notation introduced in the previous subsection specializes to our choice $\mathcal{N}$ in

(6.25). The boundary of $\mathcal{N}$ decomposes as the disjoint union $\partial\mathcal{M}_1 \cup \partial\mathcal{M}_0$. We denote restrictions of $\mathcal{U}^{\pm}$ to $\partial\mathcal{M}_0$ by

$$\mathcal{V}^{\pm} = \{(q,p) \in \mathcal{U}^{\mp} : q \in \partial\mathcal{M}_0\}. \tag{6.28}$$

Note that our sign conventions are reversed with $\mathcal{V}^{\pm} \subset \mathcal{U}^{\mp}$. This is because we want to think of vectors in $\mathcal{V}^{+}$ as pointing *into* the billiard domain $\mathcal{M}_0$, even though they point *out* of $\mathcal{N}$.

For our definition of a macro-reflection law to make sense, two additional conditions are needed:

D4. $\mathcal{U}^{+}_{\text{fin}}$ is a full-measure subset of $\mathcal{U}^{+}$, with respect to the measure μ_1.

D5. Except on a set of μ_1-measure zero, a point particle in $\mathcal{N}$ which starts with initial state $(q,p) \in \mathcal{V}^{-} \subset \mathcal{U}^{+}$ eventually returns to $\mathcal{V}^{-}$ after only finitely many collisions with $\partial\mathcal{M}_1$.

For various special cases, there are straightforward conditions which guarantee that conditions D4 and D5 hold. Typically these conditions will reduce our setting to the finite measure case, and D5 will follow from the Poincaré Recurrence Theorem. See §6.2.4, below.

As a consequence of Lemma 6.3, condition D4 implies that Φ and its iterates are well-defined on $\mathcal{U}^{+}$, except on a null set.

For $(q,p) \in \mathcal{U}^{+}$, set $M(q,p) = \min\{m \geq 1 : (q,p) \in \mathcal{U}^{+}_{m} \text{ and } \Phi^{m}(q,p) \in \mathcal{U}^{+}_{\partial C}\}$ (and set $M = \infty$ if the minimum does not exist). The *macro-reflection law associated with*

$\mathcal{M}_0$ *and* $\mathcal{M}_1$ is the mapping $P^{\mathcal{M}_1,\mathcal{M}_0} = P^{\mathcal{M}_1} : \mathcal{V}^+ \to \mathcal{V}^+ \cup \{\Delta\}$ defined by

$$P^{\mathcal{M}_1}(q,p) = \begin{cases} R \circ \Phi^{M(q,-p)}(q,-p) & \text{if } M(q,-p) < \infty, \\ \Delta & \text{otherwise,} \end{cases} \tag{6.29}$$

where Δ is a "cemetery state." To understand this definition, consider a point particle moving freely in $\mathcal{M}_1$ and reflecting specularly from $\partial\mathcal{M}_1$. If the point particle starts in initial state $(q,-p) \in \mathcal{V}^-$ and after hitting $\partial\mathcal{M}_1$ some number of times returns to $\partial\mathcal{M}_0$ in state $(q',p') \in \mathcal{V}^+$, then by definition $P^{\mathcal{M}_1}(q,p) = (q',p')$. In (6.29), the map Φ is defined for the billiard in $\mathcal{N}$ and not $\mathcal{M}_1$; thus the reason for post-composing with R is that this "undoes" the reflection from $\partial\mathcal{M}_0$ of the point particle in $\mathcal{N}$.

For $q \in \partial\mathcal{M}_0$, let $k(q) = -n(q)$ denote the outward pointing normal vector field on $\partial\mathcal{M}_0$, let $\tau_{\partial\mathcal{M}_0}$ be surface measure on $\partial\mathcal{M}_0$, and define a measure on $\mathcal{V}^+$ by

$$\Lambda(\mathrm{d}q\mathrm{d}p) = \langle p, k(q)\rangle_q \sigma_q(\mathrm{d}p)\tau_{\partial\mathcal{M}_0}(\mathrm{d}q). \tag{6.30}$$

Clearly the restriction of μ_1 to $\mathcal{V}^-$ pushes forward to Λ under negation $(q,p) \mapsto (q,-p)$. It is also easy to see that reflection R pushes forward μ_1 to Λ.

Proposition 6.5. *(i) There exists a Λ-full measure, open subset $\mathcal{V}_1^+ \subset \mathcal{V}^+$ such that $P^{\mathcal{M}_1} : \mathcal{V}_1^+ \to \mathcal{V}_1^+$ is a C^1 diffeomorphism, and $P^{\mathcal{M}_1} \circ P^{\mathcal{M}_1} = Id_{\mathcal{V}_1^+}$.*

(ii) $P^{\mathcal{M}_1}$ preserves the measure Λ.

Proof. For $m \geq 1$, let $V_m = \{(q,p) \in \mathcal{V}^+ : M(q,-p) = m\}$. For each $(q,p) \in V_m$, $M(q,p) = m < \infty$, and therefore by Lemma 6.3 there is a neighborhood N of (q,p)

such that $\Phi^m : N \to \Phi(N)$ is a diffeomorphism. Since $\mathcal{V}^-$ is an open subset of $\mathcal{U}^+$, $\Phi^{-m}(N \cap \mathcal{V}^-)$ is a neighborhood of (q,p) contained in $-V_m$. Thus V_m is open.

Let $\mathcal{V}_1^+ = \bigcup_{m\geq 1} V_m = \{(q,p) \in \mathcal{V}^+ : M(q,-p) < \infty\}$. By conditions D4 and D5, there exists a full measure set $\mathcal{F} \subset \mathcal{V}^-$ such that

$$\mathcal{F} \cap \bigcap_{m=1}^{\infty} \mathcal{U}_m^+ \subset -\mathcal{V}_1^+. \tag{6.31}$$

Since each of the sets in the intersection on the left has full μ_1-measure in $\mathcal{V}^- \subset \mathcal{U}^+$ (see Lemma 6.3), it follows that $\mathcal{V}_1^+$ has full Λ-measure in $\mathcal{V}^+$.

To see that $P^{\mathcal{M}_1}$ is involutive, let $1 \leq n \leq m$. By Lemma 6.1 and the fact that $R \circ R = \mathrm{Id}$, we have

$$-R \circ \Phi^n \circ -R \circ \Phi^m = (-R \circ \Phi \circ -R)^n \circ \Phi^m = \Phi^{m-n}. \tag{6.32}$$

Thus, on V_m,

$$R \circ \Phi^n \circ -P^{\mathcal{M}_1}(q,p) = -\Phi^{m-n}(q,-p), \tag{6.33}$$

and the first n such that $-\Phi^{m-n}(q,-p) \in \mathcal{V}^+$ is $n = m$. Thus $P^{\mathcal{M}_1}(q,p) \in V_m$, and taking $n = m$ in (6.33), we obtain $P^W \circ P^W(q,p) = (q,p)$.

Finally, since $P^{\mathcal{M}_1} = R \circ \Phi^m \circ -$ on V_m, and negation and R^{-1} pushforward Λ to μ_1, and Φ preserves μ_1 (see Lemma 6.2), we see that $P^{\mathcal{M}_1}$ preserves Λ. □

Remark 6.6. The definition (6.29) of $P^{\mathcal{M}_1}$ can be extended to cases where $\partial\mathcal{M}_1$ is allowed to intersect $\partial\mathcal{M}_0$. The only difficulty is that, if we are not careful, the space $\mathcal{N}$ defined by (6.25) may no longer belong to the class $\mathrm{CES}_0^2(\mathcal{R})$. Problems can occur

if the intersection of $\partial\mathcal{M}_1$ and $\partial\mathcal{M}_0$ is a "fat Cantor set" for example. Nonetheless, $\mathcal{N}$ will belong to the class $\mathrm{CES}_0^2(\mathcal{R})$ if appropriate assumptions are introduced.

Let $\mathcal{M} \supset \mathcal{M}_0$ and assume conditions D1 and D2 are satisfied. Then $\partial\mathcal{M}_0$ decomposes as the disjoint union $A_1 \cup A_2 \cup A_3$, where

$$A_1 = \partial\mathcal{M}_0 \cap \mathcal{N} \cap \operatorname{Int}\mathcal{M}, \qquad A_2 = \partial\mathcal{M}_0 \cap \mathcal{N} \setminus \operatorname{Int}\mathcal{M}, \qquad A_3 = \partial\mathcal{M}_0 \setminus \mathcal{N}. \tag{6.34}$$

We replace condition D3 with

D3'. $\mathcal{M} \supset \mathcal{M}_0$, and the $(d-1)$-dimensional Hausdorff measure of A_2 is zero.

This of course reduces to condition D3 when $\operatorname{Int}\mathcal{M} \supset \mathcal{M}_0$ (because A_2 is the empty set).

Under conditions D1, D2, and D3', $\mathcal{N}$ will belong to the class $\mathrm{CES}_0^2(\mathcal{R})$. To see this, replace the set $\mathcal{S}_1$ in the proof of Lemma 6.4 with $\mathcal{S}_2 := \mathcal{S}_1 \cup A_2$. Then in analogy to (6.26), we may write

$$\mathcal{N} \setminus \mathcal{S}_2 = (\operatorname{Int}\mathcal{M} \setminus \mathcal{M}_0) \cup (\partial_{\mathrm{reg}}\mathcal{M} \setminus \mathcal{M}_0) \cup A_1. \tag{6.35}$$

The rest of the argument proceeds as in the proof of the lemma, except that we replace $\partial_{\mathrm{reg}}\mathcal{M}$ with $\partial_{\mathrm{reg}}\mathcal{M} \setminus \mathcal{M}_0$ and $\partial\mathcal{M}_0$ with A_1.

From the way that we have defined $\mathcal{N}$ in (6.25), we see that $A_1 \subset \partial\mathcal{N}_{\mathrm{reg}} := \partial\mathcal{N} \setminus \mathcal{S}_2$ and $A_2 \subset \partial\mathcal{N}$, but A_3 is not a subset of $\partial\mathcal{N}$. In fact, $A_3 \subset \partial_{\mathrm{reg}}\mathcal{M} \cap \partial\mathcal{M}_0$. We extend the definition of $P^{\mathcal{M}_1}$ so that the point particle reflects specularly from $\partial\mathcal{M}_0$ whenever

it hits A_3. More precisely, we define $P^{\mathcal{M}_1} : \mathcal{V}^+ \to \mathcal{V}^+ \cup \{\Delta\}$ as follows:

$$P^{\mathcal{M}_1}(q,p) = \begin{cases} R \circ \Phi^{M(q,-p)}(q,-p) & \text{if } q \in A_1 \text{ and } M(q,-p) < \infty, \\ -p + 2\langle p, k(q)\rangle_q k(q) & \text{if } q \in A_3, \\ \Delta & \text{otherwise.} \end{cases} \tag{6.36}$$

This reduces to (6.29) when $A_1 = \partial\mathcal{M}_0$ (in which case $A_3 = \varnothing$).

Let

$$\mathcal{V}^{\pm}_{A_i} = \{(q,p) \in \mathcal{V}^{\pm} : q \in A_i\}, \quad i = 1,2,3. \tag{6.37}$$

Under conditions D1, D2, D3', D4, and D5, Proposition 6.5 still holds word for word. The proof gets modified as follows: We let $\mathcal{V}^+_1 = \mathcal{V}^+_{A_3} \cup \bigcup_{m=1}^{\infty} V_m$, where now $V_m := \{(q,p) \in \mathcal{V}^+_{A_1} : M(q,-p) = m\}$. For essentially the same reason as before $\mathcal{V}^+_1$ is open, and $\bigcup_{m=1}^{\infty} V_m$ is a Λ-full measure subset of $\mathcal{V}^+_{A_1}$, and by condition D3', $\mathcal{V}^+_{A_1} \cup \mathcal{V}^+_{A_3}$ is a full measure subset of $\mathcal{V}^+_1$. Therefore $\mathcal{V}^+$ is a full measure subset of $\mathcal{V}^+_1$. The rest of the proposition follows simply by noting that specular reflection from $\partial\mathcal{M}_0$ is involutive and preserves Λ.

6.2.2 Rough reflections

From now on we assume that $\mathcal{M}_1$ satisfies conditions D1, D2, D3', D4, and D5, stated above.

The macro-reflection law $P^{\mathcal{M}_1}$ is associated with the deterministic Markov kernel on $\mathcal{V}^+_1$,

$$\mathbb{P}^{\mathcal{M}}(q,p;\mathrm{d}q'\mathrm{d}p') := \delta_{P^{\mathcal{M}_1}(q,p)}(\mathrm{d}q'\mathrm{d}p'). \tag{6.38}$$

We call a Markov kernel $\mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')$ on $\mathcal{V}^+$ a *rough collision law* if there exists a sequence $\mathcal{M}_i \supset \mathcal{M}_0$ (satisfying hypotheses 1' and 2-5) and a sequence of positive numbers $\epsilon_i \to 0$ such that

$$\partial\mathcal{M}_i \subset \overline{N}_{\epsilon_i}(\partial\mathcal{M}_0) := \{q \in \mathcal{R} : \mathrm{dist}(q, \partial\mathcal{M}_0) \le \epsilon_i\} \tag{6.39}$$

and such that the following limit holds:

$$\mathbb{P}^{\mathcal{M}_i}(q,p;\mathrm{d}q'\mathrm{d}p')\Lambda(\mathrm{d}q\mathrm{d}p) \to \mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')\Lambda(\mathrm{d}q\mathrm{d}p), \tag{6.40}$$

weakly in the space of measures on $\mathcal{V}^+ \times \mathcal{V}^+$.

The limit (6.40) means that for any function $h \in C_c(\mathcal{V}^+ \times \mathcal{V}^+)$,

$$\begin{aligned} &\lim_{i\to\infty} \int_{\mathcal{V}^+\times\mathcal{V}^+} h(q,p;\mathrm{d}q'\mathrm{d}p')\mathbb{P}^{\mathcal{M}_i}(q,p;\mathrm{d}q'\mathrm{d}p')\Lambda(\mathrm{d}q\mathrm{d}p) \\ &\qquad = \int_{\mathcal{V}^+\times\mathcal{V}^+} h(q,p;\mathrm{d}q'\mathrm{d}p')\mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')\Lambda(\mathrm{d}q\mathrm{d}p). \end{aligned} \tag{6.41}$$

Remark 6.7. Since with respect to the uniform norm $C_c^\infty(\mathcal{V}^+ \times \mathcal{V}^+)$ is dense in $C_c(\mathcal{V}^+ \times \mathcal{V}^+)$, and the tensor product $C_c^\infty(\mathcal{V}^+) \otimes C_c^\infty(\mathcal{V}^+)$ is dense in $C_c^\infty(\mathcal{V}^+ \times \mathcal{V}^+)$, it is sufficient to verify (6.41) for functions h of form $f(q,p)g(q',p')$, where $f, g \in C_c^\infty(\mathcal{V}^+)$.

Proposition 6.8. *A rough reflection law $\mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')$ is symmetric with respect to the measure Λ, in the sense that, for any $h \in C_c(\mathcal{V}^+ \times \mathcal{V}_+)$,*

$$\begin{aligned} &\int_{\mathcal{V}^+\times\mathcal{V}^+} h(q,p,q',p')\mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')\Lambda(\mathrm{d}q\mathrm{d}p) \\ &\qquad = \int_{\mathcal{V}^+\times\mathcal{V}^+} h(q',p',q,p)\mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')\Lambda(\mathrm{d}q\mathrm{d}p). \end{aligned} \tag{6.42}$$

Proof. Suppose $\mathbb{P}^{\mathcal{M}_i} \to \mathbb{P}$, as in (6.40). We claim that an equality of form (6.42) holds if we replace $\mathbb{P}$ with $\mathbb{P}^{\mathcal{M}_i}$. This is equivalent to showing that

$$\int_{\mathcal{V}^+} h(P^{\mathcal{M}_i}(q,p),q,p)\Lambda(\mathrm{d}q\mathrm{d}p) = \int_{\mathcal{V}^+} h(q,p,P^{\mathcal{M}_i}(q,p))\Lambda(\mathrm{d}q\mathrm{d}p). \tag{6.43}$$

But making the change of variables $(q',p') = P^{W_i}(q,p)$, the above equality is a consequence of the fact that $P^{\mathcal{M}_i}$ is an involution which preserves Λ. The result follows by taking the weak limit as $i \to \infty$. □

Corollary 6.9. *A rough collision law $\mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')$ preserves the measure Λ in the sense that, for all $f \in C_c(\mathcal{V}^+)$,*

$$\int_{\mathcal{V}^+} f(q',p')\mathbb{P}(q,p;\mathrm{d}q'\mathrm{d}p')\Lambda(\mathrm{d}q\mathrm{d}p) = \int_{\mathcal{V}^+} f(q,p)\Lambda(\mathrm{d}q\mathrm{d}p). \tag{6.44}$$

Proof. Let $h(q,p,q',p') \uparrow f(q,p)$ in (6.42). □

6.2.3 Pseudometric topology

In (6.41), there is a sense in which $\mathbb{P}^{\mathcal{M}_i}$ converges to $\mathbb{P}$ as a limit with respect to a pseudometric topology. To describe this topology, let $X = T^1\mathcal{R}$, let $\mathcal{X}$ denote the Borel σ-algebra on X, and let $\mathcal{G}$ denote the set of all Markov kernels on $(X,\mathcal{X})$, i.e. the set of all functions $\mathbb{G} : X \times \mathcal{X} \to \mathbb{R}$ satisfying: (i) for each $x \in X$, $A \mapsto \mathbb{G}(x,A)$ is a Borel probability measure on X, and (ii) for each $A \in \mathcal{X}$, $x \mapsto \mathbb{G}(x,A)$ is a measurable function.

Equip $C_c(X \times X)$ with the uniform topology. Since $X \times X$ is locally compact and σ-compact, the space $C_c(X \times X)$ is separable. Let $\{f_i\}_{i\geq 1} \subset C_c(X \times X)$ be a

countable dense subset, and let $\mathcal{N}$ denote the space of all signed Borel measures on $X \times X$ which are finite on compact sets. (Since $X \times X$ is a locally compact Hausdorff space, this also implies that the measures in $\mathcal{N}$ are Radon – see [17, Theorem 7.8].) Equip $\mathcal{N}$ with the metric

$$d(\mu_1, \mu_2) := \sum_{i=1}^{\infty} 2^{-i} \frac{|\int f_i \, \mathrm{d}(\mu_1 - \mu_2)|}{1 + |\int f_i \, \mathrm{d}(\mu_1 - \mu_2)|}, \qquad \mu_1, \mu_2 \in \mathcal{N}. \tag{6.45}$$

With respect to this metric, $\mu_k \to \mu$ in $\mathcal{N}$ if and only if for all $i \geq 1$, $\lim_{k\to\infty} \int f_i \mathrm{d}\mu_k = \int f_i \mathrm{d}\mu$ if and only if $\lim_{k\to\infty} \int f \mathrm{d}\mu_k = \int f \mathrm{d}\mu$ for all $f \in C_c(X \times X)$. In particular, d restricted to the dual space $C_c(X \times X)^*$ induces the weak topology.

Given a Borel measure ν on X which is finite on compact sets and an element $\mathbb{G} \in \mathcal{G}$, let $\mathbb{G}\nu$ denote the Borel measure on $X \times X$ defined by $\int f \, \mathrm{d}(\mathbb{G}\nu) = \int (\int f(x,y)\mathbb{G}(y, \mathrm{d}x))\nu(\mathrm{d}y)$, $f \in C_c(X \times X)$. We observe that $\mathbb{G}\nu \in \mathcal{N}$; in particular if K is any compact subset of X, then since $\mathbb{G}(x, \cdot)$ is a probability measure

$$\mathbb{G}\nu(K \times X) = \nu(K) < \infty. \tag{6.46}$$

The measure ν together with the metric d induce a pseudometric on $\mathcal{G}$ defined by

$$d_{\mathcal{G}}^{\nu}(\mathbb{G}_1, \mathbb{G}_2) = d(\mathbb{G}_1\nu, \mathbb{G}_2\nu). \tag{6.47}$$

With respect to the induced pseudometric topology, $\mathbb{G}_j \to \mathbb{G}$ if and only if $\mathbb{G}_j\nu \to \mathbb{G}\nu$ in duality to $C_c(X \times X)$. Hence taking $\nu = \Lambda$ yields the desired sense of convergence.

Note that $d_{\mathcal{G}}^{\nu}$ fails to satisfy the non-degeneracy condition of a metric. If

$d^{\nu}_{\mathcal{G}}(\mathbb{G}_1, \mathbb{G}_2) = 0$, the Markov kernels $\mathbb{G}_1(\cdot, \mathrm{d}x)$ and $\mathbb{G}_2(\cdot, \mathrm{d}x)$ are still allowed to disagree on a ν-null set.

6.2.4 Some special cases

Here we discuss how conditions D1, D2, D3 (or D3'), D4, and D5 are verified for particular cases. In all of our examples, the ambient space is $\mathcal{R} = \mathbb{R}^d$ or $\mathbb{T}^{d-1} \times \mathbb{R}$ with the Euclidean metric.

Example 1: Bounded, convex bodies. Let B_0 be a closed, bounded, strictly convex subset of $\mathbb{R}^d$, and let B be some closed, connected subset of B_0 such that $B \subset \operatorname{Int} B_0$. Assume that both B_0 and B have smooth or piecewise smooth boundary; the essential conditions are that $\mathcal{M}_0 := \mathbb{R}^d \setminus \operatorname{Int} B_0$ should be an embedded C^2 submanifold of $\mathbb{R}^d$ with boundary, and $\mathcal{M}_1 := \mathbb{R}^d \setminus \operatorname{Int} B$ should belong to the class $\mathrm{CES}^2_0(\mathbb{R}^d)$. Then conditions D1-D3 of §6.2.1 are clearly satisfied.

Condition D4 holds because a linear trajectory of a point particle must eventually leave the bounded region B_0. Condition D5 holds because, by boundedness, the invariant measure on $\mathcal{U}^-$ (the set of pairs $(y, w) \in \partial B_0 \times \mathbb{S}^2$ such that w points out of B_0) is finite. Thus we may apply the Poincaré Recurrence Theorem to conclude that the point particle returns almost surely.

Hence one can define rough reflection laws on B_0 by taking weak limits, as in §6.2.2. Plakhov's work on scattering laws, reviewed in §1.2.7, falls within this setting.

Example 2: Half-space billiards. Consider the following subset of $\mathbb{R}^d$:

$$W_0 = \{(x_1, ..., x_d) : x_d \leq 0\}. \tag{6.48}$$

Let W be a subset of W_0 such that

$$\{(x_1, ..., x_d) : x_d \leq -1\} \subset W \subset \{(x_1, ..., x_d) : x_d < 0\}. \tag{6.49}$$

Assume that W has a piecewise smooth boundary; at minimum $\mathcal{M}_1 := \mathbb{R}^d \setminus W$ should belong to the class $\mathrm{CES}_0^2(\mathbb{R}^d)$. Let $\mathcal{M}_0 = \{(x_1, ..., x_d) : x_d \geq 0\}$. Then conditions D1-D3 are satisfied.

To verify condition D4, note that the only way for a billiard trajectory in $\mathcal{N} = \overline{\mathcal{M}_1 \setminus \mathcal{M}_0}$ to escape to infinity without returning to $\mathcal{M}_0$ or hitting $\partial\mathcal{M}_1$ is if the trajectory is parallel to plane $x_d = 0$. The set of states (y, w) such that w is parallel to this plane is a Λ-null.

For condition D5 to hold, additional assumptions must be imposed. One condition which implies D5 is the following:

D5a. There exist linearly independent vectors $v_1, \ldots, v_{d-1}$ spanning the plane $x_d = 0$ such that W is invariant under translation by the vectors v_i, i.e. $W + v_i = W$ for $1 \leq i \leq d-1$.

Indeed, in this case we may reduce to the quotient space obtained by identifying points which are translates of each other by integer combinations of the v_i. We can thereby view the billiard domain as a subset of the ambient space $\mathbb{T}^{d-1} \times \mathbb{R}$. The space $\mathcal{N}$ is a compact subset of $\mathbb{T}^{d-1} \times \mathbb{R}$, since it is bounded in the direction of the x_d-axis. Consequently, the invariant measure is finite, and as in the previous case we can use Poincaré's Theorem to verify condition D5.

The reflections laws considered in §1.2.2 are rough reflections in an upper half-

space billiard of dimension $d = 2$. The rough collision laws defined in this book are examples of rough reflection laws in an upper half-space of dimension $d = 3$.

Extensions. We can generalize both of the above examples to allow $\partial\mathcal{M}_1$ to intersect $\partial\mathcal{M}_0$ (thus B can intersect ∂B_0 in Example 1, and W can intersect the plane $x_d = 0$ in Example 2). However, some assumption must be imposed to ensure that condition D3' holds. One simple condition is the following: *The* $(d-1)$*-dimensional Hausdorff measure of* $\mathcal{M}_1 \cap \partial\mathcal{M}_0$ *is zero.* Since $A_2 \subset \mathcal{M}_1 \cap \partial\mathcal{M}_0$, this implies condition D3'.

In this book, both the configuration space $\mathcal{M}$ and its cylindrical approximation $\mathcal{M}_{\text{cyl}}$ satisfy condition D3' (with $\mathcal{M}_0 = \{(x_1, x_2, \alpha) : x_2 \geq 0\}$). In the former case, it is because $\mathcal{M}$ satisfies the condition stated in the previous paragraph. In the latter case, the 2-dimensional Hausdorff measure of A_2 is zero because it is a countable union of lines. (See Propositions 4.8 and 4.11 respectively.)

In the upper half-space case, if we allow W to touch the plane $x_d = 0$, then there is another condition we can impose which implies condition D5.

D5b. There exists a countable collection $\{R_j\}_{j\geq 1}$ of disjoint bounded open subsets of $\mathbb{R}^d$ such that $\operatorname{Int} W_0 \setminus W = \bigcup_{j=1}^{\infty} R_j$.

If this condition holds, then the point particle will get trapped in one of the regions R_j when it enters $\mathcal{N}$. By boundedness, the restriction of the invariant measure to any one of the regions R_j is finite, and we can once again apply the Poincaré Recurrence Theorem.

7 Ergodicity of the two-wall system

7.1 Setup

7.1.1 Configuration space and dynamics

In this section, we consider a rough unit disk D which moves freely in the interior of the strip

$$\{(x_1, x_2) : -1 \leq x_2 \leq r + 1\}, \qquad \text{where } r > 0, \tag{7.1}$$

and collides with the two walls which bound the strip. Upon making contact with either of the two walls, the trajectory of the disk is redirected back into the strip according to rough collision dynamics. Locally, during a single collision event, the projection of the phase space velocity onto the "rolling velocity" is conserved, as discussed in the introduction. But since for each of the two walls there is a different associated "rolling velocity," no corresponding quantity is conserved globally. In what follows, we aim to understand the ergodic behavior of this system.

The configuration space of the system is

$$\mathcal{M} = \{(x_1, x_2, \alpha) \in \mathbb{R}^3 : 0 \leq x_2 \leq r\}. \tag{7.2}$$

This is the set of all physically realizable states (x_1, x_2, α), where (x_1, x_2) is the center of mass of the disk and α is the counterclockwise angular orientation of the disk. Note that x_2 is bounded between 0 and r because the radius of the disk is 1.

The phase space is

$$\mathcal{M} \times \mathbb{S}^2, \tag{7.3}$$

where $\mathbb{S}^2$ is the unit sphere with respect to the kinetic energy inner product defined in §1.3.4.

To be more precise about the dynamical evolution, let $(y(t), w(t))$ denote the state of the disk at time $t \in \mathbb{R}$, and let $(y(t_-), w(t_-)) = \lim_{s \uparrow t}(y(s), w(s))$ and $(y(t_+), w(t_+)) = \lim_{s \downarrow t}(y(s), w(s))$. In the interior of $\mathcal{M}$, the trajectory $y(t)$ evolves linearly in time while the velocity $w(t)$ stays constant. Let

$$\mathbf{P}_0 = \{(x_1, x_2, \alpha) : x_2 = 0\} \quad \text{and} \quad \mathbf{P}_1 = \{(x_1, x_2, \alpha) : x_2 = r\}, \tag{7.4}$$

denote the two planes which form the boundary of $\mathcal{M}$. Note that the plane $\mathbf{P}_0$ coincides with the plane $\mathbf{P}$ defined in the introduction in §1.3.5. Rough collision dynamics means that there are Markov kernels $\mathbb{K}_0(y, w; \mathrm{d}y'\mathrm{d}w')$ and $\mathbb{K}_1(y, w, \mathrm{d}y'\mathrm{d}w')$ on $\mathbf{P}_0 \times \mathbb{S}^2_+$ which determine the collision dynamics as follows:

Assume t^* is a collision time.

- If $y(t^*) \in \mathbf{P}_0$, then the post collision state is distributed as

$$(y(t^*_+), w(t^*_+)) \sim \mathbb{K}_0(y(t^*_-), -w(t^*_-); \mathrm{d}y'\mathrm{d}w'). \tag{7.5}$$

- When the point particle hits the plane $\mathbf{P}_1$, the post-collision state is given by $\mathbb{K}_1$ conjugated with a 180 degree rotation taking $\mathbf{P}_1$ to $\mathbf{P}_0$. More precisely, let

$\overline{R} : \mathbf{P}_1 \times \mathbb{S}^2_- \to \mathbf{P}_0 \times \mathbb{S}^2_+$ be defined by

$$\overline{R}(x_1, x_2, \alpha, v_1, v_2, \omega) = (x_1, 0, \alpha, -v_1, -v_2, \omega). \tag{7.6}$$

If $y(t^*) \in \mathbf{P}_1$, then the post collision state is distributed as

$$(y(t^*_+), w(t^*_+)) \sim \overline{R}^{-1}_{\#} \mathbb{K}_1(\overline{R}(y(t^*_-), -w(t^*_-)); \mathrm{d}y' \mathrm{d}w'). \tag{7.7}$$

Here $\overline{R}^{-1}_{\#} \mathbb{K}_1(\overline{R}(y(t^*_-), -w(t^*_-)); \mathrm{d}y' \mathrm{d}w')$ denotes the pushforward under $\overline{R}^{-1}$ of $\mathbb{K}_1(R(y(t^*_-), -w(t^*_-)); \cdot)$, viewed as a measure on $\mathbf{P}_0 \times \mathbb{S}^2_+$.

We assume that $\mathbb{K}_0$ and $\mathbb{K}_1$ are rough collision laws, as defined in §1.3.5. In particular, this means that, with respect to the coordinates (y_1, y_3, θ, ψ) defined in §1.3.6, $\mathbb{K}_0$ and $\mathbb{K}_1$ have product decompositions of form expressed in Theorem 1.31. That is,

$$\mathbb{K}_0(y_1, y_3, \theta, \psi; \mathrm{d}y_1' \mathrm{d}y_3' \mathrm{d}\theta' \mathrm{d}\psi') = \delta_{(y_1, y_3)}(\mathrm{d}y_1' \mathrm{d}y_3') \widetilde{\mathbb{P}}_0(\theta, \mathrm{d}\theta') \delta_{\pi - \psi}(\mathrm{d}\psi'), \tag{7.8}$$

$$\mathbb{K}_1(y_1, y_3, \theta, \psi; \mathrm{d}y_1' \mathrm{d}y_3' \mathrm{d}\theta' \mathrm{d}\psi') = \delta_{(y_1, y_3)}(\mathrm{d}y_1' \mathrm{d}y_3') \widetilde{\mathbb{P}}_1(\theta, \mathrm{d}\theta') \delta_{\pi - \psi}(\mathrm{d}\psi'), \tag{7.9}$$

where $\widetilde{\mathbb{P}}_0(\theta, d\theta')$ and $\widetilde{\mathbb{P}}_1(\theta, d\theta')$ are some Markov kernels on $(0, \pi)$ which are symmetric with respect to the measure $\sin\theta d\theta$.

7.1.2 Spherical coordinates

In §1.3.6, we introduced the spherical coordinates $(\theta, \psi) = G : [0, 2\pi) \times (0, \pi) \to \mathbb{S}^2 \setminus \{\pm\chi\}$. Since projection of velocity onto the vector χ was conserved by the

collision dynamics, this was a natural choice of coordinates. In the case of a disk bouncing between two walls, there is a different conserved quantity associated with each wall. Namely, if we define vectors

$$\chi_0 = (m+J)^{-1/2}(-1,0,1)^T, \qquad \chi_1 = (m+J)^{-1/2}(1,0,1)^T, \tag{7.10}$$

then collisions with the plane $\mathbf{P}_0$ conserve the quantity $\langle w, \chi_0 \rangle$ while collisions with the plane $\mathbf{P}_1$ conserve the quantity $\langle w, \chi_1 \rangle$. This follows from the product decompositions (7.8) and (7.9). Note that $\chi_0 = -\chi$ as defined in section, while χ_1 is obtained by rotating χ_0 by 180 degrees about the α-axis.

These considerations motivate introducing two systems of spherical coordinates, one with north pole at χ_0 and one with north pole at χ_1. For $i = 0, 1$, let

$$\chi_i^\perp = (m^{-1} + J^{-1})^{-1/2}(-m^{-1}, 0, -J^{-1}). \tag{7.11}$$

Also, recall that $\widehat{e}_2 := m^{-1/2} e_2 = (0, m^{-1/2}, 0)$ denotes the unit normal in the direction of the $+x_2$-axis. For each i, the triple $(\chi_i, \widehat{e}_2, \chi_i^\perp)$ forms a positively oriented orthonormal frame (as usual, with respect to the metric $\langle \cdot, \cdot \rangle$), with the pair of vectors $(\chi_i, \chi_i^\perp)$ spanning the $x_1\alpha$-plane. For $i = 0, 1$, we define spherical coordinates $G_i : [0, 2\pi) \times (0, \pi) \to \mathbb{S}^2 \setminus \{\pm\chi_0\}$ by

$$G_i(\theta_i, \psi_i) = (\cos\theta_i \sin\psi_i)\chi_i^\perp + (\sin\theta_i \sin\psi_i)\widehat{e}_2 + (\cos\psi_i)\chi_i. \tag{7.12}$$

As a matter of bookkeeping, we always denote coordinates with respect to the map-

ping G_0 by $(\theta_0, \psi_0), (\theta_0', \psi_0'), (\theta_0^*, \psi_0^*)$, etc. Similarly, we denote coordinates with respect to the mapping G_1 by $(\theta_1, \psi_1), (\theta_1', \psi_1'), (\theta_1^*, \psi_1^*)$, etc.

The coordinate mapping G_0 is the same as the mapping G defined in §1.3.6 restricted to velocity coordinates, except that the roles of the north and south pole have been swapped (this is slightly reduces the number of minus signs in the arguments to follow). The mapping G_1 is obtained by composing G_0 with a 180 degree rotation about the α-axis.

We introduce one additional system of spherical coordinates with north pole at $\widehat{e}_2$, namely $\widetilde{G} : [0, 2\pi) \times (0, \pi) \to \mathbb{S}^2$, defined by

$$\widetilde{G}(\widetilde{\theta}, \widetilde{\psi}) = (\cos\widetilde{\theta}\sin\widetilde{\psi})\widehat{e}_3 + (\sin\widetilde{\theta}\sin\widetilde{\psi})\widehat{e}_1 + (\cos\widetilde{\psi})\widehat{e}_2, \tag{7.13}$$

where $\widehat{e}_1 = m^{-1/2}e_1 = (m^{-1/2}, 0, 0)$ and $\widehat{e}_3 = J^{-1/2}e_3 = (0, 0, J^{-1/2})$. This coordinate system will be used in just one proof.

7.1.3 Markov kernel on the velocity space

Let $0 \le t_0 < t_1 < t_2 < \cdots$ be the sequence of times when the disk hits one of the two walls. Since the collision laws $\mathbb{K}_0$ and $\mathbb{K}_1$ act trivially on spatial coordinates, the evolution of the configuration $\{y(t), t \ge 0\}$ is completely determined by the discrete velocity evolution $w_i := w(t_{i-})$, $i \ge 0$, in $\mathbb{S}^2$. Consider Markov kernels on $\mathbb{S}^2_+$ defined by

$$\mathbb{K}_i^{\mathrm{vel}}(w; \mathrm{d}w') := \int_{y' \in \mathbf{P}_0} \mathbb{K}_i(y, w; \mathrm{d}y'\mathrm{d}w'), \qquad i = 0, 1. \tag{7.14}$$

(Note that the right-hand side does not depend on y. This follows from (7.8) and (7.9).) Let $\widehat{R}(v_1, v_2, \omega) = (-v_1, -v_2, \omega)$. The discrete velocity evolution $\{w_i\}_{i\geq 0}$ is a Markov process on $\mathbb{S}^2$ whose transition kernel is

$$\widehat{\mathbb{K}}(w, \mathrm{d}w') := \begin{cases} \mathbb{K}_0^{\mathrm{vel}}(-w; \mathrm{d}w') & \text{if } w \in \mathbb{S}^2_-, \\ \widehat{R}_{\#}^{-1}\mathbb{K}_1^{\mathrm{vel}}(\widehat{R}(-w), \mathrm{d}w') & \text{if } w \in \mathbb{S}^2_+. \end{cases} \tag{7.15}$$

We let $\widehat{\mathbb{K}}_-$ and $\widehat{\mathbb{K}}_+$ denote the restrictions of $\widehat{\mathbb{K}}$ to $\mathbb{S}^2_-$ and $\mathbb{S}^2_+$, respectively. Thus, for each $w \in \mathbb{S}^2_\pm$, $\widehat{\mathbb{K}}_\pm(w, \mathrm{d}w')$ is a measure on $\mathbb{S}^2_\mp$.

From the coordinate representations (7.8) and (7.9) we deduce that the restricted kernels $\widehat{\mathbb{K}}_-$ and $\widehat{\mathbb{K}}_+$ may be represented in the spherical coordinates (θ_0, ψ_0) and (θ_1, ψ_1) respectively as follows:

$$\begin{aligned} \widehat{\mathbb{K}}_-(\theta_0, \psi_0; \mathrm{d}\theta_0' \mathrm{d}\psi_0') &= \widetilde{\mathbb{P}}_0(\theta_0 - \pi, \mathrm{d}\theta_0') \times \delta_{\psi_0}(\mathrm{d}\psi_0'), \\ \widehat{\mathbb{K}}_+(\theta_1, \psi_1; \mathrm{d}\theta_1' \mathrm{d}\psi_1') &= \widetilde{\mathbb{P}}_1(\theta_1 - \pi, \mathrm{d}\theta_1') \times \delta_{\psi_1}(\mathrm{d}\psi_1'). \end{aligned} \tag{7.16}$$

It is not so simple to express $\widehat{\mathbb{K}}_-$ in the coordinates (θ_1, ψ_1) or $\widehat{\mathbb{K}}_+$ in the coordinates (θ_0, ψ_0) however.

7.1.4 Invariant measures

Recall from §1.2.2 and §1.3.5 the invariant measures $\Lambda^1(\mathrm{d}x\mathrm{d}\theta) = \sin\theta \mathrm{d}x\mathrm{d}\theta$ on $\mathbb{R} \times (0, \pi)$ and $\Lambda^2(\mathrm{d}y, \mathrm{d}w) = \langle w, \widehat{e}_2\rangle \mathrm{d}y\sigma(\mathrm{d}w)$ on $\mathbf{P}_0 \times \mathbb{S}^2_+$. Here we introduce the normalized velocity marginals

$$\widehat{\Lambda}^1(\mathrm{d}\theta) = \frac{1}{2}|\sin\theta|\mathrm{d}\theta \qquad \text{on } (0, 2\pi), \tag{7.17}$$

and

$$\widehat{\Lambda}^2(\mathrm{d}w) = \frac{1}{2\pi}|\langle w, e_2\rangle|\sigma(\mathrm{d}w) \qquad \text{on } \mathbb{S}^2, \tag{7.18}$$

where as before σ is surface measure on the unit sphere. A couple of remarks are in order:

- Note that we have extended the domains of these measures to $(0, 2\pi)$ and $\mathbb{S}^2$ respectively. The reason is that $\widehat{\mathbb{K}}$ has been defined to act on the whole sphere $\mathbb{S}^2$. In the dynamics determined by $\widehat{\mathbb{K}}$, the incoming and outgoing velocities live in separate hemispheres (namely, $\mathbb{S}^2_-$ and $\mathbb{S}^2_+$, respectively, in the case of collisions with the plane $\mathbf{P}_0$, and $\mathbb{S}^2_+$ and $\mathbb{S}^2_-$, respectively, in the case of collisions with the plane $\mathbf{P}_1$); whereas the Markov kernel $\mathbb{K}$ defined in 1.3.5 acts on $\mathbf{P}_0 \times \mathbb{S}^2_+$ by mapping the *negative* of the incoming velocity to the outgoing velocity.

- The normalizing factor in (7.18) guarantees that $\widehat{\Lambda}^2$ is a probability measure on $\mathbb{S}^2$. The normalizing factor in (7.17) guarantees that the *restriction* of $\widehat{\Lambda}^1$ to $(0, \pi)$ is a probability measure; thus the total mass of $\widehat{\Lambda}^1$ is 2. The reason for this choice of normalization is that the Markov kernels $\widetilde{\mathbb{P}}_i$ are still defined on $(0, \pi)$. The measure $\widehat{\Lambda}^1|_{(0,\pi)}$ is the stationary probability measure for the Markov process determined $\widetilde{\mathbb{P}}_i$. The reason we extend the definition of $\widetilde{\Lambda}^1$ to $(0, 2\pi)$ is only because of the convenient identity (7.21), below.

By assumption $\widetilde{\mathbb{P}}_i(\theta_i, \mathrm{d}\theta_i')$ is symmetric with respect to the measure $\widehat{\Lambda}^1|_{(0,\pi)}$ for $i = 0, 1$. It follows from Proposition 1.26 that the Markov kernels $\mathbb{K}_i^{\mathrm{vel}}$, $i = 0, 1$, are symmetric with respect to the restriction of the measure $\widehat{\Lambda}^2$ to the hemisphere $\mathbb{S}^2_+$. By observing that $\widehat{\Lambda}^2$ is invariant under the mapping $\widehat{R}$, we conclude that the Markov

kernel $\widehat{\mathbb{K}}$ is symmetric with respect to the measure $\widehat{\Lambda}^2$. That is, for any test functions f and g,

$$\int_{\mathbb{S}^2\times\mathbb{S}^2} f(w')\widehat{\mathbb{K}}(w,\mathrm{d}w')g(w)\widehat{\Lambda}^2(\mathrm{d}w) = \int_{\mathbb{S}^2\times\mathbb{S}^2} g(w')\widehat{\mathbb{K}}(w,\mathrm{d}w')f(w)\widehat{\Lambda}^2(\mathrm{d}w). \tag{7.19}$$

For $i = 0, 1$, the measure $\widehat{\Lambda}^2$ has the following coordinate representation:

$$\widehat{\Lambda}^2(\mathrm{d}\theta_i\mathrm{d}\psi_i) = \frac{1}{2\pi}|\sin\theta_i|\sin^2\psi_i\mathrm{d}\theta_i\mathrm{d}\psi_i. \tag{7.20}$$

To understand where this comes from, note that the spherical volume form on $\mathbb{S}^2$ has coordinate representation $\sin\psi_i\mathrm{d}\theta_i\mathrm{d}\psi_i$, while $\langle w, e_2\rangle = v_2 = \sin\theta_i\sin\psi_i$.

We will make frequent use of the following identity:

$$\widehat{\Lambda}^2(\mathrm{d}\theta_i\mathrm{d}\psi_i) = \frac{1}{\pi}\widehat{\Lambda}^1(\mathrm{d}\theta_i)\sin^2\psi_i\mathrm{d}\psi_i. \tag{7.21}$$

7.1.5 Ergodicity

The main result of this section describes conditions under which the measure $\widehat{\Lambda}^2$ is ergodic with respect to the Markov transition kernel $\widehat{\mathbb{K}}$. Let

$$\gamma = 2\arccos\left(\frac{1}{1+J/m}\right). \tag{7.22}$$

This is twice the angle between χ_0 and χ_1, where, as usual, angles are with respect to the kinetic energy metric $\langle\cdot,\cdot\rangle$.

Theorem 7.1. *Assume that γ/π is irrational. The measure $\widehat{\Lambda}^2$ is ergodic with respect*

to $\widehat{\mathbb{K}}$ *if and only if the following two conditions hold:*

(E1) there exists an ergodic measure μ *on* $\mathbb{S}^2$ *which is not singular with respect to* $\widehat{\Lambda}^2$*, and*

(E2) for $i = 0, 1$*, the measure* $\widehat{\Lambda}^1|_{(0,\pi)}(\mathrm{d}\theta) = \frac{1}{2}\sin\theta d\theta$ *is ergodic with respect to the Markov transition kernel* $\widetilde{\mathbb{P}}_i(\theta, \mathrm{d}\theta')$*.*

The characterization provided by this theorem is not completely satisfactory in that condition (E1) may in general be difficult to verify. The following theorem provides a condition which implies (E1).

Theorem 7.2. *Assume that for* $\widehat{\Lambda}^1$*-almost every* $\theta \in (0, \pi)$*, the measure* $\widetilde{\mathbb{P}}_i(\theta, \mathrm{d}\theta')$ *on* $(0, \pi)$ *is not singular with respect to* $\widehat{\Lambda}^1$*, for* $i = 0, 1$*. Then there exists an ergodic measure* μ *on* $\mathbb{S}^2$ *which is absolutely continuous with respect to* $\widehat{\Lambda}^2$*.*

By way of illustration, recall the rough collisions from walls with microstructure consisting of elliptical arcs (see 2.3.3). We saw in this case that $\mathbb{K} = \delta_{(y_1,y_2)} \times \widetilde{\mathbb{P}}(\theta, \mathrm{d}\theta') \times \delta_{\pi-\psi}$, where $\widetilde{\mathbb{P}}$ describes point particle reflections from a microstructure consisting of circular arcs. From the formulas (2.21) and (2.24), we see that the hypothesis of Theorem 7.2 is satisfied.

We hope in a future project to provide more general conditions under which condition (i) in Theorem 7.1 holds.

The linear and angular velocity marginals of $\widehat{\Lambda}^2$ can be obtained by a straightforward calculation which we carry out in (7.2). Before proceeding with the proofs of the two theorems above, we provide a brief review of relevant notions from ergodic

theory in §7.3. We introduce in §7.4 the *Abel transform*, an integral transform which plays an important role in our proofs. We prove Theorems 7.1 and 7.2 in §7.5.

7.2 Marginals

Theorems 7.1 and 7.2 establish a kind of universality for rough collisions. Regardless of the shape of the microscopic rough features on the walls with which the disk collides, the unique non-singular ergodic measure associated with the system is $\widehat{\Lambda}^2$, under fairly general assumptions. Here we will see that the marginal of each component of linear and angular velocity is given by either a semicircle law or an absolute value function.

The marginals may be computed as follows. We use the spherical coordinates $(\widetilde{\theta}, \widetilde{\psi})$ given by (7.13). In these coordinates, $\sigma(\mathrm{d}\widetilde{\theta}\mathrm{d}\widetilde{\psi}) = \sin\widetilde{\psi}\mathrm{d}\widetilde{\theta}\mathrm{d}\widetilde{\psi}$ and $\langle w, \widehat{e}_2\rangle = \cos\widetilde{\psi}$. Thus

$$\widehat{\Lambda}^2(\mathrm{d}\widetilde{\theta}\mathrm{d}\widetilde{\psi}) = \frac{1}{2\pi}|\cos\widetilde{\psi}|\sin\widetilde{\psi}\mathrm{d}\widetilde{\theta}\mathrm{d}\widetilde{\psi}. \tag{7.23}$$

Let $\widehat{w} \in \mathbb{S}^2$ be a random vector with distribution $\widehat{\Lambda}^2$, and let $\widehat{v}_1 = \langle \widehat{w}, \widehat{e}_1\rangle$, $\widehat{v}_2 = \langle \widehat{w}, \widehat{e}_2\rangle$, and $\widehat{\omega} = \langle \widehat{w}, \widehat{e}_3\rangle$. In view of (7.13), we have

$$\begin{aligned}\mathbb{P}(\widehat{v}_1 \le u) &= \frac{1}{2\pi}\int_0^\pi \int_{\sin\widetilde{\theta}\sin\widetilde{\psi}\le u} |\cos\widetilde{\psi}|\sin\psi\mathrm{d}\widetilde{\theta}\mathrm{d}\widetilde{\psi} \\ &= \frac{1}{2\pi}\int_0^\pi F(u,\widetilde{\psi})|\cos\widetilde{\psi}|\sin\psi\mathrm{d}\widetilde{\psi},\end{aligned} \tag{7.24}$$

where

$$F(u,\widetilde{\psi}) = \begin{cases} 2\pi & \text{if } u \geq \sin\widetilde{\psi}, \\ 2\pi - 2\arccos\left(\frac{u}{\sin\widetilde{\psi}}\right) & \text{if } -\sin\widetilde{\psi} < u < \sin\widetilde{\psi}, \\ 0 & \text{if } u \leq -\sin\widetilde{\psi}. \end{cases} \tag{7.25}$$

We have

$$\frac{\partial F}{\partial u} = \begin{cases} 0 & \text{if } |u| > \sin\widetilde{\psi}, \\ \frac{2}{\sqrt{\sin^2\widetilde{\psi} - u^2}} & \text{if } |u| < \sin\widetilde{\psi}. \end{cases} \tag{7.26}$$

Differentiating (7.24), interchanging integration and differentiation, and using symmetry, we see that the density of $\widehat{v}_1$ is given by

$$\begin{aligned} f_{\widehat{v}_1}(u) &= \frac{4}{2\pi}\int_{\arcsin(|u|)}^{\pi/2} \frac{\cos\widetilde{\psi}\sin\widetilde{\psi}\mathrm{d}\widetilde{\psi}}{\sqrt{\sin^2\widetilde{\psi} - u^2}} \\ &= \frac{2}{\pi}\sqrt{\sin^2\widetilde{\psi} - u^2}\Big|_{\arcsin(|u|)}^{\pi/2} \\ &= \frac{2}{\pi}\sqrt{1-u^2}, \qquad -1 \leq u \leq 1. \end{aligned} \tag{7.27}$$

We repeat the procedure to obtain $\widehat{v}_2$. We have

$$\begin{aligned} \mathbb{P}(\widehat{v}_2 \leq u) &= \frac{1}{2\pi}\int_{\arccos u}^{\pi}\int_0^{2\pi} |\cos\widetilde{\psi}|\sin\psi\mathrm{d}\widetilde{\theta}\mathrm{d}\widetilde{\psi} \\ &= \int_{\arccos u}^{\pi} |\cos\widetilde{\psi}|\sin\widetilde{\psi}\mathrm{d}\widetilde{\psi} \\ &= \begin{cases} \frac{1}{2} - \frac{u^2}{2} & \text{if } u \leq 0, \\ \frac{1}{2} + \frac{u^2}{2} & \text{if } u > 0. \end{cases} \end{aligned} \tag{7.28}$$

Differentiating, we obtain that the density of $\widehat{v}_2$ is given by

$$f_{\widehat{v}_2}(u) = |u|, \qquad -1 \le u \le 1. \tag{7.29}$$

To obtain the angular marginal, note that the measure $\widehat{\Lambda}^2$ is invariant under rotations about the x_2-axis. Thus by symmetry, the marginal densities of $\widehat{v}_1$ and $\widehat{\omega}$ are the same:

$$f_{\widehat{\omega}}(u) = \frac{2}{\pi}\sqrt{1-u^2}, \qquad -1 \le u \le 1. \tag{7.30}$$

7.3 Results from ergodic theory

7.3.1 General dynamical systems

The results discussed here are well-known. For a good reference on ergodicity in general, discrete-time Markov processes, see [20].

Let $(\Omega, \mathcal{F})$ be measurable space, where $\mathcal{F}$ is a σ-algebra on Ω. Suppose $T : (\Omega, \mathcal{F}) \to (\Omega, \mathcal{F})$ is a measurable map and $\mathbf{p}$ is a probability measure on $(\Omega, \mathcal{F})$ such that, for all $A \in \mathcal{F}$, $\mathbf{p}(T^{-1}(A)) = \mathbf{p}(A)$. Then the 4-tuple $(\Omega, \mathcal{F}, \mathbf{p}, T)$ is referred to as a *measure preserving system.*

It is easy to see that $\mathcal{I} = \{I \in \mathcal{F} : T^{-1}(I) = I\}$ is a σ-algebra. The following is a well-known result from ergodic theory.

Theorem 7.3 (Birkhoff's ergodic theorem)**.** *Let $\xi : \Omega \to \mathbb{R}$ be a random variable with $\mathbb{E}_{\mathbf{p}}|\xi| < \infty$. Then*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \xi(T^i(\omega)) = \mathbb{E}_{\mathbf{p}}[\xi \mid \mathcal{I}], \tag{7.31}$$

for $\mathbf{p}$*-a.e.* $\omega \in \Omega$.

Here $\mathbb{E}_{\mathbf{p}}$ denotes expectation with respect to the probability measure $\mathbf{p}$, $\mathbb{E}_{\mathbf{p}}[\xi \mid \mathcal{I}]$ denotes the conditional expectation of ξ with respect to $\mathbf{p}$ and the σ-algebra $\mathcal{I}$, and $T^i = T \circ T \circ \cdots \circ T$ (i times). For a proof of this theorem, see for example [20, Theorem 5.2].

The probability measure $\mathbf{p}$ is said to be *ergodic* if for every $I \in \mathcal{I}$, $\mathbf{p}(I) \in \{0, 1\}$. In this case, Theorem 7.3 implies that

$$\frac{1}{N} \sum_{i=1}^{N} \xi(T^i(\omega)) \to \mathbb{E}_{\mathbf{p}}[\xi], \text{ as } N \to \infty \qquad \mathbf{p}\text{-a.s.} \tag{7.32}$$

7.3.2 Markov processes

Let $(X, \mathcal{X})$ be a measurable space, and let $\mathbf{k}(x, \mathrm{d}x')$ be a Markov transition kernel on $(X, \mathcal{X})$. Suppose that μ is a probability measure on $(X, \mathcal{X})$ which is invariant in the sense that, for any $A \in \mathcal{X}$,

$$\int_X \mathbf{k}(x, A)\mu(\mathrm{d}x) = \mu(A). \tag{7.33}$$

Let $\Omega = X^{\mathbb{Z}}$ be the space of X-valued, bi-infinite sequences, and let $\mathcal{F} = \sigma(\bigotimes_{i \in \mathbb{Z}} A_i : A_i \in \mathcal{X})$ be the product σ-algebra on Ω. The Markov transition kernel $\mathbf{k}$ and the invariant probability measure μ give rise to a probability measure $\mathbf{p}_\mu$ on $(\Omega, \mathcal{F})$ whose finite-dimensional distributions are given as follows: For integers $M \leq N$ and sets

$A_M, A_{M+1}, A_{M+2}, \dots A_N \in \mathcal{X}$,

$$\begin{aligned}
&\mathbf{p}_\mu(\{(x_i)_{i\in\mathbb{Z}} \in \Omega : x_i \in A_i \text{ for } M \le i \le N\}) \\
&= \int_{A_M} \cdots \int_{A_{N-1}} \int_{A_N} \mathbf{k}(x_{N-1}, \mathrm{d}x_N)\mathbf{k}(x_{N-2}, \mathrm{d}x_{N-1}) \\
&\qquad\qquad \cdots \mathbf{k}(x_M, \mathrm{d}x_{M+1})\mu(\mathrm{d}x_M).
\end{aligned} \tag{7.34}$$

Define the *shift map* $s : \Omega \to \Omega$ by

$$s((x_i)_{i\in\mathbb{Z}}) = (x_{i+1})_{i\in\mathbb{Z}}, \quad \text{for } (x_i)_{i\in\mathbb{Z}} \in \Omega. \tag{7.35}$$

Using invariance of μ, one may show that $(\Omega, \mathcal{F}, \mathbf{p}_\mu, s)$ forms a measure preserving system. We say that the probability measure μ is *ergodic* if the corresponding probability measure $\mathbf{p}_\mu$ is ergodic with respect to the action of the shift map.

The theorem we state next characterizes ergodicity of μ directly in terms of properties of the measure space $(X, \mathcal{X}, \mu)$ and the transition kernel $\mathbf{k}(x, \mathrm{d}x')$.

We say that a set $A \in \mathcal{X}$ is *μ-invariant* if for μ-a.e. $x \in A$, $\mathbf{k}(x, A) = 1$. It is easy to verify that the collection of μ-invariant subsets is a σ-algebra.

We say that $A \in \mathcal{X}$ is *μ-ergodic* if A is μ-invariant, and for every $B \in \mathcal{X}$ such that $B \subset A$, if B is μ-invariant then $\mu(B) \in \{0, \mu(A)\}$.

Theorem 7.4. *A probability measure μ on $(X, \mathcal{X})$ is ergodic if and only if the set X is μ-ergodic.*

For a proof of this theorem, we refer the reader to [20, Corollary 5.11].

7.4 The Abel Transform

A key ingredient in the proof of Theorem 7.1 is the *Abel transform*, a classic integral transform first studied by Niels Henrik Abel in the context of problems in geometric optics. This transform is defined as follows:

$$\mathcal{A}[\varphi](x) = \int_0^x \frac{\varphi(t)dt}{\sqrt{x-t}}, \qquad 0 < x < 1. \tag{7.36}$$

This transform is well-defined for $\varphi \in L^1(0,1)$. This can be seen by integrating the right-hand side of (7.36) and using Fubini's theorem to interchange the order of integration. More generally, one can apply Minkowski's and Hölder's inequalities to (7.36) show that $\mathcal{A}[\varphi]$ is a bounded operator on $L^p(0,1)$ for $p \in [1,2)$, as follows.

$$\begin{aligned}
\left(\int_0^1 \left|\int_0^x \frac{\varphi(t)dt}{\sqrt{x-t}}dt\right|^p dt\right)^{1/p} &\leq \int_0^1 \left(\int_t^1 (x-t)^{-p/2}dx\right)^{1/p} |\varphi(t)|dt \\
&= \left(1-\frac{p}{2}\right)^{-1/p} \int_0^1 (1-t)^{\frac{1}{p}-\frac{1}{2}}|\varphi(t)|dt \\
&\leq \left(1-\frac{p}{2}\right)^{-1/p} \left(\int_0^1 (1-t)^{\frac{q}{p}-\frac{q}{2}}dt\right)^{1/q} \|\varphi\|_p \\
&= \left(1-\frac{p}{2}\right)^{-1/p} \left(2-\frac{2}{p}\right)^{1-1/p} \|\varphi\|_p,
\end{aligned} \tag{7.37}$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

While there exist explicit inversion formulae for this transform, describing a sufficiently broad class of functions φ for which the inversion formulae are valid is a non-trivial task, complicated by the singularity at $t = x$. The main fact we will need

is that Abel's integral equation

$$\mathcal{A}[\varphi](x) = f(x), \qquad 0 < x < 1, \tag{7.38}$$

has at most one solution in $L^1(0,1)$. That is,

Theorem 7.5. *For every $f \in L^1(0,1)$, there is at most one $\varphi \in L^1(0,1)$ such that equation (7.38) is satisfied.*

We refer the reader to [18] for more information on the Abel transform. Theorem 7.5 was originally proved by Tonelli in [40].

7.5 Proofs of Theorems 7.1 and 7.2

All of the notions described in §7.3 specialize to our model, with $X = \mathbb{S}^2$, $\mathcal{X}$ = the completion of the Borel σ-algebra on $\mathbb{S}^2$ with respect to surface measure, and $\mathbf{k} = \widehat{\mathbb{K}}$. Most of the work in proving Theorem 7.1 is concerned with the implication ($\Leftarrow$). The main strategy will be to argue that $\mathbb{S}^2$ is $\widehat{\Lambda}^2$-ergodic and to apply Theorem 7.4.

First we introduce some new notation. Recall the two systems of spherical coordinates (θ_0, ψ_0) and (θ_1, ψ_1). In what follows, we will frequently identify pairs (θ_0, ψ_0) and (θ_1, ψ_1) with the points in $\mathbb{S}^2$ which they represent.

For $p \in (0, \pi)$ let

$$A_p = \{(\theta_0, \psi_0) \in \mathbb{S}^2 : \psi_0 = p\} \tag{7.39}$$

Geometrically $\{A_p, p \in (0, \pi)\}$ is the collection of circles in $\mathbb{S}^2$ which span planes

orthogonal to the vector χ_0. We also define the semicircles

$$A_p^{\pm} := A_p \cap \mathbb{S}^2_{\pm}. \tag{7.40}$$

Let $N_0 : \mathbb{S}^2 \to \mathbb{S}^2$ be the map whose coordinate representation is given by

$$N_0(\theta_0, \psi_0) = (\theta_0 + \pi \mod 2\pi, \psi_0).$$

If we regard N_0 as acting on points in $\mathbb{S}^2$, then N_0 is simply reflection through the line spanned by χ_0. In particular, N_0 maps $A_p^{\pm}$ onto $A_p^{\mp}$.

In a similar way, using the second system of spherical coordinates, for $p \in (0, \pi)$ we let

$$B_p = \{(\theta_1, \psi_1) \in \mathbb{S}^2 : \psi_1 = p\}, \qquad B_p^{\pm} = B_m \cap \mathbb{S}^2_{\pm}. \tag{7.41}$$

We let $N_1 : \mathbb{S}^2 \to \mathbb{S}^2$ be the map whose coordinate representation is given by

$$N_1(\theta_1, \psi_1) = (\theta_1 + \pi \mod 2\pi, \psi_1). \tag{7.42}$$

This map acts on $\mathbb{S}^2$ by reflecting points through the line spanned by χ_1. Note that N_1 maps $B_p^{\pm}$ onto $B_p^{\mp}$.

It is immediate from the coordinate representations (7.20) that the maps N_0 and N_1 preserve the measure $\widehat{\Lambda}^2$. Also, for $p \in (0, \pi)$, the restriction of N_0 to the circle A_p (identified with $(0, 2\pi)$) preserves the measure $\widehat{\Lambda}^1$, and the restriction of N_1 to the circle B_p preserves the measure $\widehat{\Lambda}^1$.

The coordinates (θ_0, ψ_0) introduce an identification of $\mathbb{S}^2 \setminus \{\pm\chi_0\}$ with the product

space $[0,2\pi)\times(0,\pi)$ in which A_p^+ corresponds to $(0,\pi)\times p$ and A_p^- corresponds to $(\pi,2\pi)\times p$. In virtue of the coordinate representations (7.16) for $\mathbb{K}_-$ and (7.21) for $\widehat{\Lambda}^2$, it is natural to regard the Markov $\widetilde{\mathbb{P}}_0$ as acting on the semicircles A_p^+ and to regard $\widehat{\Lambda}^1$ as a measure on A_p. Thus, for example, if $F\subset\mathbb{S}^2$, one can ask whether $F\cap A_p^+$ is $\widehat{\Lambda}^1$-invariant with respect to $\widetilde{\mathbb{P}}_0$, $\widehat{\Lambda}^1$-ergodic with respect to $\widetilde{\mathbb{P}}_0$, etc.

Symmetrically, with respect to the coordinates (θ_1,ψ_1), the semicircle B_p^- corresponds to set $(0,\pi)\times p$ and the semicircle B_p^+ corresponds to the set $(\pi,2\pi)\times p$. In virtue of the coordinate representation for $\mathbb{K}_+$ (7.16), $\widetilde{\mathbb{P}}_1$ acts on the semicircle B_p^-, and by (7.21), $\widehat{\Lambda}^1$ may be viewed as a measure on B_p.

If $(X,\mathcal{X},\mu)$ is a measure space and $A,B\in\mathcal{X}$, we write

$$A\overset{\mu}{=}B \quad \text{if} \quad \mu(A\triangle B)=0, \tag{7.43}$$

where $A\triangle B=(A\smallsetminus B)\cup(B\smallsetminus A)$ denotes the symmetric difference of A and B. Thus for example, $A\overset{\mu}{=}\varnothing$ means that $\mu(A)=0$.

In this subsection, "almost everywhere" (abbreviated a.e.) means either with respect to the measure $\widehat{\Lambda}^1$ or the measure $\widehat{\Lambda}^2$. This is equivalent to "almost everywhere" with respect to Lebesgue measure on the interval $(0,\pi)$ and surface measure on $\mathbb{S}^2$, respectively. If the measure is not clear from context, we will write "$\widehat{\Lambda}^1$-a.e." or "$\widehat{\Lambda}^2$-a.e."

Lemma 7.6. *Suppose $E\subset\mathbb{S}^2$ is $\widehat{\Lambda}^2$-invariant. Then for almost every $p\in(0,\pi)$,*

$$\begin{aligned}&\mathbb{P}_0(\theta_0,E\cap A_p^+)=1,\ \textit{for } \widehat{\Lambda}^1\textit{-a.e. } \theta_0\in N_0(E\cap A_p^-),\ \textit{and}\\ &\mathbb{P}_0(\theta_0,N_0(E\cap A_p^-))=1 \textit{ for } \widehat{\Lambda}^1\textit{-a.e. } \theta_0\in E\cap A_p^+.\end{aligned} \tag{7.44}$$

The same conclusion holds if we replace $\widetilde{\mathbb{P}}_0$ with $\widetilde{\mathbb{P}}_1$, $A_p^{\pm}$ with $B_p^{\mp}$, and N_0 with N_1.

Proof. Consider the following calculation.

$$\begin{aligned}\int_{(0,\pi)} \widehat{\Lambda}^1(N_0(A_{\psi_0}^- \cap E)) \sin^2 \psi_0 \mathrm{d}\psi_0 &= \widehat{\Lambda}^2(N_0(E \cap \mathbb{S}_-^2)) \\ &= \int_{N_0(E\cap\mathbb{S}_-^2)} \widehat{\mathbb{K}}(N_0(w), E \cap \mathbb{S}_+^2) \widehat{\Lambda}^2(\mathrm{d}w) \\ &= \int_{(0,\pi)} \int_{N_0(A_{\psi_0}^- \cap E)} \widetilde{\mathbb{P}}_0(\theta_0, E \cap A_\psi^+) \frac{1}{\pi} \widehat{\Lambda}^1(\mathrm{d}\theta_0) \sin^2 \psi_0 \mathrm{d}\psi_0.\end{aligned} \tag{7.45}$$

Here the first line follows from the identity (7.21) and the fact that N_0 fixes the ψ_0-coordinate. The second line follows because E is $\widehat{\Lambda}^2$-invariant with respect to $\widehat{\mathbb{K}}$, $N_0^2 = \mathrm{Id}$, and for $w \in \mathbb{S}_-^2$, $\widehat{\mathbb{K}}(w, \mathrm{d}w')$ is supported in $\mathbb{S}_+^2$. The third line uses the coordinate representation (7.16) for $\widehat{\mathbb{K}}$. Since probabilities are bounded by 1, (7.45) gives us that for almost every $p \in (0, \pi)$,

$$\widetilde{\mathbb{P}}_0(\theta_0, A_p^+ \cap E) = 1 \text{ for a.e. } \theta_0 \in N_0(E \cap A_p^-), \tag{7.46}$$

and this proves the first line in (7.44). As for the second line, we proceed as follows. By (7.46) and reversibility of $\widehat{\Lambda}^1$ with respect to $\widetilde{\mathbb{P}}_0$,

$$\begin{aligned}\int_{E\cap A_p^+} \widetilde{\mathbb{P}}_0(\theta_0, N_0(E \cap A_p^-)) \widehat{\Lambda}^1(\mathrm{d}\theta_0) &= \int_{N_0(E\cap A_p^-)} \widetilde{\mathbb{P}}_0(\theta_0, E \cap A_p^+) \widehat{\Lambda}^1(\mathrm{d}\theta_0) \\ &= \widehat{\Lambda}^1(N_0(E \cap A_p^-)).\end{aligned} \tag{7.47}$$

Again since probabilities are bounded by 1, we see that

$$\widehat{\Lambda}^1(E \cap A_p^+) \geq \widehat{\Lambda}^1(N_0(E \cap A_p^-)). \tag{7.48}$$

By (7.21), it follows that

$$\begin{aligned}\widehat{\Lambda}^2(E\cap\mathbb{S}^2_+) &= \frac{1}{\pi}\int_0^\pi \widehat{\Lambda}^1(E\cap A^+_{\psi_0})\sin^2\psi_0 \mathrm{d}\psi_0 \geq \frac{1}{\pi}\int_0^\pi \widehat{\Lambda}^1(N_0(E\cap A^-_{\psi_0}))\sin^2\psi_0 d\psi_0\\ &= \widehat{\Lambda}^2(N_0(E\cap\mathbb{S}^2_-)) = \widehat{\Lambda}^2(E\cap\mathbb{S}^2_-),\end{aligned} \tag{7.49}$$

with equality if and only if (7.48) is an equality for almost every $p\in(0,\pi)$. But since $\widehat{\mathbb{K}}$ takes $\mathbb{S}^2_\pm$ to $\mathbb{S}^2_\mp$, we have $\widehat{\mathbb{K}}(w, E\cap\mathbb{S}^2_\pm) = 1$ for almost every $w\in E\cap\mathbb{S}^2_\mp$. By reversibility with respect to $\widehat{\Lambda}^2$, we have

$$\begin{aligned}\widehat{\Lambda}^2(E\cap\mathbb{S}^2_+) &= \int_{E\cap\mathbb{S}^2_+}\widehat{\mathbb{K}}(w, E\cap\mathbb{S}^2_-)\widehat{\Lambda}^2(\mathrm{d}w)\\ &= \int_{E\cap\mathbb{S}^2_-}\widehat{\mathbb{K}}(w, E\cap\mathbb{S}^2_+)\widehat{\Lambda}^2(\mathrm{d}w)\\ &= \widehat{\Lambda}^2(E\cap\mathbb{S}^2_-).\end{aligned} \tag{7.50}$$

Therefore, (7.49) is an equality. Thus (7.48) is an equality for almost every p, and combining this observation with (7.47) we obtain

$$\int_{E\cap A^+_p}\widetilde{\mathbb{P}}_0(\theta_0, N_0(E\cap A^-_p))\widehat{\Lambda}^1(\mathrm{d}\theta_0) = \widehat{\Lambda}^1(E\cap A^+_p) \text{ for a.e. } p\in(0,\pi). \tag{7.51}$$

Hence, almost every $(p,\theta_0)\in(0,\pi)\times(E\cap A^+_p)$, $\widetilde{\mathbb{P}}_0(\theta_0, N_0(E\cap A^-_p)) = 1$, and this proves the second line in (7.44). The proof of the statement for $\widetilde{\mathbb{P}}_1$, B_p, and N_1 proceeds in exactly the same way. □

Lemma 7.7. *Assume condition (E2) from Theorem 7.1 holds. If E is a $\widehat{\Lambda}^2$-invariant subset of $\mathbb{S}^2$, then $N_0(E)$ and $N_1(E)$ are also $\widehat{\Lambda}^2$-invariant subsets of $\mathbb{S}^2$. Moreover,*

there exist measurable subsets $L_0, L_1 \subset (0,\pi)$ *such that*

$$E \cup N_0(E) \overset{\widehat{\Lambda}^2}{=} \bigcup_{p\in L_0} A_p, \qquad \textit{and} \qquad E \cup N_1(E) \overset{\widehat{\Lambda}^2}{=} \bigcup_{p\in L_1} B_p. \tag{7.52}$$

Proof. If $\widehat{\Lambda}^2(E) = 0$, the result is trivial, so assume this is not the case. We first need to show that $N_0(E)$ is $\widehat{\Lambda}^2$-invariant. Indeed, we compute

$$\begin{aligned}
&\int_{N_0(E)\cap\mathbb{S}^2_-} \widehat{\mathbb{K}}(w, N_0(E)\cap\mathbb{S}^2_+)\widehat{\Lambda}^2(\mathrm{d}w) \\
&= \int_{E\cap\mathbb{S}^2_+} \widehat{\mathbb{K}}(N_0(w), N_0(E)\cap\mathbb{S}^2_+)\widehat{\Lambda}^2(\mathrm{d}w) && (N_0 \text{ preserves the measure } \widehat{\Lambda}^2) \\
&= \frac{1}{\pi}\int_0^\pi \int_{E\cap A^+_{\psi_0}} \widetilde{\mathbb{P}}_0(\theta_0, N_0(E\cap A^-_{\psi_0}))\widehat{\Lambda}^1(\mathrm{d}\theta_0)\sin\psi_0 \mathrm{d}\psi_0 && (\text{by (7.16) and (7.21)}) \\
&= \widehat{\Lambda}^2(E\cap\mathbb{S}^2_+) && (\text{second implication in (7.44)}) \\
&= \widehat{\Lambda}^2(N_0(E)\cap\mathbb{S}^2_-) && (N_0 \text{ preserves the measure } \widehat{\Lambda}^2).
\end{aligned} \tag{7.53}$$

From the equality proved above, we see that $\widehat{\mathbb{K}}(w, N_0(E)\cap\mathbb{S}^2_+) = 1$ for $\widehat{\Lambda}^2$-a.e. $w \in$

$N_0(E) \cap \mathbb{S}^2_-$. On other hand, we also have

$$\begin{aligned}
&\int_{N_0(E)\cap\mathbb{S}^2_+} \widehat{\mathbb{K}}(w, N_0(E)\cap\mathbb{S}^2_-)\widehat{\Lambda}^2(\mathrm{d}w) \\
&= \int_{N_0(E)\cap\mathbb{S}^2_-} \widehat{\mathbb{K}}(w, N_0(E)\cap\mathbb{S}^2_+)\widehat{\Lambda}^2(\mathrm{d}w) && \text{(by reversibility)} \\
&= \widehat{\Lambda}^2(N_0(E)\cap\mathbb{S}^2_-) && \text{(by (7.53))} \\
&= \widehat{\Lambda}^2(E\cap\mathbb{S}^2_+) && (N_0 \text{ preserves the measure } \widehat{\Lambda}^2) \\
&= \widehat{\Lambda}^2(E\cap\mathbb{S}^2_-) && \text{(by (7.50))} \\
&= \widehat{\Lambda}^2(N_0(E)\cap\mathbb{S}^2_+) && (N_0 \text{ preserves the measure } \widehat{\Lambda}^2),
\end{aligned} \tag{7.54}$$

so we similarly conclude that $\widehat{\mathbb{K}}(w, N_0(E)\cap\mathbb{S}^2_-) = 1$ for $\widehat{\Lambda}^2$-a.e. $w \in N_0(E)\cap\mathbb{S}^2_+$. Since $\widehat{\mathbb{K}}$ maps $\mathbb{S}^2_-$ to $\mathbb{S}^2_+$ and $\mathbb{S}^2_+$ to $\mathbb{S}^2_-$, we conclude that $N_0(E)$ is $\widehat{\Lambda}^2$-invariant.

We define

$$L_0 = \{p \in (0,\pi) : \widehat{\Lambda}^1((E\cup N_0(E))\cap A^+_p) > 0\}. \tag{7.55}$$

To see that L_0 is measurable, let $f = \mathbf{1}_{E\cup N_0(E)} : [0,2\pi)\times(0,\pi) \to \mathbb{R}$, where we use the coordinates (θ_0, ψ_0) identify $E\cup N_0(E)$ with a (measurable) subset of $[0,2\pi)\times(0,\pi)$. For $p \in (0,\pi)$, define the "slice" $f_p : [0,2\pi) \to \mathbb{R}$ by $f_p(\theta_0) = f(\theta_0, p)$. It is well-known that the mapping $p \mapsto \int_{(0,\pi)} f_p(\theta_0)\widehat{\Lambda}^1(\theta_0)$ is measurable (see a typical proof the Fubini-Tonelli Theorem, e.g. [17, Chapter 2]), and this quantity is equal to $\widehat{\Lambda}^1((E\cup N_0(E))\cap A^+_p)$, which implies L_0 is measurable.

It follows from Lemma 7.6 that for almost every $p \in L_0$,

$$\widetilde{\mathbb{P}}_0(\theta_0, (E\cup N_0(E))\cap A^+_p) = 1 \text{ for } \widehat{\Lambda}^1\text{-a.e. } \theta_0 \in (E\cup N_0(E))\cap A^+_p. \tag{7.56}$$

Thus $E \cup N_0(E)$ is $\widehat{\Lambda}^1$-invariant with respect to the Markov kernel $\widetilde{\mathbb{P}}_0$. By condition (E2), the measure $\widehat{\Lambda}^1$ is ergodic with respect to $\widetilde{\mathbb{P}}_0$, and thus by Theorem 7.4, $(0, \pi)$ is $\widehat{\Lambda}^1$-ergodic. Hence, identifying A_p^+ with $(0, \pi)$ in the usual way,

$$\widehat{\Lambda}^1((E \cup N_0(E)) \cap A_p^+) = \widehat{\Lambda}^1(A_p^+) \text{ for } \widehat{\Lambda}^1\text{-a.e. } p \in L_0. \tag{7.57}$$

Let $U = \bigcup_{p \in L_0} A_p$, and let $U^\pm = U \cap \mathbb{S}^2_\pm = \bigcup_{p \in L_0} A_p^\pm$. By (7.21) and (7.57), we have

$$\begin{aligned} \widehat{\Lambda}^2((E \cup N_0(E)) \cap U^+) &= \frac{1}{\pi} \int_K \widehat{\Lambda}^1((E \cup N_0(E)) \cap A_p^+) \sin^2(p) \mathrm{d}p \\ &= \frac{1}{\pi} \int_K \widehat{\Lambda}^1(A_p^+) \sin^2(p) \mathrm{d}p = \widehat{\Lambda}^2(U^+). \end{aligned} \tag{7.58}$$

Recall that N_0 maps A_p^+ onto A_p^-, and preserves the measure $\widehat{\Lambda}^1$. Also note that $N_0^2 = \mathrm{Id}$. From these observations and (7.57) we obtain that

$$\widehat{\Lambda}^1((N_0(E) \cup E) \cap A_p^-) = \widehat{\Lambda}^1(A_p^-) \text{ for } \widehat{\Lambda}^1\text{-a.e. } p \in L_0. \tag{7.59}$$

Thus, by a similar calculation to (7.58), we obtain that

$$\widehat{\Lambda}^2((E \cup N_0(E)) \cap U^-) = \widehat{\Lambda}^2(U^-). \tag{7.60}$$

Combined with (7.58), we see that $\widehat{\Lambda}^2((E \cup N_0(E)) \cap U) = \widehat{\Lambda}^2(U)$, or equivalently

$$\widehat{\Lambda}^2(U \setminus (E \cup N_0(E))) = 0. \tag{7.61}$$

On the other hand, for $p \notin L_0$, $\widehat{\Lambda}^2((E \cup N_0(E)) \cap A_p^+) = 0$, and by integrating over p in

$(0,\pi)\setminus L_0$ in a similar way to (7.58), we see that that $\widehat{\Lambda}^2((E\cup N_0(E))\cap(\mathbb{S}^2_+\setminus U^+)) = 0$. Again noting that N_0 preserves $\widehat{\Lambda}^1$, we obtain that for $p\notin L_0$, $\widehat{\Lambda}^2((E\cup N_0(E))\cap A_p^-) = 0$, and thus $\widehat{\Lambda}^2((E\cup N_0(E))\cap(\mathbb{S}^2_-\setminus U^-)) = 0$. Hence,

$$\widehat{\Lambda}^2((E\cup N_0(E))\setminus U) = 0. \tag{7.62}$$

From (7.61) and (7.62), we conclude that $E\cup N_0(E) \overset{\widehat{\Lambda}^2}{=} U$. The argument for E and $N_1(E)$ proceeds in the same way. □

Lemma 7.8. *Assume condition (E2) from Theorem 7.1 holds. If E is $\widehat{\Lambda}^2$-ergodic, then $N_0(E)$ and $N_1(E)$ are also $\widehat{\Lambda}^2$-ergodic.*

Proof. We will just show that $N_0(E)$ is $\widehat{\Lambda}^2$-ergodic, as the argument for $N_1(E)$ is similar. Suppose that $F\subset N_0(E)$ is $\widehat{\Lambda}^2$-invariant. We must show that $\widehat{\Lambda}^2(F)\in\{0,\widehat{\Lambda}^2(E)\}$. By Lemma 7.7, we have that $N_0(F)$ is $\widehat{\Lambda}^2$-ergodic. But $N_0(F)\subset E$, so by $\widehat{\Lambda}^2$-ergodicity of E, $\widehat{\Lambda}^2(N_0(F))\in\{0,\widehat{\Lambda}^2(E)\}$. Since N_0 preserves the measure $\widehat{\Lambda}^2$, the result follows. □

Lemma 7.9. *Assume that γ/π is irrational and condition (E2) from Theorem 7.1 holds. Suppose that E is $\widehat{\Lambda}^2$-ergodic, and $0<\widehat{\Lambda}^2(E)<1$. Then E has the following properties:*

(A) There exists a measurable set $K\subset(-1,1)$ such that

$$E \overset{\widehat{\Lambda}^2}{=} \{w\in\mathbb{S}^2 : \langle w,\widehat{e}_2\rangle\in K\} \tag{7.63}$$

(B) $E\cup N_0(E) \overset{\widehat{\Lambda}^2}{=} \mathbb{S}^2$ and $E\cap N_0(E) \overset{\widehat{\Lambda}^2}{=} \varnothing$.

(C) $\widehat{\Lambda}^2(E) = \widehat{\Lambda}^2(N_0(E)) = \frac{1}{2}$.

Proof. Let E be such a subset of $\mathbb{S}^2$. By Lemma 7.8, $N_0(E)$ is $\widehat{\Lambda}^2$-ergodic, and $N_1(N_0(E))$ is $\widehat{\Lambda}^2$-ergodic. Now, if we regard N_0 and N_1 as acting on unit vectors in $\mathbb{S}^2$, we may write $N_0 = R_{\chi_0,2} \circ R_{13} = R_{13} \circ R_{\chi_0,2}$ and $N_1 = R_{\chi_1,2} \circ R_{13} = R_{13} \circ R_{\chi_1,2}$, where $R_{\chi_i,2}$ denotes reflection in the plane spanned by χ_i and the x_2-axis, and R_{13} denotes reflection in the x_1x_3-plane. Then,

$$N_1 \circ N_0 = R_{\chi_1,2} \circ R_{13} \circ R_{13} \circ R_{\chi_0,2} = R_{\chi_1,2} \circ R_{\chi_0,2}. \tag{7.64}$$

The composition of reflections $R_{\chi_1,2} \circ R_{\chi_0,2}$ is rotation by an angle of

$$2\angle(\chi_0, \chi_1) = 2\arccos\langle \chi_0, \chi_1 \rangle = 2\arccos\left(\frac{1}{1 + J/m}\right) = \gamma. \tag{7.65}$$

about the x_2-axis. Denote this rotation by T_γ. We have proved that

$$E \text{ is } \widehat{\Lambda}^2\text{-ergodic} \quad \Rightarrow \quad T_\gamma(E) \text{ is } \widehat{\Lambda}^2\text{-ergodic}. \tag{7.66}$$

Next, suppose for a contradiction that for all distinct positive integer pairs $k_0, k_1 \in \mathbb{Z}_+$, $\widehat{\Lambda}^2(T_\gamma^{k_0}(E) \cap T_\gamma^{k_1}(E)) = 0$. Since $T_\gamma = N_1 \circ N_0$ is a composition of maps which preserve $\widehat{\Lambda}^2$, and $\widehat{\Lambda}^2(E) > 0$,

$$\widehat{\Lambda}^2\left(\bigcup_{k=0}^{M} T_\gamma^k(E)\right) = \sum_{k=0}^{M} \widehat{\Lambda}^2(T_\gamma^k(E)) = (M+1)\widehat{\Lambda}^2(E) \to \infty \text{ as } M \to \infty, \tag{7.67}$$

contradicting the fact that $\widehat{\Lambda}^2$ is finite. We conclude that there exists $k_0 \geq 1$ such

that $\widehat{\Lambda}^2(E \cap T_\gamma^{k_0}(E)) > 0$. But E and $T_\gamma^{k_0}(E)$ are $\widehat{\Lambda}^2$-ergodic, and so $E \overset{\widehat{\Lambda}^2}{=} T_\gamma^{k_0}(E)$; otherwise $E \cap T_\gamma^{k_0}(E)$ would be an invariant subset of E with measure strictly less than $\widehat{\Lambda}^2(E)$.

Notice that $T_\gamma^{k_0} = T_{k_0\gamma}$ (rotation about the y-axis by an angle of $k_0\gamma$). By hypothesis γ/π is an irrational number. Let T be any rotation about the x_2-axis. By well-known properties of irrational rotation, there exists a sequence of integers $p_j \to \infty$ such that $T_{k_0\gamma}^{p_j} \to T$ pointwise. Then

$$\widehat{\Lambda}^2(E \triangle T(E)) = \lim_{j\to\infty} \widehat{\Lambda}^2(T_{k_0\gamma}^{p_j}(E) \triangle T(E)) = 0. \tag{7.68}$$

by bounded convergence. Thus (up to null sets) E is invariant under rotation about the x_2-axis.

To see why this implies (7.63), recall the spherical coordinates $(\widetilde{\theta}, \widetilde{\psi})$ with "north pole" at $\widehat{e}_2$, given by the formula (7.13). A rotation about the x_2-axis is represented in these coordinates by a mapping of form $\widetilde{T}(\widetilde{\theta}, \widetilde{\psi}) = (\widetilde{\theta} + \alpha \mod 2\pi, \widetilde{\psi})$, for some $\alpha \in \mathbb{R}$. To prove (A), it is enough to show that, with respect to these coordinates, $E \overset{\widehat{\Lambda}^2}{=} \{(\widetilde{\theta}, \widetilde{\psi}) : \widetilde{\psi} \in \widetilde{K}\}$ for some measurable $\widetilde{K} \subset (0, \pi)$. For $\widetilde{\psi} \in (0, \pi)$, let $E_{\widetilde{\psi}} = \{\widetilde{\theta} \in [0, 2\pi) : (\widetilde{\theta}, \widetilde{\psi}) \in E\}$. By invariance under rotations (7.68), we have

$$\begin{aligned} 0 &= \int_{\mathbb{R}} \int_0^{2\pi} \int_0^{\pi} [\mathbf{1}_E(\widetilde{\theta}, \widetilde{\psi}) - \mathbf{1}_{E \setminus \widetilde{T}(E)}(\widetilde{\theta}, \widetilde{\psi})] \mathrm{d}\widetilde{\psi} \mathrm{d}\widetilde{\theta} \mathrm{d}t \\ &= \int_0^{\pi} \int_0^{2\pi} \int_{\mathbb{R}} [\mathbf{1}_{E_{\widetilde{\psi}}}(\widetilde{\theta}) - \mathbf{1}_{E_{\widetilde{\psi}} \setminus (E_{\widetilde{\psi}} + \alpha \mod 2\pi)}(\widetilde{\theta})] \mathrm{d}t \mathrm{d}\widetilde{\theta} \mathrm{d}\widetilde{\psi}, \end{aligned} \tag{7.69}$$

where we use Fubini's theorem in the second line. Since the integrand is nonnegative,

we conclude that: For a.e. $\widetilde{\psi} \in (0, \pi)$,

$$\text{for a.e. } \widetilde{\theta} \in E_{\widetilde{\psi}}, \quad \mathbf{1}_{E_{\widetilde{\psi}}}(\widetilde{\theta} - \alpha \mod 2\pi) = 1 \text{ for a.e. } \alpha \in \mathbb{R}. \tag{7.70}$$

In particular, if the Lebesgue measure of $E_{\widetilde{\psi}}$ is positive, then there exists $\widetilde{\theta}$ such that $\mathbf{1}_{E_{\widetilde{\psi}}}(\widetilde{\theta} - \alpha \mod 2\pi) = 1$ for almost every $\alpha \in \mathbb{R}$, which in turn implies that $E_{\widetilde{\psi}}$ is a full-measure subset of $[0, 2\pi)$. Let $\widetilde{K} = \{\psi \in (0, \pi) : \mathcal{L}^1(E_{\widetilde{\psi}}) > 0\}$, where $\mathcal{L}^1$ denotes Lebesgue measure. Then $\widetilde{K}$ is measurable, and, up to null sets, $E = \{(\theta, \psi) : \psi \in \widetilde{K}\}$, as desired.

We next prove (B). Since $\widehat{\Lambda}^2(N_0(E)) = \widehat{\Lambda}^2(E) \in (0, 1)$ and by Lemma 7.8, $N_0(E)$ is $\widehat{\Lambda}^2$-ergodic, the same conclusion (A) holds for $N_0(E)$. Thus, there exists $K_2 \subset (-1, 1)$ such that

$$F := E \cup N_0(E) \overset{\widehat{\Lambda}^2}{=} \{w \in \mathbb{S}^2 : \langle w, \widehat{e}_2 \rangle \in K_2\}. \tag{7.71}$$

This implies that, up to null sets, F is invariant under rotations about the x_2-axis. But in addition, by Lemma 7.7, there exists $L_0 \subset (0, \pi)$ such that

$$F \overset{\widehat{\Lambda}^2}{=} \bigcup_{p \in L_0} A_p, \tag{7.72}$$

Recalling that each A_p is a circle contained in a plane orthogonal to χ_0, this implies that, up to null sets, F is invariant under rotations about the line spanned by χ_0. Let $T_{\alpha,\beta} : \mathbb{S}^2 \to \mathbb{S}^2$ by the composition of a counterclockwise rotation by an angle of α about the $+x_2$-axis followed by a counterclockwise rotation by an angle of β about χ_0. Since χ_0 is orthogonal to the x_2-axis, it is easy to see that, for any $w \in \mathbb{S}^2$, the

mapping $U_w : [0, 2\pi) \times [0, 2\pi) \to \mathbb{S}^2$ defined by

$$U_w(\alpha, \beta) = T_{\alpha,\beta}(w) \tag{7.73}$$

is a smooth, surjective map. By invariance of F and spherical surface measure σ under rotations about the x_2-axis and about χ_0, we have

$$\begin{aligned} 0 &= \int_0^{2\pi} \int_0^{2\pi} \int_{\mathbb{S}^2} [\mathbf{1}_F(w) - \mathbf{1}_{F \setminus T_{\alpha,\beta}^{-1}(F)}(w)] \sigma(\mathrm{d}w) \mathrm{d}\alpha \mathrm{d}\beta \\ &= \int_{\mathbb{S}^2} \int_0^{2\pi} \int_0^{2\pi} [\mathbf{1}_F(w) - \mathbf{1}_{F \setminus T_{\alpha,\beta}^{-1}(F)}(w)] \sigma(\mathrm{d}w) \mathrm{d}\alpha \mathrm{d}\beta, \end{aligned} \tag{7.74}$$

where we use Fubini's theorem in the second line. Since the integrand is non-negative, the equality above implies that for σ-a.e. $w \in F$,

$$U_w(\alpha, \beta) = T_{\alpha,\beta}(w) \in F \quad \text{for a.e. } (\alpha, \beta) \in [0, 2\pi) \times [0, 2\pi). \tag{7.75}$$

In particular, since F has strictly positive measure, there exists w such that, up to null sets,

$$F \supset U_w([0, 2\pi) \times [0, 2\pi)) = \mathbb{S}^2 \tag{7.76}$$

as desired.

Next, let $F' = E \cap N_0(E)$. Then F' is a subset of E and of $N_0(E)$, and F' is $\widehat{\Lambda}^2$-invariant since E and $N_0(E)$ are. By $\widehat{\Lambda}^2$-ergodicity of E, $\widehat{\Lambda}^2(F') \in \{0, \widehat{\Lambda}^2(E)\}$. If $\widehat{\Lambda}^2(F') = \widehat{\Lambda}^2(E)$, then $E \overset{\widehat{\Lambda}^2}{=} N_0(E)$, and therefore E is a full measure subset of $\mathbb{S}^2$, contrary to our hypothesis. Therefore, $\widehat{\Lambda}^2(F') = 0$, and this completes the proof of (B).

Finally, note that (C) follows immediately from (B) because N_0 preserves the measure $\widehat{\Lambda}^2$. $\square$

Proof of Theorem 7.1. For implication ($\Leftarrow$), suppose E is an $\widehat{\Lambda}^2$-ergodic subset with $\widehat{\Lambda}^2(E) > 0$, and assume for a contradiction that $\widehat{\Lambda}^2(E) < \widehat{\Lambda}^2(\mathbb{S}^2)$. By Lemma 7.6 and reversibility of $\widehat{\Lambda}^1$ with respect to $\widetilde{\mathbb{P}}_0$, for almost every $p \in (0, \pi)$,

$$\begin{aligned} \widehat{\Lambda}^1(N_0(E) \cap A_p^+) &= \int_{N_0(E) \cap A_p^+} \widetilde{\mathbb{P}}_0(\theta_0, E \cap A_p^+) \widehat{\Lambda}^1(\mathrm{d}\theta_0) \\ &= \int_{E \cap A_p^+} \widetilde{\mathbb{P}}_0(\theta_0, N_0(E) \cap A_p^+) \widehat{\Lambda}^1(\mathrm{d}\theta_0) = \widehat{\Lambda}^1(E \cap A_p^+). \end{aligned} \tag{7.77}$$

By Lemma 7.9 (B) $E \cap N_0(E) \overset{\widehat{\Lambda}^2}{=} \varnothing$, and $E \cup N_0(E) \overset{\widehat{\Lambda}^2}{=} \mathbb{S}^2$. Thus, by (7.21)

$$0 = \frac{1}{\pi} \int_0^\pi \widehat{\Lambda}^1((E \cap N_0(E)) \cap A_p) \sin^2(p) \mathrm{d}p, \tag{7.78}$$

and this implies that for almost every $p \in (0, \pi)$, $\widehat{\Lambda}^1((E \cap N_0(E)) \cap A_p) = 0$. As this observation and (7.77) allows us to conclude that

$$\widehat{\Lambda}^1(E \cap A_p^+) = \frac{1}{2} \widehat{\Lambda}^1(A_p^+) = \frac{1}{2} \quad \text{for a.e. } p \in (0, \pi). \tag{7.79}$$

Fix $p \in (0, \pi)$ such that the equality in (7.79) holds. Let $\pi_p : A_p^+ \to (0, \sin p)$ denote orthogonal projection onto the x_2-axis. In coordinates,

$$\pi_p(\theta_0) = \langle G_0(\theta_0, p), \widehat{e}_2 \rangle = \sin p \sin \theta_0. \tag{7.80}$$

See (7.39) and (7.12). Taking $K^+ = K \cap [0, 1)$, where K is as in Lemma 7.9, we have

$E \cap A_p^+ = \pi_p^{-1}(K^+)$. Therefore, we obtain from (7.79),

$$\begin{aligned}\frac{1}{2} = \widehat{\Lambda}^1(\pi_p^{-1}(K^+)) &= \frac{1}{2}\int_{\pi_p^{-1}(K^+)} \sin\theta_0 \mathrm{d}\theta_0 = \int_{\pi_p^{-1}(K^+)\cap(0,\pi/2)} \sin\theta_0 \mathrm{d}\theta_0 \\ &= \int_{K^+\cap(0,\sin p)} \frac{ydy}{\sin p\sqrt{\sin^2 p - y^2}}.\end{aligned} \tag{7.81}$$

The third equality in (7.81) follows by symmetry, observing in (7.80) that π_p is invariant under composition with $\theta_0 \mapsto \pi - \theta_0$. The fourth equality in (7.81) is obtained via the change of variables $y = \pi_p(\theta_0) = \sin p \sin\theta_0$. Setting $R = \sin p$, we obtain from (7.81) that

$$\begin{aligned}1 &= \frac{2}{R}\int_0^R \mathbf{1}_{K^+}(y)\frac{ydy}{\sqrt{R^2 - y^2}} \\ &= \frac{2}{R}\int_0^{R^2} \frac{1}{2}\mathbf{1}_{K^+}(t^{1/2})\frac{dt}{\sqrt{R^2 - t}} \qquad (\text{set } y = t^{1/2}) \\ &= \frac{1}{R}\mathcal{A}[\varphi_K](R^2),\end{aligned} \tag{7.82}$$

where $\mathcal{A}$ is the Abel transform defined in §7.4, and $\varphi_K(t) = \mathbf{1}_{K^+}(t^{1/2})$. Since (7.79) holds for almost every $p \in (0,\pi)$, (7.82) holds for almost every $R = \sin p \in (0,1)$. Substituting $x = R^2$, we conclude that

$$\mathcal{A}[\varphi_K](x) = \sqrt{x} \qquad \text{for a.e. } x \in (0,1). \tag{7.83}$$

But one also easily verifies that $\mathcal{A}[1/2](x) = \sqrt{x}$. Since the functions $\varphi_K(x), 1/2$, and $\sqrt{x}$ are all contained in $L^1(0,1)$, uniqueness of solutions to the Abel integral equation (7.38) implies that $\varphi_K = 1/2$ on $(0,1)$. This gives us the desired contradiction since

φ_K only takes values in $\{0,1\}$.

The converse ($\Rightarrow$) proceeds as follows. Assume that $\widehat{\Lambda}^2$ is ergodic. Then clearly (E1) holds with $\mu = \widehat{\Lambda}^2$. To prove (E2) holds, suppose that $F \subset (0,\pi)$ is $\widehat{\Lambda}^1$-invariant with respect to the Markov kernel $\widetilde{\mathbb{P}}_0$, and assume that $\widehat{\Lambda}^1(F) > 0$. We will prove that $\widehat{\Lambda}^1(F) = 1$. Under the usual identification of $(0,\pi)$ with A_p^+, let $F_p = G_0(F \times p) \subset A_p^+$ be the subset corresponding to F. Let

$$E^+ = \bigcup_{p\in(0,\pi)} F_p, \qquad E_- = \bigcup_{p\in(0,\pi)} N_0(F_p), \qquad E = E_+ \cup E_-. \tag{7.84}$$

We will prove that E is $\widehat{\Lambda}^1$-invariant. By (7.21), $\widehat{\Lambda}^1$-invariance of F_{ψ_0}, and (7.16)

$$\begin{aligned}
\widehat{\Lambda}^2(E_+) &= \frac{1}{\pi}\int_0^\pi \widehat{\Lambda}^1(F_{\psi_0})\sin^2\psi_0 \mathrm{d}\psi_0 \\
&= \frac{1}{\pi}\int_0^\pi \int_{F_{\psi_0}} \widetilde{\mathbb{P}}(\theta_0, F_{\psi_0})\widehat{\Lambda}^1(\mathrm{d}\theta_0)\sin^2\psi_0 \mathrm{d}\psi_0 \\
&= \frac{1}{\pi}\int_0^\pi \int_{N_0(F_{\psi_0})} \widetilde{\mathbb{P}}_0(\theta_0 - \pi, F_{\psi_0})\widehat{\Lambda}^1(\mathrm{d}\theta_0)\sin^2\psi_0 \mathrm{d}\psi_0 \\
&= \int_{E_-} \widehat{\mathbb{K}}(w, E_+)\widehat{\Lambda}^2(\mathrm{d}w) \\
&= \int_{E_+} \widehat{\mathbb{K}}(w, E_-)\widehat{\Lambda}^2(\mathrm{d}w),
\end{aligned} \tag{7.85}$$

where we use symmetry of $\widehat{\Lambda}^2$ with respect to $\widehat{\mathbb{K}}$ in the last line. Since probabilities are bounded by 1, we see that

$$\widehat{\mathbb{K}}(w, E_-) = 1 \quad \text{for } \widehat{\Lambda}^2\text{-a.e. } w \in E_+. \tag{7.86}$$

Noting that $E_- = N_0(E_+)$ and N_0 preserves $\widehat{\Lambda}^2$, it also follows from (7.85) that

$$\widehat{\Lambda}^2(E_-) = \widehat{\Lambda}^2(E_+) = \int_{E_-} \widehat{\mathbb{K}}(w, E_+)\widehat{\Lambda}^2(\mathrm{d}w). \tag{7.87}$$

Therefore

$$\widehat{\mathbb{K}}(w, E_+) = 1 \quad \text{for } \widehat{\Lambda}^2\text{-a.e. } w \in E_-. \tag{7.88}$$

We conclude from (7.86) and (7.88) that $E = E_+ \cup E_-$ is $\widehat{\Lambda}^2$-invariant. Note that $\widehat{\Lambda}^2(E) > 0$ since by assumption $\widehat{\Lambda}^1(F) > 0$. By ergodicity of $\widehat{\Lambda}^2$ and Theorem 7.4, it follows that E is a full measure subset of $\mathbb{S}^2$. In particular, $\widehat{\Lambda}^2(E_+) = \widehat{\Lambda}^2(E_-) = 1/2$. Thus, from the first equality in (7.85), we see that

$$\frac{1}{\pi}\int_0^\pi \widehat{\Lambda}^1(F_{\psi_0})\sin^2\psi_0 \mathrm{d}\psi_0 = \frac{1}{2} = \frac{1}{\pi}\int_0^\pi \widehat{\Lambda}^1(A^+_{\psi_0})\sin^2\psi_0 \mathrm{d}\psi_0. \tag{7.89}$$

Since $F_{\psi_0} \subset A^+_{\psi_0}$, it follows that F_{ψ_0} is a full measure subset of $A^+_{\psi_0}$ for $\widehat{\Lambda}^1$-a.e. $\psi_0 \in (0, \pi)$. Therefore F is a $\widehat{\Lambda}^1$-full measure subset of $(0, \pi)$. This proves that $(0, \pi)$ is $\widehat{\Lambda}^1$-ergodic with respect to $\widehat{\mathbb{P}}_0$. A symmetric argument shows that $(0, \pi)$ is $\widehat{\Lambda}_1$-ergodic with respect to $\widehat{\mathbb{P}}_1$. Thus property (E2) follows by Theorem 7.4. □

Before proving Theorem 7.2, we need one more lemma.

Lemma 7.10. *Suppose that for every $\widehat{\Lambda}^2$-ergodic subset $E \subset \mathbb{S}^2$, $\widehat{\Lambda}^2(E) = 0$. Then*

(i) For any $\widehat{\Lambda}^2$-invariant subset $F \subset \mathbb{S}^2$, there exists an $\widehat{\Lambda}^2$-invariant subset $F' \subset F$ with $4^{-1}\widehat{\Lambda}^2(F) \le \widehat{\Lambda}^2(F') \le 2^{-1}\widehat{\Lambda}^2(F)$.

(ii) If $F \subset \mathbb{S}^2$ is $\widehat{\Lambda}^2$-invariant and $\widehat{\Lambda}^2(F) > 0$, then there exists a partition

$\{F_1, F_2, F_3, F_4\}$ *of* F *such that each* F_j *is* $\widehat{\Lambda}^2$*-invariant and* $0 < \widehat{\Lambda}^2(F_j) < 2^{-1}\widehat{\Lambda}^2(F)$.

(iii) For each $k \geq 1$*, there exists a partition of* $\mathbb{S}^2$ *into* $\widehat{\Lambda}^2$*-invariant subsets* $\{F_1, F_2, \cdots, F_{2^{2k}}\}$ *such that* $0 < \widehat{\Lambda}^2(F_j) \leq 2^{-k}$ *for* $1 \leq j \leq 2^{2k}$.

Proof. (i) The result is obvious if $\widehat{\Lambda}^2(F) = 0$, so assume this is not the case. Define $\mathcal{T}(F) = \{F' \subset F : F' \text{ is } \widehat{\Lambda}^2\text{-invariant, and } 0 < \widehat{\Lambda}^2(F') \leq 2^{-1}\widehat{\Lambda}^2(F)\}$. Since $\widehat{\Lambda}^2(F) > 0$, F cannot be ergodic, and therefore there exists an $\widehat{\Lambda}^2$-invariant $F' \subset F$ with $0 < \widehat{\Lambda}^2(F') < \widehat{\Lambda}^2(F)$. By replacing F' with $F \setminus F'$ if necessary, we will have $\widehat{\Lambda}^2(F') \leq 2^{-1}\widehat{\Lambda}^2(F)$. Thus, $\mathcal{T}(F)$ is nonempty, the following number

$$t^* := \sup\{\widehat{\Lambda}^2(F') : F' \in \mathcal{T}(F)\} \tag{7.90}$$

exists, and $0 < t^* \leq 2^{-1}\widehat{\Lambda}^2(F)$. Suppose for a contradiction that $t^* < 4^{-1}\widehat{\Lambda}^2(F)$. For each $j \geq 1$, choose $F_j \in \mathcal{T}(F)$ such that

$$t^* - j^{-1} < \widehat{\Lambda}^2(F_j) \leq t^*. \tag{7.91}$$

Let $\widehat{F} = \bigcup_{j=1}^{\infty} F_j$, and consider the following two cases:

Case 1: $\widehat{\Lambda}^2(\widehat{F}) > t^*$. Then

$$k^* := \min\left\{k \geq 1 : \widehat{\Lambda}^2\left(\bigcup_{j=1}^{K} F_j\right) > t^*\right\} \tag{7.92}$$

exists and is finite. Let $F' = \bigcup_{j=1}^{k^*} F_j$. Then $F' \subseteq F$, and F' is $\widehat{\Lambda}^2$-invariant, and we

have

$$0 < \widehat{\Lambda}^2(F') \le \widehat{\Lambda}^2\left(\bigcup_{j=1}^{k^*-1} F_j\right) + \widehat{\Lambda}^2(F_{k^*}) \le t^* + t^* \le 2^{-1}\widehat{\Lambda}^2(F). \tag{7.93}$$

Hence $F' \in \mathcal{T}(F)$. But $\widehat{\Lambda}^2(F') > t^*$, a contradiction.

Case 2: $\widehat{\Lambda}^2(\widehat{F}) \le t^*$. This is in fact an equality, because

$$\widehat{\Lambda}^2(\widehat{F}) \ge \widehat{\Lambda}^2(F_j) \to t^* \text{ as } j \to \infty. \tag{7.94}$$

Since $\widehat{F}$ is $\widehat{\Lambda}^2$-invariant and $\widehat{\Lambda}^2(\widehat{F}) = t^* \le 2^{-1}\widehat{\Lambda}^2(F)$, the set $F \setminus \widehat{F}$ is invariant, and $\widehat{\Lambda}^2(F \setminus \widehat{F}) \ge 2^{-1}\widehat{\Lambda}^2(F) > 0$. Since $F \setminus \widehat{F}$ cannot be ergodic, there exists a $\widehat{\Lambda}^2$-invariant $G \subset F \setminus \widehat{F}$ with $0 < \widehat{\Lambda}^2(G) \le 2^{-1}\widehat{\Lambda}^2(F \setminus \widehat{F}) \le 2^{-1}\widehat{\Lambda}^2(F)$. Hence, $G \in \mathcal{T}(F)$. Define $F' = \widehat{F} \cup G$. Then F' is invariant, and since $\widehat{F}, G \in \mathcal{T}(F)$,

$$0 < \widehat{\Lambda}^2(F') = \widehat{\Lambda}^2(\widehat{F}) + \widehat{\Lambda}^2(G) \le 2t^* \le 2^{-1}\widehat{\Lambda}^2(F). \tag{7.95}$$

Therefore $F' \in \mathcal{T}(F)$. But $\widehat{\Lambda}^2(F') = t^* + \widehat{\Lambda}^2(G) > t^*$, a contradiction.

This shows that $t^* \ge 4^{-1}\widehat{\Lambda}^2(F)$, and (i) follows.

(ii) Define inductively pairs of sets $\{(F_k^1, F_k^2)\}_{k\ge1}$ as follows: F_1^1 is a $\widehat{\Lambda}^2$-invariant subset of F such that $4^{-1}\widehat{\Lambda}^2(F) \le \widehat{\Lambda}^2(F_1^1) \le 2^{-1}\widehat{\Lambda}^2(F)$, and $F_1^2 = F \setminus F_1^1$. For $k > 1$, define F_k^1 to be a $\widehat{\Lambda}^2$-invariant subset of F_{k-1}^2 such that $4^{-1}\widehat{\Lambda}^2(F_{k-1}^2) < \widehat{\Lambda}^2(F_k^1) \le 2^{-1}\widehat{\Lambda}^2(F_{k-1}^2)$, and $F_k^2 = F_{k-1}^2 \setminus F_k^1$. Clearly, for each $k \ge 1$, the collection

$$\{F_1^1, F_2^1, \ldots, F_k^1, F_k^2\} \tag{7.96}$$

is a partition of F into $\widehat{\Lambda}^2$-invariant subsets. We have $\widehat{\Lambda}^2(F_1^1) \le 2^{-1}\widehat{\Lambda}^2(F)$, and for

each $j > 1$, $\widehat{\Lambda}^2(F_j^1) \le 2^{-1}\widehat{\Lambda}^2(F_{j-1}^2) \le 2^{-1}\widehat{\Lambda}^2(F)$. Moreover, $\widehat{\Lambda}^2(F_k^2) \le (3/4)\widehat{\Lambda}^2(F_{k-1}^2)$, and by induction $\widehat{\Lambda}^2(F_k^2) \le (3/4)^k\widehat{\Lambda}^2(F)$. By taking $k = 3$, we obtain that $\widehat{\Lambda}^2(F_k^2) \le (27/64)\widehat{\Lambda}^2(F)$, and so $\{F_1^1, F_2^1, F_3^1, F_3^2\}$ is a partition of F satisfying the desired properties.

(iii) This result follows by iterating part (ii). □

Proof of Theorem 7.2. We may assume without loss of generality that for *every* $\theta \in (0,\pi)$, $\widetilde{\mathbb{P}}_i(\theta, \mathrm{d}\theta')$ is not singular with respect to $\widehat{\Lambda}^1$, for $i = 0, 1$. In the case where $\widetilde{\mathbb{P}}_i(\theta, \mathrm{d}\theta')$ is singular at points θ in a null set $N \subset (0,\pi)$, we can modify the Markov kernel so that $\widetilde{\mathbb{P}}_i(\theta, \mathrm{d}\theta') = \sin\theta' \mathrm{d}\theta'$ for $\theta \in N$. This modification does not effect the property that $\widetilde{\mathbb{P}}_i$ is symmetric with respect to $\widehat{\Lambda}^1$. It also only effects $\widehat{\mathbb{K}}$ on a $\widehat{\Lambda}^2$-null set by virtue of the coordinate representation (7.16), and it does not affect the set of invariant measures which are absolutely continuous with respect to $\widehat{\Lambda}^2$.

Because of the product decompositions (7.16), the assumption that $\widetilde{\mathbb{P}}_i(\theta, \mathrm{d}\theta')$ is not singular, for $i = 0, 1$, is equivalent to the following two statements: For $p, q \in (0,\pi)$,

$$\begin{aligned} &\text{for every } w \in A_p^- \text{ and every } M \subset A_p^+, \\ &\widehat{\Lambda}^1(M) = 1 \text{ implies } \mathbb{K}(w, M) > 0, \end{aligned} \tag{7.97}$$

and

$$\begin{aligned} &\text{for every } w \in B_q^+ \text{ and every } M \subset B_q^-, \\ &\widehat{\Lambda}^1(M) = 1 \text{ implies } \mathbb{K}(w, M) > 0. \end{aligned} \tag{7.98}$$

With these observations made, the proof proceeds in two steps.

Step 1. Let $\widehat{\mathbb{K}}^2(w, \mathrm{d}w')$ be the Markov kernel on $\mathbb{S}^2$ defined by

$$\widehat{\mathbb{K}}^2(w, B) = \int_{\mathbb{S}^2} \widehat{\mathbb{K}}(w', B)\widehat{\mathbb{K}}(w, \mathrm{d}w'), \tag{7.99}$$

for $w \in \mathbb{S}^2$ and measurable sets $B \subset \mathbb{S}^2$. Fix $w^* \in \mathbb{S}^2_-$. Then $w^* \in A^-_{p^*}$ for some $p^* \in (0, \pi)$. In this step, we will prove that $\widehat{\mathbb{K}}^2(w^*, \mathrm{d}w')$ is not singular with respect to $\widehat{\Lambda}^2$.

Define the mapping $\Pi : A^+_{p^*} \to (0, \pi)$ by

$$\Pi(w) = \arccos(\langle w, \chi_1 \rangle). \tag{7.100}$$

This mapping projects each w to its ψ_1-coordinate with respect to the coordinates (θ_1, ψ_1). Equivalently, for each w, $q = \Pi(w)$ is the unique point in $(0, \pi)$ such that w lies in the circle B_q. We claim that Π is injective. To see this, note that for each $q \in (0, \pi)$, the circles B_q and A_{p^*} can intersect in at most two points. If they intersect in two points, then one point lies in $\mathbb{S}^2_+$ and one point lies in $\mathbb{S}^2_-$. Thus there is at most one point w in $A^+_{p^*}$ such that $w \in B_q$.

By (7.12) the representation of Π in the coordinates (θ_0, ψ_0) is given by

$$\Pi(\theta_0) = \arccos((\cos\theta_0 \sin p^*)\langle \chi_0^\perp, \chi_1 \rangle + (\cos p^*)\langle \chi_0, \chi_1 \rangle), \qquad \theta_0 \in (0, \pi). \tag{7.101}$$

The above map is smooth, with first derivative given by

$$\Pi'(\theta_0) = \frac{\sin\theta_0 \sin p^*}{\sqrt{1 - \langle w, \chi_0 \rangle}}, \tag{7.102}$$

which never vanishes for $\theta_0 \in (0, \pi)$. Consequently, we see that $\Pi : A^-_{p^*} \to (0, \pi)$ is a diffeomorphism onto its image.

Let Q be a subset of $\mathbb{S}^2$ such that $\widehat{\Lambda}^2(Q) = 1$. We will be done with this step if we can show that $\widehat{\mathbb{K}}^2(w^*, Q) > 0$. By (7.16), since $w^* \in A^-_{p^*}$, $\widehat{\mathbb{K}}(w^*, \mathrm{d}w')$ is supported in $A^+_{p^*}$, and for each $w' \in A^+_{p^*}$, $\widehat{\mathbb{K}}(w', \mathrm{d}w'')$ is supported in $B^-_{\Pi(w')}$. Let

$$Q' = \{w' \in A^+_{p^*} : \widehat{\mathbb{K}}(w', Q) > 0\}. \tag{7.103}$$

Assume that Q is a $\widehat{\Lambda}^2$-full measure subset of $\mathbb{S}^2$. Then by (7.21) with $i = 1$,

$$1 = \widehat{\Lambda}^2(Q) = \frac{1}{\pi} \int_0^\pi \widehat{\Lambda}^1(B_q \cap Q) \sin^2 q \mathrm{d}q, \tag{7.104}$$

Hence, for almost every $q \in (0, \pi)$, $B_q \cap Q$ is a $\widehat{\Lambda}^1$-full measure subset of B_q, and thus $\widehat{\Lambda}^1(B^-_q \cap Q) = 1$. Since Π is a diffeomorphism onto its image, it follows that $\widehat{\Lambda}^1(B^-_{\Pi(w')} \cap Q) = 1$ for $\widehat{\Lambda}^1$-a.e. $w' \in A^+_{p^*}$. But if $\widehat{\Lambda}^1(B^-_{\Pi(w')} \cap Q) = 1$, then $\widehat{\mathbb{K}}(w, B^-_{\Pi(w')} \cap Q) > 0$ by (7.98). This proves that Q' is a full measure subset of $A^-_{p^*}$. Thus by (7.97), $\widehat{\mathbb{K}}(w^*, Q') > 0$. We conclude that

$$\widehat{\mathbb{K}}^2(w^*, Q) = \int_{Q'} \widehat{\mathbb{K}}(w', Q) \widehat{\mathbb{K}}(w^*, \mathrm{d}w') > 0, \tag{7.105}$$

as desired.

Step 2. We will prove that there exists a $\widehat{\Lambda}^2$-ergodic subset $E \subset \mathbb{S}^2$ such that $\widehat{\Lambda}^2(E) > 0$. Given such a set exists, we can define

$$\mu_E(\mathrm{d}w) = \frac{\mathbf{1}_E(w)\widehat{\Lambda}^2(\mathrm{d}w)}{\widehat{\Lambda}^2(E)}. \tag{7.106}$$

Then μ_E is an invariant measure which is absolutely continuous with respect to $\widehat{\Lambda}^2$. Moreover, if $F \subset \mathbb{S}^2$ is μ_E-invariant and $\mu_E(F) > 0$, then $F \cap E$ is $\widehat{\Lambda}^2$-invariant. Consequently, $F \cap E$ is a $\widehat{\Lambda}^2$-full measure subset of E. Hence, F is a μ_E-full measure subset of $\mathbb{S}^2$. This shows $\mathbb{S}^2$ is μ_E-ergodic, and μ_E is the desired ergodic measure by Theorem 7.4.

To prove that such an E exists, suppose for a contradiction that every $\widehat{\Lambda}^2$-ergodic subset of $\mathbb{S}^2$ is $\widehat{\Lambda}^2$-null. By the Radon-Nikodym theorem, for each $w \in \mathbb{S}^2$, we may write $\widehat{\mathbb{K}}^2(w, \mathrm{d}w') = f_w(w')\widehat{\Lambda}^2(w') + \nu_w(\mathrm{d}w')$, where $f_w \in L^1(\widehat{\Lambda}^2)$ and ν_w is singular with respect to $\widehat{\Lambda}^2$. Let $P_w = \{w' \in \mathbb{S}^2 : f_w(w') > 0\}$. By step 1, $\widehat{\Lambda}^2(P_w) > 0$. Also, note that if F is any $\widehat{\Lambda}^2$-invariant subset of $\mathbb{S}^2$, then for almost every $w \in F$, $\widehat{\Lambda}^2(F \setminus P_w) = 0$.

Fix $k \geq 1$. By Lemma 7.10(iii), we can partition $\mathbb{S}^2$ into $\widehat{\Lambda}^2$-invariant subsets $\{F_1, F_2, \cdots, F_{2^{2k}}\}$ such that $0 < \widehat{\Lambda}^2(F_j) \leq 2^{-k}$ for $1 \leq j \leq 2^{2k}$. For $\widehat{\Lambda}^2$-a.e. $w \in F_j$, $\widehat{\mathbb{K}}^2(w, \mathbb{S}^2 \setminus F_j) = 0$. This implies that for almost every $w \in F_j$, $\widehat{\Lambda}^2(P_w) \leq \widehat{\Lambda}^2(F_j) \leq 2^{-k}$. Since every point in $\mathbb{S}^2$ lies in some F_j this implies that $\widehat{\Lambda}^2(P_w) \leq 2^{-k}$ for a.e. $w \in \mathbb{S}^2$. Since k is arbitrary, we conclude that $\widehat{\Lambda}^2(P_w) = 0$ for almost every $w \in \mathbb{S}^2$, a contradiction. $\square$

8 Index of Notation

$\langle \cdot, \cdot \rangle$ – Kinetic energy inner product, §1.3.4

$\| \cdot \|$ – Norm induced by the kinetic energy inner product, §1.3.4

$\| \cdot \|_{L^p} = \| \cdot \|_{L^p_\mu(X)}$ – L^p-norm on $L^p_\mu(X)$, §1.2.1

$Y \subset\subset X$ – The closure of Y lies in X, §6.1.2

$\mathcal{A}_0$ – A special subclass of rough collision laws, §1.3.7

$\widehat{B} = \overline{W^c} + e_2$ – Base of the cylinder $\mathcal{M}_{\text{cyl}}$, §4.2

$C_c(X)$ – Continuous, compactly supported functions, §1.2.1

$C^k(X)$ – k-times continuously differentiable functions, §1.2.1

$C^\infty(X)$ – Infinitely differentiable functions, §1.2.1

$C^k_c(X) = C_c(X) \cap C^k(X)$, §1.2.1

$C^\infty_c(X) = C_c(X) \cap C^\infty(X)$, §1.2.1

$\text{CES}^2_0(\mathcal{R})$ – A special class of billiard domains with singularities, §6.1.1

$\widetilde{\chi} = (-1, 0, 1)$ – Rolling velocity, §1.3.3

$\chi = (m + J)^{-1/2}(-1, 0, 1)$ – Normalized rolling velocity, §1.3.6

$d^\nu_{\mathcal{G}}$ – A pseudometric on a space of Markov kernels $\mathcal{G}$, §6.2.3

D – Disk with satellites, §1.3.1

D_0 – Inner body of the disk D, §1.3.1

$D(y)$ – Subset of $\mathbb{R}^2$ occupied by D in configuration y, §1.3.1

∂ – Topological boundary operator, §1.2.1

$\partial_{\text{reg}}\mathcal{M}$ – Set of regular points of $\partial\mathcal{M}$, §1.3.2, §4.1

$\partial_{\text{reg}}\Sigma$ – Set of regular points of $\partial\Sigma$, §1.3.1

$\partial_s\Sigma$ – Set of singular points $\partial\Sigma$, §1.3.1

$\partial_{\text{reg}}W$ – Set of regular points of ∂W, §1.3.1

$\partial_s W$ – Set of singular points of ∂W, §1.3.1

$\delta_0 = \delta_0(\epsilon)$ – Radius of a "gap region" §4.1

e_j – j'th member of the standard basis for $\mathbb{R}^3$, §1.3.2

$\widehat{e}_j = e_j/\|e_j\|$, §1.3.6

η – "Correction" diffeomorphism, §4.3.3

$\mathcal{F}$ – Full-measure open set on which $K^{\Sigma,\epsilon}$ is defined, §1.3.5, §4.3.1

$\mathcal{F}_{\text{cyl}}$ – Full-measure open set on which $K^{\Sigma,\epsilon}_{\text{cyl}}$ is defined, §4.3.2

$\widetilde{\mathcal{F}}$ – Open set on which $\widetilde{K}^{\Sigma,\epsilon}$ is defined, §4.3.3

G_ϵ – "Perturbed" projection onto $\mathbf{Q}_0$, §5.3.2

$\Gamma_{\text{roll}} = \partial\mathcal{M} \cap \widehat{\mathcal{Z}}$, §4.1

H_1 – Smooth perturbation of configuration space, §4.2

$\overline{H}_1$ – Smooth perturbation of phase space, §4.3.3

H_ϵ – Zoomed version of H_1, §5.3.1

$\overline{H}_\epsilon$ – Zoomed version of $\overline{H}_1$, (5.212)

$\mathcal{H}^d$ – d-dimensional Hausdorff measure, §4.1, §6.1.1

Int – Topological interior operator, §1.2.1

J – Moment of inertia of the disk D, §1.3.1

$K^{\Sigma,\epsilon}$ – Collision law associated with the wall $W(\Sigma,\epsilon)$, §1.3.5, §4.3.1

$\mathbb{K}^{\Sigma,\epsilon}$ – Markov kernel representation of $K^{\Sigma,\epsilon}$, §1.3.5, §4.3.1

$\mathbb{K}$ – Rough reflection law, §1.3.5,§4.3.1

$K^{\Sigma,\epsilon}_{\text{cyl}}$ – Cylindrical collision law associated with $W(\Sigma,\epsilon)$, §4.3.2

$\mathbb{K}^{\Sigma,\epsilon}_{\text{cyl}}$ – Markov kernel representation of $K^{\Sigma,\epsilon}_{\text{cyl}}$, §4.3.2

$\widetilde{K}^{\Sigma,\epsilon}$ – Modified collision law associated with $W(\Sigma,\epsilon)$, §4.3.3

$\kappa(p)$ – Unsigned curvature of ∂W at point p, §1.3.1

$\kappa_{\max} = \sup\{\kappa(p) : p \in \partial_{\text{reg}} W\}$, §1.3.1

$\overline{\kappa}$ – Max curvature with lower cutoff, (5.101)

$\overline{L}$ – Max chord length with lower cutoff, (5.101)

$\text{Lip}(f)$ – Lipschitz constant of the function f, (5.61)

$L^p = L^p_\mu(X)$ – Set of p-integrable real functions on measure space (X, μ), §1.2.1

Λ^1 – Billiard invariant measure on $\mathbb{R} \times \mathbb{S}^1_+$, §1.2.2

Λ^2 – Billiard invariant measure on $\mathbf{P} \times \mathbb{S}^2_+$, §1.3.5

m – Mass of disk D, §1.3.1

$\mathcal{M}$ – Configuration space of the disk and wall system, §1.3.2, §4.1

$\mathcal{M}_{\text{cyl}}$ – Cylindrical configuration space, §1.3.3 §4.2

$\mathcal{M}_{\text{reg}} = \mathcal{M} \setminus \mathcal{S}$, §1.3.2, §4.1

$\mathcal{M}_{\text{roll}} = \mathcal{M} \cap \widehat{\mathcal{Z}}$, §4.1

N – Number of satellites of disk D, §1.3.1

$\mathcal{O}_\pm = H_1(\mathbb{R}^3_\pm \cap \widehat{\mathcal{Z}})$, §4.3.3

$\Omega = \Omega(\epsilon)$ – §5.2, §5.3.5

P^W – Macro-reflection law associated with wall W, §1.2.2

$P^{\Sigma,\epsilon} = P^{W(\Sigma,\epsilon)}$, §1.2.4

$\mathbb{P}^{\Sigma,\epsilon} = \mathbb{P}^{W(\Sigma,\epsilon)}$, §1.2.4

$\mathbb{P}^{W}$ – Markov kernel representation of P^{W}, §1.2.2

$\mathbb{P}$ – Rough reflection law, §1.2.2, §6.2.2

$\widetilde{\mathbb{P}}$ – Velocity component of rough reflection law, §1.2.4

$\widetilde{\mathbb{P}}^{\Sigma,\epsilon}$ – Macro-reflection law averaged over one period, §2.1

$\mathbf{P} = \{(x_1, x_2, \alpha) : x_2 = 0\}$, §1.3.2

$\widetilde{\mathbf{P}} = H_1(\mathbf{P} \cap \widehat{\mathcal{Z}})$, §4.3.3

$\mathbf{Q}_0 = \{(x_1, x_2, \alpha) : \alpha - 0\}$, §1.3.3

R, R_q – Specular reflection maps, §1.3.4, §6.1.2

R_ϵ – Narrow parallelogram, §5.2

$\mathbb{R}^2_\pm = \{(x_1, x_2) \in \mathbb{R}^2 : \pm x_2 > 0\}$, §1.2.2

$\mathbb{R}^3_\pm = \{(x_1, x_2, \alpha) \in \mathbb{R}^3 : \pm x_2 > 0\}$, §1.3.5

$\rho = \rho(\epsilon)$ – Angle between consecutive satellites in D, §1.3.1

$\breve{\rho}^{-1}$ – Generalized inverse of ρ, (5.102)

S_k – k'th satellite in D, §1.3.1

$S_k(y)$ – Point in $\mathbb{R}^2$ occupied by S_k in $D(y)$, §4.1

$\mathbb{S}^1$ – Unit circle in $\mathbb{R}^2$, §1.2.2

$\mathbb{S}^1_\pm = \mathbb{S}^1 \cap \mathbb{R}^2_\pm$, §1.2.2

$\mathbb{S}^2$ – Unit sphere in $\mathbb{R}^3$ with respect to the norm $\|\cdot\|$, §1.3.5

$\mathbb{S}^2_\pm = \mathbb{S}^2 \cap \mathbb{R}^3_\pm$, §1.3.5

$\mathcal{S}$ – Set of singular points of $\mathcal{M}$, §4.1

σ_r – Scaling map, §5.3.1

Σ – Cell for building a periodic wall, §1.2.4

$\widetilde{\Sigma}$ – Foreshortened cell, §1.3.7

$\tau_{jk}, \overline{\tau}_{jk}$ – Translation maps, §4.3.1

$\tau^{(s)}_{jk}, \overline{\tau}^{(s)}_{jk}$ – Translation maps, §4.3.2

W – Wall of billiard table in $\mathbb{R}^2$, §1.2.2

$W(\Sigma, \epsilon)$ – Periodic wall determined by Σ and ϵ, §1.2.4, §1.3.1

$\widetilde{W} = W(\widetilde{\Sigma}, \widetilde{\epsilon})$ – Foreshortened wall, §1.3.7

$\mathcal{Z}$ – Set of angular coordinates where cylindrical approximation is feasible, §4.1

$\widehat{\mathcal{Z}} = \mathbb{R}^2 \times \mathcal{Z}$, §4.1

$\mathcal{Z}^0$ – Component of $\mathcal{Z}$ containing $\alpha = 0$, (4.35)

$\widehat{\mathcal{Z}}^0 = \mathbb{R}^2 \times \mathcal{Z}^0$, (4.35)

9 References

1. O. Angel, K. Burdzy, S. Sheffield, *Deterministic approximations of random reflectors*, Trans. Amer. Math. Soc. **365**(2013), no. 12, 6367–83. MR3105755

2. V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed., Springer Science+Business, Inc., New York, 1989. MR1345386 (96c:70001)

3. P. Ballard, *The dynamics of discrete mechanical systems with perfect unilateral constraints*, Arch. Ration. Mech. Anal. **154**(2000), no. 3, 199-274. MR1785473 (2002a:70011)

4. D. S. Broomhead, E. Gutkin, *The dynamics of billiards with no-slip collisions*, Phys. D **67**(1993), no. 1-3, 188-197. MR1234441 (94g:58162)

5. A. Champneys and P. Várkonyi, *The Painlevé paradox in contact mechanics*, IMA J. Appl. Math. **81**(2016), no. 3, 538-588. MR3564666

6. N. Chernov and R. Markarian, *Chaotic Billiards*, Mathematical Surveys and Monographs, 127. American Mathematical Society, Providence, RI, 2006. MR2229799 (2007f:37050)

7. S. Cook and R. Feres, *Random billiards with wall temperature and associated Markov chains*, Nonlinearity **25**(2012), no. 9, 2503-2541. MR2967115

8. I. P. Cornfield, S. V. Fomin, and Ya. G. Sinai, *Ergodic Theory*, tr. A. B. Sossinski, Springer-Verlag, Inc., New York, 1982. MR0832433 (87f:28019)

9. C. Cox, R. Feres, *No-slip billiards in dimension two*, Contemp. Math. **698**(2017), 91-110. MR3716087

10. C. Cox, R. Feres and W. Ward, *Differential geometry of rigid body collisions and non-standard billiards*, Discrete Contin. Dyn. Syst. **36**(2016), no. 11, 6065-6099. MR3543581

11. E. D. Demaine and J. O'Rourke, *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*, Cambridge University Press, New York, 2007. MR2354878 (2008g:52001)

12. R. Feres and G. Yablonsky, *Knudsen's cosine law and random billiards*, Chemical Engineering Science, **59**(2004), 1541-1556.

13. R. Feres, Random walks derived from billiards. *Dynamics, ergodic theory, and geometry*, 179-222, *Math. Sci. Res. Inst. Publ.* **54**, Cambridge University Press, Cambridge, 2007. MR2369447 (2009c:37033)

14. R. Feres and H-K. Zhang, *The spectrum of the billiard Laplacian of a family of random billiards*, J. Stat. Phys. **141**(2010), no. 6, 1030-1054. MR2740402 (2011h:37052)

15. R. Feres and H-K. Zhang, *Spectral gap for a class of random billiards*, Commun. Math. Phys. **313**(2012), no. 2, 479-515. MR2942958

16. R. C. Fetecau, J. E. Marsden, M. Ortiz, and M. West, *Nonsmooth Lagrangian mechanics and variational collision integrators*, SIAM J. Appl. Dyn. Syst. **2**(2003), no. 3, 381-416. MR2031279 (2005e:37197)

17. G. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd Ed., John Wiley & Sons, 1999. MR1681462 (2000c:00001)

18. R. Gorenflo, S. Vessella, *Abel integral equations: analysis and applications*, Lecture Notes in Mathematics 1461, Springer-Verlag, 1991. MR1095269 (92e:45003)

19. C. Hahlweg, B. Meißner, W. Zhao, and H. Rothe, *The idea of the Lambertian surface: history, idealization, and system theoretical aspects and part 1 of a lost chapter on multiple reflection*, Proc. SPIE 7792, Reflection, Scattering, and Diffraction from Surfaces II, 779202 (2010).

20. M. Hairer, *Ergodic properties of Markov processes*, Lecture notes, University of Warwick, Available on https://www.hairer.org/notes/Markov.pdf, 2006.

21. M. Knudsen, *Kinetic Theory of Gases: Some Modern Aspects*, Methuen's Monographs on Physical Subjects, London, 1952.

22. M. Lapidus and R. Niemeyer, *Towards the Koch snowflake fractal billiard: computer experiments and mathematical conjectures*, Gems in experimental mathematics, 231–263, Contemp. Math., **517**, Amer. Math. Soc., Providence, RI, 2010. MR2731085 (2012b:37101)

23. M. Mabrouk, *A unified variational model for the dynamics of perfect unilateral constraints*, Eur. J. Mech. A Solids, **17**(1998), no. 5, 819–842. MR1650957 (99j:70019)

24. D. P. Monteiro Marques, *Chocs inélastiques standards: un rèsultat d'existence*, Sèm. Anal. Convexe, **15**(1985), no. 4, 1-32. MR0857776 (88c:70010)

25. D. P. Monteiro Marques, *Differential inclusions in nonsmooth mechanical problems: shocks and dry friction*, Progr. Nonlinear Differential Equations Appl. **9**(1993), Birkhäuser Verlag, Basel. MR1231975 (94g:34003)

26. J.-J. Moreau, *Standard inelastic shocks and dynamics of unilateral constraints*, in Unilateral Problems in Structural Analysis, G. Del Piero and F. Maceri, eds., 1985.

27. J.-J. Moreau, *Unilateral contact and dry friction in finite freedom dynamics*, in Nonsmooth Mechanics and Applications, J.-J. Moreau and P. G. Panagiotopoulos, eds., CISM Courses and Lectures 302, Springer-Verlag, Vienna, 1988.

28. P. Palffy-Muhoray, E. G. Virga, M. Wilkinson and X. Zheng, *On a paradox in the impact dynamics of smooth rigid bodies*, Math. Mech. Solids, **24**(2019), no. 3, 573-597. MR3935008

29. L. Paoli and M. Schatzman, *Schéma numérique pour un modèle de vibrations avec constraints unilatérale et perte d' énergie aux impacts, en dimension finie*, C. R. Acad. Sci. Paris Sér. I Math. **317**(1993), no. 2, 211–215. MR1231424 (94f:70022)

30. L. Paoli and M. Schatzman, *Mouvement à un nombre fini de degrés de liberté avec contraintes unilatérales: Cas avec perte d'énergie*, RAIRO Modél. Math. Anal. Numér. **27**(1993), no. 6, 673–717. MR1246995 (94m:34038)

31. A. Yu. Plakhov, *Newton's problem of the body of minimal mean resistance*, Sb. Math. **195**(2004), no. 7-8, 1017-1037. MR2101335 (2005g:49073)

32. A. Yu. Plakhov, *Billiards and two-dimensional problems of optimal resistance*, Arch. Ration. Mech. **194**(2009), no. 2, 349-382. MR2563633 (2010m:49067)

33. A. Yu. Plakhov, *Billiard scattering on rough sets: two-dimensional case*, SIAM J. Math. Anal. **40**(2009), no. 6, 2155-2178. MR2481290 (2010i:37081)

34. A. Yu. Plakhov, *Scattering in billiards and problems of Newtonian aerodynamics*, Russ. Math. Surv. **64**(2009), no. 5, 873-938. MR2588685 (2011b:37067)

35. A. Yu. Plakhov, *Exterior Billiards: Systems with Impacts Outside Bounded Domains*, Springer, 2013. MR2931647

36. L. Saint-Raymond and M. Wilkinson, *On Collision Invariants for Linear Scattering*, Comm. Pure Appl. Math., **71**(2018), no. 8, 1494-1534. MR3847748

37. S. Sasaki, *On the differential geometry of tangent bundles of Riemannian manifolds II*, Tôhoku Math. J., **14**(1962), no. 2, 146-155. MR0145456 (26 #2987)

38. J.-M. Strelcyn, *Plane billiards as smooth dynamical systems with singularities*, in Invariant Manifolds, Entropy and Billiards: Smooth Maps with Singularities, Lecture Notes in Mathematics, 1222, A. Dold and B. Eckmann, eds. Springer-Verlag, Berlin, 1986. MR0872698 (88k:58075)

39. S. Tabachnikov, *Geometry and Billiards*, Student Mathematical Library, 30. American Mathematical Society, Providence, RI; Mathematics Advanced Study Semesters, University Park, PA, 2005. MR2168892 (2006h:51001)

40. L. Tonelli, *Su un problèma di Abel*, Mathematische Annalen **99**, 183-199 (1928). MR1512447

41. M. Wilkinson, *On the non-uniqueness of physical scattering for hard non-spherical particles*, Arch. Rational Mech. Anal. **235**(2020), no. 3, 2055-2083. MR4065657

42. M. Wilkinson, *On the initial boundary value problem in the kinetic theory of hard particles I: Non-existence*, 2018. arXiv:1805.04611

43. Wikipedia, *Retroreflector*, `https://en.wikipedia.org/wiki/Retroreflector` Online; accessed 5 Nov 2021.