Asymptotic Transient Solutions of Fluid Queues Fed By A Single ON-OFF Source

Y. Shao*, J. Bai*, N. Liu*, K. Wang*, and A. D. Clark*†‡§
*Whiting School of Engineering
Johns Hopkins University, Baltimore, Maryland 21218
Email:adclark@mail.com
†Laboratory for Physical Sciences
University of Maryland, College Park, Maryland 20742
‡Lane Department of Computer Science and Electrical Engineering
West Virginia University, Morgantown, West Virginia 26505

***Abstract*—Transient analysis of fluid queues is an open research area, where matrix analytic methods (MAMs) performed in the transform domain are used to understand time-dependent behavior. However, performing the inversion – either analytically or numerically – is impractical in most cases. This work revisits the work of a fluid-queue driven by a single "ON-OFF" source where asymptotic analysis is performed to produce equivalent representations that depict short and long-time behavior. Numerical simulations illustrate the fidelity of this analysis, where the results can be leveraged to understand transient behavior in different types of fluid models. Furthermore, these results can be adapted to provide better resource estimation of fast-packet switching within various distributed networks.**

***Index Terms*—Queueing Theory, Fluid Models, Stochastic Processes, Laplace-Stieltjes Transform, Transient Analysis, Asymptotic Analysis**

I. Introduction

FLUID queues have served many applications in the computing world including understanding the performance measurements of network switches [1], peer-to-peer file sharing [2], and understanding IEEE 802.11 protocols [3]. Although many advancements have been made to understand stationary behavior, understanding transient behavior is still an open area of research. Recent techniques involve the use of either Fourier-Stieltjes Transforms (FST) and Laplace-Stieltjes Transforms (LST), where matrix analytic methods (MAMs) have been used to understand transient behavior (see [4] and [5]). However, inverting from the LST and FST transform domains to the time domain is a cumbersome process which can be impractical both analytically and computationally.

This paper revisits the special case of a fluid queue fed by single "ON-OFF" source [6], where the contributions are two-fold. First, asymptotic analysis of the solutions is performed that considers both short and long-time behavior. Next, an expansion is proposed that holistically considers these behaviors while also providing a direct correlation between the fluid flow rates and state transition rates. The numerical experiments validate our proposed schema, where the results can be incorporated into understanding various network and scheduling applications such as fast-packet switching, statistical multiplexing [7], and vacation scheduling [8].

The rest of this paper is organized in the following manner. Section II reviews the analysis of the model, where an in-depth analytical assessment of the transient behavior of the fluid queue presented in [6]. Additionally, asymptotic analysis is performed to analyze short and long-time transient behavior. Section IV provides the numerical experiments highlighting the robustness of asymptotic analysis. Section V concludes and presents some avenues of future exploration.

II. Analytical Development

Let the total state space S be comprised of a collection of possible states $S = \{S_1, \ldots, S_M\}$. Here, each state $\{S_m\}$ (for $m = 1, \ldots, M$) is either an "ON" or "OFF" state, where the duration within each state is exponentially distributed with constant transitions between each state. Suppose also that the network has a target link capacity of C PPS[1]. Within each "ON" state traffic is injected into the target link at a rate of r_m PPS, which is determined by the number of input sources in S_m. Thus, the change of in the link buffer content $Q(t)$ is described as

$$\frac{dQ}{dt} = \begin{cases} r_m - C & \text{when} \quad Q(t) > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

Equation (1) models the impact of the server to continue to receive packets. When this happens, they are held in a buffer until they can be processed. Hence, this behavior can be expressed as a Markov chain such that $Q(t)$ represents the buffer content where the buffer rate of change dQ/dt for each "ON" state S_m can be interpreted via the following cases shown in Table I.

The goal is to determine

$$W^m(x,t) = P\{Q(t) \leq x, S(t) = S_m\}, \tag{2}$$

where $x > 0$ is the buffer content. For each state S_m the probabilistic transitions can be represented as follows

[1]PPS = packets per second.

978-1-5386-7693-5/18/$31.00 © 2018 IEEE

TABLE I

Condition	Description
I	$r_m > C$. The buffer content *increases* at a rate of $r_m - C$. The server's ability to receive packets is continually being exhausted. Hence, future packets are held in the buffer $Q(t)$ (where $Q(t) > 0$) until they can be processed.
II	$r_m \leq C$ and $Q(t) > 0$. The buffer content *decreases* at a rate of $r_m - C$. The interpretation is that the server's ability to receive and process packets is still possible.
III	$r_m \leq C$ and $Q(t) = 0$. There is *no change* in the buffer content. This condition implies that there is an efficient network.

$$
\begin{aligned}
W^m(x, t+\Delta t) = (1 - \Delta t \lambda_{mm})\, W^m(x - (r_m - C)\,\Delta t, t) \\
+ \Delta t \sum_{n \neq m} \lambda_{nm} W^n(x - (r_n - C)\,\Delta t, t) \\
+ \mathcal{O}\left((\Delta t)^2\right),
\end{aligned} \tag{3}
$$

where, on the right hand side of (3), the first term describes no transition between states and the second term models the transition between the other states. Noting that

$$
\lambda_{mm} \Delta t = 1 - \Delta t \sum_{n \neq m} \lambda_{mn} \tag{4}
$$

equation (3) can be expressed as

$$
\begin{aligned}
\frac{W^m(x, t+\Delta t) - W^m(x - (r_m - C)\,\Delta t, t)}{\Delta t} = \\
\sum_{n \neq m} \lambda_{mn} W^m(x - (r_m - C)\,\Delta t, t) \\
- \sum_{n \neq m} \lambda_{nm} W^n(x - (r_n - C)\,\Delta t, t) \\
+ \mathcal{O}\left(\Delta t\right).
\end{aligned} \tag{5}
$$

It is observed that the left hand side of (5) can be written as

$$
\begin{aligned}
\frac{W^m(x, t+\Delta t) - W^m(x, t)}{\Delta t} \\
+ (r_m - C) \frac{(W^m(x, t) - W^m(x - \Delta x, t))}{\Delta x},
\end{aligned} \tag{6}
$$

where $\Delta x = (r_m - C)\Delta t$. Therefore, as $\Delta t \to 0$ equation (5) becomes

$$
\frac{\partial W^m}{\partial t} + \phi_m \frac{\partial W^m}{\partial x} = \sum_{n \neq m} \left(\lambda_{nm} W^n - \lambda_{mn} W^m\right), \tag{7}
$$

where

$$
\phi_m = r_m - C. \tag{8}
$$

Initial and Boundary Conditions: The initial conditions are based on the supposition that the buffer content at time $t = 0$ is $x \geq 0$, which occurs at the initial state of the Markov chain S_1 [7]. Hence, the initial conditions are described as

$$
W^m(x, 0) = \begin{cases} f(x) & \text{when} \quad m = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{9}
$$

where $f(x)$ depends on the initial buffer content which can be estimated via examining the initial network traffic. If the buffer content at time $t = 0$ is *almost surely* empty then $f(x) = 1$.

Analogously, the boundary conditions at $x = 0$ become

$$
W^m(0, t) = \begin{cases} q(t) & \text{when} \quad m = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{10}
$$

where $q(t)$ represents the likelihood that the buffer content can be empty. As with equation (9), $q(t)$ is only defined at S_1 due to the assumption that this behavior can be easily examined at the initial state of the Markov environment. On the other hand, if the buffer content is *almost surely* nonempty then $q(t) = 0$. The possibility that multiple "ON" sources can be on at the same time also needs to be examined, which is considered via the following relationship [7]

$$
\lim_{x \to \infty} \sum_{m=1}^{M} W^m(x, t) = 1. \tag{11}
$$

Equation (11) implies that the solutions $W^m(x, t)$ for $m = 1, \ldots, M$ need to be bounded as $x \to \infty$.

A. Special Case of a Single ON-OFF Source

For the special case of a single ON-OFF source the values are taken from two states $S = \{S_1, S_2\}$. Hence, equation (7) reduces to the coupled system

$$
\begin{aligned}
\frac{\partial W^1}{\partial t} &= -\phi_1 \frac{\partial W^1}{\partial x} + \lambda_{21} W^2 - \lambda_{12} W^1 \\
\frac{\partial W^2}{\partial t} &= -\phi_2 \frac{\partial W^2}{\partial x} + \lambda_{12} W^1 - \lambda_{21} W^2,
\end{aligned} \tag{12}
$$

where $W^1(x, t)$, $W^2(x, t)$, ϕ_1, and ϕ_2 are defined by equations (2) and (8), respectively (for $m = 1, 2$). The initial conditions and boundary conditions (at $x = 0$) are defined as

$$
W^1(x, 0) = 1 \qquad \text{and} \qquad W^2(x, 0) = 0. \tag{13}
$$

and

$$
W^1(0, t) = q(t) \qquad \text{and} \qquad W^2(0, t) = 0, \tag{14}
$$

respectively. Additionally, to consider the possibility that the sources at states S_1 and S_2 could be "ON" simultaneously, the following boundary condition must be satisfied

$$
\lim_{x \to \infty} \left\{ W^1(x, t) + W^2(x, t) \right\} = 1. \tag{15}
$$

The formulation of this model, given by equations (12) - (15), is presented in the work of [6] with the application of understanding the transient behavior of fast packet switching. This model is based on the assumptions that $\phi_1 < 0$ during the “OFF” state S_1 (Condition I) and $\phi_2 > 0$ during the “ON” state S_2 (Condition II).

Transient analysis is performed by taking the Laplace transform of equation (12) with respect to t, while also considering the initial conditions (13), resulting in the following differential system

$$\frac{d\hat{\mathbf{W}}\left(x,s\right)}{dx} = \mathbf{A} \cdot \hat{\mathbf{W}}\left(x,s\right) + \mathbf{\Psi}\left(x\right) \tag{16}$$

with the boundary conditions

$$\hat{\mathbf{W}}\left(0,s\right) = \left[\hat{q}(s), 0\right]^T \tag{17}$$

and

$$\lim_{x\to\infty} \left\{\hat{W}^1(x,s) + \hat{W}^2(x,s)\right\} = 1. \tag{18}$$

where $\hat{q}(s)$, $\hat{W}^1(x,s)$, and $\hat{W}^2(x,s)$ are the respective Laplace transforms (with respect to t) of $q(t)$, $W^1(x,t)$, and $W^2(x,t)$. In equation (16), $\hat{\mathbf{W}}(x,s) = \left[\hat{W}^1(x,s), \hat{W}^2(x,s)\right]^T$, $\mathbf{\Psi}\left(x\right) = \left[1/\phi_1, 0\right]^T$, and $\mathbf{A}$ is the 2×2 matrix given by

$$\mathbf{A} = \begin{bmatrix} -\left(\lambda_{12} + s\right)/\phi_1 & \lambda_{21}/\phi_1 \\ \lambda_{12}/\phi_2 & -\left(\lambda_{21} + s\right)/\phi_2 \end{bmatrix}. \tag{19}$$

Therefore, the general solution to the non-homogeneous boundary value problem (BVP) is of the form

$$\hat{\mathbf{W}}(x,s) = \mathbf{\Phi}_0(s) + c_1\mathbf{\Phi}_1(s)e^{w_0(s)x} + c_2\mathbf{\Phi}_2(s)e^{w_1(s)x}, \tag{20}$$

where c_1 and c_2 are constants (to be determined by the boundary conditions), $w_0(s)$ and $w_1(s)$ are the eigenvalues given by

$$w_0(s) = -\frac{\left(\phi_1 + \phi_2\right)s + \lambda_{12}\phi_2 + \lambda_{21}\phi_1}{2\phi_1\phi_2} + \frac{\sqrt{T\left(s\right)}}{2\phi_1\phi_2} \tag{21}$$

and

$$w_1(s) = -\frac{\left(\phi_1 + \phi_2\right)s + \lambda_{12}\phi_2 + \lambda_{21}\phi_1}{2\phi_1\phi_2} - \frac{\sqrt{T\left(s\right)}}{2\phi_1\phi_2}, \tag{22}$$

and $T(s)$ is given by

$$\begin{aligned} T\left(s\right) = {} & \left(\left(\phi_1 + \phi_2\right)s + \lambda_{12}\phi_2 + \lambda_{21}\phi_1\right)^2 \\ & -4\phi_1\phi_2 s\left(s + \lambda_{12} + \lambda_{21}\right). \end{aligned} \tag{23}$$

Additionally, the vectors $\mathbf{\Phi}_j(s)$ (for $j = 0, 1, 2$) are given by

$$\mathbf{\Phi}_0(s) = \psi(s)\left[\left(\lambda_{21} + s\right) + \phi_2 w_0(s)w_1(s), \lambda_{12}\right]^T, \tag{24}$$

$$\mathbf{\Phi}_1(s) = \left[\frac{\left(\lambda_{21} + s\right) + \phi_2 w_0\left(s\right)}{\phi_2}, \frac{\lambda_{12}}{\phi_2}\right]^T, \tag{25}$$

and

$$\mathbf{\Phi}_2(s) = \left[\frac{\left(\lambda_{21} + s\right) + \phi_2 w_1\left(s\right)}{\phi_2}, \frac{\lambda_{12}}{\phi_2}\right]^T, \tag{26}$$

where

$$\psi(s) = \frac{\left(w_0(s) + w_1(s)\right)}{w_0(s)w_1(s)\sqrt{T(s)}}. \tag{27}$$

It is noteworthy to mention that the boundary conditions, given by (17) and (18), need to consider the behavior of $\hat{q}(s)$ while also ensuring that $\hat{W}^1(x,s)$, and $\hat{W}^2(x,s)$ are bounded solutions as $x \to \infty$. Thus, the following relationship must hold

$$\hat{q}(s) = -\frac{1}{\phi_1 w_1\left(s\right)}. \tag{28}$$

Hence, the solutions to equations (16) - (18) are given as

$$\begin{aligned} \hat{W}^1(x,s) = {} & \left(\frac{1}{\lambda_{12} + \lambda_{21}}\right)\left[\left(\frac{\lambda_{21}}{s}\right) + \left(\frac{\lambda_{12}}{s + \lambda_{12} + \lambda_{21}}\right)\right] \\ & - \left(\frac{s + \lambda_{21} + \phi_2 w_0\left(s\right)}{s\left(s + \lambda_{12} + \lambda_{21}\right)}\right)e^{w_0(s)x} \end{aligned} \tag{29}$$

and

$$\begin{aligned} \hat{W}^2(x,s) = {} & \left(\frac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}\right)\left[\frac{1}{s} - \frac{1}{s + \lambda_{12} + \lambda_{21}}\right] \\ & \times \left(1 - e^{w_0(s)x}\right). \end{aligned} \tag{30}$$

Considerable algebraic manipulation and inversion yields the following solutions for $W^1(x,t)$ and $W^2(x,t)$

$$W^1\left(x,t\right) = \begin{cases} \dfrac{\lambda_{21} + \lambda_{12}e^{-\left(\lambda_{12} + \lambda_{21}\right)t}}{\lambda_{12} + \lambda_{21}}, \quad 0 < t < x/\phi_2 \\ \dfrac{\lambda_{21} + \lambda_{12}e^{-\left(\lambda_{12} + \lambda_{21}\right)t}}{\lambda_{12} + \lambda_{21}} - \left(\dfrac{1}{\lambda_{21} + \lambda_{12}}\right) \times \\ e^{-\frac{\lambda_{21}}{\phi_2}x}\int_0^t f_1(t - v, x)h\left(v,x\right)dv, \quad t > x/\phi_2 \end{cases} \tag{31}$$

and

$$W^2\left(x,t\right) = \begin{cases} \dfrac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}\left(1 - e^{-\left(\lambda_{12} + \lambda_{21}\right)t}\right), \; 0 < t < x/\phi_2 \\ \dfrac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}\left(1 - e^{-\left(\lambda_{12} + \lambda_{21}\right)t}\right) - \left(\dfrac{\lambda_{12}}{\lambda_{12} + \lambda_{21}}\right) \\ \times e^{-\frac{\lambda_{21}}{\phi_2}x}\int_0^t f_1(t - v, x)g\left(v,x\right)dv, \; t > x/\phi_2 \end{cases} \tag{32}$$

where

$$f_1(t,x) = 1 - e^{-(\lambda_{12}+\lambda_{21})\left(t-\frac{x}{\phi_2}\right)}, \tag{33}$$

$$h(t,x) = \frac{\lambda_{12}\lambda_{21}\phi_2}{(\phi_2-\phi_1)(\lambda_{12}+\lambda_{21})} e^{\left(-\frac{\lambda_{12}\phi_2-\lambda_{21}\phi_1}{\phi_2-\phi_1}t\right)} \times \left\{ I_0(\rho(x,t)) - \frac{1}{\kappa^2(x,t)} I_2(\rho(x,t)) \right\}, \tag{34}$$

$$g(t,x) = \delta(t) + \frac{\sqrt{-\lambda_{12}\lambda_{21}\phi_2\phi_1}}{\phi_2-\phi_1} e^{\left(-\frac{\lambda_{12}\phi_2-\lambda_{21}\phi_1}{\phi_2-\phi_1}t\right)} \times \left(\kappa(x,t) - \frac{1}{\kappa(x,t)}\right) I_1(\rho(x,t)), \tag{35}$$

$$\rho(x,t) = \frac{2t\sqrt{-\lambda_{12}\lambda_{21}\phi_1\phi_2}}{\phi_2-\phi_1}\kappa(x,t), \tag{36}$$

and

$$\kappa(x,t) = \sqrt{\left(1 - \left(\frac{1}{\phi_1} - \frac{1}{\phi_2}\right)\frac{x}{t}\right)}. \tag{37}$$

Additionally, I_n (where $n = 0, 1, 2$), in equations (34) and (35), are the modified Bessel functions of the first kind.

III. Asymptotic Expansions of Solutions

The solutions, given by equations (31) and (32), describe the transient behavior for the case of a single "ON-OFF" source; however, their usefulness is limiting due to the existence of Bessel functions in the integral expressions. To overcome these restrictions, asymptotic analysis is introduced to provide vital approximations to account for when $t \to 0$ and $t \to \infty$. Next, a comprehensive expansion is proposed in this section that fuses both series representations to provide effective approximations throughout the entire time domain.

A. Short-Time Behavior

For short-time behavior $\rho = \rho(x,t) \to 0$ as $t \to 0$. Thus, $I_n(\rho)$ can be expressed in terms of the power series expansion

$$I_n^P(\rho) = \left(\frac{\rho}{2}\right)^n \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)\Gamma(n+1+k)} \left(\frac{\rho}{2}\right)^{2k}. \tag{38}$$

Consequently, the integral expression in equation (31) becomes

$$e^{\frac{-\lambda_{21}}{\phi_2}x} \int_0^t f_1(t-v,x)h(v,x)\,dv = e^{\frac{-\lambda_{21}}{\phi_2}x} \int_0^t h(v,x)dv - e^{\left\{-(\lambda_{12}+\lambda_{21})t+\frac{\lambda_{12}}{\phi_2}x\right\}} \int_0^t e^{(\lambda_{12}+\lambda_{21})v} h(v,x)dv, \tag{39}$$

where

$$\int_0^t h(v,x)dv = \frac{\lambda_{12}\lambda_{21}\phi_2}{(\phi_2-\phi_1)(\lambda_{12}+\lambda_{21})} \times \sum_{m=0}^{\infty} \frac{\alpha^{2m}\Omega_m(a,b,a;x,t)}{\Gamma^2(m+1)2^{2m}}, \tag{40}$$

$$\int_0^t e^{(\lambda_{12}+\lambda_{21})v} h(v,x)dv = \frac{\lambda_{12}\lambda_{21}\phi_2}{(\phi_2-\phi_1)(\lambda_{12}+\lambda_{21})} \times \sum_{m=0}^{\infty} \frac{\alpha^{2m}\Omega_m(a,b,\hat{a};x,t)}{\Gamma^2(m+1)2^{2m}}, \tag{41}$$

and

$$\Omega_m(a,b,\nu;x,t) = \sum_{k=0}^{m} \binom{m}{k} \frac{(2bx)^{m-k}(m+k)!}{a^{m+k+1}} \times \left\{ \left(1 - e^{-\nu t}\sum_{l=0}^{m+k}\frac{(\nu t)^l}{l!}\right) - \frac{\alpha^2}{4(m+1)(m+2)} \times \sum_{k=0}^{m} \binom{m}{k} \frac{(2bx)^{m-k}(m+k+2)!}{a^{m+k+3}} \left(1 - e^{-\nu t}\sum_{l=0}^{m+k+2}\frac{(\nu t)^l}{l!}\right) \right\}, \tag{42}$$

and the constants a, $\hat{a}$, b, and α are defined as

$$a = \frac{\lambda_{12}\phi_2 - \lambda_{21}\phi_1}{\phi_2-\phi_1}, \qquad \hat{a} = \frac{\lambda_{12}\phi_1 - \lambda_{21}\phi_2}{\phi_2-\phi_1}, \tag{43}$$

$$b = -\frac{1}{2}\left(\frac{1}{\phi_1} - \frac{1}{\phi_2}\right), \quad \text{and} \quad \alpha^2 = -\frac{4\lambda_{12}\lambda_{21}\phi_1\phi_2}{(\phi_2-\phi_1)^2}, \tag{44}$$

respectively. Similarly, the integral expression in equation (32) becomes

$$e^{-\frac{\lambda_{21}}{\phi_2}x} \int_0^t f_1(t-v,x)g(v,x)\,dv = e^{\frac{-\lambda_{21}}{\phi_2}x} \int_0^t g(v,x)dv - e^{\left\{-(\lambda_{12}+\lambda_{21})t+\frac{\lambda_{12}}{\phi_2}x\right\}} \int_0^t e^{(\lambda_{12}+\lambda_{21})v} g(v,x)dv, \tag{45}$$

where

$$\int_0^t g(v,x)dv = 1 + \frac{\lambda_{12}\lambda_{21}x}{\phi_2-\phi_1} \sum_{m=0}^{\infty} \frac{\alpha^{2m}\chi_m(a,b,a;x,t)}{2^{2m}\Gamma(m+1)\Gamma(m+2)}, \tag{46}$$

$$\int_0^t e^{(\lambda_{12}+\lambda_{21})v} g(v,x)dv = 1 + \frac{\lambda_{12}\lambda_{21}x}{\phi_2-\phi_1} \times \sum_{m=0}^{\infty} \frac{\alpha^{2m}\chi_m(a,b,\hat{a};x,t)}{2^{2m}\Gamma(m+1)\Gamma(m+2)}, \tag{47}$$

$$\chi_m(a,b,\nu;x,t) = \sum_{k=0}^{m} \binom{m}{k} \frac{(2bx)^{m-k}(m+k)!}{a^{m+k+1}} \times \left(1 - e^{-\nu t}\sum_{l=0}^{m+k}\frac{(\nu t)^l}{l!}\right), \tag{48}$$

and the constants a, $\hat{a}$, b, and α are defined by equations (43) and (44), respectively.

B. Long-Time Behavior

For long-time behavior $\rho = \rho(x,t) \to \infty$ as $t \to \infty$. Additionally, $I_n(\rho)$ ($n \in \mathbb{Z}$) can be expressed in the following integral form

$$I_n(\rho) = \frac{1}{\pi}\int_0^{\pi} e^{\rho\cos\theta}\cos(n\theta)\,\mathrm{d}\theta. \tag{49}$$

Applying Laplace's method [9] to (49) yields the following asymptotic representation for $I_0(\rho)$

$$I_0^A(\rho(x,t)) \sim \frac{e^{\rho(x,t)}}{\sqrt{2\pi\rho(x,t)}}. \tag{50}$$

For $n \neq 0$ the asymptotic representation for $I_n(\rho)$ becomes

$$I_n^A(\rho(x,t)) \sim \frac{e^{\rho(x,t)}}{\sqrt{2\pi\rho(x,t)}}\sum_{m=0}^{\infty}\frac{(-1)^m n^{2m}}{(2m)!!\,[\rho(x,t)]^m}. \tag{51}$$

Hence, each integral in equations (39) and (45) becomes

$$\int_0^t h(v,x)dv \sim \frac{\lambda_{12}\lambda_{21}\phi_2}{\sqrt{2\pi\alpha}\,(\phi_2-\phi_1)(\lambda_{12}+\lambda_{21})} \times \left(\int_0^t \frac{e^{-av+\alpha(v^2+2bxv)^{\frac{1}{2}}}}{(v^2+2bxv)^{\frac{1}{4}}}dv - \sum_{m=0}^{\infty}\frac{(-1)^m 2^{2m}}{(2m)!!\alpha^m}\int_0^t \frac{e^{-av+\alpha(v^2+2bxv)^{\frac{1}{2}}}}{v^{\frac{m}{2}+\frac{5}{4}}(v+2bx)^{\frac{m}{2}-\frac{3}{4}}}\right)dv, \tag{52}$$

$$\int_0^t e^{(\lambda_{12}+\lambda_{21})v}h(v,x)dv \sim \frac{\lambda_{12}\lambda_{21}\phi_2}{\sqrt{2\pi\alpha}\,(\phi_2-\phi_1)(\lambda_{12}+\lambda_{21})} \times \left(\int_0^t \frac{e^{-\hat{a}v+\alpha(v^2+2bxv)^{\frac{1}{2}}}}{(v^2+2bxv)^{\frac{1}{4}}}dv - \sum_{m=0}^{\infty}\frac{(-1)^m 2^{2m}}{(2m)!!\alpha^m} \times \int_0^t \frac{e^{-\hat{a}v+\alpha(v^2+2bxv)^{\frac{1}{2}}}}{v^{\frac{m}{2}+\frac{5}{4}}(v+2bx)^{\frac{m}{2}-\frac{3}{4}}}\right)dv, \tag{53}$$

$$\int_0^t g(v,x)dv \sim 1 + \frac{bx\sqrt{\alpha}}{\sqrt{2\pi}}\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m)!!\alpha^m} \times \int_0^t \frac{e^{-av+\alpha(v^2+2bxv)^{\frac{1}{2}}}}{(v^2+2bxv)^{\frac{m}{2}+\frac{3}{4}}}dv, \tag{54}$$

and

$$\int_0^t e^{(\lambda_{12}+\lambda_{21})v}g(v,x)dv \sim 1 + \frac{bx\sqrt{\alpha}}{\sqrt{2\pi}}\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m)!!\alpha^m} \times \int_0^t \frac{e^{-\hat{a}v+\alpha(v^2+2bxv)^{\frac{1}{2}}}}{(v^2+2bxv)^{\frac{m}{2}+\frac{3}{4}}}dv, \tag{55}$$

and the constants a, $\hat{a}$, b, and α are defined by equations (43) and (44), respectively.

C. Proposed Comprehensive Expansion

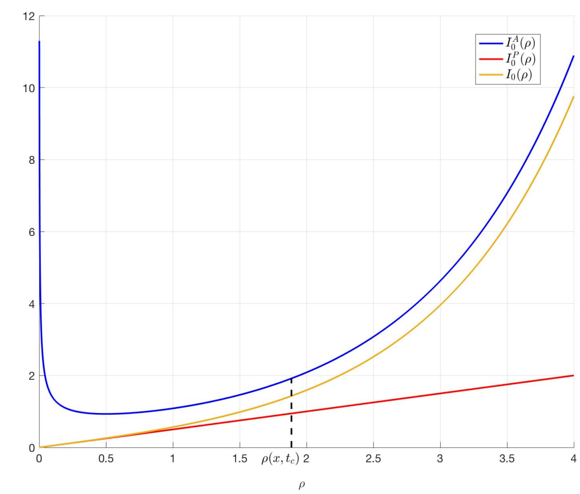

Fig. 1: Comparative behavior between the power and asymptotic representations of $I_0(\rho)$.

The series representations, described in Sections III-A and III-B, are incorporated into equations (31) and (32) to approximate the transient behavior of $W^1(x,t)$ and $W^2(x,t)$ $\forall t$. However, as Figure 1 suggests, the expansions $I_n^P(\rho)$ and $I_n^A(\rho)$ behave similarly to $I_n(\rho)$ for short and long time periods, respectively. Moreover, it is noticed that both series behave equally at a critical value t_c, which is estimated via the following expression

$$\left\| I_n(\rho(x,t_c)) - I_n^P(\rho(x,t_c)) \right\|_2 - \left\| I_n(\rho(x,t_c)) - I_n^A(\rho(x,t_c)) \right\|_2 < \epsilon, \tag{56}$$

where ϵ is the error tolerance (for $n = 0,1,2$) . This critical point t_c serves as a common transition point between the series representations of $I_n^P(\rho)$ and $I_n^A(\rho)$.

IV. Numerical Experimentation & Results

Numerical experiments were conducted to illustrate the fidelity of the proposed expansion, where the mean flow and transition rates were $\phi_1 = -1$, $\phi_2 = 4$, $\lambda_{12} = 1.3$, and $\lambda_{21} = 0.4$ to be consistent with the work of [6]. Additionally,

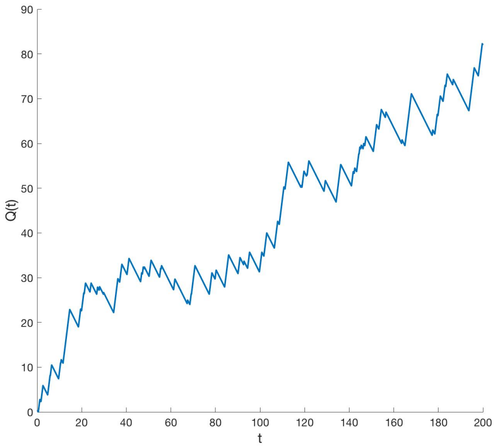

Fig. 2: Output of via Monte Carlo simulation for evaluating the fidelity of the proposed technique.

Monte Carlo simulations were performed to produce variations in the buffer content $Q(t)$ represented by equation (1) in incremental time intervals of $\Delta t = 0.1$ seconds (totaling $1,000$ trials), where an example is shown in Figure 2. As highlighted in Table II, an in-depth comparative analysis was also performed to assess the overall fidelity of the proposed expansion as well as to determine the minimum number of terms needed to characterize the transient response. Here, comparisons are made between the theoretical solutions, the leading and first order approximations, and the results from the Monte Carlo simulation. These comparisons are also used to assess the fidelity of the proposed analysis in characterizing the transient behavior of $F(x,t)$ which is defined as

$$F(x,t) = P\left(Q\left(t\right) \leq x\right) = W^1\left(x,t\right) + W^2\left(x,t\right). \quad (57)$$

The assessment presented in this section is both from visual and numerical perspectives, where absolute differences are computed as

$$\Delta_p = \|x - y\|_p \qquad \text{for} \qquad p = 1, 2, \infty, \quad (58)$$

where p is the type of Euclidean norm used and the variables x and y are the two forms of the solution to be compared. For example, in Scenario I (Table II) the difference would consist of the theoretical solution and the leading order terms in the power series representation. It is important to note that $x = 1$ is used in all evaluations for simplicity.

A. Numerical Results

Estimation of Critical Values: Equation (56) was employed to calculate the critical value t_c, to within four significant digits, where an error tolerance of $\mathcal{O}\left(10^{-5}\right)$ was used. It is interesting to note that determining t_c is based on several factors including the order of the modified Bessel function I_n and the number of terms in the comparison between both power series and asymptotic representations. For example, if one terms is used as a power and asymptotic series for I_0 $t_c = 0.2579$ seconds. However, $t_c = 2.7995$ seconds if more terms are needed in the power series expansion.

TABLE II

Scenario	Description
I	Comparing the differences between the solution and its approximation via the power series representation (first term).
II	Comparing the differences between the solution and its approximation via the combination of power series and asymptotic representation (first term from each series).
III	Comparing the differences between the solution and its approximation via the combination of power series and asymptotic representation (first two terms from each series).
IV	Comparing the differences between the theoretical solution and simulation results.
V	Comparing the differences between the simulation results and the approximation from the combination of power series and asymptotic representation (first term from each series).
VI	Comparing the differences between the simulation results and the approximation from the combination of power series and asymptotic representation (first two terms from each series).

TABLE III

Bessel Function		Estimated Critical Value t_c	
		One Term	Two Terms
I_n	$n = 0$	0.2579	2.7995
	$n = 1$	1.0320	3.7456
	$n = 2$	1.6399	6.1883

TABLE IV

Summary of Average Absolute Differences			
Scenario	Δ_1	Δ_2	Δ_∞
I	0.1806	4.483×10^{-2}	1.210×10^{-4}
II	7.495×10^{-2}	1.798×10^{-2}	4.617×10^{-5}
III	8.898×10^{-6}	5.233×10^{-5}	1.790×10^{-7}
IV	2.922×10^{-2}	1.635×10^{-2}	2.314×10^{-4}
V	0.1039	2.551×10^{-2}	2.313×10^{-4}
VI	1.11×10^{-5}	1.62×10^{-4}	3.12×10^{-4}

As illustrated in Table IV, the closest accuracy to the theoretical and the Monte Carlo simulated solution is four terms, where the first two terms from the power and asymptotic representations are incorporated into an approximation of the transient solutions W^1 and W^2. What is even more impressive is the difference in accuracy when two additional terms are added via comparing Scenarios II, III, V, and VI. This is explicitly shown when examining the behavior of Δ_1, where it is observed that the aforementioned combination expansion is within $\mathcal{O}\left(10^{-5}\right)$ in terms of numerical accuracy.

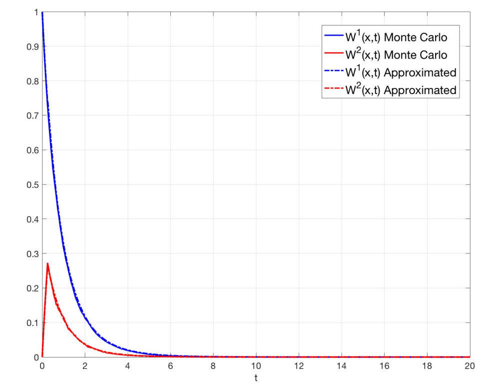

Fig. 3: Comparison of transient responses between $W^1(x,t)$ and $W^2(x,t)$, where $x = 1$, between the four-term approximated solutions and those generated by Monte Carlo simulation.

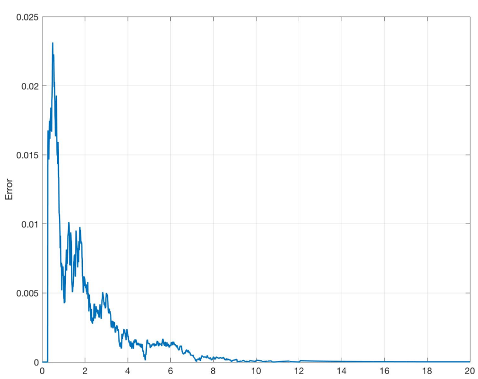

Fig. 4: Absolute error of $F(x,t)$ between the terms generated by the proposed combination expansions of $W^1(x,t)$ and $W^2(x,t)$, and those generated by Monte Carlo simulation where $x = 1$.

Figures 3 and 4 show the robustness of the proposed expansion from a graphical perspective, where Figure 3 highlights the closeness in behavior between the approximations of $W^1(x,t)$ and $W^2(x,t)$ and the Monte Carlo simulations. This is further verified in Figure 4, where the error behavior between the approximated and simulated responses of $F(x,t)$, where this trend continued when the authors investigated this behavior for all six cases.

V. Conclusion

This paper analyzes transient behavior of stochastic fluid queues fed by a single "ON-OFF" source, where asymptotic analysis is presented to provide robust approximations to depict both short and long-time behavior. Furthermore, a comprehensive approximation is proposed that shows a direct correlation between the transient behavior, the buffer flow rates, and state transition rates while also accounting for short and long-time behavior. Moreover, only a few terms are needed to approximate this behavior.

The results of this work have several applications that can benefit researchers not only studying these models, but also their applications. For example, this work can be applied to understand the behavior of fast-packet switching, from theoretical and practical viewpoints, within various distributed networks such as wireless and high-performance computing environments. This work can also be used to understand other types of fluid queues such as multi-dimensional fluid queues [4], those driven by M/M/1 queues [8], and tandem fluid queues [10].

Future explorations of this work include extending and generalizing this approach to provide robust transient solutions of fluid queues fed by multiple "ON-OFF" sources, where the applications include analyzing the effects of traffic congestion [11]. Other explorations of this work include applying these results to quantify the reliability in heterogeneous ultra-dense distributed networks [12].

Acknowledgment

This work is based upon work supported by the U.S. Army Research Laboratory (ARL) and the U.S. Department of Defense (DoD) under award numbers W911NF-13-D-0006 and FA8075-14-D-0002-0007, respectively. The views and conclusions contained in this work are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the ARL or the DoD.

References

[1] D. Anick, D. Mitra, and M. M. Sondhi, "Stochastic theory of a data-handling system with multiple sources," *Bell Labs Technical Journal*, vol. 61, no. 8, pp. 1871–1894, 1982.

[2] R. Gaeta, M. Gribaudo, D. Manini, and M. Sereno, "Analysis of resource transfers in peer-to-peer file sharing applications using fluid models," *Performance Evaluation*, vol. 63, no. 3, pp. 149–174, 2006.

[3] V. Arunachalam, V. Gupta, and S. Dharmaraja, "A fluid queue modulated by two independent birth–death processes," *Computers & Mathematics with Applications*, vol. 60, no. 8, pp. 2433–2444, 2010.

[4] N. G. Bean and M. M. O'Reilly, "A stochastic two-dimensional fluid model," *Stochastic Models*, vol. 29, no. 1, pp. 31–63, 2013.

[5] A. Samuelson, M. M. OReilly, and N. G. Bean, "Generalised reward generator for stochastic fluid models," in *Submitted to the 9th International Conference on Matrix-Analytic Methods in Stochastic Models*, 2016.

[6] P. Parthasarathy and K. V. Vijayashree, "A fluid queue fed by an on-off source," in *IEEE Annual INDICON*. IEEE, 2005, pp. 396–398.

[7] T. Tanaka, O. Hashida, and Y. Takahashi, "Transient analysis of fluid model for atm statistical multiplexer," *Performance Evaluation*, vol. 23, no. 2, pp. 145–162, 1995.

[8] K. Vijayashree and A. Anjuka, "Fluid queue driven by an m/m/1 queue subject to bernoulli-schedule-controlled vacation and vacation interruption," *Advanced in Operations Research*, 2016.

[9] F. Olver, *Asymptotics and special functions*. AK Peters/CRC Press, 1997.

[10] M. M. OReilly and W. Scheinhardt, “Analysis of tandem fluid queues,” in *Submitted to the 9th international conference on matrix-analytic methods in stochastic models*, 2016.

[11] M. Ruiz, F. Coltraro, and L. Velasco, “Cursa-sq: A methodology for service-centric traffic flow analysis,” *Journal of Optical Communications and Networking*, vol. 10, no. 9, pp. 773–784, 2018.

[12] Q. Cui, Y. Gu, W. Ni, X. Zhang, X. Tao, P. Zhang, and R. P. Liu, “Preserving reliability of heterogeneous ultra-dense distributed networks in unlicensed spectrum,” *IEEE Communications Magazine*, vol. 56, no. 6, pp. 72–78, 2018.