Fluid Limit of an Asynchronous Optical Packet Switch with Shared per Link Full Range Wavelength Conversion

Benny Van Houdt
Department of Mathematics and Computer Science, University of Antwerp - IBBT, Belgium
benny.vanhoudt@ua.ac.be

Luca Bortolussi
Department of Mathematics and Computer Science, University of Trieste, Italy
luca@dmi.units.it

ABSTRACT

We consider an asynchronous all optical packet switch (OPS) where each link consists of N wavelength channels and a pool of $C \leq N$ full range tunable wavelength converters. Under the assumption of Poisson arrivals with rate λ (per wavelength channel) and exponential packet lengths, we determine a simple closed-form expression for the limit of the loss probabilities $P_{loss}(N)$ as N tends to infinity (while the load and conversion ratio $\sigma = C/N$ remains fixed). More specifically, for $\sigma \leq \lambda^2$ the loss probability tends to $(\lambda^2 - \sigma)/\lambda(1+\lambda)$, while for $\sigma > \lambda^2$ the loss tends to zero. We also prove an insensitivity result when the exponential packet lengths are replaced by certain classes of phase-type distributions.

A key feature of the dynamical system (i.e., set of ODEs) that describes the limit behavior of this OPS switch, is that its right-hand side is discontinuous. To prove the convergence, we therefore had to generalize some existing result to the setting of piece-wise smooth dynamical systems.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Modeling Techniques; G.3 [**Probability and Statistics**]: Queueing Theory

General Terms

Performance, Theory

Keywords

Fluid limit, optical packet switch, wavelength conversion

1. INTRODUCTION

All optical packet switches (OPS) differ from traditional switches in that they avoid the need to perform any opto-electronic translations, as such they are a good candidate for future ultra-fast communications. As opposed to electronic switches where packets can be buffered (in RAM memory) for an arbitrary amount of time to avoid congestion, there exists no form of optical memory with the same capabilities for OPS switches. The lack of optical buffers may therefore be regarded as one of the main challenges faced when designing an OPS switch.

Congestion in OPS switches, which occurs whenever multiple packets want to make simultaneous use of the same wavelength on an output port, can be addressed by (a combination of) the following three methods: deflection routing, fiber delay line (FDL) buffers and tunable wavelength converters (TWCs). In case of deflection routing, part of the congested traffic is simply routed to another output port (using the same wavelength), causing additional load in the network, unordered arrival of packets at the destination nodes, and extra delays. As such it is not regarded as a viable solution except for low load scenarios [24]. FDL buffers [8, 20] provide some form of buffering, as they allow delays of optical signals by an amount of time chosen from a predefined (finite) set. TWCs on the other hand try to avoid congestion by converting an optical signal from one (congested) wavelength to another (available) wavelength. Switch architectures that rely solely on TWCs may be regarded as the simplest and more popular solutions for contention resolution in OPS networks [2].

An OPS switch using TWCs can operate either in a synchronous or asynchronous fashion, and uses either a shared per node (SPN) or shared per link (SPL) architecture. In a synchronous switch, time is slotted and packets arrive at slot boundaries, while the packet lengths are multiples of the fixed slot length. The synchronous operation may simplify the design of the switching matrix, but requires strict packet synchronization and alignment [1]. Asynchronous networks are often considered as a more natural choice for IP networks due to its variable length data packets [2]. An SPN architecture implies that there is a single pool of TWCs that is shared among all the output ports/links, while in case of SPL, each output port/link has its own set of TWCs. The shared use of the TWCs in an SPN architecture may result in a multiplexing gain, but also adds complexity to the switching matrix.

The TWCs may provide limited- or full-range wavelength conversion. A TWC is called a full-range converter, if it can convert an incoming packet from any incoming wavelength to any output wavelength. If the output wavelength range is limited (typically to a number of adjacent wavelengths), a TWC is referred to as a limited-range TWC. Finally, an OPS switch is said to support either partial or full conversion. Full conversion implies that there are as many

TWCs as there are output wavelengths, while partial conversion means that there are fewer TWCs (which is typically the case for economical reasons). When the switch uses an SPL architecture, where N represents the number of output wavelengths and C the number of TWCs per port/link, the ratio $\sigma = C/N$ is termed the conversion ratio (where $0 < \sigma < 1$).

Our contribution.

In this paper we study the loss probability of an asynchronous switch with shared per link, full-range converters that supports partial wavelength conversion. We derive a closed-form expression for the packet loss probability when the number of wavelengths N becomes large in case of Poisson arrivals with rate $\lambda < 1$ and exponential packet lengths with rate $\mu = 1$. More specifically, we prove that the loss probability tends to zero as N tends to infinity, as long as the conversion ratio $\sigma \geq \lambda^2$. Further, if $\sigma < \lambda^2$, the loss probability converges to $(\lambda^2 - \sigma)/\lambda(1 + \lambda)$ as N tends to infinity. This result is established as follows:

1. We introduce a set of ordinary differential equations (ODEs) that describe the evolution of the fluid limit of the system. One of the main characteristics of this set of ODEs is that the right-hand side is discontinuous (and therefore clearly not Lipschitz as required when relying on the work of Kurtz [11,16]).

2. We reformulate the set of ODEs as a differential inclusion and prove that it has a unique solution for any initial value.

3. We show that all the trajectories of the unique solution are regular, as such the results presented in [6] imply that the sample paths of length t for the finite system converge to the trajectory of the fluid limit.

4. We prove that the unique solution of the differential inclusion has a unique fixed point.

5. The unique fixed point is shown to be a global attractor.

6. We show that the support of the steady state measures of the finite systems converges to the unique fixed point.

In fact, to establish 6. a more general result for piece-wise smooth (PWS) dynamical systems is proven that can be used for other problem settings, as such the paper also contains a methodological contribution.

Further, the same results are also established for phase-type distributed packet lengths, except that the unique fixed point was proven to be a global attractor for a limited range of the conversion ratio σ only. These results indicate that the loss probability becomes insensitive to the packet length distribution as the number of wavelengths N grows.

Related work.

There have been various performance studies of all optical switches (see [1,2,19,22,23]), of which the two most related studies are [1,19]. The switch architecture considered in [1] is identical to the one in this paper, i.e., it considers an asynchronous switch with shared per link full-range converters that supports partial wavelength conversion. A numerical approach to compute the blocking probability was presented for both Poisson and MAP arrivals [18] (which is a class of arrival processes that contains the Markov modulated Poisson process [13] as a special case), while the packet lengths were assumed to be exponential. The sensitivity with respect to the packet length distribution was investigated by means of simulation only as the numerical approach presented does not scale well in case of more general packet lengths, e.g., phase-type [17]. Simulation results for deterministic and order-2 hyperexponential distributions, indicated that the blocking probability is nearly insensitive to the packet length distribution in case of Poisson arrivals.

In this paper we focus on the system with Poisson arrivals only and determine the loss probability in closed-form when the number of wavelengths N tends to infinity. We further prove that the switch performance does indeed become insensitive to the packet length distribution as N tends to infinity in case of order-2 hyperexponential distributions (for most of the σ values).

The study presented in [19] is also strongly related to the current paper, as it analyzes a similar synchronous switch. The arrival process considered is a discrete-time MAP arrival process and the packet lengths are multiples of the slot length and have a general distribution with finite support (which is a subclass of the discrete-time phase-type distributions). The authors also study the switch behavior as the number of wavelengths N becomes large and rely on the framework developed in [7]. More specifically, the limiting system is described by a set of difference equations that were solved numerically and for which a unique fixed point was shown to exist in case the arrival process is a superposition of N Bernoulli processes with parameter p (one for each wavelength on the output port). However, no proof was provided that this fixed point is a global attractor or that the limit of the steady state distributions for the finite systems can be expressed through this fixed point. Under the assumption that both these results can be proven, the authors showed that the loss rate decreases to zero if and only if $\sigma \geq \rho^2(1 - 1/E[L])$, where $\rho = pE[L]$ and $E[L]$ is the mean packet length expressed in slots. If we decrease the slot length to zero, meaning ρ remains fixed and $E[L]$ tends to infinity, the Bernoulli process with parameter p becomes Poisson with rate $\lambda = p$ and the formula reduces to $\sigma \geq \rho^2$, which is in agreement with the result presented in this paper as $E[L]$ was set equal to 1.

Finally, we note that, based on simulation experiments, insensitivity of the loss probability with respect to the packet length distribution for OPS switches with a large number of wavelengths was also conjectured in [23].

2. ANALYTICAL MODEL

This section discusses the analytical model used to assess the loss probability in an asynchronous OPS switch with an SPL architecture with full-range TWCs. Due to the SPL architecture, we can focus on the behavior of a single, tagged output port/link. This link contains N wavelengths and has its own pool of $C < N$ TWCs, where $\sigma = C/N$ is the conversion ratio. We assume that packets destined for the tagged link arrive according to a Poisson process with rate λN. More specifically, for all $w \in \{1, \ldots, N\}$, we assume that the packets destined for the tagged output port that arrive on wavelength w form a Poisson process with rate

$\lambda < 1$. For the packet lengths we consider both exponential as well as phase-type (PH) distributed packet lengths.

An order m phase-type distribution is characterized by a stochastic $1 \times m$ vector α (with entries $\alpha_1, \ldots, \alpha_m$) and an $m \times m$ sub-generator matrix T. Let X_{ij} represent the (i,j)-th element of a matrix X, then $T_{ij} \geq 0$ for $i \neq j$ and $T_{ii} < 0$. Further, let e denote a *column* vector with all its entries equal to one, then $T^* = -Te \geq 0$. Let Z be a phase-type random variable with representation (α, T), then $P[Z > t] = \alpha \exp(Tt)e$ and $E[Z] = \alpha(-T)^{-1}e$. In other words, Z can be regarded as the time until absorption in a continuous time Markov chain with $m+1$ states and rate matrix Q given by

$$Q = \begin{bmatrix} T & T^* \\ 0 & 0 \end{bmatrix},$$

where the initial state is sampled according to the probability vector α. The exponential distribution with mean $1/\mu$ is obtained by setting $m = 1$, $\alpha = 1$ and $T = -\mu$. Notice, any order $m > 1$ PH distribution with $T^* = \mu e$ also represents the exponential distribution with mean $1/\mu$ (in a redundant manner). Throughout the paper we set the mean packet length $1/\mu = \alpha(-T)^{-1}e$ equal to one.

Given the above assumptions on the arrival process and packet lengths, it is clear that we can analyze the switch performance by means of the following $2m$ dimensional Markov chain $\{(W^{(N)}(t), C^{(N)}(t))\}_{t\geq 0}$, where

$$W^{(N)}(t) = (W_1^{(N)}(t), \ldots, W_m^{(N)}(t))$$

and

$$C^{(N)}(t) = (C_1^{(N)}(t), \ldots, C_m^{(N)}(t)).$$

Let $W_i^{(N)}(t)$ denote the number of wavelengths in use on the tagged link that do *not* require a TWC, while the phase of the packet that is being transmitted is i at time t. Similarly, $C_i^{(N)}(t)$ represents the number of wavelengths in use that *make use* of a TWC, while the phase of the packet that is being transmitted is i at time t.

For exponential packet lengths this Markov chain becomes 2 dimensional and the following events can take place. The transmission of a packet may end on one of the wavelengths in use, causing $W^{(N)}(t)$ or $C^{(N)}(t)$ to decrease by one. These events occur at rate $W^{(N)}(t)$ and $C^{(N)}(t)$, respectively. Arrivals occur at each channel and they form a Poisson input process with rate λ. If an arrival does not require a TWC (because its incoming wavelength is still available at the output link), then $W^{(N)}(t)$ is increased by one. This happens at rate $\lambda(N - W^{(N)}(t) - C^{(N)}(t))$. On the other hand, if an arrival happens on a busy wavelength, then it is redirected to a TWC, provided that there are wavelengths available, $W^{(N)}(t) + C^{(N)}(t) < N$, and that there is a TWC available, $C^{(N)}(t) < C = \sigma N$. This kind of arrival increases $C^{(N)}(t)$ by one, and it occurs with rate $\lambda(W^{(N)}(t) + C^{(N)}(t))$, so that the total arrival rate (if channels and TWC are available) is $N\lambda$.

We can determine the steady state of the above 2 dimensional Markov chain by means of efficient numerical techniques [1, 15] for systems with as many as a few hundred wavelengths. Further, even replacing the Poisson arrivals by MAP arrivals as in [1] requires little extra effort. For (non-redundant) phase-type distributions of order 2 (or higher) these numerical techniques are only effective to analyze systems with a small number of wavelengths (which explains why [1] relied on simulation experiments to investigate the sensitivity of the packet length distribution).

Our objective is to derive a closed-form expression for the loss probability when the number of wavelengths N becomes large. Thus, we are interested in the limit behavior of the loss probability $P_{loss}(N)$, when N goes to infinity. To this end we study the scaled process $\{(W^{(N)}(t), C^{(N)}(t))/N\}_{t\geq 0}$ as N goes to infinity. Although it is not hard to show that the set of Markov chains $\{(W^{(N)}(t)/N, C^{(N)}(t)/N)\}_{t\geq 0}$, for $N \geq 1$, form a set of density dependent Markov processes as defined by Kurtz [11], it will become apparent that the right-hand side of the set of ordinary differential equations (ODEs) is in our case discontinuous (hence, not Lipschitz). This implies that we cannot rely on the results in [4, 11] to guarantee convergence and interchangeability of the limits.

3. EXPONENTIAL PACKET LENGTHS

In this section we consider exponential packet lengths with mean 1. Thus, $\{(W^{(N)}(t), C^{(N)}(t))\}_{t\geq 0}$ is a 2 dimensional Markov chain, where $W^{(N)}(t)$ is the number of the wavelengths in use at time t that do not make use of a TWC, while $C^{(N)}(t)$ the number of the wavelengths in use that also occupy a TWC. Let $w(t) = W^{(N)}(t)/N$ be the fraction of wavelengths in use that do not make use of a TWC at time t and $c(t) = C^{(N)}(t)/N$ the fraction of wavelengths in use at time t that uses a TWC (i.e., $c(t)/\sigma$ is the fraction of busy TWCs) and let N go to infinity. Given the possible events discussed in Section 2, we obtain the following set of differential equations in case of exponential packet lengths

$$\begin{aligned} \frac{d}{dt}w(t) &= \lambda(1 - w(t) - c(t)) - w(t), \\ \frac{d}{dt}c(t) &= \lambda(w(t) + c(t))1_{[w(t)+c(t)<1 \text{ and } c(t)<\sigma]} - c(t), \end{aligned} \tag{1}$$

where $1_A = 1$ if A is true and $1_A = 0$ otherwise. If we denote this system of ODEs as $\frac{d}{dt}(w(t), c(t)) = F(w(t), c(t))$, then F is clearly not Lipschitz on $S = \{(w,c) | 0 \leq w, c \leq 1, w + c \leq 1, c \leq \sigma\}$ due to the presence of the indicator function $1_{[w(t)+c(t)<1 \text{ and } c(t)<\sigma]}$. This implies that we cannot rely on the convergence results of Kurtz and Benaïm [4, 11]. Moreover, the Picard-Lindelöf theorem [21] no longer guarantees that a unique solution for $(w(t), c(t))$ exists given an arbitrary $(w(0), c(0)) \in S$.

To deal with the presence of this indicator function, we will replace the above system of discontinuous ODEs by the differential inclusion [9]

$$\frac{d}{dt}(w(t), c(t)) = \bar{F}(w(t), c(t)),$$

where $\bar{F}(w,c)$ is a set-valued function defined as

$$\bar{F}(w,c) = \begin{cases} \{F(w,c)\} & w + c \neq 1 \text{ and } c \neq \sigma, \\ \text{CO}(f_1(w,c), f_2(w,c)) & w + c = 1 \text{ or } c = \sigma, \end{cases} \tag{2}$$

where CO denotes the convex closure of a set and

$$\begin{aligned} f_1(w,c) &= (\lambda - \lambda c - (1+\lambda)w, \lambda w - (1-\lambda)c), \\ f_2(w,c) &= (\lambda - \lambda c - (1+\lambda)w, -c). \end{aligned} \tag{3}$$

Note, the *row* vector $f_1(w(t), c(t))$ and $f_2(w(t), c(t))$ denotes the rate of change of $(w(t), c(t))$ if we replace the indicator function in (1) by one and zero, respectively.

3.1 Solutions, fixed points and global attraction

In this section we prove the following theorem. We start by looking at the solution of the set of ODEs without the indicator function and discuss its influence afterwards.

THEOREM 1. *The differential inclusion* $\frac{d}{dt}(w(t), c(t)) = \bar{F}(w(t), c(t))$ *defined by* (2) *has a unique (Filippov*[1]*) solution* $(w(t), c(t))$ *for any initial value* $(w(0), c(0)) \in S$. *There exists a unique fixed point* π *in* S *given by*

$$\left(\frac{\lambda(1-\sigma)}{1+\lambda}, \sigma\right)$$

for $\sigma \leq \lambda^2$ *and*

$$(\lambda(1-\lambda), \lambda^2)$$

for $\sigma > \lambda^2$. *Further,* π *is a global attractor, i.e., all the trajectories starting from* $(w(0), c(0)) \in S$ *converge towards* π.

No boundaries.

We start by considering the set of ODEs (1) without the indicator function $1_{[w(t)+c(t)<1 \text{ and } c(t)<\sigma]}$. This equation can be written in matrix form as

$$\frac{d}{dt}(w(t), c(t)) = (w(t), c(t)) \underbrace{\begin{bmatrix} -(1+\lambda) & \lambda \\ -\lambda & \lambda - 1 \end{bmatrix}}_{\text{matrix } A} + (\lambda, 0).$$

As a result, the unique solution of the initial value problem defined by the above set of ODEs can be written as

$$(w(t), c(t)) = (\lambda, 0)(-A)^{-1}(I - e^{tA}) + (w(0), c(0))e^{tA}.$$

As -1 is an eigenvalue of A with multiplicity 2, the matrix exponential e^{tA} becomes zero as t goes to infinity. This implies that all the trajectories converge to

$$(\lambda(1-\lambda), \lambda^2) = (\lambda, 0)(-A)^{-1}.$$

Further, using

$$e^{tA} = e^{-t} \begin{bmatrix} 1 - \lambda t & -\lambda t \\ \lambda t & 1 + \lambda t \end{bmatrix},$$

we find that

$$\begin{aligned} w(t) &= \lambda(1-\lambda) - e^{-t}\left(\lambda(1 - \lambda(1+t))\right) \\ &+ e^{-t}\left((1-\lambda t)w(0) - \lambda t c(0)\right), \\ c(t) &= \lambda^2 - e^{-t}\left(\lambda^2(1+t) - \lambda t w(0) - (1+\lambda t)c(0)\right). \end{aligned} \tag{4}$$

[1]A Filippov solution is an absolutely continuous function, therefore almost everywhere differentiable, in contrast with solutions of ODE with Lipschitz continuous right-hand side, which are of class at least $\mathcal{C}^1$.

Boundary behavior.

To examine the behavior of the differential inclusion at the boundaries $\{(w, c)|w + c = 1\}$ and $\{(w, c)|c = \sigma\}$, we use the following methodology outlined in [6,9]. Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be two regions separated by a smooth boundary $\mathcal{H}$. More specifically, $\mathcal{H} = \{(w, c)|h(w, c) = 0\}$, with h a function with continuous second order derivatives such that

$$\mathcal{R}_1 = \{(w, c)|h(w, c) < 0\},$$

and

$$\mathcal{R}_2 = \{(w, c)|h(w, c) > 0\}.$$

Further, the normal vector $n(w, c)$ should be well-defined on $\mathcal{H}$ and is assumed to point into $\mathcal{R}_1$. Denote the limit of $\bar{F}$ in $(w, c) \in \mathcal{H}$ as $f_1(w, c)$ ($f_2(w, c)$) if (w, c) is approached from within $\mathcal{R}_1$ ($\mathcal{R}_2$).

According to [6, 9], the behavior of a solution of the differential inclusion that starts from $(w, c) \in \mathcal{H}$ is determined by the values of the scalar products

$$f_1(w, c)n(w, c) \text{ and } f_2(w, c)n(w, c),$$

where $n(w, c)$ was denoted as a column vector. If both are positive (negative), a *transversal crossing* is said to occur, that is, the solution will move into $\mathcal{R}_1$ ($\mathcal{R}_2$). Thus, if a trajectory hits $\mathcal{H}$ in such a point (w, c) (from within $\mathcal{R}_2$), the trajectory crosses $\mathcal{H}$. If $f_1(w, c)n(w, c)$ (or $f_2(w, c)n(w, c)$) is zero, we have a *tangential crossing* along $f_2(w, c)$ (or $f_1(w, c)$). Finally, if $f_1(w, c)n(w, c) < 0$ and $f_2(w, c)n(w, c) > 0$, we get a *(stable) sliding motion* along $\mathcal{H}$, meaning the solution follows the surface $\mathcal{H}$ (as long as these conditions remain valid). The precise trajectory of the sliding motion is discussed further on. We also note that the differential inclusion is known to have a unique solution if we start in $(w, c) \in \mathcal{H}$ provided that either $f_1(w, c)n(w, c) < 0$ or $f_2(w, c)n(w, c) > 0$ for all $(w, c) \in \mathcal{H}$ [12].

Assume we start from $(w(0), c(0)) \in S = \{(w, c)|0 \leq w, c \leq 1, w + c \leq 1, c \leq \sigma\}$. The boundary ∂S of S is the union of four line segments: $L_1 = \{(w, c)|w = 0, 0 \leq c \leq \sigma\}$, $L_2 = \{(w, c)|0 \leq w \leq 1, c = 0\}$, $L_3 = \{(w, c)|w + c = 1, 0 \leq c \leq \sigma\}$ and $L_4 = \{(w, c)|0 \leq w \leq 1 - \sigma, c = \sigma\}$. F is continuous on L_1 and L_2 and it is easy to see that the drift points towards the interior of S. The line L_3 is part of the smooth surface $\mathcal{H}$ defined by setting $h(w, c) = 1 - w - c$ and f_1 and f_2 as in (3). Clearly, the transposed normal vector $n^T(w, c) = (-1, -1)/\sqrt{2}$ and therefore $f_1(w, c)n(w, c) = (1-\lambda)/\sqrt{2} > 0$ and $f_2(w, c)n(w, c) = (w + c)/\sqrt{2} = 1/\sqrt{2} > 0$ for $(w, c) \in \mathcal{H}$. In other words, if we start on the line segment L_3, we immediately move into S and never hit this boundary again.

The behavior at the boundary $c = \sigma$ is somewhat more involved. First define $\mathcal{H}$ by letting $h(w, c) = \sigma - c$ and f_1 and f_2 as in (3). The transposed normal vector $n^T(w, c) = (0, -1)$, meaning $f_1(w, c)n(w, c) = -\lambda w + (1 - \lambda)\sigma$ and $f_2(w, c)n(w, c) = \sigma > 0$. Notice, the latter expression guarantees the uniqueness of the solution. Thus, if $w \leq (1 - \lambda)\sigma/\lambda$ we get a crossing into $\mathcal{R}_1$, otherwise we get a sliding motion along the line $\mathcal{H} = \{(w, c)|c = \sigma\}$. Notice, if $\sigma > \lambda$, then $w < (1 - \lambda)\sigma/\lambda$ holds for all $w < 1 - \sigma$, meaning there is no sliding motion on L_4 either.

Limit behavior of trajectories.

We will distinguish two cases: $\sigma \leq \lambda^2$ and $\lambda^2 < \sigma \leq 1$. In the first case, we can partition the line segment $L_4 = \{(w, c)|c = \sigma, 0 < w < 1 - \sigma\}$ into two disjoint pieces

$L_{4,1} = \{(w,c)|c = \sigma, 0 < w \leq (1-\lambda)\sigma/\lambda\}$ and $L_{4,2} = \{(w,c)|c = \sigma, (1-\lambda)\sigma/\lambda < w < 1\}$. The above discussion indicated that if we start on $L_{4,1}$, we cross into S. Further, as $f_1(w,c)n(w,c) \geq 0$ on $L_{4,1}$, we can never hit $L_{4,1}$ from the interior of S. Next, recall that the drift close to the three other boundaries points inwards and that all the trajectories starting in the interior of S move towards the fixed point $(\lambda(1-\lambda), \lambda^2)$, which is located outside of the interior of S if $\sigma \leq \lambda^2$. Thus, the line segment $L_{4,2}$ is eventually hit by all the trajectories starting in S.

When $L_{4,2}$ is hit, a sliding motion starts. As explained in [6,9], the evolution of the sliding motion is determined by the differential equation $\frac{d}{dt}(w(t), c(t)) = g(w(t), c(t))$, where $g(w,c)$ is the linear combination of $f_1(w,c)$ and $f_2(w,c)$ such that $g(w,c)$ is tangential to $\mathcal{H}$. In other words, on $L_{4,2}$ we have

$$g(w,c) = (\lambda - \lambda\sigma - (1+\lambda)w, 0).$$

Thus, if we hit $L_{4,2}$ below $w^* = \lambda(1-\sigma)/(1+\lambda)$, we slide upwards, otherwise we slide downwards. Note that (w^*, σ) is part of $L_{4,2}$ if and only if $\sigma < \lambda^2$ (it lies on the boundary between $L_{4,1}$ and $L_{4,2}$ if $\sigma = \lambda^2$). Further, if we start in $(w(0), \sigma) \in L_{4,2}$ then

$$w(t) = \frac{\lambda(1-\sigma)}{1+\lambda}(1 - \exp(-(1+\lambda)t)) + \exp(-(1+\lambda)t)w(0).$$

In conclusion, if $\sigma \leq \lambda^2$ we hit $L_{4,2}$ after a finite amount of time and start an infinite slide towards

$$(\frac{\lambda(1-\sigma)}{1+\lambda}, \sigma),$$

which is therefore a global attractor of the differential inclusion starting from any $(w(0), c(0)) \in S$.

When $\lambda^2 < \sigma \leq 1$, the fixed point $(\lambda(1-\lambda), \lambda^2)$ of the system without the indicator function, lies in the interior of S. As before we can partition the line L_4 into $L_{4,1}$ and $L_{4,2}$, where $L_{4,1}$ cannot be reached from the interior of S. The line $L_{4,2}$ can still be hit from the interior of S and a sliding motion starts when it is hit. Further, the motion is still described by the function g defined before and is downwards as (w^*, σ) lies in the interior of $L_{4,1}$. Therefore, the slide will end (after a finite amount of time) at $((1-\lambda)\sigma/\lambda, \sigma)$ by a tangential crossing into S. Now, if we set $(w(0), c(0)) = ((1-\lambda)\sigma/\lambda, \sigma)$ in the system without boundaries, (4) implies that

$$c(t) = \lambda^2 + e^{-t}(\sigma - \lambda^2)(1+t).$$

Hence, the value of $c(t)$ decreases towards λ^2 and the line $L_{4,2}$ is therefore hit at most once. The point

$$(\lambda(1-\lambda), \lambda^2),$$

is therefore a global attractor of the differential inclusion starting from any $(w(0), c(0)) \in S$.

3.2 Limit results

Our next objective is to prove the following theorem:

THEOREM 2. *For $N \geq 1$, consider the sequence of Markov chains $(W^{(N)}(t), C^{(N)}(t))$, and let π be the unique fixed point defined in Theorem 1, then*

$$\lim_{N\to\infty} \lim_{t\to\infty} \|(W^{(N)}(t)/N, C^{(N)}(t)/N) - \pi\| = 0 \text{ in probability.}$$

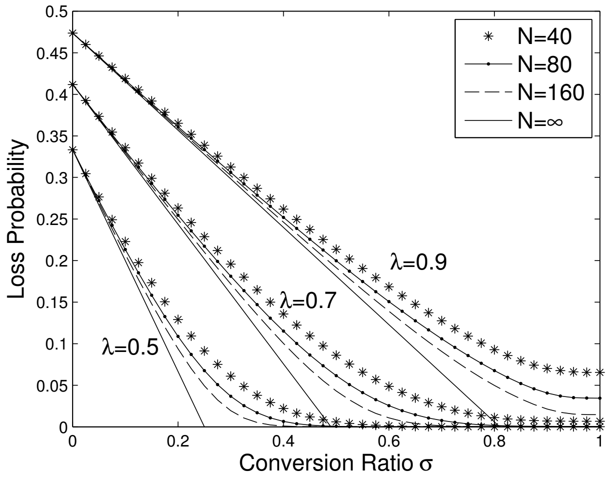

Figure 1: Comparison of loss probability $P_{loss}(N)$ for $N = 40, 80$ and 160 wavelengths per link with the fluid limit

We will prove a more general result for piece-wise smooth (PWS) dynamical systems in Section 5, which will be valid for any PWS that satisfies the four required assumptions H0 to H3 introduced in Section 5. Given the results in the previous section, it is clear that H1 to H3 are satisfied in case of exponential packet lengths and therefore Theorem 5 applies.

As the loss probability can be determined as 1 minus the ratio between the output and input rate of the switch, the previous theorem indicates that the loss probability behaves as follows:

COROLLARY 1. *Under Poisson arrivals with rate $\lambda < 1$ (per wavelength) and exponential packet lengths with mean 1, the limit of the loss probability $P_{loss}(N)$ in an asynchronous OPS switch with an SPL architecture, full range TWCs and a conversion ratio σ equals*

$$\lim_{N\to\infty} P_{loss}(N) = \frac{\lambda - \pi e}{\lambda} = 0,$$

for $\sigma > \lambda^2$ and

$$\lim_{N\to\infty} P_{loss}(N) = \frac{\lambda - \pi e}{\lambda} = \frac{\lambda^2 - \sigma}{\lambda(1+\lambda)},$$

for $0 \leq \sigma \leq \lambda^2$.

In other words, if σ exceeds λ^2, the loss probability decreases to zero as in the Erlang loss model (which corresponds to the case where $\sigma = 1$) and in the limit a fraction $1 - \lambda^2/\sigma$ of the TWCs remains idle. When $\sigma \leq \lambda^2$, all the TWC are occupied and the remaining fraction $(1-\sigma)$ of wavelengths act as a set of independent $M/M/1/1$ queues, meaning each is occupied with probability $\lambda/(1+\lambda)$. If we provision a conversion ratio of $\sigma = \eta\lambda^2$, with $\eta \leq 1$, one finds that the loss rate becomes $(1-\eta)\lambda/(1+\lambda)$. In other words, when a link is equipped with a pool of TWCs with $\sigma = \eta\lambda^2$, the loss probability diminishes by a factor $(1-\eta)$ when compared to a switch without TWCs (i.e., when $\sigma = 0$).

In Figure 1 we compare the loss probability $P_{loss}(N)$ for $N = 40, 80$ and 160 with the fluid limit for three different loads $\lambda = 0.5, 0.7$ and 0.9 (where the numerical approach

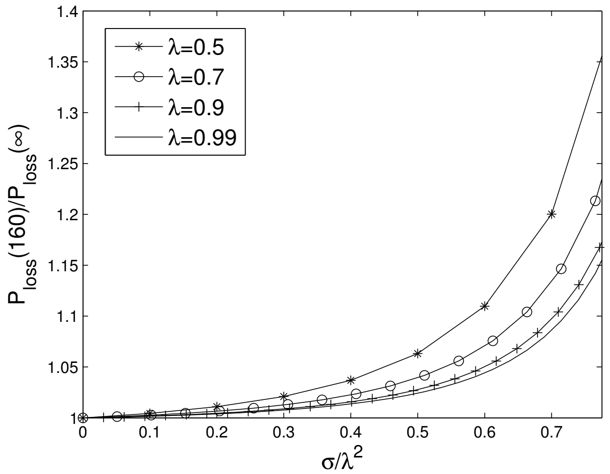

Figure 2: Comparison of ratio $P_{loss}(N)/P_{loss}(\infty)$ for $N = 160$ wavelengths per link with the fluid limit for different loads λ

of [15] was used to obtain the loss for finite N). Commercial DWDM systems with as many as 160 wavelengths are currently on the market (by the Infinera Corporation), though they still rely on opto-electronic translations to store packets in electronic RAM memory. These results confirm the convergence to the fluid limit. We also note that the convergence is slower for σ values close to λ^2. Figure 2 indicates that the accuracy tends to improve with the load λ for σ/λ^2 fixed. As σ/λ^2 approaches one, the ratio $P_{loss}(N)/P_{loss}(\infty)$ tends to infinity as $P_{loss}(\infty)$ tends to zero.

Notice that the slower convergence near λ^2 can be possibly connected to the fact that for $\sigma = \lambda^2$ the limit system undergoes a boundary equilibrium bifurcation [10], in which the limit loss probability (as a function of σ) has a discontinuity in the derivative, absent in the curves for $P_{loss}(N)$, for finite N (cf. Figure 1).

4. PHASE-TYPE PACKET LENGTHS

In this section we consider the same optical switch, but assume that the packet lengths are distributed according to a phase-type distribution with an order m representation (α, T). Recall that the matrix T is an $m \times m$ subgenerator, this implies that there exists a $\tau_1 < 0$, such that τ_1 is an eigenvalue of T and for any other eigenvalue τ_i of T, for $i = 2, \ldots, m$, the real part $Re(\tau_i) \le \tau_1$. We also assume that the mean packet length $1/\mu$, which can be computed as $\alpha(-T)^{-1}e$, is equal to 1. For technical reasons that will become apparent further on, we will assume that $T^* = -Te > 0$ in all its entries. This is for instance the case for hyperexponential distributions, but not for an Erlang-k distribution with $k > 1$. However, it should be clear that any phase-type distribution for which Te is zero in some of its entries can be approximated arbitrarily close by one for which $-Te > 0$. For further use let T_i^* be the i-th entry of the *column* vector T^*.

We denote $w_i(t)$ as the fraction of wavelengths in use that are in phase i and that do not require a converter and $c_i(t)$ as the fraction of wavelengths in use requiring a converter that are in phase i, for $i = 1, \ldots, m$. We also denote $w(t) = (w_1(t), \ldots, w_m(t))$ and $c(t) = (c_1(t), \ldots, c_m(t))$. The set of ODEs given by (1) is now replaced by

$$\begin{aligned} \frac{d}{dt} w_i(t) &= \lambda(1 - w(t)e - c(t)e)\alpha_i + \sum_{j \ne i} w_j(t) T_{ji} \\ &- \sum_{j \ne i} w_i(t) T_{ij} - w_i(t) T_i^*, \\ \frac{d}{dt} c_i(t) &= \lambda(w(t)e + c(t)e)\alpha_i 1_{[w(t)e+c(t)e<1 \text{ and } c(t)e<\sigma]} \\ &+ \sum_{j \ne i} c_j(t) T_{ji} - \sum_{j \ne i} c_i(t) T_{ij} - c_i(t) T_i^*, \end{aligned} \quad (5)$$

for $i = 1, \ldots, m$, where $T^* = -Te$. These equations can be further simplified by noting that $T_{ii} = -\sum_{j \ne i} T_{ij} - T_i^*$.

The set-valued function $\bar{F}(w_1, \ldots, w_m, c_1, \ldots, c_m)$ needed for the differential inclusion is given by

$$\bar{F}(w, c) = \begin{cases} \{F(w, c)\} & we + ce \ne 1 \text{ and } ce \ne \sigma, \\ \mathrm{CO}(f_1(w, c), f_2(w, c)) & we + ce = 1 \text{ or } ce = \sigma, \end{cases} \quad (6)$$

where $w = (w_1, \ldots, w_m)$ and $c = (c_1, \ldots, c_m)$,

$$f_1(w, c) = (\lambda\alpha + w(T - \lambda e\alpha) - c\lambda e\alpha, w\lambda e\alpha + c(T + \lambda e\alpha)) \quad (7)$$

and

$$f_2(w, c) = (\lambda\alpha + w(T - \lambda e\alpha) - c\lambda e\alpha, cT). \quad (8)$$

4.1 Solutions, fixed points and global attraction

We start by proving the theorem below for phase-type distributed packet lengths. The issue of whether the unique fixed point is also a global attractor is addressed in Appendix A, where we prove several results for certain subclasses of the set of phase-type distributions and certain ranges of the conversion ratio σ. A general proof for the complete class of phase-type distributions (with $T^* > 0$) and all possible σ values is currently still lacking.

THEOREM 3. *The differential inclusion $\frac{d}{dt}(w(t), c(t)) = \bar{F}(w(t), c(t))$ defined by (2) for phase-type distributed packet lengths with representation (α, T) and $T^* > 0$, has a unique solution $(w(t), c(t))$ for any initial value $(w(0), c(0)) \in S$. Further, there exists a unique fixed point π in S given by*

$$\left(\frac{\lambda(1-\sigma)}{1+\lambda}\theta, \sigma\theta\right)$$

for $\sigma \le \lambda^2$ and

$$(\lambda(1-\lambda)\theta, \lambda^2\theta)$$

for $\sigma > \lambda^2$, where θ is the stochastic row vector given by $\theta = \alpha(-T)^{-1}$.

No boundaries.

Consider the set of ODEs (5) after removing the indicator function $1_{[w(t)e+c(t)e<1 \text{ and } c(t)e<\sigma]}$. This equation can clearly be written in matrix form as

$$\frac{d}{dt}(w(t), c(t)) = (w(t), c(t)) \underbrace{\begin{bmatrix} T - \lambda e\alpha & \lambda e\alpha \\ -\lambda e\alpha & T + \lambda e\alpha \end{bmatrix}}_{\text{matrix } A} + (\lambda\alpha, 0),$$

where we used the fact that $T_{ii} = -\sum_{j \ne i} T_{ij} - T_i^*$. As a result, the unique solution of the initial value problem defined by the above set of ODEs can be written as

$$(w(t), c(t)) = (\lambda\alpha, 0)(-A)^{-1}(I - e^{tA}) + (w(0), c(0))e^{tA}.$$

As

$$\begin{aligned} \det(A - \tau I) &= \det\left(\begin{bmatrix} T - \tau I & \lambda e\alpha \\ T - \tau I & T + \lambda e\alpha - \tau I \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} T - \tau I & \lambda e\alpha \\ 0 & T - \tau I \end{bmatrix}\right), \end{aligned}$$

we see that if τ is an eigenvalue of T with multiplicity n, then τ is also an eigenvalue of A with multiplicity $2n$. As the real part of all the eigenvalues of T is negative, so is the real part of all the eigenvalues of A and therefore the matrix exponential e^{tA} becomes zero as t goes to infinity. This implies that all the trajectories converge to

$$(\lambda(1 - \lambda)\theta, \lambda^2\theta) = (\lambda\alpha, 0)(-A)^{-1},$$

where $\theta = \alpha(-T)^{-1}$ is a stochastic vector such that $\theta(T + T^*\alpha) = 0$ (θ is stochastic as the mean service time $\alpha(-T)^{-1}e$ equals one). The i-th entry of θ represents the probability that a wavelength is in phase i provided that it is in use, as it is the invariant vector of the rate matrix $T + T^*\alpha$. Note, the above fixed point lies in the interior of S if and only if $\sigma < \lambda^2$ (as in the exponential case).

Boundary behavior.

Assume we start from $(w(0), c(0)) \in S = \{(w, c)|0 \leq w_i, c_i \leq 1, i = 1, \ldots, m, we + ce \leq 1, ce \leq \sigma\}$. The boundary ∂S of S is the union of the surfaces: $\{(w, c)|w_i = 0\}$, $\{(w, c)|c_i = 0\}$, for $i = 1, \ldots, m$, $\{(w, c)|we + ce = 1\}$ and $\{(w, c)|ce = \sigma\}$. F is continuous on the first $2m$ surfaces and it is easy to see that the drift points towards the interior of S. The surface $\{(w, c)|we + ce = 1\}$ can be defined as a smooth surface $\mathcal{H}$ by letting $h(w, c) = 1 - we - ce$ and f_1 and f_2 as in (7) and (8). Clearly, the transposed normal vector $n^T(w, c) = (-e^T, -e^T)/\sqrt{2m}$ and therefore

$$f_1(w, c)n(w, c) = ((w + c)T^* - \lambda)/\sqrt{2m}$$

and

$$f_2(w, c)n(w, c) = (w + c)T^*/\sqrt{2m}$$

for $(w, c) \in \mathcal{H}$. By the assumption that $T^* > 0$, we see that $f_2(w, c)n(w, c) > 0$, which guarantees the uniqueness of the solution if we start on the surface $\mathcal{H}$. The value of $(w + c)T^* - \lambda$ can become negative if at least one of the entries of T^* is less than λ, that is, if the rate of completion in some phase i is less than λ. In this case $\mathcal{H}$ contains a region where a sliding motion occurs, while otherwise a (transversal) crossing occurs in each point $(w, c) \in \mathcal{H}$.

To understand the behavior during the sliding motion on $\mathcal{H}$, we note that any convex combination $\eta f_1(w, c) + (1 - \eta)f_2(w, c)$, with $we + ce = 1$, is of the form $(wT, cT + \eta\lambda\alpha)$. Such a combination is tangential to $\mathcal{H}$ if $wTe + cTe + \eta\lambda\alpha e = 0$, meaning $g(w, c)$ is obtained by setting $0 \leq \eta = (w + c)T^*/\lambda \leq 1$ (during the slide). The sliding motion is therefore described by

$$\frac{d}{dt}(w(t), c(t)) = (w(t), c(t)) \begin{bmatrix} T & T^*\alpha \\ 0 & T + T^*\alpha \end{bmatrix}.$$

Hence, if the sliding motion lasts indefinitely, $(w(t), c(t))$ converges towards $(0, \theta)$. However $\theta T^* = -\theta(Te) = \alpha e = 1$, which indicates that $(0 + \theta)T^* - \lambda = 1 - \lambda > 0$. This means that $(0, \theta)$ is a point of transversal crossing and the sliding motion must therefore end at some point where $(w + c)T^* = \lambda$, which results in a tangential crossing (along $f_2(w, c)$). We should note that we may hit the surface $\{(w, c)|ce = \sigma\}$ before this crossing takes place (if $\sigma < 1$). In the latter case, it is worth noting that no sliding motion occurs on the intersection of these two surfaces due to the assumption that $T^* > 0$ (which guarantees that $we + ce$ immediately becomes less than one when ce becomes σ).

We now proceed with the boundary behavior at $\{(w, c)| ce = \sigma\}$. We set $h(w, c) = \sigma - ce$ and f_1 and f_2 as in (7) and (8). The transposed normal vector $n^T(w, c) = (0, -e^T)/\sqrt{m}$, meaning

$$f_1(w, c)n(w, c) = (-\lambda(we) - \lambda\sigma + cT^*)/\sqrt{m}$$

and

$$f_2(w, c)n(w, c) = cT^*/\sqrt{m} > 0,$$

due to the assumption that $T^* > 0$. Note, the latter expression once again guarantees the uniqueness of the solution. If $cT^* < \lambda(we + \sigma)$ we get a sliding motion, otherwise a (transversal) crossing into $\mathcal{R}_1$ occurs. To understand the behavior of the sliding motion we note that any convex combination $\eta f_1(w, c) + (1 - \eta)f_2(w, c)$, with $ce = \sigma$, is of the form $(\lambda(1 - \sigma)\alpha + w(T - \lambda e\alpha), cT + \eta\lambda(we + \sigma)\alpha)$. Such a combination is tangential to $\mathcal{H}$ if $cTe + \eta\lambda(we + \sigma)\alpha e = 0$, meaning $g(w, c)$ is obtained by setting $0 \leq \eta = cT^*/(\lambda(we + \sigma)) \leq 1$ (during the slide). The sliding motion is therefore described by

$$\frac{d}{dt}(w(t), c(t)) = (w(t), c(t)) \begin{bmatrix} T - \lambda e\alpha & 0 \\ 0 & T + T^*\alpha \end{bmatrix} + (\lambda(1 - \sigma)\alpha, 0).$$

Thus, if the slide lasts indefinitely, $(w(t), c(t))$ converges to

$$(-\lambda(1 - \sigma)\alpha(T - \lambda e\alpha)^{-1}, \sigma\theta).$$

The vector $\alpha(T - \lambda e\alpha)^{-1}$ can be further simplified using the Sherman-Morrison formula, which states that $(A + uv^T)^{-1} = A^{-1} - (A^{-1}uv^TA^{-1})/(1 + v^TA^{-1}u)$, where u is a column vector and v^T a row vector of the appropriate size. That is, by setting $A = T$, $u = (-\lambda e)$ and $v^T = \alpha$ and using the fact that $\alpha(-T)^{-1}e = 1$, we find that $-\alpha(T - \lambda e\alpha)^{-1} = \theta/(1 + \lambda)$. Thus, if the slide lasts indefinitely, $(w(t), c(t))$ converges to

$$(\frac{\lambda(1 - \sigma)}{1 + \lambda}\theta, \sigma\theta).$$

One easily checks that the condition $f_1(w, c)n(w, c) = cT^* - \lambda(we + \sigma) \leq 0$ holds in this fixed point if and only if $\sigma \leq \lambda^2$. In other words, this point lies in the region (or on the boundary) of the area where the sliding motion occurs if and only if $\sigma \leq \lambda^2$ (as in the exponential case).

4.2 Limit results

In this section we identify some cases where the loss probability can be proven to become insensitive to the packet length distribution. Given the results in Sections 4.1 and 5, the final step exists in proving that the unique fixed point π is a global attractor. Various results for different ranges of σ and sub-classes of the class of phase-type distributions are presented in Appendix A. More specifically, the results in Appendix A show that the unique fixed point is a global attractor if $\sigma > \lambda/\min_i T_i^*$ or $\sigma < \lambda^2/(\max_i T_i^*)^2$

for any phase-type distribution with $T^* > 0$. For any hyperexponential distribution, $\sigma > \lambda^2/\min_i T_i^*$ or $\sigma < \lambda^2(1-\lambda)/(\max_i T_i^* - \lambda)$ suffices, while for order-2 hyperexponential distributions having $\sigma > \lambda^2$ is also sufficient.

In this section we will focus on the set of hyperexponential distributions of order-2 only, meaning with probability α_1 the packet length is exponential with parameter μ_1 and with probability $1-\alpha_2$ it is exponential with parameter μ_2. Note, as the mean packet length is one, we may assume without loss of generality that $\mu_1 \le 1 \le \mu_2$. Due to Theorems 3, 8, 10 and 5; as well as the equality $\theta T^* = \alpha(-T)^{-1}(-T)e = 1$, we obtain the following generalization of Corollary 1:

Corollary 2. *Under Poisson arrivals with rate $\lambda < 1$ (per wavelength) and order-2 hyperexponential packet durations with mean 1, the limit of the loss probability $P_{loss}(N)$ in an asynchronous OPS switch with an SPL architecture, full range TWCs and a conversion ratio σ equals*

$$\lim_{N\to\infty} P_{loss}(N) = \frac{\lambda - \pi\begin{pmatrix}T^*\\T^*\end{pmatrix}}{\lambda} = 0,$$

for $\sigma > \lambda^2$ and

$$\lim_{N\to\infty} P_{loss}(N) = \frac{\lambda - \pi\begin{pmatrix}T^*\\T^*\end{pmatrix}}{\lambda} = \frac{\lambda^2 - \sigma}{\lambda(1+\lambda)},$$

for $0 \le \sigma < \lambda^2(1-\lambda)/(\mu_2 - \lambda)$.

Note, as $\mu_2 > 1$ for non-exponential distributions, insensitivity remains an open issue for $\sigma \in [\lambda^2(1-\lambda)/(\mu_2-\lambda), \lambda^2]$. Extensive numerical experiments seem to indicate that π is also a global attractor when σ is part of this region, but we did not manage to come up with a formal proof thus far.

Hyperexponential distributions of order 2 are often used when fitting the first two moments of a distribution as they can match both the mean $1/\mu$ and squared coefficient of variation (SCV) for any $SCV \ge 1$. The set of order-2 hyperexponential distributions was also used in [1] to investigate the impact of the packet length distribution on the loss probability. Simulation experiments in case of Poisson arrivals indicated that for general σ the loss probability of an OPS with $N = 32$ wavelengths per link *is* sensitive to the packet length distribution as opposed to the extreme cases $\sigma = 0$ or 1. In other words, the loss probability for non-exponential packet lengths deviates from the one with exponential packet lengths especially under low load and a moderate number of TWCs. However, the difference is not substantial and the authors state that *one can still make use of exponential packet length distributions for approximating the behavior of non-exponential packet lengths.* Corollary 2 shows that these simulation results are not a coincidence. The system may not be insensitive for finite N, but it becomes insensitive as N tends to infinity.

5. STEADY STATE LIMIT OF PIECE-WISE SMOOTH DYNAMICAL SYSTEMS

The purpose of this section is to prove a theorem for the steady state limit behavior of a sequence $X^{(N)}(t)$ of CTMCs converging to the solution of a piecewise smooth dynamical system (PWS), for any finite time horizon. A PWS is characterized by an ODE of the form $\frac{d}{dt}x(t) = F(x(t))$ with $F : E \to \mathbb{R}^n$ and $E \subset \mathbb{R}^n$. Further, there exist a finite number of domains $\mathcal{R}_i$, $i = 1, \ldots, s$, such that F can be extended to a smooth (or at least Lipschitz) function on the closure $\bar{\mathcal{R}}_i$ and $\bar{E} \subseteq \cup \bar{\mathcal{R}}_i$. Note, F may be discontinuous on the union of the boundaries of $\mathcal{R}_i$.

Similarly to the ODE limit case, we will prove that, if the PWS has a unique globally attracting equilibrium, then the steady state behavior of the sequence of CTMCs will be a point mass concentrated in such an equilibrium. In case the sequence of CTMC models a constant population in terms of the occupancy measure, then the theorem will provide the equilibrium distribution of each CTMC being in each state for large N. In order to prove this theorem, we need to introduce a few concepts and some additional regularity hypothesis on the PWS.

We now fix the notation and introduce the main hypothesis on the sequence of CTMCs.

H0: $X^{(N)}(t)$ is a sequence of CTMC on normalized population counts, i.e., with increments of order $\Theta(\frac{1}{N})$ and exit rate bounded by a $\Theta(N)$ constant. If the total population remains equal to N for all $t \ge 0$, then $X^{(N)}(t)$ is the occupancy measure. Furthermore, $X^{(N)}(t)$ has density dependent rates, so that the drift $F^{(N)}(x)$ is independent of N and equal to $F(x)$.

Let $\mathcal{S}^{(N)}$ be the state space of the N-th CTMC, and let $E \subset \mathbb{R}^n$ be the closure of the union of all domains $\mathcal{S}^{(N)}$: $E = cl(\bigcup_{N\in\mathbb{N}} \mathcal{S}^{(N)})$. Let the discontinuous vector field F be such that no trajectory starting in E can ever leave E.

H1: The PWS $\frac{dx(t)}{dt} = F(x(t))$ has a unique solution for any initial point $x_0 \in E$.

The hypothesis H1 implies that solutions of the PWS are defined in E for all $t \ge 0$. This property, combined with uniqueness, makes it possible to define a notion of (*forward*) flow also for the PWS: let $\phi_t(x)$, $t \ge 0$, be the point reached at time t by the PWS starting at x at time 0. Furthermore, we can show the following:

Lemma 1. *If H0 and H1 holds, then $\phi_t(x) : E \to E$ is continuous for every $t > 0$.*

Proof. The proof relies on properties of a generalization of the notion of flow for differential inclusions [3], of which PWS are a proper subclass. Let $F : E \to \mathcal{P}(E)$ be a set valued map, where $\mathcal{P}(E)$ is the power set of E, associating a convex and compact set $F(x)$ of E with any point $x \in E$. Furthermore, assume that F is upper semi-continuous, i.e., for each open neighborhood V of $F(x)$, there exists an open neighborhood U of x such that $F(U) \subseteq V$. Notice that if $F(x) = \{y\}$ is a singleton, the definition of upper semi-continuity reduces to the standard definition of continuity. A differential inclusion is an equation of the form $\frac{dx(t)}{dt} \in F(x(t))$, whose solutions are absolute continuous functions satisfying point-wise the inclusion. If F is upper semi-continuous, then solutions exist for any initial point. Hence, we can define the semi-flow of the differential inclusion, which is a set-valued map $\phi_T(x_0) = \{x(T) \mid x(0) = x_0,\ x(t) \text{ solution of } \frac{dx(t)}{dt} \in F(x(t))\}$. It turns out that, if F is upper semi-continuous, then so is ϕ_T, provided solutions are defined up to time T [3].

Given a PWS system, it is easy to convert it to a differential inclusion, along the lines of (2). For a more precise statement, see [3, 9]. Therefore, we can conclude that the flow of our PWS system is an upper semi-continuous function, hence a posteriori continuous, as H1 implies that ϕ_T is a point-value function. □

We now require an additional hypothesis, related to the behavior of the PWS at the discontinuity surfaces, which is required to apply the convergence theorem of [6].

H2: The trajectories $x(t)$ of the PWS $\frac{dx(t)}{dt} = F(x(t))$ are *regular*, i.e., they satisfy:

1. for each initial state $x_0 \in E$, the number of *discontinuous events*, i.e., changes of vector field like transversal crossing and points in which sliding motion starts or terminates, is bounded.
2. sliding motion happens only on single discontinuity surfaces $\mathcal{H}_i$, with both vector fields having a non-null normal component to $\mathcal{H}_i$. It never happens on the intersection of more than one discontinuity surface. It can end either when the normal component of one vector field becomes zero (first order exit conditions), or when trajectories intersect another discontinuity surface. In this case, we can have transversal crossing confined to discontinuity surfaces.

Hypothesis H2 is needed as it allows us to properly define a Filippov solution for the PWS. In fact, sliding motion happening on $k > 1$ surfaces cannot be defined by the Filippov approach, as there are 2^k coefficients needed to construct the sliding vector field, but only $k + 1$ equations to constrain them. However, solutions of the PWS may still be defined in the context of differential inclusions (and convergence proved [14]), but they may be difficult to compute and not necessarily unique.

Given that the flow function $\phi_t(x)$ is well defined in each point x and continuous, hence a fortiori measurable, we can define a concept of invariant measure for $\phi_t(x)$ similarly to ODE flows: a probability measure μ on E is invariant for ϕ if and only if $\mu(A) = \mu(\phi_t^{-1}(A) \cap E)$ for each Borel measurable set $A \subseteq E$. The intersection with E is needed because flows pointing inwards E at its boundary will leave E if we reverse time. Notice that the flow of a PWS may not be time-reversible: if we invert time for stable sliding motion, we obtain an unstable sliding motion, for which trajectories are not uniquely defined.

We recall now some notions about sequences of measures that will be needed in the following. We refer the reader to [5] for any additional detail.

A sequence of probability measures $\mu^{(N)}$ on E converges weakly to a probability measure μ if and only if, for all bounded continuous functions $g : E \to \mathbb{R}$,

$$\int_E g(x)\mu^{(N)}(dx) \to \int_E g(x)\mu(dx).$$

The weak convergence provides the space of probability measures on E with a topology, called the weak topology.

A sequence of measures is *tight* if and only if, for each $\varepsilon > 0$ there exists a compact set K_ε such that, for each $N \geq 0$, $\mu_N(K_\varepsilon) > 1 - \varepsilon$. If $E \subseteq \mathbb{R}^n$ (so that it is a Polish space), then each sequence of measures on E is tight if and only if it is *relatively compact*, i.e., if and only if it has a subsequence converging to a probability measure μ. Notice that this is trivially the case whenever E is a bounded subset of $\mathbb{R}^n$.

We will now state the key lemma that provides convergence of invariant measures. It is essentially a straightforward adaption of Corollary 3.2 of [4]. Before that, we recall the following limit theorem:

THEOREM 4 (SEE [6, 14]). *Let $X^{(N)}(t)$ be the sequence of CTMCs and $\phi_t(x_0)$ be the trajectory of the PWS starting at x_0, and assume that $X^{(N)}(0)$ converges to x_0 in probability. If the PWS satisfies H0 to H2, then, for any $T > 0$, $\sup_{t \leq T} \|X^{(N)}(t) - \phi_t(x_0)\| \to 0$ in probability.*

As a corollary, it obviously holds that $\mathbb{P}\{\|X^{(N)}(T) - \phi_T(x_0)\| > \varepsilon\} \to 0$ for any fixed time T.

LEMMA 2. *Let $\mu^{(N)}$ be an invariant measure of $X^{(N)}(t)$, and assume E is bounded. Furthermore, assume that the flow $\phi_t(x)$ of the PWS defined by $X^{(N)}(t)$ satisfies H1 and H2. Then, any limit measure μ of the sequence $\mu^{(N)}$ is an invariant measure of $\phi_t(x)$.*

PROOF. The proof proceeds as in [4]. First, note that as E is bounded, the sequence of invariant measures $\mu^{(N)}$ is tight, hence relatively compact, so that the set of limits points of $\mu^{(N)}$ is not empty (that is, there exists a measure μ such that a subsequence of $\mu^{(N)}$ converges weakly to μ). Let μ be one such limit point. By possibly passing to a subsequence of $\mu^{(N)}$, we can assume that μ is the limit of $\mu^{(N)}$.

Let g be any bounded continuous function on E. As E is bounded, g is uniformly continuous on E. Now, due to the invariance of $\mu^{(N)}$, for each $T > 0$ we have:

$$\int_E \mathbb{E}_x[g(X^{(N)}(T))]\mu^{(N)}(dx) = \int_E g(x)\mu^{(N)}(dx). \quad (9)$$

Now, as g is uniformly continuous in E, for each $\delta > 0$ there is an $\eta > 0$ such that, if $\|u - v\| < \eta$ then $\|g(u) - g(v)\| < \delta$ for each $u, v \in E$. Recalling that $\phi_T(E) \subseteq E$, it holds that

$$\begin{aligned}
&\|\mathbb{E}_x[g(X^{(N)}(T)) - g(\phi_T(x))]\| = \\
&\quad \|\mathbb{E}_x[(g(X^{(N)}(T)) - g(\phi_T(x)))1_{\|X^{(N)}(T) - \phi_T(x)\| \leq \eta}] \\
&\quad \mathbb{P}\{\|X^{(N)}(T) - \phi_T(x)\| \leq \eta\} + \\
&\quad \mathbb{E}_x[(g(X^{(N)}(T)) - g(\phi_T(x)))1_{\|X^{(N)}(T) - \phi_T(x)\| > \eta}] \\
&\quad \mathbb{P}\{\|X^{(N)}(T) - \phi_T(x)\| > \eta\}\| \\
&\quad \leq \delta + 2\|g\|\mathbb{P}\{\|X^{(N)}(T) - \phi_T(x)\| > \eta\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left\| \int_E \mathbb{E}_x[g(X^{(N)}(T))]\mu^{(N)}(dx) - \int_E g(\phi_T(x))\mu^{(N)}(dx) \right\| \\
&\quad \leq \int_E \left\| \mathbb{E}_x[g(X^{(N)}(T)) - g(\phi_T(x))] \right\| \mu^{(N)}(dx) \\
&\quad \leq \delta + 2\|g\| \int_E \mathbb{P}\{\|X^{(N)}(T) - \phi_T(x)\| > \eta\}\mu^{(N)}(dx).
\end{aligned}$$

Exploiting the fact that μ is the limit of $\mu^{(N)}$, and by Theorem 4 and (9), we have that

$$\left\| \int_E g(\phi_T(x))\mu(dx) - \int_E g(x)\mu(dx) \right\| \leq \delta,$$

which, by the arbitrariness of δ and g, proves the lemma. To see why this is the case, observe that if $I_A(x)$ is the indicator function of a (measurable) set $A \subseteq E$, then $I_A(G(x))$ is the indicator function of $G^{-1}(A)$. If we could use I_A in place of g in the previous inequality, we would have concluded. To show this, fix an open set A and approximate I_A by the sequence of continuous functions $g_\varepsilon = \min\{1, d(x, A^c)/\varepsilon\}$. As g_ε converges monotonically to I_A, by the monotone convergence theorem [5] $\int_E g_\varepsilon(x) \to \int_E I_A(x)$, hence invariance holds for open sets. Furthermore, it can then be extended to all measurable sets by a straightforward application of the π-λ theorem [5]. $\square$

Notice that in the previous proof, we used continuity of ϕ only implicitly, to ensure that $g \circ \phi_T$ is integrable. However, the previous proof can be easily extended to unbounded domains E, just requiring tightness of the sequence of measures $\mu^{(N)}$. This allows to restrict the integral on a compact set K, introducing only a small error. In this case, however, continuity of ϕ_T has a crucial role, to ensure that the image of any compact set is compact, so that the uniform continuity argument can be applied on compact neighbors of $\phi_T(K)$.

Now we are almost ready to state the limit theorem about steady states. We just need an additional hypothesis on the PWS flow:

H3: The PWS system has a unique globally attracting fixed point, hence for all $x \in E$, $\lim_{t\to\infty} \phi_t(x) = x^*$.

Hypothesis H3 immediately implies the following

Proposition 1. *A PWS satisfying H0 to H3 has a unique invariant measure $\mu = \delta_{x^*}$, equal to the Dirac delta measure on the global equilibrium x^*.*

Proof. Due to $H3$, for any ball $B_\varepsilon(x^*)$ of radius ε centered in x^*, there is a $t > 0$ such that $\phi_t(E) \subseteq B_\varepsilon(x^*)$, hence $\phi_t^{-1}(B_\varepsilon(x^*)) \cap E = E$. If μ is an invariant measure, it therefore satisfies $\mu(B_\varepsilon(x^*)) = 1$ for any $\varepsilon > 0$. But the only measure with this property is δ_{x^*}. $\square$

We can now state the following result, which is a generalization of Theorem 2:

Theorem 5. *Let $X^{(N)}(t)$ be a sequence of irreducible CTMCs satisfying hypothesis H0, and let $\mu^{(N)}$ be their unique invariant measure. Let F be the drift of the sequence, defining a PWS on E with flow $\phi_t(x)$ satisfying H1 to H3, and let x^* be its globally attracting fixed point. Let $X^{(N)}(0) \to x_0 \in E$ in probability, then*

$$\lim_{N\to\infty} \lim_{t\to\infty} \|X^{(N)}(t) - x^*\| = 0 \quad \textit{in probability.}$$

Proof. Consider the sequence of CTMC invariant measures $\mu^{(N)}$ and let μ be a limit point. By Lemma 2 and Proposition 1, $\mu = \delta_{x^*}$. Therefore, the set of limit points of $\mu^{(N)}$ contains a single point, hence the whole sequence $\mu^{(N)}$ converges weakly to δ_{x^*}. Therefore, $\lim_{t\to\infty} X^{(N)}(t)$ converges in distribution, for N going to infinity, to x^*. But as x^* is deterministic, it converges also in probability [5]. $\square$

Acknowledgements

Luca Bortolussi has done part of this work while visiting on sabbatical the Laboratory for Foundations of Computer Science at the University of Edinburgh.

6. REFERENCES

[1] N. Akar, E. Karasan, and K. Dogan. Wavelength converter sharing in asynchronous optical packet/burst switching: an exact blocking analysis for markovian arrivals. *IEEE Journal on Selected Areas in Communications*, 24:69–80, 2006.

[2] N. Akar, Ezhan Karasan, and Carla Raffaelli. Fixed point analysis of limited range share per node wavelength conversion in asynchronous optical packet switching systems. *Photon. Netw. Commun.*, 18(2):255–263, 2009.

[3] J. Aubin and A. Cellina. *Differential Inclusions*. Springer-Verlag, 1984.

[4] M. Benaïm. Recursive algorithms, urn processes and chaining number of chain recurrent sets. *Ergodic Theory and Dynamical Systems*, 18:53–87, 1998.

[5] P. Billingsley. *Probability and Measure*. John Wiley and Sons, 1979.

[6] L. Bortolussi. Hybrid limits of continuous time Markov chains. In *Proc. of QEST 2011*, pages 3–12, Aachen, Germany, 2011.

[7] J. Le Boudec, D. McDonald, and J. Mundinger. A generic mean field convergence result for systems of interacting objects. In *Proc. of QEST 2007*, pages 3–15, Edinburgh, UK, 2007.

[8] F. Callegati. Optical buffers for variable length packet switching. *IEEE Communications Letters*, 4:292–294, 2002.

[9] J. Cortes. Discontinuous dynamical systems: A tutorial on solutions, nonsmooth analysis, and stability. *IEEE Control Systems Magazine*, pages 36–73, 2008.

[10] M. di Bernardo, C.J. Budd, A.R. Champneys, and P. Kowalczyk. *Piecewise-smooth dynamical systems: Theory and Applications*. Springer, 2007.

[11] S. N. Ethier and T. C. Kurtz. *Markov processes: characterization and convergence*. Wiley, 1986.

[12] A. Filippov. *Differential equations with discontinuous right-hand sides*. Mathematics and its Applications. Kluwer Academic, 1988.

[13] W. Fischer and K. Meier-Hellstern. The Markov-modulated poisson process (MMPP) cookbook. *Performance Evaluation*, 18(2):149 – 171, 1993.

[14] N. Gast and B. Gaujal. Mean field limit of non-smooth systems and differential inclusions. *SIGMETRICS Perform. Eval. Rev.*, 38:30–32, October 2010.

[15] D.P. Gaver, P.A. Jacobs, and G. Latouche. Finite Birth-and-Death models in randomly changing environments. *Adv. in Appl. Probab.*, 16:715–731, 1984.

[16] T. Kurtz. *Approximation of population processes*. Society for Industrial and Applied Mathematics, 1981.

[17] G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods and stochastic modeling*. SIAM, Philadelphia, 1999.

[18] D.M. Lucantoni. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models*, 7(1):1–46, 1991.

[19] J. F. Pérez and B. Van Houdt. A mean field model for an optical switch with a large number of wavelengths

and centralized partial conversion. *Performance Evaluation*, 67:1044–1058, Nov 2010.

[20] W. Rogiest, J. Lambert, D. Fiems, B. Van Houdt, H. Bruneel, and C. Blondia. A unified model for synchronous and asynchronous FDL buffers allowing closed-form solution. *Performance Evaluation*, 66:343–355, 2009.

[21] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[22] L. Xu, H. Perros, and G. N. Rouskas. A queueing network model of an edge optical burst switching node. In *Proceedings of the IEEE Infocom 2003*, 2003.

[23] L. Xu, H. Perros, and G. N. Rouskas. Performance analysis of an edge optical burst switching node with a large number of wavelengths. *Performance Evaluation*, 64:315–346, 2007.

[24] S. Yao, B. Mukherjee, S. J. B. Yoo, and S. Dixit. All-optical packet-switched networks: a study of contention-resolution schemes in an irregular mesh network with variable-sized packets. In *Proc. of OPTICOMM 2000*, 2000.

APPENDIX

A. PHASE-TYPE PACKET LENGTHS AND GLOBAL ATTRACTION

In this Appendix we prove that the unique fixed point π introduced in Theorem 3 is a global attractor for certain subclasses of the set of phase-type distributions and certain ranges of σ. We consider two cases: π lies in the interior of S, that is, $\sigma > \lambda^2$, and π is part of the boundary $\{(w,c)|ce = \sigma\}$, i.e., $\sigma \leq \lambda^2$. For the phase-type distribution (α, T), recall that we assumed that $T^* > 0$. Further, $\min_i T_i^* \leq 1 \leq \max_i T_i^*$ with equalities only for representations of the exponential distribution (because the mean packet length is defined as one).

A.1 Case 1: $\sigma > \lambda^2$

In this section we establish three results: one for general phase-type distributions, one for hyperexponential (HE) distributions and finally one for order-2 HE distributions. Note, the first result is trivial if $\lambda \geq \min_i T_i^*$.

THEOREM 6. *For $\sigma > \lambda/\min_i T_i^*$, the fixed point $\pi = (\lambda(1-\lambda)\theta, \lambda^2\theta)$ is a global attractor of the set of ODEs given by* (5) *for any phase-type distribution with $T^* > 0$.*

PROOF. From Section 4.1, we know that the sliding motion on the surface $\{(w,c)|we + ce = 1\}$ only occurs on the region that bounds S if one of the entries of T^* is less than λ. Thus, the motion on the part of $\{(w,c)|we + ce = 1\}$ that bounds S is transversal. Further, a sliding motion on the surface of $\{(w,c)|ce = \sigma\}$ only occurs if $cT^* \leq \lambda(we + \sigma) \leq \lambda$. As $ce = \sigma$, $cT^* \geq \sigma(\min_i T_i^*)$ which implies that a sliding motion cannot occur if $\sigma > \lambda/\min_i T_i^*$. In other words, the boundary of S cannot be reached from the interior and if we start on the boundary we immediately move into the interior of S. Therefore, all trajectories starting in S will converge to $((1-\lambda)\lambda\theta, \lambda^2\theta)$. $\square$

For a HE distribution T is diagonal and $\alpha > 0$. Further, let $\mu_i = -T_{ii}$ and assume without loss of generality that $\mu_1 < \mu_2 < \ldots < \mu_m$.

Define the function $V(w,c)$ from $\mathbb{R}^{2m}$ to $\mathbb{R}$ such that

$$V(w(t), c(t)) = (w(t) + c(t))v,$$

where v is a right-eigenvector of T such that $\alpha v > 0$, $Tv = \tau v$ with $\tau < 0$ a real eigenvalue of T. Notice, as T is a sub-generator, such an eigenvalue always exists (e.g., τ_1).

LEMMA 3. *For any phase-type distribution with $T^* > 0$ and any $\epsilon > 0$, there exists a $t_0 > 0$ such that*

$$V(w(t), c(t)) < \lambda\theta v,$$

for all $t > t_0$.

PROOF. In the interior of S, we find that

$$\frac{d}{dt}V(w(t), c(t)) = (w(t)+c(t))Tv + \lambda\alpha v = \tau(w(t)+c(t))v + \lambda\alpha v.$$

Thus, if $(w(t) + c(t))v > \lambda\theta v = -\lambda\alpha v/\tau$, we see that

$$\frac{d}{dt}V(w(t), c(t)) < -\lambda\alpha v + \lambda\alpha v = 0.$$

Similarly, $(w(t)+c(t))v < \lambda\theta v$ implies that $\frac{d}{dt}V(w(t), c(t)) > 0$. Hence, $(w(t)+c(t))v$ converges towards $\lambda\theta v$ in the interior of S. During the sliding motion on $\{(w,c)|we + ce = 1\}$ we find that

$$\frac{d}{dt}V(w(t), c(t)) = \tau(w(t) + c(t))v + (w(t) + c(t))T^*\alpha v.$$

Thus, if $(w(t) + c(t))v > \lambda\theta v = -\lambda\alpha v/\tau$ we have

$$\frac{d}{dt}V(w(t), c(t)) < ((w(t) + c(t))T^* - \lambda)\alpha v.$$

As $(w(t) + c(t))T^* - \lambda < 0$ is required during the sliding motion, we find that $(w(t) + c(t))v$ decreases towards $\lambda\theta v$ during the sliding motion. Note, when $(w(t) + c(t))v < \lambda\theta v$, the above equation only indicates that $\frac{d}{dt}V(w(t), c(t)) > ((w(t) + c(t))T^* - \lambda)\alpha v$, which is a negative number, so $(w(t) + c(t))v$ can also decrease. Finally, for the sliding motion on $\{(w,c)|ce = \sigma\}$, one readily checks that

$$\frac{d}{dt}V(w(t), c(t)) = \tau(w(t) + c(t))v + (c(t)T^* - \lambda(w(t)e + \sigma))\alpha v + \lambda\alpha v,$$

which yields

$$\frac{d}{dt}V(w(t), c(t)) < (c(t)T^* - \lambda(w(t)e + \sigma))\alpha v < 0,$$

during the sliding motion provided that $(w(t)+c(t))v > \lambda\theta v$. In conclusion, when $V(w(t), c(t)) - \lambda\theta v > 0$, for $(w(t), c(t)) \in S$, it will decrease towards zero. $\square$

In case of HE distributions we can set $v = e_i$ (a vector of zeros with a one in position i) and $\tau = T_{ii}$, for any $i = 1, \ldots, m$.

COROLLARY 3. *For any hyperexponential distribution and any $\epsilon > 0$, there exists a $t_0 > 0$ such that*

$$w_i(t) + c_i(t) < \lambda\theta_i + \epsilon/m,$$

for all $t > t_0$ and $i \in \{1, \ldots, m\}$. Hence, $w(t)e + c(t)e < \lambda + \epsilon$ for $t > t_0$.

THEOREM 7. *For $\sigma > \lambda^2/\min_i T_i^*$, the fixed point $\pi = (\lambda(1-\lambda)\theta, \lambda^2\theta)$ is a global attractor of the set of ODEs given by* (5) *for any hyperexponential distribution.*

PROOF. If we set $\epsilon < 1-\lambda$ in Corollary 3, we find that $w(t)e + c(t)e < 1$ after time t_0, meaning we cannot hit the surface $\{(w,c)|we+ce=1\}$ after time t_0. On the surface $\{(w,c)|ce=\sigma\}$ the value of $c(t)T^*$ is still lower bounded by $\sigma(\min_i T_i^*) = \sigma\mu_1$, but now for $t > t_0$, $\lambda(w(t)e+\sigma)$ can upper bounded by $\lambda(\lambda+\epsilon)$. Thus, for any $\sigma > \lambda^2/\mu_1$, we can set $\epsilon < \sigma\mu_1/\lambda - \lambda$ to guarantee that no sliding motion on $\{(w,c)|ce=\sigma\}$ can occur after time t_0. □

THEOREM 8. *For $\sigma > \lambda^2$, the fixed point $\pi = (\lambda(1-\lambda)\theta, \lambda^2\theta)$ is a global attractor of the set of ODEs given by* (5) *for any order-2 hyperexponential distribution.*

PROOF. The drift of $c_i(t)$ in the interior of S can be written as

$$\frac{d}{dt}c_i(t) = -\mu_i c_i(t) + \lambda(w(t)+c(t))e\alpha_i.$$

If $c_i(t) = \lambda^2\theta_i + \delta = \lambda^2\alpha_i/\mu_i + \delta$, for $\delta > 0$, then Corollary 3 implies

$$\frac{d}{dt}c_i(t) < -\delta + \lambda\epsilon\alpha_i,$$

after time t_0. Hence, for $c_i(t) > \lambda^2\theta_i$, $c_i(t)$ decreases after time t_0 (in the interior of S).

Further, during the slide on $\{(w,c)|ce=\sigma\}$, the drift

$$\frac{d}{dt}c_i(t) = -\mu_i c_i(t) + \left(\sum_{i=1}^{m} c_i(t)\mu_i\right)\alpha_i.$$

For HE distributions of order 2 (with $\mu_1 < \mu_2$), we have $\alpha_2 = 1-\alpha_1$ and $c_2(t) = \sigma - c_1(t)$ during the slide, which implies

$$\begin{aligned}\frac{d}{dt}c_1(t) &= \sigma\mu_2\alpha_1 - c_1(t)(\alpha_2\mu_1+\alpha_1\mu_2),\\ \frac{d}{dt}c_2(t) &= \sigma\mu_1\alpha_2 - c_2(t)(\alpha_1\mu_2+\alpha_2\mu_1).\end{aligned}$$

As $\alpha_1/\mu_1 + \alpha_2/\mu_2 = 1$ and $\theta_1 = \alpha_1/\mu_1$, we find that if $c_i(t)$ differs from $\sigma\theta_i$, $c_i(t)$ approaches $\sigma\theta_i$ during the slide on $\{(w,c)|ce=\sigma\}$ (for any t).

Hence, for $\sigma > \lambda^2$ and any $\epsilon_2 > 0$, there exists a $t_1 > t_0$ such that $c_1(t)$ is upper bounded by $\sigma\theta_1 + \epsilon_2$ for $t > t_1$. This indicates that $c(t)T^*$ can be lower bounded by $\sigma + \delta$ for any $\delta > 0$ after time t_1. Repeating the arguments in Theorem 7 with this improved lower bound on $c(t)T^*$ for order-2 HE distributions, suffices to complete the proof. □

A.2 Case 2: $\sigma \le \lambda^2$

In this section we establish a result for any phase-type distribution as well as one for the class of HE distributions.

THEOREM 9. *For $\sigma < \lambda^2/(\max_i T_i^*)^2$, the fixed point $\pi = (\lambda(1-\sigma)\theta/(1+\lambda), \sigma\theta)$ is a global attractor of the set of ODEs given by* (5) *for any phase-type distribution with $T^* > 0$.*

PROOF. During the slide on $\{(w,c)|ce=\sigma\}$, we observe that

$$\frac{d}{dt}w(t)e = -w(t)T^* - \lambda w(t)e + \lambda(1-\sigma).$$

As $w(t)T^* \le w(t)e(\max_i T_i^*)$,

$$\frac{d}{dt}w(t)e \ge -w(t)e(\lambda + \max_i T_i^*) + \lambda(1-\sigma).$$

The derivative of $w(t)e$ is therefore positive if $w(t)e < \lambda(1-\sigma)/(\lambda+\max_i T_i^*)$. For the motion in the interior of S, we see that

$$\frac{d}{dt}w(t)e = -w(t)T^* - \lambda w(t)e + \lambda(1-c(t)e),$$

where $c(t)e < \sigma$. Hence, by the same argument the derivative of $w(t)e$ is also positive in the interior of S if $w(t)e < \lambda(1-\sigma)/(\lambda+\max_i T_i^*)$. On the surface $\{(w,c)|we+ce=1\}$, we have $w(t)e \ge (1-\sigma) > \lambda(1-\sigma)/(\lambda+\max_i T_i^*)$. This allows us to conclude that for any $\epsilon > 0$, there exists a $\bar{t}$ such that for all $t > \bar{t}$

$$w(t)e > \frac{\lambda(1-\sigma)}{\lambda+\max_i T_i^*} - \epsilon.$$

When combined with the inequality $c(t)T^* \le \sigma\max_i T_i^*$, this indicates that $c(t)T^* \ge \lambda(w(t)e+\sigma)$ cannot hold after time $\bar{t}$ if

$$\begin{aligned}&\sigma(\max_i T_i^*) < \lambda\left(\frac{\lambda(1-\sigma)}{\lambda+\max_i T_i^*} - \epsilon + \sigma\right)\\ \Leftrightarrow\quad &\sigma < \left(\frac{\lambda}{\max_i T_i^*}\right)^2 (1-\epsilon(\lambda+\max_i T_i^*)/\lambda).\end{aligned}$$

Thus, with a properly chosen ϵ, we may conclude that any sliding motion on the surface $\{(w,c)|ce=\sigma\}$ that starts after time $\bar{t}$ will last indefinitely for $\sigma < (\lambda/\max_i T_i^*)^2$. □

THEOREM 10. *For $\sigma < \lambda^2(1-\lambda)/(\max_i T_i^* - \lambda)$, the fixed point $\pi = (\lambda(1-\sigma)\theta/(1+\lambda), \sigma\theta)$ is a global attractor of the set of ODEs given by* (5) *for any hyperexponential distribution.*

PROOF. The drift of $w_i(t)$ in the interior of S can be written as

$$\frac{d}{dt}w_i(t) = -\mu_i w_i(t) - \lambda(w(t)+c(t))e\alpha_i + \lambda\alpha_i.$$

Hence, if $w_i(t) = \lambda(1-\lambda)\theta_i - \delta$, for some $\delta > 0$, then

$$\frac{d}{dt}w_i(t) > \delta - \lambda\epsilon\alpha_i,$$

after time t_0, due to Corollary 3. Meaning, if $w_i(t) < \lambda(1-\lambda)\theta_i$, then $w_i(t)$ increases after time t_0 (in the interior of S). For the drift of $w_i(t)$ during the slide on $\{(w,c)|ce=\sigma\}$, we observe that

$$\frac{d}{dt}w_i(t) = -\mu_i w_i(t) - \lambda w(t)e\alpha_i + \lambda(1-\sigma)\alpha_i.$$

Hence, after time t_0 we get

$$\frac{d}{dt}w_i(t) > -\mu_i w_i(t) + \lambda(1-\lambda)\alpha_i - \lambda\epsilon\alpha_i,$$

by noting that $w(t)e < \lambda - \sigma + \epsilon$. Thus, if $w_i(t) < \lambda(1-\lambda)\theta_i$, $w_i(t)$ will increase after time t_0. This allow us to conclude that for any $\epsilon_2 > 0$ there exists a $t_2 > t_0$, such that $w(t)e$ is lower bounded by $\lambda(1-\lambda) - \epsilon_2$. The expression $\lambda(w(t)e+\sigma)$ is therefore lower bounded by $\lambda^2(1-\lambda) + \lambda\sigma - \lambda\epsilon_2$. As $c(t)T^*$ is upper bounded by $\sigma\max_i T_i^*$, we find that the unique fixed point is a global attractor for any HE distribution with $\sigma < \lambda^2(1-\lambda)/(\max_i T_i^* - \lambda)$. □