Submitted to the Annals of Applied Probability

ON THE RATE OF CONVERGENCE TO STATIONARITY OF THE M/M/N QUEUE IN THE HALFIN-WHITT REGIME

By David Gamarnik and David A. Goldberg

MIT and Georgia Institute of Technology

We prove several results about the rate of convergence to stationarity, i.e. the spectral gap, for the $M/M/n$ queue in the Halfin-Whitt regime. We identify the limiting rate of convergence to steady-state, and discover an asymptotic phase transition that occurs w.r.t. this rate. In particular, we demonstrate the existence of a constant $B^* \approx 1.85772$ s.t. when a certain excess parameter $B \in (0, B^*]$, the error in the steady-state approximation converges exponentially fast to zero at rate $\frac{B^2}{4}$. For $B > B^*$, the error in the steady-state approximation converges exponentially fast to zero at a different rate, which is the solution to an explicit equation given in terms of special functions. This result may be interpreted as an asymptotic version of a phase transition proven to occur for any fixed n by van Doorn in [33].

We also prove explicit bounds on the distance to stationarity for the $M/M/n$ queue in the Halfin-Whitt regime, when $B < B^*$. Our bounds scale independently of n in the Halfin-Whitt regime, and do not follow from the weak-convergence theory.

1. Introduction. Parallel server queueing systems can operate in a variety of regimes that balance between efficiency and quality of offered service. This is captured by the so-called Halfin-Whitt (HW) heavy-traffic regime, which can be described as critical w.r.t. the probability that an arriving job has to wait for service. Namely, in this regime the stationary probability of wait is bounded away from both zero and unity, as the number of servers grows. Although studied originally by Pollaczek [31] (see also [22]), Erlang [13], and Jagerman [21], the regime was formally introduced by Halfin and Whitt [18], who studied the $GI/M/n$ system for large n when the traffic intensity scales like $1 - Bn^{-\frac{1}{2}}$ for some strictly positive excess parameter B. They proved that, under minor technical assumptions on the inter-arrival distribution, this sequence of $GI/M/n$ queueing models has the following properties:

(i) the steady-state probability that an arriving job has to wait for service has a non-trivial limit;
(ii) the sequence of queueing processes, normalized by $n^{\frac{1}{2}}$, converges weakly to a non-trivial positive recurrent diffusion, a.k.a. the HW diffusion;
(iii) the sequence of steady-state queue length distributions, normalized by $n^{\frac{1}{2}}$, is tight and converges distributionally to the mixture of a point mass at zero and an exponential distribution.

Since the steady-state behavior of the $M/M/n$ queue in the HW regime is quite simple [18], while the transient dynamics are more complicated [18], it is common to use the steady-state approximation to the transient distribution [16]. Thus it is important to understand the quality of the steady-state approximation. The only work along these lines seems to be the recent papers [28],[27], in which the authors study the Laplace transform of the HW and related diffusions, and

AMS 2000 subject classifications: Primary 60K25

Keywords and phrases: Many-server queues, rate of convergence, spectral gap, weak convergence, orthogonal polynomials, parabolic cylinder functions

prove several results analogous to our own for these diffusions. The key difference is that in this paper we study the pre-limit diffusion-scaled $M/M/n$ queue, not the limiting diffusion. We note that the relevant transform functions were also studied in [1], although in a different context. Also, similar questions were studied for the associated sequence of fluid-scaled queues in [23].

The question of how quickly the positive recurrent $M/M/n$ queue approaches stationarity has a rich history in the queueing literature. In [30], Morse derives an explicit solution for the transient $M/M/1$ queue, and discusses implications for the exponential rate of convergence to stationarity. Similar analyses are carried out in [7] and [32]. Around the same time, both Ledermann and Reuter [29], and Karlin and McGregor (KM) [24], worked out powerful and elegant theories that could be used to give the transient distributions for large classes of birth-death processes (BDP), including the $M/M/n$ queue. The transient probabilities are expressed as integrals against a spectral measure ϕ, which is intimately related to the eigenvalues of the generator of the BDP. KM devote an entire paper [25] to the application of their theory to the $M/M/n$ queue, in which they comment explicitly on the relationship between the rate of convergence to stationarity and the support of ϕ. This relationship was later formalized in a series of papers by other authors [4],[37]. Let $P(t)$ denote the matrix of transient probabilities for the $M/M/n$ queue; i.e. $P_{i,j}(t)$ is the probability that there are j jobs in system at time t, if there are i jobs in system at time 0. Let A denote the generator matrix associated with the $M/M/n$ queue, i.e. $\frac{d}{dt}P(t) = A \cdot P(t)$ [14]. Recall that the spectral gap γ of a BDP is the absolute value of the supremum of the set of strictly negative real eigenvalues of A over an appropriate domain, and we refer the reader to [4] for details. Then it follows from the results of [4] that

THEOREM 1. *For any positive recurrent $M/M/n$ or $M/M/\infty$ queue, $\gamma \in (0, \infty)$. For all i and j, $\lim_{t\to\infty} -t^{-1} \log |P_{i,j}(t) - P_j(\infty)|$ exists, and is at least γ. For at least one pair of i and j, $\lim_{t\to\infty} -t^{-1} \log |P_{i,j}(t) - P_j(\infty)| = \gamma$. Furthermore, $\gamma = \inf\{x : x > 0, \phi(x+\epsilon) - \phi(x-\epsilon) > 0$ for all $\epsilon > 0\}$.*

We note that γ is closely related to the singularities of the Laplace transform of ϕ, and refer the reader to [25] for details. It is well-known that for the positive recurrent $M/M/1$ and $M/M/\infty$ queues, γ can be computed explicitly. In particular, it is proven in [25] that

THEOREM 2. *For the positive recurrent $M/M/1$ queue with arrival rate λ and service rate μ, $\gamma = (\lambda^{\frac{1}{2}} - \mu^{\frac{1}{2}})^2$, and the spectral measure ϕ consists of a jump at zero, and an absolutely continuous measure on $[(\lambda^{\frac{1}{2}} - \mu^{\frac{1}{2}})^2, (\lambda^{\frac{1}{2}} + \mu^{\frac{1}{2}})^2]$. For the $M/M/\infty$ queue with arrival rate λ and service rate μ, $\gamma = \mu$, and the spectral measure ϕ consists of a countably infinite number of jumps, with exactly one jump at every non-negative integer multiple of μ.*

Unfortunately, for the general positive recurrent $M/M/n$ queue, the known characterizations for γ involve computing the roots of high-degree polynomials, which may be computationally difficult. This arises from the fact that for the positive recurrent $M/M/n$ queue with arrival rate λ and service rate μ, the spectral measure ϕ consists of three parts, as described in [25]. The first part is a jump at zero, which corresponds to the steady-state distribution. The second component is an absolutely continuous measure on the interval $\left[\left(\lambda^{\frac{1}{2}} - (n\mu)^{\frac{1}{2}}\right)^2, \left(\lambda^{\frac{1}{2}} + (n\mu)^{\frac{1}{2}}\right)^2\right]$. The third component consists of a set of at most n (but possibly zero) jumps, which all exist on $\left(0, \left(\lambda^{\frac{1}{2}} - (n\mu)^{\frac{1}{2}}\right)^2\right)$. The complexity of determining γ arises from the difficulty of locating these jumps [35]. In [25], this set of jumps is expressed in terms of the zeros of a certain polynomial equation.

Significant progress towards understanding these jumps was made in a series of papers by van Doorn [33],[34],[35],[36]. Van Doorn used the KM representation and the theory of orthogonal polynomials to give several alternate characterizations and bounds for the spectral gap of a BDP, and applied these to the $M/M/n$ queue. He also showed in [33] that for each fixed n there is a transition in the nature of the spectral measure of the $M/M/n$ queue as one varies the traffic intensity, proving that

THEOREM 3. *For all $n \geq 1$, there exists $\rho_n^* \in [0,1)$ s.t. for any $M/M/n$ queue with traffic intensity at least ρ_n^*, $\gamma = \left(\lambda^{\frac{1}{2}} - (n\mu)^{\frac{1}{2}}\right)^2$; and for any $M/M/n$ queue with traffic intensity strictly less than ρ_n^*, $\gamma < \left(\lambda^{\frac{1}{2}} - (n\mu)^{\frac{1}{2}}\right)^2$.*

Unfortunately, all of the characterizations (including that of ρ_n^*) given by van Doorn are again stated in terms of the roots of high-degree polynomials, and van Doorn himself comments in [35] that one is generally better off using the approximations that he gives in the same paper. Van Doorn's work was later extended by Kijima in [26], and similar results were achieved by Zeifman using different techniques in [42]. It was also shown in [42] that $\rho_n^* \leq (1 - \frac{1}{n})^2$.

There are also some results in the literature for explicitly bounding the distance to stationarity, as opposed to just identifying the exponential rate of convergence. In [42], Zeifman used tools from the theory of differential equations to give explicit bounds on the total variational distance between the transient and steady-state distributions of a BDP, and explicitly examines the $M/M/n$ queue. In [38],[39], van Doorn and Zeifman used the techniques developed in [42] to derive explicit bounds on the distance to stationarity for a different queueing model, and examined how their bounds perform in a certain heavy-traffic regime (not HW). In [5], Chen developed very general bounds for the distance to stationarity for Markov chains, and then applied these to BDP. However, these bounds are generally not studied in the HW regime, and thus may not scale desirably with n in the HW regime. We note that the complexity of bounding the distance to stationarity uniformly for a sequence of BDP is related to the cutoff phenomenon for Markov chains [8], which has been studied in the context of queueing systems [15].

In this paper, we prove several results about the rate of convergence to stationarity for the $M/M/n$ queue in the HW regime. We identify the limiting rate of convergence to steady-state, i.e. the spectral gap, and discover an asymptotic phase transition that occurs w.r.t. this rate. Specifically, let γ_n denote the spectral gap associated with the $M/M/n$ queue with arrival rate $n - Bn^{\frac{1}{2}}$ and service rate equal to unity. Then we demonstrate the existence of a constant $B^* \approx 1.85772$ s.t. when the excess parameter $B \in (0, B^*]$, $\lim_{n\to\infty} \gamma_n = \frac{B^2}{4}$. For $B > B^*$, $\lim_{n\to\infty} \gamma_n$ exists, and can be given as the solution to an explicit equation involving special functions. This result may be interpreted as an asymptotic version of the phase transition proven to occur for any fixed n by van Doorn in [33]. Indeed, we prove that $\lim_{n\to\infty} n^{\frac{1}{2}}(1 - \rho_n^*) = B^*$. It thus follows from the results of [33] (see Theorem 3) that $\gamma_n = \left(n^{\frac{1}{2}} - (n - Bn^{\frac{1}{2}})^{\frac{1}{2}}\right)^2$ for $B < B^*$ and all sufficiently large n. Observing that $\lim_{n\to\infty} \left(n^{\frac{1}{2}} - (n - Bn^{\frac{1}{2}})^{\frac{1}{2}}\right)^2 = \frac{B^2}{4}$ links our results to those of van Doorn for the case $B < B^*$, and a similar connection exists for the case $B \geq B^*$.

We also prove explicit bounds on the distance to stationarity for the $M/M/n$ queue in the HW regime, when $B < B^*$. Our bounds scale independently of n in the HW regime, and do not follow from the weak-convergence theory.

1.1. *Outline of paper.* The rest of the paper proceeds as follows. In Section 2, we state our main results, and outline our proof technique. In Section 3, we prove a new characterization for the

spectral gap of the $M/M/n$ queue. In Sections 4 - 6, we study the asymptotic properties of this characterization. In Section 7, we compute the limiting spectral gap of the $M/M/n$ queue in the HW regime, and prove that a phase transition occurs. In Section 8, we prove our explicit bounds on the distance to stationarity. In Section 9, we compare our explicit bounds to other bounds from the literature. In Section 10 we summarize our main results and present ideas for future research. We include a technical appendix in Section 11.

2. Main Results.

2.1. *Definitions and notations.* Let $\mathcal{Q}^n$ denote the $M/M/n$ queue with arrival rate $\lambda_n \triangleq n - Bn^{\frac{1}{2}}$ and service rate $\mu \triangleq 1$, where we assume throughout that n is sufficiently large to ensure that $\lambda_n > 0$, and $n > \lambda_n + 1$. Let $Q^n(t)$ denote the number in system, i.e. the number of jobs in service plus the number of jobs waiting in queue, at time t; $Q^n(\infty)$ denote the corresponding steady-state r.v.; and γ_n denote the spectral gap of the associated Markov chain. We define $P^n_{i,j}(t) \triangleq \mathbb{P}\big(Q^n(t) = j | Q^n(0) = i\big)$, $P^n_j(\infty) \triangleq \mathbb{P}\big(Q^n(\infty) = j\big)$, $P^n_{i,\leq j}(t) \triangleq \sum_{k=0}^{j} P^n_{i,k}(t)$, and $P^n_{\leq j}(\infty) \triangleq \sum_{k=0}^{j} P^n_k(\infty)$. For a function f, we let $Z(f)\big(Z^+(f)\big)$ denote the infimum of the set of (strictly positive) real zeros of f, and set $Z(f)\big(Z^+(f)\big) = \infty$ if f has no (strictly positive) real zeros. All logarithms will be base e. Unless otherwise stated, all functions are defined only over $\mathbb{R}$. All empty products are assumed to be equal to unity, and all empty summations are assumed to be equal to zero. Also, for an event $\{E\}$, we let $I\big(\{E\}\big)$ denote the corresponding indicator function.

2.2. *The parabolic cylinder functions.* We now briefly review the two-parameter function commonly referred to as the parabolic cylinder function $D_x(z)$, since we will need these functions for the statement (and proof) of our main results. For excellent references on these functions, see [17] Section 8.31 and Section 9.24, [3] Sections 3.3-3.5, and [12] Chapter 8. Let Γ denote the Gamma function (see [19], Chapter 8.8). It is stated in [3] that $x, z \in \mathbb{R}$ implies $D_x(z) \in \mathbb{R}$, and

$$D_x(z) = \begin{cases} (\frac{2}{\pi})^{\frac{1}{2}} \exp(\frac{z^2}{4}) \int_0^\infty \exp(-\frac{y^2}{2}) cos(\frac{\pi}{2}x - zy) y^x dy & \text{if } x \geq 0; \\ \frac{\exp(-\frac{z^2}{4})}{\Gamma(-x)} \int_0^\infty \exp(-\frac{y^2}{2} - zy) y^{-(x+1)} dy & \text{if } x < 0. \end{cases} \tag{1}$$

$D_x(z)$ takes on a simpler form for integral x. In particular, it is stated in [17] that for $z \in \mathbb{R}$,

$$D_{-1}(z) = 2^{\frac{1}{2}} \exp(\frac{z^2}{4}) \int_{2^{-\frac{1}{2}}z}^\infty \exp(-y^2) dy, \quad D_0(z) = \exp(-\frac{z^2}{4}), \quad \text{and} \quad D_1(z) = z \exp(-\frac{z^2}{4}). \tag{2}$$

Note that since $\Gamma(-x) \in (0, \infty)$ for $x < 0$, (1) and (2) imply that $D_x(z) > 0$ for $z \in \mathbb{R}$ and $x \leq 0$.

The parabolic cylinder functions arise in several contexts associated with the limits of queueing models, such as the Ornstein-Uhlenbeck limit of the appropriately scaled infinite-server queue [20] and various limits associated with the Erlang loss model [40]. We note that the parabolic cylinder functions have been studied as the limits of certain polynomials under the HW scaling, using tools from the theory of differential equations [9],[10],[11],[2].

2.3. *Main results.* We now state our main results. We begin by identifying the limiting rate of convergence to steady-state, i.e. the limiting spectral gap, for the $M/M/n$ queue in the HW regime; and prove that a phase transition occurs w.r.t. this limiting rate. We define

$$\upsilon(x, y) \triangleq \begin{cases} \frac{D_x(y)}{D_{x-1}(y)} & \text{if } D_{x-1}(y) \neq 0; \\ \infty & \text{otherwise.} \end{cases}$$

Also, let $\varphi(B) \triangleq \upsilon(\frac{B^2}{4}, -B)$, $\zeta(B) \triangleq \varphi(B) + \frac{B}{2}$, and

$$\Psi_\infty(x) \triangleq \begin{cases} \upsilon(x, -B) + \frac{1}{2}\big(B + (B^2 - 4x)^{\frac{1}{2}}\big) & \text{if } x \leq \frac{B^2}{4}; \\ \infty & \text{otherwise.} \end{cases}$$

Note that $\zeta(B) = \Psi_\infty(\frac{B^2}{4})$. We include a plot of ζ.

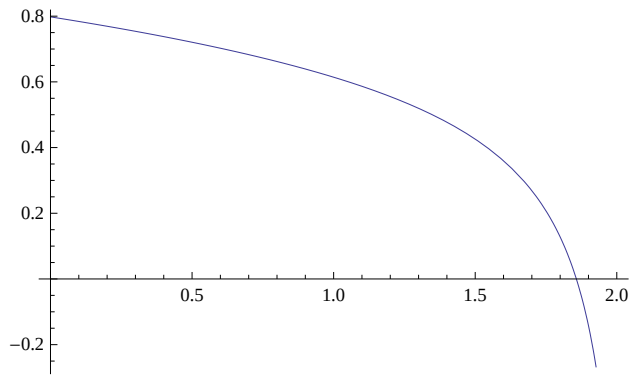

FIG 1. *Plot of* ζ

Let $B^* \triangleq Z^+(\zeta)$. Then

PROPOSITION 1. *$B^* \approx 1.85772$, and $Z^+(\Psi_\infty) \in \big(0, \min(1, \frac{B^2}{4})\big)$ for $B > B^*$.*

Our main result is that

THEOREM 4. *The limit $\gamma(B) \triangleq \lim_{n\to\infty} \gamma_n$ exists for all $B > 0$. For $0 < B \leq B^*$, $\gamma(B) = \frac{B^2}{4}$. For $B \geq B^*$, $\gamma(B) = Z^+(\Psi_\infty)$.*

We include a plot of γ.

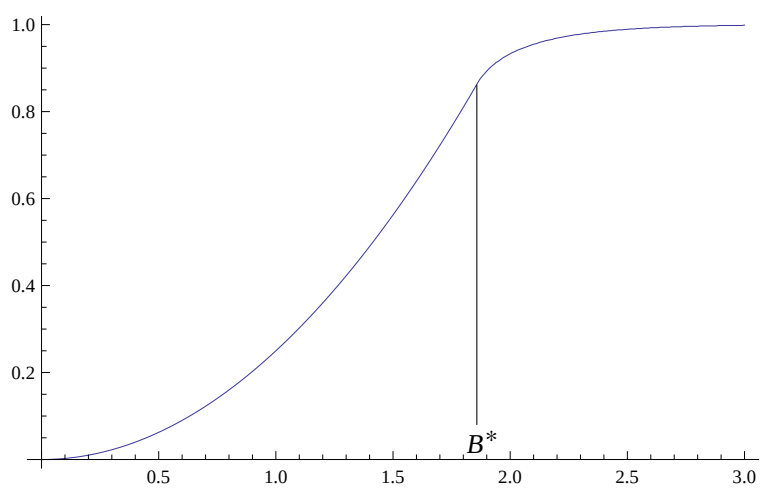

FIG 2. *Plot of* γ

Due to the non-linear manner in which the steady-state probability of wait scales in the HW regime, the case $0 < B < B^*$ actually encompasses most scenarios of practical interest. Indeed, it is proven in [18] that the limit of the steady-state probability of wait equals $\big(1+B\exp(\frac{1}{2}B^2)\int_{-\infty}^{B}\exp(-\frac{1}{2}z^2)dz\big)^{-1}$. As this limit is monotone in B, the case $0 < B < B^*$ includes all scenarios for which the steady-state probability of wait is at least .04.

We note that the results of [28] show that $\gamma(B)$ is also the spectral gap of the HW diffusion, demonstrating an interchange of limits for the $M/M/n$ queue in the HW regime. Namely, the limit of the sequence of spectral gaps equals the spectral gap of the corresponding weak limit. Interestingly, neither result implies the other, and it is an open challenge to understand this interchange more generally.

The following corollary may be interpreted as an asymptotic version of Theorem 3.

Corollary 1. *The ρ_n^* parameter of Theorem 3 satisfies*

$$\lim_{n\to\infty} n^{\frac{1}{2}}(1-\rho_n^*) = B^*.$$

We now give an interpretation of Theorem 4 and Corollary 1. The $M/M/n$ queue behaves like an $M/M/1$ queue when all servers are busy, and an $M/M/\infty$ queue when at least one server is idle. The phase transition of Theorem 4 formalizes this relationship in a new way. For $0 < B < B^*$, the KM spectral measure of the $M/M/n$ queue in the HW regime has no jumps away from the origin, and has spectral gap equal to $(\lambda_n^{\frac{1}{2}} - n^{\frac{1}{2}})^2$, two properties shared by the associated $M/M/1$ queue (see Theorem 2). For $B > B^*$, the KM spectral measure has at least one jump away from the origin, like the associated $M/M/\infty$ queue (whose spectral measure has only jumps and spectral gap equal to unity, see Theorem 2). Another interpretation is that the $M/M/n$ queue cannot approach stationarity faster than either component system would on its own.

We now state our explicit bounds on the distance to stationarity for the case $B < B^*$.

Theorem 5. *Given $B \in (0, B^*)$ and $a_1, a_2 \in \mathbb{R}$, let $a = \max(|a_1|, |a_2|, B)$. Then there exists $N_{B,a_1,a_2} < \infty$, depending only on B, a_1, and a_2, s.t. for all $n \geq N_{B,a_1,a_2}$ and $t \geq 1$,*

$$\left| n^{\frac{1}{2}} P^n_{\lceil n+a_1 n^{\frac{1}{2}}\rceil, \lceil n+a_2 n^{\frac{1}{2}}\rceil}(t) - n^{\frac{1}{2}} P^n_{\lceil n+a_2 n^{\frac{1}{2}}\rceil}(\infty) \right| \leq t^{-\frac{1}{2}} \exp\big(30(a^2+1) - \frac{B^2}{4}t\big), \tag{3}$$

and

$$\left| P^n_{\lceil n+a_1 n^{\frac{1}{2}}\rceil, \leq \lceil n+a_2 n^{\frac{1}{2}}\rceil}(t) - P^n_{\leq \lceil n+a_2 n^{\frac{1}{2}}\rceil}(\infty) \right| \leq B^{-1} t^{-\frac{1}{2}} \exp\big(30(a^2+1) - \frac{B^2}{4}t\big). \tag{4}$$

Note that Theorem 5 provides a bound for any sufficiently large fixed n and all times t greater than unity, which is independent of n, and converges to zero as $t \to \infty$. Interestingly, such uniform bounds do not follow directly from the weak-convergence theory, since the standard framework of weak convergence requires that one first fix a finite time interval of interest, and then let $n \to \infty$, in that order.

It follows from the weak-convergence theory that our explicit bounds yield corresponding bounds for the distance to stationarity of the HW diffusion. Furthermore, in light of Theorem 4, the exponent $\frac{B^2}{4}$ appearing in our bounds is the best possible. Although we were able to derive partial results for the case $B \geq B^*$, the derived bounds were considerably more complicated than those

of Theorem 5, and we leave it as an open question to derive simple explicit bounds for the case $B \geq B^*$. We note that the results of [28] suggest that the exponential dependence on a^2, and inverse dependence on $t^{\frac{1}{2}}$, of the prefactor appearing in Theorem 5 may not be tight, and it seems likely that a more refined analysis would yield sharper bounds.

2.4. *Outline of proof.* We now present an outline of the proof of our main results. To prove Theorem 4 and Corollary 1, we give a new characterization for the spectral gap γ_n, and then study its asymptotics in the HW regime. More precisely, in Section 3, we prove a new characterization for the spectral gap γ_n, in terms of a certain function Ψ_n which we define. We express γ_n in terms of three quantities: $(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$, $Z^+(\Psi_n)$, and the sign of $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)$. In Section 4, we prove that in the HW regime, Ψ_n converges to Ψ_∞, and $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)$ converges to $\zeta(B)$. In Section 5, we prove that in the HW regime, $Z^+(\Psi_n)$ converges to $Z^+(\Psi_\infty)$. In Section 6, we characterize the sign of $\zeta(B)$. In Section 7, we combine the above results to prove Theorem 4 and Corollary 1. To prove Theorem 5, we use induction arguments to bound certain polynomials which appear in the KM representation for the transient $M/M/n$ queue.

3. Characterization for γ_n. In this section we give a new characterization for γ_n. We begin by associating several functions to the $M/M/n$ queue, as in [26] and [33]. For $0 \leq k \leq n$, let $f_{n,k}(x) \triangleq \sum_{j=0}^{k} \binom{k}{j} \lambda_n^j \prod_{i=1}^{k-j}(i-x)$;

$$z_{n,k}(x) \triangleq \begin{cases} \frac{f_{n,k}(x)}{f_{n,k-1}(x)} & \text{if } f_{n,k-1}(x) \neq 0, \\ \infty & \text{otherwise;} \end{cases}$$

and $z_n(x) \triangleq z_{n,n}(x)$. We also define

$$a_n(x) \triangleq \begin{cases} \frac{1}{2}\Big(\lambda_n + n - x - \big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 - x\big)^{\frac{1}{2}}\big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - x\big)^{\frac{1}{2}}\Big) & \text{if } x \leq (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2, \\ \infty & \text{otherwise;} \end{cases}$$

and

$$\Psi_n(x) \triangleq \begin{cases} z_n(x) - a_n(x) & \text{if } z_n(x) \neq \infty \text{ or } a_n(x) \neq \infty, \\ \infty & \text{otherwise.} \end{cases}$$

We now cite some properties of $f_{n,n-1}, z_{n,k}$, and Ψ_n, as stated in [26], for use in later proofs.

Lemma 1. *(i) $f_{n,n-1}$ is strictly positive on $\big(-\infty, 1]$.*
(ii) For $k \leq n$, $z_{n,k}$ is strictly positive, continuous, and strictly decreasing on $(-\infty, 1]$.
(iii) Ψ_n is continuous and strictly decreasing on $\Big(-\infty, \min\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2, 1\big)\Big]$.

We now prove the main result of this section, a new characterization for γ_n. In particular,

Proposition 2. *(i) If $(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 < 1$ and $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) < 0$, then $\gamma_n = Z^+(\Psi_n)$.*
(ii) If $(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 < 1$ and $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) \geq 0$, then $\gamma_n = (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$.
(iii) If $(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 \geq 1$, then $Z^+(\Psi_n) \in \big(0, 1\big)$, and $\gamma_n = Z^+(\Psi_n)$.

The proof of Proposition 2 relies on the following known characterization for γ_n. Let $\sigma_n(x) \stackrel{\Delta}{=} f_{n,n}(x) - (\lambda_n n)^{\frac{1}{2}} f_{n,n-1}(x)$, and $\psi_n(x) \stackrel{\Delta}{=} f_{n,n}(x) - a_n(x) f_{n,n-1}(x)$. Then it is proven in [26] that

THEOREM 6. *If $Z(\sigma_n) \geq (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$, then $\gamma_n = (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$. If $Z(\sigma_n) < (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$, then $\gamma_n = Z(\psi_n)$.*

With Theorem 6 in hand, we now complete the proof of Proposition 2.

PROOF OF PROPOSITION 2. We begin by studying the sign of $\Psi_n(0), \Psi_n(1), \sigma_n(0)$ and $\sigma_n(1)$. Note that

$$\begin{aligned}
\Psi_n(0) &= \frac{\sum_{k=0}^{n} \binom{n}{k} \lambda_n^k (n-k)!}{\sum_{k=0}^{n-1} \binom{n-1}{k} \lambda_n^k (n-1-k)!} - \frac{1}{2}\Big(\lambda_n + n - \big((\lambda_n + n)^2 - 4\lambda_n n\big)^{\frac{1}{2}}\Big) \\
&= n \frac{\sum_{k=0}^{n} \frac{\lambda_n^k}{k!}}{\sum_{k=0}^{n-1} \frac{\lambda_n^k}{k!}} - \lambda_n \quad > \quad 0.
\end{aligned}$$

If $(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 \geq 1$, then

$$\begin{aligned}
\Psi_n(1) &= \frac{\sum_{k=0}^{n} \binom{n}{k} \lambda_n^k \prod_{i=1}^{n-k}(i-1)}{\sum_{k=0}^{n-1} \binom{n-1}{k} \lambda_n^k \prod_{i=1}^{n-1-k}(i-1)} - \frac{1}{2}\Big(\lambda_n + n - 1 - \big((\lambda_n + n - 1)^2 - 4\lambda_n n\big)^{\frac{1}{2}}\Big) \\
&= \frac{\lambda_n^n}{\lambda_n^{n-1}} - \frac{1}{2}\Big(\lambda_n + n - 1 - \big((\lambda_n + n - 1)^2 - 4\lambda_n n\big)^{\frac{1}{2}}\Big) \\
&= \frac{1}{2}\Big(\lambda_n - n + 1 + \big((\lambda_n - n + 1)^2 - 4\lambda_n\big)^{\frac{1}{2}}\Big) \quad \leq \quad 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\sigma_n(0) &= \sum_{k=0}^{n} \binom{n}{k} \lambda_n^k (n-k)! - (\lambda_n n)^{\frac{1}{2}} \sum_{k=0}^{n-1} \binom{n-1}{k} \lambda_n^k (n-1-k)! \\
&= (n-1)!\big(n \sum_{k=0}^{n} \frac{\lambda_n^k}{k!} - (\lambda_n n)^{\frac{1}{2}} \sum_{k=0}^{n-1} \frac{\lambda_n^k}{k!}\big) \quad \geq \quad (n-1)! \sum_{k=0}^{n} \frac{\lambda_n^k}{k!}(n - (\lambda_n n)^{\frac{1}{2}}) \quad > \quad 0,
\end{aligned}$$

and

$$\begin{aligned}
\sigma_n(1) &= \sum_{k=0}^{n} \binom{n}{k} \lambda_n^k \prod_{i=1}^{n-k}(i-1) - (\lambda_n n)^{\frac{1}{2}} \sum_{k=0}^{n-1} \binom{n-1}{k} \lambda_n^k \prod_{i=1}^{n-1-k}(i-1) \\
&= \lambda_n^n - (\lambda_n n)^{\frac{1}{2}} \lambda_n^{n-1} \quad < \quad 0.
\end{aligned}$$

We first prove assertion (i). Note that if $f_{n,n-1}(x) \neq 0$, then $z_n(x) - (\lambda_n n)^{\frac{1}{2}} = \frac{\sigma_n(x)}{f_{n,n-1}(x)}$. Thus Lemma 1.(i) implies that σ_n is the same sign as $z_n - (\lambda_n n)^{\frac{1}{2}}$ on $\big(-\infty, (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big]$. Recalling that $\sigma_n(0) > 0$, it follows from the continuity/monotonicity of z_n $\big($guaranteed by Lemma 1.(ii)$\big)$ and the Intermediate Value Theorem that σ_n has a zero on $\big(-\infty, (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)$ iff $z_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) - (\lambda_n n)^{\frac{1}{2}} < 0$. Since $a_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) = (\lambda_n n)^{\frac{1}{2}}$, we conclude that $Z(\sigma_n) < (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$ iff $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) < 0$.

Thus $Z(\sigma_n) < (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$, since by assumption $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) < 0$, and $\gamma_n = Z(\psi_n)$ by Theorem 6. Noting that $\Psi_n = \frac{\psi_n}{f_{n,n-1}}$ on $\big(-\infty, (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big]$, this further implies that $\gamma_n = Z(\Psi_n)$. That $\gamma_n = Z^+(\Psi_n)$ then follows from the fact that $\Psi_n(0) > 0$, and the continuity/monotonicity of Ψ_n guaranteed by Lemma 1.(iii). This completes the proof of assertion (i). The proof of assertion (ii) follows from a similar argument, and we omit the details.

We now prove assertion (iii). Since σ_n is a polynomial s.t. $\sigma_n(0) > 0$ and $\sigma_n(1) < 0$, we have that $Z(\sigma_n) < 1 \leq (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$. Thus Theorem 6 implies that $\gamma_n = Z(\psi_n)$. As in the proof of assertion (i), it follows that $\gamma_n = Z(\Psi_n)$. Since $\Psi_n(0) > 0$ and $\Psi_n(1) < 0$, the continuity/monotonicity of Ψ_n guaranteed by Lemma 1.(iii) further ensures that $\gamma_n = Z^+(\Psi_n) \in (0,1)$, completing the proof. □

4. Asymptotic Analysis of Ψ_n. In this section we derive the asymptotics of Ψ_n in the HW regime. In particular, we prove that

THEOREM 7. *For $B > 0$ and $x \in (0,1) \bigcap (0, \frac{B^2}{4}]$, $\lim_{n\to\infty} \lambda_n^{-\frac{1}{2}} \Psi_n(x) = \Psi_\infty(x)$.*

We also prove that

COROLLARY 2. *For $B \in (0,2)$, $\lim_{n\to\infty} \lambda_n^{-\frac{1}{2}} \Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) = \zeta(B)$.*

We proceed by separately analyzing the asymptotics of $\lambda_n^{-\frac{1}{2}}\big(a_n - \lambda_n\big)$ and $\lambda_n^{-\frac{1}{2}}\big(z_n - \lambda_n\big)$, beginning with a_n. Let

$$a_\infty(x) \triangleq \begin{cases} \frac{1}{2}\big(B - (B^2 - 4x)^{\frac{1}{2}}\big) & \text{if } x \leq \frac{B^2}{4}; \\ \infty & \text{otherwise.} \end{cases}$$

Then

LEMMA 2. *For $x \in [0, \frac{B^2}{4}]$, $\lim_{n\to\infty} \lambda_n^{-\frac{1}{2}}\big(a_n(x) - \lambda_n\big) = a_\infty(x)$.*

PROOF. Note that

$$\lambda_n^{-\frac{1}{2}}\big(a_n(x) - \lambda_n\big) = \Big(Bn^{\frac{1}{2}} - x - \big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - x\big)^{\frac{1}{2}}\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 - x\big)^{\frac{1}{2}}\Big)(2\lambda_n^{\frac{1}{2}})^{-1}.$$

The lemma then follows from the fact that $\lim_{n\to\infty}(Bn^{\frac{1}{2}} - x)(2\lambda_n^{\frac{1}{2}})^{-1} = \frac{B}{2}$, $\lim_{n\to\infty}\big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - x\big)^{\frac{1}{2}}\big(2\lambda_n^{\frac{1}{2}}\big)^{-1} = 1$, and $\lim_{n\to\infty}(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}}) = \frac{B}{2}$. □

We now analyze the asymptotics of z_n, and begin by proving some necessary bounds. Let us fix some $x \in (0,1)$ and integer $T \geq 3$, and define

$$R_{1,n} \triangleq \lambda_n^{\frac{x-1}{2}} \sum_{k=0}^{n-(T+1)} (n-k)^{1-x} \exp(-\lambda_n) \frac{\lambda_n^k}{k!},$$

$$R_{2,n} \triangleq \lambda_n^{\frac{x-2}{2}} \sum_{k=0}^{\lceil n - T^{-1} n^{\frac{1}{2}} \rceil} k(n-k)^{-x} \exp(-\lambda_n) \frac{\lambda_n^k}{k!}.$$

Then

LEMMA 3. *For all sufficiently large n, $\lambda_n^{-\frac{1}{2}}\big(z_n(x)-\lambda_n\big)$ is at least*

$$\exp(-4T^{-1})\frac{R_{1,n}}{R_{2,n}+4(1-x)^{-1}T^{-(1-x)}},$$

and at most

$$\exp(4T^{-1})\frac{R_{1,n}+4(1-x)^{-1}T^{-(1-x)}}{R_{2,n}}.$$

PROOF. The proof is deferred to the appendix. □

Letting $z_\infty(x) \triangleq \upsilon(x,-B)+B$, we now use Lemma 3 to demonstrate that

PROPOSITION 3. *For $x\in(0,1)$, $\lim_{n\to\infty}\lambda_n^{-\frac{1}{2}}\big(z_n(x)-\lambda_n\big)=z_\infty(x)$.*

PROOF. We proceed by relating $R_{1,n}$ and $R_{2,n}$ to the expectations of certain functions of a scaled Poisson r.v., and then analyze these expectations as $n\to\infty$ using tools from weak-convergence theory. Let X_n denote a Poisson r.v. with mean λ_n, $Z_n \triangleq \lambda_n^{-\frac{1}{2}}(X_n-\lambda_n)$,

$$Y_{1,n} \triangleq \big(B(\frac{n}{\lambda_n})^{\frac{1}{2}}-Z_n\big)^{1-x} I\Big(Z_n \le B(\frac{n}{\lambda_n})^{\frac{1}{2}}-(T+1)\lambda_n^{-\frac{1}{2}}\Big),$$

and

$$Y_{2,n} \triangleq \big(B(\frac{n}{\lambda_n})^{\frac{1}{2}}-Z_n\big)^{-x} I\Big(Z_n \le (B-T^{-1})(\frac{n}{\lambda_n})^{\frac{1}{2}}+\lambda_n^{-\frac{1}{2}}\big(\lceil n-T^{-1}n^{\frac{1}{2}}\rceil-(n-T^{-1}n^{\frac{1}{2}})\big)\Big).$$

It follows from a straightforward computation that $R_{1,n}=\mathbb{E}[Y_{1,n}]$, and $R_{2,n}=\lambda_n^{-\frac{1}{2}}\mathbb{E}[Z_nY_{2,n}]+E[Y_{2,n}]$. Let $f_1(y)\triangleq(B-y)^{1-x}I(y\le B), f_2(y)\triangleq(B-y)^{-x}I(y\le B-T^{-1})$, $f_3(y)\triangleq y(B-y)^{-x}I(y\le B-T^{-1})$, and N denote a normal r.v. with zero mean and unit variance. It may be easily verified that $\{Y_{1,n}\},\{Y_{2,n}\}$, and $\{Z_nY_{2,n}\}$ are uniformly integrable sequences of r.v.s, and converge in distribution to $f_1(N), f_2(N), f_3(N)$ respectively. It follows that $\lim_{n\to\infty}\mathbb{E}[Y_{1,n}]=\mathbb{E}[f_1(N)]=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{B}(B-y)^{1-x}\exp(-\frac{y^2}{2})dy$, $\lim_{n\to\infty}\mathbb{E}[Y_{2,n}]=\mathbb{E}[f_2(N)]=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{B-T^{-1}}(B-y)^{-x}\exp(-\frac{y^2}{2})dy$, and $\lim_{n\to\infty}\mathbb{E}[Z_nY_{2,n}]=\mathbb{E}[f_3(N)]=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{B-T^{-1}}y(B-y)^{-x}\exp(-\frac{y^2}{2})dy$. Plugging the above limits into Lemma 3, and letting $T\to\infty$, we conclude that

$$\lim_{n\to\infty}\lambda_n^{-\frac{1}{2}}\big(z_n(x)-\lambda_n\big)=\frac{\int_{-\infty}^{B}(B-y)^{1-x}\exp(-\frac{y^2}{2})dy}{\int_{-\infty}^{B}(B-y)^{-x}\exp(-\frac{y^2}{2})dy}. \tag{5}$$

We now complete the proof by relating the integrals appearing in (5) to the parabolic cylinder functions. It is stated in [17] that for all $x,z\in\mathbb{R}$,

$$D_{x+1}(z)-zD_x(z)+xD_{x-1}(z)=0. \tag{6}$$

Combining (1) and (6), we find that the r.h.s. of (5) equals

$$\frac{\int_0^\infty y^{1-x}\exp(-\frac{(B-y)^2}{2})dy}{\int_0^\infty y^{-x}\exp(-\frac{(B-y)^2}{2})dy} = \frac{\Gamma(2-x)\frac{\big(D_x(-B)+BD_{x-1}(-B)\big)}{1-x}}{\Gamma(1-x)D_{x-1}(-B)} = z_\infty(x),$$

where the final equality follows from the fact that $\frac{\Gamma(2-x)}{\Gamma(1-x)}=1-x$. □

We now complete the proofs of Theorem 7 and Corollary 2.

Proof of Theorem 7 and Corollary 2. Since $\Psi_n(x) = z_n(x) - a_n(x)$, Theorem 7 follows from Lemma 2 and Proposition 3.

We now prove Corollary 2. It follows from the monotonicity of z_n guaranteed by Lemma 1.(ii) that for any sufficiently small positive ϵ and all sufficiently large n, one has

$$\lambda_n^{-\frac{1}{2}}\big(z_n(\frac{B^2}{4}+\epsilon)-\lambda_n\big) \le \lambda_n^{-\frac{1}{2}}\Big(z_n\big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2\big)-\lambda_n\Big) \le \lambda_n^{-\frac{1}{2}}\big(z_n(\frac{B^2}{4}-\epsilon)-\lambda_n\big).$$

Thus by Proposition 3, for all sufficiently small $\epsilon > 0$,

(7)

$$z_\infty(\frac{B^2}{4}+\epsilon) \le \liminf_{n\to\infty} \lambda_n^{-\frac{1}{2}}\Big(z_n\big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2\big)-\lambda_n\Big) \le \limsup_{n\to\infty} \lambda_n^{-\frac{1}{2}}\Big(z_n\big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2\big)-\lambda_n\Big) \le z_\infty(\frac{B^2}{4}-\epsilon).$$

We now prove that z_∞ is continuous in a neighborhood of $\frac{B^2}{4}$, from which we conclude that $\lim_{n\to\infty} \lambda_n^{-\frac{1}{2}}\Big(z_n\big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2\big)-\lambda_n\Big) = z_\infty(\frac{B^2}{4})$. Indeed, since $D_x(z) > 0$ for all $z \in \mathbb{R}$ and $x \le 0$, it follows that $D_{x-1}(-B) > 0$ for $x \le 1$. The continuity of z_∞ on $(-\infty, 1]$ then follows from the fact that $D_x(-B)$ is an entire function of x [6].

Since $a_n\big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2\big) = (\lambda_n n)^{\frac{1}{2}}$, we also have that $\lim_{n\to\infty} \lambda_n^{-\frac{1}{2}}\Big(a_n\big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2\big)-\lambda_n\Big) = \frac{B}{2}$. Combining the above completes the proof, since $\zeta(B) = z_\infty(\frac{B^2}{4}) - \frac{B}{2}$. □

5. Asymptotic Analysis of $Z^+(\Psi_n)$. In this section we derive the asymptotics of $Z^+(\Psi_n)$ in the HW regime. In particular, we prove that

Theorem 8. *If $B < 2$ and $\zeta(B) \le 0$, or $B \ge 2$, then $\lim_{n\to\infty} Z^+(\Psi_n) = Z^+(\Psi_\infty)$.*

We first prove some additional properties of $Z^+(\Psi_\infty)$. Namely,

Lemma 4. *If $B < 2$ and $\zeta(B) < 0$, or $B \ge 2$, then: Ψ_∞ has a unique zero $Z^+(\Psi_\infty) \in \big(0, \min(1, \frac{B^2}{4})\big)$; Ψ_∞ is strictly positive on $[0, Z^+(\Psi_\infty))$; and Ψ_∞ is strictly negative on $\big(Z^+(\Psi_\infty), \min(1, \frac{B^2}{4})\big]$. Alternatively, if $B < 2$ and $\zeta(B) = 0$, then: Ψ_∞ is strictly positive on $[0, \min(1, \frac{B^2}{4}))$, and $Z^+(\Psi_\infty) = \frac{B^2}{4}$.*

Proof. We begin by proving that Ψ_∞ is continuous and strictly decreasing on $\big[0, \min(1, \frac{B^2}{4})\big]$. Since $\Psi_\infty = z_\infty - a_\infty$, it suffices to demonstrate the continuity and monotonicity of z_∞ and a_∞ separately. We have already shown that z_∞ is continuous on $(-\infty, 1]$, and it follows from Lemma 1.(ii) and Proposition 3 that z_∞ is non-increasing on $[0, 1]$. A straightforward calculation demonstrates that a_∞ is continuous and strictly increasing on $[0, \frac{B^2}{4}]$. Combining the above yields the desired result.

We now treat the case $B < 2$ and $\zeta(B) < 0$, or $B \ge 2$. Note that $\Psi_\infty(0) > 0$, since $\Psi_\infty(0) = \upsilon(0, -B) + B$, and by (2), $\upsilon(0, -B) > 0$. Also, $\Psi_\infty\big(\min(1, \frac{B^2}{4})\big) < 0$, which we now demonstrate by a case analysis. If $B < 2$ and $\zeta(B) < 0$, then $\min(1, \frac{B^2}{4}) = \frac{B^2}{4}$, and $\Psi_\infty(\frac{B^2}{4}) = \zeta(B) < 0$. Alternatively, if $B \ge 2$, then $\min(1, \frac{B^2}{4}) = 1$. But $\Psi_\infty(1) < 0$, since by (2), $\Psi_\infty(1) = -B + \frac{1}{2}\big(B + (B^2-4)^{\frac{1}{2}}\big) < 0$. Combining the above facts completes the proof. The case $B < 2$ and $\zeta(B) = 0$ follows similarly, and we omit the details. □

We now complete the proof of Theorem 8.

Proof of Theorem 8. We first treat the case $B < 2$ and $\zeta(B) < 0$, or $B \geq 2$, and begin by demonstrating that $\liminf_{n\to\infty} Z^+(\Psi_n) \geq Z^+(\Psi_\infty)$. Suppose for contradiction that $\liminf_{n\to\infty} Z^+(\Psi_n) < Z^+(\Psi_\infty)$. Then it follows from Lemma 4 that there exists $\epsilon > 0$ s.t. $0 < \liminf_{n\to\infty} Z^+(\Psi_n) + \epsilon < \min(1, \frac{B^2}{4})$, and $\Psi_\infty\big(\liminf_{n\to\infty} Z^+(\Psi_n) + \epsilon\big) > 0$. Thus by Theorem 7, for all sufficiently large n, $\Psi_n\big(\liminf_{n\to\infty} Z^+(\Psi_n) + \epsilon\big) > 0$, and by the monotonicity of Ψ_n $\big($see Lemma 1.(iii)$\big)$, Ψ_n is strictly positive on $\big(-\infty, \liminf_{n\to\infty} Z^+(\Psi_n) + \epsilon\big)$. But by the definition of $\liminf$, this implies the existence of an infinite strictly increasing sequence of integers $\{n_i\}$ s.t. $\Psi_{n_i}\big(Z^+(\Psi_{n_i})\big) > 0$ for all i. This is a contradiction, since $\Psi_{n_i}\big(Z^+(\Psi_{n_i})\big) = 0$ for all i, and we conclude that $\liminf_{n\to\infty} Z^+(\Psi_n) \geq Z^+(\Psi_\infty)$. The proof that $\limsup_{n\to\infty} Z^+(\Psi_n) \leq Z^+(\Psi_\infty)$, as well as the proofs for the case $B < 2$ and $\zeta(B) = 0$, follow similarly, and we omit the details. □

6. The Sign of ζ. In this section we characterize the sign of ζ on $(0, 2)$, proving that

Theorem 9. *$B^* \in (0, 2)$. ζ is strictly positive on $[0, B^*)$, and strictly negative on $(B^*, 2]$.*

We also complete the proof of Proposition 1. Although Theorem 9 seems clear from Figure 1, the formal proof of this fact is somewhat involved, since apriori it could be the case that ζ never actually becomes strictly negative at B^*, or that ζ has additional zeros on $(B^*, 2]$. We begin by proving a technical lemma about $\upsilon(x, -B)$.

Lemma 5. *For any fixed $B > 0$, $\upsilon(x, -B)$ is a concave function of x on $(0, 1)$.*

Proof. The proof is deferred to the appendix. □

We now prove some bounds for $\varphi'(B) \triangleq \frac{d}{dB}\varphi(B)$, when it exists.

Lemma 6. *φ is a differentiable function on $(0, 2)$, and*

$$\varphi'(B) < (2B^{-1} - B)\varphi(B) - \frac{B^2}{4} - \varphi^2(B) \leq B^{-2} - 1.$$

Proof. Note that $\upsilon(x, y)$ is a smooth function of y on $(-\infty, \infty)$ for any fixed $x \leq 1$, and a smooth function of x on $(-\infty, 1]$ for any fixed $y \in \mathbb{R}$. Indeed, this follows from the strict positivity of $D_{x-1}(y)$ for each fixed $x \leq 1$, and the fact that $D_x(y)$ is an entire function of y for each fixed x [12], and an entire function of x for each fixed y [6]. Thus we may apply the multivariate chain rule to φ. In light of (6), and the fact (stated in [17]) that for all $x, z \in \mathbb{R}$,

$$\frac{d}{dz}D_x(z) + \frac{1}{2}zD_x(z) - xD_{x-1}(z) = 0, \tag{8}$$

it then follows from a straightforward computation that φ is differentiable on $(-\infty, 2]$, and

$$\varphi'(B) = \frac{B}{2}\frac{d\upsilon}{dx}(\frac{B^2}{4}, -B) - \frac{B^2}{4} - \varphi^2(B) - B\varphi(B). \tag{9}$$

We now bound $\frac{d\upsilon}{dx}(\frac{B^2}{4}, -B)$. The Mean Value Theorem guarantees the existence of $c \in (0, \frac{B^2}{4})$ s.t. $\frac{d\upsilon}{dx}(c, -B) = (\frac{B^2}{4})^{-1}\big(\upsilon(\frac{B^2}{4}, -B) - \upsilon(0, -B)\big)$. In light of Lemma 5, we conclude that

$$\frac{d\upsilon}{dx}(\frac{B^2}{4}, -B) \leq (\frac{B^2}{4})^{-1}\big(\upsilon(\frac{B^2}{4}, -B) - \upsilon(0, -B)\big) < \frac{4}{B^2}\varphi(B), \tag{10}$$

where the final inequality follows from the fact that $\upsilon(0,-B)>0$ by (2). Combining (9) and (10) proves the first part of the lemma. It follows that there exists $x_B \in \mathbb{R}$ s.t. $\varphi'(B) \leq (2B^{-1}-B)x_B - x_B^2 - \frac{B^2}{4}$, which is at most $B^{-2}-1$ by elementary calculus. Combining the above completes the proof. $\square$

We now complete the proof of Theorem 9 and Proposition 1

Proof of Theorem 9 and Proposition 1. We first demonstrate that ζ is strictly positive on $[0,B^*)$. Indeed, this follows from (2), which implies that $\zeta(0) = \big(2^{\frac{1}{2}}\int_0^\infty \exp(-y^2)dy\big)^{-1} > 0$.

To complete the proof of Theorem 9, we will first show that $B^* \in (2^{\frac{1}{2}},2)$, and then apply Lemma 6 to prove that $\zeta'(B)<0$ on $(B^*,2)$. We show that $B^* \in (2^{\frac{1}{2}},2)$ in two stages, first proving that $B^* \in (0,2)$. (2) implies that $\zeta(2) = \frac{-2e^{-1}}{e^{-1}}+1<0$. That $B^* \in (0,2)$ then follows from the fact that $\zeta(0)>0$, and the Intermediate Value Theorem.

We now demonstrate that $B^* > 2^{\frac{1}{2}}$. It is proven in [35] Theorem 4.1 (i) that

$$\gamma_n \geq \inf_{k\geq 1}\Big(\lambda_n + \min(k,n) - \lambda_n^{\frac{1}{2}}\big(\min(k-1,n)^{\frac{1}{2}} + \min(k,n)^{\frac{1}{2}}\big)\Big). \tag{11}$$

Note that for $1\leq k\leq n$, $\lambda_n + \min(k,n) - \lambda_n^{\frac{1}{2}}\big(\min(k-1,n)^{\frac{1}{2}} + \min(k,n)^{\frac{1}{2}}\big)$ equals

$$(\lambda_n^{\frac{1}{2}} - k^{\frac{1}{2}})^2 + \frac{\lambda_n^{\frac{1}{2}}}{k^{\frac{1}{2}} + (k-1)^{\frac{1}{2}}} \;\geq\; \frac{1}{2}(\frac{\lambda_n}{n})^{\frac{1}{2}}.$$

For all $k \geq n+1$, the r.h.s. of (11) equals $(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$. Combining the above, we find that

$$\gamma_n \geq \min\big(\frac{1}{2}(\frac{\lambda_n}{n})^{\frac{1}{2}}, (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big).$$

Recalling that $\lim_{n\to\infty}(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 = \frac{B^2}{4}$, it follows that for any fixed $B < 2^{\frac{1}{2}}$ and all sufficiently large n, $\gamma_n \geq (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$.

Now, suppose for contradiction that $B^* < 2^{\frac{1}{2}}$. Then combining Lemma 6 with the fact that by construction $\varphi(B^*) = -\frac{B^*}{2}$, we find that

$$\varphi'(B^*) \;<\; (\frac{2}{B^*} - B^*)(-\frac{B^*}{2}) - \frac{B^{*2}}{4} - (-\frac{B^*}{2})^2 \;=\; -1.$$

It follows that $\zeta'(B^*)<0$, since $\zeta'(B^*) = \varphi'(B^*) + \frac{1}{2}$, and there exists $B' \in (0,2^{\frac{1}{2}})$ s.t. $\zeta(B')<0$. Thus if we define all relevant functions (e.g. λ_n, Ψ_n) in terms of B', Corollary 2 implies that $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) < 0$ for all sufficiently large n, and $\gamma_n < (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$ by Proposition 2.(i). But this is a contradiction, since we have already shown that $B' < 2^{\frac{1}{2}}$ implies that $\gamma_n \geq (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$ for all sufficiently large n, showing that $B^* > 2^{\frac{1}{2}}$.

We now complete the proof of Theorem 9 by demonstrating that $\zeta'(B)<0$ on $(2^{\frac{1}{2}},2)$. Indeed, for $B \in (2^{\frac{1}{2}},2)$, we have by Lemma 6 that $\zeta'(B)$ equals

$$\varphi'(B) + \frac{1}{2} \;<\; \frac{1}{B^2} - 1 + \frac{1}{2} \;= 0,$$

completing the proof of Theorem 9.

We now prove Proposition 1. In light of Theorem 9, the value of B^* may easily be evaluated numerically to the approximate value 1.85772. The second part of the proposition follows from Lemma 4. $\square$

7. Limiting Spectral Gap in the HW Regime and Asymptotic Phase Transition. In this section we complete the proofs of Theorem 4 and Corollary 1.

Proof of Theorem 4. First, suppose $0 < B < B^*$. Then it follows from Theorem 9 that $B < 2$, and $\zeta(B) > 0$. Combining with Corollary 2, we conclude that $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) > 0$ for all sufficiently large n, and $\gamma_n = (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$ by Proposition 2.(ii). Observing that $\lim_{n\to\infty}(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 = \frac{B^2}{4}$ completes the proof for this case.

Now, suppose $B = B^*$. By Proposition 2, for all sufficiently large n, either $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) < 0$, in which case $\gamma_n = Z^+(\Psi_n)$, or $\gamma_n = (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$. Let $\{n_i, i \geq 1\}$ denote the subsequence of $\{n\}$ for which $\Psi_{n_i}\big((n_i^{\frac{1}{2}} - \lambda_{n_i}^{\frac{1}{2}})^2\big) < 0$. If $\{n_i, i \geq 1\}$ is a finite set, then trivially $\gamma_n = (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$ for all sufficiently large n, and observing that $\lim_{n\to\infty}(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 = \frac{B^{*2}}{4}$ completes the proof. Alternatively, suppose $\{n_i, i \geq 1\}$ is an infinite set. Then Theorem 8 implies that $\lim_{i\to\infty} Z^+(\Psi_{n_i}) = \frac{B^{*2}}{4}$. Combining the above completes the proof for this case, since γ_n always belongs to one of two series, both of which converge to $\frac{B^{*2}}{4}$.

Next, consider the case $B \in (B^*, 2)$. It follows from Theorem 9 that $\zeta(B) < 0$. Combining with Corollary 2, we conclude that $\Psi_n\big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big) < 0$ for all sufficiently large n, and $\gamma_n = Z^+(\Psi_n)$ by Proposition 2.(i). That $\lim_{n\to\infty} \gamma_n = Z^+(\Psi_\infty)$ then follows from Theorem 8.

Finally, suppose $B \geq 2$. Then $(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 \geq 1$ for all sufficiently large n, and Proposition 2.(iii) implies that $\gamma_n = Z^+(\Psi_n)$. The proof then follows from Theorem 8. □

Proof of Corollary 1. Suppose for contradiction that $\liminf_{n\to\infty} n^{\frac{1}{2}}(1 - \rho_n^*) < B^*$. Then there exists $\epsilon > 0$, and an infinite strictly increasing sequence of integers $\{n_i, i \geq 1\}$, s.t. $\rho_{n_i}^* > 1 - (B^* - \epsilon)n_i^{-\frac{1}{2}}$ for all i. Consider the sequence $\{Z_i, i \geq 1\}$ of continuous time Markov chains, in which Z_i is an $M/M/n_i$ queueing system with $\lambda_{n_i} = n_i - (B^* - \epsilon)n_i^{\frac{1}{2}}$, $\mu = 1$. Let us define all relevant functions (e.g. $\Psi_{n_i}, \lambda_{n_i}$) w.r.t. $B^* - \epsilon$. Then since $\zeta(B^* - \epsilon) > 0$ and $B^* - \epsilon < 2$ by Theorem 9, it follows from Corollary 2 that $\Psi_{n_i}\big((n_i^{\frac{1}{2}} - \lambda_{n_i}^{\frac{1}{2}})^2\big) > 0$ for all sufficiently large i, and $\gamma_{n_i} = (n_i^{\frac{1}{2}} - \lambda_{n_i}^{\frac{1}{2}})^2$ by Proposition 2.(ii). But $\frac{\lambda_{n_i}}{n_i\mu} = 1 - (B^* - \epsilon)n_i^{-\frac{1}{2}} < \rho_{n_i}^*$ for all i. This is a contradiction, since by Theorem 3, $\frac{\lambda_{n_i}}{n_i\mu} < \rho_{n_i}^*$ implies that the spectral gap γ_{n_i} of Z_i is strictly less than $(n_i^{\frac{1}{2}} - \lambda_{n_i}^{\frac{1}{2}})^2$. Thus $\liminf_{n\to\infty} n^{\frac{1}{2}}(1 - \rho_n^*) \geq B^*$. A similar argument demonstrates that $\limsup_{n\to\infty} n^{\frac{1}{2}}(1 - \rho_n^*) \leq B^*$, and we omit the details. Combining the above completes the proof. □

8. Explicit Bounds on the Distance to Stationarity. In this section we complete the proof of Theorem 5.

8.1. *KM representation.* In this subsection we formally state the KM representation for the transient distribution of the $M/M/n$ queue, when the traffic intensity is at least ρ_n^*. Let

$$Q_{n,k}(x) \triangleq \begin{cases} 1 & \text{if k = 0,} \\ 1 - \frac{x}{\lambda_n} & \text{if k = 1,} \\ (1 - \frac{x}{\lambda_n} + \frac{\min(k-1,n)}{\lambda_n})Q_{n,k-1}(x) - \frac{\min(k-1,n)}{\lambda_n}Q_{n,k-2}(x) & \text{otherwise;} \end{cases} \tag{12}$$

and

$$c_n(x) \triangleq Q_{n,n}^2(x) - \frac{\lambda_n + n - x}{\lambda_n}Q_{n,n}(x)Q_{n,n-1}(x) + \frac{n}{\lambda_n}Q_{n,n-1}^2(x).$$

It is proven in [33] that c_n is strictly positive on $\big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2,(n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2\big)$. We also define

$$b_n(x) \triangleq \begin{cases} \big(x-(n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2\big)^{\frac{1}{2}}\big((n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2-x\big)^{\frac{1}{2}} & \text{if } (n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2 \le x \le (n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2, \\ \infty & \text{otherwise;} \end{cases}$$

and let $g_n(k) \triangleq \lambda_n^{k-n} n^{\min(n-k,0)} \prod_{i=k+1}^n i$. Then it is proven by KM in [25] (see also [33]) that

THEOREM 10. *If $\frac{\lambda_n}{n} \ge \rho_n^*$, then for all $i,j,t \ge 0$,*

$$P_{i,j}^n(t) - P_j^n(\infty) = (2\pi)^{-1} g_n(j)(\lambda_n n)^{-1} \int_{(n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2}^{(n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2} \exp(-xt) Q_{n,i}(x) Q_{n,j}(x) b_n(x) c_n(x)^{-1} dx.$$

8.2. *Bounds for $\big|Q_{n,n}(x)\big|$, $\big|Q_{n,n-1}(x)\big|$, and $\big|Q_{n,n}(x)-Q_{n,n-1}(x)\big|$.* In this subsection we prove bounds for $\big|Q_{n,n}(x)\big|$, $\big|Q_{n,n-1}(x)\big|$, and $\big|Q_{n,n}(x)-Q_{n,n-1}(x)\big|$. Let $h_n(x) \triangleq 2nb_n(x)^{-1}$. Then

LEMMA 7. *For all $x \in \big((n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2,(n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2\big)$,*

$$\big|Q_{n,n}(x)\big| \le c_n(x)^{\frac{1}{2}} h_n(x); \tag{13}$$

$$\big|Q_{n,n-1}(x)\big| \le c_n(x)^{\frac{1}{2}} h_n(x); \tag{14}$$

$$\big|Q_{n,n}(x)-Q_{n,n-1}(x)\big| \le (\frac{x}{n})^{\frac{1}{2}} c_n(x)^{\frac{1}{2}} h_n(x). \tag{15}$$

PROOF. We first prove (13). If $Q_{n,n}(x)=0$, then $\big|Q_{n,n}(x)\big| = 0 < c_n(x)^{\frac{1}{2}} h_n(x)$. Otherwise,

$$\begin{aligned} Q_{n,n}^2(x)c_n(x)^{-1} &= \big(1-\frac{\lambda_n+n-x}{\lambda_n}\frac{Q_{n,n-1}(x)}{Q_{n,n}(x)}+\frac{n}{\lambda_n}(\frac{Q_{n,n-1}(x)}{Q_{n,n}(x)})^2\big)^{-1} \\ &\le \sup_{z\in\mathbb{R}}\big((1-\frac{\lambda_n+n-x}{\lambda_n}z+\frac{n}{\lambda_n}z^2)^{-1}\big) = 4\lambda_n n b_n(x)^{-2}, \end{aligned} \tag{16}$$

where the final equality follows from elementary calculus. Taking square roots completes the proof. The proof of (14) follows from a similar argument, and we omit the details. We now prove (15). It is shown in [33] that $Q_{n,n}$ and $Q_{n,n-1}$ do not have any common zeros. Thus first suppose $Q_{n,n}(x)=0$. Then $\big(Q_{n,n}(x)-Q_{n,n-1}(x)\big)^2 c_n(x)^{-1} = \frac{\lambda_n}{n} < 1$. Combining with the fact that $4\lambda_n x b_n(x)^{-2} = 1+(\lambda_n+x-n)^2 b_n(x)^{-2} \ge 1$ completes the proof. The case $Q_{n,n-1}(x)=0$ follows from a similar argument, and we omit the details. Finally, suppose $Q_{n,n}(x)\ne 0$ and $Q_{n,n-1}(x)\ne 0$. Then

$$\begin{aligned} \big(Q_{n,n}(x)-Q_{n,n-1}(x)\big)^2 c_n(x)^{-1} &= \frac{(\frac{Q_{n,n}(x)}{Q_{n,n-1}(x)}-1)^2}{(\frac{Q_{n,n}(x)}{Q_{n,n-1}(x)})^2-\frac{\lambda_n+n-x}{\lambda_n}\frac{Q_{n,n}(x)}{Q_{n,n-1}(x)}+\frac{n}{\lambda_n}} \\ &\le \sup_{z\in\mathbb{R}}\big((z-1)^2(z^2-\frac{\lambda_n+n-x}{\lambda_n}z+\frac{n}{\lambda_n})^{-1}\big). \end{aligned} \tag{17}$$

Let $f(z) \triangleq (z-1)^2(z^2-\frac{\lambda_n+n-x}{\lambda_n}z+\frac{n}{\lambda_n})^{-1}$. It may be easily verified that $f(z)$ is a continuously differentiable rational function of z on $\mathbb{R}$, and the zeros of $\frac{d}{dz}f(z)$ occur at $z=1$ and $z=\frac{\lambda_n-n-x}{\lambda_n-n+x}$. Thus $\sup_{z\in\mathbb{R}} f(z)$ must be one of $f(1), f(\frac{\lambda_n-n-x}{\lambda_n-n+x}), \lim_{z\to-\infty} f(z), \lim_{z\to\infty} f(z)$. It follows from a straightforward computation that $f(1)=0$, $\lim_{z\to-\infty} f(z) = \lim_{z\to\infty} f(z) = 1$, and $f(\frac{\lambda_n-n-x}{\lambda_n-n+x}) = 4\lambda_n x b_n(x)^{-2}$. Combining with (17), and the fact that $4\lambda_n x b_n(x)^{-2} \ge 1$, completes the proof.

□

8.3. *Bounding* $\left|Q_{n,k}(x)\right|$ *and* $\left|Q_{n,k\pm1}(x) - Q_{n,k}(x)\right|$ *for* $k = n \pm O(n^{\frac{1}{2}})$. In this subsection, we bound $\left|Q_{n,k}(x)\right|$ and $\left|Q_{n,k\pm1}(x) - Q_{n,k}(x)\right|$ for $k = n \pm O(n^{\frac{1}{2}})$. Let $s_n(a) \triangleq a(n^{\frac{1}{2}} - a)^{-1}$, $r_n(a,x) \triangleq x(n - an^{\frac{1}{2}})^{-1}$, and $F_n(a,x) \triangleq \exp\Big(\big(1 + s_n(a)\big)\big(a + n^{-\frac{1}{2}}\big)\big(3x^{\frac{1}{2}} + a\big)\Big)$. Then we prove that

THEOREM 11. *For all* $a \geq B > 0$, $k \in [n - an^{\frac{1}{2}}, n + an^{\frac{1}{2}} + 1]$, *and* $x \in \big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2, (n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2\big)$,

$$\left|Q_{n,k}(x)\right| \leq F_n(a,x) c_n(x)^{\frac{1}{2}} h_n(x); \tag{18}$$

$$\left|Q_{n,k+1}(x) - Q_{n,k}(x)\right| \leq r_n(a,x)^{\frac{1}{2}} F_n(a,x) c_n(x)^{\frac{1}{2}} h_n(x). \tag{19}$$

We first bound $\left|Q_{n,k\pm1}(x)\right|$ and $\left|Q_{n,k\pm1}(x) - Q_{n,k}(x)\right|$ in terms of $\left|Q_{n,k}(x)\right|$ and $\left|Q_{n,k}(x) - Q_{n,k\pm1}(x)\right|$. Namely,

LEMMA 8. *For all* $a \geq B > 0$, $k \geq n - an^{\frac{1}{2}}$, $x > 0$, *and* $i \in \{1, -1\}$,

$$\left|Q_{n,k+i}(x)\right| \leq \exp\big(r_n(a,x) + s_n(a)\big)\big(\left|Q_{n,k}(x)\right| + \left|Q_{n,k}(x) - Q_{n,k-i}(x)\right|\big); \tag{20}$$

$$\left|Q_{n,k+i}(x) - Q_{n,k}(x)\right| \leq \exp\big(r_n(a,x) + s_n(a)\big)\big(r_n(a,x)\left|Q_{n,k}(x)\right| + \left|Q_{n,k}(x) - Q_{n,k-i}(x)\right|\big). \tag{21}$$

PROOF. Note that

$$\left|Q_{n,k+i}(x)\right| = \Big|\big(1 - x\min(k,n)^{\frac{i-1}{2}}\lambda_n^{\frac{-i-1}{2}}\big)Q_{n,k}(x) + \big(\frac{\min(k,n)}{\lambda_n}\big)^i\big(Q_{n,k}(x) - Q_{n,k-i}(x)\big)\Big|;$$

$$\left|Q_{n,k+i}(x) - Q_{n,k}(x)\right| = \Big|-\frac{x}{\lambda_n}Q_{n,k}(x) + \big(\frac{\min(k,n)}{\lambda_n}\big)^i\big(Q_{n,k}(x) - Q_{n,k-i}(x)\big)\Big|;$$

Since $\max\big(\left|1 - \frac{x}{\lambda_n}\right|, \left|1 - \frac{x}{\min(k,n)}\right|\big) \leq \exp\big(r_n(a,x)\big)$, $\max\big(\frac{\min(k,n)}{\lambda_n}, \frac{\lambda_n}{\min(k,n)}\big) \leq \exp\big(s_n(a)\big)$, and $\left|\frac{x}{\lambda_n}\right| \leq r_n(a,x)$, the proof then follows from the triangle inequality. □

We now use an induction argument to bound $\left|Q_{n,k}(x)\right|$ and $\left|Q_{n,k}(x) - Q_{n,k\pm1}(x)\right|$ for $k = n \pm O(n^{\frac{1}{2}})$. Let $G_n(a,x) \triangleq \exp\big(r_n(a,x) + r_n(a,x)^{\frac{1}{2}} + s_n(a)\big)$. Then we demonstrate that

LEMMA 9. *For all* $a \geq B > 0$, $k \geq n - an^{\frac{1}{2}}$, *and* $x > 0$,

$$\left|Q_{n,k}(x)\right| \leq G_n(a,x)^{|k-n|} c_n(x)^{\frac{1}{2}} h_n(x); \tag{22}$$

$$\left|Q_{n,k}(x) - Q_{n,k+1-2I(k\geq n)}(x)\right| \leq r_n(a,x)^{\frac{1}{2}} G_n(a,x)^{|k-n|} c_n(x)^{\frac{1}{2}} h_n(x). \tag{23}$$

PROOF. We first treat the case $k \geq n$. We proceed by induction on (22) and (23) simultaneously. The base case $k = n$ follows immediately from Lemma 7. Now, suppose the induction is true for some $k \geq n$. Then by Lemma 8 and the induction hypothesis, $\left|Q_{n,k+1}(x)\right|$ is at most

$$\begin{aligned}
&\exp\big(r_n(a,x) + s_n(a)\big)\big(G_n(a,x)^{k-n} c_n(x)^{\frac{1}{2}} h_n(x) + r_n(a,x)^{\frac{1}{2}} G_n(a,x)^{k-n} c_n(x)^{\frac{1}{2}} h_n(x)\big) \\
&= \exp\big(r_n(a,x) + s_n(a)\big) G_n(a,x)^{k-n} c_n(x)^{\frac{1}{2}} h_n(x)\big(1 + r_n(a,x)^{\frac{1}{2}}\big) \\
&\leq G_n(a,x)^{k+1-n} c_n(x)^{\frac{1}{2}} h_n(x).
\end{aligned}$$

Similarly, by Lemma 8 and the induction hypothesis, $\big|Q_{n,k+1}(x) - Q_{n,k}(x)\big|$ is at most

$$\begin{aligned}
&\exp\big(r_n(a,x)+s_n(a)\big)\big(r_n(a,x)G_n(a,x)^{k-n}c_n(x)^{\frac12}h_n(x) + r_n(a,x)^{\frac12}G_n(a,x)^{k-n}c_n(x)^{\frac12}h_n(x)\big)\\
&= \exp\big(r_n(a,x)+s_n(a)\big)G_n(a,x)^{k-n}c_n(x)^{\frac12}h_n(x)\big(1+r_n(a,x)^{\frac12}\big)r_n(a,x)^{\frac12}\\
&\le r_n(a,x)^{\frac12}G_n(a,x)^{k+1-n}c_n(x)^{\frac12}h_n(x).
\end{aligned}$$

This concludes the induction, proving (22) and (23) for the case $k \ge n$.

The proof for the case $k < n$ follows from a similar argument, and we omit the details. □

With Lemma 9 in hand, we now complete the proof of Theorem 11.

Proof of Theorem 11. By Lemma 9, $\big|Q_{n,k}(x)\big|$ is at most

$$\begin{aligned}
&\exp\Big(\big(an^{\frac12}+1\big)\big(r_n(a,x)+r_n(a,x)^{\frac12}+s_n(a)\big)\Big)c_n(x)^{\frac12}h_n(x)\\
&= \exp\Big(\big(a+n^{-\frac12}\big)\big(1+s_n(a)\big)\Big(xn^{-\frac12}+\big(1+s_n(a)\big)^{-\frac12}x^{\frac12}+a\Big)\Big)c_n(x)^{\frac12}h_n(x)\\
&\le \exp\Big(\big(a+n^{-\frac12}\big)\big(1+s_n(a)\big)\big(xn^{-\frac12}+x^{\frac12}+a\big)\Big)c_n(x)^{\frac12}h_n(x). \qquad (24)
\end{aligned}$$

Similarly, $\big|Q_{n,k+1}(x) - Q_{n,k}(x)\big|$ is at most

$$r_n(a,x)^{\frac12}\exp\Big(\big(a+n^{-\frac12}\big)\big(1+s_n(a)\big)\big(xn^{-\frac12}+x^{\frac12}+a\big)\Big)c_n(x)^{\frac12}h_n(x). \qquad (25)$$

Furthermore, note that $xn^{-\frac12} < 2x^{\frac12}$ for $x \in \big(0,(n^{\frac12}+\lambda_n^{\frac12})^2\big)$, since

$$\frac{xn^{-\frac12}}{2x^{\frac12}} = \frac{x^{\frac12}}{2n^{\frac12}} < \frac{n^{\frac12}+\lambda_n^{\frac12}}{2n^{\frac12}} < 1.$$

Combining with (24) and (25) completes the proof. □

8.4. *Proof of Theorem 5.* In this subsection we complete the proof of Theorem 5. We begin by deriving a variant of the KM representation for $P^n_{i,\le j}(t)$, as opposed to $P^n_{i,j}(t)$, that does not simply sum over all $j+1$ states $\le j$, but instead relies on a 'probability flow' interpretation using the Chapman-Kolmogorov (CK) differential equations.

Lemma 10. *If $\frac{\lambda_n}{n} \ge \rho_n^*$, then for all $i,j,t \ge 0$, $|P^n_{i,\le j}(t) - P^n_{\le j}(\infty)|$ is at most*

$$(2\pi)^{-1}g_n(j)n^{-1}\int_{(n^{\frac12}-\lambda_n^{\frac12})^2}^{(n^{\frac12}+\lambda_n^{\frac12})^2}\exp(-xt)x^{-1}\big|Q_{n,i}(x)\big|\big|Q_{n,j+1}(x)-Q_{n,j}(x)\big|b_n(x)c_n(x)^{-1}dx.$$

Proof. The CK differential equations imply that $\frac{d}{dt}P^n_{i,\le j}(t) = \min(j+1,n)P^n_{i,j+1}(t) - \lambda_n P^n_{i,j}(t)$. Thus for all $i,j,t \ge 0$,

$$\big|P^n_{i,\le j}(t) - P^n_{\le j}(\infty)\big| = \Big|\int_t^\infty \big(\min(j+1,n)P^n_{i,j+1}(s) - \lambda_n P^n_{i,j}(s)\big)ds\Big|. \qquad (26)$$

By detailed balance,

$$\min(j+1,n)P^n_{j+1}(\infty) - \lambda_n P^n_j(\infty) = 0.$$

Similarly,

$$\min(j+1,n)g_n(j+1) = \lambda_n g_n(j).$$

It thus follows from Theorem 10 that the r.h.s. of (26) equals

$$\Big|\int_t^\infty \Big((2\pi)^{-1}g_n(j)n^{-1}\int_{(n^{\frac12}-\lambda_n^{\frac12})^2}^{(n^{\frac12}+\lambda_n^{\frac12})^2}\exp(-xs)Q_{n,i}(x)\big(Q_{n,j+1}(x)-Q_{n,j}(x)\big)b_n(x)c_n(x)^{-1}dx\Big)ds\Big|$$
$$\le \int_t^\infty \Big((2\pi)^{-1}g_n(j)n^{-1}\int_{(n^{\frac12}-\lambda_n^{\frac12})^2}^{(n^{\frac12}+\lambda_n^{\frac12})^2}\exp(-xs)\big|Q_{n,i}(x)\big|\big|Q_{n,j+1}(x)-Q_{n,j}(x)\big|b_n(x)c_n(x)^{-1}dx\Big)ds$$
$$= (2\pi)^{-1}g_n(j)n^{-1}\int_{(n^{\frac12}-\lambda_n^{\frac12})^2}^{(n^{\frac12}+\lambda_n^{\frac12})^2}\exp(-xt)x^{-1}\big|Q_{n,i}(x)\big|\big|Q_{n,j+1}(x)-Q_{n,j}(x)\big|b_n(x)c_n(x)^{-1}dx,$$

where the final equality follows from Tonelli's Theorem. Combining the above completes the proof. □

We now prove bounds on a special type of integral that arises in the analysis of $P^n_{i,j}(t)-P^n_j(\infty)$.

LEMMA 11. *For all $B,a>0$ there exists $N_{B,a},C_{B,a}<\infty$, depending only on B and a, s.t. for all $n\ge N_{B,a}$ and $t\ge 1$,*

$$\int_{(n^{\frac12}-\lambda_n^{\frac12})^2}^{(n^{\frac12}+\lambda_n^{\frac12})^2}\exp(-xt)F_n(a,x)^2b_n(x)^{-1}dx \le (1+C_{B,a}n^{-\frac12})(\frac{\pi}{t\lambda_n})^{\frac12}\exp\big(20a^2+3aB-\frac{B^2}{4}t\big). \tag{27}$$

PROOF. The proof is deferred to the appendix. □

Finally, we complete the proof of Theorem 5.

PROOF OF THEOREM 5. Suppose $B\in(0,B^*)$, and $a_1,a_2\in\mathbb{R}$. Let $a=\max(B,|a_1|,|a_2|)$, $i=\lceil n+a_1n^{\frac12}\rceil$, and $j=\lceil n+a_2n^{\frac12}\rceil$. We first prove (3). It follows from Theorem 10 and Corollary 1 that for all sufficiently large n and all $t\ge1$, the l.h.s. of (3) is at most

$$(2\pi)^{-1}g_n(j)(\lambda_n n^{\frac12})^{-1}\int_{(n^{\frac12}-\lambda_n^{\frac12})^2}^{(n^{\frac12}+\lambda_n^{\frac12})^2}\exp\big(-xt\big)\big|Q_{n,i}(x)\big|\big|Q_{n,j}(x)\big|c_n(x)^{-1}b_n(x)dx. \tag{28}$$

Applying Theorem 11 to $|Q_{n,i}(x)|$ and $|Q_{n,j}(x)|$ in (28), we find that the l.h.s. of (3) is at most

$$2\pi^{-1}g_n(j)\frac{n^{\frac32}}{\lambda_n}\int_{(n^{\frac12}-\lambda_n^{\frac12})^2}^{(n^{\frac12}+\lambda_n^{\frac12})^2}\exp(-xt)F_n(a,x)^2b_n(x)^{-1}dx. \tag{29}$$

It then follows from Lemma 11 that there exists $N_{B,a},C_{B,a}<\infty$, depending only on B and a, s.t. for all $n\ge N_{B,a}$ and $t\ge1$, the l.h.s. of (3) is at most

$$2(\pi t)^{-\frac12}g_n(j)(\frac{n}{\lambda_n})^{\frac32}(1+C_{B,a}n^{-\frac12})\exp\big(20a^2+3aB-\frac{B^2}{4}t\big). \tag{30}$$

Since $g_n(j)\le(\frac{n}{\lambda_n})^{n-\lambda_n+1}$, combining (30) with a simple Taylor series expansion, and the fact that $B<B^*<2$, completes the proof of (3).

We now prove (4). It follows from Lemma 10 and Corollary 1 that for all sufficiently large n and all $t \geq 1$, the l.h.s. of (4) is at most

$$(31) \qquad (2\pi)^{-1} g_n(j) n^{-1} \int_{(n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2}^{(n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2} \exp(-xt) x^{-1} \big|Q_{n,i}(x)\big| \big|Q_{n,j+1}(x) - Q_{n,j}(x)\big| c_n(x)^{-1} b_n(x) dx.$$

Applying Theorem 11 to $\big|Q_{n,i}(x)\big|$ and $\big|Q_{n,j+1}(x) - Q_{n,j}(x)\big|$, we find that (31) is at most

$$2\pi^{-1} g_n(j) n (n - an^{\frac{1}{2}})^{-\frac{1}{2}} \int_{(n^{\frac{1}{2}}-\lambda_n^{\frac{1}{2}})^2}^{(n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2} \exp(-xt) x^{-\frac{1}{2}} F_n(a,x)^2 b_n(x)^{-1} dx.$$

Since $x^{-\frac{1}{2}} \leq 2B^{-1}$ for $x \geq (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$, the proof of (4) then follows from an argument similar to that used to prove (3), and we omit the details. □

9. Comparison to Other Bounds From the Literature. In this subsection we compare our bounds from Theorem 5 to two other explicit bounds given in the literature [42],[5]. In both cases we will prove that the bounds from the literature (applied to $\big|P^n_{n,\leq n}(t) - P^n_{\leq n}(\infty)\big|$ for $0 < B < B^*$) scale unfavorably in the HW regime. We begin with the bounds given in [42], which prove that for each $B \in (0, B^*)$, there exists N_B s.t. for all $n \geq N_B$ and $t \geq 0$, $\big|P^n_{n,\leq n}(t) - P^n_{\leq n}(\infty)\big|$ is at most

$$(32) \quad 4(n-1)\Big(\sum_{i=1}^{\infty} \big((\frac{n}{n-1})^i - 1\big) P^n_i(\infty) + \big((\frac{n}{n-1})^n - 1\big)\big(1 - 2P^n_n(\infty)\big)\Big) \exp\big(-(Bn^{\frac{1}{2}} - 1)(n-1)^{-1} t\big).$$

Since $\lim_{n\to\infty} \big((Bn^{\frac{1}{2}} - 1)(n-1)^{-1}\big) = 0$, the exponential rate of convergence demonstrated by (32) goes to zero as $n \to \infty$, rendering the bound in [42] ineffective. We now examine the bounds given in [5], which prove that for each $B \in (0, B^*)$, there exists N_B s.t. for all $n \geq N_B$ and $t \geq 0$, $\big|P^n_{n,\leq n}(t) - P^n_{\leq n}(\infty)\big|$ is at most

$$(33) \qquad \big(P^n_n(\infty)^{-1} - 1\big)^{\frac{1}{2}} \exp(-\gamma_n t).$$

It is well-known (see [18]) that $\liminf_{n\to\infty} \big(P^n_n(\infty)^{-1} - 1\big)^{\frac{1}{2}} n^{-\frac{1}{4}} > 0$. It follows that the prefactor appearing in (33) diverges as $n \to \infty$, rendering the bound in [5] ineffective.

It should be noted that although the bounds given in [42] and [5] are ineffective in the HW regime, both bounds hold in much greater generality, and thus remain interesting and applicable in a variety of other settings.

10. Conclusion and Open Questions. In this paper we proved several results about the rate of convergence to stationarity, i.e. the spectral gap, for the $M/M/n$ queue in the HW regime. We identified the limiting rate of convergence to steady-state, and proved that an asymptotic phase transition occurs w.r.t. this rate. In particular, we demonstrated the existence of a constant $B^* \approx 1.85772$ s.t. when a certain excess parameter $B \in (0, B^*)$, the error in the steady-state approximation converges exponentially fast to zero at rate $\frac{B^2}{4}$. For $B > B^*$, the error in the steady-state approximation converges exponentially fast to zero at a different rate, which is the solution to an explicit equation given in terms of the parabolic cylinder functions. This result may be interpreted as an asymptotic version of a phase transition proven to occur for any fixed n by van Doorn in [33]. We also proved explicit bounds on the distance to stationarity for the $M/M/n$

queue in the HW regime, when $B < B^*$. Our bounds scale independently of n in the HW regime, and do not follow from the weak-convergence theory.

This work leaves several interesting directions for future research. There are many open questions related to the interaction between weak convergence and convergence to stationarity. Although our results and those of [28] show that for the $M/M/n$ queue in the HW regime there is an 'interchange of limits' in this regard, namely the limiting rate of convergence equals the rate of convergence of the limit, it is unknown to what extent such an interchange must hold in general. Similarly, it is an open challenge to derive uniform bounds on the distance to steady-state in the HW regime for the case of non-Markovian processing times. It would also be interesting to prove that a phase transition occurs in other related models, and we refer the reader to the recent paper [27] for some results in this direction.

Acknowledgements. The authors would like to thank Hans Blanc, Ton Dieker, Erik van Doorn, Peter Glynn, Johan van Leeuwaarden, Bill Massey, Josh Reed, and Ward Whitt for their helpful discussions and insights. The authors especially thank Johan van Leeuwaarden for his insights into the parabolic cylinder functions, and sharing an early draft of his work. The authors also thank two anonymous referees, who helped to significantly improve the paper's presentation, and gratefully acknowledge support of this work by NSF grant CMMI-0726733, and a Department of Defense NDSEG fellowship.

11. Appendix.

Proof of Lemma 3. By Lemma 1.(i), $f_{n,n-1}(x) > 0$, and thus $\big(z_n(x) - \lambda_n\big)\lambda_n^{-\frac{1}{2}}$ equals

$$\begin{aligned}
&\left(\frac{\sum_{k=0}^{n}\binom{n}{k}\lambda_n^k\prod_{j=1}^{n-k}(j-x)}{\sum_{k=0}^{n-1}\binom{n-1}{k}\lambda_n^k\prod_{j=1}^{n-1-k}(j-x)} - \lambda_n\right)\lambda_n^{-\frac{1}{2}} \\
&\quad = \frac{\sum_{k=0}^{n}\binom{n}{k}\lambda_n^k\prod_{j=1}^{n-k}(j-x) - \lambda_n\sum_{k=0}^{n-1}\binom{n-1}{k}\lambda_n^k\prod_{j=1}^{n-1-k}(j-x)}{\lambda_n^{\frac{1}{2}}\sum_{k=0}^{n-1}\binom{n-1}{k}\lambda_n^k\prod_{j=1}^{n-1-k}(j-x)}.
\end{aligned} \tag{34}$$

Note that the numerator of (34) equals

$$\begin{aligned}
&\prod_{j=1}^{n}(j-x) + \sum_{k=1}^{n}\binom{n}{k}\lambda_n^k\prod_{j=1}^{n-k}(j-x) - \sum_{k=0}^{n-1}\binom{n-1}{(k+1)-1}\lambda_n^{k+1}\prod_{j=1}^{n-(k+1)}(j-x) \\
&\quad = (n-1)!\sum_{k=0}^{n}(n-k)\prod_{j=1}^{n-k}(1-\frac{x}{j})\ \frac{\lambda_n^k}{k!};
\end{aligned} \tag{35}$$

and the denominator of (34) equals

$$\begin{aligned}
\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{n-1}\binom{n-1}{(k+1)-1}\lambda_n^{k+1}\prod_{j=1}^{n-(k+1)}(j-x) &= \lambda_n^{-\frac{1}{2}}\sum_{k=1}^{n}\frac{k}{n}\binom{n}{k}\lambda_n^k\prod_{j=1}^{n-k}(j-x) \\
&= \lambda_n^{-\frac{1}{2}}(n-1)!\sum_{k=0}^{n}k\prod_{j=1}^{n-k}(1-\frac{x}{j})\ \frac{\lambda_n^k}{k!}.
\end{aligned} \tag{36}$$

Plugging (35) and (36) into (34), and multiplying through by $\frac{\exp(-\lambda_n)}{(n-1)!}$, we find that $\big(z_n(x)-\lambda_n\big)\lambda_n^{-\frac{1}{2}}$ equals

$$\frac{\sum_{k=0}^{n}(n-k)\prod_{j=1}^{n-k}(1-\frac{x}{j})\exp(-\lambda_n)\frac{\lambda_n^k}{k!}}{\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{n}k\prod_{j=1}^{n-k}(1-\frac{x}{j})\ \exp(-\lambda_n)\frac{\lambda_n^k}{k!}}. \tag{37}$$

We now demonstrate that for all sufficiently large n, the numerator of (37) is at least

$$\prod_{j=1}^{T}(1-\frac{x}{j})\sum_{k=0}^{n-(T+1)}(n-k)\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\exp(-\lambda_n)\frac{\lambda_n^k}{k!},$$

and at most

$$\prod_{j=1}^{T}(1-\frac{x}{j})\sum_{k=0}^{n-(T+1)}(n-k)\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\exp(-\lambda_n)\frac{\lambda_n^k}{k!}+(T+1)^2.$$

The numerator of (37) equals

$$\prod_{j=1}^{T}(1-\frac{x}{j})\sum_{k=0}^{n-(T+1)}(n-k)\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\exp(-\lambda_n)\frac{\lambda_n^k}{k!}+\sum_{k=n-T}^{n}(n-k)\prod_{j=1}^{n-k}(1-\frac{x}{j})\ \exp(-\lambda_n)\frac{\lambda_n^k}{k!}. \tag{38}$$

The desired lower bound follows from the fact that the second summand in (38) is non-negative. The upper bound follows from the fact that $\exp(-\lambda_n)\frac{\lambda_n^k}{k!}\le n^{-\frac{1}{2}}$ for all $k\ge 0$ by Stirling's inequality, $n-k\le T+1$ for all $k\ge n-T$, and $1-\frac{x}{j}\le 1$ for all $j\ge 1$.

It follows from a similar argument that for all sufficiently large n, the denominator of (37) is at least

$$\lambda_n^{-\frac{1}{2}}\prod_{j=1}^{T}(1-\frac{x}{j})\sum_{k=0}^{n-(T+1)}k\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\ \exp(-\lambda_n)\frac{\lambda_n^k}{k!},$$

and at most

$$\lambda_n^{-\frac{1}{2}}\prod_{j=1}^{T}(1-\frac{x}{j})\sum_{k=0}^{n-(T+1)}k\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\ \exp(-\lambda_n)\frac{\lambda_n^k}{k!}+(T+1)^2,$$

and we omit the details. Combining the above upper and lower bounds for the numerator and denominator of (37), and dividing through by $\prod_{j=1}^{T}(1-\frac{x}{j})$, we find that for all sufficiently large n, $\big(z_n(x)-\lambda_n\big)\lambda_n^{-\frac{1}{2}}$ is at least

$$\frac{\sum_{k=0}^{n-(T+1)}(n-k)\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\exp(-\lambda_n)\frac{\lambda_n^k}{k!}}{\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{n-(T+1)}k\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\ \exp(-\lambda_n)\frac{\lambda_n^k}{k!}+\frac{(T+1)^2}{\prod_{j=1}^{T}(1-\frac{x}{j})}}, \tag{39}$$

and at most

$$\frac{\sum_{k=0}^{n-(T+1)}(n-k)\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\exp(-\lambda_n)\frac{\lambda_n^k}{k!}+\frac{(T+1)^2}{\prod_{j=1}^{T}(1-\frac{x}{j})}}{\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{n-(T+1)}k\prod_{j=T+1}^{n-k}(1-\frac{x}{j})\ \exp(-\lambda_n)\frac{\lambda_n^k}{k!}}. \tag{40}$$

We now simplify the terms in (39) and (40), by proving that for all $n \geq T+1$, and $k \in [0, n-T-1]$,

$$\exp(-2T^{-1})(n-k)^{-x}T^{x} \leq \prod_{j=T+1}^{n-k}(1-\frac{x}{j}) \leq \exp(2T^{-1})(n-k)^{-x}T^{x}. \tag{41}$$

Indeed, since $0 < x < 1$, it follows from a simple Taylor series expansion that for all $j \geq 3$, $1 \leq \frac{\exp(-\frac{x}{j})}{1-\frac{x}{j}} \leq 1 + j^{-2}$. Thus for $j \geq T+1$,

$$\prod_{j=T+1}^{n-k}\frac{\exp(-\frac{x}{j})}{(1-\frac{x}{j})} \;\leq\; \prod_{j=T+1}^{n-k}(1+j^{-2}) \;\leq\; \exp(\int_{T}^{\infty}x^{-2}dx) \;=\; \exp(T^{-1}),$$

and

$$\exp(-T^{-1})\prod_{j=T+1}^{n-k}\exp(-\frac{x}{j}) \leq \prod_{j=T+1}^{n-k}(1-\frac{x}{j}) \leq \prod_{j=T+1}^{n-k}\exp(-\frac{x}{j}). \tag{42}$$

Let $H_k \triangleq \sum_{j=1}^{k}\frac{1}{j}$ denote the kth harmonic number. Then it follows from the results of [41], and the fact that $n-k > T$, that

$$\log(\frac{n-k}{T}) - (2T)^{-1} \leq H_{n-k} - H_T \leq \log(\frac{n-k}{T}) + (2T)^{-1}. \tag{43}$$

Combining (42) and (43) with the fact that $0 < \frac{x}{2T} < (2T)^{-1}$ completes the proof of (41).

It follows from (39), (40), and (41) that for all sufficiently large n, $\big(z_n(x) - \lambda_n\big)\lambda_n^{-\frac{1}{2}}$ is at least

$$\exp(-4T^{-1})\frac{\sum_{k=0}^{n-(T+1)}(n-k)^{1-x}\exp(-\lambda_n)\frac{\lambda_n^k}{k!}}{\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{n-(T+1)}k(n-k)^{-x}\;\exp(-\lambda_n)\frac{\lambda_n^k}{k!} + \frac{(T+1)^2T^{-x}}{\prod_{j=1}^{T}(1-\frac{x}{j})}}, \tag{44}$$

and at most

$$\exp(4T^{-1})\frac{\sum_{k=0}^{n-(T+1)}(n-k)^{1-x}\exp(-\lambda_n)\frac{\lambda_n^k}{k!} + \frac{(T+1)^2T^{-x}}{\prod_{j=1}^{T}(1-\frac{x}{j})}}{\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{n-(T+1)}k(n-k)^{-x}\;\exp(-\lambda_n)\frac{\lambda_n^k}{k!}}. \tag{45}$$

With Inequalities (44) and (45) in hand, we are now in a position to complete the proof of Lemma 3. We begin by proving the lower bound. The term $\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{n-(T+1)}k(n-k)^{-x}\;\exp(-\lambda_n)\frac{\lambda_n^k}{k!}$ appearing in the denominator of (44) is at most

$$\lambda_n^{-\frac{1}{2}}\sum_{k=0}^{\lceil n-T^{-1}n^{\frac{1}{2}}\rceil}k(n-k)^{-x}\exp(-\lambda_n)\frac{\lambda_n^k}{k!} + \lambda_n^{-\frac{1}{2}}\max_{0\leq k\leq n}\big(k\exp(-\lambda_n)\frac{\lambda_n^k}{k!}\big)\sum_{k=\lceil n-T^{-1}n^{\frac{1}{2}}\rceil+1}^{n-(T+1)}(n-k)^{-x}. \tag{46}$$

Recall that for all sufficiently large n, $\sup_{k\geq 0}\big(\exp(-\lambda_n)\frac{\lambda_n^k}{k!}\big) \leq n^{-\frac{1}{2}}$, and $(\frac{n}{\lambda_n})^{\frac{1}{2}} \leq 2$, from which it follows that the second summand of (46) is at most

$$(\frac{n}{\lambda_n})^{\frac{1}{2}}\sum_{k=\lceil n-T^{-1}n^{\frac{1}{2}}\rceil+1}^{n-(T+1)}(n-k)^{-x} \;\leq\; 2\int_{0}^{T^{-1}n^{\frac{1}{2}}}y^{-x}dy \;=\; 2(1-x)^{-1}T^{-(1-x)}n^{\frac{1-x}{2}}.$$

Using the above to upper-bound the denominator of (44), multiplying through by $\lambda_n^{\frac{x-1}{2}}$, and observing that $\lambda_n^{\frac{x-1}{2}} \frac{(T+1)^2 T^{-x}}{\prod_{j=1}^{T}(1-\frac{x}{j})} \leq 2(1-x)^{-1}T^{-(1-x)}$ for all sufficiently large n completes the proof of the lower bound. The upper bound follows from a similar argument, and we omit the details. $\square$

Proof of Lemma 5. We begin by demonstrating that $z_{n,k}$ is a twice-differentiable concave function on $(0,1)$ for all $k \leq n$, which will imply that z_∞, and ultimately $\upsilon(x,-B)$, are concave by taking limits. We proceed by induction on k. The base case $k=1$ is trivial, since $z_{n,1}(x) = \lambda_n + 1 - x$. Now, let us assume the statement is true for $j = 1, \ldots, k-1$ with $k-1 \leq n-1$. It may be easily verified that $f_{n,k}(x) = (\lambda_n + k - x) f_{n,k-1}(x) - \lambda_n(k-1) f_{n,k-2}(x)$. Thus since $z_{n,k-1}$ is strictly positive on $(0,1)$, which follows from Lemma 1.(ii), we find that

$$\frac{d^2}{dx^2} z_{n,k}(x) = \lambda_n(k-1)\Big(-2 z_{n,k-1}(x)^{-3}\big(\frac{d}{dx} z_{n,k-1}(x)\big)^2 + z_{n,k-1}(x)^{-2}\frac{d^2}{dx^2} z_{n,k-1}(x)\Big).$$

Since the induction hypothesis implies that $\frac{d^2}{dx^2} z_{n,k-1}(x) \leq 0$, it follows that $z_{n,k}$ is twice-differentiable on $(0,1)$ and satisfies $\frac{d^2}{dx^2} z_{n,k}(x) \leq 0$ (concavity), proving the induction.

Combining the above with Proposition 3, and the fact that pointwise limits of concave functions are concave, demonstrates that z_∞ is a concave function of x on $(0,1)$ for any fixed $B > 0$. Observing that $\upsilon(x,-B) = z_\infty(x) - B$ completes the proof. $\square$

Proof of Lemma 11. Let $d_n(a) \triangleq 6\big(1 + s_n(a)\big)\big(a + n^{-\frac{1}{2}}\big)$. Then the l.h.s. of (27) equals

$$\int_{(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2}^{2(\lambda_n n)^{\frac{1}{2}}} \exp\big(\frac{a}{3} d_n(a) - xt + d_n(a) x^{\frac{1}{2}}\big)\big(x - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)^{-\frac{1}{2}}\big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - x\big)^{-\frac{1}{2}} dx \tag{47}$$

$$+ \int_{2(\lambda_n n)^{\frac{1}{2}}}^{(n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2} \exp\big(\frac{a}{3} d_n(a) - xt + d_n(a) x^{\frac{1}{2}}\big)\big(x - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)^{-\frac{1}{2}}\big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - x\big)^{-\frac{1}{2}} dx. \tag{48}$$

Let $u_n \triangleq 2(\lambda_n n)^{\frac{1}{2}} - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2$. Since $\big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - x\big)^{-\frac{1}{2}} \leq \big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - 2(\lambda_n n)^{\frac{1}{2}}\big)^{-\frac{1}{2}}$ for $x \in \big((n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2, 2(\lambda_n n)^{\frac{1}{2}}\big)$, (47) is at most

$$\begin{aligned}
&\exp\big(\frac{a}{3} d_n(a)\big)\big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - 2(\lambda_n n)^{\frac{1}{2}}\big)^{-\frac{1}{2}} \int_{(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2}^{2(\lambda_n n)^{\frac{1}{2}}} \exp\big(-xt + d_n(a) x^{\frac{1}{2}}\big)\big(x - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)^{-\frac{1}{2}} dx \\
&= \exp\big(\frac{a}{3} d_n(a)\big)\big(\lambda_n + n\big)^{-\frac{1}{2}} \int_0^{u_n} \exp\Big(-\big(y + (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)t + d_n(a)\big(y + (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)^{\frac{1}{2}}\Big) y^{-\frac{1}{2}} dy \\
&\leq \big(\lambda_n + n\big)^{-\frac{1}{2}} \exp\big(\frac{a}{3} d_n(a) + d_n(a)(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}}) - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 t\big) \int_0^{u_n} \exp\big(-yt + d_n(a) y^{\frac{1}{2}}\big) y^{-\frac{1}{2}} dy,
\end{aligned}$$

where the final inequality follows from the fact that $\big(y + (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)^{\frac{1}{2}} \leq y^{\frac{1}{2}} + n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}}$. It may be easily verified that $-yt + d_n(a) y^{\frac{1}{2}} \leq -\frac{1}{2} yt + d_n(a)^2 (2t)^{-1}$ for all $y > 0$, and $\int_0^\infty \exp(-\frac{1}{2} yt) y^{-\frac{1}{2}} dy = (\frac{2\pi}{t})^{\frac{1}{2}}$, and we conclude that (47) is at most

$$J_1 \triangleq (\frac{\pi}{\lambda_n t})^{\frac{1}{2}} \exp\big(\frac{a}{3} d_n(a) + d_n(a)(n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}}) + d_n(a)^2 (2t)^{-1} - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2 t\big). \tag{49}$$

We now bound (48). Let $S \triangleq \big(2(\lambda_n n)^{\frac{1}{2}}, (n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2\big)$. Since $x \in S$ implies

$$\big(x - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)^{-\frac{1}{2}} \leq \big(2(\lambda_n n)^{\frac{1}{2}} - (n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}})^2\big)^{-\frac{1}{2}} \leq \big(3\lambda_n - n\big)^{-\frac{1}{2}},$$

(48) is at most

$$\begin{aligned} &\exp\big(\frac{a}{3}d_n(a)\big)\big(3\lambda_n - n\big)^{-\frac{1}{2}} \sup_{z\in S} \exp\big(-zt + d_n(a)z^{\frac{1}{2}}\big) \int_{2(\lambda_n n)^{\frac{1}{2}}}^{(n^{\frac{1}{2}}+\lambda_n^{\frac{1}{2}})^2} \big((n^{\frac{1}{2}} + \lambda_n^{\frac{1}{2}})^2 - x\big)^{-\frac{1}{2}} dx \\ = \quad &\exp\big(\frac{a}{3}d_n(a)\big)\big(3\lambda_n - n\big)^{-\frac{1}{2}} \sup_{z\in S} \exp\big(-zt + d_n(a)z^{\frac{1}{2}}\big) \int_0^{\lambda_n+n} y^{-\frac{1}{2}} dy, \end{aligned}$$

which is itself at most

$$J_2 \triangleq 2\big(\frac{\lambda_n + n}{3\lambda_n - n}\big)^{\frac{1}{2}} \exp\big(\frac{a}{3}d_n(a) + d_n(a)^2(2t)^{-1} - (\lambda_n n)^{\frac{1}{2}} t\big), \tag{50}$$

where the final inequality follows from the fact that $-zt + d_n(a)z^{\frac{1}{2}} \leq -\frac{1}{2}zt + d_n(a)^2(2t)^{-1}$, and $\int_0^{\lambda_n+n} y^{-\frac{1}{2}} dy = 2(\lambda_n + n)^{\frac{1}{2}}$. It may be easily verified that there exists $N_{B,a}, C_{B,a} < \infty$, depending only on B and a, s.t. for all $n \geq N_{B,a}$ and $t \geq 1$, one has $d_n(a) \leq 6a + C_{B,a}n^{-\frac{1}{2}}$, $n^{\frac{1}{2}} - \lambda_n^{\frac{1}{2}} \leq \frac{B}{2} + C_{B,a}n^{-\frac{1}{2}}$, and $J_2 \leq n^{-1}J_1$. The lemma then follows by using (49) to bound (47), (50) to bound (48), and applying a simple Taylor series expansion. □

REFERENCES

[1] J.D. Atkinson and T.K. Caughley, *Spectral density of piecewise linear first order systems excited by white noise.*, Int. J. Non-Linear Mechanics (1968), no. 3, 137–156.

[2] R. Bo and R. Wong, *Uniform asymptotic expansion of Charlier polynomials.*, Methods Appl. Anal. **1** (1994), no. 3, 294–313.

[3] H. Buchholz, *The confluent hypergeometric function.*, Springer-Verlag, New York, 1969.

[4] M. Chen, *Exponential l^2 convergence and l^2 spectral gap for Markov processes.*, Acta Mathematica Sinica, New Series **7** (1991), no. 1, 19–37.

[5] ———, *Estimate of exponential convergence rate in total variation by spectral gap.*, Acta Mathematica Sinica **14** (1998), no. 1, 9–16.

[6] D. Chruscinski, *Quantum mechanics of damped systems. II. damping and parabolic potential barrier.*, Journal of Mathematical Physics **45** (2004), no. 3, 841–854.

[7] A. Clarke, *A waiting line process of Markov type.*, The Annals of Mathematical Statistics **27** (1956), no. 2, 452–459.

[8] P. Diaconis, *The cutoff phenomenon in finite Markov chains.*, Proc. Nat. Acad. Sci. (1996), no. 4, 1659–1664.

[9] D. Dominici, *Asymptotic analysis of the Askey-scheme I: from Krawtchouk to Charlier.*, Cent. Eur. J. Math. **5** (2007), no. 2, 280–304.

[10] ———, *Asymptotic analysis of the Krawtchouk polynomials by the WKB method .*, Ramananujan J. **15** (2008), 303–338.

[11] T.M. Dunster, *Uniform asymptotic expansions for Charlier polynomials.*, J. Approx. Theory **112** (2001), no. 1, 93–133.

[12] A. Erdelyi, W. Magnus, F. Oberhettinger, F. Tricomi, and H. Bateman, *Higher transcendental functions, volume II*, McGraw-Hill Book Company, New York, 1953.

[13] A.K. Erlang, *On the rational determination of the number of circuits.*, The Copenhagen Telephone Company, Copenhagen, 1948.

[14] S. Ethier and T. Kurtz, *Markov processes: characterization and convergence*, Wiley and Sons, 2005.

[15] C. Fricker, P. Robert, and D. Tibi, *On the rates of convergence of Erlang's model.*, J. Appl. Probab. **36** (1999), 1167–1184.

[16] N. Gans, G. Koole, and A. Mandelbaum, *Telephone call centers: tutorial, review, and research prospects.*, Manufacturing and Service Operations Management **5** (2003), no. 2, 79–141.

[17] I.S. Gradshteyn and I.M.Ryzhik, *Table of integrals, series, and products.*, Academic Press, New York, 1980.

[18] S. Halfin and W. Whitt, *Heavy-traffic limits for queues with many exponential servers.*, Operations Research **29** (1981), no. 3, 567–588.

[19] E. Hille, *Analytic function theory*, Ginn and Company, 1959.

[20] D. Iglehart, *Limiting diffusion approximations for the many server queue and the repairman problem*, Journal of Applied Probability **2** (1965), 429–441.

[21] D. Jagerman, *Some properties of the Erlang loss function*, Bell System Techn. J. **53** (1974), no. 3, 525–551.

[22] A.J.E.M. Janssen and J.S.H. Van Leeuwaarden, *Back to the roots of the M/D/s queue and the works of Erlang, Crommelin and Pollaczek*, Statistica Neerlandica **62** (2008), no. 3, 299–313.

[23] W. Kang and K. Ramanan, *Asymptotic approximations for the stationary distributions of many-server queues*, Annals of Applied Probability. **22** (2012), no. 2, 477–521.

[24] S. Karlin and J. McGregor, *The differential equation of birth-and-death processes, and the Stieltjes moment problem*, Trans. Amer. math. Soc. **85** (1957), 489–546.

[25] ———, *Many-server queueing processes with Poisson input and exponential service times.*, Pacific J. Math. **8** (1958), 87–118.

[26] M. Kijima, *Evaluation of the decay parameter for some specialized birth-death processes.*, J.Appl.Prob. **29** (1992), 781–791.

[27] C. Knessl and J. Leeuwaarden, *Spectral gap of the Erlang A model in the Halfin-Whitt regime.*, Preprint (2011).

[28] ———, *Transient behavior of the Halfin-Whitt diffusion*, Stochastic Processes and Their Applications **121** (2011), no. 7, 1524–1545.

[29] W. Ledermann and G. Reuter, *Spectral theory for the differential equations of simple birth and death processes.*, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, **246** (1954), no. 914, 321–369.

[30] P. Morse, *Stochastic properties of waiting lines.*, Journal of the Operations Research Society of America **3** (1955), no. 3, 255–261.

[31] F. Pollaczek, *Sur l' application de la theorie des fonctions au calcul de certaines probabilites continues utilisees dans la theorie des reseaux telephoniques*, Annales de l'Institut Henri Poincare **10** (1946), 1–55.

[32] T. Saaty, *Time-dependent solution of the many-server Poisson queue.*, Operations Research **8** (1960), no. 6, 755–772.

[33] E.A. van Doorn, *Stochastic monotonicity and queueing applications of birth death processes.*, Lecture Notes in Statist. **4** (1981).

[34] ———, *On oscillation properties and the interval of orthogonality of orthogonal polynomials.*, SIAM J. Math. Anal. **15** (1984), 1031–1042.

[35] ———, *Conditions for exponential ergodicity and bounds for the decay parameter of a birth-death process.*, Advances in Applied Probability **17** (1985), 514–530.

[36] ———, *Representations and bounds for zeros of orthogonal polynomials and eigenvalues of sign-symmetric tri-diagonal matrices.*, J.Approx.Theory **51** (1987), 254–266.

[37] ———, *Representations for the rate of convergence of birth-death processes*, Theory Probab. Math. Statist. **65** (2002), 37–43.

[38] E.A. van Doorn and A.I. Zeifman, *On the speed of convergence to stationarity of the Erlang loss system*, Queueing Systems **63** (2009), 241–252.

[39] E.A. van Doorn, A.I. Zeifman, and T.L. Panfilova, *Bounds and asymptotics for the rate of convergence of birth-death processes.*, Teor. Veroyatnost. I Primenen. **54** (2009), 18–38.

[40] S. Xie and C. Knessl, *On the transient behaviour of the Erlang loss model: heavy usage asymptotics*, SIAM J. Appl. Math. **53** (1993), no. 2, 555–599.

[41] R. Young, *Euler's constant*, The Mathematical Gazette (1991), no. 75, 187–190.

[42] A.I. Zeifman, *Some estimates of the rate of convergence for birth and death processes.*, Journal of Applied Probability **28** (1991), no. 2, 268–277.

Operations Research Center and Sloan School of Management, MIT, Cambridge, MA, 02139
E-mail: gamarnik@mit.edu
Georgia Institute of Technology, Atlanta, GA, 30332
E-mail: dgoldberg9@isye.gatech.edu