

# SPDE LIMITS OF MANY-SERVER QUEUES

HAYA KASPI AND KAVITA RAMANAN

ABSTRACT. A many-server queueing system is considered in which customers with independent and identically distributed service times enter service in the order of arrival. The state of the system is represented by a process that describes the total number of customers in the system, as well as a measure-valued process that keeps track of the ages of customers in service, leading to a Markovian description of the dynamics. Under suitable assumptions, a functional central limit theorem is established for the sequence of (centered and scaled) state processes as the number of servers goes to infinity. The limit process describing the total number in system is shown to be an Itô diffusion with a constant diffusion coefficient that is insensitive to the service distribution. The limit of the sequence of (centered and scaled) age processes is shown to be a Hilbert space valued diffusion that can also be characterized as the unique solution of a stochastic partial differential equation that is coupled with the Itô diffusion. Furthermore, the limit processes are shown to be semimartingales and to possess a strong Markov property.

---

## CONTENTS

1. Introduction	1
2. Description of the Model	6
3. Fluid Limit	9
4. Certain Martingale Measures and their Stochastic Integrals	11
5. Main Results	15
6. Representation of the System Dynamics	22
7. Continuity Properties	25
8. Convergence Results	30
9. Proofs of Main Theorems	40
Appendix A. Properties of the Martingale Measure Sequence	49
Appendix B. Ramifications of Assumptions on the Service Distribution	52
Appendix C. Proof of the Representation Formula	53
Appendix D. Some Moment Estimates	53
Appendix E. Proof of Consistency	55
References	58

---

## 1. INTRODUCTION

**1.1. Background, Motivation and Results.** Many-server queues constitute a fundamental model in queueing theory and are typically harder to analyze than single-server queues. The main objective of this paper is to establish useful functional central limit theorems for many-server queues in the asymptotic regime in which the number  $N$  of servers tends to infinity and the mean arrival rate scales as  $\lambda^{(N)} = \bar{\lambda}N - \beta\sqrt{N}$  for some  $\bar{\lambda} > 0$  and  $\beta \in (-\infty, \infty)$ . For many-server queues with Poisson arrivals, this scaling was considered more than half a century ago by Erlang [7] and

*Date:* September 29, 2010.

*2000 Mathematics Subject Classification.* Primary: 60F17, 60H15; 60K25; Secondary: 90B22; 68M20.

*Key words and phrases.* Many-server queues, GI/G/N queue, fluid limits, diffusion limits, measure-valued processes, Itô diffusions, stochastic partial differential equations.

Both authors were partially supported by the US-Israel Binational Science Foundation under Grant BSF-2006379. The first author was partially supported by the Milford Bohm Chair Grant. The second author was partially supported by the National Science Foundation under Grants CMMI-0728064 (CMMI-1059967) and CMMI-0928154 (CMMI-1052750).

thereafter by Jagerman [18] for a loss system with exponential service times, but it was not until the influential work of Halfin and Whitt [15] that a general heavy traffic limit theorem was established for queues with renewal arrivals, exponential service times, normalized to have unit mean, and  $\bar{\lambda} = 1$ . As a result, this asymptotic regime is often referred to as the Halfin-Whitt regime. In contrast to conventional heavy traffic scalings, in the Halfin-Whitt regime the limiting stationary probability of a positive wait is non-trivial (i.e., it lies strictly between zero and one), which better models the behavior of many systems found in applications. Halfin and Whitt [15] showed that the limit of the sequence of processes representing the (appropriately centered and scaled) number of customers in the system is a diffusion process that behaves like an Ornstein-Uhlenbeck process below zero and like a Brownian motion with drift above zero. When  $\beta > 0$ , which ensures that each of the  $N$ -server queues is stable, this characterization of the limit process was used to establish approximations to the stationary probability of positive wait in a queue with  $N$  servers. For exponential service distributions, the work of Halfin and Whitt was subsequently generalized by Mandelbaum, Massey and Reiman [24] to the network setting and the case of inhomogeneous Poisson arrivals.

However, in many applications, statistical evidence suggests that it may be more appropriate to model the service times as being non-exponential (see, for example, the study of real call center data in Brown et al. [6] that suggests that the service times are lognormally distributed). A natural goal is then to understand the behavior of many-server queues in this scaling regime when the service distribution is not exponentially distributed. Specifically, in addition to establishing a limit theorem, the aim is to obtain a tractable representation of the limit process that makes it amenable to computation, so that the limit could be used to shed insight into performance measures of interest for an  $N$ -server queue.

In this work, we represent the state of the  $N$ -server queue by a nonnegative, integer-valued process  $X^{(N)}$  that records the total number of customers in system, as well as a measure-valued process  $\nu^{(N)}$  that keeps track of the ages of customers in service. This representation was first introduced by Kaspi and Ramanan in [22], where it was used to identify the functional strong law of large numbers limits or equivalently, fluid limits for these queues and was subsequently shown to provide a Markovian description of the dynamics (see Kang and Ramanan [20]). Under suitable assumptions, in each of the cases when the fluid limit is subcritical, critical or supercritical (which, roughly speaking, corresponds to the cases  $\bar{\lambda} < 1$ ,  $\bar{\lambda} = 1$  and  $\bar{\lambda} > 1$ ), we show (in Theorems 5.6 and 5.7) that the diffusion-scaled state sequence,  $\{(\hat{X}^{(N)}, \hat{\nu}^{(N)})\}_{N \in \mathbb{N}}$  obtained by centering the state around the fluid limit and multiplying the centered state by  $\sqrt{N}$ , converges weakly to a limit process  $(\hat{X}, \hat{\nu})$ . Moreover, the component  $\hat{X}$  is characterized as a real-valued càdlàg process that is the solution to an Itô diffusion with a constant diffusion coefficient that is *insensitive* to the service distribution, and whose drift is an adapted process that is a functional of  $\hat{\nu}_t$  (see Corollary 5.13). As for the age process, although the  $\hat{\nu}^{(N)}$  are (signed) Radon measure valued processes, the limit  $\hat{\nu}$  lies outside this space. A key challenge was to identify a suitable space in which to establish convergence without imposing restrictive assumptions on the service distribution  $G$ . Under conditions that include a large class of service distributions relevant in applications such as phase-type, Weibull, lognormal, logistic and (for a large class of parameters) Erlang and Pareto distributions, we show that the convergence of  $\hat{\nu}^{(N)}$  to  $\hat{\nu}$  holds in the space of  $\mathbb{H}_{-2}$ -valued càdlàg processes, where  $\mathbb{H}_{-2}$  is the dual of the Hilbert space  $\mathbb{H}_2$ . In particular, this immediately implies convergence of a large class of functionals of the many-server queue. In addition, we show that both processes are semimartingales with an explicit decomposition (see Theorem 5.8) and we characterize  $\hat{\nu}$  as the unique solution to a stochastic partial differential equation that is coupled with the Itô diffusion  $\hat{X}$  (see Theorem 5.11(a)). Furthermore (in Theorem 5.11(b)), we also show that the pair, along with an appended state, forms a strong Markov process.

**1.2. Relation to Prior Work.** To date, the most general results on process level convergence in the Halfin-Whitt regime were obtained in a nice pair of papers by Reed [30] and Puhalskii and

Reed [29]. Under the assumptions that  $\bar{\lambda} = 1$ , the residual service times of customers in service at time 0 are independent and identically distributed (i.i.d.) and taken from the equilibrium fluid distribution, and the total (fluid scaled) number in system converges to 1, a heavy traffic limit theorem for the sequence of processes  $\{\widehat{X}^{(N)}\}_{N \in \mathbb{N}}$  was established by Reed [30] with only a finite mean condition on the service distribution. This result was extended by Puhalskii and Reed [29] to allow for more general, possibly inhomogeneous arrival processes and residual service times of customers in service at time zero that, while still i.i.d. could be chosen from an arbitrary distribution. In this setting, convergence of finite-dimensional distributions was established in [29], and strengthened to process level convergence established when the service distribution is continuous. The general approach used in both these papers is to represent the many-server queue as a perturbation of an infinite-server queue and to establish tightness and convergence using a continuous mapping representation and estimates analogous to those obtained by Krichagina and Puhalskii [23] for the infinite-server queue under similar assumptions on the initial conditions. In both papers, the limit is characterized as the unique solution to a certain implicit stochastic convolution equation. Several previous works had also extended the Halfin-Whitt process level result for specific classes of service distributions. Noteworthy amongst them is the paper by Puhalskii and Reiman [28], which considered phase-type service distributions and characterized the heavy traffic limit theorem as a multidimensional diffusion, where each dimension corresponds to a different phase of the service distribution. Whitt [37] also established a process level result for a many-server queue with finite waiting room and a service distribution that is a mixture of an exponential random variable and a point mass at zero. Moreover, for service distributions with finite support, Mandelbaum and Momčilović [25] used a combination of combinatorial and probabilistic methods to study the limit of the virtual waiting time process. In addition to the process level results described above, interesting results on the asymptotics of steady state distributions in the Halfin-Whitt regime have been obtained by Jelenkovic, Mandelbaum and Momčilović [19] for deterministic service times and by Gamarnik and Momčilović [13] for service times that are lattice-valued with finite support.

Our work serves to complement the above mentioned results, with the focus being on establishing tractability of the limit process under reasonably general assumptions on the service distribution that includes a large class of service distributions of interest. Whereas in all the above papers only the number in system is considered, we establish convergence for a more general state process, which implies the convergence of a large class of functionals of the process and not just the number in system. In addition, our approach leads to a new characterization for the limiting number in system  $\widehat{X}$  as an Itô diffusion, which relies on an asymptotic independence result for the centered arrival and departure processes (see Proposition 8.4) that may be of independent interest. We also establish an insensitivity result showing that the diffusion coefficient depends only on the mean and variance of the interarrival times and is independent of the service distribution. As a special case, we can recover the results of Halfin and Whitt [15] and Puhalskii and Reiman [28] and (for the smaller class of service distributions that we consider) Reed [30]. Moreover, we allow in a sense more general initial conditions than those considered in Reed [30] and Puhalskii and Reed [29], both of which assume that the residual times of customers in service at the initial time are i.i.d. This property is not typically preserved at positive times. In contrast, we establish a consistency property (see Lemma 9.6) that shows that the assumptions we impose at the initial time are also satisfied at any positive time and our assumptions are trivially satisfied by a system that starts with zero initial conditions (i.e., at the fluid initial condition). As shown in Theorem 3.7 and Section 6 of Kaspi and Ramanan [22], starting from an empty system, the fluid limit does not reach the fluid equilibrium state in finite time. Therefore, the consideration of general initial conditions is useful for both capturing the transient behavior of the system as well as for establishing the (strong) Markov property for the limit process. The latter can be potentially useful as this enables the application of a wide array of tools available for Markov processes in general state spaces.

The Markovian representation of the state, though infinite-dimensional, leads to an intuitive characterization of the dynamics, which allows the framework to be extended to incorporate more

general features into the model (see, for example, the extension of this framework to include abandonments by Kang and Ramanan in [20] and [21]). In the subcritical case our results provide a characterization of the diffusion limit of the well studied infinite-server queue, which is easier to analyze due to the absence of a queue and, hence, of an interaction between those in service and those waiting in queue. A few representative works on diffusion limits of the number in system in the infinite-server queue include Iglehart [16], Borovkov [5], Whitt [36] and Glynn and Whitt [14], where the limit process is characterized as an Ornstein-Uhlenbeck process, and Krichagina and Puhalskii [23], who provided an alternative representation of the limit in terms of the so-called Kiefer process. More recently, a functional central limit theorem in the space of distribution-valued processes was established for the  $M/G/\infty$  queue by Decreusefond and Moyal [9]. In contrast to the infinite-dimensional Markovian representation in terms of residual service times used in Decreusefond and Moyal [9], the Markovian representation in terms of the age process that we use allows us to associate some natural martingales that facilitate the analysis. This perspective may be useful in the analysis of other queueing networks as well and has, for example, been recently adopted by Reed and Talreja [31] in their extension of the work of Decreusefond and Moyal [9] to establish infinite-dimensional functional central limit theorems for the  $GI/G/\infty$  queue. The work [31] adopts a semi-group approach that seems to require much stronger assumptions on the service distribution (namely that the hazard rate function  $h$  of the service distribution is infinitely differentiable and  $h$  and its derivatives are all uniformly bounded) than is imposed in our paper.

**1.3. Outline of the Paper.** Section 2 contains a precise mathematical description of the model and the state descriptor used, as well as the defining dynamical equations. A deterministic analog of the model, described by dynamical equations that are referred to as the fluid equations, is introduced in Section 3. Section 3 also recapitulates the result of Kaspi and Ramanan [22] that shows that (under fairly general conditions stated as Assumptions 1 and 2) the functional strong law of large numbers limit of the normalized (divided by  $N$ ) state of the  $N$ -server system is the unique solution to the fluid equations. In Section 4 a sequence of martingales obtained as compensated departure processes, which play an important role in the analysis, is introduced and the associated scaled martingale measures  $\widehat{\mathcal{M}}^{(N)}$ ,  $N \in \mathbb{N}$ , are shown to be orthogonal, which allows one to define certain associated stochastic convolution integrals  $\widehat{\mathcal{H}}^{(N)}$ . The main results and their corollaries are stated in Section 5, and their proofs are presented in Section 9. The proofs rely on results obtained in Sections 6, 7 and 8. Section 6 contains a succinct characterization of the dynamics and establishes a representation (see Proposition 6.4) for  $\widehat{\nu}^{(N)}$ , the diffusion-scaled age process in the  $N$ -server system, in terms of certain stochastic convolution integrals  $\widehat{\mathcal{H}}^{(N)}$ ,  $\widehat{\mathcal{K}}^{(N)}$  and the initial data. In Section 7, it is shown that the processes  $\widehat{\mathcal{K}}^{(N)}$ ,  $\widehat{X}^{(N)}$  and  $\widehat{\nu}^{(N)}$  can be obtained as a continuous mapping of the initial data sequence and the process  $\widehat{\mathcal{H}}^{(N)}$ . Section 8 is devoted to establishing convergence of the martingale measure sequence  $\{\widehat{\mathcal{M}}^{(N)}\}_{N \in \mathbb{N}}$  and the associated sequence  $\{\widehat{\mathcal{H}}^{(N)}\}_{N \in \mathbb{N}}$  of stochastic convolution integrals, jointly with the sequence of centered arrival processes and initial conditions (see Corollary 8.7). In particular, the asymptotic independence property is established. Section 8 is the most technically demanding part of the paper. To maintain the flow of the exposition, some supporting results are relegated to the Appendix. Appendix E also contains the proof of a consistency result, which shows that the assumptions on the initial conditions are reasonable. First, in Section 1.4 we introduce some common notation and terminology used in the paper.

**1.4. Notation and Terminology.** The following notation will be used throughout the paper.  $\mathbb{Z}_+$  is the set of non-negative integers,  $\mathbb{N}$  is the set of natural numbers or, equivalently, strictly positive integers,  $\mathbb{R}$  is the set of real numbers and  $\mathbb{R}_+$  the set of non-negative real numbers. For  $a, b \in \mathbb{R}$ ,  $a \vee b$  and  $a \wedge b$  denote, respectively, the maximum and minimum of  $a$  and  $b$ , and the short-hand notation  $a^+$  will also be used for  $a \vee 0$ . Given  $B \subset \mathbb{R}$ ,  $\mathbb{1}_B$  denotes the indicator function of the set  $B$  (that is,  $\mathbb{1}_B(x) = 1$  if  $x \in B$  and  $\mathbb{1}_B(x) = 0$  otherwise).

1.4.1. *Function Spaces.* Given any metric space  $\mathcal{E}$ , we denote by  $\mathcal{B}(\mathcal{E})$  the Borel sets of  $\mathcal{E}$  (with topology compatible with the metric on  $\mathcal{E}$ ), and let  $\mathcal{C}_b(\mathcal{E})$ ,  $\mathcal{AC}_b(\mathcal{E})$  and  $\mathcal{C}_c(\mathcal{E})$ , respectively, denote the space of bounded continuous functions, bounded absolutely continuous functions and the space of continuous functions with compact support defined on  $\mathcal{E}$  and taking values in the reals. We also let  $\mathcal{C}^1(\mathcal{E})$  and  $\mathcal{C}^\infty(\mathcal{E})$ , respectively, represent the space of real-valued, once continuously differentiable and infinitely differentiable functions on  $\mathcal{E}$ ,  $\mathcal{C}_c^1(\mathcal{E})$  the subspace of functions in  $\mathcal{C}^1(\mathcal{E})$  that have compact support and  $\mathcal{C}_b^1(\mathcal{E})$  the subspace of functions in  $\mathcal{C}^1(\mathcal{E})$  that, together with its first derivatives, are bounded. We let  $\mathbb{D}_{\mathcal{E}}[0, \infty)$  denote the space of  $\mathcal{E}$ -valued càdlàg functions defined on  $[0, \infty)$  and let  $\text{supp}(\varphi)$  denote the support of a function  $\varphi$ .

We will mostly be interested in the case when  $\mathcal{E} = [0, L)$  and  $\mathcal{E} = [0, L) \times \mathbb{R}_+$ , for some  $L \in (0, \infty]$ . To distinguish these cases, we will usually use  $f$  to denote generic functions on  $[0, L)$  and  $\varphi$  to denote generic functions on  $[0, L) \times \mathbb{R}_+$ . By some abuse of notation, given  $f$  on  $[0, L)$ , we will sometimes also treat it as a function on  $[0, L) \times \mathbb{R}_+$  that is constant in the second variable. Recall that given  $T < \infty$  and a continuous function  $f \in \mathcal{C}[0, T]$ , the modulus of continuity  $w_f(\cdot)$  of  $f$  is defined by

$$(1.1) \quad w_f(\delta) \doteq \sup_{s, t \in [0, T]: |t-s| < \delta} |f(t) - f(s)|, \quad \delta > 0.$$

When  $\mathcal{E} = [0, L) \times \mathbb{R}_+$ , for some  $L \leq \infty$ , we let  $\mathcal{C}^{1,1}([0, L) \times \mathbb{R}_+)$  denote the space of absolutely continuous functions  $\varphi$  on  $[0, L) \times \mathbb{R}_+$  for which the directional derivative  $\varphi_x + \varphi_s$  in the  $(1, 1)$  direction exists and is continuous and let  $\mathcal{C}_c^{1,1}([0, L) \times \mathbb{R}_+)$  (respectively,  $\mathcal{C}_b^{1,1}([0, L) \times \mathbb{R}_+)$ ) denote the subset of functions  $\varphi$  in  $\mathcal{C}^{1,1}([0, L) \times \mathbb{R}_+)$  such that  $\varphi$ , along with its directional derivative  $\varphi_x + \varphi_s$ , has compact support (respectively, is bounded). We let  $\mathbb{I}_{\mathbb{R}_+}[0, \infty)$  denote the space of non-decreasing functions  $f \in \mathbb{D}_{\mathbb{R}}[0, \infty)$  with  $f(0) = 0$ . For  $L \in [0, \infty]$ ,  $\mathbb{L}^\alpha[0, L)$ ,  $\alpha \geq 1$ , and  $\mathbb{L}^\infty[0, L)$  represent, respectively, the spaces of measurable functions  $f$  such that  $\int_{[0, L)} |f|^\alpha < \infty$  and the space of essentially bounded functions on  $[0, L)$ . Also,  $L_{loc}^i[0, L)$ ,  $i = 1, 2, \infty$ , represents the corresponding space in which the associated property holds only locally, that is, on every compact subset of  $[0, L)$ . The constant functions  $f \equiv 1$  and  $f \equiv 0$  on  $[0, L)$  will be represented by the symbols  $\mathbf{1}$  and  $\mathbf{0}$ , respectively. Given any càdlàg, real-valued function  $f$  defined on  $E$ , we define  $\|f\|_T \doteq \sup_{s \in [0, T]} |f(s)|$  for every  $T < \infty$ , and let  $\|f\|_\infty \doteq \sup_{s \in [0, \infty)} |f(s)|$ , which could possibly take the value  $\infty$ . Also, for  $f \in \mathbb{D}_{\mathbb{R}}[0, \infty)$ , we use  $\Delta f(t) = f(t) - f(t-)$  to denote the jump of  $f$  at  $t$ .

For any  $f \in \mathcal{C}^\infty[0, L)$ , let  $f^{(n)}$  denote the  $n$ th derivative of  $f$ . Also, let  $\|f\|_{\mathbb{H}_0}$  be the usual  $\mathbb{L}^2$ -norm:  $\|f\|_{\mathbb{H}_0}^2 \doteq \|f\|_{\mathbb{L}^2}^2 \doteq \left( \int_0^L f^2(x) dx \right)$ , and set

$$\|f\|_{\mathbb{H}_n}^2 \doteq \|f\|_{\mathbb{H}_0}^2 + \sum_{i=1}^n \left\| f^{(i)} \right\|_{\mathbb{H}_0}^2.$$

For  $n = 1, 2$ , and  $f$  for which the corresponding first or second (weak) derivatives are well defined, we will sometimes also use the notation  $f' = f^{(1)}$  and  $f'' = f^{(2)}$ . Note that if  $f \in \mathbb{L}^2[0, \infty)$ , then there exists a real-valued sequence  $\{x_n\}$  with  $x_n \rightarrow \infty$  and  $f(x_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Moreover,  $f^2(x_n) - f^2(0) = 2 \int_0^{x_n} f(u) f'(u) du$ . Applying the Cauchy-Schwarz inequality and taking limits as  $n \rightarrow \infty$ , this implies  $|f(0)|^2 \leq 2 \|f\|_{\mathbb{H}_0} \|f'\|_{\mathbb{H}_0} \leq 2 \|f\|_{\mathbb{H}_1}^2$ . When combined with the relation  $f^2(x) = f^2(0) + 2 \int_0^x f(u) f'(u) du$  and another application of the Cauchy-Schwarz inequality, this yields the norm inequalities

$$(1.2) \quad |f(0)| \leq \sqrt{2} \|f\|_{\mathbb{H}_1}, \quad \|f\|_\infty \leq 2 \|f\|_{\mathbb{H}_1},$$

which will be used in the sequel.

For a fixed  $[0, L)$ , we define  $\mathcal{S} = \mathcal{S}[0, L)$  (respectively,  $\mathcal{S}_c = \mathcal{S}_c[0, L)$ ) to be the vector space of  $\mathcal{C}^\infty$  functions (respectively,  $\mathcal{C}^\infty$  functions with compact support) on  $[0, L)$ , equipped with the sequence of norms  $\|\cdot\|_{\mathbb{H}_n}$ ,  $n = 0, 1, 2, \dots$ , and let  $\mathbb{H}_n = \mathbb{H}_n[0, L)$  be the completion of  $\mathcal{S}$  relative to

the norm  $\|\cdot\|_n$ . Moreover, let  $\mathcal{S}'$  be the dual of  $\mathcal{S}$  (i.e., the space of continuous linear functionals on  $\mathcal{S}$ ) equipped with the strong topology and likewise, let  $\mathcal{S}'_c$  be the dual of  $\mathcal{S}_c$ . For  $n \in \mathbb{N}$ , let  $\mathbb{H}_{-n} = \mathbb{H}_{-n}[0, L]$  be the dual of  $\mathbb{H}_n[0, L]$ , with the dual norm  $\|\cdot\|_{-n}$  defined by

$$\|f\|_{-n}^2 = \sum_{k=1}^{\infty} f(e_{nk})^2, \quad f \in \mathbb{H}_{-n},$$

where  $\{e_{nk}, k = 1, \dots, \}$  is a complete orthonormal system in  $(\mathcal{S}, \|\cdot\|_n)$ . Each  $\mathbb{H}_n$  is a Sobolev space and also a Hilbert space and it follows from Maurin's theorem (see, for example, Theorem 6.53 of [1]) that  $\|\cdot\|_{\mathbb{H}_1} \stackrel{\text{HS}}{<} \|\cdot\|_{\mathbb{H}_2}$  and it follows from Lemma 5 and Assertion 11 of [2] that  $\mathcal{S}$  is a separable, Fréchet nuclear space and consequently (see Corollary 2 and Assertion 11 of [2]), its dual  $\mathcal{S}'$  is also a separable Fréchet nuclear space. For  $\nu \in \mathcal{S}'$  and  $f \in \mathcal{S}$  and likewise, for  $\nu \in \mathbb{H}_{-n}$  and  $f \in \mathbb{H}_n$ , we let  $\nu(f)$  denote the duality pairing.

**1.4.2. Measure Spaces.** The space of Radon measures on a metric space  $\mathcal{E}$ , endowed with the Borel  $\sigma$ -algebra, is denoted by  $\mathbb{M}(\mathcal{E})$ ,  $\mathbb{M}_F(\mathcal{E})$  is the subspace of finite measures in  $\mathbb{M}(\mathcal{E})$  and  $\mathbb{M}_{\leq 1}(\mathcal{E})$  is the subspace of sub-probability measures (i.e., positive measures with total mass less than or equal to 1) on  $\mathcal{E}$ . For any Borel measurable function  $f : \mathcal{E} \rightarrow \mathbb{R}$  that is integrable with respect to  $\xi \in \mathbb{M}(\mathcal{E})$ , we often use the short-hand notation  $\langle f, \xi \rangle \doteq \int_{\mathcal{E}} f(x) \xi(dx)$ . Recall that a Radon measure on  $\mathcal{E}$  is one that assigns a finite measure to every relatively compact subset of  $\mathcal{E}$ . By identifying a Radon measure  $\mu \in \mathbb{M}(\mathcal{E})$  with the mapping on  $\mathcal{C}_c(\mathcal{E})$  defined by  $f \mapsto \langle f, \mu \rangle$ , one can equivalently define a Radon measure on  $\mathcal{E}$  as a linear mapping from  $\mathcal{C}_c(\mathcal{E})$  into  $\mathbb{R}$  such that for every compact set  $\mathcal{K} \subset \mathcal{E}$ , there exists  $L_{\mathcal{K}} < \infty$  such that

$$|\langle f, \mu \rangle| \leq L_{\mathcal{K}} \|f\|_{\infty} \quad \forall f \in \mathcal{C}_c(\mathcal{E}) \text{ with } \text{supp}(f) \subset \mathcal{K}.$$

We will equip  $\mathbb{M}_F(\mathcal{E})$  with the weak topology, i.e., a sequence  $\{\mu_n\}_{n \in \mathbb{N}}$  in  $\mathbb{M}_F(\mathcal{E})$  is said to converge to  $\mu$  in the weak topology (denoted  $\mu_n \xrightarrow{w} \mu$ ) if and only if for every  $f \in \mathcal{C}_b(\mathcal{E})$ ,  $\langle f, \mu_n \rangle \rightarrow \langle f, \mu \rangle$  as  $n \rightarrow \infty$ . The symbol  $\delta_x$  will be used to denote the measure with unit mass at the point  $x$  and we will use  $\tilde{\mathbf{0}}$  to denote the identically zero Radon measure. When  $\mathcal{E}$  is an interval, say  $[0, L]$ , for notational conciseness, we will often write  $\mathbb{M}[0, L]$  instead of  $\mathbb{M}([0, L])$ . Also, for ease of notation, given  $\xi \in \mathbb{M}[0, L]$  and an interval  $(a, b) \subset [0, L]$ , we will use  $\xi(a, b)$  and  $\xi(a)$  to denote  $\xi((a, b))$  and  $\xi(\{a\})$ , respectively.

**1.4.3. Stochastic Processes.** Given a Polish space  $\mathcal{H}$ , we denote by  $\mathbb{D}_{\mathcal{H}}[0, T]$  (respectively,  $\mathbb{D}_{\mathcal{H}}[0, \infty)$ ) the space of  $\mathcal{H}$ -valued, càdlàg functions on  $[0, T]$  (respectively,  $[0, \infty)$ ), endowed with the usual Skorokhod  $J_1$ -topology (see [4] for details on this topology). Then  $\mathbb{D}_{\mathcal{H}}[0, T]$  and  $\mathbb{D}_{\mathcal{H}}[0, \infty)$  are also Polish spaces. In this work, we will be interested in  $\mathcal{H}$ -valued stochastic processes, especially the cases when  $\mathcal{H} = \mathbb{R}$ ,  $\mathcal{H} = \mathbb{M}_F[0, L]$  for some  $L \leq \infty$ ,  $\mathcal{H} = \mathcal{S}'[0, L]$  and  $\mathcal{H} = \mathbb{H}_{-n}[0, L]$  for  $n = 1, 2$ , and products of these spaces. These are random elements that are defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and take values in  $\mathbb{D}_{\mathcal{H}}[0, \infty)$ , equipped with the Borel  $\sigma$ -algebra (generated by open sets under the Skorokhod  $J_1$ -topology). A sequence  $\{Z^{(N)}\}_{N \in \mathbb{N}}$  of càdlàg,  $\mathcal{H}$ -valued processes, with  $Z^{(N)}$  defined on the probability space  $(\Omega^{(N)}, \mathcal{F}^{(N)}, \mathbb{P}^{(N)})$ , is said to converge in distribution to a càdlàg  $\mathcal{H}$ -valued process  $Z$  defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  if and only if for every bounded, continuous functional  $F : \mathbb{D}_{\mathcal{H}}[0, \infty) \rightarrow \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E}^{(N)} \left[ F(Z^{(N)}) \right] = \mathbb{E} [F(Z)],$$

where  $\mathbb{E}^{(N)}$  and  $\mathbb{E}$  are the expectation operators with respect to the probability measures  $\mathbb{P}^{(N)}$  and  $\mathbb{P}$ , respectively. Convergence in distribution of  $Z^{(N)}$  to  $Z$  will be denoted by  $Z^{(N)} \Rightarrow Z$ .

## 2. DESCRIPTION OF THE MODEL

In Section 2.1 we describe the many-server model under consideration. In Section 2.2 we introduce the state descriptor and the dynamical equations that describe the evolution of the state.

**2.1. The  $N$ -server model.** Consider a system with  $N$  servers, where arriving customers are served in a non-idling, First-Come-First-Serve (FCFS) manner, i.e., a newly arriving customer immediately enters service if there are any idle servers or, if all servers are busy, then the customer joins the back of the queue and the customer at the head of the queue (if one is present) enters service as soon as a server becomes free. Our results are not sensitive to the exact mechanism used to assign an arriving customer to an idle server as long as the non-idling condition is satisfied. Customers are assumed to be infinitely patient, i.e., they wait in queue till they receive service. Servers are non-preemptive and serve a customer to completion before starting service of a new customer. Let  $E^{(N)}$  denote the cumulative arrival process, with  $E^{(N)}(t)$  representing the total number of customers that arrive into the system in the time interval  $[0, t]$ , and let the service requirements be given by the i.i.d. sequence  $\{v_i, i = -N + 1, -N + 2, \dots, 0, 1, \dots\}$ , with common cumulative distribution function  $G$ . Let  $X^{(N)}(0)$  represent the number of customers in the system at time 0. Due to the non-idling condition, the number of customers in service at time 0 is then  $X^{(N)}(0) \wedge N$ . The sequence  $\{v_i, i = -X^{(N)}(0) \wedge N + 1, \dots, 0\}$  represents the service requirements of customers already in service at time zero, ordered according to the amount of time they have spent in service at time zero, whereas for  $i \in \mathbb{N}$ ,  $v_i$  represents the service requirement of the  $i$ th customer to enter service after time 0.

Consider the càdlàg process  $R_E^{(N)}$  defined by

$$(2.1) \quad R_E^{(N)}(s) \doteq \inf \left\{ u > s : E^{(N)}(u) > E^{(N)}(s) \right\} - s, \quad s \in [0, \infty).$$

Note that  $R_E^{(N)}(s)$  represents the time to the next arrival. The following mild assumptions will be imposed throughout, without explicit mention.

- $E^{(N)}$  is a càdlàg non-decreasing pure jump process with  $E^{(N)}(0) = 0$  and almost surely, for  $t \in [0, \infty)$ ,  $E^{(N)}(t) < \infty$  and  $E^{(N)}(t) - E^{(N)}(t-) \in \{0, 1\}$ ;
- The process  $R_E^{(N)}$  is Markovian with respect to the augmentation of its own natural filtration;
- The cumulative arrival process is independent of the i.i.d. sequence of service requirements  $\{v_j, j = -N + 1, \dots, \}$  and, given  $R_E^{(N)}(0)$ ,  $(E^{(N)}(t), t > 0)$  is independent of  $X^{(N)}(0)$  and the ages of the customers in service at time zero, where the age of a customer is defined to be the amount of time elapsed since the customer entered service;
- $G$  has density  $g$ ;
- Without loss of generality, we can (and will) assume that the mean service requirement is 1:

$$(2.2) \quad \int_{[0, \infty)} (1 - G(x)) dx = \int_{[0, \infty)} xg(x) dx = 1.$$

Also, the right-end of the support of the service distribution is denoted by

$$L \doteq \sup\{x \in [0, \infty) : G(x) < 1\}.$$

Note that the existence of a density for  $G$  implies, in particular, that  $G(0+) = 0$ .

**Remark 2.1.** The assumptions above are fairly general, allowing for a large class of arrival processes and service distributions, and this model is sometimes referred to as the G/GI/N queueing model. When  $E^{(N)}$  is a renewal process,  $R_E^{(N)}$  is simply the forward recurrence time process, the second assumption holds (see Proposition V.1.5 of Asmussen [3]) and the model corresponds to a GI/GI/N queueing system. However, the second assumption holds more generally such as, for example, when  $E^{(N)}$  is an inhomogeneous Poisson process (see, for example, Lemma II.2.2 of Asmussen [3]).

The sequence of processes  $\{R_E^{(N)}, E^{(N)}, X^{(N)}(0), v_i, i = -N + 1, \dots, 0, 1, \dots\}_{N \in \mathbb{N}}$  are all assumed to be defined on a common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  that is large enough for the independence assumptions stated above to hold.

**2.2. State Descriptor and Dynamical Equations.** As in the study of the functional strong law of large numbers limit for this model, which was carried out in Kaspi and Ramanan [22], we will represent the state of the system by the vector of processes  $(R_E^{(N)}, X^{(N)}, \nu^{(N)})$ , where  $R_E^{(N)}$  determines the cumulative arrival process via (2.1),  $X^{(N)}(t) \in \mathbb{Z}_+$  represents the total number of customers in system (including those in service and those waiting in queue) at time  $t$  and  $\nu_t^{(N)}$  is a discrete, non-negative finite measure on  $[0, L)$  that has a unit mass at the age of each customer in service at time  $t$ . Here, the age  $a_j^{(N)}$  of the  $j$ th customer is (for each realization) a piecewise linear function that is zero till the customer enters service, then increases linearly while in service (representing the time elapsed since service began) and then remains constant (equal to its service requirement) after the customer completes service and departs the system. In order to describe the state dynamics, we will find it convenient to introduce the following auxiliary processes:

- the cumulative departure process  $D^{(N)}$ , where  $D^{(N)}(t)$  is the cumulative number of customers that have departed the system in the interval  $[0, t]$ ;
- the process  $K^{(N)}$ , where  $K^{(N)}(t)$  represents the cumulative number of customers that have entered service in the interval  $[0, t]$ .

A simple mass balance on the whole system shows that

$$(2.3) \quad D^{(N)} \doteq X^{(N)}(0) - X^{(N)} + E^{(N)}$$

Likewise, recalling that  $\langle \mathbf{1}, \nu^{(N)} \rangle = \nu^{(N)}[0, L)$  represents the total number of customers in service, an analogous mass balance on the number in service yields the relation

$$(2.4) \quad K^{(N)} \doteq \langle \mathbf{1}, \nu^{(N)} \rangle - \langle \mathbf{1}, \nu_0^{(N)} \rangle + D^{(N)}.$$

For  $j \in \mathbb{N}$ , let

$$\theta_j^{(N)} \doteq \inf\{s \geq 0 : K^{(N)}(s) \geq j\},$$

with the usual convention that the infimum of an empty set is infinity, and note that  $\theta_j^{(N)}$  denotes the time of entry into service of the  $j$ th customer to enter service after time 0. In addition, for  $j = -X^{(N)}(0) \wedge N + 1, \dots, 0$ , set  $\theta_j^{(N)} = -a_j^{(N)}(0)$  to be the amount of time that the  $j$ th customer in service at time 0 has already been in service. Then, for  $t \in [0, \infty)$  and  $j = -X^{(N)}(0) \wedge N + 1, \dots, 0, 1, \dots$ , the age process is given explicitly by

$$(2.5) \quad a_j^{(N)}(t) = \begin{cases} [t - \theta_j^{(N)}] \vee 0 & \text{if } t - \theta_j^{(N)} < v_j, \\ v_j & \text{otherwise.} \end{cases}$$

Due to the FCFS nature of the service,  $K^{(N)}(t)$  is also the highest index of any customer that has entered service and (2.5) implies that for  $j > K^{(N)}(t)$ ,  $\theta_j^{(N)} > t$  and  $a_j^{(N)}(t) = 0$ . The measure  $\nu^{(N)}$  can then be expressed as

$$(2.6) \quad \nu_t^{(N)} = \sum_{j=-\langle \mathbf{1}, \nu_0^{(N)} \rangle + 1}^{K^{(N)}(t)} \delta_{a_j^{(N)}(t)} \mathbb{1}_{\{a_j^{(N)}(t) < v_j\}},$$

where  $\delta_x$  represents the Dirac mass at the point  $x$ . The non-idling condition, which stipulates that there be no idle servers when there are more than  $N$  customers in the system, is expressed via the relation

$$(2.7) \quad N - \langle \mathbf{1}, \nu^{(N)} \rangle = [N - X^{(N)}]^+.$$

For future purposes note that (2.3), (2.4) and (2.7), together with the elementary identity  $x - x \vee 0 = x \wedge 0$ , imply the relation

$$(2.8) \quad K^{(N)} = X^{(N)} \wedge N - X^{(N)}(0) \wedge N + D^{(N)}.$$



Note that  $\langle \mathbf{1}, \nu^{(N)} \rangle \leq N$  because the maximum number of customers in service at any given time is bounded by the number of servers. In addition, if the support of  $\nu_0^{(N)}$  lies in  $[0, L)$  then it follows from (2.5) and (2.6) that  $\nu_t^{(N)}$  takes values in  $\mathbb{M}_F[0, L)$  for every  $t \in [0, \infty)$ . Thus, the state of the system is represented by the càdlàg process  $(R_E^{(N)}, X^{(N)}, \nu^{(N)})$ , which takes values in  $\mathbb{R}_+^2 \times \mathbb{M}_F[0, L)$ . For an explicit construction of the state that also shows that the state and auxiliary processes are càdlàg, see Lemma A.1 of Kang and Ramanan [20]. The results obtained in this paper are independent of the particular rule used to assign customers to stations, but for technical purposes we will find it convenient to also introduce the additional “station process”  $\sigma^{(N)} \doteq (\sigma_j^{(N)}, j \in \{-N+1, \dots, 0\} \cup \mathbb{N})$ . For each  $t \in [0, \infty)$ , if customer  $j$  has already entered service by time  $t$ , then  $\sigma_j^{(N)}(t)$  is equal to the index  $i \in \{1, \dots, N\}$  of the station at which customer  $j$  receives/received service and  $\sigma_j^{(N)}(t) \doteq 0$  otherwise. Finally, for  $t \in [0, \infty)$ , let  $\overline{\mathcal{F}}_t^{(N)}$  be the  $\sigma$ -algebra generated by  $\{R_E^{(N)}(s), a_j^{(N)}(s), \sigma_j^{(N)}(s), j \in \{-N, \dots, 0\} \cup \mathbb{N}, s \in [0, t]\}$ , and let  $\{\mathcal{F}_t^{(N)}, t \geq 0\}$  denote the associated right continuous filtration that is completed (with respect to  $\mathbb{P}$ ) so that it satisfies the usual conditions. Then it is easy to verify that  $(R_E^{(N)}, X^{(N)}, \nu^{(N)})$  is  $\{\mathcal{F}_t^{(N)}\}$ -adapted (see, for example, Section 2.2 of Kaspi and Ramanan [22]). In fact, as shown in Lemma B.1 of Kang and Ramanan [20],  $\{(R_E^{(N)}(t), X^{(N)}(t), \nu_t^{(N)}), \mathcal{F}_t^{(N)}, t \geq 0\}$  is a strong Markov process.

**Remark 2.2.** The assumed Markov property of  $R_E^{(N)}$  with respect to (the completed, right continuous version of) its natural filtration and the independence properties of  $E^{(N)}$  assumed in Section 2.1 together imply that for any  $t \in [0, \infty)$ , given  $R_E^{(N)}(t)$  the future arrivals process  $\{E^{(N)}(s), s > t\}$  is independent of  $\mathcal{F}_t^{(N)}$ .

### 3. FLUID LIMIT

We now recall the functional strong law of large numbers limit or, equivalently, fluid limit obtained in [22]. The initial data describing the system consists of  $E^{(N)}$ , the cumulative arrivals after zero,  $X^{(N)}(0)$ , the number in system at time zero, and  $\nu_0^{(N)}$ , the age distribution of customers in service at time zero. The initial data belongs to the following space:

$$(3.1) \quad \mathcal{I}_0 \doteq \{(f, x, \mu) \in \mathbb{I}_{\mathbb{R}_+}[0, \infty) \times \mathbb{R}_+ \times \mathbb{M}_{\leq 1}[0, L) : 1 - \langle \mathbf{1}, \mu \rangle = [1 - x]^+\},$$

where  $\mathbb{I}_{\mathbb{R}_+}[0, \infty)$  is the subset of non-decreasing functions  $f \in \mathbb{D}_{\mathbb{R}_+}[0, \infty)$  with  $f(0) = 0$ . Assume that  $\mathcal{I}_0$  is equipped with the product topology. Consider the “fluid scaled” versions of the processes  $H = E, X, K, D$  and measures  $H = \nu$  defined by

$$(3.2) \quad \overline{H}^{(N)} \doteq \frac{H^{(N)}}{N},$$

and let

$$\overline{R}_E^{(N)}(t) \doteq R_E^{(N)}(E^{(N)}(t)), \quad t \in [0, \infty),$$

for  $N \in \mathbb{N}$ . The fluid results in [22] were obtained under Assumptions 1 and 2 below.

**Assumption 1.** *There exists  $(\overline{E}, \overline{x}_0, \overline{\nu}_0) \in \mathcal{I}_0$  such that, as  $N \rightarrow \infty$ ,  $\mathbb{E}[\overline{X}^{(N)}(0)] \rightarrow \mathbb{E}[\overline{X}(0)]$ ,  $\mathbb{E}[\overline{E}^{(N)}(t)] \rightarrow \mathbb{E}[\overline{E}(t)]$  and, almost surely,*

$$(\overline{E}^{(N)}, \overline{x}_0^{(N)}, \overline{\nu}_0^{(N)}) \rightarrow (\overline{E}, \overline{x}_0, \overline{\nu}_0) \quad \text{in } \mathcal{I}_0.$$

Next, recall that  $G$  has density  $g$ , and let  $h$  denote its hazard rate:

$$(3.3) \quad h(x) \doteq \frac{g(x)}{1 - G(x)}, \quad x \in [0, L).$$

Observe that  $h$  is automatically locally integrable on  $[0, L)$  because for every  $0 \leq a \leq b < L$ ,

$$(3.4) \quad \int_a^b h(x) dx = \ln(1 - G(a)) - \ln(1 - G(b)) < \infty.$$

However,  $h$  is not integrable on  $[0, L)$ . In particular, when  $L < \infty$ ,  $h$  is unbounded on  $(\ell', L)$  for every  $\ell' < L$ .

**Assumption 2.** *At least one of the following two properties holds:*

- (a) *There exists  $\ell' < \infty$  such that  $h$  is bounded on  $(\ell', \infty)$ ;*
- (b) *There exists  $\ell' < L$  such that  $h$  is lower-semicontinuous on  $(\ell', L)$ .*

Note that Assumption 2(a) automatically implies that  $L = \infty$ . In Proposition 6.1 (see also Theorem 5.1 of [22]) we provide a succinct description of the dynamics of the  $N$ -server system in terms of certain integral equations. Here, we first introduce the deterministic analog of these equations, which we refer to as the fluid equations.

**Definition 3.1. (Fluid Equations)** The càdlàg function  $(\bar{X}, \bar{\nu})$  defined on  $[0, \infty)$  and taking values in  $\mathbb{R}_+ \times \mathbb{M}_{\leq 1}[0, L)$  is said to solve the *fluid equations* associated with  $(\bar{E}, \bar{x}_0, \bar{\nu}_0) \in \mathcal{I}_0$  if and only if  $\bar{X}(0) = \bar{x}_0$  and for every  $t \in [0, \infty)$ ,

$$(3.5) \quad \int_0^t \langle h, \bar{\nu}_s \rangle ds < \infty$$

and the following relations are satisfied: for every  $\varphi \in \mathcal{C}_c^{1,1}([0, L) \times \mathbb{R}_+)$ ,

$$(3.6) \quad \begin{aligned} \langle \varphi(\cdot, t), \bar{\nu}_t \rangle &= \langle \varphi(\cdot, 0), \bar{\nu}_0 \rangle + \int_0^t \langle \varphi_s(\cdot, s) + \varphi_x(\cdot, s), \bar{\nu}_s \rangle ds \\ &\quad - \int_0^t \langle h(\cdot) \varphi(\cdot, s), \bar{\nu}_s \rangle ds + \int_{[0, t]} \varphi(0, s) d\bar{K}(s), \end{aligned}$$

$$(3.7) \quad \bar{X}(t) = \bar{X}(0) + \bar{E}(t) - \int_0^t \langle h, \bar{\nu}_s \rangle ds$$

and

$$(3.8) \quad 1 - \langle \mathbf{1}, \bar{\nu}_t \rangle = [1 - \bar{X}(t)]^+,$$

where

$$(3.9) \quad \bar{K}(t) = \langle \mathbf{1}, \bar{\nu}_t \rangle - \langle \mathbf{1}, \bar{\nu}_0 \rangle + \int_0^t \langle h, \bar{\nu}_s \rangle ds.$$

We now recall the result established in Kaspi and Ramanan [22] (see Theorems 3.4 and 3.5 therein), which shows that under Assumptions 1 and 2, the fluid equations uniquely characterize the functional strong law of large numbers or mean-field limit of the  $N$ -server system, in the asymptotic regime where the number of servers and arrival rates both tend to infinity.

**Theorem 3.2** (Kaspi and Ramanan [22]). *Suppose Assumptions 1 and 2 are satisfied and  $(\bar{E}, \bar{X}(0), \bar{\nu}_0) \in \mathcal{I}_0$  is the limit of the initial data as stated in Assumption 1. Then there exists a unique solution  $(\bar{X}, \bar{\nu})$  to the associated fluid equations (3.5)–(3.8) and, as  $N \rightarrow \infty$ ,  $(\bar{X}^{(N)}, \bar{\nu}^{(N)})$  converges almost surely to  $(\bar{X}, \bar{\nu})$ . Moreover,  $(\bar{X}, \bar{\nu})$  satisfies the non-idling condition (3.8) and, for every  $f \in \mathcal{C}_b(\mathbb{R}_+)$ ,*

$$(3.10) \quad \begin{aligned} \int_{[0, M)} f(x) \bar{\nu}_t(dx) &= \int_{[0, M)} f(x+t) \frac{1 - G(x+t)}{1 - G(x)} \bar{\nu}_0(dx) \\ &\quad + \int_{[0, t]} f(t-s)(1 - G(t-s)) d\bar{K}(s), \end{aligned}$$

where  $\bar{K}$  satisfies the relation (3.9).

**Remark 3.3.** The fluid limit will be said to be critical if  $\bar{X}(t) = 1$  for all  $t \in [0, \infty)$ . In addition, it will be said to be subcritical (respectively, supercritical) if for every  $T \in [0, \infty)$ ,  $\sup_{t \in [0, T]} \bar{X}(t) < 1$  (respectively,  $\inf_{t \in [0, T]} \bar{X}(t) > 1$ ). Although, in general, the fluid limit may not stay in one regime for all  $t$  and may instead experience periods of subcriticality, criticality and supercriticality, for many natural choices of initial data, such as either starting empty or starting on the so-called “invariant manifold” of the fluid limit, the fluid limit does belong to one of these three categories. Specifically, if we define the “invariant” fluid age measure to be

$$(3.11) \quad \bar{\nu}_*(dx) = (1 - G(x)) dx, \quad x \in [0, L),$$

then it follows from Remark 3.7 and Theorem 3.8 of Kaspi and Ramanan [22] that the fluid limit associated with the initial data  $(\mathbf{1}, 1, \bar{\nu}_*)$  is critical, the fluid limit associated with the initial data  $(\mathbf{1}, a, \bar{\nu}_*)$  for some  $a > 1$  is supercritical, and if the support of  $G$  is  $[0, \infty)$ , then the fluid limit associated with the initial data  $(\bar{\lambda}\mathbf{1}, 0, \bar{\mathbf{0}})$  is subcritical whenever  $\bar{\lambda} \leq 1$ . A complete characterization of the invariant manifold of the fluid in the presence of abandonments can be found in [21].

#### 4. CERTAIN MARTINGALE MEASURES AND THEIR STOCHASTIC INTEGRALS

We now introduce some quantities that appear in the characterization of the functional central limit. The sequence of martingales obtained by compensating the departure processes in each of the  $N$ -server systems played an important role in establishing the fluid limit result in [22]. Whereas under the fluid scaling the limit of this sequence converges weakly to zero, under the diffusion scaling considered here, it converges to a non-trivial limit. This limit can be described in terms of a corresponding martingale measure, which is introduced in Section 4.1. In Section 4.2 certain stochastic convolution integrals with respect to these martingale measures are introduced, which arise in the representation formula for the centered age process in the  $N$ -server system (see Proposition 6.4). Finally, the associated “limit” quantities are defined in Section 4.3. The reader is referred to Chapter 2 of Walsh [35] for basic definitions of martingale measures and their stochastic integrals.

**4.1. A Martingale Measure Sequence.** Throughout this section, suppose that  $(E^{(N)}, x_0^{(N)}, \nu_0^{(N)})$  is an  $\mathcal{I}_0$ -valued random element representing the initial data of the  $N$ -server system, and let  $(R_E^{(N)}, X^{(N)}, \nu^{(N)})$  be the associated state process described in Section 2.2. For any measurable function  $\varphi$  on  $[0, L) \times \mathbb{R}_+$ , consider the sequence of processes  $\{Q_\varphi^{(N)}\}_{N \in \mathbb{N}}$  defined by

$$(4.1) \quad Q_\varphi^{(N)}(t) \doteq \sum_{s \in [0, t]} \sum_{j = -\langle \mathbf{1}, \nu_0^{(N)} \rangle + 1}^{K^{(N)}(t)} \mathbb{1}_{\left\{ \frac{d}{dt} a_j^{(N)}(s-) > 0, \frac{d}{dt} a_j^{(N)}(s+) = 0 \right\}} \varphi(a_j^{(N)}(s), s)$$

for  $N \in \mathbb{N}$  and  $t \in [0, \infty)$ , where  $K^{(N)}$  and  $a_j^{(N)}$  are, respectively, the cumulative entry into service process and the age process of customer  $j$  as defined by the relations (2.4) and (2.5). Note from (2.5) that the  $j$ th customer completed service (and hence departed the system) at time  $s$  if and only if

$$\frac{d}{dt} a_j^{(N)}(s-) > 0 \quad \text{and} \quad \frac{d}{dt} a_j^{(N)}(s+) = 0.$$

Thus, substituting  $\varphi = \mathbf{1}$  in (4.1), we see that  $Q_{\mathbf{1}}^{(N)}$  is equal to  $D^{(N)}$ , the cumulative departure process of (2.3). Moreover, for  $N \in \mathbb{N}$  and every bounded measurable function  $\varphi$  on  $[0, L) \times [0, \infty)$ , consider the process  $A_\varphi^{(N)}$  defined by

$$(4.2) \quad A_\varphi^{(N)}(t) \doteq \int_0^t \left( \int_{[0, L)} \varphi(x, s) h(x) \nu_s^{(N)}(dx) \right) ds, \quad t \in [0, \infty),$$

and set

$$(4.3) \quad M_\varphi^{(N)} \doteq Q_\varphi^{(N)} - A_\varphi^{(N)}.$$

It was shown in Corollary 5.5 of Kaspi and Ramanan [22] that for all bounded, continuous functions  $\varphi$  defined on  $[0, L) \times [0, \infty)$ ,  $A_\varphi^{(N)}$  is the  $\{\mathcal{F}_t^{(N)}\}$ -compensator of  $Q_\varphi^{(N)}$ , and that  $M_\varphi^{(N)}$  is a càdlàg  $\{\mathcal{F}_t^{(N)}\}$ -martingale (see also Lemma 5.2 of Kang and Ramanan [20] for a generalization of this result to a larger class of  $\varphi$ ). Moreover, from the proof of Lemma 5.9 of Kaspi and Ramanan [22], it follows that the predictable quadratic variation of  $M_\varphi^{(N)}$  takes the form

$$(4.4) \quad \langle M_\varphi^{(N)} \rangle_t = A_{\varphi^2}^{(N)}(t) = \int_0^t \left( \int_{[0, L)} \varphi^2(x) h(x) \nu_s^{(N)}(dx) \right) ds, \quad t \in [0, \infty).$$

Now, for  $B \in \mathcal{B}[0, L)$  and  $t \in [0, \infty)$ , define

$$(4.5) \quad \mathcal{M}_t^{(N)}(B) \doteq M_{\mathbb{1}_B}^{(N)}(t) = Q_{\mathbb{1}_B}^{(N)}(t) - A_{\mathbb{1}_B}^{(N)}(t).$$

Let  $\mathcal{B}_0[0, L)$  denote the algebra generated by the intervals  $[0, x]$ ,  $x \in [0, L)$ . It is easy to verify that  $\mathcal{M}^{(N)} = \{\mathcal{M}_t^{(N)}(B), \mathcal{F}_t^{(N)}, t \geq 0, B \in \mathcal{B}_0[0, L)\}$  is a martingale measure (for completeness, a proof is provided in Lemma A.1 of the Appendix). We now show that  $\mathcal{M}^{(N)}$  is in fact an orthogonal martingale measure (see page 288 of Walsh [35] for a definition). Essentially, the orthogonality property holds because almost surely, no two departures occur at the same time. First, in Lemma 4.1 below, we first state a slight generalization of this latter property, which is also used in Section 8.2 to establish an asymptotic independence result. Given  $r, s \in [0, \infty)$ , let  $D^{(N), r}(s)$  denote the cumulative number of departures during  $(r, r + s]$  of customers that entered service at or before  $r$ . In what follows, recall that the notation  $\Delta f(t) = f(t) - f(t-)$  is used to denote the jump of a function  $f$  at  $t$ .

**Lemma 4.1.** *For every  $N \in \mathbb{N}$ ,  $\mathbb{P}$  almost surely,*

$$(4.6) \quad \Delta D^{(N)}(t) \leq 1, \quad t \in [0, \infty),$$

and

$$(4.7) \quad \sum_{s \in [0, \infty)} \Delta E^{(N)}(r + s) \Delta D^{(N), r}(s) = 0, \quad r \in [0, \infty).$$

We relegate the proof of the lemma to Section A.2 and instead, now establish the orthogonality property.

**Corollary 4.2.** *For each  $N \in \mathbb{N}$ , the martingale measure  $\mathcal{M}^{(N)} = \{\mathcal{M}_t^{(N)}(B), \mathcal{F}_t^{(N)}; t \geq 0, B \in \mathcal{B}_0[0, L)\}$  is orthogonal and has covariance functional*

$$(4.8) \quad \begin{aligned} Q_t^{(N)}(B, \tilde{B}) &\doteq \langle \mathcal{M}^{(N)}(B), \mathcal{M}^{(N)}(\tilde{B}) \rangle_t = A_{\mathbb{1}_{B \cap \tilde{B}}}^{(N)}(t) \\ &= \int_0^t \left( \int_{B \cap \tilde{B}} h(x) \nu_s^{(N)}(dx) \right) ds \end{aligned}$$

for  $B, \tilde{B} \in \mathcal{B}_0[0, L)$ .

*Proof.* In order to show that the martingale measure  $\mathcal{M}^{(N)}$  is orthogonal, it suffices to show that for every  $B, \tilde{B} \in \mathcal{B}_0[0, L)$  such that  $B \cap \tilde{B} = \emptyset$ , the martingales  $\{\mathcal{M}_t^{(N)}(B); t \geq 0\}$  and  $\{\mathcal{M}_t^{(N)}(\tilde{B}); t \geq 0\}$  are orthogonal or, in other words, that

$$(4.9) \quad B \cap \tilde{B} = \emptyset \quad \Rightarrow \quad \langle \mathcal{M}^{(N)}(B), \mathcal{M}^{(N)}(\tilde{B}) \rangle \equiv 0.$$

Here,  $\langle \cdot, \cdot \rangle$  represents the predictable quadratic covariation between the two martingales. Fix two sets  $B, \tilde{B} \in \mathcal{B}_0[0, L)$  with  $B \cap \tilde{B} = \emptyset$ . By Lemma 5.2 of Kang and Ramanan [20],  $\mathcal{M}^{(N)}(B) = M_{\mathbb{1}_B}^{(N)}$  and  $\mathcal{M}^{(N)}(\tilde{B}) = M_{\mathbb{1}_{\tilde{B}}}^{(N)}$  are martingales that are compensated sums of jumps, where the jumps occur at departure times of customers whose ages lie in the sets  $B$  and  $\tilde{B}$ , respectively. Since, by (4.6) of Lemma 4.1, there are almost surely no two departures that occur at the same time, it follows that almost surely, the set of jump points of  $\mathcal{M}^{(N)}(B)$  and  $\mathcal{M}^{(N)}(\tilde{B})$  are disjoint. By Theorem

4.52 of Chapter 1 of Jacod and Shiryaev [17], it then follows that the martingales are orthogonal and (4.9) holds. Combining (4.9) with (4.4) and the biadditivity of the covariance functional, we then obtain (4.8).  $\square$

Due to the orthogonality property established in Corollary 4.2, we can now define stochastic integrals with respect to the martingale measure  $\mathcal{M}^{(N)}$ . Since  $\mathbb{E}[A_1^{(N)}(T)] < \infty$  by Lemma 5.6 of Kaspi and Ramanan [22] and  $\nu^{(N)}$  is a non-negative measure, the stochastic integral is defined for the space of deterministic, continuous and bounded functions on  $[0, L] \times [0, \infty)$  (it is in fact defined for a larger class of so-called predictable integrands satisfying a suitable integrability property, see page 292 of Walsh [35]). Moreover, by Theorem 2.5 on page 295 of Walsh [35], it follows that for all bounded and continuous  $\varphi$ , the stochastic integral  $\{\mathcal{M}_t^{(N)}(\varphi)(B), \{\mathcal{F}_t^{(N)}\}; t \geq 0, B \in \mathcal{B}_0[0, L]\}$  of  $\varphi$  with respect to  $\mathcal{M}^{(N)}$  is a càdlàg orthogonal martingale measure with covariance functional

$$(4.10) \quad \langle \mathcal{M}^{(N)}(\varphi)(B), \mathcal{M}^{(N)}(\tilde{\varphi})(\tilde{B}) \rangle_t = \int_0^t \left( \int_{B \cap \tilde{B}} \varphi(x, s) \tilde{\varphi}(x, s) h(x) \nu_s^{(N)}(dx) \right) ds$$

for bounded, continuous  $\varphi, \tilde{\varphi}$  and  $B, \tilde{B} \in \mathcal{B}_0[0, L]$ . When  $B = [0, L]$ , we will drop the dependence on  $B$  and simply write

$$\mathcal{M}^{(N)}(\varphi) = \mathcal{M}^{(N)}(\varphi)([0, L]).$$

**Remark 4.3.** For  $\varphi \in \mathcal{C}_b([0, L] \times \mathbb{R}_+)$ , the stochastic integral  $\mathcal{M}^{(N)}(\varphi)$  admits a càdlàg version. Indeed, the càdlàg martingale  $M_\varphi^{(N)}$  defined in (4.3) is a version of the stochastic integral  $\mathcal{M}^{(N)}(\varphi)$ .

It was shown in Lemma 5.9 of Kaspi and Ramanan [22] that

$$\overline{\mathcal{M}}^{(N)} \doteq \frac{\mathcal{M}^{(N)}}{N} \Rightarrow \overline{\mathcal{M}} \equiv \tilde{\mathbf{0}}$$

in the space of càdlàg finite Radon measure valued processes. Therefore, consistent with the diffusion scaling (5.1), we set

$$(4.11) \quad \widehat{\mathcal{M}}^{(N)} \doteq \frac{\mathcal{M}^{(N)}}{\sqrt{N}}.$$

It is clear from the above discussion that each  $\widehat{\mathcal{M}}^{(N)}$  is an orthogonal martingale measure with covariance functional

$$\widehat{\mathcal{Q}}_t^{(N)}(B, \tilde{B}) = \int_0^t \left( \int_{B \cap \tilde{B}} h(x) \bar{\nu}_s^{(N)}(dx) \right) ds,$$

and that for any  $\varphi$  in a suitable class of functions that includes the space of bounded and continuous functions, the stochastic integral  $\widehat{\mathcal{M}}^{(N)}(\varphi)$  is a well defined càdlàg, orthogonal  $\{\mathcal{F}_t^{(N)}\}$  martingale measure. Moreover, for every bounded, continuous  $\varphi, \tilde{\varphi}$  and  $t \in [0, \infty)$ ,

$$(4.12) \quad \langle \widehat{\mathcal{M}}^{(N)}(\varphi), \widehat{\mathcal{M}}^{(N)}(\tilde{\varphi}) \rangle_t = \int_0^t \left( \int_{[0, L]} \varphi(x, s) \tilde{\varphi}(x, s) h(x) \bar{\nu}_s^{(N)}(dx) \right) ds.$$

**4.2. Some Associated Stochastic Convolution Integrals.** In Proposition 6.4 we show that the stochastic measure-valued process  $\{\nu_t^{(N)}, t \geq 0\}$  that describes the ages of customers in the  $N$ -server system admits a convenient representation that is similar to the representation (3.10) for its fluid counterpart  $\{\bar{\nu}_t, t \geq 0\}$ , except that it contains an additional stochastic term involving the following stochastic convolution integral. For  $N \in \mathbb{N}$ ,  $\varphi \in \mathcal{C}_b([0, L] \times [0, \infty))$  and  $t \in [0, \infty)$ , define

$$(4.13) \quad \mathcal{H}_t^{(N)}(\varphi) \doteq \int \int_{[0, L] \times [0, t]} \varphi(x + t - s, s) \frac{1 - G(x + t - s)}{1 - G(x)} \mathcal{M}^{(N)}(dx, ds),$$

where the latter stochastic integral with respect to  $\mathcal{M}^{(N)}$  is well defined because  $\mathcal{M}^{(N)}$  is an orthogonal martingale measure and the function  $(x, s) \mapsto \varphi(x + t - s, s)(1 - G(x + t - s))/(1 - G(x))$

lies in  $\mathcal{C}_b([0, L] \times \mathbb{R}_+)$  for all  $\varphi \in \mathcal{C}_b([0, L] \times \mathbb{R}_+)$ . The scaled version of this quantity is then defined in the obvious manner: for  $N \in \mathbb{N}$ ,  $\varphi \in \mathcal{C}_b([0, L] \times [0, \infty))$  and  $t \in [0, \infty)$ , let

$$(4.14) \quad \widehat{\mathcal{H}}_t^{(N)}(\varphi) \doteq \int \int_{[0, L] \times [0, t]} \varphi(x+t-s, s) \frac{1-G(x+t-s)}{1-G(x)} \widehat{\mathcal{M}}^{(N)}(dx, ds).$$

**4.3. Related Limit Quantities.** We now define some additional quantities, which we subsequently show (in Corollaries 8.3 and 8.7) to be limits of the sequences  $\{\widehat{\mathcal{M}}^{(N)}\}_{N \in \mathbb{N}}$  and  $\{\widehat{\mathcal{H}}^{(N)}\}_{N \in \mathbb{N}}$ . Fix a probability space  $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$  and, on this space, let  $\widehat{\mathcal{M}} = \{\widehat{\mathcal{M}}_t(B), B \in \mathcal{B}_0[0, L], t \in [0, \infty)\}$  be a continuous martingale measure with the deterministic covariance functional

$$(4.15) \quad \widehat{\mathcal{Q}}_t(B, \tilde{B}) \doteq \langle \widehat{\mathcal{M}}(B), \widehat{\mathcal{M}}(\tilde{B}) \rangle_t = \int_0^t \left( \int_{[0, L]} \mathbb{1}_{B \cap \tilde{B}}(x) h(x) \overline{\nu}_s(dx) \right) ds$$

for  $t \in [0, \infty)$ . Thus,  $\widehat{\mathcal{M}}$  is a white noise. Let  $\mathcal{C}_{\widehat{\mathcal{M}}}$  denote the subset of continuous functions on  $[0, L] \times [0, \infty)$  that satisfies

$$(4.16) \quad \int_0^t \left( \int_{[0, L]} \varphi^2(x, s) h(x) \overline{\nu}_s(dx) \right) ds < \infty, \quad t \in [0, \infty).$$

Note that  $\mathcal{C}_{\widehat{\mathcal{M}}}$  includes, in particular, the space of bounded and continuous functions. For any such  $\varphi \in \mathcal{C}_{\widehat{\mathcal{M}}}$ , the stochastic integral  $\widehat{\mathcal{M}}^{(N)}(\varphi)$  is a well defined càdlàg, orthogonal  $\{\widehat{\mathcal{F}}_t, t \geq 0\}$  martingale measure (see page 292 of Walsh [35] for details). For any  $\varphi \in \mathcal{C}_{\widehat{\mathcal{M}}}$  and  $t \in [0, \infty)$ , the stochastic integral of  $\varphi$  with respect to  $\widehat{\mathcal{M}}$  on  $[0, L] \times [0, t]$ , denoted by

$$(4.17) \quad \widehat{\mathcal{M}}_t(\varphi) \doteq \int \int_{[0, L] \times [0, t]} \varphi(x, s) \widehat{\mathcal{M}}(dx, ds),$$

is well defined. Moreover, because  $\widehat{\mathcal{M}}$  is a continuous martingale measure,  $\widehat{\mathcal{M}}(\varphi)$  has a version as a continuous real-valued process. In fact, as Corollary 8.3 shows,  $\widehat{\mathcal{M}}$  admits a version as a continuous  $\mathbb{H}_{-2}$ -valued process.

Next, for  $t \in [0, \infty)$  and  $f \in \mathcal{C}_b[0, L]$ , let  $\widehat{\mathcal{H}}_t(f)$  be the random variable given by the following convolution integral:

$$(4.18) \quad \widehat{\mathcal{H}}_t(f) \doteq \int \int_{[0, L] \times [0, t]} f(x+t-s) \frac{1-G(x+t-s)}{1-G(x)} \widehat{\mathcal{M}}(dx, ds).$$

It is shown in Lemma 8.6 that if  $f$  is bounded and Hölder continuous then the real-valued stochastic process  $\widehat{\mathcal{H}}(f) = \{\widehat{\mathcal{H}}_t(f), t \geq 0\}$  admits a continuous version, and that  $\widehat{\mathcal{H}}$  also admits a version as a continuous  $\mathbb{H}_{-2}$ -valued process.

In order to write the convolution integrals in a more succinct fashion, we introduce the family of operators  $\{\Psi_t, t \geq 0\}$  defined, for  $t > 0$  and  $(x, s) \in [0, L] \times [0, \infty)$ , by

$$(4.19) \quad (\Psi_t f)(x, s) \doteq f(x + (t-s)^+) \frac{1-G(x + (t-s)^+)}{1-G(x)},$$

for bounded measurable  $f$ , where recall  $(t-s)^+ = \max(t-s, 0)$ . Each operator  $\Psi_t$  maps the space of bounded measurable functions on  $[0, L]$  to the space of bounded measurable functions on  $[0, L] \times [0, \infty)$  and we also have

$$(4.20) \quad \sup_{t \in [0, \infty)} \|\Psi_t \varphi\|_\infty \leq \|\varphi\|_\infty.$$

We can then write  $\widehat{\mathcal{H}}$  and  $\widehat{\mathcal{H}}^{(N)}$ , respectively, in terms of  $\widehat{\mathcal{M}}$  and  $\widehat{\mathcal{M}}^{(N)}$  as follows:

$$(4.21) \quad \widehat{\mathcal{H}}_t(f) = \widehat{\mathcal{M}}_t(\Psi_t f), \quad \widehat{\mathcal{H}}_t^{(N)}(f) = \widehat{\mathcal{M}}_t^{(N)}(\Psi_t f), \quad t \geq 0.$$

## 5. MAIN RESULTS

The main results of the paper are stated in Section 5.3. They rely on some basic assumptions and the definition of a certain map, which are first introduced in Sections 5.1 and 5.2, respectively. Corollaries of the main results are discussed in Section 5.4.

**5.1. Basic Assumptions.** We now state our assumptions on the arrival processes and initial conditions. For  $Y = E, x_0, \nu, X, K$ , let  $\bar{Y}$  be the corresponding fluid limit as described in Theorem 3.2. For  $N \in \mathbb{N}$ , the diffusion scaled quantities  $\widehat{Y}^{(N)}$  are defined as follows:

$$(5.1) \quad \widehat{Y}^{(N)} \doteq \sqrt{N} \left( \bar{Y}^{(N)} - \bar{Y} \right).$$

For simplicity, we restrict the arrival processes to be either renewal processes or time-inhomogeneous Poisson processes.

**Assumption 3.** *The sequence  $\{E^{(N)}\}_{N \in \mathbb{N}}$  of cumulative arrival processes satisfies one of the following two conditions:*

- (a) *there exist constants  $\bar{\lambda}, \sigma^2 \in (0, \infty)$  and  $\beta \in \mathbb{R}$ , such that for every  $N \in \mathbb{N}$ ,  $E^{(N)}$  is a renewal process with i.i.d. inter-renewal times  $\{\xi_j^{(N)}\}_{j \in \mathbb{N}}$  that have mean  $1/\lambda^{(N)}$  and variance  $(\sigma^2/\bar{\lambda})/(\lambda^{(N)})^2$ , where*

$$(5.2) \quad \lambda^{(N)} \doteq \bar{\lambda}N - \beta\sqrt{N}.$$

- (b) *there exist locally integrable functions  $\bar{\lambda}$  and  $\beta$  on  $[0, \infty)$  such that for every  $N \in \mathbb{N}$ ,  $E^{(N)}$  is an inhomogeneous Poisson process with intensity function*

$$(5.3) \quad \lambda^{(N)}(t) \doteq \bar{\lambda}(t)N - \beta(t)\sqrt{N}, \quad t \in [0, \infty).$$

**Remark 5.1.** Let  $\bar{\lambda}$  and  $\beta$  be the locally integrable functions defined in Assumption 3 (which are constant if Assumption 3(a) holds) and let  $\sigma(\cdot)$  be the locally square integrable function that equals the constant  $\sqrt{\sigma^2}$  if Assumption 3(a) holds, and equals  $\sqrt{\lambda(\cdot)}$  if Assumption 3(b) holds. Then, given a standard Brownian motion  $B$ , the process  $\widehat{E}$  given by

$$(5.4) \quad \widehat{E}(t) \doteq \int_0^t \sigma(s) dB(s) - \int_0^t \beta(s) ds, \quad t \in [0, \infty),$$

is a well defined diffusion and therefore a semimartingale, with  $\int_0^t \sigma(s) dB(s), t \geq 0$  being the local martingale and  $\int_0^t \beta(s) ds, t \geq 0$ , the finite variation process in the decomposition. If Assumption 3 holds then it is easy to see that  $\bar{E}$  in Assumption 1 is given by  $\bar{E}(t) = \int_0^t \bar{\lambda}(s) ds, t \geq 0$ , and  $\widehat{E}^{(N)} \Rightarrow \widehat{E}$  as  $N \rightarrow \infty$  (a proof of the latter convergence can be found in Proposition 8.4, which establishes a more general result).

We now impose a technical condition on the service distribution, which is used mainly to establish the convergence of  $\widehat{\mathcal{H}}^{(N)}(f)$  to  $\widehat{\mathcal{H}}(f)$  in  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  for Hölder continuous  $f$  in Section 8.

**Assumption 4.** *For every  $x \in [0, L)$ , the function from  $[0, L)$  to  $[0, 1]$  that takes  $y \mapsto (1 - G(x + y))/(1 - G(x))$  is Hölder continuous on  $[0, \infty)$ , uniformly in  $x$ , i.e., there exist  $C_G < \infty$  and  $\gamma_G \in (0, 1)$  and  $\delta > 0$  such that for every  $x \in [0, L)$  and  $y, y' \in [0, L)$  with  $|y - y'| < \delta$ ,*

$$(5.5) \quad \frac{|G(x + y) - G(x + y')|}{1 - G(x)} \leq C_G |y - y'|^{\gamma_G}.$$

**Remark 5.2.** As shown below, Assumption 4 is satisfied if either  $h$  is bounded, or if there exists  $l_0 < \infty$  such that  $\sup_{x \in [l_0, \infty)} h(x) < \infty$  and  $G$  is uniformly Hölder continuous on  $[0, L)$ . The hazard rate function  $h$  is locally integrable, but not integrable, on  $[0, L)$ . Thus, if  $h$  is bounded

(either uniformly or on  $[\ell_0, \infty)$  for some  $\ell_0 < \infty$ ) it must be that  $L = \infty$ . Under the first condition above, for any  $x, y, y' \in [0, \infty)$ ,  $y < y'$ ,

$$\left| \frac{G(x+y) - G(x+y')}{1 - G(x)} \right| = \int_y^{y'} \frac{g(x+u)}{1 - G(x)} du \leq \int_y^{y'} h(u) du \leq \|h\|_\infty |y - y'|,$$

and so Assumption 4 is satisfied. On the other hand, if there only exists  $\ell_0 < \infty$  such that  $\sup_{x \in [\ell_0, \infty)} h(x) < \infty$ , but  $G$  is uniformly Hölder continuous on  $[0, \infty)$ , with constant  $C < \infty$  and exponent  $\gamma > 0$ , then straightforward calculations show

$$\left| \frac{G(x+y) - G(x+y')}{1 - G(x)} \right| \leq \max \left( \frac{C}{1 - G(\ell_0)}, \|\mathbb{1}_{[\ell_0, \infty)}(x)h(x)\|_\infty \right) (y - y')^{\gamma \wedge 1},$$

and once again Assumption 4 is satisfied. A relatively easily verifiable sufficient condition for  $G$  to be uniformly Hölder continuous is that  $g \in \mathbb{L}^{1+\alpha}$  for some  $\alpha > 0$  (recall that since  $g$  is a density, we automatically have  $g \in \mathbb{L}^1$ ; thus the latter condition imposes just a little additional regularity on  $g$ ). Indeed, in this case, by Hölder's inequality,

$$|G(x) - G(y)| = \left| \int_x^y g(u) du \right| \leq \|g\|_{\mathbb{L}^{1+\alpha}} (y - x)^{\alpha/(1+\alpha)},$$

and so  $G$  is uniformly Hölder continuous with exponent  $\gamma = \alpha/(1 + \alpha) < 1$ .

Now, given  $s \geq 0$ , recall that  $\widehat{\nu}_s^{(N)}$  represents the (scaled and centered) initial age distribution at time  $s$ , and define

$$\mathcal{S}_t^{\widehat{\nu}_s^{(N)}}(f) \doteq \int_{[0, L)} f(x+t) \frac{1 - G(x+t)}{1 - G(x)} \widehat{\nu}_s^{(N)}(dx), \quad f \in \mathcal{C}_b[0, L], t \geq 0.$$

$\{\mathcal{S}_t^{\widehat{\nu}_0^{(N)}}(f), t \geq 0\}$  plays an important role in the analysis because it arises in the representation for  $\langle f, \widehat{\nu}^{(N)} \rangle$  given in Proposition 6.4. In order to write  $\mathcal{S}_t^{\widehat{\nu}_s^{(N)}}$  more concisely, it will be convenient to introduce a certain family of operators. For  $t \in [0, \infty)$ , define

$$(5.6) \quad (\Phi_t f)(x) \doteq f(x+t) \frac{1 - G(x+t)}{1 - G(x)}, \quad x \in [0, L).$$

Each  $\Phi_t$  maps the space of (bounded) measurable functions on  $[0, L)$  into itself and

$$(5.7) \quad \sup_{t \in [0, \infty)} \|\Phi_t f\|_\infty \leq \|f\|_\infty.$$

Moreover, for future purposes, we also note that  $\{\Phi_t, t \geq 0\}$  defines a semigroup, i.e.,  $\Phi_0 f = f$  and

$$(5.8) \quad \Phi_t(\Phi_s f) = \Phi_{t+s} f, \quad s, t \geq 0,$$

and, recalling the definition (4.19) of the family of operators  $\{\Psi_t, t \geq 0\}$ , it is easily verified that for every bounded, measurable function  $f$  on  $[0, L)$  and  $s, t \geq 0$ ,

$$(5.9) \quad (\Psi_s \Phi_t f)(x, u) = (\Psi_{s+t} f)(x, u), \quad (x, u) \in [0, L) \times [0, s].$$

We can now rewrite  $\mathcal{S}_t^{\widehat{\nu}_s^{(N)}}$  in terms of the operators  $\Phi_t$ ,  $t \geq 0$ , as follows:

$$(5.10) \quad \mathcal{S}_t^{\widehat{\nu}_s^{(N)}}(f) = \langle \Phi_t f, \widehat{\nu}_s^{(N)} \rangle, \quad s, t \geq 0.$$

Since  $G$  is continuous, it is clear from (5.7) that  $\Phi_t f \in \mathcal{C}_b[0, L)$  when  $f \in \mathcal{C}_b[0, L)$  and hence,  $\{\mathcal{S}_t^{\widehat{\nu}_0^{(N)}}(f), f \in \mathcal{C}_b[0, L), t \geq 0\}$  is a well defined stochastic process.

**Remark 5.3.** Since  $\widehat{\nu}_0^{(N)}$  is a signed measure with finite total mass bounded by  $\sqrt{N}$ , it can be easily verified that almost surely for  $f \in \mathbb{H}_1$ , by the norm inequality (1.2),

$$|\langle f, \widehat{\nu}_0^{(N)} \rangle| \leq 2\sqrt{N} \|f\|_\infty \leq 2\sqrt{N} \|f\|_{\mathbb{H}_1}.$$



Moreover, if Assumption 4 holds calculations similar to those in Remark 5.2, the norm inequality (1.2) and the Cauchy-Schwarz inequality show that for  $f \in \mathbb{H}_1$  and  $0 \leq s < t < \infty$ ,

$$|\mathcal{S}_t^{\widehat{\nu}_0^{(N)}}(f) - \mathcal{S}_s^{\widehat{\nu}_0^{(N)}}(f)| \leq C_G(t-s)^{\gamma_G} \|f\|_\infty + \|f\|_{\mathbb{H}_0} (t-s)^{1/2} \leq (2C_G + 1) \|f\|_{\mathbb{H}_1} |t-s|^{\gamma_G \wedge 1/2}.$$

This shows that for every  $N \in \mathbb{N}$ ,  $\mathcal{S}^{\widehat{\nu}_0^{(N)}}$  is a càdlàg process that almost surely takes values in  $\mathbb{H}_{-1}$ .

We now consider the initial conditions. We impose fairly general assumptions on the initial age sequence so as to establish the Markov property for the limit process. As shown in Lemma 9.6, these conditions are consistent in the sense that they are satisfied at any time  $s > 0$  if they are satisfied at time 0. In addition, they are trivially satisfied if each  $N$ -server system starts precisely at the fluid limit, i.e., if  $\widehat{\nu}_0^{(N)} = 0$  for every  $N$ . The reader may prefer to make the latter assumption on first reading to avoid the technicalities in the statement of this assumption. To motivate the form of the assumptions, first note that the total variation of the sequence of finite signed measures  $\{\widehat{\nu}_0^{(N)}\}_{N \in \mathbb{N}}$  tends to infinity as  $N \rightarrow \infty$ , and so it is not reasonable to expect the sequence to converge in the space of finite or Radon measures. Instead, we impose convergence in a slightly different space. As observed in Remark 5.3,  $\widehat{\nu}_0^{(N)}$  can be viewed as an  $\mathbb{H}_{-1}$ -valued (and therefore  $\mathbb{H}_{-2}$ -valued) random element and under Assumption 4,  $\{\mathcal{S}_t^{\widehat{\nu}_0^{(N)}}, t \geq 0\}$  is a càdlàg  $\mathbb{H}_{-1}$ -valued stochastic process and  $\{\mathcal{S}_t^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), t \geq 0\}$  is a càdlàg real-valued process (see Section 1.4.1 for a definition of these spaces).

**Assumption 5.** *There exists an  $\mathbb{R}$ -valued random variable  $\widehat{x}_0$  and a family of random variables  $\{\widehat{\nu}_0(f), f \in \mathcal{AC}_b[0, L]\}$ , all defined on a common probability space, such that*

- (a)  $\widehat{\nu}_0$  admits a version as an  $\mathbb{H}_{-2}$ -valued random element;
- (b) For  $f \in \mathcal{AC}_b[0, L]$ , given

$$(5.11) \quad \mathcal{S}_t^{\widehat{\nu}_0}(f) \doteq \widehat{\nu}_0(\Phi_t f) = \widehat{\nu}_0\left(f(\cdot + t) \frac{1 - G(\cdot + t)}{1 - G(\cdot)}\right), \quad t \geq 0,$$

*$\{\mathcal{S}_t^{\widehat{\nu}_0}(f), f \in \mathcal{AC}_b[0, L], t \geq 0\}$  is a family of random variables such that  $\{\mathcal{S}_t^{\widehat{\nu}_0}, t \geq 0\}$  admits a version as a continuous  $\mathbb{H}_{-2}$ -valued process,  $\{\mathcal{S}_t^{\widehat{\nu}_0}(\mathbf{1}), t \geq 0\}$  admits a version as a continuous  $\mathbb{R}$ -valued process and, for every  $t > 0$  almost surely,  $f \mapsto \mathcal{S}_t^{\widehat{\nu}_0}(f)$  is a measurable mapping from  $\mathcal{AC}_b[0, L] \subset \mathcal{C}_b[0, L] \mapsto \mathbb{R}$  (both equipped with their respective Borel  $\sigma$ -algebras);*

- (c) as  $N \rightarrow \infty$ ,  $(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \Rightarrow (\widehat{x}_0, \widehat{\nu}_0, \mathcal{S}^{\widehat{\nu}_0}, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}))$  in  $\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty)$ .

For some results, we will require the following strengthening of Assumption 5.

**Assumption 5'** *Assumption 5 holds and in addition,*

- (d) Suppose that  $\varphi \in \mathcal{C}_b([0, L] \times [0, \infty))$  is such that for every  $r > 0$ ,  $x \mapsto \varphi(x, r)$  is absolutely continuous and, for every  $T < \infty$ ,  $\varphi_x(\cdot, \cdot)$  is integrable on  $[0, L] \times [0, T]$  and  $x \mapsto \int_0^t \varphi(x, r) dr$  is Hölder continuous. Then  $\mathbb{P}$ -almost surely,  $r \mapsto \widehat{\nu}_0(\Phi_r \varphi(\cdot, r))$  is measurable and for every  $t \geq 0$ ,

$$(5.12) \quad \int_0^t \widehat{\nu}_0(\Phi_r \varphi(\cdot, r)) dr = \widehat{\nu}_0\left(\int_0^t \Phi_r \varphi(\cdot, r) dr\right).$$

Now, let  $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$  be a common probability space that supports the martingale measure  $\widehat{\mathcal{M}}$  introduced in Section 4.3, the standard Brownian motion  $B$  of Remark 5.1, the family of random variables  $\widehat{\nu}_0(f)$ ,  $f \in \mathcal{AC}_b[0, L]$ , and the random variable  $\widehat{x}_0$  of Assumption 5 such that  $\widehat{\mathcal{M}}, B$  and  $(\widehat{x}_0, \widehat{\nu}_0(f), f \in \mathcal{AC}_b[0, L])$  are mutually independent. Let  $\widehat{\mathcal{F}}_0$  be the  $\sigma$ -algebra generated by  $(\widehat{x}_0, \widehat{\nu}_0(f), f \in \mathcal{AC}_b[0, L])$ , and for  $t \geq 0$ , let  $\widehat{\mathcal{F}}_t \doteq \widehat{\mathcal{F}}_0 \vee \sigma(B_s, \widehat{\mathcal{M}}_s, s \in [0, t])$ . Then for  $t \geq 0$ ,  $\mathcal{S}_t^{\widehat{\nu}_0}(f)$ ,

$f \in \mathcal{AC}_b[0, L]$ , (and, in particular,  $\mathcal{S}_t^{\widehat{\nu}_0}(\mathbf{1})$ ) are all well defined  $\widehat{\mathcal{F}}_0$ -measurable random variables, and  $(\widehat{E}_t, \widehat{\mathcal{M}}_t)_{t \geq 0}$  are  $\{\widehat{\mathcal{F}}_t\}_{t \geq 0}$ -adapted stochastic processes. The description of the  $N$ -server model listed prior to Remark 2.1 assumes that for each  $N \in \mathbb{N}$ , given  $R_E^{(N)}(0)$ ,  $E^{(N)}$  is independent of the initial conditions  $\nu_0^{(N)}$  and  $X^{(N)}(0)$ . Together with Assumptions 3, 5 and the fact that  $R_E^{(N)}(0) \rightarrow 0$  almost surely as  $N \rightarrow \infty$ , this implies that as  $N \rightarrow \infty$ ,

$$(5.13) \quad (\widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{S}^{\widehat{\nu}_0}, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}))$$

in  $\mathcal{D}_{\mathbb{R}}[0, \infty) \times \mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty)$ .

**5.2. The Centered Many-Server Map.** We introduce a map, which we refer to as the centered many-server map, which will be used to characterize the limit. Let  $\mathcal{D}_{\mathbb{R}}^0[0, \infty)$  be the subset of functions  $f$  in  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  with  $f(0) = 0$ . The input data for this map lies in the following space:

$$\widehat{\mathcal{I}}_0 \doteq \mathcal{D}_{\mathbb{R}}^0[0, \infty) \times \mathbb{R} \times \mathcal{D}_{\mathbb{R}}[0, \infty).$$

**Definition 5.4. (Centered Many-Server Equations)** Let  $\overline{X} \in \mathcal{D}_{\mathbb{R}_+}[0, \infty)$  be fixed. Given  $(E, x_0, Z) \in \widehat{\mathcal{I}}_0$ , we say that  $(K, X, v) \in \mathcal{D}_{\mathbb{R}}^0[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty)^2$  solves the centered many-server equations (henceforth, abbreviated CMSE) associated with  $\overline{X}$  and  $(E, x_0, Z)$  if and only if for  $t \in [0, \infty)$ ,

$$(5.14) \quad v(t) = Z(t) + K(t) - \int_0^t g(t-s)K(s) ds,$$

$$(5.15) \quad K(t) = E(t) + x_0 - X(t) + v(t) - v(0)$$

and

$$(5.16) \quad v(t) = \begin{cases} X(t) & \text{if } \overline{X}(t) < 1, \\ X(t) \wedge 0 & \text{if } \overline{X}(t) = 1, \\ 0 & \text{if } \overline{X}(t) > 1. \end{cases}$$

Note that this definition automatically requires that  $E(0) = K(0) = 0$ ,  $X(0) = x_0$  and  $Z(0) = v(0)$ , which equals  $x_0$ ,  $x_0 \wedge 0$  or 0, depending on whether  $\overline{X}(0) < 1$ ,  $\overline{X}(0) = 1$  or  $\overline{X}(0) > 1$ . It is shown in Proposition 7.3 that there exists at most one solution to the CMSE for any given input data in  $\widehat{\mathcal{I}}_0$ . When a solution exists, we let  $\Lambda$  denote the associated "centered many-server" mapping (associated with  $\overline{X}$ ) that takes  $(E, x_0, Z) \in \widehat{\mathcal{I}}_0$  to the corresponding solution  $(K, X, v)$ . Let  $\text{dom}(\Lambda)$  denote the domain of  $\Lambda$ , which is defined to be the collection of input data in  $\widehat{\mathcal{I}}_0$  for which a solution to the CMSE exists.

**Remark 5.5.** Suppose  $(K, X, v) \in \Lambda(E, x_0, Z)$  for some  $(E, x_0, Z) \in \widehat{\mathcal{I}}_0$ . Then (5.14) and (5.15) together show that for  $t \geq 0$ ,

$$(5.17) \quad X(t) = x_0 + E(t) + Z(t) - \int_0^t g(t-s) [E(s) + x_0 - v(0) - X(s) + v(s)] ds.$$

Thus, if  $E$ ,  $Z$  and  $g$  are continuous then  $X$  is also continuous. If, in addition, the fluid limit is either subcritical, critical or supercritical then the continuity of  $X$  and (5.16) imply the continuity of  $v$  and, in turn, (5.15) implies the continuity of  $K$ .

The importance of the CMSE stems from the relation

$$(5.18) \quad (\widehat{K}^{(N)}, \widehat{X}^{(N)}, \widehat{\nu}^{(N)}(\mathbf{1})) = \Lambda \left( \widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1}) \right), \quad N \in \mathbb{N},$$

which is established in Lemma 7.2 under the assumption that the fluid limit is either subcritical, critical or supercritical.

**5.3. Statements of Main Results.** The first result of the paper, Theorem 5.6 below, identifies the limit of the sequence  $\{\widehat{X}^{(N)}\}_{N \in \mathbb{N}}$ . Let

$$(5.19) \quad \widehat{Y}_1^{(N)} \doteq (\widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \widehat{\mathcal{M}}^{(N)}, \widehat{\mathcal{H}}^{(N)}, \widehat{\mathcal{H}}^{(N)}(\mathbf{1})),$$

and let  $\widehat{Y}_1$  be the corresponding quantity without the superscript  $N$ , where  $\widehat{E}$ ,  $\widehat{x}_0$ ,  $\widehat{\nu}_0$ ,  $\mathcal{S}^{\widehat{\nu}_0}$ ,  $\mathcal{S}^{\widehat{\nu}_0}(\mathbf{1})$  are as defined in Remark 5.1 and Assumption 5, and  $\widehat{\mathcal{M}}$ ,  $\widehat{\mathcal{H}}$  and  $\widehat{\mathcal{H}}(\mathbf{1})$  are as defined in Section 4.3. Also, define

$$(5.20) \quad \mathcal{Y}_1 \doteq \mathcal{D}_{\mathbb{R}}[0, \infty) \times \mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty) \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty)^2 \times \mathcal{D}_{\mathbb{R}}[0, \infty).$$

**Theorem 5.6.** *Suppose Assumptions 1–5 are satisfied and suppose that the fluid limit is either subcritical, critical or supercritical. Then  $(\widehat{E}, \widehat{x}_0, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}))$  lies in the domain of the centered many-server map  $\Lambda$  and, as  $N \rightarrow \infty$ ,*

$$(5.21) \quad (\widehat{Y}_1^{(N)}, \widehat{X}^{(N)}, \widehat{K}^{(N)}, \langle \mathbf{1}, \widehat{\nu}^{(N)} \rangle) \Rightarrow (\widehat{Y}_1, \widehat{X}, \widehat{K}, \widehat{\nu}(\mathbf{1}))$$

in  $\mathcal{Y}_1 \times \mathcal{D}_{\mathbb{R}}[0, \infty)^3$ , where  $(\widehat{X}, \widehat{K}, \widehat{\nu}(\mathbf{1})) \doteq \Lambda(\widehat{E}, \widehat{x}_0, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}))$  is almost surely continuous. Furthermore, if  $g$  is continuous, then

$$(5.22) \quad \widehat{X}(t) = \widehat{X}_0 + \widehat{E}(t) - \widehat{\mathcal{M}}_t(\mathbf{1}) - \widehat{D}(t), \quad t \in [0, \infty),$$

where

$$(5.23) \quad \widehat{D}(t) \doteq \widehat{\nu}_0(\mathbf{1}) - \mathcal{S}_t^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{M}}_t(\mathbf{1}) + \widehat{\mathcal{H}}_t(\mathbf{1}) + \int_0^t g(t-s) \widehat{K}(s) ds.$$

The proof of Theorem 5.6 is presented in Section 9.1. In addition to establishing the representation (5.18), the key elements of the proof involve showing the convergence  $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$ , which is carried out in Corollary 8.7, and establishing continuity of the centered many-server map  $\Lambda$  and another auxiliary map  $\Gamma$ , which are established in Proposition 7.3 and Lemma 7.1, respectively.

We now establish a more general convergence result for the pair  $\{(\widehat{X}^{(N)}, \widehat{\nu}^{(N)})\}_{N \in \mathbb{N}}$ , which automatically yields convergence of several functionals of the process. The proof of this result is also given in Section 9.1. With  $\widehat{K}$  equal to the limit process obtained in Theorem 5.6, we define for all bounded and absolutely continuous  $f$ ,

$$(5.24) \quad \widehat{\nu}_t(f) \doteq \mathcal{S}_t^{\widehat{\nu}_0}(f) + f(0) \widehat{K}(t) + \int_0^t \widehat{K}(s) f'(t-s) (1 - G(t-s)) ds \\ - \int_0^t \widehat{K}(s) g(t-s) f(t-s) ds - \widehat{\mathcal{H}}_t(f).$$

Note that the first term on the right-hand side is well defined by the discussion following Assumption 5 (see also Lemma B.1), the next three terms are well defined because  $\widehat{K}$  is continuous and  $f'$ ,  $(1 - G)$ ,  $g$  and  $f$  are all locally integrable, and the last term is well defined since  $\widehat{\mathcal{H}}_t(f) = \widehat{\mathcal{M}}_t(\Psi_t f)$  and the continuity and boundedness of  $f$  implies  $\Psi_t f \in \mathcal{C}_b([0, L] \times \mathbb{R}_+)$ .

**Theorem 5.7.** *Suppose Assumptions 1–5 are satisfied, the fluid limit is either subcritical, critical or supercritical and  $g$  is continuous. Then, as  $N \rightarrow \infty$ ,*

$$(5.25) \quad (\widehat{Y}_1^{(N)}, \widehat{X}^{(N)}, \widehat{K}^{(N)}, \widehat{\nu}^{(N)}) \Rightarrow (\widehat{Y}_1, \widehat{X}, \widehat{K}, \widehat{\nu})$$

in  $\mathcal{Y} \doteq \mathcal{Y}_1 \times \mathcal{D}_{\mathbb{R}}[0, \infty)^2 \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty)$ .

A main focus of this paper is to show that the approximating process is a tractable process, thus demonstrating the usefulness of the approximation theorem obtained. The next two theorems show that this is indeed the case under some additional regularity conditions on the hazard rate  $h$ . First, in Theorem 5.8 we show that  $\{\widehat{X}_t, \widehat{\mathcal{F}}_t, t \geq 0\}$  is also a semimartingale. By Itô's formula this enables the description of the evolution of a large class of functionals of the process. The proof of Theorem 5.8 is given in Section 9.2.

**Theorem 5.8.** *Suppose that Assumptions 1, 3 and 5' are satisfied, the fluid limit is either subcritical, critical or supercritical and  $h$  is bounded and absolutely continuous. If  $(\widehat{K}, \widehat{X}, \widehat{\nu})$  is the limit process of Theorem 5.7, then  $\widehat{X}$  and  $\widehat{K}$  are semimartingales with decompositions  $\widehat{X} = \widehat{X}(0) + M^X + C^X$  and  $\widehat{K} = M^K + C^K$ , respectively, where*

$$M^X(t) = \int_0^t \sigma(s) dB(s) - \widehat{\mathcal{M}}_t(\mathbf{1}), \quad C^X(t) = - \int_0^t \beta(s) ds - \int_0^t \widehat{\nu}_s(h) ds, \quad t \geq 0,$$

and if  $\overline{X}$  is subcritical, then  $\widehat{K} = \widehat{E}$  and so

$$M^K(t) = \int_0^t \sigma(s) dB(s), \quad C^K(t) = - \int_0^t \beta(s) ds, \quad t \geq 0,$$

if  $\overline{X}$  is supercritical, then

$$M^K(t) = \widehat{\mathcal{M}}_t(\mathbf{1}), \quad C^K(t) = \int_0^t \widehat{\nu}_s(h) ds, \quad t \geq 0,$$

and if  $\overline{X}$  is critical, then

$$M^K(t) = \int_0^t \mathbb{1}_{\{\widehat{X}(s) \leq 0\}} \sigma(s) dB_s + \int_0^t \mathbb{1}_{\{\widehat{X}(s) > 0\}} d\widehat{\mathcal{M}}_s(\mathbf{1}), \quad t \geq 0,$$

and

$$C^K(t) = - \int_0^t \beta(s) \mathbb{1}_{\{\widehat{X}(s) \leq 0\}} ds + \int_0^t \mathbb{1}_{\{\widehat{X}(s) > 0\}} \widehat{\nu}_s(h) ds + \frac{1}{2} L_0^{\widehat{X}}(t), \quad t \geq 0,$$

where,  $L_0^{\widehat{X}}(t)$  is the local time of  $\widehat{X}$  at zero on the interval  $[0, t]$ . Moreover, for each  $t > 0$  and  $f \in \mathcal{AC}_b[0, L]$ ,  $\widehat{\nu}_t(f)$  admits the alternative representation

$$(5.26) \quad \widehat{\nu}_t(f) = \mathcal{S}_t^{\widehat{\nu}_0}(f) + \int_0^t f(t-s)(1-G(t-s)) d\widehat{K}(s) - \widehat{\mathcal{H}}_t(f),$$

where the second term is the stochastic convolution integral with respect to the semimartingale  $\widehat{K}$ .

**Remark 5.9.** If for  $f \in \mathcal{C}_b[0, L]$ ,  $\{\mathcal{S}_t^{\widehat{\nu}_0}(f), t \geq 0\}$  is a well defined stochastic process then  $\{\widehat{\nu}_t(f), t \geq 0\}$  is also a well defined stochastic process given by the right-hand side of (5.26). Moreover, under a slight strengthening of the conditions of Theorem 5.8 (specifically, of Assumption 5), we can in fact show convergence for a slightly larger class of functions than those in  $\mathbb{H}_2$ . More precisely, if for any bounded, Hölder continuous  $f$ ,  $\{\mathcal{S}_t^{\widehat{\nu}_0}(f), t \geq 0\}$  defined in (5.11) is a well defined continuous stochastic process and, as  $N \rightarrow \infty$ ,  $(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}(f), \mathcal{S}_0^{\widehat{\nu}_0^{(N)}}(f), \mathcal{S}_0^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \Rightarrow (\widehat{x}_0, \widehat{\nu}_0, \mathcal{S}_0^{\widehat{\nu}_0}(f), \mathcal{S}_0^{\widehat{\nu}_0}(\mathbf{1}))$  in  $\mathbb{R}^2 \times \mathcal{D}_{\mathbb{R}}[0, \infty)^2$ , then  $\widehat{\nu}(f)$  is also a continuous process and, as  $N \rightarrow \infty$ ,  $\widehat{\nu}^{(N)}(f) \Rightarrow \widehat{\nu}(f)$  in  $\mathcal{D}_{\mathbb{R}}[0, \infty)$ . A brief justification of this assertion is provided at the end of Section 9.2. By the independence assumptions of the model, the above conditions automatically imply the joint convergence

$$(5.27) \quad (\widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}(f), \mathcal{S}_0^{\widehat{\nu}_0^{(N)}}(f), \mathcal{S}_0^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0(f), \mathcal{S}_0^{\widehat{\nu}_0}(f), \mathcal{S}_0^{\widehat{\nu}_0}(\mathbf{1}))$$

in  $\mathcal{D}_{\mathbb{R}}[0, \infty) \times \mathbb{R}^2 \times \mathcal{D}_{\mathbb{R}}[0, \infty)^2$ .

We now show that the limiting process  $(\widehat{K}, \widehat{X}, \widehat{\nu})$  described in Theorem 5.6 can alternatively be characterized as the unique solution to a stochastic partial differential equation (SPDE), coupled with an Itô diffusion equation, and also satisfies a strong Markov property. We first introduce the SPDE, which we refer to as the stochastic age equation. In the definition of the stochastic age equation given below,  $h$  is the hazard rate function of the service distribution and  $\widehat{\nu}_0$ ,  $\widehat{\mathcal{M}}$  and  $\widehat{K}$  are the limit processes defined on the filtered probability space  $(\widehat{\Omega}, \widehat{\mathcal{F}}, \{\widehat{\mathcal{F}}_t\}, \widehat{\mathbb{P}})$ , as specified in Theorem 5.8.

**Definition 5.10. (Stochastic Age Equation)** Given  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$  defined on the filtered probability space  $(\widehat{\Omega}, \widehat{\mathcal{F}}, \{\widehat{\mathcal{F}}_t\}, \widehat{\mathbb{P}})$ ,  $\nu = \{\nu_t, t \geq 0\}$  is said to be a strong solution to the stochastic age equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$  if and only if for every  $f \in \mathcal{AC}_b[0, L]$ ,  $\nu_t(f)$  is an  $\widehat{\mathcal{F}}_t$ -measurable random variable for  $t > 0$ ,  $s \mapsto \nu_s(f)$  is almost surely measurable on  $[0, \infty)$ ,  $\{\nu_t, t \geq 0\}$  admits a version as a continuous,  $\mathbb{H}_{-2}$ -valued process and  $\mathbb{P}$ -almost surely, for every  $\varphi \in C_b^{1,1}([0, L] \times \mathbb{R})$  such that  $\varphi_t(\cdot, s) + \varphi_x(\cdot, s)$  is Lipschitz continuous for every  $s$ , and every  $t \in [0, \infty)$ ,

$$(5.28) \quad \begin{aligned} \nu_t(\varphi(\cdot, t)) &= \nu_0(\varphi(\cdot, 0)) + \int_0^t \nu_s(\varphi_x(\cdot, s) + \varphi_s(\cdot, s) - \varphi(\cdot, s)h(\cdot)) ds \\ &\quad - \int \int_{[0, L] \times [0, t]} \varphi(x, s) \widehat{\mathcal{M}}(dx, ds) + \int_0^t \varphi(0, s) d\widehat{K}_s. \end{aligned}$$

**Theorem 5.11.** Suppose Assumptions 1, 3 and 5' are satisfied, the fluid limit is either subcritical, critical or supercritical and  $h$  is absolutely continuous and bounded. Given the limit process  $(\widehat{X}, \widehat{\nu})$  of Theorem 5.7, the following assertions are true:

- (1) If the density  $g'$  of  $g$  lies in  $\mathbb{L}_{loc}^2[0, L] \cup L_{loc}^\infty[0, L]$  then  $\{\widehat{\nu}_t, \widehat{\mathcal{F}}_t, t \geq 0\}$  is the unique strong solution to the stochastic age equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ ;
- (2) if  $g'/(1-G)$  is bounded then  $\{(\widehat{X}_t, \widehat{\nu}_t, \mathcal{S}^{\widehat{\nu}_t}(\mathbf{1})), \widehat{\mathcal{F}}_t, t \geq 0\}$  is an  $\mathbb{R} \times \mathbb{H}_{-2} \times C_{\mathbb{R}}[0, \infty)$ -valued strong Markov process.

The characterization in terms of the stochastic age equation is established in Section 9.3 and the proof of the strong Markov property is given in Section 9.4. It is more natural to expect the process  $\{(\widehat{X}_t, \widehat{\nu}_t), \widehat{\mathcal{F}}_t, t \geq 0\}$  to be strong Markov with state  $\mathbb{R} \times \mathbb{H}_{-2}$ . However, due to technical reasons (see Remark 9.7 for a more detailed explanation) it was necessary to append an additional component to obtain a Markov process. We expect that this additional component should not pose too much of a problem in applications of this result.

**Remark 5.12.** Elementary calculations show that the boundedness assumptions imposed on  $g/(1-G)$  and  $g'/(1-G)$  in Theorem 5.11 (which, in particular, imply Assumption 4) are satisfied by many continuous service distributions of interest (with finite mean, normalized to have mean one) including the families of lognormal, Weibull, logistic and phase type distributions, the Gamma( $a, b$ ) distribution with shape parameter  $a-1$  or  $a \geq 2$  (and corresponding rate parameter  $b = 1/a$  to produce a mean one distribution), the Pareto distribution with shape parameter  $a > 1$  (and corresponding scale parameter  $b = (a-1)/a$  so the mean equals one), the inverted Beta( $a, b$ ) distribution when  $a > 2$  (and  $b = a+1$  so that the mean is equal to one). Note that the mean one exponential distribution is clearly also included as a special case of the Weibull and Gamma distributions.

**5.4. Corollaries of the Main Results.** As an important corollary of Theorem 5.8, when the fluid limit is critical, the limiting (scaled and centered) total number in system  $\widehat{X}$  can be characterized as an Itô diffusion.

**Corollary 5.13.** Suppose Assumption 1, Assumption 3(a) with  $\bar{\lambda} = 1$  and Assumption 5' are satisfied and  $h$  is bounded and absolutely continuous. If  $\bar{x}_0 = 1$  and  $\bar{\nu}_0(dx) = (1-G(x))dx$ , the equilibrium age measure in the critical fluid limit, then  $\widehat{X}$  satisfies the following Itô diffusion equation:

$$(5.29) \quad d\widehat{X}(t) = \widehat{x}_0 + \sigma B(t) - \widehat{M}_1(t) - \beta t - \int_0^t \langle h, \widehat{\nu}_s \rangle ds,$$

where  $\widehat{M}_1$  is a Brownian motion independent of  $B$ .

The asymptotic independence of  $\widehat{\mathcal{M}}_1$  and  $B$  in the last corollary follows from Proposition 8.4. In the particular case of an exponential service distribution, this allows us to immediately recover the form of the limiting diffusion obtained in the seminal paper of Halfin and Whitt [15].

**Corollary 5.14.** *Suppose  $G$  is the exponential distribution with parameter 1. Suppose Assumption 1 holds with  $\bar{\nu}_0(dx) = (1 - G(x))dx$  and  $\bar{x}_0 = 1$ , Assumption 3(a) holds with  $\bar{\lambda} = 1$  and Assumption 5' is satisfied. Then  $\widehat{X}$  is the unique strong solution to the stochastic differential equation*

$$(5.30) \quad \widehat{X}(t) = \widehat{x}_0 + \sqrt{1 + \sigma^2}W(t) - \beta t - \int_0^t \widehat{X}^-(s) ds$$

where  $W$  is a standard Brownian motion.

*Proof.* When  $G$  is the exponential distribution,  $h \equiv \mathbf{1}$  and therefore

$$\int_0^t \langle h, \widehat{\nu}_s \rangle ds = \int_0^t \langle \mathbf{1}, \widehat{\nu}_s \rangle ds = \int_0^t \left( \widehat{X}(s) \wedge 0 \right) ds,$$

where the last equality uses the relations (5.18), (5.16) and the fact that  $\bar{X} \equiv \mathbf{1}$ . By the independence of  $B$  and  $\widehat{\mathcal{M}}_1$ ,  $\sigma B - \widehat{\mathcal{M}}_1$  has the same distribution as  $\sqrt{1 + \sigma^2}W$ , where  $W$  is a standard Brownian motion. Substituting this back into (5.29), which is applicable since the hazard rate  $h$  of the exponential distribution is trivially bounded and absolutely continuous, we obtain (5.30). The Lipschitz continuity of the drift coefficient  $x \mapsto -\beta - x^-$  guarantees that the stochastic differential equation (5.30) has a unique strong solution.  $\square$

**Remark 5.15. (Insensitivity Result)** As a comparison of (5.29) and (5.30) reveals, under the same assumptions on the arrival process as in Corollary 5.14, the dynamical equation for  $\widehat{X}$  for general service distributions is remarkably close to the exponential case. Indeed, the ‘‘diffusion’’ coefficient is the same in both cases (and is equal to  $\sqrt{1 + \sigma^2}$ ), but the difference is that in the case of general service distributions, the drift is an  $\{\widehat{\mathcal{F}}_t\}$ -adapted process that could in general depend on the past, and not just on  $\widehat{X}_t$ , so that the resulting process is no longer Markovian.

## 6. REPRESENTATION OF THE SYSTEM DYNAMICS

In Section 6.1 we present a succinct characterization of the dynamics of the centered state process and then use that in Section 6.2 to derive an alternative representation for the centered age process.

**6.1. A Succinct Characterization of the Dynamics.** We first recall the description of the dynamics of the  $N$ -server system that was established by Kaspi and Ramanan [22].

**Proposition 6.1.** *The process  $(X^{(N)}, \nu^{(N)})$  almost surely satisfies the following coupled set of equations: for  $\varphi \in \mathcal{C}_b^{1,1}([0, L] \times \mathbb{R}_+)$  and  $t \in [0, \infty)$ ,*

$$(6.1) \quad \begin{aligned} \langle \varphi(\cdot, t), \nu_t^{(N)} \rangle &= \langle \varphi(\cdot, 0), \nu_0^{(N)} \rangle + \int_0^t \langle \varphi_x(\cdot, s) + \varphi_s(\cdot, s), \nu_s^{(N)} \rangle ds \\ &\quad - \int_0^t \langle \varphi(\cdot, s) h(\cdot), \nu_s^{(N)} \rangle ds - \mathcal{M}_t^{(N)}(\varphi) \\ &\quad + \int_{[0, t]} \varphi(0, s) dK^{(N)}(s), \end{aligned}$$

$$(6.2) \quad X^{(N)}(t) = X^{(N)}(0) + E^{(N)}(t) - \int_0^t \langle h, \nu_s^{(N)} \rangle ds - \mathcal{M}_t^{(N)}(\mathbf{1})$$

and

$$(6.3) \quad N - \langle \mathbf{1}, \nu_t^{(N)} \rangle = [N - X^{(N)}(t)]^+,$$

where  $K^{(N)}$  satisfies

$$(6.4) \quad \begin{aligned} K^{(N)} &= \langle \mathbf{1}, \nu^{(N)} \rangle - \langle \mathbf{1}, \nu_0^{(N)} \rangle + \int_0^\cdot \langle h, \nu_s^{(N)} \rangle ds + \mathcal{M}^{(N)}(\mathbf{1}) \\ &= \widehat{X}^{(N)}(0) + E^{(N)} - X^{(N)} + \langle \mathbf{1}, \nu^{(N)} \rangle - \langle \mathbf{1}, \nu_0^{(N)} \rangle. \end{aligned}$$

*Proof.* This is essentially a direct consequence of Theorem 5.1 of Kaspi and Ramanan [22]. Indeed, by subtracting and adding  $A_\varphi^{(N)}$  on the right-hand side of equations (5.4) and (5.5) in [22], and then using (4.3) above and the fact that  $M_\varphi^{(N)}$  is indistinguishable from  $\mathcal{M}^{(N)}(\varphi)$  (see Remark 4.3) one obtains (6.1) and (6.2), respectively. Equation (6.3) coincides with equation (5.6) in [22]. Finally, the first equality in (6.4) follows from (2.6) of [22] and (4.3) above, whereas the second equality in (6.4) follows from (6.2).  $\square$

Combining the characterizations of the  $N$ -server system and the fluid limit given in Proposition 6.1 and Theorem 3.2, respectively, we obtain the following representation for the centered diffusion-scaled state dynamics. In what follows, recall the centered, diffusion-scaled quantities defined in (5.1) and (4.11).

**Proposition 6.2.** *For each  $N \in \mathbb{N}$ , the process  $(\widehat{X}^{(N)}, \widehat{\nu}^{(N)})$  almost surely satisfies the following coupled set of equations: for every  $\varphi \in \mathcal{C}_c^{1,1}([0, L] \times \mathbb{R}_+)$  and  $t \in [0, \infty)$ ,*

$$(6.5) \quad \begin{aligned} \langle \varphi(\cdot, t), \widehat{\nu}_t^{(N)} \rangle &= \langle \varphi(\cdot, 0), \widehat{\nu}_0^{(N)} \rangle + \int_0^t \langle \varphi_x(\cdot, s) + \varphi_s(\cdot, s), \widehat{\nu}_s^{(N)} \rangle ds \\ &\quad - \int_0^t \langle \varphi(\cdot, s) h(\cdot), \widehat{\nu}_s^{(N)} \rangle ds - \widehat{\mathcal{M}}_t^{(N)}(\varphi) \\ &\quad + \int_{[0, t]} \varphi(0, s) d\widehat{K}^{(N)}(s), \end{aligned}$$

$$(6.6) \quad \widehat{X}^{(N)}(t) = \widehat{X}^{(N)}(0) + \widehat{E}^{(N)}(t) - \int_0^t \langle h, \widehat{\nu}_s^{(N)} \rangle ds - \widehat{\mathcal{M}}_t^{(N)}(\mathbf{1}),$$

and

$$(6.7) \quad \langle \mathbf{1}, \widehat{\nu}_t^{(N)} \rangle = \begin{cases} \widehat{X}^{(N)}(t) \wedge \sqrt{N}(1 - \overline{X}(t)) & \text{if } \overline{X}(t) < 1, \\ \widehat{X}^{(N)}(t) \wedge 0 & \text{if } \overline{X}(t) = 1, \\ \sqrt{N}(\overline{X}^{(N)}(t) - 1) \wedge 0 & \text{if } \overline{X}(t) > 1, \end{cases}$$

where

$$(6.8) \quad \begin{aligned} \widehat{K}^{(N)} &= \langle \mathbf{1}, \widehat{\nu}^{(N)} \rangle - \langle \mathbf{1}, \widehat{\nu}_0 \rangle + \int_0^\cdot \langle h, \nu_s^{(N)} \rangle ds + \widehat{\mathcal{M}}^{(N)}(\mathbf{1}) \\ &= \widehat{x}_0^{(N)} + \widehat{E}^{(N)} - \widehat{X}^{(N)} + \langle \mathbf{1}, \widehat{\nu}^{(N)} \rangle - \langle \mathbf{1}, \widehat{\nu}_0^{(N)} \rangle. \end{aligned}$$

*Proof.* Equation (6.5) is obtained by dividing each side of the equation (6.1) by  $N$ , subtracting the corresponding side of (3.6) from it and multiplying the resulting quantities by  $\sqrt{N}$ . In an exactly analogous fashion, equation (6.6) can be derived from equations (6.2) and (3.7), and equation (6.8) can be obtained from equations (6.4), (3.7) and (3.9). It only remains to justify the relation in (6.7). Dividing (6.3) by  $N$ , subtracting it from (3.8) and multiplying this difference by  $\sqrt{N}$ , we obtain

$$(6.9) \quad \langle \mathbf{1}, \widehat{\nu}_t^{(N)} \rangle = \sqrt{N} \left( [1 - \overline{X}(t)]^+ - [1 - \overline{X}^{(N)}(t)]^+ \right).$$

If  $\overline{X}(t) < 1$  then  $[1 - \overline{X}(t)]^+ = (1 - \overline{X}(t))$  and so the right-hand side above equals

$$\begin{cases} \sqrt{N}(1 - \overline{X}(t) - (1 - \overline{X}^{(N)}(t))) = \widehat{X}^{(N)}(t) & \text{if } \overline{X}^{(N)}(t) < 1, \\ \sqrt{N}(1 - \overline{X}(t)) & \text{if } \overline{X}^{(N)}(t) \geq 1, \end{cases}$$

which can be expressed as  $\widehat{X}^{(N)}(t) \wedge \sqrt{N}(1 - \overline{X}(t))$ . On the other hand, if  $\overline{X}(t) = 1$  then  $[1 - \overline{X}(t)]^+ = 0$  and the right-hand side of (6.9) equals

$$-\sqrt{N}[1 - \overline{X}^{(N)}(t)]^+ = -\sqrt{N}[\overline{X}(t) - \overline{X}^{(N)}(t)]^+ = \widehat{X}^{(N)}(t) \wedge 0.$$

Lastly, if  $\overline{X}(t) > 1$  then  $[1 - \overline{X}(t)]^+ = 0$  and so the right-hand side of (6.9) reduces to  $\sqrt{N}(\overline{X}^{(N)}(t) - 1) \wedge 0$ , and (6.7) follows.  $\square$

**Remark 6.3.** We describe conditions under which, for large  $N$ , the non-idling condition (6.7) can be further simplified and written purely in terms of  $\langle \mathbf{1}, \widehat{\nu}^{(N)} \rangle$  and  $\widehat{X}^{(N)}$ . Let  $\Omega^*$  be the set of full  $\mathbb{P}$ -measure on which the fluid limit theorem (Theorem 3.2) holds. Fix  $\omega \in \Omega^*$  (and henceforth suppress the dependence on  $\omega$ ) and let  $t \in [0, \infty)$  be a continuity point of the fluid limit. If  $\overline{X}(t) < 1$  then by Theorem 3.2 there exists  $N_0 = N_0(\omega, t) < \infty$  such that for all  $N \geq N_0$ ,  $\overline{X}^{(N)}(t) < 1$  and so

$$\widehat{X}^{(N)}(t) = \sqrt{N}(\overline{X}^{(N)}(t) - \overline{X}(t)) \leq \sqrt{N}(1 - \overline{X}(t)).$$

On the other hand, if  $\overline{X}(t) > 1$  then there exists  $N_0 = N_0(\omega, t) < \infty$  such that for all  $N \geq N_0$ ,  $\overline{X}^{(N)}(t) > 1$  and hence,

$$\sqrt{N}(\overline{X}^{(N)}(t) - 1) \geq 0.$$

Therefore, for any  $t \in [0, \infty)$  there exists  $N_0 = N_0(\omega, t) < \infty$  such that for all  $N \geq N_0$ ,

$$(6.10) \quad \langle \mathbf{1}, \widehat{\nu}_t^{(N)} \rangle = \begin{cases} \widehat{X}^{(N)}(t) & \text{if } \overline{X}(t) < 1, \\ \widehat{X}^{(N)}(t) \wedge 0 & \text{if } \overline{X}(t) = 1, \\ 0 & \text{if } \overline{X}(t) > 1. \end{cases}$$

Now, suppose the fluid limit  $\overline{X}$  is continuous and for some  $T < \infty$ , the fluid is subcritical on  $[0, T]$  in the sense of Definition 3.3. Then  $N_0$  can clearly be chosen uniformly in  $t \in [0, T]$  and so there exists  $N_0 = N_0(\omega, T) < \infty$  such that for all  $N \geq N_0(\omega, T)$ ,

$$\langle \mathbf{1}, \widehat{\nu}_t^{(N)} \rangle = \widehat{X}^{(N)}(t), \quad t \in [0, T].$$

An analogous statement holds for the supercritical case and (trivially) for the critical case.

**6.2. A Useful Representation.** Equations (6.1) and (6.5) for the age and (scaled) centered age processes, respectively, in the  $N$ -server system have a form that is analogous to the deterministic integral equation (3.6) that describes the dynamics of the age process in the fluid limit, except that they contain an additional stochastic integral term. Indeed, all three equations fall under the framework of the so-called abstract age equation introduced in Definition 4.9 of Kaspi and Ramanan [22]. Representations for solutions to the abstract age equation were obtained in Proposition 4.16 of [22]. In Corollary 6.4 below, this result is applied to obtain explicit representations for the age and centered age processes in the  $N$ -server system. Not surprisingly, these representations are similar to the corresponding representation (3.10) for solutions to the fluid age equation, except that they contain an additional stochastic integral term.

We now state the representation result, which is easily deduced from Proposition 4.16 of Kaspi and Ramanan [22]; the details of the proof are deferred to Appendix C. For conciseness of notation, for  $N \in \mathbb{N}$  and continuous  $f$ , we define

$$(6.11) \quad \widehat{\mathcal{K}}_t^{(N)}(f) \doteq \int_{[0,t]} (1 - G(t-s))f(t-s) d\widehat{K}^{(N)}(s), \quad t \in [0, \infty).$$

By applying integration by parts to the right-hand side of (6.11) and using the fact that  $\widehat{K}^{(N)}$  has jumps at at most a countable number of points, we see that for absolutely continuous  $f$ ,

$$(6.12) \quad \widehat{\mathcal{K}}_t^{(N)}(f) = f(0)\widehat{K}^{(N)}(t) + \int_0^t \widehat{K}^{(N)}(s)\xi_f(t-s) ds,$$



where

$$(6.13) \quad \xi_f \doteq (f(1-G))' = f'(1-G) - fg.$$

Also, recall the definition of the process  $\mathcal{S}_0^{\widehat{\nu}_0^{(N)}}$  given in (5.10).

**Proposition 6.4.** *For every  $N \in \mathbb{N}$ ,  $f \in \mathcal{C}_b[0, L]$  and  $t \in [0, \infty)$ ,*

$$(6.14) \quad \langle f, \nu_t^{(N)} \rangle = \mathcal{S}_t^{\nu_0^{(N)}}(f) - \mathcal{H}_t^{(N)}(f) + \mathcal{K}_t^{(N)}(f),$$

and, likewise,

$$(6.15) \quad \langle f, \widehat{\nu}_t^{(N)} \rangle = \mathcal{S}_t^{\widehat{\nu}_0^{(N)}}(f) - \widehat{\mathcal{H}}_t^{(N)}(f) + \widehat{\mathcal{K}}_t^{(N)}(f).$$

**Remark 6.5.** For subsequent use, we make the simple observation that on substituting  $\varphi = \mathbf{1}$  in (6.5) and subtracting it from (6.15), with  $f = \mathbf{1}$ , then rearranging terms and using (6.12) and the fact that  $\xi_{\mathbf{1}} = (1-G)' = -g$  by (6.13), we obtain for every  $N \in \mathbb{N}$  and  $t > 0$ ,

$$(6.16) \quad \int_0^t \langle h, \widehat{\nu}_s^{(N)} \rangle ds = \langle \mathbf{1}, \widehat{\nu}_0^{(N)} \rangle - \mathcal{S}_t^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{M}}_t^{(N)}(\mathbf{1}) + \widehat{\mathcal{H}}_t^{(N)}(\mathbf{1}) \\ + \int_0^t g(t-s) \widehat{K}^{(N)}(s-) ds.$$

## 7. CONTINUITY PROPERTIES

In Section 7.1 we establish continuity of the mapping that takes  $\widehat{K}^{(N)}$  to  $\widehat{\mathcal{K}}^{(N)}$  and in Section 7.2 we establish continuity of the centered many-server map  $\Lambda$ , which in particular shows that both  $\widehat{K}^{(N)}$  and  $\widehat{X}^{(N)}$  are obtained as continuous mappings of the initial data and  $\widehat{\mathcal{H}}^{(N)}$ .

**7.1. Continuity of an Auxiliary Map.** Here, we establish the continuity of a mapping related to the convolution integral  $\widehat{\mathcal{K}}^{(N)}$  defined in (6.11). Given any (deterministic) càdlàg function  $K$ , for absolutely continuous functions  $f$  we define

$$(7.1) \quad \mathcal{K}_t(f) \doteq f(0)K(t) + \int_0^t K(u)\xi_f(t-u) du,$$

where  $\xi_f = (f(1-G))'$ , as defined in (6.13). Since  $K$ ,  $g$  and  $f'$  are all locally integrable, for each  $t > 0$ ,  $\mathcal{K}_t$  is a well defined linear functional on the space of absolutely continuous functions. Moreover, from elementary properties of convolutions, it is clear that for any absolutely continuous  $f$ , if  $K$  is càdlàg (respectively, continuous) then so is  $\mathcal{K}(f)$ . In Lemma 7.1 below, we show that  $\mathcal{K}$  is in fact a càdlàg  $\mathbb{H}_{-2}$ -valued function and the mapping from  $K$  to  $\mathcal{K}$ , which we denote by  $\Gamma$ , is continuous. Note that by (6.12)  $\widehat{\mathcal{K}}^{(N)} = \Gamma(\widehat{K}^{(N)})$  for  $N \in \mathbb{N}$ . The continuity of  $\Gamma$  is used in the proof of Theorem 5.7 to establish convergence of  $\widehat{\mathcal{K}}^{(N)}$  to the analogous limit quantity  $\widehat{\mathcal{K}}$ , defined for absolutely continuous  $f$ , by

$$(7.2) \quad \widehat{\mathcal{K}}_t(f) \doteq f(0)\widehat{K}(t) + \int_0^t \widehat{K}(s)\xi_f(t-s) ds, \quad t \in [0, \infty).$$

The third property in Lemma 7.1 below is used in the proof of the strong Markov property in Section 9.4.

**Lemma 7.1.** *Suppose  $\Gamma$  is the map that takes  $K$  to the linear functional  $\mathcal{K}$  defined in (7.1). If  $g$  is continuous the following three properties are satisfied:*

- (1) *If  $K \in \mathcal{D}_{\mathbb{R}}[0, \infty)$  (respectively,  $\mathcal{C}_{\mathbb{R}}[0, \infty)$ ) then  $\mathcal{K} \in \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty)$  (respectively,  $\mathcal{C}_{\mathbb{H}_{-2}}[0, \infty)$ ).*
- (2)  *$\Gamma$  is a continuous map from  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  to  $\mathcal{D}_{\mathbb{H}_{-2}}[0, \infty)$ , when both domain and range are either both equipped with the topology of uniform convergence on compact sets or both equipped with the Skorokhod topology. Likewise, the map from  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  to itself that takes  $K \mapsto \mathcal{K}(\mathbf{1})$  is also continuous with respect to the Skorokhod topology on  $\mathcal{D}_{\mathbb{R}}[0, \infty)$ .*

- (3) If  $K \in \mathcal{C}_{\mathbb{R}}[0, \infty)$  then, for any  $t \in [0, \infty)$ , the real-valued function  $u \mapsto \mathcal{K}_t(\Phi_u \mathbf{1})$  on  $[0, \infty)$  is continuous and the map from  $\mathcal{C}_{\mathbb{R}}[0, \infty)$  to itself that takes  $K$  to this function is continuous (with respect to the uniform topology).

*Proof.* Let  $g$  be continuous. We first derive a general inequality (see (7.7) below) that is then used to prove both properties 1 and 2. Fix  $K, K^{(n)} \in \mathcal{D}_{\mathbb{R}}[0, \infty)$ ,  $T < \infty$  and  $t, \tau^{(n)}(t) \in [0, T]$  with  $\delta^{(n)}(t) \doteq |t - \tau^{(n)}(t)|$ ,  $n \in \mathbb{N}$ . Also, let  $\mathcal{K} \doteq \Gamma(K)$  and  $\mathcal{K}^{(n)} \doteq \Gamma(K^{(n)})$ ,  $n \in \mathbb{N}$ . For  $f \in \mathbb{H}_2$ , we can write

$$(7.3) \quad \mathcal{K}_{\tau^{(n)}(t)}^{(n)}(f) - \mathcal{K}_t(f) = f(0) \left( K^{(n)}(\tau^{(n)}(t)) - K(t) \right) + \sum_{i=1}^3 \Delta_i^{(n)}(t),$$

where

$$\begin{aligned} \Delta_1^{(n)}(t) &\doteq \int_0^{t \wedge \tau^{(n)}(t)} K(u) (\xi_f(t-u) - \xi_f(\tau_n(t) - u)) du, \\ \Delta_2^{(n)}(t) &\doteq \int_0^{t \wedge \tau^{(n)}(t)} \left( K(u) - K^{(n)}(u) \right) \xi_f(\tau^{(n)}(t) - u) du, \\ \Delta_3^{(n)}(t) &\doteq \int_{t \wedge \tau^{(n)}(t)}^t K(u) \xi_f(t-u) du + \int_{t \wedge \tau^{(n)}(t)}^{\tau^{(n)}(t)} K^{(n)}(u) \xi_f(\tau^{(n)}(t) - u) du. \end{aligned}$$

To bound the above terms, first note that by the inequality  $(1 - G) \leq 1$ , repeated application of the Cauchy-Schwarz inequality and Tonelli's theorem we obtain

$$\begin{aligned} &\int_0^s |f'(t-u)(1-G(t-u)) - f'(s-u)(1-G(s-u))| du \\ &\leq \int_0^s |f'(t-u)| (G(t-u) - G(s-u)) du + \int_0^s |f'(t-u) - f'(s-u)| du \\ &\leq w_G(|t-s|) T^{1/2} \|f'\|_{\mathbb{H}_0} + \int_0^s \left( \int_s^t |f''(w-u)| dw \right) du \\ &\leq w_G(|t-s|) T^{1/2} \|f'\|_{\mathbb{H}_0} + T|t-s|^{1/2} \|f''\|_{\mathbb{H}_0}, \end{aligned}$$

where  $w_G$  is the modulus of continuity of  $G$  as defined in (1.1). Similarly, using the continuity of  $g$  and, as usual, denoting its modulus of continuity by  $w_g$ , we have

$$\begin{aligned} &\int_0^s |f(t-u)g(t-u) - f(s-u)g(s-u)| du \\ &\leq \int_0^s |f(t-u)| |g(t-u) - g(s-u)| du + \int_0^s g(s-u) |f(t-u) - f(s-u)| du \\ &\leq T^{1/2} \|f\|_{\mathbb{H}_0} w_g(|t-s|) + \int_0^s g(s-u) \int_s^t |f'(w-u)| dw du \\ &\leq T^{1/2} \|f\|_{\mathbb{H}_0} w_g(|t-s|) + |t-s|^{1/2} \|f'\|_{\mathbb{H}_0} \leq (T^{1/2} w_g(|t-s|) + |t-s|^{1/2}) \|f\|_{\mathbb{H}_1}. \end{aligned}$$

Recalling that  $\xi_f = (f(1-G))'$ , the last two inequalities show that

$$(7.4) \quad \int_0^s |\xi_f(t-u) - \xi_f(s-u)| du \leq c_1(T, |t-s|) \|f\|_{\mathbb{H}_2},$$

where  $c_1(T, \delta) \doteq (T^{1/2}(w_G(\delta) + w_g(\delta)) + (T+1)\delta^{1/2})$  satisfies  $\lim_{\delta \rightarrow 0} c_1(T, \delta) = 0$ . On the other hand, another application of the Cauchy-Schwarz inequality and the norm inequality (1.2) shows that

$$(7.5) \quad \begin{aligned} \int_s^t |\xi_f(t-u)| du &\leq \int_s^t |f'(u)| du + \|f\|_{\infty} (G(t) - G(s)) \\ &\leq |t-s|^{1/2} \|f'\|_{\mathbb{H}_0} + \|f\|_{\infty} (G(t) - G(s)) \\ &\leq 3(|t-s|^{1/2} + w_G(|t-s|)) \|f\|_{\mathbb{H}_1} \leq c_2(T) \|f\|_{\mathbb{H}_1}, \end{aligned}$$

and

$$(7.6) \quad \|\xi_f\|_T \leq \|f'\|_\infty + \|f\|_\infty \|g\|_T \leq c_2(T) \|f\|_{\mathbb{H}_2}$$

for an appropriate finite constant  $c_2(T) < \infty$  that depends only on  $G$  and  $T$ . Substituting (7.4)–(7.6) into (7.3), we obtain

$$(7.7) \quad \frac{|\mathcal{K}_t(f) - \mathcal{K}_{\tau_n(t)}^{(n)}(f)|}{\|f\|_{\mathbb{H}_2}} \leq \left| K(t) - K^{(n)}(\tau^{(n)}(t)) \right| + c_1(T, \delta^{(n)}(t)) \|K\|_T \\ + c_2(T) \int_0^{t \wedge \tau^{(n)}(t)} \left( K(u) - K^{(n)}(u) \right) du \\ + 2\delta^{(n)}(t) c_2(T) \left( \|K\|_T \vee \left\| K^{(n)} \right\|_T \right).$$

Now, suppose  $K^{(n)} = K$  so that  $\mathcal{K}^{(n)} = \mathcal{K}$ ,  $n \in \mathbb{N}$ , and consider  $t < T$  and any sequence of points  $\tau^{(n)}(t) \in [0, T]$ ,  $n \in \mathbb{N}$ , such that  $\tau^{(n)}(t) \downarrow t$  as  $n \rightarrow \infty$ . Then the third term on the right-hand side of (7.7) vanishes, the first term converges to zero because  $K \in \mathcal{D}_{\mathbb{R}}[0, \infty)$  and the second and fourth terms converge because  $\|K\|_T < \infty$  and  $\delta^{(n)}(t) \rightarrow 0$ . This shows that  $\|\mathcal{K}_t - \mathcal{K}_{\tau^{(n)}(t)}\|_{\mathbb{H}_2} \rightarrow 0$  and hence,  $\mathcal{K} \in \mathcal{D}_{\mathbb{H}_2}[0, \infty)$ . The same argument also shows that  $\mathcal{K}$  is continuous if  $K$  is. This proves the first property.

Next, suppose that  $K^{(n)}$ ,  $n \in \mathbb{N}$ , is a sequence that converges to  $K$  in the Skorokhod topology. By the definition of the Skorokhod topology (see, for example, Chapter 3 of [4]) there exists a sequence of strictly increasing maps  $\tau^{(n)}$ ,  $n \in \mathbb{N}$ , that map  $[0, T]$  onto  $[0, T]$  and satisfy  $\|\delta^{(n)}\|_T \doteq \sup_{t \in [0, T]} |t - \tau^{(n)}(t)| \rightarrow 0$  and  $\|K^{(n)} \circ \tau^{(n)} - K\|_T \rightarrow 0$  as  $n \rightarrow \infty$ . Moreover,  $\sup_n \|K^{(n)}\|_T < \infty$  and  $K^{(n)}(u) \rightarrow K(u)$  for a.e.  $u \in [0, T]$ . Taking first the supremum over  $t \in [0, T]$  and then limits as  $n \rightarrow \infty$  in (7.7), the above properties show that the right-hand side goes to zero (where the dominated convergence theorem is used to argue that the third term vanishes). In turn, this implies that  $\sup_{t \in [0, T]} \|\mathcal{K}_t - \mathcal{K}_{\tau^{(n)}(t)}\|_{\mathbb{H}_2} \rightarrow 0$ , thus establishing convergence of  $\mathcal{K}^{(n)}$  to  $\mathcal{K}$  in the Skorokhod topology on  $\mathcal{D}_{\mathbb{H}_2}[0, \infty)$ . This establishes continuity of the map  $\Gamma$  in the Skorokhod topology. Continuity with respect to the uniform topology can be proved by setting  $\tau^{(n)}(t) = t$ ,  $n \in \mathbb{N}$ , in the argument above. The continuity of the map that takes  $K$  to  $\mathcal{K}(\mathbf{1})$  can be established in an analogous fashion. The proof is left to the reader.

To prove the last property, fix  $K \in \mathcal{C}_{\mathbb{R}}[0, \infty)$  and  $t \in [0, \infty)$ . For  $u \geq 0$ , the function  $\Phi_u \mathbf{1}$  is absolutely continuous and  $\xi_{\Phi_u \mathbf{1}} = (1 - G(\cdot + u))' = -g(\cdot + u)$ . Setting  $f = \Phi_u \mathbf{1}$  in (7.1) yields

$$\mathcal{K}_t(\Phi_u \mathbf{1}) = (1 - G(u))K(t) - \int_0^t K(s)g(t - s + u) ds.$$

The continuity of  $G$  and  $K$  and the bounded convergence theorem then show that  $u \mapsto \mathcal{K}_t(\Phi_u \mathbf{1})$  lies in  $\mathcal{C}_{\mathbb{R}}[0, \infty)$ . On the other hand, given  $K^{(i)} \in \mathcal{C}_{\mathbb{R}}[0, \infty)$  for  $i = 1, 2$  and the corresponding  $\mathcal{K}^{(i)}$ ,

$$\sup_{u \in [0, T]} |\mathcal{K}_t^{(1)}(\Phi_u \mathbf{1}) - \mathcal{K}_t^{(2)}(\Phi_u \mathbf{1})| \leq \|K^{(1)} - K^{(2)}\|_T \left( 1 + \int_0^t g(t - s + u) ds \right),$$

from which it is clear that the map from  $\mathcal{C}_{\mathbb{R}}[0, \infty)$  to itself that takes  $K$  to the function  $u \mapsto \mathcal{K}_t(\Phi_u \mathbf{1})$  is continuous.  $\square$

**7.2. Continuity of the Centered Many-Server Map.** Recall the centered many-server map  $\Lambda$  introduced in Definition 5.4. First, in Lemma 7.2, we establish the representation for  $(\widehat{X}^{(N)}, \widehat{K}^{(N)}, \widehat{\nu}^{(N)}(\mathbf{1}))$  in terms of the map  $\Lambda$  specified in (5.18), and then in Proposition 7.3 and Lemma 7.4 we establish certain continuity and measurability properties of the map  $\Lambda$ .

**Lemma 7.2.** *If the fluid limit  $\bar{X}$  of the total number in system is either subcritical, critical or supercritical and Assumption 4 holds, there exists  $\Omega^* \in \mathcal{F}$  with  $\mathbb{P}(\Omega^*) = 1$  such that for every  $\omega \in \Omega^*$ , there exists  $N^*(\omega) < \infty$  such that for all  $N \geq N^*(\omega)$ ,*

$$(\widehat{K}^{(N)}, \widehat{X}^{(N)}, \langle \mathbf{1}, \widehat{\nu}^{(N)} \rangle)(\omega) = \Lambda((\widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1}))(\omega)).$$

*Proof.* Fix  $N \in \mathbb{N}$ . By the basic definition of the model,  $\widehat{E}^{(N)}$  is càdlàg and  $\widehat{\nu}^{(N)}$  takes values in  $\mathcal{D}_{\mathcal{M}_F[0,L]}[0, \infty)$  and hence, in  $\mathcal{D}_{\mathbb{H}_{-2}}[0, \infty)$ . Due to Assumption 4 and Remark 5.3, it follows that  $\mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})$  is continuous. Moreover, as explained in the discussion right after definition (7.1),  $\widehat{K}(\mathbf{1})$  is also continuous. By the representation (6.15), it then follows that  $\widehat{\mathcal{H}}^{(N)}(\mathbf{1})$  is also càdlàg. Thus, for almost surely every  $\omega \in \Omega$ ,  $(\widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1}))(\omega) \in \widehat{\mathcal{I}}_0$ . The result then follows on comparing the three many-server equations (5.14), (5.15) and (5.16) with the second equation in (6.8), equation (6.10) of Remark 6.3 and equation (6.15).  $\square$

We now establish continuity and measurability properties of the mapping  $\Lambda$ . Recall that  $U$  denotes the renewal function associated with the distribution  $G$ .

**Proposition 7.3.** *Fix  $\bar{X} \in \mathcal{D}_{\mathbb{R}_+}[0, \infty)$ . For  $i = 1, 2$ , suppose  $(E^i, x_0^i, Z^i) \in \widehat{\mathcal{I}}_0$  and  $(K^i, X^i, v^i) \in \Lambda(E^i, x_0^i, Z^i)$ . Then, for any  $T \in [0, \infty)$ ,*

$$(7.8) \quad \|\nabla K\|_T \vee \|\nabla X\|_T \vee \|\nabla v^i\|_T \leq 3(1 + U(T))\varepsilon_T,$$

where  $\nabla S \doteq S^2 - S^1$ ,  $\|f\|_T \doteq \sup_{s \in [0, T]} |f(s)|$  and

$$(7.9) \quad \varepsilon_T \doteq (\|\nabla E\|_T \vee |\nabla x_0| \vee \|\nabla Z\|_T).$$

Hence,  $\Lambda$  is continuous with respect to the uniform topology and is single-valued on its domain.

*Proof.* Fix  $T < \infty$ . We first show that  $\|\nabla K\|_T \leq 2\varepsilon_T(1 + U(T))$ . For any  $t \in [0, T]$ , we consider two cases.

*Case 1:* Either  $\bar{X}(t) < 1$ , or both  $\bar{X}(t) = 1$  and  $X^1(t) \leq 0$ .

We claim that in this case we always have

$$(7.10) \quad \nabla v(t) - \nabla X(t) \leq 0.$$

Indeed, (5.16) shows that if  $\bar{X}(t) < 1$  then  $v^i(t) = X^i(t)$  for  $i = 1, 2$  and so the left-hand side above is identically zero. On the other hand, if  $\bar{X}(t) = 1$  and  $X^1(t) \leq 0$  then  $(X^1(t))^+ = 0$ , and so (5.16), combined with the elementary identity  $x \wedge 0 - x = -x^+$ , implies

$$\nabla v(t) - \nabla X(t) = (X^1(t))^+ - (X^2(t))^+ = -(X^2(t))^+ \leq 0,$$

and so (7.10) holds.

In turn, combining (7.10) with the fact that each solution satisfies equation (5.15), we then conclude that

$$(7.11) \quad \nabla K(t) = \nabla x_0 + \nabla E(t) - \nabla X(t) + \nabla v(t) \leq \nabla x_0 + \nabla E(t) \leq 2\varepsilon_T.$$

*Case 2:* Either  $\bar{X}(t) > 1$ , or both  $\bar{X}(t) = 1$  and  $X^1(t) > 0$ .

First, we claim that in this case,

$$(7.12) \quad \nabla v(t) = v^2(t) - v^1(t) \leq 0.$$

If either  $\bar{X}(t) > 1$ , or the relations  $\bar{X}(t) = 1$ ,  $X^1(t) > 0$  and  $X^2(t) > 0$  hold, this is trivially true since by (5.16) each term on the left-hand side of (7.12) is equal to zero. In the remaining case when  $\bar{X}(t) = 1$ ,  $X^1(t) > 0$  and  $X^2(t) \leq 0$ , (5.16) shows that  $v^1(t) = 0$  and  $v^2(t) = X^2(t) \leq 0$ , and once again (7.12) follows.

Next, considering that each solution satisfies the equation (5.14) and taking the difference, we have for every  $t \in [0, \infty)$ ,

$$(7.13) \quad \nabla K(t) = \nabla v(t) + \int_0^t g(t-s)\nabla K(s) ds - \nabla Z(t).$$

Now, define

$$\mathcal{B} \doteq \left\{ t : \int_0^t g(t-s) \nabla K(s) ds > 0 \right\}.$$

Then, combining (7.13) with (7.12), we conclude that

$$\nabla K(t) \leq 2\varepsilon_T + \mathbb{1}_{\mathcal{B}}(t) \int_0^t g(t-s) \nabla K(s) ds.$$

Applying the same inequality to  $K(s)$  for  $s \in [0, t]$  and substituting it into the last inequality, we then obtain

$$\begin{aligned} \nabla K(t) &\leq \varepsilon_T + \mathbb{1}_{\mathcal{B}}(t) \int_0^t g(t-s) \left( 2\varepsilon_T + \mathbb{1}_{\mathcal{B}}(s) \int_0^s g(s-r) \nabla K(r) dr \right) ds \\ &\leq \varepsilon_T(1 + G(t)) + \mathbb{1}_{\mathcal{B}}(t) \int_0^t g(t-s) \left( \mathbb{1}_{\mathcal{B}}(s) \int_0^s g(s-r) \nabla K(r) dr \right) ds. \end{aligned}$$

Reiterating this procedure, we obtain

$$(7.14) \quad \nabla K(t) \leq \varepsilon_T(1 + G(t) + G^{*,2}(t) + \dots) \leq \varepsilon_T U(T),$$

where  $G^{*,n}$  denotes the  $n$ -fold convolution of  $G$ .

By symmetry, the inequalities (7.11) and (7.14) obtained in Cases 1 and 2, respectively, also hold with  $\nabla K$  replaced by  $-\nabla K$ . Since  $U(T) \geq 1$ , we then have

$$|\nabla K(t)| \leq 2\varepsilon_T U(T), \quad \text{for every } t \in [0, T].$$

Taking the supremum over  $t \in [0, T]$ , we obtain

$$(7.15) \quad \|\nabla K\|_T \leq 2\varepsilon_T U(T).$$

Now, relations (5.14) and (5.15), together, show that for  $i = 1, 2$  and  $t \in [0, T]$ ,

$$X^i(t) = E^i(t) + x_0^i - v^i(0) - \int_0^t g(t-s) K^i(s) ds + Z^i(t).$$

Taking the difference and using the fact that (5.16) implies  $|\nabla(x_0^i - v^i(0))| \leq |\nabla x_0^i|$ , we obtain

$$|\nabla X(t)| \leq \|\nabla E\|_T + |\nabla x_0| + \int_0^t g(t-s) \|\nabla K\|_T ds + \|\nabla Z\|_T.$$

Taking the supremum over  $t \in [0, T]$  and using (7.15), we then conclude that

$$\|\nabla X\|_T \leq 3\varepsilon_T + 2\varepsilon_T U(T)G(T) \leq 3\varepsilon_T(1 + U(T)).$$

Together with (7.15) and the fact that (5.16) implies  $\|\nabla v(t)\|_T \leq \|\nabla X\|_T$ , this establishes (7.8). Since the Skorokhod topology coincides with the uniform topology on the space of continuous functions, (7.8) implies that the map  $\Lambda$  is continuous at points  $(E, x_0, Z) \in \mathcal{C}_{\mathbb{R}^+} \times \mathbb{R} \times \mathcal{C}_{\mathbb{R}^+}$ .  $\square$

**Lemma 7.4.** *Suppose the fluid limit is subcritical, critical or supercritical. Then the map  $\Lambda$  from  $\text{dom}(\Lambda) \subseteq \mathcal{D}_{\mathbb{R}}[0, \infty) \times \mathbb{R} \times \mathcal{D}_{\mathbb{R}}[0, \infty)$  to  $\mathcal{D}_{\mathbb{R}}[0, \infty)^3$ , with  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  equipped with the Skorokhod topology, is measurable.*

*Proof.* We first observe that it suffices to establish the measurability of the map from  $(E, x_0, Z) \in \text{dom}(\Lambda)$  to  $X$ , where  $(K, X, v) = \Lambda(E, x_0, Z)$ . Indeed, in this case, because the maps  $f \mapsto f$ ,  $f \mapsto f \wedge 0$  and  $f \mapsto 0$  from  $\mathcal{D}_{\mathbb{R}}[0, \infty)$ , equipped with the Skorokhod topology, to itself are all measurable (in fact, continuous) and addition is also a measurable mapping from  $\mathcal{D}_{\mathbb{R}}[0, \infty)^2$  to  $\mathcal{D}_{\mathbb{R}}[0, \infty)$ , it follows from (5.15) and (5.16) that the map to  $(K, v)$ , and therefore the map  $\Lambda$ , is also measurable.

By Remark 5.5, if  $(K, X, v) = \Lambda(E, x_0, Z)$  then  $X$  satisfies the integral equation

$$X(t) = R(t) + \int_0^t g(t-s) F(X(s)) ds, \quad t \geq 0,$$

where  $R(t) = R_1(t) \doteq Z(t) + E(t) - \int_0^t g(t-s)E(s) ds$  and  $F = 0$  if  $\bar{X}$  is subcritical,  $R(t) = R_2(t) \doteq R_1(t) + (1 - G(t))x_0$  and  $F(x) = x$  if  $\bar{X}$  is supercritical, and  $R(t) \doteq R_2(t)$  and  $F(x) = x^+$  if  $\bar{X}$  is critical. Note that in all cases,  $F$  is Lipschitz. Also, for fixed  $t$ , the map  $(E, Z, x_0) \mapsto R(t)$  from  $\mathcal{D}_{\mathbb{R}}[0, \infty)^2 \times \mathbb{R} \mapsto \mathbb{R}$  is clearly measurable. The latter fact implies the proof for the subcritical case. For the other two cases, by standard arguments from the theory of Volterra integral equations (see Theorem 3.2.1 of [8]) it follows that  $X(t) = \lim_{n \rightarrow \infty} (\mathcal{T}^{(n)}0)(t)$ , where  $\mathcal{T}_t^{(n)}$  is the  $n$ -fold composition of the operator  $\mathcal{T} : \mathcal{D}_{\mathbb{R}}[0, \infty) \mapsto \mathcal{D}_{\mathbb{R}}[0, \infty)$  given by  $(\mathcal{T}\xi)(t) = R_2(t) + \int_0^t g(t-s)F(\xi(s)) ds$ ,  $\xi \in \mathcal{D}_{\mathbb{R}}[0, \infty)$ . Due to the fact that convergence in the Skorokhod topology implies convergence in  $\mathbb{L}_{loc}^1$ , the map  $\xi \mapsto F(\xi)$  is a continuous mapping from  $\mathbb{L}_{loc}^1$  to itself and the Laplace convolution  $\theta \mapsto \int_0^\cdot g(\cdot - s)\theta(s) ds$  is a continuous map from  $\mathbb{L}_{loc}^1$  to  $\mathcal{C}[0, \infty)$ , it follows that for every  $t > 0$ ,  $(R, \xi) \mapsto \mathcal{T}(\xi)(t)$  is a measurable map. Because the Borel algebra associated with the Skorokhod topology is generated by cylinder sets, and measurability is preserved under compositions and limits, this implies that the map from  $R$  to  $X$  is measurable. Note that in the critical case, the above equation is of the same form as the one obtained in Theorem 3.1 of Reed [30], and a more detailed proof of measurability in the critical case can also be found in the Appendix of [30].  $\square$

## 8. CONVERGENCE RESULTS

The representation (6.15) of the pre-limit dynamics and the continuity properties established in Section 7 reduce the problem of convergence of the sequence  $\{\widehat{\nu}^{(N)}\}_{N \in \mathbb{N}}$  to that of the joint convergence of the sequence of stochastic convolution integrals  $\{\widehat{\mathcal{H}}^{(N)}\}_{N \in \mathbb{N}}$  and the sequences representing the initial data. In this section, we establish these convergence results. First, in Section 8.2 (see Proposition 8.4) we establish the joint convergence of the sequence  $\{\widehat{\mathcal{M}}^{(N)}\}_{N \in \mathbb{N}}$  and the sequence of centered arrival processes and initial conditions showing, in particular, that the centered departure process is asymptotically independent of the centered arrival process and initial conditions. Then, in Section 8.3 (see Corollary 8.7) we identify (for a suitable family of  $f$ ) the limits of the sequence  $\{\widehat{\mathcal{H}}^{(N)}(f)\}_{N \in \mathbb{N}}$ . Both limit theorems are proved using some basic estimates, which are first obtained in Section 8.1.

**8.1. Preliminary Estimates.** Let  $U$  denote the renewal function associated with the service distribution  $G$ . We begin with a useful bound, whose proof is relegated to Appendix D.

**Lemma 8.1.** *Fix  $T < \infty$ . For every  $N \in \mathbb{N}$  and positive integer  $k$ ,*

$$(8.1) \quad \mathbb{E} \left[ \left( \overline{A}_1^{(N)}(T) \right)^k \right] = \mathbb{E} \left[ \left( \int_0^T \int_{[0, L)} h(x) \overline{\nu}_s^{(N)}(dx) ds \right)^k \right] \leq k!(U(T))^k.$$

Moreover, there exists  $\overline{C}(T) < \infty$  such that for every positive integer  $k$  and measurable function  $\varphi$  on  $[0, L) \times [0, T]$ ,

$$\sup_{N \in \mathbb{N}} \mathbb{E} \left[ \left( \overline{A}_\varphi^{(N)}(T) \right)^k \right] \leq k! (\overline{C}(T))^k \left( \int_{[0, L)} \varphi^*(x) h(x) dx \right)^k,$$

where  $\varphi^*(x) \doteq \sup_{s \in [0, T]} |\varphi(x, s)|$ . Furthermore, if Assumptions 1 and 2 hold, then

$$(8.2) \quad (\overline{A}_1(T))^k = \left( \int_0^T \int_{[0, L)} h(x) \overline{\nu}_s(dx) ds \right)^k \leq k!(U(T))^k.$$

We now establish some estimates on the martingale measure  $\widehat{\mathcal{M}}^{(N)}$ , which are used in Sections 8.2 and 8.3 to establish various convergence and sample path regularity results.

**Lemma 8.2.** *For every even integer  $r$ , there exists a universal constant  $C_r < \infty$  such that for every bounded and continuous function  $\varphi$  on  $[0, L) \times \mathbb{R}$  and  $T < \infty$ ,*

$$(8.3) \quad \mathbb{E} \left[ \sup_{s \in [0, T]} \left| \widehat{\mathcal{M}}_s^{(N)}(\varphi) \right|^r \right] \leq C_r \|\varphi\|_\infty^r \left[ \left(\frac{r}{2}\right)! (U(T))^{r/2} + \frac{1}{N^{r/2}} \right], \quad N \in \mathbb{N},$$

$$(8.4) \quad \mathbb{E} \left[ \sup_{s \in [0, T]} \left| \widehat{\mathcal{M}}_s(\varphi) \right|^r \right] \leq C_r \left(\frac{r}{2}\right)! (U(T))^{r/2} \|\varphi\|_\infty^r$$

and, for  $0 \leq s \leq t$ ,

$$(8.5) \quad \mathbb{E} \left[ \left| \widehat{\mathcal{M}}_t(\varphi) - \widehat{\mathcal{M}}_s(\varphi) \right|^r \right] \leq C_r (\overline{A}_{\varphi^2}(t) - \overline{A}_{\varphi^2}(s))^{r/2}.$$

*Proof.* Since  $\widehat{\mathcal{M}}^{(N)}(\varphi)$  is a martingale, by the Burkholder-Davis-Gundy (BDG) inequality (see, for example, Theorem 7.11 of Walsh [35]) it follows that for any  $r > 1$ , there exists a universal constant  $C_r < \infty$  (independent of  $\varphi$  and  $\widehat{\mathcal{M}}^{(N)}$ ) such that

$$(8.6) \quad \mathbb{E} \left[ \sup_{s \leq T} \left| \widehat{\mathcal{M}}_s^{(N)}(\varphi) \right|^r \right] \leq C_r \mathbb{E} \left[ \left( \langle \widehat{\mathcal{M}}^{(N)}(\varphi) \rangle_T \right)^{r/2} \right] + C_r \mathbb{E} \left[ \left| \Delta \widehat{\mathcal{M}}_T^{(N),*}(\varphi) \right|^r \right],$$

where

$$\Delta \widehat{\mathcal{M}}_T^{(N),*}(\varphi) \doteq \sup_{t \in [0, T]} \left| \Delta \widehat{\mathcal{M}}_t^{(N)}(\varphi) \right| = \sup_{t \in [0, T]} \left| \widehat{\mathcal{M}}_t^{(N)}(\varphi) - \widehat{\mathcal{M}}_{t-}^{(N)}(\varphi) \right|.$$

Because the jumps of  $\widehat{\mathcal{M}}^{(N)}(\varphi)$  are bounded by  $\|\varphi\|_\infty / \sqrt{N}$ , we have

$$(8.7) \quad \mathbb{E} \left[ \left| \Delta \widehat{\mathcal{M}}_T^{(N),*}(\varphi) \right|^r \right] \leq \frac{\|\varphi\|_\infty^r}{N^{r/2}}.$$

On the other hand, by (4.12) it follows that for any  $r > 0$ ,

$$\mathbb{E} \left[ \langle \widehat{\mathcal{M}}^{(N)}(\varphi) \rangle_T^{r/2} \right] = \mathbb{E} \left[ \left( \overline{A}_{\varphi^2}^{(N)}(T) \right)^{r/2} \right] \leq \|\varphi\|_\infty^r \mathbb{E} \left[ \left( \overline{A}_1^{(N)}(T) \right)^{r/2} \right].$$

When combined with (8.1) of Lemma 8.1 this shows that if  $r = 2k$ , where  $k$  is an integer, then

$$(8.8) \quad \mathbb{E} \left[ \left( \langle \widehat{\mathcal{M}}^{(N)}(\varphi) \rangle_T \right)^{r/2} \right] \leq \|\varphi\|_\infty^r \left(\frac{r}{2}\right)! (U(T))^{r/2}.$$

Combining the estimates (8.6)–(8.8) obtained above, we obtain (8.3).

In an exactly analogous fashion, replacing  $\widehat{\mathcal{M}}^{(N)}$  and  $\overline{A}^{(N)}$ , respectively, by  $\widehat{\mathcal{M}}$  and  $\overline{A}$ , and using the continuity of  $\widehat{\mathcal{M}}(\varphi)$  and inequality (8.2) of Lemma 8.1, we obtain (8.4). Furthermore, for fixed  $s \geq 0$ , because  $\{\widehat{\mathcal{M}}_t(\varphi) - \widehat{\mathcal{M}}_s(\varphi)\}_{t \geq s}$  is a continuous martingale with quadratic variation process  $\{\overline{A}_{\varphi^2}(t) - \overline{A}_{\varphi^2}(s)\}_{t \geq s}$ , another application of the Burkholder-Davis-Gundy (BDG) inequality yields (8.5).  $\square$

As a corollary, we obtain results on the regularity of the processes  $\widehat{\mathcal{M}}^{(N)}$  and  $\widehat{\mathcal{M}}$ , which make use of the norm inequalities in (1.2).

**Corollary 8.3.** *Each  $\widehat{\mathcal{M}}^{(N)}$ ,  $N \in \mathbb{N}$ , is a càdlàg  $\mathbb{H}_{-2}$ -valued (and hence  $\mathcal{S}'$ -valued) process.  $\widehat{\mathcal{M}}$  is a continuous  $\mathbb{H}_{-2}$ -valued (and hence  $\mathcal{S}'$ -valued) process. Moreover, for any  $T < \infty$ , if for every  $f \in \mathcal{S}$ ,  $\widehat{\mathcal{M}}^{(N)}(f) \Rightarrow \widehat{\mathcal{M}}(f)$  in  $\mathcal{D}_{\mathbb{R}}[0, T]$  as  $N \rightarrow \infty$  then  $\widehat{\mathcal{M}}^{(N)} \Rightarrow \widehat{\mathcal{M}}$  in  $\mathcal{D}_{\mathbb{H}_{-2}}[0, T]$  as  $N \rightarrow \infty$ .*

*Proof.* Fix  $N \in \mathbb{N}$ . By Remark 4.3, for every  $f \in \mathcal{S}$  there exists a càdlàg version  $\widehat{\mathcal{M}}_f^{(N)}$  of  $\widehat{\mathcal{M}}^{(N)}(f)$ . Moreover, for any  $T < \infty$  it follows from (8.3) and (1.2) that given any  $\varepsilon > 0$  and  $\lambda < \infty$  there exists  $\delta > 0$  such that if  $\|f\|_{\mathbb{H}_1} \leq \delta$  then

$$(8.9) \quad \limsup_N \mathbb{P} \left( \sup_{t \in [0, T]} \left| \widehat{\mathcal{M}}_t^{(N)}(f) \right| > \lambda \right) \leq \varepsilon.$$

Thus, each  $\widehat{\mathcal{M}}^{(N)}$  is a 1-continuous stochastic process in the sense of Mitoma [27]. Since  $\mathcal{S}$  is a nuclear Fréchet space and  $\|\cdot\|_{\mathbb{H}_1} \stackrel{\text{HS}}{<} \|\cdot\|_{\mathbb{H}_2}$  (refer to the properties stated in Section 1.4.1), by Theorem 4.1 of Walsh [35] and Corollary 2 of Mitoma [27] it follows that  $\widehat{\mathcal{M}}^{(N)}$  is a càdlàg  $\mathbb{H}_{-2}$ -valued, and hence  $\mathcal{S}'$ -valued, process.

On the other hand, by Lemma 8.2  $\widehat{\mathcal{M}}(f)$  is a continuous process for every  $f \in \mathcal{S}$ . An analogous argument to the one above, that now invokes Corollary 1 of Mitoma [27] and (8.4), shows that  $\widehat{\mathcal{M}}$  is a continuous  $\mathbb{H}_{-2}$ -valued process. The last assertion of the corollary follows from (8.9) and Corollary 6.16 of Walsh [35].  $\square$

**8.2. Asymptotic Independence.** We now identify the limit of the sequence of martingale measures  $\{\widehat{\mathcal{M}}^{(N)}\}_{N \in \mathbb{N}}$  and also show that it is asymptotically independent of the centered arrival process and initial conditions. We recall from Assumption 5 and Remark 5.1 that  $\widehat{\mathcal{M}}$  is independent of the initial conditions  $(\widehat{x}_0, \widehat{\nu}_0, \mathcal{S}^{\widehat{\nu}_0}, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}))$  and  $\widehat{E}$ , where  $\widehat{E}$  is a diffusion with drift coefficient  $-\beta$  and diffusion coefficient  $\sigma^2$ .

**Proposition 8.4.** *Suppose Assumptions 1–3 and Assumption 5 hold. Then for every  $\varphi \in \mathcal{C}_b([0, L] \times \mathbb{R})$ ,  $\widehat{\mathcal{M}}^{(N)}(\varphi) \Rightarrow \widehat{\mathcal{M}}(\varphi)$  in  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  as  $N \rightarrow \infty$ . Moreover, as  $N \rightarrow \infty$ ,*

$$(\widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \widehat{\mathcal{M}}^{(N)}) \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{S}^{\widehat{\nu}_0}, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}), \widehat{\mathcal{M}})$$

in  $\mathcal{D}_{\mathbb{R}}[0, \infty) \times \mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty) \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty)$ .

*Proof.* We shall first prove the assertion under the supposition that Assumption 3(a) is satisfied, in which case  $\bar{\lambda}, \sigma^2$  are positive constants and  $\beta$  is a constant in  $\mathbb{R}$ . We start by using results of Puhalskii and Reiman [28] to recast the problem in a more convenient form. Fix  $N \in \mathbb{N}$  and define

$$\widehat{L}^{(N)}(t) \doteq \frac{1}{\sqrt{N}} \sum_{j=2}^{E^{(N)}(t)+1} \left(1 - \lambda^{(N)} \xi_j^{(N)}\right), \quad t \in [0, \infty),$$

where recall that  $\{\xi_j^{(N)}\}_{j \in \mathbb{N}}$  is the i.i.d. sequence of interarrival times of the  $N$ th renewal arrival process  $E^{(N)}$ , which has mean  $1/\lambda^{(N)}$  and variance  $(\sigma^2/\bar{\lambda})/(\lambda^{(N)})^2$ . Define

$$\widehat{\gamma}^{(N)}(t) \doteq \frac{\lambda^{(N)}}{\sqrt{N}} \left( \sum_{j=2}^{E^{(N)}(t)+1} \xi_j^{(N)} - \bar{\lambda}t \right), \quad t \geq 0.$$

Using the definition (5.2) of  $\beta$  and the fact that  $\overline{E}^{(N)}(t) = \bar{\lambda}t$ , we see that

$$(8.10) \quad \widehat{E}^{(N)}(t) = \frac{E^{(N)}(t) - N\bar{\lambda}t}{\sqrt{N}} = \frac{E^{(N)}(t) - \lambda^{(N)}t}{\sqrt{N}} + \beta t = \widehat{L}^{(N)}(t) + \widehat{\gamma}^{(N)}(t) + \beta t.$$

Puhalskii and Reiman (see page 30, Lemma A.1 and (5.15) of [28]) showed that  $\{\widehat{L}^{(N)}(t), \mathcal{F}_t^{(N)}, t \geq 0\}$  is a locally square integrable martingale and, as  $N \rightarrow \infty$ ,  $\sup_{t \leq T} |\widehat{\gamma}^{(N)}(t)| \rightarrow 0$  in probability, which implies  $\gamma^{(N)} \Rightarrow 0$ .

We will now show that for every bounded and continuous  $f$ ,

$$(8.11) \quad (\widehat{L}^{(N)}, \widehat{\mathcal{M}}_f^{(N)}) \Rightarrow (B, \widehat{\mathcal{M}}(f)) \quad \text{as } N \rightarrow \infty,$$

and for real-valued bounded, continuous functions  $\xi_1$  on  $\mathbb{R}^2$  and  $\xi_2$  on  $\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty)$ ,

$$(8.12) \quad \begin{aligned} \lim_{N \rightarrow \infty} \mathbb{E}[\xi_1(\widehat{\mathcal{M}}_f^{(N)}, \widehat{L}^{(N)}) \xi_2(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}))] \\ = \mathbb{E}[\xi_1(\widehat{\mathcal{M}}_f, B)] \mathbb{E}[\xi_2(\widehat{x}_0, \widehat{\nu}_0, \mathcal{S}^{\widehat{\nu}_0}, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}))]. \end{aligned}$$



Before presenting the proofs of these results, first note that on combining (8.11) with (8.10), the fact that  $\gamma^{(N)} \Rightarrow 0$  as  $N \rightarrow \infty$ , the relation  $\widehat{E}(t) = B(t) - \beta t$ ,  $t \geq 0$ , and the continuous mapping theorem, it follows that for every bounded and continuous  $f$ ,  $(\widehat{E}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)) \Rightarrow (\widehat{E}, \widehat{\mathcal{M}}(f))$  in  $\mathcal{D}_{\mathbb{R}}[0, \infty)^2$ . Together with (8.12) this implies that for every bounded and continuous  $f$ ,  $(\widehat{E}^{(N)}, \widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \widehat{\mathcal{M}}^{(N)}(f)) \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{S}^{\widehat{\nu}_0}, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}), \widehat{\mathcal{M}}(f))$  as  $N \rightarrow \infty$ . Together with Corollary 8.3 this implies the desired convergence stated in the proposition.

Thus, to complete the proof (when Assumption 3(a) holds), it suffices to establish (8.11) and (8.12). This is done in the following four claims; the first three claims below verify conditions of the martingale central limit theorem to establish (8.11), whereas the last claim proves (8.12).

CLAIM 1. For  $t \geq 0$ , as  $N \rightarrow \infty$ ,  $[\widehat{L}^{(N)}]_t \rightarrow \sigma^2 t$  and  $[\widehat{M}_f^{(N)}]_t \rightarrow \bar{A}_{f^2}(t)$  in probability.

PROOF OF CLAIM 1. Note that  $\widehat{L}^{(N)}$ , being a compensated sum of jumps, is a local martingale of finite variation. Thus, it is a purely discontinuous martingale (see Lemma 4.14(b) of Chapter I of Jacod and Shiryaev [17]) and hence (by Theorem 4.5.2 of Chapter 1 of [17]), its  $\{\mathcal{F}_t^{(N)}\}$  optional quadratic variation is given by

$$[\widehat{L}^{(N)}]_t = \frac{1}{N} \sum_{j=2}^{E^{(N)}(t)+1} (1 - \lambda^{(N)} \xi_j^{(N)})^2.$$

For every  $j \in \mathbb{N}$ ,  $\mathbb{E}[(1 - \lambda^{(N)} \xi_j^{(N)})^2] = \sigma^2 / \bar{\lambda} < \infty$  by Assumption 3(a), and almost surely  $\bar{E}^{(N)}(t) \rightarrow \bar{\lambda} t$  by Remark 5.1 and Theorem 3.2. Therefore, by the strong law of large numbers for triangular arrays of random variables,  $[\widehat{L}^{(N)}]_t$  converges to

$$\lim_{N \rightarrow \infty} \frac{E^{(N)}(t) - 1}{N} \mathbb{E} \left[ \left( 1 - \lambda^{(N)} \xi_j^{(N)} \right)^2 \right] = \sigma^2 t$$

as  $N \rightarrow \infty$ . This establishes the first limit of Claim 1.

Now,  $\widehat{M}_f^{(N)}$  is also a compensated sum of jumps. By the same logic we then have

$$[\widehat{M}_f^{(N)}]_t = \sum_{s \leq t} \left( \Delta \widehat{M}_f^{(N)}(s) \right)^2 = \bar{Q}_{f^2}^{(N)}(t),$$

where the last equality follows because the jumps of  $M_f^{(N)}$  coincide with those of  $Q_f^{(N)}$ . The results of Kaspi and Ramanan (see Theorem 5.4, the discussion below Theorem 5.15 and Proposition 5.17 of [22]) show that  $\bar{Q}_{f^2}^{(N)}(t) \rightarrow \bar{A}_{f^2}(t)$  in probability, and so the second limit in Claim 1 is also established.

CLAIM 2. For every  $t > 0$ ,  $[\widehat{L}^{(N)}, \widehat{M}_f^{(N)}]_t \rightarrow 0$  in probability as  $N \rightarrow \infty$ .

PROOF OF CLAIM 2. Let  $\tau_i^{(N)} \doteq \sum_{j=1}^i \xi_j^{(N)}$  be the time of the  $i$ th jump of  $E^{(N)}$ . Since  $E^{(N)}$  has unit jumps, it follows that

$$(8.13) \quad [\widehat{L}^{(N)}, \widehat{M}_f^{(N)}]_t = \frac{1}{\sqrt{N}} \sum_{i \geq 2: \tau_i^{(N)} \leq t} \left( 1 - \lambda^{(N)} \xi_{i+1}^{(N)} \right) \Delta \widehat{M}_f^{(N)} \left( \tau_i^{(N)} \right).$$

To prove the claim, it suffices to show that

$$(8.14) \quad \mathbb{E} \left[ [\widehat{L}^{(N)}, \widehat{M}_f^{(N)}]_t^2 \right] \leq \frac{\sigma^2 \|f\|_{\infty}^2}{\bar{\lambda} N} \mathbb{E} \left[ \sum_{i: \tau_i^{(N)} \leq t} \Delta \bar{D}^{(N)} \left( \tau_i^{(N)} \right) \right].$$

Indeed, then the right-hand side goes to zero as  $N \rightarrow \infty$  because the expectation on the right-hand side is bounded by  $\sup_N \mathbb{E}[\bar{D}^{(N)}(t)]$ , which is finite by Lemma 5.6 of [22] (alternatively, the

convergence to zero can also be deduced from the stronger result stated in Corollary A.2). The claim then follows by an application of Chebysev's inequality.

To establish (8.14), we first introduce the filtration  $\{\tilde{\mathcal{F}}_t^{(N)}, t \geq 0\}$ , which is defined exactly like the filtration  $\{\mathcal{F}_t^{(N)}, t \geq 0\}$ , except that the forward recurrence time process  $R_E^{(N)}$  associated with the renewal arrival process  $E^{(N)}$  is replaced by the age or backward recurrence time process  $\alpha_E^{(N)}$ , which satisfies  $\alpha_E^{(N)}(s) = s - \sup\{u < s : E^{(N)}(u) < E^{(N)}(s)\} \vee 0$  for  $s \geq 0$ . It is easy to verify that  $\widehat{M}_f^{(N)}$  is an  $\{\tilde{\mathcal{F}}_t^{(N)}\}$ -adapted process and, for each  $i \in \mathbb{N}$ ,  $\tau_i^{(N)}$  is an  $\{\tilde{\mathcal{F}}_t^{(N)}\}$ -stopping time. Moreover, for every  $i \in \mathbb{N}$ , using the independence of  $\xi_{i+1}^{(N)}$  from  $\tilde{\mathcal{F}}_{\tau_i^{(N)}}^{(N)}$  we have

$$(8.15) \quad \mathbb{E}[1 - \lambda^{(N)} \xi_{i+1}^{(N)} | \tilde{\mathcal{F}}_{\tau_i^{(N)}}^{(N)}] = \mathbb{E}[1 - \lambda^{(N)} \xi_{i+1}^{(N)}] = 0,$$

$$(8.16) \quad \mathbb{E}\left[\left(1 - \lambda^{(N)} \xi_{i+1}^{(N)}\right)^2 | \tilde{\mathcal{F}}_{\tau_i^{(N)}}^{(N)}\right] = \mathbb{E}\left[\left(1 - \lambda^{(N)} \xi_{i+1}^{(N)}\right)^2\right] = \sigma^2 / \bar{\lambda}.$$

For  $i \in \mathbb{N}$ , conditioning on  $\tilde{\mathcal{F}}_{\tau_i^{(N)}}^{(N)}$ , noting that the jump times of  $\widehat{M}_f^{(N)}$  and  $\overline{D}^{(N)}$  coincide, and using (8.16) and the estimate  $(\Delta \widehat{M}_f^{(N)})^2 \leq \|f\|_\infty^2 \Delta \overline{D}^{(N)}$ , we obtain

$$\begin{aligned} \mathbb{E}\left[\left(1 - \lambda^{(N)} \xi_{i+1}^{(N)}\right)^2 \Delta \widehat{M}_f^{(N)} \left(\tau_i^{(N)}\right)^2 | \tilde{\mathcal{F}}_{\tau_i^{(N)}}^{(N)}\right] &= \frac{\sigma^2}{\bar{\lambda}} \Delta \widehat{M}_f^{(N)} \left(\tau_i^{(N)}\right)^2 \\ &\leq \frac{\sigma^2 \|f\|_\infty^2}{\bar{\lambda}} \Delta \overline{D}^{(N)} \left(\tau_i^{(N)}\right). \end{aligned}$$

A similar conditioning argument using (8.15) shows that for  $2 \leq k < i$ ,  $i, k \in \mathbb{N}$ ,

$$\mathbb{E}\left[\left(1 - \lambda^{(N)} \xi_{k+1}^{(N)}\right) \Delta \widehat{M}_f^{(N)} \left(\tau_k^{(N)}\right) \left(1 - \lambda^{(N)} \xi_{i+1}^{(N)}\right) \Delta \widehat{M}_f^{(N)} \left(\tau_i^{(N)}\right)\right] = 0.$$

Taking first the square and then the expectation of each side of (8.13) and using the last two relations, we obtain (8.14). As argued above, this proves the claim.

CLAIM 3. The jumps of  $(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f))$  are asymptotically negligible and (8.11) holds.

PROOF OF CLAIM 3. The jumps of  $\widehat{E}^{(N)}$  and  $\widehat{L}^{(N)}$  converge to zero as  $N \rightarrow \infty$  because  $E^{(N)}$  is a counting process with unit jumps and  $\sup_{t \leq T} |\widehat{\gamma}^{(N)}(t)| \rightarrow 0$  in probability. Also, by Lemma 4.1 and the continuity of  $A_f^{(N)}$ , the jumps of  $\widehat{\mathcal{M}}^{(N)}(f) = \widehat{M}_f^{(N)}$  are uniformly bounded by  $\|f\|_\infty / \sqrt{N}$ , and so they also converge to zero in probability. Because  $\{(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f))\}_{N \in \mathbb{N}}$  is a sequence of martingales starting at zero, we can apply the martingale central limit theorem (see, e.g., Theorem 1.4 on page 339 of Ethier and Kurtz [10]). The conditions of that theorem are verified by claims 1-3 above, and (8.11) follows from the observation that  $B$  and  $\widehat{\mathcal{M}}(f)$  are independent, centered mean Gaussian processes with variance processes  $\sigma^2 t$  and  $\overline{A}_{f^2}(t)$ ,  $t \geq 0$ , respectively.

CLAIM 4. The asymptotic independence property in (8.12) holds.

PROOF OF CLAIM 4. Conditioned on  $\mathcal{F}_0^{(N)}$ ,  $\widehat{L}^{(N)}$  and  $\widehat{M}_f^{(N)}$  are still compensated sums of jumps with the same optional quadratic variation processes as without conditioning. Thus, the same argument provided above in claims 1-3 above show that, conditioned on  $\mathcal{F}_0^{(N)}$ , the sequence  $(\widehat{L}^{(N)}, \widehat{M}_f^{(N)})$  still converges weakly to  $(B, \widehat{\mathcal{M}}(f))$ . In particular, by the continuous mapping theorem, this then implies that for any bounded continuous function  $F_1$  on  $\mathbb{R}^2$ ,

$$\lim_{N \rightarrow \infty} \mathbb{E}\left[F_1(\widehat{L}^{(N)}(t), \widehat{M}_f^{(N)}(t)) | \mathcal{F}_0^{(N)}\right] = \mathbb{E}[F_1(B(t), \widehat{\mathcal{M}}_t(f))].$$

This establishes the desired asymptotic independence because for any bounded, continuous function  $F_2$  on  $\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty)$ , by the bounded convergence theorem and another

application of the continuous mapping theorem for  $F_2$ ,

$$\begin{aligned} & \lim_{N \rightarrow \infty} \mathbb{E} \left[ F_1(\widehat{L}^{(N)}(t), \widehat{M}_f^{(N)}(t)) F_2(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \right] \\ &= \lim_{N \rightarrow \infty} \mathbb{E} \left[ \mathbb{E} \left[ F_1(\widehat{L}^{(N)}(t), \widehat{M}_f^{(N)}(t)) | \mathcal{F}_0^{(N)} \right] F_2(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \right] \\ &= \mathbb{E} \left[ F_1(B(t), \widehat{\mathcal{M}}_t(f)) \mathbb{E} [F_2(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}, \mathcal{S}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}))] \right]. \end{aligned}$$

This completes the proof for the case when Assumption 3(a) is satisfied.

We now turn to the proof for the case when Assumption 3(b) is satisfied. The proof in this case is similar, and so we only elaborate on the differences. First, for  $N \in \mathbb{N}$ , define  $\widehat{L}^{(N)} = L^{(N)}/\sqrt{N}$ , where now

$$L^{(N)}(t) \doteq \left( E^{(N)}(t) - \int_0^t \lambda^{(N)}(s) ds \right), \quad t \in [0, \infty),$$

is the scaled and centered inhomogeneous Poisson process, and note that

$$(8.17) \quad \widehat{E}^{(N)}(t) = \sqrt{N} \left( \overline{E}^{(N)}(t) - \int_0^t \lambda(s) ds \right) = \widehat{L}^{(N)}(t) + \int_0^t \beta(s) ds.$$

Fix  $f$  that is bounded and continuous. To complete the proof of the proposition, it suffices to show that  $(\widehat{L}^{(N)}, \widehat{M}_f^{(N)}) \Rightarrow (\int_0^\cdot \sqrt{\lambda(s)} dB(s), \widehat{\mathcal{M}}(f))$ . Let  $\{\tilde{\mathcal{F}}_t^{(N)}\}$  be the filtration defined in Claim 2 above. Then, as is well known,  $\{\widehat{L}^{(N)}(t), \tilde{\mathcal{F}}_t^{(N)}, t \geq 0\}$  and  $\{\widehat{M}_f^{(N)}(t), \tilde{\mathcal{F}}_t^{(N)}, t \geq 0\}$  are martingales. Hence, once again, we need only verify the conditions of the martingale central limit theorem (see Theorem 1.4 on page 339 of Ethier and Kurtz [10]). Arguing exactly as in Claims 3 and 1 of the proof for case (a), it is clear that the jumps of  $\widehat{E}^{(N)}$  and  $\widehat{M}_f^{(N)}$  are uniformly bounded by  $(1 + \|f\|_\infty)/\sqrt{N}$  and for each  $t > 0$ ,  $[\widehat{M}_f^{(N)}]_t \rightarrow \overline{A}_{f^2}(t)$  in probability. Keeping in mind that the candidate limit  $(\widehat{E}, \widehat{\mathcal{M}}(f))$  is a pair of independent, continuous Gaussian martingales with respective quadratic variations  $\int_0^t \lambda(s) ds$  and  $\overline{A}_{f^2}(t)$ , to complete the proof it suffices to verify that for every  $t \in [0, \infty)$ , as  $N \rightarrow \infty$ , the following limits hold in probability:

$$(8.18) \quad [\widehat{L}^{(N)}]_t \rightarrow \int_0^t \lambda(s) ds, \quad [\widehat{L}^{(N)}, M_f^{(N)}]_t \rightarrow 0.$$

Clearly, the  $\{\tilde{\mathcal{F}}_t^{(N)}\}$ -predictable quadratic variation of  $\widehat{L}^{(N)}$  is given by  $\langle \widehat{L}^{(N)} \rangle_t = \int_0^t \lambda^{(N)} ds$ , which converges to  $\int_0^t \lambda(s) ds$  as  $N \rightarrow \infty$ . By Theorem 3.11 of Chapter VIII of [17], this implies that  $[\widehat{L}^{(N)}]_t$  converges in law to  $\int_0^t \lambda(s) ds$ . Because the limit  $\int_0^t \lambda(s) ds$  is deterministic, the convergence is also in probability. This establishes the first limit in (8.18). To establish the second limit, note that  $\widehat{L}^{(N)}$  and  $\widehat{M}_f^{(N)}$  are both compensated pure jump processes with continuous compensators, and so their optional quadratic covariation takes the form

$$[\widehat{L}^{(N)}, \widehat{M}_f^{(N)}]_t = \frac{1}{N} \sum_{s \leq t} \Delta L^{(N)}(s) \Delta \widehat{M}_f^{(N)}(s) = \frac{1}{N} \sum_{s \leq t} \Delta E^{(N)}(s) \Delta Q_f^{(N)}(s).$$

Noting that  $\Delta Q_f^{(N)} \leq \|f\|_\infty \Delta D^{(N)}$ , taking expectations of both sides above and then the limit as  $N \rightarrow \infty$ , Corollary A.2 shows that

$$\lim_{N \rightarrow \infty} \mathbb{E} \left[ [[\widehat{L}^{(N)}, \widehat{M}_f^{(N)}]_t] \right] \leq \lim_{N \rightarrow \infty} \frac{\|f\|_\infty}{N} \mathbb{E} \left[ \sum_{s \leq t} \Delta E^{(N)}(s) \Delta D^{(N)}(s) \right] = 0.$$

An application of Markov's inequality then yields the second limit in (8.18). The asymptotic independence from the initial conditions is proved exactly in the same way as when Assumption 3(a) holds (see the proof of Claim 4) and is thus omitted.  $\square$

**8.3. Convergence of Stochastic Convolution Integrals.** We now show that for suitable  $f$ ,  $\{\widehat{\mathcal{H}}^{(N)}(f)\}_{N \in \mathbb{N}}$  is a tight sequence of càdlàg processes. Since each  $\widehat{\mathcal{H}}^{(N)}(f)$  is not a martingale, the proof is more involved than the corresponding result for  $\{\widehat{\mathcal{M}}^{(N)}(f)\}_{N \in \mathbb{N}}$  and we require an additional regularity assumption (Assumption 4) on  $G$ , which holds if the hazard rate function  $h$  is bounded and  $g$  is in  $\mathbb{L}^{1+\alpha}$  for some  $\alpha > 0$  (see Remark 5.2). For conciseness, we will use the notation

$$(8.19) \quad \overline{G}(x) = 1 - G(x), \quad f\overline{G}(x) = f(x)\overline{G}(x), \quad x \in [0, \infty).$$

We first derive an elementary inequality.

**Lemma 8.5.** *Suppose Assumption 4 is satisfied, let  $f$  be a bounded, Hölder continuous function with constant  $C_f$  and exponent  $\gamma_f$ , and let  $\gamma'_f \doteq \gamma_G \wedge \gamma_f$ . The family of operators  $\{\Psi_t, t \geq 0\}$  defined in (4.19) satisfies, for all  $0 < t < t' < \infty$ ,*

$$(8.20) \quad \|\Psi_t f - \Psi_{t'} f\|_\infty \leq (C_f + C_G \|f\|_\infty) |t - t'|^{\gamma'_f}.$$

Moreover, if  $f \in \mathbb{H}_1$  then  $C_f \leq \|f\|_{\mathbb{H}_1}$  and there exists a constant  $C_0 < \infty$ , independent of  $f$ , such that the right-hand side of (8.20) can be replaced by  $C_0 \|f\|_{\mathbb{H}_1} |t - t'|^{\gamma'_f}$ .

*Proof.* Fix a bounded, Hölder continuous function  $f$ , as in the statement of the lemma. Then we can write  $\Psi_t f - \Psi_{t'} f = \varphi^{(1)} + \varphi^{(2)}$ , where

$$\varphi^{(1)}(x, s) = \frac{\overline{G}(x + (t - s)^+) - \overline{G}(x)}{\overline{G}(x)} (f(x + (t - s)^+) - f(x + (t' - s)^+))$$

and

$$\varphi^{(2)}(x, s) = f(x + (t' - s)^+) \frac{\overline{G}(x + (t - s)^+) - \overline{G}(x + (t' - s)^+)}{\overline{G}(x)}.$$

The Hölder continuity of  $f$  and the fact that  $\overline{G}$  is non-increasing show that  $\|\varphi^{(1)}\|_\infty \leq C_f |t - t'|^{\gamma_f}$ , and Assumption 4 shows that  $\|\varphi^{(2)}\|_\infty \leq C_G \|f\|_\infty |t - t'|^{\gamma_G}$ . When combined, these two inequalities yield (8.20). If  $f \in \mathcal{S}$ , the Cauchy-Schwarz inequality implies

$$|f(t) - f(t')| = \left| \int_t^{t'} f'(u) du \right| \leq \|f'\|_{\mathbb{L}^2} (t - t')^{1/2} \leq \|f\|_{\mathbb{H}_1} (t - t')^{1/2}.$$

Thus,  $C_f = \|f'\|_{\mathbb{L}^2} \leq \|f\|_{\mathbb{H}_1}$  and  $\gamma_f = 1/2$ , respectively, serve as a Hölder constant and exponent for  $f$ . When combined with (1.2) this shows that  $f$  is bounded and Hölder continuous and the second assertion of the lemma holds with  $C_0 \doteq (1 + 4C_G)$ .  $\square$

In what follows, for  $t > 0$ , let  $\Theta_t : \mathcal{C}_b([0, L]) \mapsto \mathcal{C}_b([0, L] \times [0, t])$  be the operator given by

$$(8.21) \quad (\Theta_t f)(x, u) \doteq \int_u^t (\Psi_s \varphi(\cdot, s))(x, u) ds = \int_u^t \varphi(x + s - u, s) \frac{1 - G(x + s - u)}{1 - G(x)} ds$$

for  $(x, u) \in [0, L] \times [0, t]$  and  $f \in \mathcal{C}_b[0, L]$ . The first two properties of the next lemma are used in Corollary 8.7 to establish convergence of the sequence  $\{\widehat{\mathcal{H}}^{(N)}(f)\}_{N \in \mathbb{N}}$  and regularity of the limit. The third property below is used in the proof of the Fubini type result in Lemma E.1 and the last property is used in the proof of Theorem 5.11.

**Lemma 8.6.** *If Assumption 4 is satisfied, the following properties hold:*

- (1) *Given a bounded and Hölder continuous function  $f$  on  $[0, L]$ , the sequence of processes  $\{\widehat{\mathcal{H}}^{(N)}(f)\}_{N \in \mathbb{N}}$  is tight in  $\mathbb{D}_{\mathbb{R}}[0, \infty)$  and  $\widehat{\mathcal{H}}(f)$  is  $\mathbb{P}$ -a.s. continuous.*
- (2) *There exists  $r \geq 2$  and a constant  $C_0 < \infty$  such that for any  $f \in \mathcal{S}$ ,*

$$(8.22) \quad \sup_N \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \widehat{\mathcal{H}}_t^{(N)}(f) \right|^r \right] \leq C_0 \|f\|_{\mathbb{H}_1}^r.$$

- (3) Suppose  $\varphi : [0, L] \times [0, \infty) \mapsto \mathbb{R}$  is a Borel measurable function such that for every  $x \in [0, L]$ , the function  $r \mapsto \varphi(x, r)$  is locally integrable and, for every  $t \in [0, T]$ , the function  $x \mapsto \int_0^t \varphi(x, r) dr$  is bounded and Hölder continuous with constant  $C_{\varphi, T}$  and exponent  $\gamma_{\varphi, T}$  (that is independent of  $t$ ). Then  $\mathbb{P}$  almost surely, the random field  $\{\widehat{\mathcal{H}}_s(\int_0^t(\varphi(\cdot, r)) dr), s, t \geq 0\}$  is jointly continuous in  $s$  and  $t$ .
- (4) Suppose  $\varphi \in \mathcal{C}_b([0, L] \times [0, \infty))$ . Then the process  $\{\widehat{\mathcal{M}}_t(\Theta_t \varphi), t \geq 0\}$  admits a continuous version.

*Proof.* The proof of the lemma is based on a modification of the approach used in Walsh [35] to establish convergence of stochastic convolution integrals, tailored to the present context (the proof of Theorem 7.13 in [35] works with a different space of test functions and imposes different conditions on the martingale measure  $\widehat{\mathcal{M}}^{(N)}$ , and hence does not apply directly). Fix a bounded and Hölder continuous  $f$  with constant  $C_f$  and exponent  $\gamma_f$  and fix  $T < \infty$ . Recall from (4.21) that  $\widehat{\mathcal{H}}_t^{(N)}(f) = \widehat{\mathcal{M}}_t^{(N)}(\Psi_t f)$ ,  $t \geq 0$ . The proof of the first two properties will be split into four main claims.

CLAIM 1. For each  $N \in \mathbb{N}$ ,  $\{\widehat{\mathcal{H}}_t^{(N)}(f), t \geq 0\}$  admits a càdlàg version.

PROOF OF CLAIM 1. The estimates obtained in this proof are also used to establish the other claims. Fix  $N \in \mathbb{N}$  and consider the following stochastic integral:

$$(8.23) \quad V_t^{(N)}(f) \doteq \widehat{\mathcal{M}}_T^{(N)}(\Psi_t f), \quad t \in [0, T].$$

Because  $\widehat{\mathcal{M}}^{(N)}$  is a martingale measure, we have

$$(8.24) \quad \widehat{\mathcal{H}}_t^{(N)}(f) = \widehat{\mathcal{M}}_t^{(N)}(\Psi_t f) = \mathbb{E} \left[ V_t^{(N)}(f) | \mathcal{F}_t^{(N)} \right], \quad t \in [0, T],$$

which shows that the process  $\{\widehat{\mathcal{M}}_t^{(N)}(\Psi_t f), t \geq 0\}$  is a version of the optional projection of  $V^{(N)}(f)$ . It is well known from the general theory of stochastic processes (see, for example, Theorem 7.10 of Chapter V of Rogers and Williams [33]) that the optional projection of a continuous process is an adapted càdlàg process. Therefore, to show that  $\{\widehat{\mathcal{M}}_t^{(N)}(\Psi_t f), t \in [0, T]\}$  admits a càdlàg version, it suffices to show that  $V^{(N)}(f)$  admits a continuous version. In turn, to establish continuity, it suffices to verify Kolmogorov's continuity criterion, namely, to show that there exist  $\tilde{C}_f < \infty$ ,  $\tilde{\theta} > 1$  and  $r < \infty$  such that for every  $0 \leq t' < t < T$ ,

$$(8.25) \quad \mathbb{E} \left[ \left| V_t^{(N)}(f) - V_{t'}^{(N)}(f) \right|^r \right] \leq \tilde{C}_f |t - t'|^{\tilde{\theta}}$$

(see, for example, Corollary 1.2 of Walsh [35]). Fix  $0 \leq t' \leq t \leq T$  and note that

$$(8.26) \quad \left| V_t^{(N)}(f) - V_{t'}^{(N)}(f) \right| = \left| \widehat{\mathcal{M}}_T^{(N)}(\Psi_t f - \Psi_{t'} f) \right|.$$

Let  $r$  be any positive even integer greater than  $1/\gamma_f'$ . Together with (8.3) and (8.20), this implies that (8.25) is satisfied with  $\tilde{\theta} = r\gamma_f' > 1$  and

$$(8.27) \quad \tilde{C}_f = C_r(C_f + C_G \|f\|_\infty)^r ((r/2)! U^{r/2}(T) + 1).$$

CLAIM 2.  $\widehat{\mathcal{H}}(f)$  has a continuous version.

PROOF OF CLAIM 2. Analogous to (8.23) and (8.24), we define  $\widehat{V}_t(f) \doteq \widehat{\mathcal{M}}_T(\Psi_t f)$ ,  $t \geq 0$ , and observe that

$$(8.28) \quad \widehat{\mathcal{H}}_t(f) = \widehat{\mathcal{M}}_t(\Psi_t f) = \mathbb{E} \left[ \widehat{V}_t(f) | \widehat{\mathcal{F}}_t \right], \quad t \geq 0.$$

Arguments analogous to those used in Claim 1, with the inequalities (8.4) and (8.2), respectively, now playing the role of (8.3) and (8.1), can be used to show that

$$(8.29) \quad \mathbb{E} \left[ \left| \widehat{V}_t(f) - \widehat{V}_{t'}(f) \right|^r \right] \leq \tilde{C}_f |t - t'|^{\tilde{\theta}}$$

with  $\tilde{\theta} = r\gamma'_f$ . Fix  $0 < t' < t < \infty$  with  $|t - t'| < 1$  and a bounded, Hölder continuous  $f$ . Using (8.28) and adding and subtracting  $\widehat{\mathcal{M}}_{t'}(\Psi_t f) = \mathbb{E}[V_t(f)|\widehat{\mathcal{F}}_{t'}]$ , we obtain

$$(8.30) \quad \widehat{\mathcal{H}}_t(f) - \widehat{\mathcal{H}}_{t'}(f) = \mathbb{E} \left[ V_t(f) - V_{t'}(f) | \widehat{\mathcal{F}}_{t'} \right] + \widehat{\mathcal{M}}_t(\Psi_t f) - \widehat{\mathcal{M}}_{t'}(\Psi_t f).$$

Consider any even integer  $r > 2/\gamma'_f \vee 4$  so that (8.29) holds with  $\tilde{\theta} > 2$ , let  $\bar{\theta} \doteq \lfloor r/2 \wedge \tilde{\theta} \rfloor$  and note that  $\bar{\theta}$  is an integer greater than or equal to 2. Taking first the  $r$ th power and then expectations of both sides of (8.30), and using the inequality  $(x + y)^r \leq 2^r(x^r + y^r)$  and Jensen's inequality, we obtain

$$\mathbb{E} \left[ \left| \widehat{\mathcal{H}}_{t'}(f) - \widehat{\mathcal{H}}_t(f) \right|^r \right] \leq 2^r \left( \mathbb{E} [|V_t(f) - V_{t'}(f)|^r] + \mathbb{E} \left[ \left| \widehat{\mathcal{M}}_t(\Psi_t f) - \widehat{\mathcal{M}}_{t'}(\Psi_t f) \right|^r \right] \right).$$

Applying the estimates (8.25), (8.5) and the fact that  $\|\Psi_t f\|_\infty \leq \|f\|_\infty$ , and then the inequality  $x^2 + y^2 \leq (x + y)^2$  for  $x, y \geq 0$ , this implies that

$$(8.31) \quad \begin{aligned} \mathbb{E} \left[ \left| \widehat{\mathcal{H}}_{t'}(f) - \widehat{\mathcal{H}}_t(f) \right|^r \right] &\leq 2^r \tilde{C}_f |t - t'|^{\bar{\theta}} + 2^r C_r (\bar{A}_{(\Psi_{t'} f)^2}(t) - \bar{A}_{(\Psi_{t'} f)^2}(t'))^{r/2} \\ &\leq 2^r (\tilde{C}_f \vee C_r \|f\|_\infty^2) (t + \bar{A}_1(t) - t' - \bar{A}_1(t'))^2. \end{aligned}$$

Since  $t + \bar{A}_1(t)$  is a non-negative, increasing function of  $t$ , the generalized Kolmogorov's continuity criterion (see, for example, Corollary 3 of [27]) implies that  $\widehat{\mathcal{H}}(f)$  has a continuous version.

CLAIM 3. The estimate (8.22) is satisfied.

PROOF OF CLAIM 3. From the proof of Corollary 1.2 of Walsh [35] it is straightforward to deduce that (8.25) also implies that there exists a constant  $\tilde{C}_r < \infty$ , which depends on  $r$  but is independent of  $N$  and  $f$ , such that

$$(8.32) \quad \mathbb{E} \left[ \sup_{s \in [0, T]} \left| V_s^{(N)}(f) \right|^r \right] \leq \tilde{C}_r \tilde{C}_f.$$

By (8.24) and Jensen's inequality, for every  $t \in [0, T]$ ,

$$|\widehat{\mathcal{H}}_t^{(N)}(f)|^r \leq \mathbb{E} \left[ \sup_{s \in [0, T]} \left| V_s^{(N)}(f) \right|^r | \mathcal{F}_t^{(N)} \right] \leq \mathbb{E} \left[ \sup_{s \in [0, T]} \left| V_s^{(N)}(f) \right|^r | \mathcal{F}_t^{(N)} \right].$$

By (8.32), the last term above (viewed as a process in  $t$ ) is a martingale. So Doob's inequality and (8.32) imply that

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in [0, T]} \left| \widehat{\mathcal{H}}_t^{(N)}(f) \right|^r \right] &\leq \frac{r^2}{r-1} \mathbb{E} \left[ \mathbb{E} \left[ \sup_{s \in [0, T]} \left| V_s^{(N)}(f) \right|^r | \mathcal{F}_T^{(N)} \right] \right] \\ &= \frac{r^2}{r-1} \mathbb{E} \left[ \sup_{s \in [0, T]} \left| V_s^{(N)}(f) \right|^r \right] \leq \frac{r^2}{r-1} \tilde{C}_r \tilde{C}_f. \end{aligned}$$

If  $f \in \mathcal{S}$  then by the expression for  $\tilde{C}_f$  given in (8.27) and the inequalities  $\|f\|_\infty \leq \sqrt{6} \|f\|_{\mathbb{H}_1}$  and  $C_f \leq \|f\|_{\mathbb{H}_1}$  established in (1.2) and Lemma 8.5, respectively, the right-hand side above can be replaced by  $C_0 \|f\|_{\mathbb{H}_1}^r$ , for an appropriate constant  $C_0 = C_0(G, r, T) < \infty$  that is independent of  $N$  and  $f$ . Thus, (8.22) follows.

CLAIM 4. The sequence  $\{\widehat{\mathcal{H}}_t^{(N)}(f), t \geq 0\}_{N \in \mathbb{N}}$  is tight in  $\mathcal{D}_{\mathbb{R}}[0, \infty)$ .

PROOF OF CLAIM 4. We will prove the claim by verifying Aldous' criteria for tightness of stochastic processes. A minor modification of the arguments in Claims 1-3 shows that if  $\delta_N \in (0, 1)$  and  $T_N$  is an  $\{\mathcal{F}_t^{(N)}\}$  stopping time such that  $T_N + \delta_N \leq T$ , then for any even integer  $r \geq 2$ ,

$$(8.33) \quad \mathbb{E} \left[ |V_{T_N + \delta_N}^{(N)}(f) - V_{T_N}^{(N)}(f)|^r \right] \leq \tilde{C}_f \delta_N^{r\gamma'_f}.$$

Let  $\delta_N \in (0, 1)$  and let  $T_N$  be an  $\{\mathcal{F}_t^{(N)}\}$  stopping time such that  $T_N + \delta_N \leq T$ . Using (8.24) and (8.23), the difference  $\widehat{\mathcal{H}}_{T_N + \delta_N}^{(N)}(f) - \widehat{\mathcal{H}}_{T_N}^{(N)}(f)$  can be rewritten as

$$\begin{aligned} & \mathbb{E} \left[ V_{T_N + \delta_N}^{(N)} | \mathcal{F}_{T_N + \delta_N}^{(N)} \right] - \mathbb{E} \left[ V_{T_N}^{(N)} | \mathcal{F}_{T_N}^{(N)} \right] \\ &= \mathbb{E} \left[ V_{T_N + \delta_N}^{(N)} - V_{T_N}^{(N)} | \mathcal{F}_{T_N + \delta_N}^{(N)} \right] + \mathbb{E} \left[ V_{T_N}^{(N)} | \mathcal{F}_{T_N + \delta_N}^{(N)} \right] - \mathbb{E} \left[ V_{T_N}^{(N)} | \mathcal{F}_{T_N}^{(N)} \right] \\ &= \mathbb{E} \left[ V_{T_N + \delta_N}^{(N)} - V_{T_N}^{(N)} | \mathcal{F}_{T_N + \delta_N}^{(N)} \right] + \int_{[0, L] \times (T_N, T_N + \delta_N]} \Psi_{T_N}(f)(x, s) \widehat{\mathcal{M}}^{(N)}(dx, ds). \end{aligned}$$

Recalling the covariance functional of  $\widehat{\mathcal{M}}^{(N)}$  specified in (4.12) and the fact that  $\|\Psi_{T_N}(f)\|_\infty \leq \|f\|_\infty$ , this implies that

$$\begin{aligned} & \mathbb{E} \left[ \left| \widehat{\mathcal{H}}_{T_N + \delta_N}^{(N)}(f) - \widehat{\mathcal{H}}_{T_N}^{(N)}(f) \right|^2 \right] \\ & \leq 2\mathbb{E} \left[ \left| V_{T_N + \delta_N}^{(N)}(f) - V_{T_N}^{(N)}(f) \right|^2 \right] + 2\|f\|_\infty^2 \mathbb{E} \left[ \overline{A}_1^{(N)}(T_N + \delta_N) - \overline{A}_1^{(N)}(T_N) \right] \\ & \leq 2\tilde{C}_f \delta_N^{2\gamma'_f} + 2\|f\|_\infty^2 \sup_{\tilde{N}} \mathbb{E} \left[ \sup_{t \in [0, T]} \left( \overline{A}_1^{(\tilde{N})}(t + \delta_N) - \overline{A}_1^{(\tilde{N})}(t) \right) \right], \end{aligned}$$

where the last equality uses (8.33) with  $r = 2$ . As  $\delta_N \rightarrow 0$ , the first term on the right-hand side clearly converges to zero, whereas Lemma 5.8(2) of Kaspı and Ramanan [22] shows that the second term also converges to zero. We conclude that  $\widehat{\mathcal{H}}_{T_N + \delta_N}^{(N)}(f) - \widehat{\mathcal{H}}_{T_N}^{(N)}(f)$  converges to zero in  $\mathbb{L}^2$ , and hence in probability. On the other hand, (8.22) shows that the sequence  $\{\widehat{\mathcal{H}}^{(N)}(f)\}_{N \in \mathbb{N}}$  is uniformly bounded in  $\mathbb{L}^r$ . We have thus verified Aldous' criteria (see, for example, Theorem 6.8 of Walsh [35]), and hence the sequence  $\{\widehat{\mathcal{H}}^{(N)}(f)\}_{N \in \mathbb{N}}$  is tight.

We now turn to the proof of property 3. Fix  $T < \infty$ , let  $\varphi$  be as stated in the lemma and for  $t \in [0, T]$ , define  $f_\varphi^t(x) = \int_0^t \varphi(x, r) dr$ . For  $s, t, s', t' \in [0, T]$  with  $t' < t$ , we have

$$\widehat{\mathcal{H}}_s(f_\varphi^t) - \widehat{\mathcal{H}}_{s'}(f_\varphi^{t'}) = \widehat{\mathcal{H}}_s(f_\varphi^t) - \widehat{\mathcal{H}}_{s'}(f_\varphi^t) + \widehat{\mathcal{M}}_{s'} \left( \Psi_s \left( \int_{t'}^t \varphi(\cdot, r) dr \right) \right).$$

Due to the assumed boundedness and Hölder continuity of  $f_\varphi^t$ , (8.31) and (8.4) together with the above relation imply that there exists a sufficiently large integer  $r$ , constant  $C(T, r, \varphi) < \infty$  and  $\tilde{\theta} = \tilde{\theta}(r, \varphi) > 1$  such that

$$\begin{aligned} & \mathbb{E} \left[ \left| \widehat{\mathcal{H}}_{s'} \left( \int_0^{t'} \varphi(\cdot, r) dr \right) - \widehat{\mathcal{H}}_s \left( \int_0^t \varphi(\cdot, r) dr \right) \right|^r \right] = \mathbb{E} \left[ \left| \widehat{\mathcal{H}}_{s'}(f_\varphi^{t'}) - \widehat{\mathcal{H}}_s(f_\varphi^t) \right|^r \right] \\ & \leq C(T, r, \varphi) \left( |s + \overline{A}_1(s) - s' - \overline{A}_1(s')|^{\tilde{\theta}} + |t - t'|^{\tilde{\theta}} \right). \end{aligned}$$

Property 3 then follows from the generalized Kolmogorov's criterion for continuity of random fields.

The proof of the last property of the lemma is similar to the proof of the continuity of  $\widehat{\mathcal{H}}$  given in Claim 2, and so we only provide a rough sketch. Let  $R_t(\varphi) \doteq \widehat{\mathcal{M}}_t(\Theta_t \varphi)$ , define  $\tilde{V}_t(\varphi) \doteq \widehat{\mathcal{M}}_T(\Theta_t \varphi)$  and note that  $R_t(\varphi) = \mathbb{E}[\tilde{V}_t(\varphi) | \mathcal{F}_t]$ . In a manner similar to (8.30), we can write

$$R_t(\varphi) - R_{t'}(\varphi) = \mathbb{E} \left[ \widehat{\mathcal{M}}_T \left( \int_{t'}^t \Psi_s \varphi ds \right) | \mathcal{F}_{t'} \right] + \widehat{\mathcal{M}}_t(\Theta_t \varphi) - \widehat{\mathcal{M}}_{t'}(\Theta_{t'} \varphi).$$

Using Jensen's inequality, (8.4) with  $r = 4$  and the inequalities  $\left\| \int_{t'}^t \Psi_s \varphi ds \right\|_\infty \leq |t - t'| \|\varphi\|_\infty$  and  $\|\Theta_{t'} \varphi\|_\infty \leq T \|\varphi\|_\infty$ , it follows that for  $0 < t' < t < T$ ,  $t - t' \leq 1$ ,

$$\begin{aligned} \mathbb{E} \left[ |R_t(\varphi) - R_{t'}(\varphi)|^4 \right] &\leq 2^4 \left( \mathbb{E} \left[ \left| \widehat{\mathcal{M}}_T \left( \int_{t'}^t \Psi_s \varphi ds \right) \right|^4 \right] + \|\Theta_{t'}(\varphi)\|_\infty^4 (\overline{A}_1(t) - \overline{A}_1(t'))^2 \right) \\ &\leq 2^4 \tilde{C}(T) \|\varphi\|_\infty^4 \left( |t - t'|^4 + (\overline{A}_1(t) - \overline{A}_1(t'))^2 \right) \\ &\leq 2^4 \tilde{C}(T) \|\varphi\|_\infty^4 \left( (t - t')^2 + (\overline{A}_1(t) - \overline{A}_1(t'))^2 \right) \\ &\leq 2^4 \tilde{C}(T) \|\varphi\|_\infty^4 (t - t' + \overline{A}_1(t) - \overline{A}_1(t'))^2, \end{aligned}$$

where  $\tilde{C}(T) \doteq (2C_4U(T)^2) \vee T^4$ . The claim then follows from the generalized Kolmogorov continuity criterion.  $\square$

Combining Lemma 8.6 with arguments similar to those used in the proof of Corollary 8.3, we now obtain the main convergence result of the section.

**Corollary 8.7.** *As  $N \rightarrow \infty$ ,  $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$  in  $\mathcal{Y}_1$ . Also, if for any bounded, Hölder continuous  $f$ , (5.27) holds then  $(\widehat{Y}_1^{(N)}, \widehat{\mathcal{H}}^{(N)}(f)) \Rightarrow (\widehat{Y}_1, \widehat{\mathcal{H}}(f))$  as  $N \rightarrow \infty$ .*

*Proof.* Fix  $N \in \mathbb{N}$ . For  $\tilde{k}, k \in \mathbb{N}$ ,  $i = 1, \dots, \tilde{k}$ ,  $j = 1, \dots, k$ , let  $\tilde{f}_i$  and  $f_j$ , respectively, be bounded, continuous and bounded, Hölder continuous functions. Proposition 8.4 and Lemma 8.6 imply that the sequence

$$\{(\widehat{\mathcal{M}}^{(N)}(\tilde{f}_1), \dots, \widehat{\mathcal{M}}^{(N)}(\tilde{f}_{\tilde{k}}), \widehat{\mathcal{H}}^{(N)}(f_1), \dots, \widehat{\mathcal{H}}^{(N)}(f_k))\}_{N \in \mathbb{N}}$$

is tight in  $\mathcal{D}_{\mathbb{R}}[0, \infty)^{\tilde{k}+k}$ . Since  $\widehat{\mathcal{H}}_t^{(N)}(f) = \widehat{\mathcal{M}}_t^{(N)}(\Psi_t f)$  and, likewise,  $\widehat{\mathcal{H}}_t(f) = \widehat{\mathcal{M}}_t(\Psi_t f)$ , Proposition 8.4 also shows that for  $\tilde{t}_i, t_j \in [0, \infty)$ ,  $i = 1, \dots, \tilde{k}$ ,  $j = 1, \dots, k$ , as  $N \rightarrow \infty$ , the corresponding projections converge:

$$\begin{aligned} (\widehat{\mathcal{M}}_{\tilde{t}_1}^{(N)}(\tilde{f}_1), \dots, \widehat{\mathcal{M}}_{\tilde{t}_{\tilde{k}}}^{(N)}(\tilde{f}_{\tilde{k}}), \widehat{\mathcal{H}}_{t_1}^{(N)}(f_1), \dots, \widehat{\mathcal{H}}_{t_k}^{(N)}(f_k)) &\Rightarrow \\ (\widehat{\mathcal{M}}_{\tilde{t}_1}(\tilde{f}_1), \dots, \widehat{\mathcal{M}}_{\tilde{t}_{\tilde{k}}}(\tilde{f}_{\tilde{k}}), \widehat{\mathcal{H}}_{t_1}(f_1), \dots, \widehat{\mathcal{H}}_{t_k}(f_k)). \end{aligned}$$

Since  $\mathcal{S}$  is a subset of the space of bounded and Hölder continuous functions, together the last two statements show that

$$(\widehat{\mathcal{M}}^{(N)}(\tilde{f}), \widehat{\mathcal{H}}^{(N)}(f), \widehat{\mathcal{H}}^{(N)}(f_1)) \Rightarrow (\widehat{\mathcal{M}}(\tilde{f}), \widehat{\mathcal{H}}(f), \widehat{\mathcal{H}}(f_1))$$

for  $f, \tilde{f} \in \mathcal{S}$  and  $f_1$  bounded and Hölder continuous. Because  $\mathcal{S}$  and  $\mathcal{S}'$  are nuclear Fréchet spaces, by Mitoma's theorem (see Corollary 2 of [27]) it follows that  $(\widehat{\mathcal{M}}^{(N)}, \widehat{\mathcal{H}}^{(N)}, \widehat{\mathcal{H}}^{(N)}(f_1)) \Rightarrow (\widehat{\mathcal{M}}, \widehat{\mathcal{H}}, \widehat{\mathcal{H}}(f_1))$  in  $\mathcal{D}_{\mathcal{S}'}[0, \infty)^2 \times \mathcal{D}_{\mathbb{R}}[0, \infty)$ . Since  $\|\cdot\|_{\mathbb{H}_1}^{\text{HS}} < \|\cdot\|_{\mathbb{H}_2}$  and the estimate (8.22) holds, Corollary 6.16 of Walsh [35] then shows that  $(\widehat{\mathcal{M}}^{(N)}, \widehat{\mathcal{H}}^{(N)}, \widehat{\mathcal{H}}^{(N)}(f_1)) \Rightarrow (\widehat{\mathcal{M}}, \widehat{\mathcal{H}}, \widehat{\mathcal{H}}(f_1))$  in  $\mathcal{D}_{\mathbb{H}_2}[0, \infty)^2 \times \mathcal{D}_{\mathbb{R}}[0, \infty)$ , as  $N \rightarrow \infty$ . Now,  $(\widehat{\mathcal{H}}, \widehat{\mathcal{H}}(f_1))$  is adapted to the filtration generated by  $\widehat{\mathcal{M}}$ , and  $\widehat{\mathcal{M}}$  is independent of  $(\widehat{E}, \widehat{x}_0, \widehat{v}_0, \mathcal{S}^{\widehat{v}_0}, \mathcal{S}^{\widehat{v}_0}(\mathbf{1}))$  by Assumption 5. Thus, the same argument used to establish asymptotic independence in Proposition 8.4 also shows that the convergence above can be strengthened to  $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$  and  $(\widehat{Y}_1^{(N)}, \widehat{\mathcal{H}}^{(N)}(f)) \Rightarrow (\widehat{Y}_1, \widehat{\mathcal{H}}(f))$ .  $\square$

## 9. PROOFS OF MAIN THEOREMS

**9.1. The Functional Central Limit Theorem.** Before presenting the proof of Theorem 5.6, we first establish the main convergence result.

**Proposition 9.1.** *Suppose Assumptions 1–5 are satisfied and suppose that the fluid limit is either subcritical, critical or supercritical. Then the convergence (5.21) holds and  $(\widehat{K}, \widehat{X}, \widehat{v}(\mathbf{1}))$  has almost*



surely continuous sample paths. Moreover, if  $g$  is continuous and  $\hat{\nu}$  is defined as in (5.24), then as  $N \rightarrow \infty$ ,

$$(9.1) \quad (\hat{Y}_1^{(N)}, \hat{K}^{(N)}, \hat{X}^{(N)}, \hat{\nu}^{(N)}, \hat{\mathcal{K}}^{(N)}, \hat{\mathcal{K}}^{(N)}(\mathbf{1})) \Rightarrow (\hat{Y}, \hat{K}, \hat{X}, \hat{\nu}, \hat{\mathcal{K}}, \hat{\mathcal{K}}(\mathbf{1}))$$

in  $\mathcal{Y}_1 \times \mathcal{D}_{\mathbb{R}}[0, \infty)^2 \times \mathcal{D}_{\mathbb{H}^2}^2[0, \infty) \times \mathcal{D}_{\mathbb{R}}[0, \infty)$ .

*Proof.* Corollary 8.7 shows that  $\hat{Y}_1^{(N)} \Rightarrow \hat{Y}_1$  in  $\mathcal{Y}_1$  as  $N \rightarrow \infty$ , which in particular implies that

$$(\hat{E}^{(N)}, \hat{x}_0^{(N)}, \mathcal{S}^{\hat{\nu}_0^{(N)}}(\mathbf{1}), \hat{\mathcal{H}}^{(N)}(\mathbf{1})) \Rightarrow (\hat{E}, \hat{x}_0, \mathcal{S}^{\hat{\nu}_0}(\mathbf{1}), \hat{\mathcal{H}}(\mathbf{1}))$$

as  $N \rightarrow \infty$ . By Remark 5.1, Assumption 5 and Lemma 8.6(1),  $(\hat{E}, \mathcal{S}^{\hat{\nu}_0}(\mathbf{1}), \hat{\mathcal{H}}(\mathbf{1}), \mathcal{S}^{\hat{\nu}_0}, \hat{\mathcal{H}})$  has almost surely continuous sample paths with values in  $\mathbb{R}_+^3 \times \mathbb{H}^2$ . Recalling the definition (5.19) of  $\hat{Y}_1^{(N)}$  and the fact that addition in the Skorokhod topology is continuous at points in  $\mathcal{C}[0, \infty)$ , this implies that as  $N \rightarrow \infty$ ,

$$(9.2) \quad (\hat{Y}_1^{(N)}, \hat{E}^{(N)}, \hat{x}_0^{(N)}, \mathcal{S}^{\hat{\nu}_0^{(N)}}(\mathbf{1}) - \hat{\mathcal{H}}^{(N)}(\mathbf{1})) \Rightarrow (\hat{Y}_1, \hat{E}, \hat{x}_0, \mathcal{S}^{\hat{\nu}_0}(\mathbf{1}) - \hat{\mathcal{H}}(\mathbf{1})).$$

By Lemma 7.2, almost surely  $(\hat{K}^{(N)}, \hat{X}^{(N)}, \langle \mathbf{1}, \hat{\nu}^{(N)} \rangle) = \Lambda(\hat{E}^{(N)}, \hat{x}_0^{(N)}, \mathcal{S}^{\hat{\nu}_0^{(N)}}(\mathbf{1}) - \hat{\mathcal{H}}^{(N)}(\mathbf{1}))$  for all  $N$  large enough. The continuity of  $\Lambda$  with respect to the uniform topology on  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  established in Proposition 7.3, the measurability of  $\Lambda$  with respect to the Skorokhod topology on  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  established in Lemma 7.4 and a generalized version of the continuous mapping theorem (see, for example, Theorem 10.2 of Chapter 3 of [10]) then shows that the convergence (5.21) holds with  $(\hat{K}, \hat{X}, \hat{\nu}(\mathbf{1})) \doteq \Lambda(\hat{E}, \hat{x}_0, \mathcal{S}^{\hat{\nu}_0}(\mathbf{1}) - \hat{\mathcal{H}}(\mathbf{1}))$ . By the model assumptions and Lemma 4.1, almost surely,  $\Delta E^{(N)}(t) \leq 1$  and  $\Delta D^{(N)}(t) \leq 1$  for every  $t \geq 0$ . Combining this with (2.3), (6.10) and the second equation for  $\hat{K}^{(N)}$  in (6.8), it follows that almost surely for every  $t \geq 0$ ,

$$\max(\Delta \hat{K}^{(N)}(t), \Delta \hat{X}^{(N)}(t), \Delta \langle \mathbf{1}, \hat{\nu}_t^{(N)} \rangle) \leq \frac{3}{\sqrt{N}}.$$

Because jumps are continuous in the Skorokhod topology, the weak convergence of  $(\hat{K}^{(N)}, \hat{X}^{(N)}, \hat{\nu}^{(N)}(\mathbf{1}))$  to  $(\hat{K}, \hat{X}, \hat{\nu}(\mathbf{1}))$  established in (9.1) shows that  $(\hat{K}, \hat{X}, \hat{\nu}(\mathbf{1}))$  is almost surely continuous. (Note that when  $g$  is continuous, the continuity of  $(\hat{K}, \hat{X}, \hat{\nu}(\mathbf{1}))$  is also guaranteed by Remark 5.5.)

Next, suppose  $g$  is continuous. By Lemma 7.1(2), both the map  $\Gamma$  that takes  $\hat{K}^{(N)}$  to  $\hat{\mathcal{K}}^{(N)}$  and the map that takes  $\hat{K}^{(N)}$  to  $\hat{\mathcal{K}}^{(N)}(\mathbf{1})$  is continuous (with respect to the Skorokhod topology on both the domain and range). So by (5.21) and the continuous mapping theorem, as  $N \rightarrow \infty$ ,

$$(9.3) \quad (\hat{Y}_1^{(N)}, \hat{K}^{(N)}, \hat{X}^{(N)}, \hat{\mathcal{K}}^{(N)}, \hat{\mathcal{K}}^{(N)}(\mathbf{1})) \Rightarrow (\hat{Y}, \hat{K}, \hat{X}, \hat{\mathcal{K}}, \hat{\mathcal{K}}(\mathbf{1})).$$

In turn, the representation (6.15) shows that  $\hat{\nu}^{(N)} = \mathcal{S}^{\hat{\nu}_0^{(N)}} - \hat{\mathcal{H}}^{(N)} + \hat{\mathcal{K}}^{(N)}$  and hence, is a continuous mapping of  $\hat{Y}_1^{(N)}$  and  $\hat{\mathcal{K}}^{(N)}$ . Thus, (9.3) and another application of the continuous mapping theorem show that (5.25) holds with  $\hat{\nu} = \mathcal{S}^{\hat{\nu}_0} - \hat{\mathcal{H}} + \hat{\mathcal{K}}$ . Since this coincides with the definition of  $\hat{\nu}$  given in (5.24), this establishes the proposition.  $\square$

We now prove the first two main results of the paper.

*Proof of Theorems 5.6 and 5.7.* The limit in (5.21), the continuity of  $(\hat{K}, \hat{X}, \hat{\nu}(\mathbf{1}))$  and Theorem 5.7 follow from Proposition 9.1. The relation (6.16) shows that

$$\int_0^\cdot \langle h, \hat{\nu}_s^{(N)} \rangle ds = \langle \mathbf{1}, \hat{\nu}_0^{(N)} \rangle - \mathcal{S}^{\hat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{M}}^{(N)}(\mathbf{1}) + \hat{\mathcal{H}}^{(N)}(\mathbf{1}) + \int_0^\cdot \hat{K}^{(N)}(s)g(\cdot - s) ds.$$

The last term equals  $\hat{K}^{(N)} - \hat{\mathcal{K}}^{(N)}(\mathbf{1})$ , and so by Lemma 7.1(2) the mapping from  $\hat{K}^{(N)}$  to the last term is continuous. The limit (5.21) along with the continuous mapping theorem then shows that  $\int_0^\cdot \langle h, \hat{\nu}_s^{(N)} \rangle ds \Rightarrow \tilde{D}$ , where  $\tilde{D}$  is as defined in (5.23). The relation (6.6) for  $\hat{X}^{(N)}$ , the continuity of the limit and another application of the continuous mapping theorem then yields the representation (5.22) for  $\hat{X}$ . This completes the proof of the theorem.  $\square$

**9.2. The Semimartingale Property.** In view of the representation (5.22) for  $\widehat{X}$  and the fact that  $\widehat{\mathcal{M}}_1$  and  $\widehat{E}$  are by definition semimartingales, to show that  $\widehat{X}$  is a semimartingale it suffices to show that  $\widehat{D}$  is a semimartingale. This is first carried out in Lemma 9.2 below. Throughout, we assume that Assumptions 1, 3, and 5' are satisfied, the fluid limit is subcritical, critical or supercritical and that, in addition,  $h$  is bounded and absolutely continuous. As stated in Remark 5.2, if  $h$  is bounded then Assumptions 2 and 4 are automatically satisfied. Thus, the results of Theorems 5.6 and 5.7 are valid.

**Lemma 9.2.** *Almost surely, the function  $t \mapsto \widehat{D}(t)$  is absolutely continuous and*

$$(9.4) \quad \frac{d\widehat{D}(t)}{dt} = \widehat{\nu}_t(h), \quad a.e. t \in [0, \infty).$$

*Proof.* We start by rewriting the expression (5.23) for  $\widehat{D}$  obtained in Theorem 5.6 in a more convenient form. By the definitions of  $\Phi_t$  and  $\mathcal{S}^{\widehat{\nu}_0}$  given in (5.6) and (5.11), respectively, for  $t > 0$ ,

$$\begin{aligned} \widehat{\nu}_0(\mathbf{1}) - \mathcal{S}_t^{\widehat{\nu}_0}(\mathbf{1}) &= \widehat{\nu}_0 \left( \frac{G(\cdot + t) - G(\cdot)}{1 - G(\cdot)} \right) = \widehat{\nu}_0 \left( \int_0^t h(\cdot + r) \frac{1 - G(\cdot + r)}{1 - G(\cdot)} dr \right) \\ &= \widehat{\nu}_0 \left( \int_0^t \Phi_r h(\cdot) dr \right). \end{aligned}$$

By (5.7) and the boundedness of  $h$ ,  $\Phi_r h$  is bounded (uniformly in  $r$ ) and absolutely continuous, and Assumption 4 implies that  $\int_0^t \Phi_r h dr = (1 - G(\cdot + r))/(1 - G(\cdot))$  is Hölder continuous. Therefore, applying Assumption 5'(d) with  $\varphi = \Phi_r h$ , it follows that

$$(9.5) \quad \widehat{\nu}_0(\mathbf{1}) - \mathcal{S}_t^{\widehat{\nu}_0}(\mathbf{1}) = \int_0^t \widehat{\nu}_0(\Phi_r h) dr = \int_0^t \mathcal{S}_r^{\widehat{\nu}_0}(h) dr.$$

In a similar fashion, for  $t > 0$ , using the identity  $\widehat{\mathcal{H}}_t(\mathbf{1}) = \widehat{\mathcal{M}}_t(\Psi_t \mathbf{1})$  we have

$$\begin{aligned} \widehat{\mathcal{M}}_t(\mathbf{1}) - \widehat{\mathcal{H}}_t(\mathbf{1}) &= \int_{[0, L] \times [0, t]} \frac{G(x + t - u) - G(x)}{1 - G(x)} \widehat{\mathcal{M}}(dx, du) \\ &= \int_{[0, L] \times [0, t]} \left( \int_u^t \frac{h(x + r - u)(1 - G(x + r - u))}{1 - G(x)} dr \right) \widehat{\mathcal{M}}(dx, du). \end{aligned}$$

Because  $h \in \mathcal{C}_b[0, L]$ , we can set  $\tilde{\varphi} = h$  in (E.1) of Lemma E.1 to obtain

$$(9.6) \quad \widehat{\mathcal{M}}_t(\mathbf{1}) - \widehat{\mathcal{H}}_t(\mathbf{1}) = \int_0^t \widehat{\mathcal{H}}_r(h) dr.$$

If  $h$  is absolutely continuous, then  $g$  is absolutely continuous and by the commutativity of the convolution and differentiation operations, the function  $t \mapsto \int_0^t g(t - s)\widehat{K}(s) ds$  is absolutely continuous with derivative  $g(0)\widehat{K}(t) + \int_0^t g'(t - s)\widehat{K}(s) ds$ . Together with the relations (9.5) and (9.6) and the definition (5.23) of  $\widehat{D}$ , it follows that almost surely,  $\widehat{D}$  is absolutely continuous with respect to Lebesgue measure, and has density equal to

$$\frac{d\widehat{D}_t}{dt} = \mathcal{S}_t^{\widehat{\nu}_0}(h) - \widehat{\mathcal{H}}_t(h) + g(0)\widehat{K}(t) + \int_0^t g'(t - s)\widehat{K}(s) ds.$$

The relation (9.4) then follows on comparing the right-hand side above with the right-hand side of the equation (5.24) for  $\widehat{\nu}(f)$ , setting  $f = h$  therein and using the elementary relations  $h(0) = g(0)$  and  $g' = h'(1 - G) - hg$ .  $\square$

*Proof of Theorem 5.8.* From Lemma 9.2 and the discussion prior to it, it follows that  $\widehat{X}$  is a semimartingale with the decomposition stated in Theorem 5.8. Combining the non-idling condition (6.10) with the equation (6.8) for  $\widehat{K}$ , it follows that

$$(9.7) \quad \widehat{K}(t) = \begin{cases} \widehat{E}(t) & \text{if } \overline{X} \text{ is subcritical,} \\ \widehat{E}(t) + \widehat{x}_0 - \widehat{X}(t) \vee 0 & \text{if } \overline{X} \text{ is critical,} \\ \widehat{E}(t) + \widehat{x}_0 - \widehat{X}(t) & \text{if } \overline{X} \text{ is supercritical.} \end{cases}$$

Thus, in the subcritical case, the semimartingale decomposition of  $\widehat{K}$  follows from that of  $\widehat{E}$  (see Remark 5.1), whereas in the supercritical case the semimartingale decomposition of  $\widehat{K}$  follows from those of  $\widehat{X}$  and  $\widehat{E}$ . On the other hand, when  $\overline{X}$  is critical we need the additional observation that by Tanaka's formula,

$$(9.8) \quad \widehat{X}(t) \vee 0 = \widehat{x}_0 \vee 0 + \int_0^t \mathbb{1}_{\{\widehat{X}(s) > 0\}} d\widehat{X}_s + \frac{1}{2} L_t^{\widehat{X}}(0),$$

where  $L_t^{\widehat{X}}(0)$  is the local time of  $\widehat{X}$  at zero, over the interval  $[0, t]$ . When combined with (9.7), this provides the semimartingale decomposition of  $\widehat{K}$  in the critical case. When  $\widehat{K}$  is a semimartingale, the stochastic integration by parts formula for semimartingales shows that for every  $f \in \mathcal{AC}_b(\mathcal{E})$ ,

$$(9.9) \quad \widehat{\mathcal{K}}_s(f) = \int_{[0, s]} f(s-u)(1-G(s-u)) d\widehat{K}(u), \quad s \geq 0,$$

where the latter is the convolution integral with respect to the semimartingale  $\widehat{K}$ . Thus, we obtain (5.26) from (5.24).  $\square$

*Sketch of Justification of Remark 5.9.* By Corollary 8.7, if  $f$  is bounded and Hölder continuous, then  $\widehat{\mathcal{H}}^{(N)}(f) \Rightarrow \widehat{\mathcal{H}}(f)$  in  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  and  $\{\widehat{\mathcal{H}}_t(f), t \geq 0\}$  is a continuous process. We now argue that one can, in fact, show that  $\widehat{\mathcal{K}}^{(N)}(f) \Rightarrow \widehat{\mathcal{K}}(f)$  as  $N \rightarrow \infty$  for all Hölder continuous  $f$ . Given the semimartingale decomposition  $K = M^K + A^K$ , the integral on the right-hand side of the expression (9.9) for  $\widehat{\mathcal{K}}(f)$  can be decomposed into a stochastic convolution integral with respect to the local martingale  $M^K$  and a Lebesgue-Stieltjes convolution integral with respect to the finite variation process  $A^K$ . An argument exactly analogous to the one used in Lemma 8.6(1) to analyze  $\widehat{\mathcal{H}}(f)$  can then be used to analyze the stochastic convolution integral with respect to  $M^K$  and a similar, though simpler, argument can be used to study the convolution integral with respect to  $A^K$  to show, as in Lemma 8.6 and Corollary 8.7, that for  $f$  Hölder continuous and bounded,  $\widehat{\mathcal{K}}^{(N)}(f) \Rightarrow \widehat{\mathcal{K}}(f)$  as  $N \rightarrow \infty$ , and  $\widehat{\mathcal{K}}(f)$  admits a continuous version. When combined with the convergence in (5.27), it is easy to argue as in the proof of Theorem 5.6 that, in fact, the joint convergence  $(\mathcal{S}^{\widehat{\nu}_0^{(N)}}(f), \widehat{\mathcal{K}}^{(N)}(f), \widehat{\mathcal{H}}^{(N)}(f)) \Rightarrow (\mathcal{S}^{\widehat{\nu}_0}(f), \widehat{\mathcal{K}}(f), \widehat{\mathcal{H}}(f))$  holds. Due to (6.15), by the continuous mapping theorem, this implies that  $\widehat{\nu}^{(N)}(f) \Rightarrow \widehat{\nu}(f)$  in  $\mathcal{D}_{\mathbb{R}}[0, \infty)$ , where  $\widehat{\nu}(f)$  is continuous.  $\square$

**9.3. Stochastic Age Equation.** The focus of this section is the characterization of the limiting state process in terms of a stochastic partial differential equation (SPDE), which we have called the stochastic age equation in Definition 5.10. First, in Section 9.3.1 we establish a representation for integrals of functionals of the limiting centered age process  $\{\widehat{\nu}_s, s \geq 0\}$ . This representation is then used in Section 9.3.2 to show that  $\{\widehat{\nu}_t, t \geq 0\}$  is a solution to the stochastic age equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ . The proof of uniqueness of solutions to the stochastic age equation and the proof of Theorem 5.11(1) is presented in Section 9.3.3. Throughout the section we assume that the conditions of Theorem 5.11, namely Assumptions 1, 3 and 5', the conditions on the fluid limit and the boundedness and absolute continuity of  $h$ , are satisfied and state only additional assumptions when imposed.

9.3.1. *An Integral Representation.* We start by establishing an integral representation that results from the semimartingale property for  $\widehat{K}$ . In what follows, recall the definition of the operator  $\Theta_t$  given in (8.21). Also, note that if  $h$  is bounded then Assumption 2 and (by Remark 5.2) Assumption 4 are both satisfied. Since  $h$  is also absolutely continuous, by Theorem 5.8  $\widehat{K}$  is a semimartingale and  $\widehat{\nu}_t(f)$  is given by (5.26) for every  $f \in \mathcal{AC}_b[0, L]$ .

**Lemma 9.3.** *For any  $\varphi \in \mathcal{C}_b([0, L] \times [0, \infty))$  such that  $\varphi(\cdot, t)$  is Hölder continuous uniformly in  $t$  and absolutely continuous,  $\mathbb{P}$ -almost surely for every  $t > 0$ , we have*

$$(9.10) \quad \int_0^t \widehat{\nu}_s(\varphi(\cdot, s)) ds = \widehat{\nu}_0(\Theta_t(\varphi(\cdot, s))(\cdot, 0)) - \widehat{\mathcal{M}}_t(\Theta_t \varphi) \\ + \int_0^t \left( \int_u^t \varphi(s-u, s)(1-G(s-u)) ds \right) d\widehat{K}(u).$$

*Proof.* Setting  $t = s$  and  $f = \varphi(\cdot, s)$  in (5.26), then using the identities  $\mathcal{S}_s^{\widehat{\nu}_0} = \widehat{\nu}_0(\Phi_s \cdot)$ ,  $\widehat{\mathcal{H}}_s = \widehat{\mathcal{M}}_s(\Psi_s \cdot)$  and (9.9) and lastly integrating over  $s \in [0, t]$ , we obtain

$$(9.11) \quad \int_0^t \widehat{\nu}_s(\varphi(\cdot, s)) ds = \int_0^t \widehat{\nu}_0(\Phi_s \varphi(\cdot, s)) ds - \int_0^t \widehat{\mathcal{M}}_s(\Psi_s \varphi(\cdot, s)) ds \\ + \int_{[0, t]} \left( \int_{[0, s]} \varphi(s-u, s)(1-G(s-u)) d\widehat{K}(u) \right) ds.$$

From the definition (8.21) of  $\Theta_t$  and the fact that  $(\Psi_t f)(\cdot, 0) = \Phi_t f(\cdot)$ , it follows that

$$\Theta_t(\varphi(\cdot, s))(x, 0) = \int_0^t (\Psi_s \varphi(\cdot, s))(x, 0) ds = \int_0^t \Phi_s(\varphi(\cdot, s))(x) ds.$$

Together with Assumption 5'(d), this implies that

$$(9.12) \quad \int_0^t \widehat{\nu}_0(\Phi_s \varphi(\cdot, s)) ds = \widehat{\nu}_0(\Theta_t(\varphi(\cdot, s))(\cdot, 0)),$$

which shows that the first terms on the right-hand sides of (9.10) and (9.11) are equal. The corresponding equality of the second terms on the right-hand sides of (9.10) and (9.11) follows from (E.1), whereas the equality of the third terms follows from Fubini's theorem for stochastic integrals with respect to semimartingales (see, for example, (5.17) of Revuz and Yor [32]). This completes the proof of the lemma.  $\square$

9.3.2. *A Verification Lemma.* We now show that the process  $\widehat{\nu}$  of Theorem 5.7 is a solution to the stochastic age equation. For this, it will be convenient to introduce the function  $\psi_h$  defined as follows:  $\psi_h(x, t) \doteq \exp(r_h(x, t))$  for  $(x, t) \in [0, L] \times \mathbb{R}_+$ , where

$$(9.13) \quad r_h(x, t) \doteq \begin{cases} -\int_t^x h(u) du & \text{if } 0 \leq t \leq x, \\ -\int_0^{x-t} h(u) du & \text{if } 0 \leq x \leq t. \end{cases}$$

Since  $h = g/(1-G)$ , this implies that

$$(9.14) \quad \psi_h(x, t) = \begin{cases} \frac{1-G(x)}{1-G(x-t)} & \text{if } 0 \leq t \leq x, \\ \frac{1-G(x-t)}{1-G(x)} & \text{if } 0 \leq x \leq t. \end{cases}$$

If  $g$  is absolutely continuous, then  $G$  is continuously differentiable and  $\psi_h$  is bounded, absolutely continuous and satisfies

$$(9.15) \quad \frac{\partial \psi_h}{\partial x} + \frac{\partial \psi_h}{\partial t} = -h\psi_h$$

for a.e.  $(x, t) \in [0, L] \times \mathbb{R}_+$ . Furthermore, from the definition it is easy to see that  $\psi_h(0, s) = \psi_h(x, 0) = 1$  and, for  $(x, s) \in [0, L] \times [0, \infty)$  and  $u \in [0, s]$ ,

$$(9.16) \quad \frac{\psi_h(x + s - u, s)}{\psi_h(x, u)} = \frac{1 - G(x + s - u)}{1 - G(x)} = \begin{cases} \frac{1 - G(x + s)}{1 - G(x)} & \text{if } u = 0, \\ (1 - G(s - u)) & \text{if } x = 0. \end{cases}$$

**Proposition 9.4.** *If  $h$  is Hölder continuous, then the process  $\{\widehat{\nu}_t, t \geq 0\}$  defined by (5.26) satisfies the stochastic age equation associated with  $\{\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}}\}$ .*

*Proof.* Theorem 5.11 shows that for every  $t > 0$ ,  $\{\widehat{\nu}_t(f), f \in \mathcal{AC}_b[0, L]\}$  is a family of  $\widehat{\mathcal{F}}_t$ -measurable random variables and  $\{\widehat{\nu}_t, t \geq 0\}$  admits a version as an  $\{\widehat{\mathcal{F}}_t\}$ -adapted continuous,  $\mathbb{H}_{-2}$ -valued process. Moreover, it follows from Lemma 9.6 that for every  $f \in \mathcal{AC}_b[0, L]$ , almost surely  $s \mapsto \widehat{\nu}_s(f)$  is measurable. Therefore, it only remains to show that  $\widehat{\nu}$  satisfies the equation (5.28). Fix  $t \in [0, \infty)$  and  $\varphi \in \mathcal{C}_b^{1,1}([0, L] \times [0, \infty))$  such that  $\varphi_x(\cdot, s) + \varphi_s(\cdot, s)$  is Lipschitz continuous for every  $s$ . Since  $h$  is bounded, Hölder continuous and absolutely continuous, it follows that  $\varphi_x + \varphi_s - h\varphi$  is bounded, Hölder continuous and absolutely continuous. Moreover, it is clear from (9.15) that

$$(\varphi_x + \varphi_s - h\varphi)\psi_h = (\varphi\psi_h)_x + (\varphi\psi_h)_s.$$

Substituting this and the identity (9.16) into the definition (8.21) of  $\Theta_t$ , it follows that

$$(9.17) \quad \begin{aligned} (\Theta_t(\varphi_x + \varphi_s - h\varphi))(x, u) &= \int_u^t \frac{((\varphi_x + \varphi_s - h\varphi)\psi_h)(x + s - u, s)}{\psi_h(x, u)} ds \\ &= \int_u^t \frac{((\varphi\psi_h)_x + (\varphi\psi_h)_s)(x + s - u, s)}{\psi_h(x, u)} ds \\ &= \frac{\varphi(x + t - u, t)\psi_h(x + t - u, t)}{\psi_h(x, u)} - \varphi(x, u). \end{aligned}$$

Applying Lemma 9.3 with  $\varphi$  replaced by  $\varphi_x + \varphi_s - h\varphi$ , using (9.17) and the identity  $\psi_h(0, u) = 1$ , it follows that

$$(9.18) \quad \begin{aligned} &\int_0^t \widehat{\nu}_s(\varphi_x(\cdot, s) + \varphi_s(\cdot, s) - h\varphi(\cdot, s)) ds \\ &= \widehat{\nu}_0(\varphi(\cdot + t, t)\psi_h(\cdot + t, t)) + \int_{[0, t]} \varphi(t - u, t)\psi_h(t - u, t) d\widehat{K}(u) \\ &\quad - \int_{[0, L] \times [0, t]} \varphi(x + t - u, t) \frac{\psi_h(x + t - u, t)}{\psi_h(x, u)} \widehat{\mathcal{M}}(dx, du) \\ &\quad - \widehat{\nu}_0(\varphi(\cdot, 0)) - \int_{[0, t]} \varphi(0, u) d\widehat{K}(u) + \int_{[0, L] \times [0, t]} \varphi(x, u) \widehat{\mathcal{M}}(dx, du). \end{aligned}$$

Since  $\varphi$  is bounded and  $x \mapsto \varphi(x, s)$  is absolutely continuous for every  $s$ , by the definition (5.26) of  $\widehat{\nu}_t$  and the identities in (9.16), it is clear that the sum of the first three terms on the right-hand side of (9.18) equals  $\widehat{\nu}_t(\varphi(\cdot, t))$ . With this substitution, (9.18) reduces to the stochastic age equation (5.28). This completes the proof that  $\{\widehat{\nu}_t, t \geq 0\}$  is a solution to the stochastic age equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ .  $\square$

**9.3.3. Uniqueness of Solutions to the Stochastic Age Equation.** In order to establish uniqueness, we begin with a basic “variation of constants” transformation result. Recall from Section 1.4.1 that  $\mathcal{S}_c$  is the space of  $\mathcal{C}^\infty$  functions with compact support on  $[0, L]$  equipped with the same norm as  $\mathcal{S}$ . In what follows  $g'$  is the density of  $g$ .

**Lemma 9.5.** *Suppose that  $g' \in \mathbb{L}_{loc}^2[0, L] \cup \mathbb{L}_{loc}^\infty[0, L]$ . Given a solution  $\{\nu_t, t \geq 0\}$  to the stochastic age equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ , define*

$$(9.19) \quad \mu_t(\tilde{f}) \doteq \nu_t(\tilde{f}(1 - G)^{-1}), \quad \tilde{f} \in \mathcal{S}_c.$$

Then  $\{\mu_t, t \geq 0\}$  is a continuous  $\mathcal{S}'_c$ -valued process that satisfies the following stochastic transport equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ : for every  $\tilde{f} \in \mathcal{S}'_c$ ,  $t \geq 0$ ,

$$(9.20) \quad \mu_t(\tilde{f}) = \widehat{\nu}_0(\tilde{f}(1-G)^{-1}) + \int_0^t \mu_s(\tilde{f}_x) ds + \tilde{f}(0)\widehat{K}(t) - \widehat{\mathcal{M}}_t(\tilde{f}(1-G)^{-1}).$$

*Proof.* By the definition of the stochastic age equation,  $\{\nu_t, t \geq 0\}$  is a continuous  $\mathbb{H}_{-2}$ -valued process. By the assumptions on the service distribution  $G$  and Lemma B.1, it follows that  $f = \tilde{f}(1-G)^{-1}$  is an absolutely continuous function with compact support and hence lies in  $\mathbb{H}_2$ . Therefore,  $\mu_t(\tilde{f})$  is a well defined random variable for every  $f \in \mathcal{S}_c$ ,  $t > 0$ . Moreover,  $f$  has derivative

$$(9.21) \quad f_x = \tilde{f}_x(1-G)^{-1} + hf.$$

Since  $f \in \mathcal{C}_b^1[0, L]$ , we can substitute  $\varphi = f$  in the stochastic age equation (5.28), use (9.21) and the identity  $1 - G(0) = 1$  to obtain for  $t \geq 0$ ,

$$(9.22) \quad \begin{aligned} \mu_t(\tilde{f}) = \nu_t(\tilde{f}(1-G)^{-1}) &= \nu_0(\tilde{f}(1-G)^{-1}) + \int_0^t \nu_s(\tilde{f}_x(1-G)^{-1}) ds \\ &\quad + \tilde{f}(0)\widehat{K}(t) - \widehat{\mathcal{M}}_t(\tilde{f}(1-G)^{-1}). \end{aligned}$$

Due to the continuity of  $\widehat{K}$  and  $\widehat{\mathcal{M}}$ , it follows that the right-hand side is continuous in  $t$ , which in turn implies that  $t \mapsto \mu_t(\tilde{f})$  is continuous for each  $\tilde{f} \in \mathcal{S}_c$ . Since  $\mathcal{S}'_c$  is a Fréchet nuclear space, by Mitoma's theorem  $\mu$  is a continuous  $\mathcal{S}'_c$ -space valued process. Moreover, by (9.19)  $\nu_s(\tilde{f}_x(1-G)^{-1}) = \mu_s(\tilde{f}_x)$ . Substituting this back into (9.22), it follows that  $\{\mu_t, t \geq 0\}$  satisfies (9.20).  $\square$

We can now wrap up the proof of Theorem 5.11.

*Proof of Theorem 5.11.* By assumption,  $h$  is Hölder continuous. Therefore, Proposition 9.4 shows that  $\{\widehat{\nu}_t, t \geq 0\}$  is a solution to the stochastic age equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ . Thus, in order to establish the theorem, it suffices to show that the stochastic age equation has a unique solution. Suppose that the stochastic age equation associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$  has two solutions  $\nu^{(1)}$  and  $\nu^{(2)}$  and for  $i = 1, 2$ , let  $\mu^{(i)}$  be the corresponding continuous  $\mathcal{S}'_c$ -valued process defined as in (9.19), but with  $\nu$  replaced by  $\nu^{(i)}$ . By Lemma 9.5, each  $\mu^{(i)}$  satisfies the stochastic transport equation (9.20) associated with  $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ . Define  $\eta \doteq \mu^{(1)} - \mu^{(2)}$ . It follows that for every  $\tilde{f}$  in  $\mathcal{S}_c$ ,

$$\frac{d}{dt} \langle \tilde{f}, \eta_t \rangle - \langle \tilde{f}_x, \eta_t \rangle = 0, \quad \langle \tilde{f}, \eta_0 \rangle = 0.$$

However, this is simply a deterministic transport equation and it is well known that the unique solution to this equation is the identically zero solution  $\eta \equiv 0$  (see, for example, Theorem 4 on page 408 of [11]).

Thus, for every  $\tilde{f} \in \mathcal{S}_c$ ,  $\mu^{(1)}(\tilde{f}) = \mu^{(2)}(\tilde{f})$  or equivalently,

$$(9.23) \quad \nu_t^{(1)}(\tilde{f}(1-G)^{-1}) = \nu_t^{(2)}(\tilde{f}(1-G)^{-1}).$$

Now for any  $f \in \mathcal{S}_c$ ,  $f(1-G) \in \mathbb{H}_2$ . Since  $\mathcal{S}_c$  is dense in  $\mathbb{H}_2$  (see Theorem 3.18 on page 54 of [1]), there exists a sequence  $\tilde{f}_n \in \mathcal{S}_c$  such that  $\tilde{f}_n \rightarrow f(1-G)$  in  $\mathbb{H}_2$  as  $n \rightarrow \infty$ . Replacing  $\tilde{f}$  by  $\tilde{f}_n$  in (9.23) and then letting  $n \rightarrow \infty$ , it follows that  $\nu_t^{(1)}(f) = \nu_t^{(2)}(f)$  for every  $f \in \mathcal{S}_c$ . Again using the fact that  $\mathcal{S}_c$  is dense in  $\mathbb{H}_2$ , this shows  $\nu_t^{(1)}$  and  $\nu_t^{(2)}$  are indistinguishable as  $\mathbb{H}_{-2}$ -valued elements. This proves uniqueness of solutions to the stochastic age equation and the theorem follows.  $\square$

**9.4. The Strong Markov Property.** To prove the strong Markov property, we first show that the assumptions on the initial centered age distribution imposed in Assumption 5 are consistent, in the sense that they imply that these assumptions are also satisfied at any future time  $s > 0$ . We define the following shifted processes: for  $F = \widehat{E}^{(N)}, \widehat{K}^{(N)}, \widehat{E}, \widehat{K}$ , and  $\mathcal{U} = \widehat{\mathcal{M}}^{(N)}, \widehat{\mathcal{M}}$ , and  $s \geq 0, u \geq 0$ ,

$$(9.24) \quad (\Theta_s F)(u) \doteq F(s+u) - F(s), \quad (\Theta_s \mathcal{U})_u \doteq \mathcal{U}_{s+u} - \mathcal{U}_s,$$

for  $f \in \mathcal{C}_b[0, L)$ , we define

$$(9.25) \quad (\Theta_s \widehat{\mathcal{H}}^{(N)})_t(f) \doteq (\Theta_s \widehat{\mathcal{M}}^{(N)})_t(\Psi_t f), \quad (\Theta_s \widehat{\mathcal{H}})_t(f) \doteq (\Theta_s \widehat{\mathcal{M}})_t(\Psi_t f),$$

$$(9.26) \quad (\Theta_s \widehat{\mathcal{K}}^{(N)})_t(f) = \int_{[0, t]} (1 - G(t-u)) f(t-u) d(\Theta_s \widehat{K}^{(N)})(u)$$

$$(9.27) \quad (\Theta_s \widehat{\mathcal{K}})_t(f) \doteq f(0)(\Theta_s \widehat{K})(t) + \int_0^t (\Theta_s \widehat{K})(u) \varphi_f(t-u) du,$$

and, in analogy with (5.11), for  $f \in \mathcal{AC}_b[0, L)$  we define

$$(9.28) \quad \mathcal{S}_t^{\widehat{\nu}_s}(f) \doteq \widehat{\nu}_s(\Phi_t f), \quad s, t \geq 0.$$

**Lemma 9.6.** *For every bounded and continuous  $f$ ,*

$$(9.29) \quad \widehat{\nu}_{s+t}^{(N)}(f) = \mathcal{S}_t^{\widehat{\nu}_s^{(N)}}(f) + (\Theta_s \widehat{\mathcal{K}}^{(N)})_t(f) - (\Theta_s \widehat{\mathcal{H}}^{(N)})_t(f), \quad s, t \geq 0.$$

*Likewise, if Assumptions 1–5 hold and  $g$  is continuous, then for every bounded and absolutely continuous  $f$ ,*

$$(9.30) \quad \widehat{\nu}_{s+t}(f) = \mathcal{S}_t^{\widehat{\nu}_s}(f) + (\Theta_s \widehat{\mathcal{K}})_t(f) - (\Theta_s \widehat{\mathcal{H}})_t(f), \quad s, t \geq 0.$$

*In addition, for every  $s > 0$ ,*

$$(9.31) \quad (\Theta_s \widehat{K}, \widehat{X}_{s+}, \widehat{\nu}_{s+})(\mathbf{1}) = \Lambda(\Theta_s \widehat{E}, \widehat{X}(s), \mathcal{S}^{\widehat{\nu}_s}(\mathbf{1}) - (\Theta_s \widehat{\mathcal{H}})(\mathbf{1})).$$

*Furthermore, for every  $s > 0$ , Assumption 5' holds with the sequence  $\{\widehat{\nu}_0^{(N)}\}_{N \in \mathbb{N}}$  and limit  $\widehat{\nu}_0$ , respectively, replaced by  $\{\widehat{\nu}_s^{(N)}\}_{N \in \mathbb{N}}$  and  $\widehat{\nu}_s$ .*

We defer the proof of this lemma to Appendix E, and instead now prove the strong Markov property of the state process.

*Proof of Theorem 5.11(2).* Fix  $s, t > 0$ . First, note that by Theorem 5.7 it follows that  $(\widehat{X}, \widehat{\nu})$  is an  $\mathbb{R} \times \mathbb{H}_{-2}$ -valued process. Moreover, by Lemma 9.6, Assumption 5 is satisfied with  $\widehat{\nu}_0$  replaced by  $\widehat{\nu}_s$ , which in particular implies that the random element  $\mathcal{S}_{t+}^{\widehat{\nu}_s}(\mathbf{1}) = \{\mathcal{S}_{t+u}^{\widehat{\nu}_s}(\mathbf{1}), u \geq 0\}$  almost surely takes values in  $\mathcal{C}_{\mathbb{R}}[0, \infty)$ . Also, let  $(\Theta_s \widehat{\mathcal{H}})_t(\Phi \cdot \mathbf{1})$  represent the process  $\{(\Theta_s \widehat{\mathcal{H}})_t(\Phi_u \mathbf{1}), u \geq 0\}$ . Writing  $\Phi_u \mathbf{1} = \int_0^u \Phi_r h(\cdot) dr$  and observing that  $\Phi_u \mathbf{1}$  is bounded and (due to Assumption 4) Hölder continuous uniformly in  $u$ , it follows from Lemma 8.6(3) (with  $\widehat{\mathcal{M}}$  replaced by  $\Theta_s \widehat{\mathcal{M}}$ ) that the random element  $(\Theta_s \widehat{\mathcal{H}})_t(\Phi \cdot \mathbf{1})$  takes values in  $\mathcal{C}_{\mathbb{R}}[0, \infty)$ . In addition, Assumption 3 and Corollary 8.7 show that  $\Theta_s \widehat{E}$  and  $\Theta_s \widehat{\mathcal{H}}$  are, respectively,  $\mathcal{C}_{\mathbb{R}}[0, \infty)$ -valued and  $\mathcal{C}_{\mathbb{H}_{-2}}[0, \infty)$ -valued. We now claim that there exists a continuous mapping from  $\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{C}_{\mathbb{R}}[0, \infty)^3 \times \mathcal{C}_{\mathbb{H}_{-2}}[0, \infty)$  to  $\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{C}_{\mathbb{R}}[0, \infty)$ , which we denote by  $\tilde{\Lambda} = \tilde{\Lambda}_t$ , such that  $\mathbb{P}$ -almost surely,

$$(9.32) \quad (\widehat{X}_{s+t}, \widehat{\nu}_{s+t}, \mathcal{S}^{\widehat{\nu}_{s+t}}(\mathbf{1})) = \tilde{\Lambda} \left( \widehat{X}(s), \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}_s}(\mathbf{1}), \Theta_s \widehat{E}, (\Theta_s \widehat{\mathcal{H}})_t(\Phi \cdot \mathbf{1}), \Theta_s \widehat{\mathcal{H}} \right).$$

To see why this is the case, first note that equation (9.30) shows that  $\widehat{\nu}_{s+t}$  is the sum of  $(\Theta_s \widehat{\mathcal{H}})_t, \mathcal{S}_t^{\widehat{\nu}_s}$  and  $(\Theta_s \widehat{\mathcal{K}})_t$  and, by Lemma B.1(2),  $\mathcal{S}_t^{\widehat{\nu}_s}$  is a continuous functional of  $\widehat{\nu}_s$ . Also, for  $u > 0, \Phi_u \mathbf{1}$

is bounded and absolutely continuous. Hence, by (9.30), the definition of  $\mathcal{S}^\nu$  and the semigroup property for  $\Phi$ ,  $\mathbb{P}$ -almost surely, for  $u, s, t \geq 0$ ,

$$\begin{aligned}\mathcal{S}_u^{\widehat{\nu}^{s+t}}(\mathbf{1}) &= \mathcal{S}_t^{\widehat{\nu}^s}(\Phi_u \mathbf{1}) + (\Theta_s \widehat{\mathcal{K}})_t(\Phi_u \mathbf{1}) - (\Theta_s \widehat{\mathcal{H}})_t(\Phi_u \mathbf{1}) \\ &= \mathcal{S}_{t+u}^{\widehat{\nu}^s}(\mathbf{1}) + (\Theta_s \widehat{\mathcal{K}})_t(\Phi_u \mathbf{1}) - (\Theta_s \widehat{\mathcal{H}})_t(\Phi_u \mathbf{1}).\end{aligned}$$

Next, note that (9.31) of Lemma 9.6, Proposition 7.3 and the continuity of  $\widehat{X}$  show that  $\widehat{X}_{s+t}$  is a continuous functional of  $(\Theta_s \widehat{E}, \widehat{X}(s), \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1}) - \Theta_s \widehat{\mathcal{H}}(\mathbf{1}))$ . Furthermore, due to the almost sure continuity of  $\widehat{K}$  established in Theorem 5.6, definitions (7.1) and (9.27) of  $\mathcal{K}$  and  $(\Theta_s \widehat{K})$ , respectively, and properties 2 and 3 of Lemma 7.1, it follows that  $\Theta_s \widehat{\mathcal{K}}$  and  $u \mapsto (\Theta_s \mathcal{K})_t(\Phi_u \mathbf{1})$  are almost surely obtained as continuous mappings of  $\Theta_s \widehat{K}$ . In turn, by (9.31) of Lemma 9.6,  $\Theta_s \widehat{K} = \Lambda(\Theta_s \widehat{E}, \widehat{X}(s), \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1}) - \Theta_s \widehat{\mathcal{H}}(\mathbf{1}))$ , where  $\Lambda$  is continuous by Proposition 7.3. When combined, the above observations show that the claim (9.32) holds, with  $\tilde{\Lambda}$  a suitable continuous mapping.

We now show that the claim implies the Markov property. First, from (5.4) we observe that  $\Theta_s \widehat{E}$  is adapted to the filtration generated by  $\Theta_s B$  and, likewise, (9.25) shows that  $\Theta_s \widehat{\mathcal{H}}$  is adapted to the filtration generated by  $\Theta_s \widehat{\mathcal{M}}$ . Moreover, by the definition of  $B$  as a standard Brownian motion and the definition of  $\widehat{\mathcal{M}}$  (see Section 4.3 and Remark 5.1), both  $B$  and  $\widehat{\mathcal{M}}$  are processes with independent increments with respect to the filtration  $\{\widehat{\mathcal{F}}_t, t \geq 0\}$ , which is the right continuous completion of  $\{\sigma(\widehat{\nu}_0, \widehat{X}(0)) \vee \sigma(B_s, \widehat{\mathcal{M}}_s, s \geq t), t \geq 0\}$ . In particular, this implies that  $\Theta_s B$ ,  $\Theta_s \widehat{\mathcal{H}}$  and  $u \mapsto (\Theta_s \widehat{\mathcal{H}})_t(\Phi_u \mathbf{1})$  are independent of  $\widehat{\mathcal{F}}_s$ . Therefore, for any bounded continuous function  $F$  on  $[0, \infty) \times (\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{C}_{\mathbb{R}}[0, \infty))$ ,

$$\begin{aligned}\mathbb{E}[F(s, \widehat{X}_{s+t}, \widehat{\nu}_{s+t}, \mathcal{S}^{\widehat{\nu}^{s+t}}(\mathbf{1})) | \widehat{\mathcal{F}}_s] \\ &= \mathbb{E}\left[F(s, \tilde{\Lambda}(\widehat{X}(s), \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1}), \Theta_s \widehat{E}, (\Theta_s \widehat{\mathcal{H}})_t(\Phi \cdot \mathbf{1}), \Theta_s \widehat{\mathcal{H}}(\mathbf{1}))) | \widehat{\mathcal{F}}_s\right] \\ &= \mathbb{E}\left[F(s, \tilde{\Lambda}(\widehat{X}(s), \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1}), \Theta_s \widehat{E}, (\Theta_s \widehat{\mathcal{H}})_t(\Phi \cdot \mathbf{1}), \Theta_s \widehat{\mathcal{H}}(\mathbf{1}))) | \widehat{X}(s), \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1})\right] \\ &= \mathbb{E}[F(s, \widehat{X}_{s+t}, \widehat{\nu}_{s+t}, \mathcal{S}^{\widehat{\nu}^{s+t}}(\mathbf{1})) | \widehat{X}(s), \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1})].\end{aligned}$$

This shows that  $\{(\widehat{X}_s, \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1})), \widehat{\mathcal{F}}_s, s \geq 0\}$  is a Markov process.

By Theorems 5.6 and 5.7, the sample paths  $s \mapsto (\widehat{X}(s), \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1}))$  taking values in the state space  $\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{C}_{\mathbb{R}}[0, \infty)$  are continuous. Since the state space is a complete, separable metric space, there exists a Markov kernel  $P : \mathbb{R}_+ \times (\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{C}_{\mathbb{R}}[0, \infty)) \times \mathbb{R}_+ \times \mathcal{B}(\mathbb{R} \times \mathbb{H}_{-2}) \mapsto [0, 1]$  such that for any  $(x, \nu, \psi) \in (\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{C}_{\mathbb{R}}[0, \infty))$ , and any measurable function  $F$  on  $\mathbb{R}_+ \times (\mathbb{R} \times \mathbb{H}_{-2} \times \mathcal{C}_{\mathbb{R}}[0, \infty))$ ,

$$\begin{aligned}\mathbb{E}[F(s, \widehat{X}_{s+t}, \widehat{\nu}_{s+t}, \mathcal{S}^{\widehat{\nu}^{s+t}}(\mathbf{1})) | (\widehat{X}(s), \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1})) = (x, \nu, \psi)] \\ = \int_{\mathbb{R}_+ \times \mathbb{H}_{-2}} F(s, u) P(s, (x, \nu, \psi), t, du).\end{aligned}$$

Now, when  $F$  is bounded and continuous, it follows from (9.32) and the continuity of  $\tilde{\Lambda}$  established above that the mapping from  $(x, \nu, \psi)$  to the term on the right-hand side of the last display is continuous. This implies that the Markov kernel is Feller, and it follows from Theorem 2.4 of Friedman [12] that  $\{(\widehat{X}_s, \widehat{\nu}_s, \mathcal{S}^{\widehat{\nu}^s}(\mathbf{1})), \widehat{\mathcal{F}}_s, s \geq 0\}$  is a strong Markov process (see also Theorem 7.4 of Chapter I of [34] for the time-homogeneous case, which would be applicable if the arrival process satisfies Assumption 3(a) and the fluid age measure starts at the equilibrium measure so that  $\bar{\nu}_s(dx) = (1 - G(x)) dx$  and  $\bar{X}(s) = 1$ ,  $s \geq 0$ , leading to a critical fluid limit)  $\square$

**Remark 9.7.** A more natural candidate for the (strong) Markov process would be the process  $\{(\widehat{X}_t, \widehat{\nu}_t), \mathcal{F}_t, t \geq 0\}$  taking values in  $\mathbb{R} \times \mathbb{H}_{-2}$ . However, in order to establish the Markov property, we need  $\mathcal{S}^{\widehat{\nu}^s}$  and  $\mathcal{S}^{\widehat{\nu}^s}(\mathbf{1})$  to be measurable functions of  $\widehat{\nu}_s$ . As shown in Lemma B.1, the additional boundedness assumption on  $g'/(1 - G)$  ensures that the map from  $\mathbb{H}_{-2} \mapsto \mathcal{D}_{\mathbb{H}_{-2}}[0, \infty)$  that takes  $\widehat{\nu}_s$  to  $\mathcal{S}^{\widehat{\nu}^s}$  is continuous. This is a reasonable assumption because, as noted in Remark 5.12, it



is satisfied by a large class of distributions of interest. However, unfortunately, it appears that measurability of the map from  $\mathbb{H}_{-2}$  to  $\mathcal{D}_{\mathbb{R}}[0, \infty)$  that takes  $\widehat{\nu}_s$  to  $\mathcal{S}^{\widehat{\nu}_s}(\mathbf{1}) = \widehat{\nu}_s(\Phi_s \mathbf{1})$ , which would require that  $\Phi_s \mathbf{1}$  lies in  $\mathbb{H}_2$ , cannot be obtained without imposing too severe assumptions on the service distribution  $G$ . Although  $\Phi_s \mathbf{1} \in \mathcal{AC}_b[0, L]$  and  $\{\widehat{\nu}_s(f), f \in \mathcal{AC}_b[0, L]\}$  is a well defined collection of random variables, it is not clear whether it is possible to show that  $\widehat{\nu}_s$  admits a version that takes values in the dual of some space that contains  $\Phi_s \mathbf{1}$  and such that the dual space admits a regular conditional probability so as to enable the construction of the Markov kernel. Instead, we resolve this issue by appending the  $\mathcal{D}_{\mathbb{R}}[0, \infty)$ -valued process  $\mathcal{S}^{\widehat{\nu}_s}(\mathbf{1})$  to the state descriptor.

## APPENDIX A. PROPERTIES OF THE MARTINGALE MEASURE SEQUENCE

**A.1. Proof of the Martingale Measure Property.** Recall that  $\mathcal{B}_0[0, L]$  is the algebra generated by the intervals  $[0, x]$ ,  $x \in [0, L]$ . We now show that the collection of random variables  $\{\mathcal{M}_t^{(N)}(B); t \geq 0, B \in \mathcal{B}_0[0, L]\}$  introduced in (4.5) defines a martingale measure.

**Lemma A.1.** *For each  $N \in \mathbb{N}$ ,  $\mathcal{M}^{(N)} = \{\mathcal{M}_t^{(N)}(B), \mathcal{F}_t^{(N)}; t \geq 0, B \in \mathcal{B}_0[0, L]\}$  is a martingale measure on  $[0, L]$ . Moreover, for every  $B \in \mathcal{B}_0[0, L]$  and  $t \in [0, \infty)$ ,*

$$(A.1) \quad \mathbb{E} \left[ \left( \mathcal{M}_t^{(N)}(B) \right)^2 \right] = \mathbb{E} \left[ \int_0^t \left( \int_B h(x) \nu_s^{(N)}(dx) \right) ds \right].$$

*Proof.* In order to show that  $\{\mathcal{M}_t^{(N)}(B); t \geq 0, B \in \mathcal{B}_0[0, L]\}$  defines a martingale measure on  $[0, L]$ , we verify the three properties stated in the definition of a martingale measure given on page 287 of Walsh [35]. The first property in [35], namely that  $\mathcal{M}_0^{(N)}(B) = 0$  for every  $B \in \mathcal{B}_0$ , follows trivially from the definition. Next, we verify the third property, which states that  $\{\mathcal{M}_t^{(N)}(B), \mathcal{F}_t^{(N)}, t \geq 0\}$  is a local martingale for each  $B \in \mathcal{B}_0$ . For this, first observe that any  $B \in \mathcal{B}_0[0, L]$  is the union of a finite number of disjoint intervals  $I_i$ ,  $i = 1, \dots, k$ , where each  $I_i$  is of the form  $(\alpha_i, L)$ ,  $(\alpha_i, \beta_i]$  or  $[0, \beta_i]$ , with  $0 < \alpha_i < \beta_i < L$ . For any such interval  $I_i$ , it is clear from the definition of the age process given in (2.5) that for every  $j$ , the function  $s \mapsto \mathbb{1}_{I_i}(a_j^{(N)}(s))$  defines a bounded, left continuous function on  $[0, \infty)$ . In turn, since  $\mathbb{1}_B = \sum_{i=1}^k \mathbb{1}_{I_i}$ , clearly the function  $s \mapsto \mathbb{1}_B(a_j^{(N)}(s))$  is also bounded and left continuous. By Lemma 5.2 of Kang and Ramanan [20], it then follows that for every  $B \in \mathcal{B}_0[0, L]$ ,  $\{\mathcal{M}_t^{(N)}(B), t \geq 0\}$  is an  $\{\mathcal{F}_t^{(N)}\}$ -martingale obtained as a compensated sum of jumps, where the compensator  $A_{\mathbb{1}_B}^{(N)}$  is continuous. Standard arguments (see, for example, the proof of Lemma 5.9 in Kaspi and Ramanan [22]) then show that  $\langle \mathcal{M}^{(N)}(B) \rangle$ , the predictable quadratic variation of  $\mathcal{M}^{(N)}(B)$ , equals  $A_{\mathbb{1}_B}^{(N)}$ . Since  $\mathbb{E}[A_{\mathbb{1}_B}^{(N)}(t)]$  is dominated by  $\mathbb{E}[D^{(N)}(t)]$ , which is finite by Lemma 5.6 of Kaspi and Ramanan [22], the relation (A.1) follows. On the other hand, because  $\{\nu_t^{(N)}, t \geq 0\}$  is an  $\mathbb{M}_F[0, L]$ -valued process, this shows that the set function  $B \mapsto \mathbb{E}[(\mathcal{M}_t^{(N)}(B))^2]$  is countably additive on  $\mathcal{B}_0[0, L]$ , and hence defines a finite  $\mathbb{L}^2(\Omega, \mathcal{F}^{(N)}, \mathbb{P})$ -valued measure. This verifies the second property in [35], and thus completes the proof of the lemma.  $\square$

**A.2. Proof of Lemma 4.1.** As we will show below, Lemma 4.1 is essentially a consequence of the strong Markov property of the state process, the continuity of the  $\{\mathcal{F}_t\}$ -compensator of the departure process and the independence assumptions on the service times and arrival process.

Fix  $N \in \mathbb{N}$  and, for conciseness, we suppress  $N$  from the notation. We shall first prove (4.6), namely we will show that almost surely,  $\Delta D(t) \leq 1$  for every  $t \in [0, \infty)$ . For  $k = -\langle \mathbf{1}, \nu_0 \rangle + 1, \dots$ , let  $\mathcal{E}_k$  denote the event that the departure time of customer  $k$  lies in the set of the union of departure times of customers  $j$ ,  $j < k$ . To establish (4.6), it is clearly sufficient to show that  $\mathbb{P}(\mathcal{E}_k) = 0$  for every  $k$ . Fix  $k \in \mathbb{N}$  and let  $\theta_k$  be the  $\{\mathcal{F}_t\}$ -stopping time

$$\theta_k \doteq \inf\{t : K(t) = k\}.$$

Now, consider a modified system with initial data  $\tilde{\nu}_0 = \nu_{\theta_k}$ ,  $\tilde{X}(0) = \langle \mathbf{1}, \nu_{\theta_k} \rangle$  and  $\tilde{E} \equiv 0$ . By Lemma B.1 of Kang and Ramanan [20],  $\{(R_E(t), X(t), \nu_t), t \geq 0\}$  is a strong Markov process. Therefore,

conditioned on  $\mathcal{F}_{\theta_k}$ , the departure times of customers  $j$ ,  $j \leq k$ , only depend on  $\{a_j(\theta_k), j \leq k\}$  and are independent of arrivals after  $\theta_k$ . Consequently, the probability of the event  $\mathcal{E}_k$  is equal in both the original and modified systems. In the modified system, let  $\{\tilde{a}_j(s), s \in [0, \infty)\}$  denote the age process of customer  $j$  for  $j \leq k$ , let  $\tilde{D}^{\theta_k}(s)$  denote the cumulative departures in the time  $[0, s]$  of all customers other than customer  $k$  and let  $\tilde{J}^k \doteq \{s \in [0, \infty) : \tilde{D}^{\theta_k}(s) \neq \tilde{D}^{\theta_k}(s-)\}$  be the jump times of  $\tilde{D}^{\theta_k}$ . Also, let  $\tilde{\mathcal{G}}_t^k \doteq \sigma(\tilde{a}_j(s), j < k, s \in [0, t])$  and let  $\{\mathcal{G}_t^k, t \geq 0\}$  be the right continuous completion (with respect to  $\mathbb{P}$ ) of  $\{\tilde{\mathcal{G}}_t^k, t \geq 0\}$ . By the assumed independence of the service times for different customers and the fact that  $\tilde{a}_k(0) = 0$ , the departure time  $\tilde{v}_k$  of customer  $k$  in the modified system has cumulative distribution function  $G$  and is independent of  $\tilde{J}^k$ . Therefore,

$$(A.2) \quad \begin{aligned} \mathbb{P}(\mathcal{E}_k) &= \mathbb{P}(\tilde{v}_k \in \tilde{J}^k) \\ &= \int_{[0, L)} \mathbb{P}(t \in \tilde{J}^k | \tilde{v}_k = t) dG(t) = \int_{[0, L)} \mathbb{P}(t \in \tilde{J}^k) dG(t), \end{aligned}$$

where the last equality follows from the independence of  $\tilde{v}_k$  and  $\tilde{J}^k$ . The same logic used in Lemma 5.4 of Kaspi and Ramanan [22] to identify the compensator of  $D$  also shows that the  $\{\mathcal{G}_t^k\}$ -compensator of  $\tilde{D}^{\theta_k}$  equals

$$\int_0^\cdot \left( \int_{[0, L)} \frac{g(x+s)}{1-G(x)} \nu'_0(dx) \right) ds, \quad \text{where } \nu'_0 \doteq \tilde{\nu}_0 - \delta_0,$$

where the mass at zero is deleted from the modified age measure  $\tilde{\nu}_0$  to remove customer  $k$ , which has age zero at time 0 in the modified system. By the continuity of the  $\{\mathcal{G}_t^k\}$ -compensator of  $\tilde{D}^{\theta_k}$ ,  $\tilde{D}^{\theta_k}$  is quasi-left-continuous and so  $\Delta \tilde{D}^{\theta_k}(T) = 0$  for every  $\{\mathcal{G}_t^k\}$ -predictable time  $T$  (see, for example, Theorem 4.2 and Definition 2.25 of Chapter I of Jacod and Shiryaev [17]). Choosing  $T$  to be the deterministic time  $t$ , this implies that  $\mathbb{P}(t \in \tilde{J}^k) = 0$  for every  $t \geq 0$ . When substituted into (A.2), this shows that  $\mathbb{P}(\mathcal{E}_k) = 0$ . For  $k \leq 0$ , we set  $\theta_k = 0$  and observe that, conditioned on  $\mathcal{F}_0$ , the departure time  $\tilde{v}_k$  of the  $k$ th customer has cumulative distribution function  $\tilde{G}(\cdot) \doteq (G(\cdot) - G(a_k(0)))/(1 - G(a_k(0)))$ , rather than  $G$ , so that (A.2) holds with  $G$  replaced by  $\tilde{G}$ . The rest of the proof follows exactly as in the case  $k > 0$ , and thus (4.6) holds.

We now turn to the proof of (4.7). Fix  $r, s \in [0, \infty)$ , recall that  $D^r(s)$  is the cumulative departures in the interval  $[r, r+s)$  of customers that entered service at or before time  $r$ , define  $J^r$  to be the jump times of  $D^r$  in  $[0, \infty)$  and let  $\mathcal{G}_t = \mathcal{F}_{r+t}$ ,  $t \in [0, \infty)$ . Using the same logic as in the proof of (4.6), it can be shown that  $\{D^r(t), t \geq 0\}$  has a continuous  $\{\mathcal{G}_t\}$ -compensator, given explicitly by

$$\int_0^t \left( \int_{[0, L)} \frac{g(x+s)}{1-G(x)} \nu_r(dx) \right) ds, \quad t \in [0, \infty),$$

and hence has no fixed jump times, i.e.,  $\mathbb{P}(t \in J^r | \mathcal{F}_r) = 0$  for every  $t \in \mathbb{R}_+$ . Moreover, due to the assumption of independence of the arrival processes and the service times,  $\{E(t)\}_{t \geq r}$  and  $\{D^r(t), t \geq 0\}$  are conditionally independent, given  $\mathcal{F}_r$ . Let

$$\mathcal{T} \doteq \{\bar{t} = (t_1, \dots, t_m, \dots) \in \mathbb{R}_+^\infty : 0 \leq t_1 \leq t_2 \leq \dots\},$$

and let  $\bar{T}^r$  denote the random  $\mathcal{T}$ -valued sequence of times after  $r$  at which  $E$  has a jump. Moreover, let  $\mu$  denote the conditional probability distribution of  $\bar{T}^r$ , given  $\mathcal{F}_r$ . Then, for any  $\bar{t} \in \mathcal{T}$ , using the fact that  $0 \leq \Delta E(t) \leq 1$ , we have

$$\begin{aligned} \mathbb{E} \left[ \sum_{s \in [0, \infty)} \Delta E(r+s) \Delta D^r(s) | \mathcal{F}_r, \bar{T}^r = \bar{t} \right] &= \sum_{t \in \bar{t}} \mathbb{E}[\Delta D^r(t) | \mathcal{F}_r, \bar{T}^r = \bar{t}] \\ &= \sum_{t \in \bar{t}} \mathbb{E}[\Delta D^r(t) | \mathcal{F}_r] \\ &= 0, \end{aligned}$$

where the second equality uses the conditional independence of  $\{D^r(t), t \geq 0\}$  from  $\bar{T}^r$  given  $\mathcal{F}_r$ , and the last equality follows because as argued above, conditional on  $\mathcal{F}_r$ ,  $D^r$  almost surely has no fixed jumps. In turn, integrating the left-hand side above with respect to the conditional distribution  $\mu$  and then taking expectations, it follows that

$$\mathbb{E} \left[ \sum_{s \in [0, \infty)} \Delta E(r+s) \Delta D^r(s) \right] = 0.$$

Since the term inside the expectation is non-negative, this proves (4.7).

**A.3. A Consequence of Lemma 4.1.** We now establish a consequence of Lemma 4.1, which will be used in the proof of the asymptotic independence property in Section 8.2.

**Corollary A.2.** *As  $N \rightarrow \infty$ ,*

$$\frac{1}{N} \mathbb{E} \left[ \sum_{s \leq t} \Delta E^{(N)}(s) \Delta D^{(N)}(s) \right] \rightarrow 0.$$

*Proof.* With the aim of computing the left-hand side above, using the same notation as in Lemma 4.1, for  $r > 0$  and  $s \geq r$ , let  $D^{(N),r}(s)$  denote the cumulative number of departures during  $(r, s]$  of customers that entered service at or before time  $r$ , and let  $D^{(N)+,r}(s)$  be the cumulative number of departures during  $(r, s]$  of customers that have entered service after time  $r$ . Then for  $\delta > 0$  and  $k = 1, 2, \dots$ , we have

$$\begin{aligned} \sum_{s \in (k\delta, (k+1)\delta]} \Delta E^{(N)}(s) \Delta D^{(N)}(s) &= \sum_{s \in (k\delta, (k+1)\delta]} \Delta E^{(N)}(s) \Delta D^{(N),k\delta}(s) \\ &\quad + \sum_{s \in (k\delta, (k+1)\delta]} \Delta E^{(N)}(s) \Delta D^{(N)+,k\delta}(s) \end{aligned}$$

The first summand on the right-hand side above is almost surely zero by (4.7) of Lemma 4.1. Using the fact that  $E^{(N)}$  has unit jump sizes to bound the second term, we obtain

$$\begin{aligned} \text{(A.3)} \quad \sum_{s \in (k\delta, (k+1)\delta]} \Delta E^{(N)}(s) \Delta D^{(N)}(s) &\leq \sum_{s \in (k\delta, (k+1)\delta]} \Delta D^{(N)+,k\delta}(s) \\ &\leq \sum_{j=K^{(N)}(k\delta)+1}^{K^{(N)}((k+1)\delta)} \mathbb{1}_{\{v_j \leq \delta\}}. \end{aligned}$$

Summing (A.3) over  $k = 1, \dots, \lfloor t/\delta \rfloor$  and dividing by  $N$ , we obtain

$$\frac{1}{N} \mathbb{E} \left[ \sum_{s \leq t} \Delta E^{(N)}(s) \Delta D^{(N)}(s) \right] \leq \mathbb{E} \left[ \int_0^t \mathbb{1}_{\{v_{K^{(N)}(s)} \leq \delta\}} d\bar{K}^{(N)}(s) \right] \leq \mathbb{E} \left[ \bar{Q}_{\mathbb{1}_{[0,\delta]}}^{(N)}(t + \delta) \right].$$

For each  $\delta > 0$ , let  $f_\delta$  be any continuous bounded function on  $[0, L]$  such that  $\mathbb{1}_{[0,\delta]} \leq f_\delta \leq \mathbb{1}_{[0,2\delta]}$ . Then we have

$$\frac{1}{N} \mathbb{E} \left[ \sum_{s \leq t} \Delta E^{(N)}(s) \Delta D^{(N)}(s) \right] \leq \mathbb{E} \left[ \bar{Q}_{f_\delta}^{(N)}(t + \delta) \right] = \mathbb{E} \left[ \bar{A}_{f_\delta}^{(N)}(t + \delta) \right],$$

On both sides, taking first the limit supremum as  $N \rightarrow \infty$ , and then the limit as  $\delta \downarrow 0$ , we then conclude that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \left[ \sum_{s \leq t} \Delta E^{(N)}(s) \Delta D^{(N)}(s) \right] \leq \lim_{\delta \downarrow 0} \limsup_{N \rightarrow \infty} \mathbb{E} \left[ \bar{A}_{f_\delta}^{(N)}(t + \delta) \right] = 0,$$

where the last equality follows from Lemma 5.8(3) of Kaspi and Ramanan [22].  $\square$

## APPENDIX B. RAMIFICATIONS OF ASSUMPTIONS ON THE SERVICE DISTRIBUTION

**Lemma B.1.** *Suppose  $h$  is uniformly bounded. Then Assumptions 2 and 4 are satisfied. If, in addition,  $g$  is absolutely continuous and  $g'$  is either locally essentially bounded or  $g' \in \mathbb{L}_2^{loc}$ . Then for any  $\tilde{f} \in \mathcal{S}_c$ ,  $f = \tilde{f}(1 - G)^{-1} \in \mathbb{H}_2$ . Moreover, if  $g$  is absolutely continuous and  $g'/(1 - G)$  is bounded then  $f \in \mathbb{H}_2$  implies  $\Phi_t f \in \mathbb{H}_2$  for every  $t \geq 0$  and for every  $t > 0$ , the mapping from  $\mathbb{H}_{-2}$  to  $\mathbb{H}_{-2}$  that takes  $\nu \mapsto S_t^\nu = \nu(\Phi_t \cdot)$  is Lipschitz continuous.*

*Proof.* If  $h$  is uniformly bounded, then Assumption 2 is trivially satisfied and

$$\frac{G(x+y) - G(x+\tilde{y})}{1 - G(x)} = \int_y^{\tilde{y}} \frac{g(x+u)}{1 - G(x+u)} \frac{1 - G(x+u)}{1 - G(x)} du \leq \|h\|_\infty |y - \tilde{y}|,$$

which shows that Assumption 4 is satisfied with  $C_G = \|h\|_\infty$  and  $\gamma_G = 1$ .

Now, suppose that in addition,  $g$  is absolutely continuous and  $g'$  is either locally essentially bounded or  $g' \in \mathbb{L}_2^{loc}$ . If  $f = \tilde{f}(1 - G)^{-1}$  then  $f' = \tilde{f}'_x(1 - G)^{-1} + hf$  and  $f'' = \tilde{f}''_{xx}(1 - G)^{-1} + 2h(\tilde{f}'_x + f_x) + fh^2 + fg'(1 - G)^{-1}$ . Since  $g$  is absolutely continuous,  $f$  and  $f'$  and the first three terms in the expansion of  $f''$  are continuous with compact support and hence in  $\mathbb{L}_2^{loc}$ . In addition, because  $f(1 - G)^{-1}$  is continuous with compact support, the last term lies in  $\mathbb{L}_2^{loc}$  if either  $g'$  is locally essentially bounded or, by Cauchy-Schwarz, if  $g'$  lies in  $\mathbb{L}_2^{loc}$ .

Now, suppose that  $g$  is absolutely continuous and  $g'/(1 - G)$  is bounded. Fix  $t \geq 0$  and  $f \in \mathbb{H}_2$ . For notational conciseness, let

$$r(x) \doteq r_t(x) \doteq \frac{1 - G(x+t)}{1 - G(x)}, \quad x \in [0, L].$$

Then, by the definition (4.19) of  $\Phi_t$ , for  $x \in [0, L]$ ,

$$\begin{aligned} (\Phi_t f)(x) &= r(x)f(x+t), \\ (\Phi_t f)'(x) &= r'(x)f(x+t) + r(x)f'(x+t), \\ (\Phi_t f)''(x) &= r''(x)f(x+t) + 2r'(x)f'(x+t) + r(x)f''(x+t). \end{aligned}$$

By the assumptions on  $g$ , if  $f \in \mathbb{H}_2$  then  $\Phi_t f$  has weak derivatives up to order two and elementary calculations show that

$$\begin{aligned} r'(x) &= \frac{g(x)(1 - G(x+t)) - (1 - G(x))g(x+t)}{(1 - G(x))^2} = r(x)(h(x) - h(x+t)), \\ r''(x) &= r(x) \left( \frac{g'(x)}{1 - G(x)} + h^2(x) - \frac{g'(x+t)}{1 - G(x+t)} - h^2(x+t) \right) \\ &\quad + r'(x)(h(x) - h(x+t)). \end{aligned}$$

Clearly,  $\|r\|_\infty \leq 1$  and, due to the assumed boundedness of  $h$  and  $g'/(1 - G)$ , it follows that there exists  $C \in [1, \infty)$  such that  $\|r'\|_\infty \leq C$  and  $\|r''\|_\infty \leq C$ . The above observations, when combined, show that

$$\begin{aligned} \|(\Phi_t f)\|_{\mathbb{L}_2} &\leq \|f\|_{\mathbb{L}_2} \leq \|f\|_{\mathbb{H}_2}, \\ \|(\Phi_t f)'\|_{\mathbb{L}_2} &\leq \sqrt{2}C(\|f\|_{\mathbb{L}_2} + \|f'\|_{\mathbb{L}_2}) \leq \sqrt{2}C\|f\|_{\mathbb{H}_2}, \\ \|(\Phi_t f)''\|_{\mathbb{L}_2} &\leq 4\sqrt{2}C(\|f\|_{\mathbb{L}_2} + \|f'\|_{\mathbb{L}_2} + \|f''\|_{\mathbb{L}_2}) \leq 4\sqrt{2}C\|f\|_{\mathbb{H}_2}, \end{aligned}$$

which shows that  $\|\Phi_t f\|_{\mathbb{H}_2} \leq \tilde{C}\|f\|_{\mathbb{H}_2}$  for some finite constant  $\tilde{C}$ . This shows that  $\Phi_t f \in \mathbb{H}_2$  and that, for any  $t > 0$ , the map from  $\mathbb{H}_2$  to  $\mathbb{H}_2$  that takes  $f$  to  $\Phi_t f$  is Lipschitz continuous (with constant  $\tilde{C}$ ). This, in turn, trivially implies that for  $\nu \in \mathbb{H}_{-2}$ , the linear functional on  $\mathbb{H}_2$  given by  $S^\nu : f \mapsto \nu(\Phi_t f)$  also lies in  $\mathbb{H}_{-2}$  and that the map from  $\mathbb{H}_{-2}$  to itself that takes  $\nu$  to  $S^\nu$  is also Lipschitz continuous with the same constant. This proves the second property and therefore the lemma.  $\square$

## APPENDIX C. PROOF OF THE REPRESENTATION FORMULA

Fix  $N \in \mathbb{N}$ . We first show how (6.15) can be deduced from (6.5); the proof of how to obtain (6.14) from (6.1) is analogous (in fact, a bit simpler), and is therefore omitted. Let  $\tilde{\Omega}$  be a set of full  $\mathbb{P}$ -measure such that on  $\tilde{\Omega}$ ,  $\bar{A}^{(N)}(t)$ ,  $\bar{D}^{(N)}(t)$ ,  $\bar{Q}_1^{(N)}(t)$  and  $\bar{K}^{(N)}(t)$  are finite for all  $t \in [0, \infty)$ . Fix  $\omega \in \tilde{\Omega}$  and let  $\gamma$  and  $h\hat{\nu}^{(N)}$  be the linear functionals on  $\mathcal{C}_c([0, L] \times [0, \infty))$  defined, respectively, by

$$\gamma(\varphi) \doteq \int_{[0, L]} \varphi(x, 0) \hat{\nu}_0^{(N)}(dx) - \int_{[0, L] \times [0, \infty)} \varphi(x, s) \widehat{\mathcal{M}}^{(N)}(dx, ds) + \int_{[0, \infty)} \varphi(0, s) d\widehat{K}^{(N)}(s)$$

and

$$h\hat{\nu}^{(N)}(\varphi) \doteq \int_0^\infty \langle h(\cdot)\varphi(\cdot, s), \hat{\nu}_s^{(N)} \rangle ds$$

for  $\varphi \in \mathcal{C}_c([0, L] \times [0, \infty))$ . Since the total variation of  $\hat{\nu}_0^{(N)}$  on  $[0, L]$  is bounded by  $2\sqrt{N}$ , the total variation of  $\widehat{\mathcal{M}}_t^{(N)}(\mathbf{1})$  is bounded by  $\sqrt{N}(\bar{D}^{(N)}(t) + \bar{A}_1^{(N)}(t))$  and the total variation of  $\widehat{K}^{(N)}$  on  $[0, t]$  is bounded by  $\sqrt{N}(\bar{K}^{(N)}(t) + \bar{K}(t))$ , for any  $\varphi \in \mathcal{C}_c([0, L] \times [0, \infty))$  such that  $\text{supp}(\varphi) \subset [0, L] \times [0, t]$ , we have

$$|\gamma(\varphi)| \leq \sqrt{N} \|\varphi\|_\infty \left( 2 + \bar{D}^{(N)}(t) + \bar{A}_1^{(N)}(t) + \bar{K}^{(N)}(t) + \bar{K}(t) \right)$$

and, likewise, it can be argued that

$$\left| h\hat{\nu}^{(N)}(\varphi) \right| \leq \sqrt{N} \|\varphi\|_\infty \left( \bar{A}_1^{(N)}(t) + \bar{A}_1(t) \right).$$

This shows that  $\gamma$  and  $h\hat{\nu}^{(N)}$  define Radon measures on  $[0, L] \times [0, \infty)$ . Let  $\tilde{\mathcal{C}}_c^{1,1}$  be the space of continuous functions with compact support on  $[0, L] \times [0, \infty)$  such that the directional derivative  $\varphi_x + \varphi_s$  exists and is continuous. Now, for every  $\varphi \in \tilde{\mathcal{C}}_c^{1,1}$ , sending  $t \rightarrow \infty$  in (6.5), the left-hand side of (6.5) vanishes because  $\varphi$  has compact support, and we obtain

$$- \int_0^\infty \langle \varphi_x(\cdot, s) + \varphi_s(\cdot, s), \hat{\nu}_s^{(N)} \rangle ds = -h\hat{\nu}^{(N)}(\varphi) + \gamma(\varphi).$$

Since  $\{\hat{\nu}_t^{(N)}, t \geq 0\} \in \mathcal{D}_{\mathcal{M}[0, L]}[0, \infty)$ , the last equation shows that  $\{\hat{\nu}_t^{(N)}, t \geq 0\}$  satisfies the so-called abstract age equation for  $\gamma$ , as introduced in Definition 4.9 of Kaspi and Ramanan [22]. Therefore, by Corollary 4.17 and (4.24) of [22], it follows that for every  $f \in \mathcal{C}_c[0, L)$ ,  $\langle f, \hat{\nu}_t^{(N)} \rangle = \gamma(\varphi_t^f)$ ,  $t \geq 0$ , where

$$\varphi_t^f(x, s) = \psi_h^{-1}(x, s) f(x + t - s) \psi_h(x + t - s, t), \quad (x, s) \in [0, L] \times [0, t],$$

where  $\psi_h$  is the function defined in (4.53) of [22], and reproduced as equation (9.14) of this paper. Elementary algebra (specifically combining the relations in (9.16) with the definition (4.19) of  $\Psi_t$ ) then shows that  $\varphi_t^f(x, s) = \Psi_t f(x, s)$ . For  $f \in \mathcal{C}_c[0, L)$ , the representation (6.15) is then obtained by expanding  $\gamma(\Psi_t f)$  using the definition of  $\gamma$  given above together with the relations  $(\Psi_t f)(\cdot, 0) = \Phi_t f$ ,  $(\Psi_t f)(0, \cdot) = f(t - \cdot)(1 - G(t - \cdot))$ ,  $\hat{\mathcal{H}}_t^{(N)}(f) = \widehat{\mathcal{M}}_t^{(N)}(\Psi_t f)$  and the definition (6.11) of  $\widehat{\mathcal{K}}^{(N)}$ . Since the right-hand side of (6.15) is well defined for  $f \in \mathcal{C}_b[0, L)$ , a standard approximation argument can then be used to show that the representation (6.15) holds for  $f \in \mathcal{C}_b[0, L)$ .

## APPENDIX D. SOME MOMENT ESTIMATES

In this section, we prove the estimates stated in Lemma 8.1.

*Proof of Lemma 8.1.* Fix  $N \in \mathbb{N}$  and  $T < \infty$  and, for conciseness, let  $\bar{Y}^{(N)}(s) = (R_E^{(N)}(s), \bar{X}^{(N)}(s), \bar{\nu}_s^{(N)})$ ,  $s \in [0, \infty)$ , represent the state process. Using the fact that  $\bar{M}_1^{(N)} = \bar{D}^{(N)} - \bar{A}_1^{(N)}$  is a martingale

and taking expectations of both sides of the inequality (5.30) of Kaspi and Ramanan [22], with  $t$  and  $\delta$  replaced by 0 and  $T$ , respectively, it follows that

$$(D.1) \quad \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \bar{A}_1^{(N)}(T) \right] = \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \bar{D}^{(N)}(T) \right] \leq U(T),$$

where  $U$  is the renewal function associated with  $G$ . This shows that the inequality (8.1) holds for  $k = 1$ . We proceed by induction. Suppose that (8.1) holds with  $k = j - 1$  for some integer  $j \geq 2$ . Then we can write

$$\begin{aligned} \left( \bar{A}_1^{(N)}(T) \right)^j &= \int_0^T \cdots \int_0^T \left( \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \langle h, \bar{\nu}_{s_2}^{(N)} \rangle \cdots \langle h, \bar{\nu}_{s_j}^{(N)} \rangle \right) ds_j \cdots ds_1 \\ &= \int_0^T \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \left( \int_0^T \cdots \int_0^T \left( \langle h, \bar{\nu}_{s_2}^{(N)} \rangle \cdots \langle h, \bar{\nu}_{s_j}^{(N)} \rangle \right) ds_j \cdots ds_2 \right) ds_1 \\ &= j \int_0^T \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \left( \int_{s_1}^T \cdots \int_{s_1}^T \left( \langle h, \bar{\nu}_{s_2}^{(N)} \rangle \cdots \langle h, \bar{\nu}_{s_j}^{(N)} \rangle \right) ds_j \cdots ds_2 \right) ds_1 \\ &= j \int_0^T \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \left( \bar{A}_1^{(N)}(T) - \bar{A}_1^{(N)}(s_1) \right)^{j-1} ds_1. \end{aligned}$$

Taking expectations of both sides above and applying Tonelli's theorem we obtain

$$\mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \left( \bar{A}_1^{(N)}(T) \right)^j \right] = j \int_0^T \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \left( \bar{A}_1^{(N)}(T) - \bar{A}_1^{(N)}(s_1) \right)^{j-1} \right] ds_1.$$

For each  $s_1 \in [0, T]$ , due to the Markov property of  $\bar{Y}^{(N)}$  established in Lemma B.1 of Kang and Ramanan [20] we obtain

$$\begin{aligned} &\mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \left( \bar{A}_1^{(N)}(T) - \bar{A}_1^{(N)}(s_1) \right)^{j-1} \right] \\ &= \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \mathbb{E}_{\bar{Y}^{(N)}(s_1)} \left[ \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \left( \bar{A}_1^{(N)}(T) - \bar{A}_1^{(N)}(s_1) \right)^{j-1} \mid \mathcal{F}_{s_1}^{(N)} \right] \right] \\ &= \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \mathbb{E}_{\bar{Y}^{(N)}(s_1)} \left[ \left( \bar{A}_1^{(N)}(T - s_1) \right)^{j-1} \right] \right]. \end{aligned}$$

Applying the induction assumption to the last term above, it follows that

$$\mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \left( \bar{A}_1^{(N)}(T) - \bar{A}_1^{(N)}(s_1) \right)^{j-1} \right] \leq (j-1)! U(T)^{j-1} \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \langle h, \bar{\nu}_{s_1}^{(N)} \rangle \right].$$

Combining the last three displays, applying Tonelli's theorem again and using (D.1), we obtain

$$\mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \left( \bar{A}_1^{(N)}(T) \right)^j \right] \leq j! U(T)^{j-1} \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \int_0^T \langle h, \bar{\nu}_{s_1}^{(N)} \rangle ds_1 \right] \leq j! U(T)^j.$$

This shows that (8.1) is also satisfied for  $k = j$  and hence, by induction, for all positive integers  $k$ .

We now turn to the proof of the second bound. We can assume without loss of generality that  $\varphi^* h$  is integrable on  $[0, L]$  because otherwise the inequality holds trivially. On substituting  $l = \varphi^* h$ ,  $\varphi = \mathbf{1}$ ,  $r = 0$  and  $t = T$  in (5.31) of Proposition 5.7 of Kaspi and Ramanan [22], for every  $N \in \mathbb{N}$ , we have

$$(D.2) \quad \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \bar{A}_\varphi^{(N)}(T) \right] \leq \mathbb{E}_{\bar{Y}^{(N)}(0)} \left[ \bar{A}_{\varphi^*}^{(N)}(T) \right] \leq C_1(T) \left( \int_{[0, L]} \varphi^*(x) h(x) dx \right),$$

where

$$C_1(T) \doteq \sup_N \mathbb{E}[\bar{X}^{(N)}(0) + \bar{E}^{(N)}(T)] \leq C(T) \doteq \sup_N \sup_{s \in [0, T]} \mathbb{E}[\bar{X}^{(N)}(s) + \bar{E}^{(N)}(T)],$$

which is finite by Theorem 3.2. Given (D.2), the same inductive argument used in the proof of the first assertion of the lemma can then be used to complete the proof of the second bound.

Next, note that if Assumptions 1 and 2 hold, then  $\overline{A}_1^{(N)} \Rightarrow \overline{A}_1$  by Proposition 5.17 of [22] and  $\overline{A}$  is continuous. Together with the Skorokhod representation theorem, Fatou's lemma and the inequality (D.1), this implies that

$$\overline{A}_1(T) \leq \liminf_{N \rightarrow \infty} \mathbb{E} \left[ \overline{A}_1^{(N)}(T) \right] \leq \limsup_{N \rightarrow \infty} \mathbb{E} \left[ \overline{A}_1^{(N)}(T) \right] \leq U(T).$$

The inequality (8.2) can now be deduced from this inequality exactly as the inequality (8.1) was deduced from the inequality (D.1), though the proof is in fact much simpler because  $\overline{A}_1$  is deterministic.  $\square$

#### APPENDIX E. PROOF OF CONSISTENCY

We first start by establishing some Fubini theorems.

**Lemma E.1.** *Let Assumptions 1-4 be satisfied, let  $g$  be continuous and let  $\widehat{\mathcal{H}}$  and  $\widehat{\mathcal{K}}$  be defined as in (4.18) and (7.2), respectively. Suppose  $\tilde{\varphi} \in \mathcal{C}_b([0, L] \times [0, \infty))$ . Given the family of mappings  $\Theta_t$ ,  $t \geq 0$ , defined in (8.21), we have*

$$(E.1) \quad \widehat{\mathcal{M}}_t(\Theta_t \tilde{\varphi}) = \int_0^t \widehat{\mathcal{M}}_r(\Psi_r(\tilde{\varphi}(\cdot, r))) dr = \int_0^t \widehat{\mathcal{H}}_r(\tilde{\varphi}(\cdot, r)) dr.$$

If, in addition, for every  $t \in [0, T]$ ,  $x \mapsto \int_0^t \tilde{\varphi}(x, r) dr$  is bounded and Hölder continuous, uniformly in  $t$ , then for every  $s \geq 0$ , almost surely for every  $t \geq 0$ ,

$$(E.2) \quad \widehat{\mathcal{M}}_s \left( \int_0^t \Psi_s(\tilde{\varphi}(\cdot, r)) dr \right) = \int_0^t \widehat{\mathcal{M}}_s(\Psi_s(\tilde{\varphi}(\cdot, r))) dr = \int_0^t \widehat{\mathcal{H}}_s(\tilde{\varphi}(\cdot, r)) dr.$$

Moreover, if either  $x \mapsto \tilde{\varphi}(x, r)$  is absolutely continuous for every  $r > 0$  and  $(x, r) \mapsto \tilde{\varphi}_x(x, r)(1 - G(x))$  is locally integrable on  $[0, L] \times [0, \infty)$ , or  $g$  is absolutely continuous and  $\tilde{\varphi} \in \mathcal{C}_b([0, L] \times [0, \infty))$ , then almost surely, for every  $s, t \geq 0$ ,

$$(E.3) \quad \widehat{\mathcal{K}}_s \left( \int_0^t \tilde{\varphi}(\cdot, r) dr \right) = \int_0^t \widehat{\mathcal{K}}_s(\tilde{\varphi}(\cdot, r)) dr.$$

*Proof.* Fix  $s, t \geq 0$ . Then

$$(E.4) \quad \int_0^t \Psi_s(\tilde{\varphi}(\cdot, r)) dr = \Psi_s \left( \int_0^t \tilde{\varphi}(\cdot, r) dr \right)$$

and so, by the inequality (4.20) and the boundedness assumption on  $\tilde{\varphi}$ ,  $\int_0^t \Psi_s(\tilde{\varphi}(\cdot, r)) dr$  and  $\Theta_t \tilde{\varphi}$  are uniformly bounded on  $[0, L] \times [0, t]$ . We can thus apply Fubini's theorem for stochastic integrals with respect to martingale measures (see Theorem 2.6 of [35]) to conclude that almost surely, (E.1) and (E.2) are satisfied. We now have to show that the set of measure zero on which they are not satisfied is independent of  $t$ , for which it suffices to show that the processes on both sides of (E.1) and (E.2) are continuous in  $t$ . The processes on the right-hand sides of (E.1) and (E.2) are clearly continuous in  $t$ , whereas the continuity in  $t$  of the process on the left-hand side of (E.1) follows from property 4 of Lemma 8.6. Because of (E.4), the relation  $\widehat{\mathcal{M}}_s(\Psi_s f) = \widehat{\mathcal{H}}_s(f)$  and the assumed boundedness and uniform Hölder continuity of  $x \mapsto \int_0^t \tilde{\varphi}(x, r) dr$ , the continuity in  $t$  of the left-hand side of (E.2) follows from property 3 of Lemma 8.6. Thus, for any given  $s > 0$  there exists a set of full  $\mathbb{P}$ -measure on which (E.2) and (E.1) hold simultaneously for all  $t \geq 0$ .

Next, by the definition of  $\widehat{\mathcal{K}}$  in (7.2), note that  $\widehat{\mathcal{K}}_s(\int_0^t \tilde{\varphi}(\cdot, r) dr)$  is equal to

$$\left( \int_0^t \tilde{\varphi}(0, r) dr \right) \widehat{\mathcal{K}}(s) + \int_0^s \widehat{\mathcal{K}}(u) \frac{\partial}{\partial x} \left( \int_0^t ((1 - G(x)) \tilde{\varphi}(x, r) dr) \right) \Big|_{x=s-u} du.$$

By the stated assumptions, it follows that  $g$  is continuous and for each  $r > 0$ , the function  $x \mapsto (1-G(x))\tilde{\varphi}(x, r)$  is absolutely continuous and its derivative (with respect to  $x$ ) is locally integrable. Moreover, by Theorem 5.6,  $\widehat{K}$  is almost surely continuous and thus locally bounded. Thus, we can first exchange the order of differentiation and integration and then apply Fubini's theorem for Lebesgue integrals in the last display to conclude that  $\widehat{\mathcal{K}}_s(\int_0^t \tilde{\varphi}(\cdot, r) dr)$  is equal to

$$\begin{aligned} & \int_0^t \tilde{\varphi}(0, r)\widehat{K}(s) dr + \int_0^s \widehat{K}(u) \left( \int_0^t \frac{\partial}{\partial x} ((1-G(x))\tilde{\varphi}(x, r))|_{x=s-u} dr \right) du \\ &= \int_0^t \tilde{\varphi}(0, r)\widehat{K}(s) dr + \int_0^t \left( \int_0^s \widehat{K}(u) \frac{\partial}{\partial x} ((1-G(x))\tilde{\varphi}(x, r))|_{x=s-u} du \right) dr \\ &= \int_0^t \widehat{\mathcal{K}}_s(\tilde{\varphi}(\cdot, r)) dr, \end{aligned}$$

which completes the proof of the lemma.  $\square$

We now prove the consistency lemma.

*Proof of Lemma 9.6.* Fix  $f \in \mathcal{S}$  and  $s, t \geq 0$ . Then, replacing  $t$  by  $t+s$  in (6.15), we obtain

$$(E.5) \quad \widehat{\nu}_{s+t}^{(N)}(f) = \mathcal{S}_{s+t}^{\widehat{\nu}_0^{(N)}}(f) - \widehat{\mathcal{H}}_{t+s}^{(N)}(f) + \widehat{\mathcal{K}}_{t+s}^{(N)}(f).$$

Using the shift relations introduced in (9.24)–(9.26), and recalling the definitions of  $\widehat{\mathcal{H}}^{(N)}$  and  $\widehat{\mathcal{K}}^{(N)}$  in (4.14) and (6.12), respectively, the last two terms on the right-hand side of (E.5) can be decomposed as follows:

$$(E.6) \quad \widehat{\mathcal{H}}_{t+s}^{(N)}(f) = \widehat{\mathcal{M}}_{t+s}^{(N)}(\Psi_{t+s}f) = \widehat{\mathcal{M}}_s^{(N)}(\Psi_{t+s}(f)) + (\Theta_s \widehat{\mathcal{M}}^{(N)})(\Psi_t f),$$

and, similarly,

$$(E.7) \quad \begin{aligned} \widehat{\mathcal{K}}_{t+s}^{(N)}(f) &= \int_{[0, s+t]} f(s+t-u)(1-G(s+t-u)) d\widehat{K}^{(N)}(u) \\ &= \int_{[0, s]} f(s+t-u)(1-G(s+t-u)) d\widehat{K}^{(N)}(u) \\ &\quad + \int_{[0, t]} (1-G(t-u))f(t-u) d(\Theta_s \widehat{K}^{(N)})(u). \end{aligned}$$

On the other hand, since  $\Phi_t f \in \mathcal{C}_b[0, L]$ , replacing  $f$  and  $\widehat{\nu}_0^{(N)}$  in (6.15) by  $\Phi_t f$  and  $\widehat{\nu}_s^{(N)}$ , respectively, and using the semigroup property (5.8) and the fact that  $\Psi_s \Phi_t = \Psi_{s+t}$  on the appropriate domain as specified in (5.9), we obtain

$$(E.8) \quad \begin{aligned} \mathcal{S}_t^{\widehat{\nu}_s^{(N)}}(f) = \langle \Phi_t f, \widehat{\nu}_s^{(N)} \rangle &= \langle \Phi_{s+t} f, \widehat{\nu}_0^{(N)} \rangle - \widehat{\mathcal{M}}_s^{(N)}(\Psi_s \Phi_t f) \\ &\quad + \int_{[0, s]} (\Phi_t f)(s-u)(1-G(s-u)) d\widehat{K}^{(N)}(u) \\ &= \mathcal{S}_{s+t}^{\widehat{\nu}_0^{(N)}}(f) - \widehat{\mathcal{M}}_s^{(N)}(\Psi_{s+t} f) \\ &\quad + \int_{[0, s]} f(s+t-u)(1-G(s+t-u)) d\widehat{K}^{(N)}(u). \end{aligned}$$

The relation (9.29) is then obtained by subtracting (E.8) from (E.5), rearranging terms and using the relations (E.7) and (E.6).

Now, suppose that Assumptions 1–4 are satisfied and further, assume that  $g$  is continuous. Then Theorem 5.7 shows that the limit  $\widehat{\nu}$  of  $\{\widehat{\nu}^{(N)}\}$  is a continuous  $\mathbb{H}_{-2}$ -valued process that is given explicitly by (5.24). The shifted equation (9.30) for the limit  $\widehat{\nu}$  is proved in a similar fashion as for the corresponding quantity  $\widehat{\nu}^{(N)}$  in the  $N$ -server system, except that now  $\widehat{\mathcal{K}}$  has the slightly



different representation (7.2). We fill in the details for completeness. Applying (5.24) with  $t$  replaced by  $t + s$ , we see that for bounded and absolutely continuous  $f$ ,

$$\widehat{\nu}_{t+s}(f) = \mathcal{S}_{t+s}^{\widehat{\nu}_0}(f) - \widehat{\mathcal{H}}_{t+s}(f) + f(0)\widehat{K}(s) + \int_0^{t+s} \widehat{K}(u)\xi_f(t+s-u) du.$$

On the other hand, applying (5.24) with  $f$  and  $t$ , respectively, replaced by  $\Phi_t f$  and  $s$  and using the semigroup relation (5.8) for  $\Phi_t$  and the fact that  $(\Phi_t f)(0) = f(t)(1 - G(t))$ , we obtain

$$(E.9) \quad \begin{aligned} \mathcal{S}_t^{\widehat{\nu}_s}(f) &= \widehat{\nu}_s(\Phi_t f) \\ &= \mathcal{S}_{s+t}^{\widehat{\nu}_0}(f) - \widehat{\mathcal{M}}_s(\Psi_t \Phi_t f) + f(t)(1 - G(t))\widehat{K}(s) + \int_0^s \widehat{K}(u)\xi_{\Phi_t f}(s-u) du. \end{aligned}$$

Simple calculations show that  $\xi_{\Phi_t f} = \xi_f(\cdot + t)$ . Hence,

$$\int_0^{t+s} \widehat{K}(u)\varphi_f(t+s-u) du - \int_0^s \widehat{K}(u)\xi_{\Phi_t f}(s-u) du = \int_0^t \widehat{K}(s+u)\xi_f(t-u) du$$

and, since  $\xi_f = (f(1 - G))'$ ,

$$\int_0^t \widehat{K}(s)\xi_f(t-u) du = f(0)\widehat{K}(s) - f(t)(1 - G(t))\widehat{K}(s).$$

Equation (9.30) can now be obtained by combining the last four equations with the limit analog of (E.6), in which  $\widehat{\mathcal{H}}^{(N)}$  and  $\widehat{\mathcal{M}}^{(N)}$ , respectively, are replaced by  $\widehat{\mathcal{H}}$  and  $\widehat{\mathcal{M}}$ .

To show that (9.31) is satisfied, note that by Theorem 5.6,  $(\widehat{K}, \widehat{X}, \widehat{\nu}_0(\mathbf{1})) = \Lambda(\widehat{E}, \widehat{x}_0, \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}))$ . This implies that the centered many-server equations (5.14)–(5.16) are satisfied with  $v, Z, X, K$  and  $E$ , respectively, replaced by  $\widehat{\nu}(\mathbf{1}), \mathcal{S}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}), \widehat{X}, \widehat{K}$  and  $\widehat{E}$ . Fix any  $s > 0$ . Subtracting the equation (5.15) evaluated at  $t + s$  from the same equation evaluated at  $t$ , it follows that (5.15) also holds when  $K, E, X$  and  $v$  is replaced, respectively, by  $\Theta_s \widehat{K}, \Theta_s \widehat{E}, \widehat{X}_{s+}$  and  $\widehat{\nu}_{s+}(\mathbf{1})$ . It is also clear that (5.16) is satisfied with  $v$  and  $X$  replaced by  $\widehat{\nu}_{s+t}(\mathbf{1})$  and  $\widehat{X}_{s+t}$  for all  $t \geq 0$ . Lastly, substituting  $f = \mathbf{1}$  in (9.30), using the definition (9.27) of  $\Theta_s \widehat{K}$  and the fact that  $\varphi_{\mathbf{1}} = -g$ , it follows that (5.14) holds with  $v, Z$  and  $K$ , respectively, replaced, by  $\widehat{\nu}_{s+}(\mathbf{1}), \mathcal{S}^{\widehat{\nu}_s}(\mathbf{1}) - \Theta_s \widehat{\mathcal{H}}(\mathbf{1})$  and  $\Theta_s \widehat{K}$ . This proves (9.31).

Fix  $s > 0$ . We first need to show that Assumption 3 is satisfied when  $\widehat{E}^{(N)}$  and  $\widehat{E}$ , respectively, are replaced by  $\Theta_s \widehat{E}^{(N)}$  and  $\Theta_s \widehat{E}$ . This is easily deduced using basic properties of renewal processes and Poisson processes and is thus left to the reader. Next, we show that Assumption 5' is satisfied when  $\widehat{x}_0^{(N)}, \widehat{x}_0, \widehat{\nu}_0^{(N)}$ , and  $\widehat{\nu}_0$ , respectively, are replaced by  $\widehat{X}^{(N)}(s), \widehat{X}(s), \widehat{\nu}_s^{(N)}$  and  $\widehat{\nu}_s$ . By definition (5.24),  $\widehat{\nu}_s(f)$  is a well defined random variable for all  $f \in \mathcal{AC}_b[0, L]$  and, since  $\Phi_t f \in \mathcal{AC}_b[0, L]$  if  $f \in \mathcal{AC}_b[0, L]$  it follows that  $\mathcal{S}_t^{\widehat{\nu}_s}(f) = \widehat{\nu}_s(\Phi_t f)$  is also a well defined random variable for every  $f \in \mathcal{AC}_b[0, L]$ . Now, due to (5.24), the assumed measurability of  $f \mapsto \mathcal{S}_s^{\widehat{\nu}_0}(f)$  and the joint measurability of the maps  $(s, f) \mapsto \widehat{\mathcal{M}}_s(f)$  and  $(s, f) \mapsto \widehat{\mathcal{K}}_s(f)$  for  $f \in \mathcal{C}_b[0, L]$ , which is a consequence of the definition of these stochastic integrals, it follows that almost surely the map  $f \mapsto \widehat{\nu}_s(f)$  from  $\mathcal{AC}_b[0, L] \subset \mathcal{C}_b[0, L]$  to  $\mathbb{R}$  (both equipped with their respective Borel  $\sigma$ -algebras) is also measurable. Moreover, for a given  $\varphi \in \mathcal{C}_b([0, L] \times [0, T])$  and  $t > 0$ , the maps  $r \mapsto \Phi_r \varphi(\cdot, r)$  from  $[0, T]$  to  $\mathcal{C}_b[0, L]$  and  $f \mapsto \Phi_t f$  from  $\mathcal{C}_b[0, L]$  to itself are clearly also measurable (they are in fact continuous). Since the composition of measurable maps is measurable, it follows that almost surely, both  $r \mapsto \widehat{\nu}_s(\Phi_r(\varphi(\cdot, r)))$  and  $f \mapsto \mathcal{S}_t^{\widehat{\nu}_s}(f) = \widehat{\nu}_s(\Phi_t f)$  are measurable.

In addition, due to the continuity of the limit in the convergence (9.1) established in Proposition 9.1, it follows that

$$(\widehat{X}^{(N)}(s), \widehat{\nu}_{s+}^{(N)}, \Theta_s \widehat{\mathcal{K}}^{(N)}, \Theta_s \widehat{\mathcal{H}}^{(N)}) \Rightarrow (\widehat{X}(s), \widehat{\nu}_{s+}, \Theta_s \widehat{\mathcal{K}}, \Theta_s \widehat{\mathcal{H}})$$

in the space  $\mathbb{R} \times \mathcal{D}_{\mathbb{H}_{-2}}^3[0, \infty)$ . In particular, this implies that  $\widehat{\nu}_s$  has an  $\mathbb{H}_{-2}$ -valued version, and so property (a) of Assumption 5 is satisfied. Next, by (9.29) and (9.30),  $\mathcal{S}^{\widehat{\nu}_s^{(N)}}$  can be expressed

as a linear combination of the  $\mathbb{H}_{-2}$ -valued processes  $(\widehat{\nu}_{s+}^{(N)}, \Theta_s \widehat{\mathcal{K}}^{(N)}, \Theta_s \widehat{\mathcal{H}}^{(N)})$  and, likewise,  $\mathcal{S}^{\widehat{\nu}_s}$  is the same linear combination of  $(\widehat{\nu}_{s+}, \Theta_s \widehat{\mathcal{K}}, \Theta_s \widehat{\mathcal{H}})$ . Therefore, the continuity of  $\widehat{\nu}_{s+}, \Theta_s \widehat{\mathcal{K}}$  and  $\Theta_s \widehat{\mathcal{H}}$  show that  $\mathcal{S}^{\widehat{\nu}_s}$  is a continuous  $\mathbb{H}_{-2}$ -valued process. The same logic used above then shows that the real-valued process  $\mathcal{S}^{\widehat{\nu}_s}(\mathbf{1})$  is continuous and that the limits in property (c) of Assumption 5 holds. Thus, we have established that Assumption 5 continues to hold at a shifted time.

It only remains to show that Assumption 5' is satisfied when  $\widehat{\nu}_0$  is replaced by  $\widehat{\nu}_s$ . Fix  $\varphi \in \mathcal{C}_b([0, L] \times [0, \infty))$  such that  $x \mapsto \varphi(x, r)$  is absolutely continuous and Hölder continuous and  $\varphi_x$  is integrable on  $[0, L] \times [0, T]$  for any  $T < \infty$ . We will make repeated use of the semigroup property  $\Phi_s \circ \Phi_r = \Phi_{s+r}$ , the relation  $\Psi_{s+r} = \Psi_s \circ \Phi_r$  on the appropriate domain as specified in (5.9), and the form (7.2) of  $\widehat{\mathcal{K}}$ , without explicit mention. Then, by the other assumptions on  $h$ , for any  $r > 0$ ,  $\Phi_r \varphi(\cdot, r)$  and  $\int_0^t \Phi_r \varphi(\cdot, r) dr$  are both bounded, Hölder continuous and absolutely continuous functions on  $[0, L]$ . Therefore, substituting  $f = \Phi_r \varphi(\cdot, r)$  into (5.24) we see that

$$(E.10) \quad \widehat{\nu}_s(\Phi_r \varphi(\cdot, r)) = \widehat{\nu}_0(\Phi_{s+r} \varphi(\cdot, r)) - \widehat{\mathcal{M}}_s(\Psi_{s+r} \varphi(\cdot, r)) + \widehat{\mathcal{K}}_s(\Phi_r \varphi(\cdot, r)).$$

By (5.11), it follows that  $\widehat{\nu}_0(\Phi_{s+r} \varphi(\cdot, r)) = \mathcal{S}_s^{\widehat{\nu}_0}(\Phi_r \varphi(\cdot, r))$ . Furthermore, substituting  $f = \int_0^t \Phi_r \varphi(\cdot, r) dr$  into (5.24), invoking Assumption 5'(d) with  $\varphi(\cdot, r)$  replaced by  $\Phi_s \varphi(\cdot, r)$ , and applying the Fubini theorems (E.2) and (E.3) with the absolutely continuous and uniformly bounded function  $\tilde{\varphi}(x, r) = (\Phi_r \varphi(\cdot, r))(x)$ , it follows that  $\widehat{\nu}_s \left( \int_0^t \Phi_r \varphi(\cdot, r) dr \right)$  is equal to

$$\begin{aligned} & \widehat{\nu}_0 \left( \int_0^t \Phi_{s+r} \varphi(\cdot, r) dr \right) - \widehat{\mathcal{M}}_s \left( \int_0^t \Psi_{s+r}(\varphi(\cdot, r)) dr \right) + \widehat{\mathcal{K}}_s \left( \int_0^t \Phi_r \varphi(\cdot, r) dr \right) \\ & = \int_0^t \widehat{\nu}_0(\Phi_{s+r} \varphi(\cdot, r)) dr - \int_0^t \widehat{\mathcal{M}}_s(\Psi_{s+r} \varphi(\cdot, r)) dr + \int_0^t \widehat{\mathcal{K}}_s(\Phi_r \varphi(\cdot, r)) dr. \end{aligned}$$

A comparison with (E.10) shows that the right-hand side above equals  $\int_0^t \widehat{\nu}_s(\Phi_r \varphi(\cdot, r)) dr$ . Thus, Assumption 5'(d) holds with  $\widehat{\nu}_0$  replaced by  $\widehat{\nu}_s$ .  $\square$

**Acknowledgments.** We are grateful to Leonid Mytnik for some useful discussions and to Avi Mandelbaum for bringing this open problem to our attention.

## REFERENCES

- [1] ADAMS, R.A. (1975) *Sobolev Spaces*, Academic Press, New York.
- [2] AGADZHANOV, A.N. (1989) Functional properties of Sobolev spaces of infinite order. *Soviet Math. Dokl.* **38**, 88–92.
- [3] ASMUSSEN, S. (2003) *Applied Probability and Queues*, 2nd edition ed., Springer-Verlag, New York.
- [4] BILLINGSLEY, P. (1968) *Convergence of Probability Measures*, John Wiley, New York.
- [5] (1967) BOROVKIOV, A.A. On limit laws for service processes in multichannel systems (in Russian), *Siberian Math. J.* **8** 746–763.
- [6] BROWN, L., GANS, N., MANDELBAUM, A., SAKOV, A., SHEN, H., ZELTYN, S. and ZHAO, L. (2005) Statistical analysis of a telephone call center: a queueing science perspective, *JASA* **100**, No. 469, 36–50.
- [7] ERLANG, A.K. On the rational determination of the number of circuits. (1948) In *The Life and Works of A.K. Erlang*, E. Brockmeyer, H.L. Halstrom and A. Jensen, Eds. The Copenhagen Telephone Company, Copenhagen.
- [8] BURTON, B. (2005) *Volterra Integral and Differential Equations*. Elsevier.
- [9] DECREUSEFOND, L. and MOYAL, P. (2008) A functional central limit theorem for the  $M/GI/\infty$  queue. *Annals of Applied Probability*, **18**, 2156–2178.
- [10] ETHIER, S.N. and KURTZ, T.G. (1986) *Markov Processes: Characterization and convergence*, Wiley.
- [11] EVANS, L.C. (1988) *Partial Differential Equations*, Graduate Studies in Mathematics Vol. 19, American Mathematical Society.
- [12] FRIEDMAN, A. (2006) *Stochastic Differential Equations and Applications*, Dover Publications.
- [13] GAMARNIK, D. and MOMČILOVIĆ, P. (2008). Steady-state analysis of a multiserver queue in the Halfin-Whitt regime, *Advances in Applied Probability* **40** No. 2, 548–577.
- [14] GLYNN, P. AND WHITT, W. (1991) A new view of the heavy-traffic limit theorem for infinite-server queues. *Adv. Appl. Probab.* **2** 188–209.
- [15] HALFIN, S. AND WHITT, W. (1981). Heavy-traffic limit theorems for queues with many servers, *Oper. Res.* **29**, 567–588.

- [16] IGLEHART, D.L. (1965) Limit diffusion approximations for the many server queue and the repairman problem. *J. Appl. Probab.* **2** 429–441.
- [17] JACOD, J. and SHIRYAEV, A.N. *Limit theorems for stochastic processes*, Springer-Verlag, Berlin, 1987.
- [18] JAGERMAN, D. (1974) *Some properties of Erlang loss function*. *Bell System Techn. J.* **53** no. 3, 525–551.
- [19] JELENKOVIC, P., MANDELBAUM, A. and MOMČILOVIĆ, P. (2004). Heavy traffic limits for queues with many deterministic servers, *Queueing Syst.* **47** No. 1-2, 53–69.
- [20] KANG, W. N. and RAMANAN, K. (2008) Fluid limits of many-server queues with reneging, *Annals of Applied Probability*, to appear.
- [21] KANG, W. N. and RAMANAN, K. (2010) Approximations of stationary distributions of many-server queues, *Annals of Applied Probability*, to appear.
- [22] KASPI, H. and RAMANAN, K. (2007). Law of large numbers limits for many-server queues, *Annals of Applied Probability*, to appear.
- [23] KRICHAGINA, E. and PUHALSKII, A. (1997) A heavy traffic analysis of a closed queueing system with a GI/G/∞ service center, *Queueing Syst.* **25**, 235–280.
- [24] MANDELBAUM, A., MASSEY, W. and REIMAN, M. (1998) Strong approximations for Markovian service networks, *Queueing Syst.* **30**, 149–201.
- [25] A. Mandelbaum and P. Momcilovic. *Queues with many servers: the virtual waiting-time process in the QED regime*. Math of Oper. Res. **FILL IN**
- [26] MITOMA, I. (1983) Tightness of Probabilities on  $\mathcal{C}([0, 1], S')$  and  $\mathcal{D}([0, 1], S')$ , *Annals of Probability* **11** 989-999.
- [27] MITOMA, I. (1983) On the sample continuity of  $S'$ -processes, *J. Math. Soc. Japan* **35**, No. 4, 629–636.
- [28] PUHALSKII, A.A. and REIMAN, M. (2000) The multiclass GI/PH/N queue in the Halfin-Whitt regime, *Adv. Appl. Probab.* **32**, 564-595.
- [29] PUHALSKII, A.A. and REED, J. (2010) On many servers queues in heavy traffic, *Ann. Appl. Probab.* **20** No.1, 129–195.
- [30] REED, J. (2009) The G/GI/N queue in the Halfin-Whitt regime, *Ann. Appl. Probab.* **19** No. 6, 2211-2269.
- [31] REED, J. AND TALREJA, R. Distribution-valued heavy-traffic limits for the G/GI/∞ queue. *Preprint*.
- [32] REVUZ, D. and YOR, M. (1999) *Continuous Martingales and Brownian Motion*. Third Edition, Springer, Berlin.
- [33] ROGERS, D. and WILLIAMS, D. (1994) *Diffusion, Markov Processes and Martingales, Volume 2: Foundations*. Second Edition, Wiley.
- [34] SHARPE, M. (1988) *General Theory of Markov Processes*. Academic Press.
- [35] WALSH, J.B. *An Introduction to Stochastic Partial Differential Equations*, École d'été probabilités de Saint-Flour XIV. *Lecture Notes in Math.* **1180** 265–439. Springer, Berlin.
- [36] (1982) WHITT, W. On the heavy-traffic limit theorem for GI/G/∞ queues. *Adv. Appl. Probab.* **14**, 171–190.
- [37] WHITT, W. (2005) Heavy-traffic limits for the G/H2/n/m queue, *Queueing Syst.* **30**, 1–27.

DEPARTMENT OF INDUSTRIAL ENGINEERING AND MANAGEMENT, TECHNION, HAIFA, ISRAEL  
*E-mail address:* `iehaya@techunix.technion.ac.il`

DIVISION OF APPLIED MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE, RI 02912, USA  
*E-mail address:* `Kavita.Ramanan@brown.edu`