The Annals of Applied Probability
2022, Vol. 32, No. 4, 2587–2651
https://doi.org/10.1214/21-AAP1741

HEAVY TRAFFIC SCALING LIMITS FOR SHORTEST REMAINING PROCESSING TIME QUEUES WITH HEAVY TAILED PROCESSING TIME DISTRIBUTIONS

BY SAYAN BANERJEE[1,a], AMARJIT BUDHIRAJA[1,b] AND AMBER L. PUHA[2,c]

[1]*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill,* [a]*sayan@email.unc.edu,* [b]*amarjit@unc.edu*

[2]*Department of Mathematics, California State University San Marcos,* [c]*apuha@csusm.edu*

We study a single server queue operating under the shortest remaining processing time (SRPT) scheduling policy; that is, the server preemptively serves the job with the shortest remaining processing time first. Since one needs to keep track of the remaining processing times of all jobs in the system in order to describe the evolution, a natural state descriptor for an SRPT queue is a measure valued process in which the state of the system at a given time is the finite nonnegative Borel measure on the nonnegative real line that puts a unit atom at the remaining processing time of each job in system. In this work we are interested in studying the asymptotic behavior of the suitably scaled measure valued state descriptors for a sequence of SRPT queuing systems. Gromoll, Kruk and Puha (*Stoch. Syst.* **1** (2011) 1–16) have studied this problem under diffusive scaling (time is scaled by r^2 and the mass of the measure normalized by r, where r is a scaling parameter approaching infinity). In the setting where the processing time distributions have *bounded support*, under suitable conditions, they show that the measure valued state descriptors converge in distribution to the process that at any given time is a single atom located at the right edge of the support of the processing time distribution with the size of the atom fluctuating randomly in time. In the setting where the processing time distributions have *unbounded support*, under suitable conditions, they show that the diffusion scaled measure valued state descriptors converge in distribution to the process that is identically zero. In Puha (*Ann. Appl. Probab.* **25** (2015) 3381–3404) for the setting where the processing time distributions have *unbounded support and light tails*, a nonstandard scaling of the queue length process is shown to give rise to a form of state space collapse that results in a nonzero limit.

In the current work we consider the case where processing time distributions have finite second moments and regularly varying tails. Results of Puha (*Ann. Appl. Probab.* **25** (2015) 3381–3404) suggest that the right scaling for the measure valued process is governed by a parameter c^r that is given as a certain inverse function related to the tails of the first moment of the processing time distribution. Using this parameter we consider a novel scaling for the measure valued process in which the time is scaled by a factor of r^2, the mass is scaled by the factor c^r/r and the space (representing the remaining processing times) is scaled by the factor $1/c^r$. We show that the scaled measure valued process converges in distribution (in the space of paths of measures). In a sharp contrast to results for bounded support and light tailed service time distributions, this time there is no state space collapse and the limiting measures are not concentrated on a single atom. Nevertheless, the description of the limit is simple and given explicitly in terms of a certain $\mathbb{R}_+$ valued random field which is determined from a single Brownian motion. Along the way we establish convergence of suitably scaled workload

Received March 2020; revised November 2020.

MSC2020 subject classifications. Primary 60K25, 60F17; secondary 60G57, 60G60, 68M20.

Key words and phrases. Heavy traffic, heavy tails, queueing, shortest remaining processing time, regular variation, measure valued processes, functional central limit theorem, random field, state space collapse, synchronization phenomenon, intertwined SRPT queues, derivative of the Skorohod map.

and queue length processes. We also show that as the tail of the distribution of job processing times becomes lighter in an appropriate fashion, the difference between the limiting queue length process and the limiting workload process converges to zero, thereby approaching the behavior of state space collapse.

CONTENTS

1. Introduction. We study a single-server, single-class queue operating under the shortest remaining processing time (SRPT) service discipline. Jobs arrive to the queue according to a renewal process. Each such job has associated with it a processing time, which is a random variable that represents the amount of time that the server must spend working on this job to complete its service. The processing times are assumed to be independent and identically distributed. In an SRPT queue, jobs are served one at a time such that the job with the shortest remaining processing time is served first. In particular, upon completing the service of a given job, the server then takes into service the job in system with the shortest remaining processing time. This is done with preemption so that when a job arrives with a processing time that is smaller than the remaining processing time of the job in service, the server places the job in service on hold and begins serving the job that just arrived. Processing is done in a

nonidling fashion so that the server idles only when the system is empty. While SRPT has a large memory requirement for implementation since remaining processing times of all jobs in the queue must be known, it has desirable optimality properties. In particular, it is the service discipline that minimizes queue length (see Schrage [28] and Smith [31]). Therefore, SRPT can serve as a performance benchmark (e.g., Chen and Dong [6]). The survey paper [29] by Schreiber provides a nice discussion of early works concerning SRPT.

One challenge associated with a detailed analysis of SRPT is that, due to the need to keep track of the remaining processing times of all jobs in the system, the state descriptor for an SRPT queue is infinite dimensional, even for exponentially distributed processing times. In order to describe the state of the system, Down, Gromoll and Puha [9, 10] introduce a measure valued process in which the state of the system at a given time is the finite nonnegative Borel measure on the nonnegative real line that puts a unit atom at the remaining processing time of each job in system. Under natural modeling assumptions and asymptotic conditions, they prove a fluid limit theorem (a functional law of large numbers) for this measure valued state descriptor. This yields a fluid analog for the response time of jobs in system at time zero as a function of their remaining processing times at time zero. In the critically loaded case, the rate at which this fluid analog for the response time grows as time tends to infinity is seen to be dependent on the tail behavior of the processing time distribution. These results are consistent with the growth rates obtained in [20] for steady state mean response times as the traffic intensity increases to one. In follow up work, Kruk [19] proves a fluid limit theorem for multiclass SRPT queues that includes convergence of the response times to the expression studied in [9], which justifies it as an approximation. Atar, Biswas, Kaspi and Ramanan [1] develop more general fluid limits for SRPT and other priority queues with time varying arrivals and service rates.

In this work, we consider a sequence of SRPT queues indexed by a scaling parameter r approaching infinity. We are interested in studying the asymptotic behavior of the measure valued state descriptors for this sequence of SRPT queuing systems under diffusion and other suitable scalings. This captures the performance deviation of a critically loaded SRPT queue from the fluid limit by describing the fluctuations. Gromoll, Kruk and Puha [13] provide a first step in this direction by establishing a diffusion limit theorem (a functional central limit theorem), for the sequence of measure valued processes. In [13] for the case where the processing time distributions have *bounded support*, it is shown that, with standard diffusive scaling (time is scaled by r^2 and the mass of the measure normalized by r), under natural modeling assumptions and mild asymptotic and standard heavy traffic conditions, the mass of the (scaled) measure valued state descriptors in the limit concentrates on a single atom located at the right edge of the support of the processing time distribution with the size of the atom fluctuating randomly in time. This is similar in spirit to results for static priority queues where only the queue associated with the lowest priority class is nonempty in the diffusion limit (see [5, 34]). The result for the bounded support case suggests that for processing time distributions with unbounded support, with standard diffusive scaling, one should obtain the trivial limit of the zero process for the scaled measure valued process. This is indeed true under suitable conditions as is also shown in [13]. These results are rederived by Kruk [18] via an alternative argument that leverages diffusion limits for earliest deadline first queues obtained in Kruk [17]. Although the measure valued processes under the standard diffusion scaling converge to the zero process, the workload under the diffusive scaling, which is given as the first moment of the state descriptor measure, does not converge to the zero process. Indeed, since SRPT is a nonidling service discipline, the diffusion limit for the workload process (which is independent of the scheduling policy) corresponds to a semi-martingale reflected Brownian motion (SRBM) [14]. Heuristically the above results say that, for processing time distributions with unbounded support, SRPT minimizes the queue length so

efficiently that, in the diffusion limit, the queue length process is of a smaller order than the workload process.

This raises the important problem of quantifying the precise difference in orders of the queue length and workload processes. In [25], Puha studies the case where the the processing time distributions have light tails (rapidly varying with index $-\infty$, e.g., an exponential distribution) and identifies the key quantity that determines the correct scaling for the queue length process. This quantity, denoted as c^r and defined in equation (2.9) here, is given in terms of a certain inverse function related to the tails of the first moment of the processing time distribution. Using the scaling factor c^r, [25] establishes a state space collapse result that specifies conditions under which

$$(c^r \hat{Q}^r, \hat{W}^r) \text{ converges in distribution to } (W^\infty, W^\infty) \quad \text{as } r \to \infty, \tag{1.1}$$

where $\hat{Q}^r$ and $\hat{W}^r$ are the queue length and workload processes, respectively, of the rth system with standard diffusive scaling and W^∞ is a certain SRBM on $\mathbb{R}_+$. Although [25] does not consider the convergence of the measure valued state descriptor, the result in (1.1) suggests that with an appropriate scaling, this measure valued process converges in distribution to a process of Dirac measures at one (with random weights); see Remark 2 for additional comments on this point.

In this work, we study the setting where the processing times have finite second moments and regularly varying tails (see (2.1)). Such heavy tailed processing time distributions arise naturally in various application domains, for example, file transfer models and cloud computing [7, 21], which motivates us to consider the performance of SRPT in this setting in more detail. For this, we study the asymptotic behavior of the full measure valued state descriptor under an appropriate scaling. As in [25] the quantity c^r is once more central to identifying the correct scaling. The scaled measure valued process, denoted as $\widetilde{\mathcal{Z}}^r(\cdot)$, is defined using three types of scaling: the time is scaled by a factor of r^2, the mass is scaled by the factor c^r/r and the space (representing the remaining processing times) is scaled by the factor $1/c^r$; see (2.11) for a precise definition. One of our main results (Theorem 3) gives convergence of $\widetilde{\mathcal{Z}}^r(\cdot)$ in distribution, in $\mathcal{D}([0,\infty) : \mathcal{M}_F)$ (the space of right continuous functions with left limits equipped with the usual Skorohod topology, where $\mathcal{M}_F$ is the space of finite nonnegative measures on $\mathbb{R}_+$ with the topology of weak convergence), to a limit measure valued process $\widetilde{\mathcal{Z}}(\cdot)$. In a sharp contrast to results for bounded support and light tailed service time distributions, this time there is no state space collapse and the limiting measures are not concentrated on a single atom. Nevertheless, the description of the limit is simple and given explicitly in terms of a certain $\mathbb{R}_+$ valued random field $\{W_a(t), t \in [0,\infty), a \in [0,\infty]\}$ which is determined from a single Brownian motion; see (3.2)–(3.5). Roughly speaking, $W_a(\cdot)$ can be interpreted as the asymptotic (diffusion scaled) workload process associated with jobs in the system with remaining processing times at most ac^r. In terms of $\{W_a(\cdot), a \in (0,\infty)\}$, the limiting measure valued process $\widetilde{\mathcal{Z}}(\cdot)$ is characterized as follows: for $t \in [0,\infty)$, $\widetilde{\mathcal{Z}}(t)(\{0\}) = 0$, $\widetilde{\mathcal{Z}}(t)([0,\infty)) = \int_{[0,\infty)} \frac{1}{x^2} W_x(t)\,dx$ and

$$\widetilde{\mathcal{Z}}(t)[a,b] := \int_a^b \frac{1}{x^2} W_x(t)\,dx + \frac{W_b(t)}{b} - \frac{W_a(t)}{a}, \quad 0 < a < b < \infty.$$

Along the way we also establish convergence of suitably scaled workload and queue length processes by proving in Theorem 2 that, as $r \to \infty$,

$$(c^r \hat{Q}^r(\cdot), \hat{W}^r(\cdot)) \text{ converges in distribution to } \left(\int_0^\infty \frac{1}{x^2} W_x(\cdot)\,dx, W_\infty(\cdot)\right)$$

in $\mathcal{D}([0,\infty) : \mathbb{R}_+^2)$, where $\hat{Q}^r$ and $\hat{W}^r$ are the queue length process and workload process, respectively, of the rth system with the standard diffusive scaling.

Results of [25] and Theorems 2 and 3 in the current paper suggest that the phenomenon of state space collapse is closely related to the tail behavior of the service time distributions. In Theorem 5 we make this heuristic precise by establishing that if the tail of the distribution of job processing times becomes lighter in an appropriate fashion, the difference between the limiting queue length process and the limiting workload process converges to zero, thereby approaching the behavior of state space collapse exhibited in [25] for light tailed processing time distributions. In Theorem 4, we prove another type of "asymptotic state space collapse" which roughly says that, asymptotically, the cumulative (scaled) workload due to jobs with remaining processing time more than ac^r (for large a) can be obtained by multiplying the number of such jobs present in the system with the expected value of a (full) processing time conditioned to be more than a.

The results of this work give information on response times of jobs with a given remaining processing time in SRPT queues under heavy traffic. Understanding the behavior of these response times is of interest as they quantify the "unfair" treatment of jobs with large processing times under the SRPT discipline [3, 30, 32, 33]. For Poisson arrivals, steady state mean response times have been studied by Bansal and Harchol–Balter [2] and Lin, Wierman and Zwart [20]. In [2], the steady state mean response and slowdown times are studied, with a focus on heavy tailed processing time distributions, as are characteristic of empirical workloads. In particular, [2] shows that the degree of unfairness as compared with processor sharing, a computer time sharing algorithm widely regarded as fair, is relatively small (see also Wierman and Harchol–Balter [36] for a broader discussion of fairness). Related to this, results of [9, 10] show that fluid analogs of response times in SRPT queues are sublinear for very heavy tailed processing distribution, which is a performance improvement over processor sharing. In the related work [20], expressions obtained in Perera [24] and Schassberger [27] are used to establish growth rates for the steady state mean response times as the traffic intensity increases to one (critical loading or heavy traffic). The rates that they obtain depend on the tail behavior of the processing time distribution. For instance, they grow exponentially for exponential processing times and polynomially for heavy tailed processing times. In view of the above results on dependence of key performance metrics for SRPT queues on the tail properties of processing time distributions it is of significant interest to understand the precise relationships between these tail properties and scaling limits of SRPT queues in heavy traffic. The current work contributes toward this goal.

1.1. *Methodology.* We now make some comments on the proof of one of our key results, namely Theorem 2. Central to our analysis are certain truncated workload processes $\{W_a^r(t)\}_{t\geq 0}$, $a \in [0,\infty]$, where $W_a^r(t)$ gives the amount of work (normalized by r) associated with jobs with remaining processing time at most ac^r at time r^2t in the rth system. We show in Theorem 1 that the joint distribution of $W_{a_1}^r, \ldots, W_{a_k}^r$ for finitely many threshold levels $0 \leq a_1 < \cdots < a_k \leq \infty$ converges to the joint distribution of $W_{a_1}, \ldots, W_{a_k}$ where $\{W_a(t)\}_{t\geq 0}$, $a \in [0,\infty]$, is a random field driven by a *single* Brownian motion. This novel synchronization phenomenon is a key ingredient in our proofs. It turns out that the convergence of the full measure valued state descriptor $\tilde{\mathcal{Z}}^r$ can be analyzed through the asymptotic properties of these truncated workload processes. This can be heuristically seen from an elementary integration by parts lemma (Lemma 13) that expresses the integral of any C^1 function, supported on a compact interval of $(0,\infty)$, with respect to the random measure $\tilde{\mathcal{Z}}^r(\cdot)$ in terms of the rescaled, truncated workload processes. This lemma is independent of the scheduling policy and is potentially useful for analyzing other types of policies for which one has good control over the associated truncated workload processes. Using this lemma together with Theorem 1 (which characterizes the limits of these truncated workload processes), along with appropriate tightness arguments, we then establish weak convergence of

$Z^r_f(\cdot) := \langle f, \widetilde{\mathcal{Z}}^r(\cdot)\rangle$ for piecewise C^1 functions f supported on a compact interval of $(0,\infty)$ (Theorem 14). The result is then extended to f having support which is bounded below by a positive number δ but possibly unbounded above (Lemma 15). The rest of the work is in sending $\delta \to 0$. This work, which is done in Section 5.4.1, is technically the most demanding part of the proof as is suggested by the possible singular behavior of the integrand in (3.6) near $x = 0$. The arguments are based on path decompositions of rescaled, truncated workload processes and their limiting versions into excursions and careful analysis of these excursions using martingale arguments; see additional comments at the beginning of Section 5.4.1. This is done in Lemmas 16–21, which finally lead to the proof of Theorem 2. As ingredients in the proofs, we also devise some couplings on SRPT systems started from different initial conditions (e.g., the "intertwined SRPT queueing systems" analyzed in Section 5.1), which may be of independent interest.

While the idea for the scaling involving c^r is inspired by the prior work [25], which considers lighter tailed processing time distributions, the proofs here are not variants or extensions of those in [25]. Indeed, in [25], the remaining processing times are shown to asymptotically concentrate around the spatial boosting factors c^r as r tends to infinity. This is not the case for heavier tailed processing time distribution. Instead, the remaining processing times spread out in a wider window containing c^r and the concentration arguments in [25] no longer hold. To address this, we take a different approach by rescaling the measure valued state descriptor such that mass that would otherwise shift toward infinity in a rather spread out fashion around c^r is brought back into a relevant window that spreads out around one. The asymptotic analysis of this rescaled measure-valued process requires an entirely different machinery and approach from the one used in [25] as was outlined in the previous paragraph.

We believe our techniques can be extended to SRPT systems with processing time distributions that *depend on* r, provided these distributions (indexed by r) satisfy certain uniformity conditions required by our techniques. More general r-dependence will require significant extensions of our methods and is left for future work.

1.2. *Organization.* The rest of the article is organized as follows. In Section 2, we rigorously define the sequence of SRPT systems, the heavy traffic conditions, the associated scaling and assumptions on the initial conditions. In Section 3, we state our main results. Section 4 summarizes some properties of Skorohod maps, regularly varying functions and the functional central limit theorems and tightness criteria used crucially in the proofs. Section 5 is dedicated to the proofs of our main results.

1.3. *Notation.* The following notation will be used. Let $\mathbb{N}$ denote the set of positive integers, $\mathbb{Z}$ denote the set of integers, $\mathbb{Z}_+$ denote the set of nonnegative integers, $\mathbb{R}$ the set of real numbers and $\mathbb{R}_+$ the set of nonnegative real numbers. For $a, b \in \mathbb{R}$, $a \wedge b$ and $a \vee b$ respectively denote the minimum and maximum of the set $\{a, b\}$. For a Polish space S and $T \in (0,\infty)$, we denote by $\mathcal{D}([0,T] : S)$ (resp. $\mathcal{D}([0,\infty) : S)$) the space of functions that are right continuous and have finite left limits (RCLL) from $[0,T]$ (resp. $[0,\infty)$) to S, equipped with the usual Skorohod topology. Also, denote by $\mathcal{C}([0,T] : S)$ (resp. $\mathcal{C}([0,\infty) : S)$) the space of continuous functions from $[0,T]$ (resp. $[0,\infty)$) to S, equipped with uniform (resp. local uniform) topology. Denote by $\mathcal{M}_F$ the space of finite nonnegative Borel measures on $\mathbb{R}_+$ equipped with the topology of weak convergence. For $\mu \in \mathcal{M}_F$ and a Borel measurable function f that is integrable with respect to μ or nonnegative, we write $\langle f, \mu\rangle = \int f\, d\mu$, which takes the value infinity if f is nonnegative and nonintegrable. Note that for $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_F$ and $\mu \in \mathcal{M}_F$, as $n \to \infty$, $\mu_n \to \mu$ in $\mathcal{M}_F$ if and only if $\langle f, \mu_n\rangle \to \langle f, \mu\rangle$ for every real valued, bounded, continuous function f on $\mathbb{R}_+$. The topology of weak convergence can be metrized so that $\mathcal{M}_F$ and hence $\mathcal{D}([0,T] : \mathcal{M}_F)$ are Polish

spaces. For a Borel subset $A \subseteq \mathbb{R}_+$, $\mathbf{1}_A$ denotes the indicator of set A; that is, $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ if $x \notin A$. In addition, $\mathbf{1}$ is used as a shorthand notation for $\mathbf{1}_{\mathbb{R}_+}$. For $x \in \mathbb{R}_+$, δ_x is the Dirac measure at x that puts a unit atom at x and $\delta_x^+ := \delta_x \mathbf{1}_{\{x>0\}}$ is the measure in $\mathcal{M}_F$ that equals δ_x if $x > 0$ and is the zero measure otherwise. For a real valued, bounded function f on S, we define $\|f\|_\infty := \sup_{x\in S} |f(x)|$. For $a \in \mathbb{R}_+$, a real valued function f is said to be C^1 on $[a, \infty)$ if it is defined on an open neighborhood of $[a, \infty)$ in $\mathbb{R}_+$ and is continuously differentiable on this neighborhood. For S valued random variables X_n, $n \in \mathbb{N}$ and X, we denote by $X_n \xrightarrow{d} X$ (resp. $X_n \xrightarrow{P} X$) the convergence in distribution (resp. probability) of X_n to X as $n \to \infty$. For $f \in \mathcal{D}([0, \infty) : \mathbb{R}^d)$, $0 \le s \le t \le \infty$ and $A > 0$, we will write $|f(t\#) - f(s\#)| < A$ to denote that all of the following inequalities hold: $|f(t) - f(s)| < A$, $|f(t) - f(s-)| < A$, $|f(t-) - f(s)| < A$, $|f(t-) - f(s-)| < A$.

2. Mathematical framework.

2.1. *The sequence of SRPT queues and state descriptor.* We consider a sequence of SRPT queues indexed by $\mathcal{R}$, a sequence taking values in $(1, \infty)$ tending to infinity. For each $r \in \mathcal{R}$, let $\{\check{v}_l^r, l \in \mathbb{N}\}$ be a sequence of strictly positive random variables and let $\mathbf{q}^r$ be a nonnegative integer valued random variable such that $\sum_{l=1}^{\mathbf{q}^r} \check{v}_l^r < \infty$ almost surely (with the convention that this sum is zero if $\mathbf{q}^r$ is zero). At time zero, there are $\mathbf{q}^r$ jobs in the rth system with remaining processing times $\check{v}_l^r$, $l = 1, \ldots, \mathbf{q}^r$. For $l = 1, \ldots, \mathbf{q}^r$, we refer to the job in system at time zero associated with $\check{v}_l^r$ as initial job l. Conditions on $\mathbf{q}^r$ and $\{\check{v}_l^r\}$ will be specified in Section 2.4.

Jobs arrive to the rth system according to a delayed renewal process $E^r(\cdot)$ with positive, finite rate λ^r and finite, positive initial delay. Let T^r (resp. T_1^r) denote a random variable having the distribution of a typical inter-arrival time (resp. the initial delay) in the rth system. We assume that T^r is positive and has finite standard deviation σ_A^r. We also assume that $\mathbb{E}[(T_1^r)^2] < \infty$. For $j \in \mathbb{N}$, we refer to the jth job to arrive after time zero as job j.

Upon its arrival to the rth system, each job is assigned a processing time, which is the amount of time it takes the server to process the work associated with that job. The processing times are taken to be strictly positive and independent and identically distributed. Also, the processing time distribution does not depend on r, that is, is the same for all r, and is given by a continuous distribution function F on $\mathbb{R}_+$ such that $F(0) = 0$. It is assumed that $\overline{F}(x) = 1 - F(x)$ is positive for each $x \in \mathbb{R}_+$ and that $\overline{F}$ is a regularly varying function with index $-(p + 1)$ for some $p > 1$; namely, for all $t > 0$,

$$\bar{F}(t) > 0 \quad \text{and} \quad \lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} := t^{-(p+1)}. \tag{2.1}$$

The above condition in particular implies that the processing time distribution has a finite, positive second moment. The Pareto type 1 distribution with parameters $m > 0$ and $p > 1$ (i.e., $\overline{F}(x) = \min(m^{p+1}x^{-p-1}, 1)$ for $x \in \mathbb{R}_+$) is a basic example of a processing time distribution that satisfies (2.1).

For each $r \in \mathcal{R}$, $\{\mathbf{q}^r, \check{v}_l^r, l \in \mathbb{N}\}$, $E^r(\cdot)$, and the sequence of processing times are assumed to be mutually independent of one another.

Jobs in the rth system are served in accordance with the SRPT service discipline; that is, at each time the server preemptively serves the job in system with the shortest remaining processing time. For $t \ge 0$, $l = 1, \ldots, \mathbf{q}^r$ and $j = 1, \ldots, E^r(t)$, $\check{v}_l^r(t)$ and $v_j^r(t)$ denote the remaining processing time at time t of initial job l and job j respectively. For each $r \in \mathcal{R}$ and $t \ge 0$, define

$$\mathcal{Z}^r(t) = \sum_{l=1}^{\mathbf{q}^r} \delta^+_{\check{v}_l^r(t)} + \sum_{j=1}^{E^r(t)} \delta^+_{v_j^r(t)}.$$

Then, for each $r \in \mathcal{R}$ and $t \geq 0$, $\mathcal{Z}^r(t) \in \mathcal{M}_F$ has a unit atom at the remaining processing time of each job in system. Furthermore, for each $r \in \mathcal{R}$, $\mathcal{Z}^r(\cdot)$ is a stochastic process with sample paths in $\mathcal{D}([0,\infty) : \mathcal{M}_F)$. We will find it convenient to adopt the abbreviated phrases job *size* and job *sizes* to refer to a given job's remaining processing time and the collection of all remaining processing times, respectively, at a given time. Also, a job's *initial size* refers to its processing time upon arrival with initial job $l = 1, \ldots, \mathbf{q}^r$ having initial size $\check{v}_l^r$ by convention.

2.2. *Heavy traffic conditions*. Let v denote a random variable having the distribution of the processing time of an incoming job. For each $r \in \mathcal{R}$, write

$$\rho^r := \lambda^r \mathbb{E}(v) \quad \text{and} \quad \rho_x^r := \lambda^r \mathbb{E}(v\mathbf{1}_{[v \leq x]}) \quad \text{for all } x \in \mathbb{R}_+.$$

It is assumed that there exists $\kappa \in \mathbb{R}$ and $\sigma_A, \lambda \in (0,\infty)$ such that as $r \to \infty$,

$$r\big(\rho^r - 1\big) \to \kappa, \qquad \lambda^r \to \lambda \quad \text{and} \quad \sigma_A^r \to \sigma_A. \tag{2.2}$$

Note that the first limit above implies $\lambda = 1/\mathbb{E}(v)$. Henceforth, $\kappa \in \mathbb{R}$ and $\sigma_A, \lambda \in (0,\infty)$ satisfying (2.2) are fixed. It is also assumed that

$$\limsup_{r\to\infty} \mathbb{E}\big(T_1^r\big) \leq \lambda^{-1} \quad \text{and} \quad \limsup_{r\to\infty} \mathrm{Var}\big(T_1^r\big) \vee \mathbb{E}\big[\big(T_1^r - \big(\lambda^r\big)^{-1}\big)^2\big] \leq \sigma_A^2. \tag{2.3}$$

We note here that, for our results to hold, we only need finiteness of the above lim sups. However, the above assumptions are made to treat the first inter-arrival time in a similar fashion as the later ones and thus to make the analysis less notationally cumbersome. For $r \in \mathcal{R}$ and $t \geq 0$, define

$$\overline{E}^r(t) := \frac{E^r(r^2 t)}{r^2} \quad \text{and} \quad \widehat{E}^r(t) := \frac{E^r(r^2 t) - \lambda^r r^2 t}{r} = r\big(\overline{E}^r(t) - \lambda^r t\big).$$

Assume that as $r \to \infty$,

$$\widehat{E}^r(\cdot) \xrightarrow{d} E^*(\cdot) \tag{2.4}$$

in $\mathcal{D}([0,\infty) : \mathbb{R})$, where $E^*(\cdot)$ is a one-dimensional Brownian motion starting from zero with zero drift and variance $\lambda^3 \sigma_A^2$. This also implies that as, $r \to \infty$,

$$\overline{E}^r(\cdot) \xrightarrow{d} \lambda(\cdot) \quad \text{where } \lambda(t) := \lambda t \text{ for all } t \geq 0. \tag{2.5}$$

2.3. *Scaling*. For $x \in \mathbb{R}_+$, let

$$S(x) = \frac{1}{\mathbb{E}(v\mathbf{1}_{[v>x]})}. \tag{2.6}$$

The function $S(\cdot)$ plays an important role in our analysis. As shown in [9, 19], it has the same order of magnitude as the response time of jobs with remaining processing time x in the system at time zero, in the fluid limit. Here, due to the assumptions on $F(\cdot)$, $S(\cdot)$ is a positive, nondecreasing, continuous function such that $\lim_{x\to\infty} S(x) = \infty$. In particular, the right continuous inverse $S^{-1}(\cdot)$ exists and is well defined on all of $\mathbb{R}_+$. Then, for $y \in \mathbb{R}_+$, we have

$$S^{-1}(y) := \inf\{u > 0 : S(u) > y\}, \tag{2.7}$$

and the function $y \mapsto S^{-1}(y)$ is a nonnegative, nondecreasing, right continuous function which is strictly increasing for $y \in [S(0),\infty)$. Also, for all $y \in [S(0),\infty)$,

$$S\big(S^{-1}(y)\big) = y. \tag{2.8}$$

In [9], a version of (2.7) arises as the left edge of the support of the measure valued fluid model solutions studied there. For each $r \in \mathcal{R}$, let

$$c^r := S^{-1}(r). \tag{2.9}$$

Note that $c^r = 0$ if $r \le S(0)$ and $c^r > 0$ if $r > S(0)$. As we are interested in large values of r, from now on, we will assume without loss of generality that the elements of $\mathcal{R}$ are all larger than $S(0)$. Then, (2.8) and (2.9) imply that for all $r \in \mathcal{R}$,

$$S(c^r) = r. \tag{2.10}$$

As noted in the Introduction, the quantity c^r, which was introduced in [25], identifies the correct scaling needed in order to obtain a nontrivial limit for the queue length process in the light tailed case studied there (see (1.1)). We will see that this quantity is key for the analysis of regularly varying tails as well. For each $r \in \mathcal{R}$ and $t \ge 0$, define

$$\widetilde{\mathcal{Z}}^r(t) = \frac{c^r}{r} \sum_{l=1}^{\mathbf{q}^r} \delta^+_{\check{v}^r_l(r^2 t)/c^r} + \frac{c^r}{r} \sum_{i=1}^{E^r(r^2 t)} \delta^+_{v^r_i(r^2 t)/c^r}. \tag{2.11}$$

Thus $\widetilde{\mathcal{Z}}^r(\cdot)$ is obtained from $\mathcal{Z}^r(\cdot)$ by adding three types of scaling: the time is scaled by r^2, the mass is scaled by c^r/r and the space (representing the job sizes) is scaled by $1/c^r$.

To illustrate this scaling, we consider the Pareto type 1 distribution with parameters $m > 0$ and $p > 1$ (i.e., $\overline{F}(x) = \min(m^{p+1}x^{-p-1}, 1)$ for $x \in \mathbb{R}_+$). Then for $c_p := m^{1+p}(1+p)/p$ and for each $r \in \mathcal{R}$ such that $c^r \ge m$, we find that $c^r = (c_p r)^{1/p}$ and $c^r/r = \frac{c_p^{1/p}}{r^{(p-1)/p}}$, which respectively tend to the constants $2m^2 r$ and $2m^2$ as $p \searrow 1$ and m and m/r as $p \to \infty$. The latter is traditional diffusion scaling. Upon noting that the ratio of two regularly varying functions with the same index is slowly varying, we see that for F satisfying (2.1) for some $p > 1$, c^r takes the form $L_p(r)\sqrt[p]{r}$, $r \in \mathcal{R}$, for some distribution dependent, slowly varying function L_p. See Section 4.2 for a brief summary of the relevant properties of regularly and slowly varying functions.

For each $r \in \mathcal{R}$, $t \ge 0$, and $f : \mathbb{R}_+ \to \mathbb{R}$, define

$$Z^r_f(t) := \langle f, \widetilde{\mathcal{Z}}^r(t) \rangle.$$

We will also write, for $a \in [0, \infty] := [0, \infty) \cup \{\infty\}$ and $t \ge 0$,

$$Z^r_a(t) := Z^r_{\mathbf{1}_{[0,a]}}(t) = \int_0^a \widetilde{\mathcal{Z}}^r(t)(dx). \tag{2.12}$$

For each $r \in \mathcal{R}$ and $t \ge 0$, we adopt the notation $Q^r(t) = Z^r_{\mathbf{1}}(t) = \int_0^\infty \widetilde{\mathcal{Z}}^r(t)(dx)$ so that $Q^r(t)$ represents c^r times the diffusion scaled queue length in the rth system at time instant t.

For all $x \in \mathbb{R}_+$, let $\chi(x) = x$ and $\chi_a(x) := \chi(x)\mathbf{1}_{[0,a]}(x)$ for any $a \in \mathbb{R}_+$. Also, by convention, $\chi_\infty = \chi$. For each $r \in \mathcal{R}$, $t \ge 0$ and $a \in [0, \infty]$, define

$$W^r_a(t) := Z^r_{\chi_a}(t) = \langle \chi_a, \widetilde{\mathcal{Z}}^r(t) \rangle. \tag{2.13}$$

For $r \in \mathcal{R}$, $a \in \mathbb{R}_+$ and $t \ge 0$, $W^r_a(t)$ is equal to the amount of work associated with jobs of size less or equal to ac^r at time t in the rth system under diffusion scaling. Further note that for each $r \in \mathcal{R}$, $W^r_\infty(\cdot)$ is the diffusion scaled workload process and $\lim_{a\to\infty} W^r_a(t) = W^r_\infty(t)$ for each $t \ge 0$, almost surely. Observe that for each $r \in \mathcal{R}$ and each fixed $a \in [0, \infty]$, $W^r_a(\cdot) \in \mathcal{D}([0, \infty) : \mathbb{R}_+)$. For each $r \in \mathcal{R}$, we refer to the collection $\{W^r_a(\cdot), a \in \mathbb{R}_+\}$ as the *rescaled, truncated workload processes*, which is a random field on $\mathbb{R}^2_+$ taking values in $\mathbb{R}_+$. Also note that for $r \in \mathcal{R}$ and each fixed $t \ge 0$, $W^r_\cdot(t) \in \mathcal{D}([0, \infty) : \mathbb{R}_+)$.

2.4. *Asymptotic conditions for the sequence of initial conditions.* We assume that there exists an $\mathbb{R}_+$ valued, continuous, nondecreasing stochastic process $\{w^*(a) : a \in \mathbb{R}_+\}$, with $w^*(\infty) := \lim_{a\to\infty} w^*(a)$ satisfying $\mathbb{E}(w^*(\infty)) < \infty$, such that, as $r \to \infty$,

$$\big(W^r_\cdot(0), W^r_\infty(0)\big) \xrightarrow{d} \big(w^*(\cdot), w^*(\infty)\big) \tag{2.14}$$

in $\mathcal{D}([0,\infty) : \mathbb{R}_+) \times \mathbb{R}_+$, and

$$\big\{W^r_\infty(0); r \in \mathcal{R}\big\} \text{ is uniformly integrable.} \tag{2.15}$$

Note that (2.14) and (2.15) imply that, for any $a \in \mathbb{R}_+$,

$$\lim_{r\to\infty} \mathbb{E}\big(W^r_a(0)\big) = \mathbb{E}\big(w^*(a)\big) \quad \text{and} \quad \lim_{r\to\infty} \mathbb{E}\big(W^r_\infty(0)\big) = \mathbb{E}\big(w^*(\infty)\big).$$

We further assume that there exist some $\eta^* \in (0, p-1)$, $a^* > 0$ and $\alpha^* \in (0, p]$ such that

$$\limsup_{r\to\infty} \sup_{a\in[a^*(c^r)^{-1},1]} a^{-(p-\eta^*)}\mathbb{E}\big(W^r_a(0)\big) < \infty \tag{2.16}$$

and

$$\limsup_{a\to\infty} a^{\alpha^*}\mathbb{E}\big(w^*(\infty) - w^*(a)\big) < \infty. \tag{2.17}$$

Assumption (2.16) insures that the work associated with initial jobs with remaining processing times near zero vanishes at a suitable rate as r tends to infinity. Assumption (2.17) insures that the limiting work associated with initial jobs with large remaining processing times vanishes at a suitable rate. Assumptions (2.15) and (2.16) imply that

$$\sup_{a>0} a^{-(p-\eta^*)}\mathbb{E}\big(w^*(a)\big) < \infty. \tag{2.18}$$

Finally, we assume that for any $a \in \mathbb{R}_+$,

$$Z^r_{a/c^r}(0) = \frac{c^r}{r}\sum_{l=1}^{\mathbf{q}^r} \mathbf{1}_{[\check{v}^r_l \le a]} \xrightarrow{P} 0 \quad \text{as } r \to \infty. \tag{2.19}$$

REMARK 1. A guideline for whether Assumptions (2.14)–(2.19) are natural is to check whether a sequence of systems such that each system starts from zero jobs at time zero satisfies these assumptions at any fixed positive time t. It can be checked from the proofs in Section 5 that this is indeed the case; namely, if each system starts with zero jobs then at any time $t > 0$, Assumptions (2.14)–(2.19) are satisfied with $(W^r_\cdot(0), W^r_\infty(0))$ replaced by $(W^r_\cdot(t), W^r_\infty(t))$ for each $r \in \mathcal{R}$ and $\{\check{v}^r_l\}_{1\le l\le \mathbf{q}^r}$ replaced with $\{v_i(r^2t) : 1 \le i \le E^r(r^2t), v_i(r^2t) > 0, \} \cup \{\check{v}^r_l(r^2t), 1 \le l \le \mathbf{q}^r, \check{v}^r_l(r^2t) > 0\}$. See the Appendix for a sketch of how to verify this.

2.5. *Some initial conditions satisfying Assumptions* (2.14)–(2.19). We give the following two sets of initial conditions which are easily checkable and satisfy Assumptions (2.14)–(2.19).

(I) Suppose the following hold:

(i) for each $r \in \mathcal{R}$, $\{\check{v}^r_l : l \ge 1\}$ is a sequence of independent and identically distributed random variables that is independent of $\mathbf{q}^r$;

(ii) For some $\mathbf{q}^*$ with $\mathbb{E}(\mathbf{q}^*) < \infty$, $c^r\mathbf{q}^r/r \xrightarrow{L^1} \mathbf{q}^*$ as $r \to \infty$;

(iii) $\sup_{r\in\mathcal{R}} \mathbb{E}[(\check{v}^r_1/c^r)^2] < \infty$ and $\check{v}^r_1/c^r \xrightarrow{d} \check{v}^*$ as $r \to \infty$, where $\check{v}^*$ has a continuous distribution;

(iv) there is a random variable $\underline{v}$ that stochastically lower bounds $\check{v}_1^r/c^r$ for all $r \in \mathcal{R}$ and satisfies

$$\limsup_{a \downarrow 0} a^{-(p-1-\eta^*)} \mathbb{P}(\underline{v} \le a) < \infty,$$

for some $\eta^* \in (0, p-1)$.

Then Assumptions (2.14)–(2.19) are satisfied with $\alpha^* = 1$, η^* as in part (iv) above, any $a^* > 0$, $w^*(a) = \mathbf{q}^* \mathbb{E}(\check{v}^* \mathbf{1}_{[\check{v}^* \le a]})$ for $a \in \mathbb{R}_+$ and $w^*(\infty) = \mathbf{q}^* \mathbb{E}(\check{v}^*)$. See the Appendix for a sketch of how to check that assumptions (2.14)–(2.19) hold for such a sequence of initial conditions.

(II) Another set of conditions for which Assumptions (2.14)–(2.19) hold is that along with (i) in (I) above, for some $\alpha > 0$, $(c^r)^{1+\alpha} \mathbf{q}^r / r \to 0$ in L^1 and $\{\check{v}_1^r/c^r, r \in \mathcal{R}\}$ is L^1 bounded. In particular, it can be checked that under these conditions, Assumptions (2.14)–(2.19) hold for any $\alpha^* \in (0, p]$, any $a^* > 0$, $\eta^* = (p-1-\alpha) \vee (p-1)/2$ and $w^*(a) = 0$ for $a \in (0, \infty]$. Note that these conditions are trivially satisfied if each system starts from empty, namely $\mathbf{q}^r = 0$ for all $r \in \mathcal{R}$.

3. Main results. In this section, we state the five main results in this paper. The conditions introduced in Sections 2.1, 2.2 and 2.4 will be assumed to hold throughout this work and will not be noted explicitly in statements of various results. Thus henceforth, we consider a sequence (or sequences) of SRPT queues indexed by $\mathcal{R}$ satisfying the above conditions.

3.1. *A random field governing the limiting behavior.* The first theorem (stated below) gives the important observation that for processing time distributions with regularly varying tails, the joint limiting behavior of the truncated workload processes is captured by a random field constructed from a single Brownian motion using the Skorohod map. For $f \in \mathcal{D}([0, \infty) : \mathbb{R})$ with $f(0) \ge 0$, let

$$\Gamma[f](t) := f(t) - \inf_{0 \le s \le t} \big(f(s) \wedge 0\big), \quad t \ge 0. \tag{3.1}$$

The function Γ is known as the one-dimensional Skorohod map.

THEOREM 1. *Let B be a standard real Brownian motion and $(\xi(\cdot), \xi(\infty))$ be a $\mathcal{C}([0, \infty) : \mathbb{R}_+) \times \mathbb{R}_+$ valued random variable with same distribution as $(w^*(\cdot), w^*(\infty))$ that is independent of B. For any $k \in \mathbb{N}$ and any $0 \le a_1 < \cdots < a_k \le \infty$, as $r \to \infty$,*

$$\big(W_{a_1}^r(\cdot), \ldots, W_{a_k}^r(\cdot)\big) \xrightarrow{d} \big(W_{a_1}(\cdot), \ldots, W_{a_k}(\cdot)\big)$$

in $\mathcal{D}([0, \infty) : \mathbb{R}_+^k)$, where for $a \in [0, \infty]$,

$$W_a(\cdot) := \Gamma[X_a](\cdot), \tag{3.2}$$

with Γ as in (3.1) and $\{X_a(\cdot) : a \in [0, \infty]\}$, given as follows: for $t \ge 0$,

$$X_0(t) := \xi(0) = 0, \tag{3.3}$$

$$X_a(t) := \xi(a) + \sigma B(t) + \left(\kappa - \frac{\lambda}{a^p}\right) t \quad \text{for } 0 < a < \infty, \tag{3.4}$$

$$X_\infty(t) := \xi(\infty) + \sigma B(t) + \kappa t, \tag{3.5}$$

and $\sigma^2 := \lambda \operatorname{Var}(v) + \lambda \sigma_A^2$, where v is as in Section 2.2 and σ_A is as in (2.2).

Due to (3.2)–(3.5), $\xi(a) = X_a(0) = W_a(0)$ for all $a \in [0, \infty]$. The key feature of the above result is that the Brownian motion $B(\cdot)$ that determines $X_a(\cdot)$ is the same for all $a \in [0, \infty]$. In particular, a only enters in the initial condition and the drift term. In addition, $W_\infty(\cdot)$ is the diffusion limit of the workload process as given in [14]. Theorem 1 is proved in Section 5 as a consequence of Proposition 10 and Lemma 12, stated there. In Proposition 10, upper and lower bounds on $W_a^r(t)$ and $Z_a^r(t)$ for each $a \in [0, \infty]$ and $t \geq 0$ are given by coupling it with the workload process and queue length process for a SRPT queueing system that satisfies all of the assumptions in Section 2.1, except that the renewal arrival process is thinned to only include jobs with processing time at most ac^r. A notion of ordering of two SRPT systems, which we call intertwining, is introduced in Section 5.1 and used in a crucial way to obtain the queue length bounds in Proposition 10. In Lemma 12, a functional central limit theorem (FCLT) is established for a finite collection of rescaled, truncated workload processes via the bounds obtained in Proposition 10 and establishing an FCLT for the bounding processes. Continuity properties of the Skorohod map imply Theorem 1 as a direct consequence of Lemma 12.

3.2. *Limits for the queue length process and measure valued state descriptor.* Theorem 1 can be used in describing the limiting behavior of $Z_f^r(\cdot) := \langle f, \widetilde{\mathcal{Z}}^r(\cdot)\rangle$ for a rich class of functions f as stated in the next theorem. This, in turn, gives distributional asymptotics for the scaled queue length process. Recall that $\chi(x) = x$ and $\mathbf{1}(x) = 1$ for $x \in \mathbb{R}_+$.

THEOREM 2. *Let $f : [0, \infty) \to \mathbb{R}$ be any C^1 function such that $\lim_{x\to\infty} \frac{f(x)}{x}$ exists and $\int_1^\infty \frac{|f'(x)|}{x^{\alpha^*+1}}\, dx < \infty$, where α^* is the constant appearing in Assumption (2.17). Then, as $r \to \infty$,*

$$Z_f^r(\cdot) \xrightarrow{d} Z_f(\cdot)$$

in $\mathcal{D}([0, \infty) : \mathbb{R})$, where Z_f is a real stochastic process with continuous sample paths, given by the formula

$$Z_f(t) := \int_0^\infty \left(\frac{f(x)}{x^2} - \frac{f'(x)}{x}\right) W_x(t)\, dx + \left(\lim_{x\to\infty} \frac{f(x)}{x}\right) W_\infty(t), \quad t \geq 0.$$

In particular, as $r \to \infty$,

$$W_\infty^r(\cdot) = Z_\chi^r(\cdot) \xrightarrow{d} Z_\chi(\cdot) = W_\infty(\cdot) \quad \text{and} \quad Q^r(\cdot) = Z_{\mathbf{1}}^r(\cdot) \xrightarrow{d} Z_{\mathbf{1}}(\cdot)$$

in $\mathcal{D}([0, \infty) : \mathbb{R}_+)$, where $Q(\cdot) := Z_{\mathbf{1}}(\cdot)$ satisfies

$$Q(t) = \int_0^\infty \frac{1}{x^2} W_x(t)\, dx, \quad t \geq 0. \tag{3.6}$$

Theorem 2 is proved in Section 5.4. An overview of this proof is given in Section 1.1.

The result in Theorem 2 can be strengthend to show that $\widetilde{\mathcal{Z}}^r$ converges in distribution to a measure valued process $\widetilde{\mathcal{Z}}$ in $\mathcal{D}([0, \infty) : \mathcal{M}_F)$. This is stated in the next theorem, which is proved in Section 5.5. The proof proceeds via integrating the random measure $\widetilde{\mathcal{Z}}^r$ against a class of test functions and analyzing weak convergence of the collection of processes thus obtained.

THEOREM 3. *As $r \to \infty$,*

$$\widetilde{\mathcal{Z}}^r(\cdot) \xrightarrow{d} \widetilde{\mathcal{Z}}(\cdot)$$

in $\mathcal{D}([0,\infty):\mathcal{M}_F)$, *where for each* $t \geq 0$, *the measure* $\widetilde{\mathcal{Z}}(t)$ *can be characterized as* $\widetilde{\mathcal{Z}}(t)(\{0\}) = 0$, $\widetilde{\mathcal{Z}}(t)(\mathbb{R}_+) = Q(t)$ *and*

$$\widetilde{\mathcal{Z}}(t)[a,b] := \int_a^b \frac{1}{x^2} W_x(t)\,dx + \frac{W_b(t)}{b} - \frac{W_a(t)}{a}, \quad 0 < a < b < \infty.$$

REMARK 2. The integral expression (3.6) in Theorem 2 is quite different from the main result (Theorem 3.1) in [25], which gives conditions under which light tailed processing times result in a limit theorem that states $Q(\cdot) = W_\infty(\cdot)$ (state space collapse). While the proofs given here do not cover the light tailed case, the concentration arguments given in [25] could be used to argue that the measure valued state descriptors in the light tailed case, scaled as in (2.11) above, would converge to a point mass at one with (random) total mass given by the limiting workload process $W_\infty(\cdot)$. Consequently, the rescaled, truncated workload processes $W_x^r(\cdot)$ defined in (2.13) above, in the light tailed case, would converge to $W_\infty(\cdot)$ for $x > 1$, $W_1(\cdot)$ for $x = 1$ and the process that is identically zero otherwise, and the integral given in (3.6) would be $W_\infty(t)$ for each $t \geq 0$, as it should from the results of [25]. The results in Theorem 2 and Theorem 3 demonstrate that, in contrast to the light tailed processing time distributions considered in [25], heavy tailed processing time distributions do not exhibit state space collapse and the mass of the limiting scaled measure valued state descriptor is distributed as a time-varying random profile over $\mathbb{R}_+$, as opposed to a time-varying randomly sized point mass at one.

3.3. *Tail behavior of* $\widetilde{\mathcal{Z}}$. The next result describes the asymptotic behavior of the limiting queue length and limiting workload processes defined in terms of the measure $\widetilde{\mathcal{Z}}$ when attention is restricted to the dynamics of jobs with large remaining processing times. Let

$$W'_\infty(t) := t - \sup\{s \leq t : W_\infty(s) = 0\}, \quad t \geq 0, \tag{3.7}$$

which can be recognized as the duration of the current busy period when $W_\infty(t)$ is interpreted as the work in the system at time instant t. We will see in Section 4.1 that $W'_\infty(\cdot)$ arises as the "pathwise derivative" of the Skorohod map with respect to the "drift parameter" of the process on which the map acts, which explains the notation $W'_\infty(\cdot)$. We will also assume a stronger version of (2.17) for this result, namely

$$\lim_{x\to\infty} x^p\big(\xi(\infty) - \xi(x)\big) = 0 \quad \text{almost surely.} \tag{3.8}$$

In particular, (3.8) holds when $\xi(\infty) = 0$.

THEOREM 4. *Assume* (3.8) *holds. For every* $t \geq 0$, *as* $a \to \infty$,

$$\frac{a^p}{\lambda}\langle \chi \mathbf{1}_{[a,\infty)}, \widetilde{\mathcal{Z}}(t)\rangle \to W'_\infty(t) \quad \textit{almost surely},$$

$$\frac{(p+1)a^{p+1}}{p\lambda}\widetilde{\mathcal{Z}}(t)[a,\infty) \to W'_\infty(t) \quad \textit{almost surely}.$$

In particular, for any $t \geq 0$ *such that* $W'_\infty(t) \neq 0$, *as* $a \to \infty$,

$$\frac{\langle \chi \mathbf{1}_{[a,\infty)}, \widetilde{\mathcal{Z}}(t)\rangle}{\mathbb{E}(v \mid v \geq a)\widetilde{\mathcal{Z}}(t)[a,\infty)} \to 1 \quad \textit{almost surely}. \tag{3.9}$$

Theorem 4 is proved in Section 5.5 and proceeds via connecting the tail mass processes $\{\widetilde{\mathcal{Z}}(t)[a,\infty) : t \geq 0\}$ for large a with the process $\{W'_\infty(t) : t \geq 0\}$.

REMARK 3. The above result says that if, in the diffusion limit, we restrict attention to jobs in system of size more than a (for large a), the cumulative workload due to these jobs can be approximated by multiplying the number of such jobs present in the system with the expected size of an incoming job conditional on it being more than size a. In other words in the diffusion limit, so few large jobs have entered service by a finite time t that the work associated with such jobs satisfies (3.9). This result can be heuristically understood from the SRPT dynamics under which small jobs are given priority and large jobs remain unprocessed at typical time points when the system has small jobs present. Theorem 4 can be seen as a form of asymptotic state space collapse when one restricts attention to jobs with large remaining processing times.

3.4. *Asymptotic state space collapse as $p \to \infty$.* As stated in Remark 2, the limiting scaled queue length process given in Theorem 2 differs qualitatively from its light tailed analogue treated in [25] in that, although the limiting scaled queue length and limiting scaled workload processes are driven by the same Brownian motion B, there is no state space collapse as in [25]. However, as $p \to \infty$ (i.e., the tail of the processing time distribution becomes lighter), we obtain a limiting state space collapse as described in Theorem 5 below. As we are interested in large values of p here, we will only consider $p \geq 2$. To make the dependence on p explicit, we consider a family of distributions $\{F^{(p)}(\cdot) : p \geq 2\}$ such that for each $p \geq 2$, $\overline{F}^{(p)}(\cdot) := 1 - F^{(p)}(\cdot)$ is a regularly varying function; that is, (2.1) is satisfied by $\overline{F}^{(p)}(\cdot)$. For each $p \geq 2$, consider a sequence of SRPT queues indexed by $\mathcal{R}$ such that the initial conditions $\{\mathbf{q}^{(p),r}, \check{v}_l^{(p),r}, l \in \mathbb{N}, r \in \mathcal{R}\}$ satisfy the assumptions of Section 2.4 and the arrival processes $\{E^r(\cdot), r \in \mathcal{R}\}$ do not depend on p. Consequently, $\lambda^{(p)} = 1/\mathbb{E}(v^{(p)})$, where $v^{(p)}$ is distributed as $F^{(p)}(\cdot)$, does not depend on p and we will write this quantity as λ. The processing times of jobs for each $p \geq 2$ are distributed as $F^{(p)}(\cdot)$. For each $p \geq 2$, write $\sigma(p) = \sqrt{\lambda \operatorname{Var}(v^{(p)}) + \lambda \sigma_A^2}$. For each $p \geq 2$, we let $\xi^{(p)}(\infty)$ denote the limiting initial workload (i.e., the quantity analogous to $\xi(\infty)$ in Theorem 1 for the pth system) and let $\xi^{(p)}(\cdot)$ denote the limiting initial truncated workload process (analogous to $\xi(\cdot)$ in Theorem 1 for the pth system). We will also elucidate the dependence of η^* in Assumption (2.16) by writing it as $\eta^*(p)$. We assume that

$$\sup_{p \geq 2} \mathbb{E}\big[\big(v^{(p)}\big)^2\big] < \infty, \qquad \sup_{p \geq 2} \mathbb{E}\big[\xi^{(p)}(\infty)\big] < \infty \quad \text{and} \quad \sup_{p \geq 2} C_0(p) < \infty, \tag{3.10}$$

where $C_0(p) := 2 \sup_{a>0} a^{-(p-\eta^*(p))} \mathbb{E}(\xi^{(p)}(a))$ for each $p \geq 2$. Writing $Q^{(p)}(\cdot)$ for $Q(\cdot)$ and $W_\infty^{(p)}(\cdot)$ for $W_\infty(\cdot)$ for each $p \geq 2$ to denote the limiting queue length and limiting workload processes respectively, we have that, for $p \geq 2$,

$$Q^{(p)}(t) = \int_0^\infty \frac{1}{x^2} \Gamma\big[\xi^{(p)}(x) + \sigma(p) B(\cdot) + \big(\kappa - \lambda x^{-p}\big)\iota(\cdot)\big](t)\, dx, \quad t \geq 0,$$

where ι denotes the identity map on $[0, \infty)$ and $W_\infty^{(p)}(t) = \Gamma[X_\infty^{(p)}](t)$ for $X_\infty^{(p)}(t) = \xi^{(p)}(\infty) + \sigma(p) B(t) + \kappa t$, $t \geq 0$. For the state space collapse result, we will require that $\eta^*(\cdot)$ satisfies

$$\liminf_{p \to \infty} \frac{p - 1 - \eta^*(p)}{\log p} = \infty. \tag{3.11}$$

Moreover, we will require for any $a \in (1, \infty)$,

$$\lim_{p \to \infty} \mathbb{E}\big(\xi^{(p)}(\infty) - \xi^{(p)}(a)\big) = 0. \tag{3.12}$$

Then, (3.11) implies that for large p, $\mathbb{E}(\xi^{(p)}(a))$ decreases to zero sufficiently fast with a tending to zero. Also, (3.12) implies that the main contribution to the limiting initial workload process $\xi^{(p)}(\cdot)$ for large values of p comes from initial jobs with size in $(0, 1]$. Note that if the system starts from empty, namely $\mathbf{q}^{(p),r} = 0$ for all $r \in \mathcal{R}$ and $p \geq 2$, then for any $t \geq 0$ and any $a \in (1, \infty)$, by the Lipschitz property of the Skorohod map given in (4.1) below,

$$\lim_{p\to\infty} \mathbb{E}\big(W_\infty^{(p)}(t) - W_a^{(p)}(t)\big) \leq 2\lambda t a^{-p} \to 0 \quad \text{as } p \to \infty.$$

Hence, by the discussion in Remark 1, Assumption (3.12) is indeed a natural assumption on $\xi^{(p)}(\cdot)$.

THEOREM 5. *Assume that* (3.10) *and* (3.12) *hold and we can choose* $p \mapsto \eta^*(p)$ *such that* $\eta^*(\cdot)$ *satisfies* (3.11). *Then, for any* $T > 0$,

$$\sup_{t\in[0,T]} \big|Q^{(p)}(t) - W_\infty^{(p)}(t)\big| \xrightarrow{P} 0 \quad \text{as } p \to \infty.$$

Theorem 5 is proved in Section 5.5. The proof essentially proceeds by showing that as $p \to \infty$, the time varying mass profile of the limiting measure valued state descriptor collapses onto a point mass at one. For the Pareto Type I example with $\overline{F}^{(p)}(x) = \min((\frac{\lambda x(p+1)}{p})^{-p-1}, 1)$ for $x \in \mathbb{R}_+$ and $p \geq 2$, we have $\lambda^{(p)} = \lambda$ and $\mathrm{Var}(v^{(p)}) = \lambda^{-2}/(p^2 - 1)$ for all $p \geq 2$, and the latter tends to zero as $p \to \infty$. In fact, the measure corresponding to the complementary cumulative distribution function (CCDF) $\overline{F}^{(p)}(\cdot)$ converges weakly to the point mass at λ^{-1}, that is, the service times are asymptotically deterministic, which makes the state space collapse rather intuitive. A somewhat more interesting example based on the Lomax distribution that does not have asymptotically deterministic service times has CCDFs $\overline{G}^{(p)}(x) = (1 + \frac{\lambda x}{p})^{-p-1}$, $x \in \mathbb{R}_+$, with $\lambda^{(p)} = \lambda$ and $\mathrm{Var}(v^{(p)}) = \lambda^{-2}(p + 1)/(p - 1)$ for all $p \geq 2$. This gives rise to an exponential rate λ distribution in the $p \to \infty$ limit.

REMARK 4. Consider the initial condition of the form discussed in (II) of Section 2.5, namely, along with (i) in (I) of Section 2.5, suppose for some $\alpha > 0$, $(c^r)^{1+\alpha}\mathbf{q}^r/r \to 0$ in L^1 and $\{\check{v}_1^r/c^r, r \in \mathcal{R}\}$ is L^1 bounded (note that $\mathbf{q}^r = 0$ for all $r \in \mathcal{R}$ is a special case). In this case one can replace α^* in Theorem 2 with p. Also, in this case the assumption (3.8) in Theorem 4 can be omitted. Moreover, if we consider a sequence of initial conditions indexed by $p \geq 2$ satisfying (II) of Section 2.5 such that the choice of $\alpha = \alpha(p)$ can be made such that $\alpha(p)/\log p \to \infty$ as $p \to \infty$, then the assumptions (3.10), (3.11) and (3.12) in Theorem 5 can be replaced by the single assumption $\sup_{p\geq 2} \mathbb{E}[(v^{(p)})^2] < \infty$. This applies, in particular, if $\mathbf{q}^r = 0$ for all $r \in \mathcal{R}$.

4. Preliminaries.

In this section we recall some basic facts and record some well-known results that will be used several times in this work.

4.1. *Properties of the Skorohod map.* Recall the Skorohod map Γ defined in (3.1). The properties of Γ summarized here can be found in [35], Chapter 13.5, unless noted otherwise. Then, denoting $\mathcal{D}_0([0, \infty) : \mathbb{R})$ as the space of all $f \in \mathcal{D}([0, \infty) : \mathbb{R})$ with $f(0) \geq 0$, the map Γ is a continuous map from $\mathcal{D}_0([0, \infty) : \mathbb{R})$ to $\mathcal{D}([0, \infty) : \mathbb{R}_+)$. Furthermore, the following Lipschitz property holds: for all f_1, $f_2 \in \mathcal{D}_0([0, \infty) : \mathbb{R})$ and $T \in [0, \infty)$,

$$\sup_{t\in[0,T]} \big|\Gamma[f_1](t) - \Gamma[f_2](t)\big| \leq 2 \sup_{t\in[0,T]} \big|f_1(t) - f_2(t)\big|. \tag{4.1}$$

For any $f \in \mathcal{D}_0([0,\infty)$ and any t_1, t_2 such that $0 \le t_1 \le t_2 \le T$, defining functions $g_1(s) = \Gamma[f](t_1)$ and $g_2(s) = \Gamma[f](t_1) + f(s) - f(t_1)$ for $s \in [t_1, t_2]$, note that $\Gamma[g_1](t_2) = \Gamma[f](t_1)$ and $\Gamma[g_2](t_2) = \Gamma[f](t_2)$. Using (4.1), we conclude

$$|\Gamma[f](t_2) - \Gamma[f](t_1)| \le 2 \sup_{t_1 \le s \le t_2} |g_2(s) - g_1(s)| = 2 \sup_{t_1 \le s \le t_2} |f(s) - f(t_1)|. \tag{4.2}$$

The following monotonicity property also holds. Suppose f_1, $f_2 \in \mathcal{D}_0([0,\infty) : \mathbb{R})$ are such that, for all $0 \le s \le t < \infty$ $f_1(t) - f_1(s) \le f_2(t) - f_2(s)$ and $f_1(0) \le f_2(0)$. Then, it follows that $f_1(t) \le f_2(t)$ and $\sup_{0\le s\le t}(f_1(t) - f_1(s)) \le \sup_{0\le s\le t}(f_2(t) - f_2(s))$ for all $t \ge 0$. Hence,

$$\Gamma[f_1](t) \le \Gamma[f_2](t) \quad \text{for all } t \ge 0. \tag{4.3}$$

Let $f \in \mathcal{D}_0([0,\infty) : \mathbb{R})$. For $\varepsilon \in \mathbb{R}$, let $f_\varepsilon(t) := f(t) + \varepsilon t$, $t \ge 0$. Then for every $t \ge 0$

$$\varepsilon^{-1}\big[\Gamma[f_\varepsilon](t) - \Gamma[f_0](t)\big] \to t - \sup\big\{0 \le s \le t : f_0(s) = 0\big\} \quad \text{as } \varepsilon \to 0. \tag{4.4}$$

For a proof we refer to [22], Theorem 1.1 (see also pages 1921–1922 of [8]).

4.2. *Regularly varying functions.* Recall that we assume that the complementary cumulative distribution function $\bar{F}$ of the processing time distribution is a regularly varying function with index $-(p+1)$ for some $p > 1$, namely (2.1) is satisfied. Also recall that $S(\cdot)$ is given by (2.6). A function $L : [0,\infty) \to \mathbb{R}_+$ is called a slowly varying function if

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1 \quad \text{for all } t > 0.$$

We will frequently use the following well-known properties of regularly varying functions (see [23], Theorems 1.2.1, 1.2.4, 1.2.6).

(a) From [23], Remark 1.2.3, if $L(\cdot)$ is slowly varying, then for all $\epsilon > 0$,

$$\lim_{x\to\infty} \frac{L(x)}{x^\epsilon} = 0 \quad \text{and} \quad \lim_{x\to\infty} \frac{L(x)}{x^{-\epsilon}} = \infty.$$

(b) There exists a slowly varying function L such that $\overline{F}(x) = \frac{1}{x^{p+1}} L(x)$ for all $x \ge 1$. Henceforth, such a function $L(\cdot)$ is fixed.

(c) From Karamata's theorem [23], Theorem 1.2.6(b), with $\alpha = -p - 1$,

$$\lim_{x\to\infty} \frac{\int_x^\infty \overline{F}(t)\,dt}{x\overline{F}(x)} = \frac{1}{p}. \tag{4.5}$$

In particular, the function $z \mapsto \mathbb{E}(v\mathbf{1}_{[v>z]})$ is regularly varying with index $-p$ and therefore for all $a > 0$,

$$\lim_{r\to\infty} \frac{\mathbb{E}(v\mathbf{1}_{[v>ac^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} = \frac{1}{a^p}. \tag{4.6}$$

In fact, using [23], Theorem 1.2.4 and Theorem 1.2.6(b), one has that for all $\delta > 0$

$$\lim_{r\to\infty} \frac{\mathbb{E}(v\mathbf{1}_{[v>uc^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} = \frac{1}{u^p} \quad \text{uniformly for } u \in [\delta,\infty). \tag{4.7}$$

Also, there exists a slowly varying function $\hat{L}$ such that $\mathbb{E}(v\mathbf{1}_{[v>z]}) = z^{-p}\hat{L}(z)$ for all $z > 0$. By [23], Theorem 1.2.1, $\hat{L}$ can be represented as

$$\hat{L}(z) = c(z)\exp\left(\int_1^z \frac{\epsilon(y)}{y}\,dy\right), \quad z \ge 1, \tag{4.8}$$

where c and ϵ are nonnegative Borel measurable functions satisfying $\lim_{x\to\infty} c(x) = c_0 \in (0,\infty)$ and $\lim_{x\to\infty} \epsilon(x) \to 0$.

(d) By (2.6) and (4.6), $S(\cdot)$ is regularly varying with index p. Then, by Karamata's theorem ([23], Theorem 1.2.6(b)), as $x \to \infty$ $\frac{S(x)L(x)}{x^p} \to \frac{p}{p+1}$, where $L(\cdot)$ is given in (b). Combining this with (a), it follows that for any $\epsilon > 0$, there exists $x_\epsilon > 0$ such that for all $x \geq x_\epsilon$,

$$\frac{p}{p+1}x^{p-\epsilon} < S(x) < \frac{p}{p+1}x^{p+\epsilon}. \tag{4.9}$$

By (4.9) with $x = S^{-1}(r)$, (2.10), and the fact that $S^{-1}(\cdot)$ is strictly increasing, it follows that for any $\epsilon > 0$, there exists $r_\epsilon > 0$ such that for $r \geq r_\epsilon$,

$$\left(\frac{(p+1)r}{p}\right)^{1/(p+\epsilon)} < c^r < \left(\frac{(p+1)r}{p}\right)^{1/(p-\epsilon)}. \tag{4.10}$$

In particular,

$$\lim_{r\to\infty} \frac{c^r}{r} = 0. \tag{4.11}$$

(e) $S^{-1}(\cdot)$ is regularly varying with index $1/p$.

4.3. *A functional central limit theorem.* We will need the following well-known functional central limit theorem (cf. [25], Proposition A.1). For this, recall the definitions of λ^r, $E^r(\cdot)$, $\bar{E}^r(\cdot)$, and $\hat{E}^r(\cdot)$, for $r \in \mathcal{R}$, and λ, $\lambda(\cdot)$ and $E^*(\cdot)$ given in Section 2.2. Also, for $r \in \mathcal{R}$, let $\lambda^r(t) = \lambda^r t$ for $t \geq 0$.

PROPOSITION 6. *For each $r \in \mathcal{R}$, let $\{x_k^r\}_{k=1}^\infty$ be a sequence of nonnegative independent and identically distributed random variables, with finite mean m^r and finite standard deviation s^r, that is, independent of $E^r(\cdot)$. Suppose that for some finite nonnegative constants m and s, $m^r \to m$ and $s^r \to s$, as $r \to \infty$. Further suppose that, for each $\delta > 0$,*

$$\lim_{r\to\infty} \mathbb{E}\big[(x_1^r - m^r)^2 \mathbf{1}_{|x_1^r - m^r| > r\delta}\big] = 0.$$

For $r \in \mathcal{R}$, $n \in \mathbb{N}$ and $t \in [0,\infty)$, let

$$X^r(n) = \sum_{k=1}^n x_k^r \quad \text{and} \quad \hat{X}^r(t) = \big(X^r(\lfloor r^2 t \rfloor) - \lfloor r^2 t \rfloor m^r\big)/r.$$

Then, as $r \to \infty$, $(\hat{E}^r(\cdot), \hat{X}^r(\cdot)) \xrightarrow{d} (E^*(\cdot), X^*(\cdot))$ in $\mathcal{D}([0,\infty) : \mathbb{R}^2)$, where E^* is given as in (2.4) and X^* is a Brownian motion starting from zero with zero drift and variance s^2, that is independent of E^*. Furthermore, as $r \to \infty$,

$$\big[X^r\big(r^2\bar{E}^r(\cdot)\big) - r^2\lambda^r(\cdot)m^r\big]/r \xrightarrow{d} X^*\big(\lambda(\cdot)\big) + mE^*(\cdot),$$

in $\mathcal{D}([0,\infty) : \mathbb{R})$.

4.4. *Tightness and convergence criteria.* We record here certain convenient tools for establishing tightness and proving weak convergence that will be used several times in this article.

Aldous' tightness criterion. The following criterion is a useful tool in proving tightness. Let $\{\mathbb{X}^r(\cdot) : r \in \mathcal{R}\}$ be a collection of random variables in $\mathcal{D}([0,\infty) : \mathbb{R})$. We will call a random time τ a $\mathbb{X}^r$-stopping time if for each $t \geq 0$, the event $\{\tau \leq t\}$ lies in the σ-field $\sigma(\{\mathbb{X}^r(s) : s \leq t\})$. The collection $\{\mathbb{X}^r(\cdot) : r \in \mathcal{R}\}$ is tight if and only if the following two conditions hold:

(A1) For each $t \geq 0$,

$$\lim_{a\to\infty} \limsup_{r\to\infty} \mathbb{P}\big(|\mathbb{X}^r(t)| \geq a\big) = 0.$$

(A2) For each $\epsilon, \delta, T > 0$, there exists $\eta_0 > 0$ and $r_0 \in \mathcal{R}$ such that for any $0 < \eta \leq \eta_0$ and $r \geq r_0$, if τ is a $\mathbb{X}^r$-stopping time having a discrete, finite range satisfying $\tau \leq T$, then

$$\mathbb{P}\big(|\mathbb{X}^r(\tau+\eta) - \mathbb{X}^r(\tau)| \geq \delta\big) \leq \epsilon$$

(cf. [4], Theorem 16.10 and corollary to Theorem 16.8).

The following elementary lemma will be used several times in the proofs. We provide the short proof for completeness.

LEMMA 7. *Suppose that $(\mathcal{S}, \boldsymbol{d})$ is a Polish space, S^0 is an $\mathcal{S}$-valued random variable, $\{S_m\}_{m\in\mathbb{N}}$ is a sequence of $\mathcal{S}$-valued random variables and $\epsilon^* > 0$. For each $\epsilon \in (0, \epsilon^*]$, suppose that there is $b(\epsilon) > 0$, a $\mathcal{S}$-valued random variable S^ϵ and a sequence of random variables $\{S^\epsilon_m\}_{m\in\mathbb{N}}$, with S^ϵ_m and S_m defined on the same probability space for each $m \in \mathbb{N}$, such that the following hold*:

1. $\limsup_{m\to\infty} \mathbb{P}(\boldsymbol{d}(S^\epsilon_m, S_m) > b(\epsilon)) < b(\epsilon)$ *for each* $\epsilon \in (0, \epsilon^*]$ *and* $\lim_{\epsilon\searrow 0} b(\epsilon) = 0$;
2. *for each* $\epsilon \in (0, \epsilon^*]$, $S^\epsilon_m \xrightarrow{d} S^\epsilon$ *as* $m \to \infty$;
3. $S^\epsilon \xrightarrow{d} S^0$ *as* $\epsilon \to 0$.

Then $S_m \xrightarrow{d} S^0$ *as* $m \to \infty$.

PROOF. For an $\mathcal{S}$ valued random variable X, denote its probability law as μ_X. Let d_{BL} denote the bounded-Lipschitz metric for Borel probability measures on $(\mathcal{S}, \boldsymbol{d})$. Namely, for probability measures μ, ν on $\mathcal{S}$, $d_{BL}(\mu, \nu) = \sup_g |\int g\, d\mu - \int g\, d\nu|$ where the supremum is taken over all Lipschitz functions $g : \mathcal{S} \to \mathbb{R}$ that are bounded by 1 and whose Lipschitz constant is also bounded by 1. To prove the lemma it suffices to show that $d_{BL}(\mu_{S_m}, \mu_{S_0}) \to 0$ as $m \to \infty$. By triangle inequality and (1), for all $\epsilon \in (0, \epsilon^*]$ there exists m_ϵ such that for all $m \geq m_\epsilon$,

$$\begin{aligned} d_{BL}(\mu_{S_m}, \mu_{S_0}) &\leq d_{BL}(\mu_{S_m}, \mu_{S^\epsilon_m}) + d_{BL}(\mu_{S^\epsilon_m}, \mu_{S^\epsilon}) + d_{BL}(\mu_{S^\epsilon}, \mu_{S_0}) \\ &\leq d_{BL}(\mu_{S_m}, \mu_{S^\epsilon_m}) + d_{BL}(\mu_{S^\epsilon}, \mu_{S_0}) + 2b(\epsilon). \end{aligned}$$

Taking limit as $m \to \infty$ in the above gives that, for all $\epsilon \in (0, \epsilon^*]$,

$$\limsup_{m\to\infty} d_{BL}(\mu_{S_m}, \mu_{S_0}) \leq d_{BL}(\mu_{S^\epsilon}, \mu_{S_0}) + 2b(\epsilon).$$

By sending $\epsilon \searrow 0$ and using (3) (which implies that $\lim_{\epsilon\searrow 0} d_{BL}(\mu_{S^\epsilon}, \mu_{S_0}) = 0$), the result follows. □

5. Proofs. In this section we prove the main theorems stated in Section 3. To begin, we recall that we refer to a job's remaining processing time as its *size*. In addition, we refer to a job that arrived to the system after time zero as an *external job* and a job already in the system at time zero as an *initial job*. Recall that the processing time distribution does not depended on $r \in \mathcal{R}$. For each $r \in \mathcal{R}$, we assume that the processing times are determined by a common sequence $\{v_i\}_{i=1}^\infty$ of independent and identically distributed random variables with common cumulative distribution function F such that v_i denotes the processing time of the ith external job arriving to the rth SRPT queue. Beginning in Section 5.3, F is assumed to satisfy (2.1) henceforth. For $r \in \mathcal{R}$, $t \geq 0$, and $1 \leq i \leq E^r(t)$ (resp. $1 \leq i \leq \mathbf{q}^r$), we recall that

$v_i^r(t)$ (resp. $\check{v}_i^r(t)$) denotes the remaining processing time (or size) at time t of the ith external (resp. initial) job in the rth SRPT queue.

We begin by proving some general comparison results for SRPT queueing systems that hold quite generally in that they do not require condition (2.1). These comparison results, besides being of independent interest, will be used in the proofs of our main theorems.

5.1. *Intertwined SRPT queueing systems.* In this section we consider SRPT queues as introduced in Section 2.1. We fix r and suppress it from the notation in this section. Also, as in Section 2.1, we assume that the service time distribution F is continuous, but we do not require $\bar{F}$ to be regularly varying. In fact, even a finite mean is not needed. Consider two SRPT queueing systems, say S_1 and S_2, with a common arrival process $E(\cdot)$ (which, as in Section 2.1, is a delayed renewal process), but with (possibly) different initial conditions. For each $i = 1, 2$ and $t \geq 0$, let $n^{(i)}(t)$ be the number of jobs in system S_i at time t and let $\{v^i_{(j)}(t) : 1 \leq j \leq n^i(t)\}$ be the ordered collection of job sizes in system S_i at time t, with $v^i_{(1)}(t)$ denoting the smallest job at time t, $v^i_{(2)}(t)$ denoting the second smallest job at time t, and so on. For $i = 1, 2$, define $V^i_0(t) = 0$ and $V^i_j(t) := \sum_{k=1}^{j} v^i_{(k)}(t)$, $1 \leq j \leq n^i(t)$. For each $i = 1, 2$, the state of the system S_i at time t is completely described by the vector $(V^i_0(t), \ldots, V^i_{n^i(t)}(t))$. We say that S_2 is *intertwined* in S_1 at time t if there exist integers $k(t) \geq 0$ and $l(t) \geq 1$ such that the following hold: (i) S_1 has $k(t) + l(t) - 1$ or $k(t) + l(t)$ jobs and S_2 has $k(t) + l(t)$ jobs at time t, (ii) $V^1_j(t) = V^2_j(t)$ for all $0 \leq j \leq k(t)$, and (iii) for every $1 \leq l \leq l(t)$, $V^1_{k(t)+l-1}(t) < V^2_{k(t)+l}(t) < V^1_{k(t)+l}(t)$ (where, by convention, we take $V^1_{k(t)+l(t)}(t) = \infty$ if S_1 has $k(t) + l(t) - 1$ jobs at time t). Thus, if S_2 is intertwined in S_1 at time t, we have

$$0 < V^2_1(t) = V^1_1(t) < V^2_2(t) = V^1_2(t) < \cdots < V^2_{k(t)}(t) = V^1_{k(t)}(t) \quad \text{and}$$

$$V^1_{k(t)}(t) < V^2_{k(t)+1}(t) < V^1_{k(t)+1}(t) < \cdots$$

$$< V^1_{k(t)+l(t)-1}(t) < V^2_{k(t)+l(t)}(t) < V^1_{k(t)+l(t)}(t).$$

On intervals of time when no arrival or departure takes place in either system, each V^i_j decreases at rate one as each server processes the work associated with the shortest job. Hence, intertwinement is preserved on such intervals. In the next two lemmas, we argue that intertwinement is preserved at times of a job arrival and a synchronous departure and swapped at times of an asynchronous departure in that if S_2 is intertwined in S_1 immediately before such a departure, then S_1 is intertwined in S_2 immediately following such a departure. This property, in turn, is used to compare the queue length processes of S_1 and S_2. A related, but different, notion for comparing the state of two queueing systems with a common arrival process, called work-dominance, was previously introduced by Smith [31] to establish optimality of SRPT.

To begin, we have the following lemma which states that if one system is intertwined in the other immediately before a job arrival (which is the same for both systems) then this intertwining is preserved immediately after the arrival.

LEMMA 8. *Suppose S_1 and S_2 are two SRPT queueing systems with a common arrival process. Almost surely, if at some $t > 0$ a job arrives in the two systems, and S_2 is intertwined in S_1 just before time t, then S_2 is intertwined in S_1 at time t.*

PROOF. Denote the processing time of the entering job at time t by v^*. Since F is continuous, $P(v^* = c) = 0$ for any $c \geq 0$. This property will be used without additional comments in many of the arguments below. Note that if $v^* < v^2_{(k(t-)+1)}(t-)$, then $k(t) = k(t-) + 1$ and

$l(t) = l(t-)$. In this case, for $1 \le l \le l(t)$, $V^i_{k(t)+l}(t) = V^i_{k(t-)+l}(t-) + v^*$ for $i = 1, 2$ and as S_2 was intertwined in S_1 just before time t, we obtain $V^1_{k(t)+l-1}(t) < V^2_{k(t)+l}(t) < V^1_{k(t)+l}(t)$ for all $1 \le l \le l(t)$, thus S_2 is intertwined in S_1 at time t. Otherwise, $k(t) = k(t-)$ and $l(t) = l(t-) + 1$, which we assume henceforth. For $1 \le l \le l(t)$, we consider the four possibilities as follows:

(i) $v^* > \max\{v^1_{(k(t-)+l-1)}(t-), v^2_{(k(t-)+l)}(t-)\}$, in which case,

$$V^1_{k(t)+l-1}(t) = V^1_{k(t)+l-1}(t-) \quad \text{and} \quad V^2_{k(t)+l}(t) = V^2_{k(t)+l}(t-).$$

Thus, by intertwinement before time t, $V^2_{k(t)+l}(t) > V^1_{k(t)+l-1}(t)$.

(ii) $v^1_{(k(t-)+l-1)}(t-) < v^* < v^2_{(k(t-)+l)}(t-)$, in which case, $V^1_{k(t)+l-1}(t) = V^1_{k(t)+l-1}(t-)$ and $V^2_{k(t)+l}(t) = V^2_{k(t)+l-1}(t-) + v^*$. As $v^* > v^1_{(k(t-)+l-1)}(t-) = V^1_{k(t)+l-1}(t-) - V^1_{k(t)+l-2}(t-)$, $V^1_{k(t)+l-2}(t-) < V^2_{k(t)+l-1}(t-)$ by intertwinement before time t, we obtain

$$\begin{aligned} V^2_{k(t)+l}(t) &= V^2_{k(t)+l-1}(t-) + v^* > V^2_{k(t)+l-1}(t-) + \big(V^1_{k(t)+l-1}(t-) - V^1_{k(t)+l-2}(t-)\big) \\ &> V^2_{k(t)+l-1}(t-) + \big(V^1_{k(t)+l-1}(t-) - V^2_{k(t)+l-1}(t-)\big) \\ &= V^1_{k(t)+l-1}(t-) = V^1_{k(t)+l-1}(t). \end{aligned}$$

(iii) $v^2_{(k(t-)+l)}(t-) < v^* < v^1_{(k(t-)+l-1)}(t-)$, in which case, we have $V^1_{k(t)+l-1}(t) = V^1_{k(t)+l-2}(t-) + v^*$ and $V^2_{k(t)+l}(t) = V^2_{k(t)+l}(t-)$. Also, since $k(t) = k(t-)$ and $l(t) = l(t-) + 1$, we have $l \ge 2$. As $v^* < v^1_{(k(t-)+l-1)}(t-) = V^1_{k(t)+l-1}(t-) - V^1_{k(t)+l-2}(t-)$,

$$\begin{aligned} V^1_{k(t)+l-1}(t) &= V^1_{k(t)+l-2}(t-) + v^* \\ &< V^1_{k(t)+l-2}(t-) + \big(V^1_{k(t)+l-1}(t-) - V^1_{k(t)+l-2}(t-)\big) = V^1_{k(t)+l-1}(t-). \end{aligned}$$

By intertwinement before time t, $V^2_{k(t)+l}(t-) > V^1_{k(t)+l-1}(t-)$. Hence,

$$V^2_{k(t)+l}(t) = V^2_{k(t)+l}(t-) > V^1_{k(t)+l-1}(t-) > V^1_{k(t)+l-1}(t).$$

(iv) $v^* < \min\{v^1_{(k(t-)+l-1)}(t-), v^2_{(k(t-)+l)}(t-)\}$, in which case, we have $V^1_{k(t)+l-1}(t) = V^1_{k(t)+l-2}(t-) + v^*$, $V^2_{k(t)+l}(t) = V^2_{k(t)+l-1}(t-) + v^*$ and $l \ge 2$. By intertwinement before time t,

$$V^2_{k(t)+l}(t) = V^2_{k(t)+l-1}(t-) + v^* > V^1_{k(t)+l-2}(t-) + v^* = V^1_{k(t)+l-1}(t).$$

As, almost surely, the above are the only four possibilities, we have, almost surely, $V^2_{k(t)+l}(t) > V^1_{k(t)+l-1}(t)$ for all $1 \le l \le l(t)$. By a symmetric argument, we obtain, almost surely, $V^1_{k(t)+l}(t) > V^2_{k(t)+l}(t)$ for all $1 \le l \le l(t)$. This completes the proof of the lemma. □

The following proposition compares the queue length processes for two SRPT systems started from intertwined configurations and having the same arrival process.

PROPOSITION 9. *Suppose S_1 and S_2 are two SRPT queueing systems with a common arrival process. Moreover, assume that S_2 is intertwined in S_1 at time zero. Denote the queue length process for S_i by $Q_i(\cdot)$, $i = 1, 2$, and assume $Q_2(0) = Q_1(0) + 1$. Then, almost surely, for any $t \ge 0$,*

$$Q_1(t) \le Q_2(t) \le Q_1(t) + 1.$$

PROOF. As S_2 is intertwined in S_1 at time zero, $Q_1(0) = k(0) + l(0) - 1$ and $Q_2(0) = k(0) + l(0)$. Define $\tau_0^{as} = 0$ and denote by τ_i^{as}, $i \geq 1$, the time of the ith asynchronous departure, that is, when there is a departure from one system but not the other. For any $i \geq 0$, on the time interval $[\tau_i^{as}, \tau_{i+1}^{as})$, arrivals and departures happen at the same times (synchronously) from both systems. Clearly, if S_2 is intertwined in S_1 before a synchronous departure, then it remains so after the departure. Also, after any arrival, by Lemma 8, S_2 remains intertwined in S_1 if it were the case immediately before the arrival. Thus, if S_2 is intertwined in S_1 at time τ_i^{as}, then the same property is true for every $t \in [\tau_i^{as}, \tau_{i+1}^{as})$. Then, for any $t \in [0, \tau_1^{as})$, $Q_1(t) = k(t) + l(t) - 1$ and $Q_2(t) = k(t) + l(t)$, and hence, $Q_2(t) - Q_1(t) = 1$. Moreover, as for any $t \in [0, \tau_1^{as})$, $V_{k(t)+1}^2(t) < V_{k(t)+1}^1(t)$, the first asynchronous departure happens from S_2. Thus, S_1 is intertwined in S_2 at time τ_1^{as} (i.e., the intertwinement order changes) and $Q_1(\tau_1^{as}) = Q_2(\tau_1^{as}) = k(\tau_1^{as}) + l(\tau_1^{as})$. By the same argument as above, we deduce that S_1 remains intertwined in S_2 on the time interval $[\tau_1^{as}, \tau_2^{as})$ and $Q_1(t) = Q_2(t) = k(t) + l(t)$ for all $t \in [\tau_1^{as}, \tau_2^{as})$. At time τ_2^{as}, departure happens from S_1 and the intertwinement order switches again at τ_2^{as}, and so on. Thus, we conclude that $Q_1(t) = k(t) + l(t) - 1$, $Q_2(t) = k(t) + l(t)$ for all $t \in [\tau_{2k}^{as}, \tau_{2k+1}^{as})$, $k \geq 0$, and $Q_1(t) = Q_2(t) = k(t) + l(t)$ for all $t \in [\tau_{2k+1}^{as}, \tau_{2k+2}^{as})$, $k \geq 0$. In particular, this proves the proposition. □

5.2. *Truncated SRPT queues.* For each $r \in \mathcal{R}$ and $a \in [0, \infty]$, we consider an SRPT queue with a thinned external arrival process $E_a^r(\cdot) := \sum_{i=1}^{E^r(\cdot)} \mathbf{1}_{[v_i \leq ac^r]}$, which we refer to as the rth a-truncated SRPT queue. When the ith external job arrives to the rth SRPT queue, it is an external job for the rth a-truncated SRPT queue if and only if its processing time v_i is less or equal to ac^r. Similarly, jobs in the rth a-truncated SRPT queue at time zero, namely the initial jobs, are those that are initial jobs in the rth SRPT queue such that $\breve{v}_l^r \leq ac^r$ and $1 \leq l \leq \mathbf{q}^r$. Then the rth a-truncated SRPT queue evolves in time in accordance with the SRPT service discipline by preemptively serving the job with the shortest size first. For $r \in \mathcal{R}$, $t \geq 0$ and $1 \leq i \leq E_a^r(t)$, let $v_i^{r,a}(t)$ be the size at time t in the rth a-truncated SRPT queue of the ith external arrival to the rth a-truncated SRPT queue. Similarly, for $r \in \mathcal{R}$, $t \geq 0$ and $1 \leq l \leq \mathbf{q}^r$, let $\breve{v}_l^{r,a}(t)$ be the size at time t in the rth a-truncated SRPT queue of the lth initial job in the rth a-truncated SRPT queue if $\breve{v}_l^r \leq ac^r$, and zero if $\breve{v}_l^r > ac^r$ (the latter case is vacuous if $a = \infty$).

Define for each $r \in \mathcal{R}$, $a \in [0, \infty]$ and $t \geq 0$,

$$V_a^r(t) := \sum_{i=1}^{E^r(t)} v_i \mathbf{1}_{[v_i \leq ac^r]},$$

$$\hat{V}_a^r(t) := \frac{1}{r} \sum_{i=1}^{E^r(r^2 t)} v_i \mathbf{1}_{[v_i \leq ac^r]} - r\lambda^r t \mathbb{E}(v \mathbf{1}_{[v \leq ac^r]}),$$

$$X_a^r(t) := \frac{1}{r} \sum_{l=0}^{\mathbf{q}^r} \breve{v}_l^r \mathbf{1}_{[\breve{v}_l^r \leq ac^r]} + \frac{1}{r} V_a^r(r^2 t) - rt,$$

$$Y_a^r(t) := \Gamma[X_a^r](t).$$

Also, for $r \in \mathcal{R}$, $a \in [0, \infty]$ and $t \geq 0$,

$$\widetilde{\mathcal{Q}}_a^r(t) := \frac{c^r}{r} \sum_{l=1}^{\mathbf{q}^r} \delta^+_{\breve{v}_l^{r,a}(r^2 t)/c^r} + \frac{c^r}{r} \sum_{i=1}^{E_a^r(r^2 t)} \delta^+_{v_i^{r,a}(r^2 t)/c^r}$$

denotes the scaled measure describing the state of the rth a-truncated SRPT queue at time $r^2 t$ and $Q_a^r(t) := \langle \mathbf{1}, \widetilde{\mathcal{Q}}_a^r(t) \rangle$ denotes the scaled queue length in the rth a-truncated SRPT queue

at time r^2t. Recall that, for each $r \in \mathcal{R}$, $a \in [0,\infty]$ and $t \geq 0$, $Z^r_a(t)$ and $W^r_a(t)$ are defined in (2.12) and (2.13) respectively.

We have elected to state the results in this section for truncated SRPT queues in terms of scaled processes defined above. However, since they hold for each $r \in \mathcal{R}$, one can obtain unscaled versions from these. Also, as in Section 5.1, F is required to be continuous, but $\bar{F}$ is not required to be regularly varying. The following proposition records a key observation comparing the process $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\cdot)\rangle$ with $Y^r_a(\cdot)$ and $\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\cdot)\rangle$ with $Q^r_a(\cdot)$ for $a \leq y \leq \infty$.

PROPOSITION 10. *For any $r \in \mathcal{R}$, $a \in (0,\infty)$, $a \leq y \leq \infty$ and $t \geq 0$, we have, almost surely,*

$$Y^r_a(t) \leq \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle \leq Y^r_a(t) + \frac{ac^r}{r}, \tag{5.1}$$

$$Q^r_a(t) \leq \langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle \leq Q^r_a(t) + \frac{c^r}{r}. \tag{5.2}$$

In particular, for any $r \in \mathcal{R}$, $a \in (0,\infty)$ and $t \geq 0$, we have, almost surely,

$$Y^r_a(t) \leq W^r_a(t) \leq Y^r_a(t) + \frac{ac^r}{r}, \tag{5.3}$$

$$Q^r_a(t) \leq Z^r_a(t) \leq Q^r_a(t) + \frac{c^r}{r}. \tag{5.4}$$

Moreover, almost surely, $W^r_0(t) = Y^r_0(t) = 0$, $Z^r_0(t) = Q^r_0(t) = 0$, $W^r_\infty(t) = Y^r_\infty(t)$, and $Z^r_\infty(t) = Q^r_\infty(t)$ for any $r \in \mathcal{R}$ and $t \geq 0$.

PROOF. Fix $r \in \mathcal{R}$. Note that, by definition, $W^r_0(t) = Z^r_0(t) = Q^r_0(t) = 0$ for all $t \geq 0$. Moreover, almost surely, $X^r_0(t) = -rt$ for all $t \geq 0$ and hence $Y^r_0(t) = \Gamma[X^r_0](t) = 0$ for all $t \geq 0$. Also, as $\widetilde{\mathcal{Q}}^r_\infty(t) = \widetilde{\mathcal{Z}}^r(t)$ for all $t \geq 0$, $W^r_\infty(t) = Y^r_\infty(t)$ and $Z^r_\infty(t) = Q^r_\infty(t)$ for all $t \geq 0$. Thus, the assertions in the last line of the lemma hold. Also, for each $a \in (0,\infty)$, (5.3) follows from (5.1) and (5.4) follows from (5.2) upon setting $y = \infty$, since $\widetilde{\mathcal{Z}}^r(\cdot) = \widetilde{\mathcal{Q}}^r_\infty(\cdot)$. So it suffices to verify (5.1) and (5.2).

Fix $a \in (0,\infty)$ and $a \leq y \leq \infty$. Define the stopping times $\sigma_{-1} = 0$, and for $k \in \mathbb{Z}_+$,

$$\sigma_{2k} := \inf\{s \geq \sigma_{2k-1} : \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(s)\rangle = 0\},$$

$$\sigma_{2k+1} := \inf\{s \geq \sigma_{2k} : \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(s)\rangle > 0\}.$$

To show (5.1) and (5.2), we proceed by induction. Observe that, by definition, $Y^r_a(0) = \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(0)\rangle$ and $Q^r_a(0) = \langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(0)\rangle$ since $a \leq y$. Thus, (5.1) and (5.2) hold on $[0, \sigma_{-1}]$.

First consider the case $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(0)\rangle = 0$ (which implies that $\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(0)\rangle = 0$). Then, $\sigma_0 = \sigma_{-1} = 0$ and $Y^r_a(t) = \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle = 0$ for all $t \in [0, \sigma_1)$. The map $t \mapsto \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle$ increases at $t = \sigma_1$ due to one of the following two events: (i) an external job with processing time less or equal to ac^r arrives to the system at time $r^2\sigma_1$ or (ii) an initial job with initial size in $(ac^r, yc^r]$ or an external job with processing time in $(ac^r, yc^r]$ that arrived during the time interval $(0, r^2\sigma_1)$, in course of getting served, has its size drop to ac^r at time $r^2\sigma_1$. If (i) occurs, $Q^r_a(\sigma_1) = \langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_1)\rangle = \frac{c^r}{r}$, $Y^r_a(\sigma_1) = \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_1)\rangle \leq \frac{ac^r}{r}$. If (ii) occurs, $Q^r_a(\sigma_1) = 0$, $\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_1)\rangle = \frac{c^r}{r}$, $Y^r_a(\sigma_1) = 0$, and $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_1)\rangle = \frac{ac^r}{r}$. Thus, when $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(0)\rangle = 0$, (5.1) and (5.2) hold for all $t \in [0, \sigma_1]$.

Suppose that for some $k \in \mathbb{Z}_+$ (5.1) and (5.2) hold for all $t \in [0, \sigma_{2k-1}]$ and

$$\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_{2k-1})\rangle > 0$$

(which implies that $\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_{2k-1})\rangle > 0$). We first show that (5.1) and (5.2) hold for all $t \in (\sigma_{2k-1}, \sigma_{2k}]$. By virtue of the SRPT dynamics, no job in the rth y-truncated SRPT queue at time $r^2\sigma_{2k-1}$ of size greater than ac^r at time $r^2\sigma_{2k-1}$ is served in the rth y-truncated SRPT queue during the time interval $[r^2\sigma_{2k-1}, r^2\sigma_{2k})$. Consequently, for any $t \in (\sigma_{2k-1}, \sigma_{2k})$, the following four properties are equivalent: (a) $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle - \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t-)\rangle > 0$; (b) $E^r_a(r^2t) - E^r_a(r^2t-) > 0$; (c) $X^r_a(t) - X^r_a(t-) > 0$; (d) $Y^r_a(t) - Y^r_a(t-) > 0$ and, when these equivalent properties hold, $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle - \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t-)\rangle = Y^r_a(t) - Y^r_a(t-)$. This also shows that for $t \in [\sigma_{2k-1}, \sigma_{2k})$ such that $Y^r_a(t) = 0$ and $s \in [t, \inf\{u \geq t : Y^r_a(u) > 0\} \wedge \sigma_{2k})$, $Y^r_a(s) = 0$ and $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(s)\rangle = \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle - r(s-t)$. Moreover, for $t \in [\sigma_{2k-1}, \sigma_{2k})$ such that $0 < Y^r_a(t) \leq \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle$ and $s \in [t, \inf\{u \geq t : Y^r_a(u) = 0\}]$,

$$\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(s)\rangle - \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle = \frac{1}{r}\big(V^r_a(r^2s) - V^r_a(r^2t)\big) - r(s-t) = Y^r_a(s) - Y^r_a(t).$$

From these observations, we conclude that $t \mapsto \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle - Y^r_a(t)$ is nonincreasing on the interval $[\sigma_{2k-1}, \sigma_{2k}]$ and decreases only on the set $\{u \in [\sigma_{2k-1}, \sigma_{2k}] : Y^r_a(u) = 0\}$. This also implies that either $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(u)\rangle = Y^r_a(u)$ for all $u \in (\sigma_{2k-1}, \sigma_{2k}]$ or the first $t \geq \sigma_{2k-1}$ for which $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle = Y^r_a(t)$ corresponds to σ_{2k} when $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_{2k})\rangle = Y^r_a(\sigma_{2k}) = 0$. We conclude that for any $t \in [\sigma_{2k-1}, \sigma_{2k}]$,

$$\begin{aligned} 0 &= \big\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_{2k})\big\rangle - Y^r_a(\sigma_{2k}) \leq \big\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\big\rangle - Y^r_a(t) \\ &\leq \big\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_{2k-1})\big\rangle - Y^r_a(\sigma_{2k-1}) \leq \frac{ac^r}{r}, \end{aligned}$$

where the last inequality holds by the induction hypothesis. Hence, (5.1) holds for all $t \in (\sigma_{2k-1}, \sigma_{2k}]$.

Now we show that (5.2) holds for all $t \in (\sigma_{2k-1}, \sigma_{2k}]$. If $k \in \mathbb{N}$, then, by definition of σ_{2k-1}, $\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_{2k-1}-)\rangle = 0$, and so, using the induction hypothesis, $Q^r_a(\sigma_{2k-1}-) = 0$. Moreover, the arrival times and processing times of all external jobs with processing time less than or equal to ac^r into both the rth a-truncated SRPT queue and the rth y-truncated SRPT queue on the time interval $[r^2\sigma_{2k-1}, r^2\sigma_{2k}]$ are common to both systems. Further, no job in the rth y-truncated SRPT queue at time $r^2\sigma_{2k-1}$ of size greater than ac^r at time $r^2\sigma_{2k-1}$ is served in the rth y-truncated SRPT queue during the time interval $[r^2\sigma_{2k-1}, r^2\sigma_{2k}]$. Thus, the processes $t \mapsto Q^r_a(t)$ and $t \mapsto \langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle$ on the time interval $[r^2\sigma_{2k-1}, r^2\sigma_{2k}]$ can be identified with the (scaled) queue length processes of two rth a-truncated SRPT queueing systems having the same arrival process, denoted respectively by S^r_1 and S^r_2, started at time zero and observed till S^r_2 has zero jobs. If $k = 0$ or if the increase in $t \mapsto \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle$ at time $t = \sigma_{2k-1}$ happens due to the arrival of an external job with processing time less than or equal to ac^r, then $\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(\sigma_{2k-1})\rangle = Q^r_a(\sigma_{2k-1})$. Thus, in this case, S^r_1 and S^r_2 start with the same configuration and hence, $Q^r_a(t) = \langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle$ for all $t \in [\sigma_{2k-1}, \sigma_{2k}]$. On the other hand, the increase in $t \mapsto \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\rangle$ at time $t = \sigma_{2k-1}$ may happen due to a job present in the system at a time $s < r^2\sigma_{2k-1}$, with its size in the range $(ac^r, yc^r]$ at time s, getting served in the yth truncated queue and having its size drop to ac^r at time $r^2\sigma_{2k-1}$. In this case, S^r_2 starts with one job of size ac^r and S^r_1 starts with zero jobs. Hence, S^r_2 is intertwined in S^r_1 at time zero in the sense of Section 5.1 with $k(0) = 0$ and $l(0) = 1$, and S^r_2 has one more job at time zero than S^r_1. By Proposition 9, for any $t \in [\sigma_{2k-1}, \sigma_{2k}]$,

$$Q^r_a(t) \leq \big\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}^r_y(t)\big\rangle \leq Q^r_a(t) + \frac{c^r}{r}.$$

Hence, (5.2) holds for all $t \in (\sigma_{2k-1}, \sigma_{2k}]$.

To see that (5.1) and (5.2) hold for all $t \in (\sigma_{2k}, \sigma_{2k+1}]$, first note that $Y_a^r(t) = \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}_y^r(t) \rangle = 0$ for all $t \in (\sigma_{2k}, \sigma_{2k+1})$. Moreover, observe that either $Q_a^r(\sigma_{2k+1}) = \langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}_y^r(\sigma_{2k+1}) \rangle = \frac{c^r}{r}$ and $Y_a^r(\sigma_{2k+1}) = \langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}_y^r(\sigma_{2k+1}) \rangle \le \frac{ac^r}{r}$, or $Q_a^r(\sigma_{2k+1}) = 0$, $\langle \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}_y^r(\sigma_{2k+1}) \rangle = \frac{c^r}{r}$, $Y_a^r(\sigma_{2k+1}) = 0$ and $\langle \chi \mathbf{1}_{[0,a]}, \widetilde{\mathcal{Q}}_y^r(\sigma_{2k+1}) \rangle = \frac{ac^r}{r}$. In both cases, (5.1) and (5.2) hold for all $t \in (\sigma_{2k}, \sigma_{2k+1}]$.

Thus, by induction, (5.1) and (5.2) hold for all $t \in [0, \lim_{k\to\infty} \sigma_{2k})$. To complete the proof, we show that $\lim_{k\to\infty} \sigma_{2k} = \infty$. Suppose first that $\mathbb{E}(v\mathbf{1}_{[v\le ac^r]}) > 0$. For each $k \in \mathbb{Z}_+$, let v_k^* be the processing time of the first external job to arrive to the rth y-truncated SRPT queue after time σ_{2k}. Then it is easy to see that for each $k \in \mathbb{Z}_+$, $\sigma_{2k+2} - \sigma_{2k+1} \ge r^{-2} v_k^* \mathbf{1}_{[v_k^* \le ac^r]}$. As $\{v_k^* \mathbf{1}_{[v_k^* \le ac^r]}\}_{k\ge 0}$ is a sequence of independent and identically distributed random variables, where each element has the same distribution as $v\mathbf{1}_{[v\le ac^r]}$, and since $\mathbb{E}(v\mathbf{1}_{[v\le ac^r]}) > 0$, almost surely,

$$\lim_{k\to\infty} \sigma_{2k} \ge \lim_{k\to\infty} \sum_{j=0}^{k-1} (\sigma_{2j+2} - \sigma_{2j+1}) \ge r^{-2} \lim_{k\to\infty} \sum_{j=0}^{2k-2} v_j^* \mathbf{1}_{[v_j^* \le ac^r]} = \infty.$$

If $\mathbb{E}(v\mathbf{1}_{[v\le ac^r]}) = 0$ and $\mathbb{E}(v\mathbf{1}_{[v\le yc^r]}) > 0$, then almost surely, no external job with processing time less or equal to ac^r arrives into the system and thus almost surely, $\sigma_{2k+2} - \sigma_{2k+1} = r^{-2}ac^r$ for all $k \in \mathbb{Z}_+$, and hence $\lim_{k\to\infty} \sigma_{2k} = \infty$, as desired. Finally, if $\mathbb{E}(v\mathbf{1}_{[v\le yc^r]}) = 0$, which implies that $\mathbb{E}(v\mathbf{1}_{[v\le ac^r]}) = 0$ since $a \le y$, then there exists $k_0 \in \mathbb{Z}_+$ such that $Q_y^r(\sigma_{2k_0}) = 0$ and thus $\sigma_{2k_0+1} = \infty$. Hence (5.1) and (5.2) hold for all $t \in [0, \infty)$. □

The following lemma compares queue length processes for truncated SRPT queues with different truncations.

LEMMA 11. *For all $r \in \mathcal{R}$, $0 \le x \le y \le \infty$ and $t \ge 0$,*

$$0 \le Q_y^r(t) - Q_x^r(t) \le \frac{c^r}{r} + x^{-1} Y_y^r(t).$$

PROOF. Fix $r \in \mathcal{R}$, $0 \le x \le y \le \infty$ and $t \ge 0$. Note that, almost surely,

$$\begin{aligned} 0 \le Q_y^r(t) - Q_x^r(t) &= \int_0^x \widetilde{\mathcal{Q}}_y^r(t)(dz) - Q_x^r(t) + \int_x^y \widetilde{\mathcal{Q}}_y^r(t)(dz) \\ &= \langle \mathbf{1}_{[0,x]}, \widetilde{\mathcal{Q}}_y^r(t) \rangle - Q_x^r(t) + \int_x^y \widetilde{\mathcal{Q}}_y^r(t)(dz). \end{aligned} \tag{5.5}$$

By (5.2) in Proposition 10 with $a = x$, almost surely,

$$0 \le \langle \mathbf{1}_{[0,x]}, \widetilde{\mathcal{Q}}_y^r(t) \rangle - Q_x^r(t) \le \frac{c^r}{r}.$$

Using this observation in (5.5), we obtain

$$\begin{aligned} 0 \le Q_y^r(t) - Q_x^r(t) &\le \frac{c^r}{r} + \int_x^y \widetilde{\mathcal{Q}}_y^r(t)(dz) \\ &\le \frac{c^r}{r} + x^{-1} \int_x^y z \widetilde{\mathcal{Q}}_y^r(t)(dz) \le \frac{c^r}{r} + x^{-1} Y_y^r(t), \end{aligned} \tag{5.6}$$

as desired. □

5.3. *Proof of Theorem 1.* The following lemma is a functional central limit theorem for $\{X^r_\cdot(\cdot) : r \in \mathcal{R}\}$, which is used below in conjunction with the result in Proposition 10 to prove Theorem 1. For this, recall the definition of X^r_a and X_a, $a \in [0,\infty]$, from Section 5.2 and (3.4) respectively.

LEMMA 12. *There exists a probability space on which we are given a Brownian motion B and a $\mathcal{C}([0,\infty):\mathbb{R}_+)\times\mathbb{R}_+$ valued random variable $(\xi(\cdot),\xi(\infty))$ independent of B, with the same distribution as $(w^*(\cdot),w^*(\infty))$, such that for any $k\in\mathbb{N}$ and any $0<a_1<\cdots<a_k\le\infty$, as $r\to\infty$,*

$$\big(X^r_{a_1}(\cdot),\ldots,X^r_{a_k}(\cdot)\big)\xrightarrow{d}\big(X_{a_1}(\cdot),\ldots,X_{a_k}(\cdot)\big)$$

in $\mathcal{C}([0,\infty):\mathbb{R}^k)$.

PROOF. Note that for any $r\in\mathcal{R}$, $a\in(0,\infty)$ and $t\ge 0$,

$$X^r_\infty(t)=X^r_\infty(0)+\frac{1}{r}V^r_\infty\big(r^2t\big)-rt=X^r_\infty(0)+\hat{V}^r_\infty(t)+r\big(\rho^r-1\big)t, \tag{5.7}$$

$$X^r_a(t)=X^r_a(0)+\frac{1}{r}V^r_a\big(r^2t\big)-rt=X^r_a(0)+\hat{V}^r_a(t)+r\big(\rho^r_{ac^r}-1\big)t, \tag{5.8}$$

where $X^r_a(0)=\frac{1}{r}\sum_{l=0}^{\mathbf{q}^r}\check{v}^r_l\mathbf{1}_{[\check{v}^r_l\le ac^r]}$. Note that for any $r\in\mathcal{R}$, $a\in(0,\infty)$ and $t\ge 0$,

$$\begin{aligned}
r\big(\rho^r_{ac^r}-1\big)&=r\big(\rho^r_{ac^r}-\rho^r\big)+r\big(\rho^r-1\big)=-r\lambda^r\mathbb{E}\big(v\mathbf{1}_{[v>ac^r]}\big)+r\big(\rho^r-1\big)\\
&=-\lambda^r\frac{\mathbb{E}\big(v\mathbf{1}_{[v>ac^r]}\big)}{\mathbb{E}\big(v\mathbf{1}_{[v>c^r]}\big)}r\mathbb{E}\big(v\mathbf{1}_{[v>c^r]}\big)+r\big(\rho^r-1\big)\\
&=-\lambda^r\frac{\mathbb{E}\big(v\mathbf{1}_{[v>ac^r]}\big)}{\mathbb{E}\big(v\mathbf{1}_{[v>c^r]}\big)}\frac{r}{S(c^r)}+r\big(\rho^r-1\big)=-\lambda^r\frac{\mathbb{E}\big(v\mathbf{1}_{[v>ac^r]}\big)}{\mathbb{E}\big(v\mathbf{1}_{[v>c^r]}\big)}+r\big(\rho^r-1\big),
\end{aligned} \tag{5.9}$$

where we have used (2.10) in the final equality. By (4.6),

$$\lim_{r\to\infty}\frac{\mathbb{E}\big(v\mathbf{1}_{[v>ac^r]}\big)}{\mathbb{E}\big(v\mathbf{1}_{[v>c^r]}\big)}=\frac{1}{a^p}.$$

Using this and assumption (2.2) in the above equation, we obtain that for each $a\in(0,\infty)$,

$$r\big(\rho^r_{ac^r}-1\big)\to\kappa-\frac{\lambda}{a^p}\quad\text{as } r\to\infty. \tag{5.10}$$

For $a\in(0,\infty)$, let $m^r_a=\mathbb{E}\big(v\mathbf{1}_{[v\le ac^r]}\big)$ and $(s^r_a)^2=\operatorname{Var}\big(v\mathbf{1}_{[v\le ac^r]}\big)$. Then, finiteness of the second moment of v and the fact that $\lim_{r\to\infty}c^r=\infty$ give that, for $a\in(0,\infty)$, $\lim_{r\to\infty}m^r_a=\mathbb{E}(v)$, $\lim_{r\to\infty}(s^r_a)^2=\operatorname{Var}(v)$, and for each $\delta>0$

$$\lim_{r\to\infty}\mathbb{E}\big[\big(v\mathbf{1}_{[v\le ac^r]}-m^r_a\big)^2\mathbf{1}_{|v\mathbf{1}_{[v\le ac^r]}-m^r_a|>r\delta}\big]=0.$$

Thus, by Proposition 6, for each $a\in(0,\infty)$, $\hat{V}^r_a(\cdot)\xrightarrow{d}\sigma B(\cdot)$ where $\sigma^2=\lambda\operatorname{Var}(v)+(\mathbb{E}(v))^2\lambda^3\sigma_A^2=\lambda\operatorname{Var}(v)+\lambda\sigma_A^2$ and B is a standard Brownian motion. Note that, from (2.14) and assumed mutual independence in Section 2, we in fact have that, for each $a\in(0,\infty)$,

$$\big(X^r_\cdot(0),\hat{V}^r_a(\cdot)\big)\xrightarrow{d}\big(\xi(\cdot),\sigma B(\cdot)\big) \tag{5.11}$$

in $\mathcal{D}([0,\infty):\mathbb{R}_+)\times\mathcal{D}([0,\infty):\mathbb{R})$, where ξ is distributed as w^* and is independent of B.

For each $0 < a < b \le \infty$,

$$\hat{V}_b^r(t) - \hat{V}_a^r(t) = \frac{1}{r} \sum_{i=1}^{E^r(r^2 t)} v_i \mathbf{1}_{[ac^r < v_i \le bc^r]} - r\lambda^r t \mathbb{E}(v \mathbf{1}_{[ac^r < v \le bc^r]}).$$

Note that by the finiteness of the second moment of v and $\lim_{r\to\infty} c^r = \infty$, for each $0 < a < b \le \infty$, as $r \to \infty$,

$$\mathbb{E}(v \mathbf{1}_{[ac^r < v \le bc^r]}) \to 0 \quad \text{and} \quad \mathrm{Var}(v \mathbf{1}_{[ac^r < v \le bc^r]}) \le \mathbb{E}\big(v^2 \mathbf{1}_{[ac^r < v]}\big) \to 0. \tag{5.12}$$

Thus, by Proposition 6, for each $0 < a < b \le \infty$,

$$\hat{V}_b^r(\cdot) - \hat{V}_a^r(\cdot) \xrightarrow{d} 0 \quad \text{as } r \to \infty.$$

This, combined with (5.7), (5.8) and (5.10), gives for each $0 < a < b \le \infty$,

$$\big(X_b^r(\cdot) - X_b^r(0)\big) - \big(X_a^r(\cdot) - X_a^r(0)\big) + \left(\frac{\lambda}{b^p} - \frac{\lambda}{a^p}\right)(\cdot) \xrightarrow{d} 0 \quad \text{as } r \to \infty, \tag{5.13}$$

where λ/b^p is taken to be zero if $b = \infty$.

The above convergence together with (5.10) shows that, for each $i = 1, \ldots, k$

$$X_{a_i}^r(\cdot) = X_{a_i}^r(0) + \hat{V}_{a_i}^r(\cdot) + \left(\kappa - \frac{\lambda}{a_i^p}\right)\iota(\cdot) + \eta_i^r(\cdot),$$

where $\eta_i^r(\cdot) \xrightarrow{d} 0$ as $r \to \infty$ for each i. The result now follows on combining the above convergence with (5.11). □

PROOF OF THEOREM 1. Lemma 12, continuity of the Skorohod map Γ and the continuous mapping theorem, imply that for all $k \in \mathbb{N}$ and $0 \le a_1 < a_2 < \cdot < a_k \le \infty$, $(Y_{a_1}^r, Y_{a_2}^r, \ldots, Y_{a_k}^r) \xrightarrow{d} (W_{a_1}, W_{a_2}, \ldots, W_{a_k})$. The theorem follows from this, Proposition 10 and (4.11). □

5.4. *Proof of Theorem 2.* Before proceeding, the reader may wish to review the overview of the proof of Theorem 2 given in Section 1.1. We begin by establishing the result in Lemma 13 below as a elementary consequence of integration by parts. In what follows, for $0 \le \delta < M < \infty$, we will write "$\int_\delta^M$" to denote integration over the interval $(\delta, M]$. We will also write for any function $h : (\delta, M] \to \mathbb{R}$ and any $\delta \ge 0$, $h(\delta+) := \lim_{x \searrow \delta} h(x)$, whenever this limit exists.

LEMMA 13. *Suppose that $0 < \delta < M < \infty$ and $f : (\delta, M] \to \mathbb{R}$ is a C^1 function such that $f(\delta+)$ and $f'(\delta+)$ exist. Then, writing $g(x) = f(x)/x$ for $x \in (\delta, M]$, for any $r \in \mathcal{R}$ and $t \ge 0$, the following holds*:

$$\int_\delta^M f(x) \widetilde{\mathcal{Z}}^r(t)(dx) = -\int_\delta^M g'(x) W_x^r(t)\, dx + g(M) W_M^r(t) - g(\delta+) W_\delta^r(t).$$

PROOF. Fix $r \in \mathcal{R}$ and $t \ge 0$. Define the finite nonnegative Borel measure $\mu^r(t)$ on $\mathbb{R}_+$ by $\mu^r(t)(dx) := x \widetilde{\mathcal{Z}}^r(t)(dx)$ for $x \in \mathbb{R}_+$. Then, for $0 \le a < b$, $\mu^r(t)(a, b] = W_b^r(t) - W_a^r(t)$. Therefore,

$$\int_\delta^M f(x) \widetilde{\mathcal{Z}}^r(t)(dx) = \int_\delta^M g(x) \mu^r(t)(dx) = \int_\delta^M \left(\int_\delta^x g'(y)\, dy + g(\delta+)\right) \mu^r(t)(dx)$$

$$= \int_\delta^M \int_y^M \mu^r(t)(dx) g'(y)\, dy + g(\delta+) \mu^r(t)(\delta, M]$$

$$
\begin{aligned}
&= \int_{\delta}^{M} \mu^r(t)(y, M] g'(y)\, dy + g(\delta+)\mu^r(t)(\delta, M] \\
&= \int_{\delta}^{M} \big(W_M^r(t) - W_y^r(t)\big) g'(y)\, dy + g(\delta+)\big(W_M^r(t) - W_\delta^r(t)\big) \\
&= -\int_{\delta}^{M} W_y^r(t) g'(y)\, dy + W_M^r(t)\big(g(M) - g(\delta+)\big) \\
&\quad + g(\delta+)\big(W_M^r(t) - W_\delta^r(t)\big) \\
&= -\int_{\delta}^{M} g'(y) W_y^r(t)\, dy + g(M) W_M^r(t) - g(\delta+) W_\delta^r(t),
\end{aligned}
$$

which proves the lemma. □

Next, the result in Lemma 13, along with tightness arguments, is used to establish Theorem 14, which gives convergence in distribution to the desired limit for certain compactly supported functions with support bounded away from zero.

THEOREM 14. *Suppose that* $J \in \mathbb{N}$, $0 < a_1 < b_1 \le a_2 < b_2 \cdots \le a_J < b_J < \infty$, *and* $f : [0, \infty) \to \mathbb{R}$ *is a* C^1 *function on* $(a_j, b_j]$ *for each* $1 \le j \le J$ *and zero on* $(\bigcup_{j=1}^{J} (a_j, b_j])^c$. *Also, assume* $\lim_{x \searrow a_j} f(x)$ *and* $\lim_{x \searrow a_j} f'(x)$ *exist for each* $1 \le j \le J$. *Then, writing* $g(x) = f(x)/x$ *for* $x \in (0, \infty)$, *as* $r \to \infty$,

(5.14)
$$
\int_0^{\infty} f(x) \widetilde{\mathcal{Z}}^r(\cdot)(dx) \xrightarrow{d} \sum_{j=1}^{J} \Big(-\int_{a_j}^{b_j} g'(x) W_x(\cdot)\, dx + g(b_j) W_{b_j}(\cdot) - \lim_{x \searrow a_j} g(x) W_{a_j}(\cdot) \Big)
$$

in $\mathcal{D}([0, \infty) : \mathbb{R})$. *The limiting process defined by the right side of* (5.14), *in fact, has sample paths in* $\mathcal{C}([0, \infty) : \mathbb{R})$ *almost surely.*

REMARK 5. The proof of Theorem 2 will show that we can also take $a_1 = 0$ in Theorem 14. See Remark 7 for details.

PROOF OF THEOREM 14. We will prove the theorem for $J = 1$. The proof for $J \ge 2$ follows along the same lines (with more cumbersome notation) and is, therefore, omitted. We will write the interval $(a_1, b_1]$ as $(\delta, M]$ with $0 < \delta < M < \infty$. Assume f is not identically zero (otherwise the result is trivial).

Proof of Tightness: We will use Aldous' tightness criterion stated in Section 4.4. Note that, for $r \in \mathcal{R}$ and $t \ge 0$,

$$
\Big| \int_{\delta}^{M} f(x) \widetilde{\mathcal{Z}}^r(t)(dx) \Big| \le \sup_{z \in [\delta, M]} |g(z)| \int_{\delta}^{M} x \widetilde{\mathcal{Z}}^r(t)(dx) = \sup_{z \in [\delta, M]} |g(z)| \big(W_M^r(t) - W_\delta^r(t)\big).
$$

By Theorem 1, $\{W_M^r(\cdot) - W_\delta^r(\cdot)\}_{r \in \mathcal{R}}$ is tight, which implies that $\{\int_\delta^M f(x) \widetilde{\mathcal{Z}}^r(t)(dx)\}_{r \in \mathcal{R}}$ is tight for each fixed $t \ge 0$. Thus, (A1) of Aldous' tightness criterion holds for $\{\int_\delta^M f(x) \widetilde{\mathcal{Z}}^r(\cdot)(dx)\}_{r \in \mathcal{R}}$.

Next we show that (A2) of Aldous' tightness criterion holds for the above sequence as well. Fix $T \in (0, \infty)$, $\eta \in (0, 1)$ and a stopping time τ that takes values in $[0, T]$. Then, by

Lemma 13, for $r \in \mathcal{R}$,

$$\begin{aligned}
&\left|\int_\delta^M f(x)\widetilde{\mathcal{Z}}^r(\tau+\eta)(dx) - \int_\delta^M f(x)\widetilde{\mathcal{Z}}^r(\tau)(dx)\right| \\
&\quad \le C_g\Big(\int_\delta^M \big|W_x^r(\tau+\eta) - W_x^r(\tau)\big|\,dx + \big|W_M^r(\tau+\eta) - W_M^r(\tau)\big| \\
&\qquad + \big|W_\delta^r(\tau+\eta) - W_\delta^r(\tau)\big|\Big),
\end{aligned} \tag{5.15}$$

where $C_g := (\sup_{z\in[\delta,M]}|g'(z)|) + |g(M)| + |g(\delta+)|$. By (5.3) in Proposition 10 and (4.2), for any $r \in \mathcal{R}$ and $x \in [\delta, M]$,

$$\begin{aligned}
\big|W_x^r(\tau+\eta) - W_x^r(\tau)\big| &\le \big|Y_x^r(\tau+\eta) - Y_x^r(\tau)\big| + \frac{xc^r}{r} \\
&\le 2\sup_{\tau\le s\le\tau+\eta}\big|X_x^r(s) - X_x^r(\tau)\big| + \frac{xc^r}{r}.
\end{aligned} \tag{5.16}$$

Thus, for $r \in \mathcal{R}$,

$$\begin{aligned}
\int_\delta^M \big|W_x^r(\tau+\eta) - W_x^r(\tau)\big|\,dx &\le \int_\delta^M \Big(\big|Y_x^r(\tau+\eta) - Y_x^r(\tau)\big| + \frac{xc^r}{r}\Big)dx \\
&\le 2\int_\delta^M \sup_{\tau\le s\le\tau+\eta}\big|X_x^r(s) - X_x^r(\tau)\big|\,dx + \frac{M^2c^r}{2r}.
\end{aligned} \tag{5.17}$$

Note that for $r \in \mathcal{R}$, $s \in [\tau, \tau+\eta]$ and $x \in \mathbb{R}_+$,

$$X_x^r(s) - X_x^r(\tau) = \hat{V}_x^r(s) - \hat{V}_x^r(\tau) + r\big(\rho_{xc^r}^r - 1\big)(s-\tau),$$

and hence

$$\sup_{\tau\le s\le\tau+\eta}\big|X_x^r(s) - X_x^r(\tau)\big| \le \sup_{\tau\le s\le\tau+\eta}\big|\hat{V}_x^r(s) - \hat{V}_x^r(\tau)\big| + \big|r\big(\rho_{xc^r}^r - 1\big)\big|\eta. \tag{5.18}$$

For each $r \in \mathcal{R}$, define a process $\hat{U}^r(\cdot)$ as follows:

$$\hat{U}^r(t) := \frac{1}{r}\sum_{i=1}^{\lfloor r^2 t\rfloor}\big(v_i\mathbf{1}_{[v_i>\delta c^r]} - \mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})\big), \quad \text{for } t \ge 0.$$

Note that for any $r \in \mathcal{R}$, $s \in [\tau, \tau+\eta]$ and $x \in [\delta, M]$,

$$\begin{aligned}
&\big|\hat{V}_x^r(s) - \hat{V}_x^r(\tau)\big| \\
&\quad \le \big|\hat{V}_\infty^r(s) - \hat{V}_\infty^r(\tau)\big| + \frac{1}{r}\sum_{i=E^r(r^2(\tau+\eta))+1}^{E^r(r^2(\tau+\eta))} v_i\mathbf{1}_{[v_i>\delta c^r]} + r\lambda^r\eta\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) \\
&\quad \le \big|\hat{V}_\infty^r(s) - \hat{V}_\infty^r(\tau)\big| + \big|\hat{U}^r\big(\overline{E}^r(\tau+\eta)\big) - \hat{U}^r\big(\overline{E}^r(\tau)\big)\big| \\
&\qquad + \frac{1}{r}\big(E^r\big(r^2(\tau+\eta)\big) - E^r\big(r^2\tau\big)\big)\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) + r\lambda^r\eta\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}).
\end{aligned} \tag{5.19}$$

By Proposition 6 as $r \to \infty$, $\hat{V}_\infty^r(\cdot) \xrightarrow{d} V^*(\cdot)$ in $\mathcal{D}([0, T+1] : \mathbb{R})$ for some Brownian motion V^* with zero drift and finite variance. Fix $\gamma \in (0, 1/2)$. Recall the notation $|f(t\#) - f(s\#)| < A$, from Section 1.3, for a RCLL function f, $0 \le s \le t \le \infty$ and $A > 0$. For $K > 0$, define the set

$$\Omega(K) := \big\{\big|V^*(t\#) - V^*(s\#)\big| < K\eta^\gamma \text{ for all } 0 \le s \le t \le T+1 \text{ with } t - s \le \eta\big\}.$$

Fix $\epsilon \in (0, 1/8)$. Since V^* is Holder continuous with exponent γ, there exists K_ϵ (not depending on η) large enough such that $\mathbb{P}(\Omega(K_\epsilon)) \geq 1 - \epsilon$. Since for any $K > 0$, the set

$$A(K) := \{f \in \mathcal{D}([0, T+1] : \mathbb{R}) : |f(t\#) - f(s\#)| < K\eta^\gamma \text{ for all } 0 \leq s \leq t \leq T+1 \text{ with } t - s \leq \eta\}$$

is nonempty and open in the Skorohod topology by [12], Chapter 3, Proposition 6.5, and $\hat{V}^r_\infty(\cdot) \xrightarrow{d} V^*(\cdot)$ as $r \to \infty$, the Portmanteau theorem implies that there exists $r_0 > 0$ such that for all $r \geq r_0$,

$$\mathbb{P}(\hat{V}^r_\infty(\cdot) \in A(K_\epsilon)) \geq 1 - 2\epsilon,$$

and consequently, for all $r \geq r_0$,

$$\mathbb{P}\Big(\sup_{\tau \leq s \leq \tau+\eta} |\hat{V}^r_\infty(s) - \hat{V}^r_\infty(\tau)| \geq K_\epsilon \eta^\gamma\Big) \leq 2\epsilon. \tag{5.20}$$

Recall that $\overline{E}^r(\cdot) \xrightarrow{d} \lambda(\cdot)$, where $\lambda(t) = \lambda t$ for $t \geq 0$, and by Proposition 6, $\hat{U}^r(\cdot) \xrightarrow{d} 0$ as $r \to \infty$. Therefore, as $r \to \infty$, $\hat{U}^r(\overline{E}^r(\cdot)) \xrightarrow{d} 0$ and consequently, there exists $r_1 \geq r_0$ such that for $r \geq r_1$,

$$\mathbb{P}(|\hat{U}^r(\overline{E}^r(\tau+\eta)) - \hat{U}^r(\overline{E}^r(\tau))| > \eta^\gamma) \leq 2\mathbb{P}\Big(\sup_{t \in [0, T+1]} |\hat{U}^r(\overline{E}^r(t))| > \eta^\gamma/2\Big) < \epsilon. \tag{5.21}$$

Now, using the fact that $r\mathbb{E}[v\mathbf{1}_{[v>c^r]}] = 1$ due to (2.6), (2.8) and (2.9), we write the sum of the third and the fourth terms on the right side of (5.19) as

$$\begin{aligned}
&\frac{1}{r}\big(E^r(r^2(\tau+\eta)) - E^r(r^2\tau)\big)\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) + r\lambda^r\eta\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) \\
&\quad = \frac{E^r(r^2(\tau+\eta)) - E^r(r^2\tau)}{r^2}\frac{\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} + \lambda^r\eta\frac{\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} \\
&\quad = \big(\overline{E}^r(\tau+\eta) - \overline{E}^r(\tau)\big)\frac{\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} + \lambda^r\eta\frac{\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})}.
\end{aligned} \tag{5.22}$$

As the set

$$\Omega^* := \{f \in \mathcal{D}([0, T+1] : \mathbb{R}) : |f(t\#) - f(s\#)| < 2\lambda\eta \text{ for all } 0 \leq s \leq t \leq T+1 \text{ with } t - s \leq \eta\}$$

is nonempty and open in the Skorohod topology and $\overline{E}^r(\cdot) \xrightarrow{d} \lambda(\cdot)$ as $r \to \infty$, there exists $r_2 \geq r_1$ such that for all $r \geq r_2$,

$$\mathbb{P}(\overline{E}^r(\tau+\eta) - \overline{E}^r(\tau) \geq 2\lambda\eta) < \epsilon.$$

Moreover, $\lambda^r \to \lambda$ as $r \to \infty$ and (4.6) implies

$$\lim_{r\to\infty} \frac{\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} = \frac{1}{\delta^p}.$$

Using these observations in (5.22) gives that there is an $r_3 \geq r_2$ such that for all $r \geq r_3$,

$$\mathbb{P}\Big(\frac{1}{r}\big(E^r(r^2(\tau+\eta)) - E^r(r^2\tau)\big)\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) + r\lambda^r\eta\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) > \frac{8\lambda\eta}{\delta^p}\Big) < \epsilon. \tag{5.23}$$

Using (5.19), (5.20), (5.21) and (5.23), we obtain for $r \geq r_3$,

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[\delta,M]} \sup_{\tau\leq s\leq \tau+\eta} |\hat{V}_x^r(s) - \hat{V}_x^r(\tau)| > \Big(K_\epsilon + 1 + \frac{8\lambda}{\delta^p}\Big)\eta^\gamma \Big) \\
&\quad \leq \mathbb{P}\Big(\sup_{\tau\leq s\leq \tau+\eta} |\hat{V}_\infty^r(s) - \hat{V}_\infty^r(\tau)| > K_\epsilon \eta^\gamma \Big) \\
&\quad\quad + \mathbb{P}\big(|\hat{U}^r(\overline{E}^r(\tau+\eta)) - \hat{U}^r(\overline{E}^r(\tau))| > \eta^\gamma \big) \\
&\quad\quad + \mathbb{P}\Big(\frac{1}{r}\big(E^r(r^2(\tau+\eta)) - E^r(r^2\tau)\big)\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) + r\lambda^r\eta\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]}) > \frac{8\lambda\eta}{\delta^p} \Big) \\
&\quad < 4\epsilon.
\end{aligned} \tag{5.24}
$$

Moreover, by (5.9) and the uniform convergence in (4.7), $r(\rho_{xc^r}^r - 1) \to \kappa - \frac{\lambda}{x^p}$ as $r \to \infty$ uniformly for $x \in [\delta, \infty)$. Thus, there exists $C_1 > 0$ and $r_4 \geq r_3$ such that for all $r \geq r_4$,

$$
\sup_{x\in[\delta,M]} |r(\rho_{xc^r}^r - 1)| \leq C_1. \tag{5.25}
$$

Using (5.18), (5.24) and (5.25), for some $C_2 \in (0, \infty)$ and all $r \geq r_4$,

$$
\mathbb{P}\Big(\sup_{x\in[\delta,M]} \sup_{\tau\leq s\leq \tau+\eta} |X_x^r(s) - X_x^r(\tau)| > \Big(K_\epsilon + 1 + \frac{8\lambda}{\delta^p} + C_2\Big)\eta^\gamma \Big) < 4\epsilon. \tag{5.26}
$$

Take $r_5 \geq r_4$ such that $\max\{M^2c^r/(2r), Mc^r/r\} < \eta^\gamma$ for all $r \geq r_5$ and define $C_3 := 2(M - \delta)(K_\epsilon + 1 + \frac{8\lambda}{\delta^p} + C_2) + 1$. Then, using (5.17) and (5.26), we obtain, for all $r \geq r_5$,

$$
\mathbb{P}\Big(\int_\delta^M |W_x^r(\tau+\eta) - W_x^r(\tau)|\, dx > C_3\eta^\gamma \Big) < 4\epsilon. \tag{5.27}
$$

Similarly, using (5.16) and (5.26) and writing $C_4 := 2(K_\epsilon + 1 + \frac{8\lambda}{\delta^p} + C_2) + 1$, for $r \geq r_5$, we can show that

$$
\mathbb{P}\big(|W_M^r(\tau+\eta) - W_M^r(\tau)| + |W_\delta^r(\tau+\eta) - W_\delta^r(\tau)| > C_4\eta^\gamma \big) < 4\epsilon. \tag{5.28}
$$

Finally, using (5.15), (5.27) and (5.28), and the fact that T, η, ϵ and τ were arbitrary, we conclude that for any $T > 0$, $\eta \in (0, 1)$, $\epsilon \in (0, 1/8)$, and stopping time τ taking values in $[0, T]$, there exists $C^* > 0$ and $r^* > 0$ such that for any $r \geq r^*$,

$$
\mathbb{P}\Big(\Big| \int_\delta^M f(x)\widetilde{\mathcal{Z}}^r(\tau+\eta)(dx) - \int_\delta^M f(x)\widetilde{\mathcal{Z}}^r(\tau)(dx) \Big| > C^*\eta^\gamma \Big) < 8\epsilon. \tag{5.29}
$$

For instance, $C^* = C_g(C_3 + C_4)$ and $r^* = r_5$. Equation (5.29) implies that condition (A2) of Aldous' tightness criterion also holds. Thus, $\{\int_\delta^M f(x)\widetilde{\mathcal{Z}}^r(\cdot)(dx)\}_{r\in\mathcal{R}}$ is tight in $\mathcal{D}([0, T] : \mathbb{R})$ by Aldous' tightness criterion.

Proof of finite dimensional joint convergence: For $r \in \mathcal{R}$ and $t \geq 0$, write

$$
\begin{aligned}
\Psi^r(t) &:= -\int_\delta^M g'(x)W_x^r(t)\,dx + g(M)W_M^r(t) - g(\delta+)W_\delta^r(t), \\
\Psi(t) &:= -\int_\delta^M g'(x)W_x(t)\,dx + g(M)W_M(t) - g(\delta+)W_\delta(t).
\end{aligned}
$$

Fix $k \in \mathbb{N}$, $T > 0$, and $0 \leq t_1 < \cdots < t_k \leq T$. We will use Lemma 7 and Proposition 10 to show that

$$
\mathbf{A}^r := \big(\Psi^r(t_1), \ldots, \Psi^r(t_k)\big) \xrightarrow{d} \mathbf{A} := \big(\Psi(t_1), \ldots, \Psi(t_k)\big) \tag{5.30}
$$

as $r \to \infty$. For this, for each $n \in \mathbb{N}$, let $\delta = x_0 < x_1 < \cdots < x_{K_n} = M$ be a partition of mesh n^{-1}. For $r \in \mathcal{R}$, $n \in \mathbb{N}$, and $t \geq 0$, define

$$\Psi_n^r(t) := \sum_{j=0}^{K_n-1} W_{x_j}^r(t)\big(g(x_j) - g(x_{j+1})\big) + g(M)W_M^r(t) - g(\delta+)W_\delta^r(t),$$

$$\Psi_n(t) := \sum_{j=0}^{K_n-1} W_{x_j}(t)\big(g(x_j) - g(x_{j+1})\big) + g(M)W_M(t) - g(\delta+)W_\delta(t).$$

Observe that for each $n \in \mathbb{N}$, by Theorem 1 and the continuous mapping theorem,

$$\Psi_n^r(\cdot) \xrightarrow{d} \Psi_n(\cdot) \quad \text{in } \mathcal{D}\big([0,T]:\mathbb{R}\big) \text{ as } r \to \infty. \tag{5.31}$$

By (5.31), for each $n \in \mathbb{N}$,

$$\mathbf{A}_n^r := \big(\Psi_n^r(t_1), \ldots, \Psi_n^r(t_k)\big) \xrightarrow{d} \mathbf{A}_n := \big(\Psi_n(t_1), \ldots, \Psi_n(t_k)\big) \quad \text{as } r \to \infty. \tag{5.32}$$

For each $r \in \mathcal{R}$, $n \in \mathbb{N}$ and $t \geq 0$, note that

$$\big|\Psi_n^r(t) - \Psi^r(t)\big| \leq \sum_{j=0}^{K_n-1} \int_{x_j}^{x_{j+1}} \big|g'(x)\big|\big(W_x^r(t) - W_{x_j}^r(t)\big)\,dx, \tag{5.33}$$

$$\big|\Psi_n(t) - \Psi(t)\big| \leq \sum_{j=0}^{K_n-1} \int_{x_j}^{x_{j+1}} \big|g'(x)\big|\big(W_x(t) - W_{x_j}(t)\big)\,dx. \tag{5.34}$$

By (5.33), (5.3) in Proposition 10 and the Lipschitz property (4.1) of the Skorohod map Γ, for any $r \in \mathcal{R}$, $n \in \mathbb{N}$ and $t \in [0,T]$,

$$\begin{aligned}
&\big|\Psi_n^r(t) - \Psi^r(t)\big| \\
&\leq \sum_{j=0}^{K_n-1} \int_{x_j}^{x_{j+1}} \big|g'(x)\big|\left(\big|Y_x^r(t) - Y_{x_j}^r(t)\big| + \frac{xc^r}{r}\right)dx \\
&\leq \frac{Mc^r}{r}\int_\delta^M \big|g'(x)\big|\,dx + 2\sum_{j=0}^{K_n-1} \int_{x_j}^{x_{j+1}} \big|g'(x)\big|\Big(\sup_{s\in[0,T]}\big|X_x^r(s) - X_{x_j}^r(s)\big|\Big)\,dx.
\end{aligned} \tag{5.35}$$

Now, for any $0 \leq j \leq K_n - 1$ and any $x \in [x_j, x_{j+1}]$,

$$\begin{aligned}
\sup_{s\in[0,T]}\big|X_x^r(s) - X_{x_j}^r(s)\big| &\leq \frac{1}{r}\sum_{l=1}^{\mathfrak{q}^r} \breve{v}_l^r \mathbf{1}_{[x_j c^r < \breve{v}_l^r \leq x_{j+1}c^r]} \\
&\quad + \frac{1}{r}\sum_{i=1}^{E^r(r^2T)} v_i \mathbf{1}_{[x_j c^r < v_i \leq x_{j+1}c^r]}.
\end{aligned} \tag{5.36}$$

Hence, by (5.35) and (5.36), for each $r \in \mathcal{R}$ and $n \in \mathbb{N}$,

$$\sup_{0\leq t\leq T}\big|\Psi_n^r(t) - \Psi^r(t)\big| \leq \Delta_{n,1}^r + \Delta_{n,2}^r, \tag{5.37}$$

where

$$\Delta_{n,1}^r := \frac{Mc^r}{r}\int_\delta^M \big|g'(x)\big|\,dx + 2\sum_{j=0}^{K_n-1}\int_{x_j}^{x_{j+1}}\big|g'(x)\big|\,dx\left(\frac{1}{r}\sum_{i=1}^{E^r(r^2T)} v_i \mathbf{1}_{[x_j c^r < v_i \leq x_{j+1}c^r]}\right),$$

$$\Delta_{n,2}^r := 2\sum_{j=0}^{K_n-1}\int_{x_j}^{x_{j+1}}\big|g'(x)\big|\,dx\left(\frac{1}{r}\sum_{l=1}^{\mathfrak{q}^r} \breve{v}_l^r \mathbf{1}_{[x_j c^r < \breve{v}_l^r \leq x_{j+1}c^r]}\right).$$

Observe that there exists $C > 0$ such that for all $r \in \mathcal{R}$, $n \in \mathbb{N}$ and $0 \le j \le K_n$,

$$\mathbb{E}\left(\frac{1}{r}\sum_{i=1}^{E^r(r^2T)} v_i \mathbf{1}_{[x_j c^r < v_i \le x_{j+1} c^r]}\right) \le CrT\mathbb{E}(v\mathbf{1}_{[x_j c^r < v \le x_{j+1} c^r]}). \tag{5.38}$$

Recalling that $r\mathbb{E}(v\mathbf{1}_{[v>c^r]}) = r/S(c^r) = 1$ for each $r \in \mathcal{R}$, we can write, for $r \in \mathcal{R}$ and $0 \le j \le K_n$,

$$\begin{aligned}\mathbb{E}(v\mathbf{1}_{[x_j c^r < v \le x_{j+1} c^r]}) &= \mathbb{E}(v\mathbf{1}_{[v>x_j c^r]}) - \mathbb{E}(v\mathbf{1}_{[v>x_{j+1} c^r]}) \\ &= \frac{1}{r}\left(\frac{\mathbb{E}(v\mathbf{1}_{[v>x_j c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} - \frac{\mathbb{E}(v\mathbf{1}_{[v>x_{j+1} c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})}\right).\end{aligned} \tag{5.39}$$

From the uniform convergence in (4.7), for each $n \in \mathbb{N}$, there exists $r_1(n) > 0$ such that for all $r \ge r_1(n)$,

$$\left|\frac{\mathbb{E}(v\mathbf{1}_{[v>uc^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} - \frac{1}{u^p}\right| < \frac{1}{n} \quad \text{for all } u \in [\delta, \infty). \tag{5.40}$$

By combining (5.38), (5.39) and (5.40), for any $n \in \mathbb{N}$, $r \ge r_1(n)$ and any $0 \le j \le K_n - 1$,

$$\begin{aligned}\mathbb{E}\left(\frac{1}{r}\sum_{i=1}^{E^r(r^2T)} v_i \mathbf{1}_{[x_j c^r < v_i \le x_{j+1} c^r]}\right) &\le CT\left|\frac{\mathbb{E}(v\mathbf{1}_{[v>x_j c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} - \frac{1}{x_j^p}\right| \\ &\quad + CT\left|\frac{\mathbb{E}(v\mathbf{1}_{[v>x_{j+1} c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} - \frac{1}{x_{j+1}^p}\right| \\ &\quad + CT\left(\frac{1}{x_j^p} - \frac{1}{x_{j+1}^p}\right) \le \frac{2CT}{n} + \frac{CTp}{\delta^{p+1}n}.\end{aligned}$$

Thus, for $n \in \mathbb{N}$ and $r \ge r_1(n)$,

$$\mathbb{E}[\Delta_{n,1}^r] \le \frac{Mc^r}{r}\int_\delta^M |g'(x)|\,dx + 2\left(2CT + \frac{CTp}{\delta^{p+1}}\right)\frac{1}{n}\int_\delta^M |g'(x)|\,dx. \tag{5.41}$$

Fix $\epsilon \in (0,1)$. Choose $n_\epsilon \in \mathbb{N}$ such that $2(2CT + \frac{CTp}{\delta^{p+1}})\frac{1}{n_\epsilon}\int_\delta^M |g'(x)|\,dx < \epsilon^2/(4\sqrt{k})$ and

$$\mathbb{P}\left(\sup_{\{\delta<y,z\le M:|z-y|\le n_\epsilon^{-1}\}} |\xi(z) - \xi(y)| < \frac{\epsilon}{4\sqrt{k}\int_\delta^M |g'(x)|\,dx}\right) \ge 1 - \epsilon/4,$$

which can be ensured to exist since $\xi(\cdot)$ is continuous and $[\delta, M]$ is compact. Noting that

$$\begin{aligned}S^{(\epsilon)} := \Bigg\{ f \in \mathcal{D}([\delta, M] : \mathbb{R}) : |f(z\#) - f(y\#)| &< \frac{\epsilon}{4\sqrt{k}\int_\delta^M |g'(x)|\,dx} \\ &\forall \delta \le y, z \le M \text{ with } |z - y| \le n_\epsilon^{-1}\Bigg\}\end{aligned}$$

is nonempty and open in the Skorohod topology and by assumption (2.14), we obtain $r_2 \ge r_1(n_\epsilon)$ such that for all $r \ge r_2$,

$$\mathbb{P}\left(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \breve{v}_l^r \mathbf{1}_{[x_j c^r < \breve{v}_l^r \le x_{j+1} c^r]} \ge \frac{\epsilon}{4\sqrt{k}\int_\delta^M |g'(x)|\,dx} \text{ for some } 0 \le j \le K_{n_\epsilon} - 1\right) < \epsilon/2. \tag{5.42}$$

Using (5.37), (5.41), (5.42) and the choice of n_ϵ, we conclude that for $r \geq r_2$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} \big|\Psi^r_{n_\epsilon}(t) - \Psi^r(t)\big| > \frac{\epsilon}{\sqrt{k}}\Big) \\
&\quad \leq \mathbb{P}\Big(\Delta^r_{n_\epsilon,1}(t) > \frac{\epsilon}{2\sqrt{k}}\Big) + \mathbb{P}\Big(\Delta^r_{n_\epsilon,2} > \frac{\epsilon}{2\sqrt{k}}\Big) \\
&\quad \leq \frac{2\sqrt{k}Mc^r}{\epsilon r} \int_\delta^M \big|g'(x)\big|\,dx + \frac{4\sqrt{k}}{\epsilon}\Big(2CT + \frac{CTp}{\delta^{p+1}}\Big)\frac{1}{n_\epsilon}\int_\delta^M \big|g'(x)\big|\,dx + \frac{\epsilon}{2} \\
&\quad \leq \frac{2\sqrt{k}Mc^r}{\epsilon r} \int_\delta^M \big|g'(x)\big|\,dx + \epsilon.
\end{aligned}$$

Therefore,

$$\limsup_r \mathbb{P}\big(\big\|\mathbf{A}^r_{n_\epsilon} - \mathbf{A}^r\big\|_2 > \epsilon\big) \leq \limsup_r \mathbb{P}\Big(\sqrt{k} \sup_{t\in[0,T]} \big|\Psi^r_{n_\epsilon}(t) - \Psi^r(t)\big| > \epsilon\Big) \leq \epsilon, \tag{5.43}$$

where $\|\cdot\|_2$ denotes the L^2-norm in $\mathbb{R}^k$. Thus, condition (1) of Lemma 7 holds with r in place of m, $S^\epsilon_r = \mathbf{A}^r_{n_\epsilon}$, $S_r = \mathbf{A}^r$ and $b(\epsilon) = 2\epsilon$. Condition (2) of Lemma 7 with $S^\epsilon = \mathbf{A}_{n_\epsilon}$ follows from (5.32).

Next, recall $W_a(\cdot) = \Gamma[X_a](\cdot)$ and for $\theta > 0$, $b > a > 0$, write

$$\omega\big(\xi, \theta; [a,b]\big) := \sup\big\{\big|\xi(y) - \xi(x)\big| : a \leq x, y \leq b, |y - x| \leq \theta\big\}.$$

By the continuity of $\xi(\cdot)$, for any fixed a, b, $\lim_{\theta\to 0}\omega(\xi,\theta;[a,b]) = 0$ almost surely. Using this observation along with (5.34), the Lipschitz property (4.1) of the Skorohod map and (3.3)–(3.4), for each $n \in \mathbb{N}$,

$$\begin{aligned}
&\sup_{t\in[0,T]} \big|\Psi_n(t) - \Psi(t)\big| \\
&\quad \leq 2T\lambda \sum_{j=0}^{K_n-1} \int_{x_j}^{x_{j+1}} \big|g'(x)\big|\big(x_j^{-p} - x^{-p}\big)\,dx + 2\omega\big(\xi, n^{-1}; [\delta, M]\big)\int_\delta^M \big|g'(x)\big|\,dx \\
&\quad \leq 2T\lambda \sum_{j=0}^{K_n-1} \int_{x_j}^{x_{j+1}} \big|g'(x)\big|\frac{p(x - x_j)}{x_j^{p+1}}\,dx + 2\omega\big(\xi, n^{-1}; [\delta, M]\big)\int_\delta^M \big|g'(x)\big|\,dx \\
&\quad \leq \frac{2T\lambda p}{\delta^{p+1}n}\int_\delta^M \big|g'(x)\big|\,dx + 2\omega\big(\xi, n^{-1}; [\delta, M]\big)\int_\delta^M \big|g'(x)\big|\,dx.
\end{aligned}$$

Thus, $\sup_{t\in[0,T]}|\Psi_n(t) - \Psi(t)| \to 0$ almost surely as $n \to \infty$, which implies that

$$\|\mathbf{A}_n - \mathbf{A}\| \leq \sqrt{k} \sup_{t\in[0,T]} \big|\Psi_n(t) - \Psi(t)\big| \to 0 \quad \text{almost surely as } n \to \infty. \tag{5.44}$$

Thus, condition (3) of Lemma 7 holds with $S^\epsilon = \mathbf{A}_{n_\epsilon}$ and $S^0 = \mathbf{A}$. The weak convergence in (5.30) now follows from (5.32), (5.43), (5.44) and Lemma 7.

This completes the proof of the convergence claimed in the theorem. The continuity of the limiting process in the theorem follows from that of Brownian motion and (4.2). □

To prove Theorem 2 as a consequence of Theorem 14, we need the following result.

LEMMA 15. *For $0 < \delta < \infty$, consider any C^1 function $f : [\delta, \infty) \to \mathbb{R}$ such that $\lim_{x\to\infty}\frac{f(x)}{x}$ exists and $\int_1^\infty \frac{|f'(x)|}{x^{\alpha^*+1}} < \infty$, where α^* is the constant appearing in Assumption*

(2.17). *Then, writing* $g(x) = f(x)/x$ *for* $x \in [\delta, \infty)$ *and* $g(\infty) = \lim_{x\to\infty} g(x)$, *the following distributional convergence holds in* $\mathcal{D}([0,\infty) : \mathbb{R})$:

$$\int_\delta^\infty f(x)\widetilde{\mathcal{Z}}^r(\cdot)(dx) \xrightarrow{d} -\int_\delta^\infty g'(x)W_x(\cdot)\,dx + g(\infty)W_\infty(\cdot) - g(\delta)W_\delta(\cdot), \tag{5.45}$$

as $r \to \infty$, *where the right side of* (5.45) *defines a stochastic process with sample paths in* $\mathcal{C}([0,\infty) : \mathbb{R})$.

PROOF. Fix $T, \delta > 0$. For $\delta < M < \infty$ and $t \geq 0$, define the following:

$$\Phi_M^r(t) := \int_\delta^M f(x)\widetilde{\mathcal{Z}}^r(t)(dx), \qquad \Phi_\infty^r(t) := \int_\delta^\infty f(x)\widetilde{\mathcal{Z}}^r(t)(dx),$$

and

$$\Upsilon_M(t) := -\int_\delta^M g'(x)W_x(t)\,dx + g(M)W_M(t) - g(\delta)W_\delta(t).$$

We first show that on the time interval $[0,T]$, almost surely, $\int_\delta^M g'(x)W_x(\cdot)\,dx$ converges uniformly (with respect to time) on the time interval $[0,T]$ as $M \to \infty$, and hence, the limit $\int_\delta^\infty g'(x)W_x(\cdot)\,dx$ is well defined and continuous on $[0,T]$. Note that for any $M' > M > \delta$,

$$\begin{aligned}\sup_{t\in[0,T]}\left|\int_M^{M'} g'(x)W_x(t)\,dx\right| &\leq \sup_{t\in[0,T]}\left|\int_M^{M'} g'(x)\big(W_\infty(t) - W_x(t)\big)\,dx\right| \\ &\quad + \sup_{t\in[0,T]}\left|\int_M^{M'} g'(x)W_\infty(t)\,dx\right|.\end{aligned}$$

By (3.4), (3.5), the Lipschitz property (4.1) of the Skorohod map and recalling that $\xi(\infty) := \lim_{u\to\infty}\xi(u) < \infty$ almost surely by assumption, for any $M' > M > \delta$,

$$\begin{aligned}&\sup_{t\in[0,T]}\left|\int_M^{M'} g'(x)\big(W_\infty(t) - W_x(t)\big)\,dx\right| \\ &\quad\leq 2T\lambda\int_M^\infty |g'(x)|x^{-p}\,dx + 2\int_M^\infty |g'(x)|\big(\xi(\infty) - \xi(x)\big)\,dx.\end{aligned} \tag{5.46}$$

By Assumption (2.17), $0 < \alpha^* \leq p$, and so, by the assumptions on f in the theorem,

$$\int_1^\infty |g'(x)|x^{-p}\,dx \leq \int_1^\infty \left(\frac{|f(x)|}{x^{p+2}} + \frac{|f'(x)|}{x^{p+1}}\right)dx \leq \int_1^\infty \left(\frac{|f(x)|}{x^{\alpha^*+2}} + \frac{|f'(x)|}{x^{\alpha^*+1}}\right)dx < \infty.$$

Also, by Assumption (2.17), there exists $C > 0$ such that

$$\mathbb{E}\big(\xi(\infty) - \xi(x)\big) \leq Cx^{-\alpha^*} \quad \text{for all } x \geq 1. \tag{5.47}$$

Hence, by Fubini's theorem,

$$\mathbb{E}\left(\int_1^\infty |g'(x)|\big(\xi(\infty) - \xi(x)\big)\,dx\right) \leq C\int_1^\infty \left(\frac{|f(x)|}{x^2} + \frac{|f'(x)|}{x}\right)x^{-\alpha^*}\,dx < \infty.$$

Moreover, for any $M' > M > \delta$,

$$\sup_{t\in[0,T]}\left|\int_M^{M'} g'(x)W_\infty(t)\,dx\right| = \sup_{t\in[0,T]} W_\infty(t)\big|g(M') - g(M)\big|. \tag{5.48}$$

Hence, (5.46), (5.48) and the finiteness of $\lim_{x\to\infty} g(x) = g(\infty)$ imply that, almost surely, the sequence $\{\int_\delta^{M_n} g'(x)W_x(\cdot)\,dx\}_{n=1}^\infty$ is uniformly Cauchy on $[0,T]$ for any sequence $\{M_n\}_{n=1}^\infty$ such that $\lim_{n\to\infty} M_n = \infty$, which proves the uniform convergence to the limit

$\int_\delta^\infty g'(x)W_x(\cdot)\,dx$ as $M\to\infty$. Moreover, by (3.4), (3.5) and the Lipschitz property (4.1) of the Skorohod map, for any $M>\delta$,

$$\begin{aligned}
&\sup_{t\in[0,T]}\big|g(M)W_M(t)-g(\infty)W_\infty(t)\big|\\
&\quad\le |g(M)|\sup_{t\in[0,T]}|W_M(t)-W_\infty(t)|+\Big(\sup_{t\in[0,T]}|W_\infty(t)|\Big)|g(\infty)-g(M)|\\
&\quad\le 2\Big(\sup_{x\ge\delta}|g(x)|\Big)T\lambda M^{-p}+2\Big(\sup_{x\ge\delta}|g(x)|\Big)(\xi(\infty)-\xi(M))\\
&\qquad+\Big(\sup_{t\in[0,T]}|W_\infty(t)|\Big)|g(\infty)-g(M)|.
\end{aligned}$$

The upper bound in the display immediately above tends to zero as $M\to\infty$. Thus, we conclude that

$$\Upsilon_\infty(t):=-\int_\delta^\infty g'(x)W_x(t)\,dx+g(\infty)W_\infty(t)-g(\delta)W_\delta(t),\quad t\in[0,T],$$

is well defined and $\Upsilon_\infty:[0,T]\to\mathbb{R}$ is continuous and, almost surely,

$$\lim_{M\to\infty}\sup_{t\in[0,T]}\big|\Upsilon_M(t)-\Upsilon_\infty(t)\big|=0. \tag{5.49}$$

By Theorem 14, $\Phi^r_M(\cdot)\to\Upsilon_M(\cdot)$ in $\mathcal{D}([0,T]:\mathbb{R})$ as $r\to\infty$ for each $M>\delta$. This together with (5.49) implies that conditions (2) and (3) of Lemma 7 hold. Thus, in order to show that $\Phi^r_\infty(\cdot)\to\Upsilon_\infty(\cdot)$ in $\mathcal{D}([0,T]:\mathbb{R})$ as $r\to\infty$, it suffices to show that condition (1) of Lemma 7 holds. For this, observe that as $\lim_{x\to\infty}\frac{f(x)}{x}$ exists, there exists a constant $C'>0$ such that $|f(x)|\le C'x$ for all $x\ge\delta$. Hence, for all $r\in\mathcal{R}$, $t\in[0,T]$ and $M>\delta$,

$$\begin{aligned}
|\Phi^r_\infty(t)-\Phi^r_M(t)|&\le\int_M^\infty|f(x)|\widetilde{\mathcal{Z}}^r(t)(dx)\le C'\int_M^\infty x\,\widetilde{\mathcal{Z}}^r(t)(dx)\\
&=\frac{C'}{r}\sum_{l=1}^{\mathbf{q}^r}\check{v}^r_l(r^2t)\mathbf{1}_{[\check{v}^r_l(r^2t)>Mc^r]}+\frac{C'}{r}\sum_{i=1}^{E^r(r^2t)}v_i(r^2t)\mathbf{1}_{[v_i(r^2t)>Mc^r]}\\
&\le\frac{C'}{r}\sum_{l=1}^{\mathbf{q}^r}\check{v}^r_l\mathbf{1}_{[\check{v}^r_l>Mc^r]}+\frac{C'}{r}\sum_{i=1}^{E^r(r^2T)}v_i\mathbf{1}_{[v_i>Mc^r]}.
\end{aligned}$$

Thus, due to (2.14) and (2.15) (in particular, the limits displayed after (2.15)), the independence of $E^r(\cdot)$ and $\{v_i\}_{i\in\mathbb{N}}$ for each $r\in\mathcal{R}$, Wald's lemma, and (2.5), there exists $C''>0$ such that for all $M>\delta$,

$$\begin{aligned}
&\limsup_{r\to\infty}\mathbb{E}\Big(\sup_{t\in[0,T]}\big|\Phi^r_\infty(t)-\Phi^r_M(t)\big|\Big)\\
&\quad\le C'\limsup_{r\to\infty}\mathbb{E}\left(\frac{1}{r}\sum_{l=0}^{\mathbf{q}^r}\check{v}^r_l\mathbf{1}_{[\check{v}^r_l>Mc^r]}\right)+\limsup_{r\to\infty}\frac{C'}{r}\mathbb{E}\left(\sum_{i=1}^{E^r(r^2T)}v_i\mathbf{1}_{[v_i>Mc^r]}\right)\\
&\quad=C'\limsup_{r\to\infty}\big(W^r_\infty(0)-W^r_M(0)\big)+\limsup_{r\to\infty}\frac{C'}{r}\mathbb{E}\left(\sum_{i=1}^{E^r(r^2T)}v_i\mathbf{1}_{[v_i>Mc^r]}\right)\\
&\quad\le C'\mathbb{E}\big(\xi(\infty)-\xi(M)\big)+\limsup_{r\to\infty}C''rT\mathbb{E}(v\mathbf{1}_{[v>Mc^r]}).
\end{aligned}$$

Thus, for all $M > \delta$,

$$\limsup_{r\to\infty} \mathbb{E}\Big(\sup_{t\in[0,T]} \big|\Phi^r_\infty(t) - \Phi^r_M(t)\big|\Big)$$
$$\le C'\mathbb{E}\big(\xi(\infty) - \xi(M)\big) + C''T \limsup_{r\to\infty} \frac{\mathbb{E}(v\mathbf{1}_{[v>Mc^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} \le \frac{CC'}{M^{\alpha^*}} + \frac{C''T}{M^p},$$

where we have used (2.10) in the first inequality, and (5.47) and (4.6) in the second inequality. From this bound it follows from Markov's inequality that, for all $M > \delta$,

$$\limsup_{r\to\infty} \mathbb{P}\Big(\sup_{t\in[0,T]} \big|\Phi^r_\infty(t) - \Phi^r_M(t)\big| > \frac{1}{M^{\alpha^*/2}}\Big) \le \frac{CC'}{M^{\alpha^*/2}} + \frac{C''T}{M^{p-\alpha^*/2}}. \tag{5.50}$$

As previously noted, by Theorem 14, $\Phi^r_M(\cdot) \xrightarrow{d} \Upsilon_M(\cdot)$ as $r \to \infty$ in $\mathcal{D}([0,T]:\mathbb{R})$ for each $M > \delta$. This, along with (5.49), (5.50) and Lemma 7, proves that $\Phi^r_\infty(\cdot) \xrightarrow{d} \Upsilon_\infty(\cdot)$ as $r \to \infty$ in $\mathcal{D}([0,T]:\mathbb{R})$ and that $\Upsilon_\infty(\cdot)$ is continuous in $[0,T]$, which proves the lemma since $T > 0$ was arbitrary. □

5.4.1. *Sending* $\delta \to 0$. Next we show that the result in Lemma 15 holds for $\delta = 0$. The strategy involved is again to use Lemma 7 to send $\delta \to 0$ in (5.45). In particular, by (5.45) in Lemma 15, condition (2) of Lemma 7 holds. Hence, we can apply Lemma 7 after we have shown that the left-hand side of (5.45) is close to $\int_0^\infty f(x)\widetilde{\mathcal{Z}}^r(\cdot)\,dx$ in a uniform sense as required to verify condition (1) of Lemma 7, and the right-hand side of (5.45) converges to the appropriate limit as $\delta \to 0$ to verify condition (3) of Lemma 7. However, showing that these two conditions hold becomes quite technical. To control the left-hand side of (5.45), we first show that for any $a > 0$ (not depending on r), the maximum number of jobs in the rth $a(c^r)^{-1}$-truncated queue in the time interval $[0,T]$ is small (Lemma 16) by performing an excursion analysis of the workload process. However, the estimates obtained by such an analysis turn out to be too crude to show that the number of jobs of size $\le \delta c^r$ is uniformly small on the time interval $[0,T]$ for small δ. For this, we need much more involved analysis making careful use of the SRPT dynamics. Roughly, we show that the workload process corresponding to jobs of size in the interval $[a, \delta c^r]$ can be bounded above by a (reflected) martingale with large negative drift, quantified in (5.73). We then decompose the workload process path into excursions between appropriately chosen levels and control these excursions using the upper bounding process to bound the maximum on $[0,T]$. Finally, bounding the queue-length process by a (sufficiently large) multiple of the workload process (see (5.6)), we obtain a "continuity estimate" in Lemma 18 which, in turn, gives the bound on $\sup_{t\in[0,T]} Z^r_\delta(t)$ required by Lemma 7 to control the left-hand side of (5.45) (see Lemma 19). To control the right-hand side of (5.45), we again use excursion analysis to show that the integral $\int_\delta^\infty g'(x)W_x(\cdot)\,dx$ indeed converges to a finite random variable as $\delta \to 0$ (Lemmas 20 and 21). Together, these estimates complete the proof of Theorem 2.

Recall that for $r \in \mathcal{R}$, $a > 0$, and $t \ge 0$, $r(c^r)^{-1}Q^r_{a(c^r)^{-1}}(t)$ is the queue length of the rth $a(c^r)^{-1}$-truncated SRPT queue at time r^2t. Denote by Θ the collection of all functions $\theta : \mathcal{R} \to \mathbb{R}_+$ such that $\theta(r) \to 0$ as $r \to \infty$.

LEMMA 16. *For any* $a, T > 0$, *there exist* $\theta \in \Theta$ *and* $r_0 > 0$ *such that for* $r \ge r_0$,

$$\mathbb{P}\Big(\sup_{t\in[0,T]} Q^r_{a(c^r)^{-1}}(t) > \theta(r)\Big) \le \theta(r).$$

PROOF. Fix $a, T > 0$. For $r \in \mathcal{R}$ and $i \in \mathbb{N}$, let T_i^r and v_i respectively denote the inter-arrival time and processing time of the ith external job in the SRPT queue. For $r \in \mathcal{R}$, define $T_0^r = 0$. As in Section 2.1, T_1^r is strictly positive and has a finite second moment, but does not necessarily have the same distribution as T_i^r for $i \geq 2$, and T^r is a random variable that is equal in distribution to T_i^r for each $i \geq 2$. Also, for $r \in \mathcal{R}$ and $t \geq 0$, denote by $\hat{W}_a^r(t) := rW^r_{a(c^r)^{-1}}(r^{-2}t)$, the (unscaled) workload process of the rth $a(c^r)^{-1}$-truncated SRPT queue. Finally, for $r \in \mathcal{R}$, define the stopping times $\{K_i^r\}_{i=-1}^{\infty}$ with respect to the filtration $\{\mathcal{F}_n^r\}_{n\in\mathbb{Z}_+}$, where $\mathcal{F}_0^r := \sigma(\{\mathbf{q}^r, \breve{v}_l^r : l \in \mathbb{N}\})$ and, for $n \in \mathbb{N}$, $\mathcal{F}_n^r := \sigma(\{\mathbf{q}^r, \breve{v}_l^r, T_i^r, v_i : l \in \mathbb{N}, \leq i \leq n\})$, as follows:

$$K_{-1}^r = 0, \qquad K_0^r = 0 \quad \text{if } \hat{W}_a^r(0) = 0,$$

$$\text{otherwise } K_0^r = \inf\left\{k \in \mathbb{Z}_+ : \hat{W}_a^r\left(\left(\sum_{i=0}^{k} T_i^r\right)-\right) = 0\right\},$$

and, for $j \in \mathbb{Z}_+$,

$$K_{2j+1}^r := K_{2j}^r + 1, \qquad K_{2j+2}^r := \inf\left\{k \geq K_{2j+1}^r : \hat{W}_a^r\left(\left(\sum_{i=0}^{k} T_i^r\right)-\right) = 0\right\}.$$

For $r \in \mathcal{R}, l > 0$ and $i \in \mathbb{N}$, write $T_i^{r,l} := T_i^r \wedge l$ and $v_i^a := v_i \mathbf{1}_{[v_i \leq a]}$. Recall that the processing time distribution does not depend on $r \in \mathcal{R}$. Moreover, as the distribution function F of the processing time is assumed to satisfy $F(x) < 1$ for all $x \in \mathbb{R}$, and as $a < \infty$, we have that $\lambda\mathbb{E}(v_1^a) < 1$. Also, using (2.2),

$$\limsup_{l\to\infty} \limsup_{r\to\infty} \mathbb{E}\big(T^r \mathbf{1}_{[T^r > l]}\big) = 0. \tag{5.51}$$

Using these observations, there exist $r_0 \in \mathcal{R}$ and $l, \eta > 0$ such that $\lambda_l^r := (\mathbb{E}(T^r \wedge l))^{-1}$, $\mathbb{E}(v_1^a)$, λ^r and σ_A^r satisfy that for all $r \geq r_0$

$$\lambda_l^r \mathbb{E}\big(v_1^a\big) < 1 - 4\eta, \qquad \lambda_l^r \leq 2\lambda, \qquad \lambda^r \geq \lambda/2, \qquad \big(\sigma_A^r\big)^2 \leq 2\sigma_A^2, \qquad \lambda r^2 T > 1. \tag{5.52}$$

Fix $r_0 \in \mathcal{R}$, and $l, \eta > 0$ such that (5.52) holds. Note that for any $r \geq r_0$ and $j \in \mathbb{Z}_+$ such that $K_{2j}^r \geq K_{2j-1}^r + 2$ and for any $k \in [K_{2j-1}^r + 1, K_{2j}^r - 1]$,

$$\begin{aligned}
\hat{W}_a^r\left(\sum_{i=0}^{k} T_i^r\right) &= \hat{W}_a^r\left(\sum_{i=0}^{K_{2j-1}^r} T_i^r\right) + \sum_{i=K_{2j-1}^r+1}^{k} v_i^a - \sum_{i=K_{2j-1}^r+1}^{k} T_i^r \\
&\leq \hat{W}_a^r\left(\sum_{i=0}^{K_{2j-1}^r} T_i^r\right) + \sum_{i=K_{2j-1}^r+1}^{k} v_i^a - \sum_{i=K_{2j-1}^r+1}^{k} T_i^{r,l} \\
&\leq \hat{W}_a^r\left(\sum_{i=0}^{K_{2j-1}^r} T_i^r\right) + M_1(k) - M_1\big(K_{2j-1}^r\big) + M_2^r(k) - M_2^r\big(K_{2j-1}^r\big) \\
&\quad - 2\eta\lambda^{-1}\big(k - K_{2j-1}^r\big) + \big(\big(\lambda_l^r\big)^{-1} - \mathbb{E}\big[T_1^{r,l}\big]\big)\mathbf{1}_{\{j=0\}}, \\
&\leq \hat{W}_a^r\left(\sum_{i=0}^{K_{2j-1}^r} T_i^r\right) + M_1(k) - M_1\big(K_{2j-1}^r\big) + M_2^r(k) - M_2^r\big(K_{2j-1}^r\big) \\
&\quad - 2\eta\lambda^{-1}\big(k - K_{2j-1}^r\big) + 2\lambda^{-1}\mathbf{1}_{\{j=0\}},
\end{aligned} \tag{5.53}$$

where, for $r \geq r_0$, M_1 and M_2^r are martingales (with respect to the filtration $\{\mathcal{F}_n^r\}_{n\in\mathbb{Z}_+}$ defined above) given by $M_1(0) = M_2^r(0) = 0$, and for $k \in \mathbb{N}$,

$$M_1(k) := \sum_{i=1}^{k} \big(v_i^a - \mathbb{E}(v_i^a)\big) \quad \text{and} \quad M_2^r(k) := -\sum_{i=1}^{k} \big(T_i^{r,l} - \mathbb{E}(T_i^{r,l})\big).$$

For $r \geq r_0$, write $\mathcal{T}^r(k) := \sum_{i=1}^k T_i^r$ and $\mathcal{T}_l^r(k) := \sum_{i=1}^k T_i^{r,l}$ for $k \in \mathbb{N}$. As $v_i^a \leq a$ for all $i \in \mathbb{N}$ and there are at most two jobs in the rth $a(c^r)^{-1}$-truncated SRPT queue at time $\sum_{i=0}^{K_{2j-1}^r} T_i^r$ for each $j \in \mathbb{N}$, and $r \geq r_0$, $\hat{W}_a^r(\sum_{i=0}^{K_{2j-1}^r} T_i^r) \leq 2a$ for all $j \in \mathbb{N}$. For $k \geq k_1 := 2(1 + \lambda a/\eta)$, we have $2\eta\lambda^{-1}k - 2a \geq \eta\lambda^{-1}k + \eta\lambda^{-1}k_1 - 2a \geq \eta\lambda^{-1}k + 2\eta\lambda^{-1} \geq \eta\lambda^{-1}k$. Also, observe that $M_1(\cdot)$ (resp. $\tilde{M}_2^r(\cdot) := M_2^r(\cdot + 1) - M_2^r(1)$) is equal in distribution to $M_1(\cdot + j) - M_1(j)$ (resp. $M_2^r(\cdot + j) - M_2^r(j)$) for all $j \in \mathbb{N}$. Hence, as M_1 and M_2^r, $r \geq r_0$, are martingales with bounded increments such that the bounds on the increments do not depend on $r \geq r_0$, using (5.53) and the Azuma–Hoeffding inequality, we obtain that for $k \geq k_1 := 2(1 + \lambda a/\eta)$, $j \in \mathbb{N}$ and $r \geq r_0$,

$$\begin{aligned}\mathbb{P}\big(K_{2j}^r - K_{2j-1}^r > k\big) &\leq \mathbb{P}\big(2a + M_1(k) + \tilde{M}_2^r(k) - 2\eta\lambda^{-1}k > 0\big)\\ &\leq \mathbb{P}\big(M_1(k) > \eta k/(2\lambda)\big) + \mathbb{P}\big(\tilde{M}_2^r(k) > \eta k/(2\lambda)\big) \leq 2e^{-Ck}\end{aligned}$$

for some positive constant C depending on a, η, λ and l, but not $k \geq k_1$ and $r \geq r_0$. Note that for any $r \geq r_0$ and $j \in \mathbb{N}$, the queue length of the rth a-truncated SRPT queue in the time interval $[\mathcal{T}^r(K_{2j-2}^r), \mathcal{T}^r(K_{2j-1}^r)]$ is bounded above by 2 and in the time interval $[\mathcal{T}^r(K_{2j-1}^r), \mathcal{T}^r(K_{2j}^r)]$ is bounded above by $K_{2j}^r - K_{2j-1}^r + 1$. Thus, for any $r \geq r_0$, $k \geq k_1 + 1$ and $N \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{P}\bigg(\sup_{t\in[\mathcal{T}^r(K_0^r),\mathcal{T}^r(K_{2N}^r)]} \frac{r}{c^r} Q^r_{a(c^r)^{-1}}\big(r^{-2}t\big) > k\bigg)\\ &\leq \sum_{j=1}^{N} \mathbb{P}\bigg(\sup_{t\in[\mathcal{T}^r(K_{2j-1}^r),\mathcal{T}^r(K_{2j}^r)]} \frac{r}{c^r} Q^r_{a(c^r)^{-1}}\big(r^{-2}t\big) > k\bigg) \qquad (5.54)\\ &\leq \sum_{j=1}^{N} \mathbb{P}\big(K_{2j}^r - K_{2j-1}^r > k - 1\big) \leq 2Ne^C e^{-Ck}.\end{aligned}$$

For all $r \geq r_0$, $K_{2j}^r - K_{2j-2}^r \geq 1$ for all $j \in \mathbb{N}$ and $\lambda_l^r \leq 2\lambda$. Hence, for all $r \geq r_0$ and integers $N \geq 4r^2\lambda T + 1$,

$$\begin{aligned}\mathbb{P}\big(\mathcal{T}^r\big(K_{2N}^r\big) < r^2T\big) &\leq \mathbb{P}\big(\mathcal{T}_l^r\big(K_{2N}^r\big) < r^2T\big)\\ &\leq \mathbb{P}\big(\mathcal{T}_l^r(N) - T_1^{r,l} < r^2T\big)\\ &\leq \mathbb{P}\bigg(\sum_{i=2}^{N}\big(T_i^{r,l} - \mathbb{E}\big(T_i^{r,l}\big)\big) < r^2T - (N-1)\lambda^{-1}/2\bigg) \qquad (5.55)\\ &\leq \mathbb{P}\bigg(\sum_{i=2}^{N}\big(T_i^{r,l} - \mathbb{E}\big(T_i^{r,l}\big)\big) < -\lambda^{-1}(N-1)/4\bigg)\\ &\leq \frac{16\lambda^2(2\sigma_A^2 + 4\lambda^{-2})}{N-1},\end{aligned}$$

where we have used $\mathrm{Var}(T_i^{r,l}) \leq \mathbb{E}(T^r)^2 = (\sigma_A^r)^2 + (\lambda^r)^{-2} \leq 2\sigma_A^2 + 4\lambda^{-2}$ for $i \geq 2$ in the last bound. From (5.54) and (5.55), for $r \geq r_0$ and any integers $k \geq k_1$ and $N - 1 \in$

$[4\lambda r^2 T, 5\lambda r^2 T]$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[\mathcal{T}^r(K_0^r),r^2T]}\frac{r}{c^r}Q^r_{a(c^r)^{-1}}(r^{-2}t)>k\Big)\\
&\quad\le\mathbb{P}\Big(\sup_{t\in[\mathcal{T}^r(K_0^r),\mathcal{T}^r(K_{2N}^r)]}\frac{r}{c^r}Q^r_{a(c^r)^{-1}}(r^{-2}t)>k\Big)+\mathbb{P}\big(\mathcal{T}^r(K_{2N}^r)<r^2T\big)\\
&\quad\le 2Ne^Ce^{-Ck}+\frac{16\lambda^2(2\sigma_A^2+4\lambda^{-2})}{N-1}\le 10e^C\lambda r^2Te^{-Ck}+\frac{4\lambda(2\sigma_A^2+4\lambda^{-2})}{r^2T}.
\end{aligned}$$

Taking $r_1\ge r_0$ such that $\lfloor 3\log r_1/C\rfloor+1\ge k_1$ and $k=\lfloor 3\log r/C\rfloor+1$, we obtain that for some $r_1\ge r_0$ and all $r\ge r_1$,

$$\mathbb{P}\Big(\sup_{t\in[r^{-2}\mathcal{T}^r(K_0^r),T]}Q^r_{a(c^r)^{-1}}(t)>\frac{3c^r\log r}{Cr}+\frac{c^r}{r}\Big)\le\frac{10e^C\lambda T}{r}+\frac{4\lambda(2\sigma_A^2+4\lambda^{-2})}{r^2T}. \tag{5.56}$$

Note that $\hat{W}_a^r(0)=\sum_{\ell=1}^{\mathbf{q}^r}\check{v}_\ell^r\mathbf{1}_{[\check{v}_\ell^r\le a]}$ for $r\in\mathcal{R}$. Using this in (5.53) (with $j=0$), for any $r\ge r_0$ and $k\in\mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(K_0^r>k)&=\mathbb{P}\Big(\hat{W}_a^r\Big(\sum_{i=0}^k T_i^r\Big)>0,\,K_0^r>k\Big)\\
&\le\mathbb{P}\Big(\sum_{\ell=1}^{\mathbf{q}^r}\check{v}_\ell^r\mathbf{1}_{[\check{v}_\ell^r\le a]}+M_1(k)+M_2^r(k)-2\eta\lambda^{-1}k+2\lambda^{-1}>0\Big).
\end{aligned} \tag{5.57}$$

Further, note that by Assumption (2.19), there exists $\theta\in\Theta$ such that $r\theta(r)/c^r\to\infty$ as $r\to\infty$ and for all $r\in\mathcal{R}$,

$$\mathbb{P}\Big(\sum_{\ell=1}^{\mathbf{q}^r}\mathbf{1}_{[\check{v}_\ell^r\le a]}>r\eta\lambda^{-1}\theta(r)/(a+1)c^r\Big)\le\theta(r). \tag{5.58}$$

Using these observations, we conclude that there exists $r_2\ge r_1$ such that for all $r\ge r_2$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,r^{-2}\mathcal{T}^r(K_0^r)]}\frac{r}{c^r}Q^r_{a(c^r)^{-1}}(t)>(\eta\lambda^{-1}+1)\Big(1+\frac{r\theta(r)}{c^r}\Big)\Big)\\
&\quad\le\mathbb{P}\Big(\sum_{\ell=1}^{\mathbf{q}^r}\mathbf{1}_{[\check{v}_\ell^r\le a]}+K_0^r>(\eta\lambda^{-1}+1)\Big(1+\frac{r\theta(r)}{c^r}\Big)\Big)\\
&\quad\le\mathbb{P}\Big(\sum_{\ell=1}^{\mathbf{q}^r}\mathbf{1}_{[\check{v}_\ell^r\le a]}>\eta\lambda^{-1}\Big(1+\frac{r\theta(r)}{c^r}\Big)\Big)+\mathbb{P}\Big(K_0^r>1+\frac{r\theta(r)}{c^r}\Big)\\
&\quad\le\mathbb{P}\Big(\sum_{\ell=1}^{\mathbf{q}^r}\mathbf{1}_{[\check{v}_\ell^r\le a]}>r\eta\lambda^{-1}\theta(r)/c^r\Big)\\
&\qquad+\mathbb{P}\Big(\sum_{\ell=1}^{\mathbf{q}^r}\check{v}_\ell^r\mathbf{1}_{[\check{v}_\ell^r\le a]}+M_1\Big(\Big\lfloor 1+\frac{r\theta(r)}{c^r}\Big\rfloor\Big)\\
&\qquad+M_2^r\Big(\Big\lfloor 1+\frac{r\theta(r)}{c^r}\Big\rfloor\Big)-2\eta\lambda^{-1}r\theta(r)/c^r+2\lambda^{-1}>0\Big)
\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{P}\Bigg(\sum_{\ell=1}^{\mathbf{q}^r} \mathbf{1}_{[\check{v}_\ell^r \le a]} > \eta\lambda^{-1} r\theta(r)/c^r\Bigg) + \mathbb{P}\Bigg(\sum_{\ell=1}^{\mathbf{q}^r} \check{v}_\ell^r \mathbf{1}_{[\check{v}_\ell^r \le a]} > \eta\lambda^{-1} r\theta(r)/c^r\Bigg) \\
&\quad + \mathbb{P}\Bigg(M_1\Bigg(\Bigg\lfloor 1 + \frac{r\theta(r)}{c^r}\Bigg\rfloor\Bigg) > \eta\lambda^{-1} r\theta(r)/\big(2c^r\big) + \lambda^{-1}\Bigg) \\
&\quad + \mathbb{P}\Bigg(M_2^r\Bigg(\Bigg\lfloor 1 + \frac{r\theta(r)}{c^r}\Bigg\rfloor\Bigg) > \eta\lambda^{-1} r\theta(r)/\big(2c^r\big) + \lambda^{-1}\Bigg) \\
&\le \mathbb{P}\Bigg(\sum_{\ell=1}^{\mathbf{q}^r} \mathbf{1}_{[\check{v}_\ell^r \le a]} > \eta\lambda^{-1} r\theta(r)/c^r\Bigg) + \mathbb{P}\Bigg(\sum_{\ell=1}^{\mathbf{q}^r} \mathbf{1}_{[\check{v}_\ell^r \le a]} > \eta\lambda^{-1} r\theta(r)/ac^r\Bigg) \\
&\quad + \mathbb{P}\Bigg(M_1\Bigg(\Bigg\lfloor 1 + \frac{r\theta(r)}{c^r}\Bigg\rfloor\Bigg) > \eta\lambda^{-1} r\theta(r)/\big(2c^r\big) + \lambda^{-1}\Bigg) \\
&\quad + \mathbb{P}\Bigg(M_2^r\Bigg(\Bigg\lfloor 1 + \frac{r\theta(r)}{c^r}\Bigg\rfloor\Bigg) > \eta\lambda^{-1} r\theta(r)/\big(2c^r\big) + \lambda^{-1}\Bigg) \\
&\le 2\theta(r) + 2e^{-Cr\theta(r)/c^r},
\end{aligned} \tag{5.59}
$$

where we used (5.57) in the third inequality and (5.58) and the Azuma–Hoeffding inequality in the last inequality. Since the upper bounds in (5.56) and (5.59) tend to zero as $r \to \infty$, the lemma follows from (5.56) and (5.59). □

Recall the parameter η^* specified in Section 2.4. We will need the following technical lemma in what follows.

LEMMA 17. *Let $D' \ge 8p$ and $\eta \in (\eta^*, p-1)$. There exist $M_*(\eta) > 1$, $r_*(\eta) \ge 1$ and $\delta_*(\eta) \in (0,1)$ such that for all $r \ge r_*(\eta)$ and $\delta \in [2M_*(\eta)(c^r)^{-1}, \delta_*(\eta)]$ the following hold*:

$$
\frac{2^3 - 2}{2^{2D'\log(1/\delta)} - 2} \le \delta^{D'}, \tag{5.60}
$$

$$
\lambda/2 \le \lambda^r \le 8\lambda/7, \tag{5.61}
$$

$$
\mathbb{E}\big(T_1^r\big) \le 2\lambda^{-1} \tag{5.62}
$$

$$
\sigma_A^2/2 \le \big(\sigma_A^r\big)^2 \le 2\sigma_A^2, \tag{5.63}
$$

$$
\mathbb{E}\big[\big(T_1^r - (\lambda^r)^{-1}\big)^2\big] \le 2\sigma_A^2, \tag{5.64}
$$

$$
\mathbb{E}\big[\big(v\mathbf{1}_{[v\le\delta c^r]} - \lambda^r \mathbb{E}(v\mathbf{1}_{[v\le\delta c^r]})T_i^r\big)^2\big] \le C := \mathbb{E}\big[v^2\big] + \frac{128\sigma_A^2}{49} \quad \text{for all } i \in \mathbb{N}, \tag{5.65}
$$

$$
c^r < \Bigg(\frac{(p+1)r}{p}\Bigg)^{1/(p-\eta/2)}, \tag{5.66}
$$

$$
\Bigg(\frac{p+1}{p}\Bigg)^{2(p-\eta)/(p-\eta/2)} \frac{1}{Cr^{\eta/(p-\eta/2)}} \le \min\Bigg(\frac{2}{\lambda^2\sigma_A^2}, 194\Bigg), \tag{5.67}
$$

$$
\Bigg(\frac{p}{p+1}\Bigg)^{2(p-\eta)/(p-\eta/2)} r^{\eta/(p-\eta/2)} \le r^2\delta^{2(p-\eta)}, \tag{5.68}
$$

$$
-\lambda^r \frac{\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} + r\big(\rho^r - 1\big) \le -\frac{\lambda}{4\delta^{p-\eta}}. \tag{5.69}
$$

Moreover, for any $b_0 > 0$ and any $\eta \in (\eta^*, p-1)$, there exists $\tilde{r}(\eta, b_0) \geq r_*(\eta)$ such that for any $b \geq b_0$, $r \geq \tilde{r}(\eta, b_0)$ and $\delta \in [2M_*(\eta)(c^r)^{-1}, \delta_*(\eta)]$,

$$\mathbb{P}\big(E^r\big(3br^2\delta^{2(p-\eta)}/4\big) > \lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor\big) \leq \left(\frac{p+1}{p}\right)^{2(p-\eta)/(p-\eta/2)} \frac{2^9\lambda\sigma_A^2}{br^{\eta/(p-\eta/2)}}. \tag{5.70}$$

PROOF. By (2.2), (2.3), (4.10) and other elementary considerations, there exist $M_2(\eta) > 1$, $r_2(\eta) \geq 1$ and $\delta_2(\eta) \in (0,1)$ such that (5.60)–(5.67) hold for all $r \geq r_2(\eta)$ and $\delta \in [2M_2(\eta)(c^r)^{-1}, \delta_2(\eta)]$. Then (5.68) holds for all $r \geq r_2(\eta)$ and $\delta \in [2M_2(\eta)(c^r)^{-1}, \delta_2(\eta)]$ as well, since $2M_2(\eta) \geq 2$ and (5.66) imply that for all $r \geq r_2(\eta)$ and $\delta \in [2M_2(\eta)(c^r)^{-1}, \delta_2(\eta)]$,

$$\left(\frac{p}{p+1}\right)^{\frac{2(p-\eta)}{p-\eta/2}} r^{\eta/(p-\eta/2)} = r^2\left(\frac{p}{(p+1)r}\right)^{\frac{2(p-\eta)}{p-\eta/2}} \leq r^2\big(c^r\big)^{-2(p-\eta)} \leq r^2\delta^{2(p-\eta)}.$$

From Section 4.2(c), $\mathbb{E}(v\mathbf{1}_{[v>z]}) = z^{-p}\hat{L}(z)$ for all $z > 0$, where $\hat{L}$ satisfies (4.8) for some nonnegative Borel measurable functions $c(\cdot)$ and $\epsilon(\cdot)$, with $c(\cdot)$ satisfying $\lim_{x\to\infty} c(x) = c_0 \in (0,\infty)$ and $\epsilon(\cdot)$ satisfying $\epsilon(y) \to 0$ as $y \to \infty$. For any $\eta > 0$, we can obtain $M_3(\eta) \geq M_2(\eta)$ such that for all $y, z \geq M_3(\eta)$, $\frac{c(y)}{c(z)} \geq 1/2$ and $\epsilon(y) < \eta$. Hence, for all $\delta \in (0,1)$ and $z \geq M_3(\eta)/\delta$,

$$\frac{\hat{L}(\delta z)}{\hat{L}(z)} = \frac{c(\delta z)}{c(z)}\exp\left(-\int_{\delta z}^{z}\frac{\epsilon(y)}{y}\,dy\right) \geq \frac{\exp(-\eta\int_{\delta z}^{z}\frac{1}{y}\,dy)}{2} = \frac{\exp(-\eta\log(1/\delta))}{2} = \frac{\delta^\eta}{2}.$$

Thus, for all $\delta \in (0,1)$ and $z \geq M_3(\eta)/\delta$,

$$\frac{\mathbb{E}(v\mathbf{1}_{[v>\delta z]})}{\mathbb{E}(v\mathbf{1}_{[v>z]})} = \frac{(\delta z)^{-p}\hat{L}(\delta z)}{z^{-p}\hat{L}(z)} \geq \frac{1}{2\delta^{p-\eta}}.$$

From this it follows that for some $0 < \delta_3(\eta) \leq \delta_2(\eta)$, (5.69) holds for all $r \geq r_2(\eta)$ and $\delta \in [2M_3(\eta)(c^r)^{-1}, \delta_3(\eta)]$. Setting $r_*(\eta) = r_2(\eta)$, $M_*(\eta) = M_3(\eta)$ and $\delta_*(\eta) = \delta_3(\eta)$ completes the proof of (5.60)–(5.69).

To prove (5.70), note that for any $b_0 > 0$, by (5.68), we can choose $\tilde{r}(\eta, b_0) \geq r_*(\eta)$ such that for all $r \geq \tilde{r}(\eta, b_0)$ and $\delta \in [2M_*(\eta)(c^r)^{-1}, \delta_*(\eta)]$,

$$r^2\delta^{2(p-\eta)} \geq 28(\lambda b_0)^{-1}. \tag{5.71}$$

Using (5.61) in the third line below, (5.71) in the fifth line below, and Chebychev's inequality, (5.63), and (5.68) in the sixth line below, for all $b \geq b_0$, $r \geq \tilde{r}(\eta, b_0)$, $\delta \in [2M_*(\eta)(c^r)^{-1}, \delta_*(\eta)]$,

$$\begin{aligned}
&\mathbb{P}\big(E^r\big(3br^2\delta^{2(p-\eta)}/4\big) > \lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor\big)\\
&\quad\leq \mathbb{P}\left(\sum_{i=2}^{\lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor} T_i^r < \frac{3br^2\delta^{2(p-\eta)}}{4}\right)\\
&\quad= \mathbb{P}\left(\sum_{i=2}^{\lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor} \big(T_i^r - \mathbb{E}(T_i^r)\big) < \frac{3br^2\delta^{2(p-\eta)}}{4} - \big(\lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor - 1\big)\big(\lambda^r\big)^{-1}\right)\\
&\quad\leq \mathbb{P}\left(\sum_{i=2}^{\lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor} \big(T_i^r - \mathbb{E}(T_i^r)\big) < \frac{3br^2\delta^{2(p-\eta)}}{4} - \frac{7\lambda^{-1}}{8}\big(b\lambda r^2\delta^{2(p-\eta)} - 2\big)\right)\\
&\quad= \mathbb{P}\left(\sum_{i=2}^{\lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor} \big(T_i^r - \mathbb{E}(T_i^r)\big) < -\frac{br^2\delta^{2(p-\eta)}}{8} + \frac{14\lambda^{-1}}{8}\right)
\end{aligned}$$

$$\le \mathbb{P}\Bigg(\sum_{i=2}^{\lfloor b\lambda r^2\delta^{2(p-\eta)}\rfloor} (T_i^r - \mathbb{E}(T_i^r)) < -\frac{br^2\delta^{2(p-\eta)}}{16}\Bigg)$$
$$\le \frac{2^8\lambda(\sigma_A^r)^2}{br^2\delta^{2(p-\eta)}} \le \Big(\frac{p+1}{p}\Big)^{2(p-\eta)/(p-\eta/2)} \frac{2^9\lambda\sigma_A^2}{br^{\eta/(p-\eta/2)}}.$$

Hence (5.70) holds for all $b \ge b_0$, $r \ge \tilde{r}(\eta, b_0)$, $\delta \in [2M_*(\eta)(c^r)^{-1}, \delta_*(\eta)]$. □

LEMMA 18. *Fix $T > 0$. There exist D_1, D_2, $D_3 > 0$ such that the following holds: For any $\eta \in (\eta^*, p-1)$, there exist $M(\eta) > 1$, $r(\eta) \ge 2$ and $\delta(\eta) \in (0,1)$ such that for all $r \ge r(\eta)$ and $\delta \in [2M(\eta)(c^r)^{-1}, \delta(\eta)]$,*

$$\mathbb{P}\Big(\sup_{t\in[0,T]} (Q_\delta^r(t) - Q_{\delta/2}^r(t)) > D_1\delta^{p-1-\eta}\log(\delta^{-1}) + \frac{c^r}{r}\Big)$$
$$\le 35\delta^{D_2} + \mathbb{P}\Big(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \breve{v}_l^r \mathbf{1}_{[\breve{v}_l^r \le \delta c^r]} > D_3\delta^{p-\eta}\Big).$$

PROOF. Fix $D' \ge 8p$. For $\eta \in (\eta^*, p-1)$ and $b_0 > 0$, recall $M_*(\eta)$, $r_*(\eta)$, $\delta_*(\eta)$ and $\tilde{r}(\eta, b_0)$ from Lemma 17. Set $M(\eta) = M_*(\eta)$ and take $r \ge \max\{r_*(\eta), \tilde{r}(\eta, \lambda^{-1})\}$ and $\delta \in [2M(\eta)(c^r)^{-1}, \delta_*(\eta)]$. For all $t \ge 0$, by Lemma 11 with $x = \delta/2$ and $y = \delta$,

$$0 \le Q_\delta^r(t) - Q_{\delta/2}^r(t) \le \frac{c^r}{r} + 2\delta^{-1}Y_\delta^r(t). \tag{5.72}$$

The major effort of the proof will be to obtain bounds on the probability that $\sup_{0\le t\le T'} 2\delta^{-1}Y_\delta^r(t)$ exceeds certain bounds for a suitable $T' \ge T$, which entails a detailed analysis of its excursions. To get an overview of the strategy for this, the reader may wish to look ahead to (5.96) (where $r(\eta)$, $B \ge 1$ and $\epsilon \in (0,1)$ are constants to be determined in what follows), definitions (5.79), (5.80) and (5.81), (5.95) and (5.96). In what follows, each of the three terms on the right side of (5.96) is bounded above using estimates in (5.82), (5.92) and (5.93).

Recall that $Y_\delta^r(t) = \Gamma[X_\delta^r](t)$ for $t \ge 0$. From (5.8) and (5.9), for $t \ge 0$,

$$X_\delta^r(t) = X_\delta^r(0) + \hat{V}_\delta^r(t) - \lambda^r t\frac{\mathbb{E}(v\mathbf{1}_{[v>\delta c^r]})}{\mathbb{E}(v\mathbf{1}_{[v>c^r]})} + rt(\rho^r - 1).$$

By (5.69), for all $0 \le s \le t$,

$$X_\delta^r(t) - X_\delta^r(s) \le U_\delta^r(t) - U_\delta^r(s) \quad \text{where } U_\delta^r(t) := \hat{V}_\delta^r(t) - \frac{\lambda t}{4\delta^{p-\eta}}. \tag{5.73}$$

For $k \in \mathbb{N}$, write

$$\widetilde{V}_\delta^r(k) := r\hat{V}_\delta^r\Big(r^{-2}\sum_{i=1}^k T_i^r\Big) = \sum_{i=1}^k v_i\mathbf{1}_{[v_i\le\delta c^r]} - \lambda^r\mathbb{E}(v\mathbf{1}_{[v\le\delta c^r]})\sum_{i=1}^k T_i^r.$$

By (5.65), for each $k \in \mathbb{N}$,

$$\mathbb{E}\big[(\widetilde{V}_\delta^r(k))^2\big] \le Ck.$$

Take any $B \ge 1$. Thus, as $\{\widetilde{V}_\delta^r(k)\}_{k\in\mathbb{N}}$ is a martingale (with respect to the filtration $\{\mathcal{F}_n^r\}_{n\in\mathbb{Z}_+}$, where $\mathcal{F}_0^r := \{\mathbf{q}^r, \breve{v}_l^r : l \in \mathbb{N}\}$ and $\mathcal{F}_n^r := \sigma(\mathbf{q}^r, \breve{v}_l^r, T_i^r, v_i : l \in \mathbb{N}, i \le n)$, $n \ge 1$), using Doob's

maximal inequality, (5.66), recalling $r \geq \tilde{r}(\eta, \lambda^{-1})$ and using (5.70) with $b = 16B\lambda^{-1}$,

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,12B\lambda^{-1}\delta^{2(p-\eta)}]} \hat{V}_\delta^r(t) > B\delta^{p-\eta}/2 \Big) \\
&= \mathbb{P}\left(\sup_{1\leq k\leq E^r(12B\lambda^{-1}r^2\delta^{2(p-\eta)})} \hat{V}_\delta^r\left(r^{-2}\sum_{i=1}^{k} T_i^r \right) > B\delta^{p-\eta}/2 \right) \\
&\leq \mathbb{P}\left(\sup_{1\leq k\leq \lfloor 16Br^2\delta^{2(p-\eta)}\rfloor} \hat{V}_\delta^r\left(r^{-2}\sum_{i=1}^{k} T_i^r \right) > B\delta^{p-\eta}/2 \right) \\
&\quad + \mathbb{P}\big(E^r\big(12Br^2\lambda^{-1}\delta^{2(p-\eta)}\big) > \lfloor 16Br^2\delta^{2(p-\eta)}\rfloor \big) \\
&= \mathbb{P}\Big(\sup_{1\leq k\leq \lfloor 16Br^2\delta^{2(p-\eta)}\rfloor} \widetilde{V}_\delta^r(k) > Br\delta^{p-\eta}/2 \Big) \\
&\quad + \mathbb{P}\big(E^r\big(12Br^2\lambda^{-1}\delta^{2(p-\eta)}\big) > \lfloor 16Br^2\delta^{2(p-\eta)}\rfloor \big) \\
&\leq \frac{16\mathbb{E}[(\widetilde{V}_\delta^r(\lfloor 16Br^2\delta^{2(p-\eta)}\rfloor))^2]}{B^2r^2\delta^{2(p-\eta)}} \\
&\quad + \mathbb{P}\big(E^r\big(12Br^2\lambda^{-1}\delta^{2(p-\eta)}\big) > \lfloor 16Br^2\delta^{2(p-\eta)}\rfloor \big) \\
&\leq \frac{256C}{B} + \left(\frac{p+1}{p}\right)^{2(p-\eta)/(p-\eta/2)} \frac{32\lambda^2\sigma_A^2}{Br^{\eta/(p-\eta/2)}}.
\end{aligned} \tag{5.74}
$$

Next, from (5.73), we see that for any integer $i \geq 2$ and $s \geq 0$,

$$
\begin{aligned}
&\mathbb{P}\big(X_\delta^r(\cdot + s) \text{ crosses } (i+1)B\delta^{p-\eta} \text{ before } (i-2)B\delta^{p-\eta} \mid X_\delta^r(s) = iB\delta^{p-\eta}, \\
&\qquad E^r(r^2 s) - E^r(r^2 s-) > 0 \big) \\
&\leq \mathbb{P}\Big(\sup_{t\in[0,12B\lambda^{-1}\delta^{2(p-\eta)}]} \hat{V}_\delta^r(t+s) - \hat{V}_\delta^r(s) > B\delta^{p-\eta}/2 \,\Big|\, X_\delta^r(s) = iB\delta^{p-\eta}, \\
&\qquad E^r(r^2 s) - E^r(r^2 s-) > 0 \Big) \\
&= \mathbb{P}\Big(\sup_{t\in[0,12B\lambda^{-1}\delta^{2(p-\eta)}]} \hat{V}_\delta^r(t) > B\delta^{p-\eta}/2 \Big),
\end{aligned} \tag{5.75}
$$

where, in the last step, we have used the strong Markov property of the process $\hat{V}_\delta^r(\cdot)$ at the jump times of the process $t \mapsto E^r(r^2 t)$. Combining (5.74) and (5.75), setting $B = 960C \vee 1$, and using (5.67), we conclude that for all integers $i \geq 2$, $0 \leq x \leq iB\delta^{p-\eta}$ and $s \geq 0$,

$$
\begin{aligned}
&\mathbb{P}\big(X_\delta^r(\cdot + s) \text{ crosses } (i+1)B\delta^{p-\eta} \text{ before } (i-2)B\delta^{p-\eta} \mid X_\delta^r(s) = x, \\
&\qquad E^r(r^2 s) - E^r(r^2 s-) > 0 \big) \\
&\leq \mathbb{P}\big(X_\delta^r(\cdot + s) \text{ crosses } (i+1)B\delta^{p-\eta} \text{ before } (i-2)B\delta^{p-\eta} \mid X_\delta^r(s) = iB\delta^{p-\eta}, \\
&\qquad E^r(r^2 s) - E^r(r^2 s-) > 0 \big) \\
&\leq \frac{1}{3}.
\end{aligned} \tag{5.76}
$$

Using $M(\eta)/c^r \le 2M(\eta)/c^r \le \delta$, $M(\eta) = M_*(\eta) > 1$, (5.66), (5.67), $r_*(\eta) \ge 1$ and $B \ge 960C > 2C$,

$$\begin{aligned}\delta c^r/r &= \delta^{p-\eta}\frac{c^r}{r\delta^{p-\eta-1}} \le \delta^{p-\eta}\frac{(c^r)^{p-\eta}}{M(\eta)^{p-\eta-1}r}\\ &\le \delta^{p-\eta}\left(\frac{p+1}{p}\right)^{(p-\eta)/(p-\eta/2)}\frac{r^{(p-\eta)/(p-\eta/2)}}{r}\\ &= \delta^{p-\eta}\left(\frac{p+1}{p}\right)^{(p-\eta)/(p-\eta/2)} r^{-\eta/(2p-\eta)} < B\delta^{p-\eta}/2.\end{aligned} \tag{5.77}$$

For $s \ge 0$, define the following stopping times with respect to the filtration $\{\mathcal{H}_t\}_{t\ge 0}$ given by $\mathcal{H}_t := \{\mathbf{q}^r, \check{v}_l^r, V_\delta^r(r^2 s), E^r(r^2 s) : l \in \mathbb{N}, s \le t\}$ for $t \ge 0$: $\beta_0 = s$ and for $k \in \mathbb{Z}_+$,

$$\begin{aligned}\beta_{k+1} := \inf\{t \ge \beta_k : X_\delta^r(t) - X_\delta^r(\beta_k) \ge B\delta^{p-\eta}\\ \text{or } E^r(r^2 t) - E^r(r^2 t-) > 0 \text{ and } X_\delta^r(t-) - X_\delta^r(\beta_k) \le -2B\delta^{p-\eta}\},\end{aligned}$$

and write $\widetilde{X}_\delta^r(k) := X_\delta^r(\beta_k)$. For any $k \in \mathbb{Z}_+$, note that if $X_\delta^r(\beta_{k+1}) - X_\delta^r(\beta_k) \ge B\delta^{p-\eta}$; that is, if β_{k+1} corresponds to an up-crossing of X_δ^r, then, using (5.77) and that jumps up of $X_\delta^r(\cdot)$ are at most of size $\delta c^r/r$, $X_\delta^r(\beta_{k+1}) - X_\delta^r(\beta_k) \le B\delta^{p-\eta} + \delta c^r/r \le 3B\delta^{p-\eta}/2$. Similarly, for any $k \in \mathbb{Z}_+$, if $X_\delta^r(\beta_{k+1}-) - X_\delta^r(\beta_k) \le -2B\delta^{p-\eta}$, then, by the same line of reasoning, $X_\delta^r(\beta_{k+1}) - X_\delta^r(\beta_k) \le -2B\delta^{p-\eta} + \delta c^r/r \le -3B\delta^{p-\eta}/2$. Let $\{S_\delta(k)\}_{k\in\mathbb{Z}_+}$ be a random walk with $S_\delta(0) = 9B\delta^{p-\eta}/2$ and for $k \in \mathbb{Z}_+$,

$$\begin{aligned}\mathbb{P}\big(S_\delta(k+1) - S_\delta(k) = 3B\delta^{p-\eta}/2\big) = 1/3 \quad \text{and}\\ \mathbb{P}\big(S_\delta(k+1) - S_\delta(k) = -3B\delta^{p-\eta}/2\big) = 2/3.\end{aligned}$$

Recall that $D' \ge 8p$ was fixed at the onset. Also note that (5.60) implies that $\frac{2^3-2}{2^{2D'\log(1/\delta)}-2} \le 1$, which in turn implies that $9/2 \le 3D'\log(1/\delta)$. Then, from (5.76), the above observations, (5.60), and by comparing the sequence $\{\widetilde{X}_\delta^r(k)\}_{k\in\mathbb{Z}_+}$ with $\{S_\delta(k)\}_{k\in\mathbb{Z}_+}$, it follows that, for any $t \ge 0$ and any $x_0 \in [4B\delta^{p-\eta}, 9B\delta^{p-\eta}/2]$,

$$\begin{aligned}&\mathbb{P}\Big(Y_\delta^r(t+\cdot) \text{ crosses } 3D'B\delta^{p-\eta}\log(\delta^{-1}) \text{ before } \frac{3B\delta^{p-\eta}}{2} \,\Big|\, Y_\delta^r(t) = x_0,\\ &\quad E^r(r^2 t) - E^r(r^2 t-) > 0\Big)\\ &= \mathbb{P}\Big(X_\delta^r(t+\cdot) \text{ crosses } 3D'B\delta^{p-\eta}\log(\delta^{-1}) \text{ before } \frac{3B\delta^{p-\eta}}{2} \,\Big|\, X_\delta^r(t) = x_0,\\ &\quad E^r(r^2 t) - E^r(r^2 t-) > 0\Big)\\ &\le \mathbb{P}\Big(X_\delta^r(t+\cdot) \text{ crosses } 3D'B\delta^{p-\eta}\log(\delta^{-1}) \text{ before } \frac{3B\delta^{p-\eta}}{2} \,\Big|\, X_\delta^r(t) = \frac{9B\delta^{p-\eta}}{2},\\ &\quad E^r(r^2 t) - E^r(r^2 t-) > 0\Big)\\ &\le \mathbb{P}\Big(S_\delta(\cdot) \text{ crosses } 3D'B\delta^{p-\eta}\log(\delta^{-1}) \text{ before } \frac{3B\delta^{p-\eta}}{2}\Big)\\ &\le \frac{2^3-2}{2^{2D'\log(1/\delta)}-2} \le \delta^{D'},\end{aligned} \tag{5.78}$$

where, in the second to the last inequality above, we have used the fact that, for the biased random walk S_δ, $n \mapsto 2^{2S_\delta(n)/(3B\delta^{p-\eta})}$ is a martingale (with respect to the natural filtration generated by S_δ) to compute the probability via optional stopping theorem. Define the following stopping times (with respect to the filtration $\{\mathcal{H}_t\}_{t\geq 0}$ defined above): $\tau_{-1} = 0$ and for $k \in \mathbb{Z}_+$,

$$\tau_{2k} := \inf\{t \geq \tau_{2k-1} : E^r(r^2 t) - E^r(r^2 t-) > 0 \text{ and } Y_\delta^r(t-) \leq 2B\delta^{p-\eta}\}, \tag{5.79}$$

$$\tau_{2k+1} := \inf\{t \geq \tau_{2k} : Y_\delta^r(t) \geq 4B\delta^{p-\eta}\} \tag{5.80}$$

and let

$$\mathcal{N} := \inf\Big\{k \in \mathbb{N} : \sup_{t\in[\tau_{2k-1},\tau_{2k}]} Y_\delta^r(t) \geq 3D'B\delta^{p-\eta}\log(1/\delta)\Big\}. \tag{5.81}$$

Due to (5.77) and since Y_δ^r has upward jumps of size at most $c^r\delta/r$, for each $k \in \mathbb{N}$, $Y_\delta^r(\tau_{2k-1}) \in [4B\delta^{p-\eta}, 9B\delta^{p-\eta}/2]$. As $\delta \leq \delta_*(\eta) < 1$, by (5.78),

$$\begin{aligned}\mathbb{P}\big(\mathcal{N} \leq \lfloor \delta^{-D'/2}\rfloor + 1\big) &\leq \sum_{k=1}^{\lfloor\delta^{-D'/2}\rfloor+1} \mathbb{P}\Big(\sup_{t\in[\tau_{2k-1},\tau_{2k}]} Y_\delta^r(t) \geq 3D'B\delta^{p-\eta}\log(1/\delta)\Big) \\ &\leq 2\delta^{D'/2}.\end{aligned} \tag{5.82}$$

Using (5.77), $Y_\delta^r(\tau_{2k}) \leq 3B\delta^{p-\eta}$ for all $k \in \mathbb{Z}_+$. From (5.73), it follows that, for each $k \in \mathbb{Z}_+$,

$$t \mapsto \big(\hat{V}_\delta^r(t+\tau_{2k}) - \hat{V}_\delta^r(\tau_{2k})\big) - \big(X_\delta^r(t+\tau_{2k}) - X_\delta^r(\tau_{2k})\big), \quad t \geq 0,$$

is nondecreasing in t. Thus, by the monotonicity property noted in (4.3), for each $k \in \mathbb{Z}_+$ and $t \geq 0$,

$$\begin{aligned}Y_\delta^r(t+\tau_{2k}) &= \Gamma\big[Y_\delta^r(\tau_{2k}) + \big(X_\delta^r(\cdot+\tau_{2k}) - X_\delta^r(\tau_{2k})\big)\big](t) \\ &\leq \Gamma\big[Y_\delta^r(\tau_{2k}) + \big(\hat{V}_\delta^r(\cdot+\tau_{2k}) - \hat{V}_\delta^r(\tau_{2k})\big)\big](t) \\ &\leq \Gamma\big[3B\delta^{p-\eta} + \big(\hat{V}_\delta^r(\cdot+\tau_{2k}) - \hat{V}_\delta^r(\tau_{2k})\big)\big](t).\end{aligned}$$

For each $k \in \mathbb{Z}_+$, a job arrives to the rth system at time τ_{2k}. Hence, by the strong Markov property, $\{\Gamma[3B\delta^{p-\eta} + (\hat{V}_\delta^r(\cdot+\tau_{2k}) - \hat{V}_\delta^r(\tau_{2k}))](t) : t \geq 0\}$ has the same distribution as the process $\{\Gamma[3B\delta^{p-\eta} + \hat{V}_\delta^r(\cdot)](t) : t \geq 0\}$. Thus, for each $d \in \mathbb{N}$ and $t \geq 0$,

$$\mathbb{P}\Big(\sum_{j=0}^{d}(\tau_{2j+1} - \tau_{2j}) \leq t\Big) \leq \mathbb{P}\Big(\sum_{j=0}^{d}\chi_j \leq t\Big), \tag{5.83}$$

where $\{\chi_0, \chi_1, \dots\}$ are independent and identically distributed random variables distributed as

$$\mathbb{P}(\chi_0 \leq s) = \mathbb{P}\Big(\sup_{t\in[0,s]} \Gamma\big[3B\delta^{p-\eta} + \hat{V}_\delta^r(\cdot)\big](t) \geq 4B\delta^{p-\eta}\Big), \quad s \geq 0.$$

Recalling $\delta \leq \delta_*(\eta) < 1$ and using the Lipschitz property of the Skorohod map noted in (4.1), we obtain that for all $\epsilon \in (0,1)$,

$$\mathbb{P}\big(\chi_0 \leq \epsilon\delta^{2(p-\eta)}\big) \leq \mathbb{P}\Big(\sup_{t\in[0,\epsilon\delta^{2(p-\eta)}]} |\hat{V}_\delta^r(t)| \geq B\delta^{p-\eta}/2\Big). \tag{5.84}$$

Then given $\epsilon \in (0,1)$, following the same line of reasoning used to obtain (5.74) and using (5.70) with $b = 2\epsilon$ (noting $3b/4 > \epsilon$), we obtain for $r \geq \hat{r}(\eta,\epsilon) := \max\{r_*(\eta), \tilde{r}(\eta, \lambda^{-1}),$

$\tilde{r}(\eta,\epsilon)\}$,

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,\epsilon\delta^{2(p-\eta)}]}\hat{V}_\delta^r(t)\geq B\delta^{p-\eta}/4\Big)\\
&\leq \mathbb{P}\Bigg(\sup_{1\leq k\leq\lfloor 2\epsilon\lambda r^2\delta^{2(p-\eta)}\rfloor}\hat{V}_\delta^r\Big(r^{-2}\sum_{i=1}^k T_i^r\Big)>B\delta^{p-\eta}/4\Bigg)\\
&\quad+\mathbb{P}\big(E^r\big(\epsilon r^2\delta^{2(p-\eta)}\big)>\big\lfloor 2\epsilon\lambda r^2\delta^{2(p-\eta)}\big\rfloor\big)\\
&\leq \frac{128C\lambda\epsilon}{B^2}+\Big(\frac{p+1}{p}\Big)^{2(p-\eta)/(p-\eta/2)}\frac{2^8\lambda\sigma_A^2}{\epsilon r^{\eta/(p-\eta/2)}}.
\end{aligned}
\tag{5.85}
$$

Moreover, as $\hat{V}_\delta^r(\cdot)$ decreases between successive arrivals of jobs and increases at the arrival times, for each $\epsilon\in(0,1)$, we have the following lower bound on $\hat{V}_\delta^r(\cdot)$ on the time interval $[0,\epsilon\delta^{2(p-\eta)}]$:

$$
\begin{aligned}
&\inf_{t\in[0,\epsilon\delta^{2(p-\eta)}]}\hat{V}_\delta^r(t)\\
&\geq \inf_{0\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})}\Bigg(r^{-1}\sum_{i=1}^k v_i\mathbf{1}_{[v_i\leq\delta c^r]}-\lambda^r r^{-1}\mathbb{E}(v\mathbf{1}_{[v\leq\delta c^r]})\sum_{i=1}^{k+1}T_i^r\Bigg)\\
&\geq \inf_{0\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})}\Bigg(r^{-1}\sum_{i=1}^k v_i\mathbf{1}_{[v_i\leq\delta c^r]}-\lambda^r r^{-1}\mathbb{E}(v\mathbf{1}_{[v\leq\delta c^r]})\sum_{i=1}^{k}T_i^r\Bigg)\\
&\quad-\frac{8\lambda\mathbb{E}(v)}{7r}\sup_{1\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})+1}T_k^r\\
&=\frac{1}{r}\inf_{1\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})}\widetilde{V}_\delta^r(k)-\frac{8}{7r}\sup_{1\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})+1}T_k^r,
\end{aligned}
\tag{5.86}
$$

where the bound (5.61) was used in the last term. Once again, following the arguments for obtaining (5.74) in a manner similar to those that arrive at (5.85), for $\epsilon\in(0,1)$ and $r\geq\hat{r}(\eta,\epsilon)$,

$$
\begin{aligned}
&\mathbb{P}\Big(\frac{1}{r}\inf_{1\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})}\widetilde{V}_\delta^r(k)<-B\delta^{p-\eta}/8\Big)\\
&\leq\frac{512C\lambda\epsilon}{B^2}+\Big(\frac{p+1}{p}\Big)^{2(p-\eta)/(p-\eta/2)}\frac{2^8\lambda\sigma_A^2}{\epsilon r^{\eta/(p-\eta/2)}}.
\end{aligned}
\tag{5.87}
$$

Moreover, for any $\epsilon\in(0,1)$,

$$
\begin{aligned}
&\mathbb{P}\Big(\frac{8}{7r}\sup_{1\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})+1}T_k^r>B\delta^{p-\eta}/8\Big)\\
&=\mathbb{P}\Big(\sup_{1\leq k\leq E^r(\epsilon r^2\delta^{2(p-\eta)})+1}T_k^r>\frac{7B\delta^{p-\eta}r}{64}\Big)\\
&\leq\mathbb{P}\big(E^r\big(\epsilon r^2\delta^{2(p-\eta)}\big)>\big\lfloor 2\epsilon\lambda r^2\delta^{2(p-\eta)}\big\rfloor\big)\\
&\quad+\mathbb{P}\Big(\sup_{1\leq k\leq\lfloor 2\epsilon\lambda r^2\delta^{2(p-\eta)}\rfloor+1}T_k^r>\frac{7B\delta^{p-\eta}r}{64}\Big).
\end{aligned}
\tag{5.88}
$$

Applying a union bound, Chebychev's inequality and (5.61)–(5.64), it follows that for any $\epsilon \in (0,1)$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{1\le k\le \lfloor 2\epsilon\lambda r^2\delta^{2(p-\eta)}\rfloor+1} T_k^r > \frac{7B\delta^{p-\eta}r}{64}\Big)\\
&\le \big(2\epsilon\lambda r^2\delta^{2(p-\eta)}+1\big)\max_{k=1,2}\mathbb{P}\Big(T_k^r > \frac{7B\delta^{p-\eta}r}{64}\Big)\\
&\le \big(2\epsilon\lambda r^2\delta^{2(p-\eta)}+1\big)\Big(\frac{64}{7B\delta^{p-\eta}r}\Big)^2\big(\mathbb{E}\big[(T_1^r)^2\big]\vee\mathbb{E}\big[(T^r)^2\big]\big)\\
&\le \frac{(2\lambda+(\epsilon r^2\delta^{2(p-\eta)})^{-1})\epsilon C_1}{B^2},
\end{aligned}$$

where $C_1 = 10^2(2\sigma_A^2+(2\lambda^{-1})^2)$. Thus, for $\epsilon\in(0,1)$ and $r\ge\hat r(\eta,\epsilon)$, by using the above bound and (5.70) with $b=2\epsilon$ in (5.88), we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\frac{8}{7r}\sup_{1\le k\le E^r(\epsilon r^2\delta^{2(p-\eta)})} T_k^r > B\delta^{p-\eta}/8\Big)\\
&\le\Big(\frac{p+1}{p}\Big)^{2(p-\eta)/(p-\eta/2)}\frac{2^8\lambda\sigma_A^2}{\epsilon r^{\eta/(p-\eta/2)}}+\frac{(2\lambda+(\epsilon r^2\delta^{2(p-\eta)})^{-1})\epsilon C_1}{B^2}.
\end{aligned}\tag{5.89}$$

From (5.86), (5.87) and (5.89), for $\epsilon\in(0,1)$ and $r\ge\hat r(\eta,\epsilon)$,

$$\begin{aligned}
&\mathbb{P}\Big(\inf_{t\in[0,\epsilon\delta^{2(p-\eta)}]}\hat V_\delta^r(t) < -B\delta^{p-\eta}/4\Big)\\
&\le\frac{512C\lambda\epsilon}{B^2}+2\Big(\frac{p+1}{p}\Big)^{2(p-\eta)/(p-\eta/2)}\frac{2^8\lambda\sigma_A^2}{\epsilon r^{\eta/(p-\eta/2)}}\\
&\quad+\frac{(2\lambda+(\epsilon r^2\delta^{2(p-\eta)})^{-1})\epsilon C_1}{B^2}.
\end{aligned}\tag{5.90}$$

From (5.85), (5.90) and as $B\ge1$, we can fix $\epsilon\in(0,1)$ and find $\hat r(\eta)\ge\hat r(\eta,\epsilon)$ such that for all $r\ge\hat r(\eta)$,

$$\mathbb{P}\Big(\sup_{t\in[0,\epsilon\delta^{2(p-\eta)}]}\big|\hat V_\delta^r(t)\big|\ge B\delta^{p-\eta}/2\Big)\le 1/2,$$

and hence, from (5.84),

$$\mathbb{P}\big(\chi_0\le\epsilon\delta^{2(p-\eta)}\big)\le 1/2.\tag{5.91}$$

Henceforth, we fix such an ϵ and assume $r\ge\hat r(\eta)$. Applying the Azuma–Hoeffding inequality on the martingale (with respect to its natural filtration)

$$M_\ell^\chi := \sum_{k=1}^{\ell}\big(\mathbf{1}_{[\chi_k>\epsilon\delta^{2(p-\eta)}]}-\mathbb{P}\big(\chi_0>\epsilon\delta^{2(p-\eta)}\big)\big),\quad \ell\in\mathbb{Z}_+,$$

and using (5.83) and (5.91), for any $d\ge1$, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{j=0}^{d}(\tau_{2j+1}-\tau_{2j})\le d\epsilon\delta^{2(p-\eta)}/4\Big)\\
&\le\mathbb{P}\Big(\sum_{j=0}^{d}\chi_j\le d\epsilon\delta^{2(p-\eta)}/4\Big)\le\mathbb{P}\Big(\sum_{j=1}^{d}\mathbf{1}_{[\chi_j>\epsilon\delta^{2(p-\eta)}]}\le d/4\Big)\\
&=\mathbb{P}\big(M_d^\chi+d\mathbb{P}\big(\chi_0>\epsilon\delta^{2(p-\eta)}\big)\le d/4\big)\le\mathbb{P}\big(M_d^\chi\le -d/4\big)\le e^{-d/32}.
\end{aligned}\tag{5.92}$$

Note that if $Y^r_\delta(\tilde{t}) \le 3B\delta^{p-\eta}/2$ for some $\tilde{t} < \tau_0$, then, by definition (5.79), the time of the arrival immediately following $\tilde{t}$ corresponds to τ_0. By (5.77), $Y^r_\delta(\tau_0) \le 3B\delta^{p-\eta}/2 + \frac{c^r\delta}{r} < 2B\delta^{p-\eta}$ and as $Y^r_\delta(\cdot)$ is nonincreasing in the time interval $[\tilde{t}, \tau_0)$, $\sup_{t\in[\tilde{t},\tau_0]} Y^r_\delta(t) < 2B\delta^{p-\eta}$. Consequently, if $Y^r_\delta(\cdot)$ attains any value $v > 2B\delta^{p-\eta}$ before τ_0, $Y^r_\delta(0) > 3B\delta^{p-\eta}/2$ and the time at which v is attained must be before $Y^r_\delta(\cdot)$ down crosses $3B\delta^{p-\eta}/2$. Thus, from the computation (5.78), recalling that $Y^r_\delta(0) = \frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \check{v}^r_l \mathbf{1}_{[\check{v}^r_l \le \delta c^r]}$ and using the fact that the process $Y^r_\delta(\cdot)$ started from $Y^r_\delta(0) = \frac{9B\delta^{p-\eta}}{2}$ stochastically dominates (in a pathwise fashion) the process $Y^r_\delta(\cdot)$ started from any value less than or equal to $\frac{9B\delta^{p-\eta}}{2}$,

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,\tau_0]} Y^r_\delta(t) > 3D'B\delta^{p-\eta}\log(1/\delta)\Big)\\
&\quad\le \mathbb{P}\Big(Y^r_\delta(\cdot) \text{ crosses } 3D'B\delta^{p-\eta}\log(1/\delta) \text{ before } \frac{3B\delta^{p-\eta}}{2}\Big)\\
&\quad\le \mathbb{P}\Big(Y^r_\delta(\cdot) \text{ crosses } 3D'B\delta^{p-\eta}\log(1/\delta) \text{ before } \frac{3B\delta^{p-\eta}}{2} \,\Big|\, Y^r_\delta(0) = \frac{9B\delta^{p-\eta}}{2}\Big)\\
&\qquad + \mathbb{P}\Big(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \check{v}^r_l \mathbf{1}_{[\check{v}^r_l\le\delta c^r]} > \frac{9B\delta^{p-\eta}}{2}\Big)\\
&\quad\le \delta^{D'} + \mathbb{P}\Big(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \check{v}^r_l \mathbf{1}_{[\check{v}^r_l\le\delta c^r]} > \frac{9B\delta^{p-\eta}}{2}\Big).
\end{aligned}
\tag{5.93}
$$

Let $0 < \delta(\eta) \le \delta_*(\eta)$ be such that $T < \epsilon\delta(\eta)^{-2(p-\eta)}/4$. Choose $r(\eta) \ge \hat{r}(\eta)$ such that $2M(\eta)(c^r)^{-1} < \delta(\eta)$ for all $r \ge r(\eta)$. For $r \ge r(\eta)$ and $\delta \in [2M(\eta)(c^r)^{-1}, \delta(\eta)]$, by (5.72),

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_\delta(t) - Q^r_{\delta/2}(t)\big) > 6D'B\delta^{p-1-\eta}\log(1/\delta) + \frac{c^r}{r}\Big)\\
&\quad\le \mathbb{P}\Big(\sup_{t\in[0,\epsilon\delta^{-2(p-\eta)}/4]} \big(Q^r_\delta(t) - Q^r_{\delta/2}(t)\big) > 6D'B\delta^{p-1-\eta}\log(1/\delta) + \frac{c^r}{r}\Big)\\
&\quad\le \mathbb{P}\Big(\sup_{t\in[0,\epsilon\delta^{-2(p-\eta)}/4]} Y^r_\delta(t) > 3D'B\delta^{p-\eta}\log(1/\delta)\Big)\\
&\quad\le \mathbb{P}\Big(\sup_{t\in[0,\tau_0]} Y^r_\delta(t) > 3D'B\delta^{p-\eta}\log(1/\delta)\Big)\\
&\qquad + \mathbb{P}\Big(\sup_{t\in(0,\epsilon\delta^{-2(p-\eta)}/4]} Y^r_\delta(t) > 3D'B\delta^{p-\eta}\log(1/\delta),\\
&\qquad\quad \sup_{t\in[0,\tau_0]} Y^r_\delta(t) \le 3D'B\delta^{p-\eta}\log(1/\delta)\Big).
\end{aligned}
\tag{5.94}
$$

Observe that if $\sup_{t\in[0,\tau_0]} Y^r_\delta(t) \le 3D'B\delta^{p-\eta}\log(1/\delta)$, then

$$
\sup_{t\in[0,\tau_{2\mathcal{N}-1})} Y^r_\delta(t) \le 3D'B\delta^{p-\eta}\log(1/\delta),
$$

where $\mathcal{N}$ is given in (5.81). Then, if in addition $\mathcal{N} > \lfloor\delta^{-D'/2}\rfloor + 1$ and

$$
\sup_{t\in(0,\epsilon\delta^{-2(p-\eta)}/4]} Y^r_\delta(t) > 3D'B\delta^{p-\eta}\log(1/\delta),
$$

then $\tau_{2\mathcal{N}-1} \le \epsilon\delta^{-2(p-\eta)}/4$ and hence, in this case,

$$\sum_{j=0}^{\lfloor \delta^{-D'/2}\rfloor+1} (\tau_{2j+1}-\tau_{2j}) \le \tau_{2\mathcal{N}-1} \le \epsilon\delta^{-2(p-\eta)}/4. \tag{5.95}$$

This together with (5.4.1) gives that for $r \ge r(\eta)$ and $\delta \in [2M(\eta)(c^r)^{-1}, \delta(\eta)]$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_\delta(t) - Q^r_{\delta/2}(t)\big) > 6D'B\delta^{p-1-\eta}\log(1/\delta) + \frac{c^r}{r}\Big) \\
&\quad \le \mathbb{P}\Big(\sup_{t\in[0,\tau_0]} Y^r_\delta(t) > 3D'B\delta^{p-\eta}\log(1/\delta)\Big) \\
&\quad\quad + \mathbb{P}\Bigg(\sum_{j=0}^{\lfloor \delta^{-D'/2}\rfloor+1} (\tau_{2j+1}-\tau_{2j}) \le \epsilon\delta^{-2(p-\eta)}/4, \mathcal{N} > \lfloor \delta^{-D'/2}\rfloor + 1\Bigg) \\
&\quad\quad + \mathbb{P}\big(\mathcal{N} \le \lfloor \delta^{-D'/2}\rfloor + 1\big).
\end{aligned} \tag{5.96}$$

By (5.96), (5.93), the fact that $\delta^{D'/2}\epsilon\delta^{-2(p-\eta)} \le \epsilon\delta^{2(p-\eta)}$ since $D' \ge 8p$, (5.82) and (5.92), we obtain for $r \ge r(\eta)$ and $\delta \in [2M(\eta)(c^r)^{-1}, \delta(\eta)]$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_\delta(t) - Q^r_{\delta/2}(t)\big) > 6D'B\delta^{p-1-\eta}\log(1/\delta) + \frac{c^r}{r}\Big) \\
&\quad \le \delta^{D'} + \mathbb{P}\Bigg(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \check{v}^r_l \mathbf{1}_{[\check{v}^r_l \le \delta c^r]} > \frac{9B\delta^{p-\eta}}{2}\Bigg) \\
&\quad\quad + \mathbb{P}\Bigg(\sum_{j=0}^{\lfloor \delta^{-D'/2}\rfloor+1} (\tau_{2j+1}-\tau_{2j}) \le \big(\lfloor \delta^{-D'/2}\rfloor + 1\big)\epsilon\delta^{2(p-\eta)}/4\Bigg) + 2\delta^{D'/2} \\
&\quad \le \mathbb{P}\Bigg(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \check{v}^r_l \mathbf{1}_{[\check{v}^r_l \le \delta c^r]} > \frac{9B\delta^{p-\eta}}{2}\Bigg) + e^{-\delta^{-D'/2}/32} + 3\delta^{D'/2} \\
&\quad \le \mathbb{P}\Bigg(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \check{v}^r_l \mathbf{1}_{[\check{v}^r_l \le \delta c^r]} > \frac{9B\delta^{p-\eta}}{2}\Bigg) + 35\delta^{D'/2},
\end{aligned}$$

where, in the last inequality, we used the fact that $xe^{-x/32} \le 32$ for all $x \ge 1$. This proves the lemma with $D_1 := 6D'B$, $D_2 := D'/2$ and $D_3 := 9B/2$. □

LEMMA 19. *Fix $T > 0$. Recall the constant $D_2 > 0$ from Lemma 18. For any $\eta \in (\eta^*, p-1)$, there are $\tilde{\theta}_\eta \in \Theta$, and positive constants $r'(\eta)$, $D'(\eta)$, $\widetilde{D}(\eta)$, $\delta(\eta) \in (0,1)$, $M'(\eta) > 1$ such that for all $r \ge r'(\eta)$ and $\delta \in [2M'(\eta)(c^r)^{-1}, \delta(\eta)]$,*

$$\mathbb{P}\Big(\sup_{t\in[0,T]} Z^r_\delta(t) > D'(\eta)\delta^{p-1-\eta}\big(1+\log(\delta^{-1})\big) + \tilde{\theta}_\eta(r)\Big) \le \widetilde{D}(\eta)\big(\delta^{D_2} + \delta^{\eta-\eta^*}\big) + \tilde{\theta}_\eta(r).$$

PROOF. By (5.4) in Proposition 10, for any $r \in \mathcal{R}$ and any $\delta, z \ge 0$,

$$\mathbb{P}\Big(\sup_{t\in[0,T]} Z^r_\delta(t) > z\Big) \le \mathbb{P}\Big(\sup_{t\in[0,T]} Q^r_\delta(t) > z - \frac{c^r}{r}\Big). \tag{5.97}$$

Take D_1, D_2, D_3 as in Lemma 18. Choose and fix $\eta \in (\eta^*, p-1)$ and obtain $M(\eta) > 1$ and $r(\eta) \ge 2$, $\delta(\eta) \in (0,1)$ as in Lemma 18. Define $M'(\eta) := M(\eta) \vee a^*$ where a^* appears in Assumption (2.16). Denote by θ_η and $r_0(\eta)$ the map θ and constant r_0 obtained in

Lemma 16 with $2M(\eta)$ in place of a. Define $D'(\eta) := D_1 \sum_{k=0}^{\infty} 2^{-k(p-1-\eta)}(1+k\log 2)$. For $\delta \in [2M'(\eta)(c^r)^{-1}, \delta(\eta)]$, let $K(\eta,\delta,r)$ be a nonnegative integer such that $2^{-K(\eta,\delta,r)-1}\delta < 2M'(\eta)(c^r)^{-1} \le 2^{-K(\eta,\delta,r)}\delta$. This, along with (5.66), implies that for $r \ge r(\eta)$ and $\delta \in [2M'(\eta)(c^r)^{-1}, \delta(\eta)]$,

$$K(\eta,\delta,r) \le \log_2\left(\frac{\delta c^r}{2M'(\eta)}\right) \le \log_2\left(\frac{c^r}{M'(\eta)}\right) \le C'(\eta)\log r, \tag{5.98}$$

where $C'(\eta) = 2\log_2(e)/(p-\eta/2)$ depends only on η (and p). Observe that for any $r \in \mathcal{R}$ and $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} Q^r_\delta(t) > D'(\eta)\delta^{p-1-\eta}\big[1+\log(1/\delta)\big] + C'(\eta)\frac{c^r\log r}{r} + \theta_\eta(r)\Big) \\
&\le \mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_\delta(t) - Q^r_{2M'(\eta)(c^r)^{-1}}(t)\big) \\
&\qquad > D'(\eta)\delta^{p-1-\eta}\big[1+\log(1/\delta)\big] + C'(\eta)\frac{c^r\log r}{r}\Big) \\
&\qquad + \mathbb{P}\Big(\sup_{t\in[0,T]} Q^r_{2M'(\eta)(c^r)^{-1}}(t) > \theta_\eta(r)\Big).
\end{aligned} \tag{5.99}$$

By Lemma 18, for every $r \ge r(\eta)$ and $\delta \in [2M'(\eta)(c^r)^{-1}, \delta(\eta)]$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_{2^{-k}\delta}(t) - Q^r_{2^{-k-1}\delta}(t)\big) > D_1\big(2^{-k}\delta\big)^{p-1-\eta}\log\big(2^k/\delta\big) + \frac{c^r}{r}\Big) \\
&\le 35\big(2^{-k}\delta\big)^{D_2} + \mathbb{P}\Big(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \breve{v}^r_l \mathbf{1}_{[\breve{v}^r_l \le 2^{-k}\delta c^r]} > D_3\big(2^{-k}\delta\big)^{p-\eta}\Big) \\
&\qquad \text{for all } 0 \le k \le K(\eta,\delta,r).
\end{aligned} \tag{5.100}$$

By Assumption (2.16) and (5.98) (and since $M'(\eta) \ge a^*$), there exist $C'', r'' > 0$ such that for all $r \ge r''$, $\delta \in [2M'(\eta)(c^r)^{-1}, \delta(\eta)]$, and $0 \le k \le K(\eta,\delta,r)$,

$$\mathbb{E}\Big(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \breve{v}^r_l \mathbf{1}_{[\breve{v}^r_l \le 2^{-k}\delta c^r]}\Big) \le C''\big(2^{-k}\delta\big)^{p-\eta^*}. \tag{5.101}$$

Let $r'(\eta) := \max\{r(\eta), r_0(\eta), r''\}$. For $r \ge r'(\eta)$ and $\delta \in [2M'(\eta)(c^r)^{-1}, \delta(\eta)]$, since $2^{-K(\eta,\delta,r)-1}\delta < 2M'(\eta)(c^r)^{-1}$, by Lemma 11, for any $t \ge 0$,

$$\begin{aligned}
\big(Q^r_\delta(t) - Q^r_{2M'(\eta)(c^r)^{-1}}(t)\big) &= \big(Q^r_\delta(t) - Q^r_{2^{-K(\eta,\delta,r)-1}\delta}(t)\big) \\
&\quad + \big(Q^r_{2^{-K(\eta,\delta,r)-1}\delta}(t) - Q^r_{2M'(\eta)(c^r)^{-1}}(t)\big) \\
&\le \big(Q^r_\delta(t) - Q^r_{2^{-K(\eta,\delta,r)-1}\delta}(t)\big).
\end{aligned}$$

Using this observation, along with (5.98), (5.100), (5.101), Markov's inequality and the union bound, for any $r \geq r'(\eta)$ and $\delta \in [2M'(\eta)(c^r)^{-1}, \delta(\eta)]$,

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_{\delta}(t) - Q^r_{2M'(\eta)(c^r)^{-1}}(t)\big) \\
&\qquad > D'(\eta)\delta^{p-1-\eta}\big(1+\log(1/\delta)\big) + C'(\eta)\frac{c^r \log r}{r} + \frac{c^r}{r}\Big) \\
&\leq \mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_{\delta}(t) - Q^r_{2^{-K(\eta,\delta,r)-1}\delta}(t)\big) \\
&\qquad > D'(\eta)\delta^{p-1-\eta}\big(1+\log(1/\delta)\big) + C'(\eta)\frac{c^r \log r}{r} + \frac{c^r}{r}\Big) \\
&\leq \sum_{k=0}^{K(\eta,\delta,r)} \mathbb{P}\Big(\sup_{t\in[0,T]} \big(Q^r_{2^{-k}\delta}(t) - Q^r_{2^{-k-1}\delta}(t)\big) \\
&\qquad > D_1\big(2^{-k}\delta\big)^{p-1-\eta}\log\big(2^k/\delta\big) + c^r/r\Big) \qquad (5.102) \\
&\leq \sum_{k=0}^{K(\eta,\delta,r)} 35\big(2^{-k}\delta\big)^{D_2} + \sum_{k=0}^{K(\eta,\delta,r)} \mathbb{P}\Big(\frac{1}{r}\sum_{l=1}^{\mathbf{q}^r} \check{v}^r_l \mathbf{1}_{[\check{v}^r_l \leq 2^{-k}\delta c^r]} > D_3\big(2^{-k}\delta\big)^{p-\eta}\Big) \\
&\leq \sum_{k=0}^{K(\eta,\delta,r)} 35\big(2^{-k}\delta\big)^{D_2} + \sum_{k=0}^{K(\eta,\delta,r)} \big(D_3\big(2^{-k}\delta\big)^{p-\eta}\big)^{-1} C''\big(2^{-k}\delta\big)^{p-\eta^*} \\
&\leq 35\delta^{D_2}\sum_{k=0}^{\infty} 2^{-D_2 k} + C''(D_3)^{-1}\delta^{\eta-\eta^*}\sum_{k=0}^{\infty} 2^{-(\eta-\eta^*)k} \leq \widetilde{D}(\eta)\big(\delta^{D_2} + \delta^{\eta-\eta^*}\big),
\end{aligned}
$$

where $\widetilde{D}(\eta) := 35\sum_{k=0}^{\infty} 2^{-D_2 k} + C''(D_3)^{-1}\sum_{k=0}^{\infty} 2^{-(\eta-\eta^*)k} \in (0,\infty)$. Finally, by Lemma 16, for any $r \geq r'(\eta)$),

$$
\mathbb{P}\Big(\sup_{t\in[0,T]} Q^r_{2M(\eta)(c^r)^{-1}}(t) > \theta_\eta(r)\Big) \leq \theta_\eta(r). \qquad (5.103)
$$

Taking $\tilde{\theta}_\eta(r) = C'(\eta)\frac{c^r \log r}{r} + \theta_\eta(r) + \frac{2c^r}{r}$, the lemma now follows from (5.97), (5.99), (5.102) and (5.103). □

REMARK 6. By small modifications of some of the estimates in Lemmas 16, 18 and 19, it can in fact be shown that for a sequence of systems such that each system has no jobs in system at time zero, for any $T > 0$ and any $\eta \in (0, p-1)$, there exist positive constants C, C', C'', r_0 such that for any $r \geq r_0$, $a \in [(c^r)^{-1}, 1]$ and $z \geq 0$,

$$
\mathbb{P}\Big(\sup_{t\in[0,T]} W^r_a(t) > Ca^{p-\eta}z\Big) \leq \mathbb{P}\Big(\sup_{t\in[0,T]} Z^r_a(t) > Ca^{p-\eta-1}z\Big) \leq C'e^{-C''z},
$$

where we have used the elementary bound $W^r_a(t) \leq aZ^r_a(t)$ for $t \geq 0$ to obtain the first inequality. By integrating over z, this immediately implies that, in this case, Assumption (2.16) holds with $W^r_a(0)$ replaced by $W^r_a(t)$ for any fixed $t > 0$.

The next two lemmas concern the limiting random field $\{W_a(\cdot), a \in [0,\infty]\}$. In preparation for using these two results both in the proof of Theorem 2 and in the proof of Theorem 5 (which concerns asymptotic state space collapse as $p \to \infty$), the dependence on p of various objects is made explicit in the statements of these lemmas. In this regard, we remind the

reader that $p > 1$ is presently fixed and therefore, the asymptotic conditions of Section 3.4 need not hold for the results in these lemmas to be true.

Recall $\sigma(p) = \sqrt{\lambda \operatorname{Var}(v^{(p)}) + \lambda\sigma_A^2}$, where $v^{(p)}$ denotes the job processing time distribution with highlighted dependence on p. Also recall $\eta^* = \eta^*(p)$ in Assumption (2.16).

LEMMA 20. *Let $T > 0$. Set $m_0(p) := \max\{2, \lambda, 4\kappa^2/\lambda^2, e^{\lambda/(\sigma(p))^2}, T\}$, $a_0(p) = m_0(p)^{-1/2p}$ and $H_0(p) := 8p(\sigma(p))^2/\lambda$. Then $a_0(p) \in (0,1)$ and for all $a \in (0, a_0(p))$, $\eta \in (\eta^*(p), p-1)$ and $H \geq H_0(p)$, we have*

$$\mathbb{P}\Big(\sup_{t\in[0,T]} W_a^{(p)}(t) > a^{p-\eta} + Ha^p \log(1/a)\Big)$$
$$\leq C_0(p)a^{\eta-\eta^*(p)} + e^{-\lambda/(2(\sigma(p))^2 a^\eta)} + C\big(\lambda, \sigma(p)\big)a^{2p},$$

where $C_0(p) := 2\sup_{a>0} a^{-(p-\eta^*(p))}\mathbb{E}(\xi^{(p)}(a)) < \infty$ due to (2.18) and $C(\lambda, \sigma(p)) := 2e^{\lambda/(\sigma(p))^2} + \frac{16(\sigma(p))^2}{\lambda}$.

PROOF. Since $m_0(p) > 1$, we have $a_0(p) \in (0,1)$. Fix $a \in (0, a_0(p))$, $\eta \in (\eta^*(p), p-1)$ and $H \geq H_0(p)$. To ease the notation in this proof, we suppress the dependence on p and write $m_0 = m_0(p)$, $a_0 = a_0(p)$, $H_0 = H_0(p)$, $\sigma = \sigma(p)$, $\eta^* = \eta^*(p)$ and $C_0 = C_0(p)$. Observe that since $H\log(1/a) > H\log(1/a_0) = H\log(m_0)/2p \geq \lambda H/(2p\sigma^2) \geq 4 > 1$, we have $Ha^p \log(1/a) > a^p$. Define the stopping times: $\tau_0^* := \inf\{t \geq 0 : W_a(t) = 0\}$, and for $k \in \mathbb{Z}_+$,

$$\tau_{2k+1}^* := \inf\big\{t \geq \tau_{2k}^* : W_a(t) = a^p\big\},$$
$$\tau_{2k+2}^* := \inf\big\{t \geq \tau_{2k+1}^* : W_a(t) = 0 \text{ or } W_a(t) = Ha^p\log(1/a)\big\}.$$

Define $\mathcal{N}^* := \inf\{k \in \mathbb{N} : W_a(\tau_{2k}^*) = Ha^p\log(1/a)\}$. Since $\kappa < \lambda/(2a^p)$, we have $\kappa - \frac{\lambda}{a^p} < -\frac{\lambda}{2a^p}$. Thus, by (4.3), the process $\Gamma[\overline{X}_a](\cdot)$ with $\overline{X}_a(t) := \xi(a) + \sigma B(t) - \frac{\lambda t}{2a^p}$, $t \geq 0$, dominates the process $W_a(\cdot)$ pointwise. Thus, using the fact that $t \mapsto e^{\lambda\overline{X}_a(t)/(\sigma^2 a^p)}$ is a martingale (with respect to the filtration $\{\mathcal{G}_t\}_{t\geq 0}$ given by $\mathcal{G}_t = \sigma(\overline{X}_a(0), (B(s), 0 \leq s \leq t))$ for $t \geq 0$), by the optional stopping theorem and the strong Markov property,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{[0,\tau_0^*]} W_a(t) > a^{p-\eta}\Big)\\
&\leq \mathbb{P}\big(\xi(a) > a^{p-\eta}/2\big) + \mathbb{P}\Big(\Big\{\sup_{[0,\tau_0^*]} W_a(t) > a^{p-\eta}\Big\} \cap \big\{\xi(a) \leq a^{p-\eta}/2\big\}\Big)\\
&\leq 2a^{\eta-p}\mathbb{E}\big(\xi(a)\big) + \mathbb{P}\big(\overline{X}_a(t+\cdot) \text{ crosses } a^{p-\eta} \text{ before } 0 \mid \overline{X}_a(t) = a^{p-\eta}/2\big)\\
&\leq 2a^{\eta-p}\mathbb{E}\big(\xi(a)\big) + \frac{e^{\lambda/(2\sigma^2 a^\eta)} - 1}{e^{\lambda/(\sigma^2 a^\eta)} - 1} \leq C_0 a^{\eta-\eta^*} + e^{-\lambda/(2\sigma^2 a^\eta)}.
\end{aligned} \tag{5.104}$$

As previously noted, $H\log(1/a) > 1$. This together with an argument using the optional stopping theorem in a manner similar to the above gives

$$\mathbb{P}\big(W_a(\tau_2^*) = Ha^p\log(1/a)\big) \leq \mathbb{P}\big(\overline{X}_a(t+\cdot) \text{ crosses } Ha^p\log(1/a) \text{ before } 0 \mid \overline{X}_a(t) = a^p\big)$$
$$= \frac{e^{\lambda/\sigma^2} - 1}{e^{\lambda H\log(1/a)/\sigma^2} - 1} < e^{\lambda(1-H\log(1/a))/\sigma^2} = e^{\lambda/\sigma^2} a^{\lambda H/\sigma^2}.$$

Using a union bound and the strong Markov property, this implies

$$\begin{aligned}
&\mathbb{P}\big(\mathcal{N}^* < \lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 2\big) \\
&\quad \le \big(a^{-3H\lambda/(4\sigma^2)} + 1\big) e^{\lambda/\sigma^2} a^{\lambda H/\sigma^2} \\
&\quad \le \big(1 + a^{3H\lambda/(4\sigma^2)}\big) e^{\lambda/\sigma^2} a^{\lambda H/4\sigma^2} \le 2e^{\lambda/\sigma^2} a^{H\lambda/(4\sigma^2)} \le 2e^{\lambda/\sigma^2} a^{2p}.
\end{aligned} \tag{5.105}$$

Again, by our choice of a, H and a_0, $a^{-(\frac{H\lambda}{2\sigma^2}-2p)} \ge a^{-2p} > a_0^{-2p} \ge T$ and hence

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} W_a(t) > a^{p-\eta} + Ha^p \log(1/a)\Big) \\
&\quad \le \mathbb{P}\Big(\sup_{[0,\tau_0^*\wedge T]} W_a(t) > a^{p-\eta}\Big) + \mathbb{P}\Big(\sup_{t\in[\tau_0^*\wedge T, a^{-(\frac{H\lambda}{2\sigma^2}-2p)}]} W_a(t) > Ha^p \log(1/a)\Big) \\
&\quad \le \mathbb{P}\Big(\sup_{[0,\tau_0^*\wedge T]} W_a(t) > a^{p-\eta}\Big) + \mathbb{P}\Bigg(\sum_{k=1}^{\mathcal{N}^*-1} \big(\tau_{2k}^* - \tau_{2k-1}^*\big) < a^{-(\frac{H\lambda}{2\sigma^2}-2p)}\Bigg) \\
&\quad \le \mathbb{P}\Big(\sup_{[0,\tau_0^*\wedge T]} W_a(t) > a^{p-\eta}\Big) + \mathbb{P}\big(\mathcal{N}^* < \lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 2\big) \\
&\qquad + \mathbb{P}\Bigg(\sum_{k=1}^{\lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 1} \big(\tau_{2k}^* - \tau_{2k-1}^*\big) < a^{-(\frac{H\lambda}{2\sigma^2}-2p)}, \mathcal{N}^* \ge \lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 2\Bigg).
\end{aligned} \tag{5.106}$$

Denote by σ^x the hitting time of level $x \le 0$ by the process $\{\sigma B(t) - \lambda t/2a^p, t \ge 0\}$, and let $\{\sigma_k^x\}_{k\in\mathbb{N}}$ be independent and identically distributed copies of σ^x. For each $x < 0$, by the explicit form of the moment generating function of σ^x (see Exercise 5.10 in Chapter 3.5.C of [16]),

$$\mathbb{E}(\sigma^x) = \frac{2a^p|x|}{\lambda} \quad \text{and} \quad \operatorname{Var}(\sigma^x) = \frac{8a^{3p}\sigma^2|x|}{\lambda^3}.$$

Thus, again using the strong Markov property, $a^{-H\lambda/(4\sigma^2)} \ge a^{-2p} > a_0^{-2p} \ge \lambda$ and Chebyshev's inequality,

$$\begin{aligned}
&\mathbb{P}\Bigg(\sum_{k=1}^{\lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 1} \big(\tau_{2k}^* - \tau_{2k-1}^*\big) < a^{-(\frac{H\lambda}{2\sigma^2}-2p)}, \mathcal{N}^* \ge \lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 2\Bigg) \\
&\quad \le \mathbb{P}\Bigg(\sum_{k=1}^{\lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 1} \sigma_k^{-a^p} < a^{-(\frac{H\lambda}{2\sigma^2}-2p)}\Bigg) \\
&\quad \le \mathbb{P}\Bigg(\sum_{k=1}^{\lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 1} \Big(\sigma_k^{-a^p} - \frac{2a^{2p}}{\lambda}\Big) < a^{-(\frac{H\lambda}{2\sigma^2}-2p)} - \frac{2a^{2p}a^{-3H\lambda/(4\sigma^2)}}{\lambda}\Bigg) \\
&\quad \le \mathbb{P}\Bigg(\sum_{k=1}^{\lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 1} \Big(\sigma_k^{-a^p} - \frac{2a^{2p}}{\lambda}\Big) < -\frac{a^{2p}a^{-3H\lambda/(4\sigma^2)}}{\lambda}\Bigg) \\
&\quad \le \frac{8\sigma^2 a^{4p}(\lfloor a^{-3H\lambda/(4\sigma^2)} \rfloor + 1)}{\lambda a^{4p-3H\lambda/(2\sigma^2)}} \le \frac{16\sigma^2 a^{3H\lambda/(4\sigma^2)}}{\lambda}.
\end{aligned} \tag{5.107}$$

Finally, using (5.104), (5.105) and (5.107) in (5.106), we obtain the lemma. □

LEMMA 21. *Let $T > 0$ and let $a_0(p)$, $H_0(p)$, $C_0(p)$ and $C(\lambda, \sigma(p))$ be as in Lemma 20. Then for all $\delta \in (0, a_0(p))$ and $\eta \in (\eta^*(p), p-1)$,*

$$\mathbb{P}\Big(\sup_{t\in[0,T]}\Big(\int_0^\delta x^{-2}W_x^{(p)}(t)\,dx + \delta^{-1}W_\delta^{(p)}(t)\Big) > H(p,\eta)\delta^{p-\eta-1}\big(1+\log(1/\delta)\big)\Big)$$
$$\le \tilde{C}\big(\eta, \eta^*(p), \lambda, \sigma(p)\big)\delta^{\eta-\eta^*(p)} + 3C\big(\lambda, \sigma(p)\big)\delta^{2p},$$

where

$$H(p,\eta) := H_0(p)\Big[1 + \frac{2^{p-1}}{(2^{p-1}-1)} + \frac{(\log 2)2^{p-1}}{(2^{p-1}-1)^2}\Big] + 1 + \sum_{k=1}^{\infty} 2^{-(k-1)(p-\eta-1)},$$

$$\tilde{C}\big(\eta, \eta^*(p), \lambda, \sigma(p)\big) := \Big(C_0(p) + \frac{2(\sigma(p))^2}{\lambda}\Big(\sup_{x\in\mathbb{R}_+} xe^{-x}\Big)\Big)\Big(1 + \sum_{k=1}^{\infty} 2^{-(k-1)(\eta-\eta^*(p))}\Big).$$

In particular, $\sup_{t\in[0,T]}\int_0^\infty x^{-2}W_x^{(p)}(t)\,dx < \infty$ *almost surely.*

PROOF. As in the proof of Lemma 20, we suppress the dependence on p in this proof to ease the notation in what follows. Fix $\delta \in (0, a_0)$ and $\eta \in (\eta^*, p-1)$ and set $H = H_0$. As for any $x_1 < x_2$, $X_{x_2}(t) - X_{x_1}(t)$ is nonnegative and nondecreasing in t, using the monotonicity property in (4.3), we obtain for $t \ge 0$,

$$\begin{aligned}\int_0^\delta x^{-2}W_x(t)\,dx &= \sum_{k=1}^{\infty}\int_{\delta 2^{-k}}^{\delta 2^{-(k-1)}} x^{-2}W_x(t)\,dx \\ &\le \sum_{k=1}^{\infty} W_{\delta 2^{-(k-1)}}(t)\int_{\delta 2^{-k}}^{\delta 2^{-(k-1)}} x^{-2}\,dx = \sum_{k=1}^{\infty}\frac{W_{\delta 2^{-(k-1)}}(t)}{\delta 2^{-(k-1)}}.\end{aligned} \tag{5.108}$$

By Lemma 20, for any $k \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{P}\Big(\sup_{t\in[0,T]}\frac{W_{\delta 2^{-(k-1)}}(t)}{\delta 2^{-(k-1)}} > \big(\delta 2^{-(k-1)}\big)^{p-\eta-1} + H_0\big(\delta 2^{-(k-1)}\big)^{p-1}\log\frac{1}{\delta 2^{-(k-1)}}\Big) \\ &\le C_0\big(\delta 2^{-(k-1)}\big)^{\eta-\eta^*} + e^{-\lambda/(2\sigma^2(\delta 2^{-(k-1)})^\eta)} + C(\lambda,\sigma)\big(\delta 2^{-(k-1)}\big)^{2p}.\end{aligned} \tag{5.109}$$

Also,

$$\begin{aligned}&H_0\sum_{k=1}^{\infty}\big(\delta 2^{-(k-1)}\big)^{p-1}\log\frac{1}{\delta 2^{-(k-1)}} \\ &= H_0\Big[\frac{2^{p-1}}{2^{p-1}-1}\delta^{p-1}\log(1/\delta) + \frac{(\log 2)2^{p-1}}{(2^{p-1}-1)^2}\delta^{p-1}\Big] \le H_1(p)\delta^{p-1}\big(1+\log(1/\delta)\big),\end{aligned}$$

where

$$H_1(p) := H_0\Big[\frac{2^{p-1}}{2^{p-1}-1} + \frac{(\log 2)2^{p-1}}{(2^{p-1}-1)^2}\Big].$$

Moreover,

$$\sum_{k=1}^{\infty}\big(2^{-(k-1)}\delta\big)^{p-\eta-1} = H_2(p,\eta)\delta^{p-\eta-1},$$

where $H_2(p,\eta) := \sum_{k=1}^{\infty} 2^{-(k-1)(p-\eta-1)}$. Using these observations, (5.108), a union bound and (5.109),

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]}\int_0^{\delta} x^{-2}W_x(t)\,dx > H_1(p)\delta^{p-1}\big(1+\log(1/\delta)\big)+H_2(p,\eta)\delta^{p-\eta-1}\Big)\\
&\quad\le \mathbb{P}\Big(\sup_{t\in[0,T]}\sum_{k=1}^{\infty}\frac{W_{\delta 2^{-(k-1)}}(t)}{\delta 2^{-(k-1)}} > H_1(p)\delta^{p-1}\big(1+\log(1/\delta)\big)+H_2(p,\eta)\delta^{p-\eta-1}\Big)\\
&\quad\le \sum_{k=1}^{\infty}\mathbb{P}\Big(\sup_{t\in[0,T]}\frac{W_{\delta 2^{-(k-1)}}(t)}{\delta 2^{-(k-1)}} > \big(\delta 2^{-(k-1)}\big)^{p-\eta-1}\\
&\qquad + H_0\big[\delta 2^{-(k-1)}\big]^{p-1}\log\frac{1}{\delta 2^{-(k-1)}}\Big) \qquad (5.110)\\
&\quad\le \hat{C}(\eta,\eta^*,\lambda,\sigma)\delta^{\eta-\eta^*}+C(\lambda,\sigma)\sum_{k=1}^{\infty}\big(\delta 2^{-(k-1)}\big)^{2p}\\
&\quad\le \hat{C}(\eta,\eta^*,\lambda,\sigma)\delta^{\eta-\eta^*}+2C(\lambda,\sigma)\delta^{2p},
\end{aligned}
$$

where $\hat{C}(\eta,\eta^*,\lambda,\sigma) := (C_0+\frac{2\sigma^2}{\lambda}(\sup_{x\in\mathbb{R}_+} xe^{-x}))(\sum_{k=1}^{\infty}2^{-(k-1)(\eta-\eta^*)})$. By taking $k=1$ in (5.109), we obtain

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]}\delta^{-1}W_{\delta}(t) > \delta^{p-\eta-1}+H_0\delta^{p-1}\log\Big(\frac{1}{\delta}\Big)\Big)\\
&\quad\le C_0\delta^{\eta-\eta^*}+e^{-\lambda/(2\sigma^2\delta^{\eta})}+C(\lambda,\sigma)\delta^{2p} \qquad (5.111)\\
&\quad\le \Big(C_0+\frac{2\sigma^2}{\lambda}\Big(\sup_{x\in\mathbb{R}_+}xe^{-x}\Big)\Big)\delta^{\eta-\eta^*}+C(\lambda,\sigma)\delta^{2p}.
\end{aligned}
$$

The first assertion of the lemma follows from (5.110) and (5.111) upon noting that $\tilde{C}(\eta,\eta^*,\lambda,\sigma) = \hat{C}(\eta,\eta^*,\lambda,\sigma) + (C_0 + \frac{2\sigma^2}{\lambda}(\sup_{x\in\mathbb{R}_+}xe^{-x}))$ and $H(p,\eta) = H_1(p) + H_2(p,\eta) + 1 + \frac{8\sigma^2 p}{\lambda}$.

Now, we check the last assertion. If $\mathbb{P}(\sup_{t\in[0,T]}\int_0^1 x^{-2}W_x(t)\,dx=\infty)>0$, by the finiteness of $\sup_{t\in[0,T]}\int_{\delta}^1 x^{-2}W_x(t)\,dx$ for all $\delta\in(0,1]$, there exists $\epsilon>0$ such that

$$
\mathbb{P}\Big(\sup_{t\in[0,T]}\int_0^{\delta}x^{-2}W_x(t)\,dx=\infty\Big)\ge\epsilon
$$

for all $\delta>0$, which contradicts the first assertion of the lemma. Thus,

$$
\sup_{t\in[0,T]}\int_0^1 x^{-2}W_x(t)\,dx<\infty \quad \text{almost surely.} \qquad (5.112)
$$

Moreover, by the monotonicity property noted previously

$$
\sup_{t\in[0,T]}\int_1^{\infty}x^{-2}W_x(t)\,dx\le\sup_{t\in[0,T]}W_{\infty}(t)<\infty \quad \text{almost surely.} \qquad (5.113)
$$

The last assertion of the lemma follows from (5.112) and (5.113). □

PROOF OF THEOREM 2. Fix a C^1 function $f:[0,\infty)\to\mathbb{R}$ such that $\lim_{x\to\infty}\frac{f(x)}{x}$ exists and $\int_1^{\infty}\frac{|f'(x)|}{x^{\alpha^*+1}}<\infty$. Set $g(x)=f(x)/x$ for $x>0$ and define $g(\infty)=\lim_{x\to\infty}g(x)$. By

Lemma 15, for each $\delta > 0$, as $r \to \infty$,

$$\int_\delta^\infty f(x)\widetilde{\mathcal{Z}}^r(\cdot)(dx) \xrightarrow{d} -\int_\delta^\infty g'(x)W_x(\cdot)\,dx + g(\infty)W_\infty(\cdot) - g(\delta)W_\delta(\cdot). \tag{5.114}$$

Moreover, for all $r \in \mathcal{R}$, $\int_0^\infty f(x)\widetilde{\mathcal{Z}}^r(t)(dx)$ is finite for all $t \in [0,T]$ almost surely. Fix $\eta \in (\eta^*, p-1)$. Define $C_f := \sup_{z\in[0,1]} |f(z)|$ and let D_2, $D'(\eta)$ and $\tilde{D}(\eta)$ as in Lemma 19. For each $\delta > 0$, let $b(\delta) := \max\{2C_f D'(\eta)\delta^{p-1-\eta}(1+\log(1/\delta)), 2\widetilde{D}(\eta)(\delta^{D_2} + \delta^{\eta-\eta^*})\}$. Then, by Lemma 19, for any $0 < \delta \le \delta(\eta)$,

$$\limsup_{r\to\infty} \mathbb{P}\Big(\sup_{[0,T]}\Big|\int_0^\delta f(x)\widetilde{\mathcal{Z}}^r(t)(dx)\Big| > b(\delta)\Big) < b(\delta). \tag{5.115}$$

As f is C^1 on $[0,\infty)$, $g(x) \le C_f x^{-1}$ for all $x \in (0,1]$, and $g'(x) = \frac{-f(x)}{x^2} + \frac{f'(x)}{x}$, $x > 0$, satisfies $|g'(x)| \le C'_f x^{-2}$ for all $x \in (0,1]$ for some constant $C'_f > 0$. Thus, by Lemma 21, $-\int_0^\infty g'(x)W_x(t)\,dx + g(\infty)W_\infty(t)$ is well defined and finite for all $t \in [0,T]$ almost surely, $g(\delta)W_\delta(\cdot) \to 0$ in probability uniformly over compact time intervals as $\delta \to 0$, and

$$\begin{aligned} &-\int_\delta^\infty g'(x)W_x(\cdot)\,dx + g(\infty)W_\infty(\cdot) - g(\delta)W_\delta(\cdot) \\ &\xrightarrow{d} -\int_0^\infty g'(x)W_x(\cdot)\,dx + g(\infty)W_\infty(\cdot) \end{aligned} \tag{5.116}$$

as $\delta \to 0$, in $\mathcal{D}([0,\infty):\mathbb{R})$. By Lemma 21 and the monotonicity of $\int_0^\delta x^{-2}W_x(t)\,dx$ in δ,

$$\sup_{t\in[0,T]} \int_0^\delta x^{-2}W_x(t)\,dx \to 0 \text{as } \delta \to 0 \quad \text{almost surely.}$$

This implies that, almost surely, $\int_\delta^\infty g'(x)W_x(\cdot)\,dx$ converges to $\int_0^\infty g'(x)W_x(\cdot)\,dx$ as $\delta \to 0$ uniformly in $t \in [0,T]$. Moreover, for any $\delta > 0$, by Lemma 15, $\int_\delta^\infty g'(x)W_x(\cdot)\,dx$ lies in $\mathcal{C}([0,T]:\mathbb{R})$. Thus, due to uniform convergence, $\int_0^\infty g'(x)W_x(\cdot)\,dx$ lies in $\mathcal{C}([0,T]:\mathbb{R})$ as well. The theorem follows from this observation, (5.114), (5.115), (5.116) and Lemma 7. □

REMARK 7. Along the lines of the proof of Theorem 2 one can analyze the convergence of $\int_\delta^{b_1} f(x)\widetilde{\mathcal{Z}}^r(\cdot)(dx)$ as $\delta \to 0$, where $b_1 \in (0,\infty)$, and conclude that a_1 in Theorem 14 can be taken to be 0.

5.5. *Proofs of Theorems 3 and 4.*

PROOF OF THEOREM 3. We will use Theorem 2.1 in [26]. This theorem says the following. Let $\{f_n\}_{n\ge1}$ be a countable collection of real-valued continuous functions with compact support on $\mathbb{R}_+$ which is dense in $\mathcal{C}_0(\mathbb{R}_+)$ [the space of continuous functions on $\mathbb{R}_+$ vanishing at ∞ equipped with the uniform metric]. Let $f_0 = 1$. Suppose that

$$\{Z_f^r(\cdot) = \langle f, \widetilde{\mathcal{Z}}^r(\cdot)\rangle, r \in \mathcal{R}\} \text{ is tight in } \mathcal{D}([0,T]:\mathbb{R}) \quad \text{for every } f \in \{f_n\}_{n\in\mathbb{N}_0}. \tag{5.117}$$

Then $\{\widetilde{\mathcal{Z}}^r(\cdot), r \in \mathcal{R}\}$ is tight in $\mathcal{D}([0,T]:\mathcal{M}_F)$.

By Theorem 2,

$$\int_0^\infty f_0(x)\widetilde{\mathcal{Z}}^r(\cdot)(dx) \xrightarrow{d} \int_0^\infty f_0(x)\widetilde{\mathcal{Z}}(\cdot)(dx) \quad \text{as } r \to \infty. \tag{5.118}$$

Let

$$\mathcal{C} := \Bigg\{ h = \sum_{j=1}^{J} c_j \mathbf{1}_{(a_j, b_j]} : J \in \mathbb{N}, 0 \le a_1 < b_1 \le a_2 < b_2 \cdots \le a_J < b_J < \infty, \\ c_j \in \mathbb{R} \text{ for all } 1 \le j \le J \Bigg\}.$$

By Theorem 14 and Remark 5, for any $h \in \mathcal{C}$,

$$\int_0^\infty h(x) \widetilde{\mathcal{Z}}^r(\cdot)(dx) \xrightarrow{d} \int_0^\infty h(x) \widetilde{\mathcal{Z}}(\cdot)(dx) \quad \text{as } r \to \infty. \tag{5.119}$$

Now, fix $T > 0$ and take any compactly supported real-valued continuous function f and let $\{h_k\}_{k \in \mathbb{N}}$ be a sequence in $\mathcal{C}$ such that $\|h_k - f\|_\infty \le k^{-1}$ for $k \in \mathbb{N}$. Thus, for any $k \in \mathbb{N}$,

$$\sup_{t \in [0,T]} \left| \int_0^\infty h_k(x) \widetilde{\mathcal{Z}}^r(t)(dx) - \int_0^\infty f(x) \widetilde{\mathcal{Z}}^r(t)(dx) \right| \le k^{-1} \sup_{t \in [0,T]} \int_0^\infty \widetilde{\mathcal{Z}}^r(t)(dx).$$

By Theorem 2, and the continuous mapping theorem,

$$\sup_{t \in [0,T]} \int_0^\infty \widetilde{\mathcal{Z}}^r(t)(dx) \xrightarrow{d} \sup_{t \in [0,T]} Q(t) \quad \text{as } r \to \infty,$$

where we recall that $Q(\cdot) = \int_0^\infty x^{-2} W_x(\cdot)\, dx \in \mathcal{C}([0, \infty) : \mathbb{R}_+)$ a.s. Therefore, by the Portmanteau theorem,

$$\begin{aligned} &\lim_{k \to \infty} \limsup_{r \to \infty} \mathbb{P}\Big(\sup_{t \in [0,T]} \left| \int_0^\infty h_k(x) \widetilde{\mathcal{Z}}^r(t)(dx) - \int_0^\infty f(x) \widetilde{\mathcal{Z}}^r(t)(dx) \right| > k^{-1/2} \Big) \\ &\le \lim_{k \to \infty} \limsup_{r \to \infty} \mathbb{P}\Big(\sup_{t \in [0,T]} \int_0^\infty \widetilde{\mathcal{Z}}^r(t)(dx) \ge k^{1/2} \Big) \\ &\le \lim_{k \to \infty} \mathbb{P}\Big(\sup_{t \in [0,T]} Q(t) \ge k^{1/2} \Big) = 0. \end{aligned} \tag{5.120}$$

Finally, we have that almost surely,

$$\begin{aligned} &\lim_{k \to \infty} \sup_{t \in [0,T]} \left| \int_0^\infty h_k(x) \widetilde{\mathcal{Z}}(t)(dx) - \int_0^\infty f(x) \widetilde{\mathcal{Z}}(t)(dx) \right| \\ &\le \lim_{k \to \infty} k^{-1} \sup_{t \in [0,T]} Q(t) = 0. \end{aligned} \tag{5.121}$$

By (5.119), (5.120), (5.121) and Lemma 7, we conclude that for any compactly supported real-valued continuous function f,

$$\int_0^\infty f(x) \widetilde{\mathcal{Z}}^r(\cdot)(dx) \xrightarrow{d} \int_0^\infty f(x) \widetilde{\mathcal{Z}}(\cdot)(dx) \quad \text{as } r \to \infty. \tag{5.122}$$

From (5.118) and (5.122), (5.117) is verified and hence, by Theorem 2.1 in [26], $\{\widetilde{\mathcal{Z}}^r(\cdot), r \in \mathcal{R}\}$ is tight in $\mathcal{D}([0, T] : \mathcal{M}_F)$.

Suppose along a subsequence $\widetilde{\mathcal{Z}}^r(\cdot) \Rightarrow \widetilde{\mathcal{Z}}^*(\cdot)$ as $r \to \infty$. By the continuous mapping theorem, for any $k \in \mathbb{N}$ and compactly supported real-valued continuous functions $G_1, \ldots, G_k$,

$$\big(\langle G_1, \widetilde{\mathcal{Z}}^r(\cdot)\rangle, \ldots, \langle G_k, \widetilde{\mathcal{Z}}^r(\cdot)\rangle\big) \quad \Rightarrow \quad \big(\langle G_1, \widetilde{\mathcal{Z}}^*(\cdot)\rangle, \ldots, \langle G_k, \widetilde{\mathcal{Z}}^*(\cdot)\rangle\big) \quad \text{as } r \to \infty.$$

But also, from (5.122), the Cramér–Wold theorem and using the linearity of the integral,

$$\big(\langle G_1, \widetilde{\mathcal{Z}}^r(\cdot)\rangle, \ldots, \langle G_k, \widetilde{\mathcal{Z}}^r(\cdot)\rangle\big) \quad \Rightarrow \quad \big(\langle G_1, \widetilde{\mathcal{Z}}(\cdot)\rangle, \ldots, \langle G_k, \widetilde{\mathcal{Z}}(\cdot)\rangle\big) \quad \text{as } r \to \infty.$$

Thus, $(\langle G_1, \widetilde{\mathcal{Z}}^*(\cdot)\rangle, \ldots, \langle G_k, \widetilde{\mathcal{Z}}^*(\cdot)\rangle)$ and $(\langle G_1, \widetilde{\mathcal{Z}}(\cdot)\rangle, \ldots, \langle G_k, \widetilde{\mathcal{Z}}(\cdot)\rangle)$ are equal in distribution. This shows that $\widetilde{\mathcal{Z}}^*$ has the same law as $\widetilde{\mathcal{Z}}$ (Theorem 3.1 of [15]) and so $\widetilde{\mathcal{Z}}^r$ converges to $\widetilde{\mathcal{Z}}$ in $\mathcal{D}([0,T]:\mathcal{M}_F)$ as $r \to \infty$. $\square$

PROOF OF THEOREM 4. For $x \in (0,\infty]$ and $t \geq 0$, let $\tilde{X}_x(t) = \xi(\infty) - \xi(x) + X_x(t)$. By Theorem 1,

$$0 \leq W_\infty(t) - W_x(t) = \Gamma[X_\infty](t) - \Gamma[\tilde{X}_x](t) + \Gamma[\tilde{X}_x](t) - \Gamma[X_x](t) \quad \text{for all } t \geq 0.$$

By the Lipschitz property (4.1) of Γ and Assumption (3.8),

$$\lim_{x\to\infty} \sup_{t\geq 0} \lambda^{-1} x^p \big|\Gamma[\tilde{X}_x](t) - \Gamma[X_x](t)\big| \leq \lim_{x\to\infty} 2\lambda^{-1} x^p \big|\xi(\infty) - \xi(x)\big| = 0.$$

The previous two displays together with (4.4) imply that

$$\lim_{x\to\infty} \lambda^{-1} x^p \big(W_x(t) - W_\infty(t)\big) = -W'_\infty(t) \quad \text{for all } t \geq 0.$$

Fix $t \geq 0$ and let $\epsilon > 0$. There exists $x_0 > 0$ such that for all $x \geq x_0$,

$$\big|\lambda^{-1} x^p \big(W_x(t) - W_\infty(t)\big) + W'_\infty(t)\big| < \epsilon. \tag{5.123}$$

This implies that, for $a \geq x_0$,

$$\begin{aligned} &\left|\int_a^\infty x^{-2} W_x(t)\,dx - \frac{W_\infty(t)}{a} + \frac{\lambda}{(p+1)a^{p+1}} W'_\infty(t)\right| \\ &\quad \leq \int_a^\infty \lambda x^{-p-2} \big|\lambda^{-1} x^p \big(W_x(t) - W_\infty(t)\big) + W'_\infty(t)\big|\,dx \leq \frac{\lambda\epsilon}{(p+1)a^{p+1}}. \end{aligned} \tag{5.124}$$

By Theorem 3, for all $a \in (0,\infty)$,

$$\widetilde{\mathcal{Z}}(t)[a,\infty) = \int_a^\infty \frac{1}{x^2} W_x(t)\,dx - \frac{W_a(t)}{a}.$$

Thus, from (5.123) (with $x = a$) and (5.124), for any $a \geq x_0$,

$$\begin{aligned} &\left|\widetilde{\mathcal{Z}}(t)[a,\infty) - \frac{p\lambda}{(p+1)a^{p+1}} W'_\infty(t)\right| \\ &\quad = \left|\int_a^\infty \frac{1}{x^2} W_x(t)\,dx - \frac{W_a(t)}{a} - \frac{p\lambda}{(p+1)a^{p+1}} W'_\infty(t)\right| \\ &\quad = \left|\frac{W_\infty(t)}{a} - \frac{W_a(t)}{a} - \frac{\lambda}{a^{p+1}} W'_\infty(t) + \int_a^\infty \frac{1}{x^2} W_x(t)\,dx \right. \\ &\qquad \left. - \frac{W_\infty(t)}{a} + \frac{\lambda}{(p+1)a^{p+1}} W'_\infty(t)\right| \\ &\quad \leq \lambda a^{-p-1} \big|\lambda^{-1} a^p \big(W_a(t) - W_\infty(t)\big) + W'_\infty(t)\big| \\ &\qquad + \left|\int_a^\infty x^{-2} W_x(t)\,dx - \frac{W_\infty(t)}{a} + \frac{\lambda}{(p+1)a^{p+1}} W'_\infty(t)\right| \\ &\quad \leq \frac{\lambda\epsilon}{a^{p+1}} + \frac{\lambda\epsilon}{(p+1)a^{p+1}}. \end{aligned} \tag{5.125}$$

As $\epsilon > 0$ is arbitrary, the first two limits claimed in the theorem follow from (5.123) and (5.125). To prove the last limit, note that by the first two limit results of the theorem, for any t such that $W'_\infty(t) \neq 0$,

$$\frac{p\langle \chi \mathbf{1}_{[a,\infty)}, \widetilde{\mathcal{Z}}(t)\rangle}{(p+1)a\widetilde{\mathcal{Z}}(t)[a,\infty)} \to 1 \quad \text{as } a \to \infty. \tag{5.126}$$

Moreover, for each $a > 0$,

$$\frac{p\mathbb{E}(v \mid v > a)}{(p+1)a} = \frac{p}{(p+1)a}\left(\frac{a\overline{F}(a) + \int_a^\infty \overline{F}(x)\,dx}{\overline{F}(a)}\right) = \frac{p}{(p+1)}\left(1 + \frac{\int_a^\infty \overline{F}(x)\,dx}{a\overline{F}(a)}\right).$$

This together with (4.5) gives

$$\lim_{a\to\infty} \frac{p\mathbb{E}(v \mid v > a)}{(p+1)a} = \frac{p}{(p+1)}\left(1 + \frac{1}{p}\right) = 1. \tag{5.127}$$

The last limit claimed in the theorem follows from (5.126) and (5.127). □

5.6. *Proof of Theorem 5.* In this section, we prove Theorem 5, which concerns an asymptotic relationship between of the limiting processes $Q^{(p)}$ and $W^{(p)}$ as $p \to \infty$. As in Section 3.4, we consider $p \geq 2$ and index all limiting processes (resp. parameters and constants) that depend on p with the superscript (p) (resp. an argument of p). In addition, we assume that the asymptotic conditions stated in Section 3.4 hold.

PROOF OF THEOREM 5. Recall that, for all $p \geq 2$,

$$Q^{(p)}(t) = \int_0^\infty \frac{1}{x^2} W_x^{(p)}(t)\,dx, \quad t \geq 0,$$

where

$$W_a^{(p)}(t) := \Gamma\big[X_a^{(p)}\big](t), \quad t \geq 0, a > 0,$$

with Γ denoting the Skorohod map and

$$X_a^{(p)}(t) := \xi^{(p)}(a) + \sigma(p)B(t) + \left(\kappa - \frac{\lambda}{a^p}\right)t, \quad t \geq 0, a > 0.$$

Let $T, \gamma > 0$. Take any $\vartheta > 0$. Note that, for any $p \geq 2$ and $\epsilon \in (0,1)$,

$$\sup_{t\in[0,T]} \int_{1-\epsilon}^1 x^{-2} W_x^{(p)}(t)\,dx \leq \frac{\epsilon}{1-\epsilon} \sup_{t\in[0,T]} W_\infty^{(p)}(t). \tag{5.128}$$

Using the Lipschitz property (4.1) of the Skorohod map, that $\sigma(p) = \sqrt{\lambda \operatorname{Var}(v^{(p)}) + \lambda\sigma_A^2}$ and Assumption (3.10), for all $p \geq 2$,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,T]} W_\infty^{(p)}(t)\Big] \\
&\quad \leq 2\mathbb{E}\Big[\sup_{t\in[0,T]} \big(\xi^{(p)}(\infty) + \sigma(p)|B(t)| + \kappa t\big)\Big] \\
&\quad \leq 2\sup_{p\geq 2} \mathbb{E}\big[\xi^{(p)}(\infty)\big] + 2\sqrt{\sup_{p\geq 2}\big(\lambda \operatorname{Var}\big(v^{(p)}\big) + \lambda\sigma_A^2\big)}\,\mathbb{E}\Big[\sup_{t\in[0,T]} |B(t)|\Big] + 2\kappa T \\
&\quad := \mathcal{B} < \infty,
\end{aligned} \tag{5.129}$$

where the bound $\mathcal{B}$ does not depend on p. Hence, by (5.128), (5.129) and Markov's inequality, we can choose $\epsilon \in (0,1)$ such that

$$\mathbb{P}\left(\sup_{t\in[0,T]} \int_{1-\epsilon}^1 x^{-2} W_x^{(p)}(t)\,dx > \gamma/3\right) \leq \vartheta \quad \text{for all } p \geq 2. \tag{5.130}$$

By (3.11), we obtain $p_0' \geq 2$ such that

$$\frac{p - 1 - \eta^*(p)}{\log p} > \frac{4}{\log((1-\epsilon)^{-1})} \quad \text{for all } p \geq p_0'. \tag{5.131}$$

For each $p \geq 2$, let $m_0(p)$, $a_0(p)$ be defined as in Lemma 20. Since $m_0(p) > 1$ for all $p \geq 2$, $a_0(p) \in (0,1)$ for all $p \geq 2$. Due to (3.10), $0 < \inf_{p\geq 2} \sigma(p) \leq \sup_{p\geq 2} \sigma(p) < \infty$. Thus, $\lim_{p\to\infty} a_0(p) = \lim_{p\to\infty} m_0(p)^{-1/2p} = 1$. Take $p_0 \geq p_0'$ such that for all $p \geq p_0$, $a_0(p) > 1-\epsilon$. For $p \geq 2$, write

$$H'(p) := H\big(p, (p-1+\eta^*(p))/2\big) \quad \text{and}$$
$$C'\big(p,\lambda,\sigma(p)\big) := \tilde{C}\big((p-1+\eta^*(p))/2, \eta^*(p), \lambda, \sigma(p)\big),$$

where the functions H and $\tilde{C}$ were defined in Lemma 21. Then, by Lemma 21, taking $\delta = 1-\epsilon$ and $\eta = (p-1+\eta^*(p))/2$, we obtain for any $p \geq p_0$,

$$\begin{aligned}(5.132)\quad &\mathbb{P}\Big(\sup_{t\in[0,T]} \int_0^{1-\epsilon} x^{-2} W_x^{(p)}(t)\,dx > H'(p)(1-\epsilon)^{(p-1-\eta^*(p))/2}\big(1+\log\big((1-\epsilon)^{-1}\big)\big)\Big) \\ &\leq C'\big(p,\lambda,\sigma(p)\big)(1-\epsilon)^{(p-1-\eta^*(p))/2} + 3C\big(\lambda,\sigma(p)\big)(1-\epsilon)^{2p}.\end{aligned}$$

Using the explicit forms of $C'(p,\lambda,\sigma(p))$ (defined in Lemma 21) and $C(\lambda,\sigma(p))$ (defined in Lemma 20), Assumption (3.10), and (5.131), and recalling $\inf_{p\geq 2}\sigma(p) > 0$, note that

$$\begin{aligned}C'(\lambda) &:= \sup_{p\geq 2} C'\big(p,\lambda,\sigma(p)\big) \\ &= \sup_{p\geq 2}\Big(C_0(p) + \frac{2(\sigma(p))^2}{\lambda}\Big(\sup_{x\in\mathbb{R}_+} xe^{-x}\Big)\Big)\Big(1+\sum_{k=1}^{\infty} 2^{-(k-1)(p-1-\eta^*(p))/2)}\Big) < \infty,\end{aligned}$$

and

$$C(\lambda) := \sup_{p\geq 2} C\big(\lambda,\sigma(p)\big) = \sup_{p\geq 2}\Big(2e^{\lambda/(\sigma(p))^2} + \frac{16(\sigma(p))^2}{\lambda}\Big) < \infty.$$

Using these observations in (5.132), we obtain for any $p \geq p_0$,

$$\begin{aligned}(5.133)\quad &\mathbb{P}\Big(\sup_{t\in[0,T]} \int_0^{1-\epsilon} x^{-2} W_x^{(p)}(t)\,dx > H'(p)(1-\epsilon)^{(p-1-\eta^*(p))/2}\big(1+\log\big((1-\epsilon)^{-1}\big)\big)\Big) \\ &\leq C'(\lambda)(1-\epsilon)^{(p-1-\eta^*(p))/2} + 3C(\lambda)(1-\epsilon)^{2p}.\end{aligned}$$

Using (5.131), we have that $\log p + \frac{p-1-\eta^*(p)}{2}\log(1-\epsilon) \to -\infty$ as $p\to\infty$. Exponentiating, we obtain

$$p(1-\epsilon)^{(p-1-\eta^*(p))/2}\big(1+\log\big((1-\epsilon)^{-1}\big)\big) \to 0 \quad \text{as } p\to\infty.$$

From this and the explicit form of $H(p,\eta)$ given in Lemma 21, we conclude that $H'(p)(1-\epsilon)^{(p-1-\eta^*(p))/2}(1+\log((1-\epsilon)^{-1})) \to 0$ as $p\to\infty$. Moreover, the right-hand side of (5.133) also goes to zero as $p\to\infty$. Thus,

$$(5.134)\qquad \sup_{t\in[0,T]} \int_0^{1-\epsilon} x^{-2} W_x^{(p)}(t)\,dx \xrightarrow{P} 0 \quad \text{as } p\to\infty.$$

Moreover, by the Lipschitz property (4.1) and Assumption (3.12), for each $x \in (1,\infty)$,

$$\begin{aligned}\mathbb{E}\Big(\sup_{t\in[0,T]} \big|W_x^{(p)}(t) - W_\infty^{(p)}(t)\big|\Big) &\leq 2\mathbb{E}\Big(\sup_{t\in[0,T]} \big|X_x^{(p)}(t) - X_\infty^{(p)}(t)\big|\Big) \\ &\leq 2\mathbb{E}\big(\xi^{(p)}(\infty) - \xi^{(p)}(x)\big) + \frac{2\lambda T}{x^p} \to 0 \quad \text{as } p\to\infty,\end{aligned}$$

where we recall $X_\infty^{(p)}(t) = \xi^{(p)}(\infty) + \sigma^{(p)} B(t) + \kappa t$, $t \geq 0$. By the monotonicity property noted in (4.3) and using (5.129), for all $p \geq 2$,

$$\mathbb{E}\Big(\sup_{t\in[0,T]} \big| W_x^{(p)}(t) - W_\infty^{(p)}(t) \big| \Big) \leq \mathbb{E}\Big(\sup_{t\in[0,T]} W_\infty^{(p)}(t) \Big) \leq \mathcal{B}.$$

Thus, by the dominated convergence theorem,

$$\int_1^\infty x^{-2} \mathbb{E}\Big(\sup_{t\in[0,T]} \big| W_x^{(p)}(t) - W_\infty^{(p)}(t) \big| \Big)\, dx \to 0 \quad \text{as } p \to \infty,$$

which implies

$$\sup_{t\in[0,T]} \left| \int_1^\infty x^{-2} W_x^{(p)}(t)\, dx - W_\infty^{(p)}(t) \right| \xrightarrow{P} 0 \quad \text{as } p \to \infty. \tag{5.135}$$

From (5.130), (5.134) and (5.135),

$$\begin{aligned}
&\limsup_{p\to\infty} \mathbb{P}\Big(\sup_{t\in[0,T]} \left| \int_0^\infty x^{-2} W_x^{(p)}(t)\, dx - W_\infty^{(p)}(t) \right| > \gamma \Big) \\
&\quad \leq \limsup_{p\to\infty} \mathbb{P}\Big(\sup_{t\in[0,T]} \int_0^{1-\epsilon} x^{-2} W_x^{(p)}(t)\, dx > \gamma/3 \Big) \\
&\qquad + \limsup_{p\to\infty} \mathbb{P}\Big(\sup_{t\in[0,T]} \int_{1-\epsilon}^{1} x^{-2} W_x^{(p)}(t)\, dx > \gamma/3 \Big) \\
&\qquad + \limsup_{p\to\infty} \mathbb{P}\Big(\sup_{t\in[0,T]} \left| \int_1^\infty x^{-2} W_x^{(p)}(t)\, dx - W_\infty^{(p)}(t) \right| > \gamma/3 \Big) \leq \vartheta.
\end{aligned}$$

As T, γ, $\vartheta > 0$ are arbitrary, the theorem is proved. □

APPENDIX: VERIFYING ASSUMPTIONS (2.14)–(2.19) FOR SOME INITIAL CONDITIONS

A.1. Checking Assumptions (2.14)–(2.19) at fixed time $t > 0$ for a sequence of systems with $\mathbf{q}^r = 0$ for all $r \in \mathcal{R}$. Here we sketch how to verify that if each system in the sequence starts with zero jobs then at any time $t > 0$, Assumptions (2.14)–(2.19) are satisfied with $(W_\cdot^r(0), W_\infty^r(0))$ replaced by $(W_\cdot^r(t), W_\infty^r(t))$ for each $r \in \mathcal{R}$ and $\{\check{v}_l^r\}_{1\leq l\leq \mathbf{q}^r}$ replaced with $\{v_i(r^2 t) : 1 \leq i \leq E^r(r^2 t), v_i(r^2 t) > 0\} \cup \{\check{v}_l^r(r^2 t), 1 \leq l \leq \mathbf{q}^r, \check{v}_l^r(r^2 t) > 0\}$. Fix $t > 0$ and note that since $\mathbf{q}^r = 0$ for all $r \in \mathcal{R}$, Assumptions (2.14)–(2.19) hold at time zero. Thus, Theorem 1, along with tightness arguments similar to those in the proof of Theorem 14 and the estimates in (5.12) and (5.115), can be used to show that for any fixed $t > 0$, (2.14) holds with $(W_\cdot^r(0), W_\infty^r(0))$ replaced by $(W_\cdot^r(t), W_\infty^r(t))$ for each $r \in \mathcal{R}$ and $(w^*(\cdot), w^*(\infty))$ replaced by $(W_\cdot(t), W_\infty(t))$, where W is defined in Theorem 1 with $\xi(a) = 0$ for all $a \in [0, \infty]$. The uniform integrability assumption (2.15) can be shown to hold for $\{W_\infty^r(t), r \in \mathcal{R}\}$ by first noting that for each $r \in \mathcal{R}$, $W_\infty^r(t) = \Gamma[X_\infty^r](t)$, where Γ is the Skorohod map defined in (3.1) and $X_\infty^r(\cdot)$ is defined in (5.7) (taking $X_\infty^r(0) = 0$). By (2.2), the finiteness of $\mathrm{Var}(v)$ and by applications of Doob's L^2-maximal inequality and Azuma–Hoeffding inequality, we can obtain for any $t > 0$ that $\mathbb{E}[(\sup_{0\leq s\leq t} X_\infty^r(s))^\beta] < \infty$ for any $\beta \in (1, 2)$. From this observation and the Lipschitz property of the Skorohod map stated in (4.1), we can deduce $\{W_\infty^r(t) : r \in \mathcal{R}\}$ is L^β-bounded for any $\beta \in (1, 2)$ and thus (2.15) holds. Assumption (2.16) follows along the same lines as the proof of Lemmas 16, 18 and 19 (see Remark 6). Assumption (2.17) follows by recalling that $(w^*(\cdot), w^*(\infty)) = (W_\cdot(t), W_\infty(t))$ and using the explicit form of W defined in Theorem 1 and the Lipschitz property (4.1) of the Skorohod map. Finally, (2.19) follows from Proposition 10 and Lemma 16.

A.2. Checking Assumptions (2.14)–(2.19) for initial conditions (I) given in Section 2.5. We first show that (2.14) holds. For $0 \le x \le \infty$, define $\hat{W}^r(x) := \frac{c^r \mathbf{q}^r}{r} \mathbb{E}(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[\check{v}_1^r \le x c^r]})$. For any $A \in (0,\infty)$,

$$
\begin{aligned}
&\sup_{x\in[0,A]} \left| \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[\check{v}_1^r \le x c^r]} \Big) - \mathbb{E}\big(\check{v}^* \mathbf{1}_{[\check{v}^* \le x]} \big) \right| \\
&\quad \le \sup_{x\in[0,A]} \left| \int_0^x \mathbb{P}\big(z c^r < \check{v}_1^r \le x c^r \big)\, dz - \int_0^x \mathbb{P}\big(z < \check{v}^* \le x \big)\, dz \right| \\
&\quad \le \sup_{x\in[0,A]} \left(x \big| \mathbb{P}\big(\check{v}_1^r \le x c^r \big) - \mathbb{P}\big(\check{v}^* \le x \big) \big| + \int_0^x \big| \mathbb{P}\big(\check{v}_1^r \le z c^r \big) - \mathbb{P}\big(\check{v}^* \le z \big) \big|\, dz \right) \\
&\quad \le 2A \sup_{x\in[0,A]} \big| \mathbb{P}\big(\check{v}_1^r \le x c^r \big) - \mathbb{P}\big(\check{v}^* \le x \big) \big| \to 0 \quad \text{as } r \to \infty
\end{aligned} \tag{A.1}
$$

by Pólya's theorem [11], Exercise 3.2.9, page 107, as $\check{v}^*$ has a continuous distribution. As the map $x \mapsto \mathbb{E}(\check{v}^* \mathbf{1}_{[\check{v}^* \le x]})$ is continuous by (iii), it follows from (A.1) and (iii) that for any $\epsilon > 0$, there exists $A \in (0,\infty)$, $\delta > 0$ and $r_0 \in \mathcal{R}$ such that for all $r \ge r_0$,

$$
\sup_{0\le x\le y\le A,\, y-x\le\delta} \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[x c^r < \check{v}_1^r \le y c^r]} \Big) < \epsilon \quad \text{and} \quad \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[\check{v}_1^r > A c^r]} \Big) < \epsilon. \tag{A.2}
$$

Also, by (i), for each $r \in \mathcal{R}$ and $0 \le x \le \infty$,

$$
\begin{aligned}
\mathbb{E}\big(W_x^r(0) - \hat{W}^r(x) \big)^2 &= \frac{(c^r)^2 \mathbb{E}(\mathbf{q}^r)}{r^2} \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[\check{v}_i^r \le x c^r]} - \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[\check{v}_1^r \le x c^r]} \Big) \Big)^2 \\
&\le \frac{(c^r)^2 \mathbb{E}(\mathbf{q}^r)}{r^2} \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[\check{v}_1^r \le x c^r]} \Big)^2 \le \frac{(c^r)^2 \mathbb{E}(\mathbf{q}^r)}{r^2} \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \Big)^2.
\end{aligned}
$$

This together with the conditions (ii) and (iii) and (4.11) imply that

$$
\sup_{0\le x\le\infty} \mathbb{E}\big(W_x^r(0) - \hat{W}^r(x) \big)^2 \to 0 \quad \text{as } r \to \infty. \tag{A.3}
$$

Fix any $\epsilon > 0$. Note that by (ii), $\sup_{r\in\mathcal{R}} \mathbb{E}(\frac{c^r \mathbf{q}^r}{r}) < \infty$. By (A.2) and (A.3), one can obtain a partition $0 = x_0 < x_1 < \cdots < x_k < x_{k+1} = \infty$ of $[0,\infty]$ and $r_1 \in \mathcal{R}$ such that the following hold for all $r \ge r_1$:

$$
\mathbb{E}\big| W_{x_j}^r(0) - \hat{W}^r(x_j) \big| < \frac{\epsilon}{2(k+2)} \quad \text{for all } 0 \le j \le k+1, \tag{A.4}
$$

and

$$
\mathbb{E}\Big(\frac{\check{v}_1^r}{c^r} \mathbf{1}_{[x_j c^r < \check{v}_1^r \le x_{j+1} c^r]} \Big) < \frac{\epsilon}{2 \sup_{r\in\mathcal{R}} \mathbb{E}(\frac{c^r \mathbf{q}^r}{r})} \quad \text{for all } 0 \le j \le k. \tag{A.5}
$$

By the monotonicity of the maps $x \mapsto W_x^r(0)$ and $x \mapsto \hat{W}^r(x)$, one obtains the bound

$$
\begin{aligned}
\sup_{x\in[0,\infty]} \big| W_x^r(0) - \hat{W}^r(x) \big| &\le \sup_{0\le j\le k+1} \big| W_{x_j}^r(0) - \hat{W}^r(x_j) \big| + \sup_{0\le j\le k} \big| \hat{W}^r(x_{j+1}) - \hat{W}^r(x_j) \big| \\
&\le \sum_{j=0}^{k+1} \big| W_{x_j}^r(0) - \hat{W}^r(x_j) \big| + \sup_{0\le j\le k} \big| \hat{W}^r(x_{j+1}) - \hat{W}^r(x_j) \big|.
\end{aligned}
$$

Hence, using (A.4) and (A.5) and (i), we obtain for $r \geq r_1$

$$\mathbb{E}\Big(\sup_{x\in[0,\infty]} \big|W_x^r(0) - \hat{W}^r(x)\big|\Big)$$
$$\leq \sum_{j=0}^{k+1} \mathbb{E}\big|W_{x_j}^r(0) - \hat{W}^r(x_j)\big| + \mathbb{E}\Big(\frac{c^r\mathbf{q}^r}{r}\Big) \sup_{0\leq j\leq k} \mathbb{E}\Big(\frac{\check{v}_1^r}{c^r}\mathbf{1}_{[x_j c^r < \check{v}_1^r \leq x_{j+1}c^r]}\Big) < \epsilon.$$

As $\epsilon > 0$ is arbitrary, we conclude

$$\lim_{r\to\infty} \mathbb{E}\Big(\sup_{x\in[0,\infty]} \big|W_x^r(0) - \hat{W}^r(x)\big|\Big) = 0. \tag{A.6}$$

Note that for any $\epsilon > 0$, by the second assertion of (A.2) and that fact that $\mathbb{E}(\check{v}_1^r/c^r) \to \mathbb{E}(\check{v}^*) < \infty$ as $r \to \infty$, which follows from (iii), we can obtain $A > 0$, $r_2 \in \mathcal{R}$ such that for all $r \geq r_2$

$$\sup_{x\in[A,\infty]} \Big|\mathbb{E}\Big(\frac{\check{v}_1^r}{c^r}\mathbf{1}_{[\check{v}_1^r \leq xc^r]}\Big) - \mathbb{E}\big(\check{v}^*\mathbf{1}_{[\check{v}^*\leq x]}\big)\Big|$$
$$\leq \big|\mathbb{E}(\check{v}_1^r/c^r) - \mathbb{E}(\check{v}^*)\big| + \sup_{x\in[A,\infty]} \Big|\mathbb{E}\Big(\frac{\check{v}_1^r}{c^r}\mathbf{1}_{[\check{v}_1^r > xc^r]}\Big) - \mathbb{E}\big(\check{v}^*\mathbf{1}_{[\check{v}^*>x]}\big)\Big| < \epsilon.$$

Combining this with (A.1), we obtain

$$\lim_{r\to\infty} \sup_{x\in[0,\infty]} \Big|\mathbb{E}\Big(\frac{\check{v}_1^r}{c^r}\mathbf{1}_{[\check{v}_1^r \leq xc^r]}\Big) - \mathbb{E}\big(\check{v}^*\mathbf{1}_{[\check{v}^*\leq x]}\big)\Big| = 0. \tag{A.7}$$

Defining $\tilde{W}^r(x) := \frac{c^r\mathbf{q}^r}{r}\mathbb{E}(\check{v}^*\mathbf{1}_{[\check{v}^*\leq x]})$ for $x \in [0,\infty]$, we conclude from (A.6), (A.7) and the fact $\sup_{r\in\mathcal{R}} \mathbb{E}(c^r\mathbf{q}^r/r) < \infty$ that

$$\lim_{r\to\infty} \mathbb{E}\Big(\sup_{x\in[0,\infty]} \big|W_x^r(0) - \tilde{W}^r(x)\big|\Big) = 0. \tag{A.8}$$

Finally as $c^r\mathbf{q}^r/r \xrightarrow{L^1} \mathbf{q}^*$ as $r \to \infty$ by (ii), (A.8) implies that Assumption (2.14) holds with the given choice of $w^*(\cdot)$. In fact we have shown that

$$\lim_{r\to\infty} \mathbb{E}\Big(\sup_{x\in[0,\infty]} \big|W_x^r(0) - w^*(x)\big|\Big) = 0. \tag{A.9}$$

Assumption (2.15) follows from the observation that $W_\infty^r(0) \xrightarrow{L^1} w^*(\infty)$ which holds by (A.9). Assumption (2.16) is a direct consequence of (i), (ii) and (iv). Assumption (2.17) follows from (i), (ii) and the observation that $\mathbb{E}(\check{v}^*) < \infty$ which follows from (iii) and Fatou's lemma. Assumption (2.19) follows from (ii) and (iii).

Acknowledgements. The authors acknowledge the Open Problem Session held during the Seminar on Stochastic Processes, 2019, where the addressed problem was presented as open by AP. The authors also thank the anonymous referees for very helpful advice.

Funding. Research of SB was supported in part by a Junior Faculty Development Award. AB acknowledges support from the National Science Foundation (DMS-1814894 and DMS-1853968). He is also grateful for the support from Nelder Fellowship from Imperial College, London, where part of this research was completed. Research of AP was supported in part by National Science Foundation grants DMS-1712974 and DMS-2054505 and the Charles Lee Powell Foundation.

REFERENCES

[1] ATAR, R., BISWAS, A., KASPI, H. and RAMANAN, K. (2018). A Skorokhod map on measure-valued paths with applications to priority queues. *Ann. Appl. Probab.* **28** 418–481. MR3770881 https://doi.org/10.1214/17-AAP1309

[2] BANSAL, N. and HARCHOL-BALTER, M. (2001). Analysis of SRPT scheduling: Investigating unfairness. In *ACM SIGMETRICS* 2001 *Conference on Measurement and Modeling of Computer Systems* 279–290.

[3] BENDER, M. A., CHAKRABARTI, S. and MUTHUKRISHNAN, S. (1998). Flow and stretch metrics for scheduling continuous job streams. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms* (*San Francisco*, *CA*, 1998) 270–279. ACM, New York. MR1642937

[4] BILLINGSLEY, P. (2013). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[5] BRAMSON, M. and DAI, J. G. (2001). Heavy traffic limits for some queueing networks. *Ann. Appl. Probab.* **11** 49–90. MR1825460 https://doi.org/10.1214/aoap/998926987

[6] CHEN, Y. and DONG, J. (2020). Scheduling with service-time information: The power of two priority classes. Preprint.

[7] CLAUSET, A., SHALIZI, C. R. and NEWMAN, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Rev.* **51** 661–703. MR2563829 https://doi.org/10.1137/070710111

[8] DIEKER, A. B. and GAO, X. (2014). Sensitivity analysis for diffusion processes constrained to an orthant. *Ann. Appl. Probab.* **24** 1918–1945. MR3226168 https://doi.org/10.1214/13-AAP967

[9] DOWN, D. G., GROMOLL, H. C. and PUHA, A. L. (2009). Fluid limits for shortest remaining processing time queues. *Math. Oper. Res.* **34** 880–911. MR2573501 https://doi.org/10.1287/moor.1090.0409

[10] DOWN, D., GROMOLL, H. C. and PUHA, A. (2009). State-dependent response times via fluid limits for shortest remaining processing time queues. In *San Diego ACM-Sigmetrics Performance Evaluation Review* **27** 75–76.

[11] DURRETT, R. (2019). *Probability*: *Theory and Examples*. *Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge Univ. Press, Cambridge. MR3930614 https://doi.org/10.1017/9781108591034

[12] ETHIER, S. N. and KURTZ, T. G. (2009). *Markov Processes*: *Characterization and Convergence*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR0838085 https://doi.org/10.1002/9780470316658

[13] GROMOLL, H. C., KRUK, Ł. and PUHA, A. L. (2011). Diffusion limits for shortest remaining processing time queues. *Stoch. Syst.* **1** 1–16. MR2948916 https://doi.org/10.1214/10-SSY016

[14] IGLEHART, D. L. and WHITT, W. (1970). Multiple channel queues in heavy traffic. I. *Adv. in Appl. Probab.* **2** 150–177. MR0266331 https://doi.org/10.1017/s0001867800037241

[15] KALLENBERG, O. (1974). *Lectures on Random Measures*. Consolidated University of North Carolina, Institute of Statistics.

[16] KARATZAS, I. and SHREVE, S. E. (1998). *Brownian Motion and Stochastic Calculus*. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940 https://doi.org/10.1007/978-1-4612-0949-2

[17] KRUK, Ł. (2007). Diffusion approximation for $G/G/1$ EDF queue with unbounded lead times. *Ann. Univ. Mariae Curie-Skłodowska Sect. A* **61** 51–90. MR2368922 https://doi.org/10.1088/1742-6596/61/1/011

[18] KRUK, L. (2019). Diffusion limits for SRPT and LRPT queues via EDF approximations. In *Queueing Theory and Network Applications*, 14*th International Conference*, *QTNA* 2019, *Ghent*, *Belgium*.

[19] KRUK, Ł. and SOKOŁOWSKA, E. (2016). Fluid limits for multiple-input shortest remaining processing time queues. *Math. Oper. Res.* **41** 1055–1092. MR3520764 https://doi.org/10.1287/moor.2015.0768

[20] LIN, M., WIERMAN, A. and ZWART, B. (2011). The heavy-traffic growth rate of shortest remaining processing time. *Perform. Eval.* **68** 955–966.

[21] LOBOZ, C. (2012). Cloud resource usage—Heavy tailed distributions invalidating traditional capacity planning models. *J. Grid Comput.* **10** 85–108.

[22] MANDELBAUM, A. and RAMANAN, K. (2010). Directional derivatives of oblique reflection maps. *Math. Oper. Res.* **35** 527–558. MR2724063 https://doi.org/10.1287/moor.1100.0453

[23] MIKOSCH, T. (1999). *Regular Variation*, *Subexponentiality and Their Applications in Probability Theory*. Eindhoven Univ. Technology.

[24] PERERA, R. (1993). The variance of delay time in queueing system M/G/1 with optimal strategy SRPT. *Arch. Elektron. Übertrag.tech.* **47** 110–114.

[25] PUHA, A. L. (2015). Diffusion limits for shortest remaining processing time queues under nonstandard spatial scaling. *Ann. Appl. Probab.* **25** 3381–3404. MR3404639 https://doi.org/10.1214/14-AAP1076

[26] ROELLY-COPPOLETTA, S. (1986). A criterion of convergence of measure-valued processes: Application to measure branching processes. *Stochastics* **17** 43–65. MR0878553 https://doi.org/10.1080/17442508608833382

[27] SCHASSBERGER, R. (1990). The steady-state appearance of the $M/G/1$ queue under the discipline of shortest remaining processing time. *Adv. in Appl. Probab.* **22** 456–479. MR1053240 https://doi.org/10.2307/1427545

[28] SCHRAGE, L. (1968). A proof of the optimality of the shortest remaining processing time discipline. *Oper. Res.* **16** 687–690.

[29] SCHREIBER, F. (1993). Properties and applications of the optimal queueing strategy SRPT: A survey. *Arch. Elektron. Übertrag.tech.* **47** 372–378.

[30] SILBERSCHATZ, A. and GALVIN, P. (1998). *Operating System Concepts*, 5th ed. Wiley, New York.

[31] SMITH, D. R. (1978). A new proof of the optimality of the shortest remaining processing time discipline. *Oper. Res.* **26** 197–199. MR0471112 https://doi.org/10.1287/opre.26.1.197

[32] STALLINGS, W. (1995). *Operating Systems*, 2nd ed. Prentice Hall, New York.

[33] TANENBAUM, A. S. (1992). *Modern Operating Systems*. Prentice Hall, New York.

[34] WHITT, W. (1971). Weak convergence theorems for priority queues: Preemptive-resume discipline. *J. Appl. Probab.* **8** 74–94. MR0307389 https://doi.org/10.2307/3211839

[35] WHITT, W. (2002). *Stochastic-Process Limits*: *An Introduction to Stochastic-Process Limits and Their Application to Queues*. *Springer Series in Operations Research*. Springer, New York. MR1876437

[36] WIERMAN, A. and HARCHOL-BALTER, M. (2003). Classifying scheduling policies with respect to unfairness in an M/GI/1. In *ACM SIGMETRICS* 2003 *Conference on Measurement and Modeling of Computer Systems* 238–249.