The Annals of Probability
2020, Vol. 48, No. 2, 742–759
https://doi.org/10.1214/19-AOP1373

CUTOFF FOR THE MEAN-FIELD ZERO-RANGE PROCESS WITH BOUNDED MONOTONE RATES

BY JONATHAN HERMON[1] AND JUSTIN SALEZ[2]

[1]*DPMMS, Cambridge University, jonathan.hermon@gmail.com*
[2]*LPSM, Université Paris Diderot, justin.salez@lpsm.paris*

We consider the zero-range process with arbitrary bounded monotone rates on the complete graph, in the regime where the number of sites diverges while the density of particles per site converges. We determine the asymptotics of the mixing time from any initial configuration, and establish the cutoff phenomenon. The intuitive picture is that the system separates into a slowly evolving solid phase and a quickly relaxing liquid phase: as time passes, the solid phase dissolves into the liquid phase, and the mixing time is essentially the time at which the system becomes completely liquid. Our proof uses the path coupling technique of Bubley and Dyer, and the analysis of a suitable hydrodynamic limit. To the best of our knowledge, even the order of magnitude of the mixing time was unknown, except in the special case of constant rates.

1. Introduction.

1.1. *Model.* Introduced by Spitzer in 1970 [27], the *zero-range process* is a widely studied model of interacting random walks; see, for example, [12, 21, 22] and the references therein. It describes the evolution of $m \geq 1$ indistinguishable particles randomly hopping across $n \geq 1$ sites. The interaction is specified by a function $r\colon \{1, 2, \ldots\} \to (0, \infty)$, where $r(k)$ indicates the rate at which particles are expelled from a site with k particles on it. We will here focus on the *mean-field* version of the model, where all jump destinations are uniformly distributed. More formally, we consider a continuous-time Markov chain $\mathbf{X} := (X(t)\colon t \geq 0)$ taking values in the state space

$$\Omega := \left\{ x = (x_1, \ldots, x_n) \in \mathbb{Z}_+^n \colon \sum_{i=1}^{n} x_i = m \right\}, \tag{1}$$

and whose Markov generator $\mathcal{L}$ acts on observables $\varphi\colon \Omega \to \mathbb{R}$ as follows:

$$(\mathcal{L}\varphi)(x) = \frac{1}{n} \sum_{1 \leq i, j \leq n} r(x_i)\big(\varphi(x + \delta_j - \delta_i) - \varphi(x)\big). \tag{2}$$

Here, $(\delta_i)_{1 \leq i \leq n}$ denotes the canonical basis of $\mathbb{Z}_+^n$, and we adopt the convention that $r(0) = 0$ (no jumps from empty sites). The generator $\mathcal{L}$ is irreducible and reversible with respect to the following probability measure:

$$\pi(x) \propto \prod_{i=1}^{n} \prod_{k=1}^{x_i} \frac{1}{r(k)}, \tag{3}$$

with the standard convention that an empty product is 1. The present paper is concerned with the problem of estimating the speed at which the convergence to equilibrium occurs, as

Received January 2019.
MSC2010 subject classifications. 37A25, 60J27, 82C22, 60K35.
Key words and phrases. Zero-range process, cutoff phenomenon.

quantified by the so-called *mixing times*:

$$t_{\mathrm{MIX}}(x;\varepsilon) := \min\{t \geq 0 \colon \|\mathbb{P}_x(X(t) \in \cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon\}. \tag{4}$$

In this definition, $\|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|$ denotes the total-variation distance, and the parameters $x \in \Omega$ and $\varepsilon \in (0, 1)$ specify the initial state and the desired precision, respectively. Of particular interest is the *worst-case* mixing time, obtained by maximizing over all initial states:

$$t_{\mathrm{MIX}}(\varepsilon) := \max\{t_{\mathrm{MIX}}(x;\varepsilon) \colon x \in \Omega\}. \tag{5}$$

Understanding this fundamental parameter—and in particular, its dependency in the precision $\varepsilon \in (0, 1)$—is in general a challenging task; see the books [20, 24] for a comprehensive account. Our current knowledge on the total-variation mixing time of the zero-range process is embarrassingly limited in comparison with the numerous functional-analytic estimates that have been established over the past decades [4, 6–8, 13–16, 19, 25]. In fact, to the best of our knowledge, the exact order of magnitude of the mixing time of the zero-range process has only been determined in the very special case where the rate function r is constant [15, 17, 18, 23].

1.2. *Main result*. The rate function r will remain fixed throughout the paper, and will only be assumed to be nondecreasing and bounded. Upon rescaling time by a constant factor if necessary, we may take

$$\lim_{k\to\infty} \uparrow r(k) = 1. \tag{6}$$

Our results will depend on r through the log-derivative of a certain series:

$$\Psi(z) := \frac{zR'(z)}{R(z)} \quad \text{where } R(z) := \sum_{k=0}^{\infty} \frac{z^k}{r(1)\cdots r(k)}. \tag{7}$$

It is easy to see that $\Psi \colon [0, 1) \to [0, \infty)$ is a bijection, and we write Ψ^{-1} for its inverse. All asymptotic statements will refer to the regime where the number of sites diverges while the density of particles per site stabilizes:

$$n \to \infty, \qquad \frac{m}{n} \to \rho \in [0, \infty). \tag{8}$$

To lighten the notation, we will keep the dependency upon n implicit as often as possible. Our main result is as follows.

THEOREM 1 (Worst-case mixing time). *For any fixed* $\varepsilon \in (0, 1)$,

$$\frac{t_{\mathrm{MIX}}(\varepsilon)}{n} \xrightarrow[n\to\infty]{} \gamma := \int_0^{\rho} \frac{\mathrm{d}s}{1 - \Psi^{-1}(s)}. \tag{9}$$

Although it seems intuitively clear that the worst-case mixing time should be achieved by initially placing all particles on the same site, there does not appear to be any direct justification of this fact. We will thus determine the asymptotics of the mixing time from *every* possible configuration $x \in \Omega$; see Corollary 1 below for the detailed result. The notable disappearance of ε on the right-hand side of (9) reveals a sharp transition in the convergence to equilibrium of the process, known as a *cutoff* [1, 10]. To the best of our knowledge, the occurrence of this phenomenon for the mean-field zero-range process was only known in the special *rate-one* case, where the function r is constant equal to 1 [23]. This choice trivially fits our setting (6), with

$$R(z)=\frac{1}{1-z},\qquad \Psi(z)=\frac{z}{1-z},\qquad \Psi^{-1}(s)=\frac{s}{1+s},\qquad \gamma=\rho+\frac{\rho^2}{2}. \tag{10}$$

Beyond the obvious complications raised by the nonexplicit nature of the rates, Theorem 1 requires new ideas for at least two reasons. First, the crucial spectral gap estimate of Morris [25], on which the whole argument of [23] ultimately relies, is only available in the rate-one case. Second, the stationary distribution (3) is no longer uniform, making the entropy computations from [14, 23] unapplicable. As a result, even the order of magnitude $t_{\mathrm{MIX}}(\varepsilon)=\Theta(n)$ appears to be new. We circumvent these obstacles by resorting to the powerful *path coupling* method of Bubley and Dyer [5]. This alternative route turns out to be so efficient that the proof of our generalization ends up being significantly shorter than that of the original result [23], without using anything from it.

1.3. *Proof outline.* Intuitively, the system may be viewed as consisting of two regions evolving on different time-scales:

- a slow *solid phase*, consisting of sites occupied by $\Theta(n)$ particles,
- a quick *liquid phase*, consisting of sites occupied by $o(n)$ particles.

The presence of a solid phase is a clear indication that the system is out of equilibrium (under the stationary law π, the maximum occupancy is easily seen to be $\Theta(\log n)$; see, for example, (31) below). What is less obvious, but true, is that conversely, any completely liquid system reaches equilibrium in negligible time. To prove this, we use the path coupling method of Bubley and Dyer [5]. Note that in the regime (8), there is $\overline{\rho}<\infty$, independent of n, such that

$$\frac{m}{n}\le\overline{\rho}. \tag{11}$$

By a *dimension-free constant*, we will always mean a real number that depends only on $\overline{\rho}$ and r.

THEOREM 2 (Fast mixing). *There is a dimension-free constant κ so that*

$$t_{\mathrm{MIX}}(x;\varepsilon)\le\kappa\|x\|_\infty+(\ln n)^\kappa, \tag{12}$$

for every $x\in\Omega$ and every $\varepsilon\in(0,1)$, provided $n\ge\kappa/\varepsilon$.

When combined with the worst-case bound $\|x\|_\infty\le m$, this already yields the correct order of magnitude $t_{\mathrm{MIX}}(\varepsilon)=\mathcal{O}(n)$, for any fixed $\varepsilon\in(0,1)$. However, the real interest of Theorem 2 lies in the linear dependency in $\|x\|_\infty$, which implies that the equilibrium is attained in negligible time when the initial configuration $x=x(n)$ is completely liquid: for fixed $\varepsilon\in(0,1)$,

$$\|x\|_\infty=o(n)\quad\Longrightarrow\quad t_{\mathrm{MIX}}(x;\varepsilon)=o(n). \tag{13}$$

By the Markov property, this reduces our task to that of understanding the time it takes for an arbitrary initial condition $x\in\Omega$ to become completely liquid. By symmetry, we may assume without loss of generality that the coordinates of x are arranged in decreasing order:

$$x_1\ge\cdots\ge x_n. \tag{14}$$

In the regime (8), we may also assume (upon passing to a subsequence) that

$$\frac{x_k}{n}\xrightarrow[n\to\infty]{}u_k, \tag{15}$$

for each $k \geq 1$. Note that the limiting profile $(u_k)_{k\geq 1}$ must then satisfy $u_1 \geq u_2 \geq \cdots \geq 0$ and $\sum_{k=1}^{\infty} u_k \leq \rho$. In this setting, our second step will consist in establishing a deterministic approximation of the form

$$\frac{X_k(nt)}{n} \approx \left[u_k - f(t)\right]_+, \tag{16}$$

where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a smooth increasing function describing the dissolution of the solid phase, and where we have used the notation $[a]_+ = \max(a, 0)$. This is the content of the following theorem.

THEOREM 3 (Hydrodynamic limit). *For any initial condition $x = x(n)$ satisfying* (14)–(15), *and for any fixed time horizon $T \geq 0$, we have*

$$\mathbb{E}_x\left[\sup_{k\in[n],t\in[0,T]}\left|\frac{X_k(nt)}{n} - \left[u_k - f(t)\right]_+\right|\right] \xrightarrow[n\to\infty]{} 0, \tag{17}$$

where the function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is characterized by the differential equation

$$f'(t) = 1 - \Psi^{-1}\left(\rho - \sum_{k=1}^{\infty}\left[u_k - f(t)\right]_+\right), \qquad f(0) = 0. \tag{18}$$

Moreover, for each $i \geq 1$, we have the explicit expression

$$f^{-1}(u_i) = \sum_{k=i}^{\infty} \frac{1}{k} \int_{\rho_k}^{\rho_{k-1}} \frac{\mathrm{d}s}{1 - \Psi^{-1}(s)}, \tag{19}$$

where $\rho = \rho_0 \geq \rho_1 \geq \cdots \geq 0$ are given by $\rho_k := \rho + ku_{k+1} - \sum_{i=1}^{k} u_i$.

The proof relies on a *separation of timescales* argument: the liquid phase relaxes so quickly that, on the relevant time-scale, the solid phase can be considered as inert. Consequently, the liquid phase is permanently maintained in a metastable state resembling the true equilibrium, except that the density is lower because a macroscopic number of particles are "stuck" in the solid phase. This imposes a simple asymptotic relation between the number of particles in the solid phase and the dissolution rate, from which the autonomous equation (18) arises. Note that this limiting description gives access to the dissolution time of the system: for any $t \geq 0$, we have

$$\|X(nt)\|_\infty = o(n) \quad \Longleftrightarrow \quad t \geq f^{-1}(u_1). \tag{20}$$

Combining this with (13), we readily obtain the following full description of the mixing time, of which (13) is only the special case where $u_1 = 0$.

COROLLARY 1 (Mixing time from an arbitrary initial condition). *For any fixed $\varepsilon \in (0, 1)$, we have in the regime* (14)–(15),

$$\frac{t_{\text{MIX}}(x; \varepsilon)}{n} \xrightarrow[n\to\infty]{} f^{-1}(u_1) = \sum_{k=1}^{\infty} \frac{1}{k} \int_{\rho_k}^{\rho_{k-1}} \frac{\mathrm{d}s}{1 - \Psi^{-1}(s)}. \tag{21}$$

From this detailed description, the worst-case mixing time can finally be extracted by maximizing the right-hand side of (21) over all possible profiles $(u_k)_{k\geq 1}$: we trivially always have

$$\sum_{k=1}^{\infty} \frac{1}{k} \int_{\rho_k}^{\rho_{k-1}} \frac{\mathrm{d}s}{1 - \Psi^{-1}(s)} \leq \int_0^{\rho} \frac{\mathrm{d}s}{1 - \Psi^{-1}(s)}, \tag{22}$$

and the equality is moreover attained when $u_1 = \rho$ and $u_2 = u_3 = \cdots = 0$. The maximizer corresponds to placing all particles on the same site, as anticipated. This clearly establishes Theorem 1, and the remainder of the paper is devoted to the proofs of Theorems 2 and 3, in Sections 2 and 3.

2. Fast mixing in the absence of a solid phase.

2.1. *Preliminaries.* We start with a few standard facts that will be used repeatedly in the sequel. We use the classical notation $[n] := \{1, \dots, n\}$.

Graphical construction. Let Ξ be a Poisson point process of intensity $\frac{1}{n}\,\mathrm{d}t \otimes \mathrm{d}u \otimes \mathrm{Card} \otimes \mathrm{Card}$ on $[0, \infty) \times [0, 1] \times [n] \times [n]$ (where Card denotes the counting measure), and consider the piecewise constant process $\mathbf{X} = (X(t) \colon t \geq 0)$ defined by the initial condition $X(0) = x$ and the following jumps: for each point $(t, u, i, j) \in \Xi$,

$$X(t) := \begin{cases} X(t-) + \delta_j - \delta_i & \text{if } r\big(X_i(t-)\big) \geq u, \\ X(t-) & \text{otherwise.} \end{cases} \tag{23}$$

Then $\mathbf{X}$ is a Markov process with generator $\mathcal{L}$ and initial state x. We always use this particular construction.

Monotony. Since r is nondecreasing, the above construction provides a *monotone coupling* of trajectories: if we start from two configurations $x, y \in \mathbb{Z}_+^n$ satisfying $x \leq y$ (coordinate-wise), then this property is preserved by the jumps (23), so the resulting processes $\mathbf{X}, \mathbf{Y}$ satisfy

$$\forall t \geq 0, \quad X(t) \leq Y(t). \tag{24}$$

This classical fact will play an important role in our proof.

Stochastic regularity. For any $i \in [n]$, $0 \leq s \leq t$, we have by construction

$$X_i(t) - X_i(s) \leq \Xi\big([s, t] \times [0, 1] \times [n] \times \{i\}\big), \tag{25}$$

$$X_i(s) - X_i(t) \leq \Xi\big([s, t] \times [0, 1] \times \{i\} \times [n]\big) \tag{26}$$

and the random variables on the right-hand sides are Poisson $(t - s)$.

Temperature. For any $\varphi \colon \Omega \to \mathbb{R}$, the process $\mathbf{M} = (M(t) \colon t \geq 0)$ given by

$$M(t) := \varphi\big(X(t)\big) - \varphi(x) - \int_0^t (\mathcal{L}\varphi)\big(X(s)\big)\,\mathrm{d}s \tag{27}$$

is a zero-mean martingale (see, e.g., [11]). In particular, taking $\varphi(y) = y_i$ $(i \in [n])$, we obtain

$$M(t) = X_i(t) - x_i + \int_0^t r\big(X_i(s)\big)\,\mathrm{d}s - \int_0^t \zeta(s)\,\mathrm{d}s, \tag{28}$$

where the *temperature* $\zeta(t)$ measures the average jump rate of the system:

$$\zeta(t) := \frac{1}{n} \sum_{j=1}^{n} r\big(X_j(t)\big). \tag{29}$$

Understanding the evolution of $(\zeta(t) \colon t \geq 0)$ will constitute an important step in the proof.

2.2. *Uniform downward drift*. Our first task consists in showing that the number of particles on a site cannot stay large for long.

PROPOSITION 1 (Uniform downward drift). *There are dimension-free constants $\theta, \delta > 0$ such that*

$$\mathbb{E}_x\big[e^{\theta X_i(t)}\big] \le 2\big(1 + e^{\theta(x_i - \delta t)}\big), \tag{30}$$

for all $x \in \Omega$, $i \in [n]$ and $t \in \mathbb{R}_+$. In particular, for any $a \ge 0$,

$$\mathbb{P}_x\big(X_i(t) \ge [x_i - \delta t]_+ + a\big) \le 4e^{-\theta a}. \tag{31}$$

The intuition behind this result is as follows: if $X_i(t)$ is large, then by (6) and (28), the drift of $X_i(t)$ is essentially $\zeta(t) - 1$, which is uniformly negative thanks to the following lemma.

LEMMA 1 (Temperature). *There is a dimension-free $\varepsilon > 0$ so that*

$$\mathbb{P}_x\big(\zeta(t) \ge 1 - \varepsilon\big) \le e^{-\varepsilon n},$$

for all $x \in \Omega$ and all $t \in [1, \infty)$, provided $n \ge 2$.

PROOF. Since r is $[0, 1]$-valued with $r(0) = 0$, we clearly have

$$\zeta(t) \le 1 - \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{(X_j(t)=0)}. \tag{32}$$

Now, because of (11), we can find a region $\mathcal{I} \subseteq [n]$ of size $|\mathcal{I}| = \lceil n/2 \rceil$ such that $\max_{i\in\mathcal{I}} x_i \le 2\overline{\rho}$. Note that $|\mathcal{I}^c| = \lfloor n/2 \rfloor \ge n/3$, since $n \ge 2$. For each site $i \in \mathcal{I}$, let G_i denotes the following event:

$$\big\{\Xi\big([0, 1] \times [0, 1] \times [n] \times \{i\}\big) = 0\big\} \cap \big\{\Xi\big([0, 1] \times \big[0, r(1)\big] \times \{i\} \times \mathcal{I}^c\big) \ge 2\overline{\rho}\big\}.$$

The first part forbids any new arrival at i during the time-interval $[0, 1]$, while the second ensures that the $x_i \le 2\overline{\rho}$ particles will depart (recall that $r(k) \ge r(1)$ for $k \ge 1$). Thus, $G_i \subseteq \{X_i(1) = 0\}$. Writing $\mathcal{P}(\lambda; k)$ for the probability that a Poisson variable with mean λ is at least k, we have

$$\mathbb{P}(G_i) = e^{-1}\mathcal{P}\bigg(\frac{r(1)\lfloor n/2 \rfloor}{n}; \lfloor 2\overline{\rho} \rfloor\bigg) \ge e^{-1}\mathcal{P}\bigg(\frac{r(1)}{3}; \lfloor 2\overline{\rho} \rfloor\bigg) =: q. \tag{33}$$

Since the events $(G_i)_{i\in\mathcal{I}}$ are moreover independent, we conclude that the sum $\sum_{i=1}^{n} \mathbf{1}_{(X_i(1)=0)}$ stochastically dominates a Binomial variable with parameters $\lceil n/2 \rceil$ and q. Thus, Hoeffding's inequality (see, e.g., [3]) implies

$$\mathbb{P}\Bigg(\sum_{i=1}^{n} \mathbf{1}_{(X_i(1)=0)} \le \frac{nq}{4}\Bigg) \le \exp\bigg(-\frac{nq^2}{4}\bigg),$$

so that $\varepsilon = q^2/4$ satisfies the claim for $t = 1$. Since the result is uniform in the choice of the initial state $x \in \Omega$, the claim automatically propagates to any time $t \ge 1$ by the Markov property. □

PROOF OF PROPOSITION 1. For any $\varphi : \Omega \to \mathbb{R}$, (27) implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_x\big[\varphi\big(X(t)\big)\big] = \mathbb{E}_x\big[(\mathcal{L}\varphi)\big(X(t)\big)\big]. \tag{34}$$

Taking $\varphi(y) = e^{\theta y_i}$, our plan will consist in bounding $\mathcal{L}\varphi$ in terms of φ to obtain a differential inequality. For any $y \in \Omega$, we easily compute

$$\frac{(\mathcal{L}\varphi)(y)}{\varphi(y)} = (e^\theta - 1)\left\{\frac{1}{n} \sum_{j \in [n]\setminus\{i\}} \big(r(y_j) - e^{-\theta} r(y_i)\big)\right\}.$$

The term $\{\cdot\}$ is always less than 1, and is even less than $\lambda := 1 - \varepsilon - e^{-\theta} r(k)$ if $y \in A \cap B$, where

$$A := \left\{\frac{1}{n}\sum_{j=1}^n r(y_j) < 1 - \varepsilon\right\}, \qquad B := \{y_i \geq k\}. \tag{35}$$

The parameters $\varepsilon \in (0, 1)$ and $k \in \mathbb{N}$ are arbitrary for now and will be adjusted later. Thus,

$$\begin{aligned}
\mathcal{L}\varphi &\leq (e^\theta - 1)\big(\lambda\varphi + (1-\lambda)\varphi \mathbf{1}_{(A\cap B)^c}\big) \\
&\leq (e^\theta - 1)\big(\lambda\varphi + 2\varphi(\mathbf{1}_{A^c} + \mathbf{1}_{B^c})\big) \\
&\leq (e^\theta - 1)\big(\lambda\varphi + 2e^{\theta m}\mathbf{1}_{A^c} + 2e^{\theta k}\big),
\end{aligned}$$

where we have used $\lambda \in (-1, 1)$, $\|\varphi\|_\infty = e^{\theta m}$ and $\varphi \mathbf{1}_{B^c} \leq e^{\theta k}$. By (34),

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_x\big[e^{\theta X_i(t)}\big] \leq (e^\theta - 1)\big\{\lambda \mathbb{E}_x\big[e^{\theta X_i(t)}\big] + 2e^{\theta m}\mathbb{P}_x\big(\zeta(t) \geq 1 - \varepsilon\big) + 2e^{\theta k}\big\}.$$

We now choose the dimension-free constants θ, ε, k as follows. We take ε as in Lemma 1. Since $\lambda \to -\varepsilon$ as $(k, \theta) \to (\infty, 0)$, we may then choose $k \in \mathbb{N}$ and $\theta > 0$ so that $\lambda < 0$. Upon further reducing θ if necessary, we may assume that $\theta \leq \varepsilon/\overline{\rho}$, so that $\theta m \leq \varepsilon n$. For $t \geq 1$, we then have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_x\big[e^{\theta X_i(t)}\big] \leq \alpha - \delta \mathbb{E}_x\big[e^{\theta X_i(t)}\big], \tag{36}$$

where $\alpha, \delta > 0$ are dimension-free constants. It is classical that this differential inequality implies

$$\mathbb{E}_x\big[e^{\theta X_i(t)}\big] \leq \frac{\alpha}{\delta} + \left(\mathbb{E}_x\big[e^{\theta X_i(1)}\big] - \frac{\alpha}{\delta}\right) e^{-\delta(t-1)}, \tag{37}$$

for $t \geq 1$. On the other hand, for $t \in [0, 1]$, the domination (25) implies $\mathbb{E}_x[e^{\theta X_i(t)}] \leq e^{\theta x + e^\theta - 1}$. Combining these two facts, we conclude that there is a dimension-free $\kappa \in (0, \infty)$ such that

$$\mathbb{E}_x\big[e^{\theta X_i(t)}\big] \leq \kappa\big(1 + e^{\theta x_i - \delta t}\big), \tag{38}$$

for all $t \geq 0$. By Jensen's inequality and $(1 + u)^p \leq 1 + u^p$, the conclusion still holds if we replace (κ, θ, δ) with $(\kappa^p, \theta p, \delta p)$ for any $p \in (0, 1)$. Choosing p sufficiently small so that $\kappa^p \leq 2$ and $\theta p \leq 1$ completes the proof of (30). The claim (31) is then a consequence of Chernov's bound. $\square$

2.3. *Path coupling via tagged particles.* Our proof of Theorem 2 will rely on the introduction of *tagged particles*. For $k \in \mathbb{Z}_+$, define

$$\Delta(k) := r(k + 1) - r(k) \geq 0. \tag{39}$$

Let Θ be a Poisson process of intensity $\frac{1}{n}\mathrm{Leb} \otimes \mathrm{Leb} \otimes \mathrm{Card}$ on $\mathbb{R}_+ \times [0, 1] \times [n]$, independent of the Poisson process Ξ used in the graphical construction of $\mathbf{X}$, and construct an $[n]$-valued process $\mathbf{I} = (I(t) \colon t \geq 0)$ by setting $I(0) = i$ and imposing the following jumps: for each (t, u, k) in Θ,

$$I(t) := \begin{cases} k & \text{if } \Delta\big(X_{I(t-)}(t-)\big) \geq u, \\ I(t-) & \text{else.} \end{cases} \tag{40}$$

In other words, conditionally on the *background process* $\mathbf{X}$, the *tagged particle* $\mathbf{I}$ performs a time-inhomogeneous random walk starting from i and jumping from a site ℓ to a uniformly chosen site at the time-varying rate $\Delta(X_\ell(t))$. The elementary but crucial observation is that the process

$$\big(X(t) + \delta_{I(t)} \colon t \geq 0\big) \tag{41}$$

is then distributed as a zero-range process starting from $x + \delta_i$. Now, if j is another site, we may introduce a second tagged particle $\mathbf{J} = (J(t) \colon t \geq 0)$ by setting $J(0) = j$ and for each (t, u, k) in Θ,

$$J(t) := \begin{cases} k & \text{if } \Delta\big(X_{J(t-)}(t-)\big) \geq u, \\ J(t-) & \text{else.} \end{cases} \tag{42}$$

We emphasize that we use the same processes $\mathbf{X}$, Θ to generate $\mathbf{I}$ and $\mathbf{J}$. This produces two coupled zero-range processes $(X(t) + \delta_{I(t)} \colon t \geq 0)$ and $(X(t) + \delta_{J(t)} \colon t \geq 0)$ starting from $x + \delta_i$ and $x + \delta_j$, respectively. We define their *coalescence time* as

$$\tau := \inf\big\{t \geq 0 \colon I(t) = J(t)\big\}. \tag{43}$$

Note that on the event $\{\tau \geq t\}$, we have $X(t) + \delta_{I(t)} = Y(t) + \delta_{J(t)}$ by construction. Therefore, writing P_z^t for the law of the zero-range process starting from z, we have

$$d_{\mathrm{TV}}\big(P_{x+\delta_i}^t, P_{x+\delta_j}^t\big) \leq \mathbb{P}(\tau > t). \tag{44}$$

Our goal will now consist in estimating the right-hand side.

PROPOSITION 2 (Coalescence). *There is a dimension-free κ such that*

$$\mathbb{P}\big(\tau > \kappa\big(\|x\|_\infty \vee (\ln n)^\kappa\big)\big) \leq \frac{\kappa}{n^2}.$$

Let us first quickly see how this leads to Theorem 2.

PROOF OF THEOREM 2. By stationarity of π and convexity of $d_{\mathrm{TV}}(\cdot, \cdot)$,

$$d_{\mathrm{TV}}\big(P_x^t, \pi\big) \leq \sum_{y \in \Omega} \pi(y) d_{\mathrm{TV}}\big(P_x^t, P_y^t\big). \tag{45}$$

Call $x, y \in \Omega$ *adjacent* if they differ by a single jump, that is, $y = x + \delta_j - \delta_i$ for some $1 \leq i \neq j \leq n$. When this is the case, (44) and Proposition 2 (with $m - 1$ background particles) yield

$$t \geq \kappa\big(\|x\|_\infty \vee \|y\|_\infty \vee (\ln n)^\kappa\big) \quad \Longrightarrow \quad d_{\mathrm{TV}}\big(P_x^t, P_y^t\big) \leq \frac{\kappa}{n^2}. \tag{46}$$

Now if $x, y \in \Omega$ are arbitrary, one can always connect them by a path, that is, a sequence $(w_0, w_1, \ldots, w_k)$ where $w_0 = x$, $w_k = y$ and $w_{\ell-1}$ is adjacent to w_ℓ for $1 \leq \ell \leq k$. By the triangle inequality, we have

$$d_{\mathrm{TV}}\big(P_x^t, P_y^t\big) \leq \sum_{\ell=1}^{k} d_{\mathrm{TV}}\big(P_{w_{\ell-1}}^t, P_{w_\ell}^t\big). \tag{47}$$

Choosing a shortest path further ensures that $k \le m$ and $\max_{1\le \ell<k} \|w_\ell\|_\infty \le \|x\|_\infty \vee \|y\|_\infty$, so that

$$t \ge \kappa\big(\|x\|_\infty \vee \|y\|_\infty \vee (\ln n)^\kappa\big) \quad \Longrightarrow \quad d_{\mathrm{TV}}\big(P_x^t, P_y^t\big) \le \frac{\kappa m}{n^2}. \tag{48}$$

In particular, if $t \ge \kappa(\|x\|_\infty \vee (\ln n)^\kappa)$, then the restriction of the sum in (45) to the index set $A := \{y \in \Omega \colon \|y\|_\infty \le (\ln n)^\kappa\}$ is at most $\frac{\kappa m}{n^2}$. On the other hand, the remaining part is at most $\pi(A^c) \le 4ne^{-\theta(\ln n)^\kappa}$, as can be seen by taking the $t \to \infty$ limit in (31). In conclusion, for all $x \in \Omega$,

$$t \ge \kappa\big(\|x\|_\infty \vee (\ln n)^\kappa\big) \quad \Longrightarrow \quad d_{\mathrm{TV}}\big(P_x^t, \pi\big) \le \frac{m\kappa}{n^2} + 4ne^{-\theta(\ln n)^\kappa}. \tag{49}$$

Upon replacing κ by a larger constant if necessary, we obtain the claim. □

The remainder of the section is devoted to the proof of Proposition 2. It is clear from (40), (42) that if the two tagged particles manage to jump at the same time, then they immediately coalesce. Note, however, that their jumps may be severely hindered by the background process: in the rate-one case, for example, we have $\Delta(k) = \mathbf{1}_{(k=0)}$, so that the tagged particles cannot jump unless they are alone. Our first step will consist in controlling the number of co-occupants of the tagged particles. We will then complement this by showing that, when the tagged particles do not have too many co-occupants, they have a decent chance to coalesce within a short time-interval.

LEMMA 2 (Co-occupants of the tagged particles). *There are dimension-free constants $\kappa_1, \kappa_2 < \infty$ so that for $t = \kappa_1 \|x\|_\infty$ and $a = \kappa_2 \ln(1 + \|x\|_\infty)$,*

$$\mathbb{P}\big(X_{I(t)}(t) \vee Y_{J(t)}(t) \le a\big) \ge \frac{1}{2}. \tag{50}$$

PROOF. Since $I(t), J(t)$ can only move by jumps of the form (40), (42), we necessarily have

$$I(t), J(t) \in \{i, j\} \cup \big\{k \in [n] \colon \Theta\big([0,t] \times [0,1] \times \{k\}\big) \ge 1\big\}. \tag{51}$$

Note that the random set on the right-hand side contains at most $2 + t$ elements on average. Taking a union bound over all these possibilities, and using the independence of $\Theta, \mathbf{X}$, we obtain

$$\mathbb{P}\big(X_{I(t)}(t) \vee Y_{J(t)}(t) > a\big) \le (2+t) \max_{k\in[n]} \mathbb{P}\big(X_k(t) > a\big).$$

To make this less than a half, we may choose $t = \frac{1}{\delta}\|x\|_\infty$ with δ as in (31), and $a = \frac{1}{\theta}\ln(16 + 8t)$. □

LEMMA 3 (Quick coalescence). *Set $h := \frac{3(x_i \vee x_j)+1}{r(1)}$. If $n \ge 3$, then*

$$\mathbb{P}(\tau \le h) \ge \frac{e^{-3h}}{8}. \tag{52}$$

PROOF. Write $h = t + s$ with $t = 3\frac{x_i \vee x_j}{r(1)}$ and $s = \frac{1}{r(1)}$, and note that $\{\tau \le h\} \supseteq G_i \cap G_j \cap F$, where

$$\begin{aligned}
G_i &:= \big\{\Xi\big([0,t+s] \times [0,1] \times [n] \times \{i\}\big) = 0\big\} \\
&\quad\ \cap \big\{\Xi\big([0,t] \times \big[0, r(1)\big] \times \{i\} \times [n] \setminus \{i,j\}\big) \ge x_i\big\}, \\
G_j &:= \big\{\Xi\big([0,t+s] \times [0,1] \times [n] \times \{j\}\big) = 0\big\} \\
&\quad\ \cap \big\{\Xi\big([0,t] \times \big[0, r(1)\big] \times \{j\} \times [n] \setminus \{i,j\}\big) \ge x_j\big\}, \\
F &:= \big\{\Theta\big([0,t] \times [0,1] \times [n]\big) = 0\big\} \cap \big\{\Theta\big([t,t+s] \times \big[0, r(1)\big] \times [n]\big) \ge 1\big\}.
\end{aligned}$$

Indeed, the events G_i, G_j guarantee that X_i, X_j are zero over the time-interval $[t, t+s]$, while F ensures that the tagged particles will remain in positions i, j until time t, and then make an attempt to jump over $[t, t+s]$. The first such attempt will be successful for both particles, because the conditions in (40), (42) are met (note that $\Delta(0) = r(1)$). Now, F, G_i, G_j are independent, with

$$\mathbb{P}(G_k) = e^{-t-s}\mathcal{P}\left(\frac{(n-2)r(1)t}{n}; x_k\right) \quad \text{and} \quad \mathbb{P}(F) = e^{-t}\mathcal{P}\big(r(1)s; 1\big),$$

where we recall that $\mathcal{P}(\lambda; k)$ is the probability that a Poisson variable with mean λ is at least k. The claim now easily follows from the classical estimate $\mathcal{P}(\lambda; k) > \frac{1}{2}$, valid for any $\lambda \geq k \geq 0$. $\square$

COROLLARY 2. *There is a dimension-free constant $\beta < \infty$ such that*

$$\mathbb{P}\big(\tau \leq \beta\|x\|_\infty\big) \geq \big(1 + \|x\|_\infty\big)^{-\beta}. \tag{53}$$

PROOF. Set $t = \kappa_1\|x\|_\infty$, $a = \kappa_2 \ln(1 + \|x\|_\infty)$, $h = (3a+1)/r(1)$ with κ_1, κ_2 as in Lemma 2. Then $\mathbb{P}(\tau \leq t + h)$ is at least

$$\mathbb{P}\big(X_{I(t)}(t) \vee X_{J(t)}(t) \leq a\big)\mathbb{P}\big(\tau \leq t + h | X_{I(t)}(t) \vee X_{J(t)}(t) \leq a\big).$$

The first term is at least $\frac{1}{2}$ and the second at least $\frac{1}{8}e^{-3h}$, by Lemma 3 and the Markov property. $\square$

PROOF OF PROPOSITION 2. Let β be as in the above corollary, and let $t, a \geq 0$ be parameters to be adjusted later. Consider the decreasing sequence of events $(A_k)_{k\geq 0}$ defined by

$$A_k := \{\tau > t + ka\beta\} \cap \bigcap_{\ell=0}^{k-1}\big\{\big\|X(t + \ell a\beta)\big\|_\infty \leq a\big\}. \tag{54}$$

By the above corollary and the Markov property, we have $\mathbb{P}(A_{k+1}|A_k) \leq 1 - (1+a)^{-\beta}$. Thus,

$$\mathbb{P}(A_k) \leq \big(1 - (1+a)^{-\beta}\big)^k \leq e^{-k(1+a)^{-\beta}}. \tag{55}$$

On the other hand, it is clear from the definition of A_k that

$$\mathbb{P}(\tau > t + ka\beta) \leq \mathbb{P}(A_k) + k\sup_{s \geq t}\mathbb{P}\big(\big\|X(s)\big\|_\infty > a\big). \tag{56}$$

Recalling (31), we conclude that for $t = \frac{1}{\delta}\|x\|_\infty$,

$$\mathbb{P}(\tau > t + ka\beta) \leq e^{-k(1+a)^{-\beta}} + 4kne^{-\theta a}. \tag{57}$$

Choosing $a = \frac{4}{\theta}\ln n$ and $k = \lfloor(\ln n)^{2+\beta}\rfloor$ ensures that the right-hand side is $O(\frac{1}{n^2})$, as desired. $\square$

3. Dissolution of the solid phase.

3.1. *Identification of the hydrodynamic limit.* With the setting of Theorem 3 in mind, we fix a sequence of numbers $u_1 \geq u_2 \geq \cdots \geq 0$ such that

$$\sum_{k=1}^{\infty} u_k \leq \rho. \tag{58}$$

PROPOSITION 3. *There is a unique measurable $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ satisfying*

$$f(t) = \int_0^t \left\{1 - \Psi^{-1}\left(\rho - \sum_{k=1}^{\infty} [u_k - f(s)]_+\right)\right\} \mathrm{d}s, \tag{59}$$

for all $t \geq 0$. Moreover, f is an increasing bijection and for each $i \geq 1$,

$$f^{-1}(u_i) = \sum_{k=i}^{\infty} \frac{1}{k} \int_{\rho_k}^{\rho_{k-1}} \frac{\mathrm{d}s}{1 - \Psi^{-1}(s)}, \tag{60}$$

where the numbers $\rho_0 \geq \rho_1 \geq \cdots \geq 0$ are given by $\rho_k := \rho + ku_{k+1} - \sum_{i=1}^k u_i$.

PROOF OF UNIQUENESS. Fix $t > 0$. Since $\Psi^{-1}\colon \mathbb{R}_+ \to [0, 1)$ is increasing, any solution to (59) must satisfy

$$t\big(1 - \Psi^{-1}(\rho)\big) \leq f(t) \leq t.$$

Since $\kappa(t) := \max\{k \geq 1 : u_k > t(1 - \Psi^{-1}(\rho))\}$ is finite, we deduce that

$$\sum_{k=1}^{\infty} [u_k - f(t)]_+ = \sum_{k=1}^{\kappa(t)} [u_k - f(t)]_+. \tag{61}$$

In particular, if g is another solution to (59), we have

$$\left|\sum_{k=1}^{\infty} [u_k - f(t)]_+ - \sum_{k=1}^{\infty} [u_k - g(t)]_+\right| \leq \kappa(t)\big|f(t) - g(t)\big|. \tag{62}$$

Now, Ψ^{-1} is continuously differentiable and hence α-Lipschitz on $[0, \rho]$ for some $\alpha < \infty$. Therefore,

$$\begin{aligned} &\left|\Psi^{-1}\left(\rho - \sum_{k=1}^{\infty} [u_k - f(t)]_+\right) - \Psi^{-1}\left(\rho - \sum_{k=1}^{\infty} [u_k - g(t)]_+\right)\right| \\ &\quad \leq \alpha\kappa(t)\big|f(t) - g(t)\big|. \end{aligned}$$

Integrating this and recalling (59), we obtain the differential inequality

$$\big|f(t) - g(t)\big| \leq \alpha \int_0^t \kappa(s)\big|f(s) - g(s)\big| \,\mathrm{d}s. \tag{63}$$

In order to apply Grönwall's lemma and conclude that $f = g$, we now only need to check that $\kappa \in L^1(\mathbb{R}_+)$. But

$$\int_0^{\infty} \kappa(t) \,\mathrm{d}t = \frac{1}{1 - \Psi^{-1}(\rho)} \sum_{k=1}^{\infty} u_k, \tag{64}$$

by Fubini's theorem, and the right-hand side is indeed finite. $\square$

EXPLICIT RESOLUTION. Let $\Phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ be the function defined by

$$\Phi(t) := \int_0^t \frac{1}{1 - \Psi^{-1}(s)} \,\mathrm{d}s. \tag{65}$$

Φ increases continuously from 0 to $+\infty$, so $\Phi^{-1}\colon \mathbb{R}_+ \to \mathbb{R}_+$ is well defined. Now, let $\rho = \rho_0 \geq \rho_1 \geq \cdots \geq 0$ and $t_1 \geq t_2 \geq \cdots \geq 0$ be defined by

$$\rho_k := \rho + ku_{k+1} - \sum_{i=1}^{k} u_i, \tag{66}$$

$$t_k := \sum_{i=k}^{\infty} \frac{\Phi(\rho_{i-1}) - \Phi(\rho_i)}{i}. \tag{67}$$

Finally, define a function f separately on each $[t_{k+1}, t_k), k \geq 0$ (with the convention $t_0 = +\infty$) by

$$\forall t \in [t_{k+1}, t_k), \quad f(t) := u_{k+1} + \frac{\Phi^{-1}(\Phi(\rho_k) + k(t - t_{k+1})) - \rho_k}{k}.$$

Note that $f(t_k) = u_k$ for $k \geq 1$. Moreover, the left limit of f at t_k is

$$\begin{aligned} f(t_k-) &= u_{k+1} + \frac{\Phi^{-1}(\Phi(\rho_k) + k(t_k - t_{k+1})) - \rho_k}{k} \\ &= u_{k+1} + \frac{\rho_{k-1} - \rho_k}{k} \\ &= u_k. \end{aligned}$$

This shows that f is continuous at each $t_k, k \geq 1$. Since f is clearly continuously increasing on each $[t_k, t_{k-1})$, we deduce that f is continuously increasing on the whole of $(0, \infty)$. Moreover,

$$f(0+) = \lim_{k\to\infty} \downarrow f(t_k) = \lim_{k\to\infty} \downarrow u_k = 0,$$

so setting $f(0) := 0$ extends f into a continuously increasing function on $\mathbb{R}_+$. The strict monotony together with the fact that $f(t_k) = u_k$ shows that for all $t \in \mathbb{R}_+$ and $k \geq 1$,

$$f(t) < u_k \iff t < t_k. \tag{68}$$

Finally, f is continuously differentiable on each (t_{k+1}, t_k) and for $t \in (t_{k+1}, t_k)$, we easily compute

$$\begin{aligned} f'(t) &= 1 - \Psi^{-1}\left(\rho + kf(t) - \sum_{i=1}^{k} u_i\right) \\ &= 1 - \Psi^{-1}\left(\rho - \sum_{i=1}^{\infty} [u_i - f(t)]_+\right), \end{aligned}$$

where the second equality follows from (68). Thus, f is a solution to (59), and (60) is clear. □

3.2. *Proxy for the empirical distribution.* The purpose of this section is to obtain a good approximation for the empirical measure $\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i(t)}$, out of which we will then extract a good approximation for the mean-field jump rate $\zeta(t)$. For each $z \in (0, 1)$, we define a probability distribution $\mathfrak{q}(z) = (\mathfrak{q}(z; k))_{k\geq 0}$ on $\mathbb{Z}_+$ by the formula

$$\mathfrak{q}(z; k) := \frac{1}{R(z)} \frac{z^k}{r(1)\cdots r(k)}. \tag{69}$$

We extend this definition to $z = 0$ by setting $\mathfrak{q}(0) = \delta_0$. Note that the mean of $\mathfrak{q}(z)$ is precisely $\Psi(z)$. It will be convenient to reparameterize $\mathfrak{q}$ in terms of its mean by setting $\overline{\mathfrak{q}}(s) := \mathfrak{q}(\Psi^{-1}(s))$ for $s \in (0, \infty)$. We start by showing that $\overline{\mathfrak{q}}(\rho)$ is the limiting empirical distribution at equilibrium.

LEMMA 4 (Empirical distribution at equilibrium). *In the regime* (8), *we have for any fixed* $\varepsilon > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \pi\left(\left\{x \in \Omega : d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}, \overline{\mathfrak{q}}(\rho)\right) \geq \varepsilon\right\}\right) < 0. \tag{70}$$

PROOF. In the degenerate case $\rho = 0$, the claim is trivial since any law on $\mathbb{Z}_+$ is at total-variation distance at most its mean from the Dirac mass δ_0. We henceforth assume that $\rho > 0$, and we set $z = \Psi^{-1}(\rho) \in (0, 1)$. Consider a random vector $X := (X_1, \dots, X_n)$ whose coordinates are i.i.d. with law $\mathfrak{q}(z)$. Then for any $x = (x_1, \dots, x_n) \in \Omega$, we have

$$\mathbb{P}(X = x) = \frac{z^m}{(R(z))^n} \prod_{i=1}^{n} \prod_{k=1}^{x_i} \frac{1}{r(k)}. \tag{71}$$

Thus, $x \mapsto \mathbb{P}(X = x)$ is proportional to π on Ω, and hence for any $A \subseteq \Omega$, we have the representation

$$\pi(A) = \frac{\mathbb{P}(X \in A)}{\mathbb{P}(X \in \Omega)}. \tag{72}$$

We now fix $\varepsilon > 0$ and take

$$A := \left\{ x \in \Omega : d_{\mathrm{TV}}\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}, \mathfrak{q}(z) \right) \geq \varepsilon \right\}. \tag{73}$$

Since the coordinates of X are i.i.d. with law $\mathfrak{q}(z)$, Sanov's theorem (see, e.g., [9]) implies

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(X \in A) < 0; \tag{74}$$

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(S_n \in [0, m]\big) = 0; \tag{75}$$

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\big(S_n \in [m, 2m]\big) = 0, \tag{76}$$

where $S_n = X_1 + \cdots + X_n$. On the other hand, since $\mathfrak{q}(z)$ is log-concave, and since this property is preserved under convolutions (see, e.g., [26]), the law of S_n is log-concave, and hence unimodal, so for any $a \in \mathbb{N}$,

$$(a + 1)\mathbb{P}(S_n = m) \geq \mathbb{P}\big(S_n \in [m - a, m]\big) \wedge \mathbb{P}\big(S_n \in [m, m + a]\big). \tag{77}$$

Taking $a = m$ and using (75)–(76), we get $\frac{1}{n} \log \mathbb{P}(X \in \Omega) \to 0$; (72)–(74) completes the proof. □

REMARK 1 (Monotony and regularity of $\rho \mapsto \overline{\mathfrak{q}}(\rho)$). The monotonicity (24) shows that the stationary law π is stochastically increasing in the number m of particles. In view of Lemma 4, we deduce that $\overline{\mathfrak{q}}(\rho)$ is stochastically increasing in ρ: if $\rho \leq \rho'$, then there is a coupling (Z, Z') of $\overline{\mathfrak{q}}(\rho), \overline{\mathfrak{q}}(\rho')$ such that $Z \leq Z'$ almost surely. Since Z, Z' are integer-valued, we may then write

$$d_{\mathrm{TV}}\big(\overline{\mathfrak{q}}(\rho), \overline{\mathfrak{q}}(\rho')\big) \leq \mathbb{E}\big[|Z' - Z|\big] = \mathbb{E}[Z'] - \mathbb{E}[Z] = \rho' - \rho. \tag{78}$$

In conclusion, $\rho \mapsto \overline{\mathfrak{q}}(\rho)$ is increasing and 1-Lipschitz.

We will show that the approximation $\frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \approx \overline{\mathfrak{q}}(\rho)$ remains valid out of equilibrium, provided ρ is replaced by an *effective density*, obtained by ignoring the particles in the solid phase. To formalize this, we assume that the solid phase is initially restricted to some fixed region $\{1, \dots, L\}$:

$$\max_{L < i \leq n} x_i = o(n). \tag{79}$$

Note that by (31), this property is preserved by the dynamics in the sense that for any fixed $t \geq 0$,

$$\max_{L<i\leq n} X_i(nt) = o(n), \tag{80}$$

almost surely (as long as all processes live on the same probability space).

PROPOSITION 4 (Proxy for the empirical measure). *If $x = x(n)$ satisfies (79), then for fixed $t > 0$,*

$$\mathbb{E}_x\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(nt)}, \overline{\mathfrak{q}}\left(\frac{1}{n}\sum_{i=L+1}^{n} X_i(nt)\right)\right)\right] \xrightarrow[n\to\infty]{} 0. \tag{81}$$

Since the rate function r has mean z under the law $\mathfrak{q}(z)$, we have in particular

$$\mathbb{E}_x\left[\left|\zeta(nt) - \Psi^{-1}\left(\frac{m}{n} - \frac{1}{n}\sum_{i=1}^{L} X_i(nt)\right)\right|\right] \xrightarrow[n\to\infty]{} 0. \tag{82}$$

Our proof will consist in comparing the system with one where the solid phase is removed, so that Theorem 2 becomes applicable. We will rely on the following lemma.

LEMMA 5 (Truncation). *Fix $x \in \mathbb{Z}_+^n$, and let $\widehat{x}$ be obtained by zeroing the first L coordinates. Then the processes $\mathbf{X}, \widehat{\mathbf{X}}$ obtained by applying the graphical construction to $x, \widehat{x}$ satisfy*

$$\mathbb{E}\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(t)}, \frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{X}_i(t)}\right)\right] \leq \frac{L(1+t)}{n}.$$

PROOF. By (24), we have $\widehat{X}(t) \leq X(t)$ for all $t \geq 0$. In particular,

$$\sum_{i=L+1}^{n} \left|X_i(t) - \widehat{X}_i(t)\right| = \sum_{i=L+1}^{n} X_i(t) - \sum_{i=L+1}^{n} \widehat{X}_i(t).$$

Now, observe that the right-hand side equals zero when $t = 0$, and that the only jumps of the form (23) that may increment it (by 1 unit each time) are those whose source i is in $[L]$. Consequently,

$$\mathbb{E}\left[\sum_{i=L+1}^{n} \left|\widehat{X}_i(t) - X_i(t)\right|\right] \leq \mathbb{E}\left[\Xi\left([0,t]\times[0,1]\times[L]\times[n]\right)\right] = Lt. \tag{83}$$

On the other hand, by definition of the total-variation distance, we have

$$d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(t)}, \frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{X}_i(t)}\right) \leq \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{(X_i(t)\neq\widehat{X}_i(t))}.$$

To conclude, we simply bound $\mathbf{1}_{(X_i(t)\neq\widehat{X}_i(t))}$ by 1 for $i \leq L$, and by $|X_i(t) - \widehat{X}_i(t)|$ for $i > L$. $\square$

PROOF OF PROPOSITION 4. If $L = 0$, then $\|x\| = o(n)$, so Theorem 2 ensures that for fixed $t > 0$,

$$\max_{A\subseteq\Omega}\left|\mathbb{P}_x\left(X(nt)\in A\right) - \pi(A)\right| \xrightarrow[n\to\infty]{} 0. \tag{84}$$

Since the event A in Lemma 4 satisfies $\pi(A) \to 0$, we must have $\mathbb{P}_x(X(nt) \in A) \to 0$. Thus,

$$\mathbb{E}_x\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(nt)}, \overline{\mathfrak{q}}(\rho)\right)\right] \xrightarrow[n\to\infty]{} 0.$$

On the other hand, we have $d_{\mathrm{TV}}(\overline{\mathfrak{q}}(\frac{m}{n}), \overline{\mathfrak{q}}(\rho)) \to 0$, so the case $L=0$ is proved. Now, assume that x satisfies (79) for some $L \geq 1$, and let $\widehat{x}$ be as in Lemma 5. Then $\|\widehat{x}\|_\infty = o(n)$ by construction, so the case $L=0$ with $\widehat{m} := m - (x_1 + \cdots + x_L)$ particles instead of m implies

$$\mathbb{E}\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{X}_i(nt)}, \overline{\mathfrak{q}}\left(\frac{\widehat{m}}{n}\right)\right)\right] \xrightarrow[n\to\infty]{} 0.$$

On the other hand, under the coupling of Lemma 5, we have

$$\mathbb{E}\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\widehat{X}_i(nt)}, \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(nt)}\right)\right] \leq L\left(t+\frac{1}{n}\right).$$

Finally, Remark 1 implies that

$$d_{\mathrm{TV}}\left(\overline{\mathfrak{q}}\left(\frac{\widehat{m}}{n}\right), \overline{\mathfrak{q}}\left(\sum_{i=L+1}^{n}\frac{X_i(nt)}{n}\right)\right) \leq \frac{1}{n}\left|\sum_{i=1}^{L}X_i(nt) - \sum_{i=1}^{L}x_i\right|,$$

and the right-hand side has mean at most $2Lt$ by (25). By the triangle inequality, we conclude that

$$\limsup_{n\to\infty}\mathbb{E}_x\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(nt)}, \overline{\mathfrak{q}}\left(\frac{1}{n}\sum_{i=L+1}^{n}X_i(nt)\right)\right)\right] \leq 3Lt.$$

This may seem rather weak compared to what we want to establish. However, by (80), we may apply this result with x replaced by $X(ns)$ and then invoke the Markov property to obtain

$$\limsup_{n\to\infty}\mathbb{E}_x\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(ns+nt)}, \overline{\mathfrak{q}}\left(\frac{1}{n}\sum_{i=L+1}^{n}X_i(ns+nt)\right)\right)\right] \leq 3Lt,$$

for any $s \geq 0$ and $t > 0$. Replacing t with ε and s with $t-\varepsilon$, we see that for any $0 < \varepsilon \leq t$,

$$\limsup_{n\to\infty}\mathbb{E}_x\left[d_{\mathrm{TV}}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(nt)} - \overline{\mathfrak{q}}\left(\frac{1}{n}\sum_{i=L+1}^{n}X_i(nt)\right)\right)\right] \leq 3L\varepsilon.$$

Since ε can be made arbitrarily small, the first claim (81) follows. Finally, the second claim (82) is an immediate consequence of the first claim and the general fact that

$$\left|\int h\,\mathrm{d}\mu - \int h\,\mathrm{d}\nu\right| \leq \|h\|_\infty d_{\mathrm{TV}}(\mu, \nu), \tag{85}$$

for any probability measures μ, ν and any (measurable) observable h on a measurable space. ☐

3.3. *Tightness and convergence.* We are now ready to prove Theorem 3, using the classical tightness-uniqueness strategy (see, e.g., [11]). Define

$$U_i^n(t) := \frac{X_i(nt)}{n} \quad \text{and} \quad V^n(t) := \int_0^t \bigl(1-\zeta(ns)\bigr)\,\mathrm{d}s. \tag{86}$$

The fact that $U_i^n(t) \in [0, \overline{\rho}]$ and the domination (25) suffice to guarantee the tightness of $(U_i^n)_{n\geq 1}$ in the Skorokhod space $D(\mathbb{R}_+, \mathbb{R})$, and the continuity of any weak sub-sequential limit $U_i^\star$. The same conclusion applies to $(V^n)_{n\geq 1}$, because ζ is [0, 1]-valued. Our objective is to show that necessarily

$$U_i^\star = [u_i - f]_+. \tag{87}$$

By diagonal extraction, we may find a sub-sequence along which we have the joint convergence

$$\big(V^n, U_1^n, U_2^n, \dots\big) \to \big(V^\star, U_1^\star, U_2^\star, \dots\big), \tag{88}$$

with respect to the product topology. By Skorokhod's theorem (see, e.g., [2]), we may even assume for convenience that the convergence (88) is almost sure. Our plan is to pass to the limit in the martingale

$$M_i^n(t) := U_i^n(t) - U_i^n(0) + V^n(t) - \int_0^t \big(1 - r\big(nU_i^n(s)\big)\big)\,\mathrm{d}s, \tag{89}$$

which is just a rescaled version of (28). Since U_i^n has jumps of size at most $\frac{1}{n}$ occurring at rate at most $2n$, a classical concentration estimate for martingales (see, e.g., [28], Lemma 2.1) ensures that

$$M_i^n(t) \xrightarrow[n\to\infty]{} 0, \tag{90}$$

almost surely. On the other hand, (31) easily imply that for fixed $t, h \geq 0$,

$$\limsup_{n\to\infty} \max_{i\in[n]} \big\{U_i^n(t+h) - \big[U_i^n(t) - \delta h\big]_+\big\} \leq 0, \tag{91}$$

which shows in particular that $U_i^\star(t+h) \leq [U_i^\star(t) - \delta h]_+$. Thus, $U_i^\star$ is nonincreasing. Consequently, on the event $\{U_i^\star(t) > 0\}$, we have $U_i^\star(s) > 0$ for all $s \leq t$, and hence $r(nU_i^n(s)) \to 1$, by our assumption (6). Passing to the limit in (89), we conclude that the equality

$$U_i^\star(t) = u_i - V^\star(t), \tag{92}$$

holds as long as $U_i^\star(t) > 0$. But both sides of (92) are continuous and nonincreasing, so they must reach zero at the same time. Since the left-hand side is nonnegative, we conclude that for all $t \geq 0$.

$$U_i^\star(t) = \big[u_i - V^\star(t)\big]_+. \tag{93}$$

Comparing this with (87), we now only have to show that $V^\star$ solves (59), that is,

$$V^\star(t) = \int_0^t \left\{1 - \Psi^{-1}\left(\rho - \sum_{i=1}^\infty U_i^\star(s)\right)\right\} \mathrm{d}s. \tag{94}$$

Fix a nonnegative integer L. Taking $t = 0$ in (91), we deduce that

$$\limsup_{n\to\infty} \max_{L<i\leq n} U_i^n(h) \leq [u_{L+1} - \delta h]_+. \tag{95}$$

Choosing $h = \frac{u_{L+1}}{\delta}$ ensures that the right-hand side is zero. Consequently, we may apply (82) with the initial state being $X(nh)$ and use Markov's property to obtain that for any fixed $t > h$,

$$\mathbb{E}_x\left[\left|\zeta(nt) - \Psi^{-1}\left(\frac{m}{n} - \sum_{i=1}^L U_i^n(t)\right)\right|\right] \xrightarrow[n\to\infty]{} 0. \tag{96}$$

On the other hand, by continuity of Ψ^{-1}, we have almost surely,

$$\Psi^{-1}\left(\frac{m}{n} - \sum_{i=1}^L U_i^n(t)\right) \xrightarrow[n\to\infty]{} \Psi^{-1}\left(\rho - \sum_{i=1}^L U_i^\star(t)\right). \tag{97}$$

Moreover, we can safely replace L by ∞ on the right-hand side, because (95) ensures that $U_i^\star(t) = 0$ for all $i > L$. Combining this with (96), we arrive at

$$\mathbb{E}_x\left[\left|\zeta(nt) - \Psi^{-1}\left(\rho - \sum_{i=1}^{\infty} U_i^\star(t)\right)\right|\right] \xrightarrow[n\to\infty]{} 0. \tag{98}$$

This is true for any $t > h$, but $h = \frac{u_{L+1}}{\delta}$ can be made arbitrarily small by choosing L large, so (98) holds for any $t > 0$. Integrating over t, we easily deduce (94). Finally, note that the convergence

$$U_i^n \xrightarrow[n\to\infty]{} [u_i - f]_+ \tag{99}$$

is automatically uniform on compact subsets of $\mathbb{R}_+$, because the limit is continuous. It is also uniform in i, because (95) ensures that $\max_{L<i\le n} U_i^n(h)$ can be made arbitrarily small by choosing L large enough, uniformly in n. This concludes the proof of Theorem 3.

REFERENCES

[1] ALDOUS, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability, XVII. Lecture Notes in Math.* **986** 243–297. Springer, Berlin. MR0770418 https://doi.org/10.1007/BFb0068322

[2] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. MR1700749 https://doi.org/10.1002/9780470316962

[3] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

[4] BOUDOU, A.-S., CAPUTO, P., DAI PRA, P. and POSTA, G. (2006). Spectral gap estimates for interacting particle systems via a Bochner-type identity. *J. Funct. Anal.* **232** 222–258. MR2200172 https://doi.org/10.1016/j.jfa.2005.07.012

[5] BUBLEY, R. and DYER, M. (1997). Path coupling: A technique for proving rapid mixing in Markov chains. In *Proceedings of the* 38*th Annual Symposium on Foundations of Computer Science*, *FOCS* '97 223. IEEE Computer Society, Washington, DC.

[6] CAPUTO, P. (2004). Spectral gap inequalities in product spaces with conservation laws. In *Stochastic Analysis on Large Scale Interacting Systems*. *Adv. Stud. Pure Math.* **39** 53–88. Math. Soc. Japan, Tokyo. MR2073330 https://doi.org/10.2969/aspm/03910053

[7] CAPUTO, P. and POSTA, G. (2007). Entropy dissipation estimates in a zero-range dynamics. *Probab. Theory Related Fields* **139** 65–87. MR2322692 https://doi.org/10.1007/s00440-006-0039-9

[8] DAI PRA, P. and POSTA, G. (2005). Logarithmic Sobolev inequality for zero-range dynamics. *Ann. Probab.* **33** 2355–2401. MR2184099 https://doi.org/10.1214/009117905000000332

[9] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications*. *Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. MR2571413 https://doi.org/10.1007/978-3-642-03311-7

[10] DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* **93** 1659–1664. MR1374011 https://doi.org/10.1073/pnas.93.4.1659

[11] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0838085 https://doi.org/10.1002/9780470316658

[12] EVANS, M. R. and HANNEY, T. (2005). Nonequilibrium statistical mechanics of the zero-range process and related models. *J. Phys. A* **38** R195–R240. MR2145800 https://doi.org/10.1088/0305-4470/38/19/R01

[13] FATHI, M. and MAAS, J. (2016). Entropic Ricci curvature bounds for discrete interacting systems. *Ann. Appl. Probab.* **26** 1774–1806. MR3513606 https://doi.org/10.1214/15-AAP1133

[14] GRAHAM, B. T. (2009). Rate of relaxation for a mean-field zero-range process. *Ann. Appl. Probab.* **19** 497–520. MR2521877 https://doi.org/10.1214/08-AAP549

[15] HERMON, J. and SALEZ, J. (2019). A version of Aldous' spectral-gap conjecture for the zero range process. *Ann. Appl. Probab.* **29** 2217–2229. MR3984254 https://doi.org/10.1214/18-AAP1449

[16] JANVRESSE, E., LANDIM, C., QUASTEL, J. and YAU, H. T. (1999). Relaxation to equilibrium of conservative dynamics. I. Zero-range processes. *Ann. Probab.* **27** 325–360. MR1681098 https://doi.org/10.1214/aop/1022677265

[17] LACOIN, H. (2016). The cutoff profile for the simple exclusion process on the circle. *Ann. Probab.* **44** 3399–3430. MR3551201 https://doi.org/10.1214/15-AOP1053

[18] LACOIN, H. (2016). Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *Ann. Probab.* **44** 1426–1487. MR3474475 https://doi.org/10.1214/15-AOP1004

[19] LANDIM, C., SETHURAMAN, S. and VARADHAN, S. (1996). Spectral gap for zero-range dynamics. *Ann. Probab.* **24** 1871–1902. MR1415232 https://doi.org/10.1214/aop/1041903209

[20] LEVIN, D. A. and PERES, Y. (2017). *Markov Chains and Mixing Times*, 2nd ed. Amer. Math. Soc., Providence, RI. MR3726904

[21] LIGGETT, T. M. (1999). *Stochastic Interacting Systems*: *Contact*, *Voter and Exclusion Processes*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **324**. Springer, Berlin. MR1717346 https://doi.org/10.1007/978-3-662-03990-8

[22] LIGGETT, T. M. (2005). *Interacting Particle Systems*. *Classics in Mathematics*. Springer, Berlin. MR2108619 https://doi.org/10.1007/b138374

[23] MERLE, M. and SALEZ, J. (2019). Cutoff for the mean-field zero-range process. *Ann. Probab.* **47** 3170–3201. MR4021248 https://doi.org/10.1214/19-AOP1336

[24] MONTENEGRO, R. and TETALI, P. (2006). Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.* **1** x+121. MR2341319 https://doi.org/10.1561/0400000003

[25] MORRIS, B. (2006). Spectral gap for the zero range process with constant rate. *Ann. Probab.* **34** 1645–1664. MR2271475 https://doi.org/10.1214/009117906000000304

[26] SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Stat. Surv.* **8** 45–114. MR3290441 https://doi.org/10.1214/14-SS107

[27] SPITZER, F. (1970). Interaction of Markov processes. *Adv. Math.* **5** 246–290. MR0268959 https://doi.org/10.1016/0001-8708(70)90034-4

[28] VAN DE GEER, S. (1995). Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.* **23** 1779–1801. MR1370307 https://doi.org/10.1214/aos/1176324323