The Fluid Limit of a Heavily Loaded Processor Sharing Queue

H. Christian Gromoll
EURANDOM
Den Dolech 2
5612 AZ Eindhoven
The Netherlands
gromoll@eurandom.tue.nl

Amber L. Puha*
Department of Mathematics
California State University, San Marcos
San Marcos, CA 92096–0001
apuha@csusm.edu

Ruth J. Williams†
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112
williams@math.ucsd.edu

November 27, 2001

Short title: Fluid Limit of a Processor Sharing Queue

AMS 2000 Subject Classifications: Primary 60K25, Secondary 68M20, 90B22.

Key words and phrases: Processor sharing queue, heavy traffic, fluid model, renewal equation, measure valued process.

*Research supported in part by a University of California Office of the President Postdoctoral Fellowship and an NSF Mathematical Sciences Postdoctoral Fellowship.

†Research supported in part by NSF grants DMS 9703891 & 0071408, and a gift from the David and Holly Mendel fund.

Abstract

Consider a single server queue with renewal arrivals and i.i.d. service times in which the server operates under a processor sharing service discipline. To describe the evolution of this system, we use a measure valued process that keeps track of the residual service times of all jobs in the system at any given time. From this measure valued process, one can recover the traditional performance processes, including queue length and workload. We propose and study a critical fluid model (or formal law of large numbers approximation) for a heavily loaded processor sharing queue. The fluid model state descriptor is a measure valued function whose dynamics are governed by a nonlinear integral equation. Under mild assumptions, we prove existence and uniqueness of fluid model solutions. Furthermore, we justify the critical fluid model as a first order approximation of a heavily loaded processor sharing queue by showing that, when appropriately rescaled, the measure valued processes corresponding to a sequence of heavily loaded processor sharing queues converge in distribution to a limit that is almost surely a fluid model solution.

Contents

1 Introduction

Consider a single server with an infinite capacity buffer to which jobs arrive according to a delayed renewal process. The i^{th} such arrival requires an amount of processing time that is the i^{th} member of a sequence of independent and identically distributed strictly positive random variables. The server, rather than providing service to just one job at a time, operates under a processor sharing discipline, i.e., it works simultaneously on all jobs currently in the system, providing an equal fraction of its attention to each. Thus, at any given time that the system is nonempty, each job in the system is being processed at a rate that is the reciprocal of the number of jobs in the system. When the server has fulfilled a given job's service requirement, the job exits the system. This system is known as a processor sharing queue.

The processor sharing service discipline can be viewed as an idealization of a round-robin or time-sharing protocol used in computer and communication systems. Although there is a considerable literature on processor sharing queues (see Yashkov [25] for a survey up through 1987), much of that work assumes either Poisson arrivals or exponential service times. Indeed, there are few papers on the GI/GI/1 processor sharing queue. The most general results known to us are those of Jean-Marie and Robert [20] for transient (overloaded) queues, Baccelli and Towsley [1] who establish a positive correlation type of property for the delays incurred by customers, Grishechkin [18] who obtains asymptotics of the steady state distribution as the traffic intensity approaches one, and Chen, Kella and Weiss [9] who consider a fluid approximation for the queue length process. In particular, there is no heavy traffic diffusion approximation for the queue length process of a GI/GI/1 processor sharing queue. (Note that since the workload process in a GI/GI/1 queue is the same for all nonidling service disciplines, the heavy traffic approximation for the workload process under a processor sharing service discipline is the same as the well known approximation under a FIFO (first-in-first-out) service discipline [19]. However, this simple relationship does not hold for the queue length process.)

In this paper, we study the fluid (or law of large numbers) approximation for a measure valued process that keeps track of all of the residual service times of the jobs in a heavily loaded processor sharing queue. Our main motivation for studying this so-called critical fluid model is our conjecture (to be studied in subsequent work) that an understanding of the asymptotic behavior of solutions of this model is likely to play a key role in establishing a heavy traffic diffusion approximation for the queue length process in a processor sharing queue. This conjecture represents an extrapolation of the known situation for open multiclass networks with a HL (head-of-the-line) service discipline, which we briefly summarize here. Recently, Bramson [7] showed that if the critical fluid model for an open multiclass HL-network has a certain asymptotic property, then state space collapse holds. Roughly speaking, state space collapse implies that in the heavy traffic limit, the queue length process can be recovered from the workload process by an appropriate lifting map. In a companion work to [7], Williams [24] showed that state space collapse, plus an algebraic condition on the queueing model data, is sufficient to imply a heavy traffic diffusion approximation for an open multiclass network with a HL service discipline. To illustrate this modular approach, Bramson [7] and Williams [24] applied their results to obtain new heavy traffic limit theorems for FIFO networks of Kelly type and for networks with a HLPPS (head-of-the-line proportional processor sharing) service discipline. Processor sharing, as considered in this paper, is not a HL service discipline. However, we seek to mimic the modular approach of [7, 24] for a processor sharing queue. The study of a suitable critical fluid model is the first step in this process.

The work presented in [7] indicates that it is important to use a sufficiently detailed state descriptor in formulating a fluid model. For example, in FIFO networks, one needs to keep track of the order in which customers arrive to the queue. In a processor sharing queue, each job in the system at time t has an associated residual service time. This is given by the total amount of service originally requested by the job minus the total amount of service it has received up to time t. It is necessary to keep track of all of the residual service times in order to adequately describe the state of the system at any given time. In fact, it is natural to use a measure valued process $\{\mu(t) : t \geq 0\}$ such that the measure $\mu(t)$ at time t is a finite, nonnegative Borel measure on $\mathbb{R}_+ = [0, \infty)$ that puts a unit of mass at the residual service time of each job in the system at time t. From such a measure valued process, one can recover information about the performance of the system. For example, the number of jobs in the system, or queue length, at time t is obtained by integrating $\mu(t)$ against the function that is identically one. This measure valued process will be the object of central interest in this paper, and we refer to it as the state descriptor. This terminology is slightly disingenuous since the measure is not necessarily a Markovian state descriptor. In particular, it does not include the residual interarrival time—the time remaining until the next job arrives to the system. Since the residual interarrival time will play no role in our analysis, we choose not to include it in the state descriptor. Section 2.3 contains a precise description of the state descriptor $\mu(\cdot)$. The process $\mu(\cdot)$ was previously used by Grishechkin [18], along with other measure valued descriptors, in his heavy traffic analysis of the steady state distribution of a processor sharing queue. More recently, Doytchinov, Lehoczky, and Shreve [13] used a measure valued descriptor in the context of a queueing system with deadlines.

Our critical fluid model has two parameters, $\alpha \in (0, \infty)$ and a Borel probability measure ν on $\mathbb{R}_+$ that has a finite first moment and that does not charge the origin. These parameters are limits of parameters in the queueing system, where α corresponds to the rate at which jobs arrive to the system and the probability measure ν corresponds to the distribution of the i.i.d. service times for those jobs. The qualifier *critical* refers to the fact that we are interested in the limiting regime where the service and arrival rates are equal, i.e.,

$$\alpha = \left(\int_{\mathbb{R}_+} x\nu(dx) \right)^{-1}.$$

The model is defined by considering a formal limit $\bar{\mu}(\cdot)$ of the measure valued state descriptor under law of large numbers scaling. This limit takes values in $\mathcal{M}_F$, the set of finite, nonnegative Borel measures on $\mathbb{R}_+$, which is endowed with the topology of weak convergence. For $\zeta \in \mathcal{M}_F$, the real valued projection of ζ associated with a bounded, real valued, Borel measurable function g defined on $\mathbb{R}_+$ is denoted by $\langle g, \zeta \rangle = \int_{\mathbb{R}_+} g(x)\zeta(dx)$. The fluid model equations describe the dynamics of the real valued projections of $\bar{\mu}(\cdot)$ over the class of functions

$$\mathcal{C} = \{g \in \mathbf{C}_b^1(\mathbb{R}_+) : g(0) = 0, g'(0) = 0\}.$$

Here, $\mathbf{C}_b^1(\mathbb{R}_+)$ denotes the space of once continuously differentiable real valued functions defined on $\mathbb{R}_+$ that, together with their first derivatives, are bounded on $\mathbb{R}_+$. The requirement that g and g' vanish at the origin is imposed to avoid singular behavior of $\langle g, \bar{\mu}(\cdot)\rangle$ associated with mass in the fluid model abruptly disappearing as it reaches the origin. The latter corresponds to jobs in the queueing system abruptly departing when residual service times reach zero. A fluid model solution associated with the critical data (α, ν) is a continuous function $\bar{\mu} : [0, \infty) \longrightarrow \mathcal{M}_F$ such that $\bar{\mu}(t)$ does not charge the origin for all $t \geq 0$, and which for each $g \in \mathcal{C}$ satisfies

$$\langle g, \bar{\mu}(t)\rangle = \langle g, \bar{\mu}(0)\rangle - \int_0^t \frac{\langle g', \bar{\mu}(s)\rangle}{\langle 1, \bar{\mu}(s)\rangle} ds + \alpha t \langle g, \nu \rangle,$$

for all t prior to the time that $\bar{\mu}(\cdot)$ first reaches the zero measure. Once $\bar{\mu}(\cdot)$ reaches the zero measure, it is identically equal to the zero measure thereafter. We will in fact show that under mild assumptions, a fluid model solution with $\bar{\mu}(0) \neq \mathbf{0}$ does not reach the zero measure in finite time. However, in formulating the notion of a fluid model solution, we have not assumed this a priori. (For a more detailed account, see Section 3.1.)

A different fluid approximation for a processor sharing queue, which focussed on approximating the queue length process, was proposed in [9]. The idea there was to fluid scale what is called the busy time equation (see equation (4) in [9]), and to pass to a limit to obtain equation (31) in [9]. However, passage to the limit in the proof of Theorem 16 in [9] seems to implicitly assume that all limit points are deterministic, or in other words, that all possible limit points are the same. Our approach in fact provides a proof that this assumption is valid under the mild conditions identified here (cf. Theorem 3.2).

This paper contains two main results, Theorems 3.1 and 3.2. Theorem 3.1 states that under mild assumptions on the initial condition, a fluid model solution exists and is unique. Theorem 3.2 says that, again under mild conditions, the measure valued state descriptors for a sequence of heavily loaded processor sharing queues under law of large numbers scaling converge in distribution to a measure valued process, which we refer to as the fluid limit. Moreover, sample paths of the fluid limit are almost surely fluid model solutions. Hence it is appropriate to regard the critical fluid model as a first order approximation of a heavily loaded processor sharing queue.

The paper is organized as follows. Section 2 is devoted to a precise description of the dynamics of a processor sharing queue. Two important quantities are introduced there, namely, the cumulative service process and the measure valued state descriptor. In addition, an equation that the state descriptor satisfies is presented in Section 2.3. This is used to motivate the definition of the critical fluid model, which is contained in Section 3. That section also contains the precise statements of the two main theorems, Theorems 3.1 and 3.2. Section 4 contains the proof of Theorem 3.1 and Section 5 contains the proof of Theorem 3.2. The proof of uniqueness for Theorem 3.1 and the proof of tightness, which is used to prove Theorem 3.2, benefited from some ideas in Chen, Kella and Weiss [9] (cf. Lemmas 4.4 and 5.4 below). The proof of tightness also benefited from some ideas of Grishechkin [18] (cf. the proof of Lemma 5.3 below).

The following notation will be used throughout the paper. Let $\mathbb{R}$ denote the set of real numbers. For $a, b \in \mathbb{R}$, we write $a \vee b$ for the maximum of a and b, $a \wedge b$ for the minimum of a and b, a^+ and a^- for the positive and negative parts of a respectively, $\lfloor a \rfloor$ for the largest integer less than or equal to a, and $\lceil a \rceil$ for the smallest integer greater than or equal to a. The nonnegative real numbers $[0, \infty)$ will be denoted by $\mathbb{R}_+$. For a function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$, let $\|g\|_\infty = \sup_{x \in \mathbb{R}_+} |g(x)|$ and $\|g\|_K = \sup_{x \in [0,K]} |g(x)|$ for each $K \geq 0$. We define the positive and negative parts of such a function g by $g^+(x) = g(x) \vee 0$ and $g^-(x) = (-g)(x) \vee 0$ for all $x \in \mathbb{R}_+$.

Recall that $\mathcal{M}_F$ is the set of finite, nonnegative Borel measures on $\mathbb{R}_+$. Consider $\zeta \in \mathcal{M}_F$ and a Borel measurable function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ which is integrable with respect to ζ. We define $\langle g, \zeta \rangle = \int_{\mathbb{R}_+} g(x)\zeta(dx)$. Our equations will involve expressions of the form $\int_{[a,\infty)} g(x-a)\zeta(dx)$, for $a > 0$. To ease notation throughout, we write this as $\langle g(\cdot - a), \zeta \rangle$, making the convention that such a g is always extended to be identically zero on $(-\infty, 0)$. As previously noted, $\mathcal{M}_F$ is endowed with the topology of weak convergence of measures, i.e., for $\zeta_n, \zeta \in \mathcal{M}_F$, $n = 1, 2, \ldots$, we have $\zeta_n \xrightarrow{\text{w}} \zeta$ if and only if $\langle g, \zeta_n \rangle \longrightarrow \langle g, \zeta \rangle$ as $n \to \infty$, for all $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ that are bounded and continuous. With this topology, $\mathcal{M}_F$ is a Polish space (cf. [22]). We denote the zero measure in $\mathcal{M}_F$ by $\mathbf{0}$ and the measure in $\mathcal{M}_F$ that puts one unit of mass at the point $x \in \mathbb{R}_+$ by δ_x.

For a set $B \subset \mathbb{R}_+$, we denote the indicator of the set B by 1_B. We also define the following real valued functions on $\mathbb{R}_+$: $\chi(x) = x$ for $x \in \mathbb{R}_+$, and $\varphi(x) = 1/x$ for $x \in (0, \infty)$ with $\varphi(0) = 0$. For a topological space A, denote by $\mathbf{C}_b(A)$ the set of continuous, bounded, real valued functions defined on A. In addition, for an interval $I \subset \mathbb{R}$, $\mathbf{C}_b^1(I)$ is the set of once continuously differentiable, real valued functions defined on I that together with their first derivatives are bounded on I. For $g \in \mathbf{C}_b^1(I)$ we write $g'(x) = \frac{d}{dx} g(x)$, $x \in I$.

We will use "$\Longrightarrow$" to denote convergence in distribution of random elements of a metric space. Following Billingsley [2], we will use $\mathbf{P}$ and $\mathbf{E}$ respectively to denote the probability measure and expectation operator associated with whatever space the relevant random element is defined on. All stochastic processes used in this paper will be assumed to have paths that are right continuous with finite left limits (r.c.l.l.). For a Polish space $\mathcal{S}$, we denote by $D([0, \infty), \mathcal{S})$ the space of r.c.l.l. functions from $[0, \infty)$ into $\mathcal{S}$, and we endow this space with the usual Skorohod J_1-topology (cf. [15]).

2 The Processor Sharing Queue

Here we describe the processor sharing queueing system more precisely. The primitive stochastic processes and initial condition for our model are introduced in Section 2.1. The system dynamics and performance processes are described in Section 2.2. Here an important quantity for the processor sharing queue is introduced, namely the cumulative service process. In Section 2.3, we introduce the measure valued state descriptor and a dynamic equation associated with its evolution in time. The state descriptor and the associated dynamic equation play a fundamental role in motivating the definition of the critical fluid model and in justifying it as a first order approximation of the heavily loaded processor sharing queue.

2.1 Primitive Processes and Initial Condition

The *exogenous arrival process* $E(\cdot)$ is a rate α delayed renewal process. The arrival rate α is assumed to be strictly positive and finite. Jump times of this process correspond to times at which jobs enter the system. This renewal process is defined from a sequence of *interarrival times* $\{u_i\}_{i=1}^{\infty}$, where u_1 denotes the time at which the first job to arrive after time zero enters the system and u_i, $i \geq 2$, denotes the time between the arrival of the $(i-1)^{st}$ and the i^{th} jobs to enter the system after time zero. Frequently, we will simply refer to the i^{th} job to enter the system after time zero as the i^{th} *arrival*. Thus, $U_i = \sum_{j=1}^{i} u_j$ is the time at which the i^{th} arrival enters the system, which is interpreted as zero if $i = 0$, and $E(t) = \sup\{i \geq 0 : U_i \leq t\}$ is the number of exogenous arrivals by time t. We assume that the sequence $\{u_i\}_{i=2}^{\infty}$ is an i.i.d. sequence of nonnegative random variables with $\mathbf{E}[u_2] = 1/\alpha < \infty$. The random variable u_1 is associated with an initial delay preceding the first arrival and is assumed to be strictly positive with finite mean and to be independent of $\{u_i\}_{i=2}^{\infty}$, but otherwise can have an arbitrary distribution. We refer to u_1 as the *initial residual interarrival time*. A typical situation in which this relaxed assumption on u_1 comes into play is when one would like to apply the results of this paper to a processor sharing queue that has been operating for some time in the past. Although the interarrival times of such a queue may be governed by an i.i.d. sequence, the time at which one begins to observe the system ($t = 0$) is arbitrary, and so in general does not coincide with the arrival of a job to the system. As a result, the residual interarrival time u_1 will in general have a different distribution from the subsequent interarrival times $\{u_i\}_{i=2}^{\infty}$ and, in particular, may depend on the initial state of the system.

The *service process*, $\{V(i), i = 1, 2, \ldots\}$, is such that $V(i)$ records the total amount of service required from the server by the first i arrivals. More precisely, $\{v_i\}_{i=1}^{\infty}$ denotes an i.i.d. sequence of strictly positive random variables with common distribution given by a Borel probability measure ν on $\mathbb{R}_+$. We interpret v_i as the amount of processing time that the i^{th} arrival requires from the server. The v_i's are known as the *service times* and ν is known as the *service time distribution.* Then, $V(i) = \sum_{j=1}^{i} v_j$, which is taken to be zero if $i = 0$. It is assumed that $v_1 > 0$ a.s. and $\mathbf{E}[v_1] < \infty$. In terms of ν, these assumptions are expressed by saying that ν does not charge the origin ($\nu(\{0\}) = 0$) and $\langle \chi, \nu \rangle < \infty$. Recall that $\chi(x) = x$ for all $x \in \mathbb{R}_+$.

The two processes $E(\cdot)$ and $V(\cdot)$ are called the *primitive processes*, since they provide the primitive stochastic inputs for the model. Note that $E(\cdot)$ and $V(\cdot)$ are not assumed to be independent of one another. An independence assumption is not necessary since our interest is in asymptotic behavior under fluid scaling where only laws of large numbers come into play.

The *initial condition* specifies $Z(0)$, the number of jobs present in the system at time zero, and the service requirement for each of these jobs. Here $Z(0)$ is assumed to be a nonnegative, integer valued random variable with finite mean. The service times for these jobs are taken to be the first $Z(0)$ elements of a sequence $\{\tilde{v}_j\}_{j=1}^{\infty}$ of strictly positive random variables. The job present in the system at time zero requiring $\tilde{v}_j$ units of service time will be referred to as the j^{th} *initial job.* It is assumed that the initial workload has a finite mean, i.e., that $\mathbf{E}\left[\sum_{j=1}^{Z(0)} \tilde{v}_j\right] < \infty$. The random variables $Z(0)$ and $\{\tilde{v}_j\}_{j=1}^{\infty}$ are not assumed to be independent of one another, nor are they assumed to be independent of the primitive processes.

2.2 Performance Processes and Descriptive Equations

As a processor sharing queue evolves in time, certain r.c.l.l. stochastic processes are used to track important measures of performance for the system such as queue length, workload, and idle time. Let $Z(t)$ denote the queue length at time t, which is the total number of jobs in the system at time t. Also, let $W(t)$ denote the (immediate) workload at time t, which is the total amount of time that the server must work in order to satisfy the remaining service requirement of each job present in the system at time t, ignoring future arrivals. Finally, let $Y(t)$ denote the cumulative amount of time that the server has been idle up to time t. The processes $W(\cdot)$, $Y(\cdot)$, and $Z(\cdot)$ are called *performance processes.* These processes satisfy a set of descriptive equations, which we now present.

We begin with the familiar equations for the *workload* $W(\cdot)$ and *idle time* $Y(\cdot)$ processes, which are valid for any nonidling service discipline, including processor sharing. For $t \geq 0$, we have

$$W(0) = \sum_{j=1}^{Z(0)} \tilde{v}_j, \tag{2.1}$$

$$W(t) = W(0) + V(E(t)) - t + Y(t), \tag{2.2}$$

$$Y(t) = \sup\{(W(0) + V(E(s)) - s)^- : 0 \leq s < t\}. \tag{2.3}$$

Recall that $E(t)$ is the total number of arrivals up to time t. Thus, $W(0) + V(E(t))$ is the total service time required by the initial jobs plus that required by the jobs arriving in $(0, t]$. Since the server completes $t - Y(t)$ units of work in $[0, t]$, (2.2) represents the remaining work in the system at time t.

A set of equations that describes the *queue length* process $Z(\cdot)$ under a processor sharing service discipline is the following. For $t \geq 0$,

$$Z(t) = Z(0) + E(t) - D(t), \tag{2.4}$$

$$D(t) = \sum_{j=1}^{Z(0)} 1_{\{\tilde{v}_j \leq S(t)\}} + \sum_{i=1}^{E(t)} 1_{\{v_i \leq S(t)-S(U_i)\}}, \tag{2.5}$$

$$S(t) = \int_0^t \varphi(Z(s))ds. \tag{2.6}$$

Recall that $\varphi(x) = 1/x$ if $x > 0$, and $\varphi(0) = 0$. The process $D(\cdot)$ is the *departure process*, where $D(t)$ represents the total number of jobs that have departed from the system by time t. The process $S(\cdot)$ is known as the *cumulative service process*, and $S(t)$ represents the cumulative amount of service time allocated per job up to time t.

The cumulative service process $S(\cdot)$ will play a particularly important role in our analysis. We will find it convenient to have notation for the increments of this process. For $t, h \geq 0$, define the *cumulative service per job in* $[t, t+h]$ by

$$S_{t,t+h} = S(t+h) - S(t) = \int_t^{t+h} \varphi(Z(s))ds. \tag{2.7}$$

Then the i^{th} arrival receives an amount of service equal to $v_i \wedge S_{U_i,t}$ by time t, for $t \geq U_i$. Define the *residual service times* at time $t \geq 0$ of the i^{th} arrival, $i \in \{1, \ldots, E(t)\}$, and of the j^{th} initial job, $j \in \{1, \ldots, Z(0)\}$, by

$$R_i(t) = (v_i - S_{U_i,t})^+ \qquad \text{and} \qquad \tilde{R}_j(t) = (\tilde{v}_j - S(t))^+, \tag{2.8}$$

respectively. The quantity $R_i(t)$ (resp. $\tilde{R}_j(t)$) represents the remaining amount of service time required by the i^{th} arrival (resp. j^{th} initial job) at time t. When a residual service time reaches zero, the associated job departs the system. Notice that $R_i(t)$ does not generally correspond to the amount of time that the i^{th} arrival will stay in the system beyond time t, since this job will receive service at less than full rate whenever there are other jobs in the system. The quantity $R_i(t)$ should rather be thought of as the amount of work for the system, measured in units of remaining required service time, embodied in the i^{th} arrival at time t. A similar interpretation holds for $\tilde{R}_j(t)$ as well. In fact, it can be shown that the workload at time $t \geq 0$ can be rewritten as

$$W(t) = \sum_{j=1}^{Z(0)} \tilde{R}_j(t) + \sum_{i=1}^{E(t)} R_i(t). \tag{2.9}$$

2.3 Measure Valued State Descriptor

A major difficulty in analyzing the performance of a processor sharing queue lies in the fact that it could experience large drops in its queue length process during an interval $[t, t+h]$, which cannot be predicted if the only information known about the state of the system at time t is the queue length $Z(t)$ and the workload $W(t)$. For instance, consider a processor sharing queue which, at a particular time t, has a large number of "low time" jobs in its buffer, all having very small residual service times, and a small number of "high time" jobs, all having very large residual service times. The workload $W(t)$ in this example could be made arbitrarily large, by increasing the residual service times of the high time jobs. The queue length $Z(t)$ is almost entirely embodied in the low time jobs, all of which will depart in the near future (say by time $t+h$) due to the simultaneous processing of jobs. In this situation, the imminent drop in the queue length during $[t, t+h]$ would not be evident from $Z(t)$ and $W(t)$, since both could be arbitrarily large. Thus, while these two processes are very useful measures of the overall performance of the system, they do not encode enough information about the state of the system to facilitate a proper analysis of its dynamics. We will require a richer description of the state of the system. As the above example illustrates, this description should include information about the residual service times of all jobs in the system at any given time. The use of such a state descriptor will be essential to our analysis.

An effective way to keep track of the residual service times, as well as the aforementioned performance processes and the initial condition, is to use a certain measure valued process which we call the *state descriptor*. For each $t \geq 0$, let $\mu(t)$ be the random, finite, Borel measure on $\mathbb{R}_+ = [0, \infty)$ given by

$$\mu(t) = \sum_{j=1}^{Z(0)} 1_{(0,\infty)}(\tilde{R}_j(t))\, \delta_{\tilde{R}_j(t)} + \sum_{i=1}^{E(t)} 1_{(0,\infty)}(R_i(t))\, \delta_{R_i(t)}. \tag{2.10}$$

Recall that δ_x is the measure that puts a single unit of mass at x for $x \in \mathbb{R}_+$. Thus, the random measure $\mu(t)$ has a unit of mass at the residual service time of each job that is in the system at time t, or in other words, for each $0 < a < b < \infty$, the measure that $\mu(t)$ assigns to the interval (a, b) is the number of residual service times that lie in (a, b) at time t. The indicator functions in the above definition serve to eliminate jobs with zero residual service times from the description of the system state, since such jobs have departed the system. The queue length and workload at time t can be obtained from $\mu(t)$ by integrating against an appropriate function. In particular, for $t \geq 0$,

$$Z(t) = \langle 1, \mu(t) \rangle \qquad \text{and} \qquad W(t) = \langle \chi, \mu(t) \rangle\,. \tag{2.11}$$

Recall that $\chi(x) = x$ for all $x \in \mathbb{R}_+$. Furthermore, given the primitive processes and the initial condition, one can recover $D(\cdot)$ and $S(\cdot)$ from $Z(\cdot)$, and $Y(\cdot)$ from $W(\cdot)$. Notice that the information given by the initial condition is described by the random initial measure $\mu(0)$. In particular, our assumptions on the initial condition imply that

$$\mathbf{E}\,[Z(0)] = \mathbf{E}\,[\langle 1, \mu(0) \rangle] < \infty \qquad \text{and} \qquad \mathbf{E}\,[W(0)] = \mathbf{E}\,[\langle \chi, \mu(0) \rangle] < \infty. \tag{2.12}$$

The assumptions on the initial condition together with the processor sharing dynamics imply that the random measure $\mu(t)$ takes values in the space $\mathcal{M}_F$ of finite, nonnegative Borel measures on $\mathbb{R}_+$. It is straightforward to see that $\mu(\cdot)$ is a measure valued stochastic process with sample paths in the Polish space $D\big([0,\infty), \mathcal{M}_F\big)$ of functions from $[0, \infty)$ into $\mathcal{M}_F$ that are right continuous with finite left limits. The space $D\big([0,\infty), \mathcal{M}_F\big)$ is endowed with the Skorohod J_1-topology (cf. [15]).

An equivalent formulation of (2.10) uses the real valued processes $\langle g, \mu(\cdot) \rangle$, for a suitable class of functions $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$. In fact, (2.10) holds for all $t \geq 0$ if and only if the following holds for each bounded, Borel measurable function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$,

$$\langle g, \mu(t) \rangle = \sum_{j=1}^{Z(0)} \big(1_{(0,\infty)} g\big)\,(\tilde{R}_j(t)) + \sum_{i=1}^{E(t)} \big(1_{(0,\infty)} g\big)\,(R_i(t)), \qquad t \geq 0. \tag{2.13}$$

This set of equations, or equivalently equation (2.10), will be the starting point for our analysis of processor sharing queues.

3 The Critical Fluid Model

The purpose of this section is twofold, to introduce the critical fluid model and to state our results. We begin with a precise description of a fluid model solution for critical data (α, ν) in Section 3.1. This is followed in Section 3.2 by the statement of Theorem 3.1, which gives existence and uniqueness of such fluid model solutions under mild conditions. Next our attention turns to justifying the critical fluid model as a first order approximation of a processor sharing queue operating in heavy traffic. In order to do that, we describe, in Section 3.3, a sequence of heavily loaded processor sharing queues under fluid scaling. This prepares us to state our fluid limit result, Theorem 3.2, in Section 3.4.

3.1 Definition of Fluid Model Solutions

The critical fluid model is formulated by considering a formal limit $\bar{\mu}(\cdot)$ of the measure valued state descriptors under law of large numbers scaling. The dynamics of $\bar{\mu}(\cdot)$ are prescribed through a set of equations satisfied by the real valued projections $\langle g, \bar{\mu}(t)\rangle$, $t \geq 0$, for a suitable class of functions g. To avoid singular behavior associated with the abrupt departure of mass at the origin, this class is chosen so that the functions together with their first derivatives, vanish at the origin. Specifically, we work with the class

$$\mathcal{C} = \{g \in \mathbf{C}_b^1(\mathbb{R}_+) : g(0) = 0, g'(0) = 0\}. \tag{3.1}$$

Motivated by the fact that the measure valued state descriptors do not charge the origin, we require the same of any fluid model solution (see (ii) below). Using this condition, we are able to prove existence and uniqueness of fluid model solutions under mild assumptions, despite the restriction to functions that vanish at the origin.

The critical fluid model depends on two parameters. The first is $\alpha \in (0, \infty)$, which corresponds to the renewal arrival rate to the queue. The second is a Borel probability measure ν on $\mathbb{R}_+$ satisfying $\nu(\{0\}) = 0$, which corresponds to the distribution of the service times of arrivals to the queue. It is assumed that

$$\alpha \langle \chi, \nu\rangle = 1, \tag{3.2}$$

which simply means that the arrival and service rates coincide, or equivalently that the model is critical. The pair (α, ν) is referred to as the data of the critical fluid model, or simply the *critical data.*

A *fluid model solution* for the critical data (α, ν) is a function $\bar{\mu} : [0, \infty) \longrightarrow \mathcal{M}_F$ such that

(i) $\bar{\mu}(\cdot)$ is continuous,

(ii) $\langle 1_{\{0\}}, \bar{\mu}(t)\rangle = 0$ for all $t \geq 0$,

(iii) for all $g \in \mathcal{C}$, $\bar{\mu}(\cdot)$ satisfies

$$\langle g, \bar{\mu}(t)\rangle = \langle g, \bar{\mu}(0)\rangle - \int_0^t \frac{\langle g', \bar{\mu}(s)\rangle}{\langle 1, \bar{\mu}(s)\rangle} ds + \alpha t \langle g, \nu\rangle, \tag{3.3}$$

for all $t < t^* = \inf\{s \geq 0 : \bar{\mu}(s) = \mathbf{0}\}$, and

(iv) for all $t \geq t^*$, $\bar{\mu}(t) = \mathbf{0}$.

The equations in (3.3) (one for each $g \in \mathcal{C}$) are called the *fluid model equations*.

Condition (i) is natural in light of Theorem 3.2 below, which implies that, under mild assumptions on the limiting initial condition, fluid limit points are a.s. continuous. Note that continuity of $\bar{\mu}(\cdot)$ is equivalent to continuity of $\langle g, \bar{\mu}(\cdot)\rangle$ for all $g \in \mathbf{C}_b(\mathbb{R}_+)$. The fact that the measures are precluded from charging the origin in condition (ii) stems from the fact that, at the level of the queueing system, zero residual service times correspond to jobs that have departed the system. The fluid model equations in condition (iii) will be derived from (2.13) by passing to the fluid limit. As we will see in Section 5.3, this requires some work. However, it is possible to provide an informal explanation of (3.3). We interpret the second and third terms on the right side of (3.3) as accounting for changes to the measure valued function $\bar{\mu}(\cdot)$ due to the fluid dynamics. On fluid scale, mass is being added to the system at a constant rate of α and is being distributed according to the service time probability measure ν. Thus, the third term describes changes resulting purely from arrivals. The second term describes changes resulting from the service dynamics. Recall that the server works at rate one at any given time that the system is nonempty. Since the server provides an equal fraction of its attention to each unit of mass in the system, the service rate per unit of mass in the system equals the reciprocal of the total mass in the system. Thus, the server shifts the mass towards the origin at rate $1/\langle 1, \bar{\mu}(s)\rangle$ at time s. Consequently, $\langle g(\cdot - h/\langle 1, \bar{\mu}(s)\rangle), \bar{\mu}(s)\rangle$ approximates the portion of $\langle g, \bar{\mu}(s+h)\rangle$ resulting from the service dynamics over the time interval $[s, s+h]$. In particular, $(\langle g(\cdot - h/\langle 1, \bar{\mu}(s)\rangle), \bar{\mu}(s)\rangle - \langle g(\cdot), \bar{\mu}(s)\rangle)/h$ approximates the average rate of change due to the service dynamics. Since, as h tends to zero, this difference quotient approaches the integrand of the second term on the right side of (3.3), this explains the form of that term. Condition (iv) simply reflects the fact that the data is critical, i.e., the arrival and service rates are balanced, and therefore mass shouldn't build up when starting from a zero initial measure.

We will find it convenient to refer to a fluid model solution for the critical data (α, ν) as simply a fluid model solution with the understanding that the data under consideration always satisfies (3.2). Given a fluid model solution $\bar{\mu}(\cdot)$, the fluid analogue of the queue length is defined by

$$\bar{Z}(t) = \langle 1, \bar{\mu}(t)\rangle, \qquad \text{for all } t \geq 0. \tag{3.4}$$

For obvious reasons, $\bar{Z}(t)$ is referred to as the *total mass* at time t. Due to the assumed continuity of fluid model solutions, $\bar{Z}(\cdot)$ is continuous. Also, the fluid analogue of the cumulative service per job is defined by

$$\bar{S}(t) = \int_0^t \varphi(\bar{Z}(s))ds, \qquad \text{for all } t \geq 0. \tag{3.5}$$

Since $\bar{Z}(\cdot)$ is strictly positive and continuous for $t \in [0, t^*)$, $\bar{S}(\cdot) \in \mathbf{C}^1([0, t^*))$ with

$$\bar{S}'(t) = \frac{d}{dt}\bar{S}(t) = \frac{1}{\bar{Z}(t)}, \qquad \text{for all } t \in [0, t^*). \tag{3.6}$$

Finally, the fluid analogue of the workload process is given by

$$\bar{W}(t) = \langle \chi, \bar{\mu}(t)\rangle, \qquad \text{for all } t \geq 0. \tag{3.7}$$

Since χ is not bounded, $\bar{W}(t)$ cannot be assumed to be continuous (or even finite) for an arbitrary fluid model solution. However, we will show that, under a mild assumption on the initial condition, $\bar{W}(\cdot)$ is in fact constant, although this constant equals infinity if $\langle \chi, \bar{\mu}(0)\rangle = \infty$ (cf. Theorem 3.1).

3.2 Existence and Uniqueness Result

Theorem 3.1 *Let $\xi \in \mathcal{M}_F$ be such that $\xi(\{x\}) = 0$ for all $x \in \mathbb{R}_+$. A fluid model solution $\bar{\mu}(\cdot)$ for the critical data (α, ν) with $\bar{\mu}(0) = \xi$ exists, is unique and satisfies $\bar{W}(t) = \langle \chi, \xi \rangle$ for all $t \geq 0$. In particular, if $\xi \neq \mathbf{0}$, then the associated fluid model solution never reaches the zero measure ($t^* = \infty$) and thus satisfies* (3.3) *for all time.*

Notice that Theorem 3.1 asserts that the fluid analogue of the workload process is constant and equals its initial value. This holds even if the initial measure ξ has an infinite first moment, in which case $\bar{W}(t) = \infty$ for all $t \geq 0$. The "no atoms" condition on ξ is readily explained by the assumed continuity of fluid model solutions. Indeed, if ξ were to have an atom at $x \in \mathbb{R}_+$, then since all of the mass initially at x departs the system simultaneously, the total mass would have a downward jump (i.e., a discontinuity) at that departure time. Section 4.1 contains the proof that there is at most one fluid model solution $\bar{\mu}(\cdot)$ such that $\bar{\mu}(0) = \xi$, that the workload is constant, and that $t^* = \infty$ if $\xi \neq \mathbf{0}$. Section 4.2 contains the proof of existence.

Remark. Theorem 3.1 and its proof can be extended in a straightforward manner to situations in which the fluid model data is either strictly subcritical ($\alpha\langle\chi,\nu\rangle < 1$), or strictly supercritical ($\alpha\langle\chi,\nu\rangle > 1$) and the initial measure $\xi \neq \mathbf{0}$. In particular, for strictly subcritical data ($\alpha\langle\chi,\nu\rangle < 1$) and $\xi \in \mathcal{M}_F$ that has no atoms, there is a unique $\bar{\mu} : [0,\infty) \longrightarrow \mathcal{M}_F$ satisfying $\bar{\mu}(0) = \xi$ and conditions (i)–(iv) of the definition of a fluid model solution. Moreover, $t^* = \langle\chi,\xi\rangle / (1 - \alpha\langle\chi,\nu\rangle)$ and

$$\bar{W}(t) = \bar{W}(0) + (\alpha\langle\chi,\nu\rangle - 1)t \quad \text{for } t \in [0,t^*), \qquad \bar{W}(t) = 0 \quad \text{for } t \in [t^*,\infty).$$

In the strictly supercritical case ($\alpha\langle\chi,\nu\rangle > 1$), condition (iv) is inappropriate since mass should build up from the zero initial measure. Indeed, in this case, for $\xi \in \mathcal{M}_F$ that has no atoms and satisfies $\xi \neq \mathbf{0}$, there is a unique solution $\bar{\mu} : [0,\infty) \longrightarrow \mathcal{M}_F$ satisfying conditions (i)–(iii) of the definition of a fluid model solution and $t^* = \infty$ there. The case $\alpha\langle\chi,\nu\rangle > 1$ and $\xi = \mathbf{0}$ is the only one for which the result and proof of Theorem 3.1 do not immediately generalize. In a separate paper, Puha, Stolyar, and Williams [23] formulate and analyze a strictly supercritical, measure valued fluid model. In particular, a natural fluid workload equation is added to equations (i)–(iii) of the definition of a fluid model solution, to govern how mass builds up from zero. With this additional condition, uniqueness is established for solutions of the supercritical fluid model starting from the zero measure.

3.3 A Sequence of Heavily Loaded Processor Sharing Queues

In this section, we specify the assumptions under which the fluid limit result will be proved. Consider a sequence of processor sharing queueing models indexed by r, where r increases to ∞ through a sequence in $(0,\infty)$. Each model in the sequence may be defined on a separate probability space, however we use $\mathbf{P}$ and $\mathbf{E}$ respectively for the probability and expectation operator on each of these spaces. The r^{th} model is defined as in Section 2, except that all accompanying parameters and processes have a superscript r appended to them. In particular, the performance processes associated with the r^{th} system are $W^r(\cdot)$, $Y^r(\cdot)$, and $Z^r(\cdot)$, and the state descriptor is $\mu^r(\cdot)$. Recall that (cf. (2.12)) for each $r > 0$ the initial condition $\mu^r(0)$ satisfies

$$\mathbf{E}[\langle 1, \mu^r(0)\rangle] < \infty \quad \text{and} \quad \mathbf{E}[\langle \chi, \mu^r(0)\rangle] < \infty. \tag{3.8}$$

The fluid limit result concerns the behavior of processor sharing queues on law of large numbers scale, or *fluid scale*. Accordingly, we define the fluid scaled processes

$$\bar{Z}^r(t) = \frac{1}{r}Z^r(rt) \tag{3.9}$$

$$\bar{W}^r(t) = \frac{1}{r}W^r(rt) \tag{3.10}$$

$$\bar{\mu}^r(t) = \frac{1}{r}\mu^r(rt) \tag{3.11}$$

$$\bar{E}^r(t) = \frac{1}{r}E^r(rt) \tag{3.12}$$

$$\bar{S}^r_{t,t+h} = S^r_{rt,r(t+h)} = \int_{rt}^{r(t+h)} \varphi(\langle 1, \mu^r(s)\rangle)ds = \int_t^{t+h} \varphi(\langle 1, \bar{\mu}^r(s)\rangle)ds, \tag{3.13}$$

for all $t \in [0, \infty)$, $h \geq 0$.

Let (α, ν) be critical data as in Section 3.1. In order to obtain convergence in distribution of the fluid scaled state descriptors $\bar{\mu}^r(\cdot)$ to a process that is a.s. a fluid model solution for the critical data (α, ν), we impose the following asymptotic assumptions on the sequence of processor sharing queues. For the primitive processes, assume that as $r \to \infty$,

$$\alpha^r \longrightarrow \alpha, \tag{3.14}$$

$$\nu^r \xrightarrow{\text{w}} \nu, \tag{3.15}$$

$$\langle \chi, \nu^r \rangle \longrightarrow \langle \chi, \nu \rangle, \tag{3.16}$$

$$\mathbf{E}[u_1^r]/r \longrightarrow 0, \tag{3.17}$$

$$\mathbf{E}[u_2^r; u_2^r > r] \longrightarrow 0. \tag{3.18}$$

Recall that by (3.2), $\alpha\langle \chi, \nu \rangle = 1$. Thus, assumptions (3.14) and (3.16) guarantee that as $r \to \infty$, the systems become heavily loaded. Indeed if we define the *traffic intensity parameter* $\rho^r = \alpha^r\langle \chi, \nu^r \rangle$, then the assumptions imply that $\rho^r \longrightarrow 1$ as $r \to \infty$. Assumption (3.17) implies that the initial residual interarrival time vanishes on fluid scale. We assume (3.18) in order to provide uniform control over the tail of the distribution of u_2^r, which is used to obtain a weak law of large numbers for a triangular array (cf. Lemma A.2). In applications, one often has a stronger condition than (3.18). For example, (3.18) holds if $\limsup_{r\to\infty} \mathbf{E}[(u_2^r)^{1+\epsilon}] < \infty$ for some $\epsilon > 0$.

For the fluid scaled initial measures, we assume that for some random measure Θ taking values in $\mathcal{M}_F$, we have

$$\left(\bar{\mu}^r(0), \langle \chi, \bar{\mu}^r(0)\rangle\right) \Longrightarrow \left(\Theta, \langle \chi, \Theta\rangle\right), \qquad \text{as } r \to \infty, \tag{3.19}$$

and that Θ satisfies

$$\mathbf{E}[\langle 1, \Theta\rangle] < \infty, \tag{3.20}$$

$$\mathbf{E}[\langle \chi, \Theta\rangle] < \infty, \tag{3.21}$$

$$\left\langle 1_{\{x\}}, \Theta \right\rangle = 0, \text{ for all } x \in \mathbb{R}_+, \text{a.s.} \tag{3.22}$$

In (3.19), we are assuming that the fluid scaled initial measures and their first moments converge jointly in distribution to some limit. Since, for any $g \in \mathbf{C}_b(\mathbb{R}_+)$, $\Psi_g : \mathcal{M}_F \longrightarrow \mathbb{R}$ defined by $\Psi_g(\zeta) = \langle g, \zeta\rangle$ is continuous, by the continuous mapping theorem (cf. [2], Thm. 5.1), $\langle g, \bar{\mu}^r(0)\rangle \Longrightarrow \langle g, \Theta\rangle$

as $r \to \infty$. The second component of (3.19) implies that the fluid scaled initial workload converges in distribution, i.e., $\bar{W}^r(0) \Longrightarrow \langle \chi, \Theta \rangle$ as $r \to \infty$. Assumptions (3.20) and (3.21) require the total mass and first moment of Θ to have finite expected values. Assumption (3.22) states that a.s. Θ has no atoms, which is used to prove tightness of $\{\bar{\mu}^r(\cdot)\}_{r>0}$ and to show that fluid limit points are continuous paths a.s. The "no atoms" assumption will be used in the equivalent form

$$\lim_{\kappa \downarrow 0} \mathbf{P} \left(\sup_{x \in \mathbb{R}_+} \left\langle 1_{[x,x+\kappa]}, \Theta \right\rangle < \frac{\epsilon}{4} \right) = 1, \ \text{ for all } \epsilon > 0. \tag{3.23}$$

The equivalence of (3.22) and (3.23) is proved in the Appendix (cf. Lemma A.1).

3.4 Fluid Limit Result

Theorem 3.2 *Consider a sequence of processor sharing queueing models as defined in Section 3.3, satisfying assumptions* (3.14)–(3.22). *Then the sequence of fluid scaled state descriptors* $\{\bar{\mu}^r(\cdot)\}$ *converges in distribution as* $r \to \infty$ *to a measure valued process* $\bar{\mu}^\star(\cdot)$ *taking values in* $\mathcal{M}_F$ *such that* $\bar{\mu}^\star(0)$ *is equal in distribution to* Θ *and almost surely* $\bar{\mu}^\star(\cdot)$ *is a fluid model solution for the critical data* (α, ν).

We refer to the limiting process $\bar{\mu}^\star(\cdot)$ as the *fluid limit* of the sequence $\{\bar{\mu}^r(\cdot)\}$ of fluid scaled state descriptors. Notice that, by Theorem 3.1, once an initial measure $\xi \in \mathcal{M}_F$ is specified, a fluid model solution for the critical data (α, ν) is a deterministic path taking values in $\mathcal{M}_F$. Thus we see by Theorem 3.2 that although the initial measure $\bar{\mu}^\star(0)$ of the limiting process may be random (cf. (3.19)), for each ω in some set of probability one, $\bar{\mu}^\star(t)(\omega)$ is determined for all $t > 0$ by the initial measure $\bar{\mu}^\star(0)(\omega)$. Theorem 3.2 is proved in Section 5.

4 Existence and Uniqueness of Fluid Model Solutions

This section contains the proof of Theorem 3.1 and a related result, Lemma 4.9. Before proceeding with the proofs, it will be convenient to introduce some notation. Recall that the critical data (α, ν) satisfies $\alpha \in (0, \infty)$, $\nu(\{0\}) = 0$, and $\alpha \langle \chi, \nu \rangle = 1$. Let F denote the cumulative distribution function associated with the probability measure ν. Since ν does not charge the origin, $F(0) = 0$. The cumulative distribution function F has associated with it an excess lifetime cumulative distribution function F_e, which is given by $F_e(x) = \alpha \int_0^x (1 - F(y))\, dy$, $x \in \mathbb{R}_+$. In particular, F_e has probability density function $f_e(x) = \alpha (1 - F(x))$, $x \in \mathbb{R}_+$. Here the fact that $\alpha \langle \chi, \nu \rangle = 1$ was used to simplify the form of the normalizing constant.

In Theorem 3.1, the initial measure $\xi \in \mathcal{M}_F$ is assumed to have no atoms. Denote the set of all such measures by

$$\mathcal{M}_F^c = \{\xi \in \mathcal{M}_F : \xi(\{x\}) = 0 \text{ for all } x \in \mathbb{R}_+\}.$$

Here c stands for continuous, which reflects the relationship between the "no atoms" condition and the continuity of fluid model solutions. If the initial measure $\xi = \mathbf{0}$, then clearly $\bar{\mu}(\cdot) \equiv \mathbf{0}$ is the unique fluid model solution such that $\bar{\mu}(0) = \xi$. Thus, the main interest is in proving existence and uniqueness for nonzero initial measures. For this, let

$$\mathcal{M}_F^{c,p} = \{\xi \in \mathcal{M}_F^c : \xi \neq \mathbf{0}\}.$$

Here p stands for positive. Given $\xi \in \mathcal{M}_F^{c,p}$, let

$$H_\xi(x) = \int_0^x \langle 1_{(y,\infty)}, \xi \rangle \, dy, \qquad x \in \mathbb{R}_+. \tag{4.1}$$

As we will see in Sections 4.1 and 4.2, H_ξ plays an important role in the proof of Theorem 3.1 for $\xi \in \mathcal{M}_F^{c,p}$. Since ξ has no atoms, the integrand in (4.1) is continuous. Thus, $H_\xi(\cdot)$ is continuously differentiable with

$$H'_\xi(x) = \langle 1_{(x,\infty)}, \xi \rangle, \qquad x \in \mathbb{R}_+. \tag{4.2}$$

The proof of Theorem 3.1 for $\xi \in \mathcal{M}_F^{c,p}$ takes advantage of the behavior of solutions to certain convolution equations. In order to prepare the reader, the facts that will be needed are briefly reviewed here. For a more detailed account, the reader is referred to [17], Section XI.1. Given a locally bounded Borel measurable function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ and a nondecreasing, right continuous function $U : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, let

$$(g * U)(u) = \int_{[0,u]} g(u-s) dU(s), \qquad u \geq 0.$$

Note that by convention the contribution to the above integral is $g(u)U(0)$ at $s = 0$ whenever $U(0) \neq 0$. For $\xi \in \mathcal{M}_F^{c,p}$, consider the convolution equations

$$M(u) = H_\xi(u) + (M * F_e)(u), \qquad u \geq 0, \tag{4.3}$$

$$N(u) = H'_\xi(u) + (N * F_e)(u), \qquad u \geq 0. \tag{4.4}$$

Since $H_\xi(\cdot)$ and $H'_\xi(\cdot)$ are locally bounded on $\mathbb{R}_+$, each equation has a unique bounded, Borel measurable solution on $[0,b]$ for each $b \in \mathbb{R}_+$. In fact, there is an explicit representation for these solutions in terms of the renewal function

$$U_e(u) = \sum_{i=0}^{\infty} \left(F_e^{*i}\right)(u), \qquad u \geq 0, \tag{4.5}$$

where $F_e^{*0}(\cdot) \equiv 1$ and $F_e^{*i}(\cdot) = \left(F_e^{*(i-1)} * F_e\right)(\cdot)$, $i \geq 1$. Indeed,

$$M(u) = (H_\xi * U_e)(u), \qquad u \geq 0, \tag{4.6}$$

is the unique locally bounded solution of (4.3), and

$$N(u) = \left(H'_\xi * U_e\right)(u), \qquad u \geq 0, \tag{4.7}$$

is the unique locally bounded solution of (4.4). Observe that $U_e \in \mathbf{C}(\mathbb{R}_+)$ since $F_e \in \mathbf{C}(\mathbb{R}_+)$. Also note that $(H_\xi * U_e)(0) = 0$. Since $\xi \neq \mathbf{0}$, it is easily verified that $H_\xi * U_e$ is strictly increasing and $\lim_{u\to\infty}(H_\xi * U_e)(u) = \infty$. Moreover, since $U_e \in \mathbf{C}(\mathbb{R}_+)$, $H_\xi \in \mathbf{C}^1(\mathbb{R}_+)$, and $H_\xi(0) = 0$, it follows that $H_\xi * U_e \in \mathbf{C}^1(\mathbb{R}_+)$ with $(H_\xi * U_e)'(u) = (H'_\xi * U_e)(u)$, $u \geq 0$.

We are now ready to proceed with the proofs. This section is organized in the following manner. Section 4.1 contains a proof that for $\xi \in \mathcal{M}_F^{c,p}$ there is at most one fluid model solution $\bar{\mu}(\cdot)$ such that $\bar{\mu}(0) = \xi$, and that the fluid analogue of the workload is constant for such a solution. Section 4.2 contains the proof that for $\xi \in \mathcal{M}_F^{c,p}$, there exists a fluid model solution $\bar{\mu}(\cdot)$ such that $\bar{\mu}(0) = \xi$. Finally, in Section 4.3, it is proved that the mapping from the initial measure to the fluid model solution that it determines is measurable. In fact, it will be shown that this map restricted to $\mathcal{M}_F^{c,p}$ is continuous. For a fluid model solution $\bar{\mu}(\cdot)$, the reader is reminded that the functions $\bar{Z}(\cdot)$, $\bar{S}(\cdot)$ and $\bar{W}(\cdot)$ are defined by (3.4), (3.5), and (3.7) respectively.

4.1 Proof of Uniqueness of Fluid Model Solutions

Fix $\xi \in \mathcal{M}_F^{c,p}$. Here, we will prove that there is at most one fluid model solution $\bar{\mu}(\cdot)$ with $\bar{\mu}(0) = \xi$ and that for such a solution the workload is constant. In order to do that, we will derive, for $t < t^*$, an expression for $\langle 1_{(0,w)}, \bar{\mu}(t)\rangle$, $w \in (0, \infty]$, in terms of $\bar{S}(\cdot)$ and ξ—see (4.14) below. This is used to reduce the problem of proving uniqueness of fluid model solutions to proving that $t^* = \infty$ and that $\bar{S}(\cdot)$ is uniquely determined by ξ. Then, we will utilize (4.14) below with $w = \infty$ and the assumption that $\bar{\mu}(t)(\{0\}) = 0$ for all $t \geq 0$ to obtain an equation for the total mass $\bar{Z}(t) = \langle 1, \bar{\mu}(t)\rangle$ of $\bar{\mu}(t)$ for $t < t^*$. A key aspect of the uniqueness proof will be to perform a time change in this equation, thereby reducing it to the convolution equation in (4.4) and relating the time change of $\bar{Z}(\cdot)$ to the solution of (4.4). Indeed, we will see that this relationship between the time changed $\bar{Z}(\cdot)$ and the unique solution of (4.4) forces $t^* = \infty$ and results in uniqueness of $\bar{S}(\cdot)$, given ξ.

Despite the intuitive nature of equation (4.14), it is not completely straightforward to derive it from (3.3). This is due to the fact that in (3.3) one can only consider continuously differentiable functions that together with their first derivatives vanish at the origin. In particular, the function $g = 1_{(0,w)}$ cannot be directly substituted into (3.3) and consequently must be approximated. If one were to simply substitute into (3.3) a sequence of functions in $\mathcal{C}$ that approximate $1_{(0,w)}$, one would be faced with the problem of controlling errors involving the first derivative, which can be large near the origin. In order to circumvent this difficultly, we will derive a version of (3.3) for functions that depend on both time and space—see (4.8) below. This yields a larger set of equations satisfied by any fluid model solution. In order to prove (4.14), we take $g \in \mathcal{C}$ and compose with a time-dependent spatial shift to create a time-dependent function for which the integrals in (4.8) involving the first partial derivatives cancel one another—see (4.15) and (4.16). The resulting equation (4.18) does not involve g'. Therefore, $g \in \mathcal{C}$ can increase to $1_{(0,w)}$ in (4.18) irrespective of the behavior of g' near the origin, and this yields (4.14).

We begin by deriving the version of (3.3) that holds for a class of functions of both time (denoted by $[0,\infty)$) and space (denoted by $\mathbb{R}_+$). The class will be a subset of $\mathbf{C}_b^1([0,\infty)\times\mathbb{R}_+)$, the set of once continuously differentiable functions defined on $[0,\infty)\times\mathbb{R}_+$ that together with their first partial derivatives, denoted by $f_s(s,x) = \frac{\partial}{\partial s} f(s,x)$ and $f_x(s,x) = \frac{\partial}{\partial x} f(s,x)$, are bounded. Notice that, if $\bar{\mu}(\cdot)$ is a fluid model solution such that $\bar{\mu}(0) = \xi$, then since $\xi \neq \mathbf{0}$ and $\bar{\mu}(\cdot)$ is continuous, $t^* > 0$.

Lemma 4.1 *If $\bar{\mu}(\cdot)$ is a fluid model solution such that $\bar{\mu}(0) = \xi$, then for all $f \in \mathbf{C}_b^1([0,\infty)\times\mathbb{R}_+)$ such that $f(\cdot,0) \equiv 0$ and $f_x(\cdot,0) \equiv 0$, $\bar{\mu}(\cdot)$ satisfies*

$$\langle f(t,\cdot), \bar{\mu}(t)\rangle = \langle f(0,\cdot), \xi\rangle + \int_0^t \langle f_s(s,\cdot), \bar{\mu}(s)\rangle\, ds - \int_0^t \frac{\langle f_x(s,\cdot), \bar{\mu}(s)\rangle}{\langle 1, \bar{\mu}(s)\rangle} ds + \alpha \int_0^t \langle f(s,\cdot), \nu\rangle\, ds, \tag{4.8}$$

for all $t < t^* = \inf\{s : \bar{\mu}(s) = \mathbf{0}\}$.

The next proposition will be used in the proof of Lemma 4.1.

Proposition 4.2 *Let $\bar{\mu} : [0,\infty) \longrightarrow \mathcal{M}_F$ be continuous. For each $f \in \mathbf{C}_b([0,\infty)\times\mathbb{R}_+)$,*

$$t \longrightarrow \langle f(t,\cdot), \bar{\mu}(t)\rangle$$

is a continuous function of $t \in [0,\infty)$.

Proof. Let $f \in \mathbf{C}_b([0,\infty)\times\mathbb{R}_+)$. For fixed $t \in [0,\infty)$ and $h \in (-\infty,\infty)$ such that $t+h \geq 0$,

$$\begin{aligned}&\Big|\langle f(t+h,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t)\rangle\Big| \\ &\qquad \leq \Big|\langle f(t+h,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t+h)\rangle\Big| + \Big|\langle f(t,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t)\rangle\Big|. \qquad (4.9)\end{aligned}$$

The last term on the right side of the above converges to zero as h tends to zero since $\bar{\mu}(\cdot)$ is continuous. Furthermore, by this same continuity and the fact that $\bar{\mu}(t) \in \mathcal{M}_F$, given $\epsilon > 0$, there exists an $M < \infty$, and $h_1 > 0$ such that

$$\left\langle 1_{(M,\infty)}, \bar{\mu}(t+h)\right\rangle < \epsilon, \qquad h \in [-\min(t,h_1), h_1].$$

Since $f(\cdot,\cdot)$ is uniformly continuous on $[(t-h_1)^+, t+h_1]\times[0,M]$, there exists $h_2 \in (0,h_1)$ such that $|f(t+h,x) - f(t,x)| \leq \epsilon$ for all $h \in [-\min(t,h_2),h_2]$ and $x \in [0,M]$. Therefore, for $h \in [-\min(t,h_2),h_2]$,

$$\Big|\langle f(t+h,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t+h)\rangle\Big| \leq \epsilon\left\langle 1_{[0,M]},\bar{\mu}(t+h)\right\rangle + 2\|f\|_\infty\epsilon \leq \epsilon\langle 1,\bar{\mu}(t+h)\rangle + 2\|f\|_\infty\epsilon.$$

Letting h and then ϵ tend to zero in (4.9) and using the above estimates, we see that the left side of (4.9) tends to zero as h tends to zero. ■

Proof of Lemma 4.1. Let $f \in \mathbf{C}_b^1([0,\infty)\times\mathbb{R}_+)$ be such that $f(\cdot,0)\equiv 0$ and $f_x(\cdot,0)\equiv 0$. Fix $0 \leq t < t^*$ and consider $h \in (-\infty,\infty)$ such that $0 \leq t+h < t^*$. Then

$$\begin{aligned}&\langle f(t+h,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t)\rangle \\ &= \langle f(t+h,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t+h)\rangle + \langle f(t,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t)\rangle. \qquad (4.10)\end{aligned}$$

Rewrite the first two terms on the right side of (4.10) as

$$\begin{aligned}\langle f(t+h,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t+h)\rangle &= \left\langle \int_t^{t+h} f_s(u,\cdot)du, \bar{\mu}(t+h)\right\rangle \\ &= \left\langle \int_0^1 f_s(t+hv,\cdot)h dv, \bar{\mu}(t+h)\right\rangle \\ &= h\int_0^1 \langle f_s(t+hv,\cdot),\bar{\mu}(t+h)\rangle\, dv.\end{aligned}$$

For each $v \in [0,1]$, define a function $f^v : [0,\infty)\times\mathbb{R}_+ \longrightarrow \mathbb{R}$ by $f^v(u,x) = f_s(t+(u-t)v,x)$. Then $f^v \in \mathbf{C}_b([0,\infty)\times\mathbb{R}_+)$, and so, by Proposition 4.2, $u \longrightarrow \langle f^v(u,\cdot),\bar{\mu}(u)\rangle$ is a continuous function of $u \in [0,\infty)$. Moreover, $f^v(t+h,\cdot) = f_s(t+hv,\cdot)$. Therefore, for each $v \in [0,1]$,

$$\lim_{h\to 0}\langle f_s(t+hv,\cdot),\bar{\mu}(t+h)\rangle = \lim_{h\to 0}\langle f^v(t+h,\cdot),\bar{\mu}(t+h)\rangle = \langle f^v(t,\cdot),\bar{\mu}(t)\rangle = \langle f_s(t,\cdot),\bar{\mu}(t)\rangle.$$

This together with bounded convergence implies that

$$\lim_{h\to 0}\frac{\langle f(t+h,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t+h)\rangle}{h} = \int_0^1 \langle f_s(t,\cdot),\bar{\mu}(t)\rangle\, dv = \langle f_s(t,\cdot),\bar{\mu}(t)\rangle. \qquad (4.11)$$

Now consider the last two terms on the right side of (4.10). Since $t \in [0,t^*)$ and $h \in (-\infty,\infty)$ is such that $t+h \in [0,t^*)$, and since $f(t,\cdot) \in \mathcal{C}$, we can use (3.3) with $g(\cdot) = f(t,\cdot)$ to obtain

$$\langle f(t,\cdot),\bar{\mu}(t+h)\rangle - \langle f(t,\cdot),\bar{\mu}(t)\rangle = -\int_t^{t+h}\frac{\langle f_x(t,\cdot),\bar{\mu}(s)\rangle}{\langle 1,\bar{\mu}(s)\rangle}ds + \alpha h\langle f(t,\cdot),\nu\rangle. \qquad (4.12)$$

Since $\bar{\mu}(\cdot)$ is assumed to be continuous and $\langle 1, \bar{\mu}(s)\rangle > 0$ for $s \in [0, t^*)$,

$$s \longrightarrow \frac{\langle f_x(t,\cdot), \bar{\mu}(s)\rangle}{\langle 1, \bar{\mu}(s)\rangle},$$

is a continuous function of $s \in [0, t^*)$. It follows that

$$\lim_{h\to 0} \frac{1}{h} \int_t^{t+h} \frac{\langle f_x(t,\cdot), \bar{\mu}(s)\rangle}{\langle 1, \bar{\mu}(s)\rangle} ds = \frac{\langle f_x(t,\cdot), \bar{\mu}(t)\rangle}{\langle 1, \bar{\mu}(t)\rangle}. \tag{4.13}$$

Combining (4.10), (4.11), (4.12), and (4.13), we obtain

$$\frac{d}{dt} \langle f(t,\cdot), \bar{\mu}(t)\rangle = \langle f_s(t,\cdot), \bar{\mu}(t)\rangle - \frac{\langle f_x(t,\cdot), \bar{\mu}(t)\rangle}{\langle 1, \bar{\mu}(t)\rangle} + \alpha \langle f(t,\cdot), \nu\rangle, \qquad t < t^*.$$

Since f, f_s, and f_x are continuous and bounded, and since $1/\bar{Z}(\cdot)$ is continuous on $[0, t^*)$, by Proposition 4.2, each term on the right side of the above equation is continuous for $t \in [0, t^*)$. Thus, (4.8) follows. ■

Next the fluid model equations for time-dependent functions, i.e., (4.8), are used to derive (4.14) below. Recall that this serves two purposes. Firstly, it reduces proving uniqueness of $\bar{\mu}(\cdot)$ to proving that $t^* = \infty$ and that $\bar{S}(\cdot)$ is uniquely determined by ξ. Secondly, in the proof of Lemma 4.4, (4.14) with $w = \infty$ will be used to show that a time change of $\bar{Z}(\cdot)$ is related to the unique solution of (4.4).

Lemma 4.3 *Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution such that $\bar{\mu}(0) = \xi$. Then for all $w \in (0, \infty]$,*

$$\left\langle 1_{(0,w)}, \bar{\mu}(t)\right\rangle = \left\langle 1_{(0,w)}(\cdot - \bar{S}(t)), \xi\right\rangle + \alpha \int_0^t \left\langle 1_{(0,w)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu\right\rangle ds, \tag{4.14}$$

for all $t < t^* = \inf\{s : \bar{\mu}(s) = \mathbf{0}\}$.

Proof. For $t = 0$, (4.14) holds trivially. Fix t such that $0 < t < t^*$. Consider $g \in \mathbf{C}_b^1(\mathbb{R})$ with $g(x) = 0$ for all $x \le 0$. Note that $g'(x) = 0$ for all $x \le 0$. Let

$$f(s,x) = g\left(x - \bar{S}(t) + \bar{S}(s)\right), \qquad s \in [0,t],\ x \in \mathbb{R}_+. \tag{4.15}$$

Since $\bar{S}(\cdot) \in \mathbf{C}^1([0,t^*))$ with $\bar{S}'(s) = 1/\bar{Z}(s)$ for $s \in [0, t^*)$, it follows that $f \in \mathbf{C}_b^1([0,t] \times \mathbb{R}_+)$ and

$$f_s(s,x) = \frac{g'\left(x - \bar{S}(t) + \bar{S}(s)\right)}{\bar{Z}(s)} \qquad \text{and} \qquad f_x(s,x) = g'\left(x - \bar{S}(t) + \bar{S}(s)\right), \qquad s \in [0,t],\, x \in \mathbb{R}_+.$$

Also, for $s \in [0,t]$, $f(s,0) = 0$ and $f_x(s,0) = 0$ because $g(x) = 0$ and $g'(x) = 0$ for all $x \le 0$.

Fix $0 < \epsilon < t$. We wish to construct a function f^ϵ that satisfies the conditions in Lemma 4.1 and that closely approximates f on $[0,t] \times \mathbb{R}_+$. For this, choose a cut-off function $h^\epsilon \in \mathbf{C}_b^1([0,\infty))$ such that

$$h^\epsilon(s) = \begin{cases} 1, & s \in [0, t-\epsilon], \\ 0, & s \in [t - \epsilon/2, \infty). \end{cases}$$

Extend f to be identically equal to zero on $(t,\infty) \times \mathbb{R}_+$ and define

$$f^\epsilon(s,x) = f(s,x) h^\epsilon(s), \qquad s \in [0,\infty),\ x \in \mathbb{R}_+.$$

Thus, $f^\epsilon \in \mathbf{C}_b^1([0,\infty)\times\mathbb{R}_+)$ with $f^\epsilon(\cdot,0)\equiv 0$, $f^\epsilon_x(\cdot,0)\equiv 0$, and $f^\epsilon = f$ on $[0,t-\epsilon]\times\mathbb{R}_+$.

Now substitute f^ϵ into (4.8), and use the fact that $f^\epsilon(s,\cdot)\equiv f(s,\cdot)$ for $0\le s\le t-\epsilon$ to obtain

$$\begin{aligned}\langle f(s,\cdot),\bar\mu(s)\rangle &= \langle f(0,\cdot),\xi\rangle + \int_0^s \left\langle g'\left(\cdot - \bar S(t)+\bar S(v)\right),\bar\mu(v)\right\rangle \bar Z(v)^{-1}dv \\ &\quad - \int_0^s \left\langle g'\left(\cdot - \bar S(t)+\bar S(v)\right),\bar\mu(v)\right\rangle \bar Z(v)^{-1}dv + \alpha\int_0^s \left\langle g\left(\cdot - \bar S(t)+\bar S(v)\right),\nu\right\rangle dv,\end{aligned} \tag{4.16}$$

for all $s\in[0,t-\epsilon]$, for each $\epsilon\in(0,t)$. The first and second integral terms in the above expression cancel one another. Since $\epsilon\in(0,t)$ was arbitrary, it follows that for all $s\in[0,t)$,

$$\langle f(s,\cdot),\bar\mu(s)\rangle = \left\langle g(\cdot-\bar S(t)),\xi\right\rangle + \alpha\int_0^s \left\langle g(\cdot-\bar S(t)+\bar S(v)),\nu\right\rangle dv. \tag{4.17}$$

We wish to let $s\uparrow t$ in (4.17). It is clear what effect that has on the right side. To see that the left side tends to $\langle g,\bar\mu(t)\rangle$, note that by a minor modification of the proof of Proposition 4.2, $\langle f(s,\cdot),\bar\mu(s)\rangle$ converges to $\langle f(t,\cdot),\bar\mu(t)\rangle$ as s increases to t. Moreover, by the definition of f, $\langle f(t,\cdot),\bar\mu(t)\rangle = \langle g,\bar\mu(t)\rangle$. So letting $s\uparrow t$ in (4.17) gives, for all $g\in\mathbf{C}_b^1(\mathbb{R})$ such that $g(x)=0$ for all $x\le 0$,

$$\langle g,\bar\mu(t)\rangle = \left\langle g(\cdot-\bar S(t)),\xi\right\rangle + \alpha\int_0^t \left\langle g(\cdot-\bar S(t)+\bar S(v)),\nu\right\rangle dv, \tag{4.18}$$

for all $t\in(0,t^*)$. To obtain (4.14) from (4.18), consider a sequence of nonnegative functions $\{g_n\}\subset\mathbf{C}_b^1(\mathbb{R})$ that increases to $1_{(0,w)}$ pointwise on $\mathbb{R}$ and apply monotone convergence. ∎

As a consequence of Lemma 4.3 and the assumption that $\bar\mu(t)$ does not charge the origin for all $t\ge 0$, in order to prove uniqueness, it suffices to show that $t^*=\infty$ and that $\bar S(\cdot)$ is uniquely determined by ξ. Observe that $\bar S(\cdot)$ is continuous and strictly increasing prior to the time t^* at which $\bar\mu(\cdot)$ reaches the zero measure. Thus, the continuous inverse of $\bar S(\cdot)$ on $[0,t^*)$ is given by

$$\bar T(u) = \bar S^{-1}(u) = \inf\{t\ge 0: \bar S(t)>u\}, \qquad u<u^* = \lim_{t\uparrow t^*}\bar S(t). \tag{4.19}$$

In (4.19), the superscript -1 denotes the functional inverse. Recall that, since $\bar\mu(\cdot)$ is continuous, $\bar S(\cdot)\in\mathbf{C}^1([0,t^*))$ with $\bar S'(\cdot) = 1/\bar Z(\cdot)$. This means that $\bar T(\cdot)\in\mathbf{C}^1([0,u^*))$ with

$$\bar T'(u) = \frac{1}{\bar S'(\bar T(u))}, \qquad u\in[0,u^*). \tag{4.20}$$

Our strategy is to use (4.14) with $w=\infty$ to show that $M(\cdot)=\bar T(\cdot)$ is the unique locally bounded solution of (4.3). Notice that assumption (3.2) has not yet been used, but it is needed for the statement of the following lemma.

Lemma 4.4 *Suppose that $\bar\mu(\cdot)$ is a fluid model solution for the critical data (α,ν) with $\bar\mu(0)=\xi$. Then $u^*=\infty$ and $t^*=\infty$. Moreover, for $\bar T(\cdot)$ defined by* (4.19), *$M(\cdot)=\bar T(\cdot)$ is the unique locally bounded solution of* (4.3). *In particular, $\bar T(\cdot) = (H_\xi * U_e)(\cdot)$ with $H_\xi(\cdot)$ defined by* (4.1) *and $U_e(\cdot)$ defined by (4.5).*

The statement of Lemma 4.4 is much like the statement of Theorem 17 in [9]. In fact, reading Theorem 17 in [9] suggested to us that proving something like Lemma 4.4 for our fluid model might be useful. In the remark following the proof of Lemma 4.4, we explain the connection between our work and that in [9] more precisely. Before proceeding with the proof of Lemma 4.4, we will need to verify the following proposition, which is a consequence of Lemma 4.3.

Proposition 4.5 *Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution with $\bar{\mu}(0) = \xi$. Then $u^* < \infty$ implies $t^* = \infty$.*

Proof. Suppose that $u^* < \infty$. Substituting $w = \infty$ into (4.14) and using the fact that $\bar{\mu}(t)(\{0\}) = 0$ for all $t \geq 0$ gives, for $t < t^*$,

$$\begin{aligned}
\bar{Z}(t) &= \left\langle 1_{(0,\infty)}(\cdot - \bar{S}(t)), \xi \right\rangle + \alpha \int_0^t \left\langle 1_{(0,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle ds \\
&\geq \alpha \int_0^t \left\langle 1_{(0,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle ds \\
&\geq \alpha \int_0^t \left\langle 1_{(0,\infty)}(\cdot - u^* + \bar{S}(s)), \nu \right\rangle ds.
\end{aligned}$$

Fix $0 < \epsilon < 1$ and choose $\delta > 0$ such that $\left\langle 1_{(\delta,\infty)}, \nu \right\rangle \geq \epsilon$. Choose $\tilde{t} < t^*$ such that for all $s \in [\tilde{t}, t^*)$, $u^* - \bar{S}(s) \leq \delta$. Then, for $s \in [\tilde{t}, t^*)$, $\left\langle 1_{(0,\infty)}(\cdot - u^* + \bar{S}(s)), \nu \right\rangle \geq \epsilon$. In particular, for $t \in [\tilde{t}, t^*)$,

$$\bar{Z}(t) \geq \alpha\epsilon(t - \tilde{t}).$$

Therefore, $u^* < \infty$ implies that $\bar{Z}(t)$ is bounded away from zero as t increases to t^*. If in addition $t^* < \infty$, then $\bar{Z}(\cdot)$ is discontinuous at t^* because, by definition of t^*, $\bar{Z}(t) = 0$ for all $t > t^*$. But $\bar{Z}(\cdot)$ is continuous since $\bar{\mu}(\cdot)$ is continuous, and therefore it must be the case that $t^* = \infty$ whenever $u^* < \infty$. ∎

Proof of Lemma 4.4. Observe that, since $\bar{\mu}(t)(\{0\}) = 0$ for all $t \geq 0$, (4.14) with $w = \infty$ becomes

$$\bar{Z}(t) = H'_\xi(\bar{S}(t)) + \alpha \int_0^t \left[1 - F(\bar{S}(t) - \bar{S}(s))\right] ds, \qquad t < t^*, \tag{4.21}$$

where $H'_\xi(\cdot)$ is defined by (4.2). In (4.21), let $u = \bar{S}(t)$ and perform the change of variables $v = \bar{S}(s)$ to obtain

$$\bar{Z}(\bar{T}(u)) = H'_\xi(u) + \alpha \int_0^u [1 - F(u - v)]\, d\bar{T}(v), \qquad u < u^*.$$

Since $\bar{S}'(s) = 1/\bar{Z}(s)$ for $s < t^*$ and $\bar{T}(u) < t^*$ for $u < u^*$, the left side is in fact $1/\bar{S}'(\bar{T}(u))$ for $u < u^*$. This together with (4.20) gives

$$\bar{T}'(u) = H'_\xi(u) + \alpha \int_0^u [1 - F(u - v)]\, d\bar{T}(v), \qquad u < u^*.$$

Using a change of variables and the definition of $f_e(\cdot)$, we have

$$\bar{T}'(u) = H'_\xi(u) + \int_0^u \bar{T}'(u - v) f_e(v) dv = H'_\xi(u) + \left(\bar{T}' * F_e\right)(u), \qquad u < u^*. \tag{4.22}$$

Thus, $N(\cdot) = \bar{T}'(\cdot)$ is a locally bounded solution of (4.4) on $[0, u^*)$. It follows from the uniqueness of locally bounded solutions of (4.4) on $[0, u^*)$ that $\bar{T}'(u)$ is given by the right side of (4.7) for $u \in [0, u^*)$, i.e.,

$$\bar{T}'(u) = \left(H'_\xi * U_e\right)(u), \qquad u < u^*. \tag{4.23}$$

Since $\bar{T} \in \mathbf{C}^1([0,u^*))$ and $\bar{T}(0)=0$, and since $H_\xi * U_e \in \mathbf{C}^1(\mathbb{R}_+)$, $(H_\xi * U_e)(0)=0$, and $(H_\xi * U_e)'(\cdot) = (H_\xi' * U_e)(\cdot)$, integrating (4.23) yields

$$\bar{T}(u) = (H_\xi * U_e)(u), \qquad u < u^*. \tag{4.24}$$

The final task is to show that $t^* = \infty$ and $u^* = \infty$. By (4.24) and the properties of $(H_\xi * U_e)(\cdot)$, $\lim_{u \uparrow u^*} \bar{T}(u) < \infty$ if and only if $u^* < \infty$. By definition (4.19), we have $t^* = \lim_{u \uparrow u^*} \bar{T}(u)$. Thus, $t^* < \infty$ if and only if $u^* < \infty$. But this can only be consistent with the statement of Proposition 4.5 if both $u^* = \infty$ and $t^* = \infty$. ■

Remark. A result like Lemma 4.4 holds for the fluid approximation studied in [9] as well (cf. Theorem 17 in [9]). This follows almost as an immediate consequence of one of their fluid approximation equations. In particular, they consider a fluid approximation of the so-called busy time equation given by

$$t = \left\langle \bar{S}(t) \wedge \chi, \xi \right\rangle + \alpha \int_0^t \langle \left(\bar{S}(t) - \bar{S}(s)\right) \wedge \chi, \nu \rangle ds, \qquad t < t^*, \tag{4.25}$$

(cf. (31) in [9]). The right side of (4.25) can be interpreted as the total amount of work performed in the fluid limit by the server by time t, for $t < t^*$. It is easily verified that $H_\xi(x) = \langle x \wedge \chi, \xi \rangle$ and $F_e(x) = \alpha \langle x \wedge \chi, \nu \rangle$, $x \in \mathbb{R}_+$. This, together with a time change, shows that (4.25) is equivalent to (4.3) with $M(\cdot) = \bar{T}(\cdot)$. Since our state descriptor tracks more information than just $\bar{S}(\cdot)$, it is perhaps not surprising that we are able to derive (4.25) from (3.3). However, as the proofs of Lemmas 4.1, 4.3, and 4.4 demonstrate, some work is required for this. While reading [9], we realized that deriving a version of (4.25) for our fluid model could provide an important component of our uniqueness proof.

Lemma 4.4 will now be used in conjunction with Lemma 4.3 to prove both uniqueness of fluid model solutions and the constant workload property.

Proof of Uniqueness for Theorem 3.1, $\boldsymbol{\xi \in \mathcal{M}_F^{c,p}}$. Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the critical data (α,ν) with $\bar{\mu}(0) = \xi$. By Lemma 4.3 and the fact that $\bar{\mu}(t)(\{0\}) = 0$ for all $t \geq 0$, $\langle 1_{[0,w)}, \bar{\mu}(t) \rangle$ is uniquely determined by $\bar{S}(\cdot)$, ξ, and (α,ν) for each $w \in (0,\infty]$ and for each $t \in [0,t^*)$. Since intervals of the form $[0,w)$, $w \in (0,\infty]$, generate the Borel σ-algebra on $\mathbb{R}_+$, this uniquely determines $\bar{\mu}(t)$ for each $t \in [0,t^*)$. By Lemma 4.4, $t^* = \infty$, so $\bar{\mu}(\cdot)$ is uniquely determined by $\bar{S}(\cdot)$, ξ, and (α,ν). By (4.19) and Lemma 4.4, $\bar{S}(\cdot)$ is the inverse of $(H_\xi * U_e)(\cdot)$. Since $(H_\xi * U_e)(\cdot)$, is determined by ξ and (α,ν), the function $\bar{S}(\cdot)$, and therefore the fluid model solution $\bar{\mu}(\cdot)$, is uniquely determined by ξ and (α,ν). ■

Proof that $\boldsymbol{\bar{W}(\cdot)}$ is constant for Theorem 3.1, $\boldsymbol{\xi \in \mathcal{M}_F^{c,p}}$. Suppose that $\bar{\mu}(\cdot)$ is a fluid model solution for the critical data (α,ν) with $\bar{\mu}(0) = \xi$. Fix $t \geq 0$. By definition,

$$\bar{W}(t) = \langle \chi, \bar{\mu}(t) \rangle = \int_{\mathbb{R}_+} x\, \bar{\mu}(t)(dx) = \int_{\mathbb{R}_+} \int_{[0,x]} dy\, \bar{\mu}(t)(dx).$$

By Fubini's theorem, the order of integration can be interchanged and this gives

$$\bar{W}(t) = \int_{\mathbb{R}_+} \int_{[y,\infty)} \bar{\mu}(t)(dx) dy = \int_{\mathbb{R}_+} \left\langle 1_{[y,\infty)}, \bar{\mu}(t) \right\rangle dy.$$

This together with (4.14) and the fact that $t^* = \infty$ (cf. Lemma 4.4) gives

$$\bar{W}(t) = \int_{\mathbb{R}_+} \left(\left\langle 1_{[y,\infty)}(\cdot - \bar{S}(t)), \xi \right\rangle + \alpha \int_0^t \left\langle 1_{[y,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle ds \right) dy. \tag{4.26}$$

We begin by simplifying the first term on the right side of (4.26). Since ξ has no atoms, from (4.2) it follows that

$$\int_{\mathbb{R}_+} \left\langle 1_{[y,\infty)}(\cdot - \bar{S}(t)), \xi \right\rangle dy = \int_{\mathbb{R}_+} H'_\xi(y + \bar{S}(t))dy = \int_{[\bar{S}(t),\infty)} H'_\xi(y)dy.$$

Notice that $\int_{[0,\bar{S}(t))} H'_\xi(y)dy < \infty$. Therefore, we may add and subtract this term on the right side of the above equality. Also notice that $\int_{[0,\infty)} H'_\xi(y)dy = \langle \chi, \xi \rangle$. Thus, we have

$$\int_{\mathbb{R}_+} \left\langle 1_{[y,\infty)}(\cdot - \bar{S}(t)), \xi \right\rangle dy = \bar{W}(0) - \int_{[0,\bar{S}(t))} H'_\xi(y)dy = \bar{W}(0) - H_\xi(\bar{S}(t)). \tag{4.27}$$

Next we simplify the second term on the right side of (4.26). By interchanging the order of integration, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}_+} \alpha \int_0^t \left\langle 1_{[y,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle ds\, dy \\
&\quad = \int_{\mathbb{R}_+} \alpha \int_0^t \nu\left(\left[y + \bar{S}(t) - \bar{S}(s), \infty\right)\right) ds\, dy = \int_0^t \int_{\mathbb{R}_+} \alpha\nu\left(\left[y + \bar{S}(t) - \bar{S}(s), \infty\right)\right) dy\, ds.
\end{aligned}$$

For fixed $0 \le s \le t$, $\nu\left(\left[y + \bar{S}(t) - \bar{S}(s), \infty\right)\right) = \nu\left(\left(y + \bar{S}(t) - \bar{S}(s), \infty\right)\right)$ for all $y \in \mathbb{R}_+$ such that $y + \bar{S}(t) - \bar{S}(s)$ is not an atom for ν. Since ν has at most countably many atoms, there are at most countably many exceptional $y \in \mathbb{R}_+$ (where the exceptional set may depend on s and t). This, together with the definition of F, a change of variables, and the definition of F_e gives

$$\begin{aligned}
&\int_{\mathbb{R}_+} \alpha \int_0^t \left\langle 1_{[y,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle ds\, dy \\
&\quad = \int_0^t \int_{\bar{S}(t)-\bar{S}(s)}^\infty \alpha\left(1 - F(z)\right) dz\, ds = \int_0^t \left(1 - F_e(\bar{S}(t) - \bar{S}(s))\right) ds = t - \int_0^t F_e(\bar{S}(t) - \bar{S}(s))ds.
\end{aligned}$$

Recall that $\bar{T}(\cdot)$ defined in (4.19) is the functional inverse of $\bar{S}(\cdot)$ and that $t^* = u^* = \infty$ (cf. Lemma 4.4). Using the change of variables $v = \bar{S}(s)$ gives

$$\begin{aligned}
&\int_{\mathbb{R}_+} \alpha \int_0^t \left\langle 1_{[y,\infty)}(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle dsdy \\
&\quad = t - \int_0^{\bar{S}(t)} F_e(\bar{S}(t) - v)d\bar{T}(v) = t - \left(F_e * \bar{T}\right)(\bar{S}(t)) = t - \left(\bar{T} * F_e\right)(\bar{S}(t)).
\end{aligned} \tag{4.28}$$

Combining (4.26), (4.27) and (4.28), we have

$$\bar{W}(t) = \bar{W}(0) + t - \Big(H_\xi(\bar{S}(t)) + \left(\bar{T} * F_e\right)(\bar{S}(t))\Big).$$

Observe that the third term on the right side is simply the right side of equation (4.3) for $M(\cdot) = \bar{T}(\cdot)$ evaluated at $u = \bar{S}(t)$. By Lemma 4.4, $M(\cdot) = \bar{T}(\cdot)$ is the unique locally bounded solution of (4.3). Thus,

$$\bar{W}(t) = \bar{W}(0) + t - \bar{T}(\bar{S}(t)) = \bar{W}(0) + t - t = \bar{W}(0),$$

which holds even if $\bar{W}(0) = \infty$. ■

4.2 Proof of Existence of Fluid Model Solutions

Fix $\xi \in \mathcal{M}_F^{c,p}$. We wish to prove that a fluid model solution $\bar{\mu}(\cdot)$ with $\bar{\mu}(0) = \xi$ exists. The uniqueness proof suggests that (4.14) might be used to define a suitable candidate fluid model solution. However, for that one needs to find a suitable candidate for $\bar{S}(\cdot)$. The uniqueness proof is also helpful here, since it suggests that $\bar{S}(\cdot)$ should be the inverse of the unique locally bounded solution of (4.3): see (4.19) and Lemma 4.4. Our assumptions on ξ imply that such a solution exists, so let $\bar{T}(\cdot)$ be the unique locally bounded solution of (4.3), i.e., $\bar{T}(\cdot) = (H_\xi * U_e)(\cdot)$ where $H_\xi(\cdot)$ is defined by (4.1). Then, $\bar{T}(\cdot) \in \mathbf{C}^1(\mathbb{R}_+)$ and satisfies $\bar{T}(0) = 0$, $\bar{T}(\cdot)$ is strictly increasing, and $\lim_{u\to\infty} \bar{T}(u) = \infty$. It follows that $\bar{S}(t) = \bar{T}^{-1}(t)$ for $t \geq 0$ exists, is strictly increasing, $\bar{S}(\cdot) \in \mathbf{C}^1(\mathbb{R}_+)$, and

$$\bar{S}'(t) = \frac{1}{\bar{T}'(\bar{S}(t))} \qquad t \geq 0. \tag{4.29}$$

For each $t \geq 0$, let $\bar{\mu}(t)$ be the unique element of $\mathcal{M}_F$ such that $\langle 1_{\{0\}}, \bar{\mu}(t)\rangle = 0$ and

$$\left\langle 1_{(0,w)}, \bar{\mu}(t)\right\rangle = \left\langle 1_{(0,w)}(\cdot - \bar{S}(t)), \xi\right\rangle + \alpha \int_0^t \left\langle 1_{(0,w)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu\right\rangle ds, \tag{4.30}$$

for all $w \in (0, \infty]$. Note that $\bar{\mu}(0) = \xi$ and $\langle 1_{(0,\infty)}, \bar{\mu}(t)\rangle \leq \langle 1, \xi\rangle + \alpha t$ for all $t \geq 0$.

Proposition 4.6 *Let $\bar{\mu} : [0, \infty) \longrightarrow \mathcal{M}_F$ be defined via (4.30) and let $t^* = \inf\{t \geq 0 : \bar{\mu}(t) = \mathbf{0}\}$. Then $t^* = \infty$, $\bar{\mu}(t)$ has no atoms for each $t \geq 0$, and for each $w \in (0, \infty]$, $t \longrightarrow \left\langle 1_{(0,w)}, \bar{\mu}(t)\right\rangle$ is a function in $\mathbf{C}([0, \infty))$. In particular, $\bar{\mu}(\cdot)$ is continuous.*

Proof. In order to prove that $t^* = \infty$, it suffices to show that $\bar{Z}(t) = \langle 1, \bar{\mu}(t)\rangle$ is strictly positive for all $t \geq 0$. Note that $\langle 1_{\{0\}}, \bar{\mu}(t)\rangle = 0$ and so (4.30) with $w = \infty$ becomes

$$\bar{Z}(t) = H_\xi'(\bar{S}(t)) + \alpha \int_0^t \left(1 - F(\bar{S}(t) - \bar{S}(s))\right) ds, \qquad t \geq 0, \tag{4.31}$$

where $H_\xi'(\cdot)$ is given by (4.2). Since $\xi \neq \mathbf{0}$, $\bar{Z}(0) > 0$. Equation (4.31) can be used to show that $\bar{Z}(t) > 0$ for $t > 0$. To see this note that given $\epsilon \in (0, 1)$, since $F(0) = 0$ and F is right continuous, there exists $M > 0$ such that $1 - F(x) > \epsilon$ for all $x < M$. Since $\bar{S}(\cdot)$ is continuous, for each fixed $t > 0$, there exists $\delta_t > 0$ such that $s \geq 0$ and $|t - s| < \delta_t$ imply that $|\bar{S}(t) - \bar{S}(s)| < M$. These two facts together with (4.31) yield

$$\bar{Z}(t) \geq \alpha \int_{(t-\delta_t)^+}^t \epsilon \, ds > 0, \qquad t > 0.$$

So $t^* = \infty$.

Next we show that $\bar{\mu}(t)$ has no atoms for each $t \geq 0$. Since $\bar{\mu}(0) = \xi$, this holds for $t = 0$. By (4.30), we have for each $t \geq 0$ and $0 < x < w \leq \infty$,

$$\left\langle 1_{[x,w)}, \bar{\mu}(t)\right\rangle = \left\langle 1_{[x,w)}(\cdot - \bar{S}(t)), \xi\right\rangle + \alpha \int_0^t \left\langle 1_{[x,w)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu\right\rangle ds. \tag{4.32}$$

Fix $x \in (0, \infty)$, let $w \downarrow x$, and use bounded convergence to obtain

$$\left\langle 1_{\{x\}}, \bar{\mu}(t)\right\rangle = \left\langle 1_{\{x\}}(\cdot - \bar{S}(t)), \xi\right\rangle + \alpha \int_0^t \left\langle 1_{\{x\}}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu\right\rangle ds, \tag{4.33}$$

for $t \geq 0$, $x \in (0, \infty)$. The first term on the right side of (4.33) is clearly zero since ξ has no atoms. Recall that $\bar{S}(\cdot)$ is strictly increasing and ν has at most countably many atoms. Thus, for each fixed $(t, x) \in [0, \infty) \times (0, \infty)$, $\langle 1_{\{x\}}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \rangle = 0$ for (Lebesgue) almost every $s \in [0, t]$. Therefore the second term on the right side of (4.33) is zero. Thus, $\left\langle 1_{\{x\}}, \bar{\mu}(t) \right\rangle = 0$ for all $x > 0$. By definition $\left\langle 1_{\{0\}}, \bar{\mu}(t) \right\rangle = 0$, so $\bar{\mu}(t)$ has no atoms for each $t \geq 0$.

In order to show that for each $w \in (0, \infty]$, $\langle 1_{(0,w)}, \bar{\mu}(t) \rangle$ is a continuous function of $t \in [0, \infty)$, it suffices to show that each term on the right side of (4.30) is continuous. This is straightforward for the first term since $\bar{S}(\cdot)$ is continuous and ξ has no atoms. For each $w \in (0, \infty]$ and $t \geq 0$, and for each $s \in [0, t]$ such that ν does not have an atom either at $\bar{S}(t) - \bar{S}(s)$ or at $w + \bar{S}(t) - \bar{S}(s)$,

$$\lim_{u \to t} \left\langle 1_{(0,w)}(\cdot - (\bar{S}(u) - \bar{S}(s))), \nu \right\rangle = \left\langle 1_{(0,w)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle.$$

Since $\bar{S}(\cdot)$ is strictly increasing and ν has at most countably many atoms, the above holds except perhaps for countably many values of $s \in [0, t]$, which comprise a set of Lebesgue measure zero (possibly depending on w and t). So, using bounded convergence and the fact that the integrand is bounded, one can show that, for each fixed $w \in (0, \infty]$, the second term on the right side of (4.30) is a continuous function of $t \in [0, \infty)$.

Finally, we must show that $\bar{\mu}(\cdot)$ is continuous. In order to do this, we must show that for each $g \in \mathbf{C}_b(\mathbb{R}_+)$, $\langle g, \bar{\mu}(t) \rangle$ is a continuous function of $t \in [0, \infty)$. Since $\bar{Z}(\cdot)$ is strictly positive and continuous, it suffices to show that

$$\frac{\langle g, \bar{\mu}(t) \rangle}{\bar{Z}(t)}, \tag{4.34}$$

is a continuous function of $t \in [0, \infty)$. In order to do this, for each $x \in \mathbb{R}_+$, let

$$G(t, x) = \frac{\left\langle 1_{[0,x)}, \bar{\mu}(t) \right\rangle}{\bar{Z}(t)}, \qquad t \geq 0.$$

Since, for each $t \geq 0$, $\bar{\mu}(t)$ does not have an atom at the origin, it follows that for each $x \in \mathbb{R}_+$, $\left\langle 1_{[0,x)}, \bar{\mu}(t) \right\rangle = \left\langle 1_{(0,x)}, \bar{\mu}(t) \right\rangle$ for each $t \geq 0$. Thus, from the results proved above, for each $x \in \mathbb{R}_+$, $G(t, x)$ is a continuous function of $t \in [0, \infty)$. Since $\bar{\mu}(t)/\bar{Z}(t)$ is a probability measure for each $t \geq 0$, it follows that (4.34) is a continuous function of $t \in [0, \infty)$ (cf. [2], Thm. 2.2 and use the fact that $\{[x, y) : 0 \leq x < y < \infty\} \cup \varnothing$ forms a π-system). ■

Thus far, it has been shown that the measure valued function $\bar{\mu}(\cdot)$ defined via (4.30) satisfies conditions (i) and (ii) in the definition of a fluid model solution, and that $t^* = \infty$. Thus condition (iv) holds trivially. So the remaining task is to prove that $\bar{\mu}(\cdot)$ satisfies (3.3) for all $t \geq 0$. This is addressed in Lemmas 4.7 and 4.8 below. Lemma 4.7 implies that the definitions of $\bar{S}(\cdot)$ and $\bar{\mu}(\cdot)$ give rise to the appropriate relationship between $\bar{S}'(\cdot)$ and $\bar{Z}(\cdot)$. The proof of Lemma 4.8 uses this relationship to prove that $\bar{\mu}(\cdot)$ satisfies a differential form of (3.3).

Lemma 4.7 *Let $\bar{\mu} : [0, \infty) \longrightarrow \mathcal{M}_F$ be defined via (4.30). Then, for all $t \geq 0$,*

$$\bar{S}'(t) = \frac{1}{\bar{Z}(t)}. \tag{4.35}$$

Proof. In order to verify (4.35), perform the change of variables $v = \bar{S}(s)$ in (4.31) to obtain, for $t \geq 0$,

$$\bar{Z}(t) = H'_\xi(\bar{S}(t)) + \alpha \int_0^{\bar{S}(t)} \big(1 - F(\bar{S}(t) - v)\big)\, d\bar{T}(v) = H'_\xi(\bar{S}(t)) + (\bar{T}' * F_e)(\bar{S}(t)) = \bar{T}'(\bar{S}(t)).$$

Here the final equality follows from the fact that $N(\cdot) = \bar{T}'(\cdot)$ is the unique locally bounded solution of (4.4). By (4.29), the above yields $\bar{Z}(t) = 1/\bar{S}'(t)$ for all $t \geq 0$. ■

We are now ready to prove a differential form of (3.3), which will complete the verification that $\bar{\mu}(\cdot)$ defined via (4.30) is a fluid model solution.

Lemma 4.8 *Let $\bar{\mu} : [0,\infty) \longrightarrow \mathcal{M}_F$ be defined via (4.30). For $g \in \mathcal{C}$, $t \longrightarrow \langle g, \bar{\mu}(t)\rangle$ is a function in $\mathbf{C}^1([0,\infty))$ and*

$$\frac{d}{dt}\langle g, \bar{\mu}(t)\rangle = -\frac{\langle g', \bar{\mu}(t)\rangle}{\langle 1, \bar{\mu}(t)\rangle} + \alpha \langle g, \nu\rangle, \qquad t \geq 0. \tag{4.36}$$

Proof. Observe that for each $g \in \mathcal{C}$, $\langle g, \bar{\mu}(t)\rangle$ and the terms in the expression on the right side of (4.36) are continuous functions of $t \in [0,\infty)$. To see this note that $\bar{\mu}(\cdot)$ is continuous, g and g' are bounded continuous functions, and $\bar{Z}(\cdot) = \langle 1, \bar{\mu}(\cdot)\rangle$ is continuous and strictly positive (since $t^* = \infty$). Therefore, in order to prove Lemma 4.8, it suffices to prove that for each $g \in \mathcal{C}$, $t \longrightarrow \langle g, \bar{\mu}(t)\rangle$ is a differentiable function of $t \in [0,\infty)$ and that (4.36) holds. Note that by (4.30), (4.33), and a monotone class theorem,

$$\langle g, \bar{\mu}(t)\rangle = \big\langle g(\cdot - \bar{S}(t)), \xi\big\rangle + \alpha \int_0^t \big\langle g(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu\big\rangle\, ds, \qquad t \geq 0, \tag{4.37}$$

for all $g \in \mathbf{C}_b(\mathbb{R})$ such that $g(x) = 0$ for $x \leq 0$. Fix $g \in \mathcal{C}$ and extend g to be identically equal to zero on the negative half line. Then $g \in \mathbf{C}_b^1(\mathbb{R})$ and $g(x) = 0$ and $g'(x) = 0$ for $x \leq 0$. In particular, (4.37) holds for both g and g'. We will proceed by first showing that for this fixed g, each term on the right side of (4.37) is differentiable, and then computing the derivative of each term.

By (4.35),

$$\frac{\partial}{\partial t} g(x - \bar{S}(t)) = \frac{-g'(x - \bar{S}(t))}{\bar{Z}(t)}, \qquad t \in [0,\infty),\ x \in \mathbb{R}.$$

For each $b < \infty$, the term on the right side of the above equation is continuous and bounded for $(t,x) \in [0,b] \times \mathbb{R}$, where we have used the fact that $1/\bar{Z}(\cdot)$ is continuous and bounded on $[0,b]$. This together with the fact that $\xi \in \mathcal{M}_F$ allows an interchange in the order of integration and differentiation to obtain

$$\frac{d}{dt}\big\langle g(\cdot - \bar{S}(t)), \xi\big\rangle = -\left\langle \frac{g'(\cdot - \bar{S}(t))}{\bar{Z}(t)}, \xi \right\rangle, \qquad t \geq 0, \tag{4.38}$$

where the right member is a continuous function of $t \in [0,\infty)$. Similarly, for each (fixed) $s \in [0,\infty)$,

$$\frac{\partial}{\partial t} g(x - \bar{S}(t) + \bar{S}(s)) = \frac{-g'(x - \bar{S}(t) + \bar{S}(s))}{\bar{Z}(t)}, \qquad t \in [0,\infty),\ x \in \mathbb{R},$$

which for each $b < \infty$ is again continuous and bounded (uniformly in s) for $(t, x) \in [0, b] \times \mathbb{R}$. Thus, for each (fixed) $s \in [0, \infty)$,

$$\frac{\partial}{\partial t} \left\langle g(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle = -\left\langle \frac{g'(\cdot - \bar{S}(t) + \bar{S}(s))}{\bar{Z}(t)}, \nu \right\rangle, \qquad t \geq 0. \tag{4.39}$$

For each $b < \infty$, the right side is again continuous and bounded (uniformly in s) as a function of $t \in [0, b]$. Using the facts that $\left\langle g(\cdot - \bar{S}(t) + \bar{S}(s)), \nu \right\rangle$ and the right side of (4.39) are continuous functions of $(s, t) \in [0, \infty) \times [0, \infty)$, it follows that

$$\frac{d}{dt} \int_0^t \left\langle g(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds = -\int_0^t \left\langle \frac{g'(\cdot - \bar{S}(t) + \bar{S}(s))}{\bar{Z}(t)}, \nu \right\rangle ds + \langle g, \nu \rangle, \qquad t \geq 0. \tag{4.40}$$

Thus, each term on the right side of (4.37) is differentiable and so $t \longrightarrow \langle g, \bar{\mu}(t) \rangle$ is a differentiable function of $t \in [0, \infty)$. Moreover, using (4.37), (4.38), and (4.40), and then using (4.37) again with g' in place of g gives, for $t \geq 0$,

$$\begin{aligned} \frac{d}{dt} \langle g, \bar{\mu}(t) \rangle &= -\left\langle \frac{g'(\cdot - \bar{S}(t))}{\bar{Z}(t)}, \xi \right\rangle - \alpha \int_0^t \left\langle \frac{g'(\cdot - (\bar{S}(t) - \bar{S}(s)))}{\bar{Z}(t)}, \nu \right\rangle ds + \alpha \langle g, \nu \rangle \\ &= -\left\langle \frac{g'(\cdot)}{\bar{Z}(t)}, \bar{\mu}(t) \right\rangle + \alpha \langle g, \nu \rangle, \end{aligned}$$

which verifies that (4.36) holds. ∎

Proof of Existence for Theorem 3.1, $\xi \in \mathcal{M}_F^{c,p}$. The function $\bar{\mu}(\cdot)$ defined via (4.30) clearly satisfies $\bar{\mu}(0) = \xi$ and condition (ii) in the definition of a fluid model solution. By Proposition 4.6, condition (i) holds and $t^* = \infty$, so that (iv) holds trivially. Finally, condition (iii) follows from Lemma 4.8 by integrating (4.36). ∎

4.3 Dependence on the Initial Condition

One consequence of our construction of a fluid model solution in Section 4.2 is that the mapping from the initial measure to the associated fluid model solution is measurable. We will need this rather technical fact in order to prove Theorem 3.2. Since the measurability argument depends on the construction, it is presented here. However, readers that are willing to accept this fact may forego an immediate reading of the proof in order to proceed to the convergence argument in Section 5.

Consider the Borel σ-algebra on $\mathcal{M}_F$ associated with the topology of weak convergence. Note that the sets $\mathcal{M}_F^c$ and $\mathcal{M}_F^{c,p}$ are measurable subsets of $\mathcal{M}_F$. The set $\mathcal{M}_F^{c,p}$ is endowed with the relative topology of weak convergence inherited from $\mathcal{M}_F$. Since $\mathcal{M}_F^{c,p}$ is measurable, a relatively open set in $\mathcal{M}_F^{c,p}$ is measurable in $\mathcal{M}_F$. Let $\Xi_p : \mathcal{M}_F^{c,p} \longrightarrow D([0, \infty), \mathcal{M}_F)$ be given by $\Xi_p(\xi) = \bar{\mu}(\cdot)$ for $\xi \in \mathcal{M}_F^{c,p}$, where $\bar{\mu}(\cdot)$ is defined via (4.30). Then for all $t \geq 0$ and $0 \leq x < w \leq \infty$,

$$\left\langle 1_{[x,w)}, \bar{\mu}(t) \right\rangle = \left\langle 1_{[x,w)}(\cdot - \bar{S}(t)), \xi \right\rangle + \alpha \int_0^t \left\langle 1_{[x,w)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds.$$

For $x \neq 0$, this is simply (4.32). To see that the above holds for $x = 0$, use (4.30) and the following facts: $\bar{\mu}(t)$ does not charge the origin for all $t \geq 0$, ξ has no atoms, $\bar{S}(\cdot)$ is strictly increasing, and

ν has at most countably many atoms. Since $\{[x, w) : 0 \le x < w < \infty\} \cup \varnothing$ forms a π-system, Dynkin's π-λ theorem implies that, for each $t \ge 0$

$$\langle 1_B, \bar{\mu}(t)\rangle = \langle 1_B(\cdot - \bar{S}(t)), \xi\rangle + \alpha \int_0^t \langle 1_B(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu\rangle ds, \qquad \text{for all Borel sets } B \subset \mathbb{R}_+.$$

As a shorthand notation, we write

$$\bar{\mu}(t) = \xi \circ \theta_{\bar{S}(t)} + \int_0^t \nu \circ \theta_{\bar{S}(t)-\bar{S}(s)} ds, \tag{4.41}$$

where for each Borel set $B \subset \mathbb{R}_+$, $\langle 1_B, \xi \circ \theta_{\bar{S}(t)}\rangle = \langle 1_B(\cdot - \bar{S}(t)), \xi\rangle$ and

$$\left\langle 1_B, \int_0^t \nu \circ \theta_{\bar{S}(t)-\bar{S}(s)} ds \right\rangle = \int_0^t \left\langle 1_B(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu\right\rangle ds.$$

Finally, let $\Xi : \mathcal{M}_F^c \longrightarrow D([0,\infty), \mathcal{M}_F)$ be given by $\Xi(\xi) = \bar{\mu}(\cdot)$, where $\bar{\mu}(\cdot) = \Xi_p(\xi)$ if $\xi \ne \mathbf{0}$ and $\bar{\mu}(\cdot) \equiv \mathbf{0}$ if $\xi = \mathbf{0}$.

Lemma 4.9 *The mapping Ξ_p is continuous, and Ξ is measurable.*

Before proceeding with the proof of Lemma 4.9, some comments are in order. Notice that the inverse image under the mapping Ξ of the function that is identically equal to the zero measure is simply the zero measure, which is a measurable subset of $\mathcal{M}_F^c$. Therefore, in order to show that Ξ is measurable, it suffices to show that the inverse images under the mapping Ξ of measurable subsets of $D([0,\infty), \mathcal{M}_F)$ that do not contain the function that is identically equal to the zero measure are measurable subsets of $\mathcal{M}_F^{c,p}$. This holds if Ξ_p is continuous. Thus measurability of Ξ is an immediate consequence of the fact that Ξ_p is continuous, which is verified in the proof of Lemma 4.9 below.

This naturally raises the question as to whether or not the mapping Ξ is continuous, i.e., is the mapping Ξ continuous at the zero measure. This turns out to be true, however the proof requires techniques not used elsewhere in this paper. In particular, the proof exploits a certain order preservation property of the fluid model dynamics. The interested reader can find both the statement of the order preservation property and a proof that Ξ is continuous at the zero measure in [23].

Proof of Lemma 4.9. As explained in the paragraph following the statement of Lemma 4.9, measurability of Ξ follows immediately from continuity of Ξ_p. Thus, it suffices to show that Ξ_p is continuous. In order to do that, consider the spaces

$$\begin{aligned} \mathbf{C}_{\nearrow} &= \{U : [0,\infty) \longrightarrow [0,\infty) : U \text{ is continuous, nondecreasing, and } U(0) = 0\}, \\ \mathbf{C}_{\uparrow} &= \{U \in \mathbf{C}_{\nearrow} : U \text{ is strictly increasing}\}, \text{ and} \\ \mathbf{C}_{\uparrow,\infty} &= \{U \in \mathbf{C}_{\uparrow} : \lim_{u\to\infty} U(u) = \infty\}, \end{aligned}$$

each of which is endowed with the topology of uniform convergence on compact sets. Note that, since the functions in these spaces are all nondecreasing, uniform convergence on compact sets is equivalent to pointwise convergence for each of these spaces. The proof that Ξ_p is continuous proceeds by proving continuity of various intermediate maps to and from these spaces.

Let $\Phi : \mathcal{M}_F^{c,p} \longrightarrow \mathbf{C}_{\uparrow,\infty}$ be defined by $\Phi(\xi) = \bar{S}(\cdot)$, where $\bar{S}(t) = \bar{T}^{-1}(t) = \inf\{u \ge 0 : \bar{T}(u) > t\}$, $t \ge 0$, is the functional inverse of $\bar{T}(\cdot) = (H_\xi * U_e)(\cdot)$ with $H_\xi(\cdot)$ defined by (4.1) and $U_e(\cdot)$ defined by (4.5). The first step of the proof is to show that Φ is continuous. In order to do this, we define two maps Φ_1 and Φ_2 given by

$$\Phi_1 : \mathcal{M}_F^{c,p} \longrightarrow \mathbf{C}_{\uparrow,\infty} \text{ such that } \Phi_1(\xi) = \bar{T}(\cdot),$$
$$\Phi_2 : \mathbf{C}_{\uparrow,\infty} \longrightarrow \mathbf{C}_{\uparrow,\infty} \text{ such that } \Phi_2(\bar{T}(\cdot)) = \bar{S}(\cdot).$$

Then $\Phi = \Phi_2 \circ \Phi_1$. Therefore, it suffices to show that Φ_1 and Φ_2 are continuous.

We begin by showing that Φ_1 is continuous. In order to do this, let $\Phi_{1,1} : \mathcal{M}_F^{c,p} \longrightarrow \mathbf{C}_{\nearrow}$ be given by $\Phi_{1,1}(\xi) = H_\xi(\cdot)$, with $H_\xi(\cdot)$ defined by (4.1). It is easily verified that $H_\xi(x) = \langle \chi \wedge x, \xi \rangle$, $x \in \mathbb{R}_+$. Note that for each fixed $x \in \mathbb{R}_+$, $\chi(\cdot) \wedge x \in \mathbf{C}_b(\mathbb{R}_+)$. Therefore, if $\xi_n \in \mathcal{M}_F^{c,p}$, $n \geq 1$, $\xi \in \mathcal{M}_F^{c,p}$, and $\xi_n \xrightarrow{\text{w}} \xi$, it follows that H_{ξ_n} converges pointwise to H_ξ. Thus, $\Phi_{1,1}$ is continuous since pointwise convergence of functions in $\mathbf{C}_{\nearrow}$ to a limit in $\mathbf{C}_{\nearrow}$ implies uniform convergence on compact sets. Also, let $\Phi_{1,2} : \mathbf{C}_{\nearrow} \longrightarrow \mathbf{C}_{\uparrow,\infty}$ be given by $\Phi_{1,2}(H(\cdot)) = (H * U_e)(\cdot)$. It is perhaps not surprising that $\Phi_{1,2}$ is also continuous. To see this, consider $H_n \in \mathbf{C}_{\nearrow}$, $n \geq 1$, and $H \in \mathbf{C}_{\nearrow}$ such that $H_n \longrightarrow H$ uniformly on compact sets as n tends to infinity. For each $K \geq 0$,

$$\| (H_n * U_e) - (H * U_e) \|_K \leq \|H_n - H\|_K U_e(K),$$

which tends to zero as n tends to infinity. Thus, $\Phi_{1,2}$ is continuous. Since $\Phi_1 = \Phi_{1,2} \circ \Phi_{1,1}$, it follows that Φ_1 is continuous.

Next we show that Φ_2 is continuous. To see this, consider $\bar{T}_n \in \mathbf{C}_{\uparrow,\infty}$, $n \geq 1$, and $\bar{T} \in \mathbf{C}_{\uparrow,\infty}$ such that $\bar{T}_n \longrightarrow \bar{T}$ uniformly on compact sets as n tends to infinity. Let $\bar{S}_n(\cdot) = \Phi_2(\bar{T}_n(\cdot))$, $n \geq 1$, and $\bar{S}(\cdot) = \Phi_2(\bar{T}(\cdot))$. It suffices to show that $\bar{S}_n(t) \longrightarrow \bar{S}(t)$ as n tends to infinity for each $t \geq 0$. Fix $t \geq 0$. For each $n \geq 1$, choose u_n such that $\bar{T}_n(u_n) = t$ and choose u such that $\bar{T}(u) = t$. In other words, $\bar{S}_n(t) = u_n$, $n \geq 1$, and $\bar{S}(t) = u$, and consequently it suffices to show that $u_n \to u$. Using the fact that $\bar{T}(\cdot)$ is strictly increasing, it can be verified that the sequence $\{u_n\}_{n\geq 1}$ is bounded. This together with the fact that $\bar{T}_n \longrightarrow \bar{T}$ uniformly on compact sets as n tends to infinity and $\bar{T}_n(u_n) = \bar{T}(u)$, $n \geq 1$, implies that $\lim_{n\to\infty} \bar{T}(u_n) = \bar{T}(u)$. Again using the fact that $\bar{T}(\cdot)$ is strictly increasing, it then follows that $u_n \to u$ as n tends to infinity.

We now turn our attention to proving that Ξ_p is continuous. Suppose that $\{\xi_n\}_{n\geq 1} \subset \mathcal{M}_F^{c,p}$, $\xi \in \mathcal{M}_F^{c,p}$, and $\xi_n \xrightarrow{\text{w}} \xi$. We will show that $\Xi_p(\xi_n) \longrightarrow \Xi_p(\xi)$ in the Skorohod J_1-topology. For this, note that the Prohorov metric defined on the set of Borel probability measures on $\mathbb{R}_+$ naturally extends to a metric ρ on $\mathcal{M}_F$ under which $\mathcal{M}_F$ is a Polish space. In particular, for $\zeta_1, \zeta_2 \in \mathcal{M}_F$, $\rho(\zeta_1, \zeta_2)$ is given by

$$\rho(\zeta_1,\zeta_2) = \inf\{\epsilon > 0 : \langle 1_B, \zeta_1\rangle \leq \langle 1_{B^\epsilon}, \zeta_2\rangle + \epsilon, \langle 1_B, \zeta_2\rangle \leq \langle 1_{B^\epsilon}, \zeta_1\rangle + \epsilon \text{ for all closed sets } B \subset \mathbb{R}_+\},$$

where $B^\epsilon = \{x \in \mathbb{R}_+ : \inf_{y\in B} |x-y| < \epsilon\}$. Moreover, if $\{\zeta_n\}_{n\geq 1} \subset \mathcal{M}_F$ and $\zeta \in \mathcal{M}_F$, then $\zeta_n \xrightarrow{\text{w}} \zeta$ if and only if $\lim_{n\to\infty} \rho(\zeta_n, \zeta) = 0$ (cf. [15], Ch. 3, Thms. 1.7 & 3.1, which readily generalize from probability measures to $\mathcal{M}_F$.). In order to show that $\Xi_p(\xi_n) \longrightarrow \Xi_p(\xi)$ in the Skorohod J_1-topology, it suffices to show that there exists $\{\lambda_n\}_{n\geq 1} \subset \mathbf{C}_{\uparrow,\infty}$ such that for each $K \geq 0$,

$$\lim_{n\to\infty} \sup_{0\leq t\leq K} |\lambda_n(t) - t| = 0, \tag{4.42}$$

$$\lim_{n\to\infty} \sup_{0\leq t\leq K} \rho(\xi_n \circ \theta_{\bar{S}_n(\lambda_n(t))}, \xi \circ \theta_{\bar{S}(t)}) = 0, \tag{4.43}$$

$$\lim_{n\to\infty} \sup_{0\leq t\leq K} \rho\left(\int_0^{\lambda_n(t)} \nu \circ \theta_{\bar{S}_n(\lambda_n(t)) - \bar{S}_n(s)} ds, \int_0^t \nu \circ \theta_{\bar{S}(t) - \bar{S}(s)} ds \right) = 0, \tag{4.44}$$

where $\bar{S}_n(\cdot) = \Phi(\xi_n)$, $n \geq 1$, and $\bar{S}(\cdot) = \Phi(\xi)$. Here we have used the fact that $\rho(\zeta_{11}+\zeta_{12}, \zeta_{21}+\zeta_{22}) \leq \rho(\zeta_{11}, \zeta_{21}) + \rho(\zeta_{12}, \zeta_{22})$, $\zeta_{ij} \in \mathcal{M}_F$, $i, j = 1, 2$, (4.41) above, and Prop. 5.3 and Rmk. 5.4 in Ch. 3 of [15].

For $n \geq 1$, let $\lambda_n(t) = \bar{T}_n(\bar{S}(t))$ for all $t \geq 0$, where $\bar{T}_n(\cdot) = \Phi_1(\xi_n)$. Since $\bar{T}_n \in \mathbf{C}_{\uparrow,\infty}$, $n \geq 1$, and $\bar{S} \in \mathbf{C}_{\uparrow,\infty}$, it follows that $\lambda_n \in \mathbf{C}_{\uparrow,\infty}$ for each $n \geq 1$. Since Φ_1 is continuous, for $\bar{T} = \Phi_1(\xi)$, $\bar{T}_n \longrightarrow \bar{T}$ uniformly on compact sets as $n \to \infty$. This together with $\bar{T}(\bar{S}(t)) = t$, $t \geq 0$, and the fact that $\bar{S} \in \mathbf{C}_{\uparrow,\infty}$ implies that (4.42) holds for all $K \geq 0$. Note that $\bar{S}_n(\lambda_n(t)) = \bar{S}(t)$, $t \geq 0$, which can be used to simplify the expressions in (4.43) and (4.44).

We now verify (4.43). Let $\tilde{\epsilon} > 0$ and choose $N_{\tilde{\epsilon}}$ such that $n \geq N_{\tilde{\epsilon}}$ implies that $\rho(\xi_n, \xi) \leq \tilde{\epsilon}$. For a closed set $B \subset \mathbb{R}_+$, $B + \bar{S}(t)$ is also closed, and so, for $n \geq N_{\tilde{\epsilon}}$,

$$\langle 1_{B+\bar{S}(t)}, \xi_n \rangle \leq \langle 1_{(B+\bar{S}(t))^{\tilde{\epsilon}}}, \xi \rangle + \tilde{\epsilon} \qquad \text{and} \qquad \langle 1_{B+\bar{S}(t)}, \xi \rangle \leq \langle 1_{(B+\bar{S}(t))^{\tilde{\epsilon}}}, \xi_n \rangle + \tilde{\epsilon}. \tag{4.45}$$

Note that

$$\left(B + \bar{S}(t)\right)^{\tilde{\epsilon}} \subset \left(B^{\tilde{\epsilon}} + \bar{S}(t)\right) \cup ((\bar{S}(t) - \tilde{\epsilon})^+, \bar{S}(t)). \tag{4.46}$$

Combining (4.45), (4.46), and the definitions of $\xi_n \circ \theta_{\bar{S}(t)}$ and $\xi \circ \theta_{\bar{S}(t)}$ yields that for $n \geq N_{\tilde{\epsilon}}$

$$\left\langle 1_B, \xi_n \circ \theta_{\bar{S}(t)} \right\rangle \leq \left\langle 1_{B^{\tilde{\epsilon}}}, \xi \circ \theta_{\bar{S}(t)} \right\rangle + \left\langle 1_{((\bar{S}(t)-\tilde{\epsilon})^+, \bar{S}(t))}, \xi \right\rangle + \tilde{\epsilon}, \tag{4.47}$$

$$\left\langle 1_B, \xi \circ \theta_{\bar{S}(t)} \right\rangle \leq \left\langle 1_{B^{\tilde{\epsilon}}}, \xi_n \circ \theta_{\bar{S}(t)} \right\rangle + \left\langle 1_{((\bar{S}(t)-\tilde{\epsilon})^+, \bar{S}(t))}, \xi_n \right\rangle + \tilde{\epsilon}. \tag{4.48}$$

Since $\rho(\xi_n, \xi) \leq \tilde{\epsilon}$ for $n \geq N_{\tilde{\epsilon}}$, and since $\xi(\{0\}) = 0$, it follows that $\left\langle 1_{[(\bar{S}(t)-\tilde{\epsilon})^+, \bar{S}(t)]}, \xi_n \right\rangle \leq \left\langle 1_{((\bar{S}(t)-2\tilde{\epsilon})^+, \bar{S}(t)+\tilde{\epsilon})}, \xi \right\rangle + \tilde{\epsilon}$ for $n \geq N_{\tilde{\epsilon}}$. This together with (4.48) implies that for $n \geq N_{\tilde{\epsilon}}$

$$\left\langle 1_B, \xi \circ \theta_{\bar{S}(t)} \right\rangle \leq \left\langle 1_{B^{\tilde{\epsilon}}}, \xi_n \circ \theta_{\bar{S}(t)} \right\rangle + \left\langle 1_{((\bar{S}(t)-2\tilde{\epsilon})^+, \bar{S}(t)+\tilde{\epsilon})}, \xi \right\rangle + 2\tilde{\epsilon}. \tag{4.49}$$

Since ξ has no atoms, $\lim_{\tilde{\epsilon}\to 0} \sup_{t\geq 0} \left\langle 1_{((\bar{S}(t)-2\tilde{\epsilon})^+, \bar{S}(t)+\tilde{\epsilon})}, \xi \right\rangle = 0$ (cf. proof of Lemma A.1). Therefore, given $\epsilon > 0$, there exists $\tilde{\epsilon} > 0$ such that $\left\langle 1_{((\bar{S}(t)-2\tilde{\epsilon})^+, \bar{S}(t)+\tilde{\epsilon})}, \xi \right\rangle + 2\tilde{\epsilon} \leq \epsilon$ for all $t \geq 0$. Since $\tilde{\epsilon} \leq \epsilon$ and $((\bar{S}(t) - \tilde{\epsilon})^+, \bar{S}(t)) \subset ((\bar{S}(t) - 2\tilde{\epsilon})^+, \bar{S}(t) + \tilde{\epsilon})$, from (4.47) and (4.49) it follows that, for all $t \geq 0$, $n \geq N_{\tilde{\epsilon}}$, and closed $B \subset \mathbb{R}_+$,

$$\left\langle 1_B, \xi_n \circ \theta_{\bar{S}(t)} \right\rangle \leq \left\langle 1_{B^\epsilon}, \xi \circ \theta_{\bar{S}(t)} \right\rangle + \epsilon \qquad \text{and} \qquad \left\langle 1_B, \xi \circ \theta_{\bar{S}(t)} \right\rangle \leq \left\langle 1_{B^\epsilon}, \xi_n \circ \theta_{\bar{S}(t)} \right\rangle + \epsilon.$$

Since $\bar{S}_n(\lambda_n(t)) = \bar{S}(t)$, $t \geq 0$, and since $\epsilon > 0$ was arbitrary, (4.43) holds.

We now verify (4.44). Fix $K, \epsilon > 0$. Let $\tilde{\epsilon} \in (0, \epsilon)$. Since $\bar{S}_n(\lambda_n(t)) = \bar{S}(t)$, $t \geq 0$, for a Borel set $B \subset \mathbb{R}_+$, $t \geq 0$, and $n \geq 1$,

$$\left\langle 1_B, \int_0^t \nu \circ \theta_{\bar{S}(t)-\bar{S}(s)} ds \right\rangle = \int_0^t \left\langle 1_B(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds,$$

$$\left\langle 1_B, \int_0^{\lambda_n(t)} \nu \circ \theta_{\bar{S}_n(\lambda_n(t))-\bar{S}_n(s)} ds \right\rangle = \int_0^{\lambda_n(t)} \left\langle 1_B(\cdot - (\bar{S}(t) - \bar{S}_n(s))), \nu \right\rangle ds.$$

By (4.42) and the continuity of Φ, there exists N such that $n \geq N$ implies that $\sup_{0\leq t\leq K} |\lambda_n(t) - t| < \epsilon/2$ and $\|\bar{S}_n - \bar{S}\|_{K+\epsilon/2} < \tilde{\epsilon}$. Fix such an N. Then, since $\tilde{\epsilon} < \epsilon$, it follows that for $n \geq N$,

$t \in [0,K]$, $s \in [0,t]$, and $x \in \mathbb{R}_+$ such that $x - \bar{S}(t) + \bar{S}(s) \in B$, either $x - \bar{S}(t) + \bar{S}(s) \in [0,\tilde{\epsilon})$ or $x - \bar{S}(t) + \bar{S}_n(s) \in B^\epsilon$. Similarly, for $n \geq N$, $t \in [0,K]$, $s \in [0,\lambda_n(t)]$, and $x \in \mathbb{R}_+$ such that $x - \bar{S}(t) + \bar{S}_n(s) \in B$, either $x - \bar{S}(t) + \bar{S}(s) \in (-\tilde{\epsilon},0)$ or $x - \bar{S}(t) + \bar{S}(s) \in B^\epsilon$. This gives, for $n \geq N$ and $t \in [0,K]$,

$$\begin{aligned}
\left\langle 1_B, \int_0^t \nu \circ \theta_{\bar{S}(t)-\bar{S}(s)} ds \right\rangle &\leq \int_0^t \left\langle 1_{B^\epsilon}(\cdot - (\bar{S}(t) - \bar{S}_n(s))), \nu \right\rangle ds \\
&\quad + \int_0^t \left\langle 1_{[0,\tilde{\epsilon})}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds, \\
\left\langle 1_B, \int_0^{\lambda_n(t)} \nu \circ \theta_{\bar{S}_n(\lambda_n(t))-\bar{S}_n(s)} ds \right\rangle &\leq \int_0^{\lambda_n(t)} \left\langle 1_{(-\tilde{\epsilon},0)\cup B^\epsilon}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds.
\end{aligned}$$

Since each of the integrands on the right side is bounded above by 1, since $\bar{S}(t) = \bar{S}_n(\lambda_n(t))$, and since, for $n \geq N$ and $t \in [0,K]$, $|\lambda_n(t) - t| < \epsilon/2$, it follows that for $n \geq N$ and $t \in [0,K]$,

$$\begin{aligned}
\left\langle 1_B, \int_0^t \nu \circ \theta_{\bar{S}(t)-\bar{S}(s)} ds \right\rangle &\leq \int_0^{\lambda_n(t)} \left\langle 1_{B^\epsilon}(\cdot - (\bar{S}_n(\lambda_n(t)) - \bar{S}_n(s))), \nu \right\rangle ds + \epsilon/2 \\
&\quad + \int_0^t \left\langle 1_{[0,\tilde{\epsilon})}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds, \qquad (4.50) \\
\left\langle 1_B, \int_0^{\lambda_n(t)} \nu \circ \theta_{\bar{S}_n(\lambda_n(t))-\bar{S}_n(s)} ds \right\rangle &\leq \int_0^t \left\langle 1_{B^\epsilon}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds + \epsilon/2 \\
&\quad + \int_0^t \left\langle 1_{(-\tilde{\epsilon},0)}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds. \qquad (4.51)
\end{aligned}$$

Thus, it suffices to show that there exists $\tilde{\epsilon} \in (0,\epsilon)$ such that, for all $t \in [0,K]$,

$$\int_0^t \left\langle 1_{(-\tilde{\epsilon},\tilde{\epsilon})}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds \leq \epsilon/2.$$

This requires some care since ν can have atoms. Let $A \subset \mathbb{R}_+$ denote the set containing all of the atoms of ν, which is at most countable, and let $\nu_d = \sum_{a\in A} \nu(\{a\})\delta_a$ be the Borel measure comprised of only the atoms of ν. Then $\nu_c = \nu - \nu_d$ has no atoms. Therefore, there exists $\tilde{\epsilon}_1 \in (0,\epsilon)$ such that $\sup_{y\in\mathbb{R}_+} \langle 1_{(y-\tilde{\epsilon}_1, y+\tilde{\epsilon}_1)}, \nu_c \rangle < \epsilon/6K$ (cf. the proof of Lemma A.1). Thus, for $t \in [0,K]$ and $\tilde{\epsilon} \in (0,\tilde{\epsilon}_1]$, we have

$$\begin{aligned}
&\int_0^t \left\langle 1_{(-\tilde{\epsilon},\tilde{\epsilon})}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds \\
&= \int_0^t \left\langle 1_{(\bar{S}(t)-\bar{S}(s)-\tilde{\epsilon}, \bar{S}(t)-\bar{S}(s)+\tilde{\epsilon})}, \nu_c \right\rangle ds + \int_0^t \left\langle 1_{(\bar{S}(t)-\bar{S}(s)-\tilde{\epsilon}, \bar{S}(t)-\bar{S}(s)+\tilde{\epsilon})}, \nu_d \right\rangle ds \\
&\leq \frac{\epsilon}{6K} t + \sum_{a\in A} \nu(\{a\}) \int_0^t \left\langle 1_{(\bar{S}(t)-\bar{S}(s)-\tilde{\epsilon}, \bar{S}(t)-\bar{S}(s)+\tilde{\epsilon})}, \delta_a \right\rangle ds.
\end{aligned}$$

Since $\sum_{a\in A} \nu(\{a\}) \leq 1$, there exists a finite set $A_\epsilon \subset A$ such that $\sum_{a\in A\setminus A_\epsilon} \nu(\{a\}) \leq \epsilon/6K$. For

$t \in [0,K]$ and $\tilde{\epsilon} \in (0, \tilde{\epsilon}_1]$, we have

$$
\begin{aligned}
&\int_0^t \left\langle 1_{(-\tilde{\epsilon},\tilde{\epsilon})}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds \\
&\le \frac{\epsilon}{6K} t + \sum_{a \in A \setminus A_\epsilon} \nu(\{a\}) t + \sum_{a \in A_\epsilon} \nu(\{a\}) \int_0^t \left\langle 1_{(\bar{S}(t)-\bar{S}(s)-\tilde{\epsilon}, \bar{S}(t)-\bar{S}(s)+\tilde{\epsilon})}, \delta_a \right\rangle ds \\
&\le \frac{\epsilon}{6} + \frac{\epsilon}{6} + \sum_{a \in A_\epsilon} \nu(\{a\}) \int_0^t \left\langle 1_{(\bar{S}(t)-\bar{S}(s)-\tilde{\epsilon}, \bar{S}(t)-\bar{S}(s)+\tilde{\epsilon})}, \delta_a \right\rangle ds.
\end{aligned}
$$

Note that for $t \in [0,K]$, $s \in [0,t]$, and $a \in A_\epsilon$ such that $a \in (\bar{S}(t) - \bar{S}(s) - \tilde{\epsilon}, \bar{S}(t) - \bar{S}(s) + \tilde{\epsilon})$,

$$
s \in \left[\bar{T}\left((\bar{S}(t) - a - \tilde{\epsilon})^+ \right), \bar{T}\left((\bar{S}(t) - a + \tilde{\epsilon})^+ \right) \right].
$$

If $A_\epsilon \neq \emptyset$, let $\tilde{\epsilon} \in (0, \tilde{\epsilon}_1]$ be such that

$$
\max_{a \in A_\epsilon} \sup_{t \in [0,K]} \bar{T}\left((\bar{S}(t) - a + \tilde{\epsilon})^+ \right) - \bar{T}\left((\bar{S}(t) - a - \tilde{\epsilon})^+ \right) \le \frac{\epsilon}{6|A_\epsilon|},
$$

where $|A_\epsilon|$ denotes the number of atoms in A_ϵ. Otherwise, let $\tilde{\epsilon} = \tilde{\epsilon}_1$. Then, for $t \in [0,K]$,

$$
\int_0^t \left\langle 1_{(-\tilde{\epsilon},\tilde{\epsilon})}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds \le \frac{\epsilon}{6} + \frac{\epsilon}{6} + \frac{\epsilon}{6} = \frac{\epsilon}{2}.
$$

This together with (4.50) and (4.51) gives, for $n \ge N$ and $t \in [0,K]$,

$$
\left\langle 1_B, \int_0^t \nu \circ \theta_{\bar{S}(t)-\bar{S}(s)} ds \right\rangle \le \left\langle 1_{B^\epsilon}, \int_0^{\lambda_n(t)} \nu \circ \theta_{\bar{S}_n(\lambda_n(t))-\bar{S}_n(s)} ds \right\rangle + \epsilon, \tag{4.52}
$$

$$
\left\langle 1_B, \int_0^{\lambda_n(t)} \nu \circ \theta_{\bar{S}_n(\lambda_n(t))-\bar{S}_n(s)} ds \right\rangle \le \int_0^t \left\langle 1_{B^\epsilon}(\cdot - (\bar{S}(t) - \bar{S}(s))), \nu \right\rangle ds + \epsilon. \tag{4.53}
$$

Since $K, \epsilon > 0$ were arbitrary, (4.52) and (4.53) imply (4.44). ■

5 Convergence to Fluid Model Solutions

The objective of this section is to prove the following theorem, and then apply it to prove Theorem 3.2.

Theorem 5.1 *Consider a sequence of processor sharing queues as defined in Section 3.3, satisfying assumptions* (3.14)–(3.22)*. Then the sequence of fluid scaled measure valued processes $\{\bar{\mu}^r(\cdot)\}$ is tight. Moreover, any limit point $\bar{\mu}^\star(\cdot)$ is a.s. a fluid model solution for the critical data (α, ν).*

Recall that the sequence $\{\bar{\mu}^r(\cdot)\}$ is tight if and only if the associated sequence of probability laws on $D([0,\infty), \mathcal{M}_F)$ is tight. The term "limit point" in the above statement refers to any limit in distribution along some subsequence of $\{\bar{\mu}^r(\cdot)\}$. We use this terminology because our objective is to show that all such limit points have the same law. This uniqueness in law will hinge on the almost sure characterization of limit points as fluid model solutions. Throughout this section we assume that the conditions in Section 3.3 hold. In Section 5.1 we lay the groundwork for our main proofs by compiling some basic results, and making some rather technical choices of various constants. Tightness of $\{\bar{\mu}^r(\cdot)\}$ is proved in Section 5.2, and the fact that limit points are a.s. fluid model solutions is proved in Section 5.3. Finally, the proof of Theorem 3.2 appears in Section 5.4.

5.1 Preliminaries

To expedite the proofs later on, we first establish some basic consequences of the assumptions (3.14)–(3.22) of Section 3.3, and make some judicious choices of various constants. For subsequent reference, the results of the following discussion are summarized at the end of this section, in Lemma 5.2. For the remainder of this section, let $T > 0$ and $0 < \epsilon, \eta < 1$ be fixed, and define $\tilde{\eta} = \eta/8$.

A dynamic equation. We begin by specifying a dynamic equation satisfied by the fluid scaled state descriptor $\bar{\mu}^r(\cdot)$. Starting with (2.13) and substituting in the definition of the residual service times, one obtains after some simplification that for each r, a.s. for each bounded, Borel measurable function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$, and all $t, h \geq 0$,

$$\langle g, \mu^r(t+h)\rangle = \left\langle \left(1_{(0,\infty)} g\right)(\cdot - S^r_{t,t+h}), \mu^r(t)\right\rangle + \sum_{i=E^r(t)+1}^{E^r(t+h)} \left(1_{(0,\infty)} g\right)(v^r_i - S^r_{U^r_i,t+h}). \tag{5.1}$$

Recall that we always assume g is extended to be identically zero on $(-\infty, 0)$ so that functions of the form $g(\cdot - a)$ are well defined on $\mathbb{R}_+$ for any $a > 0$. Notice that this equation provides a decomposition of the quantity $\langle g, \mu^r(t+h)\rangle$ into a component arising from jobs already in the system at time t, and a component resulting from the arrival of new jobs during the interval $(t, t+h]$. Intuitively, since $S^r_{t,t+h}$ gives the cumulative service received by each job in the system during the interval $(t, t+h]$, the measure $\mu^r(t+h)$ should be obtained from $\mu^r(t)$ by shifting the latter measure to the left by $S^r_{t,t+h}$, and removing any mass that ends up in $(-\infty, 0]$. This explains the shift and truncation of g in the first term on the right side above. Similarly, a job arriving at time $U^r_i \in (t, t+h]$ will receive a cumulative amount of service equal to $S^r_{U^r_i,t+h} \wedge v^r_i$ up to time $t+h$, and so v^r_i must be shifted to the left by this amount before contributing to the integral $\langle g, \mu^r(t+h)\rangle$. This explains the second term on the right side.

Equation (5.1) takes the following form for the fluid scaled processes. For each r, a.s. for each bounded, Borel measurable function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$, and all $t, h \geq 0$,

$$\langle g, \bar{\mu}^r(t+h)\rangle = \left\langle \left(1_{(0,\infty)} g\right)(\cdot - \bar{S}^r_{t,t+h}), \bar{\mu}^r(t)\right\rangle + \frac{1}{r}\sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+h)} \left(1_{(0,\infty)} g\right)(v^r_i - \bar{S}^r_{U^r_i/r,t+h}). \tag{5.2}$$

We refer to (5.2) as the *dynamic equation* for $\bar{\mu}^r(\cdot)$, and it is the equation we will use to analyze the behavior of $\bar{\mu}^r(\cdot)$. Frequently we will set $g \equiv 1$ in this equation, in which case the first term on the right side will look like $\left\langle 1_{(\bar{S}^r_{t,t+h},\infty)}, \bar{\mu}^r(t)\right\rangle$. Although we are ultimately only interested in functions $g \in \mathbf{C}^1_b(\mathbb{R}_+)$ when we use (5.2) to show convergence in distribution of the fluid scaled state descriptors, we will need to work with discontinuous functions at various stages along the way, which is why we include this possibility in equation (5.2).

Functional weak laws of large numbers; the constant l. We now establish some basic functional weak law of large numbers estimates which arise from our asymptotic assumptions (3.14)–(3.18) on the arrival and service processes.

First note that by Lemma A.2 (cf. Appendix) and (3.14)–(3.18), we have the following functional weak law:

$$\frac{1}{r}\sum_{i=1}^{r\bar{E}^r(\cdot)} g(v^r_i) \Longrightarrow \alpha(\cdot)\langle g, \nu\rangle, \qquad \text{as } r \to \infty, \tag{5.3}$$

for each nonnegative Borel measurable function $g : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ that is ν-a.e. continuous and satisfies $\langle g, \nu \rangle < \infty$, $\langle g, \nu^r \rangle < \infty$ for each $r > 0$, and

$$\langle g, \nu^r \rangle \longrightarrow \langle g, \nu \rangle, \qquad \text{as } r \to \infty. \tag{5.4}$$

Here $\alpha(t) = \alpha t$ for all $t \geq 0$. Moreover, since the limit in (5.3) is deterministic, the convergence in distribution there is joint with (3.19).

By (3.15), any nonnegative $g \in \mathbf{C}_b(\mathbb{R}_+)$ satisfies (5.4). In particular, for such a g and any $0 < l \leq T$ we have

$$\lim_{r\to\infty} \mathbf{P}\left(\sup_{t\in[0,T-l]} \frac{1}{r} \sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} g(v_i^r) \leq 2\alpha l \langle g, \nu \rangle \right) = 1. \tag{5.5}$$

With $l = T$ and $g \equiv 1$, this implies

$$\lim_{r\to\infty} \mathbf{P}\left(\bar{E}^r(T) \leq 2\alpha T\right) = 1. \tag{5.6}$$

For the rest of this section, we fix l such that $0 < l < \epsilon/16\alpha$ and T is an integer multiple of l. Then by (5.5), for any $g \in \mathbf{C}_b(\mathbb{R}_+)$,

$$\lim_{r\to\infty} \mathbf{P}\left(\sup_{t\in[0,T-l]} \frac{1}{r} \sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} g(v_i^r) \leq \frac{\epsilon}{8} \langle g, \nu \rangle \right) = 1. \tag{5.7}$$

We will use (5.3) repeatedly for several choices of g below, as well as (5.7) for several choices of $g \in \mathbf{C}_b(\mathbb{R}_+)$. To begin with, let $g \equiv 1$. Then our choice of l implies by (5.7) that

$$\lim_{r\to\infty} \mathbf{P}\left(\sup_{t\in[0,T-l]} \left(\bar{E}^r(t+l) - \bar{E}^r(t)\right) \leq \frac{\epsilon}{4} \right) = 1, \tag{5.8}$$

where for convenience we have relaxed the bound from $\epsilon/8$ to $\epsilon/4$ and used the fact that ν is a probability measure.

Global bounds for the initial condition; the constants M_0, M_T. By (3.19)–(3.21), we can choose an $M_0 > 0$ such that

$$\liminf_{r\to\infty} \mathbf{P}\big(\langle 1, \bar{\mu}^r(0)\rangle \vee \langle \chi, \bar{\mu}^r(0)\rangle < M_0\big) \geq 1 - \tilde{\eta}, \tag{5.9}$$

and

$$\mathbf{P}\Big(\big\langle 1_{[M_0,\infty)}, \Theta \big\rangle < \frac{\epsilon}{4}\Big) \geq 1 - \frac{\tilde{\eta}}{2}. \tag{5.10}$$

Define $M_T = M_0 + 2\alpha T$.

Fine structure bounds for the initial condition; the constant κ. Given $\kappa > 0$, let $N_\kappa = \lceil M_0/\kappa \rceil$ and define the finite set of overlapping closed intervals $\{I_n\}_{n=0}^{N_\kappa}$ by $I_n = [n\kappa, (n+2)\kappa]$, for $n = 0, 1, \ldots, N_\kappa - 1$, and $I_{N_\kappa} = [N_\kappa \kappa, \infty)$. It is evident that for every $x \in \mathbb{R}_+$, there is an $n \in \{0, 1, \ldots, N_\kappa\}$ such that

$$[x, x+\kappa] \subset I_n. \tag{5.11}$$

By (5.10) and (3.23) (using 2κ instead of κ there), we can choose $0 < \kappa < \frac{l}{2M_T}$ so that

$$\mathbf{P}\left(\max_{n=0}^{N_\kappa}\{\langle 1_{I_n}, \Theta\rangle\} < \frac{\epsilon}{4}\right) \geq 1 - \tilde{\eta}.$$

Now define the set

$$A = \left\{\zeta \in \mathcal{M}_F : \max_{n=0}^{N_\kappa}\{\langle 1_{I_n}, \zeta\rangle\} < \frac{\epsilon}{4}\right\},$$

and suppose that $\{\zeta_k\}_{k=1}^{\infty} \subset \mathcal{M}_F$ with $\zeta_k \xrightarrow{\text{w}} \zeta \in A$, as $k \to \infty$. Then since each of the sets I_n is closed, a trivial generalization of the Portmanteau theorem (cf. [2], Thm. 2.1) to finite measures yields that for $n = 0, 1, \ldots, N_\kappa$,

$$\limsup_{k\to\infty} \ \langle 1_{I_n}, \zeta_k\rangle \leq \langle 1_{I_n}, \zeta\rangle < \frac{\epsilon}{4}.$$

Thus $\zeta_k \in A$ for sufficiently large k, which implies that $A \subset \mathcal{M}_F$ is an open set. This together with (3.19) and a second application of the Portmanteau theorem yields

$$\liminf_{r\to\infty} \mathbf{P}\big(\bar{\mu}^r(0) \in A\big) \geq \mathbf{P}\big(\Theta \in A\big) \geq 1 - \tilde{\eta},$$

which implies by (5.11) that

$$\liminf_{r\to\infty} \mathbf{P}\left(\sup_{x\in\mathbb{R}_+} \left\langle 1_{[x,x+\kappa]}, \bar{\mu}^r(0)\right\rangle < \frac{\epsilon}{4}\right) \geq 1 - \tilde{\eta}. \tag{5.12}$$

Convergence of the fluid scaled workload processes; the constant γ. It is well known that, for any queue operating under a nonidling service discipline, the fluid scaled workload processes $\bar{W}^r(\cdot)$ converge in distribution, under the assumptions (3.14)–(3.22), to a process that a.s. equals its initial value for all time. So, since processor sharing is a nonidling service discipline, the sequence of processes $\{\langle\chi, \bar{\mu}^r(\cdot)\rangle\}$ converges in distribution to a process that a.s. equals its initial value for all time. For completeness, we sketch the verification of this fact below. Consider the workload equation (2.2) on fluid scale. If we define the fluid scaled *load* processes

$$\bar{L}^r(t) = \langle\chi, \bar{\mu}^r(0)\rangle + \frac{1}{r}\sum_{i=1}^{r\bar{E}^r(t)} v_i^r - t, \tag{5.13}$$

then the fluid scaled workload equation can be rewritten for all $t \geq 0$ as

$$\langle\chi, \bar{\mu}^r(t)\rangle = \bar{L}^r(t) + \sup_{0\leq s\leq t} \left(\bar{L}^r(s)\right)^-. \tag{5.14}$$

Note that by (3.16), (5.3) holds with $g = \chi$. This together with the fact that $\alpha\,\langle\chi, \nu\rangle = 1$ implies by (5.13), (5.3) and (3.19) that $\bar{L}^r(\cdot)$ converges in distribution, as $r \to \infty$, to the process that a.s. equals its initial value for all time, where that initial value has the same distribution as $\langle\chi, \Theta\rangle$. The mapping $\Upsilon : D([0,\infty), \mathbb{R}) \longrightarrow D([0,\infty), \mathbb{R}_+)$ defined by $\Upsilon(x(\cdot))(t) = x(t) + \sup\{(x(s))^- : 0 \leq s \leq t\}$ is continuous, so the continuous mapping theorem applied to (5.14) implies the result.

We make use of the above fact in the following way. Having fixed l, M_0, M_T and κ, we now choose a $0 < \gamma < \min\{\kappa\epsilon/4, l/4, \alpha T\}$. Then the fact that as $r \to \infty$, $\langle \chi, \bar{\mu}^r(\cdot)\rangle$ converges in distribution to a process that is a.s. equal to its initial value implies that

$$\lim_{r\to\infty} \mathbf{P}\left(\sup_{t\in[0,T]} \left|\langle \chi, \bar{\mu}^r(t)\rangle - \langle \chi, \bar{\mu}^r(0)\rangle\right| < \gamma/4\right) = 1. \tag{5.15}$$

Tail estimate for the state descriptor; the constant K. For any $K > 0$ such that the distribution ν does not charge K, (3.15) implies that $\left\langle \chi 1_{[0,K]}, \nu^r\right\rangle \longrightarrow \left\langle \chi 1_{[0,K]}, \nu\right\rangle$ as $r \to \infty$, which implies by (3.16) that $\left\langle \chi 1_{(K,\infty)}, \nu^r\right\rangle \longrightarrow \left\langle \chi 1_{(K,\infty)}, \nu\right\rangle$ as $r \to \infty$. So using (5.3), we have for any such K that

$$\frac{1}{r}\sum_{i=1}^{r\bar{E}^r(\cdot)} v_i^r 1_{\{v_i^r > K\}} \Longrightarrow \alpha(\cdot)\left\langle \chi 1_{(K,\infty)}, \nu\right\rangle, \qquad \text{as } r \to \infty. \tag{5.16}$$

Since, by (3.22), Θ does not charge any K a.s., we have by (3.19) that for any $K > 0$,

$$\left\langle \chi 1_{[0,K]}, \bar{\mu}^r(0)\right\rangle \Longrightarrow \left\langle \chi 1_{[0,K]}, \Theta\right\rangle, \qquad \text{as } r \to \infty.$$

Then, using the joint convergence in (3.19) as well as (3.21), we have for any $K > 0$,

$$\left\langle \chi 1_{(K,\infty)}, \bar{\mu}^r(0)\right\rangle \Longrightarrow \left\langle \chi 1_{(K,\infty)}, \Theta\right\rangle, \qquad \text{as } r \to \infty.$$

Since $\left\langle \chi 1_{(K,\infty)}, \nu\right\rangle \longrightarrow 0$ and $\mathbf{E}[\left\langle \chi 1_{(K,\infty)}, \Theta\right\rangle] \longrightarrow 0$ as $K \to \infty$, we can choose a $K > 0$ sufficiently large so that

$$\liminf_{r\to\infty} \mathbf{P}\left(\sup_{t\in[0,T]} \frac{1}{r}\sum_{i=1}^{r\bar{E}^r(t)} v_i^r 1_{\{v_i^r > K\}} + \left\langle \chi 1_{(K,\infty)}, \bar{\mu}^r(0)\right\rangle < \gamma/5\right) \geq 1 - \tilde{\eta}. \tag{5.17}$$

Fine structure estimate for the state descriptor; the constant Γ. Now we define $\Gamma = K\left(\frac{\gamma}{4} - \frac{\gamma}{5}\right)^{-1}$, and consider (5.7) for each member of a finite set of functions $\{g_n\}_{n=0}^N$, which we define in the following way. Let $N = \lceil T\Gamma/\kappa\rceil$, and for each $n = 0, 1, \ldots, N$, choose $g_n \in \mathbf{C}_b(\mathbb{R}_+)$ such that g_n is nonnegative, and for all $x \in \mathbb{R}_+$, $1_{[n\kappa,(n+1)\kappa)}(x) \leq g_n(x) \leq 1_{[(n-\frac{1}{2})\kappa,(n+\frac{3}{2})\kappa)}(x)$. Note that for $n = 0$ and $x \in \mathbb{R}_+$, $1_{[(n-\frac{1}{2})\kappa,(n+\frac{3}{2})\kappa)}(x) = 1_{[0,\frac{3}{2}\kappa)}(x)$. By (5.7), we have

$$\lim_{r\to\infty} \mathbf{P}\left(\bigcap_{n=0}^{N}\left\{\sup_{t\in[0,T-l]} \frac{1}{r}\sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} g_n(v_i^r) \leq \frac{\epsilon}{8}\langle g_n, \nu\rangle\right\}\right) = 1, \tag{5.18}$$

which implies that

$$\lim_{r\to\infty} \mathbf{P}\left(\bigcap_{n=0}^{N}\left\{\sup_{t\in[0,T-l]} \frac{1}{r}\sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} 1_{[n\kappa,(n+1)\kappa)}(v_i^r) \leq \frac{\epsilon}{8}\left\langle 1_{[(n-\frac{1}{2})\kappa,(n+\frac{3}{2})\kappa)}, \nu\right\rangle\right\}\right) = 1, \tag{5.19}$$

where for $n = 0$, we interpret the right side of the inequality as $\frac{\epsilon}{8}\left\langle 1_{[0,\frac{3}{2}\kappa)}, \nu\right\rangle$.

Finally, having chosen the constants $l, M_0, M_T, \kappa, \gamma, K, \Gamma$ in such a way that (5.6)–(5.19) hold, we define for each $r > 0$, the intersection B^r of all of the events appearing in (5.6), (5.8)–(5.10), (5.12), (5.15), (5.17) and (5.19). Since $\tilde{\eta} = \eta/8$, the above discussion implies that $\liminf_{r\to\infty} \mathbf{P}(B^r) \geq 1-\eta$. We summarize the results of this section in the following lemma.

Lemma 5.2 *Consider a sequence of processor sharing queues as defined in Section 3.3, satisfying assumptions (3.14)–(3.22). Let $T > 0$ and $0 < \epsilon, \eta < 1$ be given. Then there exist strictly positive constants $l, M_0, M_T, \kappa, \gamma, K, \Gamma, r_0$ and events $\{B^r\}_{r>0}$, such that T/l is a positive integer,*

$$\mathbf{P}(B^r) \geq 1 - \eta, \qquad \textit{for all } r > r_0,$$

and for each $r > 0$, on B^r the following hold:

$$\sup_{t\in[0,T-l]} \bar{E}^r(t+l) - \bar{E}^r(t) \leq \frac{\epsilon}{4} \tag{5.20}$$

$$\bar{E}^r(T) \leq 2\alpha T \tag{5.21}$$

$$\langle 1, \bar{\mu}^r(0)\rangle \vee \langle \chi, \bar{\mu}^r(0)\rangle < M_0 \tag{5.22}$$

$$\left\langle 1_{[M_0,\infty)}, \Theta \right\rangle < \frac{\epsilon}{4} \tag{5.23}$$

$$M_T = M_0 + 2\alpha T \tag{5.24}$$

$$\kappa < \frac{l}{2M_T} \tag{5.25}$$

$$\sup_{x\in\mathbb{R}_+} \left\langle 1_{[x,x+\kappa]}, \bar{\mu}^r(0) \right\rangle < \frac{\epsilon}{4} \tag{5.26}$$

$$\gamma < \min\{\kappa\epsilon/4, l/4, \alpha T\} \tag{5.27}$$

$$\sup_{t\in[0,T]} \left|\langle \chi, \bar{\mu}^r(t)\rangle - \langle \chi, \bar{\mu}^r(0)\rangle\right| < \gamma/4 \tag{5.28}$$

$$\sup_{t\in[0,T]} \frac{1}{r} \sum_{i=1}^{r\bar{E}^r(t)} v_i^r 1_{\{v_i^r > K\}} + \left\langle \chi 1_{(K,\infty)}, \bar{\mu}^r(0) \right\rangle < \gamma/5 \tag{5.29}$$

$$\Gamma = K\left(\frac{\gamma}{4} - \frac{\gamma}{5}\right)^{-1} \tag{5.30}$$

for $N = \lceil T\Gamma/\kappa \rceil$ and all $n \in \{0, 1, \ldots, N\}$,

$$\sup_{t\in[0,T-l]} \frac{1}{r} \sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} 1_{[n\kappa,(n+1)\kappa)}(v_i^r) \leq \frac{\epsilon}{8} \left\langle 1_{[(n-\frac{1}{2})\kappa,(n+\frac{3}{2})\kappa)}, \nu \right\rangle. \tag{5.31}$$

As a final note, we remark that by (5.2),

$$\sup_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle \leq \langle 1, \bar{\mu}^r(0)\rangle + \bar{E}^r(T), \tag{5.32}$$

so by (5.22), (5.21), (5.27), (5.28), and (5.24), we have on B^r for $r > r_0$,

$$\sup_{t\in[0,T]} \left(\langle 1, \bar{\mu}^r(t)\rangle \vee \langle \chi, \bar{\mu}^r(t)\rangle\right) \leq M_T. \tag{5.33}$$

5.2 Proof of Tightness

In this section we prove the first half of Theorem 5.1, i.e., that the sequence of measure valued processes $\{\bar{\mu}^r(\cdot)\}$ is tight. Recall that to show tightness, it suffices by Jakubowski's criterion (cf. [12], Thm. 3.6.4) to show the following two properties:

TC.1 For each $T > 0$ and $0 < \eta < 1$, there is a compact subset $C_{T,\eta}$ of $\mathcal{M}_F$ such that

$$\liminf_{r\to\infty} \mathbf{P}\big(\bar{\mu}^r(t) \in C_{T,\eta} \text{ for all } t \in [0,T]\big) \geq 1 - \eta.$$

TC.2 For each $g \in \mathbf{C}_b^1(\mathbb{R}_+)$, the sequence of real valued processes $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}$ is tight.

Note that if we define $\Psi_g : \mathcal{M}_F \longrightarrow \mathbb{R}$ by $\Psi_g(\zeta) = \langle g, \zeta\rangle$ for $g \in \mathbf{C}_b^1(\mathbb{R}_+)$, then $\{\Psi_g : g \in \mathbf{C}_b^1(\mathbb{R}_+)\}$ defines a family of continuous real valued functions on $\mathcal{M}_F$ which is closed under addition and separates points of $\mathcal{M}_F$. Hence condition TC.2 above satisfies condition (ii) of [12], Thm. 3.6.4. The proof of TC.2 will follow from [15], Ch. 3, Thm. 7.2, upon verifying in Theorem 5.6 that the usual compact containment and controlled oscillation conditions are satisfied by $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}$. Although the compact containment condition just mentioned is an immediate consequence of TC.1, we show it separately as part of Theorem 5.6, since it is elementary and it is needed before we establish TC.1. Controlling the oscillations of $\langle g, \bar{\mu}^r(\cdot)\rangle$ for Theorem 5.6 is the major difficulty in the above agenda. Proving that we have this control ultimately revolves around the fact that changes in this process over a short time interval depend on the arrival process (which is easily controlled), the magnitude of the fluid scaled queue length, $\langle 1, \bar{\mu}^r(\cdot)\rangle$, and the amount of mass concentrated near zero over that time interval (cf. (5.2) and (3.13)). We estimate $\langle 1, \bar{\mu}^r(\cdot)\rangle$ from the fluid scaled *workload*, $\langle \chi, \bar{\mu}^r(\cdot)\rangle$, and then leverage the fact that $\langle \chi, \bar{\mu}^r(\cdot)\rangle$ converges in distribution to a process that a.s. is equal for all time to its initial value, which has the same distribution as $\langle \chi, \Theta\rangle$. To do this, we consider two events for $\bar{\mu}^r(\cdot)$: the event where the initial fluid scaled workload is smaller than some threshold, and its complement. In Lemmas 5.3 and 5.4, we prove an upper bound for the fluid scaled queue length on the first event, and a lower bound for the fluid scaled queue length on the second event. Then in Lemma 5.5, we give an upper bound for the amount of mass that $\bar{\mu}^r(t)$ can have concentrated near zero. Note that this is a result about the “fine” structure of the measure $\bar{\mu}^r(t)$, as opposed to its total mass (queue length) or first moment (workload). It is an essential ingredient for completing the proof of the controlled oscillation condition for Theorem 5.6, as well as for proving the limit point properties in Section 5.3. Finally, the proof of tightness for Theorem 5.1 appears at the end of this section.

We begin by proving the three aforementioned lemmas. The first lemma provides an upper bound for the fluid scaled queue length on the event that the initial fluid scaled workload is below the threshold $\gamma/2$.

Lemma 5.3 *Let $T > 0$ and $0 < \epsilon, \eta < 1$. Let $l, M_0, M_T, \kappa, \gamma, K, \Gamma, r_0$ be the constants, and $\{B^r\}_{r>0}$ be the events, given by Lemma 5.2. Define $D_\gamma^r = \{\langle \chi, \bar{\mu}^r(0)\rangle \leq \gamma/2\}$. Then on $B^r \cap D_\gamma^r$, for $r > r_0$,*

$$\sup_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle \leq \epsilon.$$

Proof. Subdivide $[0,T]$ into time intervals of length l (recall that l was chosen so that T is an integer multiple of l). Fix $r > r_0$. We will prove by induction on n, that on $B^r \cap D_\gamma^r$,

$$\sup_{t\in[nl,(n+1)l]} \langle 1, \bar{\mu}^r(t)\rangle \leq \epsilon, \tag{5.34}$$

for $n = 0, 1, \ldots, \frac{T}{l} - 1$. We first verify the case $n = 0$. To start with we have on $B^r \cap D^r_\gamma$,

$$\begin{aligned}
\langle 1, \bar{\mu}^r(0)\rangle &= \left\langle 1_{[0,\kappa]}, \bar{\mu}^r(0)\right\rangle + \left\langle 1_{(\kappa,\infty)}, \bar{\mu}^r(0)\right\rangle \\
&\le \frac{\epsilon}{4} + \frac{1}{\kappa}\langle \chi, \bar{\mu}^r(0)\rangle \\
&\le \frac{\epsilon}{4} + \frac{\gamma/2}{\kappa} \\
&\le \frac{\epsilon}{2},
\end{aligned} \tag{5.35}$$

where the first inequality is by (5.26) and Markov's inequality, the second uses the definition of D^r_γ and the last is by (5.27). Now take $t \in [0, l]$. Then on $B^r \cap D^r_\gamma$,

$$\langle 1, \bar{\mu}^r(t)\rangle \le \langle 1, \bar{\mu}^r(0)\rangle + \bar{E}^r(l) \le \frac{\epsilon}{2} + \frac{\epsilon}{4} < \epsilon, \tag{5.36}$$

by (5.32), (5.35) and (5.20). So (5.34) holds for $n = 0$.

We now proceed by induction, having verified the first step, and assume that (5.34) holds on $B^r \cap D^r_\gamma$ for some $0 \le n < \frac{T}{l} - 1$. To show the statement holds with $(n+1)$ in place of n, we argue analogously to (5.36) by first showing that $\langle 1, \bar{\mu}^r((n+1)l)\rangle \le \epsilon/2$. To this end, we use an argument inspired in part by an idea of Grishechkin (cf. [18], p. 542). The idea is to consider two cases separately: the case where the queue length becomes zero during the interval $[nl, (n+1)l]$, and the case where it does not. We have on $B^r \cap D^r_\gamma$,

$$\langle \chi, \bar{\mu}^r(nl)\rangle \le \langle \chi, \bar{\mu}^r(0)\rangle + \left|\langle \chi, \bar{\mu}^r(nl)\rangle - \langle \chi, \bar{\mu}^r(0)\rangle\right| < \frac{\gamma}{2} + \frac{\gamma}{4} < \gamma, \tag{5.37}$$

by (5.28) and the definition of D^r_γ. Now, if $\langle 1, \bar{\mu}^r(s)\rangle$ is never zero for $s \in [nl, nl + 4\gamma]$ we can write

$$\bar{S}^r_{nl,nl+4\gamma} = \int_{nl}^{nl+4\gamma} \varphi(\langle 1, \bar{\mu}^r(s)\rangle)ds = \int_{nl}^{nl+4\gamma} \langle 1, \bar{\mu}^r(s)\rangle^{-1}\, ds \ge 4\gamma/\epsilon,$$

since we have assumed that (5.34) holds for this n and $4\gamma < l$ by (5.27). This in turn implies, by (5.37), that

$$\begin{aligned}
\left\langle 1_{(\bar{S}^r_{nl,nl+4\gamma},\infty)}, \bar{\mu}^r(nl)\right\rangle &\le \left\langle 1_{(\frac{4\gamma}{\epsilon},\infty)}, \bar{\mu}^r(nl)\right\rangle \\
&\le \frac{\epsilon}{4\gamma}\langle \chi, \bar{\mu}^r(nl)\rangle \\
&\le \frac{\epsilon}{4\gamma}\gamma = \frac{\epsilon}{4}.
\end{aligned} \tag{5.38}$$

If on the other hand $\langle 1, \bar{\mu}^r(s)\rangle = 0$ for some $s \in [nl, nl + 4\gamma]$, then all mass present in the system at time nl is gone by time s. More precisely, we have in this case, by (5.2), that

$$\left\langle 1_{(\bar{S}^r_{nl,s},\infty)}, \bar{\mu}^r(nl)\right\rangle = 0,$$

which, since $\bar{S}^r_{nl,s} \le \bar{S}^r_{nl,nl+4\gamma}$, also implies that

$$\left\langle 1_{(\bar{S}^r_{nl,nl+4\gamma},\infty)}, \bar{\mu}^r(nl)\right\rangle = 0. \tag{5.39}$$

Combining the above with (5.2), we see that in either case, on $B^r \cap D^r_\gamma$,

$$\begin{aligned}\langle 1, \bar{\mu}^r((n+1)l)\rangle &\le \left\langle 1_{(\bar{S}^r_{nl,(n+1)l},\infty)}, \bar{\mu}^r(nl)\right\rangle + \bar{E}^r((n+1)l) - \bar{E}^r(nl) \\ &\le \left\langle 1_{(\bar{S}^r_{nl,nl+4\gamma},\infty)}, \bar{\mu}^r(nl)\right\rangle + \frac{\epsilon}{4} \\ &\le \frac{\epsilon}{2}, \end{aligned} \tag{5.40}$$

by (5.27) and (5.20) for the second inequality, and by (5.39) and (5.38) for the third. Now we can complete the proof as in (5.36). Using (5.2), we have on $B^r \cap D^r_\gamma$ for any $t \in [(n+1)l, (n+2)l]$:

$$\langle 1, \bar{\mu}^r(t)\rangle \le \langle 1, \bar{\mu}^r((n+1)l)\rangle + \bar{E}^r((n+2)l) - \bar{E}^r((n+1)l) \le \frac{\epsilon}{2} + \frac{\epsilon}{4} < \epsilon,$$

where the second inequality is by (5.40) and (5.20). ■

The next lemma provides a lower bound for the fluid scaled queue length process on the event where the initial fluid scaled workload is above the threshold $\gamma/2$. A consequence of this is an upper bound for the rate at which $\bar{S}^r_{t,t+h}$ can increase as a function of h.

Lemma 5.4 *Let $T > 0$ and $0 < \epsilon, \eta < 1$. Let $l, M_0, M_T, \kappa, \gamma, K, \Gamma, r_0$ be the constants, and $\{B^r\}_{r>0}$ be the events, given by Lemma 5.2. Let $\breve{D}^r_\gamma$ be the complement of D^r_γ, i.e., $\breve{D}^r_\gamma = \{\langle \chi, \bar{\mu}^r(0)\rangle > \gamma/2\}$. Then on $B^r \cap \breve{D}^r_\gamma$, for $r > r_0$,*

$$\inf_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle \ge \frac{1}{\Gamma}, \tag{5.41}$$

and

$$\sup_{t\in[0,T-h]} \bar{S}^r_{t,t+h} \le h\Gamma, \tag{5.42}$$

for any $0 < h < T$.

Proof. For this we have adapted the proof of Lemma 4.4 in [9]. We have on $B^r \cap \breve{D}^r_\gamma$,

$$\begin{aligned}\frac{\gamma}{4} &\le \inf_{t\in[0,T]} \langle \chi, \bar{\mu}^r(t)\rangle \\ &= \inf_{t\in[0,T]} \left(\left\langle \chi 1_{[0,K]}, \bar{\mu}^r(t)\right\rangle + \left\langle \chi 1_{(K,\infty)}, \bar{\mu}^r(t)\right\rangle\right) \\ &\le \inf_{t\in[0,T]} \left(K \left\langle 1_{[0,K]}, \bar{\mu}^r(t)\right\rangle + \frac{1}{r}\sum_{i=1}^{r\bar{E}^r(t)} v^r_i 1_{\{v^r_i > K\}} + \left\langle \chi 1_{(K,\infty)}, \bar{\mu}^r(0)\right\rangle \right) \\ &\le \inf_{t\in[0,T]} K\langle 1, \bar{\mu}^r(t)\rangle + \frac{\gamma}{5},\end{aligned}$$

where the three inequalities are by (5.28) and the definition of $\breve{D}^r_\gamma$, (5.2) and (5.29) respectively. Notice that to obtain the last two terms in the second inequality above from (5.2), we have simply ignored any processing that has occurred for jobs with initial service time requirements greater than K. Now we have

$$\inf_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle \ge \frac{\frac{\gamma}{4} - \frac{\gamma}{5}}{K} = \frac{1}{\Gamma}.$$

To prove (5.42), we have by (5.41) on $B^r \cap \breve{D}^r_\gamma$, for $0 < h < T$,

$$
\begin{aligned}
\sup_{t\in[0,T-h]} \bar{S}^r_{t,t+h} &\le h \sup_{t\in[0,T]} \varphi(\langle 1, \bar{\mu}^r(t)\rangle) \\
&\le \frac{h}{\inf_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle} \\
&\le h\Gamma.
\end{aligned}
$$

■

The next lemma gives, on $B^r \cap \breve{D}^r_\gamma$, an upper bound for the amount of mass that $\bar{\mu}^r(t)$ can have concentrated near zero for $t \in [0,T]$. Note that by Lemma 5.3, on $B^r \cap D^r_\gamma$, the total mass of $\bar{\mu}^r(t)$ is bounded above by ϵ for $t \in [0,T]$.

Lemma 5.5 *Let $T > 0$ and $0 < \epsilon, \eta < 1$. Let $l, M_0, M_T, \kappa, \gamma, K, \Gamma, r_0$ be the constants, and $\{B^r\}_{r>0}$ be the events, given by Lemma 5.2. Recall that $\breve{D}^r_\gamma = \{\langle \chi, \bar{\mu}^r(0)\rangle > \gamma/2\}$. Then on $B^r \cap \breve{D}^r_\gamma$, for $r > r_0$,*

$$\sup_{t\in[0,T]} \left\langle 1_{[0,\kappa]}, \bar{\mu}^r(t)\right\rangle \le \epsilon/2.$$

Proof. In this proof we only consider realizations in $B^r \cap \breve{D}^r_\gamma$. Recall that by Lemma 5.4, $\langle 1, \bar{\mu}^r(t)\rangle > 0$ for all $t \in [0,T]$. First, consider two jobs, $i < j$ for which $v^r_i, v^r_j \in [n\kappa, (n+1)\kappa)$ for some integer $n \ge 0$. If $U^r_j/r \le t$ and $U^r_j/r - U^r_i/r \ge l$, then at time t we have

$$
\begin{aligned}
(v^r_j - \bar{S}^r_{U^r_j/r,t}) - (v^r_i - \bar{S}^r_{U^r_i/r,t}) &= \bar{S}^r_{U^r_i/r,U^r_j/r} + v^r_j - v^r_i \\
&\ge \frac{U^r_j/r - U^r_i/r}{\sup_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle} - \kappa \\
&\ge \frac{l}{M_T} - \kappa \\
&> 2\kappa - \kappa = \kappa,
\end{aligned}
$$

where the last two inequalities are by (5.33) and (5.25) respectively. This implies that at most one of

$$1_{(0,\kappa]}(v^r_i - \bar{S}^r_{U^r_i/r,t}) \quad \text{and} \quad 1_{(0,\kappa]}(v^r_j - \bar{S}^r_{U^r_j/r,t})$$

is nonzero. So for each $n \ge 0$ and $t \in [0,T]$, all jobs arriving by time t and satisfying $1_{[n\kappa,(n+1)\kappa)}(v^r_i)1_{(0,\kappa]}(v^r_i - \bar{S}^r_{U^r_i/r,t}) = 1$ must have arrived during the (fluid scaled) time interval $(s, s+l]$, for some $s \in [0, t-l]$. This gives us the following estimate at time t, for each $n = 0, 1, \ldots, \lceil T\Gamma/\kappa\rceil$:

$$\sum_{i=1}^{r\bar{E}^r(t)} 1_{[n\kappa,(n+1)\kappa)}(v^r_i)1_{(0,\kappa]}(v^r_i - \bar{S}^r_{U^r_i/r,t}) \le \sup_{s\in[0,t-l]} \sum_{i=r\bar{E}^r(s)+1}^{r\bar{E}^r(s+l)} 1_{[n\kappa,(n+1)\kappa)}(v^r_i)1_{(0,\kappa]}(v^r_i - \bar{S}^r_{U^r_i/r,t}),$$

which implies that for each $n = 0, 1, \ldots, \lceil T\Gamma/\kappa\rceil$, we have

$$\sup_{t\in[0,T]} \frac{1}{r} \sum_{i=1}^{r\bar{E}^r(t)} 1_{[n\kappa,(n+1)\kappa)}(v^r_i)1_{(0,\kappa]}(v^r_i - \bar{S}^r_{U^r_i/r,t}) \le \sup_{t\in[0,T-l]} \frac{1}{r} \sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} 1_{[n\kappa,(n+1)\kappa)}(v^r_i). \tag{5.43}$$

Next, for $t \in [0,T]$ and $U_i^r/r \le t$, we have by Lemma 5.4 that if $v_i^r \ge T\Gamma + \kappa$, then $(v_i^r - \bar{S}^r_{U_i^r/r,t}) > T\Gamma + \kappa - T\Gamma = \kappa$. Thus for $n \ge \lceil T\Gamma/\kappa \rceil + 1$, we have

$$1_{[n\kappa,(n+1)\kappa)}(v_i^r) 1_{(0,\kappa]}(v_i^r - \bar{S}^r_{U_i^r/r,t}) = 0. \tag{5.44}$$

Using (5.2), we have on $B^r \cap \breve{D}^r_\gamma$,

$$\begin{aligned}
\sup_{t\in[0,T]} \left\langle 1_{[0,\kappa]}, \bar{\mu}^r(t) \right\rangle &\le \sup_{t\in[0,T]} \left\langle 1_{(0,\kappa]}(\cdot - \bar{S}^r_{0,t}), \bar{\mu}^r(0) \right\rangle + \sup_{t\in[0,T]} \frac{1}{r} \sum_{i=1}^{r\bar{E}^r(t)} 1_{(0,\kappa]}(v_i^r - \bar{S}^r_{U_i^r/r,t}) \\
&\le \sup_{x\in\mathbb{R}_+} \left\langle 1_{[x,x+\kappa]}, \bar{\mu}^r(0) \right\rangle + \sum_{n=0}^{\infty} \sup_{t\in[0,T]} \frac{1}{r} \sum_{i=1}^{r\bar{E}^r(t)} 1_{[n\kappa,(n+1)\kappa)}(v_i^r) 1_{(0,\kappa]}(v_i^r - \bar{S}^r_{U_i^r/r,t}) \\
&\le \frac{\epsilon}{4} + \sum_{n=0}^{\lceil T\Gamma/\kappa \rceil} \sup_{t\in[0,T-l]} \frac{1}{r} \sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+l)} 1_{[n\kappa,(n+1)\kappa)}(v_i^r) \\
&\le \frac{\epsilon}{4} + \sum_{n=0}^{\lceil T\Gamma/\kappa \rceil} \frac{\epsilon}{8} \left\langle 1_{[(n-\frac{1}{2})\kappa,(n+\frac{3}{2})\kappa)}, \nu \right\rangle \\
&\le \frac{\epsilon}{4} + \frac{\epsilon}{8} \left\langle \sum_{n=0}^{\infty} 1_{[(n-\frac{1}{2})\kappa,(n+\frac{3}{2})\kappa)}, \nu \right\rangle \\
&= \frac{\epsilon}{4} + \frac{\epsilon}{4},
\end{aligned}$$

where the third line is by (5.26), (5.43) and (5.44), and the fourth line is by (5.31). ∎

We are now ready to apply the previous three lemmas in order to verify conditions that are sufficient to imply the tightness of $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}$, for $g \in \mathbf{C}_b^1(\mathbb{R}_+)$. The proof of the controlled oscillation condition splits into two cases. We use Lemma 5.3 to prove it on $B^r \cap D^r_\gamma$ and use Lemmas 5.4 and 5.5 to prove it on $B^r \cap \breve{D}^r_\gamma$.

Theorem 5.6 *Let $g \in \mathbf{C}_b^1(\mathbb{R}_+)$, $T > 0$ and $0 < \beta, \eta < 1$. Then there exist $M, \delta > 0$, and $r_0 > 0$, such that $r > r_0$ implies*

$$\mathbf{P}\left(\sup_{t\in[0,T]} \left| \langle g, \bar{\mu}^r(t) \rangle \right| \le M \right) \ge 1 - \eta, \tag{5.45}$$

$$\mathbf{P}\left(\sup_{t\in[0,T-\delta]} \sup_{h\in[0,\delta]} \left| \langle g, \bar{\mu}^r(t+h) \rangle - \langle g, \bar{\mu}^r(t) \rangle \right| \le \beta \right) \ge 1 - \eta. \tag{5.46}$$

Proof. Define $\epsilon = \beta/(2(\|g\|_\infty \vee 1))$. Then by Lemma 5.2, there exist constants l, M_0, M_T, κ, γ, K, Γ, r_0, and events $\{B^r\}_{r>0}$ such that $r > r_0$ implies $\mathbf{P}(B^r) \ge 1 - \eta$ and on B^r, (5.20)–(5.33) hold. Define

$$\begin{aligned}
M &= (\|g\|_\infty \vee 1) M_T, \\
\delta &= \min\left\{ T/2, l, \frac{\beta}{4\Gamma M_T (\|g'\|_\infty \vee 1)}, \kappa/\Gamma, 1 \right\}.
\end{aligned} \tag{5.47}$$

We only consider $r > r_0$ below. To prove (5.45), observe that on B^r, by (5.33),

$$\begin{aligned}\sup_{t\in[0,T]}\left|\langle g,\bar{\mu}^r(t)\rangle\right| &\le \|g\|_\infty \sup_{t\in[0,T]}\langle 1,\bar{\mu}^r(t)\rangle\\ &\le \|g\|_\infty M_T\\ &\le M.\end{aligned}$$

To prove (5.46), consider first D^r_γ as before. On $B^r\cap D^r_\gamma$, we have by Lemma 5.3,

$$\begin{aligned}\sup_{t\in[0,T-\delta]}\sup_{h\in[0,\delta]}\left|\langle g,\bar{\mu}^r(t+h)\rangle-\langle g,\bar{\mu}^r(t)\rangle\right| &\le 2\sup_{t\in[0,T]}\left|\langle g,\bar{\mu}^r(t)\rangle\right|\\ &\le 2\|g\|_\infty\sup_{t\in[0,T]}\langle 1,\bar{\mu}^r(t)\rangle\\ &\le 2\|g\|_\infty\,\epsilon\le\beta.\end{aligned}$$

We must show that the above estimate also holds on $B^r\cap\breve{D}^r_\gamma$. First, observe that on $B^r\cap\breve{D}^r_\gamma$, a first order Taylor expansion of g gives the following estimate for all $0<h<T$, $t\in[0,T-h]$ and $y\in(\bar{S}^r_{t,t+h},\infty)$:

$$\left|g(y-\bar{S}^r_{t,t+h})-g(y)\right| = \left|-\bar{S}^r_{t,t+h}g'(w_y)\right| \le h\Gamma\|g'\|_\infty, \tag{5.48}$$

for some $w_y\in[y-\bar{S}^r_{t,t+h},y]$, where the inequality follows by Lemma 5.4. Now subtracting $\langle g,\bar{\mu}^r(t)\rangle$ from both sides of equation (5.2) and using the fact that $(1_{(0,\infty)}g)(\cdot-\bar{S}^r_{t,t+h}) = 1_{(\bar{S}^r_{t,t+h},\infty)}(\cdot)g(\cdot-\bar{S}^r_{t,t+h})$ for $t\in[0,T-h]$ yields that on $B^r\cap\breve{D}^r_\gamma$,

$$\begin{aligned}\left|\langle g,\bar{\mu}^r(t+h)\rangle-\langle g,\bar{\mu}^r(t)\rangle\right| &= \Bigg|\left\langle 1_{(\bar{S}^r_{t,t+h},\infty)}(\cdot)\big(g(\cdot-\bar{S}^r_{t,t+h})-g(\cdot)\big),\bar{\mu}^r(t)\right\rangle - \left\langle 1_{[0,\bar{S}^r_{t,t+h}]}g,\bar{\mu}^r(t)\right\rangle\\ &\qquad + \frac{1}{r}\sum_{i=r\bar{E}^r(t)+1}^{r\bar{E}^r(t+h)}(1_{(0,\infty)}g)(v^r_i-\bar{S}^r_{U^r_i/r,t+h})\Bigg|\\ &\le \left\langle\left|1_{(\bar{S}^r_{t,t+h},\infty)}(\cdot)\left(g(\cdot-\bar{S}^r_{t,t+h})-g(\cdot)\right)\right|,\bar{\mu}^r(t)\right\rangle + \|g\|_\infty\left\langle 1_{[0,h\Gamma]},\bar{\mu}^r(t)\right\rangle\\ &\qquad + \|g\|_\infty\big(\bar{E}^r(t+h)-\bar{E}^r(t)\big)\\ &\le h\Gamma\|g'\|_\infty\langle 1,\bar{\mu}^r(t)\rangle + \|g\|_\infty\left\langle 1_{[0,h\Gamma]},\bar{\mu}^r(t)\right\rangle\\ &\qquad + \|g\|_\infty\big(\bar{E}^r(t+h)-\bar{E}^r(t)\big),\end{aligned}$$

where the second inequality is by Lemma 5.4 and the third is by (5.48).

Now taking the supremum over $h \in [0, \delta]$ and $t \in [0, T - \delta]$, we see that on $B^r \cap \breve{D}^r_\gamma$,

$$
\begin{aligned}
\sup_{t\in[0,T-\delta]} \sup_{h\in[0,\delta]} \left|\langle g, \bar{\mu}^r(t+h)\rangle - \langle g, \bar{\mu}^r(t)\rangle\right| &\le \sup_{t\in[0,T-\delta]} \Big(\delta\Gamma \|g'\|_\infty \langle 1, \bar{\mu}^r(t)\rangle + \|g\|_\infty \left\langle 1_{[0,\delta\Gamma]}, \bar{\mu}^r(t)\right\rangle \\
&\qquad + \|g\|_\infty \big(\bar{E}^r(t+\delta) - \bar{E}^r(t)\big)\Big) \\
&\le \sup_{t\in[0,T-\delta]} \left(\frac{\beta}{4M_T} \langle 1, \bar{\mu}^r(t)\rangle + \|g\|_\infty \left\langle 1_{[0,\kappa]}, \bar{\mu}^r(t)\right\rangle \right. \\
&\qquad \left. + \|g\|_\infty \big(\bar{E}^r(t+l) - \bar{E}^r(t)\big) \right) \\
&\le \frac{\beta}{4M_T} M_T + \|g\|_\infty \frac{\epsilon}{2} + \|g\|_\infty \frac{\epsilon}{4} \\
&\le \frac{\beta}{4} + \frac{\beta}{4} + \frac{\beta}{8} < \beta,
\end{aligned}
$$

where the second inequality is by (5.47) and the third is by (5.33), Lemma 5.5, and (5.20). So the desired estimate holds on both $B^r \cap D^r_\gamma$ and $B^r \cap \breve{D}^r_\gamma$. Since $\mathbf{P}(B^r) \ge 1 - \eta$, this proves (5.46). ■

Proof of Tightness for Theorem 5.1. Recall that it suffices to show conditions TC.1 and TC.2. Let $T > 0$ and $0 < \beta, \eta < 1$. To show TC.1, define

$$\tilde{C}_{T,\eta} = \{\zeta \in \mathcal{M}_F : \langle 1, \zeta\rangle \vee \langle \chi, \zeta\rangle \le M_T\}.$$

Since $\langle \chi, \zeta\rangle \le M_T$ implies $\left\langle 1_{[K,\infty)}, \zeta\right\rangle \le M_T/K$, we have

$$\sup_{\zeta\in\tilde{C}_{T,\eta}} \left\langle 1_{[K,\infty)}, \zeta\right\rangle \longrightarrow 0, \qquad \text{as } K \to \infty,$$

which implies that $\tilde{C}_{T,\eta} \subset \mathcal{M}_F$ is relatively compact (cf. [21], Thm. A 7.5). Now by (5.33),

$$\liminf_{r\to\infty} \mathbf{P}\big(\bar{\mu}^r(t) \in \tilde{C}_{T,\eta} \text{ for all } t \in [0,T]\big) \ge 1 - \eta.$$

Define $C_{T,\eta}$ to be the closure of $\tilde{C}_{T,\eta}$ and TC.1 is proved. Finally, TC.2 follows directly from Theorem 5.6 by applying a standard tightness criterion for real valued processes (cf. [15], Ch. 3, Cor. 7.4). ■

5.3 Proof of Limit Point Properties

In this section we complete the proof of Theorem 5.1 by showing that any limit point of the sequence $\{\bar{\mu}^r(\cdot)\}$ is a.s. a fluid model solution for the critical data (α, ν). In particular, we show that a.s., the sample paths of any such limit point $\bar{\mu}^\star(\cdot)$ have the properties (i)–(iv) of Section 3.1. Property (i) will follow from the a.s. continuity of the sample paths of any limit point of $\{\langle g, \bar{\mu}^r(\cdot)\rangle\}$, for $g \in \mathbf{C}^1_b(\mathbb{R}_+)$. Property (ii) will be a direct consequence of Lemmas 5.3 and 5.5. We will show en route that limit points $\bar{\mu}^\star(\cdot)$ satisfy $t^* = \infty$ on $\{\bar{\mu}^\star(0) \ne \mathbf{0}\}$ and $t^* = 0$ on $\{\bar{\mu}^\star(0) = \mathbf{0}\}$ (cf. (5.50) and (5.51) below). This is to be expected in light of Lemma 4.4. Consequently, it will suffice for the proof of property (iii) to show that a.s. on the event $\{\bar{\mu}^\star(0) \ne \mathbf{0}\}$, (3.3) holds for all $g \in \mathcal{C}$ and all $t \ge 0$. For this, we will use the dynamic equation satisfied by $\bar{\mu}^r(\cdot)$ (cf. (5.2)), together with a Riemann integral approximation, to obtain a prelimit version of (3.3), and then we pass to the limit. Similarly for the proof of property (iv), it will suffice to show that a.s. on the event $\{\bar{\mu}^\star(0) = \mathbf{0}\}$, $\bar{\mu}^\star(t) = \mathbf{0}$ for all $t \geq 0$.

Let $\bar{\mu}^\star(\cdot)$ be a limit point of $\{\bar{\mu}^r(\cdot)\}$ and suppose $\{\bar{\mu}^{r'}(\cdot)\} \subset \{\bar{\mu}^r(\cdot)\}$ is a subsequence such that $\bar{\mu}^{r'}(\cdot) \Longrightarrow \bar{\mu}^\star(\cdot)$ as $r' \to \infty$. To ease notation for the remainder of the proof, we relabel r' as r, remembering that we have passed to a subsequence which converges in distribution to $\bar{\mu}^\star(\cdot)$.

Proof of property (i). To see that a.s., $\bar{\mu}^\star(\cdot)$ has continuous sample paths, choose a countable set $V \subset \mathbf{C}_b^1(\mathbb{R}_+)$ that separates elements of $\mathcal{M}_F$ (cf. [15], Ch. 3, Prop. 4.2 ff.). It follows from Theorem 5.6 (in particular (5.46)) that for each $g \in V \subset \mathbf{C}_b^1(\mathbb{R}_+)$, the real valued process $\langle g, \bar{\mu}^\star(\cdot)\rangle$ a.s. has continuous sample paths. Since V is a countable separating class for $\mathcal{M}_F$, this implies that a.s., for every $t \geq 0$, $\bar{\mu}^\star(t-) = \bar{\mu}^\star(t)$. ■

Proof of properties (ii) and (iv). As mentioned above, it suffices to show that for each $T > 0$ we have a.s.,

$$\left\langle 1_{\{0\}}, \bar{\mu}^\star(t)\right\rangle = 0, \text{ for all } t \in [0,T), \tag{5.49}$$

$$\bar{\mu}^\star(t) \neq \mathbf{0}, \text{ for all } t \in [0,T) \text{ on } \{\bar{\mu}^\star(0) \neq \mathbf{0}\}, \tag{5.50}$$

$$\bar{\mu}^\star(t) = \mathbf{0}, \text{ for all } t \in [0,T) \text{ on } \{\bar{\mu}^\star(0) = \mathbf{0}\}. \tag{5.51}$$

To this end, let $\mathbf{Q}^r$ and $\mathbf{Q}$ be the probability laws induced on $D([0,\infty),\mathcal{M}_F)$ by $\bar{\mu}^r(\cdot)$ and $\bar{\mu}^\star(\cdot)$, respectively. For $T > 0$ fixed, define the sets

$$A' = \left\{\zeta(\cdot) \in D([0,\infty),\mathcal{M}_F) \,:\, \sup_{t\in[0,T)} \left\langle 1_{\{0\}}, \zeta(t)\right\rangle > 0\right\},$$

$$A'' = \left\{\zeta(\cdot) \in D([0,\infty),\mathcal{M}_F) \,:\, \sup_{t\in[0,T)} \langle 1, \zeta(t)\rangle > 0\right\},$$

$$A''' = \left\{\zeta(\cdot) \in D([0,\infty),\mathcal{M}_F) \,:\, \inf_{t\in[0,T)} \langle 1, \zeta(t)\rangle = 0\right\},$$

and let $A = A' \cup (A'' \cap A''')$. Clearly, any sample path of $\bar{\mu}^\star(\cdot)$ which is not an element of A satisfies (5.49)–(5.51). So it suffices to show that A is contained in a $\mathbf{Q}$-null set. For each $n = 1, 2, \ldots$, choose a pair $0 < \epsilon_n, \eta_n < 1$ such that $\sum_{n=1}^\infty \eta_n < \infty$ and $(\epsilon_n, \eta_n) \longrightarrow (0,0)$ as $n \to \infty$. Then by Lemma 5.2, for each n there exist strictly positive constants $l_n, M_{0,n}, M_{T,n}, \kappa_n, \gamma_n, K_n, \Gamma_n, r_{0,n}$ and events $\{B_n^r : r > 0\}$ such that $r > r_{0,n}$ implies $\mathbf{P}(B_n^r) \geq 1 - \eta_n$, and (5.20)–(5.31) hold on B_n^r with the above constants in place of the analogous ones appearing there. For each n, choose $0 < c_n < \kappa_n$ such that $c_n \longrightarrow 0$ as $n \to \infty$. Also define for each n

$$A'_n = \left\{\zeta(\cdot) \in D([0,\infty),\mathcal{M}_F) \,:\, \sup_{t\in[0,T)} \left\langle 1_{[0,c_n)}, \zeta(t)\right\rangle > \epsilon_n\right\},$$

$$A''_n = \left\{\zeta(\cdot) \in D([0,\infty),\mathcal{M}_F) \,:\, \sup_{t\in[0,T)} \langle 1, \zeta(t)\rangle > \epsilon_n\right\},$$

$$A'''_n = \left\{\zeta(\cdot) \in D([0,\infty),\mathcal{M}_F) \,:\, \inf_{t\in[0,T)} \langle 1, \zeta(t)\rangle < 1/\Gamma_n\right\},$$

and let $A_n = A'_n \cup (A''_n \cap A'''_n)$. Recall from the definition of the Skorohod topology, that if C is a closed subset of $\mathcal{M}_F$, then sets of the form $\{\zeta(\cdot) \in D([0,\infty),\mathcal{M}_F) : \zeta(t) \in C \text{ for all } t \in [0,T)\}$ are closed in $D([0,\infty),\mathcal{M}_F)$, and so their complements are open. Since for each n, the sets $\{\zeta \in \mathcal{M}_F : \langle 1_{[0,c_n)}, \zeta\rangle \le \epsilon_n\}$, $\{\zeta \in \mathcal{M}_F : \langle 1, \zeta\rangle \le \epsilon_n\}$, and $\{\zeta \in \mathcal{M}_F : \langle 1,\zeta\rangle \ge 1/\Gamma_n\}$ are all closed subsets of $\mathcal{M}_F$, we see that A_n', A_n'', A_n''' and A_n are open sets in $D([0,\infty),\mathcal{M}_F)$. Notice that the definition of A implies that

$$A \subset \{A_n \text{ i.o.}\} = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n,$$

so it suffices to show that $\mathbf{Q}(A_n \text{ i.o.}) = 0$. Let n be fixed for the moment. Since we have chosen c_n so that

$$\sup_{t\in[0,T]} \left\langle 1_{[0,c_n)}, \bar{\mu}^r(t)\right\rangle \le \sup_{t\in[0,T]} \left\langle 1_{[0,\kappa_n]}, \bar{\mu}^r(t)\right\rangle \le \sup_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle$$

for every r, combining Lemmas 5.3 and 5.5 yields the fact that on the entire event B_n^r, for $r > r_{0,n}$,

$$\sup_{t\in[0,T)} \left\langle 1_{[0,c_n)}, \bar{\mu}^r(t)\right\rangle \le \sup_{t\in[0,T]} \left\langle 1_{[0,c_n)}, \bar{\mu}^r(t)\right\rangle \le \epsilon_n.$$

Similarly, combining Lemmas 5.3 and 5.4 yields the fact that on B_n^r for $r > r_{0,n}$, either

$$\sup_{t\in[0,T)} \langle 1, \bar{\mu}^r(t)\rangle \le \sup_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle \le \epsilon_n,$$

or

$$\inf_{t\in[0,T)} \langle 1, \bar{\mu}^r(t)\rangle \ge \inf_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle \ge 1/\Gamma_n.$$

These three facts together imply that for $r > r_{0,n}$, $\mathbf{Q}^r(A_n) \le 1 - \mathbf{P}(B_n^r)$. So since $\mathbf{P}(B_n^r) \ge 1 - \eta_n$ for $r > r_{0,n}$, we have $\limsup_{r\to\infty} \mathbf{Q}^r(A_n) \le \eta_n$. Now since $\mathbf{Q}^r \xrightarrow{\text{w}} \mathbf{Q}$ as $r\to\infty$, and since A_n is an open set, the Portmanteau theorem (cf. [2], Thm. 2.1) yields

$$\mathbf{Q}(A_n) \le \liminf_{r\to\infty} \mathbf{Q}^r(A_n) \le \limsup_{r\to\infty} \mathbf{Q}^r(A_n) \le \eta_n,$$

which implies by choice of the η_n and the Borel-Cantelli lemma that $\mathbf{Q}(A_n \text{ i.o.}) = 0$. ∎

Proof of property (iii). Recall that since we have established that a.s., $t^* = \infty$ on $\{\bar{\mu}^\star(0) \ne \mathbf{0}\}$, and $t^* = 0$ on $\{\bar{\mu}^\star(0) = \mathbf{0}\}$, it suffices to show that a.s. on $\{\bar{\mu}^\star(0) \ne \mathbf{0}\}$, (3.3) holds for all $g \in \mathcal{C}$ and all $t \ge 0$. We begin by restricting our attention to a slightly smaller class of functions than $\mathcal{C}$. We will establish (3.3) for functions in this smaller class, and then make a simple generalization at the end of the proof. Define

$$\tilde{\mathcal{C}} = \{g \in \mathcal{C} : g' \text{ has compact support in } \mathbb{R}_+\}.$$

Recall that we always assume g is extended to be identically zero on $(-\infty, 0)$ so that functions of the form $g(\cdot - a)$ are well defined on $\mathbb{R}_+$ for any $a > 0$. In particular, for $g \in \tilde{\mathcal{C}}$ this extension yields a function in $\mathbf{C}_b^1(\mathbb{R})$ that together with its first derivative is uniformly continuous on $\mathbb{R}$. We will use this fact below to simplify certain technical details involving Taylor's formula applied near zero.

Fix $g \in \tilde{\mathcal{C}}$, and observe that since g is continuous and bounded, the assumptions of Lemma A.2 are clearly satisfied. Thus we have as $r \to \infty$,

$$\bar{X}_g^r(\cdot) \Longrightarrow \bar{X}_g(\cdot),$$

where $\bar{X}_g^r(t) = \frac{1}{r}\sum_{i=1}^{r\bar{E}^r(t)} g(v_i^r)$ and $\bar{X}_g(t) = \alpha t \langle g, \nu \rangle$ for all $t \geq 0$. The function $g \equiv 1$ also satisfies the assumptions of Lemma A.2, and so we have $\bar{E}^r(\cdot) \Longrightarrow \alpha(\cdot)$ as $r \to \infty$. Also recall that we have passed to a subsequence so that $\bar{\mu}^r(\cdot) \Longrightarrow \bar{\mu}^\star(\cdot)$ as $r \to \infty$. Now since the limits $\bar{X}_g(\cdot)$ and $\alpha(\cdot)$ are deterministic, we see that the convergence in the three results just mentioned can be taken to be joint convergence in distribution (cf. [2], Thm. 4.4). Furthermore, by invoking the Skorohod representation theorem, we may assume that there is a single underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$ on which all of the random elements in question are defined, and such that the joint convergence occurs a.s. That is, we have a.s. as $r \to \infty$,

$$(\bar{\mu}^r(\cdot), \bar{X}_g^r(\cdot), \bar{E}^r(\cdot)) \longrightarrow (\bar{\mu}^\star(\cdot), \bar{X}_g(\cdot), \alpha(\cdot)), \tag{5.52}$$

where the convergence is in the product topology of $D([0,\infty), \mathcal{M}_F) \times D([0,\infty), \mathbb{R}) \times D([0,\infty), \mathbb{R})$, where each of the terms in the product is endowed with the relevant Skorohod J_1-topology. Notice that since each of the limits in (5.52) is a.s. continuous, the convergence is in fact uniform on compact time intervals.

Consider any $\omega \in \{\bar{\mu}^\star(0) \neq \mathbf{0}\}$ such that (5.52) and properties (i) and (ii) of Section 3.1 hold, and such that $t^* = \infty$; i.e., we consider any outcome for which the limiting initial fluid scaled queue length $\langle 1, \bar{\mu}^\star(0)\rangle$ is strictly positive, and such that ω is not in any "exceptional" set. We first show that for such an ω, (3.3) holds for all $t \geq 0$ and the chosen $g \in \tilde{\mathcal{C}}$. In what follows, all random elements are understood to be evaluated at this particular ω, and most references to it are suppressed.

Our goal will be to derive a prelimit version of (3.3), which the path $\bar{\mu}^r(\cdot) \in D([0,\infty), \mathcal{M}_F)$ satisfies for sufficiently large r. We will then pass to the limit in this relation to obtain (3.3) for $\bar{\mu}^\star(\cdot)$. We start by noting that since $g \in \tilde{\mathcal{C}}$, g' is uniformly continuous on $\mathbb{R}_+$, so there is a continuous nondecreasing function $\psi_g : [0,\infty) \longrightarrow [0,\infty)$ with $\psi_g(0) = 0$, such that for any $h \in \mathbb{R}$,

$$\sup_{x \in \mathbb{R}} \left\| g'(x+h) - g'(x) \right\| \leq \psi_g(|h|). \tag{5.53}$$

Fix $T > 0$. By (5.52), the fact that $\bar{\mu}^\star(\cdot)$ is continuous, and the fact that for ω, $\bar{\mu}^\star(t) \neq \mathbf{0}$ for all $t \in [0,T]$, we may assume that r is large enough so that $\inf_{t\in[0,T]} \langle 1, \bar{\mu}^r(t)\rangle > 0$. Then for such an r we always have

$$M^r = \left\| \langle 1, \bar{\mu}^r(\cdot)\rangle \right\|_T < \infty, \tag{5.54}$$

$$m^r = \left\| \langle 1, \bar{\mu}^r(\cdot)\rangle^{-1} \right\|_T < \infty. \tag{5.55}$$

Now let $t \in [0,T]$, and for any $n = 1,2,\ldots$ and $j = 0,1,\ldots,n-1$, define $t_j = \frac{jt}{n}$ and $t^j = t_{j+1}$.

Then we have

$$
\begin{aligned}
\langle g,\bar\mu^r(t)\rangle - \langle g,\bar\mu^r(0)\rangle &= \sum_{j=0}^{n-1}\Big(\big\langle g,\bar\mu^r(t^j)\big\rangle - \langle g,\bar\mu^r(t_j)\rangle\Big)\\
&= \sum_{j=0}^{n-1}\left(\big\langle g,\bar\mu^r(t^j)\big\rangle - \left\langle g(\cdot - \bar S^r_{t_j,t^j}),\bar\mu^r(t_j)\right\rangle\right)\\
&\qquad + \sum_{j=0}^{n-1}\left(\left\langle g(\cdot - \bar S^r_{t_j,t^j}),\bar\mu^r(t_j)\right\rangle - \langle g,\bar\mu^r(t_j)\rangle\right)\\
&= \sum_{j=0}^{n-1}\frac{1}{r}\sum_{i=r\bar E^r(t_j)+1}^{r\bar E^r(t^j)} g\left(v_i^r - \bar S^r_{U_i^r/r,t^j}\right)\\
&\qquad + \sum_{j=0}^{n-1}\left\langle g(\cdot - \bar S^r_{t_j,t^j}) - g(\cdot),\bar\mu^r(t_j)\right\rangle,
\end{aligned}
\tag{5.56}
$$

where the first term in the last equality is by (5.2) and the fact that $(1_{(0,\infty)}g)\equiv g$, since $g(0)=0$ for $g\in\tilde{\mathcal{C}}$.

We handle the two terms in (5.56) separately. To begin with, since $g\in\tilde{\mathcal{C}}$ has been extended to be an element of $\mathbf{C}_b^1(\mathbb{R})$, we have the following first order Taylor expansion for each $j=0,1,\dots,n-1$ and each $x\in\mathbb{R}_+$:

$$
g(x-\bar S^r_{t_j,t^j}) - g(x) = g'(w_j^x)h_j, \tag{5.57}
$$

where $h_j = -\bar S^r_{t_j,t^j}$ and $w_j^x\in\mathbb{R}$ is in the interval $[x-\bar S^r_{t_j,t^j},x]$. Note that

$$
\max_{j<n}|h_j| = \max_{j<n}\left|\bar S^r_{t_j,t^j}\right| \le \frac{t}{n}\big\|\langle 1,\bar\mu^r(\cdot)\rangle^{-1}\big\|_T = \frac{tm^r}{n}. \tag{5.58}
$$

For each $j\in\{0,\dots,n-1\}$, let $z_j = \sup_{s\in[t_j,t^j)}\langle 1,\bar\mu^r(s)\rangle^{-1}$ and define $\tilde h_j = -z_j\frac{t}{n}$. Then

$$
\begin{aligned}
\sum_{j=0}^{n-1}\big|h_j-\tilde h_j\big| &= \sum_{j=0}^{n-1}\left|z_j\frac{t}{n} - \bar S^r_{t_j,t^j}\right|\\
&= \sum_{j=0}^{n-1}\left(z_j\frac{t}{n} - \bar S^r_{t_j,t^j}\right)\\
&= \sum_{j=0}^{n-1}\left(z_j\frac{t}{n}\right) - \bar S^r_{0,t}.
\end{aligned}
\tag{5.59}
$$

For each $n=1,2,\dots$ and $s\in[0,t)$, let $k^n(s)=\sum_{j=0}^{n-1}z_j 1_{[t_j,t^j)}(s)$ and define $k^n(t)=0$. Now we can make the following estimate for the second term in (5.56):

$$
\begin{aligned}
&\left|\sum_{j=0}^{n-1}\left\langle g(\cdot-\bar{S}^r_{t_j,t^j})-g(\cdot),\bar{\mu}^r(t_j)\right\rangle-\sum_{j=0}^{n-1}\left\langle g'(\cdot)\tilde{h}_j,\bar{\mu}^r(t_j)\right\rangle\right| \\
&\quad\le\sum_{j=0}^{n-1}\sup_{x\in\mathbb{R}_+}\left|g(x-\bar{S}^r_{t_j,t^j})-g(x)-g'(x)\tilde{h}_j\right|\langle 1,\bar{\mu}^r(t_j)\rangle, \\
&\quad=\sum_{j=0}^{n-1}\sup_{x\in\mathbb{R}_+}\left|g'(w^x_j)h_j-g'(x)\tilde{h}_j\right|\langle 1,\bar{\mu}^r(t_j)\rangle \\
&\quad\le\left\|\langle 1,\bar{\mu}^r(\cdot)\rangle\right\|_T\sum_{j=0}^{n-1}\sup_{x\in\mathbb{R}_+}\Big(\left|g'(w^x_j)-g'(x)\right|\left|h_j\right|+\left|g'(x)\right|\left|h_j-\tilde{h}_j\right|\Big), \\
&\quad\le M^r\left(n\psi_g\left(\frac{tm^r}{n}\right)\frac{tm^r}{n}+\|g'\|_\infty\left(\sum_{j=0}^{n-1}\left(z_j\frac{t}{n}\right)-\bar{S}^r_{0,t}\right)\right) \\
&\quad=M^r\left(\psi_g\left(\frac{tm^r}{n}\right)tm^r+\|g'\|_\infty\left(\int_0^t k^n(s)ds-\int_0^t\langle 1,\bar{\mu}^r(s)\rangle^{-1}ds\right)\right).
\end{aligned}
\tag{5.60}
$$

In the third line above we have used the Taylor expansion (5.57). The last inequality above then follows from (5.54), (5.53), (5.58) and (5.59). The substitution of the second integral in the last line follows by (5.55) and (3.13). Now we let $n\to\infty$ in the above inequality. By the continuity of ψ_g and the fact that $\psi_g(0)=0$, we see that the first term in the outer parentheses tends to zero. Note that $k^n(s)\longrightarrow\langle 1,\bar{\mu}^r(s)\rangle^{-1}$, as $n\to\infty$, for any $s\in[0,t)$ at which $\langle 1,\bar{\mu}^r(\cdot)\rangle^{-1}$ is continuous. Since it is continuous for almost every s (the path $\bar{\mu}^r(\cdot)$ is right continuous with finite left limits), the second term in the outer parentheses tends to zero by (5.55) and bounded convergence. Furthermore, we note that

$$
\sum_{j=0}^{n-1}\left\langle g'(\cdot)\tilde{h}_j,\bar{\mu}^r(t_j)\right\rangle=-\sum_{j=0}^{n-1}\left\langle g',\bar{\mu}^r(t_j)\right\rangle z_j\frac{t}{n},
$$

and that as $n\to\infty$,

$$
-\sum_{j=0}^{n-1}\left\langle g',\bar{\mu}^r(t_j)\right\rangle z_j\frac{t}{n}\longrightarrow-\int_0^t\frac{\langle g',\bar{\mu}^r(s)\rangle}{\langle 1,\bar{\mu}^r(s)\rangle}ds,
$$

by (5.54), (5.55) and bounded convergence, since the function $\langle g',\bar{\mu}^r(\cdot)\rangle\langle 1,\bar{\mu}^r(\cdot)\rangle^{-1}$ is also continuous for almost every s. Together with the estimate (5.60), this implies that as $n\to\infty$,

$$
\sum_{j=0}^{n-1}\left\langle g(\cdot-\bar{S}^r_{t_j,t^j})-g(\cdot),\bar{\mu}^r(t_j)\right\rangle\longrightarrow-\int_0^t\frac{\langle g',\bar{\mu}^r(s)\rangle}{\langle 1,\bar{\mu}^r(s)\rangle}ds. \tag{5.61}
$$

We handle the first term of (5.56) in a similar (although simpler) fashion. Once again we can use a first order Taylor expansion for each summand appearing in this term:

$$
g(v^r_i-\bar{S}^r_{U^r_i/r,t^j})=g(v^r_i)+g'(w^i_j)h^i_j, \tag{5.62}
$$

where $h^i_j = -\bar{S}^r_{U^r_i/r,t^j}$, and $w^i_j \in [v^r_i - \bar{S}^r_{U^r_i/r,t^j}, v^r_i]$. Since $\left|t^j - (U^r_i/r)\right| \leq t/n$ for each pair j, i in the first term of (5.56), we have as before that

$$\max_{j,i} |h^i_j| \leq \frac{t}{n} \left\| \langle 1, \bar{\mu}^r(\cdot)\rangle^{-1} \right\|_T = \frac{tm^r}{n}. \tag{5.63}$$

Now using the Taylor expansion (5.62) along with (5.63) and recalling that $\bar{X}^r_g(t) = \frac{1}{r}\sum_{i=1}^{r\bar{E}^r(t)} g(v^r_i)$, we have

$$\begin{aligned} \left| \left(\sum_{j=0}^{n-1} \frac{1}{r} \sum_{i=r\bar{E}^r(t_j)+1}^{r\bar{E}^r(t^j)} g(v^r_i - \bar{S}^r_{U^r_i/r,t^j}) \right) - \bar{X}^r_g(t) \right| &= \left| \sum_{j=0}^{n-1} \frac{1}{r} \sum_{i=r\bar{E}^r(t_j)+1}^{r\bar{E}^r(t^j)} g'(w^i_j) h^i_j \right| \\ &\leq \bar{E}^r(t) \|g'\|_\infty \frac{tm^r}{n}, \end{aligned} \tag{5.64}$$

which tends to zero as $n \to \infty$. By combining this fact with (5.61), we can let $n \to \infty$ in (5.56) to obtain the relation

$$\langle g, \bar{\mu}^r(t)\rangle = \langle g, \bar{\mu}^r(0)\rangle - \int_0^t \frac{\langle g', \bar{\mu}^r(s)\rangle}{\langle 1, \bar{\mu}^r(s)\rangle} ds + \frac{1}{r} \sum_{i=1}^{r\bar{E}^r(t)} g(v^r_i), \tag{5.65}$$

for each $t \in [0, T]$. We would like to let $r \to \infty$ in this relation to obtain

$$\langle g, \bar{\mu}^\star(t)\rangle = \langle g, \bar{\mu}^\star(0)\rangle - \int_0^t \frac{\langle g', \bar{\mu}^\star(s)\rangle}{\langle 1, \bar{\mu}^\star(s)\rangle} ds + \alpha t \langle g, \nu\rangle, \tag{5.66}$$

for each $t \in [0, T]$. For this, fix $t \in [0, T]$. By (5.52), we have that the left side, as well as each of the first and third terms on the right side of (5.65), converges to the corresponding term in (5.66). Similarly, (5.52), (5.54) and (5.55) imply that the integrands in the second term on the right side of (5.65) are uniformly bounded, and converge pointwise on $[0, t]$ to the integrand in the second term on the right side of (5.66). Thus the integrals converge by bounded convergence, and it is clear that we obtain (5.66) from (5.65) by letting $r \to \infty$.

We have shown that for the given $g \in \tilde{\mathcal{C}}$, a.s. on $\{\bar{\mu}^\star(0) \neq \mathbf{0}\}$, (3.3) holds for all $t \in [0, T]$. Since $T > 0$ was arbitrary, we may replace $t \in [0, T]$ by $t \geq 0$ in this statement. Notice that the exceptional set in the above statement may depend on g (cf. (5.52)). We now show that a.s. on $\{\bar{\mu}^\star(0) \neq \mathbf{0}\}$, (3.3) holds for all $t \geq 0$ and all $g \in \tilde{\mathcal{C}}$. For this, suppose that there are functions $g_k, g \in \tilde{\mathcal{C}}$, $k = 1, 2, \ldots$, such that $\{g_k\}_{k=1}^\infty$ and $\{g'_k\}_{k=1}^\infty$ are uniformly bounded and as $k \to \infty$,

$$g_k \longrightarrow g \qquad \text{and} \qquad g'_k \longrightarrow g', \tag{5.67}$$

pointwise on $\mathbb{R}_+$. Suppose further that for a given $\omega \in \{\bar{\mu}^\star(0) \neq \mathbf{0}\}$ such that $\bar{\mu}^\star(\cdot)$ is continuous and $t^\star = \infty$, (3.3) holds for each k, i.e., we have for each $k = 1, 2, \ldots$, and $t \geq 0$,

$$\langle g_k, \bar{\mu}^\star(t)\rangle = \langle g_k, \bar{\mu}^\star(0)\rangle - \int_0^t \frac{\langle g'_k, \bar{\mu}^\star(s)\rangle}{\langle 1, \bar{\mu}^\star(s)\rangle} ds + \alpha t \langle g_k, \nu\rangle. \tag{5.68}$$

Letting $k \to \infty$ in (5.68), it follows by (5.67) and bounded convergence that the left side as well as the first and third terms on the right side of (5.68) converge respectively to the corresponding terms of (3.3). Similarly, since $\bar{\mu}^\star(\cdot)$ is continuous and $t^\star = \infty$, the integral term also converges by

(5.67) and two applications of bounded convergence. Thus for this ω, we obtain (3.3) from (5.68) by letting $k \to \infty$.

Thus, to show that a.s. on $\{\bar{\mu}^\star(0) \neq \mathbf{0}\}$, (3.3) holds for all $t \geq 0$ and all $g \in \tilde{\mathcal{C}}$, it suffices to show that there is a countable subset $V \subset \tilde{\mathcal{C}}$ such that for any $g \in \tilde{\mathcal{C}}$, there is a sequence $\{g_k\}_{k=1}^\infty \subset V$ that together with $\{g_k'\}_{k=1}^\infty$ is uniformly bounded and satisfies (5.67). One way to construct such a set V is the following: For each $k = 1, 2, \ldots$, and $j = 1, \ldots, k^2$, let O_j^k be the open interval $((j-1)/k, (j+1)/k) \subset \mathbb{R}_+$. Then for each k, $\mathcal{O}^k = \{O_j^k\}_{j=1}^{k^2}$ is an open cover of the compact interval $[1/k, k]$. Let $\mathcal{G}^k = \{g_j^k\}_{j=1}^{k^2}$ be a partition of unity subordinate to the open cover $\mathcal{O}^k$. More precisely, $\mathcal{G}^k$ is defined such that for each k and $j = 1, 2, \ldots, k^2$, $g_j^k : \mathbb{R}_+ \longrightarrow [0,1]$ is a smooth function with support contained in O_j^k, and such that for each $x \in [1/k, k]$, $\sum_{j=1}^{k^2} g_j^k(x) = 1$. Let $\tilde{\mathcal{G}}^k$ be the set of all finite linear combinations of elements of $\mathcal{G}^k$ with rational coefficients, and let V^k be the set of all functions of the form

$$g(x) = \int_0^x \tilde{g}(y)dy,$$

where $\tilde{g} \in \tilde{\mathcal{G}}^k$. Clearly, the set $V = \bigcup_{k=1}^\infty V^k$ is a countable subset of $\tilde{\mathcal{C}}$. To see that V has the desired properties, suppose $g \in \tilde{\mathcal{C}}$. Approximate g' first by defining for each $k = 1, 2, \ldots$,

$$g_k' = \sum_{j=1}^{k^2} c_j^k g_j^k,$$

where c_j^k is a rational number chosen such that $|c_j^k - g'(j/k)| < 1/k$. Let $g_k(x) = \int_0^x g_k'(y)dy$. Then it is not difficult to verify that $\{g_k\}_{k=1}^\infty \subset V$ and that (5.67) is satisfied.

Finally, we show that in fact a.s. on $\{\bar{\mu}^\star(0) \neq \mathbf{0}\}$, (3.3) holds for all $t \geq 0$ and all $g \in \mathcal{C}$. For this, fix $g \in \mathcal{C}$. For $n = 1, 2, \ldots$, choose a function $\psi_n \in \mathbf{C}_b^1(\mathbb{R}_+)$ such that $\psi_n(x) \in [0,1]$ and $|\psi_n'(x)| \leq 2$ for all $x \in \mathbb{R}_+$, $\psi_n \equiv 1$ on $[0, n]$ and $\psi_n \equiv 0$ on $[n+1, \infty)$. Let $g_n = \psi_n g$, and note that $\{g_n\}_{n=1}^\infty$, $\{g_n'\}_{n=1}^\infty$ are uniformly bounded and that $g_n \longrightarrow g$, and $g_n' \longrightarrow g'$ pointwise on $\mathbb{R}_+$ as $n \to \infty$. Since g_n' has compact support, we see that $g_n \in \tilde{\mathcal{C}}$. Therefore, a.s. on $\{\bar{\mu}^\star(0) \neq \mathbf{0}\}$, (3.3) holds for all $t \geq 0$ and for each of the functions g_n. By the same argument as that appearing after (5.68), this implies that (3.3) also holds for g. This completes the proof of Theorem 5.1. ■

5.4 Proof of Convergence to Fluid Model Solutions

Proof of Theorem 3.2. By Theorem 5.1, the sequence $\{\bar{\mu}^r(\cdot)\}$ of measure valued processes is tight, and any limit point $\bar{\mu}^\star(\cdot)$ has sample paths which a.s. are fluid model solutions for the critical data (α, ν). By (3.19), $\bar{\mu}^\star(0)$ is equal in distribution to Θ, so it remains to show that $\bar{\mu}^\star(\cdot)$ is unique in law. Theorem 3.1 asserts that fluid model solutions are unique given an initial value $\xi \in \mathcal{M}_F^c = \{\zeta \in \mathcal{M}_F : \langle 1_{\{x\}}, \zeta \rangle = 0 \text{ for all } x \in \mathbb{R}_+\}$. More precisely, given a $\xi \in \mathcal{M}_F^c$, the unique fluid model solution $\bar{\mu}(\cdot)$ for critical data (α, ν) with initial value $\bar{\mu}(0) = \xi$ is given by $\Xi(\xi)$, where $\Xi : \mathcal{M}_F^c \longrightarrow D([0,\infty), \mathcal{M}_F)$ is the mapping introduced in Section 4.3. Since $\bar{\mu}^\star(0)$ is equal in distribution to Θ, we have by (3.22) that $\bar{\mu}^\star(0) \in \mathcal{M}_F^c$ a.s. Thus we see that a.s.,

$$\bar{\mu}^\star(\cdot) = \Xi(\bar{\mu}^\star(0)). \tag{5.69}$$

By Lemma 4.9, the mapping Ξ is measurable. So, the law of $\bar{\mu}^\star(\cdot)$ is uniquely determined by the law of the random measure Θ, which completes the proof. ■

A Appendix

Here we prove two ancillary lemmas which are needed in several places throughout the paper. The first lemma provides an equivalent formulation of assumption (3.22), which is better suited to our analysis in Section 5. The second lemma provides a basic functional weak law of large numbers, which we include for completeness.

Lemma A.1 *Let Θ be a random element of $\mathcal{M}_F$. Then the following are equivalent:*

(i) Almost surely, $\langle 1_{\{x\}}, \Theta \rangle = 0$ for all $x \in \mathbb{R}_+$.

(ii) For each $\epsilon > 0$,

$$\lim_{\kappa \downarrow 0} \mathbf{P}\left(\sup_{x \in \mathbb{R}_+} \langle 1_{[x,x+\kappa]}, \Theta \rangle < \epsilon \right) = 1.$$

Proof. The implication $(ii) \Rightarrow (i)$ is straightforward. We prove $(i) \Rightarrow (ii)$, by contradiction. Suppose that (ii) does not hold. Then there exists an $\epsilon > 0$ and a sequence κ_n such that $\kappa_n \downarrow 0$ as $n \to \infty$, and

$$\liminf_{n \to \infty} \mathbf{P}\left(\sup_{x \in \mathbb{R}_+} \langle 1_{[x,x+\kappa_n]}, \Theta \rangle \geq \epsilon \right) > 0. \tag{A.1}$$

Define

$$A_n = \left\{ \zeta \in \mathcal{M}_F : \sup_{x \in \mathbb{R}_+} \langle 1_{[x,x+\kappa_n]}, \zeta \rangle \geq \epsilon \right\},$$

and define $A = \bigcap_n A_n$. Since $A_n \supset A_{n+1}$, (A.1) implies that $\mathbf{P}(\Theta \in A) > 0$. We claim that any $\zeta \in A$ has an atom, which yields a contradiction to (i). To see this, let $\zeta \in A$, and define

$$B_n = \left\{ x \in \mathbb{R}_+ : \langle 1_{[x,x+\kappa_n]}, \zeta \rangle \geq \frac{\epsilon}{2} \right\}.$$

The sets B_n are non-empty since $\zeta \in A$. Suppose $\{x_k\}_{k=1}^{\infty} \subset B_n$ and $x_k \longrightarrow x \in \mathbb{R}_+$ as $k \to \infty$. Since there exists a sequence of closed intervals $\{I_k\}_{k=1}^{\infty}$ such that for each $k = 1, 2, \ldots$, $I_{k+1} \subset I_k$, $I_k \supset [x_k, x_k + \kappa_n] \cup [x, x + \kappa_n]$, and $\bigcap_k I_k = [x, x + \kappa_n]$, the fact that $\langle 1_{I_k}, \zeta \rangle \geq \langle 1_{[x_k,x_k+\kappa_n]}, \zeta \rangle \geq \epsilon/2$ implies that $x \in B_n$. So the sets B_n are closed. Moreover, $\zeta \in \mathcal{M}_F$ implies that the sets B_n are bounded, and therefore compact. Finally, $B_n \supset B_{n+1}$ implies that $B = \bigcap_n B_n \neq \varnothing$. So continuity from above of ζ implies that for any $x \in B$, $\langle 1_{\{x\}}, \zeta \rangle \geq \epsilon/2$, i.e., x is an atom of ζ. ∎

Lemma A.2 *Consider a sequence of real numbers $r \in (0, \infty)$ such that $r \to \infty$. For each r, let $\{u_i^r\}_{i=1}^{\infty}$ be a sequence of independent nonnegative random variables, such that $\{u_i^r\}_{i=2}^{\infty}$ is i.i.d. Assume for each r that $\mathbf{E}[u_1^r] < \infty$ and $\mathbf{E}[u_2^r]^{-1} = \alpha^r \in (0, \infty)$. For each r, let $\{v_i^r\}_{i=1}^{\infty}$ be an i.i.d. sequence of strictly positive random variables with common distribution ν^r. Let $\alpha \in (0, \infty)$, let ν be a probability distribution on $\mathbb{R}_+$, and let $g : \mathbb{R}_+ \longrightarrow \mathbb{R}$ be a Borel measurable function that is ν–a.e.*

continuous and satisfies $\langle |g|, \nu\rangle < \infty$ and $\langle |g|, \nu^r\rangle < \infty$ for every r. Assume that the following asymptotic assumptions hold as $r \to \infty$:

$$\alpha^r \longrightarrow \alpha, \tag{A.2}$$
$$\nu^r \xrightarrow{\text{w}} \nu, \tag{A.3}$$
$$\langle g^+, \nu^r\rangle \longrightarrow \langle g^+, \nu\rangle, \tag{A.4}$$
$$\langle g^-, \nu^r\rangle \longrightarrow \langle g^-, \nu\rangle, \tag{A.5}$$
$$\mathbf{E}[u_1^r]/r \longrightarrow 0, \tag{A.6}$$
$$\mathbf{E}[u_2^r; u_2^r > r] \longrightarrow 0. \tag{A.7}$$

Let $U_0^r = 0$, $U_i^r = \sum_{j=1}^i u_j^r$, $i = 1, 2, \ldots$, and for $t \geq 0$, let $E^r(t) = \sup\{i \geq 0 : U_i^r \leq t\}$, $\bar{E}^r(t) = \frac{1}{r}E^r(rt)$, $\bar{X}_g^r(t) = \frac{1}{r}\sum_{i=1}^{r\bar{E}^r(t)} g(v_i^r)$, and $\bar{X}_g(t) = \alpha t\langle g, \nu\rangle$. Then as $r \to \infty$,

$$\bar{X}_g^r(\cdot) \Longrightarrow \bar{X}_g(\cdot).$$

Proof. It suffices to show that for each $T > 0$,

$$\left\|\bar{X}_g^r(\cdot) - \bar{X}_g(\cdot)\right\|_T \Longrightarrow 0, \qquad \text{as } r \to \infty, \tag{A.8}$$

where $\|\cdot\|_T$ denotes the supremum norm over $[0, T]$. Note that both g^+ and g^- themselves satisfy the conditions of Lemma A.2, and that $\bar{X}_g^r(\cdot)$ and $\bar{X}_g(\cdot)$ are linear in g. So it is sufficient to show that (A.8) holds with g^+ in place of g.

We require a condition on the tails of the distributions of the members of the sequence $\{g^+(v_i^r)\}_{i=1}^\infty$ which will imply a weak law of large numbers. Since g^+ is ν-a.e. continuous, the distribution $\nu_{g^+}^r$ of $g^+(v_1^r)$ converges weakly as $r \to \infty$ to the distribution ν_{g^+}, where $\nu_{g^+}(A) = \nu(\{x \in \mathbb{R}_+ : g^+(x) \in A\})$, for any Borel set $A \subset \mathbb{R}_+$. By (A.4), we also have as $r \to \infty$,

$$\left\langle \chi, \nu_{g^+}^r\right\rangle = \langle g^+, \nu^r\rangle \longrightarrow \langle g^+, \nu\rangle = \langle \chi, \nu_{g^+}\rangle < \infty. \tag{A.9}$$

Since ν_{g^+} can have at most countably many atoms, we can always choose an arbitrarily large $K > 0$ so that $\chi 1_{[0,K]}$ is ν_{g^+}-a.e. continuous and therefore

$$\left\langle \chi 1_{[0,K]}, \nu_{g^+}^r\right\rangle \longrightarrow \left\langle \chi 1_{[0,K]}, \nu_{g^+}\right\rangle, \qquad \text{as } r \to \infty.$$

Combined with (A.9), this implies that

$$\mathbf{E}[g^+(v_1^r); g^+(v_1^r) > r] \longrightarrow 0, \qquad \text{as } r \to \infty. \tag{A.10}$$

Together, (A.4) and (A.10) imply by the weak law of large numbers for triangular arrays that, for $t \geq 0$ fixed,

$$\frac{1}{r}\sum_{i=1}^{\lfloor rt\rfloor} g^+(v_i^r) \Longrightarrow t\left\langle g^+, \nu\right\rangle, \qquad \text{as } r \to \infty. \tag{A.11}$$

Verification of this fact uses standard truncation arguments hinging on (A.10) (cf. [14], pp. 41–43, for example). Since the right and left members of (A.11) are nondecreasing as functions of t, and the right side defines a deterministic process that is uniformly continuous on each compact time interval, the convergence in (A.11) is actually uniform on compact time intervals, i.e., for each $T > 0$,

$$\sup_{t\in[0,T]} \left| \frac{1}{r} \sum_{i=1}^{\lfloor rt \rfloor} g^+(v_i^r) - t \left\langle g^+, \nu \right\rangle \right| \Longrightarrow 0, \qquad \text{as } r \to \infty. \tag{A.12}$$

Our assumptions (A.2), (A.6) and (A.7) on $\{u_i^r\}_{i=1}^\infty$ imply by the weak law of large numbers for renewal processes that, for each $T > 0$, $\left\| \bar{E}^r(\cdot) - \alpha(\cdot) \right\|_T \Longrightarrow 0$ as $r \to \infty$, where $\alpha(t) = \alpha t$. Since $\alpha(\cdot)$ is deterministic, we can combine this with (A.12) using the random time change theorem (cf. [2], Section 17) to obtain for each $T > 0$ as $r \to \infty$,

$$\left\| \frac{1}{r} \sum_{i=1}^{r\bar{E}^r(\cdot)} g^+(v_i^r) - \alpha(\cdot) \left\langle g^+, \nu \right\rangle \right\|_T \Longrightarrow 0. \tag{A.13}$$

■

Acknowledgement. The authors wish to thank Pat Fitzsimmons for helpful discussions during the course of this work.

References

[1] F. Baccelli and D. Towsley. The customer response times in the processor sharing queue are associated. *Queueing Systems*, **7** (1990), 269–282.

[2] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons, Inc., New York, 1968.

[3] M. Bramson. Instability of FIFO queueing networks. *Annals of Applied Probability*, **4** (1994), 414–431.

[4] M. Bramson. Instability of FIFO queueing networks with quick service times. *Annals of Applied Probability*, **4** (1994), 693–718.

[5] M. Bramson. Convergence to equilibria for fluid models of FIFO queueing networks. *Queueing Systems: Theory and Applications*, **22** (1996), 5–45.

[6] M. Bramson. Convergence to equilibria for fluid models of head-of-the-line proportional processor sharing queueing networks. *Queueing Systems: Theory and Applications*, **23** (1997), 1–26.

[7] M. Bramson. State space collapse with applications to heavy traffic limits for multiclass queueing networks. *Queueing Systems: Theory and Applications*, **30** (1998), 89–148.

[8] M. Bramson. Stability of two families of queueing networks and a discussion of fluid limits. *Queueing Systems: Theory and Applications*, **28** (1998), 7–31.

[9] H. Chen, O. Kella and G. Weiss. Fluid approximations for a processor-sharing queue. *Queueing Systems: Theory and Applications*, **27** (1997), 99–125.

[10] J. G. Dai. On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *Annals of Applied Probability*, **5** (1995), 49–77.

[11] J. G. Dai. Stability of open multiclass queueing networks via fluid models. In *Stochastic Networks, IMA Volumes in Mathematics and Its Applications*, Volume 71, F. Kelly and R. J. Williams (eds.), Springer-Verlag, New York, 1995, pp. 71–90.

[12] D. A. Dawson. Measure Valued Markov Processes. In *École d'Été de Probabilités de Saint-Flour*, Volume XXI, Lecture Notes in Math, No. 1541, P. L. Hennequin (ed.), Springer-Verlag, Berlin, 1993.

[13] B. Doytchinov, J. Lehoczky and S. Shreve. Real-time queues in heavy traffic with earliest-deadline-first-queue discipline. To appear in *Annals of Applied Probability.*

[14] R. T. Durrett. *Probability: Theory and Examples*, 2^{nd} Edition. Duxbury Press, Belmont, 1996

[15] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* John Wiley & Sons, Inc., New York, 1986.

[16] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions.* CRC Press, Boca Raton, 1992.

[17] W. Feller. *An Introduction to Probability Theory and Its Applications*, Volume 2, 2^{nd} Edition. John Wiley & Sons, Inc., New York, 1971.

[18] S. Grishechkin. GI/G/1 processor sharing queue in heavy traffic. *Advances in Applied Probability*, **26** (1994), 539–555.

[19] D. L. Iglehart and W. Whitt. Multiple channel queues in heavy traffic I. *Advances in Applied Probability*, **2** (1970), 150–177.

[20] A. Jean-Marie and P. Robert. On the transient behavior of the processor sharing queue. *Queueing Systems*, **17** (1994), 129–136.

[21] O. Kallenberg. *Random Measures.* Academic Press, New York, 1983.

[22] Yu. V. Prohorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability and its Applications*, **1** (1956), 157–214.

[23] A. L. Puha, A. L. Stoylar and R. J. Williams. The fluid limit of an overloaded processor sharing queue. Preprint.

[24] R. J. Williams. Diffusion approximations for open multiclass queueing networks: sufficient conditions involving state space collapse. *Queueing Systems: Theory and Applications*, **30** (1998), 27–88.

[25] S. F. Yashkov. Processor-sharing queues: some progress in analysis. *Queueing Systems*, **2** (1987), 1–17.