



**HAL**  
open science

# SCALING LIMITS OF SOLITONS IN THE BOX-BALL SYSTEM

Stefano Olla, Makiko Sasada, Hayate Suda

► **To cite this version:**

Stefano Olla, Makiko Sasada, Hayate Suda. SCALING LIMITS OF SOLITONS IN THE BOX-BALL SYSTEM. 2024. hal-04803830

**HAL Id: hal-04803830**

**<https://hal.science/hal-04803830v1>**

Preprint submitted on 26 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SCALING LIMITS OF SOLITONS IN THE BOX-BALL SYSTEM

STEFANO OLLA<sup>1</sup>, MAKIKO SASADA<sup>2</sup>, AND HAYATE SUDA<sup>3</sup>

ABSTRACT. We study the space-time scaling limits of solitons in the box-ball system with random initial distribution. In particular, we show that any recentered tagged soliton converges to a Brownian motion in the diffusive space-time scale, and also prove the large deviation principle for the tagged soliton under certain shift-ergodic invariant distributions, including Bernoulli product measures and two-sided Markov distributions. Furthermore, in the diffusive space-time scaling, we show that two tagged solitons converge to the same Brownian motion even if they are macroscopically far apart.

## 1. INTRODUCTION

An integrable many-body system is a deterministic dynamical system consisting of an infinite type of quasi-local conserved quantities that behave like particles interacting with each other. These quasi-local conserved quantities are called quasi-particles. Solitary waves (solitons) in solitonic systems are examples of quasi-particles. Recently, integrable many-body systems have attracted much attention from the viewpoint of non-equilibrium statistical mechanics, and in particular, generalized hydrodynamics, which describes the macroscopic behavior of quasi-particles, see the reviews [D, Sp] and references therein. In the Euler space-time scale, it is expected that the hydrodynamics is described by the following generalized hydrodynamic equation (GHD equation) for  $\mathbf{y}(u, t) = (y_a(u, t))_a$  of the universal form, regardless of models :

$$\partial_t y_a(u, t) + \partial_u (v_a^{\text{eff}}(\mathbf{y}(u, t)) y_a(u, t)) = 0,$$

where  $y_a(u, t)$  is the macroscopic density of quasi-particles of type  $a$  at macroscopic coordinate  $(u, t)$ ,  $u \in \mathbb{R}, t \geq 0$ , and  $v_a^{\text{eff}}$  is called the effective velocity of quasi-particles of type  $a$ . The specific form of the effective velocity depends on the scattering rule between quasi-particles, and this is where the differences among models arise. Although such studies have been rapidly developed in the physics literature, and the GHD theory is expected to

---

<sup>1</sup> CEREMADE, UNIVERSITÉ PARIS DAUPHINE - PSL RESEARCH UNIVERSITY AND INSTITUT UNIVERSITAIRE DE FRANCE, AND GSSI, L'AQUILA

<sup>2</sup>GRADUATE SCHOOL OF MATHEMATICAL SCIENCES, UNIVERSITY OF TOKYO

<sup>3</sup>DEPARTMENT OF PHYSICS, INSTITUTE OF SCIENCE TOKYO

2020 *Mathematics Subject Classification*. 37B15(primary) ; 82C22, 82C23 (secondary).

be applicable for a wide class of integrable systems including classical and quantum gases, chains and field theory models, very few rigorous results are obtained till now, which are for the hard rods dynamics and its generalization [BDS, FFGS], and the box-ball system (BBS) [CS].

While studies at the Euler scale have been progressing, there was no clear physical prediction on the behavior in a longer time scale. Therefore, it is important to obtain mathematically rigorous results on specific models in order to derive the universality for integrable many-body systems in the diffusive scale. Recently, by [FO], the fluctuations for hard-rods in diffusive scale has been proved rigorously. The difference from diffusive fluctuations in chaotic systems is the strong correlations between quasi-particles of the same type, i.e., quasi-particles of the same type starting at macroscopic distance converge to the same Brownian motion. However, the scattering rule in the hard-rods does not depend on the velocity of quasi-particles. This is a different feature from general integrable models, and no results have been known for the case where the scattering rule depends on velocities of the quasi-particles.

In this paper, we consider the BBS, which is a solitonic system with a scattering rule depending on velocities of quasi-particles (solitons). We rigorously show that any tagged soliton converges to a Brownian motion in diffusive space-time scale, and also prove the large deviation principle for the tagged soliton. This is the first mathematical result for the central limit theorem and the large deviation principle for quasi-particles in integrable systems with scattering rules depending on velocities of quasi-particles, unlike the hard-rods. Furthermore, we rigorously prove that solitons of the same type converge to the same Brownian motion, i.e., strong correlations between quasi-particles as observed in the hard-rods. In order to roughly describe the results, we first introduce the BBS below.

The BBS is a one-dimensional cellular automaton introduced by [TS], whose integrable structure has been extensively studied in the past, see the review [IKT] for details. The BBS exhibits solitonic behavior and is understood as a discrete counterpart of the KdV equation, which is a central example of an integrable system having solitary wave solutions. The configuration space is  $\{0, 1\}^{\mathbb{Z}}$ , where for  $\eta \in \{0, 1\}^{\mathbb{Z}}$  and  $x \in \mathbb{Z}$ ,  $\eta(x) = 1$  means that there is a ball at  $x$ , and  $\eta(x) = 0$  means that  $x$  is empty. When the total number of 1s in  $\eta \in \{0, 1\}^{\mathbb{Z}}$  is finite, the one-step time evolution  $\eta \mapsto T\eta \in \{0, 1\}^{\mathbb{Z}}$  is described by the following rules :

- An empty carrier enters the system from the left end (i.e.  $-\infty$ ) and moves to the right end (i.e.  $\infty$ );
- If there is a ball at site  $x$ , then the carrier picks up the ball;
- If the site  $x$  is empty and the carrier is not empty, then the carrier puts down a ball;
- Otherwise, the carrier just passes through.

Clearly, the total number of 1s are conserved. If we denote by  $W(x)$  the number of balls on the carrier at  $x$ , then  $W(\cdot)$  satisfies  $W(x) = 0$  for any  $|x| > L$  with sufficiently large  $L > 0$ , and

$$W(x) - W(x - 1) = \begin{cases} 1 & \text{if } \eta(x) = 1, \\ -1 & \text{if } \eta(x) = 0 \text{ and } W(x - 1) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (1.1)$$

In addition,  $T\eta$  can be represented by using  $W(\cdot)$  as

$$T\eta(x) = \eta(x) - W(x) + W(x - 1). \quad (1.2)$$

Figure 1 shows an example how  $T\eta$  can be obtained from  $\eta$ .

$\eta$ :	...	1	1	0	0	0	1	1	1	0	0	1	1	0	0	1	0	1	1	0	0	0...
$W$ :	0	1	2	1	0	0	1	2	3	2	1	2	3	2	1	2	1	2	3	2	1	0...
$T\eta$ :	...	0	0	1	1	0	0	0	0	1	1	0	0	1	1	0	1	0	0	1	1	1...

FIGURE 1.  $W$  and  $T\eta$  obtained from  $\eta = \dots 1100011100110010110000\dots$ , where  $\dots$  represents the consecutive 0s.

It is known that the above rule can be extended to  $\eta \in \Omega \subset \{0, 1\}^{\mathbb{Z}}$ , where

$$\Omega := \left\{ \eta \in \{0, 1\}^{\mathbb{Z}} ; \exists \lim_{x \rightarrow \infty} \frac{1}{x} \sum_{y=1}^x \eta(y) < \frac{1}{2}, \exists \lim_{x \rightarrow \infty} \frac{1}{x} \sum_{y=1}^x \eta(-y) < \frac{1}{2} \right\},$$

see Section 2 for details. We note that by [CKST], more detailed results are obtained for the configuration space in which the dynamics of the BBS can be defined via the Pitman transform.

In recent years, the BBS started from random initial configurations, called the randomized BBS, has been studied in terms of its statistical aspects; characterizations of classes of invariant measures for the randomized BBS [CS2, CKST, FG], limit theorems under invariant measures [CKST, FNRW, KL, KLO18, LLP, LLPS, S]. Also, the randomized BBS has been studied from the viewpoint of hydrodynamics for integrable systems [CS, KMP, KMP2, KMP3]. Currently, only the BBS and hard-rods are known to be mathematically tractable models for deriving hydrodynamics from integrable systems, and thus the BBS is recognized as an important model in statistical mechanics.

In this article, we consider the BBS on the state space  $\{0, 1\}^{\mathbb{Z}}$  under invariant measures, and derive the scaling limits of the tagged soliton. To give an overview of our results, we introduce the law of large numbers for the tagged soliton proved in [FNRW]. As mentioned at the beginning, the BBS is a soliton system and there are infinite types of solitons in the BBS labeled by positive integers  $k \in \mathbb{N}$  respectively, called  $k$ -soliton. A  $k$ -soliton consists of  $k$  1s and  $k$  0s and is identified by a certain algorithm, see Section 2.2 for details. If there are only  $k$ -solitons in the configuration, then they move to

right with velocity  $k$  at each time evolution. When solitons of different sizes exist and a  $k$ -soliton is to the left of an  $\ell$ -soliton with  $\ell < k$ , such solitons will meet at some time and *phase shift* will occur between them. In particular, during the interaction, the smaller soliton is stranded and cannot move. For example, in the following figure showing the time evolution of BBS, red **1**s and **0**s constitute a 2-soliton and blue **1** and **0** constitute a 1-soliton, and while the 2-soliton tries to overtake the 1-soliton, the 1-soliton cannot move from its position.

$$\begin{array}{rcccccccccccccccc}
\eta & \dots & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\
T\eta & \dots & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\
T^2\eta & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\
T^3\eta & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots \\
T^4\eta & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots
\end{array}$$

After the interaction, 1-soliton is shifted backward 2 sites and 2-soliton is shifted forward 2 sites from where they should have come without interaction. This is the phase shift. Thus, given a random initial configuration, even if the time evolution rule of the BBS is deterministic, the position of the tagged  $k$ -soliton at time  $n$  is randomized by the random presence of other solitons of different sizes, which is a random environment for the tagged soliton. In [FNRW], the authors show that the tagged soliton satisfies the law of large numbers (LLN) when the initial distribution  $\nu$  is invariant for  $T$  and shift-ergodic. Bernoulli product measures of uniform density  $\nu(\eta(x) = 1) = \nu(\eta(0) = 1) < 1/2$ ,  $x \in \mathbb{Z}$ , and two-sided space-homogeneous Markov distributions supported on  $\Omega$  are important examples of  $\nu$  satisfying the assumptions below.

**Theorem** ([FNRW]). *Let  $X_k(n)$  be the position of the leftmost component of a tagged  $k$ -soliton at time  $n$  and  $\nu$  be a probability measure on  $\{0, 1\}^{\mathbb{Z}}$  satisfying the following.*

- $\nu(\Omega) = 1$ .
- $\nu$  is an invariant measure of the BBS, namely  $T\nu = \nu$ .
- $\nu$  is a shift-ergodic measure.

Assume that for some  $k \in \mathbb{N}$ ,  $k$ -solitons exist with positive probability under  $\nu$ . Then, there exists some  $v_k^{\text{eff}} = v_k^{\text{eff}}(\nu) > 0$  such that

$$\lim_{n \rightarrow \infty} \frac{X_k(n)}{n} = v_k^{\text{eff}} \quad \nu\text{-a.s.}$$

The constant  $v_k^{\text{eff}}$  is called the effective velocity for  $k$ -solitons. A characterization formula for  $v_k^{\text{eff}}$ ,  $k \in \mathbb{N}$  has been obtained by [FNRW, (1.12)]. We will present an alternative formula for  $v_k^{\text{eff}}$ , see Proposition 4.4 for details. Also, we give a different proof for [FNRW, (1.12)], see Remark 4.6.

Our main results are the central limit theorem (CLT) and the large deviation principle (LDP) corresponding to the LLN. As a by-product, we also show the LLN in  $\mathbb{L}^p$ ,  $p \geq 1$ .

**Claim 1** (Limit theorems for a tagged soliton). *Assume that the initial distribution  $\nu$  is a space-homogeneous Bernoulli product measure or two-sided Markov distribution supported on  $\Omega$ . Then, for any  $k \in \mathbb{N}$ , we have the following.*

- (1) *Under the diffusive space-time scaling, the step-interpolation of the discrete-time process  $n \mapsto X_k(n) - v_k^{\text{eff}} n$  converges weakly to a centered Brownian motion  $B_k(t)$  with variance  $D_k = D_k(\nu) > 0$ .*
- (2) *The sequence  $(X_k(n)/n)_{n \in \mathbb{Z}_{\geq 0}}$  satisfies the LDP with a smooth convex rate function.*
- (3) *For any  $p \geq 1$ , we have*

$$\lim_{n \rightarrow \infty} \mathbb{E}_\nu \left[ \left| \frac{X_k(n)}{n} - v_k^{\text{eff}} \right|^p \right] = 0.$$

**Remark 1.1.** *For the tagged ball (with a certain rule to identify the positions of distinguished balls), instead of the tagged soliton, the LLN, the CLT and the LDP are shown in [CKST] under the Bernoulli product measures. However, in the case of BBS, balls are local quantities, whereas solitons are quasi-local quantities, so a more sophisticated mathematical treatment is needed to show scaling limits for solitons.*

Claim 1 will be stated precisely as Theorem 2.1, and Theorem 2.1 will be proved as a corollary of Theorems 4.5 and 4.10, where we consider more general initial distributions. Actually, when the initial distribution  $\nu$  belongs to a family of invariant measures introduced by [FG] and satisfies a good asymptotic property for existence probability of large solitons, we will show the CLT and the LDP for solitons with size  $k \geq K$ , where  $K = K(\nu)$  is a positive integer determined by  $\nu$ . We note that space-homogeneous Bernoulli product measures and two-sided Markov distributions supported on  $\Omega$  belong to this class, and  $K(\nu) = 1$ . In this paper a family of invariant measures introduced by [FG] is called  $q$ -statistics, see Section 4.1 for the precise definition.

Furthermore, we will show that two  $k$ -solitons are strongly correlated in the diffusive space-time scaling even when they are far apart at the macroscopic level.

**Claim 2** (Strong correlations between  $k$ -solitons). *Assume that the initial distribution  $\nu$  is a  $q$ -statistics with a certain second moment condition. Then, even if two  $k$ -solitons are far apart after taking the space scaling, those fluctuations converge to the same Brownian motion obtained in Claim 1 (1) in the diffusive scaling.*

In Section 4.2, we will restate Claim 2 as Theorem 4.8 with precise assumptions.

From the above results, it is expected that the macroscopic fluctuations of the density of  $k$ -solitons at diffusive time scale  $t$  can be obtained by shifting the initial fluctuation field by  $B_k(t)$  obtained in Claim 1 (1). In other

words, the macroscopic fluctuation field  $\mathcal{Y}_k(u, t)$  should follow the following stochastic partial differential equations for each  $k \in \mathbb{N}$  :

$$d\mathcal{Y}_k(u, t) = \frac{D_k}{2} \Delta_u \mathcal{Y}_k(u, t) dt + \partial_u \mathcal{Y}_k(u, t) dB_k(t), \quad (1.3)$$

It is noteworthy that the noise driving (1.3) does not depend on the spatial variables, which is in contrast to typical diffusive fluctuations for chaotic systems where an additive space-time white noise drives the macroscopic equation. We expect that (1.3) is a universal equation in completely integrable many-body systems, and was recently derived rigorously for the first time from hard-rods dynamics [FO]. We also note that diffusive corrections to Euler scale hydrodynamics for integrable many-body systems have been studied in physics literature [DBD, DBD2, DDMP, Sp]. It would be an interesting problem to derive (1.3) rigorously from the BBS, to prove that  $\{B_k(t); t \geq 0, k \in \mathbb{N}\}$  is a centered Gaussian field, and to specify the correlation between  $B_k(t)$  and  $B_\ell(t)$  with  $k \neq \ell$ .

Our approach is based on two different linearization methods for the BBS. One is called the seat number configuration, which is recently introduced by [MSSS, S], and it is a generalization of the slot configuration developed by [FNRW]. We note that the slot configuration has played an important role in the study of the dynamical aspects of BBS [CS, FG, FNRW]. The other is the  $k$ -skip map, which is a generalization of the 10-elimination introduced by [MIT] to solve the initial value problem for the BBS with periodic boundary condition. In [S], the  $k$ -skip map is considered in terms of the seat number configuration, and the relation between  $q$ -statistics and the  $k$ -skip map is studied. The results and computations in [S] are essential to obtain Lemma 5.5 and a decomposition formula (5.19) for the position of the tagged soliton, see Section 5 for details. By using Lemma 5.5, (5.19) and the property of  $q$ -statistics, the CLT/LDP for the tagged soliton can be reduced to the CLT/LDP for  $M(n)$ , respectively, where  $M(n)$  is the number of times that the tagged soliton interacts with solitons larger than itself until time  $n$ . By the same idea, we can show the LLN for the tagged soliton in  $\mathbb{L}^p$  by that for  $M(n)$ . Furthermore, Lemma 5.5 and (5.19) are also useful for showing the strong correlations between solitons of the same size. To the best of our knowledge, this is the first time that the 10-elimination is applied to the dynamical problem of the randomized BBS. A version of the 10-elimination was used in [LLP], but they considered static problems. We note that our proof strategy can be applied even if the initial distribution  $\nu$  is not a  $q$ -statistics as long as  $\nu$  has some nice property, see Remark 4.7 for details.

The rest of the paper is organized as follows. In Section 2 we briefly recall the basics of the BBS on  $\mathbb{Z}$  and introduce some terminologies used in this paper, then we describe our main results Theorem 2.1, where the initial distribution  $\nu$  is a Bernoulli product measure or two-sided Markov distribution. In Section 3, we introduce some notions that are essential for the proof as well as for describing the results under more general invariant

distributions. In Section 4 we present our main results, Theorems 4.2, 4.5, 4.8, 4.10 and 4.14, where the initial distribution is generalized to  $\mathbf{q}$ -statistics. In Section 5, we introduce the notion of the  $k$ -skip map,  $k \in \mathbb{N}$ , and we prepare some lemmas for the proofs of main results. In the subsequent sections, Section 6, 7, 8, 9 and 10, we give the proofs of the main results, Theorems 4.2, 4.5, 4.8, 4.10 and 4.14, respectively. Finally, in Section 11, we show Theorem 2.1 by using Theorems 4.2, 4.5, 4.8 and 4.10.

## 2. BOX-BALL SYSTEM

**2.1. Dynamics of the Box-Ball system.** First we recall the definition of the one-step time evolution  $\eta \mapsto T\eta$  when the total number of 1s in  $\eta \in \Omega$  is finite, presented in Introduction. A site  $x \in \mathbb{Z}$  is called a record if  $\eta(x) = T\eta(x) = 0$ . Clearly,  $T$  can be considered as the flipping 1s (resp. 0s) to 0s (resp. 1s) except for records, i.e., we can write  $T\eta$  as

$$T\eta(x) = \begin{cases} 1 - \eta(x) & \text{if } x \text{ is not a record,} \\ 0 & \text{if } x \text{ is a record.} \end{cases} \quad (2.1)$$

We note that records can be characterized as follows :

$$x \text{ is a record if and only if } \max_{z \leq x} \sum_{y=z}^x (2\eta(y) - 1) \leq -1. \quad (2.2)$$

Now we define the BBS on  $\Omega$ . The one-time step evolution  $T : \Omega \rightarrow \Omega$  can be also defined via the notion of records as follows. For  $\eta \in \Omega$ , we define a record in  $\eta$  as a site  $x \in \mathbb{Z}$  that satisfies (2.2). We note that there are infinitely many records in  $\eta$  because the asymptotic ball density as  $x \rightarrow \pm\infty$  is strictly smaller than  $1/2$ . Then, we define  $T : \Omega \rightarrow \Omega$  by (2.1).

We number records in  $\eta$  from left to right as follows. For any  $\eta \in \Omega$ , we define

$$s_\infty(\eta, 0) := \max \left\{ x \leq 0 ; \max_{z \leq x} \sum_{y=z}^x (2\eta(y) - 1) \leq -1 \right\},$$

and then we recursively define  $s_\infty(\eta, i)$  as

$$s_\infty(\eta, i) := \min \left\{ x > s_\infty(\eta, i-1) ; \max_{z \leq x} \sum_{y=z}^x (2\eta(y) - 1) \leq -1 \right\},$$

$$s_\infty(\eta, -i) := \max \left\{ x < s_\infty(\eta, -i+1) ; \max_{z \leq x} \sum_{y=z}^x (2\eta(y) - 1) \leq -1 \right\},$$

for any  $i \in \mathbb{N}$ . Notice that  $s_\infty(\eta, i) \in \mathbb{Z}$  for any  $i \in \mathbb{Z}$  because  $\eta \in \Omega$ .

We note that the dynamics of the BBS on  $\Omega$  can also be described via the carrier process  $W(\eta, x) : \mathbb{Z} \rightarrow \mathbb{Z}_{\geq 0}$  recursively defined by (1.1) and  $W(\eta, s_\infty(\eta, i)) := 0$  for any  $i \in \mathbb{Z}$ . Then, by using  $W$ ,  $T : \Omega \rightarrow \Omega$  is written as (1.2).



**2.2. Solitons in the Box-Ball configuration.** In this subsection, we explain how we can identify solitons in  $\Omega$ .

For given  $\eta \in \Omega$ , we consider the following decomposition :

$$\eta = \cup_{i \in \mathbb{Z}} \mathbf{e}^{(i)}, \quad \mathbf{e}^{(i)} := (\eta(x) ; s_{\infty}(i) \leq x < s_{\infty}(i+1)). \quad (2.3)$$

The sequence  $\mathbf{e}^{(i)}$  is called the  $i$ -th excursion of  $\eta$ . Then, for each  $\mathbf{e}^{(i)} \setminus \{s_{\infty}(i)\}$ , we can find solitons via the Takahashi-Satsuma algorithm as follows :

- Select the leftmost run of consecutive 0s or 1s such that
  - the run is consecutive 0s with finite length, or consecutive 1s.
  - the length of the subsequent run is at least as long as the length of it.
- Let  $k$  be the length of the selected run. Group the  $k$  element of the selected run and the first  $k$  elements of the subsequent run. The grouped  $2k$  elements are identified as a soliton with size  $k$ , or  $k$ -soliton.
- Remove the identified  $k$ -soliton, and repeat the above procedure until all 1s are removed.

From the definition of records, if  $\mathbf{e}^{(i)} \setminus \{s_{\infty}(i)\}$  is not empty, then all 1s and 0s in  $\mathbf{e}^{(i)} \setminus \{s_{\infty}(i)\}$  are grouped and become components of solitons. We note that from the TS-algorithm, we see that solitons of the same size do not overlap, and a larger soliton can contain a smaller soliton inside, but not vice versa. An example of applying the above algorithm to  $\eta = \dots 01100011100110010110000 \dots$  is shown in Figure 2. In this example, only two excursions have solitons, and there are one 3-soliton, three 2-solitons and one 1-soliton in total.

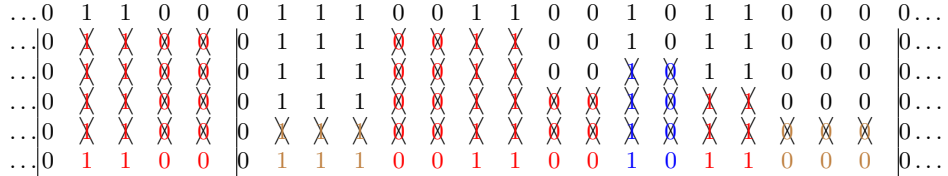


FIGURE 2. Identifying solitons in  $\eta$  by the TS Algorithm. 1-soliton is colored by blue, 2-solitons are colored by red, and 3-soliton is colored by brown.

It was discovered by [TS] that total number of  $k$ -solitons is conserved in time for each  $k \in \mathbb{N}$ , i.e., for any  $\eta \in \Omega$  with the condition  $\sum_{x \in \mathbb{Z}} \eta(x) < \infty$ , we have

$$|\{k\text{-solitons in } T\eta\}| = |\{k\text{-solitons in } \eta\}|.$$

Now we observe the behaviors of solitons in time evolution. If there are only solitons of the same size, they move to the right by  $k$  :

$$\begin{aligned} \eta &= \dots \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \dots \\ T\eta &= \dots \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ \dots \end{aligned}$$

If there are two solitons of different sizes and the larger soliton is to the left of the smaller soliton, an interaction will occur between them at some time. During the interaction, the solitons overlap each other and the shapes of the solitons are collapsed, but they return to their original shapes after the interaction is over. Furthermore, the larger soliton accelerates when overtaking the smaller soliton, while the smaller soliton stays where it is. For example, see Figure 3.

$$\begin{aligned} \eta &= \dots \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \dots \\ T\eta &= \dots \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \dots \\ T^2\eta &= \dots \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ \dots \\ T^3\eta &= \dots \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ \dots \end{aligned}$$

FIGURE 3. The 3-soliton accelerates from time 2 to 3. On the other hand, the 1-soliton does not move from time 1 to 3.

In the rest of this subsection, we introduce some notions for later use.

2.2.1. *Set of all  $k$ -solitons.* We denote by  $\Gamma_k = \Gamma_k(\eta)$  the set of all  $k$ -solitons in  $\eta$ . For any  $\gamma \in \Gamma_k$ , we define  $X(\gamma) := (\inf \gamma) - 1$ , and call  $X(\gamma)$  the position of  $\gamma$ . To obtain our key results, which are Lemma 5.5 and the decomposition formula (5.19), it is important to define  $X(\gamma)$  as  $(\inf \gamma) - 1$  instead of  $\inf \gamma$ .

2.2.2. *Natural numbering for solitons.* In this paper, since we focus on a single soliton and consider its scaling limit, it is necessary to label each soliton. For the BBS on  $\mathbb{Z}$ , it is convenient to use a record as a reference site for the detailed analysis. In particular, we are interested in the case where the origin is a record, i.e.,  $s_\infty(\eta, 0) = 0$ . However, for later use, considering the case where  $s_\infty(\eta, 0) = 0$  is not 0, we order solitons as follows. For each  $k \in \mathbb{N}$ , a  $k$ -soliton to the left of  $s_\infty(0)$  is the 0th soliton, and  $k$ -solitons are numbered in order from left to right from there. More precisely, for any  $k \in \mathbb{N}$ , we denote by  $\gamma_k^0$  the  $k$ -soliton such that

$$X(\gamma_k^0) = \max_{\gamma \in \Gamma_k, X(\gamma) < s_\infty(0)} X(\gamma).$$

Then, we recursively define  $\gamma_k^i$  as the  $k$ -soliton such that

$$X(\gamma_k^i) = \min_{\gamma \in \Gamma_k, X(\gamma) > X(\gamma_k^{i-1})} X(\gamma)$$

for  $i \geq 1$ , and

$$X(\gamma_k^i) = \max_{\gamma \in \Gamma_k, X(\gamma) < X(\gamma_k^{i+1})} X(\gamma)$$

for  $i \leq -1$ . We call  $\gamma_k^i$  the  $i$ -th  $k$ -soliton.

In this paper, the above numbering is called the natural numbering for solitons.

**2.2.3. Position of a soliton at time  $n$ .** We can track each soliton in time evolution. First, for any  $\gamma \in \Gamma_k$  we define heads  $H(\gamma)$  and tails  $T(\gamma)$  as

$$\begin{aligned} H(\gamma) &:= \{x \in \gamma; \eta(x) = 1\} = \{H_1(\gamma) < \dots < H_k(\gamma)\}, \\ T(\gamma) &:= \{x \in \gamma; \eta(x) = 0\} = \{T_1(\gamma) < \dots < T_k(\gamma)\}. \end{aligned}$$

From [FNRW, Proposition 1.3], for any  $\gamma \in \Gamma_k(\eta)$ , there exists unique  $\gamma' \in \Gamma_k(T\eta)$  such that  $T(\gamma) = H(\gamma')$ , and we write  $\gamma'$  as  $\gamma(1)$ , i.e.,  $X(\gamma(1))$  is the position of  $\gamma$  at time 1. By repeating the above, for any  $n \in \mathbb{N}$ , we can find  $\gamma(n) \in \Gamma_k(T^n\eta)$  such that  $T(\gamma(n-1)) = H(\gamma(n))$ , and we call  $X(\gamma(n))$  the position of  $\gamma$  at time  $n$ . We note that since there may be a  $k$ -soliton passing through the origin in time evolution,  $\gamma_k^i(n)$  is not always the  $i$ -th  $k$ -soliton in  $T^n\eta$ .

In the following, the position at time  $n$  of  $i$ -th  $k$ -soliton is denoted by  $X_k^i(n)$ .

**2.3. Scaling limits for solitons.** Now we state our main results on the fluctuations of  $k$ -solitons when the initial distribution  $\nu$  is given by a space-homogeneous Bernoulli product measure or two-sided Markov distribution supported on  $\Omega$ .

**Theorem 2.1.** *Assume that the initial distribution  $\nu$  is a space-homogeneous Bernoulli product measure or two-sided Markov distribution supported on  $\Omega$ . Then, for any  $k \in \mathbb{N}$  and  $i \in \mathbb{Z}$ , we have the following.*

(1) *For any  $\mathbf{T} > 0$ , the following step-interpolation process,*

$$t \mapsto \frac{1}{n} X_k^i(\lfloor n^2 t \rfloor) - n t v_k^{\text{eff}}(\nu), \quad (2.4)$$

*converges weakly in  $D[0, \mathbf{T}]$  to a Brownian motion with variance  $D_k = D_k(\nu)$ , defined in (4.15) below.*

(2) *The sequence  $(X_k^i(n)/n)_{n \in \mathbb{Z}_{\geq 0}}$  satisfies the LDP with a smooth convex rate function defined in (4.17) below.*

(3) *For any  $p \geq 1$ , we have*

$$\lim_{n \rightarrow \infty} \mathbb{E}_\nu \left[ \left| \frac{X_k(n)}{n} - v_k^{\text{eff}}(\nu) \right|^p \right] = 0.$$

**Remark 2.2.** *Formally, the variance  $D_k(\nu)$  can be obtained via the Green-Kubo type formula. We consider the following decomposition of  $X_k^i(n)$  :*

$$X_k^i(n) - X_k^i(0) - \mathbb{E}_\nu [X_k^i(n) - X_k^i(0)] = \sum_{m=0}^{n-1} \mathcal{J}_k^i(m),$$

where  $\mathcal{J}_k^i(m)$  is the centered current at time  $m$ . Then,  $D_k(\nu)$  is given by

$$D_k(\nu) = \mathbb{E}_\nu \left[ (\mathcal{J}_k^i(0))^2 \right] + 2 \sum_{n=1}^{\infty} \mathbb{E}_\nu \left[ \mathcal{J}_k^i(n) \mathcal{J}_k^i(0) \right].$$

Later in this paper we will show that when numbering solitons by a rule different from natural numbering (truncated numbering, see Section 3.1), the centered current of the  $i$ -th  $k$ -soliton  $\mathcal{J}_k^{(i)}(m)$  with the truncated numbering can be decomposed as  $\mathcal{J}_k^{(i)}(m) = \mathcal{J}_{k,1}^{(i)}(m) + \mathcal{J}_{k,2}^{(i)}(m)$ , where  $\mathcal{J}_{k,1}^{(i)}$  (resp.  $\mathcal{J}_{k,2}^{(i)}$ ) represents the effect from bigger (resp. smaller) solitons than  $k$ , see (5.30) for the precise formulas. In addition,  $\mathcal{J}_{k,1}^{(i)}, \mathcal{J}_{k,2}^{(i)}$  are not correlated, i.e.,  $\mathbb{E}_\nu \left[ \mathcal{J}_{k,1}^{(i)} \mathcal{J}_{k,2}^{(i)} \right] = 0$ , see (5.33). Moreover, since the initial distribution is homogeneous in space, the macroscopic limit of solitons does not depend on their numbering. This is the intuitive reason why  $D_k(\nu)$  has the form (4.15).

From Theorem 2.1 (1), if we focus on a single soliton, it converges to a Brownian motion whose variance depends only on the initial distribution  $\nu$  and the size of the soliton  $k$ , and not on the number  $i$ . In Section 4.2, we focus on two solitons of the same size under more general initial conditions and prove their strong correlations in diffusive space time scaling, see Theorem 4.8 and Corollary 4.9.

### 3. SEAT NUMBER CONFIGURATION

In order to prove the results described in Section 2, we need the linearization method of BBS. In this section, first we introduce the notion of a soliton with volume. This is a useful notion when looking at the correspondence between the linearization method and the position of solitons, which will be introduced later. and then recall the definition of the seat number configuration. Next, we recall the linearization method called “seat number configuration” introduced by [MSSS, S].

**3.1. A soliton with volume.** This subsection introduces the notion of volume for solitons. We then introduce a way to number solitons with volume. Finally, we explain how the increment of a soliton from time 0 to  $n$  is described.

We fix  $\eta \in \Omega$  and recall that  $X(\gamma)$  is the position of  $\gamma$ . We denote by  $\bar{X}(\gamma) := \sup \gamma$  the rightmost site of  $\gamma$ . For any  $\gamma, \gamma' \in \Gamma_k$  with  $X(\gamma) < X(\gamma')$ , we say that  $\gamma$  and  $\gamma'$  are connected if there are no  $\ell$ -solitons with  $\ell \geq k+1$  and records in  $[\bar{X}(\gamma), X(\gamma')]$ . In equation form,  $\gamma$  and  $\gamma'$  are connected if the followings hold :

- $[\bar{X}(\gamma), X(\gamma')] \cap \{s_\infty(i); i \in \mathbb{Z}\} = \emptyset$ ,
- for any  $\gamma'' \in \cup_{\ell \geq k+1} \Gamma_\ell$ ,  $[\bar{X}(\gamma), X(\gamma')] \cap \gamma'' = \emptyset$ .

Then, for any  $\gamma \in \Gamma_k$ , we define

$$\text{Con}(\gamma) := \{\gamma' \in \Gamma_k; \gamma \text{ and } \gamma' \text{ are connected}\}.$$

Note that  $\gamma' \in \text{Con}(\gamma)$  then  $\text{Con}(\gamma) = \text{Con}(\gamma')$ . For later use, we denote by  $\Gamma_k^*$  the set of  $k$ -solitons such that

- for any  $\gamma, \gamma' \in \Gamma_k^*$ ,  $\text{Con}(\gamma) \cap \text{Con}(\gamma') = \emptyset$ ,
- for any  $\gamma \in \Gamma_k^*$ ,  $X(\gamma) \leq X(\gamma'')$  for any  $\gamma'' \in \text{Con}(\gamma)$ .

In other words, the leftmost one among the connected ones is chosen as the representative and  $\Gamma_k^* = \Gamma_k^*(\eta)$  is the set of such representatives. Clearly, we have  $\Gamma_k = \cup_{\gamma \in \Gamma_k^*} \text{Con}(\gamma)$ . For any  $\gamma \in \Gamma_k^*$ , we say that the number of solitons in  $\text{Con}(\gamma)$  is the volume of  $\gamma$ , and write  $\text{Vol}(\gamma) := |\text{Con}(\gamma)|$ . Also, we say that for each  $k \in \mathbb{N}$ , an element  $\gamma \in \Gamma_k^*$  is a  $k$ -soliton with volume. For example, in the configuration used in Figure 2, there are three 2-solitons colored by red. The volume of leftmost 2-soliton is 1, and that of the middle 2-soliton is 2.

3.1.1. *Truncated numbering of solitons with volume.* We consider the truncated numbering for solitons with volume. We denote by  $\gamma_k^{(0)}$  the  $k$ -soliton with volume such that

$$X(\gamma_k^{(0)}) = \max_{\gamma \in \Gamma_k^*, X(\gamma) < s_\infty(0)} X(\gamma).$$

Then, we recursively define  $\gamma_k^{[i]}$  as the  $k$ -soliton with volume such that

$$X(\gamma_k^{(i)}) = \min_{\gamma \in \Gamma_k^*, X(\gamma) > X(\gamma_k^{(i-1)})} X(\gamma)$$

for  $i \geq 1$ , and

$$X(\gamma_k^{(i)}) = \max_{\gamma \in \Gamma_k^*, X(\gamma) < X(\gamma_k^{i+1})} X(\gamma)$$

for  $i \leq -1$ . We call  $\gamma_k^{(i)}$  the  $i$ -th  $k$ -soliton with volume. The difference from natural numbering is that the order is assigned only to the representatives in  $\Gamma_k^*$ . We note that  $\gamma_k^{(1)} = \gamma_k^1$  from the rules of numbering.

It is an abuse of notation, but for any  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we denote by  $X_k^{(i)}(\eta, n)$  the position of the  $i$ -th  $k$ -soliton with volume at time  $n$ , i.e.,  $X_k^{(i)}(\eta, n) = X(\eta, \gamma_k^{(i)}(n))$ .

3.1.2. *Interactions between solitons.* Recall that the points  $(s_\infty(\eta, i))_{i \in \mathbb{Z}}$  separate groups of solitons. For any  $\gamma \in \Gamma_k, \gamma' \in \Gamma_\ell$  with  $k < \ell$ , we say that  $\gamma$  and  $\gamma'$  are interacting if  $s_\infty(i) < X(\gamma') < X(\gamma) < s_\infty(i+1)$  for some  $i \in \mathbb{Z}$ . We say that  $\gamma \in \Gamma_k$  is free if  $\gamma$  does not interact with any  $\ell$ -soliton with  $\ell > k$ .

For any  $\gamma \in \Gamma_k, \gamma' \in \Gamma_\ell$  with  $k > \ell$ , we say that  $\gamma$  overtakes  $\gamma'$  ( or  $\gamma'$  is overtaken by  $\gamma$ ) at time  $n$  if  $X(\gamma(n-1)) < X(\gamma'(n-1))$  and  $X(\gamma(n)) > X(\gamma'(n))$ . We denote by  $N_\ell(\gamma, n)$  the number of  $\ell$ -solitons overtaken by  $\gamma$  at time  $n$ . It is shown by [FNRW, Proposition 6.4] that for any  $\gamma \in \Gamma_k$ , the increment  $X(\gamma(n)) - X(\gamma(n-1))$  can be represented as

$$\begin{aligned} & X(\gamma(n)) - X(\gamma(n-1)) \\ &= \begin{cases} k + 2 \sum_{\ell=1}^{k-1} \ell N_\ell(\gamma, n) & \text{if } \gamma(n-1) \text{ is free,} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (3.1)$$

In particular,  $X(\gamma(n)) - X(\gamma(n-1)) > 0$  if and only if  $\gamma(n-1)$  is free. For later use, we define

$$\begin{aligned} N_{k,\ell}^{(i)}(\eta, n) &:= \begin{cases} N_\ell(\gamma_k^{(i)}, n) & k > \ell, \\ 0 & \text{otherwise,} \end{cases} \\ M_k^{(i)}(\eta, n) &:= \left| \left\{ 1 \leq m \leq n ; \gamma_k^{(i)}(m) \text{ is not free} \right\} \right|, \end{aligned}$$

and

$$\begin{aligned} & M_{k,\ell}^{(i)}(\eta, n) \\ &:= \begin{cases} \left| \left\{ \gamma \in \Gamma_\ell ; \gamma \text{ overtakes } \gamma_k^{(i)} \text{ at time } m, 1 \leq m \leq n \right\} \right| & k < \ell, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Then, from (3.1), we have

$$\begin{aligned} & X_k^{(i)}(\eta, n) \\ &= X_k^{(i)}(\eta, 0) + k \left( n - M_k^{(i)}(\eta, n) \right) + 2 \sum_{m=1}^n \sum_{\ell=1}^{k-1} \ell N_{k,\ell}^{(i)}(\eta, m). \end{aligned} \quad (3.2)$$

We now observe the interactions between solitons. When a soliton  $\gamma$  is free and catches up with a smaller soliton  $\gamma'$ ,  $\gamma$  overtakes all of  $Con(\gamma')$  simultaneously. In addition, when  $\gamma$  catches up with  $Con(\gamma')$ ,  $Con(\gamma')$  are involved in the right half of  $\gamma$ , and if  $\gamma(1)$  is free, then in the next step  $Con(\gamma')$  are involved in the left half of  $\gamma$ . If  $\gamma(1)$  is not free, then both  $\gamma$  and  $Con(\gamma)$  do not move. For example, see Figure 4, 5 and 6.

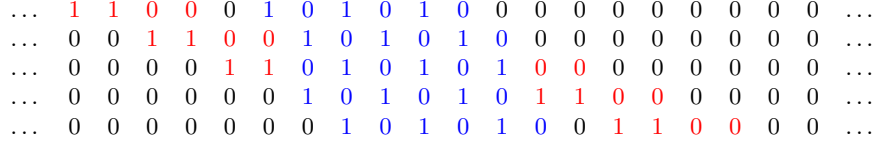


FIGURE 4. One 2-soliton with volume 1 and one 1-soliton with volume 3 are included in this figure. These solitons are interacting from the second line to fourth line. The 2-soliton overtakes the group of 1-solitons simultaneously.

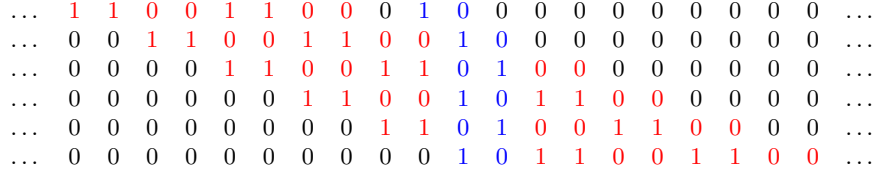


FIGURE 5. One 2-soliton with volume 2 and one 1-soliton with volume 1 are included in this figure. These solitons are interacting from the second line to fifth line. Each 2-solitons overtake the 1-soliton step by step.

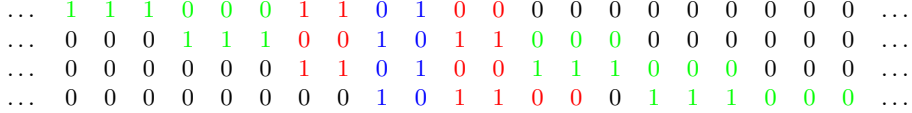


FIGURE 6. One 3-soliton with volume 1, one 2-soliton with volume 1 and one 1-soliton with volume 1 are included in this figure. The 2-soliton overtakes the 1-soliton after being overtaken by the 3-soliton.

Hence, if the  $i$ -th  $k$ -soliton is free at time  $n - 1$ , then

$$N_{k,\ell}^{(i)}(\eta, n) = \left| \left\{ \gamma \in \Gamma_\ell(\eta) ; \gamma(n-1) \subset \left[ H_1 \left( \gamma_k^{(i)}(n-1) \right), T_1 \left( \gamma_k^{(i)}(n-1) \right) \right] \right\} \right|. \quad (3.3)$$

On the other hand, we have the following inequality for  $M_k^i(n)$  :

$$2 \sum_{\ell \geq k+1} M_{k,\ell}^{(i)}(\eta, n) \leq M_k^{(i)}(\eta, n) \leq 1 + 2 \sum_{\ell \geq k+1} M_{k,\ell}^{(i)}(\eta, n). \quad (3.4)$$

We note that since the operator  $T$  and spatial shift operators are commutative, the values of  $N_{k,\ell}^{(i)}(n)$ ,  $M_k^{(i)}(n)$ ,  $M_{k,\ell}^{(i)}(n)$  are invariant under any spatial

shift that does not change the numbering of solitons. For later use, we write this fact as a lemma. We define spatial shift operators  $\tau_y : \Omega \rightarrow \Omega, y \in \mathbb{Z}$  as

$$\tau_y \eta(x) := \eta(x + y),$$

for any  $x \in \mathbb{Z}$ .

**Lemma 3.1.** *Suppose that  $\eta \in \Omega$ . Then, for any  $k, \ell \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we have*

$$X_k^{(i)}(\tau_{s_\infty(\eta, 0)} \eta, n) = X_k^{(i)}(\eta, n) - s_\infty(\eta, 0),$$

and

$$\begin{aligned} N_{k, \ell}^{(i)}(\tau_{s_\infty(\eta, 0)} \eta, n) &= N_{k, \ell}^{(i)}(\eta, n), \\ M_k^{(i)}(\tau_{s_\infty(\eta, 0)} \eta, n) &= M_k^{(i)}(\eta, n), \\ M_{k, \ell}^{(i)}(\tau_{s_\infty(\eta, 0)} \eta, n) &= M_{k, \ell}^{(i)}(\eta, n). \end{aligned}$$

Thanks to Lemma 3.1, we see that  $Y_k^{(i)}$  is independent of  $s_\infty(\eta, 0)$ . In particular,  $Y_k^{(i)}$  is a function of  $\tau_{s_\infty(\eta, 0)} \eta$ .

**3.2. Seat number configuration for the box-ball system.** To derive the limiting behaviors of solitons, it is useful to consider seat number configuration space in which the dynamics of the BBS is linearized. In this section, we briefly recall the linearization method introduced by [MSSS] and seat by [S]. The main idea of this method is to assign a different parameter to each 0, 1 in  $\eta \in \Omega$  based on the fact that  $\eta$  contains many kinds of solitons. Then, we introduce a class of invariant measures for the BBS that are defined via the seat number configuration space.

Throughout this subsection, we fix an  $\eta \in \Omega$  arbitrarily. First, we introduce the notion of *carrier with seat numbers*. We consider a situation in which the *seats* of the carrier  $W$  are indexed by  $k \in \mathbb{N}$ , i.e,  $W$  is decomposed as

$$W(\eta, x) := \sum_{k \in \mathbb{N}} \mathcal{W}_k(\eta, x), \quad \mathcal{W}_k(\eta, x) \in \{0, 1\},$$

where  $\mathcal{W}_k(\eta, x) = 1$  means that the No. $k$  seat is occupied when the carrier is at the site  $x \in \mathbb{Z}$ . Then, the refined update rule of such a carrier is given as follows:

- If there is a ball at site  $x$ , then the carrier picks up the ball and puts it at the empty seat with the smallest seat number;
- If the site  $x$  is empty, namely  $\eta(x) = 0$ , and if there is at least one occupied seat, then the carrier puts down the ball at the occupied seat with the smallest seat number;
- Otherwise, the carrier just passes through.



In other words,  $\mathcal{W}_k, k \in \mathbb{N}$  are defined as  $\mathcal{W}_k(\eta, s_\infty(\eta, i)) := 0$  for any  $i \in \mathbb{Z}$  and

$$\mathcal{W}_k(\eta, x) - \mathcal{W}_k(\eta, x-1) := \begin{cases} 1 & \text{if } \sum_{\ell=1}^{k-1} \mathcal{W}_\ell(\eta, x-1) = 1, \mathcal{W}_k(\eta, x-1) = 0 \text{ and } \eta(x) = 1, \\ -1 & \text{if } \sum_{\ell=1}^{k-1} \mathcal{W}_\ell(\eta, x-1) = 0, \mathcal{W}_k(\eta, x-1) = 1 \text{ and } \eta(x) = 0, \\ 0 & \text{otherwise,} \end{cases}$$

and we call  $\mathcal{W}_k, k \in \mathbb{N}$  the carrier with seat numbers.

By using the carrier with seat numbers, we define the *seat number configuration*  $\eta_k^\sigma \in \Omega$ ,  $\sigma \in \{\uparrow, \downarrow\}$ ,  $k \in \mathbb{N}$  and  $r \in \{0, 1\}^{\mathbb{Z}}$  as

$$\eta_k^\uparrow(x) := \begin{cases} 1 & \mathcal{W}_k(\eta, x) - \mathcal{W}_k(\eta, x-1) = 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\eta_k^\downarrow(x) := \begin{cases} 1 & \mathcal{W}_k(\eta, x) - \mathcal{W}_k(\eta, x-1) = -1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$r(\eta, x) := \begin{cases} 1 & x = s_\infty(\eta, i) \text{ for some } i \in \mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases}$$

We note that by the seat number configuration, all 1, 0 in  $\eta$  are distinguished by the parameter  $(k, \sigma)$  in the following sense : for any  $x \in \mathbb{Z}$ ,

$$r(x) + \sum_{k \in \mathbb{N}} (\eta_k^\uparrow(x) + \eta_k^\downarrow(x)) = 1.$$

In the following, if a site  $x \in \mathbb{Z}$  satisfies  $\eta_k^\sigma(x) = 1$  for some  $k \in \mathbb{N}$  and  $\sigma \in \{\uparrow, \downarrow\}$ , then we call  $x$  a  $(k, \sigma)$ -seat.

**Remark 3.2.** *We note that the seat number configuration can be described in terms of solitons as follows, see [MSSS, S] for details.*

$$\eta_k^\uparrow(x) = \begin{cases} 1 & x = H_k(\gamma) \text{ for some } \gamma \in \bigcup_{\ell \geq k} \Gamma_\ell, \\ 0 & \text{otherwise,} \end{cases}$$

$$\eta_k^\downarrow(x) = \begin{cases} 1 & x = T_k(\gamma) \text{ for some } \gamma \in \bigcup_{\ell \geq k} \Gamma_\ell, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, a  $k$ -soliton consists of exactly one  $(\ell, \sigma)$ -seat for each  $1 \leq \ell \leq k$  and  $\sigma \in \{\uparrow, \downarrow\}$ .

Then, by using the above configurations, for each  $k \in \mathbb{N}$ , we define a non-decreasing function  $\xi_k(\cdot) : \mathbb{Z} \rightarrow \mathbb{Z}$  and its inverse (for a certain sense)

$s_k(\cdot) : \mathbb{Z} \rightarrow \mathbb{Z}$  as

$$\begin{aligned} \xi_k(\eta, x) - \xi_k(\eta, x-1) &:= r(\eta, x) + \sum_{\ell \in \mathbb{N}} \sum_{\sigma \in \{\uparrow, \downarrow\}} \eta_{k+\ell}^\sigma(x), \\ \xi_k(\eta, s_\infty(\eta, 0)) &:= 0, \end{aligned}$$

and

$$s_k(\eta, x) := \min \{y \in \mathbb{Z} ; \xi_k(\eta, y) = x\}.$$

**Remark 3.3.** *The intuitive meaning of  $\xi_k$  is that it is a function that counts the number of 1s and 0s from the reference point  $s_\infty(0)$ , ignoring solitons of size  $k$  or less, and ignoring up to the  $k$ -th 1s and 0s constituting solitons, see Remark 3.2. This counting method allows us to measure the effective distance between solitons, see Remark 3.6.*

Finally, for any  $k \in \mathbb{Z}$ , we define  $\zeta_k : \mathbb{Z} \rightarrow \mathbb{Z}_{\geq 0}$  as

$$\zeta_k(\eta, i) := \sum_{y=s_k(\eta, i)+1}^{s_k(\eta, i+1)} (\eta_k^\uparrow(y) - \eta_{k+1}^\uparrow(y)).$$

We emphasize three important properties about  $\zeta_k$  as follows. The first is that the function  $\zeta_k$  and  $k$ -solitons are related via the following formula,

$$\zeta_k(\eta, i) = |\{\gamma \in \Gamma_k(\eta) ; \gamma \subset [s_k(\eta, i), s_k(\eta, i+1)]\}|,$$

i.e.,  $\zeta_k$  represents the total number of  $k$ -solitons satisfying  $\xi_k(\eta, X(\gamma)) = i$ . In particular, our  $\zeta$  coincides with the slot decomposition introduced in [FNRW], see [MSSS, Section 2.1, Proposition 2.3] and [S, Section 4.1] for details. From the same reasons as in the discussion just before Lemma 3.1, we can see that for any  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$ ,

$$\zeta_k(\tau_{s_\infty(0)}\eta, i) = \zeta_k(\eta, i). \quad (3.5)$$

The second is the bijectivity between  $\zeta$  and  $\eta$  satisfying  $s_\infty(0) = 0$ , namely the configuration such that the origin is a record. We define the space of such configurations  $\Omega_0 \subset \Omega$ , and also introduce  $\bar{\Omega} \subset \mathbb{Z}_{\geq 0}^{\mathbb{N} \times \mathbb{Z}}$  as

$$\begin{aligned} \Omega_0 &:= \Omega \cap \{s_\infty(0) = 0\} \\ \bar{\Omega} &:= \left\{ \zeta \in \mathbb{Z}_{\geq 0}^{\mathbb{N} \times \mathbb{Z}} ; \sum_{k \in \mathbb{N}} \zeta_k(i) < \infty \text{ for any } i \right\}. \end{aligned}$$

It is known that  $\zeta : \Omega_0 \rightarrow \bar{\Omega}$  is a bijection, see [FNRW, Section 3] for details. We note that we can not reconstruct the original  $\eta$  from  $(\zeta_k(\eta, \cdot))_{k \in \mathbb{N}}$  in general, because there is an arbitrariness in the choice of the position of  $s_\infty(0)$  from (3.5), see also Figure 7 for a summary of these properties, where for any  $\eta \in \Omega$ , we write

$$\tilde{\eta} := \tau_{s_\infty(0)}\eta \in \Omega_0. \quad (3.6)$$

$$\eta \longleftrightarrow (\tilde{\eta}, s_\infty(0)) \longrightarrow \tilde{\eta} \xleftarrow{\zeta} (\zeta_k(i))_{k \in \mathbb{N}, i \in \mathbb{Z}}$$

FIGURE 7. The relationship between  $\eta$  and  $\zeta$ . The arrows  $\leftrightarrow$  represent certain bijections, and the arrow  $\rightarrow$  represents the first coordinate projection.

The third is that the dynamics of the BBS can be linearized via  $\zeta_k$  with a certain offset [FNRW, S]. Here, we cite the result by [S] for later use.

**Theorem** (Theorem 4.1 in [S]). *Suppose that  $\eta \in \Omega$ . Then, for any  $k \in \mathbb{N}$  and  $i \in \mathbb{Z}$ , we have*

$$\zeta_k(T\eta, i + k + o_k(\eta)) = \zeta_k(\eta, i), \quad (3.7)$$

where the offset  $o_k(\eta)$  is given by

$$\begin{aligned} o_k(\eta) &:= s_\infty(\eta, 0) - s_\infty(T\eta, 0) + 2 \sum_{y=s_\infty(0)+1}^0 \sum_{\ell=1}^k \eta_\ell^\downarrow(y) - 2 \sum_{y=T s_\infty(0)+1}^0 \sum_{\ell=1}^k T\eta_\ell^\uparrow(y). \end{aligned}$$

**Remark 3.4.** *We note that if  $s_\infty(0) = 0$  and any soliton do not cross the origin  $x = 0$  in the evolution from  $\eta$  to  $T\eta$ , then  $o_k = 0$  for any  $k \in \mathbb{N}$ . Hence, if  $\eta(x) = 0$  for any  $x \leq 0$ , then*

$$\zeta_k(T^n \eta, i + nk) = \zeta_k(\eta, i),$$

for any  $k \in \mathbb{N}$ ,  $n \in \mathbb{N}$  and  $i \in \mathbb{Z}$ . The above equation reproduces the linearization result for the BBS on  $\{0, 1\}^{\mathbb{N}}$  shown in [MSSS].

Later in this paper we will use the following lemma, which can be considered as a version of (3.7). The proof will be given in Section A.

**Lemma 3.5.** *For any  $k \geq \ell$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{N}$ , we have*

$$\begin{aligned} &\xi_\ell(T^n \eta, X_k^{(i)}(n)) - \xi_\ell(T^{n-1} \eta, X_k^{(i)}(n-1)) \\ &= \begin{cases} k + o_\ell(T^{n-1} \eta) + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) N_{k,h}^{(i)}(\eta, n) & \text{if } \gamma_k^{(i)}(n-1) \text{ is free,} \\ \ell + o_\ell(T^{n-1} \eta) & \text{otherwise.} \end{cases} \end{aligned} \quad (3.8)$$

**Remark 3.6.** *For each  $k \in \mathbb{N}$ , we define the  $k$ -th effective distance between  $\gamma, \gamma' \in \Gamma_k$  as*

$$d_{\text{eff},k}(\eta, \gamma, \gamma') := |\xi_k(\eta, X(\gamma)) - \xi_k(\eta, X(\gamma'))|.$$

Then, we see that

$$d_{\text{eff},k}(\eta, \gamma, \gamma') = 0 \text{ if and only if } \gamma' \in \text{Con}(\gamma).$$

In addition, from Lemma 3.5, we see that the effective distance is conserved in time, i.e., for any  $\gamma, \gamma' \in \Gamma_k$  and  $n \in \mathbb{Z}_{\geq 0}$ , we have

$$\begin{aligned} d_{\text{eff},k}(T^n \eta, \gamma(n), \gamma'(n)) &= |\xi_k(T^n \eta, X(\gamma(n))) - \xi_k(T^n \eta, X(\gamma'(n)))| \\ &= d_{\text{eff},k}(\eta, \gamma, \gamma'), \end{aligned}$$

and thus we get

$$|T^n \text{Con}(\gamma(n))| = |\text{Con}(\gamma)|.$$

#### 4. GENERAL INITIAL DISTRIBUTIONS

In this section we recall a class of invariant measures for the BBS, introduced by [FG]. Then, we consider scaling limits for solitons starting from such invariant measures.

**4.1.  $\mathbf{q}$ -statistics.** We recall a class of translation-invariant stationary measures on  $\{0, 1\}^{\mathbb{Z}}$  introduced by [FG]. We define a set of infinite number of parameters as follows :

$$\mathcal{Q} := \left\{ \mathbf{q} = (q_k)_{k \in \mathbb{N}} \in [0, 1]^{\mathbb{N}} ; \sum_{k \in \mathbb{N}} k q_k < \infty \right\}.$$

From [FG, Theorem 4.4, 4.5], for given  $\mathbf{q} \in \mathcal{Q}$ , there exists a translation-invariant stationary measure  $\nu_{\mathbf{q}}$  such that  $(\zeta_k(i))_{k \in \mathbb{N}, i \in \mathbb{Z}}$  are i.i.d. for each  $k$  and independent over  $k$  under  $\nu_{\mathbf{q}}$ , and its distribution is characterized via  $\zeta$  as

$$\nu_{\mathbf{q}}(\zeta_k(i) = m) = q_k^m (1 - q_k),$$

for any  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $m \geq 0$ . In the following, we call  $\nu_{\mathbf{q}}$  the  $q$ -statistics. For later use, we recall some properties of  $\nu_{\mathbf{q}}$  in the following remark.

**Remark 4.1.** Recall that the  $i$ -th excursion  $\mathbf{e}^{(i)}(\eta)$  in  $\eta$  is defined in (2.3). These excursions are elements of the set  $\mathcal{E}$  given by

$$\begin{aligned} \mathcal{E} &:= \cup_{m \in \mathbb{Z}_{\geq 0}} \mathcal{E}(m), \\ \mathcal{E}(m) &:= \left\{ \mathbf{e} \in \{0, 1\}^{2m+1} ; \sup_{1 \leq y \leq 2m+1} \sum_{x=1}^y (2\mathbf{e}(x) - 1) \leq -1, \sum_{x=1}^{2m+1} \mathbf{e}(x) = m \right\}. \end{aligned}$$

For each  $i \in \mathbb{Z}$ ,  $\mathbf{e}^{(i)}(\eta)$  can be considered as an  $\mathcal{E}$ -valued random variable under  $\nu_{\mathbf{q}}$ . Then, from the explicit construction of  $\nu_{\mathbf{q}}$  in [FG, Section 4],  $(\mathbf{e}^{(i)})_{i \in \mathbb{Z}}$  are an i.i.d. sequence under the conditional probability measure  $\nu_{\mathbf{q}}(\cdot | s_{\infty}(0) = 0)$ . In particular, the centered configuration  $\tilde{\eta} \in \Omega_0$  defined in (3.6) is record-shift invariant under  $\nu_{\mathbf{q}}$ , i.e.,

$$\nu_{\mathbf{q}}(\tilde{\eta} \in \cdot) = \nu_{\mathbf{q}}(\tau_{s_{\infty}(x) - s_{\infty}(0)} \tilde{\eta} \in \cdot),$$

for any  $x \in \mathbb{Z}$ .

The Bernoulli product measures and stationary Markov distributions are two important classes of  $q$ -statistics. Let  $\text{Ber}(\rho)$  be the Bernoulli product measure on  $\{0, 1\}^{\mathbb{Z}}$  with intensity  $0 < \rho < 1/2$ . By choosing  $\mathbf{q}$  as

$$q_1 := \rho(1 - \rho), \quad q_k := \frac{\rho^k (1 - \rho)^k}{\prod_{\ell=1}^{k-1} (1 - q_\ell)^{2(k-\ell)}} \text{ for } k \geq 2, \quad (4.1)$$

from [FG, Theorem 3.1, Corollary 4.6], we have  $\nu_{\mathbf{q}} = \text{Ber}(\rho)$ . We denote the class of parameters  $\mathbf{q} = \mathbf{q}(\rho)$  given by (4.1) for  $0 < \rho < \frac{1}{2}$  by  $\mathcal{Q}_{\text{Ber}} \subset \mathcal{Q}$ .

Another important class of  $q$ -statistic is two-sided Markov distribution on  $\{0, 1\}^{\mathbb{Z}}$  with transition matrix  $P = (p_{ij})_{i,j=0,1}$  on  $\{0, 1\}$  satisfying  $0 < p_{01} + p_{11} < 1$ . In [FG], it is proved that such Markov distribution can be obtained by choosing  $\mathbf{q}$  as

$$q_1 := a, \quad q_k := \frac{ab^{k-1}}{\prod_{\ell=1}^{k-1} (1 - q_\ell)^{2(k-\ell)}} \text{ for } k \geq 2, \quad (4.2)$$

where

$$a := p_{01}p_{10}, \quad b := p_{00}p_{11}.$$

As shown in [S, Section 5.3], the above map  $P = (p_{ij})_{i,j=0,1} \rightarrow (a, b)$  induces a bijection between the set of transition matrix  $\{P = (p_{ij})_{i,j=0,1} ; 0 < p_{01} + p_{11} < 1\}$  and the set of the pair of parameters  $\{(a, b) ; a > 0, 0 \leq b < 1, \sqrt{a} + \sqrt{b} < 1\}$ . We define

$$\mathcal{Q}_{\text{M}} := \{\mathbf{q} ; \nu_{\mathbf{q}} \text{ is a two-sided Markov distribution}\}.$$

For each  $\mathbf{q} \in \mathcal{Q}_{\text{M}}$ , we denote by  $a(\mathbf{q}), b(\mathbf{q})$  the pair of parameters giving  $\mathbf{q}$  by (4.2). Note that by taking  $a = b = \rho(1 - \rho)$  we have  $\mathcal{Q}_{\text{Ber}} \subset \mathcal{Q}_{\text{M}}$ , and by taking  $b = 0$ , we have  $\{\mathbf{q} \in \mathcal{Q} ; q_k = 0 \text{ } k \geq 2\} \subset \mathcal{Q}_{\text{M}}$ .

In the following, we will introduce another class of  $\mathbf{q}$ -statistics. To do so, we define a shift operator  $\theta : [0, 1]^{\mathbb{N}} \rightarrow [0, 1]^{\mathbb{N}}$  as  $\theta\mathbf{q} = (q_{k+1})_{k \in \mathbb{N}}$  for any  $\mathbf{q} = (q_k)_{k \in \mathbb{N}} \in [0, 1]^{\mathbb{N}}$ . We note that  $\theta\mathcal{Q} = \mathcal{Q}$ . Moreover, in [S, Theorem 5.3], it is shown that for  $\mathbf{q} \in \mathcal{Q}_{\text{M}}$ ,  $\theta\mathbf{q} \in \mathcal{Q}_{\text{M}}$  with

$$a(\theta\mathbf{q}) = \frac{a(\mathbf{q})b(\mathbf{q})}{(1 - a(\mathbf{q}))^2}, \quad b(\theta\mathbf{q}) = \frac{b(\mathbf{q})}{(1 - a(\mathbf{q}))^2}. \quad (4.3)$$

From this, we have  $\theta\mathcal{Q}_{\text{M}} \subset \mathcal{Q}_{\text{M}}$ , but  $\theta\mathcal{Q}_{\text{M}} \neq \mathcal{Q}_{\text{M}}$ . Actually,  $\mathcal{Q}_{\text{Ber}} \not\subset \theta\mathcal{Q}_{\text{M}}$  since  $a(\mathbf{q})b(\mathbf{q}) < 1$  for any  $\mathbf{q} \in \mathcal{Q}_{\text{M}}$ . We also note that for any  $\mathbf{q} \in \mathcal{Q}_{\text{Ber}}$ ,  $\theta\mathbf{q} \notin \mathcal{Q}_{\text{Ber}}$ .

We say that  $\mathbf{q} \in \mathcal{Q}$  is asymptotically Markov if there exists some  $K \in \mathbb{N}$  such that  $\theta^{K-1}\mathbf{q} \in \mathcal{Q}_{\text{M}}$  with convention  $\theta^0\mathbf{q} = \mathbf{q}$ . We define

$$\mathcal{Q}_{\text{AM}} := \{\mathbf{q} \in \mathcal{Q} ; \mathbf{q} \text{ is asymptotically Markov}\}. \quad (4.4)$$

We note that  $\theta\mathcal{Q}_{\text{AM}} = \mathcal{Q}_{\text{AM}}$  since for any  $\mathbf{q} \in \mathcal{Q}_{\text{AM}}$ ,  $\theta\tilde{\mathbf{q}} = \mathbf{q}$  where  $\tilde{q}_1 = 0, \tilde{q}_k = q_{k-1}$  for  $k \geq 2$ . For  $\mathbf{q} \in \mathcal{Q}_{\text{AM}}$ , we define  $K(\mathbf{q})$  as

$$K(\mathbf{q}) := \min \{\ell \in \mathbb{N} ; \theta^{\ell-1}\mathbf{q} \in \mathcal{Q}_{\text{M}}\}. \quad (4.5)$$

In particular, for  $\mathbf{q} \in \mathcal{Q}_{\text{AM}}$ ,  $\mathbf{q} \in \mathcal{Q}_{\text{M}}$  if and only if  $K(\mathbf{q}) = 1$ . In summary, we have  $\mathcal{Q}_{\text{Ber}} \subsetneq \mathcal{Q}_{\text{M}} \subsetneq \mathcal{Q}_{\text{AM}} \subsetneq \mathcal{Q}$  and  $\theta\mathcal{Q}_{\text{Ber}} \not\subset \mathcal{Q}_{\text{Ber}}$ ,  $\theta\mathcal{Q}_{\text{M}} \subsetneq \mathcal{Q}_{\text{M}}$ ,  $\theta\mathcal{Q}_{\text{AM}} = \mathcal{Q}_{\text{AM}}$  and  $\theta\mathcal{Q} = \mathcal{Q}$ .

For later use, we introduce some notations. For any  $k \in \mathbb{N}$ , we define  $C_k : \mathcal{Q} \rightarrow \mathcal{Q}$  as

$$(C_k \mathbf{q})_\ell := \begin{cases} q_\ell & 1 \leq \ell \leq k, \\ 0 & \ell \geq k+1, \end{cases}$$

for any  $\mathbf{q} \in \mathcal{Q}$ . We note that under  $\nu_{C_k \mathbf{q}}$ , there are no solitons larger than  $k$  a.s. Next, for any  $k \in \mathbb{N}$  and  $\mathbf{q} \in \mathcal{Q}$ , we define  $\alpha_k(\mathbf{q})$ ,  $\beta_k(\mathbf{q})$ ,  $\bar{r}_k(\mathbf{q})$  as

$$\alpha_k(\mathbf{q}) := \mathbb{E}_{\nu_{\mathbf{q}}}[\zeta_k(0)] = \frac{q_k}{1 - q_k}, \quad (4.6)$$

$$\beta_k(\mathbf{q}) := \mathbb{E}_{\nu_{\mathbf{q}}}[(\zeta_k(0) - \alpha_k(\mathbf{q}))^2] = \frac{q_k}{(1 - q_k)^2}, \quad (4.7)$$

$$\bar{r}_k(\mathbf{q}) := \mathbb{E}_{\nu_{\theta^k \mathbf{q}}}[r(0)]. \quad (4.8)$$

We note that  $\bar{r}_k(\mathbf{q})$ ,  $k \in \mathbb{N}$  satisfies the following system,

$$\frac{1}{\bar{r}_k(\mathbf{q})} = 1 + 2 \sum_{\ell=k+1}^{\infty} \frac{(\ell - k) \alpha_\ell(\mathbf{q})}{\bar{r}_\ell(\mathbf{q})}, \quad (4.9)$$

see Section B.1 for the derivation of (4.9).

**4.2. Scaling limits for solitons under  $\mathbf{q}$ -statistics.** In this subsection we will describe our main results on the fluctuations of  $k$ -solitons under the  $\mathbf{q}$ -statistics. Since we are interested in the increment of the position of a fixed  $k$ -soliton from time 0 to  $n$ , for any  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_+$  we define

$$Y_k^{(i)}(\eta, n) := X_k^{(i)}(\eta, n) - X_k^{(i)}(\eta, 0).$$

First we recall that by [FNRW, Theorem 1.1, 1.5] and [FG, Theorem 4.5],  $Y_k^{(i)}(\cdot)$  satisfies the law of large numbers in the hyperbolic scaling under  $\nu_{\mathbf{q}}$ , and for later use we describe this fact as follows.

**Theorem** (Theorem 1.1, 1.5 in [FNRW] + Theorem 4.5 in [FG]). *Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $q_k > 0$  for some  $k \in \mathbb{N}$ . Then, for any  $i \in \mathbb{Z}$ , we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} Y_k^{(i)}(\eta, n) = v_k^{\text{eff}}(\mathbf{q}), \quad \nu_{\mathbf{q}}\text{-a.s.} \quad (4.10)$$

The constant  $v_k^{\text{eff}}(\mathbf{q})$ ,  $k \in \mathbb{N}$  is called the effective velocity of  $k$ -solitons. In this paper, we will show the  $\mathbb{L}^p$  version of the above LLN for any  $p \geq 1$ .

**Theorem 4.2.** *Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $q_k > 0$  for some  $k \in \mathbb{N}$ . Then for any  $i \in \mathbb{Z}$  and  $p \geq 1$ , we have*

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \frac{1}{n} Y_k^{(i)}(n) - v_k^{\text{eff}}(\mathbf{q}) \right|^p \right] = 0.$$

**Remark 4.3.** *If  $X_k^{(i)}(0)$  has the finite  $p$ -th moment, then one can show the  $\mathbb{L}^p$  convergence for  $X_k^{(i)}(n)/n$  instead of  $Y_k^{(i)}(n)/n$ . When  $\mathbf{q} \in \mathcal{Q}_M$ , then  $X_k^{(i)}(0)$  has the finite  $p$ -th moment for any  $p \geq 1$ , see Section 11.3.*

We will use the following relation between effective velocities. Recall that  $\alpha_k, \bar{r}_k$  are defined in (4.6) and (4.8).

**Proposition 4.4.** *Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $q_k > 0$  for some  $k \in \mathbb{N}$ . Then, we have*

$$v_k^{\text{eff}}(\mathbf{q}) = kv_1^{\text{eff}}(\theta^{k-1}\mathbf{q}) + 2 \sum_{\ell=1}^{k-1} \ell \alpha_\ell(\mathbf{q}) v_{k-\ell}^{\text{eff}}(\theta^\ell \mathbf{q}), \quad (4.11)$$

and

$$v_1^{\text{eff}}(\theta^{k-1}\mathbf{q}) = \bar{r}_k(\mathbf{q}). \quad (4.12)$$

The proof of Proposition 4.4 will be given in Section B.2.

Our purpose in this paper is to consider the fluctuations of  $Y_k^{(i)}(\cdot)$  corresponding to the law of large numbers mentioned above. The following result implies that the invariance principle(IP)/large deviations principle(LDP) for  $Y_k^{(i)}(\cdot)$  can be reduced to the IP/LDP for  $M_k^{(i)}(\cdot)$ .

**Theorem 4.5.**

- (1) *Suppose that there exist some  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$  and  $i \in \mathbb{Z}$  such that  $q_k > 0$  and the following step-interpolation process,*

$$t \mapsto \frac{1}{n} M_k^{(i)}(\eta, \lfloor n^2 t \rfloor) - t \mathbb{E}_{\nu_{\mathbf{q}}} [M_k^{(i)}(n)], \quad (4.13)$$

*converges weakly in  $D([0, \mathbf{T}])$ ,  $\mathbf{T} > 0$  to the centered Brownian motion with variance  $G_k(\mathbf{q})$  under  $\nu_{\mathbf{q}}$ . Then, the following step-interpolation process*

$$t \mapsto \frac{1}{n} Y_k^{(i)}(\eta, \lfloor n^2 t \rfloor) - nt v_k^{\text{eff}}(\mathbf{q}), \quad (4.14)$$

*also converges weakly in  $D([0, \mathbf{T}])$ ,  $\mathbf{T} > 0$  to the centered Brownian motion with variance  $D_k(\mathbf{q})$  under  $\nu_{\mathbf{q}}$ , where  $D_k(\mathbf{q})$  is given by*

$$D_k(\mathbf{q}) := \frac{v_k^{\text{eff}}(\mathbf{q})^2 G_k(\mathbf{q})}{v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})^2} + 4 \sum_{\ell=1}^{k-1} \frac{v_\ell^{\text{eff}}(\mathbf{q})^2 v_{k-\ell}^{\text{eff}}(\theta^\ell \mathbf{q}) \beta_\ell(\mathbf{q})}{v_1^{\text{eff}}(\theta^{\ell-1}\mathbf{q})^2}, \quad (4.15)$$

*and  $\beta_\ell(\mathbf{q})$  is defined in (4.7).*

- (2) *Suppose that there exist some  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$  and  $i \in \mathbb{Z}$  such that  $q_k > 0$  and the following limit*

$$\Lambda_{\mathbf{q},k}^M(\lambda) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda \left( n - M_k^{(i)}(n) \right) \right) \right] \right) \in \mathbb{R}, \quad (4.16)$$

exists for any  $\lambda \in \mathbb{R}$ , and  $\Lambda_{\mathbf{q},k}^M(\cdot)$  is essentially smooth in the sense of [DZ, Definition 2.3.5]. Then, the following limit

$$\Lambda_{\mathbf{q},k}^Y(\lambda) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda Y_k^{(i)}(n) \right) \right] \right) \in \mathbb{R} \cup \{\infty\},$$

exists for any  $\lambda \in \mathbb{R}$ , and  $\Lambda_{\mathbf{q},k}^Y(\lambda)$  satisfies (7.6). In addition, we have  $\sup_{|\lambda| \leq \delta} |\Lambda_{\mathbf{q},k}^Y(\lambda)| < \infty$  for sufficiently small  $\delta > 0$ , and  $\Lambda_{\mathbf{q},k}^Y(\cdot)$  is also essentially smooth. Consequently, thanks to the Gärtner-Ellis theorem (cf. [DZ, Theorem 2.3.6]), the sequence  $\left( \frac{Y_k^{(i)}(n)}{n} \right)_{n \in \mathbb{Z}_{\geq 0}}$  satisfies the LDP with the good rate function  $I_{\mathbf{q},k}^Y$ , where

$$I_{\mathbf{q},k}^Y(u) := \sup_{\lambda \in \mathbb{R}} \{ \lambda u - \Lambda_{\mathbf{q},k}^Y(\lambda) \}. \quad (4.17)$$

**Remark 4.6.** We note that  $v_k^{\text{eff}}(\mathbf{q})$  can be given by

$$\begin{aligned} v_k^{\text{eff}}(\mathbf{q}) &= \frac{d\Lambda_{\mathbf{q},k}^Y(\lambda)}{d\lambda} \Big|_{\lambda=0} \\ &= \bar{r}_k(\mathbf{q}) v_k^{\text{eff}}(C_k \mathbf{q}). \end{aligned} \quad (4.18)$$

In addition, (4.18) gives the same formula for the effective velocity as the formula by [FNRW, (1.12)], see Section B.3 for the proof of (4.18) and the equivalence between the formulas.

**Remark 4.7.** We note that Theorem 4.5 can be shown with initial distribution  $\nu$ , not necessarily  $\mathbf{q}$ -statics, such that  $(\zeta_k(i))_{k \in \mathbb{N}, i \in \mathbb{Z}}$  are i.i.d. for each  $k$  and independent over  $k$  and satisfy exponential moment condition by the same argument in this paper.

In the next subsection we will give sufficient conditions for  $\mathbf{q}, k$  such that the assumptions in Theorem 4.5 are satisfied.

Next, we consider the correlations between two  $k$ -solitons. Our second result implies that even if two  $k$ -solitons are macroscopically far apart, they are strongly correlated in the diffusive space-time scaling. Recall that  $\mathbf{e}^{(i)}(\eta)$  is defined in (2.3).

**Theorem 4.8.** Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $\mathbb{E}_{\nu_{\mathbf{q}}} \left[ |\mathbf{e}^{(0)}|^2 \right] < \infty$ . Then, for any  $k \in \mathbb{N}$  with  $q_k > 0$ ,  $u, v \in \mathbb{R}$  and  $0 \leq a \leq 1$  we have

$$\overline{\lim}_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \frac{1}{n} Y_k^{(\lfloor n^a u \rfloor)}(n^2) - \frac{1}{n} Y_k^{(\lfloor n^a v \rfloor)}(n^2) \right|^2 \right] = 0.$$

By combining Theorems 4.5 and 4.8, we have the following.

**Corollary 4.9.** Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $k \in \mathbb{N}$  satisfy the assumption of Theorem 4.5 (1) and  $\mathbb{E}_{\nu_{\mathbf{q}}} \left[ |\mathbf{e}^{(i)}|^2 \right] < \infty$ . Then, for any  $u, v \in \mathbb{R}$  and  $\mathbf{T} > 0$ , we



have the following weak convergence in  $D([0, \mathbf{T}])^2$  under  $\nu_{\mathbf{q}}$ .

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} \left( \frac{1}{n} Y_k^{(\lfloor n\varepsilon t \rfloor)} (\lfloor n^2 t \rfloor) - n v_k^{\text{eff}}(\mathbf{q}), \frac{1}{n} Y_k^{(\lfloor n\varepsilon t \rfloor)} (\lfloor n^2 t \rfloor) - n v_k^{\text{eff}}(\mathbf{q}) \right) \\ &= (B_k(t), B_k(t)), \end{aligned}$$

where  $B_k(\cdot)$  is the centered Brownian motion with variance  $D_k(\mathbf{q})$ .

Hence, under the assumption of Theorems 4.5 (1) and 4.8,  $k$ -solitons with volume starting at macroscopic distance converge to the same Brownian motion.

**4.3. Scaling limits for  $M_k^{(i)}(\cdot)$ .** By Theorem 4.5, we have found that for  $\mathbf{q} \in \mathcal{Q}$  and  $k \in \mathbb{N}$  such that IP/LDP for  $M_k^{(i)}(\cdot)$  hold, IP/LDP for  $k$ -solitons also hold. In this subsection, we give some sufficient conditions of such  $\mathbf{q}, k$ . To describe the results, we define  $\rho(\mathbf{q})$  as the ball density under  $\nu_{\mathbf{q}}$ , i.e.,

$$\rho(\mathbf{q}) := \nu_{\mathbf{q}}(\eta(0) = 1). \quad (4.19)$$

First we consider the case that  $\mathbf{q}$  is asymptotically Markov. Recall that  $\mathcal{Q}_{\text{AM}}, K(\mathbf{q})$  are defined in (4.4) and (4.5). If  $k$  is sufficiently large, we can show that  $M_k^{(i)}(\cdot)$  satisfies the invariance principle, and the nice regularity property of (4.16).

**Theorem 4.10.** *If  $\mathbf{q} \in \mathcal{Q}_{\text{AM}}$  and  $k \geq K(\mathbf{q})$ , then for any  $i \in \mathbb{Z}$ , (4.13) converges weakly to the Brownian motion with variance  $G_k(\mathbf{q})$  under  $\nu_{\mathbf{q}}$ , where  $G_k(\mathbf{q})$  is given by*

$$G_k(\mathbf{q}) = 4\rho(\theta^k \mathbf{q})(1 - \rho(\theta^k \mathbf{q}))(1 - 2\rho(\theta^k \mathbf{q})).$$

*In addition, for any  $i \in \mathbb{Z}$  and  $\lambda \in \mathbb{R}$ , the limit  $\Lambda_{\mathbf{q},k}^M(\lambda)$  exists. In addition,  $\Lambda_{\mathbf{q},k}^M(\lambda)$  is a smooth monotone convex function, which is explicitly given by*

$$\Lambda_{\mathbf{q},k}^M(\lambda) = \log \left( \frac{1 - 2\rho(\theta^k \mathbf{q})}{2(1 - \rho(\theta^k \mathbf{q}))} \left( e^\lambda + \sqrt{e^{2\lambda} - 1 + \frac{1}{(1 - 2\rho(\theta^k \mathbf{q}))^2}} \right) \right).$$

*In particular, the assumptions of Theorem 4.5 (1) and (2) are satisfied with  $\mathbf{q} \in \mathcal{Q}_{\text{AM}}$  and  $k \geq K(\mathbf{q})$ .*

**Remark 4.11.** *Recall that if  $\nu_{\mathbf{q}}$  is a Bernoulli product measure or two-sided Markov distribution, then  $K(\mathbf{q}) = 1$ . Hence, when the initial distribution is a Bernoulli product measure or two-sided Markov distribution supported on  $\Omega$ , then the statement of Theorem 4.10 holds for any  $k \in \mathbb{N}$ .*

**Remark 4.12.** *If  $\mathbf{q} \in \mathcal{Q}_M$ , then the ball density  $\rho(\mathbf{q})$  can be represented via  $a(\mathbf{q}), b(\mathbf{q})$  as*

$$\rho(\mathbf{q}) = \frac{1}{2} \left( 1 - \sqrt{1 - \frac{4a(\mathbf{q})}{(1 + a(\mathbf{q}) - b(\mathbf{q}))^2}} \right).$$

We note that if  $\mathbf{q} \in \mathcal{Q}_M$ , then  $\rho(\theta\mathbf{q}) < \rho(\mathbf{q})$  holds. To show this, it suffices to show that

$$\frac{a(\theta\mathbf{q})}{(1+a(\theta\mathbf{q})-b(\theta\mathbf{q}))^2} < \frac{a(\mathbf{q})}{(1+a(\mathbf{q})-b(\mathbf{q}))^2},$$

and by using (4.3), we see that

$$0 < \sqrt{a(\mathbf{q})} + \sqrt{b(\mathbf{q})} < 1 \text{ implies } \frac{a(\theta\mathbf{q})}{(1+a(\theta\mathbf{q})-b(\theta\mathbf{q}))^2} < \frac{a(\mathbf{q})}{(1+a(\mathbf{q})-b(\mathbf{q}))^2}.$$

**Remark 4.13.** By (4.12) and (B.1), under the assumption of Theorem 4.10,  $G_k(\mathbf{q})$  and  $\Lambda_{\mathbf{q},k}^M(\lambda)$  can be represented as

$$G_k(\mathbf{q}) = v_1^{\text{eff}}(\theta^{k-1}\mathbf{q}) \left(1 - v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})^2\right),$$

and

$$\Lambda_{\mathbf{q},k}^M(\lambda) = \log \left( \frac{v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})}{1 + v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})} \left( e^\lambda + \sqrt{e^{2\lambda} + \frac{1 - v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})^2}{v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})^2}} \right) \right).$$

Next, we consider the case where there are at most a finite number of nonzero elements in  $\mathbf{q}$ , i.e., there are at most a finite number of types of solitons under  $\nu_{\mathbf{q}}$ . If we denote by  $q_\ell$  the largest nonzero element, then  $M_\ell^{(i)} = 0$   $\nu_{\mathbf{q}}$ -a.s., and thus  $M_\ell^{(i)}$  trivially satisfies the assumptions in Theorem 4.5. For the second largest solitons in  $\mathbf{q}$ , we can show the following.

**Theorem 4.14.** Suppose that  $\mathbf{q} \in \mathcal{Q}$  satisfies  $q_\ell > 0$ ,  $q_h = 0$ ,  $h \geq \ell + 1$  with some  $\ell \geq 2$  and  $|\mathbf{q}| \geq 2$ . We denote by  $k = k(\mathbf{q})$  the second largest element in  $\mathbf{q}$ , i.e.,

$$k := \max \{1 \leq h \leq \ell - 1; q_h > 0\}.$$

Then, for any  $i \in \mathbb{Z}$ , (4.13) converges weakly to the Brownian motion with variance  $G_k(\mathbf{q})$  under  $\nu_{\mathbf{q}}$ , where  $G_k(\mathbf{q})$  is given by

$$G_k(\mathbf{q}) = \frac{4q_\ell}{(1-q_\ell)^{2(\ell-k)}} \left(1 + \frac{4q_\ell}{(1-q_\ell)^{2(\ell-k)}}\right)^{-\frac{3}{2}}.$$

In addition, for any  $i \in \mathbb{Z}$ , the limit  $\Lambda_{\mathbf{q},k}^M(\lambda)$  exists. In addition,  $\Lambda_{\mathbf{q},k}^M(\lambda)$  is a smooth monotone convex function, which is explicitly given by

$$\Lambda_{\mathbf{q},k}^M(\lambda) = \log \left( \frac{(1-q_\ell)^{\ell-k} e^\lambda}{2} + \sqrt{\frac{(1-q_\ell)^{2(\ell-k)} e^{2\lambda}}{4} + q_\ell} \right).$$

In particular, the assumptions of Theorem 4.5 (1) and (2) are satisfied with the above  $\mathbf{q}$  and  $k, \ell$ .

5.  $k$ -SKIP MAP FOR THE BBS

In this section we introduce the notion of  $k$ -skip map. The  $k$ -skip map is a natural generalization of the 10-elimination introduced by [MIT] in terms of the seat number configuration, and the results presented in this section are crucial for the proofs of main results. For the proofs of some known results on the  $k$ -skip map, we may refer to [S].

For any  $k \in \mathbb{N}$ , we define the  $k$ -skip map  $\Psi_k : \Omega \rightarrow \Omega$  as

$$\Psi_k(\eta)(x) := \eta(s_k(\eta, x + \xi_k(\eta, 0))).$$

First we explain the intuitive meaning of the  $k$ -skip map when  $k = 1$ . Since  $s_1(\eta, \cdot)$  is the inverse function of  $\xi_1(\eta, \cdot)$ , the subset  $\{s_1(\eta, x) ; x \in \mathbb{Z}\} \subset \mathbb{Z}$  does not include the non-increasing points of  $\xi_1(\eta, \cdot)$ , i.e.,

$$\{s_1(\eta, x) ; x \in \mathbb{Z}\} = \mathbb{Z} \setminus \{x \in \mathbb{Z} ; \eta_1^\uparrow(x) + \eta_1^\downarrow(x) = 1\}.$$

When  $0 \in \{s_1(x) ; x \in \mathbb{Z}\}$ , then  $\Psi_1(\eta)$  is obtained by removing all 1, 0 with parameter  $(1, \sigma)$ ,  $\sigma \in \{\uparrow, \downarrow\}$  from  $\eta$ , and numbering the remaining 1 and 0 from left to right with respect to the origin  $\eta(0)$ . For the case  $0 \notin \{s_1(\eta, x) ; x \in \mathbb{Z}\}$ , we first translate  $\eta$  by

$$\inf \{s_1(\eta, x) ; s_1(\eta, x) < 0\} = s_1(\eta, \xi_1(\eta, 0)),$$

so that  $0 \in \{s_1(\tau_{s_1(\eta, \xi_1(\eta, 0))}\eta, x) ; x \in \mathbb{Z}\}$ , and then we perform the same operation for  $\tau_{s_1(\eta, \xi_1(\eta, 0))}\eta$ .

$x$	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\eta(x)$	...	0	1	1	0	0	1	1	1	0	1	0	1	1	0	0	0	1	0	0	...
$\eta_1^\uparrow(x)$	...	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	...
$\eta_1^\downarrow(x)$	...	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	...
	...	0	X	1	X	0	X	1	1	X	X	X	X	1	X	0	0	X	X	0	...
$x$	-3	-2	-1	0	1	2								3	4	5				6	7
$\Psi_1(\eta)$	...	0	1	0	1	1								1	0	0				0	...

FIGURE 8. How  $\Psi_1(\eta)$  can be obtained from  $\eta$  for the case  $0 \in \{s_1(\eta, x) ; x \in \mathbb{Z}\}$ , where ... represents the consecutive 0s with infinite length.

$x$	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\eta(x)$	...	0	1	1	0	0	1	1	1	0	1	0	1	1	0	0	0	1	0	0	...
$\tau_{-1}\eta(x)$	...	0	0	1	1	0	0	1	1	1	0	1	0	1	1	0	0	0	1	0	...
$\tau_{-1}\eta_1^\uparrow(x)$	...	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	...
$\tau_{-1}\eta_1^\downarrow(x)$	...	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	...
	...	0	0	$\times$	1	$\emptyset$	0	$\times$	1	1	$\emptyset$	$\times$	$\emptyset$	$\times$	1	$\emptyset$	0	0	$\times$	$\emptyset$	...
$x$	-4	-3	-2		-1	0	1	2						3	4	5					6
$\Psi_1(\eta)$	...	0	0		1	0	1	1						1	0	0					...

FIGURE 9. How  $\Psi_1(\eta)$  can be obtained from  $\eta$  for the case  $0 \notin \{s_1(\eta, x) ; x \in \mathbb{Z}\}$ .

The above observations can be made for any  $k \in \mathbb{N}$  as well. Now, we cite some results by [S]. The following means that there is a one-to-one correspondence between sites in  $\eta$  with parameter  $(k + \ell, \sigma)$  and sites in  $\Psi_k(\eta)$  with parameter  $(\ell, \sigma)$ , for any  $k, \ell \in \mathbb{N}$  and  $\sigma \in \{\uparrow, \downarrow\}$ . This property implies that  $\Psi_k$  has the semi-group property and that  $\Psi_k$  is a shift operator for  $\zeta$ , see [S] for the details and proofs.

**Proposition** (Proposition 4.3 in [S]). *Suppose that  $\eta \in \Omega$ . Then, for any  $k, \ell \in \mathbb{N}$ ,  $\sigma \in \{\uparrow, \downarrow\}$  and  $x \in \mathbb{Z}$ , we have*

$$\Psi_k(\eta)_\ell^\sigma(x) = \eta_{k+\ell}^\sigma(s_k(\eta, x + \xi_k(\eta, 0))). \quad (5.1)$$

In addition, we have

$$\Psi_k(\Psi_\ell(\eta))(\cdot) = \Psi_{k+\ell}(\eta)(\cdot), \quad (5.2)$$

and

$$\zeta_k(\Psi_\ell(\eta), \cdot) = \zeta_{k+\ell}(\eta, \cdot). \quad (5.3)$$

In [S], the following result has been proven. We recall that  $\tilde{\eta}$  is defined in (3.6).

**Theorem** (Corollary 5.1 in [S]). *Suppose that  $\mathbf{q} \in \mathcal{Q}$ . Then, for any  $k \in \mathbb{N}$  and local function  $f : \{0, 1\}^{\mathbb{Z}} \rightarrow \mathbb{R}$ , we have*

$$\int_{\Omega} d\nu_{\mathbf{q}}(\eta) f(\Psi_k(\eta)) = \int_{\Omega} d\nu_{\theta^k \mathbf{q}}(\eta) f(\eta), \quad (5.4)$$

and

$$\int_{\Omega} d\nu_{\mathbf{q}}(\eta) f(\Psi_k(\tilde{\eta})) = \int_{\Omega} d\nu_{\theta^k \mathbf{q}}(\eta) f(\tilde{\eta}). \quad (5.5)$$

**Remark 5.1.** Notice that thanks to (5.4) or (5.5), we have

$$\alpha_k(\theta^\ell \mathbf{q}) = \alpha_{k+\ell}(\mathbf{q}), \quad \beta_k(\theta^\ell \mathbf{q}) = \beta_{k+\ell}(\mathbf{q}), \quad \bar{r}_k(\theta^\ell \mathbf{q}) = \bar{r}_{k+\ell}(\mathbf{q}),$$

for any  $k \in \mathbb{N}$ ,  $\ell \in \mathbb{Z}_{\geq 0}$ . Calculations similar to the above appear frequently in this paper.

From now on we prepare some lemmas for the proofs of main results. First we check the relation between  $\Psi_k$  and  $\tau_{s_\infty(0)}$ .

**Lemma 5.2.** Assume that  $\eta \in \Omega$ . Then, for any  $k \in \mathbb{N}$  and  $x \in \mathbb{Z}$ , we have

$$\Psi_k(\tilde{\eta})(x) = \widetilde{\Psi_k(\eta)}(x).$$

*Proof of Lemma 5.2.* First we observe that for any  $k \in \mathbb{N}$  and  $x \in \mathbb{Z}$ ,

$$s_k(\tilde{\eta}, x) + s_\infty(0) = s_k(\eta, x).$$

Next, from [S, (4.16)], for any  $k \in \mathbb{N}$ , we get

$$s_\infty(\Psi_k(\eta), 0) = -\xi_k(\eta, 0). \quad (5.6)$$

By using the above, we have

$$\begin{aligned} \Psi_k(\tilde{\eta})(x) &= \eta(s_k(\tilde{\eta}, x) + s_\infty(0)) \\ &= \eta(s_k(\eta, x)) \\ &= \Psi_k(\eta)(x - \xi_k(\eta, 0)) \\ &= \Psi_k(\eta)(x + s_\infty(\Psi_k(\eta), 0)) \\ &= \widetilde{\Psi_k(\eta)}(x). \end{aligned}$$

□

By combining (5.2), (5.3), Lemma 5.2 and the diagram in Figure 7, the relation between  $\eta, \zeta$  and the  $k$ -skip map can be expressed by the diagram, see Figure 10.

$$\begin{array}{ccccc} \eta & \xrightarrow{\tau_{s_\infty(0)}} & \tilde{\eta} & \xleftarrow[\zeta^{-1}]{\zeta} & (\zeta_k(i))_{k \in \mathbb{N}, i \in \mathbb{Z}} \\ \downarrow \Psi_1 & & \downarrow \Psi_1 & & \downarrow \Psi_1 \\ \Psi_1(\eta) & \xrightarrow{\tau_{s_\infty(\Psi_1(\eta), 0)}} & \widetilde{\Psi_1(\eta)} & \xleftarrow[\zeta^{-1}]{\zeta} & (\zeta_{k+1}(i))_{k \in \mathbb{N}, i \in \mathbb{Z}} \\ \downarrow \Psi_1 & & \downarrow \Psi_1 & & \downarrow \Psi_1 \\ \Psi_2(\eta) & \xrightarrow{\tau_{s_\infty(\Psi_2(\eta), 0)}} & \widetilde{\Psi_2(\eta)} & \xleftarrow[\zeta^{-1}]{\zeta} & (\zeta_{k+2}(i))_{k \in \mathbb{N}, i \in \mathbb{Z}} \\ \downarrow \Psi_1 & & \downarrow \Psi_1 & & \downarrow \Psi_1 \\ \dots & & \dots & & \dots \end{array}$$

FIGURE 10. The relationships between  $\eta, \zeta$  and the  $k$ -skip map.

Next we claim that for any  $k > \ell$  and  $i \in \mathbb{N}$ , there is a one-to-one correspondence between the  $i$ -th  $k$ -soliton in  $\eta$  and the  $i$ -th  $(k - \ell)$ -soliton in  $\Psi_\ell(\eta)$ .

**Lemma 5.3.** *Assume that  $\eta \in \Omega$ . For any  $k, \ell \in \mathbb{N}$ ,  $k > \ell$ ,  $h \in \mathbb{Z}_{\geq 0}$  and  $i \in \mathbb{Z}$ , we have*

$$\begin{aligned} \xi_{\ell+h}(\eta, s_{k+h}(\eta, i)) &= \xi_\ell(\Psi_h(\eta), s_k(\Psi_h(\eta), i)) \\ &= s_{k-\ell}(\Psi_{\ell+h}(\eta), i) - s_\infty(\Psi_{\ell+h}(\eta), 0). \end{aligned} \quad (5.7)$$

In particular, if  $s_{k+h}(\eta, i) = X_{k+h}^{(j)}(\eta, 0)$  for some  $j \in \mathbb{Z}$ , then

$$s_{k-\ell}(\Psi_{\ell+h}(\eta), i) = X_{k-\ell}^{(j)}(\Psi_{\ell+h}(\eta), 0), \quad (5.8)$$

and thus

$$\begin{aligned} \xi_{\ell+h}(\eta, X_{k+h}^{(j)}(\eta, 0)) &= \xi_\ell(\Psi_h(\eta), X_k^{(j)}(\Psi_h(\eta), 0)) \\ &= X_{k-\ell}^{(j)}(\Psi_{\ell+h}(\eta), 0) - s_\infty(\Psi_{\ell+h}(\eta), 0) \\ &= X_{k-\ell}^{(j)}(\Psi_{\ell+h}(\tilde{\eta}), 0). \end{aligned} \quad (5.9)$$

*Proof.* First we note that (5.7) is a direct consequence of (5.2), (5.6) and [S, (4.16)].

Next we will show (5.8). From the assumption we get

$$\sum_{x=s_{k+h}(\eta, i)+1}^{s_{k+h}(\eta, i+1)} (\eta_{k+h}^\uparrow(x) - \eta_{k+h+1}^\uparrow(x)) = \zeta_{k+h}(\eta, i) > 0.$$

Since for any  $i \in \mathbb{Z}$ ,

$$s_{\ell+h}(\eta, \xi_{\ell+h}(\eta, s_{k+h}(\eta, i))) = s_{k+h}(\eta, i),$$

we obtain

$$\begin{aligned} &\sum_{x=s_{k+h}(\eta, i)+1}^{s_{k+h}(\eta, i+1)} (\eta_{k+h}^\uparrow(x) - \eta_{k+h+1}^\uparrow(x)) \\ &= \sum_{x=\xi_{\ell+h}(\eta, s_{k+h}(\eta, i))+1}^{\xi_{\ell+h}(\eta, s_{k+h}(\eta, i+1))} (\eta_{k+h}^\uparrow(s_{\ell+h}(\eta, x)) - \eta_{k+h+1}^\uparrow(s_{\ell+h}(\eta, x))) \\ &= \sum_{x=\xi_{\ell+h}(\eta, s_{k+h}(\eta, i))+1}^{\xi_{\ell+h}(\eta, s_{k+h}(\eta, i+1)) - \xi_{\ell+h}(\eta, 0)} (\Psi_{\ell+h}(\eta)_{k-\ell}^\uparrow(x) - \Psi_{\ell+h}(\eta)_{k-\ell+1}^\uparrow(x)) \\ &= \sum_{x=s_{k-\ell}(\Psi_{\ell+h}(\eta), i)+1}^{s_{k-\ell}(\Psi_{\ell+h}(\eta), i+1)} (\Psi_{\ell+h}(\eta)_{k-\ell}^\uparrow(x) - \Psi_{\ell+h}(\eta)_{k-\ell+1}^\uparrow(x)) \\ &= \zeta_{k-\ell}(\Psi_{\ell+h}(\eta), i) > 0. \end{aligned}$$

Hence there is a  $(k - \ell)$ -soliton with volume at site  $s_{k-\ell}(\Psi_{\ell+h}(\eta), i)$  in  $\Psi_{\ell+h}(\eta)$ , and thus there exists some  $\tilde{j} \in \mathbb{Z}$  such that

$$s_{k-\ell}(\Psi_{\ell+h}(\eta), i) = X_{k-\ell}^{(\tilde{j})}(\Psi_{\ell+h}(\eta), 0).$$

Now we show  $j = \tilde{j}$ . For the case  $i \geq 0$ , from (5.3),

$$\begin{aligned} j &= \left| \{0 \leq i' \leq i ; \zeta_{k-\ell}(\Psi_{\ell+h}(\eta), i') > 0\} \right| \\ &= \left| \{0 \leq i' \leq i-1 ; \zeta_{k+\ell}(\eta, i') > 0\} \right| \\ &= \tilde{j}. \end{aligned}$$

For the case  $i < 0$ , from (5.3),

$$\begin{aligned} j-1 &= \left| \{i \leq i' \leq -1 ; \zeta_{k-\ell}(\Psi_{\ell+h}(\eta), i') > 0\} \right| \\ &= \left| \{0 \leq i' \leq i-1 ; \zeta_{k+\ell}(\eta, i') > 0\} \right| \\ &= \tilde{j} - 1. \end{aligned}$$

Thus we have  $j = \tilde{j}$ . Hence we have (5.8). □

Thanks to Lemma 5.3, we have the following.

**Lemma 5.4.** *For any  $k, \ell, h \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{N}$ , we have*

$$N_{k+h, \ell+h}^{(i)}(\eta, n) = N_{k, \ell}^{(i)}(\Psi_h(\eta), n), \quad (5.10)$$

$$M_{k+h}^{(i)}(\eta, n) = M_k^{(i)}(\Psi_h(\eta), n). \quad (5.11)$$

In particular, (5.11) implies

$$Y_1^{(i)}(\Psi_{k-1}(\eta), n) = n - M_k^{(i)}(\eta, n). \quad (5.12)$$

*Proof.* We use induction for  $n \in \mathbb{N}$ . First we consider the case  $n = 1$ . For (5.11) with  $n = 1$ , from (5.9), we see that the  $i$ -th  $k+h$ -soliton in  $\eta$  is not free at time 0 if and only if the  $i$ -th  $k$ -soliton in  $\Psi_h(\eta)$  is not free at time 0. Hence we have

$$M_{k+h}^{(i)}(\eta, 1) = M_k^{(i)}(\Psi_h(\eta), 1).$$

Next we show (5.10) with  $n = 1$ . We fix  $i \in \mathbb{Z}$  and  $k, \ell, h \in \mathbb{N}$  such that  $k > \ell$ . Then there exists some  $j \in \mathbb{Z}$  such that

$$\xi_{k+h-1}(\eta, X_{k+h}^{(i)}(\eta, 0)) = j.$$

In other words, we have

$$X_{k+h}^{(i)}(\eta, 0) = s_{k+h-1}(\eta, j).$$

In this case, we also have

$$H_{k+h}(\gamma_{k+h}^i(n-1)) = s_{k+h-1}(\eta, j+1).$$

We observe that from (3.3),

$$N_{k+h,\ell+h}^{(i)}(\eta, 1) = \begin{cases} \sum_{z=\xi_{\ell+h}(\eta, s_{k+h-1}(\eta, j+1))}^{\xi_{\ell+h}(\eta, s_{k+h-1}(\eta, j+1))} \zeta_{\ell+h}(\eta, z) & \text{if the } i\text{-th } k + \ell\text{-soliton is free,} \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, from Lemma 5.3, we get

$$\begin{aligned} \xi_{\ell+h}(\eta, X_{k+h}^{(i)}(\eta, 0)) &= \xi_{\ell}(\Psi_h(\eta), X_k^{(i)}(\Psi_h(\eta), 0)), \\ \xi_{\ell+h}(\eta, s_{k+h-1}(\eta, j+1)) &= \xi_{\ell}(\Psi_h(\eta), s_{k-1}(\Psi_h(\eta), j+1)), \end{aligned}$$

and

$$s_{k-1}(\Psi_h(\eta), j+1) = H_k(\gamma_k^{(i)}(\Psi_h(\eta), 1)).$$

Hence if the  $i$ -th  $k + \ell$ -soliton is free, we obtain

$$\begin{aligned} N_{k+h,\ell+h}^{(i)}(\eta, 1) &= \sum_{z=\xi_{\ell+h}(\eta, X_{k+h}^{(i)}(\eta))}^{\xi_{\ell+h}(\eta, s_{k+h-1}(\eta, j+1))} \zeta_{\ell+h}(\eta, z) \\ &= \sum_{\xi_{\ell}(\Psi_h(\eta), X_k^{(i)}(\Psi_h(\eta)))}^{\xi_{\ell}(\Psi_h(\eta), H_k(\gamma_k^{(i)}(\Psi_h(\eta), 1)))} \zeta_{\ell}(\Psi_h(\eta), z) \\ &= N_{k,\ell}^{(i)}(\Psi_h(\eta), 1). \end{aligned}$$

Now we assume that (5.10) and (5.11) hold up to  $n \in \mathbb{N}$ . Then,

$$\begin{aligned} N_{k+h,\ell+h}^{(i)}(\eta, n+1) &= N_{k+h,\ell+h}^{(i)}(\eta, n+1) - N_{k+h,\ell+h}^{(i)}(\eta, n) + N_{k+h,\ell+h}^{(i)}(\eta, n+1) \\ &= N_{k+h,\ell+h}^{(j)}(T^n \eta, 1) + N_{k,\ell}^{(i)}(\Psi_h(\eta), n) \\ &= N_{k,\ell}^{(j)}(\Psi_h(T^n \eta), 1) + N_{k,\ell}^{(i)}(\Psi_h(\eta), n), \end{aligned}$$

where  $j = j(\eta, n)$  is uniquely determined via

$$X_{k+h}^{(j)}(T^n \eta, 0) = X_{k+h}^{(i)}(\eta, n),$$

i.e.,  $j$  is the number assigned to the  $k$ -soliton at  $X_{k+h}^{(i)}(\eta, n)$  in  $T^n \eta$ . From [S, Proposition 4.4], we have

$$\Psi_h(T^n \eta) = \tau_{\sum_{s=1}^n \sum_{m=1}^h r(\Psi_{m-1}(T^s \eta), 0)} T^n \Psi_h(\eta),$$

and in particular, we get

$$X_k^{(j)}(\Psi_h(T^n \eta), 0) = X_k^{(i)}(\Psi_h(\eta), n) - \sum_{s=1}^n \sum_{m=1}^h r(\Psi_{m-1}(T^s \eta), 0).$$



Since the number of  $\ell$ -solitons overtaken by a tagged  $k$ -soliton from time 0 to 1 is conserved by constant spatial shift, we have

$$N_{k,\ell}^{(j)}(\Psi_h(T^n\eta), 1) = N_{k,\ell}^{(i)}(\Psi_h(\eta), n+1) - N_{k,\ell}^{(i)}(\Psi_h(\eta), n),$$

and thus (5.10) holds for  $n+1$ . By using the same argument, we can also show that (5.11) holds for  $n+1$ .  $\square$

Now we will derive some estimates for  $N_{k,\ell}, M_{k,\ell}, M_k$  by using  $\zeta$  and the  $k$ -skip map. A key observation is that from Lemma 5.3, if we apply the  $\ell$ -skip map to  $\eta$ , then  $k$ -solitons in  $\eta$  with  $k > \ell$  become  $(k-\ell)$ -solitons in  $\Psi_\ell(\eta)$ , and  $\ell$ -solitons in  $\eta$  become certain sites in  $\Psi_\ell(\eta)$ . Thus we see that a  $k$ -soliton overtaking  $\ell$ -solitons in  $\eta$  corresponds to a  $(k-\ell)$ -soliton passing a certain site in  $\Psi_\ell(\eta)$ . We note that different solitons may correspond to the same site, and if  $\xi_\ell(X_\ell^{(j)}) = x$  for some  $j \in \mathbb{Z}$  and  $x \in \mathbb{Z}$ , then the site  $x$  of  $\Psi_\ell(\tilde{\eta})$  corresponds to  $\zeta_\ell(x)$   $\ell$ -solitons. Hence, to find the total number of  $\ell$ -solitons overtaken by the  $i$ -th  $k$ -soliton in  $\eta$ , we only need to calculate the sum of  $\zeta_\ell(x)$  on  $x \in [X_{k-\ell}^{(i)}(\Psi_\ell(\eta))+1, X_{k-\ell}^{(i)}(\Psi_\ell(\eta), n)]$ . Conversely,  $M_k^{(i)}(n)$  can be calculated by counting the number of solitons passing through the site in  $\Psi_k(\eta)$  corresponding to the  $i$ -th  $k$ -soliton with volume.

**Lemma 5.5.** *Assume that  $\eta \in \Omega$ . Then, for any  $k, \ell \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we have*

$$\sum_{m=1}^n N_{k,\ell}^{(i)}(\eta, m) = \sum_{j=X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), n)} \zeta_\ell(\tilde{\eta}, j), \quad (5.13)$$

and

$$M_k^{(i)}(\eta, n) = \sum_{m=0}^{n-1} (1 - r(T^m \Psi_k(\tilde{\eta}), J_k(\tilde{\eta}, i))), \quad (5.14)$$

where

$$J_k(\eta, i) := \begin{cases} \min \left\{ j \in \mathbb{Z}_{\geq 0} ; \sum_{h=0}^j \mathbf{1}_{\{\zeta_k(\eta, h) > 0\}} = i \right\} & i \geq 1, \\ -\min \left\{ j \in \mathbb{Z}_{\geq 0} ; \sum_{h=-j}^{-1} \mathbf{1}_{\{\zeta_k(\eta, h) > 0\}} = -i + 1 \right\} & i \leq 0. \end{cases} \quad (5.15)$$

In addition,

$$\sum_{j=J_\ell(\eta, \sigma_{k,\ell}^{(i)}(\eta, n))}^{J_\ell(\eta, \sigma_{k,\ell}^{(i)}(\eta, 0))-1} \zeta_\ell(\eta, j) \leq M_{k,\ell}^{(i)}(\eta, n) \leq \sum_{j=J_\ell(\eta, \sigma_{k,\ell}^{(i)}(\eta, n))-1}^{J_\ell(\eta, \sigma_{k,\ell}^{(i)}(\eta, 0))} \zeta_\ell(\eta, j), \quad (5.16)$$

where

$$\sigma_{k,\ell}^{(i)}(\eta, n) := \inf \left\{ j \in \mathbb{Z} ; X_{\ell-k}^{(j)}(\Psi_k(\tilde{\eta}), n) \geq J_k(\tilde{\eta}, i) \right\}. \quad (5.17)$$

*Proof.* First we note that thanks to Lemma 3.1 and (5.6), without loss of generality we can assume that  $s_\infty(0) = 0$ .

First we prove (5.13). Observe that  $i$ -th  $k$ -soliton overtakes the  $j$ -th  $\ell$ -soliton up to time  $n$  if and only if

$$\xi_\ell\left(\eta, X_k^{(i)}(\eta, 0)\right) + 1 \leq \xi_\ell\left(\eta, X_\ell^{(j)}(\eta, 0)\right),$$

and

$$\xi_\ell\left(T^n\eta, X_k^{(i)}(\eta, n)\right) \geq \xi_\ell\left(T^n\eta, X_\ell^{(j)}(\eta, n)\right).$$

Hence, from Lemma 3.5, 5.3 and 5.4, we get

$$\begin{aligned} X_{k-\ell}^{(i)}(\Psi_\ell(\eta), 0) + 1 &= \xi_\ell\left(\eta, X_k^{(i)}(\eta, 0)\right) + 1 \\ &\leq \xi_\ell\left(\eta, X_\ell^{(j)}(\eta, 0)\right) \\ &\leq \xi_\ell\left(\eta, X_k^{(i)}(\eta, 0)\right) + (k-\ell)(n - M_k(\eta, n)) + 2 \sum_{m=1}^n \sum_{h=\ell+1}^{k-1} N_{k,h}^{(i)}(\eta, m) \\ &= X_{k-\ell}^{(i)}(\Psi_\ell(\eta), n). \end{aligned}$$

Hence we have (5.13).

Next we show (5.14) for  $k < \ell$ . We observe that the  $i$ -th  $k$ -soliton is free at time  $n$  if and only if the site,

$$s_k\left(T^n\eta, \xi_k\left(T^n\eta, X_k^{(i)}(\eta, n)\right)\right),$$

is a record in  $T^n\eta$ . In addition, the function  $\xi_k(T^n\eta, \cdot)$  increases at each record in  $T^n\eta$ . Hence, the  $i$ -th  $k$ -soliton is free at time  $n$  if and only if

$$\xi_k\left(T^n\eta, X_k^{(i)}(\eta, n)\right) \notin \left[\xi_k\left(T^n\eta, X_\ell^{(j)}(\eta, n)\right) + 1, \xi_k\left(T^n\eta, \bar{X}_\ell^{(j)}(\eta, n)\right)\right],$$

for any  $j \in \mathbb{Z}$  and  $\ell > k$ , where

$$\bar{X}_\ell^{(j)}(\eta, n) := \max\left\{x \in \mathbb{Z}; x \in \text{Con}\left(\gamma_\ell^{(j)}(n)\right)\right\}.$$

On the other hand, from Lemmas 3.5 and 5.3, we have

$$\xi_k\left(T^n\eta, X_k^{(i)}(\eta, n)\right) - \xi_k\left(\eta, X_k^{(i)}(\eta, 0)\right) = \sum_{m=1}^n (k + o_k(T^{m-1}\eta)),$$

and

$$\begin{aligned} &\xi_k\left(T^n\eta, X_\ell^{(j)}(\eta, n)\right) - \xi_k\left(\eta, X_\ell^{(j)}(\eta, 0)\right) \\ &= (k-\ell)\left(n - M_\ell^{(j)}(\eta, n)\right) + 2 \sum_{h=k+1}^{\ell-1} (h-k)N_{\ell,h}^{(i)}(\eta, n) \\ &\quad + \sum_{m=1}^n (k + o_k(T^{m-1}\eta)). \end{aligned}$$

By (5.9), (5.10) and (5.11) we get

$$\begin{aligned}
& \xi_k \left( T^n \eta, X_\ell^{(j)}(\eta, n) \right) \\
&= X_{\ell-k}^{(j)}(\Psi_k(\eta), 0) + (k - \ell) \left( n - M_{\ell-k}^{(j)}(\Psi_k(\eta), n) \right) \\
&\quad + 2 \sum_{h=1}^{\ell-k-1} h N_{\ell-k, h}^{(i)}(\Psi_k(\eta), n) + \sum_{m=1}^n (k + o_k(T^{m-1}\eta)) \\
&= X_{\ell-k}^{(j)}(\Psi_k(\eta), n) + \sum_{m=1}^n (k + o_k(T^{m-1}\eta)).
\end{aligned}$$

Now we consider an expression of  $\xi_k \left( T^n \eta, \bar{X}_\ell^{(j)}(\eta, n) \right)$ . Observe that there exists  $j' = j'(n) \in \mathbb{Z}$  such that

$$X_\ell^{(j')} (T^n \eta, 0) = X_\ell^{(j)} (\eta, n).$$

Since from Remark 3.6, the volume of solitons are conserved in time, we have  $|Con(\gamma_\ell^{(j')} (T^n \eta, 0))| = |Con(\gamma_\ell^{(j)} (\eta, n))|$ . In particular,  $\bar{X}_\ell^{(j')} (T^n \eta, 0) = \bar{X}_\ell^{(j)} (\eta, n)$ . Since there are no  $h$ -solitons with  $h \geq \ell + 1$  in the interval  $[X_\ell^{(j')} (T^n \eta, 0), \bar{X}_\ell^{(j')} (T^n \eta, 0)]$  at time  $n$ , from Remark 3.2, the difference of  $\xi_k \left( T^n \eta, \bar{X}_\ell^{(j')} (T^n \eta, 0) \right) - \xi_k \left( T^n \eta, X_\ell^{(j')} (T^n \eta, 0) \right)$  is equal to the total number of  $h$ -th head and tail with  $h \geq k + 1$  in  $[X_\ell^{(j')} (T^n \eta, 0), \bar{X}_\ell^{(j')} (T^n \eta, 0)]$ , i.e.,

$$\begin{aligned}
& \xi_k \left( T^n \eta, \bar{X}_\ell^{(j)}(\eta, n) \right) - \xi_k \left( T^n \eta, X_\ell^{(j)}(\eta, n) \right) \\
&= \xi_k \left( T^n \eta, \bar{X}_\ell^{(j')} (T^n \eta, 0) \right) - \xi_k \left( T^n \eta, X_\ell^{(j')} (T^n \eta, 0) \right) \\
&= 2 \sum_{h=k+1}^{\ell} \xi_h \left( T^n \eta, \bar{X}_\ell^{(j')} (T^n \eta, 0) \right) \\
&\quad \sum_{x=\xi_h(T^n \eta, X_\ell^{(j')} (T^n \eta, 0))+1} (h - k) \zeta_h (T^n \eta, x).
\end{aligned}$$

Then from (5.3) and (5.8), we get

$$\begin{aligned}
& 2 \sum_{h=k+1}^{\ell} \xi_h \left( T^n \eta, \bar{X}_\ell^{(j')} (T^n \eta, 0) \right) \\
&\quad \sum_{x=\xi_h(T^n \eta, X_\ell^{(j')} (T^n \eta, 0))+1} (h - k) \zeta_h (T^n \eta, x) \\
&= 2 \sum_{h=1}^{\ell-k} \xi_{h-k} \left( \Psi_k(T^n \eta), \bar{X}_\ell^{(j')}(\Psi_k(T^n \eta), 0) \right) \\
&\quad \sum_{x=\xi_{h-k}(\Psi_k(T^n \eta), X_{\ell-k}^{(j')}(\Psi_k(T^n \eta), 0))+1} h \zeta_h (\Psi_k(T^n \eta), x) \\
&= \bar{X}_{\ell-k}^{(j')} (\Psi_k(T^n \eta), 0) - X_{\ell-k}^{(j')} (\Psi_k(T^n \eta), 0) \\
&= \bar{X}_{\ell-k}^{(j)} (\Psi_k(\eta), n) - X_{\ell-k}^{(j)} (\Psi_k(\eta), n).
\end{aligned}$$

From the above we have

$$\xi_k \left( T^n \eta, \bar{X}_\ell^{(j)}(\eta, n) \right) = \bar{X}_{\ell-k}^{(j)}(\Psi_k(\eta), n) + \sum_{m=1}^n (k + o_k(T^{m-1}\eta)).$$

By combining the above, we see that  $i$ -th  $k$ -soliton is free at time  $n$  if and only if

$$\xi_k \left( \eta, X_k^{(i)}(\eta, 0) \right) \notin \left[ X_{\ell-k}^{(j)}(\Psi_k(\eta), n) + 1, \bar{X}_{\ell-k}^{(j)}(\Psi_k(\eta), n) \right],$$

for any  $j \in \mathbb{Z}$  and  $\ell > k$ , and this is equivalent to

$$r \left( T^n \Psi_k(\eta), \xi_k \left( \eta, X_k^{(i)}(\eta, 0) \right) \right) = 1.$$

Since

$$\xi_\ell \left( \eta, X_\ell^{(j)}(\eta, 0) \right) = J_\ell(\eta, j), \quad (5.18)$$

we have (5.14).

Finally we show (5.16). By the same computation as above, we see that  $i$ -th  $k$ -soliton will be overtaken by the  $j$ -th  $\ell$ -soliton up to time  $n$  if and only if

$$X_{\ell-k}^{(j)}(\Psi_k(\eta), 0) + 1 \leq \xi_k \left( \eta, X_k^{(i)}(\eta, 0) \right) \leq X_{\ell-k}^{(j)}(\Psi_k(\eta), n).$$

On the other hand, we see that

$$X_{\ell-k}^{(\sigma_{k,\ell}^{(i)}(\eta, 0)-1)}(\Psi_k(\eta), 0) < \xi_k \left( \eta, X_k^{(i)}(\eta, 0) \right) \leq X_{\ell-k}^{(\sigma_{k,\ell}^{(i)}(\eta, 0))}(\Psi_k(\eta)),$$

and

$$X_{\ell-k}^{(\sigma_{k,\ell}^{(i)}(\eta, n)-1)}(\Psi_k(\eta), n) < \xi_k \left( \eta, X_k^{(i)}(\eta, 0) \right) \leq X_{\ell-k}^{(\sigma_{k,\ell}^{(i)}(\eta, n))}(\Psi_k(\eta), n).$$

Hence, if  $\sigma_{k,\ell}^{(i)}(\eta, n) \leq j \leq \sigma_{k,\ell}^{(i)}(\eta, 0) - 1$ , then the  $j$ -th  $\ell$ -soliton will overtake the  $i$ -th  $k$ -soliton up to time  $n$ . Now we observe that

$$\sum_{J_\ell(\eta, j)+1}^{J_\ell(\eta, j+1)-1} \zeta_\ell(\eta, j) = 0,$$

for any  $\ell \in \mathbb{N}$  and  $j \in \mathbb{N}$ . From the above and (5.18), we have (5.16).  $\square$

The following representation of  $Y_k^{(i)}(n)$  is a key to show the main results. As we will see later in Proposition 5.9, the representation of  $Y_k^{(i)}(n)$  in Lemma 5.6 is an orthogonal decomposition of  $Y_k^{(i)}(n)$ , unlike the original formula (3.2).

**Lemma 5.6.** *For any  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$ ,  $0 \leq \ell \leq k-1$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we have*

$$\begin{aligned} Y_{k-\ell}^{(i)}(\Psi_\ell(\eta), n) &= \frac{v_{k-\ell}^{\text{eff}}(\theta^\ell \mathbf{q})}{v_1^{\text{eff}}(\theta^{k-1} \mathbf{q})} \left( n - M_k^{(i)}(\eta, n) \right) \\ &\quad + 2 \sum_{h=\ell+1}^{k-1} \frac{v_{h-\ell}^{\text{eff}}(\theta^\ell \mathbf{q})}{v_1^{\text{eff}}(\theta^{h-1} \mathbf{q})} \sum_{j=X_{k-h}^{(i)}(\Psi_h(\tilde{\eta}), 0)+1}^{X_{k-h}^{(i)}(\Psi_h(\tilde{\eta}), n)} (\zeta_h(j) - \alpha_h(\mathbf{q})), \end{aligned}$$

with convention  $\sum_{\ell=1}^0 = 0$ . In particular, we have

$$\begin{aligned} Y_k^{(i)}(\eta, n) &= \frac{v_k^{\text{eff}}(\mathbf{q})}{v_1^{\text{eff}}(\theta^{k-1} \mathbf{q})} \left( n - M_k^{(i)}(\eta, n) \right) \\ &\quad + 2 \sum_{h=1}^{k-1} \frac{v_h^{\text{eff}}(\mathbf{q})}{v_1^{\text{eff}}(\theta^{h-1} \mathbf{q})} \sum_{j=X_{k-h}^{(i)}(\Psi_h(\tilde{\eta}), 0)+1}^{X_{k-h}^{(i)}(\Psi_h(\tilde{\eta}), n)} (\zeta_h(j) - \alpha_h(\mathbf{q})). \end{aligned} \quad (5.19)$$

*Proof of Lemma 5.6.* First we note that from Lemma 3.1, for any  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we have  $Y_k^{(i)}(\eta, n) = Y_k^{(i)}(\tilde{\eta}, n)$ . Hence, without loss of generality, we can assume that  $s_\infty(0) = 0$ .

We fix  $\mathbf{q} \in \mathcal{Q}$ . From (5.2), (5.3), (5.5), Lemmas 5.4 and 5.5, for any  $k \in \mathbb{N}$ ,  $0 \leq \ell < k$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we get

$$\begin{aligned} &Y_{k-\ell}^{(i)}(\Psi_\ell(\eta), n) \\ &= (k-\ell) \left( n - M_{k-\ell}^{(i)}(\Psi_\ell(\eta), n) \right) + 2 \sum_{h=1}^{k-\ell-1} h \sum_{j=X_{k-\ell-h}^{(i)}(\Psi_{\ell+h}(\eta), 0)+1}^{X_{k-\ell-h}^{(i)}(\Psi_{\ell+h}(\eta), n)} \zeta_h(\Psi_\ell(\eta), j) \\ &= (k-\ell) \left( n - M_k^{(i)}(n) \right) + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) \sum_{j=X_{k-h}^{(i)}(\Psi_h(\eta), 0)+1}^{X_{k-h}^{(i)}(\Psi_h(\eta), n)} (\zeta_h(j) - \alpha_h(\mathbf{q})) \\ &\quad + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) \alpha_h(\mathbf{q}) Y_{k-h}^{(i)}(\Psi_h(\eta), n). \end{aligned}$$

Hence, if we write

$$\begin{aligned} A_{k-\ell, \ell} &:= Y_{k-\ell}^{(i)}(\Psi_\ell(\eta), n), \\ B_\ell &:= \sum_{j=X_{k-\ell}^{(i)}(\Psi_\ell(\eta), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_\ell(\eta), n)} (\zeta_\ell(j) - \alpha_\ell(\mathbf{q})), \\ C &:= n - M_k^{(i)}(n), \end{aligned}$$

then for any  $0 \leq \ell \leq k-1$  we have the following system.

$$A_{k-\ell, \ell} = (k-\ell)C + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) \alpha_h(\mathbf{q}) A_{k-h, h} + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) B_h, \quad (5.20)$$

with convention  $\sum_{h=k}^{k-1} = 0$ . By using (5.20) recursively starting from  $\ell = k-1$  and then  $\ell = k-2$ , and so on to  $\ell = 0$ , we can represent  $A_{k-\ell,\ell}$  as a linear combination of  $C$  and  $B_h$ ,  $\ell+1 \leq h \leq k-1$ . Hence, for any  $k \in \mathbb{N}$  and  $0 \leq \ell \leq k-1$ , there exist some positive constants  $b_{k,\ell,h}(\mathbf{q})$ ,  $\ell+1 \leq h \leq k-1$  and  $c_{k,\ell}(\mathbf{q})$  such that

$$A_{k-\ell,\ell} = c_{k,\ell}(\mathbf{q})C + \sum_{h=\ell+1}^{k-1} b_{k,\ell,h}(\mathbf{q})B_h,$$

with convention  $\sum_{h=k}^{k-1} = 0$ . In the rest of the proof, we will show that

$$b_{k,\ell,h}(\mathbf{q}) = \frac{2v_{h-\ell}^{\text{eff}}(\theta^\ell \mathbf{q})}{v_1^{\text{eff}}(\theta^{h-1} \mathbf{q})}, \quad (5.21)$$

for any  $k \geq 2$ ,  $0 \leq \ell \leq h-1 \leq k-2$ , and

$$c_{k,\ell}(\mathbf{q}) = \frac{v_{k-\ell}^{\text{eff}}(\theta^\ell \mathbf{q})}{v_1^{\text{eff}}(\theta^{\ell-1} \mathbf{q})}, \quad (5.22)$$

for any  $k \in \mathbb{N}$  and  $0 \leq \ell \leq k-1$ . By using (5.20), we have

$$\begin{aligned} A_{k-\ell,\ell} &= (k-\ell)C + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) \alpha_h(\mathbf{q}) A_{k-h,h} + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) B_h \\ &= (k-\ell)C + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) \alpha_h(\mathbf{q}) \left( c_{k,h}(\mathbf{q})C + \sum_{h'=h+1}^{k-1} b_{k,h,h'}(\mathbf{q})B_{h'} \right) \\ &\quad + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) B_h \\ &= \left( k-\ell + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) \alpha_h(\mathbf{q}) c_{k,h} \right) C \\ &\quad + 2B_{\ell+1} + 2 \sum_{h=\ell+2}^{k-1} \left( h-\ell + \sum_{h'=\ell+1}^{h-1} (h'-\ell) \alpha_{h'}(\mathbf{q}) b_{k,h',h} \right) B_h. \end{aligned}$$

Hence we have

$$b_{k,\ell,h}(\mathbf{q}) = 2(h-\ell) + 2 \sum_{h'=\ell+1}^{h-1} (h'-\ell) \alpha_{h'}(\mathbf{q}) b_{k,h',h}(\mathbf{q}), \quad (5.23)$$

and

$$c_{k,\ell} = k-\ell + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) \alpha_h(\mathbf{q}) c_{k,h}, \quad (5.24)$$

with convention  $\sum_{h'=\ell+1}^{\ell} = 0$ . On the other hand, from (4.11) and (5.5), we have

$$\begin{aligned} v_{k-\ell}^{\text{eff}}(\theta^{\ell}\mathbf{q}) &= (k-\ell)v_1^{\text{eff}}(\theta^{k-1}\mathbf{q}) + 2\sum_{h=1}^{k-\ell-1} h\alpha_{\ell+h}(\mathbf{q})v_{k-\ell-h}^{\text{eff}}(\theta^{\ell+h}\mathbf{q}) \\ &= (k-\ell)v_1^{\text{eff}}(\theta^{k-1}\mathbf{q}) + 2\sum_{h=\ell+1}^{k-1} (h-\ell)\alpha_h(\mathbf{q})v_{k-h}^{\text{eff}}(\theta^h\mathbf{q}). \end{aligned} \quad (5.25)$$

By comparing (5.23), (5.25), we see that for fixed  $k \in \mathbb{N}$ , both the sequences  $2v_{k-\ell}^{\text{eff}}(\theta^{\ell}\mathbf{q})v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})^{-1}$  and  $b_{h,\ell,k}(\mathbf{q})$  satisfy the same inductive system for  $0 \leq \ell \leq k-1$ , and these two sequence have the same value 2 with  $\ell = k-1$ . Hence we have (5.21). By the same argument for (5.24) and (5.25), we also get (5.22). Therefore Lemma 5.6 is proved.  $\square$

In the rest of this subsection, we note some consequences from Lemmas 5.5 and 5.6 and some materials in its proof. Before describing those, we consider the following remark.

**Remark 5.7.** From (5.3) and the bijectivity of  $\zeta$ ,  $\Psi_k(\tilde{\eta})$  can be described as a function of  $(\zeta_{\ell}(i))_{\ell \geq k+1, i \in \mathbb{Z}}$ . In particular, for any  $k \in \mathbb{N}$ ,  $\Psi_k(\tilde{\eta})$  and  $(\zeta_{\ell}(i))_{\ell \leq k, i \in \mathbb{Z}}$  are independent under  $\nu_{\mathbf{q}}$ ,  $\mathbf{q} \in \mathcal{Q}$ .

First we prove the exponential bound for  $Y_k^{(i)}(n)$ . To describe the result, we prepare some functions. For any  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$  and  $\lambda \in \mathbb{R}$ , we define

$$u_{\mathbf{q},k}(\lambda) := \log(\mathbb{E}_{\nu_{\mathbf{q}}}[\exp(2\lambda\zeta_k(0))]) = \begin{cases} \infty & \lambda \geq \frac{\log q_k^{-1}}{2}, \\ \log\left(\frac{1-q_k}{1-e^{2\lambda}q_k}\right) & \lambda < \frac{\log q_k^{-1}}{2}. \end{cases}$$

By using  $u_{\mathbf{q},k}(\lambda)$ , we inductively define  $U_{\mathbf{q},k}(\lambda)$  as  $U_{\mathbf{q},1}(\lambda) := \lambda$ , and

$$U_{\mathbf{q},k}(\lambda) := k\lambda + \sum_{\ell=1}^{k-1} (k-\ell)u_{\mathbf{q},\ell}(U_{\mathbf{q},\ell}(\lambda)), \quad (5.26)$$

for any  $k \geq 2$ . We note that there exists some  $\delta_{\mathbf{q},k} > 0$  depending on  $\mathbf{q}, k$  such that  $(-\infty, \delta_{\mathbf{q},k}) \subset \{\lambda \in \mathbb{R} ; U_{\mathbf{q},k}(\lambda) < \infty\}$ . In addition,  $U_{\mathbf{q},k}(\lambda)$  is a smooth monotone convex function on  $(-\infty, \delta_{\mathbf{q},k})$ .

**Lemma 5.8.** For any  $\mathbf{q} \in \mathcal{Q}$  and  $k \in \mathbb{N}$  with  $q_k > 0$ , there exists  $\lambda_{\mathbf{q},k} \in (0, \infty]$  depending on  $\mathbf{q}, k$  such that for any  $\lambda < \lambda_{\mathbf{q},k}$ ,  $i \in \mathbb{Z}$  and  $n \geq 0$ ,  $\mathbb{E}_{\nu_{\mathbf{q}}}[\exp(\lambda Y_k^{(i)}(n))] < \infty$ . Moreover, if  $\lambda < \min\{\delta_{\mathbf{q},k}, \lambda_{\mathbf{q},k}\}$ , then for any  $i \in \mathbb{Z}$  and  $n \geq 0$ , we have

$$\mathbb{E}_{\nu_{\mathbf{q}}}[\exp(\lambda Y_k^{(i)}(n))] = \mathbb{E}_{\nu_{\mathbf{q}}}[\exp(U_{\mathbf{q},k}(\lambda)(n - M_k^{(i)}(n)))] \quad (5.27)$$

*Proof.* Since  $Y_k^{(i)}(n)$  is non-negative, it suffices to consider the exponential moment of  $Y_k^{(i)}(n)$  with  $\lambda > 0$ .

We use induction on  $k \in \mathbb{N}$ . For the case  $k = 1$ , since  $|Y_1^{(i)}(n)| \leq n$ , from (5.12), we have

$$\mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda Y_1^{(i)}(n) \right) \right] = \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda n - M_1^{(i)}(n) \right) \right] < \infty,$$

for any  $\lambda \in \mathbb{R}$ .

Now we assume that the claim of this lemma holds up to  $k - 1$ . Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $q_k > 0$ . By Lemmas 3.1, 5.2, (5.12) and (5.13), we have

$$Y_k^{(i)}(\eta, n) = kY_1^{(i)}(\Psi_{k-1}(\tilde{\eta}), n) + 2 \sum_{\ell=1}^{k-1} \ell \sum_{j=X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), n)} \zeta_\ell(\tilde{\eta}, j).$$

Then from Remark 5.7, Fubini's theorem and (5.5), we have

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda Y_k^{(i)}(n) \right) \right] \\ &= \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( k\lambda Y_1^{(i)}(\Psi_{k-1}(\tilde{\eta}), n) + 2\lambda \sum_{\ell=2}^{k-1} \ell \sum_{j=X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), n)} \zeta_\ell(\tilde{\eta}, j) \right) \right. \\ & \quad \left. \times \exp \left( u_{\mathbf{q},1}(\lambda) Y_{k-1}^{(i)}(\Psi_1(\tilde{\eta}), n) \right) \right]. \end{aligned} \quad (5.28)$$

We observe that from (5.2) and (5.3),

$$\begin{aligned} & kY_1^{(i)}(\Psi_{k-1}(\tilde{\eta}), n) + 2 \sum_{\ell=2}^{k-1} \ell \sum_{j=X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), n)} \zeta_\ell(\tilde{\eta}, j) \\ &= kY_1^{(i)}(\Psi_{k-1}(\tilde{\eta}), n) + 2 \sum_{\ell=1}^{k-2} (\ell+1) \sum_{j=X_{k-1-\ell}^{(i)}(\Psi_\ell(\Psi_1(\tilde{\eta})), 0)+1}^{X_{k-1-\ell}^{(i)}(\Psi_\ell(\Psi_1(\tilde{\eta})), n)} \zeta_\ell(\Psi_1(\tilde{\eta}), j) \\ &\leq 2Y_{k-1}^{(i)}(\Psi_1(\tilde{\eta}), n). \end{aligned}$$

From the assumption of induction, there exists some  $\lambda_{\theta_{\mathbf{q},k-1}} > 0$  such that  $\mathbb{E}_{\nu_{\theta_{\mathbf{q}}}} \left[ \exp \left( \lambda Y_k^{(i)}(n) \right) \right] < \infty$  for any  $\lambda < \lambda_{\theta_{\mathbf{q},k-1}}$ ,  $i \in \mathbb{Z}$  and  $n \geq 0$ . Hence if  $\lambda$  satisfies

$$0 < 2\lambda + u_{\mathbf{q},1}(\lambda) < \min \left\{ \lambda_{\theta_{\mathbf{q},k-1}}, \frac{\log q_1^{-1}}{2} \right\},$$

then for any  $i \in \mathbb{Z}$  and  $n \geq 0$ , we have

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda Y_k^{(i)}(n) \right) \right] \\ &\leq \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( (2\lambda + u_{\mathbf{q},1}(\lambda)) Y_{k-1}^{(i)}(\Psi_1(\tilde{\eta}), n) \right) \right] \\ &= \mathbb{E}_{\nu_{\theta_{\mathbf{q}}}} \left[ \exp \left( (2\lambda + u_{\mathbf{q},1}(\lambda)) Y_{k-1}^{(i)}(n) \right) \right] \\ &< \infty. \end{aligned}$$



Since the function  $\lambda \mapsto 2\lambda + u_{\mathbf{q},1}(\lambda)$  is strictly increasing, we see that there exists some  $\lambda_{\mathbf{q},k} > 0$  such that  $\mathbb{E}_{\nu_{\theta\mathbf{q}}} \left[ \exp \left( \lambda Y_k^{(i)}(n) \right) \right] < \infty$  for any  $0 < \lambda < \lambda_{\mathbf{q},k}$ ,  $i \in \mathbb{Z}$  and  $n \geq 0$ .

From now on we show (5.27). By Lemmas 3.1, 5.2, (5.12) and (5.13), we obtain

$$\begin{aligned} & Y_{k-1}^{(i)}(\Psi_1(\tilde{\eta}), n) \\ &= (k-1) Y_1^{(i)}(\Psi_{k-1}(\tilde{\eta}), n) + 2 \sum_{\ell=2}^{k-1} (\ell-1) \sum_{j=X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), n)} \zeta_\ell(\tilde{\eta}, j). \end{aligned}$$

By substituting the above formula of  $Y_{k-1}^{(i)}(\Psi_1(\tilde{\eta}), n)$  to (5.28), we get

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda Y_k^{(i)}(n) \right) \right] \\ &= \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( (\lambda k + u_{\mathbf{q},1}(\lambda)(k-1)) Y_1^{(i)}(\Psi_{k-1}(\tilde{\eta}), n) \right) \right. \\ & \quad \left. \times \exp \left( 2 \sum_{h=2}^{k-1} (\lambda h + u_{\mathbf{q},1}(\lambda)(h-1)) \sum_{j=X_{k-h}(\Psi_h(\tilde{\eta}), 0)+1}^{X_{k-h}(\Psi_h(\tilde{\eta}), n)} \zeta_h(j) \right) \right]. \end{aligned}$$

By repeating the above computation, we have (5.27).

From the above, this lemma has been shown.  $\square$

Next, for any  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we define  $\Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, n)$ ,  $1 \leq \ell \leq k-1$  as

$$\Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, n) := \sum_{j=X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), n)} (\zeta_\ell(\tilde{\eta}, j) - \alpha_\ell(\mathbf{q})). \quad (5.29)$$

Note that from Lemma 5.2, (5.14), Lemmas 5.6 and Remark 5.7,  $Y_k^{(i)}(n)$  can be represented as

$$\begin{aligned} Y_k^{(i)}(\eta, n) - \mathbb{E}_{\nu_{\mathbf{q}}} \left[ Y_k^{(i)}(\eta, n) \right] &= -\frac{v_k^{\text{eff}}(\mathbf{q})}{v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})} \left( M_k^{(i)}(n) - \mathbb{E}_{\nu_{\mathbf{q}}} \left[ M_k^{(i)}(n) \right] \right) \\ & \quad + 2 \sum_{\ell=1}^{k-1} \frac{v_\ell^{\text{eff}}(\mathbf{q})}{v_1^{\text{eff}}(\theta^{\ell-1}\mathbf{q})} \Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, n). \quad (5.30) \end{aligned}$$

From Lemma 3.1 and Remark 5.7, we have the following proposition.

**Proposition 5.9.** *For any  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $n \in \mathbb{Z}_{\geq 0}$ , we have*

$$\mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,\ell}^{(i)}(\mathbf{q}, n) \right] = 0, \quad (5.31)$$

and

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,\ell}^{(i)}(\mathbf{q}, n) \Delta Y_{k,\ell'}^{(i)}(\mathbf{q}, n) \right] \\ &= \begin{cases} \mathbb{E}_{\nu_{\theta^\ell \mathbf{q}}} \left[ Y_{k-\ell}^{(i)}(n) \right] \beta_\ell(\mathbf{q}) & \ell = \ell', \\ 0 & \ell \neq \ell'. \end{cases} \end{aligned} \quad (5.32)$$

In addition, for any  $1 \leq \ell \leq k-1$ , we have

$$\mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,\ell}^{(i)}(\mathbf{q}, n) M_k^{(i)}(n) \right] = 0. \quad (5.33)$$

By combining the above, we have

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| Y_k^{(i)}(n) - \mathbb{E}_{\nu_{\mathbf{q}}} \left[ Y_k^{(i)}(n) \right] \right|^2 \right] \\ &= \frac{v_k^{\text{eff}}(\mathbf{q})^2}{v_1^{\text{eff}}(\theta^{k-1}\mathbf{q})^2} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| M_k^{(i)}(n) - \mathbb{E}_{\nu_{\mathbf{q}}} \left[ M_k^{(i)}(n) \right] \right|^2 \right] \\ &\quad + 4 \sum_{\ell=1}^{k-1} \frac{v_\ell^{\text{eff}}(\mathbf{q})^2 \mathbb{E}_{\nu_{\theta^\ell \mathbf{q}}} \left[ Y_{k-\ell}^{(i)}(n) \right] \beta_\ell(\mathbf{q})}{v_1^{\text{eff}}(\theta^{\ell-1}\mathbf{q})^2}. \end{aligned}$$

*Proof of Proposition 5.9.* Since the case  $k=1$  is trivial, we consider the case  $k \geq 2$ . We fix  $i \in \mathbb{Z}$ . Since  $X_{k-\ell}^{(i)}(\Psi_\ell(\tilde{\eta}), n)$  is  $\sigma(\zeta_h; h \geq \ell+1)$ -m'ble for any  $1 \leq \ell \leq k-1$  and  $n \in \mathbb{Z}_{\geq 0}$ , from Remark 5.7 we have

$$\mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, n) \mid \sigma(\zeta_h; h \geq \ell+1) \right] = 0 \quad \nu_{\mathbf{q}}\text{-a.s.}$$

Hence we obtain (5.31) and (5.32). In addition, from (5.14),  $M_k^{(i)}(n)$  is  $\sigma(\zeta_h; h \geq k)$ -m'ble for any  $n \in \mathbb{Z}_{\geq 0}$ . Hence for any  $1 \leq \ell \leq k-1$  and  $n \in \mathbb{Z}_{\geq 0}$  we get

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, n) M_k^{(i)}(n) \mid \sigma(\zeta_h; h \geq \ell+1) \right] \\ &= M_k^{(i)}(n) \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, n) \mid \sigma(\zeta_h; h \geq \ell+1) \right] \\ &= 0 \quad \nu_{\mathbf{q}}\text{-a.s.} \end{aligned}$$

Therefore we have (5.33). □

**Remark 5.10.** *The decomposition (5.30) might be useful to consider the long-time correlations between solitons with different sizes. Actually, from Remark 5.7, for any  $\mathbf{q} \in \mathcal{Q}$ ,  $k, \ell \in \mathbb{N}$ ,  $k < \ell$ ,  $i, j \in \mathbb{N}$ , we have*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,h}^{(i)}(\mathbf{q}, n) \Delta Y_{\ell,h'}^{(j)}(\mathbf{q}, n) \right] \\ &= \begin{cases} v_{k-h-1}(\theta^{h+1}\mathbf{q}) v_1^{\text{eff}}(\theta^{k-1}\mathbf{q}) \beta_{\ell+1}(\mathbf{q}) & h = h', \\ 0 & h \neq h'. \end{cases} \end{aligned}$$

In addition, for any  $0 \leq h \leq k-2$  and  $0 \leq h' \leq \ell-2$ , we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \Delta Y_{k,h}^{(i)}(\mathbf{q}, n) Y_1^{(j)}(\Psi_{\ell-1}(\eta), n) \right] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ Y_1^{(i)}(\Psi_{k-1}(\eta), n) \Delta Y_{\ell,h'}^{(i)}(\mathbf{q}, n) \right] = 0. \end{aligned}$$

Hence, if the covariance of  $Y_1^{(i)}(\Psi_{k-1}(\eta), n)$  and  $Y_1^{(j)}(\Psi_{\ell-1}(\eta), n)$  can be computed explicitly, then one can obtain the explicit correlation between the  $i$ -th  $k$ -soliton and  $j$ -th  $\ell$ -soliton, but it does not seem to be easy to compute.

## 6. PROOF OF THEOREM 4.2

Since  $Y_k^{(i)}(n)/n$  converges to  $v_k^{\text{eff}}(\mathbf{q})$  a.s., to show the  $\mathbb{L}^p$  convergence, it is sufficient to prove that  $\left( \left| Y_k^{(i)}(n)/n \right|^p \right)_{n \in \mathbb{N}}$  is uniformly integrable, i.e.,

$$\lim_{L \rightarrow \infty} \sup_{n \in \mathbb{N}} \nu_{\mathbf{q}} \left( \frac{Y_k^{(i)}(n)}{n} > L \right) = 0,$$

and

$$\lim_{L \rightarrow \infty} \sup_{n \in \mathbb{N}} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \frac{Y_k^{(i)}(n)}{n} \right|^p \mathbf{1}_{\{Y_k^{(i)}(n)/n \geq L\}} \right] = 0.$$

We recall that  $U_{\mathbf{q},k}(\lambda)$  is defined in (5.26) and is smooth on  $(-\infty, \delta_{\mathbf{q},k})$  with some  $\delta_{\mathbf{q},k} > 0$ . Thanks to (5.27), we get

$$\begin{aligned} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \frac{\lambda Y_k^{(i)}(n)}{n} \right) \right] &= \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( n U_{\mathbf{q},k} \left( \frac{\lambda}{n} \right) \frac{(n - M_k^{(i)}(n))}{n} \right) \right] \\ &\leq \exp \left( n U_{\mathbf{q},k} \left( \frac{\lambda}{n} \right) \right), \end{aligned}$$

where we use the fact  $0 \leq M_k^{(i)}(n) \leq n$ . By the smoothness of  $U_{\mathbf{q},k}(\lambda)$ , we see that if  $\lambda < \delta_{\mathbf{q},k}$ , then

$$\sup_{n \in \mathbb{N}} \exp \left( n U_{\mathbf{q},k} \left( \frac{\lambda}{n} \right) \right) < \infty.$$

Hence by the Chebyshev inequality we have

$$\begin{aligned} & \sup_{n \in \mathbb{N}} \nu_{\mathbf{q}} \left( \frac{Y_k^{(i)}(n)}{n} > L \right) \\ &\leq e^{-\lambda L} \sup_{n \in \mathbb{N}} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \frac{\lambda Y_k^{(i)}(n)}{n} \right) \right] \\ &\leq e^{-\lambda L} \sup_{n \in \mathbb{N}} \exp \left( n U_{\mathbf{q},k} \left( \frac{\lambda}{n} \right) \right) \\ &\rightarrow 0 \quad \text{as } L \rightarrow \infty. \end{aligned}$$

Moreover, from an elementary inequality  $x^p \leq ([p] + 1)!e^x$ ,  $x \geq 0$ , and the Schwartz inequality, by choosing  $0 < \lambda < \delta_{\mathbf{q},k}/2$ , we obtain

$$\begin{aligned}
& \sup_{n \in \mathbb{N}} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \frac{Y_k^{(i)}(n)}{n} \right|^p \mathbf{1}_{\{Y_k^{(i)}(n)/n \geq L\}} \right] \\
& \leq \frac{([p] + 1)!}{\lambda^p} \sup_{n \in \mathbb{N}} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \frac{\lambda Y_k^{(i)}(n)}{n} \right) \mathbf{1}_{\{Y_k^{(i)}(n)/n \geq L\}} \right] \\
& \leq \frac{([p] + 1)!}{\lambda^p} \sup_{n \in \mathbb{N}} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \frac{2\lambda Y_k^{(i)}(n)}{n} \right) \right]^{\frac{1}{2}} \nu_{\mathbf{q}} \left( \frac{Y_k^{(i)}(n)}{n} > L \right)^{\frac{1}{2}} \\
& \leq \frac{([p] + 1)!}{\lambda^p} \left( \sup_{n \in \mathbb{N}} \exp \left( n U_{\mathbf{q},k} \left( \frac{2\lambda}{n} \right) \right)^{\frac{1}{2}} \right) \left( \sup_{n \in \mathbb{N}} \nu_{\mathbf{q}} \left( \frac{Y_k^{(i)}(n)}{n} > L \right)^{\frac{1}{2}} \right) \\
& \rightarrow 0 \quad \text{as } L \rightarrow \infty.
\end{aligned}$$

Therefore Theorem 4.2 is proved.

## 7. PROOF OF THEOREM 4.5

**7.1. Proof of (1).** First we prepare the following simple lemmas.

**Lemma 7.1.** *Let  $\zeta(i)$ ,  $i \in \mathbb{N}$  be i.i.d. random variables define on any probability space with  $\mathbb{E}[\zeta(0)] = 0$  and  $\mathbb{E}[\zeta(0)^2] = 1$ , and define  $S(n) := \sum_{i=1}^n \zeta(i)$ ,  $n \in \mathbb{N}$ . Assume that  $a_\varepsilon(t)$ ,  $b_\varepsilon(t)$  are non-decreasing function on  $[0, \infty)$  such that*

$$\lim_{\varepsilon \rightarrow 0} \sup_{0 \leq t \leq T} |a_\varepsilon(t) - b_\varepsilon(t)| = 0, \quad \lim_{\varepsilon \rightarrow 0} \sup_{0 \leq t \leq T} |a_\varepsilon(t) - at| = 0,$$

with some constant  $a > 0$ . Then, for any  $T > 0$  and  $\delta > 0$ , we have

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P} \left( \sup_{0 \leq t \leq T} \varepsilon \left| S \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - S \left( \left\lfloor \frac{b_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right| > \delta \right) = 0.$$

*Proof.* Let  $B(t)$ ,  $t \geq 0$  be a standard Brownian motion defined on some probability space. Thanks to the Skorokhod embedding theorem (cf. [B, Theorem 37.7]), there exists a sequence of stopping times  $\tau(n)$ ,  $n \in \mathbb{Z}_{\geq 0}$ ,  $\tau_0 := 0$  such that  $\tau(n) - \tau(n-1)$ ,  $n \in \mathbb{N}$  are i.i.d. and

$$(B(\tau(n)), n \in \mathbb{N}) \stackrel{d}{=} (S(n), n \in \mathbb{N}).$$

Hence, we have

$$\begin{aligned} & \mathbb{P} \left( \sup_{0 \leq t \leq T} \varepsilon \left| S \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - S \left( \left\lfloor \frac{b_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right| > \delta \right) \\ &= \mathbb{P} \left( \sup_{0 \leq t \leq T} \varepsilon \left| B \left( \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right) - B \left( \tau \left( \left\lfloor \frac{b_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right) \right| > \delta \right) \\ &= \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| B \left( \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right) - B \left( \varepsilon^2 \tau \left( \left\lfloor \frac{b_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right) \right| > \delta \right). \end{aligned}$$

Now we claim that

$$\overline{\lim}_{\varepsilon \rightarrow 0} \sup_{0 \leq t \leq T} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - \varepsilon^2 \tau \left( \left\lfloor \frac{b_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right| = 0 \quad \text{a.s.} \quad (7.1)$$

Actually, for any  $t \geq 0$ , we have

$$\begin{aligned} \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) &= \varepsilon^2 \sum_{n=1}^{\lfloor a_\varepsilon(t) \varepsilon^{-2} \rfloor} (\tau(n) - \tau(n-1)) \\ &\rightarrow at \quad \text{a.s.}, \end{aligned}$$

where we use  $\mathbb{E}[\tau(1)] = 1$ . Then, for any  $n \in \mathbb{N}$ , we have

$$\overline{\lim}_{\varepsilon \rightarrow 0} \max_{0 \leq m \leq n} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon \left( \frac{mT}{n} \right)}{\varepsilon^2} \right\rfloor \right) - \frac{amT}{n} \right| = 0 \quad \text{a.s.} \quad (7.2)$$

On the other hand, by the monotonicity of  $\tau, a_\varepsilon$ , we have

$$\begin{aligned} & \sup_{t \in \left[ \frac{mT}{n}, \frac{(m+1)T}{n} \right]} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - at \right| \\ &\leq \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon \left( \frac{(m+1)T}{n} \right)}{\varepsilon^2} \right\rfloor \right) - \frac{amT}{n} \right| + \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon \left( \frac{mT}{n} \right)}{\varepsilon^2} \right\rfloor \right) - \frac{a(m+1)T}{n} \right| \\ &\leq \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon \left( \frac{(m+1)T}{n} \right)}{\varepsilon^2} \right\rfloor \right) - \frac{a(m+1)T}{n} \right| + \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon \left( \frac{mT}{n} \right)}{\varepsilon^2} \right\rfloor \right) - \frac{amT}{n} \right| + \frac{2aT}{n} \\ &\leq 2 \max_{0 \leq m \leq n} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon \left( \frac{mT}{n} \right)}{\varepsilon^2} \right\rfloor \right) - \frac{amT}{n} \right| + \frac{2aT}{n}. \end{aligned}$$

From (7.2), we see that

$$\begin{aligned} \overline{\lim}_{\varepsilon \rightarrow 0} \sup_{0 \leq t \leq T} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - at \right| &\leq \overline{\lim}_{\varepsilon \rightarrow 0} \max_{0 \leq m \leq n} \sup_{t \in \left[ \frac{mT}{n}, \frac{(m+1)T}{n} \right]} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - at \right| \\ &\leq \frac{2aT}{n} \quad \text{a.s.}, \end{aligned}$$

for any  $n \in \mathbb{N}$ . Hence we get

$$\overline{\lim}_{\varepsilon \rightarrow 0} \sup_{0 \leq t \leq T} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - at \right| = 0 \quad \text{a.s.},$$

and thus from the assumption of this lemma we obtain (7.1). From (7.1), for any  $\delta' > 0$  we have

$$\begin{aligned} & \overline{\lim}_{\varepsilon \rightarrow 0} \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| B \left( \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right) - B \left( \varepsilon^2 \tau \left( \left\lfloor \frac{b_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right) \right| > \delta \right) \\ & \leq \mathbb{P} \left( \sup_{\substack{0 \leq t, s \leq a(T+\delta'), \\ |t-s| \leq \delta'}} |B(t) - B(s)| > \delta \right) \\ & \quad + \overline{\lim}_{\varepsilon \rightarrow 0} \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - \varepsilon^2 \tau \left( \left\lfloor \frac{b_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) \right| > \delta' \right) \\ & \quad + 2 \overline{\lim}_{\varepsilon \rightarrow 0} \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| \varepsilon^2 \tau \left( \left\lfloor \frac{a_\varepsilon(t)}{\varepsilon^2} \right\rfloor \right) - at \right| > \delta' \right) \\ & = \mathbb{P} \left( \sup_{\substack{0 \leq t, s \leq a(T+\delta'), \\ |t-s| \leq \delta''}} |B(t) - B(s)| > \delta \right) \rightarrow 0 \quad \text{as } \delta' \rightarrow 0. \end{aligned}$$

From the above, Lemma 7.1 is proved.  $\square$

Recall that  $\Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, n)$  is defined in (5.29).

**Lemma 7.2.** *For any  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$ ,  $1 \leq \ell \leq k-1$ ,  $i \in \mathbb{Z}$ ,  $\mathbf{T} > 0$  and  $\delta > 0$ , we have*

$$\overline{\lim}_{\varepsilon \rightarrow 0} \nu_{\mathbf{q}} \left( \sup_{0 \leq t \leq \mathbf{T}} \frac{1}{n} \left| \Delta Y_{k,\ell}^{(i)}(\eta, \mathbf{q}, \lfloor n^2 t \rfloor) - \sum_{j=1}^{\lfloor v_{k-\ell}(\theta^\ell \mathbf{q}) n^2 t \rfloor} (\zeta_\ell(j) - \alpha_\ell(\mathbf{q})) \right| > \delta \right) = 0.$$

*Proof.* First we observe that since  $n \mapsto X_k^{(i)}(n)$  is increasing in  $n$ , by using (4.10) and the same argument used to derive (7.1) in Lemma 7.1, for any  $\mathbf{q} \in \mathcal{Q}$ ,  $k \in \mathbb{N}$ ,  $i \in \mathbb{Z}$  and  $0 \leq t \leq \mathbf{T}$ , we have

$$\lim_{\varepsilon \rightarrow 0} \sup_{0 \leq t \leq \mathbf{T}} \left| \frac{1}{n} X_k^{(i)}(\lfloor nt \rfloor) - v_k(\mathbf{q}) t \right| = 0 \quad \nu_{\mathbf{q}}\text{-a.s.} \quad (7.3)$$

Hence, from (5.5) Remark 5.7, Lemma 7.1 and (7.3), the assertion of this lemma is proved.  $\square$

Thanks to Remark 5.7, we see that the following stochastic processes,

$$t \mapsto \frac{1}{n} \sum_{j=1}^{\lfloor v_{k-\ell}^{\text{eff}}(\theta^\ell \mathbf{q}) n^2 t \rfloor} (\zeta_\ell(j) - \alpha_\ell(\mathbf{q})), \quad (7.4)$$

for  $1 \leq \ell \leq k-1$ , and

$$t \mapsto \frac{1}{n} \left( M_k^{(i)}(\eta, \lfloor n^2 t \rfloor) - \mathbb{E}_{\nu_{\mathbf{q}}} \left[ M_k^{(i)}(\lfloor n^2 t \rfloor) \right] \right), \quad (7.5)$$

are independent under  $\nu_{\mathbf{q}}$ . Moreover, since  $\nu_{\mathbf{q}}$  is translation-invariant, the distribution of (7.5) does not depend on  $i \in \mathbb{Z}$ . By following the standard way one can show that (7.4) converges weakly to the Brownian motion in  $D[0, T]$  with mean 0 and variance  $v_{k-\ell}^{\text{eff}}(\theta^\ell \mathbf{q}) \beta_\ell(\mathbf{q})$ . Hence, from the assumption of this theorem, Lemmas 5.4, 7.2 and the representation (5.30), the process (4.14) converges weakly to a sum of independent Brownian motions, and its variance  $D_k(\mathbf{q})$  is given by the sum of their variance. Therefore Theorem 4.5 (1) is proved.

**7.2. Proof of (2).** We recall that  $U_{\mathbf{q},k}(\lambda)$  is defined in (5.26). From Lemma 5.8, if  $\lambda < \min\{\delta_{\mathbf{q},k}, \lambda_{\mathbf{q},k}\}$ , then for any  $i \in \mathbb{Z}$  and  $n \geq 0$ , we have

$$\frac{1}{n} \log \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda Y_k^{(i)}(n) \right) \right] = \frac{1}{n} \log \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( U_{\mathbf{q},k}(\lambda) \left( n - M_k^{(i)}(n) \right) \right) \right].$$

Hence from the assumption of Theorem 4.5 (2), we have

$$\Lambda_{\mathbf{q},k}^Y(\lambda) = \Lambda_{\mathbf{q},k}^M(U_{\mathbf{q},k}(\lambda)). \quad (7.6)$$

We also recall that  $U_{\mathbf{q},k}(\cdot)$  is a smooth monotone convex function, and  $\sup_{\lambda \leq \delta} U_{\mathbf{q},k}(\lambda) < \infty$  for sufficiently small  $\delta > 0$ . In addition, since  $0 \leq M_k^{(i)}(n) \leq n$ , we have  $|\Lambda_{\mathbf{q},k}^M(\lambda)| \leq |\lambda|$  for any  $\lambda$ . Hence if  $\Lambda_{\mathbf{q},k}^M$  is essentially smooth, then so is  $\Lambda_{\mathbf{q},k}^Y$ . Therefore Theorem 4.5 (2) is proved.

## 8. PROOF OF THEOREM 4.8

First we prepare two lemmas. To describe them, we define  $\Xi_k(\eta, i) := \xi_k(\eta, s_\infty(\eta, i))$ , and

$$\begin{aligned} \bar{J}_k(\mathbf{q}) &:= \mathbb{E}_{\nu_{\mathbf{q}}} [J_k(1)], \\ \bar{s}_\infty(\mathbf{q}) &:= \mathbb{E}_{\nu_{\mathbf{q}}} [s_\infty(1) - s_\infty(0)], \\ \bar{\Xi}_k(\mathbf{q}) &:= \mathbb{E}_{\nu_{\mathbf{q}}} [\Xi_k(1) - \Xi_k(0)], \end{aligned}$$

where  $J_k(i)$ ,  $i \in \mathbb{Z}$  is defined in (5.15).

**Lemma 8.1.** *Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $\mathbb{E}_{\nu_{\mathbf{q}}} [|\mathbf{e}^{(0)}|^2] < \infty$ . Then, for any  $k \in \mathbb{N}$ ,  $u \in \mathbb{R}$ , we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} X_k^{(\lfloor nu \rfloor)}(0) = \frac{\bar{s}_\infty(\mathbf{q}) \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})}, \quad \nu_{\mathbf{q}}\text{-a.s. and in } \mathbb{L}_1, \quad (8.1)$$

and

$$\overline{\lim}_{L \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} X_k^{(\lfloor nu \rfloor)} - \frac{\bar{s}_\infty(\mathbf{q}) \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right| > \frac{L}{\sqrt{n}} \right) = 0. \quad (8.2)$$

*Proof of Lemma 8.1.* First we observe that

$$\begin{aligned} s_\infty(\eta, i+1) - s_\infty(\eta, i) &= |\mathbf{e}^{(i)}(\eta)| \\ &= |\mathbf{e}^{(i)}(\tilde{\eta})| \\ &= s_\infty(\tilde{\eta}, i+1) - s_\infty(\tilde{\eta}, i), \end{aligned}$$

where  $\mathbf{e}^{(i)}(\eta)$  and  $\tilde{\eta}$  are defined in (2.3) and (3.6), respectively. Similarly, we have

$$\begin{aligned} \Xi_k(\eta, i+1) - \Xi_k(\eta, i) &= 1 + \sum_{\ell \in \mathbb{N}} \sum_{x=s_\infty(\eta, i)+1}^{s_\infty(\eta, i+1)} (\eta_{k+\ell}^\uparrow(x) + \eta_{k+\ell}^\downarrow(x)) \\ &= 1 + \sum_{\ell \in \mathbb{N}} \sum_{x=s_\infty(\tilde{\eta}, i)+1}^{s_\infty(\tilde{\eta}, i+1)} (\tilde{\eta}_{k+\ell}^\uparrow(x) + \tilde{\eta}_{k+\ell}^\downarrow(x)) \\ &= 1 + \sum_{\ell \in \mathbb{N}} \left| \left\{ (k+\ell, \sigma)\text{-seats in } \mathbf{e}^{(i)}(\tilde{\eta}), \sigma \in \{\uparrow, \downarrow\} \right\} \right|. \end{aligned}$$

In particular, both  $s_\infty(\tilde{\eta}, i+1) - s_\infty(\tilde{\eta}, i)$  and  $\Xi_k(\tilde{\eta}, i+1) - \Xi_k(\tilde{\eta}, i)$  are functions of  $\mathbf{e}^{(i)}(\tilde{\eta})$ . In addition,

$$\Xi_k(\eta, i+1) - \Xi_k(\eta, i) \leq s_\infty(\eta, i+1) - s_\infty(\eta, i).$$

Hence from Remark 4.1 and the assumption of this lemma, both  $(s_\infty(\eta, i+1) - s_\infty(\eta, i))_{i \in \mathbb{Z}}$  and  $(\Xi_k(\eta, i+1) - \Xi_k(\eta, i))_{i \in \mathbb{Z}}$  are i.i.d.  $\mathbb{L}^2$  sequences under  $\nu_{\mathbf{q}}$ . In addition, since  $(\zeta_k(\eta, i))_{i \in \mathbb{Z}}$  are i.i.d. geometric random variables under  $\nu_{\mathbf{q}}$ ,  $(J_k(\eta, i+1) - J_k(\eta, i))_{i \in \mathbb{Z}}$  is also an i.i.d.  $\mathbb{L}^2$  sequence under  $\nu_{\mathbf{q}}$ . Thus we see that for any  $u \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} s_\infty(\lfloor nu \rfloor) = \bar{s}_\infty(\mathbf{q}) u, \quad \nu_{\mathbf{q}}\text{-a.s. and in } \mathbb{L}_1, \quad (8.3)$$

$$\overline{\lim}_{L \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} s_\infty \left( \left\lfloor \frac{u}{\varepsilon} \right\rfloor \right) - \bar{s}_\infty(\mathbf{q}) u \right| > \frac{L}{\sqrt{n}} \right) = 0, \quad (8.4)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} J_k(\lfloor nu \rfloor) = \bar{J}_k(\mathbf{q}) u, \quad \nu_{\mathbf{q}}\text{-a.s.}, \quad (8.5)$$

$$\overline{\lim}_{L \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} J_k(\lfloor nu \rfloor) - \bar{J}_k(\mathbf{q}) u \right| > \frac{L}{\sqrt{n}} \right) = 0,$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \Xi_k(\lfloor nu \rfloor) &= \bar{\Xi}_k(\mathbf{q}) u, \quad \nu_{\mathbf{q}}\text{-a.s.}, \\ \overline{\lim}_{L \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} \Xi_k(\lfloor nu \rfloor) - \bar{\Xi}_k(\mathbf{q}) u \right| > \frac{L}{\sqrt{n}} \right) &= 0. \end{aligned}$$



Hence we get

$$\lim_{n \rightarrow \infty} \left| \frac{1}{n} J_k(\lfloor nu \rfloor) - \frac{1}{n} \Xi_k \left( \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor \right) \right| = 0, \quad \nu_{\mathbf{q}}\text{-a.s.}, \quad (8.6)$$

$$\overline{\lim}_{L \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} J_k(\lfloor nu \rfloor) - \frac{1}{n} \Xi_k \left( \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor \right) \right| > \frac{L}{\sqrt{n}} \right) = 0. \quad (8.7)$$

First we show (8.1). We fix  $\delta > 0$  and  $\eta \in \Omega$ , such that

$$\overline{\lim}_{n \rightarrow \infty} \left| \frac{1}{n} J_k(\eta, \lfloor nu \rfloor) - \frac{1}{n} \Xi_k \left( \eta, \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor \right) \right| \leq \delta.$$

Then, since  $\xi_k(\eta, X_k^{(i)}(\eta)) = J_k(\eta, i)$  for any  $k \in \mathbb{N}$  and  $i \in \mathbb{Z}$ , and  $\xi_k(\cdot)$  increases at each record, we get

$$\begin{aligned} & \Xi_k \left( \eta, \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor - \lfloor n\delta \rfloor \right) \\ & \leq \Xi_k \left( \eta, \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor \right) - \lfloor n\delta \rfloor \\ & \leq \xi_k \left( \eta, X_k^{(\lfloor nu \rfloor)}(\eta) \right) \\ & \leq \Xi_k \left( \eta, \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor \right) + \lfloor n\delta \rfloor \\ & \leq \Xi_k \left( \eta, \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor + \lfloor n\delta \rfloor \right). \end{aligned}$$

Thus we have

$$\frac{1}{n} s_{\infty} \left( \eta, \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor - \lfloor n\delta \rfloor \right) \leq \frac{1}{n} X_k^{(\lfloor nu \rfloor)}(\eta) \leq \frac{1}{n} s_{\infty} \left( \eta, \left\lfloor \frac{n \bar{J}_k(\mathbf{q}) u}{\bar{\Xi}_k(\mathbf{q})} \right\rfloor + \lfloor n\delta \rfloor \right).$$

Hence, by using (8.3) and (8.6), we have (8.1).

By using (8.4) and (8.7) instead of (8.3) and (8.6), we can prove (8.2) from a similar computation.  $\square$

Recall that  $\sigma_{k,\ell}^{(i)}(\eta, n)$  is defined in (5.17).

**Lemma 8.2.** *Suppose that  $\mathbf{q} \in \mathcal{Q}$  and  $\mathbb{E}_{\nu_{\mathbf{q}}} [|\mathbf{e}^{(0)}|^2] < \infty$ . Then for any  $k, \ell \in \mathbb{N}$  with  $k < \ell$  and  $u, v \in \mathbb{R}$ , we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(0) = f_{k,\ell}(\mathbf{q}, u), \quad \nu_{\mathbf{q}}\text{-a.s.}, \quad (8.8)$$

$$\overline{\lim}_{L \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(0) - f_{k,\ell}(\mathbf{q}, u) \right| > \frac{L}{\sqrt{n}} \right) = 0, \quad (8.9)$$

and

$$\overline{\lim}_{L \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(n^2) - \frac{1}{n} \sigma_{k,\ell}^{(\lfloor nv \rfloor)}(n^2) - f_{k,\ell}(\mathbf{q}, u - v) \right| > \frac{L}{\sqrt{n}} \right) = 0, \quad (8.10)$$

where

$$f_{k,\ell}(\mathbf{q}, u) := \frac{\bar{\Xi}_{\ell-k}(\theta^k \mathbf{q}) \bar{J}_k(\mathbf{q}) u}{\bar{s}_{\infty}(\theta^k \mathbf{q}) \bar{J}_{\ell}(\mathbf{q})}.$$

*Proof.* We fix  $\mathbf{q}$ ,  $k < \ell$  and  $u, v$ . Without loss of generality, we can assume that  $u > v$ . For notational simplicity, we will write  $\tilde{u} = f_{k,\ell}(\mathbf{q}, u)$ .

First we show (8.8). We observe that from Lemma 8.1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} X_{\ell-k}^{(\lfloor n\tilde{u} \rfloor)}(\Psi_k(\eta)) = \bar{J}_k(\mathbf{q}) u, \quad \nu_{\mathbf{q}\text{-a. s.}},$$

where we use (5.2) and (5.4) to show  $\text{bar} J_{\ell-k}(\theta^k \mathbf{q}) = \text{bar} J_{\ell}(\mathbf{q})$ . We fix  $\delta > 0$  and  $\eta \in \Omega$  such that

$$\overline{\lim}_{n \rightarrow \infty} \left| \frac{1}{n} X_{\ell-k}^{(\lfloor n\tilde{u} \rfloor)}(\Psi_k(\eta)) - \bar{J}_k(\mathbf{q}) \tilde{u} \right| \leq \delta,$$

and

$$\overline{\lim}_{n \rightarrow \infty} \left| \frac{1}{n} J_k(\lfloor nu \rfloor) - \bar{J}_k(\mathbf{q}) u \right| \leq \delta.$$

Then, we see that

$$\frac{1}{n} X_{\ell-k}^{(\lfloor n\tilde{u} \rfloor - \lfloor n\delta/k \rfloor)}(\Psi_k(\eta)) \leq J_k(\eta, \lfloor nu \rfloor) \leq \frac{1}{n} X_{\ell-k}^{(\lfloor n\tilde{u} \rfloor + \lfloor n\delta/k \rfloor)}(\Psi_k(\eta)),$$

where we use the fact that the number of  $k$ -solitons contained in an interval  $[a, b]$ ,  $a < b$  is at most  $(b - a)/(2k)$ . Hence, from Remark 5.7 and (8.5), we obtain

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left| \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(0) - \tilde{u} \right| \leq \delta, \quad \nu_{\mathbf{q}\text{-a.s.}}$$

Therefore we have (8.8).

Next we show (8.9). We fix  $L_1 > 0$ . Then, on the following event,

$$A_{L_1} := \left\{ \left| \frac{1}{n} X_{\ell-k}^{(\lfloor n\tilde{u} \rfloor)}(\Psi_k(\eta)) - \bar{J}_k(\mathbf{q}) \tilde{u} \right| \leq L_1 n^{-\frac{1}{2}} \right\} \\ \cap \left\{ \left| \frac{1}{n} J_k(\lfloor nu \rfloor) - \bar{J}_k(\mathbf{q}) u \right| \leq L_1 n^{-\frac{1}{2}} \right\},$$

we see that

$$\left| \frac{1}{n} \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(0) - \tilde{u} \right| \leq \frac{L_1}{kn^{\frac{1}{2}}} + \frac{n\tilde{u} - \lfloor n\tilde{u} \rfloor}{n}.$$

Hence we have

$$\begin{aligned}
& \overline{\lim}_{L_2 \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \left| \frac{1}{n} \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(0) - \tilde{u} \right| > L_2 n^{-\frac{1}{2}} \right) \\
& \leq \overline{\lim}_{L_2 \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}} \left( \mathbf{1}_{A_{L_1}} \left| \frac{1}{n} \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(0) - \tilde{u} \right| > L_2 n^{-\frac{1}{2}} \right) + \nu_{\mathbf{q}}(A_{L_1}) \\
& = \overline{\lim}_{n \rightarrow \infty} \nu_{\mathbf{q}}(A_{L_1}) \rightarrow 0, \quad L_1 \rightarrow \infty.
\end{aligned}$$

Therefore we obtain (8.8).

Finally we show (8.10). We observe that  $\sigma_{k,\ell}^{(\lfloor nu \rfloor)}(n^2) - \sigma_{k,\ell}^{(\lfloor nv \rfloor)}(n^2)$  is equal to the number of  $(\ell - k)$ -solitons with volume in  $[J_k(\lfloor nv \rfloor), J_k(\lfloor nu \rfloor)]$  at time  $n^2$ , i.e.,

$$\begin{aligned}
& \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(n^2) - \sigma_{k,\ell}^{(\lfloor nv \rfloor)}(n^2) \\
& = \left| \left\{ j \in \mathbb{Z} ; J_k(\lfloor nv \rfloor) \leq X_{\ell-k}^{(j)}(T^{n^2} \Psi_k(\tilde{\eta})) \leq J_k(\lfloor nu \rfloor) \right\} \right|.
\end{aligned}$$

Since  $\nu_{\mathbf{q}}$  is an invariant measure, from Remark 5.7 we have

$$\begin{aligned}
& \left| \left\{ j \in \mathbb{Z} ; J_k(\lfloor nv \rfloor) \leq X_{\ell-k}^{(j)}(T^{n^2} \Psi_k(\tilde{\eta})) \leq J_k(\lfloor nu \rfloor) \right\} \right| \\
& \stackrel{d}{=} \left| \left\{ j \in \mathbb{Z} ; J_k(\lfloor nv \rfloor) \leq X_{\ell-k}^{(j)}(\Psi_k(\tilde{\eta})) \leq J_k(\lfloor nu \rfloor) \right\} \right| \\
& = \sigma_{k,\ell}^{(\lfloor nu \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor nv \rfloor)}(0).
\end{aligned}$$

Hence we have (8.10). □

*Proof of Theorem 4.8.* Without loss of generality, we can assume  $u > v$ . From Lemma 5.6, for any  $0 \leq \ell \leq k-1$ , we have

$$\begin{aligned}
& Y_{k-\ell}^{(\lfloor n^a u \rfloor)}(\Psi_\ell(\eta), n^2) - Y_{k-\ell}^{(\lfloor n^a v \rfloor)}(\Psi_\ell(\eta), n^2) \\
& = \frac{v_{k-\ell}^{\text{eff}}(\theta^\ell \mathbf{q})}{\bar{r}_k(\mathbf{q})} \left( M_k^{(\lfloor n^a v \rfloor)}(\tilde{\eta}, n^2) - M_k^{(\lfloor n^a u \rfloor)}(\tilde{\eta}, n^2) \right) \\
& + 2 \sum_{h=\ell+1}^{k-1} \frac{v_{h-\ell}^{\text{eff}}(\theta^\ell \mathbf{q})}{\bar{r}_h(\mathbf{q})} \left( \sum_{j=X_{k-h}^{(\lfloor n^a u \rfloor)}(\Psi_h(\tilde{\eta}), 0)+1}^{X_{k-h}^{(\lfloor n^a u \rfloor)}(\Psi_h(\tilde{\eta}), n^2)} - \sum_{j=X_{k-h}^{(\lfloor n^a v \rfloor)}(\Psi_h(\tilde{\eta}), 0)+1}^{X_{k-h}^{(\lfloor n^a v \rfloor)}(\Psi_h(\tilde{\eta}), n^2)} \right) \\
& \quad \times (\zeta_h(\tilde{\eta}, j) - \alpha_h(\mathbf{q})). \tag{8.11}
\end{aligned}$$

By using Remark 5.7 and (8.11), we get

$$\begin{aligned}
& \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \frac{1}{n} Y_{k-\ell}^{([n^a u])} (\Psi_{\ell}(\eta), n^2) - \frac{1}{n} Y_{k-\ell}^{([n^a v])} (\Psi_{\ell}(\eta), n^2) \right|^2 \right] \\
&= \frac{v_{k-\ell}^{\text{eff}} (\theta^{\ell} \mathbf{q})^2}{\bar{r}_k(\mathbf{q})^2 n^2} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| M_k^{([n^a u])} (n^2) - M_k^{([n^a v])} (n^2) \right|^2 \right] \\
&+ 4 \sum_{h=\ell+1}^{k-1} \frac{v_{h-\ell}^{\text{eff}} (\theta^{\ell} \mathbf{q})^2 \beta_h(\mathbf{q})}{\bar{r}_h(\mathbf{q})^2 n^2} \\
&\quad \times \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| Y_{k-h}^{([n^a u])} (\Psi_h(\eta), n^2) - Y_{k-h}^{([n^a v])} (\Psi_h(\eta), n^2) \right|^2 \right]. \quad (8.12)
\end{aligned}$$

where at the last line we use Lemma 3.1. For notational simplicity, we define

$$\widetilde{M}_{k,n}^{u,v} := \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| M_k^{([n^a u])} (n^2) - M_k^{([n^a v])} (n^2) \right|^2 \right].$$

Then, from (8.11) with  $\ell = k-1$  and the Schwartz inequality, we get

$$\begin{aligned}
& \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| Y_1^{([n^a u])} (\Psi_{k-1}(\eta), n^2) - Y_1^{([n^a v])} (\Psi_{k-1}(\eta), n^2) \right|^2 \right] \\
&= \frac{v_1^{\text{eff}} (\theta^{k-1} \mathbf{q})}{\bar{r}_k(\mathbf{q})} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| M_k^{([n^a v])} (\tilde{\eta}, n^2) - M_k^{([n^a u])} (\tilde{\eta}, n^2) \right|^2 \right] \\
&\leq \frac{v_1^{\text{eff}} (\theta^{k-1} \mathbf{q})}{\bar{r}_k(\mathbf{q})} \left( \widetilde{M}_{k,n}^{u,v} \right)^{\frac{1}{2}}.
\end{aligned}$$

By using this, (8.12) with  $\ell = k-2$  and the Schwartz inequality, we see that

$$\begin{aligned}
& \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| Y_2^{([n^a u])} (\Psi_{k-2}(\eta), n^2) - Y_2^{([n^a v])} (\Psi_{k-2}(\eta), n^2) \right|^2 \right] \\
&\leq \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| Y_2^{([n^a u])} (\Psi_{k-2}(\eta), n^2) - Y_2^{([n^a v])} (\Psi_{k-2}(\eta), n^2) \right|^2 \right]^{\frac{1}{2}} \\
&\leq \left( \frac{v_2^{\text{eff}} (\theta^{k-2} \mathbf{q})^2}{\bar{r}_k(\mathbf{q})^2} \widetilde{M}_{k,n}^{u,v} + \frac{4v_1^{\text{eff}} (\theta^{k-2} \mathbf{q})^2 \beta_{k-1}(\mathbf{q}) v_1^{\text{eff}} (\theta^{k-1} \mathbf{q})}{\bar{r}_{k-1}(\mathbf{q})^2 \bar{r}_k(\mathbf{q})} \left( \widetilde{M}_{k,n}^{u,v} \right)^{\frac{1}{2}} \right)^{\frac{1}{2}} \\
&\leq \frac{v_2^{\text{eff}} (\theta^{k-2} \mathbf{q})}{\bar{r}_k(\mathbf{q})} \left( \widetilde{M}_{k,n}^{u,v} \right)^{\frac{1}{2}} + \frac{2v_1^{\text{eff}} (\theta^{k-2} \mathbf{q}) \beta_{k-1}(\mathbf{q}) v_1^{\text{eff}} (\theta^{k-1} \mathbf{q})^{\frac{1}{2}}}{\bar{r}_{k-1}(\mathbf{q}) \bar{r}_k(\mathbf{q})^{\frac{1}{2}}} \left( \widetilde{M}_{k,n}^{u,v} \right)^{\frac{1}{4}}.
\end{aligned}$$

By repeating the above procedure from  $\ell = k-1$  to 0, we see that there exists some constant  $c_k = c_k(\mathbf{q})$  such that for any  $0 \leq \ell \leq k-1$ ,

$$\mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| Y_{\ell}^{([n^a u])} (\Psi_{k-\ell}(\eta), n^2) - Y_{\ell}^{([n^a v])} (\Psi_{k-\ell}(\eta), n^2) \right|^2 \right] \leq c_k \sum_{h=0}^k \left( \widetilde{M}_{k,n}^{u,v} \right)^{\left(\frac{1}{2}\right)^h}.$$

Hence, it is sufficient to show that

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| M_k^{([n^a u])} (n^2) - M_k^{([n^a v])} (n^2) \right|^2 \right] = 0. \quad (8.13)$$

From now on we prove (8.13). From (3.4), (5.16) and the triangle inequality, we have

$$\begin{aligned}
& \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| M_k^{(\lfloor n^a u \rfloor)}(n^2) - M_k^{(\lfloor n^a v \rfloor)}(n^2) \right|^2 \right]^{\frac{1}{2}} \\
& \leq 2 \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \sum_{\ell=k+1}^{\infty} M_{k,\ell}^{(\lfloor n^a u \rfloor)}(n^2) - M_{k,\ell}^{(\lfloor n^a v \rfloor)}(n^2) \right|^2 \right]^{\frac{1}{2}} \\
& \leq 2 \sum_{\ell=k+1}^{\infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| M_{k,\ell}^{(\lfloor n^a u \rfloor)}(n^2) - M_{k,\ell}^{(\lfloor n^a v \rfloor)}(n^2) \right|^2 \right]^{\frac{1}{2}} \\
& \leq 2 \sum_{\ell=k+1}^{\infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0))^{-1} \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(n^2))} \zeta_{\ell}(j) \end{array} - \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)) \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(n^2))^{-1}} \zeta_{\ell}(j) \end{array} \right|^2 \right]^{\frac{1}{2}} \\
& \quad + 2 \sum_{\ell=k+1}^{\infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0)) \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(n^2))^{-1}} \zeta_{\ell}(j) \end{array} - \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0))^{-1} \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(n^2))} \zeta_{\ell}(j) \end{array} \right|^2 \right]^{\frac{1}{2}}.
\end{aligned}$$

Hence, it is sufficient to show

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0))^{-1} \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(n^2))} \zeta_{\ell}(j) \end{array} - \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)) \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(n^2))^{-1}} \zeta_{\ell}(j) \end{array} \right|^2 \right] = 0, \tag{8.14}$$

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0)) \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(n^2))^{-1}} \zeta_{\ell}(j) \end{array} - \begin{array}{c} J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0))^{-1} \\ \sum_{j=J_{\ell}(\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(n^2))} \zeta_{\ell}(j) \end{array} \right|^2 \right] = 0, \tag{8.15}$$

for any  $u > v$  and  $\ell \geq k+1$ . In the following we will only show (8.14). We note that (8.15) can be proved by the same computation.

First we prepare an estimate for  $\sigma_{k,\ell}^{(i)}(0)$ . Since  $|X_h^{(i)}(0)| \geq 2h|i|$  for any  $i \in \mathbb{Z}$  and  $h \in \mathbb{N}$ , we have

$$0 \leq \sigma_{k,\ell}^{(i)}(0) \leq J_k(\tilde{\eta}, i), \tag{8.16}$$

for any  $i \geq 1$ , and

$$J_k(\tilde{\eta}, i) \leq \sigma_{k,\ell}^{(i)}(0) \leq 1, \tag{8.17}$$

for any  $i \leq 0$ . In addition, for notational simplicity, we define

$$I_{k,\ell}^{(i)}(\eta) := \sigma_{k,\ell}^{(i)}(0) - \sigma_{k,\ell}^{(i)}(n^2).$$

Now we estimate (8.14). Observe that

$$\begin{aligned}
& J_\ell \left( \sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, 0) \right)^{-1} \sum_{j=J_\ell \left( \sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2) \right)} \zeta_\ell(\eta, j) - J_\ell \left( \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, 0) \right) \sum_{j=J_\ell \left( \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2) \right)^{-1}} \zeta_\ell(\eta, j) \\
&= \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2)}^{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, 0)^{-1}} \zeta_\ell(\eta, J_\ell(\eta, j)) - \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2)^{-1}}^{\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, 0)} \zeta_\ell(\eta, J_\ell(\eta, j)) \\
&= \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, 0)+1}^{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, 0)^{-1}} \left( \zeta_\ell(\eta, J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \\
&\quad - \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2)}^{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2)} \left( \zeta_\ell(\eta, J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \\
&\quad + \frac{1}{1-q_\ell} \left( I_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta) - I_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta) \right). \tag{8.18}
\end{aligned}$$

For the first term in (8.18), by using (8.16) and (8.17), we have

$$\begin{aligned}
& \left| \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, 0)+1}^{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, 0)^{-1}} \left( \zeta_\ell(\eta, J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \right| \\
& \leq \sup_{J_k(\eta, -n^a(|u|+|v|)) \leq m < m' \leq J_k(\eta, n^a(|u|+|v|))} \left| \sum_{j=m+1}^{m'-1} \left( \zeta_\ell(\eta, J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \right|.
\end{aligned}$$

Since  $(\zeta_\ell(\eta, J_\ell(\eta, j)))_{j \in \mathbb{Z}}$  are i.i.d. with mean  $(1-q_\ell)^{-1}$ , and  $\zeta_\ell$  is independent of  $\zeta_k$ , by Doob's inequality we get

$$\begin{aligned}
& \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \sup_{J_k(-n^a(|u|+|v|)) \leq m < m' \leq J_k(n^a(|u|+|v|))} \left| \sum_{j=m+1}^{m'-1} \left( \zeta_\ell(J_\ell(j)) - \frac{1}{1-q_\ell} \right) \right|^2 \right] \\
& \leq 4 \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \sum_{j=J_k(-n^a(|u|+|v|))+1}^{J_k(n^a(|u|+|v|))-1} \left( \zeta_\ell(J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \right|^2 \right] \\
& \leq 8n^a(|u|+|v|) \bar{J}_k(\mathbf{q}) \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left( \zeta_\ell(J_\ell(\eta, 0)) - \frac{1}{1-q_\ell} \right)^2 \right].
\end{aligned}$$

Hence we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[ \left| \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, 0)+1}^{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, 0)^{-1}} \left( \zeta_\ell(\eta, J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \right|^2 \right] = 0.$$

For the second term in (8.18), we observe that

$$\begin{aligned}
& \sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2) \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2)} \left( \zeta_\ell(\eta, J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \\
&= \sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2) \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2)} \left( \zeta_{\ell-k}(\Psi_k(\eta), J_{\ell-k}(\Psi_k(\eta), j)) - \frac{1}{1-q_\ell} \right) \\
&= \tilde{\sigma}_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2) \sum_{j=\tilde{\sigma}_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2)} \left( \zeta_{\ell-k}(T^{n^2}\Psi_k(\eta), J_{\ell-k}(T^{n^2}\Psi_k(\eta), j)) - \frac{1}{1-q_\ell} \right),
\end{aligned}$$

where at the second line we use (5.3), and  $\tilde{\sigma}_{k,\ell}^{(i)}(\eta, m)$  is defined as

$$\tilde{\sigma}_{k,\ell}^{(i)}(\eta, m) := \inf \left\{ j \in \mathbb{Z} ; X_{\ell-k}^{(j)}(T^m \Psi_k(\tilde{\eta}), 0) \geq J_k(\tilde{\eta}, i) \right\},$$

for any  $k < \ell$ ,  $i \in \mathbb{Z}$  and  $m \in \mathbb{Z}_{\geq 0}$ . Since  $\tilde{\sigma}_{k,\ell}^{(i)}(\eta, m)$  also satisfies (8.16) and (8.17) by replacing  $\sigma_{k,\ell}^{(i)}(\eta, 0)$  with  $\tilde{\sigma}_{k,\ell}^{(i)}(\eta, m)$ , we have

$$\begin{aligned}
& \left| \tilde{\sigma}_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2) \sum_{j=\tilde{\sigma}_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2)} \left( \zeta_{\ell-k}(T^{n^2}\Psi_k(\tilde{\eta}), J_{\ell-k}(T^{n^2}\Psi_k(\tilde{\eta}), j)) - \frac{1}{1-q_\ell} \right) \right| \\
&\leq \sup_{J_k(\eta, -n^a(|u|+|v|)) \leq m < m' \leq J_k(\eta, n^a(|u|+|v|))} \left| \sum_{j=m+1}^{m'-1} \left( \zeta_{\ell-k}(T^{n^2}\Psi_k(\eta), J_{\ell-k}(T^{n^2}\Psi_k(\eta), j)) - \frac{1}{1-q_\ell} \right) \right|.
\end{aligned}$$

By the independence of  $\zeta_k$  and  $\zeta_\ell$ , and then using  $T$ -invariance of  $\nu_{\mathbf{q}}$  and Doob's inequality, we obtain

$$\begin{aligned}
& \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \sup_{J_k(-n^a(|u|+|v|)) \leq m < m' \leq J_k(n^a(|u|+|v|))} \left| \sum_{j=m+1}^{m'-1} \left( \zeta_{\ell-k}(T^{n^2}\Psi_k(\eta), J_{\ell-k}(T^{n^2}\Psi_k(\eta), j)) - \frac{1}{1-q_\ell} \right) \right|^2 \right] \\
&\leq 8n^a(|u|+|v|) \bar{J}_k(\mathbf{q}) \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left( \zeta_\ell(J_\ell(\eta, 0)) - \frac{1}{1-q_\ell} \right)^2 \right].
\end{aligned}$$

Hence we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[ \left| \sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(\eta, n^2) \sum_{j=\sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(\eta, n^2)} \left( \zeta_\ell(\eta, J_\ell(\eta, j)) - \frac{1}{1-q_\ell} \right) \right|^2 \right] = 0.$$

For the third term in (8.18), since  $(J_k(j) - J_k(j-1))_{j \in \mathbb{Z}}$  are i.i.d. and have geometric distribution with mean  $(1 - q_k)q_k^{-1}$ , by using (8.16) and (8.17) we have

$$\begin{aligned} & \overline{\lim}_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)}{n^a} \right)^4 \right] \\ & \leq 4 \overline{\lim}_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \frac{J_k(\lfloor |n^a u| \rfloor)^4 + J_k(\lfloor |n^a v| \rfloor)^4}{n^{4a}} \right] \\ & \leq \frac{4(1 - q_k)^4 (|u|^4 + |v|^4)}{q_k^4}. \end{aligned}$$

Therefore, from (8.9), by setting

$$A_{L,n} := \left\{ \left| \frac{1}{n^a} \sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \frac{1}{n^a} \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0) - f_{k,\ell}(\mathbf{q}, u - v) \right| > \frac{L}{\sqrt{n^a}} \right\},$$

with some  $L > 0$ , and using the Schwartz inequality, we have

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)}{n^a} - f_{k,\ell}(\mathbf{q}, u - v) \right)^2 \right] \\ & = \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \mathbf{1}_{A_{L,n}^c} \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)}{n^a} - f_{k,\ell}(\mathbf{q}, u - v) \right)^2 \right] \\ & + \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \mathbf{1}_{A_{L,n}} \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)}{n^a} - f_{k,\ell}(\mathbf{q}, u - v) \right)^2 \right] \\ & \leq \frac{L^2}{n^a} + 2 \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \mathbf{1}_{A_{L,n}} \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)}{n^a} \right)^2 \right] \\ & \quad + 2\nu_{\mathbf{q}}(A_{L,n}) |f_{k,\ell}(\mathbf{q}, u - v)|^2 \\ & \leq \frac{L^2}{n^a} + 2\nu_{\mathbf{q}}(A_{L,n})^{\frac{1}{2}} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)}{n^a} \right)^4 \right]^{\frac{1}{2}} \\ & \quad + 2\nu_{\mathbf{q}}(A_{L,n}) |f_{k,\ell}(\mathbf{q}, u - v)|^2, \end{aligned}$$

and thus we obtain

$$\overline{\lim}_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(0) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(0)}{n^a} - f_{k,\ell}(\mathbf{q}, -v) \right)^2 \right] = 0.$$



By the observation used in the computation for the second term in (8.18) and using (8.10), we also have

$$\overline{\lim}_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left( \frac{\sigma_{k,\ell}^{(\lfloor n^a u \rfloor)}(n^2) - \sigma_{k,\ell}^{(\lfloor n^a v \rfloor)}(n^2)}{n^a} - f_{k,\ell}(\mathbf{q}, -v) \right)^2 \right] = 0.$$

Since  $0 \leq a \leq 1$ , from the above estimates we obtain

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| I_{k,\ell}^{(\lfloor n^a u \rfloor)} - I_{k,\ell}^{(\lfloor n^a v \rfloor)} \right|^2 \right] = 0.$$

By combining the above and using the Schwartz inequality, we have (8.14), and thus Theorem 4.8 is proved.  $\square$

## 9. PROOF OF THEOREM 4.10

In this section we show Theorem 4.10. First we note that thanks to (5.4) and (5.11), it is sufficient to consider the case  $\mathbf{q} \in \mathcal{Q}_M$  and  $k = 1$ . We fix  $\mathbf{q} \in \mathcal{Q}_M$ . Next, we prepare the following lemmas. Recall that  $W(\cdot)$  is the carrier process defined in (1.1).

Next, we prepare two lemmas. Before describing the lemmas, we recall that the inverse of one-step time evolution  $T^{-1} : \Omega \rightarrow \Omega$  is given by

$$T^{-1}\eta(x) = \overleftarrow{(T\overleftarrow{\eta})}(x),$$

where  $\overleftarrow{\eta}(x) := \eta(-x)$ ,  $x \in \mathbb{Z}$ , see [CKST, (2.12)]. By using the carrier  $W$  and (1.2),  $T^{-1}\eta$  can be represented as

$$T^{-1}\eta(x) = \eta(x) - W(\overleftarrow{\eta}, -x) + W(\overleftarrow{\eta}, -x - 1).$$

We also recall that the ball density  $\rho(\mathbf{q})$  is defined in (4.19).

**Lemma 9.1.** *If the initial distribution  $\nu$  is a two-sided Markov distribution supported in  $\Omega$ , then for any  $x \in \mathbb{Z}$ ,  $(T^n\eta(x))_{n \in \mathbb{Z}}$  is an irreducible and non-periodic two-sided Markov chain on  $\{0, 1\}$  whose transition matrix is given by*

$$R = \begin{pmatrix} R_{00} & R_{01} \\ R_{10} & R_{11} \end{pmatrix} := \begin{pmatrix} 1 - \frac{\nu(\eta(0)=1)}{1-\nu(\eta(0)=1)} & \frac{\nu(\eta(0)=1)}{1-\nu(\eta(0)=1)} \\ 1 & 0 \end{pmatrix}, \quad (9.1)$$

and invariant measure  $\pi \in [0, 1]^2$  for  $R$  is the Bernoulli measure with density  $\nu(\eta(0) = 1)$ .

*Proof.* Since  $\eta \in \Omega$  is a two-sided Markov chain under  $\nu_{\mathbf{q}}$ ,  $(\eta(y))_{y \geq x}$  and  $(\eta(x))_{x \leq y}$  are independent conditional on  $\eta(x)$ . On the other hand, since the carrier  $W(\eta, x)$  is  $(\eta(y))_{y \leq x}$ -m'ble, we see that  $(T^n\eta(x))_{n \geq 0}$  (resp.  $(T^n\eta(x))_{n \leq 0}$ ) is  $(\eta(y))_{y \leq x}$ -m'ble (resp.  $(\eta(y))_{y \geq x}$ -m'ble). Hence, the processes  $(T^n\eta(x))_{n \geq 0}$  and  $(T^n\eta(x))_{n \leq 0}$  are independent conditional on  $\eta(x)$ , and this implies the Markov property at  $n = 0$ . Since  $\nu_{\mathbf{q}}$  is  $T$ -invariant,

$(T^{n+m}\eta(x))_{n \geq 0}$  and  $(T^{n+m}\eta(x))_{n \leq 0}$  are independent conditional on  $T^m\eta(x)$  for any  $m \in \mathbb{Z}$ , and thus  $(T^n\eta(x))_{n \in \mathbb{Z}}$  is a two-sided Markov chain.

Since the invariant measure for  $(T^n\eta(x))_{n \in \mathbb{Z}}$  is the Bernoulli measure with density  $\nu(\eta(0) = 1)$ , we can obtain  $R$  by direct computation.  $\square$

**Lemma 9.2.** *Suppose that  $\mathbf{q} \in \mathcal{Q}_M$ . Then, for any  $x, z \leq 0$ , the process  $r(T^n\eta, x+z)$ ,  $n \geq 2$  and the event  $\{s_\infty(0) = z\}$  are independent conditional on  $T^2\eta(x+z)$ .*

*Proof of Lemma 9.2.* Since  $r(\eta, x) = \mathbf{1}_{\{\eta(x)=T\eta(x)=0\}}$ , the event  $\{s_\infty(0) = z\}$  is  $\eta(z), T\eta(z)$ -m'ble. In addition, by taking the action  $T^{-1}$ , we see that both  $\eta(z)$  and  $T\eta(z)$  are  $(T^2\eta(y))_{y \geq z}$ -measurable. Hence  $\{s_\infty(0) = z\}$  is  $(T^2\eta(y))_{y \geq z}$ -measurable. On the other hand,  $(r(T^m\eta, x+z))$ ,  $n \geq 2$  is  $(T^2\eta(y))_{y \leq x+z}$ -measurable. Thus by the Markov property of  $T^2\eta$ , the process  $r(T^n\eta, x+z)$ ,  $n \geq 2$  and the event  $\{s_\infty(0) = z\}$  are independent conditional on  $T^2\eta(x+z)$ .  $\square$

**9.1. Convergence of (4.13).** In this subsection, we will show that (4.13) with  $k = 1$  converges to the Brownian motion with variance  $G_1(\mathbf{q})$ , and compute  $G_1(\mathbf{q})$ .

We recall the formula (5.14) with  $k = 1$ .

$$M_1^{(i)}(\eta, n) = \sum_{m=0}^{n-1} (1 - r(T^m\Psi_1(\tilde{\eta}), J_1(\tilde{\eta}, i))).$$

From Remark 5.7,  $\Psi_1(\tilde{\eta})$  is independent of  $J_1(\tilde{\eta}, i)$ . Hence by (5.5), to show the weak convergence of (4.13), it is sufficient to show that for any fixed  $x \in \mathbb{Z}$ , the following process,

$$t \mapsto \frac{1}{n} \sum_{m=0}^{\lfloor n^2 t \rfloor - 1} (r(T^m\tilde{\eta}, x) - \mathbb{E}_{\nu_{\theta\mathbf{q}}}[r(T^m\eta, x)]), \quad (9.2)$$

converges weakly to a Brownian motion under  $\nu_{\theta\mathbf{q}}$ . First we consider the case  $x \leq 0$ . Since  $T$  and  $\tau_{s_\infty(\eta, 0)}$  are commutative, we have

$$\begin{aligned} & \sum_{m=0}^{n-1} (r(T^m\tilde{\eta}, x) - \mathbb{E}_{\nu_{\theta\mathbf{q}}}[r(T^m\Psi_1(\tilde{\eta}), x)]) \\ &= \sum_{m=0}^{n-1} (r(T^m\eta, x + s_\infty(\eta, 0)) - \mathbb{E}_{\nu_{\theta\mathbf{q}}}[r(T^m\eta, x + s_\infty(\eta, 0))]). \end{aligned}$$

Then, by Lemma 9.2, the process  $(r(T^m \eta, x+z))$  and the event  $\{s_\infty(0) = z\}$ ,  $z \geq 0$  are independent conditional on  $T^2 \eta(x+z)$ . In addition, by the conditional independence and shift-invariance of  $\nu_{\theta \mathbf{q}}$ , if  $m \geq 2$ , then

$$\begin{aligned}
& \mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \eta, x+s_\infty(0))] \\
&= \sum_{z \leq 0} \mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \eta, x+z) \mathbf{1}_{\{s_\infty(0)=z\}}] \\
&= \sum_{z \leq 0} \frac{\mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \eta, x+z) \mathbf{1}_{\{T^2 \eta(x+z)=1\}}] \nu_{\theta \mathbf{q}}(s_\infty(0) = z, T^2 \eta(x+z) = 1)}{\nu_{\theta \mathbf{q}}(T^2 \eta(x+z) = 1)} \\
&\quad + \sum_{z \leq 0} \frac{\mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \eta, x+z) \mathbf{1}_{\{T^2 \eta(x+z)=0\}}] \nu_{\theta \mathbf{q}}(s_\infty(0) = z, T^2 \eta(x+z) = 0)}{\nu_{\theta \mathbf{q}}(T^2 \eta(x+z) = 0)} \\
&= \mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \eta, 0) | T^2 \eta(0) = 1] \sum_{z \leq 0} \nu_{\theta \mathbf{q}}(s_\infty(0) = z, T^2 \eta(x+z) = 1) \\
&\quad + \mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \eta, 0) | T^2 \eta(0) = 0] \sum_{z \leq 0} \nu_{\theta \mathbf{q}}(s_\infty(0) = z, T^2 \eta(x+z) = 0).
\end{aligned}$$

Since  $(T^m \eta(0))_{m \geq 2}$  is a finite ergodic Markov chain under  $\nu_{\theta \mathbf{q}}$ , and is strongly mixing with exponentially decay rate, we have

$$\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{m=2}^{\lfloor n^2 t \rfloor - 1} (\mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \eta, x+s_\infty(0))] - \bar{r}_1(\mathbf{q})) \right| = 0.$$

Also, thanks to the strong mixing property, for any  $(T^m \eta(x+z))_{m \geq 2}$ -m'ble set  $A$ , the difference  $|\nu_{\mathbf{q}}(A) - \nu_{\mathbf{q}}(T^{m'} A | T^2 \eta(x+z) = a)|$  decays exponentially fast to 0 as  $m' \rightarrow \infty$  for  $a = 0, 1$ . From the above, it suffices to show that the following process,

$$t \mapsto \frac{1}{n} \sum_{m=0}^{\lfloor n^2 t \rfloor - 1} (r(T^m \eta, x+z) - \bar{r}_1(\mathbf{q})), \quad (9.3)$$

converges weakly to a Brownian motion under  $\nu_{\theta \mathbf{q}}(\cdot)$ , and this can be shown by the invariance principle for strongly mixing stationary sequences (cf. [EK, Theorem 3.1]). Hence the convergence of (9.2) has been shown for the case  $x \leq 0$ . Next we consider the case  $x \geq 1$ . From Remark 4.1, for any  $z \in \mathbb{Z}$ , the distributions of  $\tilde{\eta}$  and  $\tau_{s_\infty(\eta, z) - s_\infty(\eta, 0)} \tilde{\eta}$  are the same under  $\nu_{\mathbf{q}}$ . Hence, we have

$$\begin{aligned}
& \sum_{m=0}^{n-1} (r(T^m \tilde{\eta}, x) - \mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \Psi_1(\tilde{\eta}), x)]) \\
& \stackrel{d}{=} \sum_{m=0}^{n-1} (r(T^m \tilde{\eta}, x - s_\infty(\eta, x)) - \mathbb{E}_{\nu_{\theta \mathbf{q}}} [r(T^m \Psi_1(\tau_{s_\infty(\eta, x)} \eta), x)]),
\end{aligned}$$

under  $\nu_{\theta \mathbf{q}}$ . Then, since  $x \leq s_\infty(\eta, x)$ , by the same argument used in Lemma 9.2, we see that the events  $\{s_\infty(\eta, x) = z\}$ ,  $x \leq z$  and  $\{s_\infty(\eta, x) = z'\}$ ,  $z' \leq 0$ , and the process  $(r(T^m \eta, x - z + z'))$  are independent conditional on

$T^2\eta(x - z + z')$ . Thus by repeating the same argument for  $x \leq 0$ , we can show that (9.2) converges weakly to a Brownian motion under  $\nu_{\theta\mathbf{q}}$ .

Now we compute the variance  $G_1(\mathbf{q})$ . We observe that by Lemma 9.1, the sum in (9.3) can be viewed as a functional of the ergodic Markov chain  $((T^m\eta(x+z), T^{m+1}\eta(x+z)))_{m \in \mathbb{Z}}$  on  $\{(0,0), (0,1), (1,0)\}$ , where its transition matrix  $R'$  and invariant measure  $\pi' \in [0,1]^3$  are given by

$$R' = \begin{pmatrix} 1 - \frac{\rho(\theta\mathbf{q})}{1-\rho(\theta\mathbf{q})} & \frac{\rho(\theta\mathbf{q})}{1-\rho(\theta\mathbf{q})} & 0 \\ 0 & 0 & 1 \\ 1 - \frac{\rho(\theta\mathbf{q})}{1-\rho(\theta\mathbf{q})} & \frac{\rho(\theta\mathbf{q})}{1-\rho(\theta\mathbf{q})} & 0 \end{pmatrix},$$

and

$$\pi'((0,0)) = 1 - 2\rho(\theta\mathbf{q}), \quad \pi'((0,1)) = \pi'((1,0)) = 2\rho(\theta\mathbf{q}).$$

Since the explicit solution of the following Poisson equation,

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \left( \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - R' \right) \mathbf{f},$$

is given by

$$\mathbf{f} = \begin{pmatrix} 1 \\ -(1 - 2\rho(\theta\mathbf{q})) \\ -2(1 - 2\rho(\theta\mathbf{q})) \end{pmatrix},$$

from [KLO12, Theorem 1.2],  $G_1(\mathbf{q})$  can be computed as

$$\begin{aligned} G_1(\mathbf{q}) &= \mathbb{E}_{\pi'}[|\mathbf{f}|^2] - \mathbb{E}_{\pi'}[|R'\mathbf{f}|^2] \\ &= 4\rho(\theta\mathbf{q})(1 - \rho(\theta\mathbf{q}))(1 - 2\rho(\theta\mathbf{q})). \end{aligned}$$

We note that thanks to (5.4), the formula of  $G_k(\mathbf{q})$  for  $k \in \mathbb{N}$  can be obtained by using  $G_k(\mathbf{q}) = G_1(\theta^{k-1}\mathbf{q})$ .

**9.2. Convergence of (4.16).** First we remark that by the same argument used in Section 9.1 and using Lemma 9.2, we see that if the following limit,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \mathbb{E}_{\nu_{\theta\mathbf{q}}} \left[ \exp \left( \lambda \sum_{m=0}^{n-1} r(T^m\eta, 0) \right) \middle| \eta(0) = w \right] \right),$$

exists and independent of  $w = 0, 1$ , then it coincides with  $\Lambda_{\mathbf{q},1}^M(\lambda)$ . By using  $r(T^m\eta, 0) = (1 - T^m\eta(0))(1 - T^{m+1}\eta(0))$ , for any  $w_0 \in \{0, 1\}$ , we have

$$\begin{aligned} & \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \exp \left( \lambda \sum_{m=0}^{n-1} r(T^m\eta, 0) \right) \middle| \eta(0) = w_0 \right] \\ &= \sum_{w_1, \dots, w_n} \prod_{i=0}^{n-1} R_{w_i w_{i+1}} e^{\lambda(1-w_i)(1-w_{i+1})} \\ &= (\tilde{R}(\lambda)^n)_{w_0 0} + (\tilde{R}(\lambda)^n)_{w_0 1}, \end{aligned}$$

where  $R_{ij}$  is defined in (9.1), and  $\tilde{R}(\lambda)$  is given by

$$\tilde{R}(\lambda) := \begin{pmatrix} \frac{1-2\rho(\theta\mathbf{q})}{1-\rho(\theta\mathbf{q})} e^\lambda & \frac{\rho(\theta\mathbf{q})}{1-\rho(\theta\mathbf{q})} \\ 1 & 0 \end{pmatrix}.$$

Hence, from [DZ, Theorem 3.1.1], we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \mathbb{E}_{\nu_{\bar{\mathbf{q}}}} \left[ \exp \left( \lambda \sum_{m=0}^{n-1} r(T^m \eta, 0) \right) \middle| \eta(0) = w \right] \right) = \log(\text{PF}(\lambda)),$$

where  $\text{PF}(\lambda)$  is the Perron-Frobenius eigenvalue of  $\tilde{R}$ . By a direct computation, we see that

$$\text{PF}(\lambda) = \frac{1-2\rho(\theta\mathbf{q})}{2(1-\rho(\theta\mathbf{q}))} \left( e^\lambda + \sqrt{e^{2\lambda} - 1 + \frac{1}{(1-2\rho(\theta\mathbf{q}))^2}} \right).$$

In particular,  $\log(\text{PF}(\lambda))$  is a smooth convex function on  $\mathbb{R}$ . The convex conjugate of  $\log(\text{PF}(\lambda))$  can be computed as

$$I_{\mathbf{q},1}^M(u) = \sup_{\lambda \in \mathbb{R}} (\lambda u - \log(\text{PF}(\lambda))) = \begin{cases} \frac{u}{2} \log \left( \frac{4\rho(\theta\mathbf{q})(1-\rho(\theta\mathbf{q}))u^2}{(1-2\rho(\theta\mathbf{q}))^2(1-u^2)} \right) \\ -\frac{1}{2} \log \left( \frac{\rho(\theta\mathbf{q})(1+u)}{(1-\rho(\theta\mathbf{q}))(1-u)} \right) & 0 \leq u < 1, \\ \log \left( \frac{2(1-\rho(\theta\mathbf{q}))}{1-2\rho(\theta\mathbf{q})} \right) & u = 1, \\ \infty & \text{otherwise.} \end{cases}$$

We note that the minimizer of  $I_{\mathbf{q},1}^M(u)$  is  $1-2\rho(\theta\mathbf{q})$ , and from (4.12), (B.1), the value of minimizer coincides with  $v_1^{\text{eff}}(\mathbf{q})$ .

## 10. PROOF OF THEOREM 4.14

We fix  $\mathbf{q} \in \mathcal{Q}$  satisfying the assumption of Theorem 4.14 and define  $k := \max\{1 \leq h \leq \ell - 1 ; q_h > 0\}$ .

First we claim that under  $\nu_{\theta^k \mathbf{q}}$ ,  $(\eta(x), W(x))$  is an ergodic Markov chain in  $x \in \mathbb{Z}$  on the state space,

$$S_{\ell-k} := \{(0,0), (0,1), \dots, (0, \ell-k-1), (1,1), (1,2), \dots, (1, \ell-k)\},$$

with transition matrix

$$P_{\ell-k} = \begin{pmatrix} P_{\ell-k}^{(1)} & P_{\ell-k}^{(2)} \\ P_{\ell-k}^{(3)} & P_{\ell-k}^{(4)} \end{pmatrix},$$

where  $P_h^{(i)}$ ,  $i = 1, \dots, 4$  are  $h \times h$  matrices given by

$$P_1^{(1)} = 1 - q_\ell, \quad P_1^{(2)} = q_\ell, \quad P_1^{(3)} = 1, \quad P_1^{(4)} = 0,$$

for  $h = 1$ , and

$$P_h^{(1)} = \begin{pmatrix} 1 - q_\ell & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad P_h^{(2)} = \begin{pmatrix} q_\ell & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix},$$

$$P_h^{(3)} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \\ 0 & \dots & 1 \end{pmatrix}, \quad P_h^{(4)} = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

for  $h \geq 2$ . Actually, if we define  $X_n = (x_n^{(1)}, x_n^{(2)})_{n \in \mathbb{Z}_{\geq 0}}$  is the Markov process on  $S_{k'}$  with the above transition matrix, and recursively define stopping times as

$$\tau_1 := \inf \{m \in \mathbb{N} ; X_{m-1} = X_m = (0, 0)\},$$

$$\tau_{n+1} := \inf \{m \geq \tau_n + 1 ; X_{m-1} = X_m = (0, 0)\},$$

then the distribution of  $(X_m^{(1)})_{\tau_1 \leq m \leq \tau_2 - 1}$  coincides with  $\nu_{\theta^k \mathbf{q}} \circ \mathbf{e}^{(i)}$ . In addition,  $(X_m^{(1)})_{\tau_n \leq m \leq \tau_{n+1} - 1}$  and  $(X_m^{(1)})_{\tau_{n'} \leq m \leq \tau_{n'+1} - 1}$  are independent if  $n \neq n'$ . Hence, from the construction of  $\mathbf{q}$ -statistics,  $(\eta(x), W(x))$ ,  $x \in \mathbb{Z}$  is the desired ergodic Markov process under  $\nu_{\theta^k \mathbf{q}}$ .

On the other hand, since there are only  $(\ell - k)$ -solitons under  $\nu_{\theta^k \mathbf{q}}$ , from (5.14), we have

$$M_k^{(0)}(\eta, n) = \sum_{m=0}^{n-1} (1 - r(T^m \Psi_k(\tilde{\eta}), J_k(\tilde{\eta}, 0)))$$

$$= \sum_{m=0}^{n-1} (1 - r(\Psi_k(\tilde{\eta}), J_k(\tilde{\eta}, 0) - (\ell - k)m)),$$

a.s. under  $\nu_{\mathbf{q}}$ . From the above, we see that  $M_k^{(0)}(\eta, n)$  is a functional of an ergodic Markov process. Therefore, by using a similar argument to Section 9, one can show that (4.13) converges weakly to a Brownian motion. Also, by using the relation  $r(x) = \mathbf{1}_{\{W(x)=0\}} \mathbf{1}_{\{W(x+1)=0\}}$ , for any  $\mathbf{s}_0 = (s_0^{(1)}, s_0^{(2)}) \in$

$S_{\ell-k}$ , we have

$$\begin{aligned}
& \mathbb{E}_{\nu_{\theta^\ell C_k \mathbf{q}}} \left[ \exp \left( \lambda \sum_{m=0}^{n-1} r((\ell-k)m) \right) \middle| (\eta(0), W(0)) = \mathbf{s}_0 \right] \\
&= \sum_{\mathbf{s}_1, \mathbf{s}_{\ell-k}, \mathbf{s}_{\ell-k+1}, \dots, \mathbf{s}_{(n-1)(\ell-k)}, \mathbf{s}_{(n-1)(\ell-k)+1}} \\
& \quad \left( \prod_{i=0}^{n-2} (P_{\ell-k})_{\mathbf{s}_{i(\ell-k)} \mathbf{s}_{i(\ell-k)+1}} e^{\lambda \delta_0(s_{i(\ell-k)}^{(2)}) \delta_0(s_{i(\ell-k)+1}^{(2)})} (P_{k'})_{\mathbf{s}_{i(\ell-k)+1} \mathbf{s}_{(i+1)(\ell-k)}}^{(\ell-k)-1} \right) \\
& \quad \times (P_{\ell-k})_{\mathbf{s}_{(n-1)k'} \mathbf{s}_{(n-1)(\ell-k)+1}} e^{\lambda \delta_0(s_{(n-1)(\ell-k)}^{(2)}) \delta_0(s_{(n-1)(\ell-k)+1}^{(2)})} \\
&= \sum_{\mathbf{s}} (\tilde{P}_{\ell-k})_{\mathbf{s}_0 \mathbf{s}}^{n-1} \sum_{\mathbf{s}'} (P_{\ell-k})_{\mathbf{s} \mathbf{s}'} e^{\lambda \delta_0(s^{(2)}) \delta_0((s')^{(2)})},
\end{aligned}$$

where

$$\tilde{P}_{\ell-k}(\lambda) := \begin{pmatrix} \tilde{P}_{\ell-k}^{(1)}(\lambda) & \tilde{P}_{\ell-k}^{(2)}(\lambda) \\ \tilde{P}_{\ell-k}^{(3)} & \tilde{P}_{\ell-k}^{(4)} \end{pmatrix},$$

and  $\tilde{P}_h^{(i)}$ ,  $i = 1, \dots, 4$  are given by

$$\tilde{P}_1^{(1)} = e^\lambda (1 - q_k), \quad \tilde{P}_1^{(2)} = q_k, \quad \tilde{P}_1^{(3)} = 1, \quad \tilde{P}_1^{(4)} = 0,$$

for  $h = 1$ , and

$$\begin{aligned}
\tilde{P}_h^{(1)}(\lambda) &:= \begin{pmatrix} e^\lambda (1 - q_\ell)^h & 0 & \dots & 0 \\ (1 - q_\ell)^{h-1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 - q_\ell & 0 & \dots & 0 \end{pmatrix}, \\
\tilde{P}_h^{(2)}(\lambda) &:= \begin{pmatrix} e^\lambda (1 - q_\ell)^{h-1} q_\ell & \dots & e^\lambda (1 - q_\ell) q_\ell & q_\ell \\ (1 - q_\ell)^{h-2} q_\ell & \dots & q_\ell & 0 \\ \vdots & \ddots & \ddots & \vdots \\ q_\ell & 0 & \dots & 0 \end{pmatrix}, \\
\tilde{P}_h^{(3)} &:= \begin{pmatrix} 0 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 0 \end{pmatrix}, \quad \tilde{P}_h^{(4)} := \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix},
\end{aligned}$$

for  $h \geq 2$ . From [DZ, Theorem 3.1.1 (e)] and the ergodicity of the Markov chain defined above, we have

$$\begin{aligned}
\Lambda_{\mathbf{q}, \ell}^M(\lambda) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \mathbb{E}_{\nu_{\theta^k \mathbf{q}}} \left[ \exp \left( \lambda \sum_{m=0}^{n-1} r((\ell-k)m) \right) \middle| (\eta(0), W(0)) = \mathbf{s}_0 \right] \right) \\
&= \log(\widetilde{\text{PF}}(\lambda)),
\end{aligned}$$

where  $\widetilde{\text{PF}}(\lambda)$  is the Perron-Frobenius eigenvalue of  $\tilde{P}_{k'}(\lambda)$ . Since

$$\det(\tilde{P}_{\ell-k}(\lambda) - xI_{2(\ell-k)}) = \det(xI_{\ell-k} (xI_{\ell-k} - \tilde{P}_{\ell-k}^{(1)}(\lambda)) - \tilde{P}_{\ell-k}^{(2)} \tilde{P}_{\ell-k}^{(3)}),$$

where  $I_h$  is the  $h \times h$  identity matrix, by direct computation we see that

$$\widetilde{\text{PF}}(\lambda) = \frac{(1 - q\ell)^{\ell-k} e^\lambda}{2} + \sqrt{\frac{(1 - q\ell)^{2(\ell-k)} e^{2\lambda}}{4} + q\ell}.$$

Hence  $D_k(\rho)$  can be computed as

$$\begin{aligned} D_k(\mathbf{q}) &= \frac{d^2 \log(\widetilde{\text{PF}}(\lambda))}{d\lambda^2} \Big|_{\lambda=0} \\ &= \frac{4q\ell}{(1 - q\ell)^{2(\ell-k)}} \left( 1 + \frac{4q\ell}{(1 - q\ell)^{2(\ell-k)}} \right)^{-\frac{3}{2}}. \end{aligned}$$

## 11. PROOF OF THEOREMS 2.1

We recall that if  $\nu$  is a space-homogeneous Bernoulli product measure or two-sided Markov distribution supported on  $\Omega$ , then there exists  $\mathbf{q} \in \mathcal{Q}_M$  such that  $\nu = \nu_{\mathbf{q}}$  and  $K(\mathbf{q}) = 1$ . In addition, under such  $\nu_{\mathbf{q}}$ , we have the exponential bound of  $|\mathbf{e}^{(0)}|$  as follows.

**Lemma 11.1.** *Suppose that  $\mathbf{q} \in \mathcal{Q}_M$ . Then, for sufficiently small  $\lambda > 0$ , we have  $\mathbb{E}_{\nu_{\mathbf{q}}} [e^{\lambda|\mathbf{e}^{(0)}|}] < \infty$ .*

The proof of Lemma 11.1 is given in Section C.

In the following subsections, we will show (1) and (2) and (3) under  $\nu_{\mathbf{q}}$ ,  $\mathbf{q} \in \mathcal{Q}_M$ .

### 11.1. Proof of (1).

*Proof of (1).* First we observe that for any  $\eta \in \Omega$ ,  $k \in \mathbb{N}$  and  $n \in \mathbb{Z}_+$ , we have

$$X_k^{(1)}(\eta, n) \leq X_k^i(\eta, n) \leq X_k^{(i)}(\eta, n), \quad (11.1)$$

for any  $i \in \mathbb{N}$ , and

$$X_k^{(i)}(\eta, n) \leq X_k^i(\eta, n) \leq X_k^{(1)}(\eta, n) \quad (11.2)$$

for any  $i \in \mathbb{Z}_{\leq 0}$ . In the following, we only consider the case  $i \geq 1$ . For the case  $i \leq 0$ , we can use the same argument. From Theorems 4.5(1), 4.8, 4.10 and Lemma 11.1, the joint distribution of the following step-interpolation processes,

$$t \mapsto \left( \frac{1}{n} X_k^{(1)}(\lfloor n^2 t \rfloor) - n t v_k^{\text{eff}}(\mathbf{q}), \frac{1}{n} X_k^{(i)}(\lfloor n^2 t \rfloor) - n t v_k^{\text{eff}}(\mathbf{q}) \right),$$

converges weakly to  $(B_k(\cdot), B_k(\cdot))$ . In particular, these processes are tight and satisfies

$$\overline{\lim}_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \sup_{0 \leq t, s \leq \mathbf{T}, |t-s| \leq \delta} \left| \frac{X_k^{(j)}(\lfloor n^2 t \rfloor) - X_k^{(j)}(\lfloor n^2 s \rfloor)}{n} - n(t-s)v_k^{\text{eff}}(\mathbf{q}) \right| > \varepsilon \right) = 0, \quad (11.3)$$



for any  $\mathbf{T}, \delta, \varepsilon > 0$  and  $j = 1, i$ . In addition, we have

$$\overline{\lim}_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \sup_{0 \leq t \leq \mathbf{T}} \left| \frac{X_k^{(i)}(\lfloor n^2 t \rfloor) - X_k^{(1)}(\lfloor n^2 t \rfloor)}{n} \right| > \varepsilon \right) = 0, \quad (11.4)$$

for any  $\mathbf{T}, \varepsilon > 0$ . On the other hand, from (11.1) we have

$$\begin{aligned} & \sup_{0 \leq t, s \leq \mathbf{T}, |t-s| \leq \delta} \left| \frac{X_k^i(\lfloor n^2 t \rfloor) - X_k^i(\lfloor n^2 s \rfloor)}{n} - n(t-s)v_k^{\text{eff}}(\mathbf{q}) \right| \\ & \leq \sup_{0 \leq t, s \leq \mathbf{T}, |t-s| \leq \delta} \left| \frac{X_k^{(i)}(\lfloor n^2 t \rfloor) - X_k^{(1)}(\lfloor n^2 s \rfloor)}{n} - n(t-s)v_k^{\text{eff}}(\mathbf{q}) \right| \\ & \quad + \sup_{0 \leq t, s \leq \mathbf{T}, |t-s| \leq \delta} \left| \frac{X_k^{(1)}(\lfloor n^2 t \rfloor) - X_k^{(i)}(\lfloor n^2 s \rfloor)}{n} - n(t-s)v_k^{\text{eff}}(\mathbf{q}) \right| \\ & \leq \sup_{0 \leq t, s \leq \mathbf{T}, |t-s| \leq \delta} \left| \frac{X_k^{(1)}(\lfloor n^2 t \rfloor) - X_k^{(1)}(\lfloor n^2 s \rfloor)}{n} - n(t-s)v_k^{\text{eff}}(\mathbf{q}) \right| \\ & \quad + \sup_{0 \leq t, s \leq \mathbf{T}, |t-s| \leq \delta} \left| \frac{X_k^{(i)}(\lfloor n^2 t \rfloor) - X_k^{(i)}(\lfloor n^2 s \rfloor)}{n} - n(t-s)v_k^{\text{eff}}(\mathbf{q}) \right| \\ & \quad + 2 \sup_{0 \leq t \leq \mathbf{T}} \left| \frac{X_k^{(i)}(\lfloor n^2 t \rfloor) - X_k^{(1)}(\lfloor n^2 t \rfloor)}{n} \right|. \end{aligned}$$

for any  $\mathbf{T}, \varepsilon > 0$ . Hence from (11.3) we have

$$\overline{\lim}_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \sup_{0 \leq t, s \leq \mathbf{T}, |t-s| \leq \delta} \left| \frac{X_k^i(\lfloor n^2 t \rfloor) - X_k^i(\lfloor n^2 s \rfloor)}{n} - n(t-s)v_k^{\text{eff}}(\mathbf{q}) \right| > \varepsilon \right) = 0,$$

for any  $\mathbf{T}, \delta, \varepsilon > 0$ . In addition, from (11.5), we get

$$\overline{\lim}_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \frac{X_k^i(0)}{n} > \varepsilon \right) = 0.$$

Thus we see that (2.4) is tight. Furthermore, from (11.1) and (11.4), we have

$$\begin{aligned} & \overline{\lim}_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \sup_{0 \leq t \leq \mathbf{T}} \left| \frac{X_k^{(i)}(\lfloor n^2 t \rfloor) - X_k^i(\lfloor n^2 t \rfloor)}{n} \right| > \varepsilon \right) \\ & \leq \overline{\lim}_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \sup_{0 \leq t \leq \mathbf{T}} \left| \frac{X_k^{(i)}(\lfloor n^2 t \rfloor) - X_k^{(1)}(\lfloor n^2 t \rfloor)}{n} \right| > \varepsilon \right) = 0. \end{aligned}$$

Therefore, (2.4) converges to the Brownian motion with variance  $D_k(\mathbf{q})$ .  $\square$

### 11.2. Proof of (2).

*Proof of (2).* First we show that for any  $k \in \mathbb{N}$  and  $i \in \mathbb{Z}$ ,  $X_k^{(i)}(\cdot)$  satisfies the LDP with rate function (4.17). Since  $X_k^{(i)}(\cdot)$  is  $\mathbb{R}$ -valued process and  $I$  is a smooth convex function, it is sufficient to show that

$$\varliminf_{n \rightarrow \infty} \frac{1}{n} \log \nu_{\mathbf{q}} \left( \frac{X_k^{(i)}(n)}{n} < x \right) \geq -\inf_{y \leq x} I(y),$$

and

$$\varlimsup_{n \rightarrow \infty} \frac{1}{n} \log \nu_{\mathbf{q}} \left( \frac{X_k^{(i)}(n)}{n} \leq x \right) \leq -\inf_{y \leq x} I(y).$$

Observe that for any  $\varepsilon > 0$ , from (11.5),

$$\begin{aligned} \varliminf_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \left| \frac{X_k^{(i)}(n) - Y_k^{(i)}(n)}{n} \right| > \varepsilon \right) &= \varliminf_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \left| \frac{X_k^{(i)}(0)}{n} \right| > \varepsilon \right) \\ &= 0. \end{aligned}$$

On the other hand, from Theorems 4.5 (2) and 4.10, we get

$$\varliminf_{n \rightarrow \infty} \frac{1}{n} \log \nu_{\mathbf{q}} \left( \frac{Y_k^{(i)}(n)}{n} < x \right) \geq -\inf_{y \leq x} I(y),$$

and

$$\varlimsup_{n \rightarrow \infty} \frac{1}{n} \log \nu_{\mathbf{q}} \left( \frac{Y_k^{(i)}(n)}{n} \leq x \right) \leq -\inf_{y \leq x} I(y).$$

Hence we see that  $X_k^{(i)}(\cdot)$  satisfies the LDP with rate function (4.17).

Now we consider the LDP for  $X_k^i$ . From Theorem 2.1 (3) and (11.1), for the case  $i \geq 1$ , we have

$$\begin{aligned} \varliminf_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \left| \frac{X_k^{(i)}(n) - X_k^i(n)}{n} \right| > \varepsilon \right) &= \varliminf_{n \rightarrow 0} \nu_{\mathbf{q}} \left( \frac{X_k^{(i)}(n) - X_k^{(1)}(n)}{n} > \varepsilon \right) \\ &\leq \varliminf_{n \rightarrow 0} \frac{1}{\varepsilon n} \mathbb{E} \left[ X_k^{(i)}(n) - X_k^{(1)}(n) \right] \\ &= 0. \end{aligned}$$

Therefore  $X_k^i(\cdot)$  satisfies the LDP with rate function (4.17). The case  $i \leq 0$  can be shown by the same argument, so we omit it.  $\square$

### 11.3. Proof of (3).

*Proof.* First We recall the inequalities (11.1) and (11.2).

In the following we only consider the case  $i \in \mathbb{N}$ , and the case  $i \in \mathbb{Z}_{\leq 0}$  can be shown by using the same strategy.

First we claim that for any  $k \in \mathbb{N}$ ,  $i \in \mathbb{N}$  and  $p \geq 1$ , we have

$$\mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| X_k^{(i)}(0) \right|^p \right] < \infty. \quad (11.5)$$

Before proving this, we recall that  $J_k(\eta, i)$ ,  $i \in \mathbb{Z}$  is defined in (5.15). From the definition of  $J_k(\eta, i)$  and the following inequality  $s_k(\eta, x) \leq s_{\infty}(\eta, x)$  for any  $x \in \mathbb{Z}_{\geq 0}$ , we get

$$X_k^{(i)}(\eta, 0) = s_k(\eta, J_k(\eta, i)) \leq s_{\infty}(\eta, J_k(\eta, i)) \leq \sum_{j=0}^{J_k(\eta, i)-1} \left| \mathbf{e}^{(j)} \right|.$$

Since  $J_k(\eta, j+1) - J_k(\eta, j)$ ,  $j \geq 1$  and  $J_k(\eta, 1)$  are i.i.d. geometric random variables with mean  $q_k^{-1}$ , we have

$$\sum_{x \in \mathbb{N}} x \nu_{\mathbf{q}}(J_k(i) = x)^{\frac{1}{2}} < \infty.$$

Then by using Remark 4.1, Lemma 11.1 and the Schwartz inequality, we obtain

$$\begin{aligned} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| X_k^{(i)}(0) \right|^p \right] &\leq \mathbb{E}_{\nu_{\mathbf{q}}} \left[ |s_{\infty}(J_k(i))|^p \right] \\ &= \sum_{x \geq i} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ |s_{\infty}(x)|^p \mathbf{1}_{\{J_k(i)=x\}} \right] \\ &\leq \sum_{x \geq i} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ |s_{\infty}(x)|^{2p} \right]^{\frac{1}{2}} \nu_{\mathbf{q}}(J_k(i) = x)^{\frac{1}{2}} \\ &\leq \sum_{x \geq i} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \sum_{j=0}^{x-1} \left| \mathbf{e}^{(j)} \right| \right|^{2p} \right]^{\frac{1}{2}} \nu_{\mathbf{q}}(J_k(i) = x)^{\frac{1}{2}} \\ &\leq \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \mathbf{e}^{(0)} \right|^{2p} \right]^{\frac{1}{2}} \sum_{x \geq i} x \nu_{\mathbf{q}}(J_k(i) = x)^{\frac{1}{2}} \\ &< \infty. \end{aligned}$$

Hence we have (11.5).

Now we show (3). From (11.1), we have

$$\begin{aligned} &\left| \frac{1}{n} X_k^i(\eta, n) - v_k^{\text{eff}}(\mathbf{q}) \right|^p \\ &\leq \left| \frac{1}{n} X_k^{(1)}(\eta, n) - v_k^{\text{eff}}(\mathbf{q}) \right|^p + \left| \frac{1}{n} X_k^{(i)}(\eta, n) - v_k^{\text{eff}}(\mathbf{q}) \right|^p. \end{aligned}$$

Then from Theorem 4.2, (11.5) and the Minkowski inequality, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \frac{1}{n} X_k^{(i)}(\eta, n) - v_k^{\text{eff}}(\mathbf{q}) \right|^p \right]^{\frac{1}{p}} \\ & \leq \lim_{n \rightarrow \infty} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| \frac{1}{n} Y_k^{(i)}(\eta, n) - v_k^{\text{eff}}(\mathbf{q}) \right|^p \right]^{\frac{1}{p}} + \frac{1}{n} \mathbb{E}_{\nu_{\mathbf{q}}} \left[ \left| X_k^{(i)}(0) \right|^p \right]^{\frac{1}{p}} \\ & = 0, \end{aligned}$$

for any  $i \in \mathbb{N}$ . Therefore Theorem 2.1 (3) is proved.  $\square$

#### ACKNOWLEDGMENT

The authors thank to Kenkichi Tsunoda for suggesting to consider the large deviation principle of the tagged soliton. The work of MS has been supported by JSPS KAKENHI GRANT No. 19H0179, 23K22414, 24K21515 and 24K00528. The work of HS has been supported by JSPS KAKENHI GRANT No. 24KJ1037, 24K16936 and 24K00528. SO appreciates the hospitality of the University of Tokyo.

#### APPENDIX A. PROOF OF LEMMA 3.5

For notational simplicity, we only consider the case  $n = 1$ . We can use the same proof presented below for any  $n \in \mathbb{N}$ .

We will quote some formulae and results from [S]. First, we recall that the carrier with capacity  $\ell \in \mathbb{N}$ , which is a variant of the carrier process, is defined as

$$W_\ell(s_\infty(i)) := 0, \\ W_\ell(x) - W_\ell(x-1) = \begin{cases} 1 & \text{if } \eta(x) = 1, W_\ell(x-1) < \ell \\ -1 & \text{if } \eta(x) = 0, W_\ell(x-1) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We note that from the construction of  $\mathcal{W}_k, k \in \mathbb{N}$ , for any  $\ell \in \mathbb{N}$  and  $x \in \mathbb{Z}$  we have the relation

$$W_\ell(x) = \sum_{k=1}^{\ell} \mathcal{W}_k(x). \quad (\text{A.1})$$

Next, from Remark 3.2 and [S, Lemma 4.2], we see that for any  $\gamma \in \Gamma_k$ ,  $k \in \mathbb{N}$ ,  $X(\gamma)$  is either a record or a  $(\ell, \sigma)$ -seat with  $\ell \geq k$  and  $\sigma \in \{0, 1\}$ . In particular, for any  $\gamma \in \Gamma_k$ , there exists some  $i \in \mathbb{Z}$  such that

$$X(\gamma) = s_k(i). \quad (\text{A.2})$$

From (A.1), (A.2) and [S, Lemma 4.2], we see that for any  $\gamma \in \Gamma_k$ ,  $k \in \mathbb{N}$  and  $1 \leq \ell \leq k$ ,

$$W_\ell(X(\gamma)) = \begin{cases} \ell & \text{if } \eta(X(\gamma)) = 1, \\ 0 & \text{if } \eta(X(\gamma)) = 0. \end{cases}$$

From the proof of [S, Theorem 4.1], for any  $\ell \in \mathbb{N}$  and  $x \in \mathbb{Z}$ , we get

$$\xi_\ell(T\eta, x) - \xi_\ell(\eta, x) = W_\ell(T\eta, x) + W_\ell(\eta, x) + o_\ell(\eta).$$

If the  $i$ -th  $k$ -soliton is free, then  $X_k^{(i)}$  is a record, and  $X_k^{(i)}(1) = T_1(\gamma_k^{(i)}) - 1$ . We observe that from the TS algorithm, if a  $\ell$ -soliton  $\gamma$  is contained in  $(H_1(\gamma_k^{(i)}), T_k(\gamma_k^{(i)}))$ , we have either  $\gamma \subset [H_1(\gamma_k^{(i)}), T_1(\gamma_k^{(i)})]$  or  $\gamma \subset (T_1(\gamma_k^{(i)}), T_k(\gamma_k^{(i)}))$ . From this observation, Remark 3.2 and [S, Lemma 4.2], for any  $1 \leq \ell \leq k$  we have

$$W_\ell(\eta, X_k^{(i)}(1)) = \sum_{h=1}^{\ell} \mathcal{W}_h(\eta, T_1(\gamma_k^{(i)}) - 1) = \sum_{h=1}^{\ell} \mathcal{W}_h(\eta, H_k(\gamma_k^{(i)})) = \ell.$$

In addition, since  $T\eta(X_k^{(i)}(1)) = 0$ , for any  $1 \leq \ell \leq k$  we have

$$W_\ell(T\eta, X_k^{(i)}(1)) = 0.$$

On the other hand, if the  $i$ -th  $k$ -soliton is not free, then  $X_k^{(i)}(1) = X_k^{(i)}$ . In addition,  $\eta(X_k^{(i)}) = 1 - T\eta(X_k^{(i)}) = 0$ . Thus for any  $1 \leq \ell \leq k$  we get

$$W_\ell(\eta, X_k^{(i)}(1)) = 0, \quad W_\ell(T\eta, X_k^{(i)}(1)) = \ell.$$

From the above, for any  $1 \leq \ell \leq k$  we have

$$\xi_\ell(T\eta, X_k^{(i)}(1)) - \xi_\ell(\eta, X_k^{(i)}(1)) = \ell + o_\ell. \quad (\text{A.3})$$

Now we assume that the  $i$ -th  $k$ -soliton is not free at time 0. Then we have  $X_k^{(i)}(1) = X_k^{(i)}$ , and thus from (A.3) we obtain (3.8) for this case. Next we assume that the  $i$ -th  $k$ -soliton is free at time 0. In this case we obtain

$$\begin{aligned} & \xi_\ell(\eta, X_k^{(i)}(1)) - \xi_\ell(\eta, X_k^{(i)}) \\ &= \sum_{y \in [X_k^{(i)+1}, X_k^{(i)}(1)]} \sum_{h \in \mathbb{N}} (\eta_{\ell+h}^\uparrow(y) + \eta_{\ell+h}^\downarrow(y)) \\ &= \sum_{y \in [X_k^{(i)+1}, X_k^{(i)}(1)] \cap \gamma_k^{(i)}} \sum_{h=\ell+1}^k (\eta_h^\uparrow(y) + \eta_h^\downarrow(y)) \\ &+ \sum_{y \in [X_k^{(i)+1}, X_k^{(i)}(1)] \cap (\gamma_k^{(i)})^c} \sum_{h=\ell+1}^{k-1} (\eta_h^\uparrow(y) + \eta_h^\downarrow(y)), \end{aligned}$$

where we use the fact that in the interval  $[H_1(\gamma_k^{(i)}), T_k(\gamma_k^{(i)})]$ , there are only  $(h, \sigma)$ -seats with  $h \leq k$ , and all  $(k, \sigma)$ -seats are elements of  $\gamma_k^{(i)}$ . For the first term, we get

$$\sum_{y \in [X_k^{(i)+1}, X_k^{(i)}(1)] \cap \gamma_k^{(i)}} \sum_{h=\ell+1}^k (\eta_{\ell+h}^\uparrow(y) + \eta_{\ell+h}^\downarrow(y)) = \sum_{y \in \gamma_k^{(i)}} \sum_{h=\ell+1}^k \eta_{\ell+h}^\uparrow(y) = k - \ell.$$

For the second term, we observe that if  $\left[X_k^{(i)} + 1, X_k^{(i)}(1)\right] \cap (\gamma_k^{(i)})^c$  is not empty, then each element is a component of some  $h$ -soliton  $\gamma$  with  $h < k$ , and  $\gamma \subset \left[X_k^{(i)} + 1, X_k^{(i)}(1)\right]$ . In addition, a  $h$ -soliton is composed by one of each  $(h', \sigma)$ -seats for  $1 \leq h' \leq h$  and  $\sigma \in \{\uparrow, \downarrow\}$ . Hence for any  $1 \leq h \leq k-1$ , we have

$$\sum_{y \in \left[X_k^{(i)} + 1, X_k^{(i)}(1)\right] \cap (\gamma_k^{(i)})^c} \eta_h^\uparrow(y) = \sum_{y \in \left[X_k^{(i)} + 1, X_k^{(i)}(1)\right] \cap (\gamma_k^{(i)})^c} \eta_h^\downarrow(y),$$

and

$$N_{k,h}^{(i)}(1) = \sum_{y \in \left[X_k^{(i)} + 1, X_k^{(i)}(1)\right] \cap (\gamma_k^{(i)})^c} (\eta_h^\uparrow(y) - \eta_{h+1}^\uparrow(y)).$$

Thus we get

$$\begin{aligned} & \sum_{y \in \left[X_k^{(i)} + 1, X_k^{(i)}(1)\right] \cap (\gamma_k^{(i)})^c} \sum_{h=\ell+1}^{k-1} (\eta_h^\uparrow(y) + \eta_h^\downarrow(y)) \\ &= 2 \sum_{y \in \left[X_k^{(i)} + 1, X_k^{(i)}(1)\right] \cap (\gamma_k^{(i)})^c} \sum_{h=\ell+1}^{k-1} \eta_h^\uparrow(y) \\ &= 2 \sum_{y \in \left[X_k^{(i)} + 1, X_k^{(i)}(1)\right] \cap (\gamma_k^{(i)})^c} \sum_{h=\ell+1}^{k-1} (h-\ell) (\eta_h^\uparrow(y) - \eta_{h+1}^\uparrow(y)) \\ &= 2 \sum_{h=\ell+1}^{k-1} (h-\ell) N_{k,h}^{(i)}(1). \end{aligned}$$

From the above, we have

$$\xi_\ell(\eta, X_k^{(i)}(1)) - \xi_\ell(\eta, X_k^{(i)}) = k - \ell + 2 \sum_{h=\ell+1}^{k-1} (h-\ell) N_{k,h}^{(i)}(1),$$

and thus from (A.3) we obtain (3.8) when the  $i$ -th  $k$ -soliton is free.

## APPENDIX B. COMPUTATIONS OMITTED IN SECTION 4

In this section we derive (4.9), (4.11), (4.12) and (4.18), whose proofs were omitted in Section 4.

**B.1. Derivation of (4.9).** First we observe that

$$\begin{aligned} \mathbb{E}_{\nu_{\mathbf{q}}} [r(0)] &= \nu_{\mathbf{q}}(\eta(x) = T\eta(x) = 0) \\ &= 1 - \nu_{\mathbf{q}}(\eta(x) = 1) - \nu_{\mathbf{q}}(T\eta(x) = 1) \\ &= 1 - 2\nu_{\mathbf{q}}(\eta(x) = 1). \end{aligned} \tag{B.1}$$

On the other hand, since  $\nu_{\mathbf{q}}$  is shift-ergodic, by the ergodic theorem (cf. [B, Theorem 6.21]), we have

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{x=1}^L \eta(x) = \nu_{\mathbf{q}}(\eta(x) = 1) \quad \nu_{\mathbf{q}}\text{-a.s.}$$

In addition, from Remark 4.1,  $s_{\infty}(\eta, i+1) - s_{\infty}(\eta, i)$ ,  $i \in \mathbb{Z}$  are i.i.d. under  $\nu_{\mathbf{q}}$ , and thus we get

$$\begin{aligned} & \lim_{L \rightarrow \infty} \frac{1}{s_{\infty}(\eta, L) - s_{\infty}(\eta, 0)} \sum_{x=s_{\infty}(\eta, 0)+1}^{s_{\infty}(\eta, L)} \eta(x) \\ &= \lim_{L \rightarrow \infty} \frac{1}{s_{\infty}(\eta, L) - s_{\infty}(\eta, 0)} \sum_{i=0}^{L-1} \frac{s_{\infty}(\eta, i+1) - s_{\infty}(\eta, i) - 1}{2} \\ &= \frac{1}{2} - \lim_{L \rightarrow \infty} \frac{L}{2(s_{\infty}(\eta, L) - s_{\infty}(\eta, 0))} \\ &= \frac{1}{2} - \frac{1}{\bar{s}_{\infty}(\mathbf{q})} \quad \nu_{\mathbf{q}}\text{-a.s.} \end{aligned}$$

where at the second line we use the fact that the number of balls in the  $i$ -th excursion is equal to  $(s_{\infty}(\eta, i+1) - s_{\infty}(\eta, i) - 1)/2$ , and the last line we use the law of large numbers for  $s_{\infty}(\eta, i+1) - s_{\infty}(\eta, i)$ ,  $i \in \mathbb{Z}$ . Hence we have

$$\bar{r}_k(\mathbf{q}) = \frac{1}{\bar{s}_{\infty}(\theta^k \mathbf{q})}.$$

Now we fix an excursion  $\mathbf{e} \in \mathcal{E}$ . Note that  $\mathbf{e}$  can be regarded as an element of  $\Omega$ , by considering  $\eta = \eta(\mathbf{e})$  as

$$\eta(x) = \begin{cases} \mathbf{e}(x+1) & 0 \leq x \leq |\mathbf{e}| - 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then, we can apply  $\Psi_{\ell}$  to  $\mathbf{e}$ , and we will write  $\Psi_{\ell}(\varepsilon)$  instead of  $\Psi_{\ell}(\eta(\mathbf{e}))$ . Then, we have the following formula for  $\mathbf{e}$ .

$$\begin{aligned} |\mathbf{e}| &= 1 + \sum_{\ell=1}^{\infty} \sum_{x=1}^{|\mathbf{e}|-1} (\eta_{\ell}^{\uparrow}(x) + \eta_{\ell}^{\downarrow}(x)) \\ &= 1 + 2 \sum_{\ell=1}^{\infty} \sum_{j=0}^{|\Psi_{\ell}(\varepsilon)|-1} \ell \zeta_{\ell}(j), \end{aligned}$$

where at the last line we use (5.1) to derive

$$\begin{aligned} \xi_{\ell}(\eta(\mathbf{e}), |\mathbf{e}| - 1) &= 1 + \sum_{h=\ell+1}^{\infty} \sum_{x=1}^{|\mathbf{e}|-1} (\eta_h^{\uparrow}(x) + \eta_h^{\downarrow}(x)) \\ &= |\Psi_{\ell}(\varepsilon)|. \end{aligned}$$

By using the above, (5.4) and Remark 5.7, we have

$$\begin{aligned} \frac{1}{\bar{r}_k(\mathbf{q})} &= \mathbb{E}_{\nu_{\theta^k \mathbf{q}}} [|\mathbf{e}|] = 1 + 2 \sum_{\ell=1}^{\infty} \ell \mathbb{E}_{\nu_{\theta^k \mathbf{q}}} [|\Psi_{\ell}(\varepsilon)|] \alpha_{\ell}(\theta^k \mathbf{q}) \\ &= 1 + 2 \sum_{\ell=1}^{\infty} (\ell - k) \frac{\alpha_{\ell}(\theta^k \mathbf{q})}{\bar{r}_{\ell}(\mathbf{q})}. \end{aligned}$$

Hence we have (4.9).

**B.2. Proof of Proposition 4.4.** First we derive (4.11) From (5.12) and (5.13), we have

$$\frac{1}{n} Y_k^{(i)}(\eta, n) = \frac{k}{n} Y_1^{(i)}(\Psi_{k-1}(\eta), n) + \frac{2}{n} \sum_{\ell=1}^{k-1} \ell \sum_{j=X_{k-\ell}^{(i)}(\Psi_{\ell}(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_{\ell}(\tilde{\eta}), n)} \zeta_{\ell}(\tilde{\eta}, j).$$

From (4.10), by taking  $n \rightarrow \infty$  we have

$$\frac{1}{n} Y_k^{(i)}(\eta, n) = v_k^{\text{eff}}(\mathbf{q}) \quad \nu_{\mathbf{q}}\text{-a.s.}$$

and

$$\frac{1}{n} Y_1^{(i)}(\Psi_{k-1}(\eta), n) = v_1^{\text{eff}}(\theta^{k-1} \mathbf{q}) \quad \nu_{\mathbf{q}}\text{-a.s.}$$

In addition, since  $X_{k-\ell}^{(i)}(\Psi_{\ell}(\tilde{\eta}), n)$  is  $\sigma(\zeta_h; h \geq \ell + 1)$ -m'ble for any  $1 \leq \ell \leq k-1$  and  $n \in \mathbb{Z}_{\geq 0}$ , by (4.10) and Remark 5.7, we have

$$\frac{1}{n} \sum_{j=X_{k-\ell}^{(i)}(\Psi_{\ell}(\tilde{\eta}), 0)+1}^{X_{k-\ell}^{(i)}(\Psi_{\ell}(\tilde{\eta}), n)} \zeta_{\ell}(\eta, j) = \alpha_{\ell}(\mathbf{q}) v_{k-\ell}^{\text{eff}}(\theta^{\ell} \mathbf{q}) \quad \nu_{\mathbf{q}}\text{-a.s.}$$

From the above, we have (4.11).

Next, we show (4.12). From (5.12) and (5.14), we have

$$\frac{1}{n} Y_1^{(i)}(\Psi_{k-1}(\eta), n) = \frac{1}{n} \sum_{m=0}^{n-1} r(T^m \Psi_k(\tilde{\eta}), J_k(\tilde{\eta}, i)).$$

Since  $Y_1^{(i)}(\Psi_{k-1}(\eta), n)$  converges to  $v_1^{\text{eff}}(\theta^{k-1} \mathbf{q})$ , by Remark 5.7, we see that if

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=0}^{n-1} r(T^m \Psi_k(\eta), x) = \bar{r}_k(\mathbf{q}), \quad (\text{B.2})$$

$\nu_{\mathbf{q}}(\cdot | s_{\infty}(0) = 0)$ -a.s. for any  $x \in \mathbb{Z}$ , then we obtain (4.12). To show (B.2), we observe that by  $T$ -invariance of  $\nu_{\mathbf{q}}$  and the ergodic theorem, we see that  $n^{-1} \sum_{m=0}^{n-1} r(T^m \Psi_k(\eta), x)$  converges a.s. to  $\mathbb{E}_{\nu_{\mathbf{q}}}[r(T^m \Psi_k(\eta), x) | \mathcal{I}]$ , where  $\mathcal{I}$  is the set of invariant sets of  $T$ . On the other hand, since  $\nu_{\mathbf{q}}$  is shift-ergodic and the limit  $\mathbb{E}_{\nu_{\mathbf{q}}}[r(T^m \Psi_k(\eta), x) | \mathcal{I}]$  is shift-invariant, we see that  $\mathbb{E}_{\nu_{\mathbf{q}}}[r(T^m \Psi_k(\eta), x) | \mathcal{I}]$  is a.s. constant. Hence we have the limit (B.2)  $\nu_{\mathbf{q}}$ -a.s., and this implies (B.2)  $\nu_{\mathbf{q}}(\cdot | s_{\infty}(0) = 0)$ -a.s. Thus (4.12) is proved.



**B.3. Proof of (4.18).** From (7.6), the derivative of  $\Lambda_{\mathbf{q},k}^Y$  with  $\lambda = 0$  is given by

$$\frac{d\Lambda_{\mathbf{q},k}^Y}{d\lambda}(0) = \frac{dU_{\mathbf{q},k}}{d\lambda}(0) \frac{d\Lambda_{\mathbf{q},k}^M}{d\lambda}(0),$$

where  $U_{\mathbf{q},k}(\lambda)$  is defined in (5.26). First we check that the expression (4.18) is the same as [FNRW, (1.12)]. We observe that  $\frac{dU_{\mathbf{q},k}}{d\lambda}(0)$  satisfies the following system,

$$\frac{dU_{\mathbf{q},k}}{d\lambda}(0) = k + 2 \sum_{\ell=1}^{k-1} (k-\ell) \frac{dU_{\mathbf{q},\ell}}{d\lambda}(0). \quad (\text{B.3})$$

On the other hand, from (4.9), (4.12) and (5.12), we have

$$\begin{aligned} \frac{d\Lambda_{\mathbf{q},k}^M}{d\lambda}(0) &= \frac{d\Lambda_{\theta^{k-1}\mathbf{q},1}^Y}{d\lambda}(0) \\ &= \bar{r}_k(\mathbf{q}). \end{aligned}$$

Then by combining (4.9) and (B.3), we see that  $\left(\frac{dU_{\mathbf{q},k}}{d\lambda}(0), \bar{r}_k^{-1}(\mathbf{q})\right)$  coincide with the quantities  $(s_k, w_k)$  in [FNRW, (1.12)], respectively, and that (4.18) and [FNRW, (1.12)] are the same.

To show (4.18), it is sufficient to prove that  $\frac{dU_{\mathbf{q},k}}{d\lambda}(0) = v_k^{\text{eff}}(C_k\mathbf{q})$ . Since  $M_k^{(i)}(\cdot) = 0$  a.s. under  $\nu_{C_k\mathbf{q}}$ , we have

$$v_k^{\text{eff}}(C_k\mathbf{q}) = \frac{dU_{C_k\mathbf{q},k}}{d\lambda}(0).$$

On the other hand, from (B.3), we get

$$\frac{dU_{C_k\mathbf{q},k}}{d\lambda}(0) = \frac{dU_{\mathbf{q},k}}{d\lambda}(0).$$

Hence  $\frac{dU_{\mathbf{q},k}}{d\lambda}(0) = v_k^{\text{eff}}(C_k\mathbf{q})$ , and thus we have (4.18).

#### APPENDIX C. PROOF OF LEMMA 11.1

From [FG, Lemma 3.7], if we write  $\tilde{\nu}_{\mathbf{q}}$  the distribution of  $\mathbf{e}^{(0)}$  on  $\mathcal{E}$  under  $\nu_{\mathbf{q}}$ , then the probability  $\tilde{\nu}_{\mathbf{q}}(\mathbf{e})$ ,  $\mathbf{e} \in \mathcal{E}$  is

$$\begin{aligned} \tilde{\nu}_{\mathbf{q}}(\mathbf{e}) &= \nu_{\mathbf{q}}(\eta(1) = 0 | \eta(0) = 0) \prod_{k \in \mathbb{N}} \left( a(\mathbf{q}) b(\mathbf{q})^{k-1} \right)^{\zeta_k(\mathbf{e})} \\ &= \nu_{\mathbf{q}}(\eta(1) = 0 | \eta(0) = 0) \left( a'(\mathbf{q}) \right)^{\sum_{k \in \mathbb{N}} \zeta_k(\mathbf{e})} b(\mathbf{q})^{\frac{|\mathbf{e}|-1}{2}}, \end{aligned}$$

where  $a'(\mathbf{q}) := a(\mathbf{q})/b(\mathbf{q})$  and  $\zeta_k(\mathbf{e})$  is the total number of  $k$ -solitons in  $\mathbf{e}$ . We observe that  $\sum_{k \in \mathbb{N}} \zeta_k(\mathbf{e})$  is equal to the number of  $1 \leq x \leq |\mathbf{e}|$  such that

$\mathbf{e}(x) = 1, \mathbf{e}(x+1) = 0$  and it is known that

$$\begin{aligned} & |\{\mathbf{e} \in \mathcal{E}(m) ; |\{1 \leq x \leq 2m+1 ; \mathbf{e}(x) = 1, \mathbf{e}(x+1) = 0\}| = z\}| \\ &= \frac{1}{m} \binom{m}{z} \binom{m}{z-1}, \end{aligned}$$

for any  $m \in \mathbb{N}$ , where the right-hand side is called the Narayana numbers. Hence, we get

$$\begin{aligned} & \mathbb{E}_{\tilde{\nu}} \left[ e^{\lambda|\mathbf{e}|} \right] \nu_{\mathbf{q}} (\eta(1) = 0 | \eta(0) = 0)^{-1} \\ &= 1 + \sum_{m=1}^{\infty} \sum_{z=1}^m \frac{1}{m} \binom{m}{z} \binom{m}{z-1} (a'(\mathbf{q}))^z (e^{2\lambda} b(\mathbf{q}))^m \\ &= 1 + \frac{a(\mathbf{q})}{1 - e^{2\lambda} (a(\mathbf{q}) + b(\mathbf{q})) + \sqrt{(1 - e^{2\lambda} (a(\mathbf{q}) + b(\mathbf{q})))^2 - 4e^{2\lambda} a(\mathbf{q}) b(\mathbf{q})}}, \end{aligned}$$

where we use the fact that the generating function of the Narayana numbers  $F(a, b)$  is given by

$$F(a, b) := \frac{1 - b(1+a) - \sqrt{(1 - b(1+a))^2 - 4ab^2}}{2b}.$$

From the above, for sufficiently small  $\lambda > 0$ , we have  $\mathbb{E}_{\tilde{\nu}} [e^{\lambda|\mathbf{e}|}] < \infty$ .

#### REFERENCES

- [BDS] C. BOLDRIGHINI, R. L. DOBRUSHIN AND Y. M. SUKHOV : One-dimensional hard rod caricature of hydrodynamics. J. Stat. Phys. **31** 577–616 (1983)
- [B] L. BREIMAN : Probability. SIAM Classics in Applied Mathematics (1992)
- [CS] D. A. CROYDON, M. SASADA : Generalized Hydrodynamic Limit for the Box-Ball System. Commun. Math. Phys. **383**, 427-463 (2021)
- [CS2] D. A. CROYDON, SASADA, M. : Invariant measures for the box-ball system based on stationary Markov chains and periodic Gibbs measures. J. Math. Phys. **60**, 083301 (2019)
- [CKST] D. A. CROYDON, T. KATO, M. SASADA, S. TSUJIMOTO : Dynamics of the box-ball system with random initial conditions via Pitman’s transformation. to appear in Mem. Amer. Math. Soc., preprint appears at arXiv:1806.02147, 2018
- [D] B. DOYON : Lecture notes on generalised hydrodynamics. SciPost Phys. Lect. Notes **18** (2020)
- [DZ] A. DEMBO AND O. ZEITOUNI : Large Deviation Techniques and Applications. Stochastic Modelling and Applied Probability, vol. 38, Springer-Verlag, Berlin, 2010, Corrected reprint of the second (1998) edition.
- [DBD] J. DE NARDIS, D. BERNARD AND B. DOYON : Hydrodynamic diffusion in integrable systems. Phys. Rev. Lett. **121** 160603 (2018).
- [DBD2] J. DE NARDIS, D. BERNARD AND B. DOYON : Diffusion in generalized hydrodynamics and quasiparticle scattering. SciPost Phys. **6** 049 (2019).
- [DDMP] J. DE NARDIS, B. DOYON, M. MEDENJAK AND M. PANFIL : Correlation functions and transport coefficients in generalised hydrodynamics. J. Stat. Mech. (2022) 014002
- [EK] S. N. ETHIER AND T.G. KURTZ : Markov Processes: Characterization and Convergence. Wiley series in probability and mathematical statistics: probability and mathematical statistics. Wiley, New York

- [FFGS] P. A. FERRARI, C. FRANCESCHINI, D. G. E. GREVINO AND AND H. SPOHN : Hard rod hydrodynamics and the Lévy Chentsov field. *Ensaos Mat.* 38. Rio de Janeiro: Sociedade Brasileira de Matemática.
- [FG] P. A. FERRARI, D. GABRIELLI : BBS invariant measures with independent soliton components. *Electron. J. Probab.* 25: 1-26 (2020)
- [FNRW] P. A. FERRARI, C. NGUYEN, L. ROLLA, AND M. WANG : Soliton decomposition of the box-ball system. *Forum of mathematics, Sigma* 9 (2021).
- [FO] P. A. FERRARI, S. OLLA : Macroscopic diffusive fluctuations for generalized hard rods dynamics. arXiv:2305.13037
- [IKT] R. INOUE, A. KUNIBA AND T. TAKAGI : Integrable structure of box–ball systems: crystal, Bethe ansatz, ultradiscretization and tropical geometry. *J. Phys. A: Math. Theor.* 45(7), 073001 (2012)
- [KLO12] T. KOMOROWSKI, C. LANDIM AND S. OLLA : Fluctuations in Markov Processes. Springer, Berlin (2012)
- [KL] A. KUNIBA, H. LYU : Large Deviations and One-Sided Scaling Limit of Randomized Multicolor Box-Ball System. *J Stat Phys* 178, 38–74 (2020)
- [KLO18] A. KUNIBA, H. LYU, M. OKADO : Randomized box-ball systems, limit shape of rigged configurations and thermodynamic Bethe ansatz. *Nuclear Physics B* 937 240–271 (2018)
- [KMP] A. KUNIBA, G. MISGUICH, V. PASQUIER : Generalized hydrodynamics in box-ball system. *J. Phys. A: Math. Theor.* 53 404001 (2020)
- [KMP2] A. KUNIBA, G. MISGUICH, V. PASQUIER : Generalized hydrodynamics in complete box-ball system for  $U_q(\mathfrak{sl}_n)$ . *Scipost phys.* 10, 095 (2021)
- [KMP3] A. KUNIBA, G. MISGUICH, V. PASQUIER : Current correlations, Drude weights and large deviations in a box-ball system. To appear in *J. Phys. A*
- [KNTW] SABURO KAKEI, JONATHAN J C NIMMO, SATOSHI TSUJIMOTO, RALPH WILLOX : Linearization of the box-ball system: an elementary approach. *Journal of Integrable Systems*. Volume 3, Issue 1, 2018, xyy002
- [KOSTY] A. KUNIBA, M. OKADO, R. SAKAMOTO, T. TAKAGI AND Y. YAMADA : Crystal interpretation of Kerov-Kirillov-Reshetikhin bijection. *Nuclear Physics B*, 740, 299–327 (2006)
- [KOY] A. KUNIBA, M. OKADO, Y. YAMADA : Box–ball system with reflecting end. *J. Nonlin. Math. Phys.* 12 475–507 (2005)
- [KTT] A. KUNIBA, T. TAKAGI AND A. TAKENOUCI : Bethe ansatz and inverse scattering transform in a periodic box-ball system. *Nucl. Phys. B* 747, 354–397 (2006)
- [LLP] L. LEVINE, H. LYU, J. PIKE : Double Jump Phase Transition in a Soliton Cellular Automaton. *Int Math Res Notices* volume 2022, Issue 1, 665–727 (2020)
- [LLPS] J. LEWIS, H. LYU, P. PYLYAVSKYY, AND A. SEN : Scaling limit of soliton lengths in a multicolor box-ball system. arXiv:1911.04458
- [MIT] JUN MADA, MAKOTO IDZUMI AND TETSUJI TOKIHIRO : On the initial value problem of a periodic box-ball system. *J. Phys. A: Math. Gen.* 39 (2006)
- [MSSS] M. MUCCICONI, M. SASADA, T. SASAMOTO AND H. SUDA : Relationships between two linearizations of the box-ball system : Kerov-Kirillov-Reshetikhin bijection and slot configuration. *Forum of Mathematics, Sigma.* 2024;12:e55.
- [Sp] H. SPOHN : Hydrodynamic Scales of Integrable Many-Body Systems. World Scientific Publishing (2024)
- [S] H. SUDA : Seat number configuration of the box-ball system, and its relation to the 10-elimination and invariant measures. arXiv:2312.15358
- [T] T. TAKAGI : Inverse scattering method for a soliton cellular automaton. *Nuclear Physics B* 707, 577–601.
- [TM] D. TAKAHASHI AND J. MATSUKIDAIRA : Box and ball system with a carrier and ultra-discrete modified KdV equation. *J. Phys. A* 30 L733–L739 (1997).

- [TS] D. TAKAHASHI AND J. SATSUMA : A soliton cellular automaton. J. Phys. Soc. Japan **59** 3514–3519 (1990)