The Annals of Applied Probability
2013, Vol. 23, No. 1, 145–229
DOI: 10.1214/11-AAP821

SPDE LIMITS OF MANY-SERVER QUEUES

BY HAYA KASPI[1] AND KAVITA RAMANAN[2]

Technion and Brown University

This paper studies a queueing system in which customers with independent and identically distributed service times arrive to a queue with many servers and 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, and 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 beyond its mean. In addition, the limit of the sequence of (centered and scaled) age processes is shown to be a diffusion taking values in a Hilbert space and is characterized as the unique solution of a stochastic partial differential equation that is coupled with the Itô diffusion describing the limiting number in system. Furthermore, the limit processes are shown to be semimartingales and to possess a strong Markov property.

CONTENTS

Received October 2010; revised October 2011.

[1]Supported in part by the Israel Science Foundation under Grant 81/09, the Milford Bohm Chair Grant and the US-Israel Binational Science Foundation under Grant BSF-2006379.

[2]Supported in part by the NSF Grants CMMI-1059967 (formerly 0728064) and CMMI-1052750 (formerly 0928154), as well as the US-Israel Binational Science Foundation under Grant BSF-2006379.

MSC2010 subject classifications. Primary 60K25, 60F17, 60H15; secondary 90B22, 68M20.

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

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 N, the number of servers, tends to infinity and $\lambda^{(N)}$, the mean arrival rate in the system with N servers, scales as $\lambda^{(N)} = \overline{\lambda}N - \beta\sqrt{N}$ for some $\overline{\lambda} > 0$ and $\beta \in (-\infty, \infty)$. For many-server queues with Poisson arrivals, this scaling was considered more than half a century ago by Erlang [8] and thereafter by Jagerman [17] for a loss system with exponential service times, but it was not until the influential work of Halfin and Whitt [14] that a heavy traffic limit theorem was established. As a result, the asymptotic regime where $\overline{\lambda} = 1$ and $\beta > 0$ 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 nontrivial (i.e., it lies strictly between zero and one), which better captures the behavior of many systems found in applications. Due to the simultaneous high utilization and good quality of service (as captured by a positive probability of no wait), this asymptotic regime is sometimes also referred to as the Quality-and-Efficiency-Driven regime. Under the assumption of renewal arrivals, exponential service times, normalized to have unit mean and limit mean arrival rate $\overline{\lambda} = 1$, Halfin and Whitt [14] 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 with a constant diffusion coefficient and a state-dependent drift that is linear and restoring (resembling an Ornstein–Uhlenbeck process) below zero and constant above zero. Under the condition $\beta > 0$, which ensures that each N-server queue is stable, this characterization of the limit process was used in [14] to establish approximations to the stationary probability of positive wait in a queue with N servers.

However, in many applications, statistical evidence suggests that it would be more appropriate to model the service times as being nonexponential (see, e.g., the study of real call center data in Brown et al. [4] that indicates 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.

When the service times are not exponentially distributed, any Markovian representation must keep track of the residual service times or the ages of customers in service. This implies that the dimension of any finite-dimensional representation of the state must grow with the number of servers (the dimension must be at least $N+1$ for an N-server system), which poses a challenge for obtaining limit theorems as $N \to \infty$. Instead, in this work a common infinite-dimensional state space is used for all N-server systems. Specifically, the state of the N-server queue is represented 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, where the age is the time elapsed since entry into service. This representation, which provides a Markovian description of the dynamics, was first introduced in Kaspi and Ramanan [21], where it was shown that the fluid-scaled sequence $(X^{(N)}, \nu^{(N)})/N$ converges almost surely to a certain deterministic process $(\overline{X}, \overline{\nu})$, referred to as the fluid limit. Fluctuations around the fluid limit can be captured by the diffusion-scaled state sequence $\{(\widehat{X}^{(N)}, \widehat{\nu}^{(N)})\}_{N\in\mathbb{N}}$, which is obtained by centering the fluid-scaled state $(X^{(N)}, \nu^{(N)})/N$ around the fluid limit $(\overline{X}, \overline{\nu})$ and multiplying the difference by $\sqrt{N}$.

In the present work, under suitable assumptions, in each of the cases when the fluid limit is subcritical, critical or supercritical (which, roughly speaking, correspond to the cases $\overline{\lambda} < 1$, $\overline{\lambda} = 1$ or $\overline{\lambda} > 1$), it is shown in Theorems 2 and 3 that the diffusion-scaled state sequence $\{(\widehat{X}^{(N)}, \widehat{\nu}^{(N)})\}_{N\in\mathbb{N}}$ converges weakly to a càdlàg stochastic process $(\widehat{X}, \widehat{\nu})$. Moreover, the $\widehat{X}$ component of the limit is shown (in Corollary 5.10) to be a real-valued Itô diffusion with a constant diffusion coefficient that is *insensitive* to the service distribution beyond its mean, and whose drift is an adapted process that is a functional of $\widehat{\nu}$. In particular, although $\widehat{X}$ is non-Markovian, it admits a fairly tractable representation. The proof of this representation relies on an asymptotic independence result for the centered arrival and departure processes (see Proposition 8.4), which may be of independent interest. As for the age process, although the $\widehat{\nu}^{(N)}$ are (signed) Radon measure valued processes, the limit $\widehat{\nu}$ lies outside this space. A key challenge was to identify a suitable space in which to establish convergence without imposing overly restrictive assumptions on the service distribution G. Under conditions that include a large class of service distributions relevant in applications such as phase-type, lognormal, logistic and (for a certain class of parameters) Erlang and Pareto distributions, it is shown that $\widehat{\nu}^{(N)}$ converges weakly to $\widehat{\nu}$ in the space of $\mathbb{H}_{-2}$-valued càdlàg processes, where $\mathbb{H}_{-2}$ is the dual of a certain Hilbert space $\mathbb{H}_2$. In addition, in Theorem 5(a) the $\widehat{\nu}$ component of the limit is characterized as the unique solution to a stochastic partial differential equation that is coupled with the Itô diffusion $\widehat{X}$. Furthermore, it is also shown in Theorem 4 that $(\widehat{X}, \widehat{\nu})$ is a semimartingale with an explicit decomposition and in Theorem 5(b) that $(\widehat{X}, \widehat{\nu})$, along with an appended state, is a strong Markov process. The proof of the strong Markov property relies on a consistency property (see Lemma 9.3 and Appendix E), which shows that the assumptions that are imposed on the initial state are also satisfied by the state at any positive time.

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 [27] and Puhalskii and Reed [25]. Under the assumptions that $\overline{\lambda} = 1$, the initial residual service times of customers in service 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 [27] with only a finite mean condition on the service distribution. This result was extended by Puhalskii and Reed [25] to allow for more general, possibly inhomogeneous arrival processes as well as more general conditions on the residual service times of customers in service at the initial time. In this setting, convergence of finite-dimensional distributions was established in [25], and strengthened to process level convergence when the service distribution is continuous. In both papers, the limit is characterized as the unique solution to a certain stochastic convolution equation.

Other 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 [26], in which phase-type service distributions are considered, and the limit is characterized as a multidimensional diffusion, where each dimension corresponds to a different phase of the service distribution. Whitt [32] 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. In addition to the process level results described above, under the stability assumption $\beta > 0$, results on the asymptotics of steady state distributions in the Halfin–Whitt regime have been obtained by Jelenkovic, Mandelbaum and Momçilović [18] for deterministic service times and by Gamarnik and Momçilović [12] 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 assumptions on the service distribution that are satisfied by 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. As a special case, we can recover the results of Halfin and Whitt [14] and Puhalskii and Reiman [26] and, for the smaller class of service distributions that we consider, Reed [27]. The Markovian representation of the state, though infinite dimensional, leads to an intuitive characterization of the dynamics, which facilitates the incorporation of more general features into the model. For example, this framework was extended by Kang and Ramanan to include abandonments in [19] and [20]. In the subcritical case the results of this paper also provide a characterization of the diffusion limit of the well-studied infinite-server queue. The latter is easier to analyze due to the absence of a queue and the consequent lack of 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 [15] and Glynn and Whitt [13], where the limit process is characterized as an Ornstein–Uhlenbeck process, and Krichagina and Puhalskii [22], who provide 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 [7]. In contrast to the infinite-dimensional Markovian representation in terms of residual service times used in Decreusefond and Moyal [7], the Markovian representation in terms of the age process that is used here allows us to associate some natural martingales that facilitate the analysis.

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 relevant functional strong law of large numbers limit results established in [21]. In Section 4 a sequence of martingales that are obtained as compensated departure processes and 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. An associated sequence of stochastic convolution integrals $\widehat{\mathcal{H}}_t^{(N)}$, $t \geq 0$, which arise in the representation of the dynamics, is also defined. 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 $\widehat{\mathcal{H}}^{(N)}$, another stochastic convolution process $\widehat{\mathcal{K}}^{(N)}$ 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 the joint 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, together with the sequence of centered arrival processes and initial conditions (see Corollary 8.7). To maintain the flow of the exposition, some supporting results are relegated to the Appendix. First, in Section 1.4 we introduce some common notation and terminology used in the paper.

1.4. *Notation and terminology.* As usual, let $\mathbb{Z}_+$ denote the set of nonnegative integers, $\mathbb{N}$ denote the set of natural numbers or, equivalently, strictly positive integers and $\mathbb{R}$ denote the set of real numbers. For $a, b \in \mathbb{R}$, let $a \vee b$ and $a \wedge b$, respectively, denote the maximum and minimum of a and b. The short-hand notation a^+ and a^- will sometimes also be used for $a \vee 0$ and $-(a \wedge 0)$, respectively. Given $B \subset \mathbb{R}$, $\mathbb{I}_B$ denotes the indicator function of the set B [i.e., $\mathbb{I}_B(x) = 1$ if $x \in B$ and $\mathbb{I}_B(x) = 0$ otherwise].

1.4.1. *Function spaces.* Given any metric space $\mathcal{E}$, $\mathbb{B}(\mathcal{E})$ denotes the Borel sets of $\mathcal{E}$ (with topology compatible with the metric on $\mathcal{E}$), and $\mathbb{C}(\mathcal{E})$ denotes the space of continuous real-valued functions defined on $\mathcal{E}$. Also, $\mathbb{C}_b(\mathcal{E})$ is the subset of bounded functions in $\mathbb{C}(\mathcal{E})$ and let $\mathbb{C}_c(\mathcal{E})$ be the subset of functions in $\mathbb{C}(\mathcal{E})$ that have compact support in $\mathcal{E}$. Let $\mathbb{D}_{\mathcal{E}}[0,\infty)$ denote the space of $\mathcal{E}$-valued càdlàg functions defined on $[0,\infty)$, and let $\mathrm{supp}(\varphi)$ denote the support of a function φ. When $\mathcal{E}$ is a domain (open connected subset) or closure of a domain in $\mathbb{R}^n$, equipped with the usual Euclidean metric and Lebesgue measure, let $\mathbb{AC}(\mathcal{E})$ represent the space of absolutely continuous functions (in the sense of Carathéodory) defined on $\mathcal{E}$, and let $\mathbb{AC}_b(\mathcal{E})$ denote the subset of bounded functions in $\mathbb{AC}(\mathcal{E})$.

We will mostly be interested in the case when $\mathcal{E} = [0, L)$ and $\mathcal{E} = [0, L) \times [0,\infty)$, for some $L \in (0,\infty]$. To distinguish these cases, f will be used to denote generic functions on $[0, L)$ and φ to denote generic functions on $[0, L) \times [0,\infty)$. By some abuse of notation, given f on $[0, L)$, it will also be treated as a function on $[0, L) \times [0,\infty)$ that is constant in the second variable. For either choice of $\mathcal{E}$, let $\mathbb{C}^1(\mathcal{E})$ and $\mathbb{C}^\infty(\mathcal{E})$, respectively, represent the space of real-valued, once continuously differentiable and infinitely differentiable functions on $\mathcal{E}$, let $\mathbb{C}^1_c(\mathcal{E})$ be the subspace of functions in $\mathbb{C}^1(\mathcal{E})$ that have compact support and $\mathbb{C}^1_b(\mathcal{E})$ the subspace of functions in $\mathbb{C}^1(\mathcal{E})$ that, together with their first derivatives, are bounded. Here, a function on $\mathcal{E}$ is said to be differentiable if it is the restriction to $\mathcal{E}$ of a differentiable function on some open neighborhood of $\mathcal{E}$. Recall that given $T < \infty$ and a continuous function $f \in \mathbb{C}[0, T]$, the modulus of continuity $w_f(\cdot)$ of f is defined by

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

For any $\varphi \in \mathbb{AC}([0, L) \times [0,\infty))$, the partial derivatives φ_x and φ_s are well defined as locally integrable functions on $[0, L) \times [0,\infty)$. The space $\mathbb{C}^{1,1}([0, L) \times [0,\infty))$ is defined as the subset of functions in $\mathbb{AC}([0, L) \times [0,\infty))$ for which $\varphi_x + \varphi_s$, the directional derivative in the $(1, 1)$ direction, is continuous. Moreover, $\mathbb{C}^{1,1}_b([0, L) \times [0,\infty))$ and $\mathbb{C}^{1,1}_c([0, L) \times [0,\infty))$ denote the subset of functions φ in $\mathbb{C}^{1,1}([0, L) \times [0,\infty))$ such that both φ and $\varphi_x + \varphi_s$ are bounded or, respectively, have compact support. Let $\mathbb{I}_0[0,\infty)$ denote the space of nondecreasing functions $f \in \mathbb{D}_{\mathbb{R}}[0,\infty)$ with $f(0) = 0$. For $L \in [0,\infty]$, $\mathbb{L}^\alpha[0, L)$, $\alpha \ge 1$, and $\mathbb{L}^\infty[0, L)$ represent, respectively, the spaces of measurable functions f on $[0, L)$ such that $\int_{[0,L)} |f(x)|^\alpha\,dx < \infty$ and the space of essentially bounded functions (with respect to Lebesgue measure) on $[0, L)$. Also, $\mathbb{L}^\alpha_{\mathrm{loc}}[0, L)$ represents the corresponding space in which the associated property holds only locally, that is, on every compact (i.e., closed and bounded) interval in $[0, L)$. 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 $[0, L)$, we define $\|f\|_T \doteq \sup_{s\in[0,T]} |f(s)|$ for every $T < L$, and let $\|f\|_\infty \doteq \sup_{s\in[0,L)} |f(s)|$, which could possibly equal infinity. As usual, for $f \in \mathbb{D}_{\mathbb{R}}[0,\infty)$ and $t > 0$, let

$f(t-) = \lim_{u \uparrow t} f(u)$ denote the left limit of f at t, with the convention that $f(0-) = f(0)$, and also let $\Delta f(t) = f(t) - f(t-)$ denote the jump of f at t.

For $f, g \in \mathbb{C}_c^\infty[0,\infty)$ and $n = 0, 1, 2, \ldots$, consider the weighted inner product defined by

$$\langle f, g \rangle_{\mathbb{H}_n} \doteq \sum_{m=0}^{n} \int_0^\infty \frac{d^m f}{dx^m}(x) \frac{d^m g}{dx^m}(x)(1+|x|^2)^n \, dx \tag{1.2}$$

and the associated norm

$$\|f\|_{\mathbb{H}_n} \doteq (\langle f, f \rangle_{\mathbb{H}_n})^{1/2}. \tag{1.3}$$

It is clear from the definition that the norm $\|\cdot\|_{\mathbb{H}_n}$ is Hilbertian (i.e., it satisfies $\|f+g\|_{\mathbb{H}_n}^2 - \|f-g\|_{\mathbb{H}_n}^2 = 2\|f\|_{\mathbb{H}_n}^2 + 2\|g\|_{\mathbb{H}_n}^2$). For each $n \in \mathbb{N}$, let $\mathbb{H}_n$ be the (weighted Sobolev) space obtained as the completion of $\mathbb{C}_c^\infty[0,\infty)$ with respect to the norm $\|\cdot\|_{\mathbb{H}_n}$. It is easy to verify that each $\mathbb{H}_n$ is a complete separable Hilbert space and, for $m < n$, $\mathbb{H}_n \subset \mathbb{H}_m$. Now, for $m, n \in \mathbb{N}$, $m < n$, $\|\cdot\|_{\mathbb{H}_m}$ is said to be HS weaker than $\|\cdot\|_{\mathbb{H}_n}$, and denoted $\|\cdot\|_{\mathbb{H}_m} \overset{\text{HS}}{<} \|\cdot\|_{\mathbb{H}_n}$, if the injection map from $\mathbb{H}_n$ to $\mathbb{H}_m$ is a Hilbert–Schmidt (or, equivalently, quasi-nuclear) operator (see, e.g., page 6 of [2] or page 330 of [31]). It follows from Theorems 3.6 and 3.7 of [2] that for all $m, n \in \mathbb{N}$,

$$n > m \quad \Rightarrow \quad \|\cdot\|_{\mathbb{H}_m} \overset{\text{HS}}{<} \|\cdot\|_{\mathbb{H}_n}. \tag{1.4}$$

Now, let $\mathbb{H}_{-n}$ be the dual of $\mathbb{H}_n$, where the dual norm $\|\cdot\|_{\mathbb{H}_{-n}}$ is given by

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

where $\{e_{nk}, k = 1, \ldots\}$ is a complete orthonormal system in $\mathbb{H}_n$. Moreover, let $\mathcal{S}$ be the Schwartz space of rapidly decreasing functions, namely the space of $\mathbb{C}^\infty$ functions on $[0,\infty)$, for which the following semi-norms are finite:

$$\|f\|_{\beta,\gamma} = \sup_{x \in [0,\infty)} \left| x^\beta \frac{d^\gamma f}{dx^\gamma}(x) \right|, \qquad \beta \in \mathbb{N}, \gamma \in \mathbb{N}.$$

Let $\mathcal{S}'$ be the topological dual of $\mathcal{S}$, which is the Schwartz space of tempered distributions. It is well known that $\mathcal{S}$ and $\mathcal{S}'$ are separable, nuclear Fréchet spaces. Moreover, the projective limit of the spaces $\mathbb{H}_n, n \in \mathbb{N}$, coincides with $\mathcal{S}$, and the dual space satisfies $\mathcal{S}' = \bigcup_{n=0}^\infty \mathbb{H}_{-n}$; see Theorem 3.8, relation (3.39) and the comment at the end of Section 3.10 of [2]. For $\nu \in \mathcal{S}'$ and $f \in \mathcal{S}$ and likewise, for $\nu \in \mathbb{H}_{-n}$ and $f \in \mathbb{H}_n$, let $\nu(f)$ denote the duality pairing. For $n = 1, 2$ and f for which the corresponding first or second (weak) derivatives are well defined, sometimes the notation $f' = f^{(1)}$ and $f'' = f^{(2)}$ will also be used. The usual $\mathbb{L}^2$ norm is denoted by

$$\|f\|_{\mathbb{L}^2} = \left(\int_{\mathbb{R}} |f(x)|^2 \, dx \right)^{1/2}.$$

It follows immediately from the definition of the norms $\|\cdot\|_n$ given above that

$$
\begin{aligned}
\|f\|_{\mathbb{L}^2} &= \|f\|_{\mathbb{H}_0}, \\
\|f\|_{\mathbb{L}_2}^2 + \|f'\|_{\mathbb{L}^2}^2 &\leq \|f\|_{\mathbb{H}_1}^2, \qquad (1.5)\\
\|f\|_{\mathbb{L}_2}^2 + \|f'\|_{\mathbb{L}^2}^2 + \|f''\|_{\mathbb{L}^2}^2 &\leq \|f\|_{\mathbb{H}_2}^2.
\end{aligned}
$$

Moreover, for any $f \in \mathbb{H}_1$, it is easy to deduce the norm inequalities

$$
|f(0)| \leq \sqrt{2}\|f\|_{\mathbb{H}_1}, \qquad \|f\|_\infty \leq 2\|f\|_{\mathbb{H}_1}, \tag{1.6}
$$

which will be used in the sequel. Indeed, if $f \in \mathbb{H}_1$, then f is absolutely continuous, and both f and f' lie in $\mathbb{L}^2$. Therefore, there exists a real-valued sequence $\{x_n\}$ with $x_n \to \infty$ and $f(x_n) \to 0$ as $n \to \infty$. Since for each $n \in \mathbb{N}$, we have $f^2(x_n) - f^2(0) = 2\int_0^{x_n} f(u)f'(u)\,du$, applying the Cauchy–Schwarz inequality and then taking limits as $n \to \infty$, we obtain $|f(0)|^2 \leq 2\|f\|_{\mathbb{L}^2}\|f'\|_{\mathbb{L}^2} \leq 2\|f\|_{\mathbb{H}_1}^2$, where the last inequality follows from the second inequality in (1.5). This yields the first inequality in (1.6). The second inequality in (1.6) can be inferred in a similar manner, applying the Cauchy–Schwarz inequality to the relation $f^2(x) = f^2(0) + 2\int_0^x f(u)f'(u)\,du$ and using the first inequality in (1.6).

Finally, let $\mathcal{D}[0,\infty)$ denote the usual space of test functions, namely the space $\mathbb{C}_c^\infty[0,\infty)$ equipped with the following notion of convergence: $f_n \to f$ in $\mathcal{D}[0,\infty)$ if the functions f_n are supported in a common compact set and, for every $m \in \mathbb{N}$, $d^m f_n/dx^m \to d^m f/dx^m$ uniformly. Also, let $\mathcal{D}'[0,\infty)$ denote its dual, the space of distributions. It is well known (see, e.g., Theorem 3.9 of [2]) that both $\mathcal{D}[0,\infty)$ and $\mathcal{D}'[0,\infty)$ are nuclear spaces.

1.4.2. *Measure spaces*. The space of Radon measures on a metric space $\mathcal{E}$, endowed with the Borel σ-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 $\mathbb{M}_{\leq N}(\mathcal{E})$ is the subspace of positive measures with total mass less than or equal to N. Note that then $\mathbb{M}_{\leq 1}(\mathcal{E})$ is the space of subprobability measures. For any Borel measurable function $f:\mathcal{E} \to \mathbb{R}$ that is integrable with respect to $\xi \in \mathbb{M}(\mathcal{E})$, the short-hand notation $\langle f, \xi\rangle \doteq \int_{\mathcal{E}} f(x)\xi(dx)$ will be used. Recall that a Radon measure on $\mathcal{E}$ is one that assigns 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 $\mathbb{C}_c(\mathcal{E})$ defined by $f \mapsto \langle f, \mu\rangle$, a Radon measure on $\mathcal{E}$ can be equivalently defined as a linear mapping from $\mathbb{C}_c(\mathcal{E})$ into $\mathbb{R}$ such that for every compact set $\mathcal{K} \subset \mathcal{E}$, there exists $L_{\mathcal{K}} < \infty$ such that

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

The space $\mathbb{M}_F(\mathcal{E})$ is equipped with the weak topology [generated by sets of the form $\{\mu : \langle f_1, \mu - \mu_o\rangle < \varepsilon_1, \ldots, \langle f_n, \mu - \mu_o\rangle < \varepsilon_n\}$, for $\mu_o \in \mathbb{M}_F(\mathcal{E})$, $n \in \mathbb{N}$, $f_i \in \mathbb{C}_b(\mathcal{E})$ and $\varepsilon_i > 0$, $i = 1, \ldots, n$]. Also, recall that a sequence $\{\mu_n\}_{n\in\mathbb{N}}$ in $\mathbb{M}_F(\mathcal{E})$ converges to $\mu \in \mathbb{M}_F(\mathcal{E})$ in the weak topology (denoted $\mu_n \stackrel{w}{\to} \mu$) if for every

$f \in \mathbb{C}_b(\mathcal{E})$, $\langle f, \mu_n \rangle \to \langle f, \mu \rangle$ as $n \to \infty$. The symbol δ_x will be used to denote the measure with unit mass at the point x, and the symbol $\tilde{\mathbf{0}}$ will be used to denote the identically zero Radon measure. When $\mathcal{E}$ is an interval, say $[0, L)$, for conciseness the notation $\mathbb{M}[0, L)$ will be used instead of $\mathbb{M}([0, L))$. Also, for ease of notation, given $\xi \in \mathbb{M}[0, L)$ and an interval $(a, b) \subset [0, L)$, $\xi(a, b)$ and $\xi(a)$ will be used to denote $\xi((a, b))$ and $\xi(\{a\})$, respectively.

1.4.3. *Stochastic processes.* Given a Polish space $\mathcal{V}$, $\mathbb{D}_{\mathcal{V}}[0, T]$ and $\mathbb{D}_{\mathcal{V}}[0, \infty)$ denote the spaces of $\mathcal{V}$-valued, càdlàg functions on $[0, T]$ and $[0, \infty)$, respectively, endowed with the usual Skorokhod J_1-topology; see [3] for details on this topology. Then $\mathbb{D}_{\mathcal{V}}[0, T]$ and $\mathbb{D}_{\mathcal{V}}[0, \infty)$ are also Polish spaces. We will be interested in $\mathcal{V}$-valued stochastic processes, especially in the cases when $\mathcal{V} = \mathbb{R}$, $\mathcal{V} = \mathbb{M}_F[0, L)$ for some $L \leq \infty$, $\mathcal{V} = \mathcal{S}'[0, L)$ and $\mathcal{V} = \mathbb{H}_{-n}[0, L)$ 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 $\mathbb{D}_{\mathcal{V}}[0, \infty)$, equipped with the Borel σ-algebra (generated by open sets under the Skorokhod J_1-topology). A sequence $\{Z^{(N)}\}_{N \in \mathbb{N}}$ of càdlàg, $\mathcal{V}$-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 $\mathcal{V}$-valued process Z defined on $(\Omega, \mathcal{F}, \mathbb{P})$ if and only if for every bounded, continuous functional $F : \mathbb{D}_{\mathcal{V}}[0, \infty) \to \mathbb{R}$,

$$\lim_{n \to \infty} \mathbb{E}^{(N)}\big[F\big(Z^{(N)}\big)\big] = \mathbb{E}[F(Z)],$$

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. The many-server model under consideration is introduced in Section 2.1. The state descriptor and the dynamical equations that describe the evolution of the state are presented in Section 2.2.

2.1. *The N-server model.* Consider a system with N servers, in which arriving customers are served in a nonidling, first-come-first-serve (FCFS) manner; that is, 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 nonidling condition is satisfied. Customers are assumed to be infinitely patient; that is, they wait in queue till they receive service. Servers are nonpreemptive 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. Then, due to the nonidling condition, the number of customers in service at time 0 is equal to $X^{(N)}(0) \wedge N$. 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 their ages at time zero, where the age of a customer that has entered service is defined to be the minimum of the amount of time elapsed since the customer entered service and the service time, as defined explicitly in (2.5) below. In particular, v_0 is the service time of the customer who has spent the least time in service amongst those in service at time zero. On the other hand, for $i \in \mathbb{N}$, v_i represents the service requirement of the ith customer to enter service after time 0.

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

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

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

- $E^{(N)}$ is a càdlàg nondecreasing pure jump process with $E^{(N)}(0) = 0$ and almost surely, $E^{(N)}(t) < \infty$ and $E^{(N)}(t) - E^{(N)}(t-) \in \{0, 1\}$ for $t \in [0, \infty)$.
- The process $R_E^{(N)}$ is Markovian with respect to the usual augmentation of its own natural filtration; see, for example, page 10 of [30] for an explicit construction of the filtration.
- The cumulative arrival process is independent of the i.i.d. sequence of service requirements $\{v_j, j = -N+1, \dots\}$. Moreover, given $\sigma(R_E^{(N)}(0))$, the σ-algebra generated by $R_E^{(N)}(0)$, the process $\{E^{(N)}(t), t > 0\}$ is independent of $X^{(N)}(0)$ and the ages of the customers that have entered service by time zero.
- G has density g.
- Without loss of generality, we assume that the mean service requirement is 1.

$$\int_{[0,\infty)} \bigl(1 - G(x)\bigr)\, dx = \int_{[0,\infty)} x g(x)\, dx = 1. \tag{2.2}$$

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. 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 [1]) 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 [1].

The processes $R_E^{(N)}$ and $E^{(N)}$ described above have trajectories in $\mathbb{D}_{\mathbb{R}}[0,\infty)$; see, for example, Appendix A of [19]. 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 [21], the state of the system will be represented 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, nonnegative 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 jth customer is (for each realization) a piecewise linear function that is zero until 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 fully describe the state dynamics, it will be convenient to introduce the following auxiliary processes:

- the cumulative departure process $D^{(N)}$, where $D^{(N)}(t)$ is the 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

$$D^{(N)}(t) \doteq X^{(N)}(0) - X^{(N)}(t) + E^{(N)}(t), \qquad t \in [0,\infty). \tag{2.3}$$

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 of customers in service yields the relation

$$K^{(N)}(t) \doteq \langle \mathbf{1}, \nu_t^{(N)}\rangle - \langle \mathbf{1}, \nu_0^{(N)}\rangle + D^{(N)}(t), \qquad t \in [0,\infty). \tag{2.4}$$

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. Note that $\theta_j^{(N)}$ denotes the time of entry into service of the jth customer to enter service after time 0. In addition, for $j = -(X^{(N)}(0) \wedge N) + 1, -X^{(N)}(0) \wedge N, \dots, 0$, set $\theta_j^{(N)} = -a_j^{(N)}(0)$ to be the amount of time that the jth 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

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

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 $\nu_t^{(N)}$ representing the age distribution at time t can then be expressed as

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

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

$$N - \langle \mathbf{1}, \nu_t^{(N)} \rangle = [N - X^{(N)}(t)]^+, \qquad t \in [0, \infty). \tag{2.7}$$

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

$$K^{(N)}(t) = X^{(N)}(t) \wedge N - X^{(N)}(0) \wedge N + D^{(N)}(t), \qquad t \in [0, \infty). \tag{2.8}$$

Clearly $\langle \mathbf{1}, \nu_t^{(N)} \rangle \leq N$ for every $t \in [0, \infty)$ 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 $[0, \infty) \times \mathbb{N} \times \mathbb{M}_F[0, L)$. For an explicit construction of the state that also shows that the state and auxiliary processes are well defined and càdlàg; see Lemma A.1 of [19]. The results obtained in this paper are independent of the particular rule used to assign customers to stations, but for technical purposes it will be 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 $\tilde{\mathcal{F}}_t^{(N)}$ be the σ-algbera 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 [21]. In fact, as shown in Lemma B.1 of [19], $\{(R_E^{(N)}(t), X^{(N)}(t), \nu_t^{(N)}), \mathcal{F}_t^{(N)}, t \geq 0\}$ is a strong Markov process.

For future purposes, we introduce some standard Markov process notation (see [30]) associated with the Markov process $(R_E^{(N)}, X^{(N)}, \nu^{(N)}) = \{(R_E^{(N)}(t), X^{(N)}(t), \nu_t^{(N)}), t \geq 0\}$. Let $\mathcal{G}_t^{(N),0}$ be the σ-algebra generated by $\{(R_E^{(N)}(s), X^{(N)}(s), \nu_s^{(N)}), s \in [0,t]\}$, and let $\mathcal{G}_\infty^{(N),0} = \sigma(\bigcup_{t\geq 0} \mathcal{G}_t^{(N),0})$. For $(r,k,\mu) \in [0,\infty) \times \mathbb{N} \times \mathbb{M}_F[0,L)$, let $\mathbb{P}_{r,k,\mu}^{(N)}$ be the law of the Markov process $(R_E^{(N)}, X^{(N)}, \nu^{(N)})$ with initial condition $(R_E^{(N)}(0), X^{(N)}(0), \nu_0^{(N)}) = (r,k,\mu)$. Also, let $\{\tilde{\mathcal{G}}_t^{(N)}, t \geq 0\}$ be the usual augmentation of the filtration $\{\mathcal{G}_t^{(N),0}, t \geq 0\}$ (as carried out, e.g., in page 25 of [30]) and let $\mathcal{G}_t^{(N)} = \bigcap_{s>t} \mathcal{G}_s^{(N),0}$, $t \geq 0$, be the associated right-continuous filtration. Note that for every $t \geq 0$, $\mathcal{G}_t^{(N)} \subseteq \mathcal{F}_t^{(N)}$ and that $\{(R_E^{(N)}, X^{(N)}, \nu^{(N)}), \mathbb{P}_{r,k,\mu}^{(N)}, (r,k,\mu) \in [0,\infty) \times \mathbb{N} \times \mathbb{M}_F[0,L)\}$ is a strong Markov family.

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

3. Fluid limit. In this section we recall the functional strong law of large numbers limit or, equivalently, fluid limit obtained in [21]. 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:

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

where we recall that $\mathbb{I}_0[0,\infty)$ is the subset of nondecreasing functions $f \in \mathbb{D}_{[0,\infty)}[0,\infty)$ with $f(0) = 0$. When $N = 1$, $\mathcal{I}_0^1$ will be denoted simply by $\mathcal{I}_0$. Assume that $\mathcal{I}_0^N$ 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

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

and let

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

for $N \in \mathbb{N}$. Observe that the fluid-scaled initial data $(\overline{E}^{(N)}, \overline{X}^{(N)}(0), \overline{\nu}_0^{(N)})$ lies in $\mathcal{I}_0$. The strong law of large numbers results in [21] 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 almost surely, as $N \to \infty$,

$$\big(\overline{E}^{(N)}, \overline{X}^{(N)}(0), \overline{\nu}_0^{(N)}\big) \to (\overline{E}, \overline{x}_0, \overline{\nu}_0) \qquad \text{in } \mathcal{I}_0$$

and moreover, as $N \to \infty$, $\mathbb{E}[\overline{X}^{(N)}(0)] \to \overline{x}_0$ and $\mathbb{E}[\overline{E}^{(N)}(t)] \to \overline{E}(t)$ for every $t \in [0,\infty)$.

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

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

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

$$\int_a^b h(x)\,dx = \ln\big(1-G(a)\big) - \ln\big(1-G(b)\big) < \infty. \tag{3.4}$$

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

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

(a) $L = \infty$ and there exists $\ell' < \infty$ such that h is bounded on (ℓ', ∞);
(b) there exists $\ell' < L$ such that h is lower-semicontinuous on (ℓ', L).

A succinct description of the dynamics of the N-server system in terms of certain integral equations is provided in Proposition 6.1; see also Theorem 5.1 of [21]. The deterministic analog of these equations, the so-called fluid equations, is introduced below.

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

$$\int_0^t \langle h, \overline{\nu}_s \rangle\, ds < \infty, \tag{3.5}$$

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

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

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

and

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

where $\overline{K}$ is a nondecreasing process that satisfies

$$\overline{K}(t) = \langle \mathbf{1}, \overline{\nu}_t\rangle - \langle \mathbf{1}, \overline{\nu}_0\rangle + \int_0^t \langle h, \overline{\nu}_s\rangle\, ds, \qquad t \in [0,\infty). \tag{3.9}$$

We now recall the result established in [21] (see Theorems 3.5 and 3.7 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 1 (Kaspi–Ramanan [21]). *Suppose Assumptions 1 and 2 are satisfied, and $(\overline{E}, \overline{x}_0, \overline{\nu}_0) \in \mathcal{I}_0$ is the limit of the initial data as stated in Assumption 1. Then there exists a unique solution $(\overline{X}, \overline{\nu})$ to the associated fluid equations (3.5)–(3.8) and, as $N \to \infty$, $(\overline{X}^{(N)}, \overline{\nu}^{(N)})$ converges almost surely to $(\overline{X}, \overline{\nu})$. Moreover, $(\overline{X}, \overline{\nu})$ satisfies for every $f \in \mathbb{C}_b([0,\infty))$,*

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

where $\overline{K}$ is a nondecreasing process that satisfies the relation (3.9). Furthermore, if $\overline{E}$ is continuous, then $(\overline{X}, \overline{\nu})$ and $\overline{K}$ are also continuous.

REMARK 3.2. In this context, the unique solution $(\overline{X}, \overline{\nu})$ to the fluid equations will also be referred to as the fluid limit. The fluid limit is said to be critical if $\overline{X}(t) = 1$ for all $t \in [0,\infty)$. In addition, it is said to be subcritical (resp., supercritical) if for every $T \in [0,\infty)$, $\sup_{t\in[0,T]} \overline{X}(t) < 1$ [resp., $\inf_{t\in[0,T]} \overline{X}(t) > 1$]. 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, that is,

$(\overline{x}_0, \overline{\nu}_0) = (0, \tilde{\mathbf{0}})$, 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 $\overline{\nu}_*$ is the "invariant" fluid age measure, defined to be

$$\overline{\nu}_*(dx) = \big(1 - G(x)\big)\,dx, \qquad x \in [0, L), \tag{3.11}$$

then it follows from Remark 3.8 and Theorem 3.9 of [21] that the fluid limit associated with the initial data $(\mathbf{1}, 1, \overline{\nu}_*)$ is critical, the fluid limit associated with the initial data $(\mathbf{1}, a, \overline{\nu}_*)$ for some $a > 1$ is supercritical and if the support of G is $[0, \infty)$, then the fluid limit associated with the initial data $(\overline{\lambda}\mathbf{1}, 0, \tilde{\mathbf{0}})$ is subcritical whenever $\overline{\lambda} \leq 1$. A complete characterization of the invariant manifold of the fluid in the presence of abandonments can be found in [20].

4. Certain martingale measures and their stochastic integrals. We now introduce some quantities that arise in the proof of the functional central limit theorem of the state process. 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 [21]. Whereas under the fluid scaling the limit of this sequence converges weakly to zero, under the diffusion scaling considered here, it converges to a nontrivial 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 [31] for basic definitions of martingale measures and their stochastic integrals.

4.1. *A martingale measure sequence.* Throughout this section, let $(E^{(N)}, X^{(N)}(0), \nu_0^{(N)})$ be an $\mathcal{I}_0^N$-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 φ on $[0, L) \times [0, \infty)$, consider the sequence of processes $\{Q_\varphi^{(N)}\}_{N\in\mathbb{N}}$ defined by

$$\begin{aligned} Q_\varphi^{(N)}(t) \doteq \sum_{s\in[0,t]} \sum_{j=-\langle \mathbf{1}, \nu_0^{(N)}\rangle+1}^{K^{(N)}(t)} & \mathbb{I}_{\{(da_j^{(N)}/dt)(s-)>0, (da_j^{(N)}/dt)(s+)=0\}} \\ & \times \varphi\big(a_j^{(N)}(s), s\big) \end{aligned} \tag{4.1}$$

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 relations (2.4) and (2.5). Note from (2.5) that the jth customer completed service (and hence departed the system) at time s if and only if

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

This is equivalent to the condition $a_j^{(N)}(s) = v_j$, and thus $\varphi(a_j^{(N)}(s), s)$ can in fact be replaced by $\varphi(v_j, s)$ in (4.1). Substituting $\varphi = \mathbf{1}$ in (4.1), it is clear 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 φ on $[0, L) \times [0, \infty)$, consider the process $A_\varphi^{(N)}$ defined by

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

and set

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

It was shown in Corollary 5.5 of [21] that for all functions $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$, $A_\varphi^{(N)}$ is the $\{\mathcal{F}_t^{(N)}\}$-compensator of $Q_\varphi^{(N)}$, and $M_\varphi^{(N)}$ is a càdlàg $\{\mathcal{F}_t^{(N)}\}$-martingale.

REMARK 4.1. In fact, $M_\varphi^{(N)}$ is a càdlàg $\{\mathcal{F}_t^{(N)}\}$-martingale for all φ in the larger class of bounded and measurable functions on $[0, L) \times [0, \infty)$. Indeed, since the mapping $(\omega, s) \to (a_j^{(N)}(\omega), s)$ on $\Omega \times [0, \infty)$ is continuous and $\{\mathcal{F}_t^{(N)}\}$-adapted, it is $\{\mathcal{F}_t^{(N)}\}$-predictable. Therefore, when φ is measurable, the mapping $(\omega, s) \mapsto \varphi(a_j^{(N)}(s, \omega), s)$ from $\Omega \times [0, \infty)$ to $\mathbb{R}$ is also $\{\mathcal{F}_t^N\}$-predictable. The assertion then follows from essentially the same argument as that used in Lemma 5.9 of [21].

From the proof of Lemma 5.9 of [21] it follows that for any bounded and measurable φ on $[0, L) \times [0, \infty)$, the $\{\mathcal{F}_t^{(N)}\}$-predictable quadratic variation of $M_\varphi^{(N)}$ takes the form

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

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

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

Let $\mathbb{B}_0[0, L)$ denote the algebra generated by the intervals $[0, x]$, $x \in [0, L)$. It is easy to verify that $\mathcal{M}^{(N)} = \{\mathcal{M}_t^{(N)}(B), \mathcal{F}_t^{(N)}, t \geq 0, B \in \mathbb{B}_0[0, L)\}$ 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 [31] for a definition). Essentially, the orthogonality property holds because almost surely, no two departures occur at the same time. In Lemma 4.2 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 4.2. *For every $N \in \mathbb{N}$, $\mathbb{P}$ almost surely,*

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

and

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

The proof of the lemma is relegated to Section A.2, and the orthogonality property is now established.

COROLLARY 4.3. *For each $N \in \mathbb{N}$, the martingale measure*

$$\mathcal{M}^{(N)} = \{\mathcal{M}_t^{(N)}(B), \mathcal{F}_t^{(N)}; t \geq 0, B \in \mathbb{B}_0[0, L)\}$$

is orthogonal and has covariance functional

$$\begin{aligned} \mathcal{Q}_t^{(N)}(B, \tilde{B}) &\doteq \langle \mathcal{M}^{(N)}(B), \mathcal{M}^{(N)}(\tilde{B}) \rangle_t = A^{(N)}_{\mathbb{I}_{B \cap \tilde{B}}}(t) \\ &= \int_0^t \left(\int_{B \cap \tilde{B}} h(x) \nu_s^{(N)}(dx) \right) ds \end{aligned} \tag{4.8}$$

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

PROOF. In order to show that the martingale measure $\mathcal{M}^{(N)}$ is orthogonal, it suffices to show that for every $B, \tilde{B} \in \mathbb{B}_0[0, L)$, such that $B \cap \tilde{B} = \varnothing$, 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

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

Here, $\langle \cdot, \cdot \rangle$ represents the $\{\mathcal{F}_t^{(N)}\}$-predictable quadratic covariation between the two martingales. Fix two sets $B, \tilde{B} \in \mathbb{B}_0[0, L)$ with $B \cap \tilde{B} = \varnothing$. By (4.1), (4.3) and Remark 4.1, it follows that $\mathcal{M}^{(N)}(B) = M^{(N)}_{\mathbb{I}_B}$ and $\mathcal{M}^{(N)}(\tilde{B}) = M^{(N)}_{\mathbb{I}_{\tilde{B}}}$ 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 4.2, there are almost surely no two departures that occur at the same time, the set of jump points of $\mathcal{M}^{(N)}(B)$ and $\mathcal{M}^{(N)}(\tilde{B})$ are almost surely disjoint. By Theorem 4.52 of Chapter 1 of Jacod and Shiryaev [16], it then follows that the martingales are orthogonal, and (4.9) holds. The relation (4.8) follows on combining (4.9) with (4.4) and the biadditivity of the covariance functional. □

The orthogonality property established in Corollary 4.3 allows us to define stochastic integrals with respect to the martingale measure $\mathcal{M}^{(N)}$. The stochastic integral is defined for a large class of so-called predictable integrands satisfying a suitable integrability property (see page 292 of Walsh [31]) which, since $\mathbb{E}[A_{\mathbf{1}}^{(N)}(T)] < \infty$ by Lemma 5.6 of [21] and $\nu^{(N)}$ is a finite nonnegative measure, includes the class of deterministic functions in $\mathbb{C}_b([0, L) \times [0, \infty))$. Moreover, by Theorem 2.5 on page 295 of Walsh [31], it follows that for all $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$, the stochastic integral $\{\mathcal{M}_t^{(N)}(\varphi)(B), \{\mathcal{F}_t^{(N)}\}; t \geq 0, B \in \mathbb{B}_0[0, L)\}$ of φ with respect to $\mathcal{M}^{(N)}$ is a càdlàg orthogonal martingale measure with covariance functional

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

for bounded, continuous $\varphi, \tilde{\varphi}$ and $B, \tilde{B} \in \mathbb{B}_0[0, L)$. 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 4.4. For $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$, 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 [21] that

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

in the space $\mathbb{D}_{\mathbb{M}_F[0,L)}[0, \infty)$. Now, let $\widehat{\mathcal{M}}^{(N)}$ be the diffusion-scaled version of the process

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

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

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

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

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

4.2. *Some associated stochastic convolution integrals.* In Proposition 6.4 it is shown 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 representation that is similar to the representation (3.10) for its fluid counterpart $\{\overline{\nu}_t, t \geq 0\}$, except that it contains an additional stochastic term involving a stochastic convolution integral with respect to the martingale measure $\mathcal{M}^{(N)}$, which is defined below. For $N \in \mathbb{N}$, $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$ and $t \in [0, \infty)$, define

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

For each $t \in [0, \infty)$, the stochastic integral with respect to $\mathcal{M}^{(N)}$ in (4.13) 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 $\mathbb{C}_b([0, L) \times [0, \infty))$ for all $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$. The scaled version of this quantity is then defined in the obvious manner: for $N \in \mathbb{N}$, $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$ and $t \in [0, \infty)$, let

$$\begin{aligned}\widehat{\mathcal{H}}_t^{(N)}(\varphi) &\doteq \frac{\mathcal{H}^{(N)}}{\sqrt{N}} \\ &= \iint_{[0,L)\times[0,t]} \varphi(x+t-s, s) \frac{1-G(x+t-s)}{1-G(x)} \widehat{\mathcal{M}}^{(N)}(dx, ds).\end{aligned} \tag{4.14}$$

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 $\widehat{\mathcal{M}} = \{\widehat{\mathcal{M}}_t(B), B \in \mathbb{B}_0[0, L), t \in [0, \infty)\}$ be a continuous martingale measure with the deterministic covariance functional

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

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

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

Note that $\mathbb{C}_{\widehat{\mathcal{M}}}$ includes, in particular, the space $\mathbb{C}_b([0, L) \times [0, \infty))$.

For any $\varphi \in \mathbb{C}_{\widehat{\mathcal{M}}}$ and $t \in [0, \infty)$, the stochastic integral of φ with respect to $\widehat{\mathcal{M}}$ on $[0, L) \times [0, t]$, denoted by

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

is well defined. In fact, for such φ, $\widehat{\mathcal{M}}(\varphi) = \{\widehat{\mathcal{M}}_t(\varphi), t \geq 0\}$ is a càdlàg, orthogonal martingale measure; see page 292 of Walsh [31] for the definition. 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 $f \in \mathbb{C}_b[0,L)$, let $\widehat{\mathcal{H}}_t(f)$ be the random variable given by the following convolution integral:

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

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

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

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

$$\sup_{t\in[0,\infty)} \|\Psi_t f\|_\infty \leq \|f\|_\infty. \tag{4.20}$$

The processes $\widehat{\mathcal{H}}$ and $\widehat{\mathcal{H}}^{(N)}$, respectively, can then be rewritten in terms of $\widehat{\mathcal{M}}$ and $\widehat{\mathcal{M}}^{(N)}$ as follows:

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

It is shown in Lemma 8.6 and Corollary 8.7 that if f is bounded and Hölder continuous then the real-valued stochastic process $\widehat{\mathcal{H}}^{(N)}(f) = \{\widehat{\mathcal{H}}_t^{(N)}(f), t \geq 0\}$ admits a càdlàg version and the process $\widehat{\mathcal{H}}(f) = \{\widehat{\mathcal{H}}_t(f), t \geq 0\}$ admits a continuous version, and, moreover, that $\widehat{\mathcal{H}}^{(N)}$ and $\widehat{\mathcal{H}}$ also admit versions as, respectively, càdlàg and continuous $\mathbb{H}_{-2}$-valued processes.

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*. For $Y = E, x_0, \nu, X, K$, let $\overline{Y}$ be the corresponding fluid limit as described in Theorem 1. For $N \in \mathbb{N}$, the diffusion scaled quantities $\widehat{Y}^{(N)}$ are defined as follows:

$$\widehat{Y}^{(N)} \doteq \sqrt{N}\big(\overline{Y}^{(N)} - \overline{Y}\big). \tag{5.1}$$

For simplicity, we only consider arrival processes that are 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 $\overline{\lambda}, \sigma^2 \in (0,\infty)$ 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 $(\sigma^2/\overline{\lambda})/(\lambda^{(N)})^2$, where

$$\lambda^{(N)} \doteq \overline{\lambda} N - \beta\sqrt{N}, \tag{5.2}$$

and the following Lindeberg condition holds: for every $\varepsilon > 0$,

$$\lim_{N\to\infty} N^2 \mathbb{E}\big[(\xi_1^{(N)})^2 \mathbb{I}_{\{\xi_1^{(N)}\sqrt{N}>\varepsilon\}}\big] = 0;$$

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

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

REMARK 5.1. Let $\overline{\lambda}(\cdot)$ and $\beta(\cdot)$ be the locally integrable functions defined in Assumption 3, and note that they are in fact constant if Assumption 3(a) holds. Also, let $\sigma(\cdot)$ be the locally square integrable function that is equal to the constant $\sqrt{\sigma^2}$ if Assumption 3(a) holds, and is equal to $(\overline{\lambda}(\cdot))^{1/2}$ if Assumption 3(b) holds. Then, given a standard Brownian motion B, the process $\widehat{E}$ given by

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

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 $\overline{E}$ in Assumption 1 is given by $\overline{E}(t) = \int_0^t \overline{\lambda}(s)\,ds$, $t \geq 0$, and $\widehat{E}^{(N)} \Rightarrow \widehat{E}$ as $N \to \infty$ (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 $\mathbb{D}_{\mathbb{R}}[0,\infty)$ for bounded and Hölder continuous functions f in Section 8.

ASSUMPTION 4. The function $y \mapsto (1 - G(x+y))/(1 - G(x))$ is Hölder continuous on $[0,\infty)$, uniformly with respect to $x \in [0, L)$, that is, there exist $C_G < \infty$, $\gamma_G \in (0,1]$ and $\delta > 0$ such that for every $x \in [0, L)$ and $y, \tilde{y} \in [0, L)$ with $|y - \tilde{y}| < \delta$,

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

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 $[0, L)$. In either case, it follows that $L = \infty$ because the hazard rate function h is locally integrable, but not integrable, on $[0, L)$. Under the first condition above, for any $x, y, \tilde{y} \in [0, \infty)$, $\tilde{y} < y$,

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

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

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

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, Hölder's inequality implies that

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

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

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

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

The process $\{\mathcal{J}_t^{\widehat{\nu}_0^{(N)}}(f), t \ge 0\}$ 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 $\mathcal{J}^{\widehat{\nu}_s^{(N)}}$ more concisely, consider the following family of operators: for $t \in [0, \infty)$, define

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

Since G is continuous, each Φ_t maps the space of bounded and measurable (resp., continuous) functions on $[0, L)$ into itself and, moreover,

$$\sup_{t\in[0,\infty)} \|\Phi_t f\|_\infty \le \|f\|_\infty. \tag{5.7}$$

For future purposes, note that $\{\Phi_t, t \ge 0\}$ defines a semigroup, that is, $\Phi_0 f = f$ and

$$\Phi_t(\Phi_s f) = \Phi_{t+s} f, \qquad s, t \ge 0. \tag{5.8}$$

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

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

The process $\mathcal{J}^{\widehat{\nu}_s^{(N)}}$ can now be rewritten in terms of the operators Φ_t, $t \geq 0$, as follows:

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

The properties stated above imply $\Phi_t f \in \mathbb{C}_b[0, L)$ when $f \in \mathbb{C}_b[0, L)$ and hence, for each $s \geq 0$, $\{\mathcal{J}_t^{\widehat{\nu}_s^{(N)}}(f), f \in \mathbb{C}_b[0, L), t \geq 0\}$ is a well-defined stochastic process. In what follows, we will refer to the spaces $\mathbb{H}_n$ and $\mathbb{H}_{-n}$ that were introduced in Section 1.4.1.

REMARK 5.3. Since $\widehat{\nu}_0^{(N)}$ is a signed measure with finite total mass bounded by $\sqrt{N}$, using the norm inequality (1.6) it is easy to see that almost surely for $f \in \mathbb{H}_1$,

$$\left|\langle f, \widehat{\nu}_0^{(N)} \rangle\right| \leq 2\sqrt{N} \|f\|_\infty \leq 4\sqrt{N} \|f\|_{\mathbb{H}_1}.$$

Moreover, if Assumption 4 holds, then calculations similar to those in Remark 5.2, the norm inequality (1.6) and the Cauchy–Schwarz inequality show that for $f \in \mathbb{H}_1$ and $0 \leq s < t < \infty$,

$$\begin{aligned} \left|\mathcal{J}_t^{\widehat{\nu}_0^{(N)}}(f) - \mathcal{J}_s^{\widehat{\nu}_0^{(N)}}(f)\right| &\leq C_G (t-s)^{\gamma_G} 2\sqrt{N} \|f\|_\infty + \|f\|_{\mathbb{H}_0} (t-s)^{1/2} \\ &\leq \left(4 C_G \sqrt{N} + 1\right) \|f\|_{\mathbb{H}_1} |t-s|^{\gamma_G \wedge 1/2}. \end{aligned}$$

This shows that for every $N \in \mathbb{N}$, $\mathcal{J}^{\widehat{\nu}_0^{(N)}}$ is a continuous (and so, in particular, 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 9.3, 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 $\widehat{\nu}_0^{(N)} = 0$ or, equivalently, $\nu_0^{(N)} = N\overline{\nu}_0$ for every N. The reader may prefer to make the latter assumption on first reading to avoid the technicalities in the statement of Assumption 5 below. To motivate the form of this assumption, 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 $N \to \infty$, 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 different space. As observed in Remark 5.3 above, $\widehat{\nu}_0^{(N)}$ can be viewed as an $\mathbb{H}_{-1}$-valued and, hence $\mathbb{H}_{-2}$-valued, random element, and under Assumption 4, $\{\mathcal{J}_t^{\widehat{\nu}_0^{(N)}}, t \geq 0\}$ is a càdlàg $\mathbb{H}_{-1}$-valued stochastic process and $\{\mathcal{J}_t^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), t \geq 0\}$ is a càdlàg real-valued process.

ASSUMPTION 5. There exists an $\mathbb{R}$-valued random variable $\widehat{x}_0$ and a family of random variables $\{\widehat{\nu}_0(f), f \in \mathbb{AC}_b[0, L)\}$, 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) there exist random variables

$$\begin{aligned} \mathcal{J}_t^{\widehat{\nu}_0}(f) &\doteq \widehat{\nu}_0(\Phi_t f) \\ &= \widehat{\nu}_0\left(f(\cdot + t)\frac{1 - G(\cdot + t)}{1 - G(\cdot)}\right), \qquad t \geq 0, f \in \mathbb{AC}_b[0, L), \end{aligned} \tag{5.11}$$

such that $(\mathcal{J}_t^{\widehat{\nu}_0} = \mathcal{J}^{\widehat{\nu}_0}(f), f \in \mathbb{H}_2), t \geq 0$, admits a version as a continuous $\mathbb{H}_{-2}$-valued process, $\{\mathcal{J}_t^{\widehat{\nu}_0}(\mathbf{1}), t \geq 0\}$ admits a version as a continuous $\mathbb{R}$-valued process and, for every $f \in \mathbb{AC}_b[0, L)$ almost surely, $t \mapsto \mathcal{J}_t^{\widehat{\nu}_0}(f)$ is a measurable function on $[0, \infty)$;

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

Some results will require the strengthening of Assumption 5 stated below.

ASSUMPTION 5′. The following property holds in addition to Assumption 5:

(d) Suppose that $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$ is such that for every $r > 0$, $x \mapsto \varphi(x, r)$ is absolutely continuous, 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 $t \geq 0$,

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

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 $\widehat{\nu}_0(f)$, $f \in \mathbb{AC}_b[0, L)$ and the random variable $\widehat{x}_0$ of Assumption 5 such that $\widehat{\mathcal{M}}$, B and $(\widehat{x}_0, \widehat{\nu}_0(f), f \in \mathbb{AC}_b[0, L))$ are mutually independent. Let $\widehat{\mathcal{F}}_0$ be the σ-algebra generated by $(\widehat{x}_0, \widehat{\nu}_0(f), f \in \mathbb{AC}_b[0, L))$ 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 $t \geq 0$, $\mathcal{J}_t^{\widehat{\nu}_0}(f)$, $f \in \mathbb{AC}_b[0, L)$, and $\mathcal{J}_t^{\widehat{\nu}_0}(\mathbf{1})$ are all well-defined $\widehat{\mathcal{F}}_0$-measurable random variables. In addition, $(\widehat{E}_t, \widehat{\mathcal{M}}_t)_{t \geq 0}$ are $\{\widehat{\mathcal{F}}_t\}$-adapted stochastic process. The description of the N-server model listed prior to Remark 2.1 assumes that for each $N \in \mathbb{N}$, given $\sigma(R_E^{(N)}(0))$, $\{E^{(N)}(t), t > 0\}$ is independent of the initial conditions

$\nu_0^{(N)}$ and $X^{(N)}(0)$. Together with Assumptions 3, 5 and the fact that $R_E^{(N)}(0) \to 0$ almost surely as $N \to \infty$, this implies that as $N \to \infty$,

$$(5.13)\quad \big(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})\big) \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}))$$

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

5.2. *The centered many-server map.* The centered many-server map defined below will be used to characterize the limit process. Let $\mathbb{D}^0_{\mathbb{R}}[0,\infty)$ be the subset of functions f in $\mathbb{D}_{\mathbb{R}}[0,\infty)$ with $f(0) = 0$. The input data for this map lies in the following space:

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

DEFINITION 5.4 (Centered many-server equations). Let $\overline{X} \in \mathbb{D}_{[0,\infty)}[0,\infty)$ be fixed. Given $(E, x_0, Z) \in \widehat{\mathcal{I}}_0$, $(K, X, \upsilon) \in \mathbb{D}^0_{\mathbb{R}}[0,\infty) \times \mathbb{D}_{\mathbb{R}}[0,\infty)^2$ is said to solve the centered many-server equations (CMSE) associated with $\overline{X}$ and (E, x_0, Z) if for $t \in [0,\infty)$,

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

$$(5.15)\qquad K(t) = E(t) + x_0 - X(t) + \upsilon(t) - \upsilon(0)$$

and

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

Note that this definition automatically requires that $E(0) = K(0) = 0$, $X(0) = x_0$, $Z(0) = \upsilon(0)$, and $\upsilon(0)$ is equal to x_0, $x_0 \wedge 0$ or 0, respectively, 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, let Λ denote the corresponding "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, υ) of the CMSE. The collection of input data in $\widehat{\mathcal{I}}_0$ for which a solution to the CMSE exists is defined to be the domain of Λ and is denoted dom(Λ).

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

$$(5.17)\qquad \begin{aligned} X(t) &= x_0 - \upsilon(0) + E(t) + Z(t) \\ &\quad - \int_0^t g(t-s)[E(s) + x_0 - \upsilon(0) - X(s) + \upsilon(s)]\,ds. \end{aligned}$$

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 υ and, in turn, (5.15) implies the continuity of K.

The importance of the CMSE stems from the relation

$$
\begin{aligned}
&\big(\widehat{K}^{(N)}, \widehat{X}^{(N)}, \langle \mathbf{1}, \widehat{\nu}^{(N)}\rangle\big) \\
(5.18)\qquad &\quad = \Lambda\big(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1})\big), \qquad N \in \mathbb{N},
\end{aligned}
$$

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 2 below, identifies the limit of the sequence $\{\widehat{X}^{(N)}\}_{N\in\mathbb{N}}$. Let

$$
(5.19)\qquad \widehat{Y}_1^{(N)} \doteq \big(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \widehat{\mathcal{M}}^{(N)}, \widehat{\mathcal{H}}^{(N)}, \widehat{\mathcal{H}}^{(N)}(\mathbf{1})\big),
$$

and let $\widehat{Y}_1$ be the corresponding quantity without the superscript N, where $\widehat{E}$, $\widehat{x}_0$, $\widehat{\nu}_0$, $\mathcal{J}^{\widehat{\nu}_0}$ and $\mathcal{J}^{\widehat{\nu}_0}(\mathbf{1})$ 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, let $\mathcal{Y}_1$ be the space given by

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

THEOREM 2. *Suppose Assumptions 1–5 are satisfied and suppose that the fluid limit is either subcritical, critical or supercritical. Then* $(\widehat{E}, \widehat{x}_0, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1})) \in \operatorname{dom}(\Lambda)$, *and as* $N \to \infty$,

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

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

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

where

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

The proof of Theorem 2 is presented in Section 9.1. In addition to establishing the relation (5.18), the key elements of the proof involve showing the convergence $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$ in $\mathcal{Y}_1$, which is carried out in Corollary 8.7, and establishing continuity of the centered many-server map Λ and another auxiliary map Γ, which are established in Proposition 7.3 and Lemma 7.1, respectively.

We now state a more general convergence result for the pair $\{(\widehat{X}^{(N)}, \widehat{\nu}^{(N)})\}_{N\in\mathbb{N}}$, whose proof is also given in Section 9.1. With $\widehat{K}$ equal to the limit process obtained in Theorem 2, for all $f \in \mathbb{AC}_b[0,L)$, let

$$
\begin{aligned}
\widehat{\nu}_t(f) \doteq{}& \mathcal{J}_t^{\widehat{\nu}_0}(f) + f(0)\widehat{K}(t) + \int_0^t \widehat{K}(s) f'(t-s)\big(1 - G(t-s)\big)\,ds \\
(5.24)\qquad &- \int_0^t \widehat{K}(s) g(t-s) f(t-s)\,ds - \widehat{\mathcal{H}}_t(f).
\end{aligned}
$$

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 $\Psi_t f \in \mathbb{C}_b([0, L) \times [0, \infty))$, where Ψ_t is the operator defined in (4.19).

THEOREM 3. *Suppose Assumptions 1–5 are satisfied, the fluid limit is either subcritical, critical or supercritical and g is continuous. Then, as $N \to \infty$,*

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

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

A main goal of this paper is to show that the limit process $(\widehat{X}, \widehat{\nu})$ is a tractable process that is amenable to analysis. The next two theorems show that this is indeed the case under some additional regularity conditions on the hazard rate h. First, Theorem 4 shows that $\{\widehat{X}_t, \widehat{\mathcal{F}}_t, t \geq 0\}$ is a semimartingale. By Itô's formula this enables the description of the evolution of suitably regular functionals of the process. The proof of Theorem 4 is given in Section 9.2.

THEOREM 4. *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 3, 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*

$$M^X(t) = \int_0^t \sigma(s)\, dB(s) - \widehat{\mathcal{M}}_t(\mathbf{1}),$$

$$C^X(t) = -\int_0^t \beta(s)\, ds - \int_0^t \widehat{\nu}_s(h)\, ds, \qquad t \geq 0,$$

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

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

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

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

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

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

and

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

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

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

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

REMARK 5.6. If for $f \in \mathbb{C}_b[0,L)$, $\{\mathcal{J}_t^{\widehat{\nu}_0}(f), t \ge 0\}$ is a well-defined stochastic process, then $\{\widehat{\nu}_t(f), t \ge 0\}$ 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 4, specifically of Assumption 5(c), the convergence in Theorem 4 can in fact be established for a slightly larger class of functions than those in $\mathbb{H}_2$. More precisely, if for any bounded and Hölder continuous function f, $\{\mathcal{J}_t^{\widehat{\nu}_0}(f), t \ge 0\}$ defined in (5.11) is a well-defined continuous stochastic process and, as $N \to \infty$, $(X^{(N)}(0), \widehat{\nu}_0^{(N)}(f), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(f), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \Rightarrow (\widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}(f), \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}))$ in $\mathbb{R}^2 \times \mathbb{D}_{\mathbb{R}}[0,\infty)^2$, then the process $\widehat{\nu}(f)$ defined by (5.26) is also continuous and, as $N \to \infty$, $\widehat{\nu}^{(N)}(f) \Rightarrow \widehat{\nu}(f)$ in $\mathbb{D}_{\mathbb{R}}[0,\infty)$. In fact, due to the independence assumptions of the model, the following joint convergence:

$$\begin{aligned} &\big(\widehat{E}^{(N)}, X^{(N)}(0), \widehat{\nu}_0^{(N)}(f), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(f), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})\big) \\ &\qquad \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0(f), \mathcal{J}^{\widehat{\nu}_0}(f), \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1})) \end{aligned} \tag{5.27}$$

in $\mathbb{D}_{\mathbb{R}}[0,\infty) \times \mathbb{R}^2 \times \mathbb{D}_{\mathbb{R}}[0,\infty)^2$ also holds. A brief justification of this assertion is provided at the end of Section 9.2.

We now show that the limit process $(\widehat{K}, \widehat{X}, \widehat{\nu})$ described in Theorem 2 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 4.

DEFINITION 5.7 (Stochastic age equation). Given $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ defined on the filtered probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \{\widehat{\mathcal{F}}_t\}, \widehat{\mathbb{P}})$, $\nu = \{\nu_t, t \ge 0\}$ is said to be a strong solution to the *stochastic age equation* associated with $(\widehat{\nu}_0, \widehat{K}, \widehat{\mathcal{M}})$ if for every $f \in \mathbb{AC}_b[0, L)$, $\nu_t(f)$ 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 $\varphi \in \mathbb{C}_b^{1,1}([0, L) \times \mathbb{R})$ such that $\varphi_t(\cdot, s) + \varphi_x(\cdot, s)$ is Lipschitz continuous for every s, and for every $t \in [0, \infty)$,

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

THEOREM 5. *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 $(\widehat{K}, \widehat{X}, \widehat{\nu})$ of Theorem 3, the following assertions are true*:

(1) *If the density g' of g lies in $\mathbb{L}^2_{\text{loc}}[0, L) \cup \mathbb{L}^\infty_{\text{loc}}[0, L)$, 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 $\{(\widehat{X}_t, \widehat{\nu}_t, \mathcal{J}^{\widehat{\nu}_t}(\mathbf{1})), \widehat{\mathcal{F}}_t, t \geq 0\}$ is an $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0, \infty)$-valued strong Markov process.*

Note that if h is absolutely continuous and bounded, this immediately implies that g is also absolutely continuous and bounded, and therefore the density g' of g exists. The characterization of $(\widehat{X}, \widehat{\nu})$ in terms of the stochastic age equation is established in Section 9.4 and the proof of the strong Markov property is given in Section 9.5. It is more natural to expect the process $\{(\widehat{X}, \widehat{\nu}_t), \widehat{\mathcal{F}}_t, t \geq 0\}$ 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.

REMARK 5.8. Elementary calculations show that the boundedness assumptions imposed on $g/(1-G)$ and $g'/(1-G)$ in Theorem 5 (which, in particular, imply Assumption 4) are satisfied by many continuous service distributions (with finite mean, normalized to have mean one) of interest. These include the families of lognormal, Pareto, 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 inverted Beta(a, b) distribution when $a > 2$ and $b = a + 1$. Note that the mean one exponential distribution is also included as a special case of the Gamma distribution.

5.4. *Corollaries of the main results*. From Theorem 4, it follows that when the fluid limit is critical, the limiting (scaled and centered) total number in system $\widehat{X}$ can be characterized as an Itô diffusion. Recall from Remark 3.2 that $\overline{\nu}_*$ is the probability measure on $[0, L)$ given by $\overline{\nu}_*(dx) = (1 - G(x))\, dx$.

COROLLARY 5.9. *Suppose Assumptions 1, 3(a) with $\overline{\lambda} = 1$ and Assumption 5 are satisfied and h is bounded and absolutely continuous. If $(\overline{x}_0, \overline{\nu}_0) = (1, \overline{\nu}_*)$, then $\widehat{X}$ satisfies the following Itô diffusion equation*:

$$\widehat{X}(t) = \widehat{x}_0 + \sigma B(t) - \tilde{B}(t) - \beta t - \int_0^t \widehat{\nu}_s(h)\,ds, \tag{5.29}$$

where $\tilde{B}$ is a Brownian motion independent of B.

PROOF. When the fluid initial condition is given by $(\overline{x}_0, \overline{\nu}_0) = (1, \overline{\nu}_*)$, the fluid limit is critical by Remark 3.2, and $\widehat{\mathcal{M}}(\mathbf{1})$ is a continuous martingale with quadratic variation t and hence a Brownian motion. Thus, using $\tilde{B}$ to denote $\widehat{\mathcal{M}}(\mathbf{1})$, Corollary 5.9 can be deduced from Remark 5.1 and the decomposition for $\widehat{X}$ given in Theorem 4. $\square$

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 [14]. In what follows, recall that $x^- = -(x \wedge 0) = -\min(x, 0)$.

COROLLARY 5.10. *Suppose $G(x) = 1 - e^{-x}$ for $x \in [0, \infty)$, and suppose Assumption 1 holds with $(\overline{x}_0, \overline{\nu}_0) = (1, \nu_*)$, Assumption 3(a) holds with $\overline{\lambda} = 1$ and Assumption 5′ is satisfied. Then $\widehat{X}$ is the unique strong solution to the stochastic differential equation*

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

where W is a standard Brownian motion.

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

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

where the last equality follows from the relations $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1})) \doteq \Lambda(\widehat{E}, \widehat{x}_0, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}))$ and (5.16) because $\overline{X} \equiv \mathbf{1}$. By the independence of B and $\tilde{B} = \widehat{\mathcal{M}}(\mathbf{1})$, $\sigma B - \tilde{B}$ has the same distribution as $\sqrt{1+\sigma^2}\,W$, where W is a standard Brownian motion. On substituting this back into (5.29), equation (5.30) is obtained. The Lipschitz continuity and local boundedness of the drift coefficient $x \mapsto -\beta + x^-$ guarantees that the stochastic differential equation (5.30) has a unique strong solution. $\square$

REMARK 5.11 (Insensitivity result). As a comparison of (5.29) and (5.30) reveals, under the same assumptions on the arrival process as in Corollary 5.10, 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. A succinct characterization of the dynamics of the centered state process is presented in Section 6.1. This is then used 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 in [21].

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

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

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

and

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

where $K^{(N)}$ is nondecreasing and

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

PROOF. This is essentially a direct consequence of Theorem 5.1 of [21]. Indeed, by subtracting and adding $A_\varphi^{(N)}$ on the right-hand side of equations (5.4) and (5.5) in [21], and then using (4.3) above and the fact that $M_\varphi^{(N)}$ is indistinguishable from $\mathcal{M}^{(N)}(\varphi)$ (see Remark 4.4), one obtains (6.1) and (6.2), respectively. Equation (6.3) coincides with equation (5.6) in [21]. Finally, the first equality in (6.4) follows from (2.6) of [21] and (4.3) above, whereas the second equality in (6.4) follows from (6.2). $\square$

The characterizations of the N-server system and the fluid limit given in Proposition 6.1 and Theorem 1, respectively, when combined, immediately yield a useful representation for the centered diffusion-scaled state dynamics. In what follows, recall the centered, diffusion-scaled quantities defined in (4.11), (4.14) and (5.1).

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 $\varphi \in \mathbb{C}_b^{1,1}([0, L) \times [0, \infty))$ and $t \in [0, \infty)$,*

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

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

and

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

where $\widehat{K}^{(N)}$ is nondecreasing and satisfies

$$
\begin{aligned}
\widehat{K}^{(N)}(t) &= \langle \mathbf{1}, \widehat{\nu}_t^{(N)} \rangle - \langle \mathbf{1}, \widehat{\nu}_0^{(N)} \rangle + \int_0^t \langle h, \nu_s^{(N)} \rangle \, ds + \widehat{\mathcal{M}}_t^{(N)}(\mathbf{1}) \\
&= \widehat{X}^{(N)}(0) + \widehat{E}^{(N)}(t) - \widehat{X}^{(N)}(t) + \langle \mathbf{1}, \widehat{\nu}_t^{(N)} \rangle - \langle \mathbf{1}, \widehat{\nu}_0^{(N)} \rangle.
\end{aligned} \tag{6.8}
$$

PROOF. Equation (6.5) is obtained by dividing each side of (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 (6.2) and (3.7), and equation (6.8) can be obtained from (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

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

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

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

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

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

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. Under suitable conditions, for large N, the nonidling 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 Ω^* be the set of full $\mathbb{P}$-measure on which the fluid limit theorem (Theorem 1) holds. Fix $\omega \in \Omega^*$ (and henceforth suppress the dependence on ω), and let $t \in [0, \infty)$ be a continuity point of the fluid limit. If $\overline{X}(t) < 1$, then by Theorem 1 there exists $N_0 = N_0(\omega, t) < \infty$ such that for all $N \geq N_0$, $\overline{X}^{(N)}(t) < 1$ and so

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

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

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

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

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

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 3.2. 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 $N \geq N_0(\omega, T)$,

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

An analogous statement holds for the supercritical case. Finally, in the critical case when $\overline{X}(t) = 1$ for $t \in [0, \infty)$, it trivially follows that almost surely, (6.10) holds for all $N \in \mathbb{N}$ and $t \in [0, \infty)$.

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 [21]. Representations for solutions to the abstract age equation were obtained in Proposition 4.16 of [21]. 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 convolution integral term.

We now state the representation result, which is easily deduced from Proposition 4.16 of [21]; the details of the proof are deferred to Appendix C. For conciseness of notation, for $N \in \mathbb{N}$ and continuous f, we define

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

and let $\mathcal{K}^{(N)}$ be defined analogously, but with $\widehat{K}^{(N)}$ replaced by $K^{(N)}$. By applying integration by parts to the right-hand side of (6.11) and using the fact that $\widehat{K}^{(N)}$ has at most a countable number of discontinuities, we see that for $f \in \mathbb{AC}[0,L)$ and $t \in [0,\infty)$,

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

where

$$\xi_f \doteq \big(f(1-G)\big)' = f'(1-G) - fg. \tag{6.13}$$

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

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

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

and, likewise,

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

REMARK 6.5. For subsequent use, we make the simple observation that on substituting $\varphi = \mathbf{1}$ in (6.5) and substracting 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 $t > 0$,

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

7. Continuity properties. Continuity of the mapping that takes $\widehat{K}^{(N)}$ to $\widehat{\mathcal{K}}^{(N)}$ is established in Section 7.1, and continuity of the centered many-server map Λ is established in Section 7.2. Together, these results show that both $\widehat{K}^{(N)}$ and $\widehat{X}^{(N)}$ can be obtained as continuous mappings of the initial data and $\widehat{\mathcal{H}}^{(N)}$.

7.1. *Continuity of an auxiliary map*. Given any (deterministic) càdlàg function K, we define

$$\mathcal{K}_t(f) \doteq f(0)K(t) + \int_0^t K(s)\xi_f(t-s)\,ds, \qquad t \in [0,\infty),\ f \in \mathbb{AC}[0,L), \tag{7.1}$$

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 $\mathbb{AC}[0,L)$. Moreover, from elementary properties of convolutions, it is clear that for any $f \in \mathbb{AC}[0,L)$, if K is càdlàg (resp., continuous) then so is $\mathcal{K}(f)$. Lemma 7.1 below shows that the mapping Γ that takes K to $\mathcal{K}$ maps $\mathcal{D}_{\mathbb{R}}[0,\infty)$ to $\mathcal{D}_{\mathbb{H}_{-2}}[0,\infty)$ and is continuous. Note that by (6.12), $\widehat{\mathcal{K}}^{(N)} = \Gamma(\widehat{K}^{(N)})$ for $N \in \mathbb{N}$. The continuity of Γ is used in the proof of Theorem 3 to establish convergence of $\widehat{\mathcal{K}}^{(N)}$ to the analogous limit quantity $\widehat{\mathcal{K}}$, defined by

$$\begin{aligned} \widehat{\mathcal{K}}_t(f) &\doteq \Gamma(\widehat{K}) \\ &= f(0)\widehat{K}(t) + \int_0^t \widehat{K}(s)\xi_f(t-s)\,ds, \qquad t \in [0,\infty),\ f \in \mathbb{AC}[0,L). \end{aligned} \tag{7.2}$$

The third property in Lemma 7.1 below is used in the proof of the strong Markov property in Section 9.5.

LEMMA 7.1. *Let Γ be the map that takes $K \in \mathcal{D}_{\mathbb{R}}[0,\infty)$ to $\mathcal{K} = \{\mathcal{K}_t, t \geq 0\}$, the family of linear functionals on $\mathbb{AC}[0,L)$ defined in* (7.1). *If g is continuous, the following three properties are satisfied*:

(1) *If $K \in \mathbb{D}_{\mathbb{R}}[0,\infty)$, then $\mathcal{K} \in \mathbb{D}_{\mathbb{H}_{-2}}[0,\infty)$. Likewise, if $K \in \mathbb{C}_{\mathbb{R}}[0,\infty)$, then $\mathcal{K} \in \mathbb{C}_{\mathbb{H}_{-2}}[0,\infty)$.*

(2) *Γ is a continuous map from $\mathbb{D}_{\mathbb{R}}[0,\infty)$ to $\mathbb{D}_{\mathbb{H}_{-2}}[0,\infty)$, when the 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 $\mathbb{D}_{\mathbb{R}}[0,\infty)$ to itself that takes $K \mapsto \mathcal{K}(\mathbf{1})$ is also continuous with respect to the Skorokhod topology on $\mathbb{D}_{\mathbb{R}}[0,\infty)$.*

(3) *If $K \in \mathbb{C}_{\mathbb{R}}[0,\infty)$, 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 $\mathbb{C}_{\mathbb{R}}[0,\infty)$ to itself that takes K to this function is continuous* (*with respect to the topology of uniform convergence on compact sets*).

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

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

where

$$\Delta_1^{(n)}(t) \doteq \int_0^{t\wedge\tau^{(n)}(t)} K(u)\big(\xi_f(t-u)-\xi_f\big(\tau^{(n)}(t)-u\big)\big)\,du,$$

$$\Delta_2^{(n)}(t) \doteq \int_0^{t\wedge\tau^{(n)}(t)} \big(K(u)-\tilde{K}^{(n)}(u)\big)\xi_f\big(\tau^{(n)}(t)-u\big)\,du,$$

$$\Delta_3^{(n)}(t) \doteq \int_{t\wedge\tau^{(n)}(t)}^{t} K(u)\xi_f(t-u)\,du + \int_{t\wedge\tau^{(n)}(t)}^{\tau^{(n)}(t)} \tilde{K}^{(n)}(u)\xi_f\big(\tau^{(n)}(t)-u\big)\,du.$$

To bound the above terms, let w_G be the modulus of continuity of G as defined in (1.1). Then, by applying the inequality $(1-G)\le 1$, the Cauchy–Schwarz inequality and Tonelli's theorem, it follows that for $0\le s\le t\le T$,

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

Similarly, using (1.5), we have

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

where w_g is the modulus of continuity of g. Recalling that $\xi_f=(f(1-G))'$, the last two inequalities and the norm inequalities (1.5) show that

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

where $c_1(T,\delta)\doteq (T^{1/2}(w_G(\delta)+w_g(\delta))+(T+1)\delta^{1/2})$ satisfies $\lim_{\delta\to 0}c_1(T,\delta)=0$. On the other hand, another application of the Cauchy–Schwarz inequality and the norm inequality (1.6) shows that

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

and

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

for an appropriate finite constant $c_2(T)<\infty$ that depends only on G and T. Substituting (7.4)–(7.6) into (7.3), for $f\in\mathbb{H}_2$, $f\neq 0$, we obtain

$$
\begin{aligned}
\frac{|\mathcal{K}_t(f)-\tilde{\mathcal{K}}^{(n)}_{\tau^{(n)}(t)}(f)|}{\|f\|_{\mathbb{H}_2}} &\le \big|K(t)-\tilde{K}^{(n)}\big(\tau^{(n)}(t)\big)\big| + c_1\big(T,\delta^{(n)}(t)\big)\|K\|_T\\
&\quad + c_2(T)\int_0^{t\wedge\tau^{(n)}(t)}\big|\tilde{K}^{(n)}(u)-K(u)\big|\,du\\
&\quad + 2\delta^{(n)}(t)c_2(T)\big(\|K\|_T\vee\|\tilde{K}^{(n)}\|_T\big).
\end{aligned}
\tag{7.7}
$$

Now, suppose $\tilde{K}^{(n)}=K$ so that $\tilde{\mathcal{K}}^{(n)}=\mathcal{K}$, $n\in\mathbb{N}$, 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 $n\to\infty$. Then the third term on the right-hand side of (7.7) vanishes, the first term converges to zero because $K\in\mathbb{D}_{\mathbb{R}}[0,\infty)$ and the second and fourth terms converge to zero because $\|K\|_T<\infty$ and $\delta^{(n)}(t)\to 0$. This shows that $\|\mathcal{K}_t-\mathcal{K}_{\tau^{(n)}(t)}\|_{\mathbb{H}_{-2}}\to 0$ and hence, $\mathcal{K}\in\mathbb{D}_{\mathbb{H}_{-2}}[0,\infty)$. The same argument also shows that $\mathcal{K}$ is continuous if K is. This proves the first property.

Next, suppose that $\tilde{K}^{(n)}$, $n\in\mathbb{N}$, is a sequence that converges to K in the Skorokhod topology. By the definition of the Skorokhod topology (see, e.g., Chapter 3 of [3]) there exists a sequence of strictly increasing maps $\tau^{(n)}$, $n\in\mathbb{N}$, that map $[0,T]$ onto itself and satisfy $\|\delta^{(n)}\|_T \doteq \sup_{t\in[0,T]}|t-\tau^{(n)}(t)|\to 0$ and $\|\tilde{K}^{(n)}\circ\tau^{(n)}-K\|_T\to 0$ as $n\to\infty$. Moreover, $\sup_n\|\tilde{K}^{(n)}\|_T<\infty$ and $\tilde{K}^{(n)}(u)\to K(u)$ for almost every $u\in[0,T]$. Taking first the supremum over $t\in[0,T]$ and then limits as $n\to\infty$ in (7.7), the above properties show that the right-hand side of (7.7) goes to zero (where the dominated convergence theorem is used to argue that the third term vanishes). In turn, this implies that $\sup_{t\in[0,T]}\|\mathcal{K}_t-\tilde{\mathcal{K}}^{(n)}_{\tau^{(n)}(t)}\|_{\mathbb{H}_{-2}}\to 0$, thus establishing convergence of $\tilde{\mathcal{K}}^{(n)}$ to $\mathcal{K}$ in the Skorokhod topology on $\mathbb{D}_{\mathbb{H}_{-2}}[0,\infty)$. This establishes continuity of the map Γ 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 $K \in \mathbb{C}_{\mathbb{R}}[0,\infty)$ 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

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

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 $\mathbb{C}_{\mathbb{R}}[0,\infty)$. On the other hand, given $K^{(i)} \in \mathbb{C}_{\mathbb{R}}[0,\infty)$ for $i = 1, 2$ and the corresponding $\mathcal{K}^{(i)}$,

$$\sup_{u \in [0,T]} \big|\mathcal{K}_t^{(1)}(\Phi_u \mathbf{1}) - \mathcal{K}_t^{(2)}(\Phi_u \mathbf{1})\big| \leq 2 \|K^{(1)} - K^{(2)}\|_T,$$

from which it is clear that the map from $\mathbb{C}_{\mathbb{R}}[0,\infty)$ 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*. First, in Lemma 7.2, the representation (5.18) for $(\widehat{X}^{(N)}, \widehat{K}^{(N)}, \widehat{\nu}^{(N)}(\mathbf{1}))$ in terms of the centered many-server map Λ introduced in Definition 5.4 is established. Then, in Proposition 7.3 and Lemma 7.4, certain continuity and measurability properties of Λ are established.

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

$$(\widehat{K}^{(N)}, \widehat{X}^{(N)}, \langle \mathbf{1}, \widehat{\nu}^{(N)} \rangle)(\omega) = \Lambda\big(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1})\big)(\omega) \tag{7.8}$$

on the interval $[0, T]$. Moreover, when the fluid limit is critical, (7.8) holds almost surely for every $N \in \mathbb{N}$ and $t \in [0,\infty)$.

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 $\mathbb{D}_{\mathcal{M}_F[0,L)}[0,\infty)$ and hence, in $\mathbb{D}_{\mathbb{H}_{-2}}[0,\infty)$. Assumption 4 and Remark 5.3 show that $\mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})$ is continuous. Moreover, since $\widehat{\mathcal{K}}^{(N)}$ and $\langle \mathbf{1}, \widehat{\nu}^{(N)} \rangle$ are càdlàg, from the representation (6.15), it follows that $\widehat{\mathcal{H}}^{(N)}(\mathbf{1})$ is also càdlàg. Thus, for almost surely every $\omega \in \Omega$, $(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1}))(\omega) \in \widehat{\mathcal{I}}_0$. Representation (7.8) 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 equations (6.12) and (6.15) with $f = \mathbf{1}$. $\square$

We now establish continuity and measurability properties of the mapping Λ.

PROPOSITION 7.3. *Fix* $\overline{X} \in \mathbb{D}_{[0,\infty)}[0,\infty)$. *For* $i = 1,2$, *suppose* $(E^i, x_0^i, Z^i) \in \widehat{\mathcal{I}}_0$ *and* $(K^i, X^i, v^i) \in \Lambda(E^i, x_0^i, Z^i)$. *Let* $\nabla S \doteq S^2 - S^1$ *for* $S = K, X, v, E$, x_0 *and* Z, *and recall* $\|f\|_T \doteq \sup_{s\in[0,T]}|f(s)|$. *Then for any* $T \in [0,\infty)$,

$$\|\nabla K\|_T \vee \|\nabla X\|_T \vee \|\nabla v\|_T \le 3\big(1 + U(T)\big)\varepsilon_T, \tag{7.9}$$

where U *is the renewal function associated with the service distribution* G *and*

$$\varepsilon_T \doteq (\|\nabla E\|_T \vee |\nabla x_0| \vee \|\nabla Z\|_T). \tag{7.10}$$

Hence, Λ *is continuous with respect to the topology of uniform convergence on compact sets and is single-valued on its domain.*

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

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

We claim that in this case we always have

$$\nabla v(t) - \nabla X(t) \le 0. \tag{7.11}$$

Indeed, (5.16) shows that if $\overline{X}(t) < 1$, 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 $\overline{X}(t) = 1$ and $X^1(t) \le 0$, then $(X^1(t))^+ = 0$, and so (5.16), combined with the elementary identity $x \wedge 0 - x = -x^+$, implies

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

and so (7.11) holds. Combining (7.11) with the fact that each solution satisfies equation (5.15) and the fact that (5.16) implies $|\nabla x_0 - \nabla v(0)| \le |\nabla x_0|$, we obtain

$$\begin{aligned}\nabla K(t) &= \nabla x_0 + \nabla E(t) - \nabla X(t) + \nabla v(t) - \nabla v_0 \\ &\le |\nabla x_0| + \nabla E(t) \\ &\le 2\varepsilon_T.\end{aligned} \tag{7.12}$$

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

First, we claim that in this case,

$$\nabla v(t) = v^2(t) - v^1(t) \le 0. \tag{7.13}$$

If either $\overline{X}(t) > 1$, or the relations $\overline{X}(t) = 1$, $X^1(t) > 0$ and $X^2(t) > 0$ hold, this is trivially true since by (5.16) each term on the left-hand side of (7.13) is equal to zero. In the remaining case when $\overline{X}(t) = 1$, $X^1(t) > 0$ and $X^2(t) \le 0$, (5.16) shows that $v^1(t) = 0$ and $v^2(t) = X^2(t) \le 0$, and once again (7.13) follows.

Next, since each solution satisfies (5.14), we have for every $t \in [0,\infty)$,

$$\nabla K(t) = \nabla v(t) + \int_0^t g(t-s)\nabla K(s)\,ds - \nabla Z(t). \tag{7.14}$$

Now, define

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

Then, combining (7.14) with (7.13), we conclude that

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

Applying the same inequality recursively to $K(s)$, $s \in [0,t]$, on the right-hand side, we obtain

$$\begin{aligned}\nabla K(t) &\le 2\varepsilon_T + \mathbb{I}_{\mathcal{B}}(t) \int_0^t g(t-s)\Big(2\varepsilon_T + \mathbb{I}_{\mathcal{B}}(s) \int_0^s g(s-r) \nabla K(r)\, dr\Big)\, ds \\ &\le 2\varepsilon_T\big(1 + G(t)\big) + \mathbb{I}_{\mathcal{B}}(t) \int_0^t g(t-s)\Big(\mathbb{I}_{\mathcal{B}}(s) \int_0^s g(s-r) \nabla K(r)\, dr\Big)\, ds.\end{aligned}$$

Reiterating this procedure, we obtain

$$\nabla K(t) \le 2\varepsilon_T\big(1 + G(t) + G^{*,2}(t) + \cdots\big) \le 2\varepsilon_T U(T), \tag{7.15}$$

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

By symmetry, inequalities (7.12) and (7.15) obtained in cases 1 and 2, respectively, also hold with ∇K replaced by $-\nabla K$. Since $U(T) \ge 1$, we then have

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

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

$$\|\nabla K\|_T \le 2\varepsilon_T U(T). \tag{7.16}$$

On the other hand, 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 $|\nabla x_0 - \nabla v(0)| \le |\nabla x_0|$, we obtain

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

Taking the supremum over $t \in [0,T]$ and using (7.16), we then conclude that

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

Together with (7.16) and the fact that (5.16) implies $\|\nabla v(t)\|_T \le \|\nabla X\|_T$, this establishes (7.9). Since the Skorokhod topology coincides with the uniform topology on the space of continuous functions, (7.9) implies that the map Λ is continuous at points $(E, x_0, Z) \in \mathbb{C}[0,\infty) \times \mathbb{R} \times \mathbb{C}[0,\infty)$. $\square$

LEMMA 7.4. *Suppose the fluid limit is subcritical, critical or supercritical. Then the centered many-server map* $\Lambda : \mathrm{dom}(\Lambda) \subseteq \mathbb{D}_{\mathbb{R}}[0,\infty) \times \mathbb{R} \times \mathbb{D}_{\mathbb{R}}[0,\infty) \mapsto \mathbb{D}_{\mathbb{R}}[0,\infty)^3$, *where* $\mathbb{D}_{\mathbb{R}}[0,\infty)$ *is equipped with the Skorokhod topology, is measurable.*

PROOF. We first establish the measurability of the mapping that takes $(E, x_0, Z) \in \mathrm{dom}(\Lambda)$ to X, where $(K, X, \nu) \in \Lambda(E, x_0, Z)$. By Remark 5.5, if $(K, X, \nu) = \Lambda(E, x_0, Z)$, then X satisfies the integral equation

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

where $R(t) = R_1(t) \doteq Z(t) + E(t) - \int_0^t g(t-s)E(s)\,ds$ and $F = 0$ if $\overline{X}$ is subcritical, $R(t) = R_2(t) \doteq R_1(t) + (1 - G(t))x_0$ and $F(x) = x$ if $\overline{X}$ is supercritical, and $R(t) = R_3(t) \doteq R_1(t) + (1 - G(t))x_0^+$ and $F(x) = x^+$ if $\overline{X}$ is critical. Note that in all cases, F is Lipschitz. Also, for fixed t, the map $(E, Z, x_0) \mapsto R(t)$ from $\mathbb{D}_{\mathbb{R}}[0, \infty)^2 \times \mathbb{R} \mapsto \mathbb{R}$ is clearly measurable. The latter fact implies the result 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 [5]) it follows that $X(t) = \lim_{n\to\infty}(\mathcal{T}^{(n)}\mathbf{0})(t)$, where $\mathcal{T}^{(n)}$ is the n-fold composition of the operator $\mathcal{T}: \mathbb{D}_{\mathbb{R}}[0, \infty) \mapsto \mathbb{D}_{\mathbb{R}}[0, \infty)$ given by $(\mathcal{T}\xi)(t) = R(t) + \int_0^t g(t-s)F(\xi(s))\,ds$, $\xi \in \mathbb{D}_{\mathbb{R}}[0, \infty)$. Due to the fact that convergence in the Skorokhod topology implies convergence in $\mathbb{L}^1_{\mathrm{loc}}$, the map $\xi \mapsto F(\xi)$ is a continuous mapping from $\mathbb{L}^1_{\mathrm{loc}}[0, \infty)$ to itself and the Laplace convolution $\theta \mapsto \int_0^{\cdot} g(\cdot - s)\theta(s)\,ds$ is a continuous map from $\mathbb{L}^1_{\mathrm{loc}}[0, \infty)$ to $\mathbb{C}[0, \infty)$, 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 [27], and a more detailed proof of measurability in the critical case can also be found in the Appendix of [27].

Now, to complete the proof, note that by (5.16), ν equals either X, $X \wedge 0$ or 0, depending on whether the fluid $\overline{X}$ is, respectively, subcritical, critical or supercritical. The maps $f \mapsto (f, f)$, $f \mapsto (f, f \wedge 0)$ and $f \mapsto (f, 0)$ from $\mathbb{D}_{\mathbb{R}}[0, \infty)$, equipped with the Skorokhod topology, to $\mathbb{D}^2_{\mathbb{R}}[0, \infty)$ are all measurable (in fact, continuous). In addition, for each t, $K(t)$ is a linear combination of $E(t)$, x_0, $X(t)$, $\nu(t)$ and $\nu(0)$ by (5.15). Therefore, it immediately follows that the map from (E, x_0, Z) to (K, X, ν) is also measurable. □

8. Convergence results. The representation (6.15) of the N-server queue dynamics and the continuity properties established in Section 7 reduce the problem of convergence of the sequence $\{\widehat{X}^{(N)}\}_{N\in\mathbb{N}}$ to that of joint convergence of the sequence of stochastic convolution integrals $\{\widehat{\mathcal{H}}^{(N)}\}_{N\in\mathbb{N}}$ and the sequences representing the initial data. The joint convergence of the sequence $\{\widehat{\mathcal{M}}^{(N)}\}_{N\in\mathbb{N}}$ and the sequence of centered arrival processes and initial conditions is first established in Section 8.2. In particular, Proposition 8.4 shows 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), the limit of the sequence $\{\widehat{\mathcal{H}}^{(N)}\}_{N\in\mathbb{N}}$ is identified. Both limit theorems are proved using some basic estimates, which are first obtained in Section 8.1.

8.1. *Preliminary estimates*. We begin with a useful bound, whose proof is relegated to Appendix D. Recall that U is the renewal function associated with the service distribution G, and also recall the definition of $\overline{A}_{\varphi}^{(N)}$, the fluid-scaled compensator of the departure process, given in (4.2).

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

$$\mathbb{E}\big[\big(\overline{A}_{\mathbf{1}}^{(N)}(T)\big)^k\big] = \mathbb{E}\bigg[\bigg(\int_0^T \int_{[0,L)} h(x)\overline{\nu}_s^{(N)}(dx)\,ds\bigg)^k\bigg] \leq k!(U(T))^k. \tag{8.1}$$

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

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

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

$$(\overline{A}_{\mathbf{1}}(T))^k = \bigg(\int_0^T \int_{[0,L)} h(x)\overline{\nu}_s(dx)\,ds\bigg)^k \leq k!(U(T))^k. \tag{8.2}$$

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 $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$ and $T < \infty$,*

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

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

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

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

PROOF. Since $\widehat{\mathcal{M}}^{(N)}(\varphi)$ is a martingale, by the Burkholder–Davis–Gundy (BDG) inequality (see, e.g., Theorem 7.11 of Walsh [31]) it follows that for any $r > 1$, there exists a universal constant $C_r < \infty$ (independent of φ and $\widehat{\mathcal{M}}^{(N)}$) such that

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

where

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

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

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

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

$$\mathbb{E}\big[\langle\widehat{\mathcal{M}}^{(N)}(\varphi)\rangle_T^{r/2}\big] = \mathbb{E}\big[\big(\overline{A}_{\varphi^2}^{(N)}(T)\big)^{r/2}\big] \le \|\varphi\|_\infty^r \mathbb{E}\big[\big(\overline{A}_{\mathbf{1}}^{(N)}(T)\big)^{r/2}\big].$$

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

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

Combining 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 $s \ge 0$, because $\{\widehat{\mathcal{M}}_t(\varphi) - \widehat{\mathcal{M}}_s(\varphi)\}_{t\ge s}$ is a continuous martingale with quadratic variation process $\{\overline{A}_{\varphi^2}(t) - \overline{A}_{\varphi^2}(s)\}_{t\ge s}$, 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 $\widehat{\mathcal{M}}$. In what follows, we will make use of the function spaces $\mathbb{H}_n$, $\mathbb{H}_{-n}$, $\mathcal{S}$, $\mathcal{S}'$ and the norm inequalities in (1.6) introduced in Section 1.4.1.

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 $\mathbb{D}_{\mathbb{R}}[0,T]$ as $N \to \infty$ then $\widehat{\mathcal{M}}^{(N)} \Rightarrow \widehat{\mathcal{M}}$ in $\mathbb{D}_{\mathbb{H}_{-2}}[0,T]$ as $N \to \infty$.*

PROOF. Fix $N \in \mathbb{N}$. By Remark 4.4, for every $f \in \mathcal{S}$, there exists a càdlàg version of $\widehat{\mathcal{M}}^{(N)}(f)$. Moreover, for any $T < \infty$ it follows from (8.3) and (1.6) that given any $\epsilon > 0$ and $\lambda < \infty$ there exists $\delta > 0$ such that if $\|f\|_{\mathbb{H}_1} \le \delta$, then

$$\limsup_N \mathbb{P}\Big(\sup_{s\in[0,T]} \big|\widehat{\mathcal{M}}_s^{(N)}(f)\big| > \lambda\Big) \le \limsup_N \frac{\mathbb{E}[\sup_{s\in[0,T]}|\widehat{\mathcal{M}}_s^{(N)}(f)|]}{\lambda} \le \varepsilon. \tag{8.9}$$

Thus, each $\widehat{\mathcal{M}}^{(N)}$ is a 1-continuous stochastic process in the sense of Mitoma [23]. Since $\mathcal{S}$ is a nuclear Fréchet space and $\|\cdot\|_{\mathbb{H}_1} \overset{\text{HS}}{<} \|\cdot\|_{\mathbb{H}_2}$, by Theorem 4.1 of Walsh [31] and Corollary 2 of Mitoma [23] it follows that $\widehat{\mathcal{M}}^{(N)}$ is a càdlàg $\mathbb{H}_{-2}$-valued, and hence $\mathcal{S}'$-valued, process.

On the other hand, it easily follows from Lemma 8.2 that $\widehat{\mathcal{M}}(f)$ is a continuous process for every $f \in \mathcal{S}$. An analogous argument to the one above, that instead invokes Corollary 1 of Mitoma [23] 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 [31]. □

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 the sequence is asymptotically independent of the sequences of centered arrival processes and initial conditions. Recall from Assumption 5, Remark 5.1 and the discussion following Assumption 5′ that $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$ is a probability space on which is defined the initial conditions $(\widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}))$, a standard Brownian motion B and a martingale measure $\widehat{\mathcal{M}}$ that is independent of B and the initial conditions and has the covariance functional specified in (4.15). Moreover, as shown in (5.4), $\widehat{E}$ is a diffusion driven by the Brownian motion B, with diffusion coefficient σ^2 and drift coefficient $-\beta$. Let $\widehat{\mathbb{E}}$ denote expectation with respect to $\widehat{\mathbb{P}}$.

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

$$\big(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \widehat{\mathcal{M}}^{(N)}\big) \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}), \widehat{\mathcal{M}})$$

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

PROOF. We shall first prove the assertion under the supposition that Assumption 3(a) is satisfied, in which case $\overline{\lambda}, \sigma^2$ are positive constants and β is a real-valued constant. We start by using results of [26] to recast the problem in a more convenient form. Fix $N \in \mathbb{N}$ and define

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

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

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

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

$$\begin{aligned} \widehat{E}^{(N)}(t) &= \frac{E^{(N)}(t) - N\overline{\lambda} t}{\sqrt{N}} = \frac{E^{(N)}(t) - \lambda^{(N)} t}{\sqrt{N}} - \beta t \\ &= \widehat{L}^{(N)}(t) + \widehat{\gamma}^{(N)}(t) - \beta t. \end{aligned} \tag{8.10}$$

In [26] [see page 30, Lemma A.1 and (5.15) therein] it was shown that $\{\widehat{L}^{(N)}(t), \mathcal{F}_t^{(N)}, t \geq 0\}$ is a locally square integrable martingale and, as $N \to \infty$, $\sup_{t \leq T} |\widehat{\gamma}^{(N)}(t)| \to 0$ in probability, which implies $\widehat{\gamma}^{(N)} \Rightarrow \mathbf{0}$.

We will now show that for every $f \in \mathbb{C}_b[0, L)$,

$$(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)) \Rightarrow (B, \widehat{\mathcal{M}}(f)) \qquad \text{as } N \to \infty \tag{8.11}$$

and for real-valued, bounded, continuous functions F_1 on $\mathbb{D}_{\mathbb{R}^2}[0, \infty)$ and F_2 on $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{D}_{\mathbb{H}_{-2}}[0, \infty) \times \mathbb{D}_{\mathbb{R}}[0, \infty)$,

$$\begin{aligned} &\lim_{N\to\infty} \mathbb{E}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)) F_2(\widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}))] \\ &\quad = \widehat{\mathbb{E}}[F_1(B, \widehat{\mathcal{M}}(f))] \widehat{\mathbb{E}}[F_2(\widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}))] \\ &\quad = \widehat{\mathbb{E}}[F_1(B, \widehat{\mathcal{M}}(f)) F_2(\widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}))]. \end{aligned} \tag{8.12}$$

Before presenting the proofs of these results, we show how these results are sufficient to establish the proposition. We first note that (8.11), together with the convergence $\widehat{\gamma}^{(N)} \Rightarrow \mathbf{0}$, implies the joint convergence $(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f), \widehat{\gamma}^{(N)}) \Rightarrow (B, \widehat{\mathcal{M}}(f), \mathbf{0})$ for every $f \in \mathbb{C}_b[0, L)$. Given (8.10) and the relation $\widehat{E}(t) = B(t) - \beta t$, an application of the continuous mapping theorem then shows that for every $f \in \mathbb{C}_b[0, L)$, $(\widehat{E}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)) \Rightarrow (\widehat{E}, \widehat{\mathcal{M}}(f))$ in $\mathbb{D}_{\mathbb{R}}[0, \infty)^2$. Similarly, the continuous mapping theorem and (8.12) imply that for every $f \in \mathbb{C}_b[0, L)$, as $N \to \infty$,

$$\begin{aligned} &(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \widehat{\mathcal{M}}^{(N)}(f)) \\ &\quad \Rightarrow (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}), \widehat{\mathcal{M}}(f)). \end{aligned}$$

Together with Corollary 8.3 this implies the desired convergence stated in the proposition.

To complete the proof in the case when Assumption 3(a) holds, it suffices to establish (8.11) and (8.12). This is done in the following five claims; the first three claims below verify conditions of the martingale central limit theorem to establish (8.11), the fourth claim establishes a slight variant of (8.11) and the last claim proves (8.12). Let $\{[\widehat{L}^{(N)}]_t\}$ and $\{[\widehat{\mathcal{M}}^{(N)}(f)]_t\}$ represent the $\{\mathcal{F}_t^{(N)}\}$-optional quadratic variation processes associated with $\{\widehat{L}_t^{(N)}, t \geq 0\}$ and $\{\widehat{\mathcal{M}}_t^{(N)}(f), t \geq 0\}$, respectively. Also, let $f \in \mathbb{C}_b[0, L)$.

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

PROOF. By (5.10) of [26], the $\{\mathcal{F}_t^{(N)}\}$-predictable quadratic variation of $\widehat{L}^{(N)}$ at time t, is given by

$$\langle \widehat{L}^{(N)} \rangle_t = E^{(N)}(t) \mathbb{E}[(1 - \lambda^{(N)} \xi_j^{(N)})^2]/N = (E^{(N)}(t)\sigma^2)/(N\overline{\lambda}), \tag{8.13}$$

where the last equality holds due to Assumption 3(a). It is easy to check that the $\{\mathcal{F}_t^{(N)}\}$-predictable jump measure of $\widehat{L}^{(N)}$, which we denote by $\pi^{(N)}(dt, dx)$, is equal to $dE_t^{(N)}\mathbb{P}((1-\lambda^{(N)}\xi_1^{(N)})/\sqrt{N} \in dx)$. By the Lindeberg condition imposed in Assumption 3(a), the sequence of predictable jump measures $\{\pi^{(N)}\}_{N\in\mathbb{N}}$ satisfies condition $[\widehat{\delta}_5 - D]$ of Theorem VIII 3.11 on page 432 of [16]. Therefore, the $\{\mathcal{F}_t^{(N)}\}$-optional quadratic variation, $[\widehat{L}^{(N)}]_t$, converges in probability to a limit if and only if $\langle \widehat{L}^{(N)}\rangle_t$, its $\{\mathcal{F}_t^{(N)}\}$-predictable quadratic variation, converges in probability to the same limit. However, by Assumption 3(a) and Remark 5.1, as $N \to \infty$, $E^{(N)}(t)/N \to \overline{\lambda}t$. When combined with (8.13), it follows that as $N \to \infty$, $\langle \widehat{L}^{(N)}\rangle_t$ converges in probability to $\sigma^2 t$. This establishes the first limit of Claim 1.

Note that $\widehat{M}_f^{(N)}$, 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 [16]) and hence (by Theorem 4.5.2 of Chapter 1 of [16]), its $\{\mathcal{F}_t^{(N)}\}$ optional quadratic variation at time t satisfies

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

where the last equality follows because the jumps of $M_f^{(N)}$ coincide with those of $Q_f^{(N)}$. The law of large numbers results in [21] (specifically, Theorem 5.4 and the discussion below Theorem 5.15 and Proposition 5.17 therein) show that $\overline{Q}_{f^2}^{(N)}(t) \to \overline{A}_{f^2}(t)$ in probability. This proves the second limit in Claim 1. $\square$

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

PROOF. For $i \in \mathbb{N}$, let $\tau_i^{(N)} \doteq \sum_{j=1}^i \xi_j^{(N)}$ be the time of the ith jump of $E^{(N)}$. Since both $\widehat{L}^{(N)}$ and $\widehat{M}_f^{(N)}$ are compensated sums of jumps, arguing as in Claim 1 the optional quadratic co-variation is given by

$$[\widehat{L}^{(N)}, \widehat{M}_f^{(N)}]_t = \frac{1}{\sqrt{N}} \sum_{i\in\mathbb{N}:\,\tau_i^{(N)}\le t} (1-\lambda^{(N)}\xi_{i+1}^{(N)})\Delta\widehat{M}_f^{(N)}(\tau_i^{(N)}), \tag{8.14}$$

where we have also used the fact that $E^{(N)}$ has unit jumps. To prove the claim, it suffices to show that

$$\mathbb{E}[[\widehat{L}^{(N)}, \widehat{M}_f^{(N)}]_t^2] \le \frac{\sigma^2\|f\|_\infty^2}{\overline{\lambda}N}\mathbb{E}\Big[\sum_{i\in\mathbb{N}:\,\tau_i^{(N)}\le t} \Delta\overline{D}^{(N)}(\tau_i^{(N)})\Big]. \tag{8.15}$$

Indeed, then the right-hand side goes to zero as $N \to \infty$ because the expectation on the right-hand side is bounded by $\sup_N \mathbb{E}[\overline{D}^{(N)}(t)]$, which is finite by Lemma 5.6 of [21]. Alternatively, the convergence to zero of the right-hand side of

(8.15) could also be deduced from the stronger result stated in Corollary A.3. The claim would then follow by an application of Chebysev's inequality.

To establish (8.15), 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

$$\mathbb{E}[1 - \lambda^{(N)}\xi_{i+1}^{(N)} | \tilde{\mathcal{F}}_{\tau_i^{(N)}}^{(N)}] = \mathbb{E}[1 - \lambda^{(N)}\xi_{i+1}^{(N)}] = 0, \tag{8.16}$$

$$\mathbb{E}[(1 - \lambda^{(N)}\xi_{i+1}^{(N)})^2 | \tilde{\mathcal{F}}_{\tau_i^{(N)}}^{(N)}] = \mathbb{E}[(1 - \lambda^{(N)}\xi_{i+1}^{(N)})^2] = \sigma^2/\overline{\lambda}. \tag{8.17}$$

Combining (8.17) with the estimate $(\Delta\widehat{M}_f^{(N)})^2 \leq \|f\|_\infty^2 \Delta\overline{D}^{(N)}$, we obtain for $i \in \mathbb{N}$,

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

A similar conditioning argument using (8.16) shows that for $2 \leq k < i$, $i, k \in \mathbb{N}$,

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

Taking first the square and then the expectation of each side of (8.14) and using the last two relations, we obtain (8.15). As argued above, this proves the second claim. $\square$

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

PROOF. The sizes of the jumps of $\widehat{E}^{(N)}$ and $\widehat{L}^{(N)}$ converge to zero as $N \to \infty$ because $E^{(N)}$ is a counting process with unit jumps and $\sup_{t \leq T} |\widehat{\gamma}^{(N)}(t)| \to 0$ in probability. Also, by Lemma 4.2 and the continuity of $A_f^{(N)}$, the sizes of 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 [9]). The conditions of that theorem are verified by Claims 1–2 above and the first assertion of Claim 3, and (8.11) follows from the observation that B and $\widehat{\mathcal{M}}(f)$ are independent, centered Gaussian processes with variance processes $\sigma^2 t$ and $\overline{A}_{f^2}(t)$, $t \geq 0$, respectively. □

For the next two claims, recall from Section 2.2 that for each $N \in \mathbb{N}$, $\{\mathcal{G}_t^{(N)}\}$ is the augmented right-continuous filtration associated with the Markov process $(R_E^{(N)}, X^{(N)}, \nu^{(N)})$ and $\mathbb{E}_{(r,k,\mu)}^{(N)}$ is the expectation with respect to the Markovian measure with initial distribution concentrated at (r, k, μ).

CLAIM 4. *For every $N \in \mathbb{N}$, the process $\{\widehat{L}_t^{(N)}, \widehat{\mathcal{M}}_t^{(N)}(f), t \geq 0\}$ is $\{\mathcal{G}_t^{(N)}\}$-adapted. Moreover, let $\{(r^{(N)}, k^{(N)}, \mu^{(N)})\}_{N\in\mathbb{N}} \subset [0,\infty) \times \mathbb{N} \times \mathbb{M}_F[0, L)$ be a deterministic sequence that satisfies $r^{(N)} \to 0$, $k^{(N)}/N \to \overline{x}_0$, $\mu^{(N)}/N \Rightarrow \overline{\nu}_0$, where $\overline{x}_0, \overline{\nu}_0$ are the limit of the fluid-scaled initial conditions, as defined in Assumption 1. Then, given any bounded and continuous functional F on $\mathbb{D}_{\mathbb{R}^2}[0,\infty)$,*

$$\lim_{N\to\infty} \mathbb{E}_{(r^{(N)},k^{(N)},\mu^{(N)})}^{(N)}\big[F\big(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)\big)\big] = \widehat{\mathbb{E}}[F(B, \widehat{\mathcal{M}}(f))]. \tag{8.18}$$

PROOF. Fix $t \in [0,\infty)$. It is easy to see from the definitions that $E^{(N)}(t), \xi_j^{(N)}$, $j \leq E^{(N)}(t)$, and hence $\widehat{L}_t^{(N)}$, are all measurable with respect to $\sigma(R_E^{(N)}(s), s \in [0,t]) \subset \mathcal{G}_t^{(N)}$. In a similar fashion, for $f \in \mathbb{C}_b[0, L)$, it follows from the definitions given in (4.1) and (4.2) that $Q_f^{(N)}(t)$ and $A_f^{(N)}(t)$ are measurable with respect to $\sigma(\nu_s^{(N)}, s \in [0,t)) \subseteq \mathcal{G}_t^{(N)}$. Thus, from (4.3), (4.5) and (4.11) it follows that $\widehat{\mathcal{M}}_t^{(N)}(f)$ is also $\mathcal{G}_t^{(N)}$-measurable and the first assertion of the claim is proved. The limit in (8.11) was established in Claim 3. Since both $\widehat{L}^{(N)}$ and $\widehat{\mathcal{M}}^{(N)}$ are adapted with respect to $\{\mathcal{G}_t^{(N)}\}$, one can use the Markov property and repeat the argument used in the proof of that claim to establish the slightly modified version of (8.11) that is stated in (8.18). □

CLAIM 5. *The asymptotic independence property in* (8.12) *holds.*

PROOF. Let F_1 be a continuous functional on $\mathbb{D}_{\mathbb{R}^2}[0,\infty)$ and F_2 a continuous functional on $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{D}_{\mathbb{H}_{-2}}[0,\infty) \times \mathbb{D}_{\mathbb{R}}[0,\infty)$. Then

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\big[F_1\big(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)\big) F_2\big(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})\big)\big] \\
&\qquad = \lim_{N\to\infty} \mathbb{E}\big[\mathbb{E}\big[F_1\big(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)\big) \\
&\qquad\qquad \times F_2\big(\widehat{x}_0^{(N)}, \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})\big)|\mathcal{G}_0^{(N)}\big]\big].
\end{aligned} \tag{8.19}$$

Now, $\widehat{X}^{(N)}(0)$ and $\widehat{\nu}_0^{(N)}$ are measurable functions of $X^{(N)}(0)$ and $\nu_0^{(N)}$ because the fluid limit quantities $\overline{x}_0$ and $\overline{\nu}_0$ are almost surely deterministic. Moreover, $\mathcal{J}^{\widehat{\nu}_0^{(N)}}$ and $\mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})$ are measurable functions of $\widehat{\nu}_0^{(N)}$. Therefore, $F_2(\widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}))$ is $\mathcal{G}_0^{(N)}$-measurable. Since $\{(\widehat{L}^{(N)}(t), \widehat{\mathcal{M}}_t^{(N)}(f)), t \geq 0\}$ is $\{\mathcal{G}_t^{(N)}\}$-adapted by Claim 4, the Markov property shows that

$$
\begin{aligned}
&\mathbb{E}[\mathbb{E}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f))F_2(\widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}))|\mathcal{G}_0^{(N)}]] \\
&\quad = \mathbb{E}[F_2(\widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \\
&\qquad \times \mathbb{E}^{(N)}_{(R_E^{(N)}(0), X^{(N)}(0), \nu_0^{(N)})}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f))]].
\end{aligned}
\tag{8.20}
$$

Since $(R_E^{(N)}(0), \overline{X}^{(N)}(0), \overline{\nu}_0^{(N)}) \to (0, \overline{x}_0, \overline{\nu}_0)$ $\mathbb{P}$-almost surely by Assumption 1, the conditions of Claim 4 are satisfied and it follows from (8.18) that $\mathbb{P}$-almost surely,

$$
\lim_{N\to\infty} \mathbb{E}^{(N)}_{(R_E^{(N)}(0), X^{(N)}(0), \nu_0^{(N)})}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f))] = \widehat{\mathbb{E}}[F_1(B, \widehat{\mathcal{M}}(f))].
$$

Since F_1 and F_2 are bounded (say by $\overline{C}$), the bounded convergence theorem implies that

$$
\begin{aligned}
&\lim_{N\to\infty} \big|\mathbb{E}[F_2(\widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1})) \\
&\qquad \times \mathbb{E}^{(N)}_{(R_E^{(N)}(0), X^{(N)}(0), \nu_0^{(N)})}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f))]] \\
&\qquad - \mathbb{E}[F_2(\widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}))\widehat{\mathbb{E}}[F_1(B, \widehat{\mathcal{M}}(f))]]\big| \\
&\quad \leq \overline{C} \lim_{N\to\infty} \mathbb{E}\big[\big|\mathbb{E}^{(N)}_{(R_E^{(N)}(0), X^{(N)}(0), \nu_0^{(N)})}[F_1(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f))] \\
&\qquad\qquad - \widehat{\mathbb{E}}[F_1(B, \widehat{\mathcal{M}}(f))]\big|\big] \\
&\quad = 0.
\end{aligned}
\tag{8.21}
$$

On the other hand, since F_2 is bounded and continuous, by Assumption 5,

$$
\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}[F_2(\widehat{X}^{(N)}(0), \widehat{\nu}_0^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}))] \\
&\quad = \widehat{\mathbb{E}}[F_2(\widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}))].
\end{aligned}
\tag{8.22}
$$

Combining (8.19)–(8.22), we obtain the first equality in (8.12). The second equality of (8.12) follows from the first due to the fact that, by construction, $(B, \widehat{\mathcal{M}}(f))$ depend only on the fluid initial conditions $\overline{x}_0$ and $\overline{\nu}_0$, which are $\mathbb{P}$ almost surely deterministic.

We now turn to the proof of (8.12) 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

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

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

$$\widehat{E}^{(N)}(t) = \sqrt{N}\left(\overline{E}^{(N)}(t) - \int_0^t \overline{\lambda}(s)\,ds\right) = \widehat{L}^{(N)}(t) + \int_0^t \beta(s)\,ds. \tag{8.23}$$

Fix f that is bounded and continuous. To complete the proof of the proposition, it suffices to show that $(\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)) \Rightarrow (\int_0^{\cdot} (\overline{\lambda}(s))^{1/2}\,dB(s), \widehat{\mathcal{M}}(f))$. 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 $\{\widehat{\mathcal{M}}_t^{(N)}(f), \tilde{\mathcal{F}}_t^{(N)}, t \geq 0\}$ are martingales. Hence, once again, we need only verify the conditions of the martingale central limit theorem. 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 $\widehat{\mathcal{M}}^{(N)}(f)$ are uniformly bounded by $(1 + \|f\|_\infty)/\sqrt{N}$ and for each $t > 0$, $[\widehat{\mathcal{M}}^{(N)}(f)]_t \to \overline{A}_{f^2}(t)$ 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 $\int_0^t \overline{\lambda}(s)\,ds$ and $\overline{A}_{f^2}(t)$, to complete the proof it suffices to verify that for every $t \in [0, \infty)$, as $N \to \infty$, the following limits hold in probability:

$$[\widehat{L}^{(N)}]_t \to \int_0^t \overline{\lambda}(s)\,ds, \qquad [\widehat{L}^{(N)}, \widehat{\mathcal{M}}^{(N)}(f)]_t \to 0. \tag{8.24}$$

Clearly, the $\{\tilde{\mathcal{F}}_t^{(N)}\}$-predictable quadratic variation of $\widehat{L}^{(N)}$ is given by $\langle \widehat{L}^{(N)} \rangle_t = \frac{1}{N}\int_0^t \lambda^{(N)}\,ds$, which converges to $\int_0^t \overline{\lambda}(s)\,ds$ as $N \to \infty$. By Theorem 3.11 of Chapter VIII of [16], this implies that the $\{\tilde{\mathcal{F}}_t^{(N)}\}$-optional quadratic variation $[\widehat{L}^{(N)}]_t$ converges in law to $\int_0^t \overline{\lambda}(s)\,ds$. Because the limit $\int_0^t \overline{\lambda}(s)\,ds$ is deterministic, the convergence is also in probability. This establishes the first limit in (8.24). To establish the second limit, note that $\widehat{L}^{(N)}$ and $\widehat{\mathcal{M}}^{(N)}(f)$ are both compensated pure jump processes with continuous compensators, and so their optional quadratic covariation takes the form

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

Noting that $\Delta Q_f^{(N)} \leq \|f\|_\infty \Delta D^{(N)}$, taking expectations of both sides above and then the limit as $N \to \infty$, Corollary A.3 shows that

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

An application of Markov's inequality then yields the second limit in (8.24). The asymptotic independence from the initial conditions is proved exactly in the same way as when Assumption 3(a) holds (see the first part of the proof of Claim 5) and is thus omitted. □

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 or if g is in $\mathbb{L}^{1+\alpha}$ for some $\alpha > 0$; see Remark 5.2. For conciseness, we will use the notation

$$\overline{G}(x) = 1 - G(x), \qquad f\overline{G}(x) = f(x)\overline{G}(x), \qquad x \in [0,\infty). \tag{8.25}$$

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 γ_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 $0 < t < t' < \infty$,*

$$\|\Psi_t f - \Psi_{t'} f\|_\infty \leq (C_f + C_G\|f\|_\infty)|t - t'|^{\gamma_f'}. \tag{8.26}$$

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 (8.26) can be replaced by $C_0\|f\|_{\mathbb{H}_1}|t - t'|^{\gamma_f'}$.

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

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

and

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

The Hölder continuity of f and the fact that $\overline{G}$ is nonincreasing 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 (8.26). If $f \in \mathbb{H}_1$, the Cauchy–Schwarz inequality and (1.5) imply

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

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 (1.6) this shows that f is bounded and Hölder continuous and the second assertion of the lemma holds with $C_0 \doteq 1 + 2C_G$. $\square$

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

$$
\begin{aligned}
(\Upsilon_t \varphi)(x, u) &\doteq \int_u^t (\Psi_s \varphi(\cdot, s))(x, u)\, ds \\
&= \int_u^t \varphi(x + s - u, s) \frac{1 - G(x + s - u)}{1 - G(x)}\, ds
\end{aligned}
\tag{8.27}
$$

for $(x, u) \in [0, L) \times [0, t]$ and $f \in \mathbb{C}_b[0, L)$. 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.

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 $\widehat{\mathbb{P}}$-almost surely continuous.*

(2) *There exists $r \geq 2$ and a constant $C_0 < \infty$ such that for any $f \in \mathcal{S}$,*

$$
\sup_N \mathbb{E}\Big[\sup_{t \in [0,T]} \big|\widehat{\mathcal{H}}_t^{(N)}(f)\big|^r \Big] \leq C_0 \|f\|_{\mathbb{H}_1}^r.
\tag{8.28}
$$

(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 are independent of t). Then $\widehat{\mathbb{P}}$-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 $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$. Then the process $\{\widehat{\mathcal{M}}_t(\Upsilon_t \varphi), t \geq 0\}$ admits a continuous version.*

PROOF. The proof of the lemma is based on a modification of the approach used in Walsh [31] to establish convergence of stochastic convolution integrals, tailored to the present context (the proof of Theorem 7.13 in [31] 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 γ_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. 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:

$$V_t^{(N)}(f) \doteq \widehat{\mathcal{M}}_T^{(N)}(\Psi_t f), \qquad t \in [0,T]. \tag{8.29}$$

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

$$\widehat{\mathcal{H}}_t^{(N)}(f) = \widehat{\mathcal{M}}_t^{(N)}(\Psi_t f) = \mathbb{E}\big[V_t^{(N)}(f)|\mathcal{F}_t^{(N)}\big], \qquad t \in [0,T], \tag{8.30}$$

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, e.g., Theorem 7.10 of Chapter V of Rogers and Williams [29]) 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 $0 \leq t' < t < T$,

$$\mathbb{E}\big[\big|V_t^{(N)}(f) - V_{t'}^{(N)}(f)\big|^r\big] \leq \tilde{C}_f |t-t'|^{\tilde{\theta}}; \tag{8.31}$$

see, for example, Corollary 1.2 of Walsh [31]. Fix $0 \leq t' \leq t \leq T$ and note that

$$\big|V_t^{(N)}(f) - V_{t'}^{(N)}(f)\big| = \big|\widehat{\mathcal{M}}_T^{(N)}(\Psi_t f - \Psi_{t'} f)\big|. \tag{8.32}$$

Let r be any positive even integer greater than $1/\gamma'_f$. Together with (8.3) and (8.26), this implies that (8.31) is satisfied with $\tilde{\theta} = r\gamma'_f > 1$ and

$$\tilde{C}_f = C_r(C_f + C_G\|f\|_\infty)^r\big((r/2)!U^{r/2}(T) + 1\big). \tag{8.33}$$

□

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

PROOF. Analogously to (8.29) and (8.30), we define $V_t(f) \doteq \widehat{\mathcal{M}}_T(\Psi_t f)$, $t \geq 0$, and observe that

$$\widehat{\mathcal{H}}_t(f) = \widehat{\mathcal{M}}_t(\Psi_t f) = \mathbb{E}[V_t(f)|\widehat{\mathcal{F}}_t], \qquad t \geq 0. \tag{8.34}$$

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

$$\mathbb{E}[|V_t(f) - V_{t'}(f)|^r] \leq \tilde{C}_f |t-t'|^{\tilde{\theta}} \tag{8.35}$$

with $\tilde{\theta} = r\gamma'_f$. Fix $0 < t' < t < \infty$ with $|t-t'| < 1$ and a bounded, Hölder continuous f. Using (8.34) and adding and subtracting $\widehat{\mathcal{M}}_{t'}(\Psi_t f) = \mathbb{E}[V_t(f)|\widehat{\mathcal{F}}_{t'}]$, we obtain

$$\widehat{\mathcal{H}}_t(f) - \widehat{\mathcal{H}}_{t'}(f) = \mathbb{E}[V_t(f) - V_{t'}(f)|\widehat{\mathcal{F}}_{t'}] + \widehat{\mathcal{M}}_t(\Psi_t f) - \widehat{\mathcal{M}}_{t'}(\Psi_t f). \tag{8.36}$$

Consider any even integer $r > 2/\gamma'_f \vee 4$ so that (8.35) holds with $\tilde{\theta} > 2$, let $\overline{\theta} \doteq \lfloor r/2 \wedge \tilde{\theta} \rfloor$ and note that $\overline{\theta}$ is an integer greater than or equal to 2. Taking first the rth power and then expectations of both sides of (8.36), and using the inequality $(x+y)^r \le 2^r(x^r + y^r)$ and Jensen's inequality, we obtain

$$\begin{aligned}\mathbb{E}[|\widehat{\mathcal{H}}_{t'}(f) - \widehat{\mathcal{H}}_t(f)|^r] \\ &\le 2^r \big(\mathbb{E}[|V_t(f) - V_{t'}(f)|^r] + \mathbb{E}[|\widehat{\mathcal{M}}_t(\Psi_{t'} f) - \widehat{\mathcal{M}}_{t'}(\Psi_{t'} f)|^r]\big).\end{aligned}$$

Applying the estimates (8.31), (8.5) and the fact that $\|\Psi_t f\|_\infty \le \|f\|_\infty$, and then the inequality $x^2 + y^2 \le (x+y)^2$ for $x, y \ge 0$, this implies that

$$\begin{aligned}&\mathbb{E}[|\widehat{\mathcal{H}}_{t'}(f) - \widehat{\mathcal{H}}_t(f)|^r] \\ &\quad\le 2^r \tilde{C}_f |t - t'|^{\tilde{\theta}} + 2^r C_r \big(\overline{A}_{(\Psi_{t'} f)^2}(t) - \overline{A}_{(\Psi_{t'} f)^2}(t')\big)^{r/2} \\ &\quad\le 2^r (\tilde{C}_f \vee C_r \|f\|_\infty^2)\big(t + \overline{A}_{\mathbf{1}}(t) - t' - \overline{A}_{\mathbf{1}}(t')\big)^2. \end{aligned}\tag{8.37}$$

Since $t + \overline{A}_{\mathbf{1}}(t)$ is a nonnegative, increasing function of t, the generalized Kolmogorov's continuity criterion (see, e.g., Corollary 3 of [23]) implies that $\widehat{\mathcal{H}}(f)$ has a continuous version. $\square$

CLAIM 3. *The estimate* (8.28) *is satisfied.*

PROOF. From the proof of Corollary 1.2 of Walsh [31] it is straightforward to deduce that (8.31) also implies that there exists a constant $\tilde{C}_r < \infty$, which depends on r but is independent of N and f, such that

$$\mathbb{E}\Big[\sup_{s\in[0,T]} \big|V_s^{(N)}(f)\big|^r\Big] \le \tilde{C}_r \tilde{C}_f. \tag{8.38}$$

By (8.30), for every $t \in [0, T]$,

$$\big|\widehat{\mathcal{H}}_t^{(N)}(f)\big| \le \mathbb{E}\Big[\sup_{s\in[0,T]} V_s^{(N)}(f) | \mathcal{F}_t^{(N)}\Big] \le \mathbb{E}\Big[\sup_{s\in[0,T]} \big|V_s^{(N)}(f)\big| | \mathcal{F}_t^{(N)}\Big].$$

By (8.38), the last term above (viewed as a processs in t) is a Doob martingale, and hence is càdlàg. Since $r > 1$, Doob's inequality and (8.38) imply that

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

If $f \in \mathcal{S}$, then by the expression for $\tilde{C}_f$ given in (8.33) and the inequalities $\|f\|_\infty \le \sqrt{6}\|f\|_{\mathbb{H}_1}$ and $C_f \le \|f\|_{\mathbb{H}_1}$ established in (1.6) and Lemma 8.5, respectively, the right-hand side above can be replaced by $C_0\|f\|_{\mathbb{H}_1}^r$, for an appopriate constant $C_0 = C_0(G, r, T) < \infty$ that is independent of N and f. Thus, (8.28) follows. $\square$

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

PROOF. 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 $T_N + \delta_N \le T$, then for any even integer $r \ge 2$,

$$\mathbb{E}\big[\big|V^{(N)}_{T_N+\delta_N}(f) - V^{(N)}_{T_N}(f)\big|^r\big] \le \tilde{C}_f \delta_N^{r\gamma'_f}. \tag{8.39}$$

Let $\delta_N \in (0,1)$, and let T_N be an $\{\mathcal{F}_t^{(N)}\}$ stopping time such that $T_N + \delta_N \le T$. Using (8.30) and (8.29), the difference $\widehat{\mathcal{H}}^{(N)}_{T_N+\delta_N}(f) - \widehat{\mathcal{H}}^{(N)}_{T_N}(f)$ can be rewritten as

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

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

$$\begin{aligned}
&\mathbb{E}\big[\big|\widehat{\mathcal{H}}^{(N)}_{T_N+\delta_N}(f) - \widehat{\mathcal{H}}^{(N)}_{T_N}(f)\big|^2\big]\\
&\quad \le 2\mathbb{E}\big[\big|V^{(N)}_{T_N+\delta_N}(f) - V^{(N)}_{T_N}(f)\big|^2\big] + 2\|f\|_\infty^2 \mathbb{E}\big[\overline{A}_{\mathbf{1}}^{(N)}(T_N+\delta_N) - \overline{A}_{\mathbf{1}}^{(N)}(T_N)\big]\\
&\quad \le 2\tilde{C}_f\delta_N^{2\gamma'_f} + 2\|f\|_\infty^2 \sup_{\tilde{N}} \mathbb{E}\Big[\sup_{t\in[0,T]}\big(\overline{A}_{\mathbf{1}}^{(\tilde{N})}(t+\delta_N) - \overline{A}_{\mathbf{1}}^{(\tilde{N})}(t)\big)\Big],
\end{aligned}$$

where the last equality uses (8.39) with $r = 2$. As $\delta_N \to 0$, the first term on the right-hand side clearly converges to zero, whereas Lemma 5.8(2) of [21] shows that the second term also converges to zero. We conclude that $\widehat{\mathcal{H}}^{(N)}_{T_N+\delta_N}(f) - \widehat{\mathcal{H}}^{(N)}_{T_N}(f)$ converges to zero in $\mathbb{L}^2$, and hence in probability. On the other hand, (8.28) 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, e.g., Theorem 6.8 of Walsh [31]), and hence the sequence $\{\widehat{\mathcal{H}}^{(N)}(f)\}_{N\in\mathbb{N}}$ is tight. $\square$

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

$$\widehat{\mathcal{H}}_s(f^t_\varphi) - \widehat{\mathcal{H}}_{s'}(f^{t'}_\varphi) = \widehat{\mathcal{H}}_s(f^t_\varphi) - \widehat{\mathcal{H}}_{s'}(f^t_\varphi) + \widehat{\mathcal{M}}_{s'}\biggl(\Psi_s\biggl(\int_{t'}^t \varphi(\cdot, r)\, dr\biggr)\biggr).$$

Due to the assumed boundedness and Hölder continuity of f^t_φ, (8.37) 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

$$\begin{aligned}
&\mathbb{E}\biggl[\biggl|\widehat{\mathcal{H}}_{s'}\biggl(\int_0^{t'} \varphi(\cdot, r)\, dr\biggr) - \widehat{\mathcal{H}}_s\biggl(\int_0^t \varphi(\cdot, r)\, dr\biggr)\biggr|^r\biggr] \\
&\qquad = \mathbb{E}[|\widehat{\mathcal{H}}_{s'}(f^{t'}_\varphi) - \widehat{\mathcal{H}}_s(f^t_\varphi)|^r] \\
&\qquad \le C(T, r, \varphi)\bigl(|s + \overline{A}_1(s) - s' - \overline{A}_1(s')|^{\tilde{\theta}} + |t - t'|^{\tilde{\theta}}\bigr).
\end{aligned}$$

Property 3 then follows from the generalized Kolmogorov criterion for continuity of random fields; see, for example, [6].

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 $R_t(\varphi) \doteq \widehat{\mathcal{M}}_t(\Upsilon_t\varphi)$, define $\tilde{V}_t(\varphi) \doteq \widehat{\mathcal{M}}_T(\Upsilon_t\varphi)$ and note that $R_t(\varphi) = \mathbb{E}[\tilde{V}_t(\varphi)|\mathcal{F}_t]$. In a manner similar to (8.36), we can write

$$R_t(\varphi) - R_{t'}(\varphi) = \mathbb{E}\biggl[\widehat{\mathcal{M}}_T\biggl(\int_{t'}^t \Psi_s\varphi(\cdot, s)\, ds\biggr)\Bigm|\mathcal{F}_{t'}\biggr] + \widehat{\mathcal{M}}_t(\Upsilon_t\varphi) - \widehat{\mathcal{M}}_{t'}(\Upsilon_t\varphi).$$

Using Jensen's inequality, (8.4) and (8.5) with $r = 4$ and the inequalities $\|\int_{t'}^t \Psi_s\varphi(\cdot, s)\, ds\|_\infty \le |t - t'|\|\varphi\|_\infty$ and $\|\Upsilon_t\varphi\|_\infty \le T\|\varphi\|_\infty$, it follows that for $0 < t' < t < T$, $t - t' \le 1$,

$$\begin{aligned}
\mathbb{E}[|R_t(\varphi) - R_{t'}(\varphi)|^4] &\le 2^4\mathbb{E}\biggl[\biggl|\widehat{\mathcal{M}}_T\biggl(\int_{t'}^t \Psi_s\varphi(\cdot, s)\, ds\biggr)\biggr|^4\biggr] \\
&\qquad + 2^4\|\Upsilon_t(\varphi)\|^4_\infty\bigl(\overline{A}_{\mathbf{1}}(t) - \overline{A}_{\mathbf{1}}(t')\bigr)^2 \\
&\le 2^4\tilde{C}(T)\|\varphi\|^4_\infty\bigl(|t - t'|^4 + \bigl(\overline{A}_{\mathbf{1}}(t) - \overline{A}_{\mathbf{1}}(t')\bigr)^2\bigr) \\
&\le 2^4\tilde{C}(T)\|\varphi\|^4_\infty\bigl((t - t')^2 + \bigl(\overline{A}_{\mathbf{1}}(t) - \overline{A}_{\mathbf{1}}(t')\bigr)^2\bigr) \\
&\le 2^4\tilde{C}(T)\|\varphi\|^4_\infty\bigl(t - t' + \overline{A}_{\mathbf{1}}(t) - \overline{A}_{\mathbf{1}}(t')\bigr)^2,
\end{aligned}$$

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. In what follows, recall the definition of the process $\widehat{Y}_1^{(N)}$ given in (5.19), and let $\widehat{Y}_1$ be the analogous limit process, given by

$$\widehat{Y}_1 \doteq (\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}), \widehat{\mathcal{M}}, \widehat{\mathcal{H}}, \widehat{\mathcal{H}}(\mathbf{1})).$$

COROLLARY 8.7. *As $N \to \infty$, $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$ 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 $N \to \infty$, and $\widehat{\mathcal{H}}$ admits a version as a continuous $\mathbb{H}_{-2}$-valued process.*

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 $\mathbb{D}_{\mathbb{R}}[0,\infty)^{\tilde{k}+k}$. 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 $N \to \infty$, the following corresponding terms converge:

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

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

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

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 Theorem 5.3(2) of Mitoma [24] 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 $\mathbb{D}_{\mathcal{S}'}[0,\infty)^2 \times \mathbb{D}_{\mathbb{R}}[0,\infty)$. Since $\|\cdot\|_{\mathbb{H}_1} \overset{\mathrm{HS}}{<} \|\cdot\|_{\mathbb{H}_2}$ and estimate (8.28) holds, Corollary 6.16 of Walsh [31] 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 $\mathbb{D}_{\mathbb{H}_{-2}}[0,\infty)^2 \times \mathbb{D}_{\mathbb{R}}[0,\infty)$, as $N \to \infty$. In fact, from Corollary 8.3, Lemma 8.6(1) and the proofs of Corollary 1 of [24] and Corollary 6.16 of [31], it follows that the sample paths of $(\widehat{\mathcal{M}}, \widehat{\mathcal{H}}, \widehat{\mathcal{H}}(f_1))$ lie in $\mathbb{C}_{\mathbb{H}_{-2}}[0,\infty)^2 \times \mathbb{C}_{\mathbb{R}}[0,\infty)$, which proves the last assertion of the corollary. 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 $(\widehat{E}, \widehat{x}_0, \widehat{\nu}_0, \mathcal{J}^{\widehat{\nu}_0}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}))$ by (8.12) and the construction of $\widehat{\mathcal{M}}$ and B described after Assumption 5$'$. 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 2, 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 limit in (5.21) holds and $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1}))$ has almost surely continuous sample paths. Moreover, if g is continuous, and $\widehat{\nu}$ is defined as in (5.24), then as $N \to \infty$,*

$$(9.1)\qquad \big(\widehat{Y}_1^{(N)}, \widehat{K}^{(N)}, \widehat{X}^{(N)}, \widehat{\nu}^{(N)}, \widehat{\mathcal{K}}^{(N)}, \widehat{\mathcal{K}}^{(N)}(\mathbf{1})\big) \Rightarrow (\widehat{Y}, \widehat{K}, \widehat{X}, \widehat{\nu}, \widehat{\mathcal{K}}, \widehat{\mathcal{K}}(\mathbf{1}))$$

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

PROOF. Corollary 8.7 shows that $\widehat{Y}_1^{(N)} \Rightarrow \widehat{Y}_1$ in $\mathcal{Y}_1$ as $N \to \infty$, which in particular implies that

$$\big(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}), \widehat{\mathcal{H}}^{(N)}(\mathbf{1})\big) \Rightarrow (\widehat{E}, \widehat{x}_0, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}), \widehat{\mathcal{H}}(\mathbf{1}))$$

as $N \to \infty$. By Remark 5.1, Assumption 5, Lemma 8.6(1) and Corollary 8.7, $(\widehat{E}, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}), \widehat{\mathcal{H}}(\mathbf{1}), \mathcal{J}^{\widehat{\nu}_0}, \widehat{\mathcal{H}})$ has almost surely continuous sample paths with values in $\mathbb{R}^3 \times \mathbb{H}^2_{-2}$. Since addition in the Skorokhod topology is continuous at points in $\mathbb{C}[0,\infty)$, as $N \to \infty$,

$$(9.2)\quad \big(\widehat{Y}_1^{(N)}, \widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1})\big) \Rightarrow \big(\widehat{Y}_1, \widehat{E}, \widehat{x}_0, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1})\big).$$

By Lemma 7.2, almost surely $(\widehat{K}^{(N)}, \widehat{X}^{(N)}, \langle \mathbf{1}, \widehat{\nu}^{(N)}\rangle) = \Lambda(\widehat{E}^{(N)}, \widehat{X}^{(N)}(0), \mathcal{J}^{\widehat{\nu}_0^{(N)}}(\mathbf{1}) - \widehat{\mathcal{H}}^{(N)}(\mathbf{1}))$ for all N large enough. The continuity of Λ with respect to the uniform topology on $\mathbb{D}_{\mathbb{R}}[0,\infty)$ established in Proposition 7.3, the measurability of Λ with respect to the Skorokhod topology on $\mathbb{D}_{\mathbb{R}}[0,\infty)$ established in Lemma 7.4 and a generalized version of the continuous mapping theorem (see, e.g., Theorem 10.2 of Chapter 3 of [9]), then shows that convergence (5.21) holds with $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1})) \doteq \Lambda(\widehat{E}, \widehat{x}_0, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}))$. By the model assumptions and Lemma 4.2, almost surely, $\Delta E^{(N)}(t) \le 1$ and $\Delta D^{(N)}(t) \le 1$ for every $t \ge 0$. Combining this with (2.3), (6.10) and the second equation for $\widehat{K}^{(N)}$ in (6.8), it follows that almost surely for every $t \ge 0$,

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

Because the jump size functional (at some fixed time t) is continuous in the Skorokhod topology, the weak convergence of the process $(\widehat{K}^{(N)}, \widehat{X}^{(N)}, \widehat{\nu}^{(N)}(\mathbf{1}))$ to the process $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1}))$, which was established in (9.1), shows that $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1}))$ is almost surely continuous. Note that when g is continuous, the continuity of $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1}))$ is also guaranteed by Remark 5.5.

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

$$(9.3)\qquad \big(\widehat{Y}_1^{(N)}, \widehat{K}^{(N)}, \widehat{X}^{(N)}, \widehat{\mathcal{K}}^{(N)}, \widehat{\mathcal{K}}^{(N)}(\mathbf{1})\big) \Rightarrow (\widehat{Y}, \widehat{K}, \widehat{X}, \widehat{\mathcal{K}}, \widehat{\mathcal{K}}(\mathbf{1})).$$

In turn, (6.15) shows that $\widehat{\nu}^{(N)} = \mathcal{J}^{\widehat{\nu}_0^{(N)}} - \widehat{\mathcal{H}}^{(N)} + \widehat{\mathcal{K}}^{(N)}$, and hence, $\widehat{\nu}^{(N)}$ can be obtained as a continuous mapping of $\widehat{Y}_1^{(N)}$ and $\widehat{\mathcal{K}}^{(N)}$. Thus, (9.3) and another application of the continuous mapping theorem show that (5.25) holds with $\widehat{\nu} = \mathcal{J}^{\widehat{\nu}_0} - \widehat{\mathcal{H}} + \widehat{\mathcal{K}}$. That this coincides with the definition of $\widehat{\nu}$ given in (5.24) can be seen on recalling the definition of $\mathcal{K}$ given in (7.1). $\square$

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

PROOF OF THEOREMS 2 AND 3. The limit in (5.21), the continuity of $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1}))$ and Theorem 3 follow from Proposition 9.1. Relation (6.16) and the fact that $\widehat{K}^{(N)}$ is almost everywhere continuous because it is càdlàg, shows that

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

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

9.2. *The semimartingale property.* In view of representation (5.22) for $\widehat{X}$ and the fact that $\widehat{\mathcal{M}}_{\mathbf{1}}$ and $\widehat{E}$ are by definition semimartingales, to show that $\widehat{X}$ is a semimartingale it suffices to show that $\tilde{D}$ is a process of almost surely finite variation (on every bounded interval), and hence a semimartingale. This is 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. If h is bounded, then Assumptions 2 and 4 are also satisfied by Remark 5.2, and so the results of Theorems 2 and 3 are valid.

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

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

PROOF. We start by rewriting the expression (5.23) for $\tilde{D}$ in a more convenient form. For $t > 0$, definitions (5.6) and (5.11) of Φ_t and $\mathcal{J}^{\widehat{\nu}_0}$, respectively, show that

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

By (5.7) and the boundedness of h, $\Phi_r h$ is bounded (uniformly in r) and absolutely continuous, and Assumption 4 implies that $\int_0^t \Phi_r h\,dr = (G(\cdot + t) - G(\cdot))/(1 - G(\cdot))$ is Hölder continuous. Therefore, applying Assumption 5′(d) with $\varphi = \Phi_r h$, it follows that

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

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

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

where Υ_t is the operator defined in (8.27). Substituting $\tilde{\varphi} = h \in \mathbb{C}_b[0, L)$ in (E.1) of Lemma E.1 then yields the equality

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

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 relations (9.5) and (9.6) and definition (5.23) of $\tilde{D}$, it follows that almost surely, $\tilde{D}$ is absolutely continuous with respect to Lebesgue measure, and has density equal to

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

Relation (9.4) then follows on comparing the right-hand side above with the right-hand side of (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 4. From (5.22), (5.4) and Lemma 9.2 (see also the discussion prior to the lemma), it follows that $\widehat{X}$ is a semimartingale with the decomposition stated in Theorem 4. Using the relation $(\widehat{K}, \widehat{X}, \widehat{\nu}(\mathbf{1})) = \Lambda(\widehat{E}, \widehat{x}_0, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}))$ established in Theorem 3 along with (5.15) and (5.16), it follows that

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

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,

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

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 $f \in \mathbb{AC}_b[0,\infty)$,

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

where the latter is the convolution integral with respect to the semimartingale $\widehat{K}$. Thus, we obtain (5.26) from (5.24) and (7.1). □

SKETCH OF JUSTIFICATION OF REMARK 5.6. 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 $N \to \infty$ for all Hölder continuous f. Given the semimartingale decomposition $\widehat{K} = M^K + C^K$ established in Theorem 4, the integral on the right-hand side of 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 C^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 C^K to show, as in Lemma 8.6 and Corollary 8.7, that for functions f that are Hölder continuous and bounded, $\widehat{\mathcal{K}}^{(N)}(f) \Rightarrow \widehat{\mathcal{K}}(f)$ as $N \to \infty$, and $\widehat{\mathcal{K}}(f)$ admits a continuous version. When combined with Assumptions 3 and 5, it is easy to argue as in the proof of Theorem 2 that, in fact, the joint convergence $(\widehat{E}^{(N)}, \widehat{X}^{(N)}, \mathcal{J}^{\widehat{\nu}_0^{(N)}}(f), \widehat{\mathcal{K}}^{(N)}(f), \widehat{\mathcal{H}}^{(N)}(f)) \Rightarrow (\widehat{E}, \widehat{X}, \mathcal{J}^{\widehat{\nu}_0}(f), \widehat{\mathcal{K}}(f), \widehat{\mathcal{H}}(f))$ 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, and in fact the joint convergence specified in (5.27) holds. □

9.3. *A consistency property.* In this section a certain consistency property is established. This consistency property will be used in Section 9.4.2 to show that $\widehat{\nu}$ satisfies the stochastic age equation and in Section 9.5 to establish the strong Markov property. Roughly speaking, the consistency property states that if the age distribution satisfies the conditions stated in Assumption 5 at the initial time, then these conditions are also satisfied at any future time $s > 0$. For a precise statement, consider the following shifted processes: for $F = \widehat{E}^{(N)}, \widehat{K}^{(N)}, \widehat{E}, \widehat{K}$, and $\mathcal{R} = \widehat{\mathcal{M}}^{(N)}, \widehat{\mathcal{M}}$, and $s \geq 0$, $u \geq 0$,

$$(\Theta_s F)(u) \doteq F(s+u) - F(s), \qquad (\Theta_s \mathcal{R})_u \doteq \mathcal{R}_{s+u} - \mathcal{R}_s, \tag{9.10}$$

and for f bounded and continuous, we define

$$\begin{aligned} (\Theta_s \widehat{\mathcal{H}}^{(N)})_t(f) &\doteq \big(\Theta_s \widehat{\mathcal{M}}^{(N)}\big)_t(\Psi_{s+t} f), \\ (\Theta_s \widehat{\mathcal{H}})_t(f) &\doteq (\Theta_s \widehat{\mathcal{M}})_t(\Psi_{s+t} f), \end{aligned} \tag{9.11}$$

$$(\Theta_s \widehat{\mathcal{K}}^{(N)})_t(f) = \int_{[0,t]} \big(1 - G(t-u)\big) f(t-u)\, d\big(\Theta_s \widehat{K}^{(N)}\big)(u), \tag{9.12}$$

$$(\Theta_s \widehat{\mathcal{K}})_t(f) = f(0)(\Theta_s \widehat{K})(t) + \int_0^t (\Theta_s \widehat{K})(u) \varphi_f(t-u)\, du. \tag{9.13}$$

LEMMA 9.3. *For every bounded and continuous f*,

$$\widehat{\nu}_{s+t}^{(N)}(f) = \mathcal{J}_t^{\widehat{\nu}_s^{(N)}}(f) + \big(\Theta_s \widehat{\mathcal{K}}^{(N)}\big)_t(f) - \big(\Theta_s \widehat{\mathcal{H}}^{(N)}\big)_t(f), \qquad s, t \geq 0. \tag{9.14}$$

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

$$\widehat{\nu}_{s+t}(f) = \mathcal{J}_t^{\widehat{\nu}_s}(f) + (\Theta_s \widehat{\mathcal{K}})_t(f) - (\Theta_s \widehat{\mathcal{H}})_t(f), \qquad s, t \geq 0. \tag{9.15}$$

In addition, for every $s > 0$,

$$\big(\Theta_s \widehat{K}, \widehat{X}_{s+\cdot}, \widehat{\nu}_{s+\cdot}(\mathbf{1})\big) = \Lambda\big(\Theta_s \widehat{E}, \widehat{X}(s), \mathcal{J}^{\widehat{\nu}_s}(\mathbf{1}) - (\Theta_s \widehat{\mathcal{H}})(\mathbf{1})\big). \tag{9.16}$$

Furthermore, if Assumption $5'$ holds, then 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.

9.4. *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 5.7. First, in Section 9.4.1 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 9.4.2 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(1) is presented in Section 9.4.3. Throughout the section we assume that the conditions of Theorem 5 (namely, Assumptions 1, 3 and 5, and the conditions on the fluid limit), are satisfied, and that h is bounded and absolutely continuous, and state only additional assumptions when imposed.

9.4.1. *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 Υ_t given in (8.27). 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 4 $\widehat{K}$ is a semimartingale, and $\widehat{\nu}_t(f)$ is given by (5.26) for every $f \in \mathbb{AC}_b[0, L)$.

LEMMA 9.4. *For any $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$ such that $\varphi(\cdot, t)$ is Hölder continuous uniformly in t and absolutely continuous, $\mathbb{P}$-almost surely for every $t > 0$, we have*

$$\begin{aligned}\int_0^t \widehat{\nu}_s(\varphi(\cdot, s)\, ds &= \widehat{\nu}_0((\Upsilon_t\varphi)(\cdot, 0)) - \widehat{\mathcal{M}}_t(\Upsilon_t\varphi) \\ &\quad + \int_{[0,t]} \left(\int_u^t \varphi(s-u, s)\big(1 - G(s-u)\big)\, ds \right) d\widehat{K}(u).\end{aligned} \tag{9.17}$$

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

$$\begin{aligned}&\int_0^t \widehat{\nu}_s(\varphi(\cdot, s))\, ds \\ &\quad = \int_0^t \widehat{\nu}_0(\Phi_s\varphi(\cdot, s))\, ds - \int_0^t \widehat{\mathcal{M}}_s(\Psi_s\varphi(\cdot, s))\, ds \\ &\qquad + \int_0^t \left(\int_{[0,s]} \varphi(s-u, s)\big(1 - G(s-u)\big)\, d\widehat{K}(u) \right) ds.\end{aligned} \tag{9.18}$$

From the definition (8.27) of Υ_t and the fact that $(\Psi_s f)(\cdot, 0) = \Phi_s f(\cdot)$, it follows that

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

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

$$\begin{aligned}\int_0^t \widehat{\nu}_0(\Phi_s\varphi(\cdot, s))\, ds &= \widehat{\nu}_0\left(\int_0^t (\Phi_s\varphi(\cdot, s))(\cdot)\, ds \right) \\ &= \widehat{\nu}_0((\Upsilon_t\varphi)(\cdot, 0)),\end{aligned} \tag{9.19}$$

which shows that the first terms on the right-hand sides of (9.17) and (9.18) are equal. Equality of the second terms on the right-hand sides of (9.17) and (9.18) follows from (E.1) of Lemma E.1, whereas 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 [28]. This completes the proof of the lemma. $\square$

9.4.2. *A verification lemma.* We now show that the process $\{\widehat{\nu}_t, t \geq 0\}$ of Theorem 3 is a solution to the stochastic age equation. For this, it will be convenient to introduce the function ψ_h defined as follows: $\psi_h(x,t) \doteq \exp(r_h(x,t))$ for $(x,t) \in [0,L) \times [0,\infty)$, where

$$r_h(x,t) \doteq \begin{cases} -\displaystyle\int_{x-t}^{x} h(u)\,du, & \text{if } 0 \leq t \leq x, \\ -\displaystyle\int_{0}^{x} h(u)\,du, & \text{if } 0 \leq x \leq t. \end{cases} \tag{9.20}$$

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

$$\psi_h(x,t) = \begin{cases} \dfrac{1-G(x)}{1-G(x-t)}, & \text{if } 0 \leq t \leq x, \\ 1-G(x), & \text{if } 0 \leq x \leq t. \end{cases} \tag{9.21}$$

If g is absolutely continuous, then G is continuously differentiable, and ψ_h lies in $\mathcal{C}_b^{1,1}([0,L) \times [0,\infty))$ and satisfies

$$\frac{\partial \psi_h}{\partial x} + \frac{\partial \psi_h}{\partial t} = -h\psi_h \tag{9.22}$$

for almost every $(x,t) \in [0,L) \times [0,\infty)$. 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 $u \in [0,s]$,

$$\begin{aligned} \frac{\psi_h(x+s-u,s)}{\psi_h(x,u)} &= \frac{1-G(x+s-u)}{1-G(x)} \\ &= \begin{cases} \dfrac{1-G(x+s)}{1-G(x)}, & \text{if } u = 0, \\ \big(1-G(s-u)\big), & \text{if } x = 0. \end{cases} \end{aligned} \tag{9.23}$$

PROPOSITION 9.5. *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 shows that for every $t > 0$, $\{\widehat{\nu}_t(f), f \in \mathbb{AC}_b[0,L)\}$ 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 9.3 that for every $f \in \mathbb{AC}_b[0,L)$, almost surely $s \mapsto \widehat{\nu}_s(f)$ is measurable. Therefore, it only remains to show that $\widehat{\nu}$ satisfies equation (5.28). Fix $t \in [0,\infty)$ and $\varphi \in \mathbb{C}_b^{1,1}([0,L) \times [0,\infty))$ 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 (9.22) that

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

Substituting this and identity (9.23) into the definition (8.27) of Υ_t, it follows that

$$
\begin{aligned}
&\big(\Upsilon_t(\varphi_x+\varphi_s-h\varphi)\big)(x,u)\\
&\quad=\int_u^t \frac{((\varphi_x+\varphi_s-h\varphi)\psi_h)(x+s-u,s)}{\psi_h(x,u)}\,ds\\
&\quad=\int_u^t \frac{((\varphi\psi_h)_x+(\varphi\psi_h)_s)(x+s-u,s)}{\psi_h(x,u)}\,ds\\
&\quad=\frac{\varphi(x+t-u,t)\psi_h(x+t-u,t)}{\psi_h(x,u)}-\varphi(x,u).
\end{aligned}
\tag{9.24}
$$

Note that the integrand in the last integral on the right-hand side of (9.17) of Lemma 9.4 is $\Upsilon_t(\varphi(\cdot,s))(0,u)$. Therefore, applying Lemma 9.4 with φ replaced by $\varphi_x+\varphi_s-h\varphi$, and using (9.24) and the identity $\psi_h(x,0)=\psi_h(0,u)=1$, it follows that

$$
\begin{aligned}
&\int_0^t \widehat{\nu}_s\big(\varphi_x(\cdot,s)+\varphi_s(\cdot,s)-h\varphi(\cdot,s)\big)\,ds\\
&\quad=\widehat{\nu}_0\big(\varphi(\cdot+t,t)\psi_h(\cdot+t,t)\big)+\int_{[0,t]}\varphi(t-u,t)\psi_h(t-u,t)\,d\widehat{K}(u)\\
&\qquad-\iint_{[0,L)\times[0,t]}\varphi(x+t-u,t)\frac{\psi_h(x+t-u,t)}{\psi_h(x,u)}\widehat{\mathcal{M}}(dx,du)\\
&\qquad-\widehat{\nu}_0(\varphi(\cdot,0))-\int_{[0,t]}\varphi(0,u)\,d\widehat{K}(u)+\iint_{[0,L)\times[0,t]}\varphi(x,u)\widehat{\mathcal{M}}(dx,du).
\end{aligned}
\tag{9.25}
$$

Since φ 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 (9.23), it is clear that the sum of the first three terms on the right-hand side of (9.25) is equal to $\widehat{\nu}_t(\varphi(\cdot,t))$. With this substitution, (9.25) reduces to the stochastic age equation (5.28) associated with $(\widehat{\nu}_0,\widehat{K},\widehat{\mathcal{M}})$. $\square$

9.4.3. *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 $\mathcal{D}[0,\infty)$ is the space of test functions, and $\mathcal{D}'[0,\infty)$ is the space of distributions. Also, recall that g' denotes the density of g (which is well defined since we have assumed g is absolutely continuous).

LEMMA 9.6. *Suppose that $g'\in\mathbb{L}^\infty_{\mathrm{loc}}[0,L)$. Given a solution $\{\nu_t,t\ge 0\}$ to the stochastic age equation associated with $(\widehat{\nu}_0,\widehat{K},\widehat{\mathcal{M}})$, define*

$$
\mu_t(\tilde{f})\doteq\nu_t\big(\tilde{f}(1-G)^{-1}\big),\qquad \tilde{f}\in\mathcal{D}[0,\infty).
\tag{9.26}
$$

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

$\mathcal{D}[0,\infty)$ *and* $t \geq 0$,

$$\mu_t(\tilde{f}) = \widehat{\nu}_0\big(\tilde{f}(1-G)^{-1}\big) + \int_0^t \mu_s(\tilde{f}_x)\,ds + \tilde{f}(0)\widehat{K}(t) \\ - \widehat{\mathcal{M}}_t\big(\tilde{f}(1-G)^{-1}\big). \tag{9.27}$$

PROOF. Fix $\tilde{f} \in \mathcal{D}[0,\infty)$, and let $f \doteq \tilde{f}(1-G)^{-1}$. Under the given assumptions on G, it follows from Lemma B.1 that $f \in \mathbb{H}_2$. Because $\{\nu_t, t \geq 0\}$ is a solution of the stochastic age equation, it is a continuous $\mathbb{H}_{-2}$-valued processes. Therefore $\{\nu_t(f), t \geq 0\}$, and hence $\{\mu_t(\tilde{f}), t \geq 0\}$, are continuous real-valued stochastic processes. Since $\mathcal{D}[0,\infty)$ and $\mathcal{D}'[0,\infty)$ are nuclear spaces, by Mitoma's theorem [23] it follows in fact that $\{\mu_t, t \geq 0\}$ is a continuous $\mathcal{D}'[0,\infty)$-valued process.

We now show that $\{\mu_t, t \geq 0\}$ solves the transport equation (9.27). Indeed, by property 2 of Lemma B.1 it also follows that for $\tilde{f} \in \mathcal{D}[0,\infty)$, $f_x = \tilde{f}_x(1-G)^{-1} + fh$ is bounded and Lipschitz continuous. Therefore, we can substitute $\varphi = f$ in the stochastic age equation (5.28) and use the identity $1 - G(0) = 1$ to obtain, for $t \geq 0$,

$$\begin{aligned} \mu_t(\tilde{f}) &= \nu_t\big(\tilde{f}(1-G)^{-1}\big) \\ &= \nu_0\big(\tilde{f}(1-G)^{-1}\big) + \int_0^t \nu_s\big(\tilde{f}_x(1-G)^{-1}\big)\,ds \\ &\quad + \tilde{f}(0)\widehat{K}(t) - \widehat{\mathcal{M}}_t\big(\tilde{f}(1-G)^{-1}\big). \end{aligned} \tag{9.28}$$

However, (9.26) implies that $\nu_s(\tilde{f}_x(1-G)^{-1}) = \mu_s(\tilde{f}_x)$. Substituting this back into (9.28), it follows that $\{\mu_t, t \geq 0\}$ satisfies (9.27). □

We can now complete the proof of Theorem 5(1).

PROOF OF THEOREM 5(1). By assumption, h is Hölder continuous. Therefore, Proposition 9.5 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 $\mathcal{D}'[0,\infty)$-valued process defined as in (9.26), but with ν replaced by $\nu^{(i)}$. By Lemma 9.6, each $\mu^{(i)}$ satisfies the stochastic transport equation (9.27) 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 $\mathcal{D}[0,\infty)$,

$$\frac{d}{dt}\eta_t(\tilde{f}) - \eta_t(\tilde{f}_x) = 0, \qquad \eta_0(\tilde{f}) = 0.$$

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 [10].

Thus, for every $\tilde{f} \in \mathcal{D}[0,\infty) = \mathbb{C}_c^\infty[0,\infty)$, $\mu^{(1)}(\tilde{f}) = \mu^{(2)}(\tilde{f})$ or equivalently,

$$\nu_t^{(1)}\big(\tilde{f}(1-G)^{-1}\big) = \nu_t^{(2)}\big(\tilde{f}(1-G)^{-1}\big). \tag{9.29}$$

Now, by Lemma B.1 for any $f \in \mathbb{C}_c^\infty[0,\infty)$, the assumptions on G imply that $f(1-G) \in \mathbb{H}_2$. Since $\mathbb{C}_c^\infty[0,\infty)$, equipped with the $\|\cdot\|_{\mathbb{H}_2}$ norm, is dense in $\mathbb{H}_2$ (by the very definition of $\mathbb{H}_2$) there exists a sequence $\tilde{f}_n \in \mathbb{C}_c^\infty[0,\infty)$ such that $\tilde{f}_n \to f(1-G)$ in $\mathbb{H}_2$ as $n \to \infty$. Replacing $\tilde{f}$ by $\tilde{f}_n$ in (9.29) and then letting $n \to \infty$, it follows that $\nu_t^{(1)}(f) = \nu_t^{(2)}(f)$ for every $f \in \mathbb{C}_c^\infty[0,\infty)$. Once again using the fact that $\mathbb{C}_c^\infty[0,\infty)$ 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. $\square$

9.5. *The strong Markov property.* We now establish the strong Markov property stated as Theorem 5(2). We will use the notation introduced in Section 9.3, and the consistency property stated as Lemma 9.3.

PROOF OF THEOREM 5(2). Fix $s, t > 0$. First, note that by Theorem 3, $(\widehat{X}, \widehat{\nu})$ is a continuous $\mathbb{R} \times \mathbb{H}_{-2}$-valued process. Moreover, by Lemma 9.3, Assumption 5 is satisfied with $\widehat{\nu}_0$ replaced by $\widehat{\nu}_s$, which in particular implies that the random element $\mathcal{J}_{t+\cdot}^{\widehat{\nu}_s}(\mathbf{1}) = \{\mathcal{J}_{t+u}^{\widehat{\nu}_s}(\mathbf{1}), u \geq 0\}$ almost surely takes values in $\mathbb{C}_{\mathbb{R}}[0,\infty)$. Next, observe that $\Phi_u \mathbf{1} = \mathbf{1} - \int_0^u (\Phi_r h)(\cdot)\,dr$ and that $x \mapsto \int_0^u (\Phi_r h)(x)\,dr$ is bounded and (due to Assumption 4) Hölder continuous, uniformly with respect to u. Properties 1 and 3 of Lemma 8.6 (with $\widehat{\mathcal{H}}$ replaced by $\Theta_s\widehat{\mathcal{H}}$) then show that the random processes $(\Theta_s\widehat{\mathcal{H}})_\cdot(\mathbf{1}) = \{(\Theta_s\widehat{\mathcal{H}})_t(\mathbf{1}), t \geq 0\}$ and $(\Theta_s\widehat{\mathcal{H}})_t(\Phi_\cdot\mathbf{1}) = \{(\Theta_s\widehat{\mathcal{H}})_t(\Phi_u\mathbf{1}), u \geq 0\}$ take values in $\mathbb{C}_{\mathbb{R}}[0,\infty)$. In addition, Assumption 3 and Corollary 8.7 show that $\Theta_s\widehat{E}$ and $\Theta_s\widehat{\mathcal{H}}$ are, respectively, $\mathbb{C}_{\mathbb{R}}[0,\infty)$-valued and $\mathbb{C}_{\mathbb{H}_{-2}}[0,\infty)$-valued. We now claim that there exists a continuous mapping from $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty)^4 \times \mathbb{C}_{\mathbb{H}_{-2}}[0,\infty)$ to $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty)$, which we denote by $\tilde{\Lambda} = \tilde{\Lambda}_t$, such that $\widehat{\mathbb{P}}$-almost surely,

$$\begin{aligned}&\big(\widehat{X}(s+t), \widehat{\nu}_{s+t}, \mathcal{J}^{\widehat{\nu}_{s+t}}(\mathbf{1})\big)\\ &\quad = \tilde{\Lambda}\big(\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_\cdot^{\widehat{\nu}_s}(\mathbf{1}), \Theta_s\widehat{E}, (\Theta_s\widehat{\mathcal{H}})_t(\Phi_\cdot\mathbf{1}), (\Theta_s\widehat{\mathcal{H}})_\cdot(\mathbf{1}), \Theta_s\widehat{\mathcal{H}}\big).\end{aligned} \tag{9.30}$$

To see why this is the case, first note that equation (9.15) shows that $\widehat{\nu}_{s+t}$ is the sum of $(\Theta_s\widehat{\mathcal{H}})_t$, $\mathcal{J}_t^{\widehat{\nu}_s}$ and $(\Theta_s\widehat{\mathcal{K}})_t$ and, by Lemma B.1(2), the map from $\mathbb{H}_{-2}$ to $\mathbb{D}_{\mathbb{H}_{-2}}[0,\infty)$ that takes $\widehat{\nu}_s$ to $\{\mathcal{J}_t^{\widehat{\nu}_s}, t \geq 0\}$ is continuous. Also, for $u > 0$, $\Phi_u\mathbf{1}$ is bounded and absolutely continuous. Hence, by (9.15), the definition of $\mathcal{J}^{\widehat{\nu}_{s+t}}$ and the semigroup property for Φ, $\widehat{\mathbb{P}}$-almost surely, for $u, s, t \geq 0$,

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

Due to the almost sure continuity of $\widehat{K}$ established in Theorem 2, definitions (7.1) and (9.13) 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 (9.16) of Lemma 9.3, $(\Theta_s\widehat{K}, \widehat{X}(s+\cdot))$ are equal to the first and second marginals of $\Lambda(\Theta_s\widehat{E}, \widehat{X}(s), \mathcal{J}^{\widehat{\nu}_s}(\mathbf{1}) - \Theta_s\widehat{\mathcal{H}}(\mathbf{1}))$, where Λ is the centered many-server map, which is continuous by Proposition 7.3. Since $\widehat{X}$ almost surely has continuous sample paths, for any given $t \geq 0$, the pair $(\Theta_s\widehat{K}, \widehat{X}(s+t))$ can also be obtained as a continuous mapping of $(\Theta_s\widehat{E}, \widehat{X}(s), \mathcal{J}^{\widehat{\nu}_s}(\mathbf{1}) - \Theta_s\widehat{\mathcal{H}}(\mathbf{1}))$. When combined, the above observations show that claim (9.30) 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, (9.11) 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 $\{\widehat{\mathcal{F}}_t, t \geq 0\}$. 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 $\overline{F}$ on $[0,\infty) \times (\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty))$,

$$
\begin{aligned}
&\widehat{\mathbb{E}}[\overline{F}(s, \widehat{X}_{s+t}, \widehat{\nu}_{s+t}, \mathcal{J}_{\cdot}^{\widehat{\nu}_{s+t}}(\mathbf{1}))|\widehat{\mathcal{F}}_s] \\
&\quad = \widehat{\mathbb{E}}[\overline{F}(s, \tilde{\Lambda}(\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_{\cdot}^{\widehat{\nu}_s}(\mathbf{1}), \Theta_s\widehat{E}, (\Theta_s\widehat{\mathcal{H}})_t(\Phi_{\cdot}\mathbf{1}), \Theta_s\widehat{\mathcal{H}}(\mathbf{1}), \Theta_s\widehat{\mathcal{H}}))|\widehat{\mathcal{F}}_s] \\
&\quad = \widehat{\mathbb{E}}[\overline{F}(s, \tilde{\Lambda}(\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_{\cdot}^{\widehat{\nu}_s}(\mathbf{1}), \Theta_s\widehat{E}, (\Theta_s\widehat{\mathcal{H}})_t(\Phi_{\cdot}\mathbf{1}), \Theta_s\widehat{\mathcal{H}}(\mathbf{1}), \Theta_s\widehat{\mathcal{H}}))| \\
&\qquad\qquad \widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_{\cdot}^{\widehat{\nu}_s}(\mathbf{1})] \\
&\quad = \widehat{\mathbb{E}}\big[\overline{F}\big(s, \widehat{X}(s+t), \widehat{\nu}_{s+t}, \mathcal{J}_{\cdot}^{\widehat{\nu}_{s+t}}(\mathbf{1})\big)|\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}^{\widehat{\nu}_s}(\mathbf{1})\big].
\end{aligned}
$$

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

By Theorems 2 and 3, the sample paths $s \mapsto (\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_{\cdot}^{\widehat{\nu}_s}(\mathbf{1}))$ of the Markov process take values in the state space $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty)$ and are continuous. Since the state space is a Polish space, there exists a Markov stochastic kernel $P_{s,u}$ on $(\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty), \mathbb{B}(\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty)))$ such that for each $(x, \nu, \psi) \in \mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty)$, and measurable function F on $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty)$,

$$
\begin{aligned}
T_{s,u}F(x,\nu,\psi) &\doteq \widehat{\mathbb{E}}[F(\widehat{X}(u), \widehat{\nu}_u, \mathcal{J}_{\cdot}^{\widehat{\nu}_u}(\mathbf{1}))|(\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_{\cdot}^{\widehat{\nu}_s}(\mathbf{1})) = (x,\nu,\psi)] \\
&= \int_{\mathbb{R}\times\mathbb{H}_{-2}\times\mathbb{C}_{\mathbb{R}}[0,\infty)} F(w) P_{s,u}((x,\nu,\psi), dw).
\end{aligned}
$$

Note that for $0 \leq s \leq u \leq v < \infty$, $T_{s,u}T_{u,v} = T_{s,v}$ and $T_{s,s}$ is the identity. It now follows from (9.30), the continuity of $\tilde{\Lambda}$ established above and the continuity of the sample paths of the deterministic fluid processes $\overline{E}$ and $\overline{\nu}$, which, respectively, determine the distribution of $\Theta_s\widehat{E}$ and the quadratic variation of the martingale $\Theta_s\widehat{\mathcal{M}}$ (and hence the distribution of $\Theta_s\widehat{\mathcal{H}}$) that for every continuous functional F on $\mathbb{R} \times \mathbb{H}_{-2} \times \mathbb{C}_{\mathbb{R}}[0,\infty)$ and $t \geq 0$, the map $(s,(x,\nu,\psi)) \mapsto T_{s,s+t}F(x,\nu,\psi)$ is continuous. This implies that the Markov process $\{(\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_{\cdot}^{\widehat{\nu}_s}(\mathbf{1})), s \geq 0\}$ satisfies the Feller property (as defined, e.g., on page 23 of [11]). Hence, by Theorem 2.4 of Friedman [11], $\{(\widehat{X}(s), \widehat{\nu}_s, \mathcal{J}_{\cdot}^{\widehat{\nu}_s}(\mathbf{1})), s \geq 0\}$ is a strong Markov process. (The Feller property and Theorem 2.4 of [11] are stated for Euclidean spaces, but carry over with no changes to state spaces that are complete, separable metric spaces, as is the case in the present setting.) Furthermore, note that the strong Markov process is time-homogeneous if the arrival process satisfies Assumption 3(a) with $\overline{\lambda} = 1$ and $(\overline{x}_0, \overline{\nu}_0) = (1, \overline{\nu}_*)$, in which case the fluid limit is critical and constant. $\square$

REMARK 9.7. A more natural candidate for the (strong) Markov process would be the process $\{(\widehat{X}_t, \widehat{\nu}_t), \widehat{\mathcal{F}}_t, t \geq 0\}$ that takes values in $\mathbb{R} \times \mathbb{H}_{-2}$. However, in order to establish the Markov property, we need $\mathcal{J}^{\widehat{\nu}_s}$ and $\mathcal{J}^{\widehat{\nu}_s}(\mathbf{1})$ to be expressed as measurable mappings of $\widehat{\nu}_s$. As shown in Lemma B.1, the additional boundedness assumption on $g'/(1-G))$ ensures that the map from $\mathbb{H}_{-2} \mapsto \mathbb{D}_{\mathbb{H}_{-2}}[0, \infty)$ that takes $\widehat{\nu}_s$ to $\mathcal{J}^{\widehat{\nu}_s} = \{\mathcal{J}_t^{\widehat{\nu}_s}, t \geq 0\}$ is continuous. This is a reasonable assumption because, as noted in Remark 5.8, 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 $\mathbb{C}_{\mathbb{R}}[0, \infty)$ that takes $\widehat{\nu}_s$ to $\mathcal{J}^{\widehat{\nu}_s}(\mathbf{1}) = \{\mathcal{J}_t^{\widehat{\nu}_s}(\mathbf{1}), t \geq 0\}$, which would require that $\Phi_s \mathbf{1}$ lie in $\mathbb{H}_2$, cannot be obtained without imposing rather severe assumptions on the service distribution G. While $\widehat{\nu}_s$ does extend as a random linear functional to act on continuous and bounded functions such as $\Phi_s f$, the space of random linear functionals appears not to be a sufficiently nice space to support a Markov process. In particular, it is not clear that a regular conditional probability would exist so as to enable the construction of the Markov kernel. We chose to resolve this issue by adding the $\mathbb{C}_{\mathbb{R}}[0, \infty)$-valued process $\mathcal{J}^{\widehat{\nu}_s}(\mathbf{1}) = \{\mathcal{J}_t^{\widehat{\nu}_s}(\mathbf{1}), t \geq 0\}$ to the state descriptor.

APPENDIX A: PROPERTIES OF THE MARTINGLE MEASURE SEQUENCE

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

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

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

PROOF. In order to show that $\{\mathcal{M}_t^{(N)}(B); t \geq 0, B \in \mathbb{B}_0[0, L)\}$ defines a martingale measure on $[0, L)$, it suffices to verify the three properties stated in the definition of a martingale measure given on page 287 of Walsh [31]. The first property in [31], namely that $\mathcal{M}_0^{(N)}(B) = 0$ for every $B \in \mathbb{B}_0[0,\infty)$, follows trivially from the definition and the third property, which states that $\{\mathcal{M}_t^{(N)}(B), \mathcal{F}_t^{(N)}, t \geq 0\}$ is a local martingale for each $B \in \mathbb{B}_0[0,\infty)$, is an immediate consequence of Remark 4.1. In addition, from (4.4) it follows that $\langle \mathcal{M}^{(N)}(B)\rangle$, the predictable quadratic variation of $\mathcal{M}^{(N)}(B)$, is equal to $A_{\mathbb{I}_B}^{(N)}$. Since $\mathbb{E}[A_{\mathbb{I}_B}^{(N)}(t)]$ is dominated by $\mathbb{E}[D^{(N)}(t)]$, which is finite by Lemma 5.6 of [21], 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 $\mathbb{B}_0[0,L)$, and hence defines a finite $\mathbb{L}^2(\Omega, \mathcal{F}^{(N)}, \mathbb{P})$-valued measure. This shows that the second property in [31] is also satisfied, and thus completes the proof of the lemma. $\square$

A.2. Proof of Lemma 4.2. We fix $N \in \mathbb{N}$ and, for conciseness, suppress the superscript N from the notation. As shown below, Lemma 4.2 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.

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 $k = -\langle \mathbf{1}, \nu_0\rangle + 1, -\langle \mathbf{1}, \nu_0\rangle + 2, \ldots,$ 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 θ_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 [19], $\{(R_E(t), X(t), \nu_t), t \geq 0\}$ is a strong Markov process. Therefore, conditioned on $\mathcal{F}_{\theta_k}$, the departure times of customers j, $j \leq k$, are independent of the arrivals after θ_k and, moreover, the distributions of their departure times only depend on $\{a_j(\theta_k), j \leq k\}$. Consequently, the probability of the event $\mathcal{E}_k$ is the same in 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 with index $j < k$ 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,

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

where the last equality follows from the independence of $\tilde{v}_k$ and $\tilde{J}^k$. The logic that was used in Lemma 5.4 of [21] to identify the compensator of D can also be used to show that the $\{\mathcal{G}_t^k\}$-compensator of $\tilde{D}^{\theta_k}$ equals

$$\int_0^{\cdot} \left(\int_{[0,L)} \frac{g(x+s)}{1-G(x)} \nu_0'(dx) \right) ds \qquad \text{where } \nu_0' \doteq \tilde{\nu}_0 - \delta_0,$$

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 [16]. 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 kth 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 Lemma 5.4 of [21], it can be shown that $\{D^r(t), t \geq 0\}$ has a continuous $\{\mathcal{G}_t\}$-compensator, given explicitly by

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

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

$$\mathcal{T} \doteq \{\bar{t} = (t_1, \ldots, t_m, \ldots) \in [0, \infty)^{\infty} : 0 \leq t_1 \leq t_2 \leq \cdots\},$$

and let $\overline{T}^r$ denote the random $\mathcal{T}$-valued sequence of times after r at which E has a jump. Moreover, let μ denote the conditional probability distribution of $\overline{T}^r$, given $\mathcal{F}_r$. For any $\bar{t} \in \mathcal{T}$, using the fact that $0 \leq \Delta E(t) \leq 1$, the conditional independence of $\{D^r(t), t \geq 0\}$ from $\overline{T}^r$ given $\mathcal{F}_r$ and the property established above that, conditional of $\mathcal{F}_r$, D^r has almost surely no fixed jumps, we have

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

In turn, integrating the left-hand side above with respect to the conditional distribution μ and then taking expectations, it follows that

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

Since the term inside the expectation is nonnegative, this proves (4.7).

REMARK A.2. From the proof, it is clear that relations (4.6) and (4.7) also hold almost surely with respect to $\mathbb{P}^{(N)}_{r,k,\mu}$ for every $(r,k,\mu) \in [0,\infty) \times \mathbb{N} \times \mathbb{M}_F[0,L)$.

A.3. A consequence of Lemma 4.2. We now establish a consequence of Lemma 4.2, which will be used in the proof of the asymptotic independence property in Section 8.2.

COROLLARY A.3. *As $N \to \infty$,*

$$\frac{1}{N}\mathbb{E}\Bigg[\sum_{s\le t} \Delta E^{(N)}(s)\Delta D^{(N)}(s)\Bigg] \to 0.$$

PROOF. With the aim of computing the left-hand side above, and using the same notation as in Lemma 4.2, for $r, s \in [0,\infty)$, let $D^{(N),r}(s)$ denote the cumulative number of departures during $(r, r+s]$ of customers that entered service at or before time r, and let $D^{(N)+,r}(s)$ be the cumulative number of departures during $(r, r+s]$ of customers that have entered service after time r. For $\delta > 0$ and $k = 0, 1, 2, \ldots,$ we have

$$\begin{aligned}\sum_{s\in(k\delta,(k+1)\delta]} \Delta E^{(N)}(s)\Delta D^{(N)}(s) &= \sum_{s\in(0,\delta]} \Delta E^{(N)}(k\delta+s)\Delta D^{(N),k\delta}(s)\\ &\quad+ \sum_{s\in(0,\delta]} \Delta E^{(N)}(k\delta+s)\Delta D^{(N)+,k\delta}(s).\end{aligned}$$

The first summand on the right-hand side above is almost surely equal to zero by (4.7) of Lemma 4.2. Since $E^{(N)}$ has unit jump sizes, the second term can be bounded as follows:

$$\begin{aligned}\sum_{s\in(k\delta,(k+1)\delta]} \Delta E^{(N)}(s)\Delta D^{(N)}(s) &\le \sum_{s\in(0,\delta]} \Delta D^{(N)+,k\delta}(s)\\ &\le \sum_{j=K^{(N)}(k\delta)+1}^{K^{(N)}((k+1)\delta)} \mathbb{I}_{\{v_j\le\delta\}}.\end{aligned}\tag{A.3}$$

Summing (A.3) over $k = 0, 1, \ldots, \lfloor t/\delta \rfloor$ and dividing by N, we obtain

$$\begin{aligned}\frac{1}{N}\mathbb{E}\Big[\sum_{s\le t}\Delta E^{(N)}(s)\Delta D^{(N)}(s)\Big] &\le \mathbb{E}\Big[\int_0^{t+\delta}\mathbb{I}_{\{v_{K^{(N)}(s)}\le\delta\}}\,d\overline{K}^{(N)}(s)\Big]\\ &\le \mathbb{E}\big[\overline{Q}^{(N)}_{\mathbb{I}_{[0,\delta]}}(t+2\delta)\big]\\ &= \mathbb{E}\big[\overline{A}^{(N)}_{\mathbb{I}_{[0,\delta]}}(t+2\delta)\big],\end{aligned}$$

where the last equality follows from the martingale property stated in Remark 4.1. On both sides, taking first the limit supremum as $N \to \infty$, and then the limit as $\delta \downarrow 0$, we obtain

$$\limsup_{N\to\infty}\frac{1}{N}\mathbb{E}\Big[\sum_{s\le t}\Delta E^{(N)}(s)\Delta D^{(N)}(s)\Big] \le \lim_{\delta\downarrow 0}\limsup_{N\to\infty}\mathbb{E}\big[\overline{A}^{(N)}_{\mathbb{I}_{[0,\delta]}}(t+2\delta)\big] = 0,$$

where the last equality follows from Lemma 5.8(3) of [21]. Since $\Delta E^{(N)}(s)$ and $\Delta D^{(N)}(s)$ are always nonnegative this establishes the corollary. $\square$

APPENDIX B: RAMIFICATIONS OF ASSUMPTIONS ON THE SERVICE DISTRIBUTION

LEMMA B.1. *The following properties hold*:

(1) *If h is bounded*, *then Assumptions 2 and 4 are satisfied.*

(2) *Suppose g is absolutely continuous and either $g' \in \mathbb{L}^\infty_{\text{loc}}[0, L)$ or $g' \in \mathbb{L}^2_{\text{loc}}[0, L)$. Then for any $\tilde{f} \in \mathcal{D}[0,\infty)$, the corresponding functions $\overline{f} \doteq \tilde{f}(1-G)$ and $f \doteq \tilde{f}(1-G)^{-1}$ lie in $\mathbb{H}_2$. Furthermore*, *in the case when $g' \in \mathbb{L}^\infty_{\text{loc}}[0, L)$, f' is Lipschitz continuous.*

(3) *If h is bounded*, *g is absolutely continuous and $g'/(1-G) \in \mathbb{L}^\infty[0, L)$, then $f \in \mathbb{H}_2$ implies $\Phi_t f \in \mathbb{H}_2$ for every $t \ge 0$ and*, *moreover*, *for every $t > 0$*, *the mapping from $\mathbb{H}_{-2}$ to $\mathbb{H}_{-2}$ that takes $\nu \mapsto \mathcal{J}^\nu_t = \nu(\Phi_t \cdot)$ is Lipschitz continuous.*

PROOF. If h is uniformly bounded, then Assumption 2 is trivially satisfied and

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

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

Now, suppose that g is absolutely continuous. Then for any $\tilde{f} \in \mathcal{D}[0,\infty)$, $\tilde{f}$, $\tilde{f}'$ and $\tilde{f}''$ lie in $\mathbb{C}^\infty_c[0,\infty)$ and so $\overline{f} = \tilde{f}(1-G)$ is absolutely continuous with density $f' = \tilde{f}'(1-G) - \tilde{f}g$ and f' is also absolutely continuous, with density $f'' = \tilde{f}''(1-G) - 2\tilde{f}'g - \tilde{f}g'$. Thus f, f' and the first two terms on the right-hand side of the expression for f'' are continuous and bounded with compact support. Moreover, if $g' \in \mathbb{L}^\infty_{\text{loc}}[0, L)$, then $\tilde{f}g'$ is also bounded and has compact support.

Thus, $f \in \mathbb{H}_2$ in this case. On the other hand, if $g' \in \mathbb{L}^2_{\text{loc}}[0, L)$ and the support of f is contained in the interval $[0, R]$, for some $R < \infty$, then

$$\int_0^\infty \tilde{f}^2(x)(g'(x))^2(1+x^2)^2\,dx$$
$$\leq \sup_{x\in[0,R]} |f^2(x)(1+x^2)^2| \left| \int_0^R |g'(x)|^2\,dx \right| < \infty,$$

where the finiteness follows from the fact that f, and therefore the function $x \mapsto f^2(x)(1+x^2)^2$, is continuous and the fact that $g' \in \mathbb{L}^2_{\text{loc}}[0, L)$. Thus, in this case too, $f \in \mathbb{H}_2$.

Next, consider $f = \tilde{f}(1-G)^{-1}$. Then $f' = \tilde{f}'(1-G)^{-1} + hf$ is also absolutely continuous with compact support (and hence lies in $\mathbb{L}^2_{\text{loc}}[0, L)$). Moreover, elementary calculations show that $f'' = \tilde{f}''(1-G)^{-1} + 2hf' + + \tilde{f}g'(1-G)^{-2}$. Since G and g are absolutely continuous and $\tilde{f}''$ and f' are continuous with compact support, the first two terms on the right-hand side of the expression for f'' are also continuous with compact support, and hence are bounded and lie in $\mathbb{L}^2_{\text{loc}}[0, L)$. In addition, if $g' \in \mathbb{L}^\infty_{\text{loc}}[0, L)$, then the last term is also bounded and with compact support and hence lies in $\mathbb{L}^2_{\text{loc}}[0, L)$. Thus, in this case $f \in \mathbb{H}_2$ and f' is Lipschitz continuous. On the other hand, if g' lies in $\mathbb{L}^2_{\text{loc}}[0, L)$ then an argument similar to that given above shows that the last term lies in $\mathbb{L}^2_{\text{loc}}[0, L)$ and hence $f'' \in \mathbb{L}^2_{\text{loc}}[0, L)$ and so it follows that $f \in \mathbb{H}_2$ in this case as well.

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

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

Then, by the definition (4.19) of Φ_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 $\Phi_t f$ is continuously differentiable, has an absolutely continuous derivative and elementary calculations show that

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

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

$$\begin{aligned}\|(\Phi_t f)\|_{\mathbb{H}_2}^2 &\leq 10C^2\biggl(\int_0^\infty [f(x+t)^2 + f'(x+t)^2 + f''(x+t)^2](1+x^2)^2\,dx\biggr)\\ &= 10C^2\biggl(\int_t^\infty [f(x)^2 + f'(x)^2 + f''(x)^2]\bigl(1+(x-t)^2\bigr)^2\,dx\biggr)\\ &\leq 10C^2\|f\|_{\mathbb{H}_2}^2.\end{aligned}$$

This shows that for any $t > 0$, $\Phi_t f \in \mathbb{H}_2$ and the map from $\mathbb{H}_2$ to $\mathbb{H}_2$ that takes f to $\Phi_t f$ is Lipschitz continuous (with constant $\sqrt{10}C$). This, in turn, trivially implies that for $\nu \in \mathbb{H}_{-2}$, the linear functional on $\mathbb{H}_2$ given by $\mathcal{J}_t^\nu : f \mapsto \nu(\Phi_t f)$ also lies in $\mathbb{H}_{-2}$ and that the map from $\mathbb{H}_{-2}$ to itself that takes ν to $\mathcal{J}^\nu$ is also Lipschitz continuous with the same constant. Indeed,

$$\begin{aligned}\|\mathcal{J}_t^\nu\|_{\mathbb{H}_{-2}} &= \sup_{f:\|f\|_{\mathbb{H}_2}\leq 1} |\mathcal{J}_t^\nu(f)| = \sup_{f:\|f\|_{\mathbb{H}_2}\leq 1} |\nu(\Phi_t f)|\\ &\leq \sup_{f:\|f\|_{\mathbb{H}_2}\leq 1} \|\nu\|_{\mathbb{H}_{-2}}\|\Phi_t f\|_{\mathbb{H}_2} \leq \sqrt{10}C\|\nu\|_{\mathbb{H}_{-2}}.\end{aligned}$$

This completes the proof of the lemma. □

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 $\tilde{\Omega}$, $\overline{A}^{(N)}(t)$, $\overline{D}^{(N)}(t)$, $\overline{Q}_{\mathbf{1}}^{(N)}(t)$ and $\overline{K}^{(N)}(t)$ are finite for all $t \in [0,\infty)$. Fix $\omega \in \tilde{\Omega}$, and let γ and $h\widehat{\nu}^{(N)}$ be the linear functionals on $\mathbb{C}_c([0,L) \times [0,\infty))$ defined, respectively, by

$$\begin{aligned}\gamma(\varphi) &\doteq \int_{[0,L)} \varphi(x,0)\widehat{\nu}_0^{(N)}(dx) - \iint_{[0,L)\times[0,\infty)} \varphi(x,s)\widehat{\mathcal{M}}^{(N)}(dx,ds)\\ &\quad + \int_{[0,\infty)} \varphi(0,s)\,d\widehat{K}^{(N)}(s)\end{aligned}$$

and

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

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

$$|\gamma(\varphi)| \leq \sqrt{N}\|\varphi\|_\infty\bigl(2 + \overline{D}^{(N)}(t) + \overline{A}_{\mathbf{1}}^{(N)}(t) + \overline{K}^{(N)}(t) + \overline{K}(t)\bigr)$$

and, likewise, it can be argued that

$$\bigl|h\widehat{\nu}^{(N)}(\varphi)\bigr| \leq \sqrt{N}\|\varphi\|_\infty\bigl(\overline{A}_{\mathbf{1}}^{(N)}(t) + \overline{A}_{\mathbf{1}}(t)\bigr).$$

This shows that γ and $h\widehat{\nu}^{(N)}$ define Radon measures on $[0, L) \times [0, \infty)$. Now, for every $\varphi \in \mathbb{C}_c^{1,1}([0, L) \times [0, \infty))$, sending $t \to \infty$ in (6.5), the left-hand side of (6.5) vanishes because φ has compact support, and we obtain

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

Since $\{\widehat{\nu}_t^{(N)}, t \geq 0\} \in \mathbb{D}_{\mathcal{M}_F[0,L)}[0, \infty)$, the last equation shows that $\{\widehat{\nu}_t^{(N)}, t \geq 0\}$ satisfies the so-called abstract age equation for γ introduced in Definition 4.9 of [21]. Therefore, by Corollary 4.17 and (4.24) of [21], for every $f \in \mathbb{C}_c[0, L)$, $\langle f, \widehat{\nu}_t^{(N)}\rangle = \gamma(\varphi_t^f)$, $t \geq 0$, where

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

and ψ_h is the function defined in (4.55) of [21], and introduced as (9.21) in the present paper. Elementary algebra [specifically combining the relations in (9.23) with the definition (4.19) of Ψ_t] then shows that $\varphi_t^f(x, s) = \Psi_t f(x, s)$.

Using the definition of γ given above together with the relations $(\Psi_t f)(\cdot, 0) = \Phi_t f$, $(\Psi_t f)(0, \cdot) = f(t - \cdot)(1 - G(t - \cdot))$, $\widehat{\mathcal{H}}_t^{(N)}(f) = \widehat{\mathcal{M}}_t^{(N)}(\Psi_t f)$ and the definition (6.11) of $\widehat{\mathcal{K}}^{(N)}$, for $f \in \mathbb{C}_c[0, L)$, it can be shown that $\gamma(\Psi_t f)$ is equal to the right-hand side of representation (6.15). This establishes (6.15) for $f \in \mathbb{C}_c[0, L)$. A standard approximation argument can then be used to show that representation (6.15) holds for all $f \in \mathbb{C}_b[0, L)$.

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 $T < \infty$. For brevity, let the state process be represented by $\overline{Y}^{(N)}(s) = (R_E^{(N)}(s), \overline{X}^{(N)}(s), \overline{\nu}_s^{(N)})$, $s \in [0, \infty)$. Recall from (4.3) that $\overline{M}_{\mathbf{1}}^{(N)} = \overline{D}^{(N)} - \overline{A}_{\mathbf{1}}^{(N)}$ is a martingale. Therefore, taking expectations of both sides of inequality (5.30) of [21], with t and δ replaced by 0 and T, respectively, it follows that

$$\mathbb{E}_{\overline{Y}^{(N)}(0)}\bigl[\overline{A}_{\mathbf{1}}^{(N)}(T)\bigr] = \mathbb{E}_{\overline{Y}^{(N)}(0)}\bigl[\overline{D}^{(N)}(T)\bigr] \leq U(T), \tag{D.1}$$

where U is the renewal function associated with G. This shows that 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

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

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

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

For each $s_1 \in [0, T]$, due to the Markov property of $\overline{Y}^{(N)}$ established in Lemma B.1 of [19], we obtain

$$\begin{aligned}&\mathbb{E}_{\overline{Y}^{(N)}(0)}[\langle h, \overline{\nu}_{s_1}^{(N)}\rangle(\overline{A}_{\mathbf{1}}^{(N)}(T) - \overline{A}_{\mathbf{1}}^{(N)}(s_1))^{j-1}] \\ &\quad = \mathbb{E}_{\overline{Y}^{(N)}(0)}[\mathbb{E}_{\overline{Y}^{(N)}(0)}[\langle h, \overline{\nu}_{s_1}^{(N)}\rangle(\overline{A}_{\mathbf{1}}^{(N)}(T) - \overline{A}_{\mathbf{1}}^{(N)}(s_1))^{j-1}|\mathcal{F}_{s_1}^{(N)}]] \\ &\quad = \mathbb{E}_{\overline{Y}^{(N)}(0)}[\langle h, \overline{\nu}_{s_1}^{(N)}\rangle\mathbb{E}_{\overline{Y}^{(N)}(s_1)}[(\overline{A}_{\mathbf{1}}^{(N)}(T - s_1))^{j-1}]].\end{aligned}$$

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

$$\begin{aligned}&\mathbb{E}_{\overline{Y}^{(N)}(0)}[\langle h, \overline{\nu}_{s_1}^{(N)}\rangle(\overline{A}_{\mathbf{1}}^{(N)}(T) - \overline{A}_{\mathbf{1}}^{(N)}(s_1))^{j-1}] \\ &\quad \leq (j-1)!U(T)^{j-1}\mathbb{E}_{\overline{Y}^{(N)}(0)}[\langle h, \overline{\nu}_{s_1}^{(N)}\rangle].\end{aligned}$$

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

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

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. Recall that given $\varphi \in \mathbb{C}[[0, L) \times [0, T])$, φ^* is the function defined by $\varphi^*(x) \doteq \sup_{s\in[0,T]} \varphi(x, s)$. 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 [21], for every $N \in \mathbb{N}$, we have

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

where

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

which is finite by Theorem 1. 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 $\overline{A}_{\mathbf{1}}^{(N)} \Rightarrow \overline{A}_{\mathbf{1}}$ by Proposition 5.17 of [21] and $\overline{A}_{\mathbf{1}}$ is continuous. Together with the Skorokhod representation theorem, Fatou's lemma and the inequality (D.1), this implies that

$$\overline{A}_{\mathbf{1}}(T) \le \liminf_{N\to\infty} \mathbb{E}\big[\overline{A}_{\mathbf{1}}^{(N)}(T)\big] \le \limsup_{N\to\infty} \mathbb{E}\big[\overline{A}_{\mathbf{1}}^{(N)}(T)\big] \le U(T).$$

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

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 $\tilde{\varphi} \in \mathbb{C}_b([0,L) \times [0,\infty))$. Then, almost surely, for every $t \ge 0$, we have*

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

where Υ_t, $t \ge 0$, is the family of mappings defined in (8.27). Moreover, if for every $T < \infty$, $x \mapsto \int_0^t \tilde{\varphi}(x,r)\,dr$ is bounded and Hölder continuous, uniformly in $t \in [0,T]$, then for every $s \ge 0$, almost surely for every $t \ge 0$,

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

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 $\tilde{\varphi} \in \mathbb{C}_b([0,L) \times [0,\infty))$, then almost surely, for every $s,t \ge 0$,

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

PROOF. Fix $s, t \geq 0$. Then $\int_0^t \Psi_s(\tilde{\varphi}(\cdot, r))\,dr = \Psi_s(\int_0^t \tilde{\varphi}(\cdot, r)\,dr)$ and so, by inequality (4.20) and the boundedness assumption on φ, $\int_0^t \Psi_s(\tilde{\varphi}(\cdot, r))\,dr$ and $\Upsilon_t \tilde{\varphi}$ 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 [31]) to conclude that almost surely, (E.2) and (E.1) are satisfied. The processes on the right-hand sides of (E.2) and (E.1) are clearly continuous in t, whereas the continuity of the processes on the left-hand sides of (E.2) and (E.1) follows from properties 1 and 4 of Lemma 8.6. Thus, there exists a set of full $\widehat{\mathbb{P}}$-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

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

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 2, $\widehat{K}$ 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

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

which completes the proof of the lemma. □

We now prove the consistency lemma.

PROOF OF LEMMA 9.3. Fix $f \in \mathcal{S}$ and $s, t \geq 0$. Then, replacing t by $t + s$ in (6.15), we obtain

$$\widehat{\nu}^{(N)}_{s+t}(f) = \mathcal{J}^{\widehat{\nu}^{(N)}_0}_{s+t}(f) - \widehat{\mathcal{H}}^{(N)}_{t+s}(f) + \widehat{\mathcal{K}}^{(N)}_{t+s}(f). \tag{E.4}$$

Using the shift relations introduced in (9.10)–(9.12), 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.4) can be decomposed as follows:

$$\widehat{\mathcal{H}}^{(N)}_{t+s}(f) = \widehat{\mathcal{M}}^{(N)}_{t+s}(\Psi_{t+s} f) = \widehat{\mathcal{M}}^{(N)}_s(\Psi_{t+s}(f)) + (\Theta_s \widehat{\mathcal{M}}^{(N)})(\Psi_{t+s} f) \tag{E.5}$$

and, similarly,

$$\begin{aligned}\widehat{\mathcal{K}}_{t+s}^{(N)}(f) &= \int_{[0,s+t]} f(s+t-u)\big(1-G(s+t-u)\big)\,d\widehat{K}^{(N)}(u)\\ &= \int_{[0,s]} f(s+t-u)\big(1-G(s+t-u)\big)\,d\widehat{K}^{(N)}(u)\\ &\quad + \int_{[0,t]} \big(1-G(t-u)\big) f(t-u)\,d\big(\Theta_s \widehat{K}^{(N)}\big)(u).\end{aligned} \tag{E.6}$$

On the other hand, since $\Phi_t f \in \mathbb{C}_b[0,L)$, 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

$$\begin{aligned}\mathcal{J}_t^{\widehat{\nu}_s^{(N)}}(f) &= \langle \Phi_t f, \widehat{\nu}_s^{(N)}\rangle\\ &= \langle \Phi_{s+t} f, \widehat{\nu}_0^{(N)}\rangle - \widehat{\mathcal{M}}_s^{(N)}(\Psi_s \Phi_t f)\\ &\quad + \int_{[0,s]} (\Phi_t f)(s-u)\big(1-G(s-u)\big)\,d\widehat{K}^{(N)}(u)\\ &= \mathcal{J}_{s+t}^{\widehat{\nu}_0^{(N)}}(f) - \widehat{\mathcal{M}}_s^{(N)}(\Psi_{s+t} f)\\ &\quad + \int_{[0,s]} f(s+t-u)\big(1-G(s+t-u)\big)\,d\widehat{K}^{(N)}(u).\end{aligned} \tag{E.7}$$

Relation (9.14) is then obtained by subtracting (E.7) from (E.4), rearranging terms and using the relations (E.6) and (E.5).

Now, suppose that Assumptions 1–4 are satisfied and further, assume that g is continuous. Then Theorem 3 shows that the limit $\widehat{\nu}$ of $\{\widehat{\nu}^{(N)}\}_{N\in\mathbb{N}}$ is a continuous $\mathbb{H}_{-2}$-valued process that is given explicitly by (5.24). The shifted equation (9.15) 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,

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

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 Φ_t and the fact that $(\Phi_t f)(0) = f(t)(1-G(t))$, we obtain

$$\begin{aligned}\mathcal{J}_t^{\widehat{\nu}_s}(f) &= \widehat{\nu}_s(\Phi_t f)\\ &= \mathcal{J}_{s+t}^{\widehat{\nu}_0}(f) - \widehat{\mathcal{M}}_s(\Psi_s \Phi_t f) + f(t)\big(1-G(t)\big)\widehat{K}(s)\\ &\quad + \int_0^s \widehat{K}(u)\xi_{\Phi_t f}(s-u)\,du.\end{aligned} \tag{E.8}$$

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

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

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

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

Equation (9.15) can now be obtained by combining the last four equations with the limit analog of (E.5), 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 (9.16) is satisfied, note that by Theorem 2, $(\widehat{K}, \widehat{X}, \widehat{\nu}_0(\mathbf{1})) = \Lambda(\widehat{E}, \widehat{x}_0, \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}))$. This implies that the centered many-server equations (5.14)–(5.16) are satisfied with v, Z, X, K and E, respectively, replaced by $\widehat{\nu}(\mathbf{1}), \mathcal{J}^{\widehat{\nu}_0}(\mathbf{1}) - \widehat{\mathcal{H}}(\mathbf{1}), \widehat{X}, \widehat{K}$ and $\widehat{E}$. Fix any $s > 0$. Subtracting (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 $\Theta_s\widehat{K}, \Theta_s\widehat{E}, \widehat{X}_{s+\cdot}$ and $\widehat{\nu}_{s+\cdot}(\mathbf{1})$. 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 $t \geq 0$. Finally, substituting $f = \mathbf{1}$ in (9.15), using the definition (9.13) of $\Theta_s\widehat{\mathcal{K}}$ and the fact that $\xi_{\mathbf{1}} = -g$, it follows that (5.14) holds with v, Z and K, respectively, replaced, by $\widehat{\nu}_{s+\cdot}(\mathbf{1}), \mathcal{J}^{\widehat{\nu}_s}(\mathbf{1}) - \Theta_s\widehat{\mathcal{H}}(\mathbf{1})$ and $\Theta_s\widehat{K}$. This proves (9.16).

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 $\widehat{X}^{(N)}(0), \widehat{x}_0, \widehat{\nu}_0^{(N)}$, 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), $\widehat{\nu}_s$ is a random linear functional on the space of bounded and absolutely continuous functions. In addition, due to the continuity of the limit in the convergence (9.1) established in Proposition 9.1, it follows that

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

in the space $\mathbb{R} \times \mathcal{D}^3_{\mathbb{H}_{-2}}[0,\infty)$. 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 (9.14) and (9.15), $\mathcal{J}^{\widehat{\nu}_s^{(N)}}$ can be expressed as a linear combination of the $\mathbb{H}_{-2}$-valued processes $(\widehat{\nu}_{s+\cdot}^{(N)}, \Theta_s\widehat{\mathcal{K}}^{(N)}, \Theta_s\widehat{\mathcal{H}}^{(N)})$ and, likewise, $\mathcal{J}^{\widehat{\nu}_s}$ is the same linear combination of $(\widehat{\nu}_{s+\cdot}, \Theta_s\widehat{\mathcal{K}}, \Theta_s\widehat{\mathcal{H}})$. Therefore, the continuity of $\widehat{\nu}_{s+\cdot}, \Theta_s\widehat{\mathcal{K}}$ and $\Theta_s\widehat{\mathcal{H}}$ show that $\mathcal{J}^{\widehat{\nu}_s}$ is a continuous $\mathbb{H}_{-2}$-valued process. The same logic used above then shows that the real-valued process $\mathcal{J}^{\widehat{\nu}_s}(\mathbf{1})$ 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.

Now, for every $s \geq 0$, $\widehat{\nu}_s^{(N)}$ satisfies Assumption 5′(d) because it is a finite signed measure. Thus, it only remains to show that Assumption 5′(d) is satisfied when $\widehat{\nu}_0$

is replaced by $\widehat{\nu}_s$. Fix $\varphi \in \mathbb{C}_b([0, L) \times [0, \infty))$ such that $x \mapsto \varphi(x, r)$ is absolutely continuous and Hölder continuous, and φ_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

$$\widehat{\nu}_s(\Phi_r \varphi(\cdot, r)) = \widehat{\nu}_0(\Phi_{s+r} \varphi(\cdot, r)) - \widehat{\mathcal{M}}_s(\Psi_{s+r} \varphi(\cdot, r)) + \widehat{\mathcal{K}}_s(\Phi_r \varphi(\cdot, r)). \tag{E.9}$$

By (5.11), it follows that $\widehat{\nu}_0(\Phi_{s+r} \varphi(\cdot, r)) = \mathcal{J}_s^{\widehat{\nu}_0}(\Phi_r \varphi(\cdot, r))$. First, note that due to (E.9) and (5.24), the almost sure measurability of $s \mapsto \widehat{\nu}_s(\Phi_r \varphi(\cdot, r))$ follows from the assumed measurability of $s \mapsto \mathcal{J}_s^{\widehat{\nu}_0}(f)$ 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 $f \in \mathbb{C}_b[0, L)$, which is a consequence of the definition of these stochastic integrals. Essentially the same argument shows that for $f \in \mathbb{AC}_b[0, L)$ and $s > 0$, almost surely $t \mapsto \mathcal{J}_t^{\widehat{\nu}_s}(f)$ is measurable. 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-type relations in (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 $\widehat{\nu}_s(\int_0^t \Phi_r \varphi(\cdot, r)\, dr)$ is equal to

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

A comparison with (E.9) 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 $\widehat{\nu}_s$. □

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] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd ed. Springer, New York.

[2] BEREZANSKII, Y. M. (1986). *Selfadjoint Operators in Spaces of Functions of Infinitely Many Variables*. *Translations of Mathematical Monographs* **63**. Amer. Math. Soc., Providence, RI.

[3] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[4] 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** 36–50.

[5] BURTON, B. (2005). *Volterra Integral and Differential Equations*. Elsevier, Amsterdam.

[6] CENTSOV, N. N. (1956). Wiener random fields depending on several parameters. *Dokl. Akad. Nauk., SSSR* **106** 607–609.

[7] DECREUSEFOND, L. and MOYAL, P. (2008). A functional central limit theorem for the $M/GI/\infty$ queue. *Ann. Appl. Probab.* **18** 2156–2178.

[8] ERLANG, A. K. (1948). On the rational determination of the number of circuits. In *The Life and Works of A. K. Erlang* (E. Brockmeyer, H. L. Halstrom and A. Jensen, eds.). The Copenhagen Telephone Company, Copenhagen.

[9] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

[10] EVANS, L. C. (1988). *Partial Differential Equations. Graduate Studies in Mathematics* **19**. Amer. Math. Soc., Providence, RI.

[11] FRIEDMAN, E. (2006). *Stochastic Differential Equations and Applications*. Dover, New York.

[12] GAMARNIK, D. and MOMČILOVIĆ, P. (2008). Steady-state analysis of a multiserver queue in the Halfin–Whitt regime. *Adv. in Appl. Probab.* **40** 548–577.

[13] GLYNN, P. and WHITT, W. (1991). A new view of the heavy-traffic limit theorem for infinite-server queues. *Adv. in Appl. Probab.* **2** 188–209.

[14] HALFIN, S. and WHITT, W. (1981). Heavy-traffic limit theorems for queues with many servers. *Oper. Res.* **29** 567–588.

[15] IGLEHART, D. L. (1965). Limit diffusion approximations for the many server queue and the repairman problem. *J. Appl. Probab.* **2** 429–441.

[16] JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Berlin.

[17] JAGERMAN, D. (1974). Some properties of Erlang loss function. *Bell System Techn. J.* **53** 525–551.

[18] JELENKOVIC, P., MANDELBAUM, A. and MOMČILOVIĆ, P. (2004). Heavy traffic limits for queues with many deterministic servers. *Queueing Syst.* **47** 53–69.

[19] KANG, W. N. and RAMANAN, K. (2010). Fluid limits of many-server queues with reneging. *Ann. Appl. Probab.* **20** 2204–2260.

[20] KANG, W. N. and RAMANAN, K. (2012). Asymptotic approximations for stationary distributions of many-server queues with abandonment. *Ann. Appl. Probab.* **22** 477–521.

[21] KASPI, H. and RAMANAN, K. (2011). Law of large numbers limits for many-server queues. *Ann. Appl. Probab.* **21** 33–114.

[22] 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.

[23] MITOMA, I. (1983). On the sample continuity of $\mathcal{S}'$-processes. *J. Math. Soc. Japan* **35** 629–636.

[24] MITOMA, I. (1983). Tightness of probabilities on $\mathcal{C}([0,1], \mathcal{S}')$ and $\mathcal{D}([0,1], \mathcal{S})$. *Ann. Probab.* **11** 989–999.

[25] PUHALSKII, A. A. and REED, J. (2010). On many servers queues in heavy traffic. *Ann. Appl. Probab.* **20** 129–195.

[26] PUHALSKII, A. A. and REIMAN, M. (2000). The multiclass GI/PH/N queue in the Halfin–Whitt regime. *Adv. in Appl. Probab.* **32** 564–595.

[27] REED, J. (2009). The G/GI/N queue in the Halfin–Whitt regime. *Ann. Appl. Probab.* **19** 2211–2269.

[28] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, Berlin.

[29] ROGERS, D. and WILLIAMS, D. (1994). *Diffusion, Markov Processes and Martingales, Volume 2: Foundations*, 2nd ed. Wiley, New York.

[30] SHARPE, M. (1988). *General Theory of Markov Processes*. Academic Press, San Diego.

[31] WALSH, J. B. (1986). An introduction to stochastic partial differential equations. In *École d'Été Probabilités de Saint-Flour XIV*. *Lecture Notes in Math*. **1180** 265–439. Springer, Berlin.

[32] 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: iehaya@techunix.technion.ac.il

DIVISION OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912
USA
E-MAIL: Kavita_Ramanan@brown.edu