Random processes with long memory

outline of PhD thesis

Illés Horváth

Institute of Mathematics, Budapest University of Technology and Economics

Supervisor: prof. Bálint Tóth
Advisor: prof. Miklós Telek

2015

Introduction

Markov processes have been widely examined; the theory is well developed and applications are abundant. Different fields of applications include statistical mechanics, chemistry, economics, population dynamics and queueing theory.

As models become more and more complicated, a natural need arises to extend results available for Markov processes to systems where the Markov property does not fully hold, that is, to random processes with long memory. The exact nature of the memory in such systems can be very different; in mathematical physics, examples include interacting particle systems or a single particle moving in a random environment. In these cases, the memory corresponds to the state of the environment. In queueing theory, non-exponential service or interarrival times lead to M/G/1 and G/M/1 queues respectively; in such cases, the memory corresponds to the age of non-exponential clocks.

Non-Markovian behaviour can be handled using several different approaches. First, the state space may be extended to include more information about the process in order to make it Markovian. The difficulty of this approach is that the state space may end up being extremely large and difficult to handle. Nevertheless, this approach works for many physical systems, and theory has been constantly developed over the last decades. Initiated by Kipnis and Varadhan [16], martingale approximation techniques provide central limit theorem for a wide class of interacting particle systems and random walks in random environments. Traditionally, sufficient conditions that guarantee central limit theorem have been called *sector conditions*. One specific type of random walks of interest is the *"true" (myopic) self-avoiding walk*, where the walker is pushed to areas less-visited by the random walk. It has been introduced in the physics literature in the 1980's [1].

For basic queueing models, matrix analytic methods are available, and direct calculations are also possible using Laplace–Stieltjes transform. These are established and straightforward methods. For more involved queuing models, another way to handle non-Markovian behaviour is via approximation by Markovian processes. General distributions may be approximated by specific classes of distributions that result in Markovian models. One of the most relevant classes of distributions for Markovian modelling is *phase-type distributions*.

Behaviour of Markov population models with a large population size has been examined extensively in the last few decades. One of the first results is Kurtz's theorem [20], which describes the mean-field limit (the limit of the system as population size goes to infinity) as the solution of a system of ordinary differential equations. Since then, a lot of effort from people from many different backgrounds (e.g. computer science, biology, chemistry) has been spent to extend the classic model to include a varying degree of non-Markovian behaviour. One such class of population models is the so-called *population generalized semi-Markov process*.

The thesis is divided into four chapters.

The topic of Chapter 2 is sufficient conditions for the martingale approximation and central limit theorem for functionals of stationary and ergodic Markov chains. It is joint work with B. Tóth and B. Vető and is based on [15] and parts of [14].

The topic of Chapter 3 is the behaviour of the "true" (myopic) self-avoiding walk in $d \geq 3$. It is

joint work with B. Tóth and B. Vető and is based on parts of [14].

The topic of Chapter 4 is O'Cinneide's characterization theorem for phase-type distributions. The main result is a new, constructive proof of the sufficient direction of the characterization theorem. It is joint work with M. Telek. The chapter is based on [13].

The topic of Chapter 5 is the mean-field limit for a class of population generalized semi-Markov processes (PGSMPs). It is joint work with M. Telek and R. Hayden and is based on [11].

In the rest of this outline, we first give a background on each of the four topics, then present the results of the thesis in four numbered sections.

Central limit theorem for functionals of stationary and ergodic Markov processes; sector conditions

Let $(\Omega, \mathcal{F}, \pi)$ be a probability space: Ω is the state space of a *stationary and ergodic* Markov process $t \mapsto \eta(t)$. We put ourselves in the Hilbert space $\mathcal{H} := \mathcal{L}^2(\Omega, \pi)$. Denote the *infinitesimal generator* of the semigroup of the process by G, which is a well-defined (possibly unbounded) closed linear operator on $\mathcal{H}$.

Let $f \in \mathcal{H}$, such that $(f, \mathbb{1}) = \int_\Omega f \, \mathrm{d}\pi = 0$. We ask about central limit theorem/invariance principle for

$$N^{-1/2} \int_0^{Nt} f(\eta(s)) \, \mathrm{d}s \tag{1}$$

as $N \to \infty$.

The main tool to prove martingale approximation and central limit theorem is the following theorem by Kipnis and Varadhan [16]:

Theorem KV. *With the notation and assumptions as before, if the following two limits hold in $\mathcal{H}$:*

$$\lim_{\lambda \to 0} \lambda^{1/2} \big(\lambda I - G\big)^{-1} f = 0,$$
$$\lim_{\lambda \to 0} S^{1/2} \big(\lambda I - G\big)^{-1} f =: v \in \mathcal{H},$$

then there exists a zero mean $\mathcal{L}^2$-martingale $M(t)$ adapted to the filtration of the Markov process $\eta(t)$, with stationary and ergodic increments such that

$$\lim_{N \to \infty} N^{-1} \mathbf{E} \left(\Big(\int_0^N f(\eta(s)) \, \mathrm{d}s - M(N) \Big)^2 \right) = 0.$$

Theorem KV is an abstract result that might be difficult to apply to specific models; hence the need to formulate consequences that are easier to check.

In [16], Kipnis and Varadhan proved that in the reversible case (when G is self-adjoint), Theorem KV can be applied and thus martingale approximation and central limit theorem hold with no extra assumptions other than finiteness of the asymptotic variance of f (which is clearly necessary):

Theorem. *If $G = G^*$, then martingale approximation and central limit theorem applies to every function f with*

$$\overline{\lim_{t\to\infty}}\, t^{-1}\mathbf{E}\left((\int_0^t f(\eta(s))\,\mathrm{d}s)^2\right) < \infty. \tag{2}$$

Later, more general sufficient conditions have been formulated, known collectively as *sector conditions*.

In 1996, Varadhan introduced the *strong sector condition (SSC)* [37]; using the notation

$$S := -\frac{1}{2}(G + G^*), \qquad A := \frac{1}{2}(G - G^*).$$

for the symmetric and antisymmetric parts of the operator G.

Theorem (SSC)**.** *If*

$$\left\| S^{-1/2} A S^{-1/2} \right\| < \infty.$$

holds, then martingale approximation and central limit theorem applies to every function

$$f \in \mathrm{Ran}(S^{1/2}). \tag{3}$$

(3) is dubbed the H_{-1}*-condition*; it guarantees finiteness of the asymptotic variance of f in 2.

In 2000, Sethuraman, Varadhan and Yau proved the *graded sector condition (GSC)* [30]. Assume that the space has an orthogonal decomposition (in other words, the space is graded):

$$\mathcal{H} = \overline{\oplus_{n=1}^{\infty} \mathcal{H}_n} \tag{4}$$

such that the grading is consistent with S and A in the sense that S is diagonal (with respect to the grading):

$$S = \sum_{n=0}^{\infty} S_{n,n}, \qquad S_{n,n} : \mathcal{H}_n \to \mathcal{H}_n \tag{5}$$

and A changes the grade by 1:

$$A = \sum_{n=0}^{\infty} A_{n,n+1} + A_{n,n-1}, \qquad A_{n,n+1} : \mathcal{H}_n \to \mathcal{H}_{n+1},\ A_{n,n-1} : \mathcal{H}_n \to \mathcal{H}_{n-1},\ A^*_{n,n+1} = -A_{n+1,n}. \tag{6}$$

Theorem (GSC)**.**

$$If \|S^{-1/2}_{n+i,n+i} A_{n,n+i} S^{-1/2}_{n,n}\| \le cn^{\beta} \qquad (i = \pm 1)$$

where $\beta < 1$ or $\beta = 1$ and c is small enough, then martingale approximation and central limit theorem applies to every function $f \in \mathrm{Ran}(S^{1/2})$.

For some applications, the GSC has been improved to allow some parts of A to be diagonal and some parts of S to be offdiagonal.

For applications for the various sector conditions, see the survey [17].

The results in Section 1 and [15] are the introduction of a new sector condition called the *relaxed sector condition* and an improved version of the graded sector condition.

"True" (or myopic) self-avoiding random walk

The *'true' (or myopic) self-avoiding walk* model (TSAW) was introduced in the physics literature by Amit, Parisi and Peliti in [1]. This is a nearest neighbor non-Markovian random walk in $\mathbb{Z}^d$ which prefers to jump to those neighbors which were less visited in the past. Long memory effects are caused by a path-wise self-repellence of the trajectories due to a push by the negative gradient of (softened) local time.

Let $t \mapsto X(t) \in \mathbb{Z}^d$ be a continuous time nearest neighbor jump process on the integer lattice $\mathbb{Z}^d$ whose law is given as follows:

$$\mathbf{P}\big(X(t+\mathrm{d}t) = y \;\big|\; \mathcal{F}_t, X(t) = x\big) = \mathbb{1}_{\{|\,x-y\,|=1\}} w(\ell(t,x) - \ell(t,y))\,\mathrm{d}t + o(\mathrm{d}t) \tag{7}$$

where

$$\ell(t,z) := \ell(0,z) + |\,\{0 \le s \le t : X(s) = z\}\,| \qquad z \in \mathbb{Z}^d \tag{8}$$

is the occupation time measure of the walk $X(t)$ with some initial values $\ell(0,z) \in \mathbb{R}$, $z \in \mathbb{Z}^d$, and the self-interaction rate function w is assumed to be increasing. This is a continuous time version of the *'true' self-avoiding random walk* defined in [1].

Non-rigorous (but nevertheless convincing) scaling and renormalization group arguments suggest the following dimension-dependent asymptotic scaling behaviour (see e.g. [1], [25], [28]):

– In $d = 1$: $X(t) \sim t^{2/3}$ with intricate, non-Gaussian scaling limit.

– In $d = 2$: $X(t) \sim t^{1/2}(\log t)^{\zeta}$ and Gaussian (that is Wiener) scaling limit expected. (We note that actually there is some controversy in the physics literature about the value of the exponent ζ in the logarithmic correction.)

– In $d \ge 3$: $X(t) \sim t^{1/2}$ with Gaussian (i.e. Wiener) scaling limit expected.

In $d = 1$, for some particular cases of the model (discrete time TSAW with edge, rather than site repulsion and continuous time TSAW with site repulsion, as defined above), the limit theorem for $t^{-2/3}X(t)$ was established in [33], respectively, [35] with the truly intricate limiting distribution identified. The limit of the *process* $t \mapsto N^{-2/3}X(Nt)$ was constructed and analyzed in [36].

In $d = 2$, superdiffusive lower bounds of order $t^{1/2}(\log\log t)^{1/2}$ for the isotropic case, respectively, of order $t^{1/2}(\log t)^{1/4}$ for the anisotropic case, have been proved for these two-dimensional models, cf. [34].

A closely related model to the TSAW is the so-called *self-repelling Brownian polymer*, which is essentially the continuous-space counterpart of TSAW. For diffusive bounds for the self-repelling Brownian polymer in 1-dimension, see [31], and for dimensions $d \geq 3$, see [14] and the PhD thesis of Bálint Vető [38]. *This model is not part of the present thesis.*

We address the $d \geq 3$ case in Section 2 and in [14]. We identify a natural stationary (in time) and ergodic distribution of the environment (the local time profile) as seen from the moving particle. For a wide class of self-interaction functions, we establish diffusive lower and upper bounds for the displacement and for a particular, more restricted class of interactions, we prove full CLT for the finite dimensional distributions of the displacement.

These results settle part of the conjectures in [1]. The proof of the CLT follows the non-reversible version of Kipnis – Varadhan theory, using an improved version of the graded sector condition.

Phase-type distributions

Consider a continuous-time Markov chain on $n+1$ states with exactly one absorbing state. We assume that the initial probability distribution of the absorbing state is 0. Let X denote the time of absorption; its probability density function (pdf) is the following function $f : \mathbb{R}^+ \to \mathbb{R}^+$:

$$f(t) = -\alpha \mathbf{A} e^{t\mathbf{A}} \mathbf{1}, \quad t \geq 0, \tag{9}$$

where α is the initial row vector of size n (not including the absorbing state), and $\mathbf{A}$ is the vanishing infinitesimal generator; it is essentially the infinitesimal generator of the Markov chain, with the absorbing state removed. That is, $\mathbf{A}$ is a substochastic matrix of size $n \times n$, where the sum of row i is equal to the negative of the rate of absorption from state i. $\mathbf{1}$ is the column vector of size n whose elements are all equal to 1. The 0 initial probability of absorption corresponds to $\alpha\mathbf{1} = 1$; equivalently, X does not have a probability mass at 0.

Distributions that can be obtained in the above form are called *phase-type distributions*; their class is denoted by PH.

Phase-type distributions can be used to approximate general distributions; PH is dense in total variation distance among all absolutely continuous positive distributions [5].

The pdf of a phase-type distribution is always analytic and takes the form

$$f(t) = \sum_i \sum_{j=1}^{n_i} c_{\lambda_i, j} t^{j-1} e^{-\lambda_i t}$$

where $-\lambda_i$ are the eigenvalues of $\mathbf{A}$, n_i is the multiplicity of λ_i and $c_{\lambda_i, j}$ are constants.

For a given f in PH, α and $\mathbf{A}$ are not unique; not even their dimensions are unique. Hence it makes sense to call the pair $(\alpha, \mathbf{A})$ a representation for f if (9) holds. The dimension of α (and $\mathbf{A}$) is called the order of the representation.

The class of *matrix exponential functions* (ME) is defined as follows:

Definition 1. *A nonnegative random variable X with probability density function f is in the class ME if there exists a vector α of size n and a matrix $\mathbf{A}$ of size $n \times n$ for some finite n such that*

$$f(t) = -\alpha \mathbf{A} e^{t\mathbf{A}} \mathbf{1}, \quad t \geq 0,$$

In this case, we will also say that f (and X) is ME($\alpha, \mathbf{A}$)-distributed.

The difference of an ME pdf compared to a PH pdf is that we do not pose nonnegativity conditions on α and $\mathbf{A}$. While the formula is exactly the same, if either α has negative or $\mathbf{A}$ has negative offdiagonal elements, the stochastic interpretation that X is the time of absorption of a Markov chain is no longer available.

Clearly, PH is a subclass of ME. The difference between the two classes is characterized due to O'Cinneide [26]. We need two more definitions.

Definition 2. *f satisfies the* positive density condition *if*

$$f(t) > 0 \quad \forall t > 0.$$

Note that the definition allows the density at 0 to be equal to 0.

Definition 3. *f satisfies the* dominant eigenvalue condition *if for some ME representation of minimal order ($\alpha, \mathbf{A}$) of f, $\mathbf{A}$ has a single eigenvalue with maximal real part.*

The dominant eigenvalue is always real to avoid oscillation of f around 0; the above definition excludes the case when a is the dominant real eigenvalue and there is a pair of complex eigenvalues with the same real part. However, the multiplicity of a may be higher than 1. We also remark that if the dominant eigenvalue condition holds for some minimal ME representation, it holds for all minimal ME representations of f.

Theorem 1 (O'Cinneide)**.** *If f_X is ME(α, $\mathbf{A}$) distributed, then f_X has a finite dimensional PH(β, $\mathbf{B}$) representation iff the following two conditions hold:*

- *f_X satisfies the dominant eigenvalue condition and*
- *f_X satisfies the positive density condition.*

The main importance of the theorem is the sufficient direction; that is, if the dominant eigenvalue condition and the positive density condition hold, then a PH representation always exists. O'Cinneide's original proof of the theorem is not constructive; Maier [22] offers another approach using automata-theoretic algorithms.

In Section 3 and in [11] we provide a constructive proof for the sufficient part of the characterization theorem. We propose an explicit procedure for computing a PH representation of a matrix exponential function and showing that the procedure always terminates successfully if the matrix exponential function satisfies the positive density condition and the dominant eigenvalue condition.

Compared to existing results, one of the main advantages of the presented constructive proof is that it is rather elementary, using basic function and matrix theory and stochastic interpretation of Markov processes. It also links more recent results (such as the sparse monocyclic representation of [8]) to the characterization theorem.

Generalized semi-Markovian population models

A (homogeneous) Markov population model is defined as follows. Fix a positive integer N. Each of N individuals is inhabiting a state from a finite set $\mathcal{S}$. Each individual performs Markov transitions in continuous time: an individual in state i transitions to state j with rate r_{ij}^N. The rates may depend on the *global state of* the system; the global state of the system is defined as the total number of individuals in each state, that is, a vector $\mathbf{x}^N \in (\{0, 1, \ldots, N\})^{|\mathcal{S}|}$ with $x_1^N + \cdots + x_{|\mathcal{S}|}^N = N$. It is easy to see that the global state of the system $\mathbf{x}^N(t)$ is a continuous-time Markov chain.

We are interested in the behaviour of such a system for large values of N. A usual assumption is that a family of Markov population models is *density-dependent*; this means that the transition rates depend only on the *normalized global state* of the sytem, independent of N. The normalized global state of the system is defined as $\bar{\mathbf{x}}^N = \frac{\mathbf{x}^N}{N}$.

Density-dependence commonly occurs in real-life scenarios in the field of chemistry (chemical reaction speed may be affected by concentration), biology and many computer network applications.

While the global state of the system is Markovian, an explicit analysis of this Markov chain is infeasible because the size of the state space increases exponentially in N.

The classic result of Kurtz [20] says that, upon some further regularity conditions (r_{ij} are Lipschitz-continuous and the initial conditions converge), the evolution of a density-dependent Markov population model converges to the solution of a system of ordinary differential equations (ODEs) as $N \to \infty$. The main advantage of Kurtz's approach is that the size of the system of equations is $|\mathcal{S}|$ regardless of N, thus avoiding the state-space explosion issue. Another consequence is that the limit is deterministic: for large values of N, the behaviour of the global state of the system is very close to deterministic. The deterministic limit is called the *mean-field limit* of the system.

Just like for the Markov population model described above, a *population generalized semi-Markov process (PGSMP)* has a finite local state space $\mathcal{S}$; each of N individuals is inhabiting a state from $\mathcal{S}$, but apart from each individual making Markov transitions, some of the states have a so-called *active clock*. When an individual enters a state with an active clock, a generally-timed clock starts. The distribution of the time before the clock goes off may depend on the state. Once the clock goes off, the individual makes a transition to another state. Such non-exponential waiting times between transitions are usually called *delays* since they are assumed to be non-interruptible by Markovian transitions.

There has been a recent interest for PGSMPs; in 2012, Hayden [12] and Bortolussi and Hillston [4] both proved the mean-field limit for PGSMPs with *deterministic* delays in different ways using different methods.

In Section 4 and [13], we determine and prove the mean-field limit for PGSMPs with generally-

timed delays; the main difference compared to Kurtz's theorem is that the mean-field limit is the solution of a system of delayed differential equations (DDEs), where the evolution of the system depends not just on the current state of the system, but also on its entire past. The change from ODEs to DDEs corresponds to the fact that a "memory" has been introduced to the system by the generally-timed clocks.

1 The relaxed sector condition and the improved graded sector condition

The main results of this section are the so-called relaxed sector condition and an improved version of the graded sector condition. It is joint work with B. Tóth and B. Vető and is based on [15] and parts of [14].

Relaxed sector condition

Let $\mathcal{C} \subset \mathcal{H}$ be a common core for the operators G, G^*, S and A and let

$$B_\lambda := (\lambda I + S)^{-1/2} A (\lambda I + S)^{-1/2}, \qquad \lambda > 0. \tag{10}$$

B_λ are densely defined and skew-self-adjoint on $(\lambda I + S)^{1/2}\mathcal{C}$.

The main point of the relaxed sector condition is that if there exists another essentially skew-self-adjoint operator B, *formally* identified as

$$B := S^{-1/2} A S^{-1/2}, \tag{11}$$

and a sufficiently large subspace on which the sequence of operators B_λ converges pointwise (strongly) to B, as $\lambda \to 0$, then, the H_{-1}-condition (3) implies that the martingale approximation and CLT of Theorem KV follow.

Theorem 2 (**Relaxed sector condition).** *Assume that there exist a subspace $\widetilde{\mathcal{C}} \subseteq \cap_{\lambda>0} \mathrm{Dom}(B_\lambda)$ which is still dense in $\mathcal{H}$ and an operator $B : \widetilde{\mathcal{C}} \to \mathcal{H}$ which is essentially skew-self-adjoint and such that for any vector $\varphi \in \widetilde{\mathcal{C}}$*

$$\lim_{\lambda \to 0} \| B_\lambda \varphi - B\varphi \| = 0. \tag{12}$$

Then the martingale approximation and CLT of Theorem KV follow.

The proof of this theorem shows close similarities with the Trotter-Kurtz theorem. See Theorem 2.12 in [21].

The main point in applications is the proper choice of $\widetilde{\mathcal{C}}$ and proving that B is not just skew-symmetric (which follows from its formal definition) but actually essentially skew-self-adjoint.

Theorem SSC follows directly. Assume now the existence of a grading as in (4)–(6); the following diagonal version of the GSC also follows:

Theorem (GSC from RSC)**.** *If there exists a positive nondecreasing sequence c_n such that*

$$\sum_{n=1}^{\infty} c_n^{-1} = \infty$$

such that

$$\|S_{n+i,n+i}^{-1/2} A_{n,n+i} S_{n,n}^{-1/2}\| \leq c_n \qquad (i = \pm 1),$$

then martingale approximation and central limit theorem applies to every function $f \in \mathrm{Ran}(S^{1/2})$.

The key to the proof is proving

$$\mathrm{Ran}(I \pm S^{-1/2} A S^{-1/2})^{\perp} = 0, \tag{13}$$

which is the counterpart of *the basic criterion of self-adjointness*. See e.g. Theorem VIII.3. of [29].

Improved graded sector condition

We reformulate the graded sector condition from [27] and [17] in a somewhat enhanced version.

Assume that the Hilbert space $\mathcal{H} = \mathcal{L}^2(\Omega, \pi)$ is graded

$$\mathcal{H} = \overline{\oplus_{n=0}^{\infty} \mathcal{H}_n}, \tag{14}$$

and the infinitesimal generator is consistent with this grading in the following sense:

$$S = \sum_{n=0}^{\infty} \sum_{j=-r}^{r} S_{n,n+j}, \qquad S_{n,n+j} : \mathcal{H}_n \to \mathcal{H}_{n+j}, \qquad S_{n,n+j}^* = S_{n+j,n}, \tag{15}$$

$$A = \sum_{n=0}^{\infty} \sum_{j=-r}^{r} A_{n,n+j}, \qquad A_{n,n+j} : \mathcal{H}_n \to \mathcal{H}_{n+j}, \qquad A_{n,n+j}^* = -A_{n+j,n} \tag{16}$$

for some finite positive integer r. Here and in the sequel, the double sum $\sum_{n=0}^{\infty} \sum_{j=-r}^{r} \cdots$ is meant as
$\sum_{n=0}^{\infty} \sum_{j=-r}^{r} \mathbb{1}_{\{n+j \geq 0\}} \cdots$.

Theorem 3 (Improved GSC)**.** *Let the Hilbert space and the infinitesimal generator be graded in the sense specified above. Assume that there exists an operator $D = D^* \geq 0$ which acts diagonally on the grading of $\mathcal{H}$:*

$$D = \sum_{n=0}^{\infty} D_{n,n}, \qquad D_{n,n} : \mathcal{H}_n \to \mathcal{H}_n \tag{17}$$

such that

$$0 \le D \le S. \tag{18}$$

Assume also that, with some $C < \infty$ and $2 \le \kappa < \infty$, the following bounds hold:

$$\left\| D_{n,n}^{-1/2}(S_{n,n} + A_{n,n})D_{n,n}^{-1/2} \right\| \le Cn^{\kappa}, \tag{19}$$

$$\left\| D_{n+j,n+j}^{-1/2} A_{n,n+j} D_{n,n}^{-1/2} \right\| \le \frac{n}{12r^2\kappa} + C, \qquad j = \pm 1, \dots, \pm r, \tag{20}$$

$$\left\| D_{n+j,n+j}^{-1/2} S_{n,n+j} D_{n,n}^{-1/2} \right\| \le \frac{n^2}{6r^3\kappa^2} + C, \qquad j = \pm 1, \dots, \pm r, \tag{21}$$

Under these conditions on the operators, for any function $f \in \oplus_{n=0}^{N} \mathcal{H}_n$, with some $N < \infty$, if

$$D^{-1/2} f \in \mathcal{H}, \tag{22}$$

then the martingale approximation and CLT of Theorem KV hold.

In the original formulation of the graded sector condition (see [30], [17] and [27]), the bound imposed in (21) on the symmetric part of the generator was of the same form as that imposed in (20) on the skew-symmetric part. We can go up to the bound of order n^2 (rather than of order n) in (21) due to decoupling of the estimates of the self-adjoint and skew self-adjoint parts. The proof follows the main lines of the original one with one extra observation which allows for a more precise estimate of the order of the term corresponding to (21).

2 Central limit theorem for the "true" (or myopic) self-avoiding random walk in $d \ge 3$

The main results of this section and [14] is diffusive bounds for the TSAW for a wide class of self-interaction functions in $d \ge 3$, and for a more restricted class of interactions, we prove full CLT for the finite dimensional distributions of the displacement.

Let $t \mapsto X(t) \in \mathbb{Z}^d$ be a continuous time nearest neighbor jump process on the integer lattice $\mathbb{Z}^d$ whose law is defined by (7), with $w : \mathbb{R} \to (0, \infty)$ a fixed smooth "rate function" for which

$$\inf_{u \in \mathbb{R}} w(u) := \gamma > 0, \tag{23}$$

and denote by s and r its even, respectively, odd part:

$$s(u) := \frac{w(u) + w(-u)}{2} - \gamma, \qquad r(u) := \frac{w(u) - w(-u)}{2}. \tag{24}$$

Beside (23), we make the following assumptions: there exist constants $c > 0$, $\varepsilon > 0$ and $C < \infty$ such

that

$$\inf_{u\in\mathbb{R}} r'(u) > c, \tag{25}$$

$$s(u) < C\exp\{(c-\varepsilon)u^2/2\}, \tag{26}$$

and, finally, we make the technical assumption that $r(\cdot)$ is an analytic function which satisfies:

$$\sum_{n=0}^{\infty}\left(\frac{2}{c-\varepsilon}\right)^{n/2}\left|\,r^{(n)}(0)\,\right| < \infty. \tag{27}$$

Condition (23) is *ellipticity* which ensures that the jump rates of the random walk considered are *minorated* by an ordinary simple symmetric walk. Condition (25) ensures sufficient self-repellence of the trajectories and sufficient log-convexity of the stationary measure identified later. Conditions (26) and (27) are of technical nature.

We consider the $d \geq 3$ cases. First, we identify a rather natural stationary and ergodic (in time) distribution of the environment (essentially: the local time profile) as seen from the position of the moving point. In this particular stationary and ergodic regime, we prove diffusive (that is of order t) bounds on the variance of $X(t)$ and *diffusive limit* (that is non-degenerate CLT with normal scaling) for the displacement.

It is natural to consider the local time profile as seen from the position of the random walker

$$\eta(t) = \big(\eta(t,x)\big)_{x\in\mathbb{Z}^d} \qquad \eta(t,x) := \ell(t, X(t)+x). \tag{28}$$

It is obvious that $t \mapsto \eta(t)$ is a c.a.d.l.a.g. Markov process on the state space

$$\Omega := \{\omega = \big(\omega(x)\big)_{x\in\mathbb{Z}^d} \,:\, \omega(x)\in\mathbb{R},\ \ (\forall\varepsilon>0)\lim_{|\,x\,|\to\infty}|\,x\,|^{-\varepsilon}\,|\,\omega(x)\,| = 0\}. \tag{29}$$

Note that we allow initial values $\ell(0,x) \in \mathbb{R}$ for the occupation time measure and thus $\ell(t,x)$ need not be non-negative. The group of spatial shifts

$$\tau_z : \Omega\to\Omega, \qquad \tau_z\omega(x) := \omega(z+x), \qquad z\in\mathbb{Z}^d \tag{30}$$

acts naturally on Ω.

Let

$$\mathcal{U} := \{e\in\mathbb{Z}^d : |\,e\,| = 1\}. \tag{31}$$

We will denote by e the $2d$ unit vectors from $\mathcal{U}$ and by e_l, $l = 1,\ldots,d$, the unit vectors pointing in the positive coordinate directions.

The infinitesimal generator of the process $t \mapsto \eta(t)$, $f : \Omega \to \mathbb{R}$, is

$$Gf(\omega) = \sum_{e \in \mathcal{U}} w(\omega(0) - \omega(e))\big(f(\tau_e \omega) - f(\omega)\big) + \mathcal{D}f(\omega) \tag{32}$$

where the (unbounded) linear operator

$$\mathcal{D}f(\omega) := \frac{\partial f}{\partial \omega(0)}(\omega) \tag{33}$$

is well-defined for smooth cylinder functions.

The meaning of the various terms on the right-hand side of (32) is clear: the terms in the sum are due to the random shifts of the environment caused by the jumps of the random walker while the last term on the right-hand side is due to the deterministic linear growth of local time at the site actually occupied by the random walker.

Next, we define a probability measure on Ω which will turn out to be stationary and ergodic for the Markov process $t \mapsto \eta(t)$. Let

$$R : \mathbb{R} \to [0, \infty), \qquad R(u) := \int_0^u r(v)\, \mathrm{d}v. \tag{34}$$

R is strictly convex and even. We denote by $\mathrm{d}\pi(\omega)$ *the unique centered Gibbs measure* (Markov field) on Ω defined by the conditional specifications for $\Lambda \subset \mathbb{Z}^d$ finite:

$$\mathrm{d}\pi\left(\omega_\Lambda \,\middle|\, \omega_{\mathbb{Z}^d \setminus \Lambda}\right) = Z_\Lambda^{-1} \exp\left\{ -\frac{1}{2} \sum_{\substack{x,y \in \Lambda \\ |\, x-y \,| = 1}} R(\omega(x) - \omega(y)) - \sum_{\substack{x \in \Lambda, y \in \Lambda^c \\ |\, x-y \,| = 1}} R(\omega(x) - \omega(y)) \right\} \mathrm{d}\omega_\Lambda \tag{35}$$

where ω_Λ is the Lebesgue measure on Λ. Note that the (translation invariant) Gibbs measure given by the specifications (35) exists only in three and more dimensions. For information about gradient measures of this type, see [10]. The measure π is invariant under the spatial shifts and the dynamical system $(\Omega, \pi, \tau_z : z \in \mathbb{Z}^d)$ is *ergodic*.

In the particular case when $r(u) = u$, $R(u) = u^2/2$, the measure $\mathrm{d}\pi(\omega)$ is the distribution of the massless free Gaussian field on $\mathbb{Z}^d$, $d \geq 3$ with expectations and covariances

$$\int_\Omega \omega(x)\, \mathrm{d}\pi(\omega) = 0, \qquad \int_\Omega \omega(x)\omega(y)\, \mathrm{d}\pi(\omega) = (-\Delta)^{-1}_{x,y} =: C(y - x) \tag{36}$$

where Δ is the lattice Laplacian: $\Delta_{x,y} = \mathbb{1}_{\{|\, x-y \,| = 1\}} - 2d\mathbb{1}_{\{|\, x-y \,| = 0\}}$. We will refer to this special setup as *the Gaussian case*.

Proposition 1. *The probability measure $\pi(\omega)$ is stationary and ergodic for the Markov process $t \mapsto \eta(t) \in \Omega$.*

The proof of stationarity relies on the so-called *Yaglom-reversibility* property: with the notation

$$Jf(\omega) := f(-\omega), \tag{37}$$

we get

$$JSJ = S, \quad JAJ = -A, \qquad JGJ = G^*. \tag{38}$$

Actually, (38) means slightly more than stationarity: the time-reversed and flipped process

$$t \mapsto \widetilde{\eta}(t) := -\eta(-t) \tag{39}$$

is equal in law to the process $t \mapsto \eta(t)$. This time reversal symmetry is called *Yaglom reversibility* and it appears in many models with physical symmetries. See e.g. [40], [41].

Ergodicity comes from ergodicity of the shifts on (Ω, π).

The law of large numbers for the displacement of the random walker follows:

Corollary 1. *For π-almost all initial profiles $\ell(0, \cdot)$, almost surely*

$$\lim_{t\to\infty} \frac{X(t)}{t} = 0. \tag{40}$$

The main results of this section refer to the diffusive scaling limit of the displacement.

Theorem 4. *(1) If conditions* (23), (25), (26) *and* (27) *hold for the rate function, then*

$$0 < \gamma \leq \inf_{|\,e\,|=1} \varliminf_{t\to\infty} t^{-1}\mathbf{E}\left((e\cdot X(t))^2\right) \leq \sup_{|\,e\,|=1} \varlimsup_{t\to\infty} t^{-1}\mathbf{E}\left((e\cdot X(t))^2\right) < \infty. \tag{41}$$

(2) Assume that

$$r(u) = u, \qquad s(u) = s_4u^4 + s_2u^2 + s_0, \tag{42}$$

and we also make the technical assumption that s_4/γ is sufficiently small. Then the matrix of asymptotic covariances

$$\sigma_{kl}^2 := \lim_{t\to\infty} t^{-1}\mathbf{E}\left(X_k(t)X_l(t)\right) \tag{43}$$

exists and it is non-degenerate. The finite dimensional distributions of the rescaled displacement process

$$X_N(t) := N^{-1/2}X(Nt) \tag{44}$$

converge to those of a d dimensional Brownian motion with covariance matrix σ^2.

The proof uses the martingale + integral decomposition of the displacement $X(t)$:

$$X(t) = N(t) + \int_0^t \overline{\varphi}(\eta(s))\,\mathrm{d}s + \int_0^t \widetilde{\varphi}(\eta(s))\,\mathrm{d}s. \tag{45}$$

Here, $N(t)$ is the martingale part due to the jump rates and $\varphi : \Omega - > \mathbb{R}^d$ is the conditional speed of the walker depending on the environment.

The diffusive lower bound is due to a part of $N(t)$ being decorrelated with other terms.

The diffusive upper bound utilizes calculations in the Fourier transform domain and the following consequence of the Brascamp-Lieb inequality (see e.g. Proposition 2.1 in [3]):

Lemma 5. *For any smooth cylinder function $F : \Omega \to \mathbb{R}$ and $0 \le \lambda < c/2$:*

$$\begin{aligned} &\frac{1}{Z(\lambda)}\mathbf{E}\left(F(\omega)^2 \exp\{\lambda(\omega(0)-\omega(e))^2\}\right) \\ &\qquad \le \frac{1}{c}\frac{1}{Z(\lambda)}\mathbf{E}\left(\sum_{x,y\in\mathbb{Z}^d} \partial_x F(\omega)(-\Delta)^{-1}_{xy}\partial_y F(\omega)\exp\{\lambda(\omega(0)-\omega(e))^2\}\right) \\ &\qquad\qquad + \frac{1}{Z(\lambda)^2}\mathbf{E}\left(F(\omega)\exp\{\lambda(\omega(0)-\omega(e))^2\}\right)^2. \end{aligned} \tag{46}$$

∂_x denotes $\frac{\partial}{\partial\omega(x)}$.

The lemma guarantees finiteness of the asymptotic variance for any polynomial choice of w by induction, and also for series with a fast enough decay of coefficients.

Finally, in the special case

$$r(u) = u, \qquad s(u) = s_4u^4 + s_2u^2 + s_0,$$

the improved version of the GSC is applied. We set ourselves in the Fock space where $\mathcal{H}_n$ is equivalent to polynomials of degree n (orthogonalized); to prove the upper bounds in Theorem 3, calculations are carried out in this space, with one crucial part being

$$\left\| D^{-1/2}_{n+j,n+j} S D^{-1/2}_{n,n} \right\| = O(n^{\deg s/2}),$$

which allows us to go up to degree 4 in the even part of the jump rate function.

3 Constructive proof of the phase-type characterization theorem

The main result of this section is a new, constructive proof of the sufficient direction of Theorem 1. The publication containing the results of this section is [13].

The structure of the representation obtained is relatively simple and transparent, and also comes with an explicit formula for the order of the representation. The algorithm consists of five main steps.

Steps 1 and 2 are preparatory, and Step 5 is just correction related to Step 2. Step 3 has been proposed in [8]; the main novelty here is Step 4, which is slightly more detailed in the following sketch of the algorithm.

- Step 1. We find an equivalent minimal ME representation $(\alpha_1, \mathbf{A}_1)$ for $(\alpha, \mathbf{A})$ if it is not minimal by eliminating any "extra" eigenvalues of $\mathbf{A}$, which does not contribute to the pdf (the pdf will be denoted by f_X).

- Step 2. This step applies only if density is zero at 0, that is, $f_X(0) = 0$. This step is essentially what may be called "deconvolution": we represent f_X as the convolution of some f_Y matrix exponential density function with $f_Y(0) > 0$ and an appropriate Gamma (Erlang) distribution Erlang(k, μ); if f_Y has a Markovian representation, then it gives a straightforward Markovian representation for f_X as well. Thus we only need to find a Markovian representation for f_Y (and the corresponding representation), where $f_Y(0) > 0$. If this step is applied, Steps 3 and 4 are applied for f_Y instead of f_X, and we switch back to f_X in Step 5.

- Step 3. An equivalent ME representation $(\gamma, \mathbf{G})$ is given with Markovian matrix $\mathbf{G}$, while γ may still have negative elements. The main tool of this step is the so-called monocyclic structure (with Feedback-Erlang blocks) [8]. Typically, the size of $\mathbf{G}$ is larger than that of $\mathbf{A}_1$ (because each pair of complex conjugate eigenvalues is represented with at least 3 phases); that said, $\mathbf{G}$ is a sparse matrix with a simple block bi-diagonal structure. For this step only the dominant eigenvalue condition is necessary.

- Step 4. γ and $\mathbf{G}$ are transformed further into β and $\mathbf{B}$ where β is Markovian (and the Markovity of $\mathbf{B}$ is also preserved) essentially by adding an "Erlang-tail" (a number of sequentially connected exponential phases with identical rates) of proper order and rate to the monocyclic structure described by the Markovian matrix $\mathbf{G}$. The main mathematical tool of this step is the approximation of elementary functions. The skeleton of this step is composed of the following elements:

  - Find τ such that $\gamma e^{\tau \mathbf{G}} > 0$ (element-wise). Such τ always exists if the dominant eigenvalue and the positive density conditions hold and $\mathbf{G}$ has Feedback-Erlang structure.
  - Find λ' such that
    $$\gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^{\tau\lambda} > 0 \quad \forall \lambda \geq \lambda'$$
    which is always possible since $\left\| \gamma(\mathbf{I} + \frac{\mathbf{G}}{\lambda})^{\tau\lambda} - \gamma e^{\tau \mathbf{G}} \right\| \to 0$ as $\lambda \to \infty$.
  - Let $\epsilon = \inf_{t \in (0,\tau)} f_X(t)$. $\epsilon > 0$ because of the positive density condition and the result of Step 2. Find λ'' such that
    $$\left| -\gamma e^{\mathbf{G}\tau} \mathbf{G}\mathbf{1} + \gamma \left(\mathbf{I} + \frac{\mathbf{G}}{\lambda} \right)^{\tau\lambda} \mathbf{G}\mathbf{1} \right| < \epsilon \quad \forall \lambda \geq \lambda''.$$

This ensures that $-\gamma\left(\mathbf{I}+\frac{\mathbf{G}}{\lambda}\right)^k \mathbf{G}\mathbf{1} > 0$ for $k = 1, \ldots, n$ where $n = \tau\lambda''$. This is always possible when $\epsilon > 0$.

– Extend the $(\gamma, \mathbf{G})$ representation with an Erlang tail of rate $\lambda = \max(\lambda', \lambda'')$ and order $n = \lceil \lambda\tau \rceil$.

- Step 5. If Step 2 was applied, at this point we have a Markovian representation for f_Y, and we switch back to f_X. If Step 2 was not applied, Step 5 does not apply either.

Theorem 6. *If the positive density condition and the dominant eigenvalue condition hold, then the above algorithm always successfully finds a PH representation for f_X.*

The proof for the necessary direction is essentially a generalization of the Perron–Frobenius theorem as in e.g. [23].

4 Mean-field limit for a class of population models with generally-timed transitions

The main result of this section is the mean-field limit for a class of population generalized semi-Markov processes (PGSMPs). It is joint work with M. Telek and R. Hayden and is based on [11].

We extend the homogeneous Markov population model. N is the population size, each individual is inhabiting a state from a finite set $\mathcal{S}$ and each individual in state i performs Markov transitions from transitions to other states j with rate r_{ij}^N. The global state of the system is the total number of individuals in each state, that is, a vector $\mathbf{x}^N \in \{0, 1, \ldots, N\}^{(S)}$ with $x_1^N + \cdots + x_{|\mathcal{S}|}^N = N$. The *normalized global state* is $\mathbf{x} = \frac{\mathbf{x}^N}{N}$.

We also assume that the density-dependent aggregate rates $r_{ij}(\mathbf{x}) = x_i r_{ij}^N(N\mathbf{x})$ are Lipschitz-continuous with common Lipschitz-constant R. R will be assumed to also be an upper bound on r_{ij}.

We introduce generally-timed transitions. Partition the states into $\mathcal{S} = \mathcal{S}_0 \cup \mathcal{S}_1$, where $\mathcal{S}_0$ contains the states where no generally-timed transitions are allowed, while $\mathcal{S}_1$ contains the states where generally-timed transitions are allowed (these will be called *active states*). For each $i \in \mathcal{S}_1$, a distribution function F_i is given; our only assumption on F_i is that it is concentrated on $[0, \infty)$.

Whenever an individual enters a state $i \in \mathcal{S}_1$, it generates a random time according to F_i independent of everything (we say that the clock is initialized). After that time has elapsed, the individual makes a transition to some other state (these types of transitions will be called non-Markovian transitions). Upon leaving state i, the clock is disabled and will be resampled according to F_i if the individual returns to i later.

To summarize: a single individual may have at most one active clock at any given time; however, there is no restriction on the total number of simultaneously active clocks in the entire system.

We have a number of assumptions. We assume that the system is delay-only, that is, if i is an active state then $r_{ij} = 0 \, \forall j$. We also assume that the non-Markovian transition from an active state

i always targets the same state; we will formulate this by saying that the distribution $\mathbf{p}^i$ of the target state is deterministic: p^i_j is equal to 1 for a unique $j \in \mathcal{S}$ and 0 otherwise. We also assume that the target state is non-active, that is, $j \in \mathcal{S}_0$. This is a technical restriction ensuring that non-Markovian transitions do not follow each other directly. However, it is not a modelling restriction as the state space may be reconfigured so the generally-timed transition is followed by a (very fast) Markovian transition sampling from an arbitrary distribution. We set $p^i_j = 0$ for all other pairs (i, j).

We also assume that the initial state of the system is concentrated on $\mathcal{S}_1$; in other words, no generally-timed clocks are active initially.

We are looking to construct the above model via Poisson-representation. Let $P_{ij}(.)$ be independent Poisson-processes with rate 1 for every $i \neq j \in \mathcal{S}$. Let $\{T^{ij}_k\}_{k=1}^{\infty}$ be mutually independent sequences of identically-distributed random variables distributed according to F_i for each $i \in \mathcal{S}_1$ and $j \in \mathcal{S}_0$. (Taking a separate sample for each j will make formulating the Poisson-representation easier.) For easier formulation, F_i will be included in the notation for $i \in \mathcal{S}_0$ as well, but with 0 coefficients, so F_i is arbitrary for $i \in \mathcal{S}_0$.

The Poisson-representation of $\bar{\mathbf{x}}^N(t)$ is

$$\begin{aligned}\bar{x}^N_i(t) =& \bar{x}^N_i(0) - \sum_{j:j\neq i} \frac{1}{N} P_{ij}\left(N\int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) + \sum_{j:j\neq i} \frac{1}{N} P_{ji}\left(N\int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\\ &+ \sum_{h\in\mathcal{S}_0}\sum_{j\in\mathcal{S}_1}\int_{z=0}^t p^i_j \mathbf{1}\left(T^{ji}_{P_{hj}\left(N\int_0^z r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)} \leq t-z\right)\frac{1}{N}\mathrm{d}P_{hj}\left(N\int_0^z r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)\\ &- \sum_{h\in\mathcal{S}_0}\sum_{j:j\in\mathcal{S}_0}\int_{z=0}^t p^j_i \mathbf{1}\left(T^{ij}_{P_{hi}\left(N\int_0^z r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)} \leq t-z\right)\frac{1}{N}\mathrm{d}P_{hi}\left(N\int_0^z r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u\right) \quad (47)\end{aligned}$$

for $i \in \mathcal{S}$.

The first term is the initial condition; the second and third terms correspond to Markovian jumps when an individual transitions from and to state i, respectively.

The first of the last two terms in the formula (47) should be understood as follows. If i is active, the term is 0 by our assumptions on p^i_j. If i is inactive, consider an active state j with $p^i_j = 1$ and an inactive state h. If a Markov transition from h to j occurs at time z, a non-Markovian clock distributed according to F_j starts. The clock samples from the list $\{T^{ji}_k\}_{k=1}^{\infty}$; to ensure that a new k is used for each clock, k is set to $P_{hj}\left(\int_0^z r_{hj}((\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)$ (which increases with each arrival of P_{hj}). When the indicator variable is 1, the clock has already set off before time t and needs to be counted among the actual transitions; when the indicator variable is 0, the clock has not yet set off by time t, so the corresponding $j \to i$ transition has not yet occurred, and the contribution of the integral is 0.

Conversely, the last term of (47) is nonzero only if i is active; consider an inactive state j with $p^j_i = 1$ and an inactive state h. If a Markov transition from h to i occurs at time z, a non-Markovian clock distributed according to F_i starts. The clock samples from the list $\{T^{ij}_k\}_{k=1}^{\infty}$ with $k = P_{ij}\left(\int_0^z r_{ij}((\bar{\mathbf{x}}^N(u))\mathrm{d}u\right)$ (which increases with each arrival of P_{ij}). When the indicator variable is 1, the clock has already set off before time t and needs to be counted among the actual transitions (which decrease the number of individuals in state i, hence the negative sign); when the indicator

variable is 0, the clock has not yet set off by time t, so the corresponding $i \to j$ transition has not yet occurred.

The mean-field limit is defined by the following delayed differential equations (written in integral form):

$$\begin{aligned} v_i(t) =& v_i(0) - \sum_{j:j\neq i} \int_0^t r_{ij}(\mathbf{v}(u))\mathrm{d}u + \sum_{j:j\neq i} \int_0^t r_{ji}(\mathbf{v}(u))\mathrm{d}u \\ &+ \sum_{h\in\mathcal{S}_0} \sum_{j\in\mathcal{S}_1} \int_{u=0}^t p_j^i F_j(t-u) r_{hj}(\mathbf{v}(u))\mathrm{d}u \\ &- \sum_{h\in\mathcal{S}_0} \sum_{j:j\in\mathcal{S}_0} \int_{u=0}^t p_i^j F_i(t-u) r_{hi}(\mathbf{v}(u))\mathrm{d}u \end{aligned} \tag{48}$$

for $i \in \mathcal{S}$.

Lipschitz continuity of r_{ij} guarantees that the solution of (48) uniquely exists.

Again, we assume convergence of the initial condition:

$$\lim_{N\to\infty} \mathbf{P}(\|\mathbf{v}(0) - \bar{\mathbf{x}}^N(0)\| > \epsilon) = 0 \quad \forall \varepsilon > 0.$$

Theorem 7. *Under the assumptions and setup given above, we have, for any $T > 0$ and $\epsilon > 0$:*

$$\lim_{N\to\infty} \mathbf{P}\left\{ \sup_{t\in[0,T]} \|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| > \epsilon \right\} = 0$$

A few words about the proof. Define the auxiliary process $\mathbf{y}^N(t)$ via

$$\begin{aligned} y_i^N(t) :=& v_i(0) - \sum_{j:j\neq i} \int_0^t r_{ij}(\bar{\mathbf{x}}^N(u))\mathrm{d}u + \sum_{j:j\neq i} \int_0^t r_{ji}(\bar{\mathbf{x}}^N(u))\mathrm{d}u \\ &+ \sum_{h\in\mathcal{S}_0} \sum_{j\in\mathcal{S}_1} \int_{u=0}^t p_j^i F_j(t-u) r_{hj}(\bar{\mathbf{x}}^N(u))\mathrm{d}u \\ &- \sum_{h\in\mathcal{S}_0} \sum_{j:j\in\mathcal{S}_0} \int_{u=0}^t p_i^j F_i(t-u) r_{hi}(\bar{\mathbf{x}}^N(u))\mathrm{d}u \end{aligned} \tag{49}$$

for $i \in \mathcal{S}$.

Then

$$|\bar{x}_i^N(t) - v_i(t)| \le |\bar{x}_i^N(t) - y_i^N(t)| + |y_i^N(t) - v_i(t)|$$

for any $i \in \mathcal{S}$.

Denote

$$D_i^N(T) = \sup_{t\in[0,T]} |\bar{x}_i^N(t) - y_i^N(t)|.$$

We estimate $\|\mathbf{y}^N(t) - \mathbf{v}(t)\|$ by

$$|y_i^N(t) - v_i(t)| \leq C \int_0^t \|\mathbf{x}^N(u) - \mathbf{v}(u)\| \mathrm{d}u$$

for some finite C ($\|.\|$ is the maximum norm on $\mathbb{R}^{\mathcal{S}}$). We aim to show that $D_i^N(T) \to 0$ in probability as $N \to \infty$ for each $i \in \mathcal{S}$; once we have that, we have

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \leq \max_{i \in \mathcal{S}} D_i^N(T) + ZR \int_0^t \|\bar{\mathbf{x}}^N(u) - \mathbf{v}(u)\| \mathrm{d}u \tag{50}$$

and an application of Grönwall's lemma ([9], page 498) readily yields

$$\|\bar{\mathbf{x}}^N(t) - \mathbf{v}(t)\| \leq \max_{i \in \mathcal{S}} D_i^N(T) \exp(ZRT),$$

proving the theorem.

$D_i^N(T) \to 0$ in probability is proved by applying a number of probability concentration results, notably the functional strong law of large numbers for the Poisson process ([39], Section 3.2) and Azuma's inequality [7, 2] after a proper setup.

References

[1] D. Amit, G. Parisi, and L. Peliti. Asymptotic behavior of the 'true' self-avoiding walk. *Phys. Rev. B*, 27:1635–1645, 1983.

[2] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

[3] S. G. Bobkov and M. Ledoux. From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. and Funct. Anal.*, 10:1028–1052, 2000.

[4] Luca Bortolussi and Jane Hillston. Fluid approximation of ctmc with deterministic delays. *Int. Conf. on Quantitative Evaluation of Systems*, pages 53–62, 2012.

[5] L. Breuer and D. Baum. An Introduction to Queueing Theory and Matrix-Analytic Methods Springer, 2005

[6] Peter Buchholz and Miklós Telek. On minimal representation of rational arrival processes. *Annals of Operations Research*, 202(1):35–58, 2013.

[7] Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: A survey. *Internet Mathematics*, 3(1):79–127, 2006.

[8] C. Commault and S. Mocanu. Phase-type distributions and representations: some open problems for system theory. *Int. J. Control*, 76(6):566–580, 2003.

[9] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 2005.

[10] T. Funaki. *Stochastic Interface Models. Lectures on Probability Theory and Statistics, Lecture Notes in Mathematics*, volume 1869, 103–274, 2005.

[11] R. Hayden, I. Horváth, and M. Telek. Mean field for performance models with generally distributed-timed transitions. In W. Sanders G. Norman, editor, *11th International Conference on Quantitative Evaluation of Systems, QEST*, volume 8657 of *Lecture Notes in Computer Science*, pages 90–105. Springer, 2014.

[12] R. A. Hayden. Mean field for performance models with deterministically-timed transitions. In *9th International Conference on Quantitative Evaluation of Systems, QEST*, pages 63–73, Sept 2012.

[13] I. Horváth and M. Telek. A constructive proof of the phase-type characterization theorem. *Stochastic Models*, to appear, 2015.

[14] I. Horváth, B. Tóth, and B. Vető. Diffusive limits for "true" (or myopic) self-avoiding random walks and self-repellent Brownian polymers in dimensions 3 and higher. *Probability Theory and Related Fields*, 153(3-4), 2012.

[15] I. Horváth, B. Tóth, and B. Vető. Relaxed sector condition. *Bulletin of the Institute of Mathematics, Academia Sinica (N.S.)*, 7:463–476, 2012.

[16] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes with applications to simple exclusion. *Communications in Mathematical Physics*, 106:1–19, 1986.

[17] T. Komorowski, C. Landim, and S. Olla. *Fluctuations in Markov Processes – Time Symmetry and Martingale Approximation*, volume 345 of *Grundlehren der mathematischen Wissenschaften*, Springer, 2012.

[18] T. Komorowski and S. Olla. On the sector condition and homogenization of diffusions with a gaussian drift. *Journal of Functional Analysis*, 197:179–211, 2003.

[19] G. Kozma and B. Tóth. Central limit theorem for random walks in divergence-free random drift field: H_{-1} suffices. preprint, 2014, `http://arxiv.org/abs/1411.4171`

[20] Thomas G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, 1978.

[21] T. L. Liggett. *Interacting Particle Systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften*.

[22] R. S. Maier. The algebraic construction of phase-type distributions. *Commun. Stat., Stochastic Models*, 7(4):573–602, 1991.

[23] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2004.

[24] S. Mocanu and C. Commault. Sparse representations of phase-type distributions. *Commun. Stat., Stochastic Models*, 15(4):759–778, 1999.

[25] S. P. Obukhov and L. Peliti. Renormalisation of the "true" self-avoiding walk. *J. Phys. A*, 16:L147–L151, 1983.

[26] Colm Art O'Cinneide. Characterization of phase-type distributions. *Communications in Statistics. Stochastic Models*, 6(1):1–57, 1990.

[27] S. Olla. Central limit theorems for tagged particles and for diffusions in random environment. In É. Pardoux F. Comets, editor, *Milieux aléatoires*, volume 12 of *Panoramas et Synthses*. Societé Mathématique de France, Paris, 2002.

[28] L. Peliti and L. Pietronero. Random walks with memory. *Riv. Nuovo Cimento*, 10:1–33, 1987.

[29] M. Reed and B. Simon. *Methods of Modern Mathematical Physics Vol 1, 2.*, volume 1, 2. Academic Press New York, 1975.

[30] S. Sethuraman, S. R. S. Varadhan, and H-T. Yau. Diffusive limit of a tagged particle in asymmetric simple exclusion processes. *Communications in Pure and Applied Mathematics*, 53:972–1006, 2000.

[31] P. Tarrès, B. Tóth and B. Valkó. Diffusivity bounds for 1d Brownian polymers. Ann. Probab., 40(2):695–713, 2012.

[32] B. Tóth. Persistent random walk in random environment. *Probability Theory and Related Fields*, 71:615–625, 1986.

[33] B. Tóth. The 'true' self-avoiding walk with bond repulsion on $\mathbb{Z}$: limit theorems. *Ann. Probab.*, 23:1523–1556, 1995.

[34] B. Tóth and B. Valkó. Superdiffusive bounds on self-repellent Brownian polymers and diffusion in the curl of the gaussian free field in d=2. *Journal of Stat. Phys.*, 147(1):113–131, 2012.

[35] B. Tóth and B. Vető. Continuous time 'true' self-avoiding random walk on $\mathbb{Z}$. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 8:59–75, 2011.

[36] B. Tóth and W. Werner. The true self-repelling motion. *Probab. Theory Rel. Fields*, 111:375–452, 1998.

[37] S. R. S. Varadhan. Self-diffusion of a tagged particle in equilibrium of asymmetric mean zero random walks with simple exclusion. *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques*, 31:273–285, 1996.

[38] B. Vető. Asymptotic behaviour of random walks with long memory. PhD Thesis, 2011.

[39] Ward Whitt. Internet supplement to Stochastic-Process Limits, 2002.

[40] A. M. Yaglom. On the statistical treatment of Brownian motion (russian). *Dokl. Akad. Nauk SSSR*, 56:691–694, 1947.

[41] A. M. Yaglom. On the statistical reversibility of Brownian motion (russian). *Mat. Sbornik*, 24:457–492, 1949.