J Stat Phys (2010) 138: 160–269
DOI 10.1007/s10955-009-9871-5

Deterministic Approach to the Kinetic Theory of Gases

József Beck

Received: 4 June 2009 / Accepted: 4 November 2009 / Published online: 1 December 2009

Abstract In the so-called Bernoulli model of the kinetic theory of gases, where (1) the particles are dimensionless points, (2) they are contained in a cube container, (3) no attractive or exterior forces are acting on them, (4) there is no collision between the particles, (5) the collision against the walls of the container are according to the law of elastic reflection, we deduce from Newtonian mechanics two local probabilistic laws: a Poisson limit law and a central limit theorem. We also prove some global law of large numbers, justifying that "density" and "pressure" are constant. Finally, as a byproduct of our research, we prove the surprising super-uniformity of the typical billiard path in a square.

Keywords Typical billiard path in a square and in a cube · Elastic reflection · Large billiard systems · Poisson limit law · Fourier analysis · Parseval's formula

1 Time-evolution in the Local Case: Poisson Limit Theorem

1.1 Where Does Randomness Come from?

This paper provides rigorous mathematical proofs to support the postulate in statistical mechanics that the particles are represented by independent random variables. I recall that the kinetic theory of gases describes gas as an accumulation of a very large (but finite) number N of rapidly moving tiny particles (N is at the order of 10^{20} per cm^3, the average speed is roughly around 10^3 meter per second at room temperature, depending on the gas, and the size of the particles is about 10^{-9}–10^{-10} meter). The particles ($=$ molecules) are colliding with one another and against the wall of the container. Hence, if for the time-point $t = 0$ we know the space coordinates

$$(x_j(0), y_j(0), z_j(0)), \quad j = 1, 2, \ldots, N \tag{1.1}$$

J. Beck (✉)
Rutgers University, New Brunswick, NJ, USA
e-mail: jbeck@math.rutgers.edu

and the velocities

$$(\dot{x}_j(0), \dot{y}_j(0), \dot{z}_j(0)), \quad j = 1, 2, \ldots, N \tag{1.2}$$

of the particles (we call (1.1) and (1.2) the initial condition), the state of the system is *theoretically* determined for the entire future $t > 0$ too. Theoretically yes, but practically no: an effective determination of even the simplest properties of gas is completely hopeless to achieve in that way. Indeed, in order to compute the time evolution of the system of N particles, we would have to deal with $6N$ equations in $6N$ variables, which is of course a totally unrealistic task if N is in the range of 10^{20}.

It is the general view among physicists, therefore, that the basic properties of gas cannot be deduced from the principles of classical mechanics alone, and this impossibility was the basis for a probabilistic treatment called "statistical mechanics". Statistical mechanics is based on (often implicit) postulates involving non-Newtonian concepts such as *probability* and *statistical independence*—the later usually combined with uniform distribution. The physicists prefer to call it "equal a priori probabilities in the phase space"; see any textbook, e.g., Tolman [18], or Thompson [17], or Uhlenbeck–Ford [19].

As an illustration, consider the famous Maxwell–Boltzmann energy law

$$\text{Probability}(\text{energy} = E_j) = \frac{e^{-\beta E_j}}{\sum_i e^{-\beta E_i}}$$

(where $1/\beta = kT$, k is the Boltzmann's constant and T is the temperature), which is generally considered the single most important law in statistical mechanics. Every known mathematical "proof" of the Maxwell–Boltzmann energy law is based on the postulate of equal a priori probabilities in the phase space.

How can we justify the Equiprobability Postulate? How does probability enter Classical Mechanics? Unfortunately, the task of finding a rigorous mathematical foundation for Statistical Mechanics remains largely unsolved. The objective of this paper is exactly to give a new insight to this long-standing open problem.

We have to admit, however, that the lack of rigorous mathematical foundations is not such a big headache for the physicists: the majority of them are pragmatists anyway. They are perfectly satisfied with the fact that Statistical Mechanics works: it can correctly predict the outcomes of (most of) the experiments. Agreement with experiment is the best substitute for a rigorous mathematical proof of the Equiprobability Postulate.

Physicists say: "try this; if it works (with reasonable level of accuracy) that will justify the postulate". In this paper I represent the viewpoint of a mathematician. With all due respect (and admiration!) to the physicists, a mathematician by training is obliged to point out the characteristic fallacy: "inductive experience that the postulate works is not a rigorous mathematical proof".

From Physics to Mathematics: Probability Theory What the physicists call *equal a priori probabilities in the phase space* is nothing else than the mathematical term (*statistical*) *independence with uniformly distributed components.* In other words, the simplest rigorous mathematical model in statistical mechanics describes the ideal gas in terms of independent and uniformly distributed random variables. More precisely, the physical system of an ideal gas of N particles in a cube container—say, the unit cube $[0, 1]^3$—is represented by N mutually independent random variables $X_1, X_2, \ldots, X_N$, where each X_j is uniformly distributed in $[0, 1]^3$, meaning that for any measurable subset $A \subset [0, 1]^3$, $\Pr[X_j \in A] = \text{volume}(A)$.

Here is a simple but important question that we can easily answer in this probabilistic model. What is the distribution of the number of particles of an ideal gas lying in a given fixed domain $A \subset [0,1]^3$ of very small volume $\mathrm{vol}(A) = \frac{1}{N}$?

Let X_A denote the number of particles lying in A; it is a random variable. The expected value of X_A is clearly 1:

$$\mathbf{E}X_A = N \cdot \frac{1}{N} = 1,$$

and for any integer $0 \le k \le N$, we have the probability

$$\Pr[X_A = k] = \binom{N}{k} \frac{1}{N^k} \left(1 - \frac{1}{N}\right)^{N-k}. \tag{1.3}$$

(Of course, (1.3) is 0 if $k > N$.) If k is fixed and $N \to \infty$, then we have the well-known limit

$$\Pr[X_A = k] = \frac{1}{k!} \left(1 - \frac{1}{N}\right)^N \cdot \frac{N(N-1)\cdots(N-k+1)}{N^k} \to \frac{1}{k!} e^{-1}, \tag{1.4}$$

which is a special case of the Poisson Limit Theorem.

If we switch the (mathematical) expectation from 1 to an arbitrary positive constant $\lambda > 0$, that is, $\mathrm{vol}(A) = \frac{\lambda}{N}$, then

$$\lim_{N\to\infty} \Pr[X_A = k] = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{1.5}$$

which is the general case of the Poisson Limit Theorem. (Note that for the Poisson distribution with parameter $\lambda > 0$ (see (1.5)) the expectation and the variance are both equal to λ.)

Next let $A_1, A_2, \ldots, A_r$ be a finite sequence of disjoint measurable subsets of the unit cube $[0,1]^3$, and assume that $\mathrm{vol}(A_i) = \lambda_i/N$, $i = 1, 2, \ldots, r$. We study the distribution of the vector-valued random variable $(X_{A_1}, X_{A_2}, \ldots, X_{A_r})$, where X_{A_i} denotes the number of particles lying in A_i (I recall that the N particles are represented by N mutually independent random variables $X_1, X_2, \ldots, X_N$, where each X_j is uniformly distributed in $[0,1]^3$). Let $k_1, k_2, \ldots, k_r$ be an arbitrary sequence of non-negative integers with $0 \le k_1 + k_2 + \cdots + k_r \le N$. We have

$$\begin{aligned}
&\Pr[(X_{A_1}, X_{A_2}, \ldots, X_{A_r}) = (k_1, k_2, \ldots, k_r)] \\
&= \binom{N}{k_1}\binom{N-k_1}{k_2}\cdots\binom{N-k_1-\cdots-k_{r-1}}{k_r}\left(\frac{\lambda_1}{N}\right)^{k_1}\left(\frac{\lambda_2}{N}\right)^{k_2}\cdots\left(\frac{\lambda_r}{N}\right)^{k_r} \\
&\quad \cdot \left(1 - \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_r}{N}\right)^{N-k_1-\cdots-k_r}.
\end{aligned}$$

If $k_1, k_2, \ldots, k_r$ are fixed and $N \to \infty$, then we have the simple limit

$$\begin{aligned}
&\lim_{N\to\infty} \Pr[(X_{A_1}, X_{A_2}, \ldots, X_{A_r}) = (k_1, k_2, \ldots, k_r)] \\
&= \frac{\lambda_1^{k_1}}{k_1!} e^{-\lambda_1} \cdot \frac{\lambda_2^{k_2}}{k_2!} e^{-\lambda_2} \cdots \frac{\lambda_r^{k_r}}{k_r!} e^{-\lambda_r}.
\end{aligned} \tag{1.6}$$

Comparing (1.5) to (1.6), it is natural to call (1.6) the Poisson *product* formula.

Let's return to (1.3). If we switch from constant to $\lambda = \lambda(N)$ with $\lambda(N) \to \infty$ but $\lambda(N)/N \to 0$ as $N \to \infty$, then (1.3) remains true with λ/N instead of $1/N$, but the limit in (1.5) is replaced with the De Moivre–Laplace limit

$$\lim_{N\to\infty} \sum_{c_1\sqrt{\lambda}\le k-\lambda\le c_2\sqrt{\lambda}} \Pr[X_A = k] = \frac{1}{\sqrt{2\pi}} \int_{c_1}^{c_2} e^{-u^2/2}\, du \tag{1.7}$$

for any fixed real numbers $-\infty < c_1 < c_2 < \infty$. Notice that (1.7) is a special case of the Central Limit Theorem (or "normal law", or "bell curve law", or "Gaussian distribution law") in the special case of the asymmetric binomial distribution.

The Poisson Limit Theorem and the Central Limit Theorem are the two most important limit theorems; they are the trademarks of probability theory.

Now we leave the probabilistic model, and return to Newtonian mechanics. The objective of this paper is to show that, in the case of a very simple deterministic model, namely, where the following five properties hold:

(1) the particles are dimensionless points of equal mass,
(2) they are contained in a cube container,
(3) no attractive or exterior forces are acting on them,
(4) there is no collision between the particles,
(5) the collisions against the walls of the container are according to the law of elastic reflection (i.e., the angle of incidence equals the angle of reflection),

we can deduce from the fundamental principles of mechanics the two probabilistic laws described in (1.3)–(1.7). More precisely, we prove that the time-evolution of the deterministic model exhibits a local Poisson Limit Theorem and a local Central Limit Theorem; see Theorem 1 below.

Also, we will prove a global *law of large numbers* implying that "the density is constant"; see Theorems 2 and 3 in Sect. 3. (The fourth main result, Theorem 4 in Sect. 4, is about the "super-uniformity" of the typical billiard paths in a square or a rectangle. The surprising message is that, the "ugliness" of the measurable subset $A \subset [0,1]^2$ we test the uniformity with, is basically irrelevant!)

We may call the model described by (1)–(5) the "Bernoulli model", after Daniel Bernoulli who introduced a similar model around 1738. What we do in this paper is a *quantitative* theory of the Bernoulli model, providing explicit error terms. Instead of relying on ergodic theory—which is considered the traditional mathematical approach to rigorous statistical mechanics—our approach is built around the Kronecker–Weyl equidistribution theorem and the use of hard Fourier analysis.

The five properties (1)–(5) of our simple deterministic model ("Bernoulli model") can be restated in an illuminating way in terms of point-billiard: N non-interacting billiard balls—each represented by a point mass—move freely inside a cube container, each one along a straight line, until one hits the wall (i.e., one of the six faces of the cube). The reflection off the wall is elastic; after the reflection the point (= billiard ball) continues its linear motion with the new velocity (but the speed remains the same; we ignore friction, air resistance, etc.) until its hits the wall again, and so forth. The same applies for all N billiard balls (= points = "molecules").

The initial condition, i.e., the starting point of the billiard path and the initial direction, uniquely determine an infinite piecewise linear billiard path $\mathbf{x}(t) = (x_1(t), x_2(t), x_3(t))$, $0 < t < \infty$ in the unit cube. For simplicity, consider first the billiard path in the unit square; the law of reflection implies that there are at most four different directions along the billiard path: the initial direction is preserved modulo $\pi/2$ (= the angle of the square), which is one-fourth of the whole angle 2π. If we switch from the unit square to the d-dimensional unit cube with any $d \geq 3$, then again the law of reflection implies that there are only a bounded number of different directions along the billiard path (of course the bound depends on the dimension d).

Let's return to the point-billiard in the unit cube. The vague term of "typical billiard path" can be made precise very easily: we just have to define a measure on the set of all initial conditions of the billiard paths. The initial condition consists of a starting point $\mathbf{y} \in [0, 1)^3$ and an initial direction $\mathbf{u} \in S^2$ (here $\mathbf{u}$ is a 3-dimensional unit vector and S^2 is the unit sphere; note that the speed remains constant as the time passes). Therefore, the corresponding measure is simply the product of the 3-dimensional Lebesgue measure in the unit cube ("volume") and the normalized surface area on the unit sphere S^2.

This way a vague term such as "$1 - \varepsilon$ part of all billiard paths" becomes perfectly precise. Similar argument works for a large system of N point-billiards (we take the product measure, which is the natural measure in the phase space).

1.2 The Trick of Unfolding

Next we explain the well-known trick of *unfolding* the billiard path inside the unit cube to a straight line in the entire 3-space. The idea is very simple and elegant: we keep reflecting the unit cube in the respective face (where the path hits the boundary) and unfold the piecewise linear billiard path ("broken line") to a straight line. We strongly recommend the reader to draw a picture in the plane, and see how the "broken" billiard path becomes a straight line via unfolding (of course in the plane the *cube* is replaced by the *square*, and the *face* is replaced by the *side*).

Two straight lines in the 3-space correspond to the same billiard path if and only if they differ by a translation through an integral vector where both coordinates are even, i.e., where the vector is from the lattice $2\mathbb{Z}+2\mathbb{Z}+2\mathbb{Z}$. In other words, the problem of the distribution of a billiard path in the unit cube is equivalent to the distribution of the corresponding torus-line in the $2 \times 2 \times 2$ cube.

As far as I know, the first appearance of the geometric trick of unfolding is in a paper of D. König and A. Szücs from 1913, and it became widely known after Hardy and Wright included it in their famous book on number theory [5]. König and Szücs used the trick of unfolding (combined with the Kronecker–Weyl theorem) to prove the following elegant property of the billiard path in a square: if the slope of the initial direction is rational, then the billiard path is periodic, and if the slope of the initial direction is irrational, then the billiard path is dense, and what is more, it is uniformly distributed in the unit square (see [5]). Notice that the analog statement for torus-lines is the famous Kronecker–Weyl equidistribution theorem (I will return to the Kronecker–Weyl theorem later in Sects. 2–4).

1.3 Time-evolution in the Deterministic Bernoulli Model: Theorem 1

In our simplistic Bernoulli model the particles don't collide with one another, so the speed v_k of the kth particle remains constant, and, as I said above, the velocity is also basically constant: the velocity (= time-derivative)

$$\dot{\mathbf{x}}_j(t) = (\dot{x}_{j,1}(t), \dot{x}_{j,2}(t), \dot{x}_{j,3}(t))$$

of the jth particle can have only a few different values as $0 < t < \infty$ (due to the elastic collisions against the walls, a consequence of the right angles in the cube). We can say, therefore, that, unlike the position, the velocity of a fixed particle (= billiard ball) does not "mix" as $0 < t < \infty$. To make our deterministic model more realistic, we could easily assume that the speeds $v_1, \dots, v_N$ of the N particles satisfy the Maxwellian distribution (= normal distribution). But because the proof is rather complicated and the notation is quite messy, for simplicity we decided to restrict ourselves to the special case where all speeds are equal:

$$v_1 = v_2 = \cdots = v_N = v \geq 1.$$

Theorem 1 *Assume that N non-interacting billiard balls, each represented by a point mass, move freely inside the unit cube $I^3 = [0,1]^3$ such that the reflection off the wall (= side of the cube) is elastic. Let $\mathbf{x}_j(t) = (x_{j,1}(t), x_{j,2}(t), x_{j,3}(t))$ describe the trajectory of the jth billiard ball (= point) in the time interval $0 \leq t \leq T$, where $\mathbf{x}_j(0) = \mathbf{y}_j$ is the initial position, $\dot{\mathbf{x}}_j(0) = v \cdot \mathbf{u}_j$ is the initial velocity and $v > 0$ is the common speed ($\mathbf{u}_j$ is a unit vector, i.e., $\mathbf{u}_j \in S^2 =$ unit sphere; $N \geq 1$, $T > 1$ and $v > 1$ are arbitrary, but the theorem becomes interesting only if N and vT are both large).*

Let $A \subset I^3$ be an arbitrary Lebesgue measurable subset of the unit cube with volume $\mathrm{vol}(A) = \lambda/N$ (the range of parameter $\lambda > 0$ will be given in (1.9) below). Let $Y_A(t)$ denote the point-counting function:

$$Y_A(t) = \sum_{\substack{1 \leq j \leq N: \\ \mathbf{x}_j(t) \in A}} 1.$$

Let

$$m = \min\left\{ \frac{e^{\frac{1}{2}\sqrt{\log(vT)}}}{101}, \sqrt{N} \right\} \quad \text{and} \quad \varepsilon = \frac{1}{\sqrt{m}}, \tag{1.8}$$

where log *denotes the natural (i.e., base e) logarithm.*

Then for more than $1 - \varepsilon$ part of the initial conditions

$$\omega = (\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N) \in \left(I^3\right)^N \times \left(S^2\right)^N = \Omega$$

(in the sense of the product measure on Ω), the distribution of the point-counting function

$$Y_A(\omega; t) = Y_A(\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N; t)$$

is very close to the Poisson distribution with parameter λ (assuming N and vT are both large) in the following quantitative sense: for every real number

$$0 < \lambda \leq \frac{\log m}{8} \tag{1.9}$$

and every integer $k \geq 0$,

$$\left| \frac{1}{T} \mathrm{measure}\{0 \leq t \leq T : Y_A(\omega; t) = k\} - \frac{\lambda^k}{k!} e^{-\lambda} \right| < \varepsilon. \tag{1.10}$$

Finally, we can generalize (1.10) to get the following analog of the product formula (1.6). Let $A_1, A_2, \dots, A_r$ be an arbitrary finite sequence of disjoint measurable subsets of the unit cube $[0,1]^3$ with $\mathrm{vol}(A_i) = \lambda_i/N$, $i = 1, 2, \dots, r$. Then for more than $1 - \varepsilon$ part of the *initial conditions* $\omega \in \Omega$, (*in the sense of the product measure on* Ω), *the distribution of the point-counting function*

$$\left| \frac{1}{T}\text{measure}\{0 \le t \le T : (Y_{A_1}(\omega; t), \ldots, Y_{A_r}(\omega; t)) = (k_1, \ldots, k_r)\} - \frac{\lambda_1^{k_1}}{k_1!}e^{-\lambda_1} \cdots \frac{\lambda_r^{k_r}}{k_r!}e^{-\lambda_r} \right| < r \cdot \varepsilon \tag{1.11}$$

holds for all $r \ge 1$, *all vectors* $(k_1, \ldots, k_r)$ *of non-negative integers*, *and all*

$$0 < \lambda_i \le \frac{\log m}{8} \quad (1 \le i \le r).$$

Remarks (a) In statistical mechanics one usually studies the limit process, sometimes called thermodynamics, where the ratio *particle/volume* remains a fixed constant as $N \to \infty$. More precisely, we replace the unit cube with a large cube of volume N/λ (i.e., the side length is $(N/\lambda)^{1/3}$) with some fixed constant $\lambda > 0$, and consider the limit $N \to \infty$ (i.e., the number of particles tends to infinity). For every N, let $A = A(N)$ be an arbitrary measurable subset of volume$(A(N)) = 1$ (a subset of the large cube of volume N/λ), and we study the distribution of the number of particles in $A = A(N)$ during a long time-interval $0 < t < T$. Theorem 1 makes it possible to carry out the limit process *precisely*. Theorem 1 implies that, independently of the way we take the double limit $N \to \infty$ and $T \to \infty$ (i.e., the relative relation of N and T is totally irrelevant), the number of particles in $A = A(N)$ has a definite *limit distribution*, which is the Poisson distribution with parameter $\lambda > 0$. We emphasize that the *relevant* limit in statistical mechanics is when first T is fixed and $N \to \infty$, and then, in the second step, $T \to \infty$.

(b) It is remarkable that the "complexity" of the given subset $A \subset [0, 1]^3$ does *not* play any role in the theorem. Of course we cannot say anything nontrivial about *all* possible measurable $A \subset [0, 1]^3$ simultaneously (since the volume of a billiard path is zero). We can easily generalize, however, Theorem 1 for an arbitrary infinite sequence $A_1, A_2, A_3, \ldots$ of measurable subsets of the unit cube with $\text{vol}(A_i) = \lambda_i/N$. The only necessary modification in (1.10) is to insert a *weight factor* in the upper bound:

$$\left| \frac{1}{T}\text{measure}\{0 \le t \le T : Y_{A_i}(\omega; t) = k\} - \frac{\lambda_i^k}{k!}e^{-\lambda_i} \right| < i\varepsilon$$

for all $i = 1, 2, 3, \ldots$.

(c) Theorem 1 is about a single time-interval $0 \le t \le T$, where T is arbitrary but fixed. It is natural to ask what happens in the sequential case, that is, when we study the distribution of the point-counting function $Y_A(t)$ as t runs in $0 \le t \le T$ simultaneously for all $0 < T < T_0$, where T_0 is some large real number. It is not too difficult to prove such a sequential version of Theorem 1 by using a straightforward adaptation of the so-called *dyadic method*, originally developed for orthogonal series.

We can "sequentialize" Theorem 1 as follows. For every

$$2 < T < e^{N^{1/8}} \tag{1.12}$$

write

$$m(T) = \min\left\{\frac{e^{\frac{1}{2}\sqrt{\log(vT)}}}{101}, \sqrt{N}\right\} \quad \text{and} \quad \varepsilon(T) = \frac{1}{\sqrt{m(T)}},$$

then (say) for more than 99.99 percent of the initial conditions

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \left(I^3\right)^N \times \left(S^2\right)^N = \Omega$$

the distribution of the point-counting function

$$Y_A(\omega; t) = Y_A(\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N; t)$$

is very close to the Poisson distribution with parameter λ in the following sense:

$$\left|\frac{1}{T}\text{measure}\{0 \le t \le T : Y_A(\omega; t) = k\} - \frac{\lambda^k}{k!}e^{-\lambda}\right| < \varepsilon(T) \tag{1.13}$$

holds for every T in (1.12), for every real number $0 < \lambda \le \frac{\log m(T)}{8}$ and for every integer $k \ge 0$.

Note that in the kinetic theory of gases the number of particles is $N \approx 10^{24}$, so the upper bound for T in (1.12) is in the range of e^{10^3}, which is "effectively infinite".

The message of this sequential version of Theorem 1 is the following: as more and more time passes, the distribution of the point-counting function $Y_A(t)$ gets closer and closer to the Poisson distribution, and the speed of convergence is basically independent of the number of particles. Nevertheless, the number of particles is crucial in an indirect way: it gives a natural limitation to the Poisson approximation.

(d) We can give an "ergodic theorem type" interpretation of Theorem 1 in the sense of the equality

$$\text{space-average} = \text{time-average}.$$

Indeed, at the beginning $t = 0$, the initial positions $\mathbf{x}_j(0) = \mathbf{y}_j$, $1 \le j \le N$ of the N point-billiards are independent and uniformly distributed random variables (uniformly distributed in the unit cube $[0, 1]^3$). So the number of points $Y_A(\omega; 0)$ at the start $t = 0$ in a given (measurable) subset $A \subset [0, 1]^3$ of volume $\text{vol}(A) = \lambda/N$ (ω is the complete initial condition) has the binomial distribution (let $0 \le k \le N$):

$$\Pr[Y_A(\omega; 0) = k] = \binom{N}{k}\left(\frac{\lambda}{N}\right)^k \cdot \left(1 - \frac{\lambda}{N}\right)^{N-k} \to \frac{\lambda^k}{k!}e^{-\lambda}$$

that approximates the Poisson distribution with parameter $\lambda > 0$ as λ is fixed and $N \to \infty$. On the other hand, by (1.10) and (1.13), for a typical but *fixed* initial condition $\omega_0 \in \Omega$, the time evolution of the counting function $Y_A(\omega_0; t)$ approximates the same Poisson distribution as $0 \le t \le T \to \infty$. Therefore, we can roughly say that

$$\text{space-average} = \text{Poisson distribution} = \text{time-average},$$

which is indeed in the spirit of the ergodic theorem. There is, however, a fundamental difference between Theorem 1 and (say) Birkhoff's individual ergodic theorem. The ergodic theorem is a soft/qualitative result: it does not say anything about the speed of convergence; it does not give any error term. Theorem 1, on the other hand, is a hard/quantitative result:

it gives an explicit (and not too bad) error term describing the speed of convergence to the Poisson distribution.

(e) The condition (1.9) means that parameter λ can be arbitrarily large. It is a classical result in probability theory that for large values of parameter λ the Poisson distribution can be well-approximated with the normal distribution: for any fixed real numbers $-\infty < c_1 < c_2 < \infty$,

$$\lim_{\lambda\to\infty} \sum_{c_1\sqrt{\lambda}\le k-\lambda\le c_2\sqrt{\lambda}} \frac{\lambda^k}{k!}e^{-\lambda} = \frac{1}{\sqrt{2\pi}}\int_{c_1}^{c_2} e^{-u^2/2}\,du. \tag{1.14}$$

The proof of (1.14) is a routine application of Stirling's formula; the following variant is particularly precise and useful:

$$\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\cdot e^{\frac{1}{12n+1}} < n! < \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\cdot e^{\frac{1}{12n}}. \tag{1.15}$$

Also, by using (1.15) we can easily obtain a very good estimation on the speed of convergence in (1.14).

(f) Notice that vT is the distance made by either billiard ball (= point) in the time interval $0 \le t \le T$. Theorem 1 becomes interesting when both N and vT are large. In kinetic theory of gases N is enormous (the number of molecules in a standard unit volume is around $N = 10^{23}$) and vT is also "large". Indeed, the hydrogen gas at room temperature has mean speed around $v = 10^3$ meter per second. If the container is a cube of side length (say) 10^{-1} meter, then in one-second time (i.e., $T = 1$) a hydrogen molecule travels a distance about 10^3 meters, which is ten-thousand times the size of the container. It is fair to call 10^4 "large".

In (1.8) we defined the key parameter m in terms of a subpolynomial function

$$e^{\sqrt{\log(vT)}} \tag{1.16}$$

of vT. I conjecture that the subpolynomial function in (1.16) can be substantially improved; probably up to a small power of vT; perhaps even to (say) $\sqrt{vT}$. (Needless to say, the constant factor 101 in (1.8) is "accidential"; we didn't make any serious effort to find a better constant.)

(g) In our simplistic Bernoulli model the "mixing" of the system comes exclusively from the effect of point particles elastically reflected by a flat wall (one of the six faces of a cube). This is not too realistic for several reasons. One reason is that the wall, though it looks flat at a macroscopic level, certainly shows a complicated/detailed non-flat structure at a microscopic level (i.e., at the level of the gas molecules).

Another reason is that real world molecules are not dimensionless points; they have a well-defined size roughly around 10^{-10} meter (this remarkable fact is known since the 1860s). This size is not negligible, and this is why in normal conditions (say, at room temperature) the collision of a particle with another particle is several thousand times more likely than the collision against the wall. (This explains why, despite the large speed of the molecules, gases mix relatively slowly. For example, suppose hydrogen sulfide is generated at one end of a room; it may take a couple of minutes before the odor is noticed at the other end.) The collision of the molecules one another is the source of "mixing" the individual velocities. An attempt to understand and describe the effect of molecule-molecule collision was the main motivation for the theory of "dispersing and semi-dispersing billiards"—initiated by Sinai and his school—developed in the last 40 years.

The subject of dispersing (or scattering, or chaotic) billiards has a large literature now. I am certainly not an expert, so I just briefly mention that the most important tool they use is the “Markov partition”, which is very different from what I am doing here. I refer the reader to the book *Chaotic Billiards* by N. Chernov and R. Makarian [2]; see also the volume *Hard Balls and the Lorentz Gas* (Encyclopedia of Mathematical Sciences, vol. 101, Mathematical Physics II, editor: D. Szász, Springer, 2000) and the recent papers of N. Simányi [11, 12], which are about arbitrary number of particles, and are regarded the strongest known results. (Unfortunately, these results assume that the mass-distribution of the particles is random (“generic”), which is not too realistic in physics.)

(h) Theorem 1 works in any dimension $d \geq 1$. The obvious reason why I stated the result in the special case $d = 3$ is the motivation from physics.

(i) Summarizing, we can say that, our simplistic flat-wall-collision model (“Bernoulli model”) has a very limited source of “mixing” (in particular, the individual velocities do not mix at all—since we are ignoring the scattering molecule-molecule collisions). Nevertheless, even this very restrained model can exhibit remarkable randomness in the form of an explicit Poisson Limit Theorem and a Central Limit Theorem. This demonstrates that rigorous mathematical reasoning based strictly on classical mechanics can prove some of the signature laws of the probabilistic model. (Needless to say, the extreme simplicity of the Bernoulli model greatly helps to overcome the mathematical difficulties.)

(j) It is interesting to point out that Theorem 1 remains true if we replace the Bernoulli model with the so-called *exchange-velocity-at-impact hard ball* model, in which we represent the particles by solid balls of equal size with small but positive radius and equal mass. We permit collisions of two—but not three or more—particles (= tiny balls), and make the simplifying assumption that at impact they simply exchange their velocities. In other words, by changing the labels of the two particles, they simply continue their way as if no collision had happened. This gives back the Bernoulli model, and this is why Theorem 1 applies. (Note, however, that this way—i.e., by changing the labels of the particles involved in the collision—we lose to some extent the deterministic character of the model.)

Note that the possibility of a *deterministic* kinetic theory of gases fascinated mathematicians for a long time. I refer the reader to the early works of Steinhaus [15, 16], Egerváry and Turán [4], and Hlawka [6] (are there other works in the subject that I don’t know about?).

There is also a rather vague and informal argument in Lecture 8 (“Ergodic theory of an ideal gas”) in Sinai’s book [13] about the one-dimensional ideal gas that the reader may find interesting, and want to compare to what we are doing in this paper. See also the paper of Sinai and Volkovsky [14], and perhaps we can also mention the somewhat related paper of Prosser [10].

I should point out that there are three more theorems (see Sects. 3–4 for the details) that are worthwhile to be briefly mentioned here. Theorem 1 clarifies the probabilistic aspects of the time-evolution in the local case, i.e., when the expected number of point-billiards in $A \subset [0, 1]^3$ is just a constant (or it tends to infinity very slowly compared to the number of points N). But what happens in the global case when $\mathrm{vol}(A)$ jumps up from $O(1/N)$ to the constant range such as (say) $1/10 < \mathrm{vol}(A) < 9/10$? That is, when the expected number of point-billiards in A is as large as constant times N? I will discuss this question in Sect. 3 (see Theorems 2 and 3). Also, in Sect. 4 I will discuss a surprising byproduct of this research: the astonishing “super-uniformity” of the typical billiard paths in a square (or rectangle). Roughly speaking, the set of typical billiard paths represents the family of most uniformly distributed curves in the square. I will prove a quantitative result justifying this vague statement (see Theorem 4).

Before talking about these global results in Sects. 3–4, in the next section I give some intuition behind the complicated proof of Theorem 1.

2 Some Guiding Intuitions

2.1 First Guiding Intuition: Kronecker–Weyl Equidistribution Theorem

The proof of Theorem 1 is very long, so it is crucial to emphasize the main reason why it works. If I have to put it in one sentence, then I would say: it is the continuous version of the well-known Kronecker–Weyl equidistribution theorem (a fundamental result in uniform distribution) that I consider the soft/qualitative reason behind the hard/quantitative Theorem 1. Let $\alpha_1, \alpha_2, \ldots, \alpha_d$ be linearly independent over the rationals (this is a "typical" property) and let $R = \prod_{i=1}^{d}[a_i, b_i]$ be an arbitrary rectangular box (i.e., Cartesian product of intervals) in the d-dimensional unit cube $[0,1]^d$ (where $d \geq 1$ is an arbitrary integer). The Kronecker–Weyl theorem states that

$$\lim_{T\to\infty} \frac{1}{T}\text{meas}\{0 \leq t \leq T : (t\alpha_1, t\alpha_2, \ldots, t\alpha_d) \in R \pmod 1\} = \text{vol}(R) = \prod_{i=1}^{d}(b_i - a_i), \quad (2.1)$$

where of course "meas" stands for the one-dimensional Lebesgue measure and "vol" stands for the d-dimensional volume. The message of (2.1) is that the d events "$t\alpha_i \in [a_i, b_i] \pmod 1$", $1 \leq i \leq d$ become independent in the limit as the length T of the interval of consideration $0 \leq t \leq T$ tends to infinity. We can say, therefore, that (2.1) is an "asymptotic product rule"; on the other hand, (statistical) independence is perfectly characterized by the product rule.

The next step is to recall the well-known fact that the σ-algebra of Lebesgue measurable sets is generated by the family of rectangular boxes (Cartesian product of intervals) $\prod_{i=1}^{d}[a_i, b_i]$. Putting these together, we obtain the chain of implications

$$\text{Kronecker–Weyl theorem} \Longrightarrow \text{weak independence} \Longrightarrow \text{Poisson's limit theorem}, \quad (2.2)$$

which is the guiding intuition of the long proof of Theorem 1.

I have to point out that the connection between the Kronecker–Weyl theorem and (statistical) independence was already noticed by M. Kac (see his argument in Sect. 3.5 in his well-known book [7]), and also it was implicitly stated in Sinai's book [13]. It is fair to say, therefore, that the first half of (2.2):

$$\text{Kronecker–Weyl theorem} \Longrightarrow \text{weak independence}$$

is a folklore observation/intuition. Of course, there is a world of difference between a vague intuition and a rigorous mathematical proof. To illustrate this, note that Theorem 1 is a local result, and the problem of finding a global analog of Theorem 1 (i.e., a global central limit theorem for the time-evolution) leads to some unexpected results. I will return to this exciting question in Sect. 3.

It is worth while to briefly discuss here a quantitative version of the qualitative (2.1). For simplicity I restrict myself to the discrete sequence $n\mathbf{a} \pmod 1$, $n = 1, 2, 3, \ldots$ (where $\mathbf{a} = (\alpha_1, \ldots, \alpha_d)$), and we test uniformity with respect to the family of all axes-parallel boxes $R = I_1 \times \cdots \times I_d$ where $I_i = [a_i, b_i] \subset [0, 1]$: let

$$\Delta(\mathbf{a}; N) = \max_{\substack{1\leq m\leq N \\ I_i\subset[0,1]:1\leq i\leq d}} \left| \sum_{\substack{1\leq k\leq n: \\ k\mathbf{a}\in I_1\times\cdots\times I_d \pmod 1}} 1 - n|I_1|\cdots|I_d| \right|$$

denote the *discrepancy* function. What can we say about the asymptotic behavior of the discrepancy function $\Delta(\mathbf{a}; N)$ as $N \to \infty$ for almost every $\mathbf{a} = (\alpha_1, \ldots, \alpha_d)$?

In 1923 Khinchine [8] solved the one-dimensional problem: he proved that for almost every α, the discrepancy function $\Delta(\alpha; N)$ is between $\log N \cdot \log\log N$ and $\log N \cdot (\log\log N)^{1+\varepsilon}$. His proof made use of the theory of continued fractions (see also his book [9]).

Unfortunately, the classical theory of continued fraction does not seem to generalize in higher dimensions, so the multidimensional generalization of Khinchine's theorem remained an open problem for a long time. Finally, in 1994 I [1] succeeded to prove it by using Fourier analysis (instead of continued fractions): for every $d \ge 1$ and for almost every $\mathbf{a} = (\alpha_1, \ldots, \alpha_d)$, the discrepancy function

$$\Delta(\mathbf{a}; N) \text{ is between } (\log N)^d \cdot \log\log N \text{ and } (\log N)^d \cdot (\log\log N)^{1+\varepsilon}. \tag{2.3}$$

What is more, we can upgrade (2.3) to a precise convergence-divergence criterion.

Despite the big differences between (2.3) and Theorem 1 (one is about the *family* of all axes-parallel boxes and the other one is about an arbitrary but *fixed* measurable subset A), there is an important similarity: both proofs use "hard" Fourier analysis (and both are very long).

2.2 Comparing Theorem 1 to the Erdős–Kac Theorem

It is very instructive to compare the "asymptotic independence" in (2.1) and our Theorem 1 to the well-known Erdős–Kac theorem about the number of prime factors of typical integers. Note that beyond games of chance it is very hard to find "natural" examples of perfect independence in mathematics, and the unique factorization property of integers gives an example where at least some kind of an *almost* independence arises in a natural way. Let p be a prime number, and let $X_p = X_p(n)$ be a function defined on the set of natural numbers as follows: $X_p(n) = 1$ if n is divisible by p, and 0 otherwise. In other words, $X_p = X_p(n)$ is the characteristic function of the set of multiples of p. For any integer $r \ge 2$ and for any set $2 \le p_1 < p_2 < \cdots < p_r$ of r different primes we have

$$|\{1 \le n \le N : X_{p_1}(n) X_{p_2}(n) \cdots X_{p_r}(n) = 1\}| = \left\lfloor \frac{N}{p_1 p_2 \cdots p_r} \right\rfloor \tag{2.4}$$

(where $\lfloor y \rfloor$ denotes the lower integral part of y), and so

$$\lim_{N\to\infty} \frac{1}{N} |\{1 \le n \le N : X_{p_1}(n) X_{p_2}(n) \cdots X_{p_r}(n) = 1\}| = \frac{1}{p_1 p_2 \cdots p_r}, \tag{2.5}$$

where I used the standard notation $|\cdots|$ to denote the number of elements of a finite set. Equation (2.5) is an "asymptotic product rule" (similarly to (2.1)), expressing the vague intuition "the primes are independent" in a precise statement. Of course (2.5) is not a deep result, but it is important: it was the starting point of the fascinating study of describing the number of prime factors of typical integers—a subject where we can supplement the "apparent randomness" with rigorous proofs. Using the traditional notation in number theory, we define $\omega(n)$ as the number of different prime factors of n, and $\Omega(n)$ as the total number of prime factors (i.e., each prime is counted with multiplicity). Thus, for $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$ we have

$$\omega(n) = r \quad \text{and} \quad \Omega(n) = \alpha_1 + \alpha_2 + \cdots + \alpha_r.$$

Both $\omega(n)$ and $\Omega(n)$ behave very irregularly as $n \to \infty$. The minimum is 1, and it is attained for the primes. On the other hand, $\Omega(n)$ can be as large as $\log_2 n = \log n / \log 2$, and it happens for $n = 2^k$; $\omega(n)$ can be as large as $(1 + o(1)) \log n / \log\log n$, and asymptotic equality is attained for the products $n = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdots p_r$ of the small primes.

Both $\omega(n)$ and $\Omega(n)$ show "apparent randomness", which is very plausible from the fact that

$$\omega(n) = \sum_p X_p(n), \tag{2.6}$$

that is, ω is the sum of the almost independent X_ps (see (2.4)–(2.5)). Of course, in (2.6) we can restrict the infinite summation to $p \le n$.

By using standard number theory, it is easy to show that the average order of both $\omega(n)$ and $\Omega(n)$ is $\log\log n$. The pioneering result about $\omega(n)$ (and $\Omega(n)$) was proved by Hardy and Ramanujan in 1917. They showed that, for the overwhelming majority of n, $\omega(n)$ falls into the interval

$$\log\log n - (\log\log n)^{1/2+\varepsilon} < \omega(n) < \log\log n + (\log\log n)^{1/2+\varepsilon}. \tag{2.7}$$

That is, the typical fluctuation of $\omega(n)$ (and $\Omega(n)$) around the expected value $\log\log n$ is (roughly speaking) at most square root size. This line of research culminated in the following elegant central limit theorem proved in 1939.

Theorem (Erdős–Kac Theorem) *Let $-\infty < a < b < \infty$ be arbitrary reals; then the density of integers n for which*

$$\log\log n + a\sqrt{\log\log n} < \omega(n) < \log\log n + b\sqrt{\log\log n}$$

is given by the integral

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx.$$

The same holds for $\Omega(n)$.

Note that these results are capable of far-reaching generalizations: this led to the developments of a new branch called probabilistic number theory.

Let me summarize the similarity between Theorem 1 and the Erdős–Kac theorem. Both are based on relatively simple "asymptotic product rules" (see (2.1) and (2.5)), but the switch from the "soft intuition" to the "hard theorem" is far from easy—it fits to say that in both cases the devil is in the details. (For example, in the case of the Erdős–Kac theorem, Kac had the right intuition, but got stuck on the technical details, and was helped out by the number-theoretic expertise of Paul Erdős.)

2.3 Second Guiding Intuition: The Poisson Point Process as a "Fixpoint"

In the well-known book on ergodic theory [3] there is an infinite model that the authors call the "ideal gas in the d-dimensional space". This construction is not used in our proof of Theorem 1; nevertheless, it is definitely worth while to be mentioned here as a guiding intuition. The reason is that in this infinite model the Poisson point process turns out to be *invariant under the time evolution*. This crucial fact gives a new insight to the central role of the Poisson distribution.

The infinite model in [3] is rather different from our Bernoulli model of “finitely many non-interacting point particles in a box container”. The infinite model represents the ideal gas as “infinitely many non-interacting point particles in the whole d-dimensional space $\mathbb{R}^d$ with no walls”. The precise definition goes as follows (see Sect. 1 in Chap. 9 of [3]).

We work with the usual space coordinates $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, the velocity $\mathbf{v} = (v_1, \ldots, v_d) \in \mathbb{R}^d$, and consider the pair

$$(\mathbf{x}, \mathbf{v}) = (x_1, \ldots, x_d, v_1, \ldots, v_d) \in \mathbb{R}^d_x \oplus \mathbb{R}^d_v = \mathbb{R}^{2d}.$$

The constant speed linear motion is described by the simple differential equations

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}, \qquad \frac{d\mathbf{v}}{dt} = \mathbf{0}. \tag{2.8}$$

This defines the following one-parameter group (“time evolution”):

$$S^t(\mathbf{x}, \mathbf{v}) = (\mathbf{x} + \mathbf{v}t, \mathbf{v}). \tag{2.9}$$

A well-known theorem of Liouville yields that every product measure of the form

$$d\rho = d\mathbf{x}\, f(\mathbf{v}) d\mathbf{v} \tag{2.10}$$

is invariant under the one-parameter group S^t (“time evolution”) defined in (2.9). The function $f(\mathbf{v})$ in (2.10) is positive; it means any density function with

$$\int_{-\infty}^{\infty} f(\mathbf{v})\, d\mathbf{v} = 1. \tag{2.11}$$

Of course, in statistical mechanics the natural choice for $f(\mathbf{v})$ is the Maxwell distribution

$$f(\mathbf{v}) = \text{const} \cdot e^{-\beta \mathbf{v}\cdot\mathbf{v}},$$

where $\beta > 0$ and $\mathbf{v} \cdot \mathbf{v} = v_1^2 + \cdots + v_d^2$ is the usual dot product.

Let

$$Y = \{(\mathbf{x}_1, \mathbf{v}_1), (\mathbf{x}_2, \mathbf{v}_2), (\mathbf{x}_3, \mathbf{v}_3), \ldots\} \subset \mathbb{R}^{2d}$$

be a *locally finite* set, which means that for every bounded subset $B \subset \mathbb{R}^{2d}$, the intersection $Y \cap B$ is finite; formally, $|Y \cap B| < \infty$. Note that Y represents the set of point-particles in a given instant. Let $\mathcal{Y}$ denote the space of all locally finite sets $Y \subset \mathbb{R}^{2d}$. Let $\mathcal{C}$ denote the smallest σ-algebra such that all counting functions $\chi_B = \chi_B(Y) = |Y \cap B|$ are measurable, where B runs through the bounded Borel sets in $\mathbb{R}^{2d}$.

In other words, $\mathcal{C}$ is the smallest σ-algebra containing all sets

$$C_{B,k} = \{Y \in \mathcal{Y} : |Y \cap B| = k\},$$

where $B \subset \mathbb{R}^{2d}$ is any bounded Borel set and $k \geq 0$ is any integer.

To define the Poisson point process, we introduce a measure $\mu = \mu(\rho)$ on the σ-algebra $\mathcal{C}$ as follows. For any bounded Borel set $B \subset \mathbb{R}^{2d}$ and any integer $k \geq 0$, let

$$\mu\left(C_{B,k}\right) = \frac{(\rho(B))^k}{k!} e^{-\rho(B)}, \tag{2.12}$$

and if two bounded Borel set $B_1, B_2 \subset \mathbb{R}^{2d}$ are disjoint, then

$$\mu\left(C_{B_1,k_1} \cap C_{B_2,k_2}\right) = \mu\left(C_{B_1,k_1}\right)\mu\left(C_{B_2,k_2}\right) \tag{2.13}$$

holds for all integers $k_1 \geq 0$ and $k_2 \geq 0$. Of course, (2.12) is the Poisson distribution, and (2.13) means that the random variables $\chi_{B_1} = \chi_{B_1}(Y)$ and $\chi_{B_2} = \chi_{B_2}(Y)$ (where $Y \in \mathcal{Y}$) are independent. Finally, note that, by using (2.12) and (2.13), there is a unique way to extend $\mu = \mu(\rho)$ to the whole σ-algebra generated by the sets $C_{B,k}$, i.e., to the σ-algebra $\mathcal{C}$. Thus we obtain a probability space $(\mathcal{Y}, \mathcal{C}, \mu = \mu(\rho))$, where $d\rho = d\mathbf{x} f(\mathbf{v}) d\mathbf{v}$ (see (2.10)) is a special product measure on $\mathbb{R}^{2d}$. We call $\mu = \mu(\rho)$ the Poisson measure.

The key property of the Poisson measure $\mu = \mu(\rho)$ is that it is invariant under the "time flow". More precisely, for every locally finite set

$$Y = \{(\mathbf{x}_1, \mathbf{v}_1), (\mathbf{x}_2, \mathbf{v}_2), (\mathbf{x}_3, \mathbf{v}_3), \ldots\} \in \mathcal{Y},$$

and for every real number t (= time), write

$$\begin{aligned} T^t Y &= \left\{S^t(\mathbf{x}_1, \mathbf{v}_1), S^t(\mathbf{x}_2, \mathbf{v}_2), S^t(\mathbf{x}_3, \mathbf{v}_3), \ldots\right\} \\ &= \{(\mathbf{x}_1 + \mathbf{v}_1 t, \mathbf{v}_1), (\mathbf{x}_2 + \mathbf{v}_2 t, \mathbf{v}_2), (\mathbf{x}_3 + \mathbf{v}_3 t, \mathbf{v}_3), \ldots\}. \end{aligned}$$

We call T^t the *time flow*. (Note that T^t is well-defined on $\mathcal{Y}$: T^t takes a locally finite set Y into another locally finite set.)

The key fact that the Poisson measure $\mu = \mu(\rho)$ is invariant under the time flow T^t is a simple consequence of Liouville's theorem. Indeed, we have

$$\mu\left(T^t C_{B,k}\right) = \mu\left(C_{S^t B,k}\right) = \mu\left(C_{B,k}\right),$$

where in the last step we used that $d\rho = d\mathbf{x} f(\mathbf{v}) d\mathbf{v}$ is invariant under the one-parameter group S^t ("time evolution"), i.e., Liouville's theorem.

(It resembles the case of the central limit theorem, where the key property of the normal distribution is that it is the fixpoint of the Fourier transform. Here in our case the Poisson measure is also a "fixpoint": it is invariant under the time flow.)

The book [3] calls the dynamical system

$$(\mathcal{Y}, \mathcal{C}, \mu(\rho), T^t) \tag{2.14}$$

the "ideal gas in the d-dimensional space $\mathbb{R}^d$". The dynamical system (2.14) turns out to be *ergodic* (in fact, it has the stronger property of mixing). Ergodicity means that we can apply Birkhoff's individual ergodic theorem: it says that, for almost every initial condition (say, at $t = 0$), the time average of the event "a bounded Borel set $B \subset \mathbb{R}^{2d}$ contains k point-particles" converges to the Poisson distribution

$$\frac{(\rho(B))^k}{k!} e^{-\rho(B)}$$

as $t \to \infty$. This is similar to the statement of Theorem 1, and thus gives a deep insight of the problem.

2.4 Ergodic Theory Versus Fourier Analysis: Soft vs. Hard

I have to emphasize, however, that the last arguments—the Poisson point process and the ergodic theorem—are *not* used in our proof of Theorem 1. This is not surprising at all: the ergodic theorem is a "soft" result in the sense that it does not say anything about the speed of convergence. The main point in Theorem 1 is *exactly* that it is a hard/quantitative result with an explicit error term, where the error term—surprisingly!—does not depend on the ugliness of the test set. This kind of quantitative statement is beyond the power of the ergodic theorem.

We can thus say that Theorem 1—a result about the time evolution of a system—is basically a "hard ergodic theorem" (well, at least in a special case). The ergodic theorem is generally considered a difficult/deep result; it is not surprising, therefore, that our proof of Theorem 1 is complicated.

In Theorem 1 the number of point-particles N can be arbitrarily large, and we can carry out the limit where the time interval T is fixed and the number of particles N tends to infinity. This *is* the relevant limit in statistical mechanics, where the number of particles is astronomically large (in the range of the Avogadro number $\approx 10^{24}$), but the time is relatively short. The fundamental weakness of applying Birkhoff's ergodic theorem in statistical mechanics is exactly the "wrong limit". In the ergodic theorems the *time tends to infinity*, and these statements are "soft" by nature: they simply cannot provide any information on the speed of convergence. In other words, the ergodic theorem cannot help to carry out the relevant limit in statistical mechanics. This explains the importance of using hard Fourier analysis instead.

Note that in the proof of Theorem 1 we apply some standard tools from Fourier analysis (repeated applications of Parseval's formula, Cauchy–Schwarz inequality, etc.), and also some measure-theoretic and combinatorial arguments (e.g., Chebyshev's inequality, inclusion-exclusion formula). The tools are standard; the real difficulty is how to put these standard tools together (the devil is in the details).

3 Time-evolution in the Global Case

3.1 What about the Central Limit Theorem?

Theorem 1 clarifies the probabilistic aspects of the *local* case: when the volume of the subset $A \subset [0,1]^3 = I^3$ is *very small*. So small that the expected number of point-billiards in $A \subset [0,1]^3$ is just a constant, or it tends to infinity very slowly compared to the number of points N as $N \to \infty$. But what happens in the *global* case when $\text{vol}(A)$ jumps up from $O(1/N)$ to the constant range such as (say) $1/10 < \text{vol}(A) < 9/10$? That is, what happens when the expected number of point-billiards in A is as large as constant times N?

Well, then the vague intuition in (2.2) has the form

$$\text{Kronecker–Weyl theorem} \Longrightarrow \text{weak independence} \Longrightarrow \text{central limit theorem},$$

which makes it plausible to guess that perhaps the time-evolution of the point-counting function

$$Y_A(t) = \sum_{\substack{1 \le j \le N: \\ \mathbf{x}_j(t) \in A}} 1 \tag{3.1}$$

satisfies a central limit theorem. This leads to the following natural question: Is it true that the local Theorem 1 generalizes to a global central limit theorem?

What I *can* prove is a global central limit theorem for the time-evolution of the point-counting function in "nice" sets $A \subset [0,1]^3$. (Here "nice" means, e.g., the family of axes-parallel boxes, or the family of solid spheres, or even the large class of all convex sets.) However, in the general case of an arbitrary Lebesgue measurable subset $A \subset [0,1]^3$ we *cannot* expect a global central limit theorem; I will explain this unexpected fact below. In the general case I can only prove some weaker results: namely some analogs of the law of large numbers; see the second half of this section.

Basic Intuition in the Global Case If $A \subset [0,1]^3$ is a "large" measurable subset with, say, $1/10 < \mathrm{vol}(A) < 9/10$, then the time-evolution $0 < t < T$ of the point-counting function $Y_A(\omega; t)$ (see (3.1)), where

$$\omega = (\mathbf{x}_j(0), \dot{\mathbf{x}}_j(0) : 1 \le j \le N)$$

is a fixed typical initial condition, can be well approximated by a sum of independent and identically distributed random variables

$$Y_A(\omega; t) \approx \xi_1 + \xi_2 + \cdots + \xi_N \tag{3.2}$$

with $\Pr[\xi_j = 1] = \mathrm{vol}(A) = p$, $\Pr[\xi_j = 0] = 1 - \mathrm{vol}(A) = 1 - p = q$, $1 \le j \le N$.

Note in advance that the proof of Theorem 1—a local Poisson limit theorem—can be briefly summarized as follows: it is a repeated application of the second moment method to the *sequence* of sets A, $A \times A$, $A \times A \times A$, ... (Cartesian products in higher and higher dimensional spaces). The restriction to the second moment method explains why we are able to stay/work in the L_2 space, and can prove the general Theorem 1 about *arbitrary* measurable subsets A.

Central Limit Theorem: Why "Nice" Sets? The proof of a global central limit theorem for the time-evolution of the point-counting function inevitably leads to a study of the *higher moments*, and it requires the convergence of numerical series such as (say)

$$\sum_{\mathbf{r}_1+\mathbf{r}_2+\mathbf{r}_3+\mathbf{r}_4+\mathbf{r}_5+\mathbf{r}_6=\mathbf{0}} \frac{|a_{\mathbf{r}_1}a_{\mathbf{r}_2}a_{\mathbf{r}_3}a_{\mathbf{r}_4}a_{\mathbf{r}_5}a_{\mathbf{r}_6}|}{|\mathbf{r}_1|+|\mathbf{r}_2|+|\mathbf{r}_3|+|\mathbf{r}_4|+|\mathbf{r}_5|+|\mathbf{r}_6|}, \tag{3.3}$$

where $a_{\mathbf{r}}$ is the Fourier coefficient of the 0-1-valued characteristic function χ_A of the set $A \subset I^3 = [0,1)^3$:

$$\chi_A(\mathbf{w}) = \sum_{\mathbf{r}\in\mathbb{Z}^3} a_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{w}} \quad \text{with } a_{\mathbf{r}} = \int_A e^{-2\pi i \mathbf{r}\cdot\mathbf{z}}\, d\mathbf{z}. \tag{3.4}$$

Note that $\mathbf{r}\cdot\mathbf{w} = r_1w_1 + r_2w_2 + r_3w_3$ denotes the usual inner product of vectors, and we have

$$a_{-\mathbf{r}} = \overline{a_{\mathbf{r}}},$$

where "overline" denotes the complex conjugates. Clearly $a_{\mathbf{0}} = \mathrm{vol}(A)$ ($=$ the volume of A), and by Parseval's formula,

$$\sum_{\mathbf{r}\in\mathbb{Z}^3} |a_{\mathbf{r}}|^2 = \int_{I^3} \chi_A^2(\mathbf{w})\, d\mathbf{w} = \mathrm{vol}(A). \tag{3.5}$$

But just because

$$\sum_{\mathbf{r}\in\mathbb{Z}^3} |a_{\mathbf{r}}|^2 < \infty \tag{3.6}$$

(note that, by the Riesz–Fisher theorem, the L_2-space is characterized by (3.6)), the convergence in (3.6) does *not* guarantee the convergence in (3.3).

As an illustration, consider the one-dimensional sequence

$$a_{-n} = a_n = \frac{1}{\sqrt{n}\log(n+1)}, \quad n = 1, 2, 3, \ldots. \tag{3.7}$$

Clearly

$$\sum_{r\neq 0} a_r^2 = 2\sum_{n=1}^{\infty} \frac{1}{n\log^2(n+1)} < \infty. \tag{3.8}$$

On the other hand, the analog of (3.3)

$$\sum_{r_1+r_2+r_3+r_4+r_5+r_6=0:r_j\neq 0} \frac{a_{r_1}a_{r_2}a_{r_3}a_{r_4}a_{r_5}a_{r_6}}{|r_1|+|r_2|+|r_3|+|r_4|+|r_5|+|r_6|} \tag{3.9}$$

is a divergent series. Indeed, for any 5-tuple $(r_1, r_2, r_3, r_4, r_5)$ with $N \le r_1, \ldots, r_5 \le 2N$ there is a unique r_6 in the interval $-10N \le r_6 \le -5N$ such that $r_1 + r_2 + r_3 + r_4 + r_5 + r_6 = 0$, and the corresponding contribution in the sum (3.9) is

$$\ge \frac{N^5}{(\sqrt{10N}\log(10N+1))^6 \cdot 20N} \ge \text{const} \cdot \frac{N}{\log^6 N},$$

which tends to infinity as $N \to \infty$.

Next I explain how the proof of a global central limit theorem inevitably leads to the convergence of sum (3.3) (and to the convergence of some even more complicated sums), demanding that the Fourier coefficients $a_{\mathbf{r}}$ of the characteristic function χ_A of the set $A \subset I^3 = [0, 1)^3$ (see (3.4)) must tend to zero "rapidly"—a property that is satisfied only by "nice" sets $A \subset [0, 1)^3$. Of course, I have to tell what "rapidly" and "nice" actually mean.

According to the Basic Intuition in (3.2), for a typical initial condition

$$\omega = (\mathbf{x}_j(0), \dot{\mathbf{x}}_j(0) : 1 \le j \le N)$$

"the third central moment is very close to $o(N^{3/2})$", that is,

$$\frac{1}{T}\int_0^T (Y_A(\omega; t) - Np)^3\, dt \approx o(N^{3/2}), \tag{3.10}$$

assuming T is large. (We could also work with the mth moment for any integer $m \ge 3$.)

To justify (3.10), I recall that we have N non-interacting billiard balls, each represented by a point mass, which move freely inside the unit cube $I^3 = [0, 1]^3$ such that the reflection off the wall (= side of the cube) is elastic. The trajectory of the jth billiard ball (= point) in the time interval $0 \le t \le T$ is described by $\mathbf{x}_j(t) = (x_{j,1}(t), x_{j,2}(t), x_{j,3}(t))$; $\mathbf{x}_j(0) = \mathbf{y}_j$ is the initial position, $\dot{\mathbf{x}}_j(0) = v \cdot \mathbf{u}_j$ is the initial velocity and $v > 0$ is the common speed ($\mathbf{u}_j$ is a unit vector, i.e., $\mathbf{u}_j \in S^2$ = unit sphere).

In view of the geometric trick of *unfolding* the billiard paths to straight lines in the 3-space, it suffices to deal with N torus lines $\mathbf{x}_k(t) = (x_{k,1}(t), x_{k,2}(t), x_{k,3}(t))$ (mod 1), $k = 1, 2, \ldots, N$ where

$$x_{k,1}(t) = u_{k,1}tv + y_{k,1}, \qquad x_{k,2}(t) = u_{k,2}tv + y_{k,2}, \qquad x_{k,3}(t) = u_{k,3}tv + y_{k,3} \tag{3.11}$$

and

$$u_{k,1}^2 + u_{k,2}^2 + u_{k,3}^2 = 1, \tag{3.12}$$

i.e., $\mathbf{u}_k = (u_{k,1}, u_{k,2}, u_{k,3})$ is a unit vector. Since $v \geq 1$ is the common speed of the particles, the length of the straight line segment $\mathbf{x}_k(t)$, $0 < t < T$ is clearly vT. The pair

$$(\mathbf{x}_k(0), \dot{\mathbf{x}}_k(0)) = (\mathbf{y}_k, v\mathbf{u}_k)$$

is the initial condition of the kth torus-line $\mathbf{x}_k(t)$. We use the short notation ω for the initial condition of the whole system.

Let $A \subset I^3 = [0, 1)^3$ be an arbitrary Lebesgue measurable subset. The trick of unfolding means that we consider the union of 8 copies of A, and then we shrink the corresponding $2 \times 2 \times 2$ cube to the unit cube; the resulted set is also denoted—for notational convenience—by A.

By applying the Fourier series (3.4), we have ($p = \mathrm{vol}(A) = a_{\mathbf{0}}$)

$$\begin{aligned} Y_A(\omega; t) - Np &= \sum_{k=1}^{N} \chi_A(\mathbf{x}_k(t)) = \sum_{k=1}^{N} \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3: \\ \mathbf{r}\neq\mathbf{0}}} a_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{x}_k(t)} \\ &= \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3: \\ \mathbf{r}\neq\mathbf{0}}} \sum_{k=1}^{N} a_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{y}_k} \cdot e^{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vt}. \end{aligned} \tag{3.13}$$

Thus by (3.13) we have (we break the long expression below into two lines)

$$\begin{aligned} &\frac{1}{T}\int_0^T (Y_A(\omega; t) - Np)^3 \, dt \\ &= \sum_{\substack{\mathbf{r}_1\in\mathbb{Z}^3: \\ \mathbf{r}_1\neq\mathbf{0}}} \sum_{\substack{\mathbf{r}_2\in\mathbb{Z}^3: \\ \mathbf{r}_2\neq\mathbf{0}}} \sum_{\substack{\mathbf{r}_3\in\mathbb{Z}^3: \\ \mathbf{r}_3\neq\mathbf{0}}} a_{\mathbf{r}_1} a_{\mathbf{r}_2} a_{\mathbf{r}_3} \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\sum_{j_3=1}^{N} e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot\mathbf{y}_{j_1} + \mathbf{r}_2\cdot\mathbf{y}_{j_2} + \mathbf{r}_3\cdot\mathbf{y}_{j_3})} \\ &\quad \cdot \frac{1}{T}\int_0^T e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot\mathbf{u}_{j_1} + \mathbf{r}_2\cdot\mathbf{u}_{j_2} + \mathbf{r}_3\cdot\mathbf{u}_{j_3})vt} \, dt \\ &= \sum_{\substack{\mathbf{r}_1\in\mathbb{Z}^3: \\ \mathbf{r}_1\neq\mathbf{0}}} \sum_{\substack{\mathbf{r}_2\in\mathbb{Z}^3: \\ \mathbf{r}_2\neq\mathbf{0}}} \sum_{\substack{\mathbf{r}_3\in\mathbb{Z}^3: \\ \mathbf{r}_3\neq\mathbf{0}}} a_{\mathbf{r}_1} a_{\mathbf{r}_2} a_{\mathbf{r}_3} \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\sum_{j_3=1}^{N} e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot\mathbf{y}_{j_1} + \mathbf{r}_2\cdot\mathbf{y}_{j_2} + \mathbf{r}_3\cdot\mathbf{y}_{j_3})} \\ &\quad \cdot \frac{e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot\mathbf{u}_{j_1} + \mathbf{r}_2\cdot\mathbf{u}_{j_2} + \mathbf{r}_3\cdot\mathbf{u}_{j_3})vT} - 1}{2\pi \mathrm{i}(\mathbf{r}_1\cdot\mathbf{u}_{j_1} + \mathbf{r}_2\cdot\mathbf{u}_{j_2} + \mathbf{r}_3\cdot\mathbf{u}_{j_3})vT} \\ &= \sum\nolimits_1(\omega). \end{aligned} \tag{3.14}$$

To justify (3.10), we compute the quadratic average of $\sum_1(\omega)$ (see (3.14)) over all initial conditions ω (the common speed v remains fixed): we want to show that the quadratic average is $o(N^3)$. First we keep the unit vectors (initial directions) $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N \in S^2$ fixed, and evaluate the square integral $\sum_2$ below by integrating over the initial positions $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N \in I^3 = [0,1)^3$:

$$\sum\nolimits_2(\mathbf{u}_1, \ldots, \mathbf{u}_N) = \int_{I^3} \cdots \int_{I^3} \Big(\sum\nolimits_1(\omega)\Big)^2 \, d\mathbf{y}_1 \cdots d\mathbf{y}_N. \tag{3.15}$$

By definition,

$$\begin{aligned}
&\sum\nolimits_2(\mathbf{u}_1, \ldots, \mathbf{u}_N) \\
&\quad = \sum_{\substack{\mathbf{r}_m \in \mathbb{Z}^3: \\ \mathbf{r}_m \neq \mathbf{0},\ 1 \le m \le 6}} \ \sum_{\substack{1 \le j_m \le N: \\ 1 \le m \le 6}} a_{\mathbf{r}_1} a_{\mathbf{r}_2} a_{\mathbf{r}_3} a_{\mathbf{r}_4} a_{\mathbf{r}_5} a_{\mathbf{r}_6} \frac{e^{2\pi \mathrm{i}(\mathbf{r}_1 \cdot \mathbf{u}_{j_1} + \mathbf{r}_2 \cdot \mathbf{u}_{j_2} + \mathbf{r}_3 \cdot \mathbf{u}_{j_3}) vT} - 1}{2\pi \mathrm{i}(\mathbf{r}_1 \cdot \mathbf{u}_{j_1} + \mathbf{r}_2 \cdot \mathbf{u}_{j_2} + \mathbf{r}_3 \cdot \mathbf{u}_{j_3}) vT} \\
&\qquad \cdot \frac{e^{2\pi \mathrm{i}(\mathbf{r}_4 \cdot \mathbf{u}_{j_4} + \mathbf{r}_5 \cdot \mathbf{u}_{j_5} + \mathbf{r}_6 \cdot \mathbf{u}_{j_6}) vT} - 1}{2\pi \mathrm{i}(\mathbf{r}_4 \cdot \mathbf{u}_{j_4} + \mathbf{r}_5 \cdot \mathbf{u}_{j_5} + \mathbf{r}_6 \cdot \mathbf{u}_{j_6}) vT} \\
&\qquad \cdot \int_{I^3} \cdots \int_{I^3} e^{2\pi \mathrm{i}(\mathbf{r}_1 \cdot \mathbf{y}_{j_1} + \mathbf{r}_2 \cdot \mathbf{y}_{j_2} + \mathbf{r}_3 \cdot \mathbf{y}_{j_3} + \mathbf{r}_4 \cdot \mathbf{y}_{j_4} + \mathbf{r}_5 \cdot \mathbf{y}_{j_5} + \mathbf{r}_6 \cdot \mathbf{y}_{j_6})} \, d\mathbf{y}_1 \cdots d\mathbf{y}_N.
\end{aligned} \tag{3.16}$$

Due to the last integral, the nonzero contribution in $\sum_2$ (see (3.16)) comes from the terms with the property that in the sum

$$\mathbf{r}_1 \cdot \mathbf{y}_{j_1} + \mathbf{r}_2 \cdot \mathbf{y}_{j_2} + \mathbf{r}_3 \cdot \mathbf{y}_{j_3} + \mathbf{r}_4 \cdot \mathbf{y}_{j_4} + \mathbf{r}_5 \cdot \mathbf{y}_{j_5} + \mathbf{r}_6 \cdot \mathbf{y}_{j_6}$$

every $\mathbf{y}_j$ has zero coefficient. There are a few such cases; one of them is the following:

Case A: $j_1 = j_2 = j_3 = j_4 = j_5 = j_6$ and $\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3 + \mathbf{r}_4 + \mathbf{r}_5 + \mathbf{r}_6 = \mathbf{0}$

Using Case A in (3.16), we have

$$\begin{aligned}
&\sum\nolimits_2(\mathbf{u}_1, \ldots, \mathbf{u}_N) \\
&\quad = \sum_{\substack{(\mathbf{r}_1, \ldots, \mathbf{r}_6, j): \\ \mathbf{r}_1, \ldots, \mathbf{r}_6 \in \mathbb{Z}^3 \setminus \mathbf{0},\ 1 \le j \le N \\ \mathbf{r}_1 + \cdots + \mathbf{r}_6 = \mathbf{0}}} a_{\mathbf{r}_1} \cdots a_{\mathbf{r}_6} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3) \cdot \mathbf{u}_j vT} - 1}{2\pi (\mathbf{r}_1 + \mathbf{r}_2 + \mathbf{r}_3) \cdot \mathbf{u}_j vT} \right|^2 + \cdots
\end{aligned} \tag{3.17}$$

where the $\cdots$ at the end of (3.17) indicates that there are a few more cases beyond Case A.

If $\mathbf{r} \in \mathbb{Z}^3 \setminus \mathbf{0}$ and $\tau \geq 1$ is an arbitrary constant, then a simple calculation gives

$$\int_{S^2} \left| \frac{e^{2\pi \mathrm{i} \mathbf{r} \cdot \mathbf{u} \tau} - 1}{2\pi \mathbf{r} \cdot \mathbf{u} \tau} \right|^2 d\mathbf{u} = \frac{\text{const}}{|\mathbf{r}| \tau}. \tag{3.18}$$

I will show a detailed proof of (3.18) later; see (5.15).

Next we integrate $\sum_2$ over the direction vectors $\mathbf{u}_k \in S^2$, $k = 1, 2, \ldots, N$, and apply (3.18) with $\tau = vT$; this leads us to the following (in order to normalize, we have to divide by 4π, which is the surface area of the unit sphere S^2):

$$\sum\nolimits_3 = \frac{1}{4\pi} \int_{S^2} \cdots \frac{1}{4\pi} \int_{S^2} \sum\nolimits_2(\mathbf{u}_1, \ldots, \mathbf{u}_N) \, d\mathbf{u}_1 \ldots d\mathbf{u}_N$$

$$= N \sum_{\substack{(\mathbf{r}_1,\ldots,\mathbf{r}_6):\\ \mathbf{r}_1,\ldots,\mathbf{r}_6\in\mathbb{Z}^3\setminus\mathbf{0}\\ \mathbf{r}_1+\mathbf{r}_2+\mathbf{r}_3=-(\mathbf{r}_4+\mathbf{r}_5+\mathbf{r}_6)\neq\mathbf{0}}} a_{\mathbf{r}_1}\cdots a_{\mathbf{r}_6}\cdot\frac{\text{const}}{|\mathbf{r}_1+\mathbf{r}_2+\mathbf{r}_3|vT}+\cdots \tag{3.19}$$

where again the $\cdots$ at the end of (3.18) indicates that there are some extra terms that I skipped.

In view of (3.19), it is perfectly reasonable to assume the convergence of the series

$$\sum_{\substack{(\mathbf{r}_1,\ldots,\mathbf{r}_6):\\ \mathbf{r}_1,\ldots,\mathbf{r}_6\in\mathbb{Z}^3\setminus\mathbf{0}\\ \mathbf{r}_1+\mathbf{r}_2+\mathbf{r}_3=-(\mathbf{r}_4+\mathbf{r}_5+\mathbf{r}_6)\neq\mathbf{0}}} \frac{|a_{\mathbf{r}_1}a_{\mathbf{r}_2}a_{\mathbf{r}_3}a_{\mathbf{r}_4}a_{\mathbf{r}_5}a_{\mathbf{r}_6}|}{|\mathbf{r}_1+\mathbf{r}_2+\mathbf{r}_3|}<\infty, \tag{3.20}$$

since otherwise we don't have a chance to prove $\sum_3 = o(N^3)$, which is "basically" equivalent to (3.10). Since (3.20) implies the convergence of the numerical series in (3.3), now we see why the proof of a global central limit theorem "inevitably requires" the convergence of sum (3.3).

To guarantee (3.3) (or (3.20)), we certainly need some *extra* condition of the type "$a_{\mathbf{r}}$ tends to zero rapidly as $|\mathbf{r}|\to\infty$" (besides the Parseval formula $\sum_{\mathbf{r}}|a_{\mathbf{r}}|^2<\infty$ which always holds). Note that the classes of axes-parallel boxes, or solid spheres, or even the large class of convex sets, all satisfy this "rapid decreasing of the Fourier coefficients $a_{\mathbf{r}}$" type condition.

How to prove a global central limit theorem for these "nice" sets? The basic idea is to apply the "moment method". To explain this, I recall the *Basic Intuition* in the global case (see (3.2)): if $A\subset[0,1]^3$ is a "large" measurable subset with, say, $1/10<\text{vol}(A)<9/10$, then the time-evolution $0<t<T$ of the point-counting function $Y_A(\omega;t)$ (see (3.1)), where ω is a fixed typical initial condition, can be well approximated by a sum of independent and identically distributed random variables

$$Y_A(\omega;t)\approx\xi_1+\xi_2+\cdots+\xi_N \tag{3.21}$$

with $\Pr[\xi_j=1]=\text{vol}(A)=p$, $\Pr[\xi_j=0]=1-\text{vol}(A)=1-p=q$, $1\le j\le N$. The sum $\xi_1+\xi_2+\cdots+\xi_N$ in (3.21) has the binomial distribution $B(N,p)$ with parameters N ("large integer") and p ("probability"):

$$\Pr[\xi_1+\xi_2+\cdots+\xi_N=k]=\binom{N}{k}p^kq^{N-k},\quad 0\le k\le N. \tag{3.22}$$

To prove a global central limit theorem for the time-evolution, we compare the higher moments: we will show that the mth central moment σ_m^*

$$\sigma_m^*=\sigma_m^*(N,p)=\mathbf{E}\left(\sum_{j=1}^N(\xi_j-p)\right)^m \tag{3.23}$$

of the binomial distribution $B(N,p)$ is "very close" to the integral

$$\sigma_m(\omega)=\sigma_m(\omega;T;N,p)=\frac{1}{T}\int_0^T(Y_A(\omega;t)-Np)^m\,dt \tag{3.24}$$

for the overwhelming majority of the initial conditions ω. Here $m=1,2,3,\ldots$, and we assume that N and T are both "large". The basic idea is simple, but the execution requires long and delicate calculations, so I stop here. This paper is already far too long; I postpone the "global central limit theorem for the time-evolution of the point-counting function in nice sets" to another paper.

What I will prove here is a simpler and weaker result: a global analog of the law of large numbers (with much shorter proof).

3.2 The "Density" Is Constant

The "global case" means that $\mathrm{vol}(A)$ is in the constant range such as (say) $\mathrm{vol}(A) = 1/3$ or $\mathrm{vol}(A) = 2/3$. So what happens when the expected number of point-billiards in A is as large as constant times N? The time-evolution of the point-counting function exhibits a weak analog of the law of large numbers.

Theorem 2 *Similarly to Theorem 1, assume that N non-interacting point-billiards move freely inside the unit cube $I^3 = [0, 1]^3$ such that the reflection off the wall ($=$ side of the cube) is elastic. Let $\mathbf{x}_j(t) = (x_{j,1}(t), x_{j,2}(t), x_{j,3}(t))$ describe the trajectory of the jth billiard ball ($=$point) in the time interval $0 \le t \le T$, where $\mathbf{x}_j(0) = \mathbf{y}_j$ is the initial position, $\dot{\mathbf{x}}_j(0) = v \cdot \mathbf{u}_j$ is the initial velocity and $v > 0$ is the common speed.*

Let $A \subset [0, 1]^3$ be an arbitrary Lebesgue measurable subset of the unit cube with volume $0 < \mathrm{vol}(A) < 1$, and let $Y_A(t)$ denote the point-counting function:

$$Y_A(t) = \sum_{\substack{1 \le j \le N: \\ \mathbf{x}_j(t) \in A}} 1.$$

Let $0 < \varepsilon < 1$ and $0 < \eta < 1$. Then for more than $1 - \varepsilon$ part of the initial conditions

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \left(I^3\right)^N \times \left(S^2\right)^N = \Omega$$

(in the sense of the product measure on Ω), the point-counting function $Y_A(\omega; t)$ is very close to the "expectation" $N \cdot \mathrm{vol}(A)$ in the sense that

$$\frac{1}{T}\text{measure}\{0 \le t \le T : |Y_A(\omega; t) - N \cdot \mathrm{vol}(A)| > \eta \cdot N \cdot \mathrm{vol}(A)\}$$
$$< \frac{2\sqrt{3}}{\sqrt{\varepsilon}\eta^2\sqrt{N \cdot \mathrm{vol}(A)}} \left(\frac{2}{vT} + \frac{1}{vTN \cdot \mathrm{vol}(A)} + \frac{2}{vT(N \cdot \mathrm{vol}(A))^2} + \frac{8}{N \cdot \mathrm{vol}(A)}\right)^{1/2}.$$

Remarks (1) I note without proof that one can easily prove a simultanous version of Theorem 2 (like how (1.11) is a simultanous version of (1.10)), and also one can easily prove sequential versions, such as (b) and (c) in the Remarks after Theorem 1.

(2) Again the main point is that Theorem 2 holds for *arbitrary* measurable sets A in the unit cube.

(3) Theorem 2 is the most effective if $N \cdot \mathrm{vol}(A)$ is "large", and the estimation becomes useless if $N \cdot \mathrm{vol}(A)$ is in the constant range. But this is perfectly natural, since $N \cdot \mathrm{vol}(A)$ is the average number of point-particles in subset A, and of course the law of large numbers *fails* to work for "small numbers", i.e., when $N \cdot \mathrm{vol}(A)$ is in the constant range.

(4) *An illustration of Theorem 2*. The following example is motivated by the kinetic theory of gases. Suppose that the unit cube $[0, 1]^3$ represents a cube-shaped container of side 1 meter, $N = 10^{27}$ (roughly the number of gas molecules), $v = 10^3$ meter per second (typical speed of a gas molecule in room temperature), $T =$ hundred years $= 100 \cdot 365 \cdot 24 \cdot 60 \cdot 60 \approx$

$3 \cdot 10^9$ seconds, $\varepsilon = 10^{-4}$, $A \subset [0, 1]^3$ is an arbitrarily complicated (but measurable) subset with $\mathrm{vol}(A) = 1/2$, and $\eta = 10^{-3}$.

In this special case, Theorem 2 gives the following information. Consider a long time interval of 100 years: for more than 99.99 percent of the initial conditions $\omega \in \Omega$, the number of particles $Y_A(\omega; t)$ in the given subset $A \subset [0, 1]^3$ of volume $1/2$ stays very close to the expected number $N/2$ in the quantitative sense that the deviation from $N/2$ is less than one-tenth of a percent except possibly for a set of time-points t with total time less than 10 seconds!

Well, 10 seconds is a remarkably short time compared to 100 years. The message of this striking result is that in the deterministic Bernoulli model the "density" is constant (which easily yields that the "pressure" is also constant).

The proof of Theorem 2 is postponed to Sect. 8.

3.3 Quick Evolution from Extremely Jammed Start to Uniform Distribution: Theorem 3

Just like in Theorems 1 and 2, we study the time evolution of an ideal gas of N molecules—represented by N point-billiards—in a cube-shaped container that we identify with the unit cube $I^3 = [0, 1]^3$, but we have a novelty: the initial position of the system is *extremely non-uniform*. More precisely, assume that at $t = 0$ the N point-billiards are all concentrated in the first octant $[0, 1/2]^3$ of the unit cube. Of course, instead of $[0, 1/2]$ we could choose (say) $[0, 1/3]$ or $[0, 1/4]$; it doesn't really make any difference in the proof below. Note that the event "the system of N point-particles is concentrated in the first octant" has probability 8^{-N}, which is *inconceivably* small if N is in the range of the Avogadro number $N \approx 10^{24}$.

What we are interested in is the following question: How long does it take for the system, starting from this extremely "jammed" position (i.e., where all points are in the first octant), to achieve uniformity in the whole unit cube?

We test uniformity in the following way. Let $A \subset [0, 1]^3$ be an arbitrary Lebesgue measurable subset of the unit cube with volume $0 < \mathrm{vol}(A) < 1$, and let $Y_A(t)$ denote the point-counting function:

$$Y_A(t) = \sum_{\substack{1 \le j \le N: \\ \mathbf{x}_j(t) \in A}} 1,$$

where, as usual, $\mathbf{x}_j(t) = (x_{j,1}(t), x_{j,2}(t), x_{j,3}(t))$ describes the trajectory of the jth point-billiard in the time interval $0 \le t \le T$, where $\mathbf{x}_j(0) = \mathbf{y}_j$ is the initial position, $\dot{\mathbf{x}}_j(0) = v \cdot \mathbf{u}_j$ is the initial velocity and $v > 0$ is the common speed of the point-billiards. What we want is that, for some "large" T, the overwhelming majority of the time-points t in $0 < t < T$ exhibit uniformity in the sense

$$Y_A(t) = N \cdot \mathrm{vol}(A) + \text{negligible error}. \tag{3.25}$$

What is the shortest time-interval T for which we can guarantee (3.25)? Our goal here is to discuss this question.

Theorem 3 *Similarly to Theorems 1 and 2, assume that N non-interacting point-billiards move freely inside the unit cube $I^3 = [0, 1]^3$ such that the reflection off the wall (= side of the cube) is elastic. Let $\mathbf{x}_j(t) = (x_{j,1}(t), x_{j,2}(t), x_{j,3}(t))$ describe the trajectory of the jth billiard ball (= point) in the time interval $0 \le t \le T$, where $\mathbf{x}_j(0) = \mathbf{y}_j$ is the initial position, $\dot{\mathbf{x}}_j(0) = v \cdot \mathbf{u}_j$ is the initial velocity and $v > 0$ is the common speed.*

Assume that the initial position of the system is concentrated in the first octant $[0, 1/2]^3$ *of the unit cube* ($=$ *container*), *that is*, *the initial condition* ω *satisfies*

$$\omega = (\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N) \in \left([0, 1/2]^3\right)^N \times \left(S^2\right)^N = \widetilde{\Omega}.$$

Let $A \subset [0, 1]^3$ *be an arbitrary Lebesgue measurable subset of the unit cube with volume* $0 < \mathrm{vol}(A) < 1$, *and let* $Y_A(t)$ *denote the point-counting function*:

$$Y_A(t) = \sum_{\substack{1 \le j \le N: \\ \mathbf{x}_j(t) \in A}} 1.$$

Then the L_2*-norm of the discrepancy of the point-counting function* $Y_A(\omega; t)$ *from its "expectation"* $N \cdot \mathrm{vol}(A)$ *is estimated from above as follows*:

$$\begin{aligned} &\left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\frac{1}{T} \int_0^T \left(\frac{Y_A(\omega; t) - N \cdot \mathrm{vol}(A)}{N}\right)^2 dt\right)^2 d\omega \\ &\quad \le 3 \cdot 10^4 \cdot \mathrm{vol}^2(A) \left(\frac{\log(vT)}{vT}\right)^2 + 12 \cdot 10^2 \cdot \mathrm{vol}^3(A) \left(\frac{\log(vT)}{vT}\right)^2 \\ &\qquad + \frac{4 \cdot 10^4 \cdot \mathrm{vol}^3(A)}{vTN} + \frac{32 \cdot 10^{13} \cdot \mathrm{vol}^2(A)}{vTN} + \frac{24 \cdot 10^8 \cdot \mathrm{vol}^2(A)}{vTN^2} + \frac{3}{N^2}. \end{aligned} \tag{3.26}$$

Remarks At first sight the upper bound (2.2) seems very complicated, but it is in fact simpler than it looks, since the terms with N in the denominator are usually negligible.

If the square-integral

$$\left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\frac{1}{T} \int_0^T \left(\frac{Y_A(\omega; t) - N \cdot \mathrm{vol}(A)}{N}\right)^2 dt\right)^2 d\omega$$

is "small" for an arbitrary Lebesgue-measurable subset $A \subset [0, 1]^3$ of the unit cube, then we can intuitively say that, the distribution of a typical system of N point-billiards with an extremely jammed start evolves to globally uniform in time T. How long does it take to achieve uniformity? The following example is motivated by the kinetic theory of gases. Suppose that the unit cube $[0, 1]^3$ represents a cube-shaped container of side 1 meter, $N = 10^{27}$ (roughly the number of gas molecules), $v = 10^3$ meter per second (typical speed of a gas molecule in room temperature), $A \subset [0, 1]^3$ is an arbitrarily complicated (but measurable) subset with $\mathrm{vol}(A) = 1/2$, and let $T = 10$ seconds. By evaluating the right hand side of (2.2), we obtain the upper bound

$$\left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\frac{1}{T} \int_0^T \left(\frac{Y_A(\omega; t) - N \cdot \mathrm{vol}(A)}{N}\right)^2 dt\right)^2 d\omega < \frac{1}{100},$$

which is "small". We can say, therefore, that during the relatively short time-interval of 10 seconds the system evolves from extremely jammed to uniformly distributed—at least this is the typical behavior.

What is more, as the time goes by, the system becomes more and more uniform—this is the message of Theorem 3; see the asymptotic behavior of the right hand side of (3.26) as $T \to \infty$.

My (complicated) estimation in (3.26) is certainly not optimal: I am convinced that in reality the system in the example above achieves uniformity in a fraction of a second.

The proof of Theorem 3, similarly to Theorem 2, is postponed to Sect. 8. Both proofs are based on a second moment argument.

4 Super-uniformity of the Typical Billiard Path

4.1 What is Super-uniformity?

This section is a detour: I discuss a surprising byproduct of my research in deterministic kinetic theory of gases. By using the basic technique of this paper—Fourier analysis—I show that the typical billiard paths in a square (or any rectangle) are extremely uniform far beyond "common sense". What most experts would consider a "common sense expectation" is the square-root size ("random") error; what we can prove is the much smaller *square-root logarithm*(!). Since square-root logarithm is "almost" constant, our upper bound is essentially independent of the complexity of the test set. We can say, therefore, that (roughly speaking) "the ugliness of the test set is irrelevant". Or we can say that the set of typical billiard paths represents the family of most uniformly distributed curves in the square. Theorem 4 below makes these vague statements precise.

More precisely, we study the trajectory of a single point-billiard, and for simplicity we restrict ourselves to the unit square (2-dimensional case). We want to compare the *actual time*—i.e., the time the point-billiard spends in a given (measurable) subset A of the unit square—to the *expected time*. The *expected time* is area(A) times the total time, which reflects "perfect uniformity" (we assume that the speed is one).

In view of the trick of *unfolding* the billiard path to a straight line in the plane (explained in Sect. 1), it suffices to deal with torus-lines (of course we shrink the corresponding 2×2 square to the unit square). Let $A \subset I^2 = [0, 1)^2$ be an arbitrary Lebesgue measurable subset (via unfolding it corresponds to the union of four copies of the given test-set), and consider the torus-line $\mathbf{x}(t) = (x_1(t), x_2(t))$ (mod 1) where

$$x_1(t) = \alpha_1 t + y_1, \qquad x_2(t) = \alpha_2 t + y_2 \quad \text{and} \quad \alpha_1^2 + \alpha_2^2 = 1. \tag{4.1}$$

The second part of (4.1) means that the speed is one, so the length of the straight line segment $\mathbf{x}(t)$, $0 < t < T$ is clearly T, i.e., time = arc-length. The initial condition $(\mathbf{y}, (\alpha_1, \alpha_2))$ describes the starting point $\mathbf{y} \in [0, 1)^2$ and the angle (by the point (α_1, α_2) on the unit circle) of the torus-line $\mathbf{x}(t)$. Let $A(T) = A(T; \mathbf{y}, (\alpha_1, \alpha_2))$ denote the total time the torus-line $\mathbf{x}(t)$ (defined in (4.1)) spends in subset A during the given time interval $0 < t < T$. That is,

$$A(T; \mathbf{y}, (\alpha_1, \alpha_2)) = \text{actual time}, \quad \text{and} \quad \text{area}(A) \cdot T = \text{expected time},$$

and we want to compare the two.

I begin with the continuous form of the Kronecker–Weyl equidistribution theorem.

Theorem (Continuous Kronecker–Weyl) *If the slope α_2/α_1 is irrational, then for every starting point* $\mathbf{y} \in [0, 1)^2$,

$$\lim_{T\to\infty} \frac{A(T; \mathbf{y}, (\alpha_1, \alpha_2))}{T} = \text{area}(A) \tag{4.2}$$

for all Jordan measurable sets $A \subset [0, 1)^2$.

Note that a set in a euclidean space is Jordan measurable if and only if the characteristic function χ_A of the set is Riemann integrable.

The billiard path (4.1) has 2-dimensional Lebesgue measure zero, so (4.2) *cannot* be true for all Lebesgue measurable sets $A \subset [0,1)^2$. However, by involving the ergodic theorem, we can formulate a result, similar to (4.2), which holds for every Lebesgue measurable set $A \subset [0,1)^2$. First we have to define a dynamical system: for every real t we define the map

$$\Phi_t : \mathbf{y} \to \mathbf{y} + t(\alpha_1, \alpha_2) \pmod 1. \tag{4.3}$$

Clearly Φ_t is a mapping of the unit square $\mathbf{y} \in [0,1)^2$ into itself, and Φ_t preserves the Lebesgue measure ("area"). Since $\Phi_t(\mathbf{x}(0)) = \mathbf{x}(t)$ (see (4.1)), it is customary to call Φ_t the "time-shift". The quadruplet

$$\left([0,1)^2, \mathcal{L}, \lambda, \Phi_t\right), \tag{4.4}$$

where $\mathcal{L}$ denotes the σ-algebra of all Lebesgue measurable sets $A \subset [0,1)^2$ and λ is the 2-dimensional Lebesgue measure, is an *ergodic* dynamical system *if* the slope α_2/α_1 is irrational. Note that ergodicity follows from (4.2).

Next we apply Birkhoff's ergodic theorem: it states that, given any Lebesgue measurable set $A \subset [0,1)^2$, for *almost every* starting point $\mathbf{y} \in [0,1)^2$,

$$\lim_{T\to\infty} \frac{A(T; \mathbf{y}, (\alpha_1, \alpha_2))}{T} = \text{area}(A) = \lambda(A). \tag{4.5}$$

Since the classes of Jordan and Lebesgue measurable sets are both very large, and contain arbitrarily "ugly" (= complicated) sets $A \subset [0,1)^2$, it is not too surprising that neither the continuous Kronecker–Weyl Theorem (4.2), nor the ergodic Theorem (4.5) can say anything about the speed of convergence. In both cases (4.2) and (4.5),

$$|A(T; \mathbf{y}, (\alpha_1, \alpha_2)) - \text{area}(A) \cdot T| = o(T), \tag{4.6}$$

but we know nothing beyond that.

However, if we replace "every irrational (= ergodic) slope" with "almost every slope", then we can upgrade the weak (4.6) to a shockingly strong upper bound for the discrepancy:

$$|A(T; \mathbf{y}, (\alpha_1, \alpha_2)) - \text{area}(A) \cdot T| = O\left(\sqrt{\log T}\right). \tag{4.7}$$

Theorem 4 below is exactly an explicit/precise version of (4.7).

Throughout $\log x$ and $\log_2 x$ stand for the natural (i.e., base e) and the binary (i.e., base 2) logarithms (I don't use ln at all).

Theorem 4 *Let A be an arbitrary Lebesgue measurable subset of the unit square* $[0,1)^2$ *with two-dimensional Lebesgue measure* area(A), *and let $T > 100$ be an arbitrarily large (but fixed) real number. Let $\mathbf{x}(t) = (x_1(t), x_2(t))$, $0 \le t \le T$ be a billiard path of length T (= time) in the unit square, and let $A(T)$ denote the time the billiard path spends in subset A*:

$$A(T) = \text{measure}\,\{t \in [0,T] : \mathbf{x}(t) \in A\}.$$

Let $0 < \varepsilon < 1/2$ be arbitrary. Then for $1 - \varepsilon$ part of all billiard paths of length T in the square,

$$|A(T) - T \cdot \text{area}(A)| < \frac{10}{\varepsilon}\sqrt{\text{area}(A)(1 - \text{area}(A))} \cdot \sqrt{\log_2 T} \cdot \log_2 \log_2 T. \tag{4.8}$$

4.2 Remarks

(1) It is astonishing that, given an arbitrarily complicated subset $A \subset [0, 1)^2$, the discrepancy in (4.8) is just *square-root logarithmic*, that is, "almost constant" (the "ugliness" of A plays no role in (4.8)). On the other hand, if Theorem 4 is restricted to extremely "nice" subsets, say to the narrow family of axis-parallel subsquares, then constant discrepancy $O(1)$ in (4.8) is still unavoidable (to be explained below). In other words, in Theorem 4 the "ugliness" (= complexity) of the test set $A \subset [0, 1)^2$ is basically irrelevant! We can summarize the message in the following statement: *the most uniformly distributed curves in the unit square are the typical billiard paths*. Perhaps the greatest surprise here is that such a vague question has a definite quantitative answer.

(2) As I promised, I explain the almost trivial fact that even for the narrow class of axis-parallel subsquares we must have constant discrepancy $O(1)$ in (4.8). Consider the two subsquares $A_1 = [0, 1/3]^2$ and $A_2 = [2/3, 1]^2$ that are "far" from each other; the distance between them is $\sqrt{2}/3$. Let $\mathbf{x}(t)$ be an arbitrary continuous curve in the unit square; we always assume that the arc-length of every segment $\mathbf{x}(t)$, $T_1 < t < T_2$ is exactly $T_2 - T_1$ (meaning: t is the time and a point-mass moves along the curve with unit speed). For any real number $\tau > 0$ write

$$A_i(\tau) = \text{measure}\,\{t \in [0, \tau] : \mathbf{x}(t) \in A_i\}\,, \quad i = 1, 2$$

where A_i, $i = 1, 2$ are the two subsquares mentioned above. We show that the following four discrepancies:

$$|A_i(T) - T \cdot \text{area}(A_i)|, \quad |A_i(T + c) - (T + c) \cdot \text{area}(A_i)|, \quad i = 1, 2, \tag{4.9}$$

where $c = \sqrt{2}/3$ is the distance between the two given subsquares A_1 and A_2 (computed for the same curve!), cannot be all $o(1)$. Indeed, the middle segment $\mathbf{x}(t)$, $T < t < T + c$ of the curve cannot visit both subsquares (because the arc-length is exactly the distance between A_1 and A_2); consequently, at least one of the four discrepancies in (4.9) must be

$$\geq \frac{1}{2} c \cdot \text{area}(A_i) = \frac{1}{2} \cdot \frac{\sqrt{2}}{3} \cdot \frac{1}{9} = \frac{\sqrt{2}}{54}.$$

This proves that in Theorem 4 we cannot hope for discrepancy $o(1)$ in (4.8) even for the simplest families of subsets.

(3) The vague term of "typical billiard path" in Theorem 4 can be made precise in the usual way: by defining a measure on the set of all initial conditions of the billiard paths. Since the initial condition consists of a starting point $\mathbf{y} \in [0, 1)^2$ and an initial direction (angle) $\theta \in [0, 2\pi)$, the corresponding measure is simply the product of the two-dimensional Lebesgue measure on the unit square and the normalized one-dimensional Lebesgue measure.

(4) Theorem 4 gives an interesting new insight to the general question of *discrete versus continuous*. The Kronecker–Weyl equidistribution theorem has two forms: a discrete form and a continuous form.

Theorem (Kronecker–Weyl Theorem (discrete)) *Let $d \geq 1$ be any integer, and let $\mathbf{a} = (\alpha_1, \ldots, \alpha_d)$ be an arbitrary d-dimensional vector with real coordinates. The sequence $n\mathbf{a}$* (mod 1), *$n = 1, 2, 3, \ldots$ is uniformly distributed in the unit cube $[0, 1)^d$, meaning*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{\substack{1 \leq n \leq N: \\ n\mathbf{a} \in R \ (\text{mod } 1)}} 1 = \text{volume}(R)$$

for any rectangular box $R = I_1 \times \cdots \times I_d \subset [0,1)^d$ (*i.e., Cartesian product of intervals*) *if and only if* $1, \alpha_1, \ldots, \alpha_d$ *are linearly independent over the rationals.*

Theorem (Kronecker–Weyl Theorem (continuous)) *We have*

$$\lim_{T\to\infty} \frac{1}{T} \text{measure}\{0 < t < T : t\mathbf{a} \in R \pmod 1\} = \text{volume}(R)$$

for any rectangular box $R = I_1 \times \cdots \times I_d \subset [0,1)^d$ *if and only if* $\alpha_1, \ldots, \alpha_d$ *are linearly independent over the rationals.*

(Note that the continuous form was already mentioned in (2.1) and in (4.2).) These two forms show that the discrete sequence

$$n\mathbf{a} \pmod 1, \quad n = 1, 2, 3, \ldots \tag{4.10}$$

and the continuous torus line passing through the origin

$$t\mathbf{a} \pmod 1, \quad 0 < t < \infty \tag{4.11}$$

have the same equidistribution property—at least from a *qualitative* viewpoint. (Also, the two versions have almost identical proofs based on the famous Weyl's Criterion.)

The surprising message of Theorem 4 is that, in spite of these similarities, the *quantitative* aspects of (4.10) and (4.11) are very different. Indeed, Theorem 4 states that a typical billiard path—which is just a general torus line $t\mathbf{a} + \mathbf{b} \pmod 1$ $0 < t < \infty$ via unfolding—estimates the area of an arbitrary but fixed (measurable) subset $A \subset [0,1]^2$ with "error" ($=$ discrepancy) $O(\sqrt{\log T})$, where T is the length of the time-interval. On the other hand, a typical discrete sequence of the form (which is the discrete version of the general torus line $t\mathbf{a} + \mathbf{b}$ in dimension $d = 1$)

$$n\alpha + \beta \pmod 1, \quad n = 1, 2, \ldots, N$$

cannot estimate the one-dimensional Lebesgue measure of an arbitrary but fixed $A \subset [0,1]$ with "error" $o(\sqrt{N})$. This is the message of the following result.

Proposition 4.1 *For every integer* $N \geq 1$, *there is a measurable subset* $A = A_N \subset [0,1]$ (*in fact, A is a finite union of intervals*) *of measure* $1/2$ *such that, for the majority of the pairs* $(\alpha, \beta) \in [0,1)^2$,

$$\left| \sum_{\substack{1 \leq n \leq N: \\ n\alpha+\beta \in A \pmod 1}} 1 - N \cdot \text{measure}(A) \right| > \frac{\sqrt{N}}{5}. \tag{4.12}$$

Note without proof that Proposition 4.1 is best possible: the error term $\sqrt{N}$ in (4.12) cannot be replaced by any larger function of N.

Comparing Theorem 4 to Proposition 4.1, we see a huge difference between the sizes of the "errors"

$$\sqrt{\log T} \quad \text{and} \quad \sqrt{N}$$

(for simplicity I ignored the negligible iterated logarithmic factor of T). This shows that the quantitative aspects of (4.10) and (4.11) are dramatically different, and explains why I used the strong term *super-uniformity* in the title of this section.

(5) We can view Theorem 4 and Proposition 4.1 as the starting points of a new direction in the study of the classical subject of uniform distribution. There are many natural questions here, which are worth while pursuing. I just mention a few: What happens in dimensions $d \geq 3$? (I mention without proof that the curve-discrepancy $< T^{\frac{1}{2}-\frac{1}{2(d-1)}}$; here for simplicity I ignored the logarithmic factors.) Another question: estimating the volume and integral with point sampling, which one is better: regular sampling or random sampling (= Monte Carlo method)? Can we "beat" the Monte Carlo method?

Since this paper is already far too long, I will address these exciting questions in another paper. Here I just briefly mention one more related result without proof. The huge difference between the sizes of the "errors"

$$\sqrt{N} \quad \text{and} \quad \sqrt{\log T}$$

in Proposition 4.1 and Theorem 4 does not stop in dimension two: if we "increase both dimensions by one", the "error" becomes constant(!), which is clearly best possible. The term "increase both dimensions by one" means that we replace the torus line-segment $\mathbf{x}(t)$ (mod 1), $0 < t < T$, where $\mathbf{x}(t) = (x_1(t), x_2(t))$,

$$x_1(t) = \alpha_1 t + y_1, \qquad x_2(t) = \alpha_2 t + y_2,$$

with the torus-parallelogram $\mathbf{x}(t_1, t_2)$ (mod 1), $0 < t_1 < T_1$, $0 < t_2 < T_2$, where $\mathbf{x}(t_1, t_2) = (x_1(t_1, t_2), x_2(t_1, t_2), x_3(t_1, t_2)) \in \mathbb{R}^3$ and

$$x_1(t_1, t_2) = \alpha_{1,1} t_1 + \alpha_{1,2} t_2 + y_1, \qquad x_2(t_1, t_2) = \alpha_{2,1} t_1 + \alpha_{2,2} t_2 + y_2,$$
$$x_3(t_1, t_2) = \alpha_{3,1} t_1 + \alpha_{3,2} t_2 + y_3.$$

A typical torus-parallelogram (with respect to an arbitrary 3-dimensional subset of the unit cube) is even more uniform than a typical torus line-segment (with respect to an arbitrary 2-dimensional subset in the unit square): the "error" (= discrepancy) becomes bounded—less than an *absolute constant* (independent of the values of T_1, T_2)—which is of course smaller than the $\sqrt{\log T}$ in Theorem 4 (which tends to infinity, though very slowly, as $T \to \infty$).

(6) *A novel application to volume computation.* The message of Theorem 4 in a nutshell is that the "ugliness" of the subset is irrelevant, and the error term is shockingly small. Shortly speaking: *line sampling* is much more efficient than the traditional point sampling! Note that basically the same "almost constant discrepancy" result holds in the 3-dimensional space for convex sets. This motivates the following completely new way of computing, or rather approximating, the volume of 3-dimensional convex sets (= solids). Convex sets are special in the sense that the intersection with a (straight) line is a line segment (so its length is determined by the two endpoints).

For simplicity, assume that A is a convex subset of the unit cube $[0, 1]^3$; we want to approximate the volume vol(A). First we extend the subset $A \subset [0, 1]^3$ *periodically* over the whole 3-space (of course, the period is 1 in each of the three coordinate directions). Then we choose a "typical" straight line segment of length n in the 3-space (n is "large"). It means, more precisely, that first we choose a starting point $\mathbf{y}$ in (say) the unit cube $[0, 1]^3$, and then we choose a direction $\mathbf{u}$, represented by a point on the unit sphere S^2. The starting point $\mathbf{y} \in [0, 1]^3$, the direction $\mathbf{u} \in S^2$, and the length n uniquely determine a straight line segment $L(n) = L(\mathbf{y}; \mathbf{u}; n)$ in the ordinary 3-space. We consider the intersection of the line segment $L(n) = L(\mathbf{y}; \mathbf{u}; n)$ with the periodic extension of $A \subset [0, 1]^3$ over the whole 3-space: it consists of $O(n)$ "pieces", where each "piece" is a line segment itself. For notational simplicity, let *total*$[L(n) \cap A]$ denote the total length of the $O(n)$ "pieces". The mathematical theorem mentioned above states that, for the overwhelming majority of the initial conditions (initial condition = starting point $\mathbf{y} \in [0, 1]^3$ and direction $\mathbf{u} \in S^2$ together), *total*$[L(n) \cap A]$ is shockingly close to the "expected value" $n \cdot \text{vol}(A)$: the discrepancy

$$|total[L(n) \cap A] - n \cdot \text{vol}(A)|$$

is negligible, and the "ugliness" of the convex set $A \subset [0, 1]^3$ is irrelevant.

The obvious benefit of working with a convex set $A \subset [0, 1]^3$ is that the total length of the $O(n)$ "pieces" *total*$[L(n) \cap A]$ can be easily computed. Indeed, each piece is a line segment itself, so we just need to know the coordinates of the two endpoints: the distance comes from a straightforward application of Pythagorean theorem. Finally, we just add up the $O(n)$ distances. It is relatively easy, therefore, to determine the exact value of *total*$[L(n) \cap A]$, which happens to be very close to n times the volume vol(A). Dividing by n, we obtain a very good approximation of the volume of the convex set $A \subset [0, 1]^3$.

It is an important technical question how to actually determine the two endpoints of a "piece" (= short line segment). The answer heavily depends on the way the convex set $A \subset [0, 1]^3$ is described. A typical way to describe a complicated convex set is to represent it as the intersection of a few simpler convex sets. As an illustration, consider the case when $A \subset [0, 1]^3$ is the intersection of a ball B, a cube C, and a tetrahedron D: $A = B \cap C \cap D$. Of course, the volumes vol(B), vol(C), vol(D) are easy to be computed, but the volume of the intersection vol$(B \cap C \cap D) = $ vol(A) has nothing to do with the simplicity of vol(B), vol(C), vol(D); the computation of vol$(B \cap C \cap D)$ is hard! But here comes the advantage of the line sampling: the endpoints of a "piece" (= intersection of A with a line) are either on B, or C, or D, and the intersection of a straight line with a ball (or cube, or tetrahedron) is a trivial calculation. This gives a practical solution for the actual computation of the endpoints of the "pieces" for very large classes of convex sets.

I just outlined the 3-dimensional case, but this procedure clearly works in any dimension, and seems very promising even if the set is not convex. What we really need is that the intersection of the line segment $L(n)$ and the periodic extension of A consists of "not too many pieces (= short line-segments)".

Note that the 2-dimensional case (= area computation) is particularly simple: then we have a lucky shortcut. Indeed, "area under the curve" is just the definite integral of a function in one variable. This leads us to the classical field of numerical integration and the classical quadrature formulas (e.g., midpoint rule, trapezoidal rule, Simpson's rule).

Unfortunately, in higher dimensions—i.e., for functions of at least two variables—the classical quadrature formulas all break down, mainly because there is no natural analog of the "equidistant set in the unit interval" $0, 1/n, 2/n, \ldots, (n-1)/n$ in higher dimensions. This is why in higher dimensions the only practical solution is the Monte Carlo method ("random sampling"). What we demonstrate here is that, in the classical subject of volume computation, a novel way of regular sampling—namely, line sampling—can beat the Monte Carlo method.

I conclude Sect. 4 with a proof of Proposition 4.1 (the proof of Theorem 4 is postponed to Sect. 6).

4.3 Proof of Proposition 4.1

The basic idea is to involve probability theory (I assume that the reader is familiar with the basic concepts such as probability space, event, and random variable).

Given any $0 < \alpha < 1$ and $0 < \beta < 1$, let $\mathcal{X}(\alpha, \beta; N)$ denote the arithmetic progression $\beta + \alpha, \beta + 2\alpha, \ldots, \beta + N\alpha$ (mod 1). Here (mod 1) means that we take the fractional part of $\beta + j\alpha$ $(1 \le j \le N)$, so the elements of $\mathcal{X}(\alpha, \beta; N)$ are all in the unit interval $[0, 1)$. We write the elements of $\mathcal{X}(\alpha, \beta; N)$ in increasing order:

$$\mathcal{X}(\alpha, \beta; N) = \{x_1, x_2, \ldots, x_N\} \quad \text{where}$$
$$0 \le x_1 = \beta + j_1\alpha < x_2 = \beta + j_2\alpha < \cdots < x_N = \beta + j_N\alpha < 1$$

and $j_1, j_2, \ldots, j_N$ is a permutations of $1, 2, \ldots, N$.

In general, given an arbitrary increasing sequence

$$0 \le x_1 < x_2 < x_3 < \cdots < x_m < 1$$

in the unit interval, let $gap(\mathcal{X})$ denote the smallest gap:

$$gap(\mathcal{X}) = \min_{1 \le j \le N} (x_{j+1} - x_j) \quad \text{where } x_{N+1} = 1 + x_1$$

and $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$. We choose a sufficiently large positive integer r such that $0 < \frac{1}{r} < gap(\mathcal{X})$. We define 2^r modified ± 1-valued Rademacher functions as follows. For every vector

$$\mathbf{v} = (v_1, v_2, \ldots, v_r) \quad \text{with } v_j \in \{-1, 1\},\ j = 1, 2, \ldots, r,$$

we define a modified Rademacher function $R_{\mathbf{v}}(x)$ in $0 < x < 1$ by the following rule:

$$\text{if } \frac{2j-2}{2r} \le x < \frac{2j-1}{2r} \quad \text{then } R_{\mathbf{v}}(x) = v_j, \quad \text{and}$$
$$\text{if } \frac{2j-1}{2r} \le x < \frac{2j}{2r} \quad \text{then } R_{\mathbf{v}}(x) = -v_j.$$

The idea behind this construction is to associate with the given point set

$$\mathcal{X} = \{x_1, x_2, x_3, \ldots, x_N\}$$

(having minimum gap $> 1/r$) a sequence of N independent random variables: namely, a sequence of Heads and Tails of length N. Indeed, we consider the set

$$\Omega^* = \{\mathbf{v} = (v_1, v_2, \ldots, v_r) : v_j \in \{-1, 1\},\ j = 1, 2, \ldots, r\}$$

a discrete probability space (or sample space); define the events

$$H_i = \{\mathbf{v} \in \Omega^* : R_{\mathbf{v}}(y_i) = 1\}, \qquad T_i = \{\mathbf{v} \in \Omega^* : R_{\mathbf{v}}(y_i) = -1\}$$

(here I used the letters "H" and "T" on purpose to indicate Heads and Tails), and define the random variables X_i, $i = 1, 2, \ldots, N$:

$$X_i = X_i(\mathbf{v}) = R_{\mathbf{v}}(x_i),$$

which has value 1 or -1 depending on whether $\mathbf{v} \in H_i$ or $\mathbf{v} \in T_i$ (Heads and Tails). For every $\mathbf{v} \in \Omega^*$ the average

$$\frac{1}{N} \sum_{i=1}^{N} X_i(\mathbf{v}) = \frac{1}{N} \sum_{i=1}^{N} R_{\mathbf{v}}(x_i)$$

is a Riemann sum approximating the integral $\int_0^1 R_{\mathbf{v}}(x)\,dx = 0$ which equals zero. It follows from the construction, based on modified Rademacher functions and the gap condition, that $X_1, X_2, \ldots, X_N$ is a sequence of *independent* random variables with common distribution $\Pr[X_i = 1] = \Pr[X_i = -1] = 1/2$. That is, $X_1, X_2, \ldots, X_N$ is indeed a sequence of Heads and Tails.

Therefore, we can apply the central limit theorem, which says that for any fixed real number $\lambda > 0$,

$$\Pr[|X_1 + X_2 + \cdots + X_N| \le \lambda\sqrt{N}] = \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-u^2/2}\,du + O(N^{-1/2}). \tag{4.13}$$

Here "Pr" means *equiprobability* in the discrete probability space Ω^*. Since

$$\begin{aligned} &|X_1 + X_2 + \cdots + X_N| \\ &\quad = \left|\sum_{i=1}^{N} R_{\mathbf{v}}(x_i)\right| = \left|\sum_{i=1}^{N} R_{\mathbf{v}}(x_i) - N\int_0^1 R_{\mathbf{v}}(x)\,dx\right| = \mathrm{error}(\mathbf{v}; \mathcal{X}), \end{aligned}$$

we can rewrite (4.13) as

$$\Pr[\mathrm{error}(\mathbf{v}; \mathcal{X}) \le \lambda\sqrt{X}] = c_0(\lambda) + O(N^{-1/2}), \tag{4.14}$$

where

$$c_0(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-u^2/2}\,du.$$

Equation (4.14) is clearly equivalent to

$$2^{-r}\left|\left\{\mathbf{v} \in \Omega^* : \mathrm{error}(\mathbf{v}; \mathcal{X}) \le \lambda\sqrt{N}\right\}\right| = c_0(\lambda) + O(N^{-1/2}), \tag{4.15}$$

where $\mathrm{error}(\mathbf{v}; \mathcal{X})$ stands for the approximation error

$$\left|\sum_{i=1}^{N} R_{\mathbf{v}}(x_i) - N\int_0^1 R_{\mathbf{v}}(x)\,dx\right|,$$

based on the N-element set $\mathcal{X} = \{x_1, x_2, \ldots, x_m\}$ of division points.

Next we return to the special case $\mathcal{X} = \mathcal{X}(\alpha, \beta; N)$:

$$\begin{aligned} &\mathcal{X}(\alpha, \beta; N) = \{x_1, x_2, \ldots, x_N\} \quad \text{where} \\ &\quad 0 \le x_1 = \beta + j_1\alpha < x_2 = \beta + j_2\alpha < \cdots < x_N = \beta + j_N\alpha < 1 \end{aligned}$$

and $j_1, j_2, \ldots, j_N$ is a permutations of $1, 2, \ldots, N$. We show that for the majority of the pairs $(\alpha, \beta) \in [0, 1]^2$,

$$gap(\mathcal{X}(\alpha, \beta; N)) > \frac{1}{N^2}, \tag{4.16}$$

implying that for these pairs (α, β) we can specify the value of parameter r (used in the construction of the independent random variables $X_1, \ldots, X_N$ above) as $r = N^2$.

To prove (4.16), first note that the minimum gap $gap(\mathcal{X}(\alpha, \beta; N))$ is independent of the value of β, so it suffices to study the sequence $\|j\alpha\|$, $1 \le j < N$, where $\|y\|$ denotes the distance of a real number y from the nearest integer. Thus we have

$$gap(\mathcal{X}(\alpha, \beta; N)) = gap(\mathcal{X}(\alpha, 0; N)) = \min_{1 \le k \le N-1} \|k\alpha\|. \tag{4.17}$$

The "bad" case

$$\min_{1 \le k \le N-1} \|k\alpha\| \le \frac{1}{N^2}$$

means that, there is an integer $m = m(\alpha)$ in $1 \le m \le N - 1$ such that

$$\|m\alpha\| \le \frac{1}{N^2},$$

or equivalently,

$$\alpha \in \left[\frac{q}{m} - \frac{1}{mN^2}, \frac{q}{m} + \frac{1}{mN^2}\right] = I(m; q; N)$$

for some integer $q = q(\alpha)$ in $0 \le q \le m$. We form the union set

$$B = B(N) = \bigcup_{1 \le m < N} \bigcup_{0 \le q \le m} I(m; q; N),$$

which is considered the set of "bad" α's. We can easily estimate the Lebesgue measure (i.e., length) of the set B:

$$\begin{aligned} \text{measure}(B) = \text{measure}(B(N)) &\le \sum_{1 \le m < N} \sum_{0 \le q \le m} \text{length}\,(I(m; q; N)) \\ &= \sum_{1 \le m < N} \sum_{0 \le q \le m} \frac{2}{mN^2} = \frac{2}{N^2} \sum_{1 \le m < N} \frac{m+1}{m} < \frac{4}{N}. \end{aligned} \tag{4.18}$$

Consider the set of "good" pairs

$$\mathcal{G}(N) = \left\{(\alpha, \beta) \in [0, 1]^2 : gap(\mathcal{X}(\alpha, \beta; N)) > \frac{1}{N^2}\right\}.$$

By (4.16)–(4.18),

$$\text{area}(\mathcal{G}(N)) > 1 - \frac{4}{N}, \tag{4.19}$$

where of course "area" means the two-dimensional Lebesgue measure.

Applying (4.15) for every pair $(\alpha, \beta) \in \mathcal{G}(N)$ with $r = N^2$, we have

$$\begin{aligned} &\frac{1}{\text{area}(\mathcal{G}(N))} \int_{\mathcal{G}(N)} 2^{-r} \left|\left\{\mathbf{v} \in \Omega^* : \text{error}(\mathbf{v}; \mathcal{X}(\alpha, \beta; N)) \le \lambda\sqrt{N}\right\}\right| \, d\alpha \, d\beta \\ &\quad = c_0(\lambda) + O(N^{-1/2}). \end{aligned} \tag{4.20}$$

Since (4.20) is an average, there must exist a $\mathbf{v}_0 \in \Omega^*$ such that

$$\frac{\text{area}\{(\alpha, \beta) \in \mathcal{G}(N) : \text{error}(\mathbf{v}_0; \mathcal{X}(\alpha, \beta; N)) \le \lambda\sqrt{N}\}}{\text{area}(\mathcal{G}(N))} \le c_0(\lambda) + O(N^{-1/2}). \tag{4.21}$$

We choose the value of $\lambda > 0$ such that

$$c_0(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-u^2/2}\, du = \frac{1}{3}.$$

The tables of the normal distribution give the explicit value $\lambda = \lambda_0 = 0.43$. Then, by (4.21), for the majority of the pairs $(\alpha, \beta) \in \mathcal{G}(N)$,

$$\text{error}(\mathbf{v}_0; \mathcal{X}) = \left| \sum_{i=1}^{N} R_{\mathbf{v}_0}(x_i) - N \int_0^1 R_{\mathbf{v}_0}(x)\, dx \right| > \lambda_0 \sqrt{N} = 0.43\sqrt{N}. \tag{4.22}$$

The modified Rademacher function $R_{\mathbf{v}_0}(x)$ has values ± 1, but Proposition 4.1 is about a subset $A = A_N \subset [0, 1]$ of measure $1/2$. The switch from $R_{\mathbf{v}_0}(x)$ to the desired $A = A_N \subset [0, 1]$ is obvious: $A = A_N$ is defined as the set of $x \in [0, 1]$ for which $R_{\mathbf{v}_0}(x) = 1$. The switch from ± 1 to $1, 0$ (i.e., the characteristic function of A) means a “halving” in (4.22), which completes the proof of Proposition 4.1.

5 Proof of Theorem 1: Starting with Fourier Analysis

5.1 Using Parseval’s Formula

In view of the geometric trick of *unfolding* the billiard paths to straight lines in the 3-space, it suffices to deal with N torus lines $\mathbf{x}_k(t) = (x_{k,1}(t), x_{k,2}(t), x_{k,3}(t))$ (mod 1), $k = 1, 2, \ldots, N$ where

$$x_{k,1}(t) = u_{k,1} t v + y_{k,1}, \qquad x_{k,2}(t) = u_{k,2} t v + y_{k,2}, \qquad x_{k,3}(t) = u_{k,3} t v + y_{k,3} \tag{5.1}$$

and

$$u_{k,1}^2 + u_{k,2}^2 + u_{k,3}^2 = 1, \tag{5.1'}$$

i.e., $\mathbf{u}_k = (u_{k,1}, u_{k,2}, u_{k,3})$ is a unit vector. Since $v \geq 1$ is the common speed of the particles, the length of the straight line segment $\mathbf{x}_k(t)$, $0 < t < T$ is clearly vT. The pair $(\mathbf{y}_k, \mathbf{u}_k)$ describes the starting point $\mathbf{y}_k = \mathbf{x}_k(0) \in I^3 = [0, 1)^3$ and the direction $\mathbf{u}_k \in S^2$ (S^2 is the unit sphere) of the kth torus-line $\mathbf{x}_k(t)$. We call the pair $(\mathbf{y}_k, \mathbf{u}_k)$ the initial condition of the kth torus-line $\mathbf{x}_k(t)$.

Let $A \subset I^3 = [0, 1)^3$ be an arbitrary Lebesgue measurable subset. Via unfolding it corresponds to the union of 8 copies of the given subset in Theorem 1, where we shrink the corresponding $2 \times 2 \times 2$ cube to the unit cube. Consider the Fourier series of the 0-1-valued characteristic function χ_A of the set A:

$$\chi_A(\mathbf{w}) = \sum_{\mathbf{r} \in \mathbb{Z}^3} a_{\mathbf{r}} e^{2\pi i \mathbf{r} \cdot \mathbf{w}} \quad \text{with } a_{\mathbf{r}} = \int_A e^{-2\pi i \mathbf{r} \cdot \mathbf{z}}\, d\mathbf{z}, \tag{5.2}$$

where $\mathbf{r} \cdot \mathbf{w} = r_1 w_1 + r_2 w_2 + r_3 w_3$ denotes the standard inner product of vectors. Clearly $a_{\mathbf{0}} = \text{vol}(A)$ ($=$ the volume of A), and by Parseval’s formula,

$$\sum_{\substack{\mathbf{r} \in \mathbb{Z}^3: \\ \mathbf{r} \neq \mathbf{0}}} |a_{\mathbf{r}}|^2 = \int_{I^3} \chi_A^2(\mathbf{w})\, d\mathbf{w} - |a_{\mathbf{0}}|^2 = \text{vol}(A) - \text{vol}^2(A). \tag{5.3}$$

Let $T \geq 1$ be a real number. We denote the total time that the kth torus-line $\mathbf{x}_k(t)$ (defined in (5.1)) spends in subset A during $0 < t < T$ by $A_k(T) = A_k(T; \mathbf{y}_k, \mathbf{u}_k)$: we have

$$\begin{aligned} A_k(T) = A_k(T; \mathbf{y}_k, \mathbf{u}_k) &= \text{measure}\,\{t \in [0, T] : \mathbf{x}_k(t) \in A \ (\text{mod } 1)\} \\ &= \int_0^T \chi_A(\mathbf{x}_k(t))\,dt = \int_0^T \sum_{\mathbf{r}\in\mathbb{Z}^3} a_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{x}_k(t)}\,dt \\ &= \sum_{\mathbf{r}\in\mathbb{Z}^3} a_{\mathbf{r}} \int_0^T e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{x}_k(t)}\,dt = \sum_{\mathbf{r}\in\mathbb{Z}^3} a_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{y}_k} \int_0^T e^{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vt}\,dt \\ &= a_{\mathbf{0}}T + \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}} a_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{y}_k} \cdot \frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vT} - 1}{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)v}. \end{aligned} \tag{5.4}$$

Let M be an arbitrary integer in the range $1 \leq M \leq N$, and consider the sum

$$F_M = F_M(\mathbf{y}_k, \mathbf{u}_k : k = 1, 2, \ldots, M) = \sum_{k=1}^{M} \left(\frac{1}{T} A_k(T; \mathbf{y}_k, \mathbf{u}_k) - \text{vol}(A) \right). \tag{5.5}$$

Fix the M unit vectors $\mathbf{u}_k \in S^2$, $k = 1, 2, \ldots, M$, and evaluate the square integral

$$\sum\nolimits_1 (\mathbf{u}_k : k = 1, 2, \ldots, M) = \int_{I^3} \cdots \int_{I^3} (F_M(\mathbf{y}_k, \mathbf{u}_k : k = 1, 2, \ldots, M))^2\, d\mathbf{y}_1 \ldots d\mathbf{y}_M. \tag{5.6}$$

Note that (5.6) is a multiple integral, which consists of M single integrals.

The evaluation of (5.6) is rather simple if we multiply out the square F_M^2—where the sum F_M is defined in (5.4)–(5.5)—and apply the two orthogonality relations:

$$\int_{I^3} \int_{I^3} e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot\mathbf{y}_j - \mathbf{r}_2\cdot\mathbf{y}_k)}\, d\mathbf{y}_j\, d\mathbf{y}_k = 0 \tag{5.7}$$

for any $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{Z}^3 \setminus \mathbf{0}$ and $j \neq k$, and

$$\int_{I^3} e^{2\pi \mathrm{i}(\mathbf{r}_1 - \mathbf{r}_2)\cdot\mathbf{y}}\, d\mathbf{y} = 0, \tag{5.8}$$

unless $\mathbf{r}_1 = \mathbf{r}_2$. Then the majority of the terms (= sub-integrals) turn out to be zero, and we obtain the relatively simple sum

$$\begin{aligned} \sum\nolimits_1 &= \sum\nolimits_1 (\mathbf{u}_k : k = 1, 2, \ldots, M) \\ &= \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}} |a_{\mathbf{r}}|^2 \cdot \sum_{k=1}^{M} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vT} - 1}{2\pi(\mathbf{r}\cdot\mathbf{u}_k)vT} \right|^2. \end{aligned} \tag{5.9}$$

We can also say that (5.9) is a consequence of Parseval's formula.

Next we integrate $\sum_1$ over the direction vectors $\mathbf{u}_k \in S^2$, $k = 1, 2, \ldots, M$, which leads to another multiple integral consisting of M single integrals (in order to normalize, we have to divide by 4π, which is the surface area of the unit sphere S^2):

$$\sum\nolimits_1^* = \frac{1}{4\pi} \int_{S^2} \cdots \frac{1}{4\pi} \int_{S^2} \sum\nolimits_1 (\mathbf{u}_k : k = 1, 2, \ldots, M)\, d\mathbf{u}_1 \ldots d\mathbf{u}_M$$

$$= \frac{1}{4\pi}\int_{S^2}\dots\frac{1}{4\pi}\int_{S^2}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}}|a_{\mathbf{r}}|^2\cdot\sum_{k=1}^{M}\left|\frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vT}-1}{2\pi(\mathbf{r}\cdot\mathbf{u}_k)vT}\right|^2 d\mathbf{u}_1\dots d\mathbf{u}_M. \tag{5.10}$$

Let's focus on the last factor in (5.10): we have the obvious upper bound

$$\left|\frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vT}-1}{2\pi(\mathbf{r}\cdot\mathbf{u}_k)vT}\right| \le \min\left\{\frac{1}{\pi|\mathbf{r}\cdot\mathbf{u}_k|vT},1\right\}. \tag{5.11}$$

I recall that the dot product $\mathbf{r}\cdot\mathbf{u}$ has magnitude $|\mathbf{r}\cdot\mathbf{u}| = |\mathbf{r}|\cos\theta$, where θ is the angle between the unit vector $\mathbf{u}\in S^2$ and $\mathbf{r}\in\mathbb{Z}^3\setminus\mathbf{0}$.

In order to apply (5.11) in (5.10), we need a well-known fact about the surface area of some spherical domains. Let $\mathbf{e}\in S^2$ be an arbitrary but fixed unit vector, and let $0<\delta<1$ be an arbitrary real number. Consider the "spherical belt" (often called the "spherical zone"):

$$S^2(\mathbf{e};\delta) = \left\{\mathbf{u}\in S^2 : |\mathbf{e}\cdot\mathbf{u}|\le\delta\right\}.$$

On one hand it is trivial that the surface area of the "belt" $S^2(\mathbf{e};\delta)$ is independent of $\mathbf{e}$; it is very surprising, on the other hand, that the dependence on δ is just plain linear:

$$\text{SurfaceArea}\left(S^2(\mathbf{e};\delta)\right) = 4\pi\delta. \tag{5.12}$$

Equation (5.12) implies that, for any $\mathbf{r}\in\mathbb{Z}^3\setminus\mathbf{0}$ and any real number ϱ with $0<\varrho<|\mathbf{r}|$,

$$\text{SurfaceArea}\left(\left\{\mathbf{u}\in S^2 : |\mathbf{r}\cdot\mathbf{u}|\le\varrho\right\}\right) = 4\pi\cdot\frac{\varrho}{|\mathbf{r}|}. \tag{5.13}$$

By using (5.11) and (5.13), we can easily estimate (5.10) as follows. First note that

$$\min\left\{\frac{1}{\pi|\mathbf{r}\cdot\mathbf{u}|vT},1\right\} \iff |\mathbf{r}\cdot\mathbf{u}|\le\frac{1}{\pi vT}. \tag{5.14}$$

By (5.13)–(5.14),

$$\begin{aligned}
&\frac{1}{4\pi}\int_{S^2}\min\left\{\frac{1}{(\pi|\mathbf{r}\cdot\mathbf{u}|vT)^2},1\right\}d\mathbf{u}\\
&= \frac{1}{\pi vT|\mathbf{r}|}+\frac{1}{(\pi|\mathbf{r}|vT)^2}\int_{\delta(\mathbf{r})}^{1}\frac{dx}{x^2} = \frac{1}{\pi vT|\mathbf{r}|}+\frac{1}{(\pi|\mathbf{r}|vT)^2}(\pi vT|\mathbf{r}|-1)\\
&= \frac{2}{\pi vT|\mathbf{r}|}-\frac{1}{(\pi vT|\mathbf{r}|)^2},
\end{aligned} \tag{5.15}$$

where $\delta(\mathbf{r}) = (\pi vT|\mathbf{r}|)^{-1}$ (see (5.14)).

By (5.10)–(5.11) and (5.15),

$$\begin{aligned}
\sum\nolimits_1^{*} &\le \sum_{k=1}^{M}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}}|a_{\mathbf{r}}|^2\cdot\left(\frac{1}{4\pi}\int_{S^2}\dots\frac{1}{4\pi}\int_{S^2}\min\left\{\frac{1}{(\pi|\mathbf{r}\cdot\mathbf{u}_k|vT)^2},1\right\}d\mathbf{u}_1\dots d\mathbf{u}_M\right)\\
&= M\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}}|a_{\mathbf{r}}|^2\cdot\left(\frac{2}{\pi vT|\mathbf{r}|}-\frac{1}{(\pi vT|\mathbf{r}|)^2}\right)
\end{aligned}$$

$$\leq \frac{2M}{\pi v T} \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}} |a_{\mathbf{r}}|^2 \cdot \frac{1}{|\mathbf{r}|} \leq \frac{2M}{\pi v T} \cdot \mathrm{vol}(A), \tag{5.16}$$

where in the last step we applied (5.3). Therefore, we just completed the proof of the following result.

Lemma 5.1 *We have*

$$\sum\nolimits_1^* \leq \frac{2}{\pi v T} \cdot M \cdot \mathrm{vol}(A), \tag{5.17}$$

where the square-integral $\sum_1^*$ (*see* (5.10)) *is the following multiple integral*

$$(4\pi)^{-M} \int_{I^3} \cdots \int_{I^3} \int_{S^2} \cdots \int_{S^2} \left(\sum_{k=1}^{M} \left(\frac{1}{T} A_k(T; \mathbf{y}_k, \mathbf{u}_k) - \mathrm{vol}(A) \right) \right)^2 d\mathbf{y}_1 \ldots d\mathbf{y}_M \, d\mathbf{u}_1 \ldots d\mathbf{u}_M, \tag{5.18}$$

which consists of $2M$ *single integrals.*

Notice that in (5.18) we had to break up the long integral into two lines; we keep up doing this practice for the rest of the paper.

A Technical Note The argument in (5.4) (and in (5.9), and in many similar cases below) is rather informal: for example, we changed the order of infinite summation and integration, but didn't say anything about under what condition can we really do that. Note, for example, that for an "ugly" measurable set A the Fourier series of the characteristic function χ_A can be divergent in many points, but this kind of technical nuisance is totally irrelevant for us: what we really care about is the Parseval formula. It is a well-known fact that the Parseval formula characterizes the L_2 space = the class of functions for which the Lebesgue square integral $\int_0^1 f^2(x)\,dx$ exists (and finite); see the Riesz–Fisher theorem. The point is that the characteristic function $f = \chi_A$ clearly belongs to L_2, and we can *safely work* in the L_2 space ("Lebesgue square-integrable"). The precise proof of this is a standard argument that I very briefly outline here. If $f = g$ is a sufficiently smooth function—say, twice differentiable with continuous derivative—then its Fourier series is absolutely convergent, and so every manipulation that we carried out (such as, changing the order of summation and integration) is perfectly justified and legitimate. The last step is to approximate a Lebesgue square-integrable function f with a sequence of smooth functions g_k, $k = 1, 2, 3, \ldots$: we want the approximation error $|f(x) - g_k(x)|$ to be arbitrarily small (as $k \to \infty$) except on a set of x's with small Lebesgue measure (the measure tends to zero as $k \to \infty$). This last step is just another routine argument in the theory of the Lebesgue integral.

Now we are ready to give

5.2 A Brief Outline of the Rest of the Proof of Theorem 1

We introduce the random variable $E = E(T)$ (see (5.5)):

$$E = E(N; [1, M]; T) = E(T; \mathbf{y}_k, \mathbf{u}_k : 1 \leq k \leq M) = \frac{1}{T} \sum_{k=1}^{M} A_k(T; \mathbf{y}_k, \mathbf{u}_k). \tag{5.19}$$

Combining Lemma 5.1 with Chebyshev's well-known inequality, we obtain that the *typical* size of E is

$$E = M \cdot \mathrm{vol}(A) + O\left(\sqrt{\frac{1}{vT} \cdot M \cdot \mathrm{vol}(A)}\right), \tag{5.20}$$

and (5.20) holds for the majority of the initial conditions $\mathbf{y}_k, \mathbf{u}_k : 1 \le k \le M$. For notational simplicity assume that the ratio $N/M = m \ge 1$ is an integer; the optimal choice of parameter m will be specified later. Let Γ denote the set of all $N!$ permutations of $1, 2, \ldots, N$, and let $\gamma \in \Gamma$ be an arbitrary permutation, i.e., $\gamma(1), \gamma(2), \ldots, \gamma(N)$ is a rearrangement of the first N integers. (The reason behind introducing the permutations is to validate the combinatorial calculations in (5.27)–(5.31) below.) Write

$$\begin{aligned} &E_1(\gamma) = E(N; [1, M]; T; \gamma), \qquad E_2(\gamma) = E(N; [M+1, 2M]; T; \gamma), \\ &E_3(\gamma) = E(N; [2M+1, 3M]; T; \gamma), \quad \ldots, \quad E_m(\gamma) = E(N; [N-M+1, N]; T; \gamma), \end{aligned} \tag{5.21}$$

where for any integer h in $1 \le h \le m$,

$$\begin{aligned} E_h(\gamma) &= E_h(T; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : (h-1)M + 1 \le k \le hM; \gamma) \\ &= \frac{1}{T} \sum_{k=(h-1)M+1}^{hM} A_{\gamma(k)}(T; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)}) \\ &= \frac{1}{T} \int_0^T Z_h(\gamma; t)\, dt, \end{aligned} \tag{5.22}$$

where

$$\begin{aligned} Z_h(\gamma; t) &= Z_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : (h-1)M + 1 \le k \le hM; \gamma; t) \\ &= \sum_{k=(h-1)M+1}^{hM} \chi_A(\mathbf{x}_{\gamma(k)}(t)). \end{aligned} \tag{5.23}$$

In view of (5.22)–(5.23) we can say that, the " time average" (or "expectation") of $Z_h(\gamma; t)$, as t runs in $0 < t < T$, equals $E_h(\gamma)$; or formally,

$$\frac{1}{T} \int_0^T Z_h(\gamma; t)\, dt = E_h(\gamma). \tag{5.24}$$

The values of $Z_h(\gamma; t)$, as t runs in $0 < t < T$, are non-negative integers $0, 1, 2, 3, \ldots$; now for every integer $\ell \ge 0$ we define the set

$$\begin{aligned} W_h(\gamma; \ell) &= W_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : (h-1)M + 1 \le k \le hM; \gamma; \ell) \\ &= \{t \in [0, T] : Z_h(\gamma; t) = \ell\}. \end{aligned} \tag{5.25}$$

Then we have the following disjoint decomposition of the interval $0 \le t \le T$:

$$\begin{aligned} [0, T] &= W_h(\gamma; 0) \cup W_h(\gamma; 1) \cup W_h(\gamma; 2) \cup W_h(\gamma; 3) \cup \cdots \\ &= W_h(\gamma; 0) \cup W_h(\gamma; 1) \cup W_h(\gamma; \ge 2). \end{aligned}$$

Write

$$V_h(\gamma;\ell)=V_h(\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:(h-1)M+1\le k\le hM;\gamma;\ell)=\frac{1}{T}\text{measure}\,(W_h(\gamma;\ell))\,,$$

implying

$$0\le V_h(\gamma;\ell)\le 1.$$

We need to give an upper bound for the size $V_h(\gamma;\ell)$ of a "typical" set $W_h(\gamma;\ell)$ with $\ell\ge 2$ ("typical" means the majority of the initial conditions $\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:(h-1)M+1\le k\le hM$ and a "typical" permutation γ). In fact, we will estimate a whole power-of-two group

$$\sum_{\ell=2^j}^{2^{j+1}-1} V_h(\gamma;\ell)\quad\text{for any integer } j\ge 1.$$

We will obtain such an upper bound by using a second moment argument that I am outlining below. For notational simplicity I omit the fixed permutation $\gamma\in\Gamma$ in the notation.

Since $\ell\ge 2$, we can write $\ell=\ell_1+\ell_2$ with $1\le\ell_1=\lfloor\ell/2\rfloor$ and $1\le\ell_2=\lceil\ell/2\rceil$. For simplicity assume that M is even; let $I_1\cup I_2$ be an arbitrary *halving split* of the set $\{(h-1)M+1,(h-1)M+2,\dots,hM\}$ of M integers into two disjoint subsets of size $M/2$ each. There are exactly $\binom{M}{M/2}$ such halving splits. For any fixed halving split (I_1,I_2), write

$$Z_{I_1}(t)Z_{I_2}(t)=\left(\sum_{k_1\in I_1}\chi_A(\mathbf{x}_{k_1}(t))\right)\left(\sum_{k_2\in I_2}\chi_A(\mathbf{x}_{k_2}(t))\right).$$

Let $\ell\ge 2$; if $t_0\in W_h(\ell)\Leftrightarrow Z_h(t_0)=\ell$ for some $0\le t_0\le T$, then at least $\binom{\ell}{\ell_1}2^{-\ell}$ part of the $\binom{M}{M/2}$ halving splits have the property that

$$Z_{I_1}(t_0)=\ell_1\quad\text{and}\quad Z_{I_2}(t_0)=\ell_2,\quad\text{implying } Z_{I_1}(t_0)Z_{I_2}(t_0)=\ell_1\ell_2. \tag{5.26}$$

Indeed, this fact is a corollary of a standard combinatorial problem as follows. Suppose a box contains ℓ red balls and $M-\ell$ white balls, where $2\le\ell\le M(=\text{even})$. If we randomly choose half of the balls from the box, what is the probability that we have exactly ℓ_1 red balls among them? The probability in question is clearly

$$Prob(\ell,M)=\frac{\binom{\ell}{\ell_1}\binom{M-\ell}{M/2-\ell_1}}{\binom{M}{M/2}}. \tag{5.27}$$

If ℓ is fixed and $M\to\infty$ then $\lim Prob(\ell,M)=\binom{\ell}{\ell_1}2^{-\ell}$, and a simple calculation shows that for any $2\le\ell\le M(=\text{even})$ we have the inequality

$$Prob(\ell,M)\ge\binom{\ell}{\ell_1}2^{-\ell}\ge\frac{1}{2\sqrt{\ell}}. \tag{5.28}$$

Similarly, we extend the case

$$Z_{I_1}(t_0)=\ell_1=\lfloor\ell/2\rfloor\quad\text{and}\quad Z_{I_2}(t_0)=\ell_2=\lceil\ell/2\rceil$$

in (5.26) to all

$$Z_{I_1}(t_0) = \ell' \quad \text{and} \quad Z_{I_2}(t_0) = \ell'' \quad \text{where } \frac{\ell - \sqrt{2\ell}}{2} < \ell' < \frac{\ell + \sqrt{2\ell}}{2} \tag{5.29}$$

and $\ell'' = \ell - \ell'$. This leads to the following extension of (5.28): for any $\ell \geq 2$ we have

$$2^{-\ell} \sum_{\left|\frac{\ell}{2} - \ell'\right| < \sqrt{\ell/2}} \binom{\ell}{\ell'} \geq \frac{1}{2}. \tag{5.30}$$

(The inequality in (5.30) is a well-known elementary fact about the binomial distribution in probability theory: it can be proved by Chebyshev's inequality, or simply by estimating the binomial coefficients via Stirling's formula; note that for large ℓ the central limit theorem gives a limit constant which is better than $1/2$.)

Let's return to (5.29): we clearly have

$$Z_{I_1}(t_0) Z_{I_2}(t_0) = \ell' \ell'' > \frac{\ell - \sqrt{2\ell}}{2} \cdot \frac{\ell + \sqrt{2\ell}}{2} = \frac{\ell^2 - 2\ell}{4} \tag{5.31}$$

for all ℓ', ℓ'' satisfying (5.29).

It follows from (5.30) that, in order to estimate the sum of $V_h(\ell)$ with $2^j \leq \ell < 2^{j+1}$ in the typical case, it suffices to have an analog of Lemma 5.1 for the term

$$E(I_1, I_2) = \frac{1}{T} \int_0^T Z_{I_1}(t) Z_{I_2}(t)\, dt.$$

We will prove later that typically

$$E(I_1, I_2) = \left(\frac{M}{2} \cdot \mathrm{vol}(A) \right)^2 + O\left(\frac{1}{\sqrt{vT}} M \cdot \mathrm{vol}(A) \right),$$

see Lemma 7.1 at the end of Sect. 7. For the general case; see Lemma 9.1 at the end of Sect. 9. Finally, by putting these facts together, we will be able to complete the proof of Theorem 1 by repeated applications of a basically Chebyshev's type inequality; see Sects. 10–12.

We conclude this section with a brief preview of how the proof of Theorem 1 actually ends (see Sect. 12 after (12.9)). Let's return to (5.25); for notational convenience, we omit the initial condition and the permutation. We write

$$W_h(\geq 1) = \bigcup_{\ell=1}^{\infty} W_h(\ell).$$

Let μ denote the usual one-dimensional Lebesgue measure ("length"). By the inclusion-exclusion principle ("sieve argument")

$$\begin{aligned} &\mu\{0 \leq t \leq T : \mathbf{x}_k(t) \notin A \text{ for all } k = 1, 2, \ldots, N\} \\ &= \mu\left(\bigcap_{h=1}^{m} W_h(0) \right) = \mu([0, T]) - \sum_{h=1}^{m} \mu\left(W_h(\geq 1)\right) \end{aligned}$$

$$+\sum_{1\le h_1<h_2\le m}\mu\left(W_{h_1}(\ge 1)\cap W_{h_2}(\ge 1)\right)$$
$$-\sum_{1\le h_1<h_2<h_3\le m}\mu\left(W_{h_1}(\ge 1)\cap W_{h_2}(\ge 1)\cap W_{h_3}(\ge 1)\right)$$
$$+\sum_{1\le h_1<h_2<h_3<h_4\le m}\mu\left(W_{h_1}(\ge 1)\cap W_{h_2}(\ge 1)\cap W_{h_3}(\ge 1)\cap W_{h_4}(\ge 1)\right)\mp\cdots. \tag{5.32}$$

We will show that "typically"

$$\frac{1}{T}\mu\left(W_{h_1}(\ge 1)\cap W_{h_2}(\ge 1)\cap\cdots\cap W_{h_j}(\ge 1)\right)=(M\cdot\mathrm{vol}(A))^j+\text{negligible error} \tag{5.33}$$

holds for all $1\le j\le m$ and $1\le h_1<\cdots<h_j\le m$. Let

$$\mathrm{vol}(A)=\frac{\lambda}{N}, \tag{5.34}$$

where $\lambda>0$ is a fixed constant. Using (5.33)–(5.34) in (5.32), we obtain that

$$\begin{aligned}\frac{1}{T}\mu\left(\bigcap_{h=1}^m W_h(0)\right)\\ &=1-m\cdot\frac{\lambda}{m}+\binom{m}{2}\cdot\left(\frac{\lambda}{m}\right)^2\\ &\quad-\binom{m}{3}\cdot\left(\frac{\lambda}{m}\right)^3+\binom{m}{4}\cdot\left(\frac{\lambda}{m}\right)^4\mp\cdots+\text{negligible error.}\end{aligned} \tag{5.35}$$

If m is "large", then

$$\binom{m}{j}\approx\frac{m^j}{j!},$$

and so (5.35) basically gives the Taylor series of e^x with $x=-\lambda$:

$$\begin{aligned}&\frac{1}{T}\mu\{0\le t\le T:\mathbf{x}_k(t)\notin A\ \text{ for all } k=1,2,\ldots,N\}\\ &=\frac{1}{T}\mu\left(\bigcap_{h=1}^m W_h(0)\right)=1-\lambda+\frac{\lambda^2}{2!}-\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}\mp\cdots+\text{negligible error}\\ &=e^{-\lambda}+\text{negligible error,}\end{aligned} \tag{5.36}$$

which is a special case of the Poisson distribution. The main difficulty is how to prove (5.33) (we are guided by the Kronecker–Weyl Theorem (2.1) and the intuition (2.2)).

Now I interrupt the long proof of Theorem 1, and include the proof of Theorem 4. I do it here, because the proof of Theorem 4 is very similar to that of Lemma 5.1; the main novelty is to involve a new diophantine approximation type lemma (see Lemma 6.1 below). Also, in the proof of Theorem 4 we have to be more careful with the estimations (for example, in the last line of (5.16) we just used the trivial lower bound $|\mathbf{r}|\ge 1$, instead of estimating a more complicated sum).

6 Proof of Theorem 4

In view of the trick of *unfolding* the billiard path to a straight line in the plane (explained in Sect. 1), it suffices to deal with torus-lines (of course we shrink the corresponding 2×2 square to the unit square). Let $A \subset I^2 = [0, 1)^2$ be an arbitrary Lebesgue measurable subset (via unfolding it corresponds to the union of four copies of the given subset A in Theorem 4), and consider the Fourier series of the 0-1 valued characteristic function χ_A of the set A:

$$\chi_A(\mathbf{u}) = \sum_{\mathbf{r} \in \mathbb{Z}^2} a_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{u}} \quad \text{with } a_{\mathbf{r}} = \int_A e^{-2\pi i \mathbf{r}\cdot\mathbf{y}}\, d\mathbf{y}, \tag{6.1}$$

where $\mathbf{r}\cdot\mathbf{u} = r_1 u_1 + r_2 u_2$ denotes the standard inner product of vectors. Clearly $a_{\mathbf{0}} = \text{area}(A)$ ($=$ the 2-dimensional Lebesgue measure of A), and by Parseval's formula,

$$\sum_{\substack{\mathbf{r} \in \mathbb{Z}^2: \\ \mathbf{r} \neq \mathbf{0}}} |a_{\mathbf{r}}|^2 = \text{area}(A) - \text{area}^2(A). \tag{6.2}$$

Consider the torus-line $\mathbf{x}(t) = (x_1(t), x_2(t))$ (mod 1) where

$$x_1(t) = \alpha_1 t + y_1, \qquad x_2(t) = \alpha_2 t + y_2 \quad \text{and} \quad \alpha_1^2 + \alpha_2^2 = 1. \tag{6.3}$$

The length of the straight line segment $\mathbf{x}(t)$, $0 < t < T$ is clearly T, i.e., time $=$ arc-length. The pair $(\mathbf{y}, (\alpha_1, \alpha_2))$ describes the starting point $\mathbf{y} \in [0, 1)^2$ and the angle (by the point (α_1, α_2) on the unit circle) of the torus-line $\mathbf{x}(t)$. The total time $A(T) = A(T; \mathbf{y}, (\alpha_1, \alpha_2))$ that the torus-line $\mathbf{x}(t)$ (defined in (6.3)) spends in subset A during $0 < t < T$ equals

$$\begin{aligned}
A(T) &= A(T; \mathbf{y}, (\alpha_1, \alpha_2)) = \text{measure}\,\{t \in [0, T] : \mathbf{x}(t) \in A \text{ (mod 1)}\} \\
&= \int_0^T \chi_A(\mathbf{x}(t))\, dt = \int_0^T \sum_{\mathbf{r} \in \mathbb{Z}^2} a_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{x}(t)}\, dt \\
&= \sum_{\mathbf{r} \in \mathbb{Z}^2} a_{\mathbf{r}} \int_0^T e^{2\pi i \mathbf{r}\cdot\mathbf{x}(t)}\, dt = \sum_{\mathbf{r} \in \mathbb{Z}^2} a_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{y}} \int_0^T e^{2\pi i(\alpha_1 r_1 + \alpha_2 r_2)t}\, dt \\
&= a_{\mathbf{0}} T + \sum_{\substack{\mathbf{r} \in \mathbb{Z}^2: \\ \mathbf{r} \neq \mathbf{0}}} a_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{y}} \cdot \frac{e^{2\pi i(\alpha_1 r_1 + \alpha_2 r_2)T} - 1}{2\pi i(\alpha_1 r_1 + \alpha_2 r_2)}.
\end{aligned} \tag{6.4}$$

Since $a_{\mathbf{0}} = \text{area}(A)$ ($=$ Lebesgue measure of A), by (6.4) we have

$$\begin{aligned}
\text{discrepancy} &= A(T; \mathbf{y}, (\alpha_1, \alpha_2)) - T \cdot \text{area}(A) \\
&= \sum_{\substack{\mathbf{r} \in \mathbb{Z}^2: \\ \mathbf{r} \neq \mathbf{0}}} a_{\mathbf{r}} \cdot \frac{e^{2\pi i(\alpha_1 r_1 + \alpha_2 r_2)T} - 1}{2\pi i(\alpha_1 r_1 + \alpha_2 r_2)} \cdot e^{2\pi i \mathbf{r}\cdot\mathbf{y}}.
\end{aligned} \tag{6.5}$$

Fix any point (α_1, α_2) on the unit circle $\alpha_1^2 + \alpha_2^2 = 1$, and run the starting point $\mathbf{y}$ through the unit square; then by Parseval's formula ($I^2 = [0, 1]^2$)

$$\int_{I^2} (A(T; \mathbf{y}, (\alpha_1, \alpha_2)) - T \cdot \text{area}(A))^2\, d\mathbf{y}$$

$$= \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ \mathbf{r}\neq\mathbf{0}}} |a_{\mathbf{r}}|^2 \cdot \left| \frac{e^{2\pi \mathrm{i}(\alpha_1 r_1+\alpha_2 r_2)T}-1}{2\pi(\alpha_1 r_1+\alpha_2 r_2)} \right|^2. \tag{6.6}$$

Let's study the last factor in (6.6): we have the obvious upper bound

$$\left| \frac{e^{2\pi \mathrm{i}(\alpha_1 r_1+\alpha_2 r_2)T}-1}{2\pi(\alpha_1 r_1+\alpha_2 r_2)} \right| \le \min\left\{ \frac{1}{\pi|\alpha_1 r_1+\alpha_2 r_2|}, T \right\}. \tag{6.7}$$

Key Definition Let $0<\varepsilon<1/2$; we say that a point (α_1,α_2) on the unit circle $\alpha_1^2+\alpha_2^2=1$ is ε-bad if there exists an $\mathbf{r}\in\mathbb{Z}^2$ such that

$$|\alpha_1 r_1+\alpha_2 r_2| \le \frac{\varepsilon}{40|\mathbf{r}|\cdot\log_2|\mathbf{r}|\cdot(\log_2\log_2|\mathbf{r}|)^2} \tag{6.8}$$

for some $|\mathbf{r}|\ge 8$ or

$$|\alpha_1 r_1+\alpha_2 r_2| \le \frac{\varepsilon}{40|\mathbf{r}|} \tag{6.9}$$

for some $1\le|\mathbf{r}|<8$, where $|\mathbf{r}|=\sqrt{r_1^2+r_2^2}$.

Note that the complicated denominator in (6.8) is motivated by the fact that the numerical series

$$\sum_{n=3}^{\infty} \frac{1}{n(\log n)^2} \tag{6.10}$$

is very close to the border of convergence-divergence: the slightly larger series

$$\sum_{n=3}^{\infty} \frac{1}{n\log n}$$

is already divergent, but (6.10) is still convergent (see (6.15) below; of course we could replace the exponent 2 in (6.10) with $1+\varepsilon$, but the gain would be negligible).

Next I show that the set $\mathcal{B}$ of all ε-bad points (α_1,α_2) on the unit circle $\alpha_1^2+\alpha_2^2=1$ forms a small minority: the measure of $\mathcal{B}$ is negligible compared to the circumference 2π of the unit circle. (Note in advance that at the end we will throw out all initial conditions having ε-bad angles.)

Lemma 6.1 *The set $\mathcal{B}$ of ε-bad points (see the Key Definition) is small in the sense that*

$$\frac{\text{measure}(\mathcal{B})}{2\pi} < \frac{\varepsilon}{2}. \tag{6.11}$$

Proof Notice that $\alpha_1 r_1+\alpha_2 r_2$ is a dot product of two vectors, so the absolute value $|\alpha_1 r_1+\alpha_2 r_2|$ equals $|\mathbf{r}|\sin\theta$, where θ is the angle between the unit vector (α_1,α_2) and the vector $(-r_2,r_1)$ perpendicular to $\mathbf{r}=(r_1,r_2)$. Therefore, given any $\mathbf{r}\in\mathbb{Z}^2$ with $|\mathbf{r}|\ge 8$, inequality (6.8) defines two short diametrically opposite arcs on the unit circle $\alpha_1^2+\alpha_2^2=1$ with total arc-length

$$4\arcsin\left(\frac{\varepsilon}{40|\mathbf{r}|^2\cdot\log_2|\mathbf{r}|\cdot(\log_2\log_2|\mathbf{r}|)^2}\right),$$

where of course arcsin is the inverse of sin. Similarly, given any $\mathbf{r} \in \mathbb{Z}^2$ with $1 \le |\mathbf{r}| < 8$, inequality (6.9) defines two short diametrically opposite arcs on the unit circle $\alpha_1^2 + \alpha_2^2 = 1$ with total arc-length

$$4 \arcsin\left(\frac{\varepsilon}{40|\mathbf{r}|^2}\right).$$

It follows that

$$\begin{aligned} \text{measure}(\mathcal{B}) < & \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ 1\le|\mathbf{r}|<8}} 4\arcsin\left(\frac{\varepsilon}{40|\mathbf{r}|^2}\right) \\ & + \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ |\mathbf{r}|\ge 8}} 4\arcsin\left(\frac{\varepsilon}{40|\mathbf{r}|^2 \cdot \log_2|\mathbf{r}| \cdot (\log_2\log_2|\mathbf{r}|)^2}\right). \end{aligned} \tag{6.12}$$

By using the trivial inequality

$$\arcsin(x) < x + x^2 \quad \text{for } 0 < x < 1, \tag{6.13}$$

we can easily estimate the sums in (6.12). We begin with the auxiliary sum

$$\sum\nolimits_1 = \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ |\mathbf{r}|\ge 8}} \frac{1}{|\mathbf{r}|^2 \cdot \log_2|\mathbf{r}| \cdot (\log_2\log_2|\mathbf{r}|)^2}. \tag{6.14}$$

We estimate (6.14) by applying a standard power-of-two decomposition:

$$\begin{aligned} \sum\nolimits_1 &= \sum_{k=3}^{\infty} \sum_{2^k \le |\mathbf{r}| < 2^{k+1}} \frac{1}{|\mathbf{r}|^2 \cdot \log_2|\mathbf{r}| \cdot (\log_2\log_2|\mathbf{r}|)^2} \\ &< \sum_{k=3}^{\infty} \pi 4^{k+1} \cdot \frac{1}{4^k \cdot k \cdot (\log_2 k)^2} \\ &= 4\pi \sum_{k=3}^{\infty} \frac{1}{k \cdot (\log_2 k)^2}. \end{aligned} \tag{6.15}$$

Note that in (6.15) we used the trivial fact that the number of lattice points in the annulus $2^k \le |\mathbf{r}| < 2^{k+1}$ is less than the area of the big circle $\pi \cdot 4^{k+1}$.

Returning to (6.15), we can estimate the infinite series with the corresponding definite integral:

$$\sum_{k=3}^{\infty} \frac{1}{k \cdot (\log_2 k)^2} < \int_2^{\infty} \frac{dx}{x(\log_2 x)^2} = \log 2,$$

and using this in (6.15), we have

$$\sum\nolimits_1 < 4\pi \log 2. \tag{6.16}$$

Similarly,

$$\sum\nolimits_2 = \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ |\mathbf{r}|\geq 8}} \frac{1}{|\mathbf{r}|^4\cdot(\log_2|\mathbf{r}|)^2\cdot(\log_2\log_2|\mathbf{r}|)^4}$$
$$= \sum_{k=3}^{\infty}\sum_{2^k\leq|\mathbf{r}|<2^{k+1}} \frac{1}{|\mathbf{r}|^4\cdot(\log_2|\mathbf{r}|)^2\cdot(\log_2\log_2|\mathbf{r}|)^4}$$
$$< \sum_{k=3}^{\infty}\pi 4^{k+1}\cdot\frac{1}{16^k\cdot k^2\cdot(\log_2 k)^4} < \frac{\pi}{100}. \tag{6.17}$$

We also need the simple numerical facts

$$\sum\nolimits_4 = \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ 1\leq|\mathbf{r}|<8}} \frac{1}{|\mathbf{r}|^4} < \sum\nolimits_3 = \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ 1\leq|\mathbf{r}|<8}} \frac{1}{|\mathbf{r}|^2} < 6\pi. \tag{6.18}$$

Combining (6.12)–(6.18), we have

$$\frac{\text{measure}(\mathcal{B})}{2\pi} < \frac{\varepsilon}{20\pi}\sum\nolimits_1 + \frac{\varepsilon^2}{800\pi}\sum\nolimits_2 + \frac{\varepsilon}{20\pi}\sum\nolimits_3 + \frac{\varepsilon^2}{800\pi}\sum\nolimits_4 < \frac{\varepsilon}{2},$$

completing the proof of Lemma 6.1. □

Let $\mathcal{A}$ denote the complement of $\mathcal{B}$, that is, $\mathcal{A}$ is the set of points (α_1,α_2) on the unit circle $\alpha_1^2+\alpha_2^2=1$ which are *not* ε-bad (see the Key Definition). We want to give an upper bound to the integral

$$\int_{\mathcal{A}}\left(\int_{I^2}(A(T;\mathbf{y},(\alpha_1,\alpha_2))-T\cdot\text{area}(A))^2\,d\mathbf{y}\right)ds, \tag{6.19}$$

where in the outer integral of (6.19) "ds" indicates integration with respect to the arc-length (since $\mathcal{A}$ is a "large" subset of the unit circle). We prove the following result.

Lemma 6.2 *We have*

$$\int_{\mathcal{A}}\left(\int_{I^2}(A(T;\mathbf{y},(\alpha_1,\alpha_2))-T\cdot\text{area}(A))^2\,d\mathbf{y}\right)ds$$
$$\leq \text{area}(A)(1-\text{area}(A))\cdot\frac{2688}{\pi^2}\cdot\frac{1}{\varepsilon}\log_2 T\cdot(\log_2\log_2 T)^2.$$

Proof By using (6.6)–(6.7), we have

$$\text{integral}(6.19)\leq\sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ \mathbf{r}\neq\mathbf{0}}}|a_{\mathbf{r}}|^2\cdot\int_{\mathcal{A}}\min\left\{\frac{1}{\pi^2(\alpha_1r_1+\alpha_2r_2)^2},T^2\right\}ds. \tag{6.20}$$

If $(\alpha_1,\alpha_2)\in\mathcal{A}$ then by definition (see (6.8)–(6.9))

$$|\alpha_1r_1+\alpha_2r_2|>\frac{\varepsilon}{40|\mathbf{r}|\cdot\log_2|\mathbf{r}|\cdot(\log_2\log_2|\mathbf{r}|)^2} \tag{6.21}$$

for all $|\mathbf{r}| \geq 8$ and

$$|\alpha_1 r_1 + \alpha_2 r_2| > \frac{\varepsilon}{40|\mathbf{r}|} \tag{6.22}$$

for all $1 \leq |\mathbf{r}| < 8$. Let $\mathbf{r} \in \mathbb{Z}^2$ with $|\mathbf{r}| = \sqrt{r_1^2 + r_2^2} \geq 8$ be arbitrary but fixed; to estimate the integral at the end of (6.20), we apply a standard power-of-two decomposition of the set

$$\mathcal{A}(\mathbf{r}) = \{(\alpha_1, \alpha_2) : \alpha_1^2 + \alpha_2^2 = 1,\ (6.21) \text{ holds}\} \supset \mathcal{A} \tag{6.23}$$

as follows: let ℓ be an arbitrary integer in the range

$$0 \leq \ell \leq L(\mathbf{r}) = \log_2 \left(\frac{40}{\varepsilon} |\mathbf{r}| \cdot \log_2 |\mathbf{r}| \cdot (\log_2 \log_2 |\mathbf{r}|)^2 \right), \tag{6.24}$$

and write

$$\mathcal{A}_\ell(\mathbf{r}) = \left\{ (\alpha_1, \alpha_2) : \alpha_1^2 + \alpha_2^2 = 1,\ 2^{-\ell-1} < |\alpha_1 r_1 + \alpha_2 r_2| \leq 2^{-\ell} \right\}. \tag{6.25}$$

Finally, write

$$\mathcal{A}_{-1}(\mathbf{r}) = \left\{ (\alpha_1, \alpha_2) : \alpha_1^2 + \alpha_2^2 = 1,\ |\alpha_1 r_1 + \alpha_2 r_2| > 1 \right\}, \tag{6.26}$$

and so we have the disjoint decomposition

$$\mathcal{A}(\mathbf{r}) = \bigcup_{-1 \leq \ell \leq L(\mathbf{r})} \mathcal{A}_\ell(\mathbf{r}) \supset \mathcal{A}. \tag{6.27}$$

For every $\ell \geq 0$ we have the estimation

$$\text{measure}(\mathcal{A}_\ell(\mathbf{r})) \leq 4 \arcsin\left(\frac{1}{|\mathbf{r}| 2^\ell} \right) \leq 4 \left(\frac{1}{|\mathbf{r}| \cdot 2^\ell} + \frac{1}{|\mathbf{r}|^2 \cdot 4^\ell} \right), \tag{6.28}$$

where (6.28) is just an easy adaptation of the argument at the beginning of the proof of Lemma 6.1.

Motivated by (6.20) and (6.27), we need to estimate the sum

$$\begin{aligned}
&\sum_{\substack{\mathbf{r} \in \mathbb{Z}^2: \\ |\mathbf{r}| \geq 8}} |a_\mathbf{r}|^2 \cdot \int_{\mathcal{A}(\mathbf{r})} \min\left\{ \frac{1}{\pi^2 (\alpha_1 r_1 + \alpha_2 r_2)^2}, T^2 \right\} ds \\
&\quad = \sum_{\substack{\mathbf{r} \in \mathbb{Z}^2: \\ |\mathbf{r}| \geq 8}} |a_\mathbf{r}|^2 \sum_{\ell=-1}^{L(\mathbf{r})} \int_{\mathcal{A}_\ell(\mathbf{r})} \min\left\{ \frac{1}{\pi^2 (\alpha_1 r_1 + \alpha_2 r_2)^2}, T^2 \right\} ds \\
&\quad \leq \sum_{\substack{\mathbf{r} \in \mathbb{Z}^2: \\ |\mathbf{r}| \geq 8}} |a_\mathbf{r}|^2 \sum_{\ell=-1}^{L(\mathbf{r})} \text{measure}(\mathcal{A}_\ell(\mathbf{r})) \cdot \min\left\{ \frac{4^{\ell+1}}{\pi^2}, T^2 \right\} \\
&\quad \leq \sum_{\substack{\mathbf{r} \in \mathbb{Z}^2: \\ |\mathbf{r}| \geq 8}} |a_\mathbf{r}|^2 \left(\sum_{\ell=0}^{L(\mathbf{r})} 4 \left(\frac{1}{|\mathbf{r}| \cdot 2^\ell} + \frac{1}{|\mathbf{r}|^2 \cdot 4^\ell} \right) \min\left\{ \frac{4^{\ell+1}}{\pi^2}, T^2 \right\} + 2\pi \cdot \frac{1}{\pi^2} \right),
\end{aligned} \tag{6.29}$$

where in the last step we used (6.28); the last term in (6.29) is a trivial bound for the special case $\ell = -1$ in the summation; and finally, $L(\mathbf{r})$ is defined in (6.24).

To estimate (6.29), we need some rather long but totally routine calculations. For any $\mathbf{r} \in \mathbb{Z}^2$ with $|\mathbf{r}| \geq 8$, we have (δ is a 0-1 valued indicator function to be defined below)

$$\begin{aligned}
&\sum_{\ell=0}^{L(\mathbf{r})} \left(\frac{1}{|\mathbf{r}| \cdot 2^\ell} + \frac{1}{|\mathbf{r}|^2 \cdot 4^\ell} \right) \min\left\{ \frac{4^{\ell+1}}{\pi^2}, T^2 \right\} \\
&\quad = \frac{1}{|\mathbf{r}|} \sum_{\ell=0}^{L(\mathbf{r})} \min\left\{ \frac{2^{\ell+2}}{\pi^2}, \frac{T^2}{2^\ell} \right\} + \frac{1}{|\mathbf{r}|^2} \sum_{\ell=0}^{L(\mathbf{r})} \min\left\{ \frac{4}{\pi^2}, \frac{T^2}{4^\ell} \right\} \\
&\quad \leq \frac{1}{|\mathbf{r}|} \sum_{\substack{0 \leq \ell \leq L(\mathbf{r}): \\ 2^{\ell+1} \leq \pi T}} \frac{2^{\ell+2}}{\pi^2} + \frac{T^2}{|\mathbf{r}|} \sum_{\substack{0 \leq \ell \leq L(\mathbf{r}): \\ 2^{\ell+1} > \pi T}} 2^{-\ell} + \frac{\log_2 T}{|\mathbf{r}|^2} \\
&\quad \leq \frac{1}{|\mathbf{r}|} \cdot \frac{8}{\pi^2} \min\left\{ 2^{L(\mathbf{r})}, \pi T/2 \right\} + \frac{T^2}{|\mathbf{r}|} \cdot 2\delta\left(2^{L(\mathbf{r})} \geq \pi T/2 \right) \cdot \frac{2}{\pi T} + \frac{\log_2 T}{|\mathbf{r}|^2},
\end{aligned} \tag{6.30}$$

where $\delta(2^{L(\mathbf{r})} \geq \pi T/2) = 1$ if $2^{L(\mathbf{r})} \geq \pi T/2$ and $\delta(2^{L(\mathbf{r})} \geq \pi T/2) = 0$ if $2^{L(\mathbf{r})} < \pi T/2$.

By (6.24), if $2^{L(\mathbf{r})} < \pi T/2$ then

$$\begin{aligned}
&\frac{1}{|\mathbf{r}|} \cdot \frac{8}{\pi^2} \min\left\{ 2^{L(\mathbf{r})}, \pi T/2 \right\} + \frac{T^2}{|\mathbf{r}|} \cdot 2\delta\left(2^{L(\mathbf{r})} \geq \pi T/2 \right) \cdot \frac{2}{\pi T} \\
&\quad = \frac{8}{\pi^2} \cdot \frac{40}{\varepsilon} \cdot \log_2 |\mathbf{r}| \cdot (\log_2 \log_2 |\mathbf{r}|)^2,
\end{aligned} \tag{6.31}$$

and if $2^{L(\mathbf{r})} \geq \pi T/2$ then

$$\begin{aligned}
&\frac{1}{|\mathbf{r}|} \cdot \frac{8}{\pi^2} \min\left\{ 2^{L(\mathbf{r})}, \pi T/2 \right\} + \frac{T^2}{|\mathbf{r}|} \cdot 2\delta\left(2^{L(\mathbf{r})} \geq \pi T/2 \right) \cdot \frac{2}{\pi T} \\
&\quad = \frac{4T}{\pi |\mathbf{r}|} + \frac{4T}{\pi |\mathbf{r}|} = \frac{8T}{\pi |\mathbf{r}|}.
\end{aligned} \tag{6.32}$$

If $2^{L(\mathbf{r})} < \pi T/2$ and $|\mathbf{r}| \geq 8$ then of course

$$\log_2 |\mathbf{r}| < L(\mathbf{r}) < \log_2(\pi T),$$

and so the last term in (6.31) can be estimated from above as follows:

$$\frac{8}{\pi^2} \cdot \frac{40}{\varepsilon} \cdot \log_2 |\mathbf{r}| \cdot (\log_2 \log_2 |\mathbf{r}|)^2 < \frac{8}{\pi^2} \cdot \frac{40}{\varepsilon} \log_2(\pi T) \cdot (\log_2 \log_2(\pi T))^2. \tag{6.33}$$

On the other hand, if we have the equality

$$\pi T/2 = 2^{L(\mathbf{r})} = \frac{40}{\varepsilon} |\mathbf{r}| \cdot \log_2 |\mathbf{r}| \cdot (\log_2 \log_2 |\mathbf{r}|)^2 \quad \text{and} \quad |\mathbf{r}| \geq 8, \tag{6.34}$$

then clearly

$$\frac{T}{|\mathbf{r}|} = \frac{2}{\pi} \cdot \frac{40}{\varepsilon} \cdot \log_2 |\mathbf{r}| \cdot (\log_2 \log_2 |\mathbf{r}|)^2 \leq \frac{2}{\pi} \cdot \frac{40}{\varepsilon} \cdot \log_2 T \cdot (\log_2 \log_2 T)^2, \tag{6.35}$$

and (6.35) remains true if we go beyond the equality (6.34) to the whole range $2^{L(\mathbf{r})} \geq \pi T/2$.

Summarizing, by (6.30)–(6.35) for any $\mathbf{r} \in \mathbb{Z}^2$ with $|\mathbf{r}| \geq 8$ we have

$$\sum_{\ell=0}^{L(\mathbf{r})} \left(\frac{1}{|\mathbf{r}| \cdot 2^\ell} + \frac{1}{|\mathbf{r}|^2 \cdot 4^\ell} \right) \min \left\{ \frac{4^{\ell+1}}{\pi^2}, T^2 \right\}$$
$$\leq \frac{16}{\pi^2} \cdot \frac{41}{\varepsilon} \log_2 T \cdot (\log_2 \log_2 T)^2. \tag{6.36}$$

Applying (6.36) in (6.29), we obtain

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ |\mathbf{r}|\geq 8}} |a_{\mathbf{r}}|^2 \cdot \int_{\mathcal{A}(\mathbf{r})} \min \left\{ \frac{1}{\pi^2(\alpha_1 r_1 + \alpha_2 r_2)^2}, T^2 \right\} ds$$
$$\leq \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ |\mathbf{r}|\geq 8}} |a_{\mathbf{r}}|^2 \cdot \frac{64}{\pi^2} \cdot \frac{42}{\varepsilon} \log_2 T \cdot (\log_2 \log_2 T)^2. \tag{6.37}$$

Similarly,

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ 1\leq|\mathbf{r}|<8}} |a_{\mathbf{r}}|^2 \cdot \int_{\mathcal{A}(\mathbf{r})} \min \left\{ \frac{1}{\pi^2(\alpha_1 r_1 + \alpha_2 r_2)^2}, T^2 \right\} ds$$
$$\leq \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ |\mathbf{r}|\geq 8}} |a_{\mathbf{r}}|^2 \cdot \frac{64}{\pi^2} \cdot \frac{42}{\varepsilon}. \tag{6.38}$$

Returning to (6.19)–(6.27), and using (6.37)–(6.38),

$$\frac{1}{2\pi} \int_{\mathcal{A}} \left(\int_{I^2} (A(T; \mathbf{y}, (\alpha_1, \alpha_2)) - T \cdot \mathrm{area}(A))^2 \, d\mathbf{y} \right) ds$$
$$\leq \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ |\mathbf{r}|\geq 8}} |a_{\mathbf{r}}|^2 \cdot \frac{1344}{\pi^3} \cdot \frac{1}{\varepsilon} \cdot \log_2 T \cdot (\log_2 \log_2 T)^2 + \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ 1\leq|\mathbf{r}|<8}} |a_{\mathbf{r}}|^2 \cdot \frac{1344}{\pi^3} \cdot \frac{1}{\varepsilon}$$
$$\leq \sum_{\substack{\mathbf{r}\in\mathbb{Z}^2:\\ \mathbf{r}\neq\mathbf{0}}} |a_{\mathbf{r}}|^2 \cdot \frac{1344}{\pi^3} \cdot \frac{1}{\varepsilon} \log_2 T \cdot (\log_2 \log_2 T)^2$$
$$= \mathrm{area}(A)(1 - \mathrm{area}(A)) \cdot \frac{1344}{\pi^3} \cdot \frac{1}{\varepsilon} \log_2 T \cdot (\log_2 \log_2 T)^2, \tag{6.39}$$

where in the last step we used (6.2). Equation (6.39) gives Lemma 6.2. □

Now we are ready to finish the proof of Theorem 4: we just throw out the "bad" initial conditions and apply Chebyshev's inequality. First a definition: for any $\lambda > 0$ let

$$\Omega(\lambda) = \left\{ (\mathbf{y}, (\alpha_1, \alpha_2)) \in [0, 1)^2 \times \mathcal{A} : |A(T; \mathbf{y}, (\alpha_1, \alpha_2)) - T \cdot \mathrm{area}(A)| \geq \lambda \right\}. \tag{6.40}$$

Combining (6.39)–(6.40) with Chebyshev's inequality,

$$\frac{1}{2\pi}\text{measure}(\Omega(\lambda)) \le \text{area}(A)(1-\text{area}(A)) \cdot \frac{1344}{\pi^3} \cdot \frac{1}{\varepsilon} \log_2 T \cdot (\log_2 \log_2 T)^2 \cdot \lambda^{-2}, \quad (6.41)$$

where "measure" stands for the 3-dimensional Lebesgue measure.

By making the choice

$$\lambda = \lambda_0 = \frac{10\sqrt{\text{area}(A)(1-\text{area}(A))}}{\varepsilon}\sqrt{\log_2 T} \cdot \log_2 \log_2 T \quad (6.42)$$

in (6.41), we conclude

$$\frac{1}{2\pi}\text{measure}(\Omega(\lambda_0)) \le \frac{\varepsilon}{2}. \quad (6.43)$$

If we throw out the set of initial conditions (starting point and angle) ($\mathbf{y}$, (α_1,α_2)) contained in $\Omega(\lambda_0)$, and also throw out those initial conditions ($\mathbf{y}$, (α_1,α_2)) for which the angle (α_1,α_2) is ε-bad (i.e., $(\alpha_1,\alpha_2)\in\mathcal{B}$), then by (6.43) and Lemma 6.1 the total loss is $\le \varepsilon/2+\varepsilon/2=\varepsilon$. Combining this fact with (6.42)–(6.43), Theorem 4 follows.

Now we return to the long proof of Theorem 1.

7 Proof of Theorem 1: The Simplest Simultaneous Case

Let j,k be arbitrary integers with $1\le j<k\le N$, and let

$$A_{j,k}(T) = A_{j,k}(T;\mathbf{y}_j,\mathbf{u}_j;\mathbf{y}_k,\mathbf{u}_k)$$

denote the total time between $0<t<T$ when the jth torus-line $\mathbf{x}_j(t)$ and the kth torus-line $\mathbf{x}_k(t)$ are both in subset A simultaneously; in other words, when the two torus lines are in A *at the same time*.

The key observation is that we can describe $A_{j,k}(T)$ in terms of the Cartesian product $A\times A\subset I^6=[0,1]^6$ of $A\subset I^3$ with itself. Indeed, we have

$$\begin{aligned} A_{j,k}(T) &= A_{j,k}(T;\mathbf{y}_j,\mathbf{u}_j;\mathbf{y}_k,\mathbf{u}_k) \\ &= \text{measure}\left\{t\in[0,T]:\mathbf{x}_j(t)\in A \pmod 1 \text{ and } \mathbf{x}_k(t)\in A \pmod 1\right\} \\ &= \int_0^T \chi_A(\mathbf{x}_j(t))\chi_A(\mathbf{x}_k(t))\,dt = \int_0^T \chi_{A\times A}(\mathbf{x}_j(t),\mathbf{x}_k(t))\,dt, \end{aligned} \quad (7.1)$$

where $\chi_{A\times A}$ is the 0-1 valued characteristic function of $A\times A\subset I^6$. Write $B=A\times A$; we need the Fourier series of the characteristic function $\chi_B=\chi_{A\times A}$:

$$\chi_B(\mathbf{w}) = \chi_{A\times A}(\mathbf{w}) = \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{w}} \quad \text{with } b_{\mathbf{r}} = \int_{A\times A} e^{-2\pi i\mathbf{r}\cdot\mathbf{z}}\,d\mathbf{z}, \quad (7.2)$$

where $\mathbf{r}\cdot\mathbf{w}=r_1w_1+\cdots+r_6w_6$ denotes the standard inner product. Clearly $b_{\mathbf{0}}=\text{vol}(A\times A)=\text{vol}^2(A)$ (= the volume of $A\times A$), and by Parseval's formula,

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\ne\mathbf{0}}} |b_{\mathbf{r}}|^2 = \text{vol}^2(A)-\text{vol}^4(A), \quad (7.3)$$

which is the analog of (5.3). Let's return to (7.1): by using the Fourier series (7.2), we have

$$\begin{aligned} A_{j,k}(T) &= A_{j,k}(T;\mathbf{y}_j,\mathbf{u}_j;\mathbf{y}_k,\mathbf{u}_k) \\ &= \int_0^T \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot(\mathbf{x}_j(t),\mathbf{x}_k(t))}\,dt \\ &= \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} \int_0^T e^{2\pi \mathrm{i}\mathbf{r}\cdot(\mathbf{x}_j(t),\mathbf{x}_k(t))}\,dt = \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot(\mathbf{y}_j,\mathbf{y}_k)} \int_0^T e^{2\pi \mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vt}\,dt \\ &= b_{\mathbf{0}}T + \sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}} b_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot(\mathbf{y}_j,\mathbf{y}_k)} \cdot \frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}-1}{2\pi \mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))v}. \end{aligned} \tag{7.4}$$

To clarify the notation here, note that (say) $(\mathbf{y}_j,\mathbf{y}_k)$ means a 6-dimensional vector for which the first 3 coordinates are given by $\mathbf{y}_j$ and the last 3 coordinates are given by $\mathbf{y}_k$.

Let M be an arbitrary integer in the range $1\le M\le N/2$, and consider the double sum

$$\begin{aligned} F_{1,2} &= F_{1,2}(\mathbf{y}_j,\mathbf{u}_j : 1\le j\le M;\mathbf{y}_k,\mathbf{u}_k : M+1\le k\le 2M) \\ &= \sum_{j=1}^{M}\sum_{k=M+1}^{2M}\left(\frac{1}{T}A_{j,k}(T)-\mathrm{vol}^2(A)\right) \\ &= \sum_{j=1}^{M}\sum_{k=M+1}^{2M}\frac{1}{T}\int_0^T \chi_A(\mathbf{x}_j(t))\chi_A(\mathbf{x}_k(t))\,dt - M^2\cdot\mathrm{vol}^2(A) \\ &= E_{1,2}-M^2\cdot\mathrm{vol}^2(A), \end{aligned} \tag{7.5}$$

where

$$E_{1,2}=\frac{1}{T}\int_0^T Z_1(t)Z_2(t)\,dt,$$

$$Z_1(t)=\sum_{j=1}^{M}\chi_A(\mathbf{x}_j(t)) \quad\text{and}\quad Z_2(t)=\sum_{k=M+1}^{2M}\chi_A(\mathbf{x}_k(t)).$$

By (7.4) we have

$$F_{1,2}=\sum_{j=1}^{M}\sum_{k=M+1}^{2M}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}} b_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot(\mathbf{y}_j,\mathbf{y}_k)} \cdot \frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}-1}{2\pi \mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))v}. \tag{7.6}$$

Fix the $2M$ unit vectors $\mathbf{u}_j\in S^2$, $j=1,2,\ldots,M$ and $\mathbf{u}_k\in S^2$, $k=M+1,M+2,\ldots,2M$, and evaluate the square integral

$$\sum\nolimits_{1,2}=\int_{I^3}\cdots\int_{I^3}\left(F_{1,2}(\mathbf{y}_j,\mathbf{u}_j : 1\le j\le M;\mathbf{y}_k,\mathbf{u}_k : M+1\le k\le 2M)\right)^2 d\mathbf{y}_1\ldots d\mathbf{y}_{2M}. \tag{7.7}$$

Note that (7.7) is a multiple integral, which consists of $2M$ single integrals.

To evaluate (7.7), we multiply out the square $F_{1,2}^2$ (where for $F_{1,2}$ we use (7.6)) and apply some orthogonality relations, leading to huge cancellations. To understand the cancellations, we study the following sub-problem: When does the multiple integral

$$\begin{aligned} &\mathit{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2) \\ &= \int_{I^3} \cdots \int_{I^3} e^{2\pi \mathrm{i}(\mathbf{r}_1 \cdot (\mathbf{y}_{j_1}, \mathbf{y}_{k_1}) - \mathbf{r}_2 \cdot (\mathbf{y}_{j_2}, \mathbf{y}_{k_2}))}\, d\mathbf{y}_{j_1} \ldots d\mathbf{y}_{k_2}, \end{aligned} \tag{7.8}$$

where $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{Z}^6 \setminus \mathbf{0}$ and $1 \le j_1, j_2 \le M < k_1, k_2 \le 2M$, equal to zero?

Well, the first challenge is that the integral (7.8) can be a double, or a triple, or a quadruple integral; it depends on whether the index set $\{j_1, j_2, k_1, k_2\}$ consists of 2 or 3 or 4 different integers. Accordingly, we distinguish three cases.

Case 1: quadruple integral: $j_1 \ne j_2$ and $k_1 \ne k_2$

Then clearly

$$\mathit{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2) = 0.$$

Case 2: triple integral: either $j_1 = j_2$ and $k_1 \ne k_2$, or $j_1 \ne j_2$ and $k_1 = k_2$

Then with $\mathbf{r}_1 = (\mathbf{r}_{1,1}, \mathbf{r}_{1,2})$ and $\mathbf{r}_2 = (\mathbf{r}_{2,1}, \mathbf{r}_{2,2})$,

$$\begin{aligned} &\mathit{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2) \\ &= \int_{I^3} \int_{I^3} \left(\int_{I^3} e^{2\pi \mathrm{i}(\mathbf{r}_{1,1} - \mathbf{r}_{2,1}) \cdot \mathbf{y}_{j_1}}\, d\mathbf{y}_{j_1} \right) e^{2\pi \mathrm{i}(\mathbf{r}_{1,2} \cdot \mathbf{y}_{k_1} - \mathbf{r}_{2,2} \cdot \mathbf{y}_{k_2})}\, d\mathbf{y}_{k_1}\, d\mathbf{y}_{k_2}, \end{aligned}$$

and this integral is always 0, unless $\mathbf{r}_{1,1} = \mathbf{r}_{2,1}$ and $\mathbf{r}_{1,2} = \mathbf{r}_{2,2} = (0,0,0)$, and then of course the integral is 1. Similar result holds for the other case $j_1 \ne j_2$ and $k_1 = k_2$. Therefore, Case 2 gives non-zero contribution (namely, one) if and only if $\mathbf{r}_1 = \mathbf{r}_2$ and either the first 3 coordinates are zero, or the last three coordinates are zero.

Case 3: double integral: $j_1 = j_2$ and $k_1 = k_2$

Then

$$\mathit{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2) = \int_{I^6} e^{2\pi \mathrm{i}(\mathbf{r}_1 - \mathbf{r}_2) \cdot \mathbf{y}}\, d\mathbf{y},$$

which is always 0, unless $\mathbf{r}_1 = \mathbf{r}_2$.

Now we are ready to evaluate $\sum_{1,2}$ (see (7.7)): squaring (7.6) and applying Cases 1–3 above, we obtain

$$\begin{aligned} \sum\nolimits_{1,2} &= \sum\nolimits_{1,2}(\mathbf{u}_j : 1 \le j \le M; \mathbf{u}_k : M+1 \le k \le 2M) \\ &= \sum_{\substack{\mathbf{r} \in \mathbb{Z}^6: \\ \mathbf{r} \ne \mathbf{0}}} |b_{\mathbf{r}}|^2 \cdot \sum_{j=1}^{M} \sum_{k=M+1}^{2M} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r} \cdot (\mathbf{u}_j, \mathbf{u}_k)) vT} - 1}{2\pi (\mathbf{r} \cdot (\mathbf{u}_j, \mathbf{u}_k)) vT} \right|^2 \\ &\quad + \sum_{\substack{\mathbf{r} = (\mathbf{r}_1, \mathbf{0}) \in \mathbb{Z}^6: \\ \mathbf{r}_1 \in \mathbb{Z}^3 \setminus \mathbf{0}}} |b_{\mathbf{r}}|^2 \cdot \sum_{j=1}^{M} \sum_{k_1=M+1}^{2M} \sum_{k_2=M+1}^{2M} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}_1 \cdot \mathbf{u}_j) vT} - 1}{2\pi (\mathbf{r}_1 \cdot \mathbf{u}_j) vT} \right|^2 \end{aligned}$$

$$+\sum_{\substack{\mathbf{r}=(\mathbf{0},\mathbf{r}_2)\in\mathbb{Z}^6:\\ \mathbf{r}_2\in\mathbb{Z}^3\setminus\mathbf{0}}} |b_{\mathbf{r}}|^2\cdot\sum_{j_1=1}^{M}\sum_{j_2=1}^{M}\sum_{k=M+1}^{2M}\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot\mathbf{u}_k)vT}-1}{2\pi(\mathbf{r}_2\cdot\mathbf{u}_k)vT}\right|^2. \tag{7.9}$$

Next we integrate $\sum_{1,2}$ over the $2M$ direction vectors $\mathbf{u}_j\in S^2$, $j=1,2,\dots,M$ and $\mathbf{u}_k\in S^2$, $k=M+1,M+2,\dots,2M$ (this is another multiple integral consisting of $2M$ single integrals):

$$\sum\nolimits_{1,2}^{*}=(4\pi)^{-2M}\int_{S^2}\dots\int_{S^2}\sum\nolimits_{1,2}(\mathbf{u}_j,\mathbf{u}_k:1\le j\le M, M+1\le k\le 2M)\,d\mathbf{u}_1\dots d\mathbf{u}_{2M}. \tag{7.10}$$

Let's return to the first sum in (7.9): we have the obvious upper bound

$$\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}-1}{2\pi(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}\right|\le\min\left\{\frac{1}{\pi|\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k)|vT},1\right\}. \tag{7.11}$$

We need to estimate the integral

$$(4\pi)^{-2}\int_{S^2}\int_{S^2}\min\left\{\frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k)vT)^2},1\right\}d\mathbf{u}_j\,d\mathbf{u}_k. \tag{7.12}$$

By (5.12)–(5.13), for any real numbers $c_1<c_2$ we have,

$$\mathrm{SurfaceArea}\left(\left\{\mathbf{u}\in S^2:c_1\le\mathbf{r}\cdot\mathbf{u}\le c_2\right\}\right)=4\pi\cdot\frac{\min\{c_2,r\}-\max\{c_1,-r\}}{r}, \tag{7.13}$$

where $r=|\mathbf{r}|$ and $\mathbf{r}\in\mathbb{Z}^3\setminus\mathbf{0}$.

Now let $\mathbf{r}\in\mathbb{Z}^6\setminus\mathbf{0}$, and write $\mathbf{r}=(\mathbf{r}_1,\mathbf{r}_2)$; then clearly $\mathbf{r}_1\in\mathbb{Z}^3\setminus\mathbf{0}$ or $\mathbf{r}_2\in\mathbb{Z}^3\setminus\mathbf{0}$. Suppose that (say) $\mathbf{r}_1\in\mathbb{Z}^3\setminus\mathbf{0}$; then we can estimate the integral in (7.12) as follows:

$$\begin{aligned}&(4\pi)^{-2}\int_{S^2}\int_{S^2}\min\left\{\frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k)vT)^2},1\right\}d\mathbf{u}_j\,d\mathbf{u}_k\\ &\quad=(4\pi)^{-2}\int_{S^2}\left(\int_{S^2}\min\left\{\frac{1}{(\pi(\mathbf{r}_1\cdot\mathbf{u}_j+\mathbf{r}_2\cdot\mathbf{u}_k)vT)^2},1\right\}d\mathbf{u}_j\right)d\mathbf{u}_k.\end{aligned} \tag{7.14}$$

For any fixed value of $\mathbf{u}_k$, the inner integral in (7.14) can be estimated from above by repeating the argument in (5.14)–(5.15) and using (7.13): with $c_0=\mathbf{r}_2\cdot\mathbf{u}_k$ we have

$$\int_{S^2}\min\left\{\frac{1}{(\pi(\mathbf{r}_1\cdot\mathbf{u}_j+c_0)vT)^2},1\right\}d\mathbf{u}_j\le\frac{2}{\pi vT|\mathbf{r}_1|}, \tag{7.15}$$

and using it in (7.14), we obtain

$$(4\pi)^{-2}\int_{S^2}\int_{S^2}\min\left\{\frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k)vT)^2},1\right\}d\mathbf{u}_j\,d\mathbf{u}_k\le\frac{2}{\pi vT|\mathbf{r}_1|}. \tag{7.16}$$

Let's return to (7.9)–(7.10). By using (7.11) and (7.16), we have

$$(4\pi)^{-2M}\int_{S^2}\dots\int_{S^2}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}}|b_{\mathbf{r}}|^2\cdot\sum_{j=1}^{M}\sum_{k=M+1}^{2M}\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}-1}{2\pi(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}\right|^2 d\mathbf{u}_1\dots d\mathbf{u}_{2M}$$

$$\le (4\pi)^{-2M}\int_{S^2}\cdots\int_{S^2}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}}|b_{\mathbf{r}}|^2\cdot\sum_{j=1}^{M}\sum_{k=M+1}^{2M}\min\left\{\frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k)vT)^2},1\right\}d\mathbf{u}_1\ldots d\mathbf{u}_{2M}$$

$$\le\frac{2}{\pi vT}\cdot M^2\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}}|b_{\mathbf{r}}|^2\le\frac{2}{\pi vT}\cdot M^2\cdot\mathrm{vol}^2(A),\tag{7.17}$$

where in the last step we used (7.3). This settles the contribution of first one of the three big sums on the right hand side of (7.9).

Next we deal with the second big sum on the right hand side of (7.9). By repeating the argument of (5.14)–(5.15), we have

$$(4\pi)^{-2M}\int_{S^2}\cdots\int_{S^2}\sum_{\substack{\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^3\setminus\mathbf{0}}}\sum_{j=1}^{M}\sum_{k_1=M+1}^{2M}\sum_{k_2=M+1}^{2M}$$

$$|b_{\mathbf{r}}|^2\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot\mathbf{u}_j)vT}-1}{2\pi(\mathbf{r}_1\cdot\mathbf{u}_j)vT}\right|^2d\mathbf{u}_1\ldots d\mathbf{u}_{2M}$$

$$\le(4\pi)^{-2M}\int_{S^2}\cdots\int_{S^2}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}}\sum_{j=1}^{M}\sum_{k=M+1}^{2M}$$

$$|b_{\mathbf{r}}|^2\cdot\min\left\{\frac{1}{(\pi(\mathbf{r}_1\cdot\mathbf{u}_j)vT)^2},1\right\}d\mathbf{u}_1\ldots d\mathbf{u}_{2M}$$

$$\le\frac{2}{\pi vT}\cdot M^3\sum_{\substack{\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^3\setminus\mathbf{0}}}|b_{\mathbf{r}}|^2.\tag{7.18}$$

A Key Technical Detail Note that for any $\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^6$ with $\mathbf{r}_1\in\mathbb{Z}^3\setminus\mathbf{0}$,

$$b_{\mathbf{r}}=\int_{A\times A}e^{-2\pi\mathrm{i}\mathbf{r}\cdot\mathbf{z}}\,d\mathbf{z}$$

$$=\mathrm{vol}(A)\int_A e^{-2\pi\mathrm{i}\mathbf{r}_1\cdot\mathbf{w}}\,d\mathbf{w}=\mathrm{vol}(A)\cdot a_{\mathbf{r}_1},\tag{7.19}$$

where $a_{\mathbf{r}_1}$ is defined in (5.2). Therefore, by (5.3),

$$\sum_{\substack{\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^3\setminus\mathbf{0}}}|b_{\mathbf{r}}|^2=\mathrm{vol}^2(A)\sum_{\mathbf{r}_1\in\mathbb{Z}^3\setminus\mathbf{0}}|a_{\mathbf{r}_1}|^2\le\mathrm{vol}^3(A).\tag{7.20}$$

Since $\mathrm{vol}(A)$ is very small, it is absolutely crucial that in (7.20) we got the cube of the volume (instead of the trivial upper bound $\mathrm{vol}(A\times A)=\mathrm{vol}^2(A)$).

Of course, the same argument works for the third big sum on the right hand side of (7.9).

Summarizing, by (7.9)–(7.10) and (7.17)–(7.20), we obtain the following analog of Lemma 5.1.

Lemma 7.1 *We have*

$$\sum\nolimits_{1,2}^{*} \leq \frac{2}{\pi v T} \cdot \left(M^2 \cdot \mathrm{vol}^2(A) + 2M^3 \cdot \mathrm{vol}^3(A)\right), \tag{7.21}$$

where the square-integral $\sum_{1,2}^{*}$ (*see* (7.5), (7.7), (7.10)) *equals the multiple integral*

$$(4\pi)^{-2M} \int_{I^3} \cdots \int_{I^3} \int_{S^2} \cdots \int_{S^2}$$
$$\left(\sum_{j=1}^{M} \sum_{k=M+1}^{2M} \left(\frac{1}{T} A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k) - \mathrm{vol}^2(A) \right) \right)^2 d\mathbf{y}_1 \ldots d\mathbf{y}_{2M}\, d\mathbf{u}_1 \ldots d\mathbf{u}_{2M}, \tag{7.22}$$

which consists of $4M$ *single integrals.*

Again we interrupt the proof of Theorem 1.

8 Proving Theorems 2 and 3

8.1 Proof of Theorem 2

We follow the notation introduced in Sects. 1–7. Again we use the geometric trick of *unfolding* the billiard paths to straight lines in the 3-space, so it suffices to deal with N torus lines $\mathbf{x}_k(t) = (x_{k,1}(t), x_{k,2}(t), x_{k,3}(t))$ (mod 1), $k = 1, 2, \ldots, N$ where

$$x_{k,1}(t) = u_{k,1} t v + y_{k,1}, \qquad x_{k,2}(t) = u_{k,2} t v + y_{k,2}, \qquad x_{k,3}(t) = u_{k,3} t v + y_{k,3}$$

and

$$u_{k,1}^2 + u_{k,2}^2 + u_{k,3}^2 = 1,$$

i.e., $\mathbf{u}_k = (u_{k,1}, u_{k,2}, u_{k,3})$ is a unit vector ($v \geq 1$ is the common speed of the particles).

Let $A \subset I^3 = [0, 1)^3$ be an arbitrary Lebesgue measurable subset (via unfolding it corresponds to the union of 8 copies of the given subset in Theorem 2; we shrink the corresponding $2 \times 2 \times 2$ cube to the unit cube).

To prove Theorem 2 we use the second moment method: we want to give a "good" upper bound for the square integral

$$(4\pi)^{-N} \int_{\Omega} \left(\int_0^T (Y_A(\omega; t) - N \cdot \mathrm{vol}(A))^2\, dt \right)^2 d\omega.$$

Let

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \left(I^3\right)^N \times \left(S^2\right)^N = \Omega$$

be an arbitrary initial condition, and we start with the equality

$$\int_0^T (Y_A(\omega; t) - N \cdot \mathrm{vol}(A))^2\, dt$$

$$= \int_0^T \left(\sum_{k=1}^N (\chi_A(\mathbf{x}_k(\omega; t)) - \mathrm{vol}(A)) \right)^2 dt$$

$$= 2\sum\nolimits^{(1)}(\omega; T) - 2(N-1) \cdot \mathrm{vol}(A) \cdot \sum\nolimits^{(2)}(\omega; T) + \sum\nolimits^{(3)}(\omega; T), \tag{8.1}$$

where

$$\sum\nolimits^{(1)}(\omega; T) = \int_0^T \sum_{1 \le j < k \le N} \left(\chi_A(\mathbf{x}_j(\omega; t)) \chi_A(\mathbf{x}_k(\omega; t)) - \mathrm{vol}^2(A) \right) dt, \tag{8.2}$$

$$\sum\nolimits^{(2)}(\omega; T) = \int_0^T \sum_{k=1}^N (\chi_A(\mathbf{x}_k(\omega; t)) - \mathrm{vol}(A))\, dt, \tag{8.3}$$

and

$$\sum\nolimits^{(3)}(\omega; T) = \int_0^T \sum_{k=1}^N (\chi_A(\mathbf{x}_k(\omega; t)) - \mathrm{vol}(A))^2\, dt. \tag{8.4}$$

For $\sum^{(2)}(\omega; T)$ we apply Lemma 5.1 with $M = N$: since

$$\sum\nolimits^{(2)}(\omega; T) = \sum_{k=1}^N (A_k(T; \mathbf{y}_k, \mathbf{u}_k) - T \cdot \mathrm{vol}(A)),$$

by Lemma 5.1 we have

$$(4\pi)^{-N} \int_\Omega \left(\sum\nolimits^{(2)}(\omega; T) \right)^2 d\omega \le \frac{TN \cdot \mathrm{vol}(A)}{v}. \tag{8.5}$$

For $\sum^{(1)}(\omega; T)$ we start with the obvious equality

$$\sum\nolimits^{(1)}(\omega; T) = \sum_{1 \le j < k \le N} \left(A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k) - T \cdot \mathrm{vol}^2(A) \right),$$

which leads us to Lemma 7.1. Unfortunately, we cannot directly apply Lemma 7.1. Instead we *repeat the whole proof:* thus we obtain the following result, which is just a slightly modified version of Lemma 7.1:

$$(4\pi)^{-N} \int_\Omega \left(\sum\nolimits^{(1)}(\omega; T) \right)^2 d\omega \le \frac{T}{v} \left(N^2 \cdot \mathrm{vol}^2(A) + N^3 \cdot \mathrm{vol}^3(A) \right). \tag{8.6}$$

We also need the following well-known inequality: for any real numbers C_1, C_2, C_3,

$$(C_1 + C_2 + C_3)^2 \le 3 \left(C_1^2 + C_2^2 + C_3^2 \right). \tag{8.7}$$

Combining (8.1)–(8.7), we have

$$(4\pi)^{-N} \int_\Omega \left(\int_0^T (Y_A(\omega; t) - N \cdot \mathrm{vol}(A))^2\, dt \right)^2 d\omega$$

$$\leq 3 \cdot (4\pi)^{-N} \int_\Omega \left(\sum^{(3)} (\omega; T) \right)^2 d\omega$$
$$+ 3 \left(4N^2 \cdot \mathrm{vol}^2(A) \frac{TN \cdot \mathrm{vol}(A)}{v} + 4\frac{T}{v} \left(N^2 \cdot \mathrm{vol}^2(A) + N^3 \cdot \mathrm{vol}^3(A) \right) \right). \tag{8.8}$$

To estimate the term

$$(4\pi)^{-N} \int_\Omega \left(\sum^{(3)} (\omega; T) \right)^2 d\omega$$

in (8.8), we first go back to (8.4), and apply the trivial inequality

$$(\chi - V)^2 \leq 2\chi^2 + 2V^2 \leq 2\chi + V^2 \quad \text{if } \chi = 1, 0$$

twice:

$$\left(\sum^{(3)} (\omega; T) \right)^2$$
$$\leq \left(2 \sum_{k=1}^N A_k(T; \omega) + 2NT \cdot \mathrm{vol}^2(A) \right)^2$$
$$= 4 \left(\left(\sum_{k=1}^N A_k(T; \omega) - NT \cdot \mathrm{vol}(A) \right) + NT \cdot (\mathrm{vol}(A) + \mathrm{vol}^2(A)) \right)^2$$
$$\leq 8 \left(\sum_{k=1}^N A_k(T; \omega) - NT \cdot \mathrm{vol}(A) \right)^2 + 8N^2T^2 \cdot (\mathrm{vol}(A) + \mathrm{vol}^2(A))^2.$$

Thus we have

$$(4\pi)^{-N} \int_\Omega \left(\sum^{(3)} (\omega; T) \right)^2 d\omega$$
$$\leq 8 \cdot (4\pi)^{-N} \int_\Omega \left(\sum_{k=1}^N A_k(T; \omega) - NT \cdot \mathrm{vol}(A) \right)^2 d\omega$$
$$+ 8N^2T^2 \cdot (\mathrm{vol}(A) + \mathrm{vol}^2(A))^2$$
$$\leq \frac{8NT \cdot \mathrm{vol}(A)}{v} + 8N^2T^2 \cdot (\mathrm{vol}(A) + \mathrm{vol}^2(A))^2, \tag{8.9}$$

where in the last line of (8.9) we applied (8.5) (i.e., Lemma 5.1).

Using (8.9) in (8.8), we have

$$(4\pi)^{-N} \int_\Omega \left(\int_0^T (Y_A(\omega; t) - N \cdot \mathrm{vol}(A))^2 \, dt \right)^2 d\omega$$
$$\leq 3 \cdot \left(\frac{8NT \cdot \mathrm{vol}(A)}{v} + 8N^2T^2 \cdot (\mathrm{vol}(A) + \mathrm{vol}^2(A))^2 \right)$$
$$+ 3 \left(4N^2 \cdot \mathrm{vol}^2(A) \frac{TN \cdot \mathrm{vol}(A)}{v} + 4\frac{T}{v} \left(N^2 \cdot \mathrm{vol}^2(A) + N^3 \cdot \mathrm{vol}^3(A) \right) \right)$$

$$\leq 12T^2N^2 \cdot \mathrm{vol}^2(A)\left(\frac{2}{vTN \cdot \mathrm{vol}(A)} + 8 + \frac{1+2N \cdot \mathrm{vol}(A)}{vT}\right). \tag{8.10}$$

Let $0 < \varepsilon < 1$ be arbitrary. It follows from (8.10) that there is a (measurable) subset $\Omega(\varepsilon; T; bad)$ of Ω such that,

$$\mathrm{measure}(\Omega(\varepsilon; T; bad)) < \varepsilon \cdot \mathrm{measure}(\Omega), \tag{8.11}$$

and for all

$$\omega \in \Omega \setminus \Omega(\varepsilon; T; bad), \tag{8.12}$$

$$\int_0^T (Y_A(\omega; t) - N \cdot \mathrm{vol}(A))^2 \, dt$$
$$\leq 2\sqrt{\frac{3}{\varepsilon}} \cdot TN \cdot \mathrm{vol}(A)\left(\frac{2}{vTN \cdot \mathrm{vol}(A)} + 8 + \frac{1+2N \cdot \mathrm{vol}(A)}{vT}\right)^{1/2}. \tag{8.13}$$

Similarly, by (8.13) we have for any $\eta > 0$,

$$\frac{1}{T}\mathrm{measure}\{0 \leq t \leq T : |Y_A(\omega; t) - N \cdot \mathrm{vol}(A)| > \eta N \cdot \mathrm{vol}(A)\}$$
$$< \frac{2}{\eta^2 N \cdot \mathrm{vol}(A)}\sqrt{\frac{3}{\varepsilon}}\left(\frac{2}{vTN \cdot \mathrm{vol}(A)} + 8 + \frac{1+2N \cdot \mathrm{vol}(A)}{vT}\right)^{1/2}. \tag{8.14}$$

Finally, note that (8.14) completes the proof of Theorem 2.

8.2 Proof of Theorem 3

Similarly to Theorem 2, our approach is based on the second moment method. For convenience I introduce the notation $\tilde{I} = [0, 1/2]$, so $[0, 1/2]^3 = \tilde{I}^3$. Our basic assumption is that the initial conditions

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \tag{8.15}$$

are uniformly distributed in the product set

$$\left([0, 1/2]^3\right)^N \times \left(S^2\right)^N = \left(\tilde{I}^3\right)^N \times \left(S^2\right)^N = \widetilde{\Omega}, \tag{8.16}$$

which is 8^{-N} part of the whole space Ω. The second moment method means that we want to give a "good" upper bound for the square integral

$$\left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\int_0^T (Y_A(\omega; t) - N \cdot \mathrm{vol}(A))^2 \, dt\right)^2 d\omega, \tag{8.17}$$

where $2/\pi$ comes from the ratio of 8 and 4π (here 4π is the surface area of the unit sphere S^2 and $1/8$ is the volume of the octant $[0, 1/2]^3 = \tilde{I}^3$).

Again we use (8.1)–(8.4): let

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \widetilde{\Omega} \tag{8.18}$$

be an arbitrary "jammed" initial condition, and we use the equality

$$\int_0^T (Y_A(\omega;t) - N \cdot \mathrm{vol}(A))^2\,dt$$
$$= \int_0^T \left(\sum_{k=1}^{N} (\chi_A(\mathbf{x}_k(\omega;t)) - \mathrm{vol}(A))\right)^2 dt$$
$$= 2\sum\nolimits^{(1)}(\omega;T) - 2(N-1)\cdot \mathrm{vol}(A)\cdot \sum\nolimits^{(2)}(\omega;T) + \sum\nolimits^{(3)}(\omega;T), \tag{8.19}$$

where

$$\sum\nolimits^{(1)}(\omega;T) = \int_0^T \sum_{1\le j<k\le N} \big(\chi_A(\mathbf{x}_j(\omega;t))\chi_A(\mathbf{x}_k(\omega;t)) - \mathrm{vol}^2(A)\big)\,dt, \tag{8.20}$$

$$\sum\nolimits^{(2)}(\omega;T) = \int_0^T \sum_{k=1}^{N} (\chi_A(\mathbf{x}_k(\omega;t)) - \mathrm{vol}(A))\,dt, \tag{8.21}$$

and

$$\sum\nolimits^{(3)}(\omega;T) = \int_0^T \sum_{k=1}^{N} (\chi_A(\mathbf{x}_k(\omega;t)) - \mathrm{vol}(A))^2\,dt. \tag{8.22}$$

For $\sum^{(3)}(\omega;T)$ it suffices to use the trivial bound

$$0 \le \sum\nolimits^{(3)}(\omega;T) \le NT, \tag{8.23}$$

which holds for all $\omega \in \Omega$ and all $T > 0$.

Since our subspace $\widetilde{\Omega}$ is just a tiny part of Ω (namely, $\widetilde{\Omega}$ is 8^{-N} part of Ω), we cannot directly apply Lemma 5.1 or Lemma 7.1 (i.e., we cannot simply repeat the proof of Theorem 2). Instead, we are going to develop a corresponding analog for both Lemma 5.1 and Lemma 7.1. The first step is to prove an

Analog of Lemma 5.1 I begin with recalling (5.2)–(5.6). As usual, we apply the geometric trick of *unfolding* the billiard paths to straight lines in the 3-space, and so it suffices to deal with N torus lines $\mathbf{x}_k(t) = (x_{k,1}(t), x_{k,2}(t), x_{k,3}(t))$ (mod 1), $k = 1, 2, \dots, N$ where

$$x_{k,1}(t) = u_{k,1}tv + y_{k,1}, \qquad x_{k,2}(t) = u_{k,2}tv + y_{k,2}, \qquad x_{k,3}(t) = u_{k,3}tv + y_{k,3}$$

and

$$u_{k,1}^2 + u_{k,2}^2 + u_{k,3}^2 = 1,$$

i.e., $\mathbf{u}_k = (u_{k,1}, u_{k,2}, u_{k,3})$ is a unit vector ($v \ge 1$ is the common speed).

Let $A \subset I^3 = [0,1)^3$ be an arbitrary Lebesgue measurable subset (via unfolding it corresponds to the union of 8 copies of the given subset; we shrink the corresponding $2\times 2\times 2$ cube to the unit cube). We work with the Fourier series of the 0-1-valued characteristic function χ_A of $A \subset I^3 = [0,1)^3$:

$$\chi_A(\mathbf{w}) = \sum_{\mathbf{r}\in\mathbb{Z}^3} a_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{w}} \quad \text{with } a_{\mathbf{r}} = \int_A e^{-2\pi i \mathbf{r}\cdot\mathbf{z}}\,d\mathbf{z}. \tag{8.24}$$

Clearly $a_{\mathbf{0}} = \text{vol}(A)$, and by Parseval's formula,

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}} |a_{\mathbf{r}}|^2 = \int_{I^3} \chi_A^2(\mathbf{w})\, d\mathbf{w} - |a_{\mathbf{0}}|^2 = \text{vol}(A) - \text{vol}^2(A). \tag{8.25}$$

Given a real number $T \geq 1$, we denote the total time that the kth torus-line $\mathbf{x}_k(t)$ (representing the kth point-billiard) spends in subset A during $0 < t < T$ by $A_k(T) = A_k(T; \mathbf{y}_k, \mathbf{u}_k)$: we have

$$\begin{aligned} A_k(T) = A_k(T; \mathbf{y}_k, \mathbf{u}_k) &= \text{measure}\{t \in [0, T] : \mathbf{x}_k(t) \in A \ (\text{mod } 1)\} \\ &= \int_0^T \chi_A(\mathbf{x}_k(t))\, dt = \int_0^T \sum_{\mathbf{r}\in\mathbb{Z}^3} a_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{x}_k(t)}\, dt \\ &= a_{\mathbf{0}} T + \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}} a_{\mathbf{r}} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{y}_k} \cdot \frac{e^{2\pi\mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vT} - 1}{2\pi\mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)v}. \end{aligned} \tag{8.26}$$

Let

$$F_N = F_N(\mathbf{y}_k, \mathbf{u}_k : 1 \leq k \leq N) = \sum_{k=1}^N \left(\frac{1}{T} A_k(T; \mathbf{y}_k, \mathbf{u}_k) - \text{vol}(A)\right). \tag{8.27}$$

Fix the N unit vectors $\mathbf{u}_k \in S^2$, $1 \leq k \leq N$, and evaluate the square integral

$$\widetilde{\sum}_1(\mathbf{u}_k : 1 \leq k \leq N) = \int_{\tilde{I}^3} \cdots \int_{\tilde{I}^3} (F_N(\mathbf{y}_k, \mathbf{u}_k : 1 \leq k \leq N))^2\, d\mathbf{y}_1 \ldots d\mathbf{y}_N. \tag{8.28}$$

Note that (8.28) is a multiple integral, which consists of N single integrals.

To evaluate (8.28), we multiply out the square F_N^2, where the sum F_N is defined in (8.26)–(8.27). The orthogonality relations (5.7)–(5.8) fail for the half $\tilde{I} = [0, 1/2]$ of the unit interval. Instead we use the following elementary fact: for $r \in \mathbb{Z} \setminus 0$ we have

$$\frac{1}{|\tilde{I}|} \int_{\tilde{I}} e^{2\pi\mathrm{i}ry}\, dy = 2\int_0^{1/2} e^{2\pi\mathrm{i}ry}\, dy = \frac{2\mathrm{i}}{\pi r} \text{ or } 0, \tag{8.29}$$

depending on whether r is odd or even. Now let $\mathbf{r} = (r_1, r_2, r_3) \in \mathbb{Z}^3 \setminus \mathbf{0}$, then

$$\frac{1}{\text{vol}(\widetilde{I}^3)} \int_{\tilde{I}^3} e^{2\pi\mathrm{i}\mathbf{r}\cdot\mathbf{y}}\, d\mathbf{y} = \prod_{j=1}^3 \left(\frac{2\mathrm{i}}{\pi r_j} \text{ or } 0 \text{ or } 1\right), \tag{8.30}$$

where the three cases indicated in (8.30) depend on whether $r_j \neq 0$ is odd or $r_j \neq 0$ is even or $r_j = 0$ ($j = 1, 2, 3$).

It is convenient to introduce the following new notation:

$$\Theta(\mathbf{r}) = \prod_{j=1}^d \frac{1}{\max\{|r_j|, 1\}} \tag{8.31}$$

for any real vector $\mathbf{r} = (r_1, \ldots, r_d)$ of any dimension $d \geq 1$.

By (8.30)–(8.31) we have

$$\frac{1}{\mathrm{vol}(\widetilde{I}^3)}\left|\int_{\widetilde{I}^3} e^{2\pi \mathrm{i}\mathbf{r}\cdot\mathbf{y}}\,d\mathbf{y}\right| \le \Theta(\mathbf{r}) \tag{8.32}$$

for all $\mathbf{r}\in\mathbb{Z}^3$. We consider (8.32) an "analog" of (5.8).

The following corollary of (8.32) is considered a similar "analog" of (5.7):

$$\frac{1}{\mathrm{vol}^2(\widetilde{I}^3)}\left|\int_{\widetilde{I}^3}\int_{\widetilde{I}^3} e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot\mathbf{y}_j-\mathbf{r}_2\cdot\mathbf{y}_k)}\,d\mathbf{y}_j\,d\mathbf{y}_k\right| \le \Theta(\mathbf{r}_1)\Theta(\mathbf{r}_2), \tag{8.33}$$

which holds for any $\mathbf{r}_1,\mathbf{r}_2\in\mathbb{Z}^3\setminus\mathbf{0}$ and $j\ne k$.

By using (8.32)–(8.33) instead of (5.7)–(5.8), we obtain the following (much longer!) "analog" of (5.9):

$$\begin{aligned}\widetilde{\sum}_1 &= \widetilde{\sum}_1(\mathbf{u}_k : 1\le k\le N)\\ &\le \widetilde{\sum}_1(1)+\widetilde{\sum}_1(2)+\widetilde{\sum}_1(3),\end{aligned} \tag{8.34}$$

where

$$\begin{aligned}\widetilde{\sum}_1(1) = &\sum_{\substack{\mathbf{r}_1\in\mathbb{Z}^3:\\ \mathbf{r}_1\ne\mathbf{0}}}\sum_{\substack{\mathbf{r}_2\in\mathbb{Z}^3:\\ \mathbf{r}_2\ne\mathbf{0}}}\sum_{j=1}^{N}\sum_{\substack{k=1:\\ k\ne j}}^{N}|a_{\mathbf{r}_1}|\\ &\cdot\Theta(\mathbf{r}_1)\cdot|a_{\mathbf{r}_2}|\cdot\Theta(\mathbf{r}_2)\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot\mathbf{u}_j)vT}-1}{2\pi(\mathbf{r}_1\cdot\mathbf{u}_j)vT}\right|\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot\mathbf{u}_k)vT}-1}{2\pi(\mathbf{r}_2\cdot\mathbf{u}_k)vT}\right|,\end{aligned} \tag{8.35}$$

$$\begin{aligned}\widetilde{\sum}_1(2) = &\sum_{\substack{\mathbf{r}_1\in\mathbb{Z}^3:\\ \mathbf{r}_1\ne\mathbf{0}}}\sum_{\substack{\mathbf{r}_2\in\mathbb{Z}^3:\\ \mathbf{r}_2\notin\{\mathbf{0},\mathbf{r}_1\}}}\sum_{k=1}^{N}|a_{\mathbf{r}_1}|\cdot|a_{\mathbf{r}_2}|\\ &\cdot\Theta(\mathbf{r}_1-\mathbf{r}_2)\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot\mathbf{u}_k)vT}-1}{2\pi(\mathbf{r}_1\cdot\mathbf{u}_k)vT}\right|\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot\mathbf{u}_k)vT}-1}{2\pi(\mathbf{r}_2\cdot\mathbf{u}_k)vT}\right|,\end{aligned} \tag{8.36}$$

and finally, $\widetilde{\sum}_1(3)$ is (5.9):

$$\widetilde{\sum}_1(3) = \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\ne\mathbf{0}}}|a_{\mathbf{r}}|^2\cdot\sum_{k=1}^{N}\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}\cdot\mathbf{u}_k)vT}-1}{2\pi(\mathbf{r}\cdot\mathbf{u}_k)vT}\right|^2. \tag{8.37}$$

Next we integrate $\widetilde{\sum}_1$ over the direction vectors $\mathbf{u}_k\in S^2$, $1\le k\le N$, which leads to another multiple integral consisting of N single integrals (to normalize, we have to divide by $4\pi =$ the surface area of the unit sphere S^2):

$$\begin{aligned}\widetilde{\sum}_1^{*} &= \frac{1}{4\pi}\int_{S^2}\cdots\frac{1}{4\pi}\int_{S^2}\widetilde{\sum}_1(\mathbf{u}_k : 1\le k\le N)\,d\mathbf{u}_1\ldots d\mathbf{u}_N\\ &\le (4\pi)^{-N}\int_{S^2}\cdots\int_{S^2}\left(\widetilde{\sum}_1(1)+\widetilde{\sum}_1(2)+\widetilde{\sum}_1(3)\right)d\mathbf{u}_1\ldots d\mathbf{u}_N\end{aligned}$$

$$= \widetilde{\sum}_1^*(1) + \widetilde{\sum}_1^*(2) + \widetilde{\sum}_1^*(3), \tag{8.38}$$

where for $i = 1, 2, 3$,

$$\widetilde{\sum}_1^*(i) = (4\pi)^{-N} \int_{S^2} \cdots \int_{S^2} \widetilde{\sum}_1(i)\, d\mathbf{u}_1 \ldots d\mathbf{u}_N.$$

Again we use the trivial inequality (5.11):

$$\left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot\mathbf{u})vT} - 1}{2\pi(\mathbf{r}\cdot\mathbf{u})vT} \right| \le \min\left\{ \frac{1}{\pi|\mathbf{r}\cdot\mathbf{u}|vT}, 1 \right\}. \tag{8.39}$$

Repeating the arguments in (5.12)–(5.15), we have with $\delta(\mathbf{r}) = (\pi vT|\mathbf{r}|)^{-1}$,

$$\begin{aligned}
&\frac{1}{4\pi}\int_{S^2} \min\left\{ \frac{1}{\pi|\mathbf{r}\cdot\mathbf{u}|vT}, 1 \right\} d\mathbf{u} \\
&\quad = \delta(\mathbf{r}) + \delta(\mathbf{r}) \int_{\delta(\mathbf{r})}^1 \frac{dx}{x} = \delta(\mathbf{r}) + \delta(\mathbf{r}) \log\left(\frac{1}{\delta(\mathbf{r})}\right) \\
&\quad = \frac{1 + \log(\pi vT|\mathbf{r}|)}{\pi vT|\mathbf{r}|}.
\end{aligned} \tag{8.40}$$

By (8.35), (8.38)–(8.40) we have

$$\begin{aligned}
\widetilde{\sum}_1^*(1) &\le N^2 \left(\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\ne\mathbf{0}}} |a_{\mathbf{r}}| \cdot \Theta(\mathbf{r}) \cdot \frac{1 + \log(\pi vT|\mathbf{r}|)}{\pi vT|\mathbf{r}|} \right)^2 \\
&\le N^2 \cdot \left(\frac{\log(vT)}{vT} \right)^2 \cdot \left(\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\ne\mathbf{0}}} |a_{\mathbf{r}}|^2 \right) \\
&\quad \cdot \left(\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\ne\mathbf{0}}} \Theta^2(\mathbf{r}) \cdot \left(\frac{1 + \log|\mathbf{r}|}{|\mathbf{r}|} \right)^2 \right),
\end{aligned} \tag{8.41}$$

where in the last step we used the Cauchy–Schwarz inequality. By (8.25) (Parseval's formula):

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\ne\mathbf{0}}} |a_{\mathbf{r}}|^2 \le \mathrm{vol}(A). \tag{8.42}$$

Furthermore, by using the definition of $\Theta(\mathbf{r})$ (see (8.31)), we have the upper bound

$$\begin{aligned}
&\sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\ne\mathbf{0}}} \Theta^2(\mathbf{r}) \cdot \left(\frac{1 + \log|\mathbf{r}|}{|\mathbf{r}|} \right)^2 \\
&\quad \le \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\ne\mathbf{0}}} \Theta^2(\mathbf{r})
\end{aligned}$$

$$\le \left(\sum_{r \in \mathbb{Z}} \frac{1}{\max\{r^2, 1\}} \right)^3 = \left(1 + 2 \sum_{r=1}^{\infty} \frac{1}{r^2} \right)^3$$

$$= \left(1 + 2 \cdot \frac{\pi^2}{6} \right)^3 < 100. \tag{8.43}$$

Combining (8.41)–(8.43), we obtain

$$\widetilde{\sum}_1^*(1) \le 100 N^2 \cdot \mathrm{vol}(A) \cdot \left(\frac{\log(vT)}{vT} \right)^2. \tag{8.44}$$

Next we study $\widetilde{\sum}_1^*(2)$. We need the following upper bound based on the Cauchy–Schwarz inequality for integrals of functions (we assume $\mathbf{r}_1 \ne \mathbf{0} \ne \mathbf{r}_2$):

$$\frac{1}{4\pi} \int_{S^2} \min\left\{ \frac{1}{\pi |\mathbf{r}_1 \cdot \mathbf{u}| vT}, 1 \right\} \cdot \min\left\{ \frac{1}{\pi |\mathbf{r}_2 \cdot \mathbf{u}| vT}, 1 \right\} d\mathbf{u}$$

$$\le \left(\frac{1}{4\pi} \int_{S^2} \min\left\{ \frac{1}{(\pi |\mathbf{r}_1 \cdot \mathbf{u}| vT)^2}, 1 \right\} d\mathbf{u} \right)^{1/2} \cdot \left(\frac{1}{4\pi} \int_{S^2} \min\left\{ \frac{1}{(\pi |\mathbf{r}_2 \cdot \mathbf{u}| vT)^2}, 1 \right\} d\mathbf{u} \right)^{1/2}$$

$$\le \left(\frac{2}{\pi vT} \right) \cdot (|\mathbf{r}_1| \cdot |\mathbf{r}_2|)^{-1/2}, \tag{8.45}$$

where in the last step we used (5.15).

By (8.36), (8.38)–(8.39) and (8.45) we have

$$\widetilde{\sum}_1^*(2) \le N \cdot \left(\frac{2}{\pi vT} \right) \cdot \sum_{\substack{\mathbf{r}_1 \in \mathbb{Z}^3: \\ \mathbf{r}_1 \ne \mathbf{0}}} \sum_{\substack{\mathbf{r}_2 \in \mathbb{Z}^3: \\ \mathbf{r}_2 \notin \{\mathbf{0}, \mathbf{r}_1\}}} |a_{\mathbf{r}_1}| \cdot |a_{\mathbf{r}_2}| \cdot \Theta(\mathbf{r}_1 - \mathbf{r}_2) \cdot (|\mathbf{r}_1| \cdot |\mathbf{r}_2|)^{-1/2}. \tag{8.46}$$

Using the simple inequality

$$|a_{\mathbf{r}_1}| \cdot |a_{\mathbf{r}_2}| \le \frac{|a_{\mathbf{r}_1}|^2 + |a_{\mathbf{r}_2}|^2}{2}$$

in (8.46), we have

$$\widetilde{\sum}_1^*(2) \le N \cdot \left(\frac{2}{\pi vT} \right) \cdot \sum_{\substack{\mathbf{r}_1 \in \mathbb{Z}^3: \\ \mathbf{r}_1 \ne \mathbf{0}}} \sum_{\substack{\mathbf{r}_2 \in \mathbb{Z}^3: \\ \mathbf{r}_2 \notin \{\mathbf{0}, \mathbf{r}_1\}}} \frac{|a_{\mathbf{r}_1}|^2 + |a_{\mathbf{r}_2}|^2}{2} \cdot \Theta(\mathbf{r}_1 - \mathbf{r}_2) \cdot (|\mathbf{r}_1| \cdot |\mathbf{r}_2|)^{-1/2}. \tag{8.47}$$

We estimate the coefficient of $|a_{\mathbf{r}_1}|^2$ in (8.47) as follows (write $\mathbf{r}_3 = \mathbf{r}_1 - \mathbf{r}_2$):

$$|a_{\mathbf{r}_1}|^2 \cdot \left(\sum_{\substack{\mathbf{r}_2 \in \mathbb{Z}^3: \\ \mathbf{r}_2 \notin \{\mathbf{0}, \mathbf{r}_1\}}} \Theta(\mathbf{r}_1 - \mathbf{r}_2) \cdot (|\mathbf{r}_1| \cdot |\mathbf{r}_2|)^{-1/2} \right)$$

$$= |a_{\mathbf{r}_1}|^2 \cdot \left(\sum_{\substack{\mathbf{r}_3 \in \mathbb{Z}^3: \\ \mathbf{r}_3 \notin \{\mathbf{0}, -\mathbf{r}_2\}}} \Theta(\mathbf{r}_3) \cdot (|\mathbf{r}_2 + \mathbf{r}_3| \cdot |\mathbf{r}_2|)^{-1/2} \right)$$

$$\le |a_{\mathbf{r}_1}|^2 \cdot \sqrt{2}\left(\sum_{r\in\mathbb{Z}} \frac{1}{\max\{|r|^{7/6}, 1\}}\right)^3$$

$$= |a_{\mathbf{r}_1}|^2 \cdot \sqrt{2}\left(1 + 2\sum_{r=1}^{\infty} \frac{1}{r^{7/6}}\right)^3 < 4999|a_{\mathbf{r}_1}|^2. \tag{8.48}$$

Note that in (8.48) we used the triangle inequality $|\mathbf{r}_2 + \mathbf{r}_3| + |\mathbf{r}_2| \ge |\mathbf{r}_3|$, which implies

$$\sqrt{2}\max\{\sqrt{|\mathbf{r}_2 + \mathbf{r}_3|}, \sqrt{|\mathbf{r}_2|}\}$$

$$\ge \sqrt{|\mathbf{r}_3|} = \left(r_{3,1}^2 + r_{3,2}^2 + r_{3,3}^2\right)^{1/4} \ge \prod_{j=1}^{3} \left(\max\{|r_{3,j}|, 1\}\right)^{1/6} = (\Theta(\mathbf{r}_3))^{-1/6}, \tag{8.49}$$

where $\mathbf{r}_3 = (r_{3,1}, r_{3,2}, r_{3,3})$. Inequality (8.49) and (8.31) together clearly justify (8.48).

By (8.47)–(8.48) we have

$$\widetilde{\sum}_1^*(2) \le N \cdot \left(\frac{2}{\pi v T}\right) \cdot 4999 \sum_{\substack{\mathbf{r}\in\mathbb{Z}^3:\\ \mathbf{r}\neq\mathbf{0}}} |a_{\mathbf{r}}|^2 \le \frac{9998N \cdot \mathrm{vol}(A)}{\pi v T}, \tag{8.50}$$

where in the last step we used (8.42) (i.e., Parseval's formula).

Finally, for $\widetilde{\sum}_1^*(3) = (5.9)$ we use (5.16):

$$\widetilde{\sum}_1^*(3) \le \frac{2N \cdot \mathrm{vol}(A)}{\pi v T}. \tag{8.51}$$

Returning to (8.38), and applying (8.44), (8.50)–(8.51), we have

$$\left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\frac{1}{T}\int_0^T \sum_{k=1}^{N} (\chi_A(\mathbf{x}_k(\omega; t)) - \mathrm{vol}(A))\, dt\right)^2 d\omega$$

$$= \widetilde{\sum}_1^* \le \widetilde{\sum}_1^*(1) + \widetilde{\sum}_1^*(2) + \widetilde{\sum}_1^*(3)$$

$$\le \left(10N\frac{\log(vT)}{vT}\right)^2 \cdot \mathrm{vol}(A) + \frac{10^4 \cdot N \cdot \mathrm{vol}(A)}{\pi v T}. \tag{8.52}$$

We consider (8.52) an "analog" of Lemma 5.1.

By (8.21) and (8.52) we obtain the following "analog" of (8.5):

$$\left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\sum{}^{(2)}(\omega; T)\right)^2 d\omega$$

$$\le \left(\frac{10N\log(vT)}{v}\right)^2 \cdot \mathrm{vol}(A) + \frac{10^4 \cdot TN \cdot \mathrm{vol}(A)}{\pi v}. \tag{8.53}$$

Next we prove an

Analog of Lemma 7.1 As usual, for any integers $1 \le j < k \le N$ let

$$A_{j,k}(T) = A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k)$$

denote the total time between $0 < t < T$ when the jth torus-line $\mathbf{x}_j(t)$ and the kth torus-line $\mathbf{x}_k(t)$ are both in subset A simultaneously. We describe $A_{j,k}(T)$ in terms of the Cartesian product $A \times A \subset I^6 = [0, 1]^6$ of $A \subset I^3$ with itself:

$$\begin{aligned} A_{j,k}(T) &= A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k) \\ &= \text{measure}\left\{t \in [0, T] : \mathbf{x}_j(t) \in A \pmod 1 \text{ and } \mathbf{x}_k(t) \in A \pmod 1\right\} \\ &= \int_0^T \chi_A(\mathbf{x}_j(t))\chi_A(\mathbf{x}_k(t))\,dt = \int_0^T \chi_{A\times A}(\mathbf{x}_j(t), \mathbf{x}_k(t))\,dt, \end{aligned} \tag{8.54}$$

where $\chi_{A\times A}$ is the 0-1 valued characteristic function of $A \times A \subset I^6$. Write $B = A \times A$; we work with the Fourier series of the characteristic function $\chi_B = \chi_{A\times A}$:

$$\chi_B(\mathbf{w}) = \chi_{A\times A}(\mathbf{w}) = \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{w}} \quad \text{with } b_{\mathbf{r}} = \int_{A\times A} e^{-2\pi i \mathbf{r}\cdot\mathbf{z}}\,d\mathbf{z}. \tag{8.55}$$

Clearly $b_{\mathbf{0}} = \text{vol}(A \times A) = \text{vol}^2(A)$, and by Parseval's formula,

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}} |b_{\mathbf{r}}|^2 = \text{vol}^2(A) - \text{vol}^4(A). \tag{8.56}$$

We have

$$\begin{aligned} A_{j,k}(T) &= A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k) \\ &= \int_0^T \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot(\mathbf{x}_j(t),\mathbf{x}_k(t))}\,dt \\ &= \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} \int_0^T e^{2\pi i \mathbf{r}\cdot(\mathbf{x}_j(t),\mathbf{x}_k(t))}\,dt = \sum_{\mathbf{r}\in\mathbb{Z}^6} b_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot(\mathbf{y}_j,\mathbf{y}_k)} \int_0^T e^{2\pi i (\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vt}\,dt \\ &= b_{\mathbf{0}} T + \sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}} b_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot(\mathbf{y}_j,\mathbf{y}_k)} \cdot \frac{e^{2\pi i (\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT} - 1}{2\pi i (\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))v}. \end{aligned} \tag{8.57}$$

As usual, $(\mathbf{y}_j, \mathbf{y}_k)$ means a 6-dimensional vector for which the first 3 coordinates are given by $\mathbf{y}_j$ and the last 3 coordinates are given by $\mathbf{y}_k$.

Consider the large sum

$$\begin{aligned} F_{N,N} &= F_{N,N}(\mathbf{y}_i, \mathbf{u}_i : 1 \le i \le N) \\ &= \sum_{1\le j<k\le N} \left(\frac{1}{T} A_{j,k}(T) - \text{vol}^2(A)\right) \\ &= \sum_{1\le j<k\le N} \frac{1}{T}\int_0^T \chi_A(\mathbf{x}_j(t))\chi_A(\mathbf{x}_k(t))\,dt \; - M^2 \cdot \text{vol}^2(A). \end{aligned} \tag{8.58}$$

We have

$$F_{N,N} = \sum_{1\le j<k\le N} \sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}} b_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot(\mathbf{y}_j,\mathbf{y}_k)} \cdot \frac{e^{2\pi i (\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT} - 1}{2\pi i (\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))v}. \tag{8.59}$$

Fix the N unit vectors $\mathbf{u}_i \in S^2$, $1 \le i \le N$, and evaluate the square integral

$$\widetilde{\sum}_{1,1} = 8^N \int_{\tilde{I}^3} \dots \int_{\tilde{I}^3} \left(F_{N,N}(\mathbf{y}_i, \mathbf{u}_i : 1 \le i \le N)\right)^2 d\mathbf{y}_1 \dots d\mathbf{y}_N, \tag{8.60}$$

where $\tilde{I} = [0, 1/2]$, and so $\tilde{I}^3 = [0, 1/2]^3$ is the first octant of the unit cube.

To evaluate (8.60), we multiply out the square $F_{N,N}^2$, which leads us to the following sub-problem: we have to estimate the multiple integral

$$\begin{aligned} &\widetilde{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2) \\ &\quad = 8 \int_{\tilde{I}^3} \dots 8 \int_{\tilde{I}^3} e^{2\pi \mathrm{i}(\mathbf{r}_1 \cdot (\mathbf{y}_{j_1}, \mathbf{y}_{k_1}) - \mathbf{r}_2 \cdot (\mathbf{y}_{j_2}, \mathbf{y}_{k_2}))}\, d\mathbf{y}_{j_1} \dots d\mathbf{y}_{k_2}, \end{aligned} \tag{8.61}$$

where $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{Z}^6 \setminus \mathbf{0}$, $1 \le j_1 < k_1 \le N$ and $1 \le j_2 < k_2 \le N$.

The first challenge is that the integral (8.61) can be a double, or a triple, or a quadruple integral; it depends on whether the index set $\{j_1, j_2, k_1, k_2\}$ consists of 2 or 3 or 4 different integers. Accordingly, we distinguish several cases.

Case 1: quadruple integral: j_1, k_1, j_2, k_2 are four distinct integers

Then by (8.29),

$$\left|\widetilde{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2)\right| \le \Theta(\mathbf{r}_1)\Theta(\mathbf{r}_2).$$

Case 2a: triple integral: $j_1 = j_2$ is the only coincidence

Then with $\mathbf{r}_1 = (\mathbf{r}_{1,1}, \mathbf{r}_{1,2})$ and $\mathbf{r}_2 = (\mathbf{r}_{2,1}, \mathbf{r}_{2,2})$,

$$\begin{aligned} &\left|\widetilde{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2)\right| \\ &\quad = 8^3 \left| \int_{\tilde{I}^3} \int_{\tilde{I}^3} \left(\int_{\tilde{I}^3} e^{2\pi \mathrm{i}(\mathbf{r}_{1,1} - \mathbf{r}_{2,1}) \cdot \mathbf{y}_{j_1}}\, d\mathbf{y}_{j_1} \right) e^{2\pi \mathrm{i}(\mathbf{r}_{1,2} \cdot \mathbf{y}_{k_1} - \mathbf{r}_{2,2} \cdot \mathbf{y}_{k_2})}\, d\mathbf{y}_{k_1}\, d\mathbf{y}_{k_2} \right| \\ &\quad \le \Theta(\mathbf{r}_{1,1} - \mathbf{r}_{2,1})\Theta(\mathbf{r}_{1,2})\Theta(\mathbf{r}_{2,2}), \end{aligned}$$

and this covers even the extreme case when $\mathbf{r}_{1,1} = \mathbf{r}_{2,1}$ and $\mathbf{r}_{1,2} = \mathbf{r}_{2,2} = (0, 0, 0)$ (then of course the integral is 1). Similar result holds for the other cases, such as

Case 2b: triple integral: $k_1 = j_2$ is the only coincidence

Then

$$\begin{aligned} &\left|\widetilde{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2)\right| \\ &\quad \le \Theta(\mathbf{r}_{1,2} - \mathbf{r}_{2,1})\Theta(\mathbf{r}_{1,1})\Theta(\mathbf{r}_{2,2}). \end{aligned}$$

Case 2c: triple integral: $j_1 = k_2$ is the only coincidence

Then

$$\begin{aligned} &\left|\widetilde{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2)\right| \\ &\quad \le \Theta(\mathbf{r}_{1,1} - \mathbf{r}_{2,2})\Theta(\mathbf{r}_{1,2})\Theta(\mathbf{r}_{2,1}). \end{aligned}$$

Case 2d: triple integral: $k_1 = k_2$ is the only coincidence

Then

$$\begin{aligned}&\left|\widetilde{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2)\right|\\&\quad\le \Theta(\mathbf{r}_{1,2}-\mathbf{r}_{2,2})\Theta(\mathbf{r}_{1,1})\Theta(\mathbf{r}_{2,1}).\end{aligned}$$

Finally, we have

Case 3: double integral: $j_1 = j_2$ and $k_1 = k_2$

Then

$$\left|\widetilde{Int}(\mathbf{r}_1, j_1, k_1; \mathbf{r}_2, j_2, k_2)\right| = 8^2\left|\int_{\tilde{I}^6} e^{2\pi \mathrm{i}(\mathbf{r}_1-\mathbf{r}_2)\cdot\mathbf{y}}\,d\mathbf{y}\right| \le \Theta(\mathbf{r}_1-\mathbf{r}_2),$$

which covers even the extreme case $\mathbf{r}_1 = \mathbf{r}_2$.

Now we are ready to evaluate $\widetilde{\sum}_{1,1}$ (see (8.60)): squaring (8.59) and applying Cases 1, 2a–2d, 3 above, we obtain

$$\begin{aligned}&8^N\int_{\tilde{I}^3}\cdots\int_{\tilde{I}^3}\big(F_{N,N}(\mathbf{y}_i,\mathbf{u}_i : 1\le i\le N)\big)^2\,d\mathbf{y}_1\ldots d\mathbf{y}_N\\&\quad=\widetilde{\sum}_{1,1}\\&\quad\le \widetilde{\sum}_{1,1}(1)+\widetilde{\sum}_{1,1}(2a)+\widetilde{\sum}_{1,1}(2b)\\&\qquad+\widetilde{\sum}_{1,1}(2c)+\widetilde{\sum}_{1,1}(2d)+\widetilde{\sum}_{1,1}(3),\end{aligned}\tag{8.62}$$

where

$$\begin{aligned}\widetilde{\sum}_{1,1}(1) = &\sum_{\substack{\mathbf{r}_1\in\mathbb{Z}^6:\\ \mathbf{r}_1\ne\mathbf{0}}}\ \sum_{\substack{\mathbf{r}_2\in\mathbb{Z}^6:\\ \mathbf{r}_2\ne\mathbf{0}}}\ \sum_{1\le j_1<k_1\le N}\ \sum_{\substack{1\le j_2<k_2\le N:\\ (j_2,k_2)\ne(j_1,k_1)}} |b_{\mathbf{r}_1}|\cdot\Theta(\mathbf{r}_1)\cdot|b_{\mathbf{r}_2}|\cdot\Theta(\mathbf{r}_2)\\&\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}-1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}\right|\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot(\mathbf{u}_{j_2},\mathbf{u}_{k_2}))vT}-1}{2\pi(\mathbf{r}_2\cdot(\mathbf{u}_{j_2},\mathbf{u}_{k_2}))vT}\right|,\end{aligned}\tag{8.63}$$

and

$$\begin{aligned}\widetilde{\sum}_{1,1}(2a) = &\sum_{\substack{\mathbf{r}_1=(\mathbf{r}_{1,1},\mathbf{r}_{1,2})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{\substack{\mathbf{r}_2=(\mathbf{r}_{2,1},\mathbf{r}_{2,2})\in\mathbb{Z}^6:\\ \mathbf{r}_2\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{1\le j_1<k_1\le N}\ \sum_{\substack{j_1<k_2\le N:\\ k_2\ne k_1}} |b_{\mathbf{r}_1}|\cdot|b_{\mathbf{r}_2}|\\&\cdot\Theta(\mathbf{r}_{1,1}-\mathbf{r}_{2,1})\cdot\Theta(\mathbf{r}_{1,2})\Theta(\mathbf{r}_{2,2})\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}-1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}\right|\\&\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_2}))vT}-1}{2\pi(\mathbf{r}_2\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_2}))vT}\right|,\end{aligned}\tag{8.64}$$

and similarly,

$$\widetilde{\sum}_{1,1}(2b) = \sum_{\substack{\mathbf{r}_1=(\mathbf{r}_{1,1},\mathbf{r}_{1,2})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{\substack{\mathbf{r}_2=(\mathbf{r}_{2,1},\mathbf{r}_{2,2})\in\mathbb{Z}^6:\\ \mathbf{r}_2\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{1\le j_1<k_1\le N}\ \sum_{k_1<k_2\le N} |b_{\mathbf{r}_1}|\cdot|b_{\mathbf{r}_2}|$$

$$\cdot\,\Theta(\mathbf{r}_{1,2}-\mathbf{r}_{2,1})\cdot\Theta(\mathbf{r}_{1,1})\Theta(\mathbf{r}_{2,2})\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}-1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}\right|$$

$$\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot(\mathbf{u}_{k_1},\mathbf{u}_{k_2}))vT}-1}{2\pi(\mathbf{r}_2\cdot(\mathbf{u}_{k_1},\mathbf{u}_{k_2}))vT}\right|, \tag{8.65}$$

$$\widetilde{\sum}_{1,1}(2c)=\sum_{\substack{\mathbf{r}_1=(\mathbf{r}_{1,1},\mathbf{r}_{1,2})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{\substack{\mathbf{r}_2=(\mathbf{r}_{2,1},\mathbf{r}_{2,2})\in\mathbb{Z}^6:\\ \mathbf{r}_2\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{1\le j_1<k_1\le N}\ \sum_{1\le j_2<j_1}|b_{\mathbf{r}_1}|\cdot|b_{\mathbf{r}_2}|$$

$$\cdot\,\Theta(\mathbf{r}_{1,1}-\mathbf{r}_{2,2})\cdot\Theta(\mathbf{r}_{1,2})\Theta(\mathbf{r}_{2,1})\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}-1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}\right|$$

$$\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot(\mathbf{u}_{j_2},\mathbf{u}_{j_1}))vT}-1}{2\pi(\mathbf{r}_2\cdot(\mathbf{u}_{j_2},\mathbf{u}_{j_1}))vT}\right|, \tag{8.66}$$

$$\widetilde{\sum}_{1,1}(2d)=\sum_{\substack{\mathbf{r}_1=(\mathbf{r}_{1,1},\mathbf{r}_{1,2})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{\substack{\mathbf{r}_2=(\mathbf{r}_{2,1},\mathbf{r}_{2,2})\in\mathbb{Z}^6:\\ \mathbf{r}_2\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{1\le j_1<k_1\le N}\ \sum_{\substack{1\le j_2<k_1:\\ j_2\ne j_1}}|b_{\mathbf{r}_1}|\cdot|b_{\mathbf{r}_2}|$$

$$\cdot\,\Theta(\mathbf{r}_{1,2}-\mathbf{r}_{2,2})\cdot\Theta(\mathbf{r}_{1,1})\Theta(\mathbf{r}_{2,1})\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}-1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1}))vT}\right|$$

$$\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot(\mathbf{u}_{j_2},\mathbf{u}_{k_1}))vT}-1}{2\pi(\mathbf{r}_2\cdot(\mathbf{u}_{j_2},\mathbf{u}_{k_1}))vT}\right|, \tag{8.67}$$

and finally,

$$\widetilde{\sum}_{1,1}(3)=\sum_{\substack{\mathbf{r}_1\in\mathbb{Z}^6:\\ \mathbf{r}_1\ne\mathbf{0}}}\ \sum_{\substack{\mathbf{r}_2\in\mathbb{Z}^6:\\ \mathbf{r}_2\ne\mathbf{0}}}\ \sum_{1\le j<k\le N}|b_{\mathbf{r}_1}|\cdot|b_{\mathbf{r}_2}|$$

$$\cdot\,\Theta(\mathbf{r}_1-\mathbf{r}_2)\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}-1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}\right|\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}_2\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}-1}{2\pi(\mathbf{r}_2\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}\right|. \tag{8.68}$$

Next we integrate $\widetilde{\sum}_{1,1}$ over the N direction vectors $\mathbf{u}_i\in S^2$, $1\le i\le N$:

$$\begin{aligned}
\widetilde{\sum}^*_{1,1}&=(4\pi)^{-N}\int_{S^2}\cdots\int_{S^2}\widetilde{\sum}_{1,1}(\mathbf{u}_i:1\le i\le N)\,d\mathbf{u}_1\ldots d\mathbf{u}_N\\
&\le(4\pi)^{-N}\int_{S^2}\cdots\int_{S^2}\Bigg(\widetilde{\sum}_{1,1}(1)+\widetilde{\sum}_{1,1}(2a)+\widetilde{\sum}_{1,1}(2b)\\
&\qquad+\widetilde{\sum}_{1,1}(2c)+\widetilde{\sum}_{1,1}(2d)+\widetilde{\sum}_{1,1}(3)\Bigg)\,d\mathbf{u}_1\ldots d\mathbf{u}_N\\
&=\widetilde{\sum}^*_{1,1}(1)+\widetilde{\sum}^*_{1,1}(2a)+\widetilde{\sum}^*_{1,1}(2b)\\
&\qquad+\widetilde{\sum}^*_{1,1}(2c)+\widetilde{\sum}^*_{1,1}(2d)+\widetilde{\sum}^*_{1,1}(3).
\end{aligned}\tag{8.69}$$

Again we use the obvious upper bound

$$\left|\frac{e^{2\pi i(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}-1}{2\pi(\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k))vT}\right| \le \min\left\{\frac{1}{\pi|\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k)|vT},1\right\}. \tag{8.70}$$

Repeating the arguments in (7.13)–(7.16), we obtain the following analog of (8.40):

$$\begin{aligned}(4\pi)^{-2}\int_{S^2}\int_{S^2}\min\left\{\frac{1}{\pi|\mathbf{r}\cdot(\mathbf{u}_j,\mathbf{u}_k)|vT},1\right\}d\mathbf{u}_j d\mathbf{u}_k \\ \le \frac{1+\log(\pi vT\max\{|\mathbf{r}|,1\})}{\pi vT\max\{|\mathbf{r}|_\infty,1\}},\end{aligned} \tag{8.71}$$

where, as usual,

$$|\mathbf{r}|_\infty = \max_{1\le i\le d}|r_i| \quad \text{if } \mathbf{r}=(r_1,\ldots,r_d)$$

denotes the sup-norm; clearly $\sqrt{d}|\mathbf{r}|_\infty \ge |\mathbf{r}|$. Thus have

$$\begin{aligned}\widetilde{\sum}^*_{1,1}(1) &\le \binom{N}{2}^2\left(\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\ne\mathbf{0}}}|b_\mathbf{r}|\cdot\Theta(\mathbf{r})\cdot\frac{1+\log(\pi vT|\mathbf{r}|)}{\pi vT|\mathbf{r}|_\infty}\right)^2 \\ &\le \frac{N^4}{4}\cdot\left(\frac{\log(vT)}{vT}\right)^2\cdot\left(\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\ne\mathbf{0}}}|b_\mathbf{r}|^2\right) \\ &\quad\cdot\left(\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\ne\mathbf{0}}}\Theta^2(\mathbf{r})\cdot\left(\frac{1+\log|\mathbf{r}|}{|\mathbf{r}|_\infty}\right)^2\right),\end{aligned} \tag{8.72}$$

where in the last step we used the Cauchy–Schwarz inequality. By (8.56) (Parseval's formula):

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\ne\mathbf{0}}}|b_\mathbf{r}|^2 \le \mathrm{vol}^2(A). \tag{8.73}$$

Furthermore, by using the definition of $\Theta(\mathbf{r})$ (see (8.31)), and the trivial inequality

$$\frac{1+\log|\mathbf{r}|}{|\mathbf{r}|_\infty} \le \sqrt{6}\frac{1+\log|\mathbf{r}|}{|\mathbf{r}|} \le \sqrt{6},$$

we have the following analog of (8.43):

$$\begin{aligned}&\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\ne\mathbf{0}}}\Theta^2(\mathbf{r})\cdot\left(\frac{1+\log|\mathbf{r}|}{|\mathbf{r}|_\infty}\right)^2 \\ &\quad\le \sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\ne\mathbf{0}}}6\Theta^2(\mathbf{r})\end{aligned}$$

$$\le 6\left(\sum_{r\in\mathbb{Z}}\frac{1}{\max\{r^2,1\}}\right)^6 = 6\left(1+2\sum_{r=1}^{\infty}\frac{1}{r^2}\right)^6$$
$$= \left(1+2\cdot\frac{\pi^2}{6}\right)^6 < 4\cdot 10^4. \tag{8.74}$$

Note that a more careful estimation gives a better constant factor in (8.74):

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}} \Theta^2(\mathbf{r})\cdot\left(\frac{1+\log|\mathbf{r}|}{|\mathbf{r}|_\infty}\right)^2 < 10^4. \tag{8.75}$$

(I challenge the reader to double-check (8.75).) Combining these facts, we obtain

$$\widetilde{\sum}^*_{1,1}(1) \le 10^4\cdot\frac{N^4}{4}\cdot\mathrm{vol}^2(A)\cdot\left(\frac{\log(vT)}{vT}\right)^2, \tag{8.76}$$

which is an analog of (8.44).

Next we study $\widetilde{\sum}^*_{1,1}(2a)$. The following inequality is an analog of (8.45) (we assume $\mathbf{r}_1\neq\mathbf{0}\neq\mathbf{r}_2$):

$$(4\pi)^{-3}\int_{S^2}\int_{S^2}\int_{S^2}\min\left\{\frac{1}{\pi|\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1})|vT},1\right\}$$
$$\cdot\min\left\{\frac{1}{\pi|\mathbf{r}_2\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_2})|vT},1\right\}d\mathbf{u}_{j_1}d\mathbf{u}_{k_1}d\mathbf{u}_{k_2}$$
$$\le\left((4\pi)^{-3}\int_{S^2}\int_{S^2}\int_{S^2}\min\left\{\frac{1}{\pi|\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_1})|vT},1\right\}d\mathbf{u}_{j_1}d\mathbf{u}_{k_1}d\mathbf{u}_{k_2}\right)^{1/2}$$
$$\cdot\left((4\pi)^{-3}\int_{S^2}\int_{S^2}\int_{S^2}\min\left\{\frac{1}{\pi|\mathbf{r}_1\cdot(\mathbf{u}_{j_1},\mathbf{u}_{k_2})|vT},1\right\}d\mathbf{u}_{j_1}d\mathbf{u}_{k_1}d\mathbf{u}_{k_2}\right)^{1/2}$$
$$\le\left(\frac{2}{\pi vT}\right)\cdot(|\mathbf{r}_1|_\infty\cdot|\mathbf{r}_2|_\infty)^{-1/2}, \tag{8.77}$$

where in the last step we used (7.16) twice.

Thus we have

$$\widetilde{\sum}^*_{1,1}(2a)\le\frac{N^3}{3}\sum_{\substack{\mathbf{r}_1=(\mathbf{r}_{1,1},\mathbf{r}_{1,2})\in\mathbb{Z}^6:\\ \mathbf{r}_1\in\mathbb{Z}^6\setminus\mathbf{0}}}\ \sum_{\substack{\mathbf{r}_2=(\mathbf{r}_{2,1},\mathbf{r}_{2,2})\in\mathbb{Z}^6:\\ \mathbf{r}_2\in\mathbb{Z}^6\setminus\mathbf{0}}}|b_{\mathbf{r}_1}|\cdot|b_{\mathbf{r}_2}|$$
$$\cdot\Theta(\mathbf{r}_{1,1}-\mathbf{r}_{2,1})\cdot\Theta(\mathbf{r}_{1,2})\Theta(\mathbf{r}_{2,2})\cdot\frac{2}{\pi vT}\cdot(|\mathbf{r}_1|_\infty\cdot|\mathbf{r}_2|_\infty)^{-1/2}. \tag{8.78}$$

Similarly to (8.46)–(8.48), we use the simple inequality

$$|b_{\mathbf{r}_1}|\cdot|b_{\mathbf{r}_2}|\le\frac{|b_{\mathbf{r}_1}|^2+|b_{\mathbf{r}_2}|^2}{2}$$

in (8.78), and then estimate the coefficient of $|b_{\mathbf{r}_1}|^2$ as follows (write $\mathbf{r}_1 - \mathbf{r}_2 = \mathbf{r}_3 = (\mathbf{r}_{3,1}, \mathbf{r}_{3,2})$):

$$|b_{\mathbf{r}_1}|^2 \cdot \left(\sum_{\substack{\mathbf{r}_2=(\mathbf{r}_{2,1},\mathbf{r}_{2,2})\in\mathbb{Z}^6: \\ \mathbf{r}_2\in\mathbb{Z}^6\setminus\mathbf{0}}} \Theta(\mathbf{r}_{1,1}-\mathbf{r}_{2,1}) \cdot \Theta(\mathbf{r}_{1,2})\Theta(\mathbf{r}_{2,2}) \cdot \frac{2}{\pi v T} \cdot (|\mathbf{r}_1|_\infty \cdot |\mathbf{r}_2|_\infty)^{-1/2} \right)$$

$$= |b_{\mathbf{r}_1}|^2 \cdot \left(\sum_{\substack{\mathbf{r}_{3,1}\in\mathbb{Z}^3: \\ \mathbf{r}_{3,1}\neq\mathbf{0}}} \sum_{\mathbf{r}_{1,2}\in\mathbb{Z}^3} \sum_{\mathbf{r}_{2,2}\in\mathbb{Z}^3} \Theta(\mathbf{r}_{3,1}) \cdot \Theta(\mathbf{r}_{1,2})\Theta(\mathbf{r}_{2,2}) \right.$$
$$\left. \cdot (|\mathbf{r}_2+\mathbf{r}_3|_\infty \cdot |\mathbf{r}_2|_\infty)^{-1/2} \right). \tag{8.79}$$

By repeating the argument in (8.49), we have

$$2\max\{|\mathbf{r}_2+\mathbf{r}_3|_\infty, |\mathbf{r}_2|_\infty\}$$
$$\geq \max\{|\mathbf{r}_3|_\infty, 2\} \geq \prod_{j=1}^{3} \left(\max\{|r_{3,j}|, 1\}\right)^{1/3} = \left(\Theta(\mathbf{r}_{3,1})\right)^{-1/3}, \tag{8.80}$$

where $\mathbf{r}_3 = (r_{3,1}, r_{3,2}, \ldots, r_{3,6})$.

Also,

$$|\mathbf{r}_1|_\infty \geq \prod_{j=4}^{6} \left(\max\{|r_{1,j}|, 1\}\right)^{1/3} = \left(\Theta(\mathbf{r}_{1,2})\right)^{-1/3}, \tag{8.81}$$

and

$$|\mathbf{r}_2|_\infty \geq \prod_{j=4}^{6} \left(\max\{|r_{2,j}|, 1\}\right)^{1/3} = \left(\Theta(\mathbf{r}_{2,2})\right)^{-1/3}. \tag{8.82}$$

We combine the factorization

$$(|\mathbf{r}_1|_\infty \cdot |\mathbf{r}_2|_\infty)^{1/2} = (|\mathbf{r}_2+\mathbf{r}_3|_\infty \cdot |\mathbf{r}_2|_\infty)^{1/4} \cdot |\mathbf{r}_1|_\infty^{1/4} \cdot |\mathbf{r}_2|_\infty^{1/4} \tag{8.83}$$

with (8.80), (8.81), (8.82) as follows: we apply (8.80), (8.81), (8.82) in this order to the three factors on the right hand side of (8.83). This gives the following upper bound to (8.79):

$$(8.79) \leq |b_{\mathbf{r}_1}|^2 \cdot 2^{1/4} \left(\sum_{r\in\mathbb{Z}} \frac{1}{\max\{|r|^{13/12}, 1\}} \right)^9$$
$$= |b_{\mathbf{r}_1}|^2 \cdot 2^{1/4} \left(1 + 2\sum_{r=1}^{\infty} \frac{1}{r^{13/12}} \right)^9 < 10^{13}|b_{\mathbf{r}_1}|^2. \tag{8.84}$$

Thus we have

$$\widetilde{\sum}^*_{1,1}(2a) \leq \frac{N^3}{3} \cdot \frac{2}{\pi v T} \cdot 10^{13} \sum_{\mathbf{r}\in\mathbb{Z}^6\setminus\mathbf{0}} |b_{\mathbf{r}}|^2$$

$$\le \frac{2\cdot 10^{13}}{3\pi vT}\cdot N^3\cdot \mathrm{vol}^2(A), \tag{8.85}$$

where in the last step we used (8.56).

Of course, the same argument—and the same upper bound—applies for "2b", "2c" and "2d" instead of "2a" in (8.85).

It remains to estimate $\widetilde{\sum}^*_{1,1}(3)$: see (8.68). The argument is basically the same as the proof of (8.85) for $\widetilde{\sum}^*_{1,1}(2a)$; in fact, this case is somewhat simpler. We have

$$\begin{aligned}\widetilde{\sum}^*_{1,1}(3) &= \binom{N}{2}\sum_{\substack{\mathbf{r}_1\in\mathbb{Z}^6:\\ \mathbf{r}_1\neq\mathbf{0}}}\sum_{\substack{\mathbf{r}_2\in\mathbb{Z}^6:\\ \mathbf{r}_2\neq\mathbf{0}}}\frac{|b_{\mathbf{r}_1}|^2+|b_{\mathbf{r}_2}|^2}{2}\cdot\Theta(\mathbf{r}_1-\mathbf{r}_2)\cdot\frac{2}{\pi vT}\cdot(|\mathbf{r}_1|_\infty\cdot|\mathbf{r}_2|_\infty)^{-1/2}\\ &\le \binom{N}{2}\cdot\frac{2}{\pi vT}\cdot\left(\sum_{\substack{\mathbf{r}\in\mathbb{Z}^6:\\ \mathbf{r}\neq\mathbf{0}}}|b_{\mathbf{r}}|^2\right)\cdot 2^{1/2}\left(\sum_{r\in\mathbb{Z}}\frac{1}{\max\{|r|^{13/12},1\}}\right)^6\\ &< \frac{6\cdot 10^8}{\pi vT}\cdot N^2\cdot\mathrm{vol}^2(A).\end{aligned} \tag{8.86}$$

Summarizing, we have

$$\begin{aligned}&\left(\frac{2}{\pi}\right)^N\int_{\widetilde{\Omega}}\left(\frac{1}{T}\int_0^T\sum_{1\le j<k\le N}(\chi_A(\mathbf{x}_j(\omega;t))\chi_A(\mathbf{x}_k(\omega;t))-\mathrm{vol}^2(A))\,dt\right)^2 d\omega\\ &\quad=\widetilde{\sum}^*_{1,1}\le\widetilde{\sum}^*_{1,1}(1)+\widetilde{\sum}^*_{1,1}(2a)+\cdots+\widetilde{\sum}^*_{1}(2d)+\widetilde{\sum}^*_{1,1}(3)\\ &\quad\le\frac{10^4}{4}N^4\cdot\mathrm{vol}^2(A)\left(\frac{\log(vT)}{vT}\right)^2+4\cdot\frac{2\cdot 10^{13}\cdot N^3\cdot\mathrm{vol}^2(A)}{\pi vT}\\ &\qquad+\frac{6\cdot 10^8 N^2\cdot\mathrm{vol}^2(A)}{\pi vT}.\end{aligned} \tag{8.87}$$

We consider (8.87) an "analog" of Lemma 7.1.

By (8.20) and (8.87) we obtain the following "analog" of (8.6):

$$\begin{aligned}&\left(\frac{2}{\pi}\right)^N\int_{\widetilde{\Omega}}\left(\sum{}^{(1)}(\omega;T)\right)^2 d\omega\\ &\quad\le\frac{10^4}{4}N^4\cdot\mathrm{vol}^2(A)\left(\frac{\log(vT)}{v}\right)^2+4\cdot\frac{2\cdot 10^{13}\cdot TN^3\cdot\mathrm{vol}^2(A)}{\pi v}\\ &\qquad+\frac{6\cdot 10^8 TN^2\cdot\mathrm{vol}^2(A)}{\pi v}.\end{aligned} \tag{8.88}$$

Now we are ready to prove Theorem 3. Using (8.19) and the simple inequality (8.7), we have

$$\left(\frac{2}{\pi}\right)^N\int_{\widetilde{\Omega}}\left(\int_0^T(Y_A(\omega;t)-N\cdot\mathrm{vol}(A))^2\,dt\right)^2 d\omega$$

$$\leq 3 \cdot 2^2 \left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\sum^{(1)}(\omega; T)\right)^2 d\omega$$
$$+ 3 \cdot (2N \cdot \mathrm{vol}(A))^2 \left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\sum^{(2)}(\omega; T)\right)^2 d\omega + 3N^2T^2, \tag{8.89}$$

where in the last step we used (8.23).

Using (8.53) and (8.88) in (8.89), we have

$$\left(\frac{2}{\pi}\right)^N \int_{\widetilde{\Omega}} \left(\frac{1}{T}\int_0^T \left(\frac{Y_A(\omega; t) - N \cdot \mathrm{vol}(A)}{N}\right)^2 dt\right)^2 d\omega$$
$$\leq 3 \cdot 10^4 \cdot \mathrm{vol}^2(A) \left(\frac{\log(vT)}{vT}\right)^2 + 12 \cdot 10^2 \cdot \mathrm{vol}^3(A) \left(\frac{\log(vT)}{vT}\right)^2 + \frac{4 \cdot 10^4 \cdot \mathrm{vol}^3(A)}{vTN}$$
$$+ \frac{32 \cdot 10^{13} \cdot \mathrm{vol}^2(A)}{vTN} + \frac{24 \cdot 10^8 \cdot \mathrm{vol}^2(A)}{vTN^2} + \frac{3}{N^2}, \tag{8.90}$$

which completes the proof of Theorem 3.

In the rest of the paper we complete the proof of Theorem 1.

9 Proof of Theorem 1: The General Simultaneous Case

Let $\ell \geq 2$ be an arbitrary integer. Section 7 was about the special case $\ell = 2$; now we discuss the general case in a very similar way.

Assume that $2 \leq \ell \leq m = N/M$. Let $1 \leq k(1) < k(2) < \cdots < k(\ell) \leq N$ be arbitrary integers, and let

$$A_{k(1),\ldots,k(\ell)}(T) = A_{k(1),\ldots,k(\ell)}(T; \mathbf{y}_{k(j)}, \mathbf{u}_{k(j)} : 1 \leq j \leq \ell)$$

denote the total time between $0 < t < T$ when the ℓ torus-lines $\mathbf{x}_{k(j)}(t)$, $1 \leq j \leq \ell$ are all in subset A simultaneously; in other words, when these ℓ torus lines are in A *at the same time*.

The key observation is that we can describe $A_{k(1),\ldots,k(\ell)}(T)$ in terms of the ℓth Cartesian power $A^\ell = A \times \cdots \times A \subset I^{3\ell} = [0, 1]^{3\ell}$ of $A \subset I^3 = [0, 1]^3$. Indeed, we have

$$\begin{aligned} A_{k(1),\ldots,k(\ell)}(T) &= A_{k(1),\ldots,k(\ell)}(T; \mathbf{y}_{k(j)}, \mathbf{u}_{k(j)} : 1 \leq j \leq \ell) \\ &= \text{measure}\left\{t \in [0, T] : \mathbf{x}_{k(j)}(t) \in A \ (\mathrm{mod}\ 1) \text{ for all } 1 \leq j \leq \ell\right\} \\ &= \int_0^T \chi_A(\mathbf{x}_{k(1)}(t)) \cdots \chi_A(\mathbf{x}_{k(\ell)}(t))\, dt \\ &= \int_0^T \chi_{A\times\cdots\times A}(\mathbf{x}_{k(1)}(t), \ldots, \mathbf{x}_{k(\ell)}(t))\, dt, \end{aligned} \tag{9.1}$$

where $\chi_{A\times\cdots\times A}$ is the 0-1 valued characteristic function of $A^\ell = A \times \cdots \times A \subset I^{3\ell}$. Write $B = A^\ell$; we need the Fourier series of the characteristic function $\chi_B = \chi_{A^\ell}$:

$$\chi_B(\mathbf{w}) = \chi_{A\times\cdots\times A}(\mathbf{w}) = \sum_{\mathbf{r}\in\mathbb{Z}^{3\ell}} b_{\mathbf{r}} e^{2\pi i \mathbf{r}\cdot\mathbf{w}} \quad \text{with } b_{\mathbf{r}} = \int_{A\times\cdots\times A} e^{-2\pi i \mathbf{r}\cdot\mathbf{z}}\, d\mathbf{z}, \tag{9.2}$$

where $\mathbf{r}\cdot\mathbf{w} = r_1 w_1 + \cdots + r_{3\ell} w_{3\ell}$ denotes the standard inner product. Clearly $b_{\mathbf{0}} = \mathrm{vol}(A^\ell) = (\mathrm{vol}(A))^\ell$ (= the volume of A^ℓ), and by Parseval's formula,

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}\neq\mathbf{0}}} |b_{\mathbf{r}}|^2 = (\mathrm{vol}(A))^\ell - (\mathrm{vol}(A))^{2\ell}, \tag{9.3}$$

which is the analog of (5.3) and (7.3). Let's return to (9.1): by using the Fourier series (9.2), we have

$$\begin{aligned}
A_{k(1),\dots,k(\ell)}(T) &= A_{k(1),\dots,k(\ell)}(T;\mathbf{y}_{k(j)},\mathbf{u}_{k(j)} : 1\le j\le \ell)\\
&= \int_0^T \sum_{\mathbf{r}\in\mathbb{Z}^{3\ell}} b_{\mathbf{r}} e^{2\pi i\mathbf{r}\cdot(\mathbf{x}_{k(1)}(t),\dots,\mathbf{x}_{k(\ell)}(t))}\,dt\\
&= \sum_{\mathbf{r}\in\mathbb{Z}^{3\ell}} b_{\mathbf{r}} \int_0^T e^{2\pi i\mathbf{r}\cdot(\mathbf{x}_{k(1)}(t),\dots,\mathbf{x}_{k(\ell)}(t))}\,dt\\
&= \sum_{\mathbf{r}\in\mathbb{Z}^{3\ell}} b_{\mathbf{r}} e^{2\pi i\mathbf{r}\cdot(\mathbf{y}_{k(1)},\dots,\mathbf{y}_{k(\ell)})} \int_0^T e^{2\pi i\mathbf{r}\cdot(\mathbf{u}_{k(1)},\dots,\mathbf{u}_{k(\ell)})vt}\,dt\\
&= b_{\mathbf{0}}T + \sum_{\substack{\mathbf{r}\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}\neq\mathbf{0}}} b_{\mathbf{r}} e^{2\pi i\mathbf{r}\cdot(\mathbf{y}_{k(1)},\dots,\mathbf{y}_{k(\ell)})}\cdot\frac{e^{2\pi i(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\dots,\mathbf{u}_{k(\ell)})vT}-1}{2\pi i(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\dots,\mathbf{u}_{k(\ell)}))v}.
\end{aligned}\tag{9.4}$$

To clarify the notation here, note that (say) $(\mathbf{y}_{k(1)},\dots,\mathbf{y}_{k(\ell)})$ means a 3ℓ-dimensional vector for which the first 3 coordinates are given by $\mathbf{y}_{k(1)}$, and so on, and finally, the last 3 coordinates are given by $\mathbf{y}_{k(\ell)}$.

Let M be an arbitrary integer in the range $1\le M\le N/\ell$, and consider the sum

$$\begin{aligned}
F_{1,\dots,\ell} &= F_{1,\dots,\ell}(\mathbf{y}_{k(1)},\mathbf{u}_{k(1)} : 1\le k(1)\le M;\dots,\mathbf{y}_{k(\ell)},\mathbf{u}_{k(\ell)} : (\ell-1)M+1\le k(\ell)\le \ell M)\\
&= \sum_{k(1)=1}^{M}\cdots\sum_{k(\ell)=(\ell-1)M+1}^{\ell M}\left(\frac{1}{T}A_{k(1),\dots,k(\ell)}(T)-(\mathrm{vol}(A))^\ell\right)\\
&= \sum_{k(1)=1}^{M}\cdots\sum_{k(\ell)=(\ell-1)M+1}^{\ell M}\frac{1}{T}\int_0^T \chi_A(\mathbf{x}_{k(1)}(t))\cdots\chi_A(\mathbf{x}_{k(\ell)}(t))\,dt \;-M^\ell\cdot(\mathrm{vol}(A))^\ell\\
&= E_{1,\dots,\ell}-M^\ell\cdot(\mathrm{vol}(A))^\ell,
\end{aligned}\tag{9.5}$$

where

$$E_{1,\dots,\ell} = \frac{1}{T}\int_0^T Z_1(t)\cdots Z_\ell(t)\,dt,$$

$$Z_1(t) = \sum_{k=1}^{M}\chi_A(\mathbf{x}_k(t)),\dots,Z_\ell(t) = \sum_{k=(\ell-1)M+1}^{\ell M}\chi_A(\mathbf{x}_k(t)).$$

By (9.4) we have

$$F_{1,\dots,\ell} = \sum_{k(1)=1}^{M} \cdots \sum_{k(\ell)=(\ell-1)M+1}^{\ell M} \sum_{\substack{\mathbf{r}\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}\neq\mathbf{0}}} b_{\mathbf{r}} e^{2\pi i\mathbf{r}\cdot(\mathbf{y}_{k(1)},\dots,\mathbf{y}_{k(\ell)})} \cdot \frac{e^{2\pi i(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\dots,\mathbf{u}_{k(\ell)}))vT} - 1}{2\pi i(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\dots,\mathbf{u}_{k(\ell)}))v}. \quad (9.6)$$

Fix the ℓM unit vectors $\mathbf{u}_{k(1)} \in S^2$, $k(1) = 1, 2, \dots, M, \dots, \mathbf{u}_{k(\ell)} \in S^2$, $k(\ell) = (\ell-1)M + 1, \dots, \ell M$, and evaluate the square integral

$$\sum\nolimits_{1,\dots,\ell} = \int_{I^3} \cdots \int_{I^3} \big(F_{1,\dots,\ell}(\mathbf{y}_{k(1)}, \mathbf{u}_{k(1)} : 1 \le k(1) \le M; \dots; \mathbf{y}_{k(\ell)}, \mathbf{u}_{k(\ell)} : (\ell-1)M + 1 \le k(\ell) \le \ell M)\big)^2 \, d\mathbf{y}_1 \dots d\mathbf{y}_{\ell M}. \quad (9.7)$$

Note that (9.7) is a multiple integral, which consists of ℓM single integrals.

To evaluate (9.7), we multiply out the square $F^2_{1,\dots,\ell}$ (where for $F_{1,\dots,\ell}$ we use (9.6)) and apply some orthogonality relations, leading to huge cancellations. To understand the cancellations, we study the following sub-problem: When does the multiple integral

$$\begin{aligned} &Int(\mathbf{r}_1, k(1)_1, \dots, k(\ell)_1; \mathbf{r}_2, k(1)_2, \dots, k(\ell)_2) \\ &= \int_{I^3} \cdots \int_{I^3} e^{2\pi i(\mathbf{r}_1\cdot(\mathbf{y}_{k(1)_1},\dots,\mathbf{y}_{k(\ell)_1}) - \mathbf{r}_2\cdot(\mathbf{y}_{k(1)_2},\dots,\mathbf{y}_{k(\ell)_2}))} \, d\mathbf{y}_{k(1)_1} \dots d\mathbf{y}_{k(\ell)_2}, \end{aligned} \quad (9.8)$$

where $\mathbf{r}_1, \mathbf{r}_2 \in \mathbb{Z}^{3\ell} \setminus \mathbf{0}$ and $1 \le k(1)_1, k(1)_2 \le M < \dots \le (\ell-1)M < k(\ell)_1, k(\ell)_2 \le \ell M$, equal to zero?

Well, the first challenge is that in the multiple integral (9.8) the number of integrations depends on whether the index set $\{k(1)_1, \dots, k(\ell)_1, \dots, k(1)_2, \dots, k(\ell)_2\}$ consists of ℓ or $\ell+1$ or $\dots$ or 2ℓ different integers. Accordingly, we distinguish $\ell+1$ cases.

Case 1: $\{k(1)_1, \dots, k(\ell)_1, \dots, k(1)_2, \dots, k(\ell)_2\}$ consists of 2ℓ different integers

Then clearly

$$Int(\mathbf{r}_1, k(1)_1, \dots, k(\ell)_1; \mathbf{r}_2, k(1)_2, \dots, k(\ell)_2) = 0.$$

Case 2: $\{k(1)_1, \dots, k(\ell)_1, \dots, k(1)_2, \dots, k(\ell)_2\}$ consists of $2\ell - 1$ different integers

For notational simplicity, assume that $k(1)_1 = k(1)_2$. Write $\mathbf{r}_1 = (\mathbf{r}_{1,1}, \mathbf{r}_{1,2})$ and $\mathbf{r}_2 = (\mathbf{r}_{2,1}, \mathbf{r}_{2,2})$, where $\mathbf{r}_{1,1}, \mathbf{r}_{2,1} \in \mathbb{Z}^3$ and $\mathbf{r}_{1,2}, \mathbf{r}_{2,2} \in \mathbb{Z}^{3\ell-3}$. Then

$$\begin{aligned} &Int(\mathbf{r}_1, k(1)_1, \dots, k(\ell)_1; \mathbf{r}_2, k(1)_2, \dots, k(\ell)_2) \\ &= \int_{I^3} \cdots \int_{I^3} \left(\int_{I^3} e^{2\pi i(\mathbf{r}_{1,1}-\mathbf{r}_{2,1})\cdot\mathbf{y}_{k(1)_1}} \, d\mathbf{y}_{k(1)_1} \right) \\ &\quad \times e^{2\pi i(\mathbf{r}_{1,2}\cdot(\mathbf{y}_{k(2)_1},\dots,\mathbf{y}_{k(\ell)_1}) - \mathbf{r}_{2,2}\cdot(\mathbf{y}_{k(2)_2},\dots,\mathbf{y}_{k(\ell)_2}))} \, d\mathbf{y}_{k(2)_1} \dots d\mathbf{y}_{k(\ell)_2}, \end{aligned}$$

and this integral is always 0, unless $\mathbf{r}_{1,1} = \mathbf{r}_{2,1}$ and $\mathbf{r}_{1,2} = \mathbf{r}_{2,2} = (0, \dots, 0) = \mathbf{0}$, and then of course the integral is 1. Similar result holds for the other cases when $k(j)_1 = k(j)_2$ for exactly one index j in $1 \le j \le \ell$.

Case 3: $\{k(1)_1, \dots, k(\ell)_1, \dots, k(1)_2, \dots, k(\ell)_2\}$ consists of $2\ell - 2$ different integers

For notational simplicity, assume that $k(1)_1 = k(1)_2$ and $k(2)_1 = k(2)_2$. Write $\mathbf{r}_1 = (\mathbf{r}_{1,1}, \mathbf{r}_{1,2})$ and $\mathbf{r}_2 = (\mathbf{r}_{2,1}, \mathbf{r}_{2,2})$, where $\mathbf{r}_{1,1}, \mathbf{r}_{2,1} \in \mathbb{Z}^6$ and $\mathbf{r}_{1,2}, \mathbf{r}_{2,2} \in \mathbb{Z}^{3\ell-6}$. Then

$$\begin{aligned}
&Int(\mathbf{r}_1, k(1)_1, \ldots, k(\ell)_1; \mathbf{r}_2, k(1)_2, \ldots, k(\ell)_2) \\
&\quad = \int_{I^3} \cdots \int_{I^3} \left(\int_{I^6} e^{2\pi \mathrm{i}(\mathbf{r}_{1,1} - \mathbf{r}_{2,1})\cdot \mathbf{y}}\, d\mathbf{y} \right) \\
&\qquad \times e^{2\pi \mathrm{i}(\mathbf{r}_{1,2}\cdot(\mathbf{y}_{k(3)_1}, \ldots, \mathbf{y}_{k(\ell)_1}) - \mathbf{r}_{2,2}\cdot(\mathbf{y}_{k(3)_2}, \ldots, \mathbf{y}_{k(\ell)_2}))}\, d\mathbf{y}_{k(3)_1} \ldots d\mathbf{y}_{k(\ell)_2},
\end{aligned}$$

and this integral is always 0, unless $\mathbf{r}_{1,1} = \mathbf{r}_{2,1}$ and $\mathbf{r}_{1,2} = \mathbf{r}_{2,2} = (0, \ldots, 0) = \mathbf{0}$, and then of course the integral is 1. Similar result holds for the other cases when $k(j)_1 = k(j)_2$ for exactly two indexes $j = j_1$ and $j = j_2$ in $1 \le j_1 < j_2 \le \ell$.

Cases 4, 5, ... go similarly; for illustration I just include the last case.

Case $\ell + 1$: $k(j)_1 = k(j)_2$ holds for all $1 \le j \le \ell$

Then

$$\begin{aligned}
&Int(\mathbf{r}_1, k(1)_1, \ldots, k(\ell)_1; \mathbf{r}_2, k(1)_2, \ldots, k(\ell)_2) \\
&\quad = \int_{I^{3\ell}} e^{2\pi \mathrm{i}(\mathbf{r}_1 - \mathbf{r}_2)\cdot \mathbf{y}}\, d\mathbf{y},
\end{aligned}$$

which is always 0, unless $\mathbf{r}_1 = \mathbf{r}_2$.

Now we are ready to evaluate $\sum_{1,\ldots,\ell}$ (see (9.7)): squaring (9.6) and applying Cases 1, 2, $\ldots, \ell, \ell+1$ above (in fact, we apply them in reverse order), we obtain

$$\begin{aligned}
\sum\nolimits_{1,\ldots,\ell} &= \sum\nolimits_{1,2} (\mathbf{u}_{k(1)} : 1 \le k(1) \le M; \ldots; \mathbf{u}_{k(\ell)} : (\ell-1)M + 1 \le k(\ell) \le \ell M) \\
&= \sum_{\substack{\mathbf{r}\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}\neq\mathbf{0}}} |b_{\mathbf{r}}|^2 \cdot \sum_{k(1)=1}^{M} \cdots \sum_{k(\ell)=(\ell-1)M+1}^{\ell M} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_{k(1)}, \ldots, \mathbf{u}_{k(\ell)}))vT} - 1}{2\pi(\mathbf{r}\cdot(\mathbf{u}_{k(1)}, \ldots, \mathbf{u}_{k(\ell)}))vT} \right|^2 \\
&\quad + \sum_{\substack{\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}_1\in\mathbb{Z}^{3\ell-3}\setminus\mathbf{0}}} |b_{\mathbf{r}}|^2 \cdot \sum_{k(1)=1}^{M} \cdots \sum_{k(\ell-1)=(\ell-2)M+1}^{(\ell-1)M} \\
&\quad \sum_{k(\ell)_1=(\ell-1)M+1}^{\ell M} \sum_{k(\ell)_2=(\ell-1)M+1}^{\ell M} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_{k(1)}, \ldots, \mathbf{u}_{k(\ell-1)})vT} - 1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_{k(1)}, \ldots, \mathbf{u}_{k(\ell-1)}))vT} \right|^2 + \cdots \\
&\quad + \sum_{\substack{\mathbf{r}=(\mathbf{0},\mathbf{r}_\ell)\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}_\ell\in\mathbb{Z}^{3\ell-3}\setminus\mathbf{0}}} |b_{\mathbf{r}}|^2 \cdot \sum_{k(1)_1=1}^{M} \sum_{k(1)_2=1}^{M} \sum_{k(2)=M+1}^{2M} \\
&\quad \sum_{k(3)=2M+1}^{3M} \cdots \sum_{k(\ell)=(\ell-1)M+1}^{\ell M} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}_\ell\cdot(\mathbf{u}_{k(2)}, \ldots, \mathbf{u}_{k(\ell)})vT} - 1}{2\pi(\mathbf{r}_2\cdot(\mathbf{u}_{k(2)}, \ldots, \mathbf{u}_{k(\ell)}))vT} \right|^2 \\
&\quad + \sum_{\substack{\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}_1\in\mathbb{Z}^{3\ell-6}\setminus\mathbf{0}}} |b_{\mathbf{r}}|^2 \cdot \sum_{k(1)=1}^{M} \cdots \sum_{k(\ell-2)=(\ell-3)M+1}^{(\ell-2)M} \sum_{k(\ell-1)_1=(\ell-2)M+1}^{(\ell-1)M} \sum_{k(\ell-1)_2=(\ell-2)M+1}^{(\ell-1)M}
\end{aligned}$$

$$\sum_{k(\ell)_1=(\ell-1)M+1}^{\ell M} \sum_{k(\ell)_2=(\ell-1)M+1}^{\ell M} \left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}_1\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell-2)})vT} - 1}{2\pi(\mathbf{r}_1\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell-2)}))vT} \right|^2 + \cdots . \tag{9.9}$$

Next we integrate $\sum_{1,\ldots,\ell}$ over the ℓM direction vectors $\mathbf{u}_{k(1)} \in S^2$, $k(1) = 1, \ldots, M, \ldots,$ $\mathbf{u}_{k(\ell)} \in S^2$, $k(\ell) = (\ell-1)M+1, \ldots, \ell M$ (this multiple integral consists of ℓM single integrals):

$$\begin{aligned} \sum\nolimits_{1,\ldots,\ell}^{*} = (4\pi)^{-\ell M} \int_{S^2} \cdots \int_{S^2} \sum\nolimits_{1,\ldots,\ell} (\mathbf{u}_{k(1)} : 1 \le k(1) \le M; \ldots; \mathbf{u}_{k(\ell)} : \\ (\ell-1)M+1 \le k(\ell) \le \ell M)\, d\mathbf{u}_1 \ldots d\mathbf{u}_{\ell M}. \end{aligned} \tag{9.10}$$

Let's return to the first sum in (9.9): we have the obvious upper bound

$$\left| \frac{e^{2\pi \mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)}))vT} - 1}{2\pi(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)}))vT} \right| \le \min\left\{ \frac{1}{\pi|\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)})|vT}, 1 \right\}. \tag{9.11}$$

We need to estimate the integral

$$(4\pi)^{-\ell} \int_{S^2} \cdots \int_{S^2} \min\left\{ \frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)})vT)^2}, 1 \right\} d\mathbf{u}_{k(1)} \ldots d\mathbf{u}_{k(\ell)}. \tag{9.12}$$

I recall (7.13): for any real numbers $c_1 < c_2$ we have,

$$\mathrm{SurfaceArea}\left(\left\{\mathbf{u} \in S^2 : c_1 \le \mathbf{r}\cdot\mathbf{u} \le c_2\right\}\right) = 4\pi \cdot \frac{\min\{c_2, r\} - \max\{c_1, -r\}}{r}, \tag{9.13}$$

where $r = |\mathbf{r}|$ and $\mathbf{r} \in \mathbb{Z}^3 \setminus \mathbf{0}$.

Now let $\mathbf{r} \in \mathbb{Z}^{3\ell} \setminus \mathbf{0}$, and write $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_\ell)$; then clearly $\mathbf{r}_1 \in \mathbb{Z}^3 \setminus \mathbf{0}$ or $\ldots$ or $\mathbf{r}_\ell \in \mathbb{Z}^3 \setminus \mathbf{0}$. Suppose that (say) $\mathbf{r}_1 \in \mathbb{Z}^3 \setminus \mathbf{0}$; then we can estimate the integral in (9.12) as follows:

$$\begin{aligned} &(4\pi)^{-\ell} \int_{S^2} \cdots \int_{S^2} \min\left\{ \frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)})vT)^2}, 1 \right\} d\mathbf{u}_{k(1)} \ldots d\mathbf{u}_{k(\ell)} \\ &= (4\pi)^{-\ell} \int_{S^2} \cdots \int_{S^2} \\ &\left(\int_{S^2} \min\left\{ \frac{1}{(\pi(\mathbf{r}_1\cdot\mathbf{u}_{k(1)} + \cdots + \mathbf{r}_\ell\cdot\mathbf{u}_{k(\ell)})vT)^2}, 1 \right\} d\mathbf{u}_{k(1)} \right) d\mathbf{u}_{k(2)} \ldots d\mathbf{u}_{k(\ell)}. \end{aligned} \tag{9.14}$$

For any fixed value of $\mathbf{u}_{k(2)}, \ldots, \mathbf{u}_{k(\ell)}$, the inner integral in (9.14) can be estimated from above by repeating the argument in (5.15) and using (9.13), and thus we obtain the upper bound (let $c_0 = \mathbf{r}_2\cdot\mathbf{u}_{k(2)} + \cdots + \mathbf{r}_\ell\cdot\mathbf{u}_{k(\ell)}$)

$$\int_{S^2} \min\left\{ \frac{1}{(\pi(\mathbf{r}_1\cdot\mathbf{u}_{k(1)} + c_0)vT)^2}, 1 \right\} d\mathbf{u}_{k(1)} \le \frac{2}{\pi vT|\mathbf{r}_1|},$$

which is an analog of (7.15). Using this in (9.14), we obtain

$$\begin{aligned} &(4\pi)^{-\ell} \int_{S^2} \cdots \int_{S^2} \min\left\{ \frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)})vT)^2}, 1 \right\} d\mathbf{u}_{k(1)} \ldots d\mathbf{u}_{k(\ell)} \\ &\le \frac{2}{\pi vT|\mathbf{r}_1|} \le \frac{2}{\pi vT}. \end{aligned} \tag{9.15}$$

Let's return now to (9.9)–(9.10). By using (9.11) and (9.15), we have

$$(4\pi)^{-\ell M}\int_{S^2}\cdots\int_{S^2}$$

$$\sum_{\substack{\mathbf{r}\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}\neq\mathbf{0}}}\sum_{k(1)=1}^{M}\cdots\sum_{k(\ell)=(\ell-1)M+1}^{\ell M}|b_{\mathbf{r}}|^2\cdot\left|\frac{e^{2\pi\mathrm{i}(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)}))vT}-1}{2\pi(\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)}))vT}\right|^2 d\mathbf{u}_1\ldots d\mathbf{u}_{\ell M}$$

$$\le(4\pi)^{-\ell M}\int_{S^2}\cdots\int_{S^2}\sum_{k(1)=1}^{M}\cdots\sum_{k(\ell)=(\ell-1)M+1}^{\ell M}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}\neq\mathbf{0}}}|b_{\mathbf{r}}|^2$$

$$\cdot\min\left\{\frac{1}{(\pi\mathbf{r}\cdot(\mathbf{u}_{k(1)},\ldots,\mathbf{u}_{k(\ell)})vT)^2},1\right\}d\mathbf{u}_1\ldots d\mathbf{u}_{\ell M}$$

$$\le\frac{2}{\pi vT}\cdot M^{\ell}\sum_{\substack{\mathbf{r}\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}\neq\mathbf{0}}}|b_{\mathbf{r}}|^2\le\frac{2}{\pi vT}\cdot M^{\ell}\cdot(\mathrm{vol}(A))^{\ell},\tag{9.16}$$

where in the last step we used (9.3). This settles the contribution of the first big sum on the right hand side of (9.9).

The rest of (9.9) can be handled similarly; the only novelty is to replace (9.3) with the following argument (see (9.17) below). Suppose that (say) $\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^{3\ell}$ with $\mathbf{r}_1\in\mathbb{Z}^{3j}\setminus\mathbf{0}$ and $1\le j<\ell$; then

$$b_{\mathbf{r}}=\int_{A^\ell}e^{-2\pi\mathrm{i}\mathbf{r}\cdot\mathbf{z}}\,d\mathbf{z}$$

$$=(\mathrm{vol}(A))^{\ell-j}\int_{A^j}e^{-2\pi\mathrm{i}\mathbf{r}_1\cdot\mathbf{w}}\,d\mathbf{w}=(\mathrm{vol}(A))^{\ell-j}\cdot a_{\mathbf{r}_1},$$

where $a_{\mathbf{r}_1}$ (i.e., the Fourier coefficient of the characteristic function of $A^j=A\times\cdots\times A$ j-times) satisfies the Parseval formula:

$$\sum_{\mathbf{r}_1\in\mathbb{Z}^{3j}}|a_{\mathbf{r}_1}|^2=\mathrm{vol}(A^j)=(\mathrm{vol}(A))^j.$$

Thus we have

$$\sum_{\substack{\mathbf{r}=(\mathbf{r}_1,\mathbf{0})\in\mathbb{Z}^{3\ell}:\\ \mathbf{r}_1\in\mathbb{Z}^{3j}\setminus\mathbf{0}}}|b_{\mathbf{r}}|^2\le(\mathrm{vol}(A))^{2\ell-2j}\cdot(\mathrm{vol}(A))^j=(\mathrm{vol}(A))^{2\ell-j}.\tag{9.17}$$

This is the analog of (7.19)–(7.20).

Therefore, we obtain the following analog of Lemma 7.1.

Lemma 9.1 *For every integer $1\le\ell\le m=N/M$ we have*

$$\sum\nolimits^*_{1,\ldots,\ell}\le\frac{2}{\pi vT}\cdot\left(M^{\ell}\cdot(\mathrm{vol}(A))^{\ell}+\ell\cdot M^{\ell+1}\cdot(\mathrm{vol}(A))^{\ell+1}+\binom{\ell}{2}M^{\ell+2}\cdot(\mathrm{vol}(A))^{\ell+2}\right.$$

$$\left.+\binom{\ell}{3}M^{\ell+3}\cdot(\mathrm{vol}(A))^{\ell+3}+\cdots+\binom{\ell}{\ell-1}M^{2\ell-1}\cdot(\mathrm{vol}(A))^{2\ell-1}\right)$$

$$= \frac{2}{\pi v T} \cdot M^{2\ell} \cdot (\mathrm{vol}(A))^{2\ell} \left(\left(\frac{1}{M \cdot \mathrm{vol}(A)} + 1 \right)^{\ell} - 1 \right), \tag{9.18}$$

where the square-integral $\sum^{*}_{1,\dots,\ell}$ *(see* (9.5), (9.7), (9.10)) *equals the multiple integral*

$$(4\pi)^{-\ell M} \int_{I^3} \cdots \int_{I^3} \int_{S^2} \cdots \int_{S^2}$$
$$\left(\sum_{k(1)=1}^{M} \cdots \sum_{k(\ell)=(\ell-1)M+1}^{\ell M} \left(\frac{1}{T} A_{k(1),\dots,k(\ell)}(T) - (\mathrm{vol}(A))^{\ell} \right) \right)^2 d\mathbf{y}_1 \dots d\mathbf{y}_{\ell M}\, d\mathbf{u}_1 \dots d\mathbf{u}_{\ell M}, \tag{9.19}$$

which consists of $2\ell M$ *single integrals.*

10 Proof of Theorem 1: Applying Lemmas 5.1 and 7.1

The objective of this section is to prove Lemma 10.1; see at the end. We recall that $A_k(T) = A_k(T; \mathbf{y}_k, \mathbf{u}_k)$ denotes the time the kth torus-line $\mathbf{x}_k(t)$—defined in (5.1), noting that $\mathbf{y}_k$ is the initial position and $\mathbf{u}_k$ is the direction—spends in subset A during $0 < t < T$, that is,

$$\begin{aligned} A_k(T) = A_k(T; \mathbf{y}_k, \mathbf{u}_k) &= \text{measure}\,\{t \in [0, T] : \mathbf{x}_k(t) \in A \ (\mathrm{mod}\ 1)\} \\ &= \int_0^T \chi_A(\mathbf{x}_k(t))\, dt. \end{aligned}$$

Also I use Γ to denote the set of all $N!$ permutation of $1, 2, \dots, N$. (Note in advance that the permutations play a crucial role in the applications of Lemmas 7.1 and 9.1; see below.) Let $\gamma \in \Gamma$ be an arbitrary permutation, and for any $1 \le h \le m = N/M$ write

$$\begin{aligned} E_h(\gamma) &= E_h(T; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma) \\ &= \frac{1}{T} \sum_{k=(h-1)M+1}^{hM} A_{\gamma(k)}(T; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)}) = \frac{1}{T} \int_0^T Z_h(\gamma; t)\, dt, \end{aligned} \tag{10.1}$$

where

$$\begin{aligned} Z_h(\gamma; t) &= Z_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma; t) \\ &= \sum_{k=(h-1)M+1}^{hM} \chi_A(\mathbf{x}_{\gamma(k)}(t)), \end{aligned} \tag{10.2}$$

where for convenience we use the short notation

$$\mathcal{I}(h) = \{(h-1)M + 1, (h-1)M + 2, \dots, hM\}.$$

By Lemma 5.1, for every h in $1 \le h \le m$ we have (here we break the long integral into three lines)

$$(4\pi)^{-N}\int_{I^3}\cdots\int_{I^3}\int_{S^2}\cdots\int_{S^2}\frac{1}{N!}\sum_{\gamma\in\Gamma}$$

$$\big(E_h(T;\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:k\in\mathcal{I}(h);\gamma)-M\cdot\mathrm{vol}(A)\big)^2\,d\mathbf{y}_1\ldots d\mathbf{y}_N\,d\mathbf{u}_1\ldots d\mathbf{u}_N$$

$$\leq\frac{2}{\pi vT}\cdot M\cdot\mathrm{vol}(A). \tag{10.3}$$

Let

$$\Omega=I^3\times\cdots\times I^3\times S^2\times\cdots\times S^2=I^{3N}\times\big(S^2\big)^N \tag{10.4}$$

denote the space of all initial conditions

$$(\mathbf{y}_1,\ldots,\mathbf{y}_N,\mathbf{u}_1,\ldots,\mathbf{u}_N)\in\Omega, \tag{10.5}$$

where $I=[0,1]$ and S^2 is the unit sphere (in the usual 3-space). Clearly measure$(\Omega)=(4\pi)^N$, since the surface area of the unit sphere S^2 is 4π (of course *measure* means the product measure).

Let $0<\varepsilon<1$ be arbitrary but fixed; its value will be specified later. We obtain from (10.3) by a standard *average argument* that, there is a measurable subset $\Omega_1(bad)$ of Ω such that, for all initial conditions

$$(\mathbf{y}_1,\ldots,\mathbf{y}_N,\mathbf{u}_1,\ldots,\mathbf{u}_N)\in\Omega\setminus\Omega_1(bad), \tag{10.6}$$

we have that

$$\frac{1}{N!}\sum_{\gamma\in\Gamma}\big(E_h(T;\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:k\in\mathcal{I}(h);\gamma)-M\cdot\mathrm{vol}(A)\big)^2$$

$$\leq\frac{4m}{\varepsilon}\cdot\frac{2}{\pi vT}\cdot M\cdot\mathrm{vol}(A) \tag{10.7}$$

holds for all $1\leq h\leq m$, and $\Omega_1(bad)$ is "negligible":

$$\text{measure}\,(\Omega_1(bad))<\frac{\varepsilon}{4}\text{measure}(\Omega). \tag{10.8}$$

Indeed, the *average argument* goes as follows. For each $h=1,2,\ldots,m$ we delete the set of initial conditions which violate (10.7); each such set forms a small minority, because the violation of (10.7) means "much larger than the average". Thus we altogether delete less than $m\cdot\frac{\varepsilon}{4m}=\frac{\varepsilon}{4}$ part, proving (10.8). This kind of *average argument* will be used repeatedly in the rest of the paper.

Needless to say, we consider $\Omega_1(bad)$ a "bad" subset of Ω; we will throw it away at the end of the proof of the theorem.

Let j,k be arbitrary integers with $1\leq j<k\leq N$; I recall that

$$A_{j,k}(T)=A_{j,k}(T;\mathbf{y}_j,\mathbf{u}_j;\mathbf{y}_k,\mathbf{u}_k)$$

denotes the total time between $0<t<T$ when the jth torus-line $\mathbf{x}_j(t)$ and the kth torus-line $\mathbf{x}_k(t)$ are both in subset A *simultaneously*. We have

$$\begin{aligned}
A_{j,k}(T) &= A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k) \\
&= \text{measure}\left\{t \in [0, T] : \mathbf{x}_j(t) \in A \ (\text{mod } 1) \text{ and } \mathbf{x}_k(t) \in A \ (\text{mod } 1)\right\} \\
&= \int_0^T \chi_A(\mathbf{x}_j(t))\chi_A(\mathbf{x}_k(t))\, dt.
\end{aligned} \tag{10.9}$$

Again let $\gamma \in \Gamma$ be an arbitrary permutation of $1, 2, \ldots, N$, and for any $1 \le h \le m = N/M$ write

$$\begin{aligned}
E^{(half)}_{2h-1,2h}(\gamma) &= E^{(half)}_{2h-1,2h}(T; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma) \\
&= \frac{1}{T} \sum_{j=(h-1)M+1}^{(h-\frac{1}{2})M} \sum_{k=(h-\frac{1}{2})M+1}^{hM} A_{\gamma(j),\gamma(k)}(T; \mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)}; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)}) \\
&= \frac{1}{T} \int_0^T Z^{(half)}_{2h-1}(\gamma; t) Z^{(half)}_{2h}(\gamma; t)\, dt,
\end{aligned} \tag{10.10}$$

where

$$\begin{aligned}
Z^{(half)}_{2h-1}(\gamma; t) &= Z^{(half)}_{2h-1}\left(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}\left(h - \frac{1}{2}\right); \gamma; t\right) \\
&= \sum_{k=(h-1)M+1}^{(h-\frac{1}{2})M} \chi_A(\mathbf{x}_{\gamma(k)}(t))
\end{aligned} \tag{10.11}$$

and

$$\begin{aligned}
Z^{(half)}_{2h}(\gamma; t) &= Z^{(half)}_{2h}(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h; 1/2); \gamma; t) \\
&= \sum_{k=(h-\frac{1}{2})M+1}^{hM} \chi_A(\mathbf{x}_{\gamma(k)}(t)),
\end{aligned} \tag{10.12}$$

where for convenience we use the short notation

$$\mathcal{I}(h'; 1/2) = \left\{\left(h' - \frac{1}{2}\right)M + 1, \left(h' - \frac{1}{2}\right)M + 2, \ldots, h'M\right\},$$

noting that h' can be any integer or half-integer.

Applying Lemma 7.1 with $M = M/2$, for every h in $1 \le h \le m$ we have

$$\begin{aligned}
&(4\pi)^{-N} \int_{I^3} \cdots \int_{I^3} \int_{S^2} \cdots \int_{S^2} \frac{1}{N!} \sum_{\gamma \in \Gamma} \\
&\quad \left(E^{(half)}_{2h-1,2h}(T; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma) - \frac{M^2}{4} \cdot \text{vol}^2(A)\right)^2 d\mathbf{y}_1 \ldots d\mathbf{y}_N\, d\mathbf{u}_1 \ldots d\mathbf{u}_N \\
&\le \frac{2}{\pi v T} \cdot \left(\frac{M^2}{4} \cdot \text{vol}^2(A) + 2\frac{M^3}{8} \cdot \text{vol}^3(A)\right).
\end{aligned} \tag{10.13}$$

Let $0 < \varepsilon < 1$; it follows from (10.13) by a standard average argument (similar to (10.7)–(10.8)) that, there is a measurable subset $\Omega_2(bad)$ of Ω such that, for all initial conditions

$$(\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N) \in \Omega \setminus \Omega_2(bad), \tag{10.14}$$

we have that

$$\frac{1}{N!}\sum_{\gamma\in\Gamma}\left(E^{(half)}_{2h-1,2h}(T; \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma) - \frac{M^2}{4}\cdot \mathrm{vol}^2(A)\right)^2$$
$$\le \frac{4m}{\varepsilon}\cdot\frac{2}{\pi v T}\cdot\left(\frac{M^2}{4}\cdot\mathrm{vol}^2(A) + 2\frac{M^3}{8}\cdot\mathrm{vol}^3(A)\right) \tag{10.15}$$

holds for all $1 \le h \le m$, and again $\Omega_2(bad)$ is “negligible”:

$$\text{measure}\,(\Omega_2(bad)) < \frac{\varepsilon}{4}\text{measure}(\Omega). \tag{10.16}$$

Again $\Omega_2(bad)$ is a “bad” subset of Ω that we will throw away at the end.

Consider a “good” initial condition

$$(\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N) \in \Omega \setminus (\Omega_1(bad) \cup \Omega_2(bad))\,. \tag{10.17}$$

It means that we can use (10.7)–(10.8) and (10.15)–(10.16).

The “time average” of

$$Z_h(\gamma; t) = Z_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma; t)$$
$$= \sum_{k=(h-1)M+1}^{hM} \chi_A(\mathbf{x}_{\gamma(k)}(t)),$$

as t runs in $0 < t < T$, equals

$$\frac{1}{T}\int_0^T Z_h(\gamma; t)\,dt = E_h(\gamma). \tag{10.18}$$

The values of $Z_h(\gamma; t)$, as t runs in $0 < t < T$, are non-negative integers $0, 1, 2, 3, \dots$; now for every integer $\ell \ge 0$ we define the set

$$W_h(\gamma; \ell) = W_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma; \ell) = \{t \in [0, T] : Z_h(\gamma; t) = \ell\}. \tag{10.19}$$

Then we have the following disjoint decomposition of the interval $0 \le t \le T$:

$$[0, T] = W_h(\gamma; 0) \cup W_h(\gamma; 1) \cup W_h(\gamma; 2) \cup W_h(\gamma; 3) \cup \cdots$$
$$= W_h(\gamma; 0) \cup W_h(\gamma; 1) \cup W_h(\gamma; \ge 2).$$

Write

$$V_h(\gamma; \ell) = V_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h); \gamma; \ell) = \frac{1}{T}\text{measure}\,(W_h(\gamma; \ell))\,, \tag{10.20}$$

implying

$$0 \le V_h(\gamma; \ell) \le 1.$$

We need to give an upper bound for the size $V_h(\gamma;\ell)$ of a "typical" set $W_h(\gamma;\ell)$ with $\ell \geq 2$ ("typical" means the majority of the initial conditions $\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h)$ and the majority of the permutations $\gamma \in \Gamma$). In fact, we will estimate a whole power-of-two group

$$\sum_{\ell=2^j}^{2^{j+1}-1} V_h(\gamma;\ell) \quad \text{for any integer } j \geq 1.$$

We will obtain such an upper bound by using a second moment argument.

For simplicity assume that M is even; let $I_1 \cup I_2$ be an arbitrary *halving split* of the set $\{(h-1)M+1, (h-1)M+2, \ldots, hM\} = \mathcal{I}(h)$ of M consecutive integers into two disjoint subsets of size $M/2$ each. There are exactly $\binom{M}{M/2}$ such halving splits. For any fixed halving split (I_1, I_2), write

$$Z_{I_1}(\gamma;t)Z_{I_2}(\gamma;t) = \left(\sum_{k_1 \in I_1} \chi_A(\mathbf{x}_{\gamma(k_1)}(t))\right)\left(\sum_{k_2 \in I_2} \chi_A(\mathbf{x}_{\gamma(k_2)}(t))\right).$$

Let $\ell \geq 2$; if $t_0 \in W_h(\gamma;\ell) \Leftrightarrow Z_h(\gamma;t_0) = \ell$ for some $0 \leq t_0 \leq T$, then at least *one-half* of the $\binom{M}{M/2}$ halving splits have the property that

$$Z_{I_1}(\gamma;t_0)Z_{I_2}(\gamma;t_0) \geq 1 \quad \text{for } \ell = 2,3 \quad \text{and} \quad Z_{I_1}(\gamma;t_0)Z_{I_2}(\gamma;t_0) > \frac{\ell^2 - 2\ell}{4} \quad \text{for } \ell \geq 4. \tag{10.21}$$

Indeed, this is exactly the argument in (5.27)–(5.31).

Let $\mathcal{M}$ be an arbitrary M-element subset of the first N integers $\{1, 2, \ldots, N\}$, and let $\mathcal{P}(h;\mathcal{M}) \subset \Gamma$ denote the set of all permutations $\gamma \in \Gamma$ such that

$$\{\gamma(k) : k \in \mathcal{I}(h)\} = \mathcal{M}. \tag{10.22}$$

Note that

$$V_h(\gamma;\ell) = V_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h);\gamma;\ell)$$

depends only on the whole class $\gamma \in \mathcal{P}(h;\mathcal{M})$, and we denote this common value with

$$V_h(\mathcal{P}(h;\mathcal{M});\ell) = V_h(\mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h);\mathcal{P}(h;\mathcal{M});\ell). \tag{10.23}$$

We need the following trivial inequality: if $\alpha \geq 0$ and $\beta \geq 0$ are arbitrary positive real numbers, then

$$0 \leq \alpha < 2\beta \quad \text{or } (\alpha - \beta)^2 \geq \frac{1}{4}\alpha^2. \tag{10.24}$$

Let $j \geq 1$ be an integer, and consider the values of ℓ in the power-of-two group

$$2^j \leq \ell < 2^{j+1}. \tag{10.25}$$

Combining (10.15), (10.21), inequality (10.24), and using notation (10.23), we obtain that

$$\frac{1}{\binom{N}{M}} \sum^{**}_{\substack{\mathcal{M} \subset \{1,\ldots,N\}: \\ |\mathcal{M}|=M}} \frac{1}{4}\left(\left(\sum_{2^j \leq \ell < 2^{j+1}} V_h(\mathcal{P}(h;\mathcal{M});\ell)\right) \frac{(2^j)^2 - 2 \cdot 2^j}{4}\right)^2$$

$$\leq 2 \cdot \frac{4m}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(\frac{M^2}{4} \cdot \mathrm{vol}^2(A) + 2\frac{M^3}{8} \cdot \mathrm{vol}^3(A) \right) \tag{10.26}$$

holds for all $1 \leq h \leq m$ and all $j \geq 2$ (see (10.25)). Note that the double asterisk $**$ in $\sum^{**}$ means that the summation is restricted to the terms satisfying

$$\left(\left(\sum_{2^j \leq \ell < 2^{j+1}} V_h(\mathcal{P}(h; \mathcal{M}); \ell) \right) \frac{(2^j)^2 - 2 \cdot 2^j}{4} \right)^2 \geq 2 \cdot \frac{M^2}{4} \cdot \mathrm{vol}^2(A), \tag{10.27}$$

see (10.24). Also the first factor "2" in the last line in (10.26) comes from the *one-half* right before (10.21).

In the special case $j = 1$, covering the values $\ell = 2$ and 3, we have

$$\frac{1}{\binom{N}{M}} \sum^{**}_{\substack{\mathcal{M} \subset \{1,\ldots,N\}: \\ |\mathcal{M}| = M}} \frac{1}{4} \left(\sum_{2 \leq \ell \leq 3} V_h(\mathcal{P}(h; \mathcal{M}); \ell) \right)^2$$

$$\leq 2 \cdot \frac{4m}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(\frac{M^2}{4} \cdot \mathrm{vol}^2(A) + 2\frac{M^3}{8} \cdot \mathrm{vol}^3(A) \right). \tag{10.28}$$

Let's return to (10.26): we divide both sides with the factor

$$\left(\frac{(2^j)^2 - 2 \cdot 2^j}{4} \right)^2 = 2^{2j-2} \left(2^{j-1} - 1 \right)^2,$$

and obtain

$$\frac{1}{\binom{N}{M}} \sum^{**}_{\substack{\mathcal{M} \subset \{1,\ldots,N\}: \\ |\mathcal{M}| = M}} \left(\sum_{2^j \leq \ell < 2^{j+1}} V_h(\mathcal{P}(h; \mathcal{M}); \ell) \right)^2$$

$$\leq \frac{4}{2^{2j-2}(2^{j-1} - 1)^2} \cdot 2 \cdot \frac{4m}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(\frac{M^2}{4} \cdot \mathrm{vol}^2(A) + 2\frac{M^3}{8} \cdot \mathrm{vol}^3(A) \right) \tag{10.29}$$

for all $1 \leq h \leq m$ and all $j \geq 2$ (it remains true even for $j = 1$ if the first factor in the last line of (10.29) is replaced with "4").

It follows from (10.28)–(10.29) by a standard average argument (similar to (10.7)–(10.8)) that, there is a subset $\Gamma_1(bad)$ of Γ (= the set of all $N!$ permutations of the first N integers) such that, for all permutations

$$\gamma \in \Gamma \setminus \Gamma_1(bad) \tag{10.30}$$

and all integers $1 \leq h \leq m$ and $j \geq 1$,

$$\sum_{2^j \leq \ell < 2^{j+1}} V_h(\gamma; \ell) \leq \frac{4\sqrt{m} \cdot j}{\max\{2^{j-1}(2^{j-1} - 1), 1\}} \cdot C(10.29)$$

$$+ \frac{2}{\max\{2^{j-1}(2^{j-1} - 1), 1\}} \cdot \frac{M^2}{4} \cdot \mathrm{vol}^2(A), \tag{10.31}$$

and

$$\frac{1}{N!}|\Gamma_1(bad)| < \frac{1}{8}, \tag{10.32}$$

where

$$C(10.29) = \left(2 \cdot \frac{4m}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(\frac{M^2}{4} \cdot \mathrm{vol}^2(A) + 2\frac{M^3}{8} \cdot \mathrm{vol}^3(A)\right)\right)^{1/2}, \tag{10.33}$$

the extra factor of j in the numerator right before $C(10.29)$ in (10.31) comes from the convergent series

$$\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6} < 2,$$

the last term in (10.31) comes from (10.27), and finally, in (10.32) I used the standard notation $|\cdots|$ to denote the number of elements of a finite set.

Since we assumed a "good" initial condition (see (10.17)):

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \Omega \setminus (\Omega_1(bad) \cup \Omega_2(bad)), \tag{10.34}$$

we can use (10.7)–(10.8) and (10.15)–(10.16). We already used (10.15) in (10.26); now we use (10.7). A simple average argument (similar to (10.7)–(10.8)) gives that, there is a subset $\Gamma_2(bad)$ of Γ (= the set of $N!$ permutations of $1, 2, \ldots, N$) such that, for all permutations

$$\gamma \in \Gamma \setminus \Gamma_2(bad) \tag{10.35}$$

and all integers $1 \le h \le m$, we have

$$|E_h(\omega; \gamma) - M \cdot \mathrm{vol}(A)| \le 2\sqrt{m} \cdot C(10.7), \tag{10.36}$$

and

$$\frac{1}{N!}|\Gamma_2(bad)| < \frac{1}{8}, \tag{10.37}$$

where

$$C(10.7) = \left(\frac{4m}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot M \cdot \mathrm{vol}(A)\right)^{1/2}. \tag{10.38}$$

Next we turn to (10.17)–(10.20). We have

$$\begin{aligned} E_h(\omega; \gamma) &= \frac{1}{T}\int_0^T Z_h(\omega; \gamma; t)\, dt = \sum_{\ell=1}^{\infty} V_h(\omega; \gamma; \ell) \cdot \ell \\ &= V_h(\omega; \gamma; 1) + \sum_{j=1}^{\infty} \sum_{2^j \le \ell < 2^{j+1}} V_h(\omega; \gamma; \ell) \cdot \ell. \end{aligned} \tag{10.39}$$

Let

$$\gamma \in \Gamma \setminus (\Gamma_1(bad) \cup \Gamma_2(bad)), \tag{10.40}$$

then by (10.31), (10.33), (10.36), (10.38) and (10.39),

$$\begin{aligned}
&|V_h(\omega;\gamma;1) - M\cdot \mathrm{vol}(A)| \\
&\quad\le 2\sqrt{m}\cdot C(10.7) + \sum_{j=1}^{\infty}\left(\sum_{2^j\le \ell<2^{j+1}} V_h(\omega;\gamma;\ell)\right)\cdot 2^{j+1} \\
&\quad\le 2\sqrt{m}\cdot C(10.7) + C(10.29)\cdot \sum_{j=1}^{\infty}\frac{4\sqrt{m}\cdot 2^{j+1}\cdot j}{\max\{2^{j-1}(2^{j-1}-1),1\}} \\
&\qquad + \frac{M^2}{4}\cdot \mathrm{vol}^2(A)\cdot \sum_{j=1}^{\infty}\frac{2\cdot 2^{j+1}}{\max\{2^{j-1}(2^{j-1}-1),1\}}
\end{aligned} \tag{10.41}$$

for every $1\le h\le m$.

Let

$$\mathrm{vol}(A) = \frac{\lambda}{N}. \tag{10.42}$$

Since $m = N/M$, we have $M\cdot\mathrm{vol}(A) = M\cdot\lambda/N = \lambda/m$. Using this in (10.41), and estimating the two infinite series, we obtain

$$\begin{aligned}
&\left|V_h(\omega;\gamma;1) - \frac{\lambda}{m}\right| \\
&\quad\le 4m\left(\frac{1}{\varepsilon}\cdot\frac{1}{vT}\cdot\frac{\lambda}{m}\right)^{1/2} \\
&\qquad + 200m\left(\frac{1}{\varepsilon}\cdot\frac{1}{vT}\cdot\left(\left(\frac{\lambda}{m}\right)^2+\left(\frac{\lambda}{m}\right)^3\right)\right)^{1/2} + \frac{25}{4}\left(\frac{\lambda}{m}\right)^2.
\end{aligned} \tag{10.43}$$

We will also need the following estimate, which can be proved exactly the same way as (10.43)):

$$\begin{aligned}
V_h(\omega;\gamma;\ge 2) &= \sum_{j=1}^{\infty}\left(\sum_{2^j\le\ell<2^{j+1}} V_h(\omega;\gamma;\ell)\right) \\
&\le C(10.29)\cdot\sum_{j=1}^{\infty}\frac{4\sqrt{m}\cdot j}{\max\{2^{j-1}(2^{j-1}-1),1\}} \\
&\quad + \frac{M^2}{4}\cdot\mathrm{vol}^2(A)\cdot\sum_{j=1}^{\infty}\frac{2}{\max\{2^{j-1}(2^{j-1}-1),1\}} \\
&\le 10m\left(\frac{1}{\varepsilon}\cdot\frac{1}{vT}\cdot\left(\left(\frac{\lambda}{m}\right)^2+\left(\frac{\lambda}{m}\right)^3\right)\right)^{1/2} + \left(\frac{\lambda}{m}\right)^2
\end{aligned} \tag{10.44}$$

for every $1\le h\le m$.

Therefore, we have just proved (see (10.42)–(10.44))

Lemma 10.1 *Assume that the initial condition satisfies the requirement*

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \Omega \setminus (\Omega_1(bad) \cup \Omega_2(bad)),$$

and also, the permutation γ satisfies

$$\gamma \in \Gamma \setminus (\Gamma_1(bad) \cup \Gamma_2(bad)),$$

then for every $1 \le h \le m$ we have

$$\left| V_h(\omega; \gamma; 1) - \frac{\lambda}{m} \right| \le \frac{25}{4} \left(\frac{\lambda}{m}\right)^2 + \frac{1}{\sqrt{vT}} \cdot \frac{1}{\sqrt{\varepsilon}} \left(4\sqrt{\lambda m} + 200\lambda \sqrt{\left(1 + \frac{\lambda}{m}\right)} \right), \tag{10.45}$$

and

$$V_h(\omega; \gamma; \ge 2) = \sum_{j=1}^{\infty} \left(\sum_{2^j \le \ell < 2^{j+1}} V_h(\omega; \gamma; \ell) \right) \le \left(\frac{\lambda}{m}\right)^2 + \frac{1}{\sqrt{vT}} \cdot \frac{10}{\sqrt{\varepsilon}} \cdot \lambda \sqrt{\left(1 + \frac{\lambda}{m}\right)}, \tag{10.46}$$

where

$$\mathrm{vol}(A) = \frac{\lambda}{N}.$$

Finally, note that

$$\mathrm{measure}(\Omega_1(bad)) < \frac{\varepsilon}{4}\mathrm{measure}(\Omega), \qquad \mathrm{measure}(\Omega_2(bad)) < \frac{\varepsilon}{4}\mathrm{measure}(\Omega),$$

and similarly, $|\Gamma| = N!$ and

$$|\Gamma_1(bad)| < \frac{N!}{8}, \qquad |\Gamma_2(bad)| < \frac{N!}{8}.$$

The message of Lemma 10.1 is the following: if vT is "large", then the terms in the last line of (10.45) and (10.46) are negligible compared to $(\lambda/m)^2$, and so we have

$$V_h(\omega; \gamma; 1) = \frac{\lambda}{m} + O\left(\left(\frac{\lambda}{m}\right)^2\right) \quad \text{and} \quad V_h(\omega; \gamma; \ge 2) = O\left(\left(\frac{\lambda}{m}\right)^2\right). \tag{10.47}$$

The objective of the next section is to prove a simultaneous version of (10.47).

11 Simultaneous Generalization of Lemma 10.1

Let j, k be arbitrary integers with $1 \le j < k \le N$; I recall that

$$A_{j,k}(T) = A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k)$$

denotes the total time between $0 < t < T$ when the jth torus-line $\mathbf{x}_j(t)$ and the kth torus-line $\mathbf{x}_k(t)$ are both in subset A *simultaneously*:

$$\begin{aligned} A_{j,k}(T) &= A_{j,k}(T; \mathbf{y}_j, \mathbf{u}_j; \mathbf{y}_k, \mathbf{u}_k) \\ &= \text{measure}\left\{t \in [0, T] : \mathbf{x}_j(t) \in A \text{ (mod 1) and } \mathbf{x}_k(t) \in A \text{ (mod 1)}\right\} \\ &= \int_0^T \chi_A(\mathbf{x}_j(t))\chi_A(\mathbf{x}_k(t))\, dt. \end{aligned} \tag{11.1}$$

Again let $\gamma \in \Gamma$ be an arbitrary permutation of $1, 2, \dots, N$, and for any $1 \le h_1 < h_2 \le m = N/M$ write

$$\begin{aligned} &E_{h_1,h_2}(\gamma) \\ &\quad = E_{h_1,h_2}(\mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h_2); \gamma) \\ &\quad = \sum_{j=(h_1-1)M+1}^{h_1 M} \sum_{k=(h_2-1)M+1}^{h_2 M} \frac{1}{T} A_{\gamma(j),\gamma(k)}(T) \\ &\quad = \frac{1}{T}\int_0^T Z_{h_1}(\gamma; t) Z_{h_2}(\gamma; t)\, dt. \end{aligned} \tag{11.2}$$

Note that we are constantly using the short notation (which was already introduced in Sect. 10)

$$\mathcal{I}(h) = \{(h-1)M+1, (h-1)M+2, \dots, hM\},$$

where h is always an integer; but later we will also use

$$\mathcal{I}(h'; 1/2) = \left\{\left(h' - \frac{1}{2}\right)M + 1, \left(h' - \frac{1}{2}\right)M + 2, \dots, h'M\right\},$$

where h' can be any integer or half-integer.

By Lemma 7.1, for every $1 \le h_1 < h_2 \le m$ we have

$$\begin{aligned} &(4\pi)^{-N} \int_{I^3} \cdots \int_{I^3} \int_{S^2} \cdots \int_{S^2} \frac{1}{N!} \sum_{\gamma \in \Gamma} \\ &\qquad \Big(E_{h_1,h_2}(T; \mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h_2); \gamma) \\ &\qquad - M^2 \cdot \text{vol}^2(A) \Big)^2 d\mathbf{y}_1 \dots d\mathbf{y}_N\, d\mathbf{u}_1 \dots d\mathbf{u}_N \\ &\le \frac{2}{\pi v T} \cdot \left(M^2 \cdot \text{vol}^2(A) + 2M^3 \cdot \text{vol}^3(A)\right). \end{aligned} \tag{11.3}$$

Let $0 < \varepsilon < 1$; it follows from (11.3) by a standard average argument (similar to (10.7)–(10.8)) that, there is a measurable subset $\Omega_3(bad)$ of Ω (= the set of all initial conditions) such that, for all initial conditions

$$(\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N) \in \Omega \setminus \Omega_3(bad), \tag{11.4}$$

we have that

$$\frac{1}{N!}\sum_{\gamma\in\Gamma}\bigg(E_{h_1,h_2}(T;\mathbf{y}_{\gamma(j)},\mathbf{u}_{\gamma(j)}:j\in\mathcal{I}(h_1);\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:k\in\mathcal{I}(h_2);\gamma)$$

$$-M^2\cdot\mathrm{vol}^2(A)\bigg)^2\le\frac{4m^2}{\varepsilon}\cdot\frac{2}{\pi vT}\cdot\left(M^2\cdot\mathrm{vol}^2(A)+2M^3\cdot\mathrm{vol}^3(A)\right)\qquad(11.5)$$

holds for all $1\le h_1<h_2\le m$, and $\Omega_3(bad)$ is "negligible":

$$\text{measure}\,(\Omega_3(bad))<\frac{\varepsilon}{8}\text{measure}(\Omega).\qquad(11.6)$$

As usual, $\Omega_3(bad)$ is a "bad" subset of Ω that we will throw away at the end.

Again let $1\le h_1<h_2\le m=N/M$, let $\delta=0$ or 1, and define (see also (9.1))

$$\begin{aligned}
&E^{(half)}_{2h_1-1,2h_1,2h_2-\delta}(\gamma)\\
&=E^{(half)}_{2h_1-1,2h_1,2h_2-\delta}\bigg(T;\mathbf{y}_{\gamma(j)},\mathbf{u}_{\gamma(j)}:j\in\mathcal{I}(h_1);\\
&\quad\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:k\in\mathcal{I}\left(h_2-\frac{\delta}{2};1/2\right);\gamma\bigg)\\
&=\frac{1}{T}\sum_{j_1=(h_1-1)M+1}^{(h_1-\frac{1}{2})M}\ \sum_{j_2=(h_1-\frac{1}{2})M+1}^{h_1M}\ \sum_{k=(h_2-\frac{1+\delta}{2})M+1}^{(h_2-\frac{\delta}{2})M}A_{\gamma(j_1),\gamma(j_2),\gamma(k)}(T)\\
&=\frac{1}{T}\int_0^T Z^{(half)}_{2h_1-1}(\gamma;t)Z^{(half)}_{2h_1}(\gamma;t)Z^{(half)}_{2h_2-\delta}(\gamma;t)\,dt,
\end{aligned}\qquad(11.7)$$

where

$$\begin{aligned}
Z^{(half)}_{2h-1}(\gamma;t)&=Z^{(half)}_{2h-1}(\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:k\in\mathcal{I}(h);\gamma;t)\\
&=\sum_{k=(h-1)M+1}^{(h-\frac{1}{2})M}\chi_A(\mathbf{x}_{\gamma(k)}(t))
\end{aligned}\qquad(11.8)$$

and

$$\begin{aligned}
Z^{(half)}_{2h}(\gamma;t)&=Z^{(half)}_{2h}(\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}:k\in\mathcal{I}(h;1/2);\gamma;t)\\
&=\sum_{k=(h-\frac{1}{2})M+1}^{hM}\chi_A(\mathbf{x}_{\gamma(k)}(t)).
\end{aligned}\qquad(11.9)$$

Of course, we can similarly define

$$E^{(half)}_{2h_1-\delta,2h_2-1,2h_2}(\gamma),\qquad(11.10)$$

where $1\le h_1<h_2\le m$ and $\delta=0$ or 1; and also the quadruple version

$$E^{(half)}_{2h_1-1,2h_1,2h_2-1,2h_2}(\gamma).\qquad(11.11)$$

Applying Lemma 9.1 with $M = M/2$ and $\ell = 3$, for every $1 \le h_1 < h_2 \le m$ and $\delta = 0$ or 1 we have

$$(4\pi)^{-N} \int_{I^3} \cdots \int_{I^3} \int_{S^2} \cdots \int_{S^2} \frac{1}{N!} \sum_{\gamma \in \Gamma}$$

$$\left(E^{(half)}_{2h_1-1,2h_1,2h_2-\delta}\Big(T; \mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \right.$$

$$\left. \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}\Big(h_2 - \frac{\delta}{2}; 1/2\Big); \gamma \Big) - \frac{M^3}{8} \cdot \mathrm{vol}^3(A) \right)^2 d\mathbf{y}_1 \ldots d\mathbf{y}_N \, d\mathbf{u}_1 \ldots d\mathbf{u}_N$$

$$\le \frac{2}{\pi v T} \cdot \left(\frac{M^3}{8} \cdot \mathrm{vol}^3(A) + 3\frac{M^4}{16} \cdot \mathrm{vol}^4(A) + 3\frac{M^5}{32} \cdot \mathrm{vol}^5(A) \right)$$

$$= \frac{2}{\pi v T} \cdot \left(\frac{M}{2} \mathrm{vol}(A) \right)^6 \left(\left(\frac{2}{M \cdot \mathrm{vol}(A)} + 1 \right)^3 - 1 \right). \tag{11.12}$$

Of course, we have a perfect analog of (11.12) for (11.10).

And also, we have a similar result for (11.11) which goes as follows. Applying Lemma 9.1 with $M = M/2$ and $\ell = 4$, for every $1 \le h_1 < h_2 \le m$ we have

$$(4\pi)^{-N} \int_{I^3} \cdots \int_{I^3} \int_{S^2} \cdots \int_{S^2} \frac{1}{N!} \sum_{\gamma \in \Gamma}$$

$$\left(E^{(half)}_{2h_1-1,2h_1,2h_2-1,2h_2}(T; \mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \right.$$

$$\left. \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h_2); \gamma) - \frac{M^4}{16} \cdot \mathrm{vol}^4(A) \right)^2 d\mathbf{y}_1 \ldots d\mathbf{y}_N \, d\mathbf{u}_1 \ldots d\mathbf{u}_N$$

$$\le \frac{2}{\pi v T} \cdot \left(\frac{M^4}{16} \cdot \mathrm{vol}^4(A) + 4\frac{M^5}{32} \cdot \mathrm{vol}^5(A) + 6\frac{M^6}{64} \cdot \mathrm{vol}^6(A) + 4\frac{M^7}{128} \cdot \mathrm{vol}^7(A) \right)$$

$$= \frac{2}{\pi v T} \cdot \left(\frac{M}{2} \mathrm{vol}(A) \right)^8 \left(\left(\frac{2}{M \cdot \mathrm{vol}(A)} + 1 \right)^4 - 1 \right). \tag{11.13}$$

Let $0 < \varepsilon < 1$; it follows from (11.12)–(11.13) by a standard average argument (similar to (10.7)–(10.8)) that, there is a measurable subset $\Omega_4(bad)$ of Ω such that, for all initial conditions

$$(\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \Omega \setminus \Omega_4(bad), \tag{11.14}$$

we have that

$$\frac{1}{N!} \sum_{\gamma \in \Gamma} \left(E^{(half)}_{2h_1-1,2h_1,2h_2-\delta}\Big(T; \mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \right.$$

$$\left. \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}\Big(h_2 - \frac{\delta}{2}; 1/2\Big) M; \gamma \Big) - \frac{M^3}{8} \cdot \mathrm{vol}^3(A) \right)^2$$

$$\le \frac{8m^2}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(\frac{M}{2} \mathrm{vol}(A) \right)^6 \left(\left(\frac{2}{M \cdot \mathrm{vol}(A)} + 1 \right)^3 - 1 \right) \tag{11.15}$$

holds for all $1 \le h_1 < h_2 \le m$ and $\delta = 0$ or 1, and similarly,

$$\frac{1}{N!}\sum_{\gamma\in\Gamma}\Bigg(E^{(half)}_{2h_1-\delta,2h_2-1,2h_2}\Big(T;\Big(\mathbf{y}_{\gamma(j)},\mathbf{u}_{\gamma(j)}: j\in\mathcal{I}\Big(h_1-\frac{\delta}{2};1/2\Big);$$

$$\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}: k\in\mathcal{I}(h_2);\gamma\Big)-\frac{M^3}{8}\cdot\mathrm{vol}^3(A)\Bigg)^2$$

$$\le\frac{8m^2}{\varepsilon}\cdot\frac{2}{\pi vT}\cdot\left(\frac{M}{2}\mathrm{vol}(A)\right)^6\left(\left(\frac{2}{M\cdot\mathrm{vol}(A)}+1\right)^3-1\right) \tag{11.16}$$

holds for all $1 \le h_1 < h_2 \le m$ and $\delta = 0$ or 1, and similarly,

$$\frac{1}{N!}\sum_{\gamma\in\Gamma}\Bigg(E^{(half)}_{2h_1-1,2h_1,2h_2-1,2h_2}(T;\mathbf{y}_{\gamma(j)},\mathbf{u}_{\gamma(j)}: j\in\mathcal{I}(h_1);$$

$$\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}: k\in\mathcal{I}(h_2);\gamma)-\frac{M^4}{16}\cdot\mathrm{vol}^4(A)\Bigg)^2$$

$$\le\frac{8m^2}{\varepsilon}\cdot\frac{2}{\pi vT}\cdot\left(\frac{M}{2}\mathrm{vol}(A)\right)^8\left(\left(\frac{2}{M\cdot\mathrm{vol}(A)}+1\right)^4-1\right) \tag{11.17}$$

holds for all $1 \le h_1 < h_2 \le m$, and finally, $\Omega_4(bad)$ is "negligible":

$$\text{measure}\,(\Omega_4(bad)) < \frac{\varepsilon}{8}\text{measure}(\Omega). \tag{11.18}$$

Again $\Omega_4(bad)$ represents a "bad" subset of Ω that we will throw away at the end.

Consider a "good" initial condition

$$(\mathbf{y}_1,\ldots,\mathbf{y}_N,\mathbf{u}_1,\ldots,\mathbf{u}_N)\in\Omega\setminus(\Omega_3(bad)\cup\Omega_4(bad))\,. \tag{11.19}$$

It means that we can use (11.5)–(11.6) and (11.15)–(11.18).

Let's return to (11.2): for any $1 \le h_1 < h_2 \le m$ the "time average" of the product

$$Z_{h_1}(\gamma;t)Z_{h_2}(\gamma;t),$$

as t runs in $0 < t < T$, equals

$$\frac{1}{T}\int_0^T Z_{h_1}(\gamma;t)Z_{h_2}(\gamma;t)\,dt$$

$$=E_{h_1,h_2}(\gamma)=E_{h_1,h_2}(\mathbf{y}_{\gamma(j)},\mathbf{u}_{\gamma(j)}: j\in\mathcal{I}(h_1);\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}: k\in\mathcal{I}(h_2);\gamma).$$

The values of $Z_h(\gamma;t)$, as t runs in $0 < t < T$, are non-negative integers $0, 1, 2, 3, \ldots$; now for every pair of non-negative integers (ℓ_1,ℓ_2) we define the set

$$W_{h_1,h_2}(\gamma;\ell_1;\ell_2)$$

$$=W_{h_1,h_2}(\mathbf{y}_{\gamma(j)},\mathbf{u}_{\gamma(j)}: j\in\mathcal{I}(h_1);\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)}: k\in\mathcal{I}(h_2);\gamma;\ell_1;\ell_2)$$

$$=\{t\in[0,T]: Z_{h_1}(\gamma;t)=\ell_1 \;\text{ and }\; Z_{h_2}(\gamma;t)=\ell_2\}. \tag{11.20}$$

Then we have the following disjoint decomposition of the interval $0 \le t \le T$:

$$[0, T] = \bigcup_{\ell_1=0}^{\infty} \bigcup_{\ell_2=0}^{\infty} W_{h_1,h_2}(\gamma; \ell_1; \ell_2).$$

Write

$$\begin{aligned} &V_{h_1,h_2}(\gamma; \ell_1; \ell_2) \\ &\quad = V_{h_1,h_2}(\mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h_2); \gamma; \ell_1; \ell_2) \\ &\quad = \frac{1}{T} \text{measure} \left(W_{h_1,h_2}(\gamma; \ell_1; \ell_2) \right), \end{aligned} \tag{11.21}$$

implying

$$0 \le V_{h_1,h_2}(\gamma; \ell_1; \ell_2) \le 1.$$

We need to give an upper bound for the size $V_{h_1,h_2}(\gamma; \ell_1; \ell_2)$ of a "typical" set $W_{h_1,h_2}(\gamma; \ell_1; \ell_2)$ with $\ell_1 \ell_2 \ge 2$ ("typical" means the majority of the initial conditions and the majority of the permutations). In fact, we will estimate whole power-of-two groups such as

$$\sum_{\ell_1=2^j}^{2^{j+1}-1} V_{h_1,h_2}(\gamma; \ell_1; 1) \tag{11.22}$$

and

$$\sum_{\ell_2=2^j}^{2^{j+1}-1} V_{h_1,h_2}(\gamma; 1; \ell_2) \tag{11.23}$$

for all integers $j \ge 1$, and also

$$\sum_{\ell_1=2^{j_1}}^{2^{j_1+1}-1} \sum_{\ell_2=2^{j_2}}^{2^{j_2+1}-1} V_{h_1,h_2}(\gamma; \ell_1; \ell_2) \tag{11.24}$$

for all integers $j_1 \ge 1$ and $j_2 \ge 1$. We will obtain an upper bound by using a second moment argument—similarly to what we did in Sect. 10.

For simplicity assume that M is even; let $h = h_1$ or h_2. Let $I_1 \cup I_2$ be an arbitrary *halving split* of the set $\{(h-1)M + 1, (h-1)M + 2, \ldots, hM\} = \mathcal{I}(h)$ of M consecutive integers into two disjoint subsets of size $M/2$ each. There are exactly $\binom{M}{M/2}$ such halving splits. For any fixed halving split (I_1, I_2), write

$$Z_{I_1}(\gamma; t) Z_{I_2}(\gamma; t) = \left(\sum_{k_1 \in I_1} \chi_A(\mathbf{x}_{\gamma(k_1)}(t)) \right) \left(\sum_{k_2 \in I_2} \chi_A(\mathbf{x}_{\gamma(k_2)}(t) \right).$$

First we discuss the case (11.24). Let $\ell_1 \ge 2$ and $\ell_2 \ge 2$. If

$$t_0 \in W_{h_1,h_2}(\gamma; \ell_1; \ell_2) \quad \Longleftrightarrow \quad Z_{h_1}(\gamma; t_0) = \ell_1 \quad \text{and} \quad Z_{h_2}(\gamma; t_0) = \ell_2$$

for some $0 \le t_0 \le T$, then at least *one-half* of the $\binom{M}{M/2}$ halving splits $(I_1^{(1)}, I_2^{(1)})$ of $\{(h_1-1)M+1, (h_1-1)M+2, \dots, h_1 M\} = \mathcal{I}(h_1)$ have the property that

$$\begin{aligned} &Z_{I_1^{(1)}}(\gamma; t_0) Z_{I_2^{(1)}}(\gamma; t_0) \ge 1 \quad \text{for } \ell_1 = 2, 3 \quad \text{and} \\ &Z_{I_1^{(1)}}(\gamma; t_0) Z_{I_2^{(1)}}(\gamma; t_0) > \frac{\ell_1^2 - 2\ell_1}{4} \quad \text{for } \ell_1 \ge 4, \end{aligned} \tag{11.25}$$

and similarly, at least *one-half* of the $\binom{M}{M/2}$ halving splits $(I_1^{(2)}, I_2^{(2)})$ of $\{(h_2-1)M+1, (h_2-1)M+2, \dots, h_2 M\} = \mathcal{I}(h_2)$ have the property that

$$\begin{aligned} &Z_{I_1^{(2)}}(\gamma; t_0) Z_{I_2^{(2)}}(\gamma; t_0) \ge 1 \quad \text{for } \ell_2 = 2, 3 \quad \text{and} \\ &Z_{I_1^{(2)}}(\gamma; t_0) Z_{I_2^{(2)}}(\gamma; t_0) > \frac{\ell_2^2 - 2\ell_2}{4} \quad \text{for } \ell_2 \ge 4. \end{aligned} \tag{11.26}$$

Indeed, just like in Sect. 8, we applied here the argument in (5.27)–(5.31).

Let $\mathcal{M}_1$ be an arbitrary M-element subset of the first N integers $\{1, 2, \dots, N\}$, and let $\mathcal{P}(h_1; \mathcal{M}_1) \subset \Gamma$ denote the set of all permutations $\gamma \in \Gamma$ such that

$$\{\gamma(j) : j \in \mathcal{I}(h_1)\} = \{\gamma(j) : (h_1-1)M+1 \le j \le h_1 M\} = \mathcal{M}_1, \tag{11.27}$$

and let $\mathcal{M}_2$ be an arbitrary M-element subset of $\{1, 2, \dots, N\}$ disjoint from $\mathcal{M}_1$, and let $\mathcal{P}(h_2; \mathcal{M}_2) \subset \Gamma$ denote the set of all permutations $\gamma \in \Gamma$ such that

$$\{\gamma(k) : k \in \mathcal{I}(h_2)\} = \{\gamma(k) : (h_2-1)M+1 \le k \le h_2 M\} = \mathcal{M}_2. \tag{11.28}$$

Note that

$$\begin{aligned} &V_{h_1,h_2}(\gamma; \ell_1; \ell_2) \\ &\quad = V_{h_1,h_2}(\mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h_2); \gamma; \ell_1; \ell_2) \end{aligned}$$

depends only on the whole class $\gamma \in \mathcal{P}(h_1; \mathcal{M}_1) \cap \mathcal{P}(h_2; \mathcal{M}_2)$, and we denote this common value with

$$\begin{aligned} &V_{h_1,h_2}(\mathcal{P}(h_1; \mathcal{M}_1); \mathcal{P}(h_2; \mathcal{M}_2); \ell_1; \ell_2) \\ &\quad = V_{h_1,h_2}(\mathbf{y}_{\gamma(j)}, \mathbf{u}_{\gamma(j)} : j \in \mathcal{I}(h_1); \mathbf{y}_{\gamma(k)}, \mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h_2); \\ &\qquad \mathcal{P}(h_1; \mathcal{M}_1); \mathcal{P}(h_2; \mathcal{M}_2); \ell_1; \ell_2). \end{aligned} \tag{11.29}$$

Again we use inequality (10.24): if $\alpha \ge 0$ and $\beta \ge 0$ are arbitrary positive real numbers, then

$$0 \le \alpha < 2\beta \quad \text{or} \quad (\alpha - \beta)^2 \ge \frac{1}{4}\alpha^2. \tag{11.30}$$

Let $j_1 \ge 1$ and $j_2 \ge 1$ be integers, and consider the power-of-two groups

$$\sum_{\ell_1 = 2^{j_1}}^{2^{j_1+1}-1} V_{h_1,h_2}(\gamma; \ell_1; 1), \qquad \sum_{\ell_2 = 2^{j_2}}^{2^{j_2+1}-1} V_{h_1,h_2}(\gamma; 1; \ell_2) \tag{11.31}$$

and

$$\sum_{\ell_1=2^{j_1}}^{2^{j_1+1}-1} \sum_{\ell_2=2^{j_2}}^{2^{j_2+1}-1} V_{h_1,h_2}(\gamma;\ell_1;\ell_2). \tag{11.32}$$

Combining (11.14)–(11.16), (11.25)–(11.26), inequality (11.30), and using notation (11.29), we obtain that

$$\frac{1}{\binom{N}{M}\binom{N-M}{M}} \sum^{***}_{\substack{\mathcal{M}_1\subset\{1,\ldots,N\}:\\ |\mathcal{M}_1|=M}} \sum^{***}_{\substack{\mathcal{M}_2\subset\{1,\ldots,N\}:\\ |\mathcal{M}_2|=M,\ |\mathcal{M}_1\cap\mathcal{M}_2|=0}}$$
$$\frac{1}{4}\left(\left(\sum_{2^{j_1}\le\ell_1<2^{j_1+1}} V_{h_1,h_2}(\mathcal{P}(h_1;\mathcal{M}_1);\mathcal{P}(h_2;\mathcal{M}_2);\ell_1;1)\right)\max\left\{\frac{(2^{j_1})^2-2\cdot 2^{j_1}}{4},1\right\}\right)^2$$
$$\le 2\cdot\frac{8m^2}{\varepsilon}\cdot\frac{2}{\pi v T}\cdot\left(\frac{M}{2}\mathrm{vol}(A)\right)^6\left(\left(\frac{2}{M\cdot\mathrm{vol}(A)}+1\right)^3-1\right) \tag{11.33}$$

holds for all $1\le h_1<h_2\le m$ and for all $j_1\ge 1$ (see the first sum in (11.31)), and similarly,

$$\frac{1}{\binom{N}{M}\binom{N-M}{M}} \sum^{***}_{\substack{\mathcal{M}_1\subset\{1,\ldots,N\}:\\ |\mathcal{M}_1|=M}} \sum^{***}_{\substack{\mathcal{M}_2\subset\{1,\ldots,N\}:\\ |\mathcal{M}_2|=M,\ |\mathcal{M}_1\cap\mathcal{M}_2|=0}}$$
$$\frac{1}{4}\left(\left(\sum_{2^{j_2}\le\ell_2<2^{j_2+1}} V_{h_1,h_2}(\mathcal{P}(h_1;\mathcal{M}_1);\mathcal{P}(h_1;\mathcal{M}_2);1;\ell_2)\right)\max\left\{\frac{(2^{j_2})^2-2\cdot 2^{j_2}}{4},1\right\}\right)^2$$
$$\le 2\cdot\frac{8m^2}{\varepsilon}\cdot\frac{2}{\pi v T}\cdot\left(\frac{M}{2}\mathrm{vol}(A)\right)^6\left(\left(\frac{2}{M\cdot\mathrm{vol}(A)}+1\right)^3-1\right) \tag{11.34}$$

holds for all $1\le h_1<h_2\le m$ and for all $j_2\ge 1$ (see the second sum in (11.31)), and finally,

$$\frac{1}{\binom{N}{M}\binom{N-M}{M}} \sum^{***}_{\substack{\mathcal{M}_1\subset\{1,\ldots,N\}:\\ |\mathcal{M}_1|=M}} \sum^{***}_{\substack{\mathcal{M}_2\subset\{1,\ldots,N\}:\\ |\mathcal{M}_2|=M,\ |\mathcal{M}_1\cap\mathcal{M}_2|=0}}$$
$$\frac{1}{4}\left(\left(\sum_{2^{j_1}\le\ell_1<2^{j_1+1}}\ \sum_{2^{j_2}\le\ell_2<2^{j_2+1}} V_{h_1,h_2}(\mathcal{P}(h_1;\mathcal{M}_1);\mathcal{P}(h_2;\mathcal{M}_2);\ell_1;\ell_2)\right)\right.$$
$$\left.\cdot\max\left\{\frac{(2^{j_1})^2-2\cdot 2^{j_1}}{4},1\right\}\cdot\max\left\{\frac{(2^{j_2})^2-2\cdot 2^{j_2}}{4},1\right\}\right)^2$$
$$\le 4\cdot\frac{8m^2}{\varepsilon}\cdot\frac{2}{\pi v T}\cdot\left(\frac{M}{2}\mathrm{vol}(A)\right)^8\left(\left(\frac{2}{M\cdot\mathrm{vol}(A)}+1\right)^4-1\right) \tag{11.35}$$

holds for all $1 \leq h_1 < h_2 \leq m$ and for all $j_1 \geq 1$, $j_2 \geq 1$ (see (11.32)). Note that the triple asterisk *** in (11.33) means that the summation is restricted to the terms satisfying

$$\left(\left(\sum_{2^{j_1} \leq \ell_1 < 2^{j_1+1}} V_{h_1,h_2}(\mathcal{P}(h_1;\mathcal{M}_1);\mathcal{P}(h_2;\mathcal{M}_2);\ell_1;1)\right)\max\left\{\frac{(2^{j_1})^2 - 2\cdot 2^{j_1}}{4},1\right\}\right)^2$$
$$\geq 2\cdot\frac{M^3}{8}\cdot \mathrm{vol}^3(A), \tag{11.36}$$

see (11.30). Also the first factor "2" in the last line of (11.33) comes from the *one-half* before (11.25).

Similarly, the triple asterisk *** in (11.34) means that the summation is restricted to the terms satisfying

$$\left(\left(\sum_{2^{j_2} \leq \ell_2 < 2^{j_2+1}} V_{h_1,h_2}(\mathcal{P}(h_1;\mathcal{M}_1);\mathcal{P}(h_2;\mathcal{M}_2);1;\ell_2)\right)\max\left\{\frac{(2^{j_2})^2 - 2\cdot 2^{j_2}}{4},1\right\}\right)^2$$
$$\geq 2\cdot\frac{M^3}{8}\cdot \mathrm{vol}^3(A), \tag{11.37}$$

and finally, the triple asterisk *** in (11.35) means that the summation is restricted to the terms satisfying

$$\left(\left(\sum_{2^{j_1} \leq \ell_1 < 2^{j_1+1}}\ \sum_{2^{j_2} \leq \ell_2 < 2^{j_2+1}} V_{h_1,h_2}(\mathcal{P}(h_1;\mathcal{M}_1);\mathcal{P}(h_2;\mathcal{M}_2);\ell_1;\ell_2)\right)\right.$$
$$\left.\cdot\max\left\{\frac{(2^{j_1})^2 - 2\cdot 2^{j_1}}{4},1\right\}\cdot\max\left\{\frac{(2^{j_2})^2 - 2\cdot 2^{j_2}}{4},1\right\}\right)^2 \geq 2\cdot\frac{M^4}{16}\cdot\mathrm{vol}^4(A). \tag{11.38}$$

Note that the first factor "4" in the last line of (11.35) comes from the product of the *one-half* before (11.25) and the other *one-half* before (11.26).

Next we divide both sides of (11.33) with the factor

$$\max\left\{\left(\frac{(2^j)^2 - 2\cdot 2^j}{4}\right)^2,1\right\} = \max\left\{2^{2j-2}\left(2^{j-1}-1\right)^2,1\right\} \tag{11.39}$$

where $j = j_1$, divide both sides of (11.34) with the factor (11.39) where $j = j_2$, and divide both sides of (11.35) with the product

$$\max\left\{2^{2j_1-2}\left(2^{j_1-1}-1\right)^2,1\right\}\cdot\max\left\{2^{2j_2-2}\left(2^{j_2-1}-1\right)^2,1\right\}.$$

Then by using the standard *average argument*, we obtain that, there is a subset $\Gamma_3(bad)$ of Γ such that, for all permutations

$$\gamma \in \Gamma \setminus \Gamma_3(bad) \tag{11.40}$$

and all integers $1 \leq h_1 < h_2 \leq m$ and $j_1 \geq 1$, $j_2 \geq 1$ we have:

$$\sum_{2^{j_1} \leq \ell_1 < 2^{j_1+1}} V_{h_1,h_2}(\gamma;\ell_1;1)$$

$$\leq \frac{4m^2 \cdot j_1}{\max\{2^{j_1-1}(2^{j_1-1}-1), 1\}} \cdot C(11.33)$$

$$+ \frac{2}{\max\{2^{j_1-1}(2^{j_1-1}-1), 1\}} \cdot \frac{M^3}{8} \cdot \mathrm{vol}^3(A), \tag{11.41}$$

where the extra factor of j_1 in the numerator right before $C(11.33)$ comes from the convergent series $\sum_{j\geq 1} j^{-2} < 2$, the factor $C(11.33)$ itself is defined as

$$C(11.33) = \left(2 \cdot \frac{8m^2}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(\frac{M}{2}\mathrm{vol}(A)\right)^6 \left(\left(\frac{2}{M \cdot \mathrm{vol}(A)} + 1\right)^3 - 1\right)\right)^{1/2}, \tag{11.42}$$

and the last term in (11.41) comes from (11.36); similarly, we have:

$$\sum_{2^{j_2} \leq \ell_1 < 2^{j_2+1}} V_{h_1,h_2}(\gamma; 1; \ell_2)$$

$$\leq \frac{4m^2 \cdot j_2}{\max\{2^{j_2-1}(2^{j_2-1}-1), 1\}} \cdot C(11.33)$$

$$+ \frac{2}{\max\{2^{j_2-1}(2^{j_2-1}-1), 1\}} \cdot \frac{M^3}{8} \cdot \mathrm{vol}^3(A), \tag{11.43}$$

which is a perfect analog of (11.41); moreover, we have:

$$\sum_{2^{j_1} \leq \ell_1 < 2^{j_1+1}} \sum_{2^{j_2} \leq \ell_2 < 2^{j_2+1}} V_{h_1,h_2}(\gamma; \ell_1; \ell_2)$$

$$\leq \frac{4m^2 \cdot j_1 j_2}{\max\{\frac{(2^{j_1})^2 - 2\cdot 2^{j_1}}{4}, 1\} \cdot \max\{\frac{(2^{j_2})^2 - 2\cdot 2^{j_2}}{4}, 1\}} \cdot C(11.35)$$

$$+ \frac{2}{\max\{\frac{(2^{j_1})^2 - 2\cdot 2^{j_1}}{4}, 1\} \cdot \max\{\frac{(2^{j_2})^2 - 2\cdot 2^{j_2}}{4}, 1\}} \cdot \frac{M^4}{16} \cdot \mathrm{vol}^4(A), \tag{11.44}$$

where

$$C(11.35) = \left(4 \cdot \frac{8m^2}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(\frac{M}{2}\mathrm{vol}(A)\right)^8 \left(\left(\frac{2}{M \cdot \mathrm{vol}(A)} + 1\right)^4 - 1\right)\right)^{1/2}, \tag{11.45}$$

and finally,

$$\frac{1}{N!}|\Gamma_3(bad)| < \frac{1}{16} \tag{11.46}$$

(where, as usual, $|\cdots|$ denotes the number of elements of a finite set).

Since we assumed a "good" initial condition (see (11.19)):

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \Omega \setminus (\Omega_3(bad) \cup \Omega_4(bad)), \tag{11.47}$$

we can use (11.5)–(11.6) and (11.15)–(11.18). We already used (11.15)–(11.17) in (11.33); now we use (11.5). The standard average argument gives that, there is a subset $\Gamma_4(bad)$ of

Γ (= the set of $N!$ permutations of $1, 2, \ldots, N$) such that, for all permutations

$$\gamma \in \Gamma \setminus \Gamma_4(bad) \tag{11.48}$$

and all integers $1 \le h_1 < h_2 \le m$, we have

$$|E_{h_1,h_2}(\omega;\gamma) - M^2 \cdot \mathrm{vol}^2(A)| \le 4m \cdot C(11.5), \tag{11.49}$$

and

$$\frac{1}{N!}|\Gamma_4(bad)| < \frac{1}{16}, \tag{11.50}$$

where

$$C(11.5) = \left(\frac{4m^2}{\varepsilon} \cdot \frac{2}{\pi v T} \cdot \left(M^2 \cdot \mathrm{vol}^2(A) + 2M^3 \cdot \mathrm{vol}^3(A)\right)\right)^{1/2}. \tag{11.51}$$

Next we turn to (11.19)–(11.21). We have

$$\begin{aligned} E_{h_1,h_2}(\omega;\gamma) &= \frac{1}{T}\int_0^T Z_{h_1}(\omega;\gamma;t) Z_{h_2}(\omega;\gamma;t)\,dt \\ &= \sum_{\ell_1=1}^{\infty}\sum_{\ell_2=1}^{\infty} V_{h_1,h_2}(\omega;\gamma;\ell_1;\ell_2)\cdot \ell_1\ell_2 \\ &= V_{h_1,h_2}(\omega;\gamma;1;1) + \sum_{j=1}^{\infty}\sum_{\ell_1=2^j}^{2^{j+1}-1} V_{h_1,h_2}(\omega;\gamma;\ell_1;1)\cdot\ell_1 \\ &\quad + \sum_{j=1}^{\infty}\sum_{\ell_2=2^j}^{2^{j+1}-1} V_{h_1,h_2}(\omega;\gamma;1;\ell_2)\cdot\ell_2 \\ &\quad + \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\sum_{\ell_1=2^{j_1}}^{2^{j_1+1}-1}\sum_{\ell_2=2^{j_2}}^{2^{j_2+1}-1} V_{h_1,h_2}(\omega;\gamma;\ell_1;\ell_2)\cdot\ell_1\ell_2. \end{aligned} \tag{11.52}$$

Let

$$\gamma \in \Gamma \setminus (\Gamma_3(bad) \cup \Gamma_4(bad)), \tag{11.53}$$

then by (11.49) and (11.52),

$$\begin{aligned} &|V_{h_1,h_2}(\omega;\gamma;1;1) - M^2 \cdot \mathrm{vol}^2(A)| \\ &\quad \le 4m \cdot C(11.5) + \sum_{j=1}^{\infty}\left(\sum_{\ell_1=2^j}^{2^{j+1}-1} V_{h_1,h_2}(\omega;\gamma;\ell_1;1)\right)\cdot 2^{j+1} \\ &\qquad + \sum_{j=1}^{\infty}\left(\sum_{\ell_2=2^j}^{2^{j+1}-1} V_{h_1,h_2}(\omega;\gamma;1;\ell_2)\right)\cdot 2^{j+1} \end{aligned}$$

$$+\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\left(\sum_{\ell_1=2^{j_1}}^{2^{j_1+1}-1}\sum_{\ell_2=2^{j_2}}^{2^{j_2+1}-1} V_{h_1,h_2}(\omega;\gamma;\ell_1;\ell_2)\right)\cdot 2^{j_1+1}\cdot 2^{j_2+1}. \tag{11.54}$$

Applying (11.41), (11.43) and (11.44), we have

$$\begin{aligned}
&|V_{h_1,h_2}(\omega;\gamma;1;1) - M^2\cdot \mathrm{vol}^2(A)|\\
&\quad\le 4m\cdot C(11.5) + C(11.33)\cdot\sum_{j=1}^{\infty}\frac{8m^2\cdot 2^{j+1}\cdot j}{\max\{2^{j-1}(2^{j-1}-1),1\}}\\
&\qquad + \frac{M^3}{8}\cdot \mathrm{vol}^3(A)\cdot\sum_{j=1}^{\infty}\frac{4\cdot 2^{j+1}}{\max\{2^{j-1}(2^{j-1}-1),1\}}\\
&\qquad + C(11.35)\cdot\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\frac{4m^2\cdot 2^{j_1+1}\cdot 2^{j_2+1}\cdot j_1 j_2}{\max\{\frac{(2^{j_1})^2-2\cdot 2^{j_1}}{4},1\}\cdot\max\{\frac{(2^{j_2})^2-2\cdot 2^{j_2}}{4},1\}}\\
&\qquad + \frac{M^4}{16}\cdot \mathrm{vol}^4(A)\cdot\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\frac{2\cdot 2^{j_1+1}\cdot 2^{j_2+1}}{\max\{\frac{(2^{j_1})^2-2\cdot 2^{j_1}}{4},1\}\cdot\max\{\frac{(2^{j_2})^2-2\cdot 2^{j_2}}{4},1\}}
\end{aligned} \tag{11.55}$$

for every $1\le h_1<h_2\le m$.

Let

$$\mathrm{vol}(A)=\frac{\lambda}{N}. \tag{11.56}$$

Since $m=N/M$, we have $M\cdot\mathrm{vol}(A)=M\cdot\lambda/N=\lambda/m$. Using this in (11.55), and estimating the infinite series, we obtain

$$\begin{aligned}
&\left|V_{h_1,h_2}(\omega;\gamma;1;1)-\left(\frac{\lambda}{m}\right)^2\right|\\
&\quad\le\frac{1}{\sqrt{vT}}\cdot\frac{15}{\sqrt{\varepsilon}}\cdot\lambda m\left(1+2\frac{\lambda}{m}\right)^{1/2}+\frac{1}{\sqrt{vT}}\cdot\frac{800}{\sqrt{\varepsilon}}\cdot(\lambda m)^{3/2}\\
&\qquad+\frac{1}{\sqrt{vT}}\cdot\frac{10^4}{\sqrt{\varepsilon}}\cdot\lambda^2 m+6\left(\frac{\lambda}{m}\right)^3+30\left(\frac{\lambda}{m}\right)^4.
\end{aligned} \tag{11.57}$$

We will also need the following estimate, which can be proved exactly the same way as (11.57):

$$\begin{aligned}
V_{h_1,h_2}(\omega;\gamma;>(1,1)) &= \sum_{j=1}^{\infty}\sum_{\ell_1=2^j}^{2^{j+1}-1}V_{h_1,h_2}(\omega;\gamma;\ell_1;1)\\
&\quad+\sum_{j=1}^{\infty}\sum_{\ell_2=2^j}^{2^{j+1}-1}V_{h_1,h_2}(\omega;\gamma;1;\ell_2)\\
&\quad+\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\sum_{\ell_1=2^{j_1}}^{2^{j_1+1}-1}\sum_{\ell_2=2^{j_2}}^{2^{j_2+1}-1}V_{h_1,h_2}(\omega;\gamma;\ell_1;\ell_2)
\end{aligned}$$

$$\leq \frac{1}{\sqrt{vT}} \cdot \frac{100}{\sqrt{\varepsilon}} \cdot (\lambda m)^{3/2} + \left(\frac{\lambda}{m}\right)^3 + \frac{1}{\sqrt{vT}} \cdot \frac{80}{\sqrt{\varepsilon}} \cdot \lambda^2 m + \left(\frac{\lambda}{m}\right)^4. \tag{11.58}$$

Therefore, we have just proved (see (11.56)–(11.58))

Lemma 11.1 *Assume that the initial condition satisfies the requirement*

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \Omega \setminus (\Omega_3(bad) \cup \Omega_4(bad)),$$

and also, the permutation γ satisfies

$$\gamma \in \Gamma \setminus (\Gamma_3(bad) \cup \Gamma_4(bad)),$$

then for every $1 \leq h_1 < h_2 \leq m$ we have

$$\begin{aligned} &\left| V_{h_1,h_2}(\omega; \gamma; 1; 1) - \left(\frac{\lambda}{m}\right)^2 \right| \\ &\leq 6\left(\frac{\lambda}{m}\right)^3 + 30\left(\frac{\lambda}{m}\right)^4 \\ &\quad + \frac{1}{\sqrt{vT}} \cdot \frac{1}{\sqrt{\varepsilon}} \left(15\lambda m \left(1 + 2\frac{\lambda}{m}\right)^{1/2} + 800(\lambda m)^{3/2} + 10^4 \lambda^2 m\right), \end{aligned} \tag{11.59}$$

and

$$\begin{aligned} &V_{h_1,h_2}(\omega; \gamma; > (1,1)) \\ &= \sum_{\ell_1=1}^{\infty} \sum_{\substack{1 \leq \ell_2 < \infty: \\ \ell_1 \ell_2 > 1}} V_{h_1,h_2}(\omega; \gamma; \ell_1; \ell_2) \\ &\leq \left(\frac{\lambda}{m}\right)^3 + \left(\frac{\lambda}{m}\right)^4 + \frac{1}{\sqrt{vT}} \cdot \frac{1}{\sqrt{\varepsilon}} \left(100(\lambda m)^{3/2} + 80\lambda^2 m\right), \end{aligned} \tag{11.60}$$

where

$$\operatorname{vol}(A) = \frac{\lambda}{N}.$$

Finally, note that

$$\text{measure}(\Omega_3(bad)) < \frac{\varepsilon}{8}\text{measure}(\Omega), \qquad \text{measure}(\Omega_4(bad)) < \frac{\varepsilon}{8}\text{measure}(\Omega),$$

and similarly, $|\Gamma| = N!$ and

$$|\Gamma_3(bad)| < \frac{N!}{16}, \qquad |\Gamma_4(bad)| < \frac{N!}{16}.$$

The message of Lemma 11.1 is the following: if vT is "large", then the terms in the last line of (11.59) and (11.60) are negligible compared to $(\lambda/m)^3$, and so we have

$$\begin{aligned} V_{h_1,h_2}(\omega;\gamma;1;1) &= \left(\frac{\lambda}{m}\right)^2 + O\left(\left(\frac{\lambda}{m}\right)^3\right) \quad \text{and} \\ V_{h_1.h_2}(\omega;\gamma;>(1,1)) &= O\left(\left(\frac{\lambda}{m}\right)^3\right). \end{aligned} \tag{11.61}$$

12 Completing the Proof of Theorem 1

Notice that Lemma 11.1 at the end of the last section is a generalization of Lemma 10.1. Applying the same method—namely, a repeated application of Lemma 9.1, combined with a standard average argument—we can easily prove the following more general result; see Lemma 12.1 below. We start with some notation.

Let r be an arbitrary integer in the range $2 \le r \le m = N/M$ (note that $r = 1$ is covered by Lemma 10.1, and $r = 2$ is Lemma 11.1; here we treat the general case). Let $1 \le k_1 < k_2 < \cdots < k_r \le N$ be arbitrary integers and, as usual, let

$$A_{k_1,\ldots,k_r}(T) = A_{k_1,\ldots,k_r}(T;\mathbf{y}_{k_i},\mathbf{u}_{k_i} : 1 \le i \le r)$$

denote the total time between $0 < t < T$ when the r torus-lines $\mathbf{x}_{k_i}(t)$, $1 \le i \le r$ are all in subset A simultaneously:

$$\begin{aligned} A_{k_1,\ldots,k_r}(T) &= A_{k_1,\ldots,k_r}(T;\mathbf{y}_{k_i},\mathbf{u}_{k_i} : 1 \le i \le r) \\ &= \text{measure}\left\{t \in [0,T] : \mathbf{x}_{k_i}(t) \in A \ (\text{mod } 1) \text{ for all } 1 \le i \le r\right\} \\ &= \int_0^T \chi_A(\mathbf{x}_{k_1}(t)) \cdots \chi_A(\mathbf{x}_{k_r}(t))\,dt. \end{aligned} \tag{12.1}$$

Let $\gamma \in \Gamma$ be an arbitrary permutation of $1, 2, \ldots, N$, and for any sequence $1 \le h_1 < h_2 < \cdots < h_r \le m = N/M$ write

$$\begin{aligned} &E_{h_1,\ldots,h_r}(\gamma) \\ &= E_{h_1,\ldots,h_r}(\mathbf{y}_{\gamma(k_1)},\mathbf{u}_{\gamma(k_1)} : k_1 \in \mathcal{I}(h_1);\ldots;\mathbf{y}_{\gamma(k_r)},\mathbf{u}_{\gamma(k_r)} : k_r \in \mathcal{I}(h_r);\gamma) \\ &= \sum_{k_1=(h_1-1)M+1}^{h_1 M} \cdots \sum_{k_r=(h_r-1)M+1}^{h_r M} \frac{1}{T} A_{\gamma(k_1),\ldots,\gamma(k_r)}(T) \\ &= \frac{1}{T}\int_0^T Z_{h_1}(\gamma;t)\cdots Z_{h_r}(\gamma;t)\,dt, \end{aligned} \tag{12.2}$$

where we used the short notation (h is an integer)

$$\mathcal{I}(h) = \{(h-1)M+1, (h-1)M+2, \ldots, hM\},$$

and also

$$Z_h(\gamma;t) = Z_h(\mathbf{y}_{\gamma(k)},\mathbf{u}_{\gamma(k)} : k \in \mathcal{I}(h);\gamma;t)$$

$$= \sum_{k=(h-1)M+1}^{hM} \chi_A(\mathbf{x}_{\gamma(k)}(t)).$$

The values of $Z_h(\gamma; t)$, as t runs in $0 < t < T$, are non-negative integers $0, 1, 2, 3, \dots$; now for every sequence of non-negative integers $(\ell_1, \dots, \ell_r)$ we define the set

$$\begin{aligned}
&W_{h_1,\dots,h_r}(\gamma; \ell_1; \dots; \ell_r)\\
&\quad = W_{h_1,\dots,h_r}(\mathbf{y}_{\gamma(k_1)}, \mathbf{u}_{\gamma(k_1)} : k_1 \in \mathcal{I}(h_1); \dots; \mathbf{y}_{\gamma(k_r)}, \mathbf{u}_{\gamma(k_r)} : k_r \in \mathcal{I}(h_r); \gamma; \ell_1; \dots; \ell_r)\\
&\quad = \{t \in [0, T] : Z_{h_i}(\gamma; t) = \ell_i \ \text{ for all } i = 1, \dots, r\}.
\end{aligned} \tag{12.3}$$

Then we have the following disjoint decomposition of the interval $0 \le t \le T$:

$$[0, T] = \bigcup_{\ell_1=0}^{\infty} \bigcup_{\ell_2=0}^{\infty} W_{h_1,\dots,h_r}(\gamma; \ell_1; \dots; \ell_r).$$

Write

$$\begin{aligned}
&V_{h_1,\dots,h_r}(\gamma; \ell_1; \dots; \ell_r)\\
&\quad = V_{h_1,\dots,h_r}(\mathbf{y}_{\gamma(k_1)}, \mathbf{u}_{\gamma(k_1)} : k_1 \in \mathcal{I}(h_1); \dots; \mathbf{y}_{\gamma(k_r)}, \mathbf{u}_{\gamma(k_r)} : k_r \in \mathcal{I}(h_r); \gamma; \ell_1; \dots; \ell_r)\\
&\quad = \frac{1}{T} \text{measure} \left(W_{h_1,\dots,h_r}(\gamma; \ell_1; \dots; \ell_r) \right),
\end{aligned} \tag{12.4}$$

implying

$$0 \le V_{h_1,\dots,h_r}(\gamma; \ell_1; \dots; \ell_r) \le 1.$$

Write

$$V_{h_1,\dots,h_r}(\gamma; > (1, \dots, 1)) = \sum_{\substack{\ell_i \ge 1,\ 1 \le i \le r: \\ \ell_1 \cdots \ell_r > 1}} V_{h_1,\dots,h_r}(\gamma; \ell_1; \dots; \ell_r).$$

Now we are ready to formulate a generalization of Lemma 11.1 (which represents the case $r = 2$) for arbitrary $r \ge 2$.

Lemma 12.1 *For every integer $r \ge 2$ there is a subset $\Omega^{(r)}(bad)$ of Ω with*

$$\text{measure}(\Omega^{(r)}(bad)) < \frac{1}{4(r-1)^2} \cdot \text{measure}(\Omega), \tag{12.5}$$

and there is a subset $\Gamma^{(r)}(bad)$ of Γ with

$$|\Gamma^{(r)}(bad)| < \frac{N!}{8(r-1)^2} \tag{12.6}$$

such that, for any initial condition

$$\omega = (\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N) \in \Omega \setminus \Omega^{(r)}(bad)$$

and for any permutation γ satisfying

$$\gamma \in \Gamma \setminus \Gamma^{(r)}(bad),$$

we have for all $1 \le h_1 < \cdots < h_r \le m$:

$$\left| V_{h_1,\ldots,h_r}(\omega;\gamma;1;\ldots;1) - \left(\frac{\lambda}{m}\right)^r \right| \\ \le 10\left(\frac{\lambda}{m}\right)^{r+1} + \frac{1}{\sqrt{vT}} \cdot \frac{10^{4r}}{\sqrt{\varepsilon}} \left((\lambda m)^{r-\frac{1}{2}} + \lambda^r \cdot m^{r-1}\right), \tag{12.7}$$

and

$$V_{h_1,\ldots,h_r}(\omega;\gamma;>(1,\ldots,1)) \\ \le 2\left(\frac{\lambda}{m}\right)^{r+1} + \frac{1}{\sqrt{vT}} \cdot \frac{10^{2r}}{\sqrt{\varepsilon}} \left((\lambda m)^{r-\frac{1}{2}} + \lambda^r \cdot m^{r-1}\right), \tag{12.8}$$

where

$$\mathrm{vol}(A) = \frac{\lambda}{N}.$$

The choice $\frac{1}{4(r-1)^2}$ in (12.5)–(12.6) is motivated by the numerical fact

$$\frac{1}{2} + \sum_{r=2}^{\infty} \frac{1}{4(r-1)^2} < 1. \tag{12.9}$$

By using Lemma 12.1 for $r \ge 2$ and Lemma 10.1 for $r = 1$, we can easily execute the proof plan outlined at the end of Sect. 5.

Let $Y_A(t)$ denote the point-counting function:

$$Y_A(;t) = Y_A(\omega;t) = \sum_{\substack{1 \le k \le N: \\ \mathbf{x}_k(t) \in A}} 1,$$

where the N torus lines $\mathbf{x}_k(t)$, $1 \le k \le N$ are determined by a given initial condition

$$\omega = (\mathbf{y}_1,\ldots,\mathbf{y}_N,\mathbf{u}_1,\ldots,\mathbf{u}_N) \in \Omega.$$

For typical initial conditions we are able to describe the distribution of

$$Y_A(t) = Y_A(\omega;t) = Y_A(\mathbf{y}_1,\ldots,\mathbf{y}_N,\mathbf{u}_1,\ldots,\mathbf{u}_N;t)$$

as t runs in $0 \le t \le T$. We begin with value 0, that is, we describe the density of $Y_A(\omega;t) = 0$ as $0 \le t \le T$. Note that

$$\{0 \le t \le T : Y_A(\omega;t) = 0\} = \bigcap_{h=1}^{m} W_h(\omega;\gamma;0)$$

holds for any permutation $\gamma \in \Gamma$. We recall that

$$W_h(\omega;\gamma;\ge 1) = \{t \in [0,T] : Z_h(\omega;\gamma;t) \ge 1\}$$

and

$$W_{h_1,\ldots,h_r}(\omega;\gamma;\ge 1) = W_{h_1,\ldots,h_r}(\omega;\gamma;1;\ldots;1) \cup W_{h_1,\ldots,h_r}(\omega;\gamma;>(1,\ldots,1))$$

$$= \left\{ t \in [0,T] : \prod_{i=1}^{r} Z_{h_i}(\omega;\gamma;t) \geq 1 \right\} = \bigcap_{i=1}^{r} W_{h_i}(\omega;\gamma;\geq 1).$$

Next we return to (5.32), but instead of using the complete inclusion-exclusion equality, we apply the following *truncated inequality*.

Lemma 12.2 ("Brunn's sieve") *Let $(\Omega_0, \mathcal{F}_0, \mu_0)$ be any probability space (i.e., $\mathcal{F}_0$ is a σ-algebra of the subsets of Ω_0 and μ_0 is a σ-additive measure on $\mathcal{F}_0$), let $B_1, B_2, \ldots, B_m \in \mathcal{F}_0$ be arbitrary events, and let $1 \leq q < m$ be an arbitrary odd integer. Then*

$$\begin{aligned}
&1 - \sum_{i=1}^{m} \mu_0(B_i) + \sum_{1 \leq i_1 < i_2 \leq m} \mu_0(B_{i_1} \cap B_{i_2}) \mp \cdots - \sum_{1 \leq i_1 < \cdots < i_q \leq m} \mu_0(B_{i_1} \cap \cdots \cap B_{i_q}) \\
&\quad \leq \mu_0\left(\bigcap_{i=1}^{m} B_i\right) \\
&\quad \leq 1 - \sum_{i=1}^{m} \mu_0(B_i) + \sum_{1 \leq i_1 < i_2 \leq m} \mu_0(B_{i_1} \cap B_{i_2}) \mp \cdots + \sum_{1 \leq i_1 < \cdots < i_{q+1} \leq m} \mu_0(B_{i_1} \cap \cdots \cap B_{i_{q+1}}).
\end{aligned}$$

Remark This inequality, well-known to number-theorists as the "Brunn's sieve", and also known as the Bonferroni inequalities, has a simple proof that is based on the fact that the partial sums of the alternating series

$$\binom{m}{0} - \binom{m}{1} + \binom{m}{2} - \binom{m}{3} \pm$$

are alternating in sign:

$$\binom{m}{0} - \binom{m}{1} + \binom{m}{2} \mp \cdots + (-1)^q \binom{m}{q} = (-1)^q \binom{m-1}{q}.$$

The plan is to combine Lemma 12.1 (and Lemma 10.1 for $r = 1$) with Lemma 12.2. Choose an initial condition

$$\omega = (\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N) \in \Omega \setminus \left(\bigcup_{r=1}^{q} \Omega^{(r)}(bad) \right) \tag{12.10}$$

and a permutation

$$\gamma \in \Gamma \setminus \left(\bigcup_{r=1}^{q} \Gamma^{(r)}(bad) \right). \tag{12.11}$$

In view of Lemma 12.1 (and Lemma 10.1 for $r = 1$), more than $1 - \varepsilon$ part of Ω is available in (12.10), and more than half of Γ is available in (12.11).

By Lemma 12.2, we have with $Y_A(\omega;t) = Y_A(\mathbf{y}_1, \ldots, \mathbf{y}_N, \mathbf{u}_1, \ldots, \mathbf{u}_N; t)$ (μ denotes the one-dimensional Lebesgue measure):

$$\frac{1}{T}\mu\{0 \leq t \leq T : Y_A(\omega;t) = 0\}$$

$$= \frac{1}{T}\mu\left(\bigcap_{h=1}^{m} W_h(\omega;0)\right)$$

$$\geq 1 - \sum_{h=1}^{m} V_h(\omega;\gamma;\geq 1) + \sum_{1\leq h_1<h_2\leq m} V_{h_1,h_2}(\omega;\gamma;\geq 1) \mp \cdots$$

$$- \sum_{1\leq h_1<\cdots<h_q\leq m} V_{h_1,\ldots,h_q}(\omega;\gamma;\geq 1), \tag{12.12}$$

and similarly, we have the other inequality:

$$\frac{1}{T}\mu\{0\leq t\leq T : Y_A(\omega;t)=0\}$$

$$= \frac{1}{T}\mu\left(\bigcap_{h=1}^{m} W_h(\omega;0)\right)$$

$$\leq 1 - \sum_{h=1}^{m} V_h(\omega;\gamma;\geq 1) + \sum_{1\leq h_1<h_2\leq m} V_{h_1,h_2}(\omega;\gamma;\geq 1) \mp \cdots$$

$$+ \sum_{1\leq h_1<\cdots<h_{q+1}\leq m} V_{h_1,\ldots,h_{q+1}}(\omega;\gamma;\geq 1), \tag{12.13}$$

where $1\leq q<m$ is any odd integer. We will specify the optimal value of parameter q very soon.

Let $1\leq r\leq q$; since

$$V_{h_1,\ldots,h_r}(\omega;\gamma;\geq 1) = V_{h_1,\ldots,h_r}(\omega;\gamma;1;\ldots;1) + V_{h_1,\ldots,h_r}(\omega;\gamma;>(1,\ldots,1)),$$

by Lemma 12.1 (and Lemma 10.1 for $r=1$) we have:

$$\left| V_{h_1,\ldots,h_r}(\omega;\gamma;\geq 1) - \left(\frac{\lambda}{m}\right)^r \right|$$

$$\leq 12\left(\frac{\lambda}{m}\right)^{r+1} + \frac{1}{\sqrt{\upsilon T}}\cdot\frac{10^{4r}+10^{2r}}{\sqrt{\varepsilon}}\left((\lambda m)^{r-\frac{1}{2}} + \lambda^r\cdot m^{r-1}\right). \tag{12.14}$$

Therefore, we obtain

$$\left| \frac{1}{T}\mu\{0\leq t\leq T : Y_A(\omega;t)=0\} - 1 + m\frac{\lambda}{m} - \binom{m}{2}\left(\frac{\lambda}{m}\right)^2 \right.$$

$$\left. + \binom{m}{3}\left(\frac{\lambda}{m}\right)^3 \mp \cdots + \binom{m}{q}\left(\frac{\lambda}{m}\right)^q \right|$$

$$\leq \sum_{r=1}^{q+1} 12\binom{m}{r}\left(\frac{\lambda}{m}\right)^{r+1}$$

$$+ \sum_{r=1}^{q+1}\binom{m}{r}\frac{1}{\sqrt{\upsilon T}}\cdot\frac{10^{4r}+10^{2r}}{\sqrt{\varepsilon}}\left((\lambda m)^{r-\frac{1}{2}} + \lambda^r\cdot m^{r-1}\right). \tag{12.15}$$

First we estimate the inner sum in (12.15), and show that it is close to the Taylor series of $e^{-\lambda}$:

$$
\begin{aligned}
&1 - m\frac{\lambda}{m} + \binom{m}{2}\left(\frac{\lambda}{m}\right)^2 - \binom{m}{3}\left(\frac{\lambda}{m}\right)^3 \pm \cdots - \binom{m}{q}\left(\frac{\lambda}{m}\right)^q \\
&= \sum_{r=0}^{q} (-1)^r \frac{\lambda^r}{r!} + \sum_{r=0}^{q} (-1)^r \frac{\lambda^r}{r!}\left(\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{r-1}{m}\right) - 1\right) \\
&= e^{-\lambda} - \sum_{r=q+1}^{\infty} (-1)^r \frac{\lambda^r}{r!} \\
&\quad + \sum_{r=0}^{q} (-1)^r \frac{\lambda^r}{r!}\left(\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{r-1}{m}\right) - 1\right).
\end{aligned} \tag{12.16}
$$

Applying (12.16) in (12.15), we have

$$
\begin{aligned}
&\left|\frac{1}{T}\mu\{0 \le t \le T : Y_A(\omega; t) = 0\} - e^{-\lambda}\right| \\
&\quad \le \sum_{r=q+1}^{\infty} (-1)^r \frac{\lambda^r}{r!} \\
&\quad\quad + \sum_{r=0}^{q} (-1)^r \frac{\lambda^r}{r!}\left(\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{r-1}{m}\right) - 1\right) \\
&\quad\quad + \sum_{r=1}^{q+1} 12\binom{m}{r}\left(\frac{\lambda}{m}\right)^{r+1} \\
&\quad\quad + \sum_{r=1}^{q+1} \binom{m}{r}\frac{1}{\sqrt{vT}} \cdot \frac{10^{4r} + 10^{2r}}{\sqrt{\varepsilon}}\left((\lambda m)^{r-\frac{1}{2}} + \lambda^r \cdot m^{r-1}\right).
\end{aligned} \tag{12.17}
$$

Using the well-known facts

$$
j! \ge \left(\frac{j}{e}\right)^j \quad \text{and} \quad 1 - \frac{r}{m} \ge e^{-2r/m} \quad \text{for } 1 \le r \le m/4,
$$

we have

$$
\sum_{r=q+1}^{\infty} \frac{\lambda^r}{r!} \le \sum_{r=q+1}^{\infty} \left(\frac{e\lambda}{r}\right)^r \tag{12.18}
$$

and

$$
\left(\sum_{r=0}^{q} \frac{\lambda^r}{r!}\right)\left|\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{r-1}{m}\right) - 1\right| \le e^{\lambda}\left(1 - e^{-q^2/m}\right). \tag{12.19}
$$

Moreover, we have

$$
12\sum_{r=1}^{q+1} \binom{m}{r}\left(\frac{\lambda}{m}\right)^{r+1} \le \frac{12\lambda}{m}\sum_{r=1}^{q+1} \frac{\lambda^r}{r!} \le \frac{12\lambda \cdot e^{\lambda}}{m}, \tag{12.20}
$$

and

$$\frac{1}{\sqrt{vT}}\sum_{r=1}^{q+1}\binom{m}{r}\frac{10^{4r}+10^{2r}}{\sqrt{\varepsilon}}\left((\lambda m)^{r-\frac{1}{2}}+\lambda^r\cdot m^{r-1}\right)$$

$$\leq \frac{1}{\sqrt{vT}}\cdot\frac{2}{\sqrt{\varepsilon}}\sum_{r=1}^{q+1}\frac{\lambda^r}{r!}\cdot(101m)^{2q+2}\leq\frac{1}{\sqrt{vT}}\cdot\frac{2}{\sqrt{\varepsilon}}\cdot e^{\lambda}\cdot(101m)^{2q+2}. \quad (12.21)$$

Using (12.18)–(12.21) in (12.17), we obtain

$$\left|\frac{1}{T}\mu\{0\leq t\leq T: Y_A(\omega;t)=0\}-e^{-\lambda}\right|$$

$$\leq\sum_{r=q+1}^{\infty}\left(\frac{e\lambda}{r}\right)^r+e^{\lambda}\left(1-e^{-q^2/m}\right)+\frac{12\lambda\cdot e^{\lambda}}{m}+\frac{1}{\sqrt{vT}}\cdot\frac{2}{\sqrt{\varepsilon}}\cdot e^{\lambda}\cdot(101m)^{2q+2}.$$

(12.22)

Now we are ready to specify the values of the key parameters m and q (where $1\leq q=$ odd $<m<N$): let

$$m=\min\left\{\frac{e^{\frac{1}{2}\sqrt{\log(vT)}}}{101},\sqrt{N}\right\}\quad\text{and}\quad q=\log m, \quad (12.23)$$

where log denotes the natural (i.e., base e) logarithm. Note that m and q ($=$ odd) are integers, so in (12.23) we actually take the nearest integers.

We have

$$(101m)^{2q+2}\leq\frac{e^{\frac{1}{2}\log(vT)}}{m^2}=\frac{\sqrt{vT}}{m^2},$$

$$1-e^{-q^2/m}\leq 2\frac{q^2}{m}=2\frac{(\log m)^2}{m},$$

and

$$\sum_{r=q+1}^{\infty}\left(\frac{e\lambda}{r}\right)^r\leq e^{-q}=\frac{1}{m}\quad\text{if } 0<\lambda\leq\frac{\log m}{e^2}.$$

Using these facts in (12.22), we conclude that

$$\left|\frac{1}{T}\mu\{0\leq t\leq T: Y_A(\omega;t)=0\}-e^{-\lambda}\right|$$

$$\leq\frac{1}{m}+\frac{2e^{\lambda}(\log m)^2}{m}+\frac{12\lambda\cdot e^{\lambda}}{m}+\frac{2}{\sqrt{\varepsilon}}\cdot\frac{e^{\lambda}}{m^2}$$

$$\leq\frac{1}{m}+\frac{2m^{e^{-2}}(\log m)^2}{m}+\frac{12\lambda\cdot m^{e^{-2}}}{m}+\frac{2}{\sqrt{\varepsilon}}\cdot\frac{m^{e^{-2}}}{m^2}$$

$$\leq\frac{1}{m^{3/4}}+\frac{2}{\sqrt{\varepsilon}}\cdot\frac{1}{m} \quad (12.24)$$

if

$$0 < \lambda \le \frac{\log m}{e^2}. \tag{12.25}$$

This basically proves the special case $k = 0$ in (1.10).

Next let $k \ge 1$ be an arbitrary positive integer, and we describe the density of the point-counting function $Y_A(\omega; t) = k$ as $0 \le t \le T$. We basically repeat the argument of the special case $k = 0$ above with some natural modifications.

Our starting point is the following inequality, which can be considered as an analog of (5.32): for any $1 \le k \le m$ we have

$$\begin{aligned}
&\Bigg| \frac{1}{T}\mu\{0 \le t \le T : Y_A(\omega; t) = k\} \\
&\quad - \sum_{1\le j_1<\cdots<j_k\le m} \mu\Bigg(\bigcap_{i=1}^{k} W_{j_i}(\omega;\gamma;1) \cap \bigcap_{\substack{1\le h\le m:\\ h\ne j_i,\ 1\le i\le k}} W_h(\omega;\gamma;0)\Bigg)\Bigg| \\
&\le \sum_{h=1}^{m} \mu\left(W_h(\omega;\gamma;\ge 2)\right).
\end{aligned} \tag{12.26}$$

For notational convenience, for any fixed sequence $1 \le j_1 < \cdots < j_k \le m$ of length k we write

$$\mathcal{J} = \{j_1, \ldots, j_k\} \quad \text{and} \quad W^*(\mathcal{J}) = \bigcap_{i=1}^{k} W_{j_i}(\omega;\gamma;1).$$

By using Brunn's sieve (Lemma 12.2), we have

$$\begin{aligned}
&\mu\Bigg(\bigcap_{i=1}^{k} W_{j_i}(\omega;\gamma;1) \cap \bigcap_{\substack{1\le h\le m:\\ h\ne j_i,\ 1\le i\le k}} W_h(\omega;\gamma;0)\Bigg) \\
&\ge \mu\big(W^*(\mathcal{J})\big) - \sum_{\substack{1\le h\le m:\\ h\notin\mathcal{J}}} \mu\big(W^*(\mathcal{J}) \cap W_h(\omega;\gamma;\ge 1)\big) \\
&\quad + \sum_{\substack{1\le h_1<h_2\le m:\\ h_i\notin\mathcal{J},\ 1\le i\le 2}} \mu\big(W^*(\mathcal{J}) \cap W_{h_1}(\omega;\gamma;\ge 1) \cap W_{h_2}(\omega;\gamma;\ge 1)\big) \mp \cdots \\
&\quad + (-1)^{q-k} \sum_{\substack{1\le h_1<\cdots<h_{q-k}\le m:\\ h_i\notin\mathcal{J},\ 1\le i\le q-k}} \mu\big(W^*(\mathcal{J}) \cap W_{h_1}(\omega;\gamma;\ge 1) \cap \cdots \\
&\quad \cap W_{h_{q-k}}(\omega;\gamma;\ge 1)\big),
\end{aligned} \tag{12.27}$$

assuming $(q - k)$ is *odd*, and similarly, under the same condition, we have the other inequality:

$$\mu\Bigg(\bigcap_{i=1}^{k} W_{j_i}(\omega;\gamma;1) \cap \bigcap_{\substack{1\le h\le m:\\ h\ne j_i,\ 1\le i\le k}} W_h(\omega;\gamma;0)\Bigg)$$

$$\le \mu\left(W^*(\mathcal{J})\right) - \sum_{\substack{1\le h\le m:\\ h\notin\mathcal{J}}} \mu\left(W^*(\mathcal{J})\cap W_h(\omega;\gamma;\ge 1)\right)$$

$$+ \sum_{\substack{1\le h_1<h_2\le m:\\ h_i\notin\mathcal{J},\ 1\le i\le 2}} \mu\left(W^*(\mathcal{J})\cap W_{h_1}(\omega;\gamma;\ge 1)\cap W_{h_2}(\omega;\gamma;\ge 1)\right) \mp \cdots$$

$$+ (-1)^{q-k+1} \sum_{\substack{1\le h_1<\cdots<h_{q-k+1}\le m:\\ h_i\notin\mathcal{J},\ 1\le i\le q-k+1}} \mu\Big(W^*(\mathcal{J})\cap W_{h_1}(\omega;\gamma;\ge 1)\cap\cdots$$

$$\cap W_{h_{q-k+1}}(\omega;\gamma;\ge 1)\Big). \tag{12.28}$$

If $(q-k)$ is *even*, then of course in (12.27)–(12.28) we have to make a switch between $\ge$ and $\le$.

Notice that (12.27)–(12.28) are analogs of (12.12)–(12.13). By repeating the arguments after (12.13), we obtain the following analog of (12.15): for any $1\le k\le q$ we have

$$\left|\frac{1}{T}\mu\{0\le t\le T: Y_A(\omega;t)=k\} - \binom{m}{k}\left(\frac{\lambda}{m}\right)^k \right.$$

$$\left. + \binom{m}{k}\binom{m-k}{1}\left(\frac{\lambda}{m}\right)^{k+1} \mp\cdots + (-1)^{q-k-1}\binom{m}{k}\binom{m-k}{q-k}\left(\frac{\lambda}{m}\right)^q\right|$$

$$\le m\left(\frac{\lambda}{m}\right)^2 + \sum_{r=1}^{q+1} 12\binom{m}{r}\left(\frac{\lambda}{m}\right)^{r+1}$$

$$+ \sum_{r=1}^{q+1}\binom{m}{r}\frac{1}{\sqrt{vT}}\cdot\frac{10^{4r}+10^{2r}}{\sqrt{\varepsilon}}\left((\lambda m)^{r-\frac12}+\lambda^r\cdot m^{r-1}\right), \tag{12.29}$$

where the term $m(\frac{\lambda}{m})^2$ comes from the contribution of the last line in (12.26).

Similarly to (12.15), first we estimate the inner sum in (12.29):

$$\sum(*) = \binom{m}{k}\left(\frac{\lambda}{m}\right)^k - \binom{m}{k}\binom{m-k}{1}\left(\frac{\lambda}{m}\right)^{k+1} \pm\cdots$$

$$+ (-1)^{q-k}\binom{m}{k}\binom{m-k}{q-k}\left(\frac{\lambda}{m}\right)^q$$

$$= \binom{m}{k}\left(\frac{\lambda}{m}\right)^k\left(1-\left(1-\frac{k}{m}\right)\lambda + \left(1-\frac{k}{m}\right)\left(1-\frac{k+1}{m}\right)\frac{\lambda^2}{2!}\right.$$

$$-\left(1-\frac{k}{m}\right)\left(1-\frac{k+1}{m}\right)\left(1-\frac{k+2}{m}\right)\frac{\lambda^3}{3!}\pm\cdots$$

$$\left.+(-1)^{q-k}\left(1-\frac{k}{m}\right)\cdots\left(1-\frac{q-1}{m}\right)\frac{\lambda^{q-k}}{(q-k)!}\right). \tag{12.30}$$

Using the arguments in (12.18)–(12.19), we have

$$\left|\sum(*) - \frac{\lambda^k}{k!}e^{-\lambda}\right| \le \frac{e^{2\lambda}q^2}{m}. \tag{12.31}$$

Repeating the arguments in (12.20)–(12.24), by (12.29)–(12.31) we obtain the following analog of (12.24): for every k in $0 \le k \le q = \log m$,

$$\begin{aligned}&\left|\frac{1}{T}\mu\{0 \le t \le T : Y_A(\omega;t) = k\} - \frac{\lambda^k}{k!}e^{-\lambda}\right|\\ &\quad \le \frac{e^{2\lambda}q^2}{m} + \frac{1}{m^{3/4}} + \frac{2}{\sqrt{\varepsilon}}\cdot\frac{1}{m} < \frac{2}{m^{3/4}} + \frac{2}{\sqrt{\varepsilon}}\cdot\frac{1}{m}\end{aligned} \tag{12.32}$$

if

$$0 < \lambda \le \frac{\log m}{e^2}. \tag{12.33}$$

Summarizing, by (12.23)–(12.25) and (12.32) we have that, with

$$m = \min\left\{\frac{e^{\frac{1}{2}\sqrt{\log(vT)}}}{101}, \sqrt{N}\right\} \quad \text{and} \quad \varepsilon = \frac{1}{\sqrt{m}},$$

for more than $1-\varepsilon$ part of the initial conditions

$$\omega = (\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{u}_1, \dots, \mathbf{u}_N) \in \Omega,$$

the estimation

$$\left|\frac{1}{T}\mu\{0 \le t \le T : Y_A(\omega;t) = k\} - \frac{\lambda^k}{k!}e^{-\lambda}\right| < \frac{1}{\sqrt{m}} \tag{12.34}$$

holds for all integers k in $0 \le k \le q = \log m$.

Note that the remaining case $k > \log m$ is trivial, since then the main term becomes much smaller than the error term. Indeed, taking the sum of (12.32) for $k = 0, 1, \dots, q$, and subtracting it from 1, we obtain

$$\begin{aligned}&\frac{1}{T}\mu\{0 \le t \le T : Y_A(\omega;t) > q = \log m\}\\ &\quad \le \left(1 - \sum_{k=0}^{q}\frac{\lambda^k}{k!}e^{-\lambda}\right) + \frac{2q+2}{m^{3/4}} + \frac{2q+2}{\sqrt{\varepsilon}}\cdot\frac{1}{m}\\ &\quad < e^{-\lambda}\sum_{k>\log m}\frac{\lambda^k}{k!} + \frac{2\log m + 2}{m^{3/4}} + \frac{2\log m+2}{\sqrt{\varepsilon}}\cdot\frac{1}{m} < \frac{1}{m^{2/3}},\end{aligned} \tag{12.35}$$

where in the last line we used the upper bound (12.33) for λ. Combining (12.34)–(12.35), (1.10) follows.

The proof of the product formula (1.11) is a straightforward adaptation of the proof of (1.10) above. The obvious difference is that we work with the Cartesian product $A_1 \times \cdots \times A_r$ instead of A, and consequently, we begin the proof with an application of Lemma 9.1 instead of an application of Lemma 5.1. Apart from this, the rest of the argument is basically the same. This completes the proof of Theorem 1.

Concluding Remark: How to Sequentialize Theorem 1? Here we briefly outline how to modify the proof of Theorem 1 to obtain (1.12)–(1.13). The basic idea is very simple. Instead of working with a fixed time interval $[0,T]$ (what we did in Theorem 1), we have to work with *all* intervals $[0,T']$ simultaneously, where T' runs through all integers in the range (1.12) (i.e., $2<T<e^{N^{1/8}}$) with "short" binary representation

$$T' = 2^{n_1} + 2^{n_2} + 2^{n_3} + \cdots + 2^{n_p} \tag{12.36}$$

where $p \ge 1$, $n_1 > n_2 > n_3 > \cdots > n_p \ge n_1 - \ell(n_1)$. The only new idea in the proof is how to specify the value of the new parameter $\ell = \ell(n_1)$. Note that, if $\ell = \ell(n_1)$ is "small" compared to n_1, then, for fixed n_1, there are relatively few integers T' of the short binary form (12.36)—indeed, the exact number is $2^\ell = 2^{\ell(n_1)}$. On the other hand, we want the set of integers of the form (12.36) to be relatively "dense", so $\ell = \ell(n_1)$ cannot be too small.

I recall (1.12) and the line after:

$$m(T) = \min\left\{\frac{e^{\frac{1}{2}\sqrt{\log(vT)}}}{101}, \sqrt{N}\right\} \quad \text{and} \quad \varepsilon(T) = \frac{1}{\sqrt{m(T)}}$$

if $2<T<e^{N^{1/8}}$. Motivated by this, the best compromise for $\ell = \ell(n_1)$ is the following: let $\ell = \ell(n_1)$ be in the range of $\sqrt{n_1}$ if

$$2 < T' < N^{\log N} = e^{(\log N)^2},$$

and let $\ell = \ell(n_1)$ be in the range of $\log N$ if

$$N^{\log N} \le T' < e^{N^{1/8}}.$$

With this choice, there are relatively few integers T' of the short binary form (12.36), and, at the same time, the set of integers of the form (12.36) is relatively "dense" in the range (1.12) (i.e., in the interval $[2, e^{N^{1/8}}]$). This is how we obtain, by a straightforward adaptation of the proof of Theorem 1, the sequential result (1.13).

Also, we can easily extend Theorem 1 to large families of time intervals $[T_1, T_2]$ in the range $0 \le T_1 < T_2 < e^{N^{1/8}}$. That is, the starting point T_1 can also be a variable. The only new idea that we need here is to involve Liouville's classical theorem: the time-flow of a Hamiltonian system preserves the standard volume (Lebesgue measure) in the phase space. (Or as the physicists like to put it: the "ensemble fluid" moves as if it were an incompressible fluid.) We need Liouville's theorem for the following reason. Theorem 1 (and any other key result in this paper) is a statement about "$1-\varepsilon$ part of the initial conditions", and because of Liouville's theorem, it doesn't matter that we choose the time-point $t=0$ or any other $t=T_1>0$. The time-flow is measure-preserving, so "$1-\varepsilon$ part" at $t=0$ remains "$1-\varepsilon$ part" later at any other time-point $t=T_1>0$.

References

1. Beck, J.: Probabilistic diophantine approximation, I. Kronecker sequences. Ann. Math. **140**, 451–502 (1994)
2. Chernov, N., Makarian, R.: Chaotic Billiards. Mathematical Surveys and Monographs, vol. 127. American Mathematical Society, Providence (2006)
3. Cornfeld, I.P., Fomin, S.V., Sinai, Ya.G.: Ergodic Theory. Springer, New York (1981)

4. Egerváry, E., Turán, P.: On a certain point of the kinetic theory of gases. Stud. Math. **12**, 170–180 (1951)
5. Hardy, G.H., Wright, E.M.: An Introduction to the Theory of Numbers, 5th edn. Clarendon, Oxford (1979)
6. Hlawka, E.: Mathematische Modelle der kinetischen Gastheorie. In: Gruber, P.M., Schmidt, W.M. (eds.) Edmund Hlawka Selecta, pp. 361–375. Springer, Berlin (2000) (the original paper was published in 1974)
7. Kac, M.: Statistical Independence in Probability, Analysis and Number Theory. The Carus Mathematical Monographs, vol. 12. The Mathematical Association of America, Washington (1959)
8. Khinchine, A.: Eins Satz über Kettenbrüche, mit arithmetischen Anwendungen. Math. Z. **18**, 289–306 (1923)
9. Khinchine, A.: Continued Fractions. Groningen, Noordhoff (1963), English translation
10. Prosser, R.T.: Spectral analysis of classical central force motion. J. Math. Phys. **10**(12), 2233–2239 (1969)
11. Simányi, N.: Proof of the Boltzmann–Sinai ergodic hypothesis for typical hard disk systems. Invent. Math. **154**, 123–178 (2003)
12. Simányi, N.: Proof of the ergodic hypothesis for typical hard ball systems. Ann. Inst. Henri Poincare **5**, 203–233 (2004)
13. Sinai, Ya.G.: Introduction to Ergodic Theory. Mathematical Notes. Princeton University Press, Princeton (1976)
14. Sinai, Ya.G., Volkovsky, K.L.: Ergodic properties of an ideal gas with infinitely many degrees of freedom. Funct. Anal. Appl. **5**(4), 19–21 (1971)
15. Steinhaus, H.: Sur les fonctions indépendantes (VII). Stud. Math. **10**, 1–20 (1948)
16. Steinhaus, H.: Sur les fonctions indépendantes (X) (Equipartition de molécules dans un récipient cubique). Stud. Math. **13**, 1–17 (1953)
17. Thompson, C.J.: Mathematical Statistical Mechanics. The Macmillan Company, London (1972)
18. Tolman, R.C.: The Principles of Statistical Mechanics. Oxford University Press, London (1938) (reprinted several times since)
19. Uhlenbeck, G.E., Ford, G.W.: Lectures in Statistical Mechanics. American Mathematical Society, Providence (1963)