



# Quantitative control of Wasserstein distance between Brownian motion and the Goldstein–Kac telegraph process

Gerardo Barrera<sup>a</sup> and Jani Lukkarinen<sup>b</sup>

*Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68, Pietari Kalmin katu 5, FI-00014, Helsinki, Finland,*

<sup>a</sup>*gerardo.barreravargas@helsinki.fi,* <sup>b</sup>*jani.lukkarinen@helsinki.fi*

Received 27 February 2022; revised 19 May 2022; accepted 30 May 2022

**Abstract.** In this manuscript, we provide a non-asymptotic process level control between the telegraph process and the Brownian motion with suitable diffusivity constant via a Wasserstein distance with quadratic average cost. In addition, we derive non-asymptotic estimates for the corresponding time average  $p$ -th moments. The proof relies on coupling techniques such as coin-flip coupling, synchronous coupling and the Komlós–Major–Tusnády coupling.

**Résumé.** Dans cet article, nous fournissons un contrôle au niveau de processus et non asymptotique entre le processus télégraphique et le mouvement brownien avec une constante de diffusivité appropriée par rapport à la distance de Wasserstein et avec un coût moyen quadratique. De plus, nous dérivons des estimations non asymptotiques pour les  $p$ -ièmes moments correspondants. La preuve repose sur des techniques de couplage telles que le couplage pile ou face, le couplage synchrone et le couplage Komlós–Major–Tusnády.

*MSC2020 subject classifications:* Primary 60G50; 60J65; 60J99; 60K35; secondary 35L99; 60K37; 60K40

*Keywords:* Brownian motion; Coin-flip coupling; Decoupling; Free velocity flip model; Komlós–Major–Tusnády coupling; Random evolutions; Synchronous coupling; Telegraph process; Wasserstein distance

## Contents

1. Introduction . . . . .	934
2. The setting and the main results . . . . .	935
2.1. Construction of the free velocity flip model . . . . .	935
2.2. Related works . . . . .	936
2.3. Comparing two stochastic processes . . . . .	937
2.3.1. Coupling and Wasserstein metric . . . . .	937
2.3.2. Global estimate for moments . . . . .	937
2.4. Main result and consequences . . . . .	938
3. Outline of the proof of Theorem 2.1 . . . . .	940
3.1. Definition of the process $Y$ . . . . .	941
3.2. Definition of the process $Z$ . . . . .	942
4. Proof of Proposition 3.1: Coin-flip coupling . . . . .	943
4.1. Expectation for continuous observables of the free velocity flip model . . . . .	943
4.2. Coin flip type coupling . . . . .	945
4.3. Comparison between $L^{-1}\mathbb{X}_{[0,T]}^{\delta}$ and $\mathbb{Y}_{[0,T]}$ . . . . .	946
5. Proof of Proposition 3.2: Synchronous coupling . . . . .	953
6. Proof of Lemma 3.3: Komlós–Major–Tusnády coupling . . . . .	960
Appendix A: Moment estimates . . . . .	966
A.1. Poisson moment estimates . . . . .	966
A.2. $\delta$ -constraint moment estimates . . . . .	968
Appendix B: Free velocity flip moment estimates . . . . .	970
Appendix C: Definition and basic properties of the $\delta$ -constraints . . . . .	975
Appendix D: Tools . . . . .	980

Acknowledgements . . . . .	981
Funding . . . . .	981
CRedit authorship contribution statement . . . . .	981
Declaration of competing interest . . . . .	981
References . . . . .	981

## 1. Introduction

The so-called (Goldstein–Kac) telegraph process is perhaps the simplest example of a random evolution, see [21] and [24]. It describes the movement of a particle which starts at time zero from the origin and moves with a finite constant speed  $v_0$  on the line. The initial direction of the motion, positive or negative, is chosen at random with the same probability. The changes of direction are driven by a homogeneous Poisson process of a positive constant intensity  $\lambda$ . In other words, when a jump occurs in the Poisson process, the particle instantaneously takes the opposite direction and keeps moving with the same speed (it just flips the sign of its velocity) until the next jump in the Poisson process happens, then it takes the opposite direction again, and so on. This is in fact an example of a zigzag process considered in [3]. Later on, we provide a precise mathematical definition in (2.5). In [21] and [24], it is shown that the family of probability densities of the particle position  $(f(x, t) : x \in \mathbb{R}, t \geq 0)$  solves the hyperbolic partial differential equation of the type

$$(1.1) \quad \frac{\partial^2}{\partial t^2} f(x, t) + 2\lambda \frac{\partial}{\partial t} f(x, t) = v_0^2 \frac{\partial^2}{\partial x^2} f(x, t).$$

Since (1.1) appears in electrical wave propagation models, it is known as the telegraph equation or the hyperbolic heat equation. An explicit solution of (1.1) can be carried out in terms of special functions, see for instance Chapter 2 in [29].

The telegraph process has been proposed as an alternative to diffusion models and, as such, extensively studied by the probability and physics communities. Its generalisations are ubiquitous in applications, including transport phenomena in physical and biological systems, and it has produced a vast mathematics literature; standard references include [5, 12, 14, 15, 19–24, 26–30, 37, 40, 46] and further references may be found from therein. It is also used in the context of risk theory and to model financial markets, see [16, 29] and [38]. Statistics for the telegraph process are done in [22]. More recently, in [10] the authors compute the distribution of the maximum of the asymmetric telegraph process.

Despite its simplicity and the already existing results about the agreement with Brownian motion using the above marginals, it is not easy to connect these two processes on the level of realisations, i.e., on the process level. The present work is motivated by the need to understand a similar connection to diffusion in an extension of the telegraph process where the stochastic process is augmented into an interacting particle system, ordered along a circle, and each particle follows an independent telegraph process but there additionally is an interaction potential acting between the nearest neighbour particles. The resulting system is called a *velocity flip model*: the connection between them and suitably chosen diffusion processes have been explored, for instance, in [2, 34, 36, 45] but process level understanding of the connection is still lacking, even in the simplest case of a harmonic interaction potential.

Our aim is to provide a process level control between the telegraph process and the Brownian motion with suitable diffusivity constant via a Wasserstein distance with quadratic average cost. Often, one can rely on generators to see if a Markov process converges to another process, see for instance [18] and [42]. There are several ways to quantify the preceding convergence, for instance, by the Wasserstein distance (Kantorovich distance). For the setting of two pure jump Markov  $\mathbb{R}^d$ -valued processes with bounded intensity of jumps, the time evolution of the Wasserstein distance between them can be written as an integro-differential equation in terms of their generators and Kantorovich potentials, see for instance Theorem 3.1 in [1]. In dimension one, the latter remains true for piecewise deterministic Markov processes, see Theorem 6.4 in [1]. However, we do not follow this approach since the relevant Kantorovich potentials are hard to compute and also the process level control, in which we are interested, is with respect to an average quadratic cost function given in (2.10) below. This is in fact a natural distance between process; we refer to [4] for further details. For diffusions, under regular assumptions on their coefficients, by analytical approach the Wasserstein distance and total variation distance can be estimated, see Theorem 1.1 in [6]. By coupling approach, in [17] the authors obtain bounds of the total variation distance between two multidimensional Itô processes with different drifts. We point out that in general, it is difficult to obtain rates of convergence by the generator approach or for functional limit theorems. One possible approach is offered by the so-called Stein’s method, see [9, 11], although also this tool is difficult to implement in the present case.

In this manuscript we present non-asymptotic estimates between the telegraph process and the Brownian motion with suitable diffusivity constant. The main mathematical tools for the rigorous control of the estimates are coupling techniques. We do not require explicit knowledge of the marginals of the corresponding processes. Roughly speaking, the

idea is to compare the telegraph process with a decoupled process (perturbed random walk) via the coin-flip coupling (see [33] or Appendix A in [44]). Then we compare the decoupled process and the Brownian motion with suitable diffusivity constant. Since the decoupled process has zero mean at even jumps, the latter is done in two steps via the synchronous coupling and the celebrated Komlós–Major–Tusnády coupling (see [31] and [32]). The argument is detailed in Section 3. As a consequence of our main result, Theorem 2.1, we obtain non-asymptotic estimates for the time average of  $p$ -th moments.

The paper is organised as follows: Section 2 lays out the setting, the main result Theorem 2.1 and its consequences formulated as Corollary 2.7 and Corollary 2.9. Section 3 sketches the steps leading to the proof of the main result Theorem 2.1, which boils down to the proofs of Proposition 3.1, Proposition 3.2 and Lemma 3.3. Proposition 3.1 is proved in Section 4, Proposition 3.2 is shown in Section 5 and Lemma 3.3 is proved Section 6. Finally, there is an Appendix which collects main technical results and estimates used in the main text. It is divided in four subsections: Appendix A yields quantitative estimates of moments, Appendix B establishes coupling estimates between the free velocity flip model and a suitable Brownian motion, Appendix C gives integration formulas for the Dirac  $\delta$ -constrain probability measure, and Appendix D provides basic auxiliary results.

## 2. The setting and the main results

This section is divided in four subsections. We start by introducing the model and the notation in Section 2.1. Then, we discuss related works in Section 2.2. Next, we define the metric in which we compare our processes in Section 2.3. Finally, the main result and its consequences are given in Section 2.4.

### 2.1. Construction of the free velocity flip model

Let  $\mathbb{N} = \{1, 2, \dots\}$ ,  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$  and  $\mathbb{R}_* := (0, \infty)$ . For any  $\mathfrak{d} \in \mathbb{R}_*^{\mathbb{N}} := \{z = (z_1, z_2, \dots) : z_j \in \mathbb{R}_* \text{ for all } j \in \mathbb{N}\}$ ,  $\mathfrak{d} = (\delta_1, \delta_2, \dots)$ , we set  $t_0(\mathfrak{d}) := 0$  and

$$(2.1) \quad t_n(\mathfrak{d}) := \sum_{k=1}^n \delta_k \quad \text{for any } n \in \mathbb{N}.$$

We define the set of admissible sequences  $\Delta_\infty$  by

$$(2.2) \quad \Delta_\infty := \left\{ \mathfrak{d} \in \mathbb{R}_*^{\mathbb{N}} : \sum_{k=1}^{\infty} \delta_k = \infty \right\}.$$

In other words,  $\mathfrak{d} \in \Delta_\infty$  if and only if  $t_n(\mathfrak{d}) \rightarrow \infty$  as  $n \rightarrow \infty$ . For each  $\mathfrak{d} \in \Delta_\infty$  we set

$$(2.3) \quad \begin{aligned} N : \Delta_\infty &\longrightarrow \mathcal{D}([0, \infty), \mathbb{R}) \\ \mathfrak{d} &\longmapsto N(\cdot; \mathfrak{d}) : t \mapsto N(t; \mathfrak{d}) = \sup\{n \in \mathbb{N}_0 \mid t_n(\mathfrak{d}) \leq t\}, \end{aligned}$$

where  $\mathcal{D}([0, \infty), \mathbb{R})$  denote the set of Càdlàg paths defined on  $[0, \infty)$  and taking values on  $\mathbb{R}$ . We observe that the function  $t \mapsto N(t; \mathfrak{d})$  is non-decreasing.

The sequence  $\mathfrak{d}$  is taken to encode the waiting times between jumps for a certain piecewise linear path  $(X(t; \mathfrak{d}) : 0 \leq t < \infty)$ . We point out that then  $N(t; \mathfrak{d})$  corresponds to the number of jumps which have occurred by the time  $t$ . The process  $X$ , which can be identified with a “free velocity flip process path in one dimension”, is defined by the following properties:

1. Instantaneous position  $X(t; \mathfrak{d}) \in \mathbb{R}$ ,
2. starting velocity is  $v_0 \in \mathbb{R}$ , and
3. at each “jump” the velocity flips its sign,  $V(t; \mathfrak{d}) \mapsto -V(t; \mathfrak{d})$ . More precisely,

$$V(t; \mathfrak{d}) := v_0(-1)^{N(t; \mathfrak{d})} \quad \text{for all } t \geq 0.$$

For simplicity, we assume that the initial position is zero, that is,  $X(0; \mathfrak{d}) = 0$ . Then we explicitly define the free velocity flip path  $X$  as follows:

$$(2.4) \quad \begin{aligned} X : \Delta_\infty &\longrightarrow \mathcal{C}([0, \infty), \mathbb{R}) \\ \mathfrak{d} &\longmapsto X(\cdot; \mathfrak{d}) : t \mapsto X(t; \mathfrak{d}) := \int_0^t ds V(s; \mathfrak{d}), \end{aligned}$$

where  $\mathcal{C}([0, \infty), \mathbb{R})$  denotes the set of continuous paths defined on  $[0, \infty)$  and taking values on  $\mathbb{R}$ . It is not hard to see that for any  $\mathfrak{d} \in \Delta_\infty$  and  $t \geq 0$ ,

$$(2.5) \quad X(t; \mathfrak{d}) = v_0 \sum_{k=1}^M (-1)^{k-1} \delta_k + v_0 (-1)^M (t - t_M(\mathfrak{d})), \quad M = N(t; \mathfrak{d}),$$

with the convention that  $\sum_{k=a}^b (\dots) = 0$  whenever  $a > b$ ,  $a, b \in \mathbb{Z}$ . In particular, for  $t \in [0, \delta_1)$ , we have  $X(t; \mathfrak{d}) = v_0 t$ . Also, by definition of  $N$ , we have

$$|v_0 (-1)^M (t - t_M(\mathfrak{d}))| \leq |v_0| (t_{M+1}(\mathfrak{d}) - t_M(\mathfrak{d})) = |v_0| \delta_{M+1}.$$

*Notation.* When the sequence  $\mathfrak{d} = (\delta_k : k \in \mathbb{N})$  is distributed according to the infinite product measure of exponential distributions with parameter  $\lambda > 0$ , from here on we always write  $\mathfrak{s} = (s_1, s_2, \dots)$  instead of  $\mathfrak{d}$ .

We note that  $\mathfrak{s} \in \Delta_\infty$  almost surely and also  $N(t; \mathfrak{s}) < \infty$  for any  $t \geq 0$  almost surely. In fact,  $N(t; \mathfrak{s})$  is the unique (random) value  $n \in \mathbb{N}_0$  for which  $t_n(\mathfrak{s}) \leq t < t_{n+1}(\mathfrak{s})$  and the process

$$(2.6) \quad (N(t; \mathfrak{s}) : t \geq 0) \quad \text{is called a Poisson process of intensity } \lambda.$$

For further details about Poisson processes we recommend the monographies [8] and [39]. We denote by  $(\Omega, \mathcal{F}, \mathbb{P})$  the complete probability space where  $X^\mathfrak{s} := (X(t; \mathfrak{s}) : t \geq 0)$  is defined and by  $\mathbb{E}$  the expectation with respect to  $\mathbb{P}$ . The vector process  $((X(t; \mathfrak{s}), V(t; \mathfrak{s})) : t \geq 0)$  is Markovian on the state space  $\mathbb{R} \times \{-v_0, v_0\}$ , see for instance Section 12.1, p. 469 in [18]. However, the marginal process  $X^\mathfrak{s}$  is not itself Markovian for any non-zero initial velocity  $v_0$ .

Note that for  $v_0 = 0$  we have  $X(t; \mathfrak{s}) = 0$  for all  $t \geq 0$ . In the rest of the manuscript we always assume that  $v_0 \in \mathbb{R} \setminus \{0\}$ . We point out that the free velocity flip model corresponds to the so-called Goldstein–Kac telegraph process when the initial velocity  $v_0$  is chosen at random from the uniform distribution on  $\{-1, 1\}$  (Rademacher distribution) and independent from the Poisson process  $(N(t; \mathfrak{s}) : t \geq 0)$ . In this scenario, a straightforward computation yields

$$\mathbb{E}[e^{-zV(t)}] = \cosh(z) \quad \text{for any } z \in \mathbb{R} \text{ and } t \geq 0,$$

where  $\cosh$  denotes the hyperbolic cosine function, and hence the velocity process  $(V(t) : t \geq 0)$  is stationary.

In what follows we do not assume that  $v_0$  is random. Nevertheless, our main result and its consequences hold true as soon as we assume that  $v_0$  is a random variable with finite inverse second moment and independent of the Poisson process  $(N(t; \mathfrak{s}) : t \geq 0)$ . Hence, it covers the Goldstein–Kac telegraph process. The variance of  $X(t; \mathfrak{s})$ ,  $\text{Var}(X(t; \mathfrak{s}))$ , is given by

$$(2.7) \quad \text{Var}[X(t; \mathfrak{s})] = \frac{v_0^2}{\lambda^2} \left( \lambda t + e^{-2\lambda t} - \frac{e^{-4\lambda t}}{4} - \frac{3}{4} \right) \quad \text{for all } t \geq 0.$$

For details, see Item 2 of Lemma B.1 in Appendix B.

We also choose a target spatial scale,  $L > 0$ , and time scale,  $T > 0$ , in which we wish to control the position of the process  $X$ . The evolution of the process  $X$  will then be most conveniently described in terms of the following scaling parameters:

$$(2.8) \quad T_* := \lambda T \quad \text{and} \quad L_* = |v_0|^{-1} \lambda L.$$

By formula (2.7) we have

$$\text{Var}[L^{-1} X(T; \mathfrak{s})] = \frac{T_*}{L_*^2} + \frac{1}{L_*^2} \left( e^{-2T_*} - \frac{e^{-4T_*}}{4} - \frac{3}{4} \right).$$

The latter suggests that for  $T_* \rightarrow \infty$ ,  $L_* \rightarrow \infty$ ,  $T_*/L_*^2 \rightarrow \zeta$  for some  $\zeta > 0$ , the scaling process  $(L^{-1} X(t; \mathfrak{s}) : 0 \leq t \leq T)$  should behave as the Brownian motion with suitable diffusivity coefficient.

### 2.2. Related works

It is well-known that, under a suitable scaling, the telegraph process satisfies a functional central limit theorem. To be more precise, if  $\lambda \rightarrow \infty$  and  $v_0 \rightarrow \infty$  such that  $v_0^2/\lambda \rightarrow 1$ , then the telegraph process  $(X(t; \mathfrak{s}) : 0 \leq t < \infty)$  converges weakly in  $\mathcal{C}([0, \infty), \mathbb{R})$  to a standard Brownian motion  $W = (W(t) : t \geq 0)$ , see for instance Theorem 1.1 in [5] and the references therein. Using the fact that the evolution equation associated to  $(X(t; \mathfrak{s}) : 0 \leq t < \infty)$  is given by the telegraph

partial differential equation (1.1), asymptotics for the transition probability can be carried out in terms of Bessel functions, see [40]. However, up to our knowledge, the unique reference discussing about rate of convergence of the fixed rescaled time marginal to the corresponding Gaussian distribution is [23], in which an expansion of length two is carried out for the marginal  $X(t; \mathfrak{s})$  based on the marginal  $W(t)/\sqrt{\lambda}$ . The error term is of order  $\mathcal{O}(t)$  as  $t \rightarrow \infty$ .

We emphasize that the main Theorem 2.1 of this paper states a non-asymptotic estimate in the average quadratic Wasserstein metric between the processes  $(L^{-1}X(t; \mathfrak{s}) : t \geq 0)$  and a Brownian motion (not necessarily standard). This, in particular, implies Theorem 1.1 in [5] when  $\lambda \rightarrow \infty$  and  $v_0 \rightarrow \infty$  in such a way that  $v_0^2/\lambda \rightarrow 1$ . The latter is referred as a singular perturbation scale in [23].

### 2.3. Comparing two stochastic processes

#### 2.3.1. Coupling and Wasserstein metric

The Wasserstein metric is used to measure the distance between two probability measures on a Radon space. Polish spaces are standard examples of Radon spaces. Let  $d \in \mathbb{N}$  and consider two stochastic processes  $(X(t) : t \geq 0)$  and  $(Y(t) : t \geq 0)$  defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and taking values in  $\mathbb{R}^d$ . The Skorokhod space (Radon space) consisting of the set of right-continuous with left limits functions taking values in  $\mathbb{R}^d$  is denoted by  $\mathcal{D}([0, \infty), \mathbb{R}^d)$  and assume that the trajectories of the processes  $(X(t) : t \geq 0)$  and  $(Y(t) : t \geq 0)$  belong to  $\mathcal{D}([0, \infty), \mathbb{R}^d)$ . Let  $T > 0$  be a *target time scale*, and denote by  $\mu_1(dX)$  and  $\mu_2(dY)$  the probability measures of the processes  $(X(t) : t \geq 0)$  and  $(Y(t) : t \geq 0)$ , respectively, restricted to times  $t$  with  $0 \leq t \leq T$ . Then we define the *average quadratic cost function*

$$(2.9) \quad c_2(X, Y) := \frac{1}{T} \int_0^T ds \|X(s) - Y(s)\|^2,$$

where  $\|\cdot\|$  denotes the Euclidean norm on  $\mathbb{R}^d$ . The corresponding Wasserstein distance of order 2 with cost function  $c_2$  between  $\mu_1$  and  $\mu_2$ ,  $\mathcal{W}_2(\mu_1, \mu_2)$ , is defined by

$$(2.10) \quad \mathcal{W}_2(\mu_1, \mu_2) := \left( \inf_{\gamma} \int \gamma(dX, dY) c_2(X, Y) \right)^{1/2} = \left( \inf_{\gamma} \int \gamma(dX, dY) \frac{1}{T} \int_0^T ds \|X(s) - Y(s)\|^2 \right)^{1/2},$$

where the infimum is taken over all couplings  $\gamma$  between the probability measures  $\mu_1$  and  $\mu_2$ . In other words, for all continuous and bounded observables  $h : \mathcal{D}([0, T], \mathbb{R}^d) \rightarrow \mathbb{C}$  it follows

$$\int \gamma(dX, dY) h(X) = \int \mu_1(dX) h(X) \quad \text{and} \quad \int \gamma(dX, dY) h(Y) = \int \mu_2(dY) h(Y).$$

For shorthand, we write  $\mathcal{W}_2(X, Y)$  in place of  $\mathcal{W}_2(\mu_1, \mu_2)$ . We point out that (2.10) defines a natural distance between stochastic processes, see for instance [4]. Indeed,  $\mathcal{W}_2$  defines a metric that metrizes the weak topology on the set of probability measures  $\eta$  on  $L^2([0, T], \mathbb{R}^d)$  such that

$$\int_{L^2([0, T], \mathbb{R}^d)} \eta(dX) \int_0^T ds \|X(s)\|^2 < \infty.$$

Unfortunately, the numerical computation of  $\mathcal{W}_2$  or any other Wasserstein metric on an infinite dimensional space is very difficult. For basic definitions, properties and notions related to couplings and Wasserstein metrics, we refer to [41] and [47].

#### 2.3.2. Global estimate for moments

Let  $w : \mathbb{R}^d \rightarrow [0, \infty)$  be a weight function which yields an  $L^2$ -integrable observable with respect to  $\mu_1$  and  $\mu_2$ . To be more precise, we require that

$$(2.11) \quad C_1^2 := \int \mu_1(dX) \frac{1}{T} \int_0^T ds (w(X(s)))^2 < \infty \quad \text{and} \quad C_2^2 := \int \mu_2(dY) \frac{1}{T} \int_0^T ds (w(Y(s)))^2 < \infty.$$

Consider then any measurable function  $f : \mathbb{R}^d \rightarrow \mathbb{C}$  which is “ $w$ -Lipschitz”, in the following precise sense: there is a Lipschitz constant  $K \geq 0$  such that

$$(2.12) \quad |f(x) - f(y)| \leq K \frac{w(x) + w(y)}{2} \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^d,$$

where  $|\cdot|$  denotes the modulus of a complex number. The preceding condition (2.12) is chosen to obtain upper bounds for the error as we see in (2.14). The preceding definition (2.12) is motivated by the following observation: for any  $p > 0$  it follows that

$$(2.13) \quad \left| |x|^p - |y|^p \right| \leq K \frac{w(x) + w(y)}{2} |x - y| \quad \text{for all } x, y \in \mathbb{R},$$

where  $K = 2p$  and  $w : \mathbb{R} \rightarrow [0, \infty)$  is given by  $w(z) = |z|^{p-1}$  for all  $z \in \mathbb{R}$ . For details see Lemma B.2 in Appendix B. Inequality (2.13) implies that for the  $p$ -th moment observables, the estimation of the corresponding  $C_1$  and  $C_2$  needs implicitly good estimates of the  $2(p - 1)$ -th moments.

For short, we use the following standard notation  $\langle h \rangle_\mu$  to denote  $\int \mu(dX)h(X)$  for any  $h : \mathcal{D}([0, T], \mathbb{C}) \rightarrow \mathbb{C}$ . Then, in particular, for any coupling  $\gamma$  between  $\mu_1$  and  $\mu_2$  we have

$$\begin{aligned} & \left| \left\langle \frac{1}{T} \int_0^T ds f(X(s)) \right\rangle_{\mu_1} - \left\langle \frac{1}{T} \int_0^T ds f(Y(s)) \right\rangle_{\mu_2} \right| \\ &= \left| \int \gamma(dX, dY) \frac{1}{T} \int_0^T ds (f(X(s)) - f(Y(s))) \right| \\ &\leq \int \gamma(dX, dY) \frac{1}{T} \int_0^T ds \frac{K}{2} (w(X(s)) + w(Y(s))) \|X(s) - Y(s)\| \\ &\leq \frac{K}{2} \left[ \int \gamma(dX, dY) \frac{1}{T} \int_0^T ds (w(X(s)) + w(Y(s)))^2 \right]^{1/2} \left[ \int \gamma(dX, dY) \frac{1}{T} \int_0^T ds \|X(s) - Y(s)\|^2 \right]^{1/2} \\ &\leq \frac{K}{\sqrt{2}} [C_1^2 + C_2^2]^{1/2} \left[ \int \gamma(dX, dY) c_2(X, Y) \right]^{1/2}, \end{aligned}$$

where we have used Hölder’s inequality twice. Since this is true for any coupling  $\gamma$ , we obtain an *explicit estimate for time-averages* using the Wasserstein distance  $\mathcal{W}_2$  defined in (2.10). In other words, for any  $f$  satisfying (2.12) we have the global estimate

$$(2.14) \quad \left| \left\langle \frac{1}{T} \int_0^T ds f(X(s)) \right\rangle_{\mu_1} - \left\langle \frac{1}{T} \int_0^T ds f(Y(s)) \right\rangle_{\mu_2} \right| \leq K \max\{C_1, C_2\} \mathcal{W}_2(X, Y).$$

Moments of random variables are one of the most interesting and useful observables both from theoretical and practical points of view. Hence, the right-hand side of (2.14) can be read as follows: good approximations in terms of the above  $\mathcal{W}_2$ -distance imply similar approximations for a large class of physically relevant time-averages.

2.4. Main result and consequences

Given any stochastic process  $S = (S(t) : t \geq 0)$  and a target time  $T > 0$ , for convenience we define the projection process up to time  $T$  as follows:

$$\mathbb{S}_{[0, T]} := (S(t) : 0 \leq t \leq T).$$

Recall that  $X^\mathfrak{s} = (X(t; \mathfrak{s}) : t \geq 0)$  denotes the velocity flip model defined in (2.5), where  $\mathfrak{s}$  is sampled from an infinite product of exponential distribution of parameter  $\lambda$ . For any  $L > 0$  and  $T > 0$  we consider the corresponding projection process  $L^{-1}\mathbb{X}_{[0, T]}^\mathfrak{s}$  and we compare it with the projection process  $\mathbb{B}_{[0, T]}$  of a Brownian motion  $B = (B(t) : t \geq 0)$  with diffusivity constant given by  $\sigma^2 = L^{-2}v_0^2\lambda^{-1}$ . The independent coupling between  $L^{-1}\mathbb{X}_{[0, T]}^\mathfrak{s}$  and  $\mathbb{B}_{[0, T]}$  yields the following crude estimate: for any  $L > 0, T > 0, \lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that

$$(2.15) \quad \mathcal{W}_2(L^{-1}\mathbb{X}_{[0, T]}^\mathfrak{s}, \mathbb{B}_{[0, T]}) \leq \left( \frac{1}{4T_*L_*^2} (1 - e^{-2T_*}) - \frac{1}{2L_*^2} + \frac{T_*}{L_*^2} \right)^{1/2},$$

where the scaling parameters are given by  $T_* = \lambda T$  and  $L_* = |v_0|^{-1}\lambda L$ . For details for we refer to Lemma B.4 in Appendix B. If we do not use the independent coupling and we just apply Hölder’s inequality for the cross term  $\mathbb{E}[L^{-1}X(s; \mathfrak{s})B(s)]$ , we pay the price of a factor  $\sqrt{2}$  in the right-hand side of the preceding estimate. Nevertheless, for  $T_* \rightarrow \infty$  and  $L_* \rightarrow \infty$  in such way that  $T_*/L_*^2 \rightarrow \zeta > 0$  for some  $\zeta > 0$ , the right-hand side of (2.15) tends to  $\zeta^{1/2}$  and hence it is not informative.

The main result of this manuscript is the following.

**Theorem 2.1 (Brownian motion approximation of the free velocity flip model).** *There exists a (pure) positive constant  $C$  such that for any  $L > 0$ ,  $T > 0$ ,  $\lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$(2.16) \quad \mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^s, \mathbb{B}_{[0,T]}) \leq C\sqrt{T_*L_*^{-2}T_*^{-1/4}}(\sqrt{\ln(T_*+3)} + T_*^{-3/4}) + CL_*^{-1},$$

where the constants  $T_*$  and  $L_*$  are given by

$$(2.17) \quad T_* := \lambda T, \quad L_* = |v_0|^{-1}\lambda L,$$

and the diffusivity constant of the Brownian motion  $B := (B(t) : t \geq 0)$  is defined by

$$(2.18) \quad \sigma^2 := L^{-2}\frac{v_0^2}{\lambda}.$$

We continue to rely on the notations and assumptions in Theorem 2.1. We point out that the average cost and the time-one cost are linked by a deterministic time-change of the processes.

**Remark 2.2 (The average cost vs. the time-one cost).** The cost function  $c_2$  defined in (2.9) depends on the target time  $T$  and hence the corresponding Wasserstein distance  $\mathcal{W}_2$  defined in (2.10) also depends on  $T$ . For any  $t \geq 0$  we set

$$\tilde{X}(t; \mathfrak{s}) := L^{-1}X(Tt; \mathfrak{s}) \quad \text{and} \quad \tilde{B}(t) := B(Tt).$$

Then we have

$$\mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^s, \mathbb{B}_{[0,T]}) = \widetilde{\mathcal{W}}_2(\tilde{\mathbb{X}}_{[0,1]}^s, \tilde{\mathbb{B}}_{[0,1]}),$$

where the cost function  $\tilde{c}_2$  of  $\widetilde{\mathcal{W}}_2$  is given by

$$\tilde{c}_2(\tilde{X}, \tilde{B}) = \int_0^1 |\tilde{X}(t; \mathfrak{s}) - \tilde{B}(t)|^2 dt.$$

We point out that the cost function  $\tilde{c}_2$  does not depend on the target time  $T$  and hence the distance  $\widetilde{\mathcal{W}}_2$  also does not depend on  $T$ .

Using the scale invariance property of the Brownian motion, we stress the dependence of (2.18) in (2.16).

**Remark 2.3 (The diffusivity constant  $\sigma^2$ ).** We note that (2.16) can be rewritten as

$$\mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^s, \sigma\mathbb{W}_{[0,T]}) \leq C\sqrt{T_*L_*^{-2}T_*^{-1/4}}(\sqrt{\ln(T_*+3)} + T_*^{-3/4}) + CL_*^{-1},$$

where  $(W(t) : t \geq 0)$  is a standard Brownian motion.

We point out that the diffusive scaling limit is deduced from (2.16).

**Remark 2.4 (Diffusive scaling limit).** For  $T_* \rightarrow \infty$  and  $L_* \rightarrow \infty$  in such way that  $T_*/L_*^2 \rightarrow \zeta > 0$  for some  $\zeta > 0$ , the right-hand side of (2.16) is  $\mathcal{O}(T_*^{-1/4}\sqrt{\ln(T_*+3)})$  and therefore tends to zero. For instance, choosing the diffusive scaling  $L = \sqrt{\frac{Tv_0^2}{\lambda\zeta}}$  and taking  $T \rightarrow \infty$  produces such a limit.

Now, we describe the meaning of the parameters (2.17).

**Remark 2.5 (The meaning of the parameters).** Note that  $T_*$  is the mean of the number of jumps at time  $T$  and that  $L_* = \frac{1}{|v_0|\lambda^{-1}}L$  where  $|v_0|\lambda^{-1}$  is the mean free path of the unscaled process  $X^s$ . In other words,  $\langle |X(t_1(\mathfrak{s}); \mathfrak{s})| \rangle_{\mathfrak{s}} = |v_0|\lambda^{-1}$ , where  $t_1(\mathfrak{s}) = s_1$ .

**Remark 2.6 (A word about the constant  $C$ ).** The pure constant  $C$  of Theorem 2.1 can be estimated explicitly from the proof. However, the estimates given in the proof are not optimal and the most likely lead to overestimation.

Next corollary yields a Brownian motion approximation of the telegraph process.

**Corollary 2.7 (Brownian motion approximation of the telegraph process).** *Suppose that all assumptions made in Theorem 2.1 hold. Let  $L > 0$ ,  $T > 0$  and  $\lambda > 0$ . In addition, assume that the initial velocity  $v_0$  is chosen with the uniform distribution on the set  $\{-c, c\}$  for some  $c > 0$  and independent of the Poisson process  $(N(t; \mathfrak{s}) : t \geq 0)$  given in (2.6). Then there exists a positive constant  $C$  such that for any  $L > 0$ ,  $T > 0$  and  $\lambda > 0$  it follows that*

$$\mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^{\mathfrak{s}}, \mathbb{B}_{[0,T]}) \leq C\sqrt{T_*L_*^{-2}T_*^{-1/4}}(\sqrt{\ln(T_*+3)} + T_*^{-3/4}) + CL_*^{-1}.$$

We emphasize that the large flip rate limit is implied from (2.16).

**Remark 2.8 (Large flip rate limit).** If we choose  $L = 1$  and  $T$  arbitrary but fixed,  $\lambda \rightarrow \infty$  and  $|v_0| = c \rightarrow \infty$  in such a way that  $c^2/\lambda \rightarrow 1$  we reproduce the limit used in Theorem 1.1 of [5]. On the other hand, then we have

$$C\sqrt{T_*L_*^{-2}T_*^{-1/4}}(\sqrt{\ln(T_*+3)} + T_*^{-3/4}) + CL_*^{-1} = \mathcal{O}(\lambda^{-1/4}\sqrt{\ln(\lambda)}).$$

Therefore, Corollary 2.7 with the help of Remark 2.3 implies that the process  $\mathbb{X}_{[0,T]}^{\mathfrak{s}}$  converges weakly in  $\mathcal{C}[0, T]$  to the distribution of a standard Brownian motion  $\mathbb{W}_{[0,T]}$ .

In addition, Theorem 2.1 allows us to derive estimates for the time average of the  $p$ -th moments as follows.

**Corollary 2.9 (Time average of  $p$ -th moments).** *Suppose that all assumptions made in Theorem 2.1 hold. Let  $p > 0$  and consider the observable  $f(x) = |x|^p$  for any  $x \in \mathbb{R}$ . Then for any  $L > 0$ ,  $T > 0$ ,  $\lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$\left| \left\langle \frac{1}{T} \int_0^T ds f(L^{-1}X(s; \mathfrak{s})) \right\rangle_{\mu_1} - \left\langle \frac{1}{T} \int_0^T ds f(B(s)) \right\rangle_{\mu_2} \right| \leq 2p \max\{C_1, C_2\} \mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^{\mathfrak{s}}, \mathbb{B}_{[0,T]}),$$

where  $C_1^2 = C_2^2 = 1$  if  $p = 1$ , whereas if  $p \neq 1$  we have

$$C_1^2 \leq \frac{\tilde{C}T_*^{p-1}}{pL_*^{2(p-1)}} + \frac{C'T_*^{p-2}}{(p-1)L_*^{2(p-1)}}, \quad C_2^2 = \frac{2^{p-1}\Gamma(\frac{2p-1}{2})}{p\sqrt{\pi}} \frac{T_*^{p-1}}{L_*^{2(p-1)}},$$

the constants  $T_*$  and  $L_*$  are the scaling parameters that appear in (2.17),  $\tilde{C}$  and  $C'$  are the constants that appear for  $r = 2(p-1)$  in Item 2, (B.1), of Lemma B.1 in Appendix B and  $\Gamma$  denotes the Gamma function.

In addition, if  $p \in \mathbb{N}$  and  $g(x) = x^p$  for  $x \in \mathbb{R}$ , then

$$\left| \left\langle \frac{1}{T} \int_0^T ds g(L^{-1}X(s; \mathfrak{s})) \right\rangle_{\mu_1} - \left\langle \frac{1}{T} \int_0^T ds g(B(s)) \right\rangle_{\mu_2} \right| \leq 2p \max\{C_1, C_2\} \mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^{\mathfrak{s}}, \mathbb{B}_{[0,T]}).$$

**Proof.** The proof follows directly from inequality (2.14), Lemma B.2 in Appendix B and Lemma B.3 in Appendix B.  $\square$

Additionally, Theorem 2.1 holds true in  $\mathcal{W}_p$  for  $p \in [1, 2]$ .

**Remark 2.10 (Convergence in  $\mathcal{W}_p$ ,  $p \in [1, 2]$ ).** Analogously to (2.10), for any  $p \geq 1$  we define the  $p$ -th average distance  $\mathcal{W}_p$  distance by

$$\mathcal{W}_p(\mu_1, \mu_2) := \left( \inf_{\gamma} \int \gamma(dX, dY) (c_2(X, Y))^{p/2} \right)^{1/p}.$$

By Hölder's inequality we obtain that Theorem 2.1, Corollary 2.7 and Corollary 2.9 are valid for the  $\mathcal{W}_p$  for  $p \in [1, 2]$ .

### 3. Outline of the proof of Theorem 2.1

The idea of the proof is to compare a suitable rescaled free velocity flip process  $L^{-1}X = (L^{-1}X(t; \mathfrak{s}) : t \geq 0)$  with a simplified process (decoupled process)  $Y := (Y(t) : t \geq 0)$  in which computations can be carried out. However, the



process  $Y$  only possesses zero mean at even jump times. Hence, it is reasonable to compare the simplified process  $Y$  at even jump times with a Brownian motion  $B = (B(t) : t \geq 0)$  with a suitable diffusivity constant.

By the triangle inequality, the estimation of the left-hand side of (2.16) is divided in three main components as follows:

$$(3.1) \quad \mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^s, \mathbb{B}_{[0,T]}) \leq \mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^s, \mathbb{Y}_{[0,T]}) + \mathcal{W}_2(\mathbb{Y}_{[0,T]}, \mathbb{Z}_{[0,T]}) + \mathcal{W}_2(\mathbb{Z}_{[0,T]}, \mathbb{B}_{[0,T]}).$$

where the process  $Z = (Z(t) : t \geq 0)$  is the Càdlàg process obtained from the process  $Y$  at even jumps by a constant piecewise interpolation between even jump times. The explicit construction will be given below.

### 3.1. Definition of the process $Y$

Decoupling techniques are very useful in the theory and applications of Probability and Statistics to replace (or reduce) the dependence of the system. For further discussion we recommend the monograph [13]. We note that (2.5) can be seen as a perturbed random walk. However, we point out that  $n = N(t; \mathfrak{s})$  and  $X(t; \mathfrak{s})$  are highly dependent. To overcome this difficulty, we define a suitable decoupled process  $(Y(t) : t \geq 0)$  which is indeed, a perturbed random walk. Then using the cost function (2.9) and the metric (2.10) we quantify the error term between  $L^{-1}\mathbb{X}_{[0,T]}^s$  and  $\mathbb{Y}_{[0,T]}$ .

Let  $\mathbf{u} = (u_j : j \in \mathbb{N})$  be sequence of independent and identically distributed (i.i.d. for short) random variables with exponential distribution of parameter one defined in the probability space  $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ . For each (deterministic)  $n \in \mathbb{N}$ , we define  $\mathbf{u}^{(n)} := (u_k^{(n)} : k \in \mathbb{N})$  as follows

$$(3.2) \quad u_k^{(n)} := u_k \quad \text{for } k \in \{1, \dots, n\} \quad \text{and} \quad u_k^{(n)} := 2\beta_n \quad \text{for } k \in \{n+1, n+2, \dots\}.$$

where  $\beta_n = n$  (the value of  $\beta_n$  will later be fixed by an optimisation argument, hence the somewhat peculiar notation here). Then we also define

$$(3.3) \quad \delta_k^{(n)} := \frac{T}{\beta_n} u_k^{(n)} \quad \text{for } k \in \mathbb{N}.$$

Note that then  $\delta_k^{(n)} = 2T$  if  $k \geq n+1$ . We consider the corresponding ‘‘jump-times’’

$$(3.4) \quad t_m^{(n)} := \sum_{k=1}^m \delta_k^{(n)} \quad \text{for all } m \in \mathbb{N} \text{ and } t_0^{(n)} = 0,$$

and collect its increments as  $\mathfrak{d}^{(n)} = (\delta_1^{(n)}, \delta_2^{(n)}, \dots)$ . Since  $\mathfrak{d}^{(n)} \in \Delta_\infty$ , where  $\Delta_\infty$  is defined in (2.2), using the map (2.5) we can define the stochastic process  $(Y(t; n, \mathfrak{d}^{(n)}) : t \geq 0)$  by the following rule: for any  $t \geq 0$

$$(3.5) \quad Y(t; n, \mathfrak{d}^{(n)}) := L^{-1}X(t; \mathfrak{d}^{(n)}) = L^{-1}v_0 \sum_{k=1}^M (-1)^{k-1} \delta_k^{(n)} + L^{-1}v_0 (-1)^M (t - t_M(\mathfrak{d}^{(n)})), \quad M = N(t; \mathfrak{d}^{(n)}),$$

where  $(N(t; \mathfrak{d}^{(n)}) : t \geq 0)$  is given in (2.3). In addition, we take a random variable  $K$  with Poisson distribution with parameter  $T_* = \lambda T$  defined in the probability space  $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$  and independent of  $\mathbf{u}$ . For short, for any  $\lambda > 0$ ,  $(p_n(\lambda) : n \in \mathbb{N}_0)$  denotes the Poisson distribution  $\text{Po}(\lambda)$ , that is,

$$p_n(\lambda) := \frac{1}{n!} \lambda^n e^{-\lambda} \quad \text{for any } n \in \mathbb{N}_0.$$

From now on, we will use a convenient generic indicator function notation  $\mathbb{1}_{\{P\}}$  where  $P$  is some condition; more precisely, given a condition  $P(x)$  which depends on the variable  $x$ , the corresponding indicator function is defined as the map  $x \mapsto \mathbb{1}_{\{P\}}(x) \in \{0, 1\}$  where  $\mathbb{1}_{\{P\}}(x) = 1$ , if  $P(x)$  is true, and  $\mathbb{1}_{\{P\}}(x) = 0$  otherwise. We will also often use the alternative notation  $\mathbb{1}_{\{P(x)\}}$  to denote  $\mathbb{1}_{\{P\}}(x)$ .

Let  $(\Omega, \mathcal{F}, \mathbb{Q}) := (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$  be the product probability space of  $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$  and  $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$  and denote by  $\mathbb{E}_{\mathbb{Q}}$  the expectation with respect to  $\mathbb{Q}$ . On  $(\Omega, \mathcal{F}, \mathbb{Q})$  we define the stochastic process  $(Y(t; K, \mathbf{u}) : t \geq 0)$  as follows: for each  $\omega = (\omega_1, \omega_2) \in \Omega$  we set

$$(3.6) \quad Y(t; K, \mathbf{u})(\omega) := \sum_{n=0}^{\infty} Y(t; n, \mathfrak{d}^{(n)})(\omega_1) \mathbb{1}_{\{K=n\}}(\omega_2) \quad \text{for any } t \geq 0,$$

where for each (not random)  $n \in \mathbb{N}$ ,  $(Y(t; n, \mathfrak{d}^{(n)}) : t \geq 0)$  is defined in (3.5) and for  $n = 0$  we define  $Y(t; 0, \mathfrak{d}^{(0)}) := L^{-1}v_0t$  for all  $t \geq 0$ . Let  $F : \mathcal{C}([0, T], \mathbb{R}) \rightarrow \mathbb{R}$  be a bounded measurable function. In particular, it only depends on the values of the process up to time  $T$  such as  $c_1(X)$  and  $c_2(X)$  given in (2.11). By Fubini's theorem we have

$$\mathbb{E}_{\mathbb{Q}}[F(Y(\cdot; K, \mathbf{u}))] = \sum_{n=0}^{\infty} \mathbb{E}_{\mathbb{Q}}[F(Y(\cdot; K, \mathbf{u}))\mathbb{1}_{\{K=n\}}],$$

where

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[F(Y(\cdot; K, \mathbf{u}))\mathbb{1}_{\{K=n\}}] &= \mathbb{E}_{\mathbb{Q}}[F(Y(\cdot; n, \mathbf{u}))\mathbb{1}_{\{K=n\}}] \\ (3.7) \qquad \qquad \qquad &= p_n(T_*)\mathbb{E}_{\mathbb{Q}}[F(Y(\cdot; n, \mathbf{u}))] \\ &= p_n(T_*) \int_{\mathbb{R}_*^n} d^n \mathbf{u} e^{-\sum_{k=1}^n u_k} F(L^{-1}X(\cdot; T\beta_n^{-1}\mathbf{u}^{(n)})), \end{aligned}$$

$\mathbf{u}^{(n)}$  is given in (3.2), and the map  $X$  is defined in (2.4). As evident in the last equality, the function  $\mathbf{u} \mapsto F(Y(\cdot; n, \mathbf{u}))$  only depends on the components  $u_1, \dots, u_n$ .

Bearing all this in mind, we compare  $L^{-1}\mathbb{X}_{[0,T]}^{\mathfrak{s}}$  with the decoupled process  $\mathbb{Y}_{[0,T]}$ .

**Proposition 3.1 (Coin-flip coupling).** *There exists a (pure) constant  $\kappa_1 > 0$  such that for any  $L > 0, T > 0, \lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$(3.8) \qquad \mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^{\mathfrak{s}}, \mathbb{Y}_{[0,T]}) \leq \kappa_1 \sqrt{T_* L_*^{-2} T_*^{-1/4}} \sqrt{\ln(T_* + 3)} + \kappa_1 L_*^{-1},$$

where the constants  $T_*$  and  $L_*$  are given in (2.17).

The complete proof is given in Section 4 and it is based on the coin-flip coupling (see [33] or Appendix A in [44]).

### 3.2. Definition of the process $Z$

In the sequel, we define the process  $Z$ . We note that (2.1) and (2.3) implies  $N(t_n; \mathfrak{s}) = n$  for all  $n \in \mathbb{N}_0$ . Then (2.5) yields for all  $n \in \mathbb{N}$

$$L^{-1}X(t_{2n}(\mathfrak{s}); \mathfrak{s}) - L^{-1}X(t_{2(n-1)}(\mathfrak{s}); \mathfrak{s}) = -L^{-1}v_0(s_{2n} - s_{2n-1}).$$

That is, at even jumps, the increments of the process  $L^{-1}X$  are i.i.d. random variables which can be described explicitly. Recall the notations (3.2), (3.3) and (3.4) used to define the process  $Y$ . In what follows, we define the stochastic process  $Z$  as the constant Càdlàg interpolation between the values of  $Y$  at even jumps.

We recall that  $\lfloor \cdot \rfloor$  denotes the so-called floor function, that is, for any  $x \in \mathbb{R}$ ,  $\lfloor x \rfloor$  gives the largest integer less than or equal than  $x$ . For  $n \in \mathbb{N} \setminus \{0, 1\}$  we define the Càdlàg version of the random walk  $(\eta_k^*(\mathfrak{s}) : k \in \mathbb{N}_0)$  given by

$$Z(t; n, \mathfrak{s}) = \sum_{k=1}^{\tilde{n}} \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{s})\}} \eta_k^*(\mathfrak{s}) \quad \text{for any } t \geq 0,$$

where  $\tilde{n} = \lfloor n/2 \rfloor$ ,

$$(3.9) \qquad \eta_k^*(\mathfrak{s}) = -L^{-1}v_0(s_{2k} - s_{2k-1}), \quad k \in \mathbb{N}, \eta_0^*(\mathfrak{s}) = 0,$$

and  $\mathfrak{s} = (s_1, s_2, \dots) = \lambda^{-1}\mathbf{u} = \lambda^{-1}(u_1, u_2, \dots)$  is distributed according to the infinitely product measure of exponential distributions with parameter  $\lambda > 0$ . For  $n \in \{0, 1\}$  we set  $Z(t; n, \mathfrak{s}) = 0$  for all  $t \geq 0$ . Analogously to (3.6) we define on  $(\Omega, \mathcal{F}, \mathbb{Q})$  the stochastic process  $(Z(t; K, \mathbf{u}) : t \geq 0)$  as follows: for each  $\omega = (\omega_1, \omega_2) \in \Omega$  it is given by

$$(3.10) \qquad Z(t; K, \mathfrak{s})(\omega) := \sum_{n=0}^{\infty} Z(t; n, \mathfrak{s})(\omega_1)\mathbb{1}_{\{K=n\}}(\omega_2) \quad \text{for any } t \geq 0.$$

We remark that the process  $Y$  and the process  $Z$  are constructed in the same probability space  $(\Omega, \mathcal{F}, \mathbb{Q})$ .

**Proposition 3.2 (Synchronous coupling).** *There exists a (pure) constant  $\kappa_2 > 0$  such that for any  $L > 0$ ,  $T > 0$ ,  $\lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$(3.11) \quad \mathcal{W}_2(\mathbb{Y}_{[0,T]}, \mathbb{Z}_{[0,T]}) \leq \kappa_2 \sqrt{T_* L_*^{-2} T_*^{-1/4}} (1 + T_*^{-3/4}),$$

where the constants  $T_*$  and  $L_*$  are given in (2.17).

The complete proof can be found in Section 5 and it is based on the synchronous coupling of the processes  $Y$  and  $Z$  which are constructed with the same sequence of random variables as an input.

**Lemma 3.3 (Komlós–Major–Tusnády coupling).** *There exists a (pure) constant  $\kappa_3 > 0$  such that for any  $L > 0$ ,  $T > 0$ ,  $\lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$\mathcal{W}_2(\mathbb{Z}_{[0,T]}, \mathbb{B}_{[0,T]}) \leq \kappa_3 \sqrt{T_* L_*^{-2} T_*^{-1/4}} (1 + T_*^{-3/4}),$$

where the constants  $T_*$  and  $L_*$  are given in (2.17) and the diffusivity constant of the Brownian motion  $B := (B(t) : t \geq 0)$  is defined in (2.18).

The proof is given in Section 6 and it relies on the Komlós–Major–Tusnády coupling (see [31] and [32]).

We stress the fact that Theorem 2.1 is just a consequence of what we have already stated up to here.

**Proof of Theorem 2.1.** It follows directly from inequality (3.1) with the help of Proposition 3.1, Proposition 3.2 and Lemma 3.3.  $\square$

#### 4. Proof of Proposition 3.1: Coin-flip coupling

In this section we prove Proposition 3.1. We start with some preliminaries.

##### 4.1. Expectation for continuous observables of the free velocity flip model

Since then  $N(t; \mathfrak{s}) < \infty$  almost surely, for any (regular enough, e.g., positive and measurable) observable  $F(\mathbb{X}_{[0,T]}^{\mathfrak{s}})$  which depends only on the path up to times  $T > 0$ , such as  $c_1$  and  $c_2$  defined in (2.11), we can decompose the events depending on the realisation of  $N(T; \mathfrak{s})$ . We recall the generic indicator function notation  $\mathbb{1}_{\{P\}}$  from Section 3.1, and obtain

$$\mathbb{E}[F(\mathbb{X}_{[0,T]}^{\mathfrak{s}})] = \sum_{n=0}^{\infty} \mathbb{E}[F(\mathbb{X}_{[0,T]}^{\mathfrak{s}}) \mathbb{1}_{\{N(T; \mathfrak{s})=n\}}].$$

This partition helps, since for any  $n$  the path up to time  $T$  and thus also the value of  $F(\mathbb{X}_{[0,T]}^{\mathfrak{s}})$ , only depends on  $s_k$ , for  $0 \leq k \leq n+1$  which is crucial in our argument. First, for  $n=0$ , we can directly evaluate

$$\mathbb{E}[F(\mathbb{X}_{[0,T]}^{\mathfrak{s}}) \mathbb{1}_{\{N(T; \mathfrak{s})=0\}}] = F(I) e^{-\lambda T}, \quad \text{where } I(t) = v_0 t \text{ for all } t \geq 0.$$

Therefore, in the following computation we only need to consider values  $n \in \mathbb{N}$ .

For such  $n$ , we have the following explicit integral representation

$$\mathbb{E}[F(\mathbb{X}_{[0,T]}^{\mathfrak{s}}) \mathbb{1}_{\{N(T; \mathfrak{s})=n\}}] = \lambda^{n+1} \int_{\mathbb{R}_*^{n+1}} d^{n+1} \mathfrak{s} \prod_{k=1}^{n+1} e^{-\lambda s_k} F(\mathbb{X}_{[0,T]}^{\mathfrak{s}}) \mathbb{1}_{\{t_n(\mathfrak{s}) \leq T\}} \mathbb{1}_{\{s_{n+1} > T - t_n(\mathfrak{s})\}},$$

where  $(t_n(\mathfrak{s}) : n \in \mathbb{N}_0)$  are the jump times defined in (2.1). The notation is somewhat formal, since the full path  $\mathbb{X}^{\mathfrak{s}}$  depends on the full sequence  $\mathfrak{s}$ . However, whenever both indicator functions are non-zero in the above integrand, we note that  $\mathbb{X}_{[0,T]}^{\mathfrak{s}}$  only depends on the values  $s_k$ ,  $k=1, 2, \dots, n$ . Explicitly, defining  $v_k := s_k/T$ ,  $k=1, 2, \dots, n$ , and setting somewhat arbitrarily  $v_k = 2$  for  $k > n$ , we find that then  $\mathbb{X}_{[0,T]}^{\mathfrak{s}} = X(\cdot; T\mathfrak{v})_{[0,T]}$  since the conditions in the integrand guarantee that the jump  $n+1$  occurs after  $T$  which is also ensured by the above choice  $Tv_{n+1} = 2T$ . In this case, for simplicity, we also drop the restriction to the time-interval  $[0, T]$  from the notation, and simply write  $F(\mathbb{X}_{[0,T]}^{\mathfrak{s}}) = F(X(\cdot; T\mathfrak{v}))$  in the above integral.

Performing the implied change of variables, setting  $\tilde{r} = s_{n+1}/T$ , and using Fubini’s theorem to reorder the integrals, we obtain

$$\mathbb{E}[F(\mathbb{X}_{[0,T]}^s) \mathbb{1}_{\{N(T; \mathfrak{s})=n\}}] = T_*^{n+1} \int_{\mathbb{R}_*^n} d^n \mathbf{v} \left( \int_{\mathbb{R}_*} d\tilde{r} e^{-T_*(\tilde{r}+t_n(\mathbf{v}))} F(X(\cdot; T\mathbf{v})) \mathbb{1}_{\{t_n(\mathbf{v}) \leq 1\}} \mathbb{1}_{\{\tilde{r} > 1-t_n(\mathbf{v})\}} \right),$$

where  $T_* = \lambda T$ , and we have used the obvious scaling property  $t_n(T\mathbf{v}) = T t_n(\mathbf{v})$ . The  $\tilde{r}$ -integral can then be evaluated explicitly using  $\int_{1-t_n(\mathbf{v})}^\infty d\tilde{r} e^{-T_*(\tilde{r}+t_n(\mathbf{v}))} = T_*^{-1} e^{-T_*}$ . The resulting integral representation can be further simplified by relying on Dirac  $\delta$ -functions whose precise construction as non-negative Radon measures is explained in Appendix C. In particular, applying item iv) of Lemma C.2 justifies the result from the following formal computation:

$$\begin{aligned} \mathbb{E}[F(\mathbb{X}_{[0,T]}^s) \mathbb{1}_{\{N(T; \mathfrak{s})=n\}}] &= T_*^n e^{-T_*} \int_{\mathbb{R}_*^n} d^n \mathbf{v} F(X(\cdot; T\mathbf{v})) \mathbb{1}_{\{t_n(\mathbf{v}) \leq 1\}} \\ (4.1) \quad &= T_*^n e^{-T_*} \int_{\mathbb{R}_*^n} d^n \mathbf{v} F(X(\cdot; T\mathbf{v})) \mathbb{1}_{\{t_n(\mathbf{v}) \leq 1\}} \int_{\mathbb{R}_*} dr \delta(t_n(\mathbf{v}) - r) \\ &= T_*^n e^{-T_*} \int_{\mathbb{R}_*} dr \mathbb{1}_{\{r \leq 1\}} \left( \int_{\mathbb{R}_*^n} d^n \mathbf{v} F(X(\cdot; T\mathbf{v})) \delta(t_n(\mathbf{v}) - r) \right) \\ &= \frac{1}{n!} T_*^n e^{-T_*} \int_0^1 dr nr^{n-1} (n-1)! \int_{\mathbb{R}_*^n} d^n \mathbf{u} \delta(t_n(\mathbf{u}) - 1) F(X(\cdot; T\mathbf{u})), \end{aligned}$$

with a slight abuse of notation since the values of the sequence “Tr  $\mathbf{u}$ ” for  $k > n$  are not rescaled by  $r$ , i.e., they are still equal to  $2T$ .

Since  $\int_{\mathbb{R}_*^n} d^n \mathbf{u} \delta(t_n(\mathbf{u}) - 1) = \frac{1}{(n-1)!}$ , the remaining integrals form a product of two probability measures for  $r$  and  $\mathbf{u}$ , i.e.,

$$(4.2) \quad d\mathbf{r} \cdot nr^{n-1} \mathbb{1}_{\{r \in (0,1]\}}$$

and

$$(4.3) \quad d^n \mathbf{u} \cdot (n-1)! \delta(t_n(\mathbf{u}) - 1).$$

By straightforward computations we have

$$\langle r \rangle = 1 - \frac{1}{n+1}, \quad \langle r^2 \rangle - \langle r \rangle^2 = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n}{(n+1)^2(n+2)} \leq \frac{1}{(n+1)^2},$$

where the expectations are computed with respect to the probability measure (4.2). Hence,  $r$  is distributed close to 1 with standard deviation  $\mathcal{O}(1/n)$  as  $n \rightarrow \infty$ . Using the permutation invariance of (4.3), we also find the following statistics for components of  $\mathbf{u}$ ,

$$\langle u_i \rangle = \frac{1}{n}, \quad \langle u_i^2 \rangle = \frac{2}{n(n+1)}, \quad \langle u_i^3 \rangle = \frac{6}{n(n+1)(n+2)}, \quad \langle u_i u_j \rangle = \frac{1}{n(n+1)}, \quad i \neq j,$$

where the expectations are computed with respect to the probability measure (4.3). For details see Lemma A.3 in Appendix A. In particular, the standard deviation of each component  $u_i$  is also  $\mathcal{O}(1/n)$ .

We now rescale the jumps for the process  $(Y(t; n, \mathfrak{s}) : 0 \leq t \leq T)$  so that a good coupling can be formed between it and  $(L^{-1}X(t; n, \mathfrak{s}) : 0 \leq t \leq T)$ . Assume that  $n \in \mathbb{N}$  is deterministic, and for shorthand we set  $f(T\mathbf{v}) := F(L^{-1}X(\cdot; T\mathbf{v}))$ . Then (3.7) implies

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[F(Y(\cdot; n, \mathbf{u}))] &= \beta_n^n \int_{\mathbb{R}_*^n} d^n \mathbf{v} e^{-\beta_n \sum_{k=1}^n v_k} f(T\mathbf{v}) = \beta_n^n \int_{\mathbb{R}_*} dr e^{-\beta_n r} \int_{\mathbb{R}_*^n} d^n \mathbf{v} \delta\left(\sum_{k=1}^n v_k - r\right) f(T\mathbf{v}) \\ (4.4) \quad &= \beta_n^n \int_{\mathbb{R}_*} dr r^{n-1} e^{-\beta_n r} \int_{\mathbb{R}_*^n} d^n \mathbf{u} \delta\left(\sum_{k=1}^n u_k - 1\right) f(T\mathbf{u}) \\ &= \frac{1}{(n-1)!} \beta_n^n \int_{\mathbb{R}_*} dr r^{n-1} e^{-\beta_n r} (n-1)! \int_{\mathbb{R}_*^n} d^n \mathbf{u} \delta\left(\sum_{k=1}^n u_k - 1\right) f(T\mathbf{u}). \end{aligned}$$

## 4.2. Coin flip type coupling

Let

$$(4.5) \quad \nu_1(dr) = nr^{n-1} \mathbb{1}_{\{0 < r \leq 1\}} dr \quad \text{and} \quad \nu_2(dr) = \frac{1}{(n-1)!} \beta^n r^{n-1} e^{-\beta r} \mathbb{1}_{\{r > 0\}} dr$$

denote probability measures on  $\mathbb{R}_*$ . We note that  $\nu_2$  is the Gamma distribution of parameters  $n$  and  $\beta$ . Moreover, it has mean  $n/\beta$  and variance  $n/\beta^2$ . There is a choice of  $\beta$  which makes the densities agree “with a high probability”, and the following computations yield that the original choice  $\beta = \beta_n = n$  suffices. By (4.1) and (4.4) we only need to generate a good coupling between  $\nu_1$  and  $\nu_2$ . For this case, one good choice is using a “coin-flip type” coupling. Let

$$(4.6) \quad g(r) := n! \mathbb{1}_{\{0 < r \leq 1\}} \beta^{-n} e^{\beta r}$$

and define a coupling  $\gamma_0$  by setting the action for any observable  $h : \mathbb{R}_*^2 \rightarrow \mathbb{R}$  as follows

$$\begin{aligned} \int_{\mathbb{R}_*^2} \gamma_0(dr_1, dr_2) h(r_1, r_2) &= \int_{\mathbb{R}_*} \nu_2(dr) \min\{1, g(r)\} h(r, r) \\ &\quad + \int_{\mathbb{R}_*} \nu_2(dr_1) \int_{\mathbb{R}_*} \nu_2(dr_2) \frac{1}{Z_{\beta, n}} (1 - g(r_2))_+ (g(r_1) - 1)_+ h(r_1, r_2), \end{aligned}$$

where  $(x)_+ := x \mathbb{1}_{\{x > 0\}}$  and

$$(4.7) \quad Z_{\beta, n} = \int_{\mathbb{R}_*} \nu_2(dr) (1 - g(r))_+ = \int_{\mathbb{R}_*} \nu_2(dr) (g(r) - 1)_+ > 0.$$

This is indeed a coupling between  $\nu_1$  and  $\nu_2$ , see for instance relation (5.11) of [35], where it has been applied the techniques from Appendix A in [44]. We stress that  $\nu_1, \nu_2, g$  and  $\gamma_0$  depend on  $n$ , however, for ease of notation we drop its dependence.

Finally, using this  $\gamma_0$  we define a coupling  $\gamma$  between  $L^{-1}\mathbb{X}_{[0, T]}^5$  and  $\mathbb{Y}_{[0, T]}$  as follows

$$(4.8) \quad \begin{aligned} &\mathbb{E}_\gamma[F(L^{-1}\mathbb{X}_{[0, T]}^5, \mathbb{Y}_{[0, T]})] \\ &= e^{-T_*} F(L^{-1}I, L^{-1}I) \\ &\quad + \sum_{n=1}^{\infty} p_n(T_*) \int_{\mathbb{R}_*^2} \gamma_0(dr_1, dr_2) (n-1)! \int_{\mathbb{R}_*^n} d^n u \delta\left(\sum_{k=1}^n u_k - 1\right) F(L^{-1}X(\cdot; Tr_1u), L^{-1}X(\cdot; Tr_2u)), \end{aligned}$$

where  $I = I(t) = v_0 t$  for all  $t \geq 0$ . One can now check that this has the right marginals. Thus for the choice  $F = c_2$  given in (2.9), we need to control

$$c_2(L^{-1}X(\cdot; Tr_1u), L^{-1}X(\cdot; Tr_2u)) = \frac{1}{L^2 T} \int_0^T dt |X(t; Tr_1u) - X(t; Tr_2u)|^2.$$

For  $r_2 = r_1$ , the value of the preceding cost is clearly zero whatever  $n$  and  $u$  are. Therefore,

$$(4.9) \quad \begin{aligned} &\mathbb{E}_\gamma[c_2(L^{-1}\mathbb{X}_{[0, T]}^5, \mathbb{Y}_{[0, T]})] \\ &= \sum_{n=1}^{\infty} p_n(T_*) (n-1)! \int_{\mathbb{R}_*^n} d^n u \delta\left(\sum_{k=1}^n u_k - 1\right) \\ &\quad \times \int_{\mathbb{R}_*} \nu_2(dr_1) \int_{\mathbb{R}_*} \nu_2(dr_2) \frac{1}{Z_{\beta, n}} (1 - g(r_2))_+ (g(r_1) - 1)_+ c_2(L^{-1}X(\cdot; Tr_1u), L^{-1}X(\cdot; Tr_2u)). \end{aligned}$$

Then for  $t \in [0, T]$  we rewrite  $X(t; Tr_1u) - X(t; Tr_2u)$  as follows

$$(4.10) \quad \begin{aligned} X(t; Tr_1u) - X(t; Tr_2u) &= v_0 D_1(t, u, r_1, r_2, T) + v_0 D_2(t, u, r_1, r_2, T) \\ &\quad + v_0 D'_1(t, u, r_1, r_2, T) + v_0 D'_2(t, u, r_1, r_2, T), \end{aligned}$$

where, writing  $n_i = n_i(t) = N(t; Tr_i u) \leq n$ , we have

$$(4.11) \quad D_1(t, u, r_1, r_2, T) := T(r_1 - r_2) \sum_{k=1}^{n_1} (-1)^{k-1} u_k,$$

$$(4.12) \quad |D_2(t, u, r_1, r_2, T)| := Tr_2 \left| \sum_{\min\{n_1, n_2\} < k \leq \max\{n_1, n_2\}} (-1)^{k-1} u_k \right|,$$

$$(4.13) \quad D'_1(t, u, r_1, r_2, T) := (-1)^{n_1} (t - t_{n_1}(Tr_1 u)) \quad \text{and} \quad D'_2(t, u, r_1, r_2, T) := (-1)^{n_2} (t - t_{n_2}(Tr_2 u)).$$

For short, for each  $j \in \{1, 2\}$  we write  $D_j, D'_j$  instead of  $D_j(t, u, r_1, r_2, T), D'_j(t, u, r_1, r_2, T)$ , respectively. By Hölder's inequality we obtain the following upper bound

$$|X(t; Tr_1 u) - X(t; Tr_2 u)|^2 \leq 4v_0^2 (|D_1|^2 + |D_2|^2 + |D'_1|^2 + |D'_2|^2) \quad \text{for all } t \in [0, T].$$

In what follows, we find a positive constant  $\kappa_1$  that does not depend on  $v_0, \lambda, L$  and  $T$  and such that

$$(4.14) \quad \mathcal{W}_2^2(L^{-1} \mathbb{X}_{[0, T]}^5, \mathbb{Y}_{[0, T]}) \leq |v_0|^2 \mathbb{E}_\gamma [c_2(L^{-1} X, Y)] \leq \kappa_1 |v_0|^2 (\lambda^{-3/2} L^{-2} T^{1/2} \ln(\lambda T + 3) + \lambda^{-2} L^{-2}),$$

where  $\gamma$  is the coupling measure defined in (4.8). By (2.8) we have that  $T_* = \lambda T$  and  $L_* = |v_0|^{-1} \lambda L$  and hence inequality (4.14) reads as follows

$$\mathcal{W}_2^2(L^{-1} \mathbb{X}_{[0, T]}^5, \mathbb{Y}_{[0, T]}) \leq \kappa_1 T_* L_*^{-2} T_*^{-1/2} \ln(T_* + 3) + \kappa_1 L_*^{-2}.$$

### 4.3. Comparison between $L^{-1} \mathbb{X}_{[0, T]}^5$ and $\mathbb{Y}_{[0, T]}$

By (4.9) and (4.10) we have

$$(4.15) \quad \begin{aligned} \mathbb{E}_\gamma [c_2(L^{-1} \mathbb{X}_{[0, T]}^5, \mathbb{Y}_{[0, T]})] &\leq 4v_0^2 \sum_{n=1}^\infty p_n(T_*) (n-1)! \int_{\mathbb{R}_*^n} d^n u \delta \left( \sum_{k=1}^n u_k - 1 \right) \\ &\times \int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2) \frac{1}{Z_{\beta, n}} (1 - g(r_2))_+ (g(r_1) - 1)_+ (L^2 T)^{-1} \\ &\times \int_0^T dt (D_1^2 + D_2^2 + (D'_1)^2 + (D'_2)^2). \end{aligned}$$

We start with the apparently worst term  $D_2$ .

*Estimates on  $D_2$ .* We point out that the term  $D_2$  given in (4.12) is complicated since  $n, n_1, n_2$  and  $u$  are not independent, and the components of  $u$  satisfy the constraint  $u_1 + \dots + u_n = 1$ .

**Lemma 4.1 (Remainder  $D_2$ ).** *There is a positive constant  $C_2$  such that for all  $T > 0, \lambda > 0$  and  $L > 0$  it follows that*

$$(4.16) \quad \begin{aligned} \mathcal{R}_2 &:= v_0^2 \sum_{n=1}^\infty p_n(T_*) (L^2 T)^{-1} \int_{\mathbb{R}_*^n} (n-1)! d^n u \delta \left( \sum_{k=1}^n u_k - 1 \right) \int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2) \int_0^T dt \\ &\times |D_2|^2 \frac{1}{Z_{\beta, n}} (1 - g(r_2))_+ (g(r_1) - 1)_+ \leq C_2 T_* L_*^{-2} T_*^{-1/2} \ln(T_* + 3), \end{aligned}$$

where the probability measure  $v_2$  is given in (4.4), the function  $g$  is given in (4.6) and the normalisation constant  $Z_{\beta, n}$  is given in (4.7).

**Proof.** By Fubini's theorem the left hand-side of (4.16) can be written as

$$(4.17) \quad \begin{aligned} \mathcal{R}_2 &= v_0^2 \sum_{n=1}^{\infty} p_n(T_*) \frac{L^{-2}T}{Z_{\beta,n}} \int_0^T dt \int_{\mathbb{R}_*} v_2(dr_1) (g(r_1) - 1)_+ \int_{\mathbb{R}_*} v_2(dr_2) r_2^2 (1 - g(r_2))_+ \\ &\quad \times \int_{\mathbb{R}_*^n} (n-1)! d^n u \delta \left( \sum_{k=1}^n u_k - 1 \right) \Big|_{\min\{n_1, n_2\} < k \leq \max\{n_1, n_2\}} \sum_{k=1}^n (-1)^{k-1} u_k \Big|^2. \end{aligned}$$

For each  $i \in \{1, 2\}$  recall that  $n_i = n_i(t) = N(t; Tr_i u) \leq n$  and note that if  $n_1 = n_2$ , then  $D_2 = 0$ . By definition of  $n_1$  and  $n_2$  we have

$$t_{n_i}(Tr_i u) \leq t < t_{n_i+1}(Tr_i u) \quad \text{for each } i \in \{1, 2\},$$

which reads as

$$Tr_i \sum_{k=1}^{n_i} u_k \leq t < Tr_i \sum_{k=1}^{n_i+1} u_k \quad \text{for each } i \in \{1, 2\}.$$

Since the integration with respect to the measures  $v_2(dr_1)$  and  $v_2(dr_2)$  are on the set  $\mathbb{R}_*$ , we assume that  $r_1 > 0$  and  $r_2 > 0$  and then we have

$$(4.18) \quad \sum_{k=1}^{n_i} u_k \leq \frac{t}{Tr_i} < \sum_{k=1}^{n_i+1} u_k \quad \text{for each } i \in \{1, 2\}.$$

Without loss of generality we may assume that  $n_1 < n_2$ . We rewrite the preceding inequalities as follows

$$(4.19) \quad \sum_{k=1}^{n_1+1} u_k + \sum_{k=n_1+2}^{n_2} u_k \leq \frac{t}{Tr_2} < \sum_{k=1}^{n_1+1} u_k + \sum_{k=n_1+2}^{n_2} u_k + u_{n_2+1}.$$

The idea is to do a change of variables for suitable indexes as follows. For the indexes  $\{1, 2, \dots, n_1, n_1 + 1\} \cup \{n_2 + 1, n_2 + 2, \dots\}$  we do not do a change of variables. Also, in the case that  $n_2 - n_1$  is an even number, we do not do a change of variable for the index  $k = n_2$ . To be more precise, let  $\mathcal{I} := \{n_1 + 2, n_1 + 3, \dots, n_2\}$  be a set of indexes and define  $K(n_1, n_2) = \lfloor \frac{n_2 - n_1 - 1}{2} \rfloor$ . For  $K \geq 1$  we set

$$(4.20) \quad \hat{u}_k := u_{2k+n_1} + u_{2k+n_1+1} \quad \text{and} \quad \omega_k := \frac{u_{2k+n_1} - u_{2k+n_1+1}}{\hat{u}_k}$$

for each  $k \in \{1, \dots, K\}$ . Observe that

$$(4.21) \quad \sum_{k \in \mathcal{I}} u_k = \sum_{k=1}^K \hat{u}_k + u_{n_2} \mathbb{1}_{\{n_2 - n_1 - 1 \text{ is odd}\}}.$$

In fact, the preceding equality holds even if  $K = 0$  recalling the convention about empty sums. We point out that right-hand side of (4.21) does not depend on  $(\omega_k : 1 \leq k \leq K)$ .

In the sequel, for  $j_1 < j_2$  we show that

$$(4.22) \quad \left\langle \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \right\rangle_{\nu(d^n u)} = \frac{1}{3} \int_{\mathbb{R}_*^n} (n-1)! d^n u \delta \left( \sum_{k=1}^n u_k - 1 \right) \sum_{k=1}^K \hat{u}_k^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}},$$

where  $\nu(d^n u) = d^n u \cdot (n-1)! \delta(t_n(u) - 1)$ . We stress that  $n_i = N(t; Tr_i u)$ . It is clear that the restriction  $n_1 = j_1$  does not depend on  $\omega_1, \dots, \omega_K$  since we do not do any change of variables for  $k \leq j_1 + 1$ . However, a priori the condition  $n_2 = j_2$  could depend on  $\omega_1, \dots, \omega_K$ . By (4.19) and (4.21) we find that it is not the case. Using the symmetry, in the case  $n_2 < n_1$  we set  $\mathcal{I} := \{n_2 + 2, n_2 + 3, \dots, n_1\}$  be a set of indexes and define  $K(n_1, n_2) = \lfloor \frac{n_1 - n_2 - 1}{2} \rfloor$ .

On both sides of (4.22) we can replace the random variable  $K(n_1, n_2)$  by the fixed value  $K(j_1, j_2)$ . For convenience, we denote  $K(j_1, j_2)$  by  $K$  in the following. We note that (4.22) is trivially true for  $K = 0$ . Hence, we only need to prove

for  $K \geq 1$ . We start by observing

$$\begin{aligned} & \left\langle \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \right\rangle_{\nu(\mathbb{d}^n \mathbf{u})} \\ &= \int_{\mathbb{R}_+^n} (n-1)! \mathbb{d}^n \mathbf{u} \delta \left( \sum_{k=1}^K \hat{u}_k + \sum_{k \in \mathcal{J}} u_k - 1 \right) \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}}, \end{aligned}$$

where

$$\mathcal{J} := \begin{cases} (\{1, 2, \dots, n\} \setminus \mathcal{I}) \cup \{n_2\} & \text{if } n_2 - n_1 \text{ is an even number,} \\ \{1, 2, \dots, n\} \setminus \mathcal{I} & \text{if } n_2 - n_1 \text{ is an odd number.} \end{cases}$$

The following computations are based on properties of the  $\delta$ -constraint measures whose construction and main properties are discussed in detail in Appendix C. We apply Lemma C.1, in particular, we wish to approximate the measure by ordinary Lebesgue integrals using the mollifier functions  $\Phi_\varepsilon(\cdot; 1)$  defined there. To do this rigorously, it is necessary that the rest of the integrand is continuous: for this reason, we first need to approximate the indicator function  $\mathbb{1}_{\{n_1(u)=j_1, n_2(u)=j_2\}}$  by a sequence of continuous functions (for fixed  $t$ ,  $r_1$ , and  $r_2$ ). This can be done even monotonously using non-negative functions and without introducing a dependence on the variables  $\omega_1, \dots, \omega_K$ , and thus the limits of the sequence can be taken in the integrands, in the beginning and at the end of the computation, using monotone convergence theorem. To simplify the notation, we make the computations using the notation  $\mathbb{1}_{\{n_1(u)=j_1, n_2(u)=j_2\}}$  for one of its continuous approximants.

Replacing the  $\delta$ -function by the mollifier  $\Phi_\varepsilon$  and after using the above continuous approximant for the integrand, we may perform the change of variables (4.20) using standard rules of Lebesgue measures. This yields

$$\begin{aligned} & \int_{\mathbb{R}_+^n} \mathbb{d}^n \mathbf{u} \Phi_\varepsilon \left( \sum_{k=1}^K \hat{u}_k + \sum_{k \in \mathcal{J}} u_k - 1 \right) \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \\ &= \int_{\mathbb{R}^n} \mathbb{d}\hat{u}_1 \mathbb{d}\omega_1 \cdots \mathbb{d}\hat{u}_K \mathbb{d}\omega_K \prod_{k \in \mathcal{J}} \mathbb{d}u_k \Phi_\varepsilon \left( \sum_{k=1}^K \hat{u}_k + \sum_{k \in \mathcal{J}} u_k - 1 \right) \\ & \quad \times \prod_{k=1}^K \left( \mathbb{1}_{\{\hat{u}_k > 0\}} \mathbb{1}_{\{|\omega_k| < 1\}} \frac{\hat{u}_k}{2} \right) \prod_{k \in \mathcal{J}} \mathbb{1}_{\{u_k > 0\}} \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \\ &= \frac{1}{3} \int_{\mathbb{R}_+^K \times \mathbb{R}_+^{|\mathcal{J}|}} \mathbb{d}\hat{u}_1 \cdots \mathbb{d}\hat{u}_K \prod_{k \in \mathcal{J}} \mathbb{d}u_k \Phi_\varepsilon \left( \sum_{k=1}^K \hat{u}_k + \sum_{k \in \mathcal{J}} u_k - 1 \right) \\ & \quad \times \prod_{k=1}^K (\mathbb{1}_{\{\hat{u}_k > 0\}} \hat{u}_k) \prod_{k \in \mathcal{J}} \mathbb{1}_{\{u_k > 0\}} \sum_{k=1}^K \hat{u}_k^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}}, \end{aligned}$$

where the last equality holds by Fubini's theorem, the fact that the variables  $\hat{u}_1, \dots, \hat{u}_K$  and  $\omega_1, \dots, \omega_K$  are independent, and that the argument in the function  $\Phi_\varepsilon$  does not depend on  $\omega_1, \dots, \omega_K$  (see (4.21)). Also,  $|\mathcal{J}|$  denotes the cardinality of  $\mathcal{J}$ . Now, we reintroduce the variables  $\omega_1, \dots, \omega_K$  yielding that final integral is equal to

$$\begin{aligned} & \frac{1}{3} \int_{\mathbb{R}^n} \mathbb{d}\hat{u}_1 \mathbb{d}\omega_1 \cdots \mathbb{d}\hat{u}_K \mathbb{d}\omega_K \prod_{k \in \mathcal{J}} \mathbb{d}u_k \Phi_\varepsilon \left( \sum_{k=1}^K \hat{u}_k + \sum_{k \in \mathcal{J}} u_k - 1 \right) \\ & \quad \times \prod_{k=1}^K (\mathbb{1}_{\{\hat{u}_k > 0\}} \hat{u}_k) \prod_{k=1}^K \left( \mathbb{1}_{\{|\omega_k| < 1\}} \frac{1}{2} \right) \prod_{k \in \mathcal{J}} \mathbb{1}_{\{u_k > 0\}} \sum_{k=1}^K \hat{u}_k^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}}. \end{aligned}$$

Inverting the change of variables and taking  $\varepsilon \rightarrow 0$  followed by the limit for the continuous approximants, we deduce (4.22).



For notational convenience we set  $K = 0$  when  $n_1 = n_2$ . By the Law of total probability we have

$$\begin{aligned}
(4.23) \quad \left\langle \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \right\rangle_{\nu(\mathbb{d}^n \mathbf{u})} &= \sum_{j_1=1, j_2=1}^n \left\langle \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \right\rangle_{\nu(\mathbb{d}^n \mathbf{u})} \\
&= \frac{1}{3} \sum_{j_1=1, j_2=1}^n \int_{\mathbb{R}_*^n} \nu(\mathbb{d}^n \mathbf{u}) \left( \sum_{k=1}^K \hat{u}_k^2 \right) \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \\
&\leq \frac{1}{3} \sum_{j_1=1, j_2=1}^n \int_{\mathbb{R}_*^n} \nu(\mathbb{d}^n \mathbf{u}) \max_{1 \leq k \leq K} \hat{u}_k \left( \sum_{k=1}^K \hat{u}_k \right) \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \\
&\leq \frac{2}{3} \sum_{j_1=1, j_2=1}^n \int_{\mathbb{R}_*^n} \nu(\mathbb{d}^n \mathbf{u}) \max_{1 \leq k \leq n} u_k \left( \sum_{k=1}^K \hat{u}_k \right) \mathbb{1}_{\{n_1=j_1, n_2=j_2\}}.
\end{aligned}$$

Recall that  $r_1 > 0$  and  $r_2 > 0$ . For  $n_1 < n_2$ , (4.18) reads

$$t_{n_1}(\mathbf{u}) \leq \frac{t}{Tr_1} < t_{n_1+1}(\mathbf{u}) \quad \text{and} \quad t_{n_2}(\mathbf{u}) \leq \frac{t}{Tr_2} < t_{n_2+1}(\mathbf{u}),$$

which with the help of (4.21) implies

$$\sum_{k=1}^K \hat{u}_k + u_{n_2} \mathbb{1}_{\{n_2 - n_1 - 1 \text{ is odd}\}} = \sum_{k \in \mathcal{I}} u_k = t_{n_2}(\mathbf{u}) - t_{n_1+1}(\mathbf{u}) \leq \frac{t}{Tr_2} - \frac{t}{Tr_1}.$$

Analogously, for  $n_2 < n_1$  we obtain

$$\sum_{k \in \mathcal{I}} u_k = t_{n_1}(\mathbf{u}) - t_{n_2+1}(\mathbf{u}) \leq \frac{t}{Tr_1} - \frac{t}{Tr_2}.$$

Combining the preceding inequality with (4.23) and Fubini's theorem yields

$$\begin{aligned}
(4.24) \quad \left\langle \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \right\rangle_{\nu(\mathbb{d}^n \mathbf{u})} &\leq \frac{2}{3} \sum_{j_1=1, j_2=1}^n \int_{\mathbb{R}_*^n} \nu(\mathbb{d}^n \mathbf{u}) \max_{1 \leq k \leq n} u_k \left| \frac{t}{Tr_2} - \frac{t}{Tr_1} \right| \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \\
&= \frac{2t}{3T} \left| \frac{1}{r_2} - \frac{1}{r_1} \right| \sum_{j_1=1, j_2=1}^n \int_{\mathbb{R}_*^n} \nu(\mathbb{d}^n \mathbf{u}) \left( \max_{1 \leq k \leq n} u_k \right) \mathbb{1}_{\{n_1=j_1, n_2=j_2\}} \\
&= \frac{2t}{3T} \left| \frac{1}{r_2} - \frac{1}{r_1} \right| \left\langle \max_{1 \leq k \leq n} u_k \right\rangle_{\nu(\mathbb{d}^n \mathbf{u})}.
\end{aligned}$$

By Lemma A.5 there exists a positive constant  $C$  such that

$$\left\langle \max_{1 \leq k \leq n} u_k \right\rangle_{\nu(\mathbb{d}^n \mathbf{u})} \leq \frac{C \ln(n+1)}{n} \quad \text{for all } n \in \mathbb{N}.$$

Combining the preceding inequalities in (4.24) we obtain

$$(4.25) \quad \left\langle \left( \sum_{k=1}^K \hat{u}_k \omega_k \right)^2 \right\rangle_{\nu(\mathbb{d}^n \mathbf{u})} \leq \frac{2Ct}{3T} \left| \frac{1}{r_2} - \frac{1}{r_1} \right| \frac{\ln(n+1)}{n}.$$

Inserting (4.25) in (4.17) we have

$$(4.26) \quad \mathcal{R}_2 \leq \frac{Cv_0^2 L^{-2} T^2}{3} \sum_{n=1}^{\infty} p_n(T_*) \frac{\ln(n+1)}{n} \frac{1}{Z_{\beta, n}} \int_{\mathbb{R}_*} \nu_2(\mathrm{d}r_1) (g(r_1) - 1)_+ \int_{\mathbb{R}_*} \nu_2(\mathrm{d}r_2) r_2^2 (1 - g(r_2))_+ \left| \frac{1}{r_2} - \frac{1}{r_1} \right|.$$

Note that

$$(4.27) \quad \int_{\mathbb{R}_*} v_2(dr_2)r_2^2(1-g(r_2))_+ \leq \langle r^2 \rangle_{v_2} = \frac{1}{n} + 1 \leq 2.$$

We need to estimate

$$\frac{1}{Z_{\beta,n}} \int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2)(1-g(r_2))_+(g(r_1)-1)_+ r_2 \left| 1 - \frac{r_2}{r_1} \right|.$$

Here, it is important that  $\beta = n$ . Since  $r \mapsto g(r)$  is strictly monotone increasing from  $g(0) < 1$  to  $g(1) > 1$ , there is a unique  $r_* \in (0, 1)$  such that  $g(r_*) = 1$ , and then  $g(r) < 1$  if and only if  $r < r_*$ . Since  $g(r) = 0$  for  $r > 1$ , the above integrand can be non-zero only if  $0 < r_2 < r_*$  and  $r_* < r_1 \leq 1$ . In particular, then  $r_2 < r_1$ , and we have

$$0 < 1 - \frac{r_2}{r_1} \leq 1 - r_2.$$

Since  $Z_{\beta,n} = \int_{\mathbb{R}_*} v_2(dr)(g(r)-1)_+ > 0$  and  $\beta = n$ , we have  $\langle r \rangle_{v_2} = 1$  and  $\langle (1-r)^2 \rangle_{v_2} = \frac{1}{n}$ . Then

$$(4.28) \quad \begin{aligned} & \frac{1}{Z_{\beta,n}} \int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2)(1-g(r_2))_+(g(r_1)-1)_+ r_2 \left| 1 - \frac{r_2}{r_1} \right| \\ & \leq \int_{\mathbb{R}_*} v_2(dr_2)(1-g(r_2))_+ r_2 |1-r_2| \leq \langle r^2 \rangle_{v_2}^{1/2} \langle (1-r)^2 \rangle_{v_2}^{1/2} \leq \frac{\sqrt{2}}{\sqrt{n}}. \end{aligned}$$

Combining (4.27) and (4.28) in (4.26) we deduce

$$\mathcal{R}_2 \leq \frac{\sqrt{2}Cv_0^2L^{-2}T^2}{3} \sum_{n=1}^{\infty} p_n(T_*) \frac{\ln(n+1)}{n^{3/2}}.$$

By Lemma A.1 in Appendix A and the fact that  $T_* = \lambda T$  and  $L_* = |v_0|^{-1}\lambda L$  yields the existence of a constant  $\tilde{C}$  such that

$$\mathcal{R}_2 \leq \frac{\sqrt{2}Cv_0^2L^{-2}T^2}{3} \frac{\tilde{C} \ln(\lambda T + 3)}{(\lambda T)^{3/2}} = 3^{-1}C_2L_*^{-2}T_*^{1/2} \ln(T_* + 3),$$

where  $C_2 = \sqrt{2}C\tilde{C}$ . This finishes the proof of Lemma 4.1. □

*Estimates on  $D'_1$  and  $D'_2$ .* In this subsection we estimate the contribution of the incomplete jumps given in (4.13).

**Lemma 4.2 (Incomplete jumps).** *There is a positive constant  $C_{1,2}$  such that for all  $T > 0$ ,  $\lambda > 0$  and  $L > 0$  it follows that*

$$(4.29) \quad \begin{aligned} \mathcal{R}_{1,2} & := v_0^2 \sum_{n=1}^{\infty} p_n(T_*) (L^2T)^{-1} \int_{\mathbb{R}_*^n} (n-1)! d^n u \delta \left( \sum_{k=1}^n u_k - 1 \right) \int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2) \int_0^T dt \\ & \times (|D'_1|^2 + |D'_2|^2) \frac{1}{Z_{\beta,n}} (1-g(r_2))_+(g(r_1)-1)_+ \leq C_{1,2} T_* L_*^{-2} T_*^{-1/2}, \end{aligned}$$

where the probability measure  $v_2$  is given in (4.4), the function  $g$  is given in (4.6) and the normalisation constant  $Z_{\beta,n}$  is given in (4.7).

**Proof.** We start by noticing that (4.8) and (4.5) yield  $0 < r_1 \leq 1$  and  $r_2 > 0$ . If  $n_1, n_2 < n$ , the remaining ‘‘incomplete jumps’’ can be estimated by

$$|D'_1| \leq Tr_1 u_{n_1+1}, \quad |D'_2| \leq Tr_2 u_{n_2+1}.$$

By (2.6) and the fact that we are integrating over the measure  $\delta(\sum_{k=1}^n u_k - 1)$ , we observe that  $n_i(t) = n$  occurs if and only if  $t \geq t_n(Tr_i u) = Tr_i \sum_{k=1}^n u_k = Tr_i$ . Then  $0 \leq t - t_n(Tr_i u) = t - Tr_i \leq T(1 - r_i)$ . Therefore, in this case

$|D'_i| \leq T(1 - r_i)$  which yields  $0 < r_2 \leq 1$ . Hence, for  $n_1(t) = n$  we have

$$\frac{1}{L^2 T} \int_{Tr_1}^T dt |D'_1|^2 \leq \frac{T^2}{L^2} (1 - r_1)^3.$$

Similarly, for  $n_2(t) = n$  we obtain

$$\mathbb{1}_{\{r_2 \leq 1\}} \frac{1}{L^2 T} \int_{Tr_2}^T dt |D'_2|^2 \leq \mathbb{1}_{\{r_2 \leq 1\}} \frac{T^2}{L^2} (1 - r_2)^3.$$

For all other times, we estimate by splitting the integral at the points  $0 = t_0(Tr_1 u) < t_1(Tr_1 u) < \dots < t_n(Tr_1 u) = Tr_1$ ,

$$\frac{1}{L^2 T} \int_0^{Tr_1} dt |D'_1|^2 = \frac{1}{L^2 T} \sum_{k=0}^{n-1} \int_{t_k(Tr_1 u)}^{t_{k+1}(Tr_1 u)} dt |D'_1|^2 \leq \frac{1}{L^2 T} \sum_{k=0}^{n-1} T^3 r_1^3 u_{k+1}^3 \leq \frac{T^2}{L^2} \sum_{k=1}^n u_k^3.$$

Analogously, for  $r_2 \leq 1$  we have

$$\frac{1}{L^2 T} \int_0^{Tr_2} dt |D'_2|^2 \leq \frac{1}{L^2 T} \sum_{k=0}^{n-1} T^3 r_2^3 u_{k+1}^3 \leq \frac{T^2}{L^2} r_2^3 \sum_{k=1}^n u_k^3.$$

Finally, if  $r_2 > 1$  and denoting  $n' := N(T; Tr_2 u) < n$ , we estimate

$$\frac{1}{L^2 T} \int_0^T dt |D'_2|^2 \leq \frac{1}{L^2 T} \left( \sum_{k=0}^{n'-1} \int_{t_k(Tr_2 u)}^{t_{k+1}(Tr_2 u)} dt T^2 r_2^3 u_{k+1}^2 + \int_{t_{n'}(Tr_2 u)}^{t_{n'+1}(Tr_2 u)} dt T^2 r_2^2 u_{n'+1}^2 \right) \leq \frac{T^2}{L^2} r_2^3 \sum_{k=1}^n u_k^3.$$

Collecting all pieces together we obtain

$$\frac{1}{L^2 T} \int_0^T dt (|D'_1|^2 + |D'_2|^2) \leq \frac{T^2}{L^2} \sum_{k=1}^n u_k^3 (1 + r_2^3) + \frac{T^2}{L^2} ((1 - r_1)^3 + (1 - r_2)_+^3).$$

This needs to be integrated over the measure for  $d^n u$ ,  $v_2(dr_1)$ ,  $v_2(dr_2)$  in (4.29). Since each term is non-negative and depends only on either  $r_1$  or  $r_2$ , we use their known simple marginal distributions. By Lemma A.3 in Appendix A we have

$$\langle u_k^3 \rangle = \frac{6}{n(n+1)(n+2)} \quad \text{for all } k \in \{1, \dots, n\}.$$

Moreover, since  $v_2$  has Gamma distribution with parameters  $n$  and  $\beta$ , straightforward computations yields

$$\langle (1 - r_1)_+^3 \rangle_{v_1} = \frac{6}{(n+1)(n+2)(n+3)} \quad \text{and} \quad \langle r_2^3 \rangle_{v_2} = \frac{n(n+1)(n+2)}{\beta^3}.$$

By the Cauchy–Schwarz inequality we have  $\langle (1 - r_2)_+^3 \rangle_{v_2}^2 \leq \langle (1 - r_2)_+^6 \rangle_{v_2} \leq \langle (1 - r_2)^6 \rangle_{v_2}$ . Since  $v_2$  is the Gamma distribution with parameters  $n$  and  $\beta$ , we point out that  $\langle (1 - r_2)^6 \rangle_{v_2}$  can be evaluated explicitly, and is  $O(n^{-3})$  for the choice  $\beta = n$  which localises the mean  $\langle r \rangle_{v_2} = 1$ , see Item 5 of Lemma D.2 in Appendix D. That is, there exists a constant  $C > 0$  such that

$$\langle (1 - r_2)_+^3 \rangle_{v_2} = \int_0^1 dr \frac{1}{(n-1)!} (1-r)^3 n^n r^{n-1} e^{-nr} \leq \frac{C}{n^{3/2}}.$$

Hence,

$$\begin{aligned} & (L^2 T)^{-1} \int_{\mathbb{R}_*^n} (n-1)! d^n u \delta\left(\sum_{k=1}^n u_k - 1\right) \int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2) \int_0^T dt \\ & \quad \times (|D'_1|^2 + |D'_2|^2) \frac{1}{Z_{\beta,n}} (1 - g(r_2))_+ (g(r_1) - 1)_+ \\ & \leq \frac{T^2}{L^2} \frac{6}{(n+1)(n+2)} (1 + n^{-3}(n+2)(n+1)n) + \frac{T^2}{L^2} \left( \frac{6}{(n+1)(n+2)(n+3)} + \frac{C}{n^{3/2}} \right). \end{aligned}$$

We stress that we have already set  $\beta = n$ . Finally, we take the final missing expectation over the Poisson distribution of  $n$  with parameter  $T_*$  and with the help of Lemma A.1 in Appendix A we obtain the existence of a positive constant  $C_{1,2}$  such that

$$\mathcal{R}_{1,2} \leq C_{1,2} T_* L_*^{-2} T_*^{-1/2}.$$

This finishes the proof of Lemma 4.2. □

*Estimates on  $D_1$ .* In this subsection we estimate the contribution of the term  $D_1$  given in (4.11).

**Lemma 4.3 (Complete jumps).** *There is a positive constant  $C_1$  such that for all  $T > 0$ ,  $\lambda > 0$  and  $L > 0$  it follows that*

$$\begin{aligned} \mathcal{R}_1 &:= v_0^2 \sum_{n=1}^{\infty} p_n(T_*) (L^2 T)^{-1} \int_{\mathbb{R}_*^n} (n-1)! d^n u \delta\left(\sum_{k=1}^n u_k - 1\right) \int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2) \int_0^T dt \\ &\times |D_1|^2 \frac{1}{Z_{\beta,n}} (1 - g(r_2))_+ (g(r_1) - 1)_+ \leq C_1 L_*^{-2}, \end{aligned}$$

where the probability measure  $v_2$  is given in (4.4), the function  $g$  is given in (4.6) and the normalisation constant  $Z_{\beta,n}$  is given in (4.7).

**Proof.** By (4.11) we have

$$D_1^2 = (r_1 - r_2)^2 T^2 \sum_{k,k'=1}^n \mathbb{1}_{\{k,k' \leq N(t; T r_1 u)\}} (-1)^{k+k'} u_k u_{k'}.$$

Note that

$$\begin{aligned} \frac{1}{T} \int_0^T dt \mathbb{1}_{\{k,k' \leq N(t; T r_1 u)\}} &= \sum_{m=1}^n \mathbb{1}_{\{k,k' \leq m\}} \frac{1}{T} \int_0^T dt \mathbb{1}_{\{N(t; T r_1 u) = m\}} \\ &= \sum_{m=1}^{n-1} \mathbb{1}_{\{k,k' \leq m\}} r_1 u_{m+1} + 1 - r_1 = 1 - r_1 + \sum_{m=2}^n \mathbb{1}_{\{k,k' \leq m-1\}} r_1 u_m. \end{aligned}$$

Therefore,

$$\frac{1}{T} \int_0^T dt D_1^2 = (r_1 - r_2)^2 T^2 r_1 \sum_{m=2}^n u_m \sum_{k,k'=1}^{m-1} (-1)^{k+k'} u_k u_{k'} + (r_1 - r_2)^2 T^2 (1 - r_1) \sum_{k,k'=1}^n (-1)^{k+k'} u_k u_{k'}.$$

Taking an expectation over  $(n-1)! d^n u \delta(\sum_{k=1}^n u_k - 1)$  and using Lemma A.3 in Appendix A we obtain

$$\begin{aligned} \mathcal{J}_1 &:= \int_{\mathbb{R}_*^n} (n-1)! d^n u \delta\left(\sum_{k=1}^n u_k - 1\right) \frac{1}{L^2 T} \int_0^T dt D_1^2 \\ &= (r_1 - r_2)^2 T^2 L^{-2} r_1 \sum_{m=2}^n \sum_{k,k'=1}^{m-1} (-1)^{k+k'} \frac{\mathbb{1}_{\{k'=k\}} + 1}{n(n+1)(n+2)} \\ &\quad + (r_1 - r_2)^2 T^2 L^{-2} (1 - r_1) \sum_{k,k'=1}^n (-1)^{k+k'} \frac{\mathbb{1}_{\{k'=k\}} + 1}{n(n+1)}. \end{aligned}$$

Note that for any  $m$  we have  $\sum_{k=1}^{m-1} (-1)^k = \pm 1$ . Thus,

$$\mathcal{J}_1 \leq (r_1 - r_2)^2 T^2 L^{-2} r_1 \frac{2}{n+2} + (r_1 - r_2)^2 T^2 L^{-2} (1 - r_1) \frac{2}{n+1} \leq \frac{2}{n+1} (r_1 - r_2)^2 T^2 L^{-2}.$$

Since

$$(r_1 - r_2)^2 \leq 2((1 - r_2)^2 + (1 - r_1)^2),$$

$$\langle (1 - r_1)^2 \rangle_{v_1} = \frac{2}{(n+1)(n+2)} \quad \text{and} \quad \langle (1 - r_2)^2 \rangle_{v_2} = \frac{1}{n},$$

we obtain the existence of a positive constant  $C$  such that

$$\int_{\mathbb{R}_*} v_2(dr_1) \int_{\mathbb{R}_*} v_2(dr_2) \frac{1}{Z_{\beta,n}} (1 - g(r_2))_+ (g(r_1) - 1)_+ \mathcal{J}_1 \leq C \frac{T^2 L^{-2}}{n^2}.$$

Finally, we take the final missing expectation over the Poisson distribution of  $n$  with parameter  $T_*$  and with the help of Lemma A.1 in Appendix A we obtain the existence of a positive constant  $C_1$  such that

$$\mathcal{R}_1 \leq C_1 L_*^{-2}.$$

This finishes the proof of Lemma 4.3.  $\square$

In the sequel, we stress the fact that Proposition 3.1 is just a consequence of what we have already proved up to here.

**Proof of Proposition 3.1.** Combining Lemma 4.1, Lemma 4.2 and Lemma 4.3 in (4.15) implies (3.8). This finishes the proof of Proposition 3.1.  $\square$

## 5. Proof of Proposition 3.2: Synchronous coupling

In this section we prove of Proposition 3.2. Recall that  $\mathbf{u} = (u_j : j \in \mathbb{N})$  is a sequence of i.i.d. random variables with exponential distribution of parameter one defined in the probability space  $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ . For each (deterministic)  $n \in \mathbb{N}$ , (3.5) defines

$$Y(t; n, \mathfrak{d}^{(n)}) = L^{-1} v_0 \sum_{k=1}^M (-1)^{k-1} \delta_k^{(n)} + L^{-1} v_0 (-1)^M (t - t_M(\mathfrak{d}^{(n)})), \quad \text{with } M := N(t; \mathfrak{d}^{(n)})$$

for any  $t \geq 0$ , where  $\mathbf{u}^{(n)} := (u_k^{(n)} : k \in \mathbb{N})$ ,  $\delta^{(n)} := (\delta_k^{(n)} : k \in \mathbb{N})$  and  $t_M(\mathfrak{d}^{(n)})$  are defined in (3.2), (3.3) and (3.4), respectively. Moreover, for  $n \in \mathbb{N} \setminus \{1\}$  we define the Càdlàg version of the random walk  $(\eta_k^*(\mathfrak{s}) : k \in \mathbb{N}_0)$  given in (3.9). That is,

$$Z(t; n, \mathfrak{s}) = \sum_{k=1}^{\tilde{n}} \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{s})\}} \eta_k^*(\mathfrak{s}) \quad \text{for any } t \geq 0,$$

where  $\tilde{n} = \lfloor n/2 \rfloor$  and  $\mathfrak{s} = (s_1, s_2, \dots) = (1/\lambda)(u_1, u_2, \dots)$  is distributed according to the infinitely product measure of exponential distributions with parameter  $\lambda > 0$ .

We define the auxiliary process

$$\tilde{Z}(t; n, \mathfrak{s}) = \frac{T_*}{n} \sum_{k=1}^{\tilde{n}} \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)}(\mathfrak{s}))\}} \eta_k^*(\mathfrak{s}) \quad \text{for any } t \geq 0,$$

where the sequence  $(w_j^{(n)}(\mathfrak{s}) : j \in \mathbb{N})$  is defined by setting

$$w_j^{(n)}(\mathfrak{s}) = \frac{T}{n} u_j^{(n)} = \frac{T_*}{n} s_j^{(n)} \quad \text{for any } j \in \mathbb{N}.$$

For notational convenience we write  $(w_j^{(n)} : j \in \mathbb{N})$  instead of  $(w_j^{(n)}(\mathfrak{s}) : j \in \mathbb{N})$ . For  $n \in \{0, 1\}$  we set  $Z(t; n, \mathfrak{s}) = \tilde{Z}(t; n, \mathfrak{s}) = 0$  for all  $t \geq 0$ . Analogously to (3.6) and (3.10) we define  $\tilde{Z} := (\tilde{Z}(t; K, \mathfrak{s}) : t \geq 0)$ , where  $K$  is a random variable with Poisson distribution with parameter  $T_* = \lambda T$  defined in the probability space  $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$  and independent of  $\mathbf{u}$ . Then we let  $(\Omega, \mathcal{F}, \mathbb{Q}) := (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$  be the product probability space of  $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$  and  $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ .

We compare the processes  $Y, Z$  and  $\tilde{Z}$  as follows:

$$(5.1) \quad \mathcal{W}_2(\mathbb{Y}_{[0,T]}, \mathbb{Z}_{[0,T]}) \leq \mathcal{W}_2(\mathbb{Y}_{[0,T]}, \tilde{\mathbb{Z}}_{[0,T]}) + \mathcal{W}_2(\tilde{\mathbb{Z}}_{[0,T]}, \mathbb{Z}_{[0,T]}).$$

The following lemma provides an estimate for the second term of the right-hand side of (5.1).

**Lemma 5.1.** *For any  $L > 0, T > 0, \lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$\mathcal{W}_2^2(\tilde{\mathbb{Z}}_{[0,T]}, \mathbb{Z}_{[0,T]}) \leq 28T_*L_*^{-2}T_*^{-1/2}(1 + T_*^{-1/2}).$$

In particular, for  $T_* \geq 1$  it follows that

$$\mathcal{W}_2^2(\tilde{\mathbb{Z}}_{[0,T]}, \mathbb{Z}_{[0,T]}) \leq 56T_*L_*^{-2}T_*^{-1/2}.$$

**Proof.** By definition (2.10), the tower property (Item (iv)) of Theorem 8.14 in [25] and Fubini’s theorem we have

$$(5.2) \quad \begin{aligned} \mathcal{W}_2^2(\mathbb{Z}_{[0,T]}, \tilde{\mathbb{Z}}_{[0,T]}) &\leq \frac{1}{T} \sum_{n=2}^{\infty} p_n(T_*) \int_0^T \mathbb{E}_{\mathbb{Q}}[|Z(t; n, \mathfrak{s}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 | K = n] dt \\ &= \frac{1}{T} \sum_{n=2}^{\infty} p_n(T_*) \int_0^T \mathbb{E}_{\mathbb{P}_1}[|Z(t; n, \mathfrak{s}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2] dt \\ &= \frac{1}{T} \sum_{n=2}^{\infty} p_n(T_*) \mathbb{E}_{\mathbb{P}_1} \left[ \int_0^T |Z(t; n, \mathfrak{s}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 dt \right]. \end{aligned}$$

In the sequel, we estimate

$$\frac{1}{T} \int_0^T dt |Z(t; n, \mathfrak{s}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2.$$

Since  $|x + y|^2 \leq 2x^2 + 2y^2$  for any  $x, y \in \mathbb{R}$ , we have

$$(5.3) \quad \begin{aligned} |Z(t; n, \mathfrak{s}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 &\leq 2 \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) (\mathbb{1}_{\{t \geq t_{2k}(\mathfrak{s})\}} - \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}}) \right|^2 \\ &\quad + 2 \left( 1 - \frac{T_*}{n} \right)^2 \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \right|^2. \end{aligned}$$

We start with the estimate of second term of the right-hand side of the preceding inequality. For any  $\ell \in \{1, \dots, \tilde{n}\}$  set

$$(5.4) \quad \hat{s}_\ell := s_{2\ell} + s_{2\ell-1} \quad \text{and} \quad \omega_\ell := \frac{s_{2\ell} - s_{2\ell-1}}{\hat{s}_\ell}$$

and observe that

$$(5.5) \quad \eta_\ell^*(\mathfrak{s}) = -\frac{v_0}{L} \hat{s}_\ell \omega_\ell \quad \text{for any } \ell \in \{1, \dots, \tilde{n}\}.$$

We stress that for any  $k \in \mathbb{N}$ ,  $2k$  is an even number and, therefore, it is possible to apply pairwise Lemma D.1 in Appendix D. We find that  $t_{2k}(\mathfrak{w}^{(n)})$  is independent of all  $\omega_\ell, \ell \in \{1, \dots, \tilde{n}\}$ . Since

$$\left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \right|^2 = \sum_{k=1}^{\tilde{n}} \sum_{k'=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) \eta_{k'}^*(\mathfrak{s}) \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \mathbb{1}_{\{t \geq t_{2k'}(\mathfrak{w}^{(n)})\}},$$

Item (i) and Item (iii) of Lemma D.1 in Appendix D imply

$$\left( 1 - \frac{T_*}{n} \right)^2 \left\langle \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \right|^2 \right\rangle_\omega = \left( 1 - \frac{T_*}{n} \right)^2 \sum_{k=1}^{\tilde{n}} \sum_{k'=1}^{\tilde{n}} \langle \eta_k^*(\mathfrak{s}) \eta_{k'}^*(\mathfrak{s}) \rangle_\omega \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \mathbb{1}_{\{t \geq t_{2k'}(\mathfrak{w}^{(n)})\}}$$

$$\begin{aligned}
&= \left(1 - \frac{T_*}{n}\right)^2 \sum_{k=1}^{\tilde{n}} \langle |\eta_k^*(\mathfrak{s})|^2 \rangle_{\omega} \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \\
&\leq \left(1 - \frac{T_*}{n}\right)^2 \langle |\eta_1^*(\mathfrak{s})|^2 \rangle_{\omega} \tilde{n},
\end{aligned}$$

where  $\omega$  denotes the product measure given by the law of the random vector  $(\omega_1, \dots, \omega_{\tilde{n}})$ . Combining the above inequality with Fubini's theorem, (5.5), Lemma D.1 in Appendix D, and Item 4 of Lemma D.2 in Appendix D we obtain

$$\begin{aligned}
\left(1 - \frac{T_*}{n}\right)^2 \frac{1}{T} \int_0^T dt \left\langle \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \right|^2 \right\rangle_{\hat{s}, \omega} &\leq \left(1 - \frac{T_*}{n}\right)^2 \langle |\eta_1^*(\mathfrak{s})|^2 \rangle_{\hat{s}, \omega} \tilde{n} \\
&\leq \left(1 - \frac{T_*}{n}\right)^2 \frac{v_0^2}{L^2 \lambda^2} n = \frac{T_*}{L_*^2} \left(1 - \frac{T_*}{n}\right)^2 \frac{n}{T_*},
\end{aligned}$$

where  $\hat{s}$  denote the product measure given by the law of the random vector  $(\hat{s}_1, \dots, \hat{s}_{\tilde{n}})$ . We observe that

$$\left(1 - \frac{T_*}{n}\right)^2 \frac{n}{T_*} = \left(\frac{n}{T_*} - 1\right) + \left(\frac{T_*}{n} - 1\right) \leq \frac{1}{T_*} |n - T_*| + \frac{1}{n} |T_* - n|.$$

Then by the Cauchy Schwarz inequality, Lemma A.1 in Appendix A and Lemma D.3 in Appendix D we deduce

$$\begin{aligned}
\left\langle \left(1 - \frac{T_*}{n}\right)^2 \frac{n}{T_*} \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} &\leq \frac{1}{T_*} \langle |n - T_*| \mathbb{1}_{\{n \geq 2\}} \rangle_{\text{Po}(T_*)} + \left\langle |n - T_*| \frac{1}{n} \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \\
&\leq \frac{1}{T_*} \langle (n - T_*)^2 \rangle_{\text{Po}(T_*)}^{1/2} + \langle (n - T_*)^2 \rangle_{\text{Po}(T_*)}^{1/2} \langle n^{-2} \mathbb{1}_{\{n \geq 1\}} \rangle_{\text{Po}(T_*)}^{1/2} \\
&\leq \frac{T_*^{1/2}}{T_*} + T_*^{1/2} \sqrt{\frac{16}{T_*^2}} = 5T_*^{-1/2}.
\end{aligned}$$

Collecting all pieces together we obtain

$$(5.6) \quad \left\langle \left(1 - \frac{T_*}{n}\right)^2 \frac{1}{T} \int_0^T dt \left\langle \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}} \right|^2 \right\rangle_{\hat{s}, \omega} \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \leq 5 \frac{T_*}{L_*^2} T_*^{-1/2}.$$

We continue with the estimate of

$$\frac{1}{T} \int_0^T dt \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) (\mathbb{1}_{\{t \geq t_{2k}(\mathfrak{s})\}} - \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}}) \right|^2.$$

By (5.4), (5.5), Item (i) and Item (iii) of Lemma D.1 in Appendix D we have

$$\begin{aligned}
&\left\langle \frac{1}{T} \int_0^T dt \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s}) (\mathbb{1}_{\{t \geq t_{2k}(\mathfrak{s})\}} - \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}}) \right|^2 \right\rangle_{\omega} \\
&= \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \sum_{\ell'=1}^{\tilde{n}} \hat{s}_{\ell} \hat{s}_{\ell'} \langle \omega_{\ell} \omega_{\ell'} \rangle_{\omega} \frac{1}{T} \int_0^T dt (\mathbb{1}_{\{t > t_{2\ell}(\mathfrak{s})\}} - \mathbb{1}_{\{t > t_{2\ell}(\mathfrak{w}^{(n)})\}}) (\mathbb{1}_{\{t > t_{2\ell'}(\mathfrak{s})\}} - \mathbb{1}_{\{t > t_{2\ell'}(\mathfrak{w}^{(n)})\}}) \\
&= \frac{1}{3} \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_{\ell}^2 \frac{1}{T} \int_0^T dt (\mathbb{1}_{\{t > t_{2\ell}(\mathfrak{s})\}} - \mathbb{1}_{\{t > t_{2\ell}(\mathfrak{w}^{(n)})\}})^2 \\
(5.7) \quad &= \frac{1}{3} \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_{\ell}^2 \frac{1}{T} \int_0^T dt \mathbb{1}_{\{\min\{t_{2\ell}(\mathfrak{s}), t_{2\ell}(\mathfrak{w}^{(n)})\} < t < \max\{t_{2\ell}(\mathfrak{s}), t_{2\ell}(\mathfrak{w}^{(n)})\}\}}
\end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{3} \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} (\max\{t_{2\ell}(\mathfrak{s}), t_{2\ell}(\mathfrak{w}^{(n)})\} - \min\{t_{2\ell}(\mathfrak{s}), t_{2\ell}(\mathfrak{w}^{(n)})\}) \\ &\leq \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})|. \end{aligned}$$

We note that  $\hat{s}_1$  has Gamma distribution with parameters 2 and  $\lambda$  and then Lemma D.2 in Appendix D yields  $\langle \hat{s}_1^4 \rangle_{\hat{s}} = 5!/\lambda^4$ . Then we integrate with respect to  $\hat{s}$  and using the Cauchy Schwarz inequality we obtain

$$\begin{aligned} \left\langle \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})| \right\rangle_{\hat{s}} &\leq \frac{v_0^2}{L^2 T} \sum_{\ell=1}^{\tilde{n}} \langle \hat{s}_\ell^4 \rangle_{\hat{s}}^{1/2} \langle |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})|^2 \rangle_{\hat{s}}^{1/2} \\ &= \frac{v_0^2}{L^2 T} \langle \hat{s}_1^4 \rangle_{\hat{s}}^{1/2} \sum_{\ell=1}^{\tilde{n}} \langle |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})|^2 \rangle_{\hat{s}}^{1/2} \\ &= \frac{v_0^2}{L^2 T} \frac{\sqrt{120}}{\lambda^2} \sum_{\ell=1}^{\tilde{n}} \langle |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})|^2 \rangle_{\hat{s}}^{1/2} \\ &= \frac{\sqrt{120}}{L_*^2} \frac{1}{T} \sum_{\ell=1}^{\tilde{n}} \langle |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})|^2 \rangle_{\hat{s}}^{1/2}. \end{aligned}$$

By Item (ii) of Lemma D.1 in Appendix D we have that each  $\hat{s}_j$  has Gamma distribution with parameters 2 and  $\lambda$ . Since  $\hat{s}_1, \dots, \hat{s}_{\tilde{n}}$  are i.i.d. random variables, we obtain that  $\sum_{j=1}^{\ell} \hat{s}_j = \sum_{j=1}^{2\ell} s_j$  has Gamma distribution with parameters  $2\ell$  and  $\lambda$ . Hence, Item 5 of Lemma D.2 in Appendix D yields

$$\left\langle \left| \sum_{j=1}^{2\ell} s_j \right|^2 \right\rangle_{\hat{s}} = (2\ell + 1)(2\ell)/\lambda^2.$$

We then obtain

$$\langle |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})|^2 \rangle_{\hat{s}} = \left(1 - \frac{T_*}{n}\right)^2 \left\langle \left| \sum_{j=1}^{2\ell} s_j \right|^2 \right\rangle_{\hat{s}} = \left(1 - \frac{T_*}{n}\right)^2 \frac{(2\ell + 1)(2\ell)}{\lambda^2}.$$

Therefore,

$$\left\langle \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} |t_{2\ell}(\mathfrak{s}) - t_{2\ell}(\mathfrak{w}^{(n)})| \right\rangle_{\hat{s}} \leq \frac{\sqrt{120}}{L_*^2} \frac{1}{T} \left|1 - \frac{T_*}{n}\right| \sum_{\ell=1}^{\tilde{n}} \frac{\sqrt{6\ell}}{\lambda} \leq \frac{T_*}{L_*^2} \left|1 - \frac{T_*}{n}\right| \frac{n^2}{T_*^2}.$$

Now, we integrate the right-hand side of the preceding inequality over  $n$  for  $n \geq 2$ . With the help of the Cauchy–Schwarz inequality and Lemma D.3 in Appendix D, we obtain

$$\begin{aligned} \left\langle 9 \frac{T_*}{L_*^2} \left|1 - \frac{T_*}{n}\right| \frac{n^2}{T_*^2} \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} &\leq 9 \frac{T_*}{L_*^2} \frac{1}{T_*^2} \langle |n - T_*|^2 \rangle_{\text{Po}(T_*)}^{1/2} \langle n^2 \mathbb{1}_{\{n \geq 2\}} \rangle_{\text{Po}(T_*)}^{1/2} \\ &\leq 9 \frac{T_*}{L_*^2} \frac{1}{T_*^2} T_*^{1/2} (T_*^2 + T_*)^{1/2} \leq 9 \frac{T_*}{L_*^2} T_*^{-1/2} (1 + T_*^{-1/2}), \end{aligned}$$

where in the last step we used subadditivity inequality for  $x \mapsto x^{1/2}$ . Combining all pieces together we deduce

$$(5.8) \quad \sum_{n \geq 2} p_n(T_*) \left\langle \frac{1}{T} \int_0^T dt \left| \sum_{k=1}^{\tilde{n}} \eta_k^*(\mathfrak{s})(\mathbb{1}_{\{t \geq t_{2k}(\mathfrak{s})\}} - \mathbb{1}_{\{t \geq t_{2k}(\mathfrak{w}^{(n)})\}}) \right|^2 \right\rangle_{\hat{s}, \omega} \leq 9 \frac{T_*}{L_*^2} T_*^{-1/2} (1 + T_*^{-1/2}).$$



Now, by (5.3), (5.6) and (5.8) we obtain

$$\sum_{n \geq 2} p_n(T_*) \frac{1}{T} \int_0^T \mathbb{E}_{\mathbb{P}_1} [ |Z(t; n, \mathfrak{s}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 ] dt \leq 28 \frac{T_*}{L_*^2} T_*^{-1/2} (1 + T_*^{-1/2}).$$

This concludes the proof of Lemma 5.1. □

The following lemma provides an estimate for the first term of the right-hand side of (5.1).

**Lemma 5.2.** *For any  $L > 0$ ,  $T > 0$ ,  $\lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$\mathcal{W}_2^2(\mathbb{Y}_{[0,T]}, \tilde{\mathbb{Z}}_{[0,T]}) \leq 39T_*L_*^{-2}T_*^{-1/2} + 38T_*L_*^{-2}e^{-T_*}(T_* + T_*^2).$$

In addition, for  $T_* > 0$  we obtain

$$\mathcal{W}_2^2(\mathbb{Y}_{[0,T]}, \tilde{\mathbb{Z}}_{[0,T]}) \leq 115T_*L_*^{-2}T_*^{-1/2}.$$

**Proof.** Analogously to (5.2) we obtain

$$(5.9) \quad \begin{aligned} \mathcal{W}_2^2(\mathbb{Y}_{[0,T]}, \tilde{\mathbb{Z}}_{[0,T]}) &\leq \frac{1}{T} \sum_{n=2}^{\infty} p_n(T_*) \mathbb{E}_{\mathbb{P}_1} \left[ \int_0^T |Y(t; n, \mathfrak{s}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 dt \right] \\ &\quad + \frac{1}{T} p_0(T_*) \int_0^T L^{-2} v_0^2 t^2 dt + \frac{1}{T} p_1(T_*) \int_0^T \mathbb{E}_{\mathbb{P}_1} [ |Y(s; 1, \mathfrak{s})|^2 ] ds. \end{aligned}$$

We start with the following claim:

$$(5.10) \quad Y(t_{2k}(\mathfrak{w}^{(n)}); n, \mathfrak{w}^{(n)}) = \tilde{Z}(t_{2k}(\mathfrak{w}^{(n)}); n, \mathfrak{w}^{(n)})$$

for all  $k \in \mathbb{N}$  and  $n \in \mathbb{N} \setminus \{0, 1\}$ . Indeed, note that

$$\begin{aligned} Y(t_{2k}(\mathfrak{w}^{(n)}); n, \mathfrak{w}^{(n)}) &= L^{-1} v_0 \sum_{j=1}^{2k} (-1)^{j-1} w_j^{(n)} \\ &= L^{-1} v_0 \sum_{j=1}^{2k} (-1)^{j-1} \frac{T}{n} u_j^{(n)} \\ &= L^{-1} v_0 \sum_{j=1}^{2k} (-1)^{j-1} \frac{T\lambda}{n} s_j = \frac{T_*}{n} \sum_{j=1}^k \eta_j^*(\mathfrak{s}) \\ &= \tilde{Z}(t_{2k}(\mathfrak{w}^{(n)}); n, \mathfrak{w}^{(n)}), \end{aligned}$$

which yields the claim. Bearing the preceding claim in mind, we consider the following split that overestimate our desired estimate: for all  $n \in \mathbb{N} \setminus \{0, 1\}$

$$(5.11) \quad \begin{aligned} &\frac{1}{T} \int_0^T \mathbb{E}_{\mathbb{P}_1} [ |Y(t; n, \mathfrak{w}^{(n)}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 ] dt \\ &\leq \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \sum_{\ell=1}^{\tilde{n}} \int_{t_{2\ell-2}(\mathfrak{w}^{(n)})}^{t_{2\ell-1}(\mathfrak{w}^{(n)})} |Y(t; n, \mathfrak{w}^{(n)}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 dt \right] \\ &\quad + \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \sum_{\ell=1}^{\tilde{n}} \int_{t_{2\ell-1}(\mathfrak{w}^{(n)})}^{t_{2\ell}(\mathfrak{w}^{(n)})} |Y(t; n, \mathfrak{w}^{(n)}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 dt \right] \\ &\quad + \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathfrak{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathfrak{w}^{(n)})}^T |Y(t; n, \mathfrak{w}^{(n)}) - \tilde{Z}(t; n, \mathfrak{w}^{(n)})|^2 dt \right]. \end{aligned}$$

where  $\tilde{n} = \lfloor n/2 \rfloor$ . For each  $\ell \in \{1, \dots, \tilde{n}\}$  observe that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \int_{t_{2\ell-2}(\mathbf{w}^{(n)})}^{t_{2\ell-1}(\mathbf{w}^{(n)})} |Y(t; n, \mathbf{w}^{(n)}) - \tilde{Z}(t; n, \mathbf{w}^{(n)})|^2 dt \right] &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \int_0^{t_1(\mathbf{w}^{(n)})} |Y(t; n, \mathbf{w}^{(n)})|^2 dt \right] \\ &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \int_0^{w_1^{(n)}} L^{-2} v_0^2 t^2 dt \right] = T^{-1} \frac{1}{3} L^{-2} v_0^2 \langle (w_1^{(n)})^3 \rangle \\ &= T^{-1} \frac{1}{3} L^{-2} v_0^2 \frac{T^3}{n^3} \langle u_1^3 \rangle = \frac{1}{2} \frac{T_*}{L_*^2} \frac{T_*}{n^3}, \end{aligned}$$

due to  $w_1^{(n)} := \frac{T}{n} u_1$  and  $u_1$  having an exponential distribution with parameter one. Analogously,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \int_{t_{2\ell-1}(\mathbf{w}^{(n)})}^{t_{2\ell}(\mathbf{w}^{(n)})} |Y(t; n, \mathbf{w}^{(n)}) - \tilde{Z}(t; n, \mathbf{w}^{(n)})|^2 dt \right] &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \int_{t_1(\mathbf{w}^{(n)})}^{t_2(\mathbf{w}^{(n)})} |L^{-1} v_0 w_1^{(n)} + L^{-1} v_0 (-1)(t - w_1^{(n)})|^2 dt \right] \\ &= \mathbb{E}_{\mathbb{P}_1} \left[ T^{-1} \frac{1}{3} L^{-2} v_0^2 ((w_2^{(n)} - w_1^{(n)})^3 + (w_1^{(n)})^3) \right] \\ &= T^{-1} \frac{1}{3} L^{-2} v_0^2 \langle (w_1^{(n)})^3 \rangle = \frac{1}{2} \frac{T_*}{L_*^2} \frac{T_*}{n^3}. \end{aligned}$$

Hence, for  $\mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) > T\}}$

$$\frac{1}{T} \int_0^T \mathbb{E}_{\mathbb{P}_1} [|Y(t; n, \mathbf{w}^{(n)}) - \tilde{Z}(t; n, \mathbf{w}^{(n)})|^2] dt \leq 2\tilde{n} \frac{1}{2} \frac{T_*}{L_*^2} \frac{T_*}{n^3} \leq \frac{1}{2} \frac{T_*}{L_*^2} \frac{T_*}{n^2}.$$

In what follows, we estimate the possible remainder error given by the last jump, that is,

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}^{(n)}) - \tilde{Z}(t; n, \mathbf{w}^{(n)})|^2 dt \right] \\ &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}^{(n)}) - \tilde{Z}(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right] \\ &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}^{(n)}) - Y(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right], \end{aligned}$$

where the last equality follows by (5.10).

First, we assume that  $n$  is an even number. We note that  $2\tilde{n} = n$  and then

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}^{(n)}) - Y(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right] \\ &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T L^{-2} v_0^2 |t - t_{2\tilde{n}}(\mathbf{w}^{(n)})|^2 dt \right] \\ (5.12) \quad &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \frac{1}{3} L^{-2} v_0^2 (T - t_{2\tilde{n}}(\mathbf{w}^{(n)}))^3 \right] \\ &= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \frac{1}{3} L^{-2} v_0^2 |T - t_{2\tilde{n}}(\mathbf{w}^{(n)})|^3 \right] \\ &\leq \frac{1}{3T} L^{-2} v_0^2 (\mathbb{E}_{\mathbb{P}_1} [|T - t_{2\tilde{n}}(\mathbf{w}^{(n)})|^4])^{3/4}, \end{aligned}$$

where in the last inequality we used Hölder's inequality. We note that

$$(5.13) \quad \frac{1}{3T} L^{-2} v_0^2 (\mathbb{E}_{\mathbb{P}_1} [|T - t_{2\tilde{n}}(\mathbf{w}^{(n)})|^4])^{3/4} \leq C_1 \frac{T_*}{L_*^2} \frac{T_*}{n^{3/2}},$$

where  $C_1 := \langle (u_1 - 1)^4 \rangle^{3/4} = 9^{3/4}$  and  $u_1$  has exponential distribution with parameter one.

Now, we assume that  $n$  is an odd number. We note that  $2\tilde{n} = n - 1$  and then

$$\begin{aligned}
& \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}) - Y(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right] \\
&= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \mathbb{1}_{\{t_n(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}^{(n)}) - Y(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right] \\
&\quad + \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \mathbb{1}_{\{t_n(\mathbf{w}^{(n)}) > T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T L^{-2} v_0^2 |t - t_{2\tilde{n}}(\mathbf{w}^{(n)})|^2 dt \right] \\
&= \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \mathbb{1}_{\{t_n(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^{t_{2\tilde{n}+1}(\mathbf{w}^{(n)})} |Y(t; n, \mathbf{w}^{(n)}) - Y(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right] \\
&\quad + \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \mathbb{1}_{\{t_n(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}+1}(\mathbf{w}^{(n)})}^T L^{-2} v_0^2 |t - t_{2\tilde{n}+1}(\mathbf{w}^{(n)})|^2 dt \right] \\
&\quad + \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \mathbb{1}_{\{t_n(\mathbf{w}^{(n)}) > T\}} \frac{1}{3} L^{-2} v_0^2 (T - t_{2\tilde{n}}(\mathbf{w}^{(n)}))^3 \right] \\
&\leq \frac{1}{2} \frac{T_* T_*}{L_*^2 n^3} + \frac{1}{3T} L^{-2} v_0^2 (\mathbb{E}_{\mathbb{P}_1} [|T - t_{2\tilde{n}+1}(\mathbf{w}^{(n)})|^4])^{3/4} \\
&\quad + \frac{4}{3T} L^{-2} v_0^2 T^3 \left| 1 - \frac{2\tilde{n}}{n} \right|^3 + \frac{4}{3T} L^{-2} v_0^2 \left( \mathbb{E}_{\mathbb{P}_1} \left[ \left| \frac{2\tilde{n}T}{n} - t_{2\tilde{n}}(\mathbf{w}^{(n)}) \right|^4 \right] \right)^{3/4}.
\end{aligned}$$

We note that

$$\frac{1}{3T} L^{-2} v_0^2 (\mathbb{E}_{\mathbb{P}_1} [|T - t_{2\tilde{n}+1}(\mathbf{w}^{(n)})|^4])^{3/4} \leq C_1 \frac{T_* T_*}{L_*^2 n^{3/2}},$$

where  $C_1 = \langle (u_1 - 1)^4 \rangle^{3/4}$ , where  $u_1$  has exponential distribution with parameter one. A straightforward computation yields

$$\frac{4}{3T} L^{-2} v_0^2 T^3 \left| 1 - \frac{2\tilde{n}}{n} \right|^3 = \frac{4}{3} \frac{T_* T_*}{L_*^2 n^3}.$$

We also have

$$\frac{4}{3T} L^{-2} v_0^2 \left( \mathbb{E}_{\mathbb{P}_1} \left[ \left| \frac{2\tilde{n}T}{n} - t_{2\tilde{n}}(\mathbf{w}^{(n)}) \right|^4 \right] \right)^{3/4} \leq 4C_1 \frac{T_* T_*}{L_*^2 n^{3/2}},$$

$$(5.14) \quad \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}^{(n)}) - Y(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right] \leq (2 + C_1) \frac{T_* T_*}{L_*^2 n^3} + 5C_1 \frac{T_* T_*}{L_*^2 n^{3/2}}.$$

Therefore (5.12), (5.13) and (5.14) yield for any  $n \in \mathbb{N} \setminus \{0, 1\}$

$$(5.15) \quad \mathbb{E}_{\mathbb{P}_1} \left[ \frac{1}{T} \mathbb{1}_{\{t_{2\tilde{n}}(\mathbf{w}^{(n)}) \leq T\}} \int_{t_{2\tilde{n}}(\mathbf{w}^{(n)})}^T |Y(t; n, \mathbf{w}^{(n)}) - Y(t_{2\tilde{n}}(\mathbf{w}^{(n)}); n, \mathbf{w}^{(n)})|^2 dt \right] \leq (2 + C_1) \frac{T_* T_*}{L_*^2 n^3} + 5C_1 \frac{T_* T_*}{L_*^2 n^{3/2}}.$$

By (5.9), (5.11), (5.15) and Lemma A.1 in Appendix A we obtain

$$\begin{aligned}
\mathcal{W}_2^2(\mathbb{Y}_{[0, T]}, \tilde{\mathbb{Z}}_{[0, T]}) &\leq \sum_{n=2}^{\infty} p_n(T_*) \left( \frac{T_* T_*}{L_*^2 n^2} + (2 + C_1) \frac{T_* T_*}{L_*^2 n^3} + 5C_1 \frac{T_* T_*}{L_*^2 n^{3/2}} \right) \\
&\quad + \frac{L^{-2} v_0^2 T^2}{3} p_0(T_*) + \frac{1}{T} p_1(T_*) \int_0^T \mathbb{E}_{\mathbb{P}_1} [|Y(s; 1, \mathfrak{s})|^2] ds \\
&\leq 8C_2 \frac{T_* T_*}{L_*^2} T_*^{-1/2} + \frac{T_*^2}{3L_*^2} e^{-T_*} + e^{-T_*} T_* \frac{1}{T} \int_0^T \mathbb{E}_{\mathbb{P}_1} [|Y(s; 1, \mathfrak{s})|^2] ds.
\end{aligned}$$

where  $C_2 := 6C_1 + 3$ . We point out that for  $n = 1$  (5.14) implies

$$\frac{1}{T} \int_0^T \mathbb{E}_{\mathbb{P}_1} [ |Y(s; 1, \mathfrak{s})|^2 ] ds \leq (2 + 6C_1) \frac{T_*}{L_*^2} T_*.$$

This concludes the proof of Lemma 5.2. □

In the sequel, we stress the fact that Proposition 3.2 is just a consequence of what we have already shown up to here.

**Proof of Proposition 3.2.** Combining Lemma 5.1 and Lemma 5.2 in (5.1) implies (3.11). This finishes the proof of Proposition 3.2. □

**6. Proof of Lemma 3.3: Komlós–Major–Tusnády coupling**

In this section we apply the Komlós–Major–Tusnády (KMT) coupling to prove Lemma 3.3. That is, the existence of a constant  $\mathcal{K}$  satisfying

$$(6.1) \quad \mathcal{W}_2^2(\mathbb{Z}_{[0,T]}, \mathbb{B}_{[0,T]}) \leq \mathcal{K} T_* L_*^{-2} T_*^2 e^{-T_*} + \mathcal{K} T_* L_*^{-2} T_*^{-1/2} (1 + T_*^{-3/2}).$$

Recall that

$$(6.2) \quad \mathcal{W}_2^2(\mathbb{Z}_{[0,T]}, \mathbb{B}_{[0,T]}) \leq \frac{1}{T} \int_0^T ds \mathbb{E}_\pi [ |Z(s) - B(s)|^2 ]$$

for any coupling  $\pi$  between  $\mathbb{Z}_{[0,T]}$  and  $\mathbb{B}_{[0,T]}$ . Let  $\pi^*$  be the so-called Komlós–Major–Tusnády coupling given in Theorem 1 of [32] (or Lemma B.5 in Appendix B) gluing with an independent Poisson random variable  $K$  with parameter  $T_* = \lambda T$ . Since  $K$  has Poisson distribution with parameter  $T_*$ , inequality (6.2), the tower property (Item (iv)) of Theorem 8.14 in [25]) and Fubini’s theorem imply

$$(6.3) \quad \begin{aligned} \mathcal{W}_2^2(\mathbb{Z}_{[0,T]}, \mathbb{B}_{[0,T]}) &\leq p_0(T_*) \frac{1}{T} \int_0^T ds \mathbb{E}_{\pi^*} [ |Z(s) - B(s)|^2 | K = 0 ] \\ &\quad + p_1(T_*) \frac{1}{T} \int_0^T ds \mathbb{E}_{\pi^*} [ |Z(s) - B(s)|^2 | K = 1 ] + J_0, \end{aligned}$$

where

$$(6.4) \quad J_0 := \frac{1}{T} \sum_{n=2}^\infty p_n(T_*) \int_0^T ds \mathbb{E}_{\pi^*} [ |Z(s) - B(s)|^2 | K = n ].$$

In the sequel we estimate  $J_0$ . To make the proof easier to follow, we have divided it in 8 steps.

*Step 1: First splitting.* Let  $\tilde{n} = \lfloor n/2 \rfloor$  and split  $J_0$  as follows:

$$J_0 := \frac{1}{T} \sum_{n=2}^\infty p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \mathbb{E}_{\pi^*} [ |Z(s) - B(s)|^2 | K = n ].$$

For each  $j \in \{1, \dots, \tilde{n}\}$  we write  $\Theta_{j-1} B(s) := B(s) - B((j-1)T/\tilde{n})$  for any  $s \geq 0$  and note that  $B(s) = B((j-1)T/\tilde{n}) + \Theta_{j-1} B(s)$ . Since  $(x - y)^2 \leq 2x^2 + 2y^2$  for any  $x, y \in \mathbb{R}$ , we have

$$(6.5) \quad J_0 \leq J_1 + J_2,$$

where

$$J_1 := \frac{2}{T} \sum_{n=2}^\infty p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \mathbb{E}_{\pi^*} [ |\Theta_{j-1} B(s)|^2 | K = n ]$$

and

$$J_2 := \frac{2}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \mathbb{E}_{\pi^*} [|Z(s) - B((j-1)T/\tilde{n})|^2 | K = n].$$

*Step 2.* We start with the estimate of  $J_1$ . Since for each  $j \in \{1, \dots, \tilde{n}\}$  the random variable  $\Theta_{j-1}B(s)$  has Gaussian distribution with zero mean and variance  $\sigma^2(s - (j-1)T/\tilde{n})$ , with  $\sigma^2$  being defined in (2.18), we deduce

$$J_1 \leq \frac{2}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \sigma^2 \left( s - \frac{(j-1)T}{\tilde{n}} \right) = T \sigma^2 \langle \tilde{n}^{-1} \mathbb{1}_{\{n \geq 2\}} \rangle_{\text{Po}(T_*)}.$$

The preceding inequality with the help of a slight shift in Lemma A.1 in Appendix A, (2.18), the fact that  $\tilde{m} \geq m/2 - 1$ ,  $m \geq 2$  and the definition of  $\sigma^2$  given in (2.18) yields

$$\begin{aligned} (6.6) \quad J_1 &\leq T \sigma^2 \langle \tilde{n}^{-1} \mathbb{1}_{\{n=2\}} \rangle_{\text{Po}(T_*)} + T \sigma^2 \langle \tilde{n}^{-1} \mathbb{1}_{\{n \geq 3\}} \rangle_{\text{Po}(T_*)} \\ &\leq T \sigma^2 p_2(T_*) + 2T \sigma^2 \langle (n-2)^{-1} \mathbb{1}_{\{n \geq 3\}} \rangle_{\text{Po}(T_*)} \\ &\leq 2^{-1} \frac{T_*}{L_*^2} T_*^2 e^{-T_*} + 2^{7/2} \frac{T_*}{L_*^2} \frac{1}{T_*}. \end{aligned}$$

This finishes the estimate of  $J_1$ .

*Step 3: Second splitting.* We continue with the estimate of  $J_2$ . The idea is to apply the Komlós–Major–Tusnády coupling between a suitable random walk and the Brownian motion with diffusivity constant  $\sigma^2$ . We define the random walk  $(S_m(\mathfrak{s}) : m \in \mathbb{N}_0)$  as follows:

$$S_m(\mathfrak{s}) := \sum_{j=1}^m \eta_j(\mathfrak{s}) \quad \text{for } m \in \mathbb{N} \text{ and } S_0(\mathfrak{s}) := 0,$$

where the sequence  $(\eta_j(\mathfrak{s}) : j \in \mathbb{N})$  are defined by

$$\eta_j(\mathfrak{s}) = -v_0(s_{2j} - s_{2j-1}) \quad \text{for any } j \in \mathbb{N}.$$

Hence, the sequence  $(\eta_j(\mathfrak{s}) : j \in \mathbb{N})$  are i.i.d. random variables with zero mean and variance  $2v_0^2 \text{Var}[\mathfrak{s}_1] = 2v_0^2 \frac{1}{\lambda^2}$ . Set

$$(6.7) \quad \eta_j^*(\mathfrak{s}) := \frac{\eta_j(\mathfrak{s})}{L} = \frac{-v_0(s_{2j} - s_{2j-1})}{L} \quad \text{for any } j \in \mathbb{N}$$

and note that

$$\frac{S_j(\mathfrak{s})}{L} := \sum_{k=1}^j \eta_k^*(\mathfrak{s}) \quad \text{for any } j \in \mathbb{N}.$$

Then  $\mathbb{E}[\eta_j^*(\mathfrak{s})] = 0$  and  $\text{Var}[\eta_j^*(\mathfrak{s})] = \frac{v_0^2}{L^2} 2 \text{Var}[\mathfrak{s}_1] = \frac{2v_0^2}{L^2 \lambda^2} = \frac{2}{L_*^2}$  for all  $j \in \mathbb{N}$ .

Similarly as (6.5) we obtain

$$(6.8) \quad J_2 \leq J_2^{(1)} + J_2^{(2)},$$

where

$$J_2^{(1)} := \frac{4}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \mathbb{E}_{\pi^*} [|S_{j-1}(\mathfrak{s})/L - B((j-1)T/\tilde{n})|^2 | K = n]$$

and

$$J_2^{(2)} := \frac{4}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \mathbb{E}_{\pi^*} [|Z(s) - S_{j-1}(\mathfrak{s})/L|^2 | K = n].$$

*Step 4: Third splitting.* In what follows we estimate  $J_2^{(1)}$ . We start noticing that

$$(6.9) \quad J_2^{(1)} \leq J_2^{(1,1)} + J_2^{(1,2)},$$

with

$$J_2^{(1,1)} := \frac{8}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 \middle| K = n \right]$$

and

$$J_2^{(1,2)} := \frac{8}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds \mathbb{E}_{\pi^*} \left[ \left| \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, \sigma^2 T/\tilde{n}) \right|^2 \middle| K = n \right],$$

where  $(\mathbf{N}_\ell(0, \eta^2) : \ell \in \mathbb{N})$  are i.i.d. Gaussian random variables with zero mean and variance  $\eta^2 > 0$ . We stress that the preceding sequence  $(\mathbf{N}_\ell(0, \eta^2) : \ell \in \mathbb{N})$  of random variables is the sequence used to construct the Komlós–Major–Tusnády coupling  $\pi^*$ .

*Step 5.* Now, we estimate the simplest term  $J_2^{(1,2)}$ . It is straightforward to see that

$$\begin{aligned} J_2^{(1,2)} &= \frac{8}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} \int_{\frac{j-1}{\tilde{n}}T}^{\frac{j}{\tilde{n}}T} ds (j-1) \left| \sqrt{2L_*^{-2}} - \sqrt{\sigma^2 T/\tilde{n}} \right|^2 \\ &= \frac{8}{T} \sum_{n=2}^{\infty} p_n(T_*) \sum_{j=1}^{\tilde{n}} (j-1) \frac{T}{\tilde{n}} \left| \sqrt{2L_*^{-2}} - \sqrt{\sigma^2 T/\tilde{n}} \right|^2 \\ &= 4 \sum_{n=2}^{\infty} p_n(T_*) (\tilde{n}-1) \left| \sqrt{2L_*^{-2}} - \sqrt{\sigma^2 T/\tilde{n}} \right|^2. \end{aligned}$$

Since  $\sigma^2 = L_*^{-2}\lambda$ , we have

$$\begin{aligned} J_2^{(1,2)} &\leq 4 \sum_{n=2}^{\infty} p_n(T_*) \tilde{n} \left| \sqrt{2L_*^{-2}} - \sqrt{\sigma^2 T/\tilde{n}} \right|^2 = 4L_*^{-2} \sum_{n=2}^{\infty} p_n(T_*) \tilde{n} \left| \sqrt{2} - \sqrt{T_*/\tilde{n}} \right|^2 \\ &= 4T_* L_*^{-2} \sum_{n=2}^{\infty} p_n(T_*) \left| \sqrt{\frac{2\tilde{n}}{T_*}} - 1 \right|^2 \leq 4T_* L_*^{-2} \sum_{n=2}^{\infty} p_n(T_*) \left| \frac{2\tilde{n}}{T_*} - 1 \right| \\ &\leq 4T_* L_*^{-2} \left( \sum_{n=2}^{\infty} p_n(T_*) \left| \frac{2\tilde{n}}{T_*} - 1 \right|^2 \right)^{1/2}, \end{aligned}$$

where we have used the following subadditivity inequality:  $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|}$  for any non-negative numbers  $x$  and  $y$ , and in the last inequality we applied the Cauchy–Schwarz inequality. The preceding inequality with the help of Lemma A.2 in Appendix A implies

$$(6.10) \quad J_2^{(1,2)} \leq 4T_* L_*^{-2} \left( \sum_{n=2}^{\infty} p_n(T_*) \left| \frac{2\tilde{n}}{T_*} - 1 \right|^2 \right)^{1/2} \leq 4\sqrt{35} T_* L_*^{-2} T_*^{-1/2}.$$

*Step 6.* In the sequel we estimate  $J_2^{(1,1)}$ . We note that

$$(6.11) \quad J_2^{(1,1)} = 8 \sum_{n=2}^{\infty} \frac{p_n(T_*)}{\tilde{n}} \sum_{j=1}^{\tilde{n}} \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 \middle| K = n \right].$$

Let  $\eta > 0$  be fixed. For any  $j \in \{1, \dots, \tilde{n}\}$  we consider the following split

$$\begin{aligned}
 & \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 \middle| K = n \right] \\
 (6.12) \quad &= \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 \mathbb{1}_{A_\eta^c} \middle| K = n \right] + \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 \mathbb{1}_{A_\eta} \middle| K = n \right] \\
 &\leq \eta^2 + \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 \mathbb{1}_{A_\eta} \middle| K = n \right],
 \end{aligned}$$

where

$$A_\eta := \left\{ \sup_{1 \leq j \leq \tilde{n}} \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 > \eta^2 \right\}.$$

By the Cauchy–Schwarz inequality we have

$$\begin{aligned}
 & \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^2 \mathbb{1}_{A_\eta} \middle| K = n \right] \\
 (6.13) \quad &\leq \left( \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^4 \middle| K = n \right] \right)^{1/2} (\pi^*(A_\eta | K = n))^{1/2}.
 \end{aligned}$$

By the KMT coupling (see Theorem 1 of [32]) there exist positive constants  $C_{\text{KMT}}^{(1)}$ ,  $C_{\text{KMT}}^{(2)}$  and  $\vartheta$  such that for every  $\tilde{n}$  it follows that

$$(6.14) \quad \pi^*(A_{\eta_*(n)} | K = n) \leq C_{\text{KMT}}^{(2)} e^{-\vartheta \ln(n)} = \frac{C_{\text{KMT}}^{(2)}}{n^\vartheta},$$

where

$$(6.15) \quad \eta_*(n) := \frac{\sqrt{2}}{L_*} (C_{\text{KMT}} \ln(\tilde{n}) + \ln(n)).$$

We point out that the constants  $C_{\text{KMT}}^{(1)}$ ,  $C_{\text{KMT}}^{(2)}$  and  $\vartheta$  only depend on the law of the random variable (6.7). Later on, we choose  $\vartheta = 4$ . Combining (6.12), (6.13), (6.14) and (6.15) in (6.11) we obtain

$$(6.16) \quad J_2^{(1,1)} \leq 8 \sum_{n=2}^{\infty} p_n(T_*) \eta_*^2(n) + 8 \sum_{n=2}^{\infty} \frac{p_n(T_*)}{\tilde{n}} \left( \frac{C_{\text{KMT}}^{(2)}}{n^\vartheta} \right)^{1/2} \sum_{j=1}^{\tilde{n}} a_j,$$

where

$$a_j := \left( \mathbb{E}_{\pi^*} \left[ \left| S_{j-1}(\mathfrak{s})/L - \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2}) \right|^4 \middle| K = n \right] \right)^{1/2} \quad \text{for any } j \in \{1, \dots, \tilde{n}\}.$$

Note that for any  $j \in \{1, \dots, \tilde{n}\}$  we have

$$b_j := \left( \mathbb{E}_{\pi^*} \left[ \sum_{\ell=1}^{j-1} \mathbf{N}_\ell(0, 2L_*^{-2})^4 \right] \right)^{1/2} = (\mathbb{E}_{\pi^*} [\mathbf{N}_\ell(0, 2(j-1)L_*^{-2})^4])^{1/2} = 2(j-1)L_*^{-2}$$

and

$$c_j := (\mathbb{E}_{\pi^*} [|S_{j-1}(\mathfrak{s})/L|^4])^{1/2}.$$

Moreover, straightforward computations yields  $c_j = \frac{1}{L_*^2}(\mathbb{E}[|S_{j-1}|^4])^{1/2} = L_*^{-2}\sqrt{22j + 2j^2} \leq 5L_*^{-2}j$ . Since  $a_j \leq 3(b_j + c_j)$ , (6.16) with the chose of  $\vartheta = 4$  and Lemma A.1 in Appendix A implies

$$\begin{aligned} J_2^{(1,1)} &\leq 32 \max\{1, C_{\text{KMT}}^2\} L_*^{-2} \sum_{n=2}^{\infty} p_n(T_*) \ln^2(n) + 84\sqrt{C_{\text{KMT}}^{(2)}} L_*^{-2} \sum_{n=2}^{\infty} p_n(T_*) n^{-1} \\ &\leq 32 \max\{1, C_{\text{KMT}}^2\} L_*^{-2} \sum_{n=2}^{\infty} p_n(T_*) \ln^2(n) + 252\sqrt{C_{\text{KMT}}^{(2)}} L_*^{-2} T_*^{-1}. \end{aligned}$$

Let  $\gamma \in (0, 1/2)$  be fixed and note that there exists a constant  $C(\gamma) > 0$  such that  $\ln(n) \leq C(\gamma)n^\gamma$  for all  $n \in \mathbb{N}$ . Hence, Hölder’s inequality yields

$$J_2^{(1,1)} \leq 32C^2(\gamma) \max\{1, C_{\text{KMT}}^2\} T_* L_*^{-2} T_*^{2\gamma-1} + 252\sqrt{C_{\text{KMT}}^{(2)}} T_* L_*^{-2} T_*^{-2}.$$

The choice of  $\gamma = 1/4$  implies the existence of a positive constant  $C_2^{(1)}$  such that

$$(6.17) \quad J_2^{(1,1)} \leq C_2^{(1)} T_* L_*^{-2} T_*^{-1/2} (1 + T_*^{-3/2}).$$

Combining (6.10) and (6.17) in (6.9) we obtain

$$(6.18) \quad J_2^{(1)} \leq (C_2^{(1)} + 4\sqrt{35}) T_* L_*^{-2} T_*^{-1/2} (1 + T_*^{-3/2}).$$

*Step 7.* Finally, we estimate  $J_2^{(2)}$ . Let  $(S(t) : t \geq 0)$  be the Càdlàg piecewise constant extension of the random walk  $(\eta_\ell^*(\mathfrak{s}) : \ell \in \mathbb{N}_0)$  given in (6.7). Since

$$Z(t; n, \mathfrak{s}) = \sum_{\ell=1}^{\tilde{n}} \eta_\ell^*(\mathfrak{s}) \mathbb{1}_{\{t > t_{2\ell}(\mathfrak{s})\}} \quad \text{and} \quad S(t; n, \mathfrak{s}) = \sum_{\ell=1}^{\tilde{n}} \eta_\ell^*(\mathfrak{s}) \mathbb{1}_{\{t > \frac{\ell T}{\tilde{n}}\}},$$

we have

$$\begin{aligned} &\frac{1}{T} \int_0^T dt |Z(t; n, \mathfrak{s}) - S(t; n, \mathfrak{s})|^2 \\ &= \sum_{\ell=1}^{\tilde{n}} \sum_{\ell'=1}^{\tilde{n}} \eta_\ell^*(\mathfrak{s}) \eta_{\ell'}^*(\mathfrak{s}) \frac{1}{T} \int_0^T dt (\mathbb{1}_{\{t > t_{2\ell}(\mathfrak{s})\}} - \mathbb{1}_{\{t > \frac{\ell T}{\tilde{n}}\}})(\mathbb{1}_{\{t > t_{2\ell'}(\mathfrak{s})\}} - \mathbb{1}_{\{t > \frac{\ell' T}{\tilde{n}}\}}). \end{aligned}$$

In the sequel, we recall the change of variable introduced in (5.4) and (5.5). That is, for any  $\ell \in \{1, \dots, \tilde{n}\}$  we set

$$\hat{s}_\ell := s_{2\ell} + s_{2\ell-1} \quad \text{and} \quad \omega_\ell := \frac{s_{2\ell} - s_{2\ell-1}}{\hat{s}_\ell}$$

and observe that

$$\eta_\ell^*(\mathfrak{s}) = -\frac{v_0}{L} \hat{s}_\ell \omega_\ell \quad \text{for any } \ell \in \{1, \dots, \tilde{n}\}.$$

An analogous reasoning using in (5.7) with the help of Item (i) and Item (iii) of Lemma D.1 in Appendix D implies

$$\begin{aligned} &\left\langle \frac{1}{T} \int_0^T dt |Z(t; n, \mathfrak{s}) - S(t; n, \mathfrak{s})|^2 \right\rangle_\omega \\ &= \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \sum_{\ell'=1}^{\tilde{n}} \hat{s}_\ell \hat{s}_{\ell'} \langle \omega_\ell \omega_{\ell'} \rangle_\omega \frac{1}{T} \int_0^T dt (\mathbb{1}_{\{t > t_{2\ell}(\mathfrak{s})\}} - \mathbb{1}_{\{t > \frac{\ell T}{\tilde{n}}\}})(\mathbb{1}_{\{t > t_{2\ell'}(\mathfrak{s})\}} - \mathbb{1}_{\{t > \frac{\ell' T}{\tilde{n}}\}}) \\ (6.19) \quad &= \frac{1}{3} \left(\frac{v_0}{L}\right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} \int_0^T dt (\mathbb{1}_{\{t > t_{2\ell}(\mathfrak{s})\}} - \mathbb{1}_{\{t > \frac{\ell T}{\tilde{n}}\}})^2 \end{aligned}$$



$$\begin{aligned}
&= \frac{1}{3} \left( \frac{v_0}{L} \right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} \int_0^T dt \mathbb{1}_{\{\min\{t_{2\ell}(\mathfrak{s}), \frac{\ell T}{\tilde{n}}\} < t < \max\{t_{2\ell}(\mathfrak{s}), \frac{\ell T}{\tilde{n}}\}\}} \\
&\leq \frac{1}{3} \left( \frac{v_0}{L} \right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} \left( \max \left\{ t_{2\ell}(\mathfrak{s}), \frac{\ell T}{\tilde{n}} \right\} - \min \left\{ t_{2\ell}(\mathfrak{s}), \frac{\ell T}{\tilde{n}} \right\} \right) \\
&\leq \left( \frac{v_0}{L} \right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} \left| t_{2\ell}(\mathfrak{s}) - \frac{\ell T}{\tilde{n}} \right|,
\end{aligned}$$

where  $\omega$  denotes the product measure given by the law of the random vector  $(\omega_1, \dots, \omega_{\tilde{n}})$ . We note that

$$(6.20) \quad \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \left| t_{2\ell}(\mathfrak{s}) - \frac{\ell T}{\tilde{n}} \right| \leq \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \left| t_{2\ell}(\mathfrak{s}) - \frac{2\ell}{\lambda} \right| + \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \left| \frac{2\ell}{\lambda} - \frac{\ell T}{\tilde{n}} \right|$$

and let  $\hat{s}$  denote the product measure given by the law of the random vector  $(\hat{s}_1, \dots, \hat{s}_{\tilde{n}})$ . On the one hand, Item (ii) of Lemma D.1 in Appendix D with the help of Fubini's theorem yields

$$\begin{aligned}
(6.21) \quad &\left\langle \left( \frac{v_0}{L} \right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} \left| \frac{2\ell}{\lambda} - \frac{\ell T}{\tilde{n}} \right| \right\rangle_{\hat{s}} = \left( \frac{v_0}{L} \right)^2 \sum_{\ell=1}^{\tilde{n}} \langle \hat{s}_\ell^2 \rangle_{\hat{s}} \frac{1}{T} \left| \frac{2\ell}{\lambda} - \frac{\ell T}{\tilde{n}} \right| \\
&= 2 \left( \frac{v_0}{\lambda L} \right)^2 \sum_{\ell=1}^{\tilde{n}} 2\ell \left| \frac{1}{T_*} - \frac{1}{2\tilde{n}} \right| \leq 2 \left( \frac{v_0}{\lambda L} \right)^2 \tilde{n}^2 \left| \frac{1}{T_*} - \frac{1}{2\tilde{n}} \right| \leq \frac{T_*}{L_*^2} \frac{\tilde{n}}{T_*} \left| \frac{2\tilde{n}}{T_*} - 1 \right|.
\end{aligned}$$

On the other hand, the Cauchy–Schwarz inequality implies

$$\begin{aligned}
(6.22) \quad &\left\langle \left( \frac{v_0}{L} \right)^2 \sum_{\ell=1}^{\tilde{n}} \hat{s}_\ell^2 \frac{1}{T} \left| t_{2\ell}(\mathfrak{s}) - \frac{2\ell}{\lambda} \right| \right\rangle_{\hat{s}} = \left\langle \frac{1}{L_*^2 T} \sum_{\ell=1}^{\tilde{n}} \lambda^2 \hat{s}_\ell^2 \left| t_{2\ell}(\mathfrak{s}) - \frac{2\ell}{\lambda} \right| \right\rangle_{\hat{s}} \\
&\leq \frac{1}{L_*^2 T} \sqrt{\sum_{\ell=1}^{\tilde{n}} \lambda^4 \langle \hat{s}_\ell^4 \rangle_{\hat{s}}} \sqrt{\sum_{\ell=1}^{\tilde{n}} \left\langle \left| t_{2\ell}(\mathfrak{s}) - \frac{2\ell}{\lambda} \right|^2 \right\rangle_{\hat{s}}} \leq \frac{1}{L_*^2 T} \sqrt{120\tilde{n}} \sqrt{\sum_{\ell=1}^{\tilde{n}} \frac{2\ell}{\lambda^2}} \\
&\leq \frac{1}{L_*^2 T} \sqrt{120\tilde{n}} \frac{\tilde{n}}{\lambda} \leq 11 \frac{T_*}{L_*^2} \frac{n^{3/2}}{T_*}.
\end{aligned}$$

Combining (6.20), (6.21) and (6.22) in (6.19) with the help of Fubini's theorem implies

$$\left\langle \frac{1}{T} \int_0^T dt |Z(t; n, \mathfrak{s}) - S(t; n, \mathfrak{s})|^2 \right\rangle_{\hat{s}, \omega} \leq \frac{T_*}{L_*^2} \frac{\tilde{n}}{T_*} \left| \frac{2\tilde{n}}{T_*} - 1 \right| + 11 \frac{T_*}{L_*^2} \frac{n^{3/2}}{T_*}.$$

Recall that  $n$  has Poisson distribution of parameter  $T_*$ . Then we have

$$\begin{aligned}
(6.23) \quad &\sum_{n=2}^{\infty} p_n(T_*) \left\langle \frac{1}{T} \int_0^T dt |Z(t; n, \mathfrak{s}) - S(t; n, \mathfrak{s})|^2 \right\rangle_{\hat{s}, \omega} \\
&\leq \frac{T_*}{L_*^2} \left\langle \frac{\tilde{n}}{T_*} \left| \frac{2\tilde{n}}{T_*} - 1 \right| \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} + 11 \frac{T_*}{L_*^2} \frac{\langle n^{3/2} \mathbb{1}_{\{n \geq 2\}} \rangle_{\text{Po}(T_*)}}{T_*^2}.
\end{aligned}$$

Now, we estimate the first term of the right-hand side of (6.23). By the Cauchy–Schwarz inequality we obtain

$$\frac{T_*}{L_*^2} \left\langle \frac{\tilde{n}}{T_*} \left| \frac{2\tilde{n}}{T_*} - 1 \right| \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \leq \frac{T_*}{L_*^2} \left( \left\langle \frac{\tilde{n}^2}{T_*^2} \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \right)^{1/2} \left( \left\langle \left| \frac{2\tilde{n}}{T_*} - 1 \right|^2 \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \right)^{1/2}.$$

We note that

$$\left\langle \frac{\tilde{n}^2}{T_*^2} \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \leq \frac{1}{4T_*^2} \langle n^2 \rangle_{\text{Po}(T_*)} \leq \frac{1}{4T_*} + \frac{1}{4}.$$

Also, by Lemma A.2 in Appendix A we obtain

$$\left\langle \left| \frac{2\tilde{n}}{T_*} - 1 \right|^2 \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \leq \frac{35}{T_*}.$$

Then

$$(6.24) \quad \frac{T_*}{L_*^2} \left\langle \frac{\tilde{n}}{T_*} \left| \frac{2\tilde{n}}{T_*} - 1 \right| \mathbb{1}_{\{n \geq 2\}} \right\rangle_{\text{Po}(T_*)} \leq 3 \frac{T_*}{L_*^2} T_*^{-1/2} (T_*^{-1/2} + 1).$$

Now, we estimate the second term of the right-hand side of (6.23). Then Hölder's inequality yields

$$(6.25) \quad \begin{aligned} \frac{T_*}{L_*^2} \frac{\langle n^{3/2} \mathbb{1}_{\{n \geq 2\}} \rangle_{\text{Po}(T_*)}}{T_*^2} &\leq \frac{T_*}{L_*^2} \frac{\langle n^{3/2} \rangle_{\text{Po}(T_*)}}{T_*^2} \leq \frac{T_*}{L_*^2} \frac{1}{T_*^2} (\langle n^2 \rangle_{\text{Po}(T_*)})^{3/4} \\ &= \frac{T_*}{L_*^2} \frac{1}{T_*^2} (T_* + T_*^2)^{3/4} \leq \frac{T_*}{L_*^2} (T_*^{-1/2} + T_*^{-5/4}). \end{aligned}$$

Combining (6.24) and (6.25) in (6.23) we obtain

$$(6.26) \quad J_2^{(2)} \leq 3 \frac{T_*}{L_*^2} T_*^{-1/2} (T_*^{-1/2} + 1) + 11 \frac{T_*}{L_*^2} T_*^{-1/2} (1 + T_*^{-3/4}).$$

Therefore, (6.8), (6.18) and (6.26) yields the existence of a constant  $K_1 > 0$  satisfying

$$(6.27) \quad \begin{aligned} J_2 &\leq (C_2^{(1)} + 24) T_* L_*^{-2} T_*^{-1/2} (1 + T_*^{-3/2}) + 3 T_* L_*^{-2} T_*^{-1/2} (1 + T_*^{-1/2}) + 11 T_* L_*^{-2} T_*^{-1/2} (1 + T_*^{-3/4}) \\ &\leq K_1 T_* L_*^{-2} T_*^{-1/2} (1 + T_*^{-1/2} + T_*^{-3/4} + T_*^{-3/2}). \end{aligned}$$

*Step 8: Concluding step.* As a consequence of (6.5), (6.6) and (6.27) we obtain an upper bound for  $J_0$  defined in (6.4). The latter with the help of (6.3) yields (6.1).

## Appendix A: Moment estimates

This section contains derivation of the moment estimates used in the main text.

### A.1. Poisson moment estimates

**Lemma A.1 (Bounds for the inverse moments of Poisson r.v.).** *Let  $N$  be a Poisson random variable with parameter  $\lambda > 0$ . Then for any  $p > 0$  it follows that*

$$(A.1) \quad \langle N^{-p} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} \leq \frac{C_p}{\lambda^p}, \quad \text{where } C_p := 2^{p(p+2)/2}.$$

In addition,

$$(A.2) \quad \langle N^{-p} \ln(N+1) \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} \leq \frac{C_{2p}^{1/2} \ln(\lambda+3)}{\lambda^p}.$$

**Proof.** We start with the case  $p \in \mathbb{N}$ . By definition we have

$$(A.3) \quad \langle N^{-p} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} = \sum_{j \geq 1} \frac{1}{j^p} e^{-\lambda} \frac{\lambda^j}{j!} = \frac{e^{-\lambda}}{\lambda^p} \sum_{j \geq 1} \frac{\lambda^{j+p}}{j^p (j!)}.$$

Note that

$$\frac{1}{j} \leq \frac{2}{j+1} \quad \text{for any } j \in \mathbb{N}.$$

Hence,

$$(A.4) \quad \frac{1}{j^p} \leq \prod_{k=1}^p \frac{2^k}{j+k} \quad \text{for any } j \in \mathbb{N}.$$

By (A.3) and (A.4) we have

$$(A.5) \quad \langle N^{-p} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} \leq K_p \frac{e^{-\lambda}}{\lambda^p} \sum_{j \geq 1} \frac{\lambda^{j+p}}{(j+p)!} = K_p \frac{e^{-\lambda}}{\lambda^p} \left( e^{-\lambda} - \sum_{j=0}^p \frac{\lambda^j}{j!} \right) \leq \frac{K_p}{\lambda^p},$$

where  $K_p := \prod_{k=1}^p 2^k = 2^{p(p+1)/2} \leq C_p$ .

We continue with the case  $p > 1$  and  $p \notin \mathbb{N}$ . Let  $\lceil p \rceil$  be the least integer greater than or equal to  $p$  and set  $p_1 = \lceil p \rceil / p$ . We observe that  $p_1 > 1$  and let  $p_2$  be the Hölder conjugate of  $p_1$ , that is,  $1/p_1 + 1/p_2 = 1$ . By Hölder's inequality we have

$$(A.6) \quad \begin{aligned} \langle N^{-p} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)}^{1/p} &= \langle N^{-p} \mathbb{1}_{\{N \geq 1\}} \cdot 1 \rangle_{\text{Po}(\lambda)}^{1/p} \leq \left( \langle (N^{-p} \mathbb{1}_{\{N \geq 1\}})^{p_1} \rangle_{\text{Po}(\lambda)}^{1/p_1} \langle 1^{p_2} \rangle_{\text{Po}(\lambda)}^{1/p_2} \right)^{1/p} \\ &\leq \langle (N^{-p} \mathbb{1}_{\{N \geq 1\}})^{p_1} \rangle_{\text{Po}(\lambda)}^{1/p_1} = \langle N^{-\lceil p \rceil} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)}^{1/\lceil p \rceil}. \end{aligned}$$

Since  $\lceil p \rceil \in \mathbb{N} \setminus \{1\}$ , (A.5) yields

$$(A.7) \quad \langle N^{-\lceil p \rceil} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)}^{1/\lceil p \rceil} \leq \left( \frac{K_{\lceil p \rceil}}{\lambda^{\lceil p \rceil}} \right)^{1/\lceil p \rceil} = \frac{K_{\lceil p \rceil}^{1/\lceil p \rceil}}{\lambda},$$

where  $K_{\lceil p \rceil} = 2^{\lceil p \rceil(\lceil p \rceil+1)/2}$ . By (A.6) and (A.7) we obtain

$$\langle N^{-p} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} \leq \frac{K_{\lceil p \rceil}^{p/\lceil p \rceil}}{\lambda^p} \leq \frac{2^{p(p+2)/2}}{\lambda^p} = \frac{C_p}{\lambda^p}.$$

Finally, for  $p \in (0, 1)$ , let  $q := 1/p > 1$ . By Hölder's inequality and (A.5) (for  $p = 1$ ) we obtain

$$\langle N^{-p} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} \leq \langle (N^{-p})^q \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)}^{1/q} = \langle N^{-1} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)}^{1/q} \leq \left( \frac{K_1}{\lambda} \right)^p \leq \frac{C_p}{\lambda^p}.$$

Now, we prove (A.2). Note that the Jensen inequality implies

$$\langle (\ln(N+3))^2 \rangle_{\text{Po}(\lambda)} \leq (\ln(\lambda+3))^2.$$

By the Cauchy–Schwarz inequality, inequality (A.1) and the Jensen inequality we have

$$\begin{aligned} \langle N^{-p} \ln(N+1) \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} &\leq \left( \langle N^{-2p} \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} \right)^{1/2} \left( \langle (\ln(N+1))^2 \mathbb{1}_{\{N \geq 1\}} \rangle_{\text{Po}(\lambda)} \right)^{1/2} \\ &\leq \frac{C_{2p}^{1/2}}{\lambda^p} \left( \langle (\ln(N+1))^2 \rangle_{\text{Po}(\lambda)} \right)^{1/2} \leq \frac{C_{2p}^{1/2} \ln(\lambda+3)}{\lambda^p}. \end{aligned}$$

□

**Lemma A.2.** *Let  $N$  be a Poisson random variable with parameter  $\lambda > 0$ . Then*

$$(A.8) \quad \left\langle \left| \frac{2\lfloor N/2 \rfloor}{\lambda} - 1 \right|^2 \mathbb{1}_{\{N \geq 2\}} \right\rangle_{\text{Po}(\lambda)} \leq \frac{e^{-\lambda}(2-\lambda)^2}{2} + \frac{9}{\lambda}.$$

In particular,

$$\left\langle \left| \frac{2\lfloor N/2 \rfloor}{\lambda} - 1 \right|^2 \mathbb{1}_{\{N \geq 2\}} \right\rangle_{\text{Po}(\lambda)} \leq \frac{35}{\lambda}.$$

**Proof.** We observe that

$$(A.9) \quad \left\langle \left| \frac{2\lfloor N/2 \rfloor}{\lambda} - 1 \right|^2 \mathbb{1}_{\{N \geq 2\}} \right\rangle_{\text{Po}(\lambda)} = \left| \frac{2}{\lambda} - 1 \right|^2 p_2(\lambda) + \left\langle \left| \frac{2\lfloor N/2 \rfloor}{\lambda} - 1 \right|^2 \mathbb{1}_{\{N \geq 3\}} \right\rangle_{\text{Po}(\lambda)}.$$

An straightforward computation yields

$$(A.10) \quad \begin{aligned} \left\langle \left| \frac{2\lfloor N/2 \rfloor}{\lambda} - 1 \right|^2 \mathbb{1}_{\{N \geq 3\}} \right\rangle_{\text{Po}(\lambda)} &= \left\langle \frac{4\lfloor N/2 \rfloor^2}{\lambda^2} \mathbb{1}_{\{N \geq 3\}} \right\rangle_{\text{Po}(\lambda)} - \left\langle \frac{4\lfloor N/2 \rfloor}{\lambda} \mathbb{1}_{\{N \geq 3\}} \right\rangle_{\text{Po}(\lambda)} + \langle \mathbb{1}_{\{N \geq 3\}} \rangle_{\text{Po}(\lambda)} \\ &\leq \left\langle \frac{N^2}{\lambda^2} \right\rangle_{\text{Po}(\lambda)} + \left\langle \left( \frac{4 - 2N}{\lambda} \right) \mathbb{1}_{\{N \geq 3\}} \right\rangle_{\text{Po}(\lambda)} + 1 \\ &\leq \frac{5}{\lambda} + 2 - \frac{2}{\lambda} \langle N \mathbb{1}_{\{N \geq 3\}} \rangle_{\text{Po}(\lambda)} \\ &\leq \frac{5}{\lambda} + 2 - \frac{2}{\lambda} (\lambda - p_1(\lambda) - 2p_2(\lambda)) \\ &\leq \frac{5}{\lambda} + \frac{2}{\lambda} (p_1(\lambda) + 2p_2(\lambda)) \leq \frac{9}{\lambda}. \end{aligned}$$

Finally, (A.9) and (A.10) imply (A.8). □

A.2.  $\delta$ -constraint moment estimates

**Lemma A.3 (Polynomial moments and correlations for the  $\delta$ -function).** *Let  $n \in \mathbb{N}$ . For  $p \in \mathbb{N}$  and any  $i \in \{1, \dots, n\}$  it follows that*

$$\langle u_i^p \rangle = \frac{p!}{\prod_{k=0}^{p-1} (n+k)},$$

where the integration is respect to the probability measure  $d^n \mathbf{u} \cdot (n-1)! \delta(t_n(\mathbf{u}) - 1)$ . In particular,

$$\langle u_i \rangle = \frac{1}{n}, \quad \langle u_i^2 \rangle = \frac{2}{n(n+1)} \quad \text{and} \quad \langle u_i^3 \rangle = \frac{6}{n(n+1)(n+2)}.$$

Moreover, for  $n \in \mathbb{N} \setminus \{1\}$  and any  $i, j \in \{1, \dots, n\}$   $i \neq j$  it follows that

$$\langle u_i u_j \rangle = \frac{1}{n(n+1)}.$$

**Proof.** For  $n = 1$  the result is obvious. We assume  $n \in \mathbb{N} \setminus \{1\}$ .

The estimates rely on the properties of the measure  $d^n \mathbf{u} \cdot (n-1)! \delta(t_n(\mathbf{u}) - 1)$  proven in Appendix C. In particular, we know that it is permutation invariant. Therefore, we have that  $u_i, i \in \{1, \dots, n\}$  are identically distributed and then

$$(A.11) \quad \begin{aligned} \langle u_i^p \rangle &= \langle u_1^p \rangle \quad \text{for all } i \in \{1, \dots, n\} \text{ and } p \in \mathbb{N}, \\ \langle u_i u_j \rangle &= \langle u_1 u_2 \rangle \quad \text{for all } i, j \in \{1, \dots, n\} \text{ satisfying } i \neq j. \end{aligned}$$

Let  $p \in \mathbb{N}$ . We next compute  $\langle u_1^p \rangle$ . By Lemma C.2, then

$$\langle u_1^p \rangle = (n-1) \int_0^1 u_1^p (1-u_1)^{n-2} du_1.$$

Recall that the density  $\text{Beta}(\alpha, \beta)(dx)$  of the Beta distribution  $\text{Beta}(\alpha, \beta)$  with parameters  $\alpha > 0$  and  $\beta > 0$  is given by

$$(A.12) \quad \text{Beta}(\alpha, \beta)(dx) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{1}_{\{x \in [0,1]\}} dx.$$

As a consequence,

$$(n-1) \int_0^1 u_1^p (1-u_1)^{n-2} du_1 = (n-1) \frac{\Gamma(p+1)\Gamma(n-1)}{\Gamma(n+p)} = \frac{p!(n-1)!}{(n-1)! \prod_{k=0}^{p-1} (n+k)} = \frac{p!}{\prod_{k=0}^{p-1} (n+k)}.$$

Hence,

$$\langle u_1^p \rangle = \frac{p!}{\prod_{k=0}^{p-1} (n+k)}.$$

Next, we continue with the computation of  $\langle u_1 u_2 \rangle$ . Assume that  $n \geq 3$ . By Lemma C.2, then

$$\langle u_1 u_2 \rangle = (n-1)(n-2) \int_0^1 u_1 \int_0^{1-u_1} u_2 (1-u_1-u_2)^{n-3} du_2 du_1.$$

Let  $x = \frac{u_2}{1-u_1}$ . Then using the fact that  $\text{Beta}(\alpha, \beta)(dx)$  given in (A.12) is probability measure for any  $\alpha > 0$  and  $\beta > 0$ , we obtain

$$\begin{aligned} \langle u_1 u_2 \rangle &= (n-1)(n-2) \int_0^1 u_1 (1-u_1)^{n-1} du_1 \int_0^1 x (1-x)^{n-3} dx \\ &= (n-1)(n-2) \frac{\Gamma(2)\Gamma(n)}{\Gamma(n+2)} \frac{\Gamma(2)\Gamma(n-2)}{\Gamma(n)} = \frac{1}{n(n+1)}. \end{aligned}$$

Finally, for  $n = 2$  recall the constraint  $u_1 + u_2 = 1$  and observe that

$$\langle u_1 u_2 \rangle = \langle u_1(1-u_1) \rangle = \langle u_1 \rangle - \langle u_1^2 \rangle = \frac{1}{n} - \frac{2}{n(n+1)} = \frac{n-1}{n(n+1)} = \frac{1}{6}. \quad \square$$

**Lemma A.4 (Exponential moments formula for the  $\delta$ -function).** *Let  $n \in \mathbb{N} \setminus \{1\}$ . For all  $i \in \{1, \dots, n\}$  and  $\theta \in \mathbb{R}$  it follows that*

$$\langle e^{\theta u_i} \rangle = (n-1)e^\theta \int_0^1 x^{n-2} e^{-\theta x} dx,$$

where the integration is respect to the probability measure  $d^n \mathbf{u} \cdot (n-1)! \delta(t_n(\mathbf{u}) - 1)$ . In addition,

$$\langle e^{(n-1)u_i} \rangle \leq 3\sqrt{n-1}.$$

**Proof.** By Lemma C.2 and using (A.11), we find

$$\langle e^{\theta u_1} \rangle = \langle e^{\theta u_i} \rangle = (n-1) \int_0^1 e^{\theta u_1} (1-u_1)^{n-2} du_1$$

for all  $i \in \{1, \dots, n\}$ . Using the change of variable to  $x = 1 - u_1$ , we obtain

$$\langle e^{\theta u_1} \rangle = (n-1) \int_0^1 e^{\theta(1-x)} x^{n-2} dx = (n-1)e^\theta \int_0^1 e^{-\theta x} x^{n-2} dx.$$

Let  $y = \theta x$ . Then

$$\langle e^{\theta u_1} \rangle = \frac{(n-1)e^\theta}{\theta^{n-1}} \int_0^\theta e^{-y} y^{n-2} dy \leq \frac{(n-1)e^\theta}{\theta^{n-1}} \int_0^\infty e^{-y} y^{(n-1)-1} dy = \frac{(n-1)e^\theta}{\theta^{n-1}} \Gamma(n-1) = \frac{(n-1)!e^\theta}{\theta^{n-1}}.$$

By [43] we have

$$j! \leq \sqrt{2\pi} e^{1/12} \sqrt{j} \frac{j^j}{e^j} \quad \text{for all } j \in \mathbb{N}.$$

Then the choice  $\theta = n-1$  yields

$$\langle e^{\theta u_1} \rangle \leq \frac{(n-1)!e^{n-1}}{(n-1)^{n-1}} \leq \sqrt{2\pi} e^{1/12} \sqrt{n-1} < 3\sqrt{n-1}. \quad \square$$

**Lemma A.5 (Moment estimates for the maximum).** *For any measurable and bound observable  $f : \mathbb{R}_*^n \rightarrow \mathbb{R}$ , we denote by  $\langle f \rangle$  the integral of  $f$  with respect to the probability measure  $d^n \mathbf{u} \cdot (n-1)! \delta(t_n(\mathbf{u}) - 1)$ . Then there exists a positive*

constant  $C$  such that for all  $n \in \mathbb{N}$  it follows that

$$(A.13) \quad \left\langle \max_{i \in \{1, \dots, n\}} u_i \right\rangle \leq C \frac{\ln(n+1)}{n}.$$

**Proof.** The statement for  $n = 1$  is immediate. In the sequel, we assume  $n \in \mathbb{N} \setminus \{1\}$ . We start with the proof of (A.13). We use the so-called Chernoff's trick explained in Section 2.5 of [7]. Recall that the random variables  $u_1, \dots, u_n$  are identically distributed. By the Jensen inequality for the convex function  $f(x) = \exp(\theta x)$ ,  $x \in \mathbb{R}$  and  $\theta > 0$  yields

$$\begin{aligned} \exp\left(\theta \left\langle \max_{i \in \{1, \dots, n\}} u_i \right\rangle\right) &= f\left(\left\langle \max_{i \in \{1, \dots, n\}} u_i \right\rangle\right) \leq \left\langle f\left(\max_{i \in \{1, \dots, n\}} u_i\right)\right\rangle = \left\langle \exp\left(\theta \max_{i \in \{1, \dots, n\}} u_i\right)\right\rangle \\ &= \left\langle \max_{i \in \{1, \dots, n\}} \exp(\theta u_i) \right\rangle \leq \left\langle \sum_{i=1}^n \exp(\theta u_i) \right\rangle = n \langle \exp(\theta u_1) \rangle. \end{aligned}$$

Hence,

$$(A.14) \quad \left\langle \max_{i \in \{1, \dots, n\}} u_i \right\rangle \leq \frac{\ln(n)}{\theta} + \frac{\ln(\langle \exp(\theta u_1) \rangle)}{\theta} \quad \text{for all } \theta > 0.$$

Inequality (A.14) for  $\theta = n - 1$  with the help of Lemma A.4 in Appendix A implies

$$\left\langle \max_{i \in \{1, \dots, n\}} u_i \right\rangle \leq \frac{\ln(n)}{n-1} + \frac{\ln(3\sqrt{n})}{n-1} \leq 3 \frac{\ln(n+1)}{n} + 2 \frac{\ln(3)}{n} \leq C_1 \frac{\ln(n+1)}{n},$$

where  $C_1 := \max\{6, 4 \ln(3)/\ln(2)\}$ . □

## Appendix B: Free velocity flip moment estimates

In this section, we estimate the moments of the free velocity flip model via Theorem 1 in [23].

**Lemma B.1 (Moment estimates for the free velocity flip model).** *The following moment estimates are valid.*

1. *The first-moment of  $X(t; \mathfrak{s})$  is given by*

$$\mathbb{E}[X(t; \mathfrak{s})] = \frac{v_0}{2\lambda} (1 - e^{-2\lambda t}) \quad \text{for all } t \geq 0.$$

2. *The second-moment and the variance of  $X(t; \mathfrak{s})$  are given by*

$$\mathbb{E}[(X(t; \mathfrak{s}))^2] = \frac{v_0^2}{2\lambda^2} (2\lambda t - (1 - e^{-2\lambda t})) \quad \text{and} \quad \text{Var}[X(t; \mathfrak{s})] = \frac{v_0^2}{\lambda^2} \left( \lambda t + e^{-2\lambda t} - \frac{e^{-4\lambda t}}{4} - \frac{3}{4} \right).$$

*for all  $t \geq 0$ , respectively.*

3. *For any  $r > 0$  there exists a constant  $C(r)$  (does not depend on  $\lambda$ ) such that*

$$(B.1) \quad \mathbb{E}[|X(t; \mathfrak{s})|^r] \leq |v_0|^r (\tilde{C}(r) \lambda^{-r/2} t^{r/2} + C(r) \lambda^{-r/2-1} t^{r/2-1}) \quad \text{for all } t \geq 0,$$

*where*

$$(B.2) \quad \tilde{C}(r) := \int_{\mathbb{R}} dz \varphi(z) |z|^r = 2^{r/2} \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi}} \quad \text{with } \varphi(z) := \frac{e^{-z^2/2}}{\sqrt{2\pi}} \mathbb{1}_{\mathbb{R}}(z)$$

*and  $\Gamma$  denotes the Gamma function. The constant  $C(r)$  can be estimated from the proof of Theorem 1 in [23]. The following crude bound is also true*

$$\mathbb{E}[|X(t; \mathfrak{s})|^r] \leq |v_0|^r t^r \quad \text{for all } t \geq 0.$$

*The last bound is good for times  $t \ll 1$ .*

**Proof.** We start with the proof of Item 1. By (2.4) and Fubini's theorem we have

$$\mathbb{E}[X(t; \mathfrak{s})] = v_0 \int_0^t ds \mathbb{E}[(-1)^{N(s; \mathfrak{s})}] = v_0 \int_0^t ds \sum_{k=0}^{\infty} (-1)^k e^{-\lambda s} \frac{(\lambda s)^k}{k!} = v_0 \int_0^t ds e^{-2\lambda s} = \frac{v_0}{2\lambda} (1 - e^{-2\lambda t})$$

for all  $t \geq 0$ , which yields Item 1. The proof of Item 2 follows directly from Lemma 1 in [40]. In the sequel, we prove Item 3. To prove Item 3 we apply the estimate given in Theorem 1 in [23]. For  $p = 1$ ,  $\delta = 1$ ,  $M = 1$  in Theorem 1 in [23] we have the existence of a positive constant  $C$  such that

$$(B.3) \quad \int_{\mathbb{R}} dx e^{|x|} \left| t^{1/2} h(t, t^{1/2}x, 1) - \varphi(x) - t^{-1/2} \frac{x\varphi(x)}{2} \right| \leq \frac{C}{t} \quad \text{for all } t > 0,$$

where for  $t > 0$ ,  $\lambda > 0$ ,  $x \in \mathbb{R}$  the function  $h$  is given by

$$h(t, x, \lambda) := \frac{\lambda}{2} e^{-\lambda t} \left\{ \mathcal{I}_0(\lambda(t^2 - x^2)^{1/2}) + ((t+x)/(t-x))^{1/2} \mathcal{I}_1(\lambda(t^2 - x^2)^{1/2}) \right\} \mathbb{1}_{(-t, t)}(x).$$

Here,  $\mathcal{I}_0$  and  $\mathcal{I}_1$  denote the modified Bessel functions of order 0 and 1, respectively. In other words,

$$\mathcal{I}_0(z) = \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{(k!)^2} \quad \text{and} \quad \mathcal{I}_1(z) = \frac{z}{2} \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k!(k+1)!} \quad \text{for any } z \in \mathbb{R}.$$

Let  $r > 0$  be fixed. Note that there exists a positive constant  $K_r$  such that

$$(B.4) \quad |x|^r \leq K_r e^{|x|} \quad \text{for all } x \in \mathbb{R}.$$

Recall that

$$X(t; \mathfrak{s}) = v_0 \int_0^t ds (-1)^{N(s; \mathfrak{s})},$$

where the entries of  $\mathfrak{s} = (s_1, s_2, \dots)$  are i.i.d. with exponential distribution of parameter  $\lambda$ . Denote by  $\stackrel{\mathcal{D}}{=}$  equality in distribution. Note that

$$X(t; \mathfrak{s}) \stackrel{\mathcal{D}}{=} v_0 \int_0^t ds (-1)^{N(\lambda s; \mathfrak{u})},$$

where the entries of  $\mathfrak{u} = (u_1, u_2, \dots)$  are i.i.d. with exponential distribution of parameter 1. Hence for any  $r > 0$  we have

$$\mathbb{E}[|X(t; \mathfrak{s})|^r] = \frac{1}{\lambda^r} \mathbb{E} \left[ \left| v_0 \int_0^{\lambda t} ds (-1)^{N(s; \mathfrak{u})} \right|^r \right].$$

Consequently,

$$(B.5) \quad \int_{\mathbb{R}} dz h(\lambda t, z, \lambda) |z|^r = \frac{1}{\lambda^r} \int_{\mathbb{R}} dz h(\lambda t, z, 1) |z|^r.$$

We continue with the estimate of the  $r$ -th moment. The change of variable  $z = \tilde{t}^{1/2}x$  with  $\tilde{t} := \lambda t$  yields

$$(B.6) \quad \begin{aligned} \int_{\mathbb{R}} dz h(\lambda t, z, \lambda) |z|^r &= \int_{\mathbb{R}} dx \tilde{t}^{1/2} h(\tilde{t}, \tilde{t}^{1/2}x, 1) |\tilde{t}^{1/2}x|^r \\ &= \int_{\mathbb{R}} dx \left( \tilde{t}^{1/2} h(\tilde{t}, \tilde{t}^{1/2}x, 1) - \varphi(x) - \tilde{t}^{-1/2} \frac{x\varphi(x)}{2} \right) |\tilde{t}^{1/2}x|^r \\ &\quad + \int_{\mathbb{R}} dx \left( \varphi(x) + \tilde{t}^{-1/2} \frac{x\varphi(x)}{2} \right) |\tilde{t}^{1/2}x|^r. \end{aligned}$$

By (B.3) and (B.4) we have

$$(B.7) \quad \left| \int_{\mathbb{R}} dx \left( \tilde{t}^{1/2} h(\tilde{t}, \tilde{t}^{1/2}x, 1) - \varphi(x) - \tilde{t}^{-1/2} \frac{x\varphi(x)}{2} \right) |\tilde{t}^{1/2}x|^r \right| \leq \tilde{t}^{r/2-1} K_r C = t^{r/2-1} \lambda^{r/2-1} K_r C.$$

Since  $\varphi(x) = \varphi(-x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$  for all  $x \in \mathbb{R}$ , we have  $\int_{\mathbb{R}} dx x \varphi(x) |x|^r = 0$ . Then we observe that

$$(B.8) \quad \int_{\mathbb{R}} dx \left( \varphi(x) + \tilde{t}^{-1/2} \frac{x\varphi(x)}{2} \right) |\tilde{t}^{1/2} x|^r = \int_{\mathbb{R}} dx \varphi(x) |\tilde{t}^{1/2} x|^r = \tilde{C}(r) \tilde{t}^{r/2} = \tilde{C}(r) \lambda^{r/2} t^{r/2},$$

where  $\tilde{C}(r)$  is defined in (B.2). Relations (B.5) and (B.6) with the help of (B.7) and (B.8) imply Item 3. □

The next lemma shows that the  $p$ -th moment and the  $p$ -th absolute moment are observables that satisfy (2.12).

**Lemma B.2 (Moments and absolute moments).** *For any  $p \in \mathbb{N}$  it follows that*

$$(B.9) \quad \left| \|x\|^p - \|y\|^p \right| \leq K \frac{w(x) + w(y)}{2} \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^d,$$

where  $K = 2p$  and  $w : \mathbb{R}^d \rightarrow [0, \infty)$  is given by  $w(z) = \|z\|^{p-1}$  for all  $z \in \mathbb{R}^d$ . For  $d = 1$  let  $K = 2p$  and  $w(z) = |z|^{p-1}$  for all  $z \in \mathbb{R}$ . The following inequalities are also valid.

(i) For any  $p > 0$  it follows that

$$(B.10) \quad \left| |x|^p - |y|^p \right| \leq K \frac{w(x) + w(y)}{2} |x - y| \quad \text{for all } x, y \in \mathbb{R},$$

(ii) For any  $p \in \mathbb{N}$  it follows that

$$(B.11) \quad \left| x^p - y^p \right| \leq K \frac{w(x) + w(y)}{2} |x - y| \quad \text{for all } x, y \in \mathbb{R},$$

**Proof.** We start with the proof of (B.9). We first note that if  $\|x\| = 0$  or  $\|y\| = 0$ , then (B.9) holds. Hence, we assume that  $\|x\| \neq 0$  and  $\|y\| \neq 0$ . By the trichotomy property we have the following cases:

- (a)  $0 \neq \|x\| = \|y\|$ .
- (b)  $0 < \|x\| < \|y\|$ .
- (c)  $0 < \|y\| < \|x\|$ .

The case a) is immediate. We start with the case b). Let  $r := \|x\|/\|y\| \in (0, 1)$  and note that

$$\begin{aligned} \left| \|x\|^p - \|y\|^p \right| &= \|y\|^p - \|x\|^p = \|y\|^p (1 - r^p) = \|y\|^p (1 - r) \sum_{k=0}^{p-1} r^k \leq p \|y\|^p (1 - r) = p \|y\|^{p-1} (\|y\| - \|x\|) \\ &= p w(y) \left| \|x\| - \|y\| \right| \leq K \frac{w(x) + w(y)}{2} \|x - y\|. \end{aligned}$$

The case c) is completely analogous to the case b). The proof of (B.9) is complete.

We continue with the proof of (B.10). For  $p = 1$  it is easy to see that (B.10) holds. In the sequel, we assume that  $p \neq 1$ . Let  $x, y \in \mathbb{R}$  be fixed. For  $|x| = |y|$  it is easy to see that (B.10) holds. Without loss of generality, we assume that  $|x| \neq |y|$ . By the Fundamental Theorem of Calculus we have

$$\begin{aligned} \left| |y|^p - |x|^p \right| &= \left| p \int_{|x|}^{|y|} dz z^{p-1} \right| \leq \left| p \int_{|x|}^{|y|} dz |z|^{p-1} \right| \leq p \max_{|x| \wedge |y| \leq z \leq |x| \vee |y|} w(z) \left| |x| - |y| \right| \\ &\leq p (w(x) + w(y)) \left| |x| - |y| \right| \leq K \frac{w(x) + w(y)}{2} |x - y|, \end{aligned}$$

where the third inequality follows from the fact that the function  $(0, \infty) \ni z \mapsto w(z)$  is increasing for  $p > 1$  and decreasing for  $p \in (0, 1)$ . The proof of (B.10) is complete.

The proof of (B.11) follows from an analogous reasoning as used in the proof of (B.10). □

Lemma B.2 in Appendix B yields that for the observables ( $p$ -th moment and the  $p$ -th absolute moment), the weighted function  $w$  that appears in (2.12) can be chosen as  $w(z) = \|z\|^{p-1}$ ,  $z \in \mathbb{R}^d$ .



In the sequel, we estimate the constants  $C_1$  and  $C_2$  (for  $w(z) = |z|^{p-1}$ ,  $z \in \mathbb{R}$ ) that appears in (2.11), i.e.,

$$(B.12) \quad C_1^2 := \int \mu_1(dX) \frac{1}{T} \int_0^T ds (w(X(s)))^2 < \infty, \quad C_2^2 := \int \mu_2(dY) \frac{1}{T} \int_0^T ds (w(Y(s)))^2 < \infty,$$

for the processes  $X$  and  $Y$  being replaced by the rescaled telegraph process  $L^{-1}X$  and a Brownian motion  $B$  with a suitable diffusivity constant  $\sigma^2$ , respectively. Hence the constant  $\max\{C_1, C_2\}$ , which is needed in the estimate given in the right-hand side of (2.14), is also estimated.

**Lemma B.3 (Time average estimates for the  $w$ -weight).** *For  $p = 1$  it follows that  $C_1^2 = C_2^2 = 1$ . For any  $p > 0$  with  $p \neq 1$ , the following estimates are valid*

$$C_1^2 \leq \frac{\tilde{C} T_*^{p-1}}{p L_*^{2(p-1)}} + \frac{C T_*^{p-2}}{(p-1) L_*^{2(p-1)}} \quad \text{and} \quad C_2^2 = \frac{2^{p-1} \Gamma(\frac{2p-1}{2})}{p \sqrt{\pi}} \frac{T_*^{p-1}}{L_*^{2(p-1)}},$$

where the constants  $T_*$  and  $L_*$  are the scaling parameters that appears in (2.8),  $\tilde{C}$  and  $C$  are the constants that appears for  $r = 2(p-1)$  in Item 3, (B.1), of Lemma B.1 in Appendix B and  $\Gamma$  denotes the Gamma function.

**Proof.** We observe that  $C_1^2 = C_2^2 = 1$  for  $p = 1$ . We start with the estimate of  $C_1^2$  for  $p \neq 1$ . In this case, the process  $X$  is replaced by  $L^{-1}X$  in (B.12). A direct computation yields

$$\begin{aligned} C_1^2 &= \int \mu_1(dX) \frac{1}{T} \int_0^T ds (w(L^{-1}X(s; \mathfrak{s})))^2 = \int \mu_1(dX) \frac{1}{T} \int_0^T ds |L^{-1}X(s; \mathfrak{s})|^{2(p-1)} \\ &= \frac{1}{L^{2(p-1)} T} \int_0^T ds \mathbb{E}[|X(s; \mathfrak{s})|^{2(p-1)}] \leq \frac{|v_0|^{2(p-1)}}{L^{2(p-1)} T} \int_0^T ds (\tilde{C} \lambda^{-(p-1)} s^{p-1} + C \lambda^{-p} s^{p-2}) \\ &= \frac{|v_0|^{2(p-1)} \tilde{C} \lambda^{-(p-1)} T^{p-1}}{p L^{2(p-1)}} + \frac{|v_0|^{2(p-1)} C \lambda^{-p} T^{p-2}}{(p-1) L^{2(p-1)}} = \frac{\tilde{C} T_*^{p-1}}{p L_*^{2(p-1)}} + \frac{C T_*^{p-2}}{(p-1) L_*^{2(p-1)}}, \end{aligned}$$

where the constants  $\tilde{C}$  and  $C$  are the constants that appears for  $r = 2(p-1)$  in Item 3, (B.1), of Lemma B.1 in Appendix B.

We continue with the estimate of  $C_2^2$ . Since the process  $Y$  in (B.12) is the Brownian motion with diffusivity constant  $\sigma^2 = \lambda L_*^{-2}$  given in (2.18), we have

$$\begin{aligned} C_2^2 &= \int \mu_2(dY) \frac{1}{T} \int_0^T ds (w(B(s)))^2 = \int \mu_2(dY) \frac{1}{T} \int_0^T ds |B(s)|^{2(p-1)} = \frac{1}{T} \int_0^T ds \mathbb{E}[|B(s)|^{2(p-1)}] \\ &= \mathbb{E}[|\sigma B(1)|^{2(p-1)}] \frac{1}{T} \int_0^T ds s^{p-1} = \frac{\sigma^{2(p-1)} 2^{p-1} \Gamma(\frac{2p-1}{2})}{\sqrt{\pi}} \frac{T^{p-1}}{p} = \frac{2^{p-1} \Gamma(\frac{2p-1}{2})}{p \sqrt{\pi}} \frac{T_*^{p-1}}{L_*^{2(p-1)}}, \end{aligned}$$

where in the last inequality we used relation (B.2) given in Item 3 of Lemma B.1 in Appendix B.  $\square$

In the following we estimate the distance between  $L^{-1}\mathbb{X}_{[0,T]}^{\mathfrak{s}} = (L^{-1}X(t; \mathfrak{s}) : 0 \leq t \leq T)$  and  $\mathbb{B}_{[0,T]} = (B(t) : 0 \leq t \leq T)$  by the independent coupling between them.

**Lemma B.4 (Independent coupling bound).** *For any  $L > 0$ ,  $T > 0$ ,  $\lambda > 0$  and  $v_0 \in \mathbb{R} \setminus \{0\}$  it follows that*

$$\mathcal{W}_2(L^{-1}\mathbb{X}_{[0,T]}^{\mathfrak{s}}, \mathbb{B}_{[0,T]}) \leq \left( \frac{1}{4T_* L_*^2} (1 - e^{-2T_*}) - \frac{1}{2L_*^2} + \frac{T_*}{L_*^2} \right)^{1/2},$$

where  $T_* = \lambda T$  and  $L_* = |v_0|^{-1} \lambda L$ .

**Proof.** We estimate the following expectation

$$\mathbb{E}_{\mathcal{Y}} \left[ \frac{1}{T} \int_0^T dt (L^{-1}X(t; \mathfrak{s}) - B(t))^2 \right]$$

for a suitable coupling  $\gamma$  between  $L^{-1}\mathbb{X}_{[0,T]}^s$  and  $\mathbb{B}_{[0,T]}$ . We recall that the diffusivity constant of  $B$  is given by  $\sigma^2 = \lambda L_*^{-2}$  given in (2.18). By Fubini's theorem and Pythagoras's theorem we obtain

$$\begin{aligned} \mathbb{E}_\gamma \left[ \frac{1}{T} \int_0^T dt (L^{-1}X(t; \mathfrak{s}) - B(t))^2 \right] &= \frac{1}{T} \int_0^T dt \mathbb{E}_\gamma [(L^{-1}X(t; \mathfrak{s}) - B(t))^2] \\ (B.13) \qquad \qquad \qquad &= \frac{1}{T} \int_0^T dt (\mathbb{E}_\gamma [(L^{-1}X(t; \mathfrak{s}))^2] + \mathbb{E}_\gamma [(B(t))^2]) \\ &\quad - \frac{2}{T} \int_0^T dt \mathbb{E}_\gamma [L^{-1}X(t; \mathfrak{s})B(t)]. \end{aligned}$$

By Lemma B.1 in Appendix B we have

$$(B.14) \qquad \qquad \mathbb{E}_\gamma [(X(t; \mathfrak{s}))^2] = \frac{v_0^2}{2\lambda^2} (2\lambda t - (1 - e^{-2\lambda t})) \quad \text{for all } t \geq 0$$

and for any coupling  $\gamma$ . Since  $(B(t) : t \geq 0)$  is a Brownian motion with diffusivity  $\sigma^2$ , the second moment of its marginal  $B(t)$  is given by

$$(B.15) \qquad \qquad \mathbb{E}_\gamma [(B(t))^2] = \sigma^2 t \quad \text{for all } t \geq 0$$

for any coupling  $\gamma$ . Next, we estimate the contribution of the term  $\mathbb{E}_\gamma [L^{-1}X(t; \mathfrak{s})B(t)]$  for a suitable coupling  $\gamma$ . By the use of the independent coupling  $\gamma = \pi$  we obtain

$$(B.16) \qquad \qquad \mathbb{E}_\pi [L^{-1}X(t; \mathfrak{s})B(t)] = \mathbb{E}_\pi [L^{-1}X(t; \mathfrak{s})] \mathbb{E}_\pi [B(t)] = 0.$$

Combining (B.14), (B.15) and (B.16) in (B.13) implies

$$\begin{aligned} &\inf_\gamma \mathbb{E}_\gamma \left[ \frac{1}{T} \int_0^T dt (L^{-1}X(t; \mathfrak{s}) - B(t))^2 \right] \\ &\leq \mathbb{E}_\pi \left[ \frac{1}{T} \int_0^T dt (L^{-1}X(t; \mathfrak{s}) - B(t))^2 \right] \\ (B.17) \qquad \qquad \qquad &= \frac{1}{T} \int_0^T dt (\mathbb{E}_\pi [(L^{-1}X(t; \mathfrak{s}))^2] + \mathbb{E}_\pi [(B(t))^2]) \\ &= \frac{1}{T} \int_0^T dt \left( \frac{v_0^2}{2\lambda^2 L^2} (2\lambda t - 1 + e^{-2\lambda t}) + \sigma^2 t \right) \\ &= \frac{1}{T} \left( \frac{v_0^2}{2\lambda^2 L^2} \lambda T^2 - \frac{v_0^2}{2\lambda^2 L^2} T + \frac{v_0^2}{2\lambda^2 L^2} \frac{1}{2\lambda} (1 - e^{-2\lambda T}) + \frac{\sigma^2 T^2}{2} \right) \\ &= \frac{T_*}{2L_*^2} - \frac{1}{2L_*^2} + \frac{1}{4L_*^2 T_*} (1 - e^{-2T_*}) + \frac{1}{2} \sigma^2 T, \end{aligned}$$

where  $T_* = \lambda T$  and  $L_* = |v_0|^{-1} \lambda L$ . Since  $\sigma^2 = \lambda L_*^{-2}$ , (B.17) yields

$$\mathcal{W}_2^2(L^{-1}\mathbb{X}_{[0,T]}^s, \mathbb{B}_{[0,T]}) \leq \frac{1}{4T_* L_*^2} (1 - e^{-2T_*}) - \frac{1}{2L_*^2} + \frac{T_*}{L_*^2},$$

which yields the statement. □

For completeness of the presentation, we recall the Komlós–Major–Tusnády coupling (see Theorem 1 of [32]) and we state it as a lemma.

**Lemma B.5 (Komlós–Major–Tusnády Theorem).** *Let  $F : \mathbb{R} \rightarrow [0, 1]$  be a distribution function satisfying*

$$\int_{\mathbb{R}} e^{r \cdot z} F(dz) < \infty \quad \text{for any } r \in (-r_0, r_0)$$

for some positive number  $r_0 > 0$ . In addition, we assume

$$\int_{\mathbb{R}} zF(dz) = 0 \quad \text{and} \quad \int_{\mathbb{R}} z^2F(dz) = 1.$$

Then there exists a probability space  $(\Omega_*, \mathcal{F}_*, \pi^*)$  for which we can construct sequences  $X := (X_n : n \in \mathbb{N})$  and  $Y := (Y_n : n \in \mathbb{N})$  of random variables satisfying

- (a)  $X$  is i.i.d. with common distribution function  $F$ ,
- (b)  $Y$  is i.i.d. with standard Gaussian distribution

in a way that for all  $x > 0$  and every  $n \in \mathbb{N}$  the following estimate is valid

$$\pi^* \left( \max_{1 \leq m \leq n} \left| \sum_{j=1}^m X_j - \sum_{j=1}^m Y_j \right| \geq C \ln(n) + x \right) \leq K e^{-\vartheta x}.$$

Here, the positive constants  $C, K, \vartheta$  depend only on  $F$ , and  $\vartheta$  can be taken as large as desired by choosing  $C$  large enough. Consequently,  $|\sum_{j=1}^n X_j - \sum_{j=1}^n Y_j| = \mathcal{O}(\ln(n))$  almost surely.

### Appendix C: Definition and basic properties of the $\delta$ -constraints

In the text, we frequently integrate with respect to measures which are identified using the somewhat formal Dirac  $\delta$ -function notation. As explained in Appendix A in [35], this slightly formal notion can often be given a fully rigorous definition as a positive Radon measure obtained via the Riesz–Markov–Kakutani representation theorem applied to a limit of a sequence of approximations where the  $\delta$ -function is replaced by a suitably chosen positive ordinary function. The benefit of using the approximations becomes apparent when one needs to perform operations which are standard to Lebesgue measures, such as changes of integration variables or splitting with aid of Fubini’s theorem.

In this appendix, we explain the definition and derive the properties used in the text related to integration with respect to the probability measure  $\nu(d^n u)$  which is defined on  $\mathbb{R}_*^n, n \in \mathbb{N} \setminus \{1\}$ , by the formula

$$\nu(d^n u) = d^n u (n-1)! \delta(t_n(u) - 1),$$

where  $t_n(u) = \sum_{j=1}^n u_j$  and  $d^n u$  denotes the restriction of the Lebesgue measure to  $\mathbb{R}_*^n$ . These measures are obtained by scaling from the more general family of positive measures

$$\tilde{\nu}_r(d^n u) = d^n u \delta(t_n(u) - r), \quad r > 0,$$

since  $\nu(d^n u) = (n-1)! \tilde{\nu}_1(d^n u)$ .

The most direct definition of the measure  $\tilde{\nu}_r$  is obtained by formal integration of the  $\delta$ -function over the last coordinate,  $u_n$ . We define a positive semidefinite linear functional  $\Lambda(\cdot; r)$  on  $C_c(\mathbb{R}_*^n)$  by setting for any continuous function  $f : \mathbb{R}_*^n \rightarrow \mathbb{C}$  which has a compact support,

$$(C.1) \quad \Lambda(f; r) := \int_{X'(n,r)} d^{n-1} u f \left( u, r - \sum_{j=1}^{n-1} u_j \right),$$

where  $X'(n, r) := \{u \in \mathbb{R}_*^{n-1} : \sum_{j=1}^{n-1} u_j < r\}$  is a bounded Borel set. Since  $\mathbb{R}_*^n$  is a locally compact Hausdorff space, the Riesz–Markov–Kakutani representation theorem implies that there exists a unique Radon measure  $\tilde{\nu}_r$  for which  $\Lambda(f; r) = \int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) f(u)$ . Now  $K_m := \{u \in \mathbb{R}_*^n : \frac{1}{m} \leq u_j \leq m \text{ for all } j\}, m \in \mathbb{N}$ , form an increasing covering sequence of compact subsets of  $\mathbb{R}_*^n$ . Thus by Urysohn’s lemma, for each  $m \in \mathbb{N}$ , we can find a continuous function  $\phi_m : \mathbb{R}_*^n \rightarrow \mathbb{R}$  such that  $0 \leq \phi_m \leq 1$  and  $\phi_m(u) = 1$ , if  $u \in K_m$ , and  $\phi_m(u) = 0$ , if  $u \notin K_{m+1}$ . Then each  $\phi_m$  has a compact support and, by the dominated convergence theorem, we find that  $\Lambda(\phi_m; r) \rightarrow \int_{X'(n,r)} d^{n-1} u 1$ , as  $m \rightarrow \infty$ . Here, by induction in  $n$ , one can check the well-known simplex volume result,

$$\int_{X'(n,r)} d^{n-1} u 1 = \int_{\mathbb{R}_*^{n-1}} d^{n-1} u \mathbb{1}_{\{\sum_{i=1}^{n-1} u_i < r\}} = \frac{r^{n-1}}{(n-1)!},$$

which is valid for all  $r > 0$  and  $n \in \mathbb{N}$  with  $n \geq 2$ . Since clearly also  $|\Lambda(f; r)| \leq \frac{r^{n-1}}{(n-1)!} \|f\|_\infty$  for any  $f \in C_c(\mathbb{R}_*^n)$ , we find that  $\int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) 1 = \frac{r^{n-1}}{(n-1)!}$ . Therefore, each  $\tilde{\nu}_r$  is a bounded, positive semidefinite Radon measure, and  $\nu(d^n u) = (n-1)! \tilde{\nu}_1(d^n u)$  is a Radon probability measure on  $\mathbb{R}_*^n$ .

The following technical lemma shows that the formula (C.1) indeed applies to a much larger class of functions. In addition, it also gives a convenient approximation result by ordinary integrals for integration of continuous functions. This can be used to derive many further properties about integrals over  $\tilde{\nu}_r(d^n u)$ , similar to what is done in Appendix A in [35]. We derive the main properties used in the present work from these, as summarized later in Lemma C.2.

**Lemma C.1.** *Suppose  $n \geq 2$  and  $r > 0$ . Denote  $E_r := \{u \in \mathbb{R}_*^n : t_n(u) \neq r\}$ . Then all of the statements below hold.*

1.  $\tilde{\nu}_r(E_r) = 0$ .
2. If  $A \subset \mathbb{R}_*^{n-1}$  has Lebesgue measure zero, then  $\tilde{\nu}_r(A \times \mathbb{R}_*) = 0$ .
3. Suppose  $f : \mathbb{R}_*^n \rightarrow \mathbb{C}$  is a Borel measurable function. If  $f$  is non-negative or  $f \in L^1(\tilde{\nu}_r)$ , we have

$$(C.2) \quad \int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) f(u) = \int_{\mathbb{R}_*^{n-1}} d^{n-1} u \mathbb{1}_{\{t_{n-1}(u) < r\}} f(u, r - t_{n-1}(u)) = \frac{r^{n-1}}{(n-1)!} \int_{\mathbb{R}_*^n} \nu(d^n u) f(ru).$$

4. Suppose  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is a continuous function with a compact support for which  $\varphi \geq 0$  and  $\int_{\mathbb{R}} dx \varphi(x) = 1$ . Define  $\Phi_\varepsilon(u; r) := \varepsilon^{-1} \varphi((t_n(u) - r)/\varepsilon)$  for  $u \in \mathbb{R}_*^n$  and  $\varepsilon > 0$ . Then for any function  $F : [0, \infty)^n \rightarrow \mathbb{C}$  which is continuous, we have

$$\int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) F(u) = \lim_{\varepsilon \rightarrow 0^+} \int_{\mathbb{R}_*^n} d^n u \Phi_\varepsilon(u; r) F(u).$$

5. The measure  $\tilde{\nu}_r$  is permutation invariant: if  $P$  is a permutation of  $\{1, 2, \dots, n\}$ , then for any measurable function  $f$  the function  $u \mapsto f(u_P)$ ,  $(u_P)_j := u_{P(j)}$ ,  $j = 1, 2, \dots, n$ , has the same integral as  $f$ .

**Proof.** For the first item, consider the set  $S = \mathbb{R}_*^n \setminus E_r$ . We choose a cutoff function  $\eta : \mathbb{R} \rightarrow \mathbb{R}$  which is continuous,  $0 \leq \eta \leq 1$ , and  $\eta(x) = 1$ , if  $|x| \leq 1$ , and  $\eta(x) = 0$ , if  $|x| \geq 2$ . Recall the earlier cutoff functions  $\phi_m$ , and define  $F_m(u) := \phi_m(u) \eta(m(t_n(u) - r))$ . If  $u \in S$ , then  $F_m(u) = \phi_m(u) \rightarrow 1$  as  $m \rightarrow \infty$ . If  $u \notin S$ , then  $\eta(m(t_n(u) - r)) \rightarrow 0$  as  $m \rightarrow \infty$ , and thus also  $\lim_{m \rightarrow \infty} F_m(u) = 0$ . Therefore, by the dominated convergence theorem,  $\tilde{\nu}_r(S) = \lim_{m \rightarrow \infty} \int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) F_m(u)$ . Since  $F_m \in C_c(\mathbb{R}_*^n)$ , we have

$$\int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) F_m(u) = \int_{X'(n,r)} d^{n-1} u \phi_m\left(u, r - \sum_{j=1}^{n-1} u_j\right) \eta(0) = \int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) \phi_m(u),$$

which goes to  $\tilde{\nu}_r(\mathbb{R}_*^n) < \infty$  as  $m \rightarrow \infty$ . Therefore,  $S$  has full measure, and thus its complement  $E_r$  must have measure zero. This implies that for any bounded measurable function  $f$  we have

$$\int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n u) f(u) = \int_S \tilde{\nu}_r(d^n u) f(u) = \int_S \tilde{\nu}_r(d^n u) f(u, r - t_{n-1}(u))|_{u=P_{n-1}u}$$

where  $P_{n-1}u := (u_1, u_2, \dots, u_{n-1}) \in \mathbb{R}_*^{n-1}$ .

For the next item, let us first suppose that  $A \subset \mathbb{R}_*^{n-1}$  is open. Then it can be written as a limit of an increasing sequence of compact sets. Applying Urysohn’s lemma with open sets  $A \times (0, m + 1)$  and an increasing sequence of compact sets obtained by taking a product with the above sets and  $[1/m, m]$ , we find a sequence of functions  $f_m \in C_c(\mathbb{R}_*^n)$  which pointwise monotone increase to the indicator function of  $A \times \mathbb{R}_*$ . Thus by using the monotone convergence theorem, we conclude that  $\tilde{\nu}_r(A \times \mathbb{R}_*) = \int_A d^{n-1} u$ . If  $A \subset \mathbb{R}_*^{n-1}$  has zero Lebesgue measure, it can be covered with countably many open sets so that the sum of their measures is arbitrarily small. Hence, we conclude that then  $\tilde{\nu}_r(A \times \mathbb{R}_*) = 0$ .

For the third item, we construct a suitable regularisation so that the defining functional can be used to evaluate the integral. In fact, we only need to do this assuming that  $f : \mathbb{R}_*^n \rightarrow \mathbb{C}$  is bounded and Borel measurable. Namely, suppose that (C.2) holds for such  $f$ . If  $f$  is non-negative, we can then apply the result to every  $f_N(u) := \mathbb{1}_{\{f(u) \leq N\}} f(u)$ ,  $N \in \mathbb{N}$ . This is a monotone non-decreasing sequence converging to  $f$ , and thus the monotone convergence theorem may be applied to each of the three integrals in (C.2). This shows that (C.2) holds for all measurable, non-negative functions. Finally, if  $g \in L^1(\tilde{\nu}_r)$ , then  $f = |g|$  is a non-negative function for which the first integral in (C.2) is finite. Thus by the previous result also the other two integrals are then finite. Therefore,  $f$  is a dominant for the function sequence  $g_N$ ,

defined by  $g_N(\mathbf{u}) := \mathbb{1}_{\{|g(\mathbf{u})| \leq N\}} g(\mathbf{u})$ ,  $N \in \mathbb{N}$ . Since each  $g_N$  is bounded and  $g_N \rightarrow g$  pointwise, dominated convergence theorem, applied to each of the three integrals separately, implies that (C.2) holds also for  $g$ .

Let us thus assume that  $f : \mathbb{R}_*^n \rightarrow \mathbb{C}$  is bounded and Borel measurable. We then choose continuous cutoff functions  $\eta_\delta : \mathbb{R} \rightarrow \mathbb{R}$ ,  $0 < \delta \leq 1$ , which satisfy  $0 \leq \eta_\delta \leq 1$ ,  $\eta_\delta(x) = 0$  if  $|x| \geq 1 + \delta$ , and  $\eta_\delta(x) = 1$  if  $|x| \leq 1$ . For  $\varepsilon > 0$ , we define

$$\chi_{\delta,\varepsilon}(x) := \eta_\delta(|x|/\varepsilon), \quad x \in \mathbb{R}^{n-1},$$

and set, for all  $x \in \mathbb{R}^{n-1}$ ,

$$h(x) := \begin{cases} f(x, r - t_{n-1}(x)), & \text{if } x \in \mathbb{R}_*^{n-1} \text{ and } t_{n-1}(x) < r, \\ 0, & \text{otherwise.} \end{cases}$$

Then  $h$  is bounded,  $h \in L^1(\mathbb{R}^{n-1})$  and

$$\int_{\mathbb{R}^{n-1}} d^{n-1}x h(x) = \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \mathbb{1}_{\{t_{n-1}(u) < r\}} f(u, r - t_{n-1}(u)).$$

The regularisation we need will be provided by the functions

$$h_{\delta,\varepsilon}(x) := \frac{1}{Z_{\delta,\varepsilon}} \int_{\mathbb{R}^{n-1}} d^{n-1}y \chi_{\delta,\varepsilon}(x-y) h(y),$$

where

$$Z_{\delta,\varepsilon} := \int_{\mathbb{R}^{n-1}} d^{n-1}y \chi_{\delta,\varepsilon}(y) < \infty.$$

By applying the Lebesgue dominated convergence theorem, we then find that each  $h_{\delta,\varepsilon}$  is continuous function on  $\mathbb{R}^{n-1}$ . Also,  $\|h_{\delta,\varepsilon}\|_\infty \leq \|f\|_\infty$  and the support of each  $h_{\delta,\varepsilon}$  is contained in the compact set  $\{x \in [0, \infty)^{n-1} : |x| \leq 2 + r\}$ .

Let  $X_0 \subset \mathbb{R}^{n-1}$  collect all Lebesgue points of  $h$ . Since  $h \in L^1(\mathbb{R}^{n-1})$ , we know that its complement has Lebesgue measure zero. By definition, for every  $x \in X_0$  we then have

$$\lim_{\varepsilon \rightarrow 0^+} \frac{1}{|B_\varepsilon|} \int_{B_\varepsilon} dy |h(x-y) - h(x)| = 0,$$

where  $B_\varepsilon$  denotes the closed ball with a radius  $\varepsilon$ , centred at the origin. Now, if  $\varepsilon > 0$ ,

$$\lim_{\delta \rightarrow 0^+} \chi_{\delta,\varepsilon}(y) = \mathbb{1}_{\{|y| \leq \varepsilon\}},$$

and thus by applying the dominated convergence theorem twice, we find that for all  $x$

$$\lim_{\delta \rightarrow 0^+} h_{\delta,\varepsilon}(x) = \frac{1}{|B_\varepsilon|} \int_{B_\varepsilon} dy h(x-y).$$

Therefore,

$$\lim_{\varepsilon \rightarrow 0^+} \left( \lim_{\delta \rightarrow 0^+} h_{\delta,\varepsilon}(x) \right) = h(x), \quad x \in X_0.$$

For notational convenience, let us denote from now on the above double limit by “ $\lim_{\varepsilon,\delta}$ ”. Finally, we recall the increasing sequence of cutoff functions  $\phi_m$  used earlier. For  $m \in \mathbb{N}$  and  $\mathbf{u} \in \mathbb{R}_*^n$  we define

$$f_{m,\delta,\varepsilon}(\mathbf{u}) := h_{\delta,\varepsilon}(P_{n-1}\mathbf{u})\phi_m(\mathbf{u}),$$

where the projection  $P_{n-1}$  was defined above. By construction,  $f_{m,\delta,\varepsilon} \in C_c(\mathbb{R}_*^n)$ , and thus

$$\begin{aligned} \int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n \mathbf{u}) f_{m,\delta,\varepsilon}(\mathbf{u}) &= \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \mathbb{1}_{\{t_{n-1}(u) < r\}} f_{m,\delta,\varepsilon}(u, r - t_{n-1}(u)) \\ (C.3) \qquad \qquad \qquad &= \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \mathbb{1}_{\{t_{n-1}(u) < r\}} h_{\delta,\varepsilon}(u)\phi_m(u, r - t_{n-1}(u)). \end{aligned}$$

The dominated convergence theorem allows us to conclude that the limit of the right hand side is given by

$$\int_{\mathbb{R}_*^{n-1}} d^{n-1}u \mathbb{1}_{\{t_{n-1}(u) < r\}} h(u) = \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \mathbb{1}_{\{t_{n-1}(u) < r\}} f(u, r - t_{n-1}(u)).$$

Again, by the dominated convergence theorem, the limit of the left-hand side of (C.3), as  $m \rightarrow \infty$ , is given by  $\int_{\mathbb{R}_*^n} \tilde{v}_r(d^n u) h_{\delta, \varepsilon}(P_{n-1}u)$ . Furthermore,

$$\lim_{\varepsilon, \delta} \int_{X_0 \times \mathbb{R}_*} \tilde{v}_r(d^n u) h_{\delta, \varepsilon}(P_{n-1}u) = \int_{X_0 \times \mathbb{R}_*} \tilde{v}_r(d^n u) h(P_{n-1}u) = \int_{X_0 \times \mathbb{R}_*} \tilde{v}_r(d^n u) f(u),$$

where in the last equality we used the observation that if  $t_n(u) = r$ , then also  $t_{n-1}(P_{n-1}u) < r$  and thus  $f(u) = f(P_{n-1}u, r - \sum_{j=1}^{n-1} u_j) = h(P_{n-1}u)$ . However, by the second item, now  $\tilde{v}_r(X_0^c \times \mathbb{R}_*) = 0$ , and thus

$$\lim_{\varepsilon, \delta} \int_{X_0 \times \mathbb{R}_*} \tilde{v}_r(d^n u) h_{\delta, \varepsilon}(P_{n-1}u) = \int_{\mathbb{R}_*^n} \tilde{v}_r(d^n u) f(u).$$

We can conclude that the first equality in (C.2) holds for bounded Borel measurable  $f$ . But then for any such function  $f$ , we also have

$$\frac{1}{(n-1)!} \int_{\mathbb{R}_*^n} v(d^n u) f(ru) = \int_{\mathbb{R}_*^n} \tilde{v}_1(d^n u) f(ru) = \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \mathbb{1}_{\{t_{n-1}(u) < 1\}} f(ru, r - rt_{n-1}(u)).$$

A change of variables to  $v = ru$  in the last integral yields

$$r^{-(n-1)} \int_{\mathbb{R}_*^{n-1}} d^{n-1}v \mathbb{1}_{\{t_{n-1}(v) < r\}} f(v, r - t_{n-1}(v)) = r^{-(n-1)} \int_{\mathbb{R}_*^n} \tilde{v}_r(d^n v) f(v).$$

Therefore, (C.2) holds for all bounded Borel measurable  $f$ . This completes the proof of the third item.

We next prove the fourth statement. Given  $0 < \varepsilon \leq 1$ , let us denote

$$I_\varepsilon = \int_{\mathbb{R}_*^n} d^n u \Phi_\varepsilon(u) F(u).$$

Since  $\varphi$  has a compact support, there is  $R > 0$  such that  $\varphi(x) = 0$  for all  $|x| \geq R$ . Thus  $\Phi_\varepsilon(u) = 0$  if  $t_n(u) \geq r + R$ , using the assumption  $\varepsilon \leq 1$ . Denote  $Y := \{x \in [0, \infty)^n : t_n(x) \leq r + R\}$  and  $Y' := Y \cap \mathbb{R}_*^n$ . Since  $F$  is continuous and  $Y$  is compact,  $M := \sup_{x \in Y} |F(x)| < \infty$ . We thus find that  $\Phi_\varepsilon(u) |F(u)| \leq M \Phi_\varepsilon(u)$ .

We can thus use Fubini's theorem to reorder the integrals, resulting in

$$I_\varepsilon = \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \left( \int_0^\infty du_n \varepsilon^{-1} \varphi((u_n + t_{n-1}(u) - r)/\varepsilon) F(u, u_n) \right).$$

A change of variables from  $u_n$  to  $x := (u_n + t_{n-1}(u) - r)/\varepsilon$  yields

$$I_\varepsilon = \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \left( \int_{-\infty}^\infty dx \mathbb{1}_{\{x > (\sum_{j=1}^{n-1} u_j - r)/\varepsilon\}} \varphi(x) F\left(u, r - \sum_{j=1}^{n-1} u_j + \varepsilon x\right) \right).$$

If  $\sum_{j=1}^{n-1} u_j \geq r + R\varepsilon$ , then either the indicator function or  $\varphi(x)$  is zero, implying that the integrand is then zero. On the other hand, the measure of the set  $\{u \in \mathbb{R}_*^{n-1} : r \leq \sum_{j=1}^{n-1} u_j < r + R\varepsilon\}$  goes to zero as  $\varepsilon \rightarrow 0$ . Therefore, by the dominated convergence theorem and the continuity of  $F$  we have

$$\lim_{\varepsilon \rightarrow 0} I_\varepsilon = \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \mathbb{1}_{\{\sum_{j=1}^{n-1} u_j < r\}} \left( \int_{-\infty}^\infty dx \varphi(x) F\left(u, r - \sum_{j=1}^{n-1} u_j\right) \right) = \int_{\mathbb{R}_*^n} \tilde{v}_r(d^n u) F(u),$$

where the last equality follows from the normalisation of  $\varphi$  and the first part of the Lemma.

The final item is a corollary of the previous one: if  $f \in C_c(\mathbb{R}_*^n)$ , we can apply the previous item to approximate its expectation. However,  $d^n u \Phi_\varepsilon(u; r)$  is obviously permutation invariant, and we obtain the statement for such functions  $f$ .

By the uniqueness statement in the Riesz–Markov–Kakutani theorem, this is sufficient to conclude that the Radon measure  $\tilde{\nu}_r$  itself is permutation invariant.  $\square$

Essentially as a corollary, we obtain the following technical estimates used in the paper.

**Lemma C.2 ( $\delta$ -integration).** *The following formulas are valid.*

(i) *Marginal moments.* Suppose  $n \geq 2$  and  $f : [0, \infty) \rightarrow \mathbb{C}$  is a continuous function. Then

$$(C.4) \quad \langle f(u_1) \rangle_{\nu(d^n \mathbf{u})} = (n - 1) \int_0^1 du_1 f(u_1) (1 - u_1)^{n-2}.$$

(ii) *Correlations.* Suppose  $n \geq 3$  and  $g : [0, \infty)^2 \rightarrow \mathbb{C}$  is a continuous function. Then

$$\langle g(u_1, u_2) \rangle_{\nu(d^n \mathbf{u})} = (n - 1)(n - 2) \int_0^1 du_1 \int_0^{1-u_1} du_2 g(u_1, u_2) (1 - u_1 - u_2)^{n-3}.$$

(iii) *Concentration.* We have  $t_n(\mathbf{u}) = 1$  almost surely under the probability measure  $\nu(d^n \mathbf{u})$ .

(iv) “Fubini’s theorem”. Suppose  $n \geq 2$  and  $f : \mathbb{R}_*^n \rightarrow \mathbb{C}$  is a Borel measurable function. If  $f$  is also either non-negative, or Lebesgue integrable, then

$$(C.5) \quad \int_{\mathbb{R}_*^n} d^n \mathbf{v} f(\mathbf{v}) = \int_{\mathbb{R}_*} dr \left( \int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n \mathbf{v}) f(\mathbf{v}) \right) = \int_{\mathbb{R}_*} dr \frac{r^{n-1}}{(n - 1)!} \left( \int_{\mathbb{R}_*^n} \nu(d^n \mathbf{u}) f(r\mathbf{u}) \right),$$

where the integrals are either all absolutely convergent or all infinite.

**Proof.** The third item is a direct corollary of Item 1 of Lemma C.1. Therefore, the assumed continuity properties of  $f$  and  $g$  imply that the observables  $\mathbf{u} \mapsto f(u_1)$  and  $\mathbf{u} \mapsto g(u_1, u_2)$  are bounded on the support of the measure  $\nu$ . Hence, they are integrable, and we can apply Item 3 of Lemma C.1.

Assuming  $n \geq 3$ , we first obtain

$$\langle g(u_1, u_2) \rangle_{\nu(d^n \mathbf{u})} = (n - 1)! \int_{\mathbb{R}_*^{n-1}} d^{n-1} \mathbf{u} \mathbb{1}_{\{t_{n-1}(\mathbf{u}) < 1\}} g(u_1, u_2).$$

If  $n = 3$ , this is equal to  $\int_0^1 du_1 \int_0^{1-u_1} du_2 g(u_1, u_2)$ . If  $n > 3$ , the indicator function is still zero if  $u_1 + u_2 \geq 1$ , and thus we have

$$\int_{\mathbb{R}_*^{n-1}} d^{n-1} \mathbf{u} \mathbb{1}_{\{t_{n-1}(\mathbf{u}) < 1\}} g(u_1, u_2) = \int_0^1 du_1 \int_0^{1-u_1} du_2 g(u_1, u_2) \int_{\mathbb{R}_*^{n-3}} d^{n-3} \mathbf{v} \mathbb{1}_{\{\sum_{i=1}^{n-3} v_i < 1 - u_1 - u_2\}}.$$

Hence, by using the earlier simplex volume result, we can conclude that for all  $n \geq 3$ ,

$$\langle g(u_1, u_2) \rangle_{\nu(d^n \mathbf{u})} = (n - 1)(n - 2) \int_0^1 du_1 \int_0^{1-u_1} du_2 g(u_1, u_2) (1 - u_1 - u_2)^{n-3}.$$

Similarly, if  $n \geq 2$ , we have

$$\langle f(u_1) \rangle_{\nu(d^n \mathbf{u})} = (n - 1)! \int_{\mathbb{R}_*^{n-1}} d^{n-1} \mathbf{u} \mathbb{1}_{\{t_{n-1}(\mathbf{u}) < 1\}} f(u_1).$$

If  $n = 2$ , this evaluates to  $\int_0^1 du_1 f(u_1) = (n - 1) \int_0^1 du_1 f(u_1) (1 - u_1)^{n-2}$ . If  $n > 2$ , as above, we obtain

$$\langle f(u_1) \rangle_{\nu(d^n \mathbf{u})} = (n - 1)! \int_0^1 du_1 f(u_1) \int_{\mathbb{R}_*^{n-2}} d^{n-2} \mathbf{v} \mathbb{1}_{\{\sum_{i=1}^{n-2} v_i < 1 - u_1\}}.$$

and evaluating the volume of the simplex results in (C.4).

Therefore, only the fourth item remains to be proven. Let  $n \geq 2$  and suppose  $f : \mathbb{R}_*^n \rightarrow \mathbb{C}$  is Borel measurable. Let us first consider the case where  $f$  is also non-negative. Applying Item 3 of Lemma C.1 we find that for any  $r > 0$ ,

$$\int_{\mathbb{R}_*^n} \tilde{\nu}_r(d^n \mathbf{u}) f(\mathbf{u}) = \int_{\mathbb{R}_*^{n-1}} d^{n-1} \mathbf{u} \mathbb{1}_{\{t_{n-1}(\mathbf{u}) < r\}} f(\mathbf{u}, r - t_{n-1}(\mathbf{u})).$$

The integrand on the right hand side is non-negative and measurable under  $dr \times (d^{n-1}u)$ . Thus by Fubini's theorem, we find that

$$\int_{\mathbb{R}_*} dr \left( \int_{\mathbb{R}_*^n} \tilde{v}_r(d^n u) f(u) \right) = \int_{\mathbb{R}_*^{n-1}} d^{n-1}u \left( \int_{\mathbb{R}_*} dr \mathbb{1}_{\{t_{n-1}(u) < r\}} f(u, r - t_{n-1}(u)) \right).$$

We change the integration variable  $r$  to  $u_n = r - t_{n-1}(u)$ , yielding

$$\int_{\mathbb{R}_*^{n-1}} d^{n-1}u \left( \int_0^\infty du_n f(u, u_n) \right).$$

By Fubini's theorem, this is equal to  $\int_{\mathbb{R}_*^n} d^n u f(u)$ . Hence we have proven the first equality in (C.5) for non-negative  $f$  and the second equality follows then immediately from Item 3 of Lemma C.1.

Let us then assume that  $f \in L^1(\mathbb{R}^n)$ . Then (C.5) holds for  $|f|$ , as well as for positive and negative parts of  $\operatorname{Re} f$  and  $\operatorname{Im} f$ . Each of these functions is bounded by  $|f|$ , and thus all of the resulting integrals are finite. Thus by taking the appropriate complex linear combination, we conclude that (C.5) holds also for  $f$ , with absolutely convergent integrals. This concludes the proof of the lemma.  $\square$

### Appendix D: Tools

The following section collects a few standard results, given here to facilitate following the argument in the main text.

**Lemma D.1.** *Let  $U$  and  $V$  be independent random variables with exponential distribution of parameter  $\lambda > 0$ . Set  $X = (U - V)/(U + V)$  and  $Y = U + V$ . Then*

- (i) *the random variable  $X$  has continuous uniform distribution in  $[-1, 1]$ .*
- (ii) *the random variable  $Y$  has Gamma distribution with parameter 2 and  $\lambda$ .*
- (iii) *the random variables  $X$  and  $Y$  are independent.*

**Proof.** The proof of Item (i) and Item (ii) are straightforward. We leave the details to the interested reader. In the sequel, we prove Item (iii). For any continuous and bounded function  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  we have

$$\begin{aligned} \mathbb{E}[F(U, V)] &= \lambda^2 \int_0^\infty \int_0^\infty du dv e^{-\lambda u - \lambda v} F(u, v) \\ (D.1) \quad &= \frac{\lambda^2}{2} \int_{\mathbb{R}^2} dy dz e^{-\lambda y} \mathbb{1}_{\{y+z>0\}} \mathbb{1}_{\{y-z>0\}} F\left(\frac{y+z}{2}, \frac{y-z}{2}\right) \\ &= \frac{\lambda^2}{2} \int_0^\infty dy e^{-\lambda y} \int_{-y}^y dz F\left(\frac{y+z}{2}, \frac{y-z}{2}\right), \end{aligned}$$

where in the second equality we apply the Change of Variable theorem for  $y = u + v$  and  $z = u - v$ . Equality (D.1) with the help of the change of variable  $x = z/y$  yields

$$\mathbb{E}[F(U, V)] = \int_0^\infty dy \underbrace{\lambda^2 y e^{-\lambda y}}_{\text{density of } Y} \int_{-1}^1 \underbrace{\frac{1}{2}}_{\text{density of } X} dx F\left(y \frac{1+x}{2}, y \frac{1-x}{2}\right).$$

In particular, for continuous and bounded functions  $f$  and  $g$  the choice  $F(u, v) := f((u - v)/(u + v))g(u + v)$ ,  $u, v \in \mathbb{R}$  yields

$$\mathbb{E}[F(U, V)] = \mathbb{E}[f(X)g(Y)] = \int_0^\infty dy \lambda^2 y e^{-\lambda y} g(y) \int_{-1}^1 dx \frac{1}{2} f(x).$$

Hence, the proof of Item (iii) is finished.  $\square$

The following lemma provides distributional properties for the sum and difference of i.i.d. exponential random variables. Since the proof is straightforward, we omit the details here.



**Lemma D.2.** Let  $X$  and  $Y$  be independent random variables having exponential distribution of parameter  $\theta > 0$ . Then the following is valid.

1.  $X - Y$  has Laplace distribution with location parameter zero and scale parameter  $\frac{1}{\theta}$ .
2.  $X + Y$  has Gamma distribution with parameters 2 and  $\theta$ .
3.  $\mathbb{E}[X - Y] = 0$  and  $\mathbb{E}[X + Y] = \frac{2}{\theta}$ .
4.  $\text{Var}[X - Y] = \text{Var}[X + Y] = \frac{2}{\theta^2}$ .
5. In general, if  $Z$  is distributed according to a Gamma distribution of parameters  $m \in \mathbb{N}$  and  $\theta > 0$ , then for any  $k \in \mathbb{N}$  we have  $\mathbb{E}[Z^k] = \theta^{-k} \frac{(m+k-1)!}{(m-1)!}$ . In addition, for the case  $m = \theta$  we have

$$\mathbb{E}[|1 - Z|^6] = \frac{15m^2 + 130m + 120}{m^5}.$$

In the next lemma, we recall the exact value for the mean and the second moment of a Poisson random variable.

**Lemma D.3 (First and second moment of Poisson distribution).** Let  $X$  be a random variable with Poisson distribution of parameter  $\lambda > 0$ . Then it follows that

$$\langle X \rangle_{\text{Po}(\lambda)} = \lambda \quad \text{and} \quad \langle X^2 \rangle_{\text{Po}(\lambda)} = \lambda^2 + \lambda.$$

## Acknowledgements

The authors are grateful to the reviewer for the thorough examination of the manuscript, which has led to a significant improvement.

## Funding

The research has been supported by the Academy of Finland, via the Matter and Materials Profi4 university profiling action, an Academy project (project No. 339228) and the Finnish centre of excellence in Randomness and Structures (project No. 346306). JL would also like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme Frontiers in kinetic theory: connecting microscopic to macroscopic (KineCon 2022) where partial work on this paper was undertaken. This work was supported by EPSRC grant no EP/R014604/1.

## CRedit authorship contribution statement

All authors have contributed equally to the paper.

## Declaration of competing interest

The authors declare that they have no conflict of interest.

## References

- [1] A. Alfonsi, J. Corbetta and B. Jourdain. Evolution of the Wasserstein distance between the marginals of two Markov processes. *Bernoulli* **24** (4A) (2018) 2461–2498. [MR3779692](#)
- [2] C. Bernardin and S. Olla. Transport properties of a chain of anharmonic oscillators with random flip of velocities. *J. Stat. Phys.* **145** (5) (2011) 1224–1255. [MR2863732](#)
- [3] J. Bierkens, G. Roberts and P. Zitt. Ergodicity of the zigzag process. *Ann. Appl. Probab.* **29** (4) (2019) 2266–2301. [MR3983339](#)
- [4] J. Bion-Nadal and D. Talay. On a Wasserstein-type distance between solutions to stochastic differential equations. *Ann. Appl. Probab.* **29** (3) (2019) 1609–1639. [MR3914552](#)
- [5] L. Bogachev and N. Ratanov. Occupation time distributions for the telegraph process. *Stochastic Process. Appl.* **121** (8) (2011) 1816–1844. [MR2811025](#)
- [6] V. Bogachev, M. Röckner and S. Shaposhnikov. Distances between transition probabilities of diffusions and applications to nonlinear Fokker-Planck-Kolmogorov equations. *J. Funct. Anal.* **271** (5) (2016) 1262–1300. [MR3522009](#)
- [7] S. Boucheron, G. Lugosi and P. Massart. *Concentration Inequalities*. Oxford University Press, Oxford, 2013. [MR3185193](#)

- [8] P. Brémaud. *Point Process Calculus in Time and Space*. Springer, Cham, 2020. [MR4212534](#)
- [9] P. Briand, C. Geiss and S. Geiss. Donsker-type theorem for BSDEs: Rate of convergence. *Bernoulli* **27** (2) (2021) 899–929. [MR4255220](#)
- [10] F. Cinque and E. Orsingher. On the exact distributions of the maximum of the asymmetric telegraph process. *Stochastic Process. Appl.* **142** (1–b) (2021) 601–633. [MR4324351](#)
- [11] L. Coutin and L. Decreusefond. Donsker’s theorem in Wasserstein-1 distance. *Electron. Commun. Probab.* **25** (27) (2020) 1–13. [MR4089734](#)
- [12] A. De Gregorio and F. Iafate. Telegraph random evolutions on a circle. *Stochastic Process. Appl.* **141** (2021) 79–108. [MR4293769](#)
- [13] V. de la Peña and G. Decoupling *Probability and Its Applications*. Springer Verlag, New York, 1999. [MR1666908](#)
- [14] A. Di Crescenzo, A. Iuliano, B. Martinucci and S. Zacks. Generalized telegraph process with random jumps. *J. Appl. Probab.* **50** (2) (2013) 450–463. [MR3102492](#)
- [15] A. Di Crescenzo, B. Martinucci and S. Zacks. Telegraph process with elastic boundary at the origin. *Methodol. Comput. Appl. Probab.* **20** (1) (2018) 333–352. [MR3760349](#)
- [16] G. Di Masi, Y. Kabanov and W. Runggaldier. Mean-variance hedging of options on stocks with Markov volatilities. *Theory Probab. Appl.* **39** (1) (1994) 172–182. [MR1348196](#)
- [17] A. Eberle and R. Zimmer. Sticky couplings of multidimensional diffusions with different drifts. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** (4) (2019) 2370–2394. [MR4029157](#)
- [18] S. Ethier and T. Kurtz. *Markov Processes*. John Wiley & Sons Inc., New York, 1986. [MR0838085](#)
- [19] S. Foong and S. Kanno. Properties of the telegrapher’s random process with or without a trap. *Stochastic Process. Appl.* **53** (1) (1994) 147–173. [MR1290711](#)
- [20] A. Ghosh, R. Rastegar and A. Roitershtein. On a directionally reinforced random walk. *Proc. Amer. Math. Soc.* **142** (9) (2014) 3269–3283. [MR3223382](#)
- [21] S. Goldstein. On diffusion by discontinuous movements and on the telegraph equation. *Quart. J. Mech. Appl. Math.* **4** (2) (1951) 129–156. [MR0047963](#)
- [22] S. Iacus and N. Yoshida. Estimation for the discretely observed telegraph process. *Theory Probab. Math. Statist.* **78** (2009) 37–47. [MR2446847](#)
- [23] A. Janssen. The distance between the Kac process and the Wiener process with applications to generalized telegraph equations. *J. Theoret. Probab.* **3** (2) (1990) 349–360. [MR1046338](#)
- [24] M. Kac. A stochastic model related to the telegrapher’s equation. *Rocky Mountain J. Math.* **4** (3) (1974) 497–510. [MR0510166](#)
- [25] A. Klenke. *Probability Theory: A Comprehensive Course*, 2nd edition. Springer Verlag, London, 2014. [MR3112259](#)
- [26] A. Kolesnik. The equations of Markovian random evolution on the line. *J. Appl. Probab.* **35** (1) (1998) 27–35. [MR1622442](#)
- [27] A. Kolesnik. The explicit probability distribution of the sum of two telegraph processes. *Stoch. Dyn.* **15** (2) (2015) 1550013. [MR3332273](#)
- [28] A. Kolesnik. Linear combinations of the telegraph random processes driven by partial differential equations. *Stoch. Dyn.* **18** (4) (2018) 1850020. [MR3842248](#)
- [29] A. Kolesnik and N. Ratanov. *Telegraph Processes and Option Pricing*. Springer, Berlin, Heidelberg, 2013. [MR3115087](#)
- [30] A. Kolesnik and A. Turbin. The equation of symmetric Markovian random evolution in a plane. *Stochastic Process. Appl.* **75** (1) (1998) 67–87. [MR1629022](#)
- [31] Komlós, P. Major and G. Tusnády. An approximation of partial sums of independent RV’s and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** (1–2) (1975) 111–131. [MR0375412](#)
- [32] J. Komlós, P. Major and G. Tusnády. An approximation of partial sums of independent RV’s, and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34** (1) (1976) 33–58. [MR0402883](#)
- [33] K. Koskinen and J. Lukkarinen. Estimation of local microcanonical averages in two lattice mean-field models using coupling techniques. *J. Stat. Phys.* **180** (1) (2020) 1206–1251. [MR4131030](#)
- [34] J. Lukkarinen. Thermalization in harmonic particle chains with velocity flips. *J. Stat. Phys.* **155** (6) (2014) 1143–1177. [MR3207733](#)
- [35] J. Lukkarinen. Multi-state condensation in Berlin–Kac spherical models. *Comm. Math. Phys.* **373** (1) (2020) 389–433. [MR4050098](#)
- [36] J. Lukkarinen, M. Marozzi and A. Nota. Harmonic chain with velocity flips: Thermalization and kinetic theory. *J. Stat. Phys.* **165** (5) (2016) 809–844. [MR3572498](#)
- [37] B. Martinucci and A. Meoli. Certain functionals of squared telegraph processes. *Stoch. Dyn.* **20** (1) (2020) 2050005. [MR4066800](#)
- [38] C. Mazza and A. Rullière. A link between wave governed random motions and ruin processes. *Insurance Math. Econom.* **35** (2) (2004) 205–222. [MR2095886](#)
- [39] T. Mikosch. *Non-life Insurance Mathematics*, 2nd edition. Springer, Berlin, Heidelberg, 2009. [MR2503328](#)
- [40] E. Orsingher. Probability law, flow function, maximum distribution of wave-governed random motions and their connections with Kirchoff’s laws. *Stochastic Process. Appl.* **34** (1) (1990) 49–66. [MR1039562](#)
- [41] V. Panaretos and Y. Zemel. An invitation to statistics in Wasserstein space. *SpringerBriefs in Probabil. Math. Statist.* (2020). [MR4350694](#)
- [42] M. Pinsky. *Lectures on Random Evolution*. World Scientific Publishing Co. Inc., River Edge, New Jersey, 1991. [MR1143780](#)
- [43] H. Robbins. A remark on Stirling’s formula. *Amer. Math. Monthly* **62** (1) (1955) 26–29. [MR0069328](#)
- [44] E. Saksman and C. Webb. The Riemann zeta function and Gaussian multiplicative chaos: Statistics on the critical line. *Ann. Probab.* **48** (6) (2020) 2680–2754. [MR4164452](#)
- [45] M. Simon. Hydrodynamic limit for the velocity-flip model. *Stochastic Process. Appl.* **123** (10) (2013) 3623–3662. [MR3084154](#)
- [46] W. Stadje and S. Zacks. Telegraph processes with random velocities. *J. Appl. Probab.* **41** (1) (2004) 665–678. [MR2074815](#)
- [47] C. Villani. *Optimal Transport*. Springer, Berlin, Heidelberg, 2009. [MR2459454](#)