

# MEAN FIELD LIMITS OF LARGE JACKSON NETWORKS IN HEAVY TRAFFIC

BY RAMI ATAR <sup>1,a</sup>

<sup>1</sup>*Technion–Israel Institute of Technology*, <sup>a</sup>[rami@technion.ac.il](mailto:rami@technion.ac.il)

We consider an open Jackson network with  $n$  exchangeable single-server stations and weak all-to-all interaction through routing: upon service completion at station  $i$ , a job is routed to station  $j$  with probability  $p/n$ , where  $p \in (0, 1)$ , or leaves the system with probability  $q = 1 - p$ . We study a joint asymptotic regime in which the number of stations tends to infinity while the system approaches heavy traffic. Under the critical-load condition and diffusive scaling of time and queue length, we prove propagation of chaos for the queue-length and cumulative-idleness processes. The limiting McKean–Vlasov dynamics are described by the nonlinear reflected Brownian motion

$$\bar{X}(t) = \bar{X}_0 + \bar{W}(t) + \hat{\gamma}t + \bar{L}(t) - p\mathbb{E}\bar{L}(t),$$

where  $\bar{W}$  is a Brownian motion with variance parameter 2 and  $\bar{L}$  is the reflection term at the origin. The proof proceeds by coupling the rescaled queueing network to a Brownian particle system interacting through boundary local times. A key step is a decoupling construction that replaces the correlated routing noise in the prelimit by asymptotically equivalent independent driving noises.

**1. Introduction.** This paper studies a Jackson network with exchangeable nodes under a joint McKean–Vlasov and heavy-traffic scaling. The McKean–Vlasov, or many-station, limit corresponds to letting the number of nodes tend to infinity, whereas the diffusion limit corresponds to imposing a critical-load condition and applying diffusive scaling to the queue lengths. We show that, under this joint scaling, the diffusively scaled queue-length and idleness processes satisfy a propagation-of-chaos property and converge to a nonlinear reflected Brownian motion.

*1.1. Network structure and scaling.* The model consists of  $n$  single-server queueing stations. Server  $i$  is dedicated to queue  $i$  and serves the job at the head of the line whenever the queue is nonempty. The queue length at station  $i$  is the number of jobs in queue  $i$ , including the job in service, if any. The queue-length process is the  $n$ -dimensional process whose  $i$ th coordinate is the queue length at station  $i$ . Exogenous arrivals to each station are Poisson, and service times are exponential. Upon completion of service at station  $i$ , a job is routed to station  $j$  with probability  $p/n$ , or leaves the system with probability  $q$ , where  $p > 0$ ,  $q > 0$ , and  $p + q = 1$ . Self-routing is allowed.

On a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we are given, for each  $n$ , a  $\mathbb{Z}_+^n$ -valued random variable  $X_0^n = (X_{0i}^n)_{i \in [n]}$ , representing the initial condition, and a collection of Poisson processes  $(\Pi_{ij}^n)_{(i,j) \in J^n}$ , where

$$J^n = (\{0\} \times [n]) \cup [n]^2 \cup ([n] \times \{0\}).$$

---

*MSC2020 subject classifications.* Primary 60K25, 60K35, 60J60, 60H10.

*Keywords and phrases.* Jackson networks, McKean–Vlasov limits, reflected Brownian motion,

*Date:* May 31, 2026.

For  $i \in [n]$ , the process  $\Pi_{0i}^n$  has intensity  $\gamma_n$  and represents the stream of exogenous arrivals to station  $i$ . We assume that  $\gamma_n$  is asymptotic to  $\gamma + \hat{\gamma}n^{-1/2}$ , where  $\gamma > 0$  and  $\hat{\gamma} \in \mathbb{R}$ ; that is,

$$\lim_{n \rightarrow \infty} \hat{\gamma}_n = \hat{\gamma}, \quad \hat{\gamma}_n = n^{1/2}(\gamma_n - \gamma).$$

For  $i \in [n]$ , the process  $\Pi_{ij}^n$  has rate  $p/n$  for  $j \in [n]$ , including the case  $j = i$ , and has rate  $q$  for  $j = 0$ . These processes represent potential service completions at station  $i$ , followed either by routing to station  $j \in [n]$  or by departure from the system when  $j = 0$ . Such a potential completion is counted as an actual completion only if queue  $i$  is nonempty at the event time. The Poisson processes are assumed to be mutually independent and independent of  $X_0^n$ . For each  $i \in [n]$ , the total rate of the potential service-completion processes  $\Pi_{ij}^n$ ,  $j = 0, 1, \dots, n$ , is  $p + q = 1$ , corresponding to a unit service rate at each station.

Let  $X^n = (X_i^n)_{i \in [n]}$  be the queue-length process, modeled as the unique solution of

(1)

$$X_i^n(t) = X_{0i}^n + \Pi_{0i}^n(t) - \sum_{j=0}^n \int_0^t B_i^n(s-) d\Pi_{ij}^n(s) + \sum_{j=1}^n \int_0^t B_j^n(s-) d\Pi_{ji}^n(s), \quad i \in [n],$$

$$B_i^n(t) = 1_{\{X_i^n(t) > 0\}}, \quad i \in [n].$$

The first two terms on the right are the initial condition and the exogenous arrivals to station  $i$ . The third term counts actual service completions from station  $i$ , including both routed jobs and departures from the network, while the fourth term counts jobs routed into station  $i$  from other stations. The busyness indicators  $B_i^n$  ensure that potential service completions are counted only when the corresponding queue is nonempty. Define also the idleness indicator  $I_i^n = 1 - B_i^n$  and the cumulative idleness process

$$L_i^n(t) = \int_0^t I_i^n(s) ds.$$

To obtain a critically loaded system, we impose the heavy-traffic condition

(2)

$$\gamma = q.$$

We study the diffusively scaled queue-length and cumulative-idleness processes

$$(3) \quad \hat{X}_i^n(t) = n^{-1/2} X_i^n(nt), \quad \hat{L}_i^n(t) = n^{-1/2} L_i^n(nt), \quad i \in [n].$$

Finally, define the empirical measure of the scaled queue-length/idleness pairs and the empirical mean of the scaled cumulative idleness by

$$(4) \quad \mu_t^n = \frac{1}{n} \sum_{i \in [n]} \delta_{(\hat{X}_i^n(t), \hat{L}_i^n(t))}, \quad \lambda^n(t) = \frac{1}{n} \sum_{i \in [n]} \hat{L}_i^n(t).$$

1.2. *The limiting McKean–Vlasov process.* A one-dimensional Brownian motion with drift  $b$  and variance  $\sigma^2$  will be referred to as a  $(b, \sigma^2)$ -BM. The McKean–Vlasov process relevant to our setting, also referred to as the *nonlinear reflected Brownian motion* (RBM), is given by

$$(5) \quad \begin{aligned} \bar{X}(t) &= \bar{X}_0 + \bar{W}(t) + \hat{\gamma}t + \bar{L}(t) - p\lambda(t), \\ \lambda(t) &= \mathbb{E}\bar{L}(t). \end{aligned}$$

Here,  $\bar{W}$  is a  $(0, 2)$ -BM independent of the initial condition  $\bar{X}_0$ , and  $\bar{L}$  is the reflection term at 0. Thus  $\bar{X}$  takes values in  $\mathbb{R}_+ := [0, \infty)$ ,  $\bar{L}$  is nondecreasing,  $\bar{L}(0) = 0$ , and  $\bar{L}$  increases

only when  $\bar{X} = 0$ . By [1, Theorem 2.4], (5) admits a pathwise unique strong solution, and the processes  $\bar{X}$  and  $\bar{L}$  have continuous paths.

The form of (5) is natural in view of classical heavy-traffic theory for Jackson networks. Indeed, a classical result of Reiman [10] shows that, when  $n$  is fixed, the diffusion-scaled queue-length process converges in heavy traffic to a semimartingale RBM in  $\mathbb{R}_+^n$ , with reflection matrix determined by the routing probabilities. In the present model, routing probabilities are of order  $1/n$ , and therefore the corresponding reflection interaction is weak at the level of each pair of stations but is spread over all stations. The nonlinear RBM (5) is the mean-field analogue of this fixed-dimensional reflected Brownian approximation.

The proof uses recent results about Brownian particle systems on the half-line. Let  $(\tilde{X}^n, \tilde{L}^n)$  be the solution of

$$(6) \quad \tilde{X}_i^n(t) = \tilde{X}_{0i}^n + \tilde{W}_i(t) + \hat{\gamma}t + \tilde{L}_i^n(t) - \frac{p}{n} \sum_{j \in [n]} \tilde{L}_j^n(t), \quad i \in [n],$$

where  $(\tilde{W}_i)_{i \in [n]}$  are i.i.d.  $(0, 2)$ -BMs and the processes  $\tilde{L}_i^n$  are the corresponding reflection terms. The results of both [1] and [3] show that this Brownian system undergoes propagation of chaos, with (5) as its McKean–Vlasov limit (a minor normalization difference is addressed in Section 2.3). Our main task is to show that, under the joint heavy-traffic and many-station scaling, the queueing network can be coupled to this Brownian particle system with an error that vanishes as  $n \rightarrow \infty$ .

The connection with (6), however, is not automatic from the fixed-dimensional heavy-traffic theorem. The semimartingale RBM describing the fixed-dimensional limit is driven by a Brownian motion with dependent components, whereas (6) is driven by i.i.d. BMs. One of the main steps of the proof is to overcome this dependence by an asymptotically negligible perturbation of the prelimit driving noise. The mechanism behind this perturbation is described after Theorem 1.1 and carried out in Lemma 3.2.

**1.3. Main result.** Some notation needed to present the main result is as follows. For a Polish space  $E$ , let  $C(\mathbb{R}_+, E)$  be the space of continuous paths  $\mathbb{R}_+ \rightarrow E$ , equipped with the local uniform topology, and let  $D(\mathbb{R}_+, E)$  be the space of càdlàg paths  $\mathbb{R}_+ \rightarrow E$ , equipped with the  $J_1$  topology. In particular, for  $n \in \mathbb{N}$  let

$$C^{(n)} = C(\mathbb{R}_+, \mathbb{R}^n), \quad D^{(n)} = D(\mathbb{R}_+, \mathbb{R}^n).$$

Let  $\mathcal{P}(E)$  denote the set of probability measures on  $E$ , equipped with the topology of weak convergence. Write  $\Rightarrow$  for convergence in distribution.

The initial conditions are assumed to satisfy the following.

**ASSUMPTION 1.** For each  $n$ , the initial conditions  $(\hat{X}_{0i}^n)_{i \in [n]}$  are exchangeable. Moreover, there exists  $\mu_0^{(1)} \in \mathcal{P}(\mathbb{R}_+)$  such that

$$\frac{1}{n} \sum_{i \in [n]} \delta_{\hat{X}_{0i}^n} \rightarrow \mu_0^{(1)}$$

in  $\mathcal{P}(\mathbb{R}_+)$  in probability.

Our main result is the following.

**THEOREM 1.1.** *Let Assumption 1 hold. Let  $k \in \mathbb{N}$ , and let  $(\bar{X}, \bar{L}, \lambda)$  be the solution of (5) with  $\bar{X}_0 \sim \mu_0^{(1)}$ . Then*

$$((\hat{X}_i^n)_{i \in [k]}, (\hat{L}_i^n)_{i \in [k]}) \Rightarrow ((\bar{X}_i)_{i \in [k]}, (\bar{L}_i)_{i \in [k]})$$

in  $D^{(k)} \times C^{(k)}$ , where  $(\bar{X}_i, \bar{L}_i)$  are  $k$  independent copies of  $(\bar{X}, \bar{L})$ . Moreover,

$$(\mu^n, \lambda^n) \rightarrow (\mu, \lambda)$$

in  $D(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+^2)) \times C^{(1)}$ , in probability, where

$$\mu_t = \mathbb{P} \circ (\bar{X}(t), \bar{L}(t))^{-1}, \quad t \geq 0.$$

The theorem shows that the joint heavy-traffic and many-station limit is described by independent copies of the nonlinear RBM (5), and the empirical distribution of the scaled queue-length/idleness pairs converges to the law of the McKean–Vlasov process. As mentioned above, the proof proceeds by coupling the queueing network to the Brownian particle system (6). Two estimates are needed for this purpose. The first shows that an idleness-dependent martingale term in the queueing dynamics is negligible. The second addresses the dependence in the prelimit driving noise. This dependence comes from the row and column sums of a common Poisson routing array. We construct an independent replacement for the column-sum process, with the same law as the original, whose difference from the original column-sum process is negligible after diffusion scaling. This turns the correlated prelimit driving processes into independent martingales whose limits are independent BMs. These two ingredients reduce the proof to the propagation-of-chaos result for (6) established in [1].

REMARK 1. Unlike standard McKean–Vlasov diffusions, where the interaction acts through the state variable, the interaction in (5) is carried by the reflection process. This originates from the queueing network:  $L_i^n(t)$  is the cumulative idle time of server  $i$ , and the routing mechanism couples stations through the empirical average  $\lambda^n(t)$  (this is apparent in equation (8)). In the heavy-traffic limit, idleness is encoded by the reflection term, leading to the interaction term  $p\mathbb{E}[\bar{L}(t)]$  in (5).

REMARK 2. One common use of diffusion approximations for queueing networks is to study stationary behavior by first passing to a scaling limit and then sending time to infinity in the limiting model. This work is not aimed at deriving stationary approximations. Indeed, for the model studied here, the stationary distribution of the prelimit is explicit. If  $\hat{\gamma}_n < 0$ , or equivalently  $\gamma_n < q$ , then the traffic intensity at each station is

$$\rho_n = \frac{\gamma_n}{q},$$

and the product-form stationary distribution is given by

$$\pi^n(x) = \prod_{i \in [n]} (1 - \rho_n) \rho_n^{x_i}, \quad x = (x_1, \dots, x_n) \in \mathbb{Z}_+^n.$$

In particular, when  $\hat{\gamma}_n \rightarrow \hat{\gamma} < 0$ , the stationary law of  $n^{-1/2} X_i^n$  converges to the exponential distribution with rate  $|\hat{\gamma}|/q$ . Thus the interest here is not in deriving new stationary distributions, but rather in describing the time-dependent, or transient, behavior of the network under the scaling.

1.4. *Related work.* The fixed-dimensional heavy-traffic theory for generalized Jackson networks goes back to the aforementioned work [10], which proves convergence of diffusion-scaled queue-length processes to semimartingale RBM in the orthant. We refer to [5] for background on RBM in orthants, the associated Skorohod problem, queueing networks, and their diffusion approximations.

McKean–Vlasov limits and propagation of chaos originate with the pioneering work of McKean [9] and were developed systematically in [7] and [14]. McKean–Vlasov limits for

reflected diffusions were first studied in [13], and many extensions have since been considered. In much of this literature, the interaction acts through the empirical law in the drift or diffusion coefficients, whereas in the present setting the interaction arises through the boundary, via the cumulative idleness or local-time terms.

Particle systems with interaction through local times are closer in spirit to the limit appearing here. Such models appear in [14, Chapter II], where particles interact through their collision local times, and in recent works on systems with constraints on the empirical law; see [6] and references therein.

The propagation-of-chaos input used in the present paper comes from [1], which studies Brownian particle systems on the half-line with interaction through boundary local times and proves the corresponding limit theorem for the reference system (6), and also treats a more general model with random reflection directions. The system (6) had previously been studied in [3], where its McKean–Vlasov limit was established; the emphasis there is on regimes in which solutions may exist only locally in time, including cases such as  $p \geq 1$ .

There is also a substantial body of work on mean-field limits for queueing networks. In this direction, one lets the number of servers, stations, or nodes tend to infinity at the queueing-network level, without applying diffusion scaling to the queue lengths. Examples include [2, 11, 12]. These works are distinct from the present heavy-traffic regime and therefore from the Brownian approximation obtained here.

*1.5. Further notation.* For  $x \in \mathbb{R}$ , let  $x^\pm = \max(\pm x, 0)$ . Let  $\iota$  denote the identity map on  $\mathbb{R}_+$ ,  $\iota(t) = t$ . For  $n \in \mathbb{N}$ , equip  $\mathbb{R}^n$  with the Euclidean norm  $|\cdot|$ .

For a Polish space  $(E, d)$  and  $G \subset E$ , let

$$C_G(\mathbb{R}_+, E) = \{x \in C(\mathbb{R}_+, E) : x(0) \in G\},$$

and define  $D_G(\mathbb{R}_+, E)$  similarly. Let  $C_+(\mathbb{R}_+, \mathbb{R}_+)$  denote the set of functions in  $C(\mathbb{R}_+, \mathbb{R}_+)$  that are nondecreasing and start at 0, and define  $D_+(\mathbb{R}_+, \mathbb{R}_+)$  similarly.

Let  $\mathcal{C}^{(n)}$  and  $\mathcal{D}^{(n)}$  denote the Borel  $\sigma$ -algebras on  $C^{(n)}$  and  $D^{(n)}$ , respectively.

For  $f : \mathbb{R}_+ \rightarrow \mathbb{R}^n$  and  $0 \leq \delta \leq t$ , let

$$\|f\|_t = \sup\{|f(s)| : s \in [0, t]\}$$

and

$$w_t(f, \delta) = \sup\{|f(s) - f(u)| : s, u \in [0, t], |s - u| \leq \delta\}.$$

Finally, for any  $K = (K_i)_{i \in [n]}$ , write

$$\langle K \rangle := \frac{1}{n} \sum_{i \in [n]} K_i.$$

*1.6. Organization of the paper.* Section 2 gives several tools used in the proof. Section 2.1 derives an equation for the diffusively scaled queueing network, Section 2.2 recalls the Skorohod map on the half-line, and Section 2.3 states the propagation-of-chaos result for the Brownian particle system from [1]. Section 3 proves the main theorem. The proof is based on two ingredients: an estimate showing that an idleness-dependent martingale term is negligible, and a decoupling construction that replaces the correlated routing noise by independent BMs. These ingredients are then combined to prove Theorem 1.1.

## 2. Toolbox.

2.1. *An equation for the rescaled system.* Denote accelerated processes by

$$\check{\Pi}_{ij}^n(t) = \Pi_{ij}^n(nt), \quad \check{B}_i(t) = B_i(nt), \quad \check{I}_i(t) = I_i(nt).$$

Note that  $\check{\Pi}_{ij}^n$  are Poisson processes with rates  $\gamma_n n$ ,  $p$ ,  $qn$ . Let compensated versions be given by

$$\begin{aligned} \hat{\Pi}_{0i}^n(t) &= n^{-1/2}(\check{\Pi}_{0i}(t) - \gamma_n nt) & i \in [n] \\ \hat{\Pi}_{ij}^n(t) &= n^{-1/2}(\check{\Pi}_{ij}(t) - pt) & i \in [n], j \in [n] \\ \hat{\Pi}_{i0}^n(t) &= n^{-1/2}(\check{\Pi}_{i0}(t) - qnt) & i \in [n]. \end{aligned}$$

Our goal here is to write the dynamics of  $\hat{X}_i$  in terms of  $\hat{\Pi}_{ij}$ ,  $\check{B}_i$ , and  $\hat{L}_i$  alone. This is achieved in equations (8)-(11) below. To this end, note that by (1),

$$(7) \quad \begin{aligned} \hat{X}_i(t) &= n^{-1/2} X_i(nt) = \hat{X}_{0i} + n^{-1/2} \check{\Pi}_{0i}(t) - n^{-1/2} \sum_{j=0}^n \int_0^t \check{B}_i(s-) d\check{\Pi}_{ij}(s) \\ &\quad + n^{-1/2} \sum_{j=1}^n \int_0^t \check{B}_j(s-) d\check{\Pi}_{ji}(s). \end{aligned}$$

If we let  $G_{ij}(t) = n^{-1/2} \int_0^t \check{B}_i(s-) d\check{\Pi}_{ij}(s)$  then for  $i \in [n]$ ,  $j \in [n]$ ,

$$\begin{aligned} G_{ij}(t) &= \hat{\Pi}_{ij}(t) + pn^{-1/2}t - n^{-1/2} \int_0^t \check{I}_i(s-) d\check{\Pi}_{ij}(s) \\ &= \hat{\Pi}_{ij}(t) + pn^{-1/2}t - \int_0^t \check{I}_i(s-) d\hat{\Pi}_{ij}(s) - pn^{-1} \hat{L}_i(t), \end{aligned}$$

whereas

$$\begin{aligned} G_{i0}(t) &= \hat{\Pi}_{i0}(t) + qn^{1/2}t - n^{-1/2} \int_0^t \check{I}_i(s-) d\check{\Pi}_{i0}(s) \\ &= \hat{\Pi}_{i0}(t) + qn^{1/2}t - \int_0^t \check{I}_i(s-) d\hat{\Pi}_{i0}(s) - q\hat{L}_i(t). \end{aligned}$$

Now, the sum of the first two terms on the right of (7) equals

$$\hat{X}_{0i} + \hat{\Pi}_{0i}(t) + \gamma_n n^{1/2}t = \hat{X}_{0i} + \hat{\Pi}_{0i}(t) + \hat{\gamma}_n t + n^{1/2} \gamma t.$$

For the remaining terms on the right of (7),

$$- \sum_{j \in [n]} G_{ij} = - \sum_{j \in [n]} \hat{\Pi}_{ij}(t) - pn^{1/2}t + \sum_{j \in [n]} \int_0^t \check{I}_i(s-) d\hat{\Pi}_{ij}(s) + p\hat{L}_i(t),$$

while

$$\sum_{j \in [n]} G_{ji} = \sum_{j \in [n]} \hat{\Pi}_{ji}(t) + pn^{1/2}t - \sum_{j \in [n]} \int_0^t \check{I}_j(s-) d\hat{\Pi}_{ji}(s) - pn^{-1} \sum_{j \in [n]} \hat{L}_j(t).$$

Recalling (2), by which  $\gamma = 1 - p = q$ , this gives

$$(8) \quad \hat{X}_i = \hat{X}_{0i} + \hat{W}_i + \hat{\gamma}_n t + \hat{L}_i - p\langle \hat{L} \rangle + \hat{M}_i,$$

where

$$(9) \quad \hat{W}_i = \hat{\Pi}_{0i} - \sum_{j \geq 0} \hat{\Pi}_{ij} + \sum_{j \geq 1} \hat{\Pi}_{ji},$$

$$(10) \quad \hat{M}_i = \sum_{j \geq 0} \hat{M}_{ij} - \sum_{j \geq 1} \hat{M}_{ji}, \quad \hat{M}_{ij}(t) = \int_0^t \check{I}_i(s-) d\hat{\Pi}_{ij}(s).$$

Also, by the definition of  $\hat{L}_i$ , one has the complementarity condition

$$(11) \quad \int \hat{X}_i d\hat{L}_i = 0.$$

**2.2. The Skorohod map.** The Skorohod problem on the half line is stated as follows:

Given  $w \in D_{\mathbb{R}_+}(\mathbb{R}_+, \mathbb{R})$ , find  $x \in D(\mathbb{R}_+, \mathbb{R}_+)$  and  $\ell \in D_+(\mathbb{R}_+, \mathbb{R}_+)$  such that  $x = w + \ell$  and  $\int_{[0, \infty)} x d\ell = 0$ .

An elementary lemma provides its unique solution.

**LEMMA 2.1.** For any  $w \in D_{\mathbb{R}_+}(\mathbb{R}_+, \mathbb{R})$ , the Skorohod problem on the half line has a unique solution  $(x, \ell) \in D(\mathbb{R}_+, \mathbb{R}_+) \times D_+(\mathbb{R}_+, \mathbb{R}_+)$ . It is given by

$$(12) \quad x(t) = w(t) + \sup_{s \leq t} w^-(s), \quad \ell(t) = \sup_{s \leq t} w^-(s), \quad t \geq 0.$$

**Proof.** See [5, Theorem 6.1]. □

Denote the solution map by  $\Gamma : D_{\mathbb{R}_+}(\mathbb{R}_+, \mathbb{R}) \rightarrow D(\mathbb{R}_+, \mathbb{R}_+) \times D_+(\mathbb{R}_+, \mathbb{R}_+)$ . Here,  $\Gamma(w) = (\Gamma_1(w), \Gamma_2(w)) = (x, \ell)$  is given by (12).

Note that the map is Lipschitz in the sense that, for  $w_1, w_2 \in D_{\mathbb{R}_+}(\mathbb{R}_+, \mathbb{R})$  and  $t \in \mathbb{R}_+$ ,

$$(13) \quad \|\Gamma_1(w_1) - \Gamma_1(w_2)\|_t \leq 2\|w_1 - w_2\|_t, \quad \|\Gamma_2(w_1) - \Gamma_2(w_2)\|_t \leq \|w_1 - w_2\|_t.$$

That  $\Gamma_2$  is 1-Lipschitz will play an important role in our proofs.

Note that with the notation just introduced we can write (8)–(11) as follows. Let

$$(14) \quad \hat{Y}_i = \hat{W}_i + \hat{\gamma}_i \ell + \hat{M}_i.$$

Then

$$(15) \quad (\hat{X}_i, \hat{L}_i) = \Gamma(\hat{X}_{0i} + \hat{Y}_i - p\langle \hat{L} \rangle), \quad i \in [n].$$

We now show that (15) has a unique solution  $(\hat{X}_i^n, \hat{L}_i^n)$  for given data  $\hat{X}_{i0} + \hat{Y}_i$ .

**LEMMA 2.2.** Fix  $n \in \mathbb{N}$  and  $\beta \in (-1, 1)$ . Let  $w \in D_{\mathbb{R}_+^n}(\mathbb{R}_+, \mathbb{R}^n)$ . Then the system of equations

$$(x_i, \ell_i) = \Gamma(w_i + \beta\langle \ell \rangle), \quad i \in [n]$$

has a unique solution  $(x, \ell) \in D(\mathbb{R}_+, \mathbb{R}_+)^n \times D_+(\mathbb{R}_+, \mathbb{R}_+)^n$ .

**Proof.** A proof of this result was given in [8] for a more general class of Skorohod problems in the orthant, assuming continuous data  $w$ . Their argument is based on a contraction mapping. The same argument applies without change to càdlàg paths  $w$ , and is sufficient for our purposes here. For completeness, we include the proof in the present special case, where the argument is particularly simple.

It suffices to consider the problem on a finite time interval  $[0, T]$ . Let  $D_T^{(n)} = D([0, T], \mathbb{R}^n)$ . Fix  $w$  as in the statement of the lemma and let  $\Phi : D_T^{(n)} \rightarrow D_T^{(n)}$  be the map

$$\Phi_i(\ell) = \Gamma_2(w_i + \beta\langle \ell \rangle) = \sup_{s \leq \cdot} [(w_i(s) + \beta\langle \ell \rangle(s))^-], \quad \Phi(\ell) = (\Phi_i(\ell))_{i \in [n]}.$$

For  $\ell, \bar{\ell} \in D_T^{(n)}$ , we have

$$\max_i \|\Phi_i(\ell) - \Phi_i(\bar{\ell})\|_T \leq |\beta| \|\langle \ell \rangle - \langle \bar{\ell} \rangle\|_T \leq |\beta| \|\ell - \bar{\ell}\|_T \leq |\beta| \max_i \|\ell_i - \bar{\ell}_i\|_T.$$

Thus  $\Phi$  is a contraction on  $D_T^{(n)}$  equipped with the uniform metric, which is a complete metric space, and therefore  $\Phi$  has a unique fixed point  $\ell$ . The components of  $\ell$  are nondecreasing owing to the running supremum expression in (12). Finally, again by (12), it is easy to see that  $x = w + \ell$  takes values in  $\mathbb{R}_+^n$ .  $\square$

**2.3. The Brownian system.** Here we present a result from [1] for a Brownian particle system.

Let  $(\tilde{X}_{0i}^n)_{i \in [n]}$  be exchangeable  $\mathbb{R}_+$ -valued RVs for every  $n$ , serving as initial conditions, and satisfying  $n^{-1} \sum_{i \in [n]} \delta_{\tilde{X}_{0i}^n} \rightarrow \mu_0^{(1)}$  in  $\mathcal{P}(\mathbb{R}_+)$  in probability. Let  $(\tilde{W}_i)_{i \in [n]}$  be mutually independent  $(0, 2)$ -BMs, independent of the initial conditions. Applying Lemma 2.2, let  $(\tilde{X}_i^n, \tilde{L}_i^n)$  be the unique solution to

$$(\tilde{X}_i^n, \tilde{L}_i^n) = \Gamma(\tilde{X}_{0i}^n + \tilde{W}_i^n + \hat{\gamma}t - p\langle \tilde{L}^n \rangle).$$

Clearly, the sample paths of  $\tilde{X}^n$  and  $\tilde{L}^n$  lie in  $C(\mathbb{R}_+, \mathbb{R}_+)^n$  and  $C_+(\mathbb{R}_+, \mathbb{R}_+)^n$ , respectively. Define, similar to (4),

$$\tilde{\mu}_t^n = \frac{1}{n} \sum_{i \in [n]} \delta_{(\tilde{X}_i^n(t), \tilde{L}_i^n(t))}, \quad \tilde{\lambda}^n(t) = \frac{1}{n} \sum_{i \in [n]} \tilde{L}_i^n(t).$$

**THEOREM 2.3 ([1]).** *Let  $(\bar{X}, \bar{L}, \lambda)$  be a solution to (5) with  $\bar{X}_0 \sim \mu_0^{(1)}$ . Let  $k \in \mathbb{N}$ . Then  $(\tilde{X}_i^n, \tilde{L}_i^n)_{i \in [k]} \Rightarrow (\bar{X}_i, \bar{L}_i)_{i \in [k]}$  in  $C^{(2k)}$ , where  $(\bar{X}_i, \bar{L}_i)$  are  $k$  independent copies of  $(\bar{X}, \bar{L})$ . Moreover,  $(\tilde{\mu}^n, \tilde{\lambda}^n) \rightarrow (\mu, \lambda)$  in  $C(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+^2)) \times C^{(1)}$ , in probability, where  $\mu_t = \mathbb{P} \circ (\bar{X}(t), \bar{L}(t))^{-1}$  for  $t \geq 0$ .*

**Proof.** In [1], a Brownian particle system is considered in which the boundary terms are averaged in a slightly different way, namely

$$(16) \quad Y_i^n(t) = \tilde{X}_{0,i}^n + \tilde{W}_i^n(t) + \hat{\gamma}t + K_i^n(t) - \frac{p}{n-1} \sum_{j \neq i} K_j^n(t), \quad i \in [n], t \geq 0,$$

with  $K_i$  the reflection terms. The results show precisely the statement of the theorem for  $(Y_i^n, K_i^n)$ . Hence it suffices to show that the difference is negligible. By exchangeability, it suffices to prove that  $\|\tilde{X}_1^n - Y_1^n\|_t \rightarrow 0$  and  $\|\tilde{L}_1^n - K_1^n\|_t \rightarrow 0$  in probability. To this end, using (13),

$$\begin{aligned} \|\tilde{L}_i - K_i\|_t &\leq p \left\| \frac{1}{n-1} \sum_{j \neq i} \tilde{L}_j - \frac{1}{n} \sum_j K_j \right\|_t \\ &= p \left\| \frac{1}{n} \sum_j (\tilde{L}_j - K_j) + \left( \frac{1}{n} - \frac{1}{n-1} \right) \sum_j \tilde{L}_j - \frac{1}{n-1} \tilde{L}_i \right\|_t. \end{aligned}$$

Hence, with  $\varepsilon_n = (n-1)^{-1}$ ,  $\|\tilde{L}_i - K_i\|_t \leq p(\|\tilde{L} - K\|_t) + \varepsilon_n \langle \tilde{L} \rangle(t) + \varepsilon_n \tilde{L}_i(t)$ . Thus

$$\mathbb{E} \|\tilde{L}_1 - K_1\|_t \leq 2\varepsilon_n(1-p)^{-1} \mathbb{E} \tilde{L}_1(t) \leq 2\varepsilon_n(1-p)^{-1} \{\mathbb{E} \|\tilde{L}_1 - K_1\|_t + \mathbb{E} K_1(t)\}.$$

It follows from [1, Lemma 3.6] that  $\mathbb{E} K_1(t)$  remains bounded as  $n \rightarrow \infty$ . The above thus shows that  $\mathbb{E} \|\tilde{L}_1 - K_1\|_t \rightarrow 0$ . The result  $\mathbb{E} \|\tilde{X}_1 - Y_1\|_t \rightarrow 0$  follows.  $\square$

**3. Proof of main result.** We now state the two main lemmas on which the proof of Theorem 1.1 is based. First is an estimate on the martingale term in (8).

LEMMA 3.1.  $\lim_n \mathbb{E}[\|\hat{M}_1\|_t^2] = 0$ .

Next, let us rewrite  $\hat{W} = \hat{W}^n$  of (9) in a way that emphasizes that its structure involves the row-sum and column-sum of an array of compensated Poissons:

$$(17) \quad \hat{W}_i = \hat{\Pi}_{0i} - \hat{\Pi}_{i0} - \hat{\Pi}_i^{\text{rs}} + \hat{\Pi}_i^{\text{cs}}, \quad \hat{\Pi}_i^{\text{rs}} := \sum_{j \in [n]} \hat{\Pi}_{ij}, \quad \hat{\Pi}_i^{\text{cs}} := \sum_{j \in [n]} \hat{\Pi}_{ji}.$$

Clearly, the components  $\hat{W}_i$  are dependent. The following lemma will allow us construct a perturbation of  $\hat{W}_i$  whose components are independent.

LEMMA 3.2. *After augmenting the space with additional Poisson processes, if necessary, there exists, for each  $n$ , a tuple  $(\Pi_i^*)_{i \in [n]}$  of i.i.d. rate- $pn$  Poisson processes that are independent of  $(\hat{\Pi}_{0i}, \hat{\Pi}_{i0}, \hat{\Pi}_i^{\text{rs}})_{i \in [n]}$ , such that, denoting  $\check{\Pi}_i = n^{-1/2}(\Pi_i^* - pn)$ , the processes  $(\check{\Pi}_i - \hat{\Pi}_i^{\text{cs}})_{i \in [n]}$  are exchangeable and  $\mathbb{E}[\|\check{\Pi}_1 - \hat{\Pi}_1^{\text{cs}}\|_t] \rightarrow 0$  as  $n \rightarrow \infty$ .*

In Sections 3.1 and 3.2, we present the proofs of Lemmas 3.1 and 3.2, respectively. In Section 3.3, we give the proof of Theorem 1.1.

3.1. *Proof of Lemma 3.1 (Estimating the martingale term).* We fix  $t$  and let  $c$  denote a positive constant that does not depend on  $n$  (but may depend on  $t$ ), whose value may change from one appearance to another.

The proof proceeds in two steps, establishing the following two estimates:

$$(18) \quad \mathbb{E}[\|\hat{M}_1\|_t^2] \leq cn^{-1/2} \mathbb{E}\hat{L}_1(t)$$

$$(19) \quad \mathbb{E}\hat{L}_1(t) \leq c + c\mathbb{E}\|\hat{M}_1\|_t.$$

Together, these estimates imply the result. Indeed, the inequality  $x^2 \leq cn^{-1/2}(1+x)$  implies  $x^2 \leq cn^{-1/2}$ .

Toward proving (18), we have by (10),

$$\begin{aligned} \hat{M}_1(t) &= \hat{M}_{10}(t) + \sum_{j \in [n], j \neq 1} (\hat{M}_{1j}(t) - \hat{M}_{j1}(t)) \\ &= \int_0^t \check{I}_1(s-) d\hat{\Pi}_{10}(s) + \sum_{j \in [n], j \neq 1} \left( \int_0^t \check{I}_1(s-) d\hat{\Pi}_{1j}(s) - \int_0^t \check{I}_j(s-) d\hat{\Pi}_{j1}(s) \right). \end{aligned}$$

Because  $\hat{\Pi}_{10}, \hat{\Pi}_{1j}, \hat{\Pi}_{j1}, j \neq 1$  do not have common jump times, we have

$$\begin{aligned} [\hat{M}_1](t) &= n^{-1} \int_0^t \check{I}_1(s-) d\check{\Pi}_{10}(s) \\ &\quad + n^{-1} \sum_{j \in [n], j \neq 1} \left( \int_0^t \check{I}_1(s-) d\check{\Pi}_{1j}(s) + \int_0^t \check{I}_j(s-) d\check{\Pi}_{j1}(s) \right). \end{aligned}$$

Recalling that the compensators of  $\check{\Pi}_{10}$ ,  $\check{\Pi}_{1j}$ ,  $\check{\Pi}_{j1}$  are  $qnt$ ,  $pt$  and  $pt$ , respectively, and using exchangeability,

$$\begin{aligned} \mathbb{E}\{\hat{M}_1(t)\} &= \mathbb{E} \int_0^t \check{I}_1(s)q ds + n^{-1} \sum_{j \in [n], j \neq 1} \mathbb{E} \int_0^t (\check{I}_1(s) + \check{I}_j(s))p ds \\ &= (q + 2pn^{-1}(n-1))\mathbb{E} \int_0^t \check{I}_1(s)ds \\ &\leq 2\mathbb{E} \int_0^t \check{I}_1(s)ds. \end{aligned}$$

Noting that  $\int_0^t \check{I}_1(s)ds = n^{-1/2}\hat{L}_1(t)$  and using Doob's inequality gives (18).

Next, to show (19), note that by (15), Lemma 2.1, and then (14),

$$\begin{aligned} \hat{L}_1(t) &= \Gamma_2(\hat{X}_{01} + \hat{Y}_1 - p\langle \hat{L} \rangle)(t) \\ &= \sup_{s \leq t} [(\hat{X}_{01} + \hat{Y}_1(s) - p\langle \hat{L} \rangle(s))^-] \\ &\leq \sup_{s \leq t} [(\hat{Y}_1(s) - p\langle \hat{L} \rangle(s))^-] \\ &\leq \|\hat{W}_1\|_t + |\hat{\gamma}_n|t + \|\hat{M}_1\|_t + p\langle \hat{L} \rangle(t), \end{aligned}$$

where  $\|\langle \hat{L} \rangle\|_t = \langle \hat{L} \rangle(t)$  because  $\hat{L}_i$  are nondecreasing. Taking expectation, using exchangeability and the fact that  $p < 1$  gives

$$(20) \quad \mathbb{E}\hat{L}_1(t) \leq c + c\mathbb{E}\|\hat{W}_1\|_t + c\mathbb{E}\|\hat{M}_1\|_t.$$

By (9),

$$\hat{W}_1 = \hat{\Pi}_{01} - \hat{\Pi}_{10} + \sum_{j \in [n], j \neq 1} (\hat{\Pi}_{j1} - \hat{\Pi}_{1j}).$$

In particular,  $\hat{W}_1$  is a martingale and

$$\{\hat{W}_1\}(t) = n^{-1} \left( \check{\Pi}_{01}(t) + \check{\Pi}_{10}(t) + \sum_{j \in [n], j \neq 1} (\check{\Pi}_{j1}(t) + \check{\Pi}_{1j}(t)) \right).$$

The expression in parentheses is a Poisson RV of parameter  $(n\gamma_n + nq + 2(n-1)p)t$ , implying that  $\sup_n \mathbb{E}\{\|\hat{W}_1\}(t)\} < \infty$ . As a result,  $\sup_n \mathbb{E}\|\hat{W}_1\|_t < \infty$ . Using this in (20) gives (19).  $\square$

**3.2. Proof of Lemma 3.2 (Decoupling the routing noise).** The proof is based on a construction stated in Lemma 3.3 below. To present this lemma, let, for each  $n \in \mathbb{N}$ ,  $N_{ij} = N_{ij}^{(n)}$ ,  $(i, j) \in [n]^2$ , be an array of i.i.d. rate- $\beta$  Poisson processes, for some constant  $\beta > 0$ . Let

$$U_i = \sum_j N_{ij}, \quad i \in [n], \quad V_j = \sum_i N_{ij}, \quad j \in [n].$$

Then  $(U_i)_{i \in [n]}$  are independent rate- $\beta n$  Poisson processes, and so are  $V_j$ ,  $j \in [n]$ . The families  $(U_i)_{i \in [n]}$  and  $(V_j)_{j \in [n]}$  are each composed of independent Poisson processes, but they are not independent of one another since they are built from the same Poisson array. The purpose of the lemma is to construct a new column-sum process  $Z$ , independent of the row sums  $U$ , which has the same law as  $V$  and remains asymptotically close to  $V$  at the scale needed for the diffusion approximation.

LEMMA 3.3. *After augmenting the probability space to support additional Poisson processes, if necessary, there exist, for each  $n \in \mathbb{N}$ , independent rate- $\beta n$  Poisson processes  $Z_j^{(n)}$ ,  $j \in [n]$  such that  $Z^{(n)} = \{Z_j^{(n)}\}$  is independent of  $U^{(n)} = \{U_i^{(n)}\}$ . Moreover, the processes  $(V_j^{(n)}, Z_j^{(n)})$ ,  $j \in [n]$ , are exchangeable, and for any  $t$  and any  $\kappa > 0$ ,*

$$n^{-\kappa} \mathbb{E}[\|Z_1^{(n)} - V_1^{(n)}\|_t] \rightarrow 0.$$

**Proof.** Let  $S = \sum_{ij} N_{ij}$ . It is a rate- $\beta n^2$  Poisson process. Let  $\sigma_k$ ,  $k \geq 1$  denote the jump times of  $S$  in increasing order. Let  $(R_k, C_k)$  be the row and column labels associated with the  $k$ -th arrival. Namely,  $(R_k, C_k)$  is the unique  $(i, j)$  for which  $N_{ij}$  makes a jump at time  $\sigma_k$ . Clearly, both  $\{R_k\}$  and  $\{C_k\}$  are i.i.d. uniform on  $[n]$ . Moreover,  $\{R_k\}$ ,  $\{C_k\}$  and  $S$  are mutually independent.

The processes  $N_{ij}$  can be recovered from  $S$  and the labels via

$$N_{ij}(t) = \sum_{k: \sigma_k \leq t} 1_{\{(R_k, C_k) = (i, j)\}}.$$

Consequently,  $U$  and  $V$  can be written as

$$U_i(t) = \sum_{k: \sigma_k \leq t} 1_{\{R_k = i\}}, \quad V_j(t) = \sum_{k: \sigma_k \leq t} 1_{\{C_k = j\}}.$$

To construct  $Z = Z^{(n)}$ , let  $T = T^{(n)}$  be another rate- $\beta n^2$  Poisson process, independent of  $\{N_{ij}\}$ . Let  $\tau_k$ ,  $k \geq 1$  be its jump times in increasing order. Let

$$Z_j(t) = \sum_{k: \tau_k \leq t} 1_{\{C_k = j\}}.$$

Then obviously,  $Z$  has same law as  $V$ . Moreover, because  $Z$  is constructed in terms of  $T$  and  $\{C_k\}$  alone, while  $U$ , as shown above, can be expressed in terms of  $S$  and  $\{R_k\}$  alone, the independence of  $(T, \{C_k\})$  and  $(S, \{R_k\})$  implies that of  $U$  and  $Z$ . Note that the exchangeability of the  $n$  pairs  $(V_j, Z_j)$ ,  $j \in [n]$ , is clear from the construction.

It remains to show the estimate on  $Z_1 - V_1$ . Let  $\pi(m) = \sum_{k=1}^m 1_{\{C_k=1\}}$ . Then

$$V_1(t) = \pi(S(t)), \quad Z_1(t) = \pi(T(t)).$$

Thus

$$\|Z_1 - V_1\|_t = \sup_{s \leq t} |\pi(S(s)) - \pi(T(s))|.$$

Let  $\Theta = \Theta(n, t) = \|T - S\|_t$ . Fix  $0 < \alpha < \kappa \wedge 2$ . For a bound on  $\Theta$ , note that  $T - S$  is a martingale with jumps of size one and quadratic variation  $T + S$ . By a standard exponential martingale inequality for compensated Poisson processes, there are constants  $c_1, c_2 > 0$ , depending on  $t$  and  $\beta$ , such that

$$\mathbb{P}(\Theta > n^{1+\alpha/2}) \leq c_1 e^{-c_2 n^\alpha}.$$

A Chernoff bound for  $S(t)$ , equivalently  $T(t)$ , gives  $\mathbb{P}(S(t) > 2\beta t n^2) \leq c_1 e^{-c_2 n^2}$ . Hence, the event

$$A_n = \left\{ \max(S(t), T(t)) \leq 2\beta t n^2, \Theta \leq n^{1+\alpha/2} \right\}$$

satisfies

$$\mathbb{P}(A_n^c) \leq c_1 e^{-c_2 n^\alpha}.$$

Let  $l_n = \lceil n^{1+\alpha/2} \rceil$ . On  $A_n$ ,

$$\|Z_1 - V_1\|_t \leq H(n, t) := \sup\{\pi(r + l_n) - \pi(r) : 0 \leq r \leq 2\beta t n^2\}.$$

For fixed  $r$ ,  $\pi(r + l_n) - \pi(r)$  is binomial with parameters  $(l_n, n^{-1})$ . Hence, by a Chernoff bound,

$$\mathbb{P}(\pi(r + l_n) - \pi(r) > n^\alpha) \leq e^{-n^\alpha + en^{\alpha/2}} \leq e^{-\frac{1}{2}n^\alpha}$$

for all sufficiently large  $n$ . Since there are at most  $cn^2$  possible values of  $r$ , it follows that

$$\mathbb{P}(H(n, t) > n^\alpha) \leq cn^2 e^{-\frac{1}{2}n^\alpha}.$$

Moreover,  $H(n, t) \leq l_n$ . Therefore

$$\mathbb{E}H(n, t) \leq n^\alpha + l_n \mathbb{P}(H(n, t) > n^\alpha) \leq n^\alpha + o(1).$$

Finally,

$$\|Z_1 - V_1\|_t \leq H(n, t) + (V_1(t) + Z_1(t))1_{A_n^c}.$$

Since  $V_1(t)$  and  $Z_1(t)$  are rate- $\beta n$  Poisson random variables,

$$\mathbb{E}[(V_1(t) + Z_1(t))^2] \leq cn^2.$$

By Cauchy–Schwarz and the exponential bound on  $\mathbb{P}(A_n^c)$ ,

$$\mathbb{E}[(V_1(t) + Z_1(t))1_{A_n^c}] \leq cn \mathbb{P}(A_n^c)^{1/2} = o(1).$$

Consequently,

$$\mathbb{E}\|Z_1 - V_1\|_t \leq cn^\alpha + o(1).$$

Since  $\alpha < \kappa$ , this gives

$$\mathbb{E}[n^{-\kappa}\|Z_1 - V_1\|_t] \rightarrow 0,$$

as required. □

Based on this we can present the following.

**Proof of Lemma 3.2.** Consider the processes  $\check{\Pi}_{ij}$ ,  $i, j \in [n]$ , an  $n \times n$  array of i.i.d. rate- $p$  Poisson processes. Apply Lemma 3.3 to this array with  $\beta = p$ , denoting the corresponding tuple  $(Z_i)_{i \in [n]}$  by  $(\Pi_i^*)_{i \in [n]}$ . Let

$$\dot{\Pi}_i(t) = n^{-1/2}(\Pi_i^*(t) - pnt), \quad i \in [n].$$

Then the components are i.i.d., and this tuple is independent of  $(\hat{\Pi}_{0i}, \hat{\Pi}_{i0}, \hat{\Pi}_i^{\text{rs}})_{i \in [n]}$ . Moreover, recalling the definition of  $\check{\Pi}_{ij}$  and denoting by  $\check{\Pi}_i^{\text{cs}}$  the column-sum  $\sum_{j \in [n]} \check{\Pi}_{ji}$ , we have  $\hat{\Pi}_i^{\text{cs}}(t) = \check{\Pi}_i^{\text{cs}}(t) - pnt$ . Thus

$$\dot{\Pi}_1 - \hat{\Pi}_1^{\text{cs}} = n^{-1/2}(\Pi_1^* - \check{\Pi}_1^{\text{cs}}).$$

Taking now  $\kappa = 1/2$  in Lemma 3.3 shows  $\mathbb{E}[\|\dot{\Pi}_1 - \hat{\Pi}_1^{\text{cs}}\|_t] \rightarrow 0$ . □

3.3. *Proof of Theorem 1.1.* Going back to equations (14)-(15), our first goal is to couple  $\{\hat{Y}_i^n\}_{i \in [n]}$ , with i.i.d.  $(\hat{\gamma}, 2)$ -BMs  $(\tilde{Y}_i)_{i \in \mathbb{N}}$ , in such a way that, for each  $n$ ,  $\hat{Y}_i - \tilde{Y}_i$  are exchangeable and  $\hat{Y}_1^n \rightarrow \tilde{Y}_1$  in probability as  $n \rightarrow \infty$ . To this end, rewrite  $(\hat{Y}_i)_{i \in [n]}$ , as

$$(21) \quad \begin{aligned} \hat{Y}_i &= \dot{W}_i + \hat{\gamma}_n \iota + \hat{E}_i, \\ \dot{W}_i &:= \hat{\Pi}_{0i} - \hat{\Pi}_{i0} - \hat{\Pi}_i^{\text{rs}} + \dot{\Pi}_i, \\ \hat{E}_i &:= \hat{\Pi}_i^{\text{cs}} - \dot{\Pi}_i + \hat{M}_i. \end{aligned}$$

By Lemma 3.2, the replacement of  $\hat{\Pi}_i^{\text{cs}}$  by  $\dot{\Pi}_i$  makes the processes  $\dot{W}^i$  i.i.d. martingales, while the error is absorbed into  $\hat{E}_i$ . The quadratic variation of  $\dot{W}_1$  is

$$[\dot{W}_1] = n^{-1}(\check{\Pi}_{01} + \check{\Pi}_{10} + \check{\Pi}_1^{\text{rs}} + \Pi_1^*)$$

where  $\check{\Pi}_i^{\text{rs}} = \sum_j \check{\Pi}_{ij}$ . The expression in parentheses on the right of the above display is a Poisson process of rate  $\gamma_n n + qn + 2pn$ . Since  $\gamma_n \rightarrow q$ , we have that  $[\dot{W}_1] \rightarrow 2\iota$  in probability, and thus  $\dot{W}_1$  converges in distribution to a  $(0, 2)$ -BM. Moreover, by Lemmas 3.1 and 3.2, the processes  $\hat{E}_i$  are exchangeable and, in view of Lemmas 3.1 and 3.2,

$$(22) \quad \mathbb{E}[\|\hat{E}_1\|_t] \rightarrow 0.$$

We now invoke Skorohod's representation to realize the prelimit and limit processes on a common space. With a slight abuse of notation, we continue to denote the prelimit processes by the same symbols as before. The construction uses the one-dimensional convergence  $\dot{W}_1^n \Rightarrow \dot{W}_1$  together with the estimate  $\hat{E}_1^n \rightarrow 0$  to obtain a coupling of the full triangular array  $(\dot{W}_i^n, \tilde{W}_i, \hat{E}_i^n)_{i \in [n]}$ , in such a way that the joint law of  $(\dot{W}_i^n, \hat{E}_i^n)_{i \in [n]}$  is preserved.

In particular, since  $D^{(1)}$  is a Polish space,  $\dot{W}_1$  can be coupled with a  $(0, 2)$ -BM  $\tilde{W}_1$  in such a way that  $\dot{W}_1 \rightarrow \tilde{W}_1$  in probability. Let  $K_n(w, dy)$  be a transition kernel on  $(D^{(1)}, \mathcal{D}^{(1)})$  such that  $K_n(w, \cdot)$  is a version of the regular conditional law  $\mathcal{L}(\dot{W}_1 | \tilde{W}_1 = w)$  for  $\mathcal{L}(\tilde{W}_1)$ -a.e.  $w$  [4, Section 10.4]. Let  $G_n(\bar{w}, d\bar{e})$  be a transition kernel on  $(D^{(n)}, \mathcal{D}^{(n)})$  such that  $G_n(\bar{w}, \cdot)$  is a version of the regular conditional law  $\mathcal{L}(\{\hat{E}_i\}_{i \in [n]} | \{\dot{W}_i\}_{i \in [n]} = \bar{w})$  for  $\mathcal{L}(\{\dot{W}_i\})$ -a.e.  $\bar{w}$ .

Now construct on one probability space the following random elements. Let  $(\tilde{W}_i)_{i \in \mathbb{N}}$  be i.i.d.  $(0, 2)$ -BMs. Construct  $\{\dot{W}_i^n\}_{i \in [n]}$  so that, conditionally on  $\{\tilde{W}_j\}_{j \in \mathbb{N}}$ , they are independent with

$$\mathcal{L}(\dot{W}_i^n | \{\tilde{W}_j\}) = \mathcal{L}(\dot{W}_i^n | \tilde{W}_i) = K_n(\tilde{W}_i, \cdot), \quad i \in [n].$$

This makes  $\{\dot{W}_i^n\}_{i \in [n]}$  independent and preserves their original law. Since  $\dot{W}_1^n \rightarrow \tilde{W}_1$  in probability under this coupling, and since  $\sup_n \mathbb{E}\|\dot{W}_1^n\|_t^2 < \infty$  and  $\mathbb{E}\|\tilde{W}_1\|_t^2 < \infty$ , the family  $\{\|\dot{W}_1^n - \tilde{W}_1\|_t\}_n$  is uniformly integrable. Hence

$$(23) \quad \mathbb{E}\|\dot{W}_1^n - \tilde{W}_1\|_t \rightarrow 0.$$

Next, construct  $\{\hat{E}_i^n\}_{i \in [n]}$  so that their conditional law given  $\{\dot{W}_i^n\}_{i \in [n]}$  is given by  $G_n(\{\dot{W}_i^n\}_{i \in [n]}, \cdot)$ . This preserves the joint law of  $\{\dot{W}_i^n, \hat{E}_i^n\}_{i \in [n]}$ . Finally, construct  $\{\hat{X}_{i0}^n\}_{i \in [n]}$  according to their original law, independently of all  $\tilde{W}_i, \dot{W}_i^n$  and  $\hat{E}_i^n$ .

In agreement with (21), on the new probability space, set  $\hat{Y}_i^n = \dot{W}_i^n + \hat{\gamma}_n \iota + \hat{E}_i^n$ . Recalling relation (15) and the uniqueness statement of Lemma 2.2, apply that lemma with  $\beta = -p$  and  $w_i = \hat{X}_{i0}^n + \hat{Y}_i^n$  to define  $(\hat{X}_i^n, \hat{L}_i^n)$  as the unique solution to the system of equations

$$(\hat{X}_i^n, \hat{L}_i^n) = \Gamma(\hat{X}_{i0}^n + \hat{Y}_i^n - p\langle \hat{L}_i^n \rangle), \quad i \in [n].$$

This preserves the original law of  $(\hat{X}^n, \hat{L}^n)$ .

The next step is to construct the Brownian system and estimate the distance between the two systems. Let

$$\tilde{Y}_i = \tilde{W}_i + \hat{\gamma}_i, \quad i \in [n],$$

and let  $(\tilde{X}^n, \tilde{L}^n)$  be the unique (again by Lemma 2.2) tuple satisfying

$$(\tilde{X}_i^n, \tilde{L}_i^n) = \Gamma(\hat{X}_{0i}^n + \tilde{Y}_i - p\langle \tilde{L}^n \rangle), \quad i \in [n].$$

We now use the 1-Lipschitz property of  $\Gamma_2$  stated in (13). Namely, for any  $t > 0$ ,

$$\begin{aligned} \|\hat{L}_i^n - \tilde{L}_i^n\|_t &\leq \|(\hat{X}_{0i}^n + \hat{Y}_i^n - p\langle \hat{L}^n \rangle) - (\hat{X}_{0i}^n + \tilde{Y}_i - p\langle \tilde{L}^n \rangle)\|_t \\ (24) \quad &\leq \|\hat{W}_i^n - \tilde{W}_i\|_t + |\hat{\gamma}_i - \tilde{\gamma}_i|_t + \|\hat{E}_i^n\|_t + p\langle \|\hat{L}^n - \tilde{L}^n\|_t \rangle. \end{aligned}$$

Averaging,

$$\langle \|\hat{L}^n - \tilde{L}^n\|_t \rangle \leq q^{-1}(\langle \|\hat{W}^n - \tilde{W}\|_t \rangle + |\hat{\gamma}_n - \tilde{\gamma}_n|_t + \langle \|\hat{E}^n\|_t \rangle).$$

Substituting back in (24),

$$\begin{aligned} \|\hat{L}_i^n - \tilde{L}_i^n\|_t &\leq \|\hat{W}_i^n - \tilde{W}_i\|_t + (1 + pq^{-1})|\hat{\gamma}_i - \tilde{\gamma}_i|_t + \|\hat{E}_i^n\|_t \\ &\quad + pq^{-1}(\langle \|\hat{W}^n - \tilde{W}\|_t \rangle + \langle \|\hat{E}^n\|_t \rangle). \end{aligned}$$

By (22) and (23), the expression on the right converges to zero in  $L^1$ . Hence we obtain  $\mathbb{E}[\|\hat{L}_1^n - \tilde{L}_1^n\|_t] \rightarrow 0$ . Using

$$\hat{X}_i^n - \tilde{X}_i^n = (\hat{X}_{0i}^n + \hat{Y}_i^n - p\langle \hat{L}^n \rangle + \hat{L}_i^n) - (\hat{X}_{0i}^n + \tilde{Y}_i - p\langle \tilde{L}^n \rangle + \tilde{L}_i^n),$$

and once again (24), shows that  $\mathbb{E}[\|\hat{X}_1^n - \tilde{X}_1^n\|_t] \rightarrow 0$  as well.

The convergence of a  $k$ -tuple to  $k$  independent copies of  $(\bar{X}, \bar{L})$  now follows from that of the Brownian system, stated in Theorem 2.3.

The convergence of the empirical measures also follows now from Theorem 2.3, but some further details are required. Let  $\hat{Z}_i^n = (\hat{X}_i^n, \hat{L}_i^n)$  and  $\tilde{Z}_i^n = (\tilde{X}_i^n, \tilde{L}_i^n)$ . Recall that we have  $\tilde{\mu}_t^n \rightarrow \mu_t$  in  $C(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^2))$  in probability. The pairs  $(\hat{Z}_i^n, \tilde{Z}_i^n)_{i \in [n]}$  are exchangeable and one has  $\|\hat{Z}_1^n - \tilde{Z}_1^n\|_t \rightarrow 0$  in probability. Fix  $t > 0$  and  $\varepsilon > 0$ . By exchangeability,

$$\frac{1}{n} \mathbb{E} \sum_{i \in [n]} \mathbf{1}_{\{\|\hat{Z}_i^n - \tilde{Z}_i^n\|_t > \varepsilon\}} = \mathbb{P}(\|\hat{Z}_1^n - \tilde{Z}_1^n\|_t > \varepsilon) \rightarrow 0.$$

Hence  $\frac{1}{n} \sum_{i \in [n]} \mathbf{1}_{\{\|\hat{Z}_i^n - \tilde{Z}_i^n\|_t > \varepsilon\}} \rightarrow 0$  in probability. Now take a 1-Lipschitz, 1-bounded function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Then for  $s \leq t$ ,

$$\left| \int f d\mu_s^n - \int f d\tilde{\mu}_s^n \right| \leq \frac{1}{n} \sum_{i \in [n]} |f(\hat{Z}_i^n(s)) - f(\tilde{Z}_i^n(s))| \leq \varepsilon + \frac{2}{n} \sum_{i \in [n]} \mathbf{1}_{\{\|\hat{Z}_i^n - \tilde{Z}_i^n\|_t > \varepsilon\}}.$$

Noting that the expression on the right does not depend on  $s$ , and taking  $n \rightarrow \infty$  followed by  $\varepsilon \rightarrow 0$ , shows that the bounded-Lipschitz distance satisfies

$$\sup_{s \leq t} d_{\text{BL}}(\mu_s^n, \tilde{\mu}_s^n) \rightarrow 0 \text{ in probability.}$$

The result now follows by the triangle inequality.

As for the convergence of  $\lambda^n$ , we have

$$\|\lambda^n - \tilde{\lambda}^n\|_t = \|\langle \hat{L}^n - \tilde{L}^n \rangle\|_t \leq \langle \|\hat{L}^n - \tilde{L}^n\|_t \rangle \rightarrow 0$$

in probability, hence again the result follows from Theorem 2.3.  $\square$

*Acknowledgment.* Research was supported by ISF grant 3240/25.

## REFERENCES

- [1] ATAR, R. (2026). High-dimensional limits for reflected Brownian motion in the orthant. *arXiv preprint arXiv:2605.01958*.
- [2] BACCELLI, F., KARPELEVICH, F. I., KELBERT, M. Y., PUHALSKII, A. A., RYBKO, A. and SUHOV, Y. M. (1992). A mean-field limit for a class of queueing networks. *Journal of statistical physics* **66** 803–825.
- [3] BAKER, G., HAMBLY, B. and JETTKANT, P. (2025). Particle Systems and McKean–Vlasov Dynamics with Singular Interaction through Local Times. *arXiv preprint arXiv:2503.08837*.
- [4] BOGACHEV, V. I. (2007). *Measure Theory I-II*. Springer.
- [5] CHEN, H. and YAO, D. D. (2001). *Fundamentals of Queueing Networks. Applications of Mathematics (New York)* **46**. Springer-Verlag, New York.
- [6] COGHI, M., DREYER, W., FRIZ, P. K., GAJEWSKI, P., GUHLKE, C. and MAURELLI, M. (2022). A McKean–Vlasov SDE and Particle System with Interaction from Reflecting Boundaries. *SIAM Journal on Mathematical Analysis* **54** 2251–2294.
- [7] GÄRTNER, J. (1988). On the McKean–Vlasov limit for interacting diffusions. *Mathematische Nachrichten* **137** 197–248.
- [8] HARRISON, J. M. and REIMAN, M. I. (1981). Reflected Brownian motion on an orthant. *Ann. Probab.* **9** 302–308.
- [9] MCKEAN JR, H. P. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences* **56** 1907–1911.
- [10] REIMAN, M. I. (1984). Open queueing networks in heavy traffic. *Mathematics of Operations Research* **9** 441–458.
- [11] RYBKO, A. and SHLOSMAN, S. (2016). Stationary States of the Generalized Jackson Networks. *Markov Processes and Related Fields* **22** 759–774.
- [12] RYBKO, A. N. and SHLOSMAN, S. B. (2005). Poisson hypothesis for information networks. II. *Moscow mathematical journal* **5** 927–959.
- [13] SZNITMAN, A.-S. (1984). Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *Journal of Functional Analysis* **56** 311–336.
- [14] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *Ecole d’été de probabilités de Saint-Flour XIX—1989* 165–251. Springer.