ON SIMULATION OF NON-MARKOVIAN STOCHASTIC PETRI NETS
WITH HEAVY-TAILED FIRING TIMES

Peter W. Glynn

Peter J. Haas

Stanford University
Stanford, CA 94305, USA

IBM Research - Almaden
San Jose, CA 95120-6099, USA

ABSTRACT

Long-run stochastic stability is a precondition for applying steady-state simulation output analysis methods to a stochastic Petri Net (SPN), and is of interest in its own right. A fundamental stability requirement for an irreducible SPN is that the markings of the net be recurrent, in that the marking process visits each marking infinitely often with probability 1. We study recurrence properties of irreducible non-Markovian SPNs with finite marking set. Our focus is on the "clocks" that govern the transition firings, and we consider SPNs in which zero, one, or at least two simultaneously-enabled transitions can have very heavy-tailed clock-setting distributions. We establish positive recurrence, null recurrence, and, perhaps surprisingly, possible transience of markings for these respective regimes. The transience result stands in strong contrast to Markovian or semi-Markovian SPNs, where irreducibility and finiteness of the marking set guarantee positive recurrence.

1 INTRODUCTION

Stochastic Petri nets (SPNs) have proved to be a powerful and enduring framework for modeling and performance analysis of complex systems characterized by concurrency, synchronization, precedence, and contention for shared resources. In particular, SPNs with general clock-setting distributions ("non-Markovian" SPNs) have long been recognized as an appealing framework for modeling and simulation of discrete-event stochastic systems (Balbo and Chiola 1989; Chiola 1991; Chiola and Ferscha 1993; Haas 2002; Haas and Shedler 1985). The "marking" of a net (assignment of "token" counts to "places") corresponds to the system state and the "firing" of "transitions" corresponds to the occurrence of state-changing events. The "marking process" of an SPN is the stochastic process that records the marking as it evolves over continuous time, thus capturing the continuous-time evolution of the modeled system. SPNs admit a graphical representation (see Figures 1–3) that is well suited to top-down and bottom-up modeling of complex systems. The state-transition and event-scheduling mechanisms for a discrete-event system are apparent from the SPN graph, and the graph structure can be exploited to develop efficient algorithms for generating a sample path of the marking process during a simulation run, both in sequential and distributed simulation settings (Chiola 1991; Chiola and Ferscha 1993). Symmetries in the SPN graph can be exploited, via extensions such as "colored" SPNs or "stochastic well-formed nets", to provide compact model representations and to facilitate simulation output analysis by yielding shorter cycle lengths for the regenerative method and by increasing overall simulation efficiency; see, e.g., Gaeta and Chiola (1995), Chapter 9 in Haas (2002), and Prisgrove and Shedler (1986). Finally, Petri net tools can sometimes be applied to compute qualitative properties of a model—such as reachability of markings or liveness of transitions—that complement quantitative properties estimated via simulation (Chiola et al. 1995).

In the context of steady-state simulation output analysis for SPNs, a fundamental question is whether the marking process of an SPN is stochastically stable over time. Such long-run stochastic stability ensures that questions about the "steady state" of a given system are well posed, and is a precondition for applying steady-state simulation output analysis techniques such as regenerative, spectral, or standardized time series

methods. Steady-state stability is important in its own right, since lack of such stability can indicate problematic behavior in the system of interest or deficiencies in the modeling process.

A basic requirement for steady-state output analysis is that the markings of the net be "recurrent", in that the marking process visits each marking infinitely often with probability 1. For example, the regenerative method for simulation output analysis—see, e.g., Shedler (1993) and references therein—requires that each regeneration point be almost surely (a.s.) finite, which often amounts to the requirement that a specified marking (having a "single state" property) is recurrent. Other methods, such as the method of standardized time series, assume that the marking process obeys a functional central limit theorem (FCLT); see Glynn and Iglehart (1990). Such an FCLT is implied by a "Harris recurrence" condition that in turn implies recurrence for each marking (Glynn and Haas 2006). In this paper, we study recurrence properties of non-Markovian SPNs. We focus on SPNs having a finite marking set and in which any marking can potentially be reached from any other marking via a sequence of transition firings; SPNs having this latter property are called "irreducible" (see Definition 1 below).

Lack of recurrence for an SPN with an infinite marking set often means that the marking process drifts off to the far reaches of the set, never to return. Such outward drift can correspond, for example, to overflowing queues or buffers in telecommunication, service, and manufacturing models. In SPNs with finite or infinite marking sets, transience of certain states can correspond to "starvation" (withholding of needed resources) or over-provisioning of some process or activity within a complex system—see Section 5—or to effective reducibility in a nominally "irreducible" system. In the latter scenario, any state can be reached from any other state in principle, but if such reachability hinges on passage through transient states, then certain states will eventually become unreachable from each other with probability 1. Consequently, the steady-state or limiting distribution of the marking process, if it exists, may depend upon the initial state of the system in a potentially complex manner, making steady-state estimation difficult.

An SPN changes marking in accordance with the firing of one or more transitions that are "enabled" in the current marking. The firing of transitions is in turn governed by a set of *clocks*, with one clock per transition. The clock corresponding to a transition indicates the time until a transition is scheduled to fire. When each clock is set according to an exponential distribution, the marking process reduces to a continuous-time Markov chain (Haas 2002, Sec. 3.4), and classical theory applies; see, e.g., Asmussen (2003). In particular, if the SPN is irreducible with finite marking set, then recurrence of each marking is assured. Indeed, each marking is "positive recurrent"—in that the expected number of transition firings between visits is finite—and the expected hitting times of each marking (in continuous time) are finite. Similarly, if at most one transition is enabled in each marking, then the marking process is a semi-Markov process, and again the finiteness of the marking set and the irreducibility property together guarantee that each marking is positive recurrent. If each clock-setting distribution has finite mean, then the expected hitting times of each marking are finite.

When the clock-setting distribution functions can be arbitrary and multiple transitions can be enabled in a marking, the behavior of the SPN becomes much more complex and sometimes counterintuitive. For example, it is possible to construct an SPN having an infinite marking set in which the expected time between transition firings increases linearly but, with probability 1, an infinite number of marking changes occur in a finite time interval (Haas 2002, p. 90). We show here that, even when the marking set is constrained to be finite, non-intuitive behavior can occur in a non-Markovian SPN when the means of the clock-setting distributions can be infinite. Specifically, we study SPNs in which zero, one, or at least two simultaneously-enabled transitions can have such "(very) heavy-tailed" clock-setting distributions. We establish discrete-time positive recurrence, null recurrence, and, perhaps surprisingly, possible transience of markings for these respective regimes. These results, which have a rough analogy to the theory of random walks on multidimensional integer lattices, illustrate the richly complex behavior of non-Markovian SPNs relative to their Markovian or semi-Markovian counterparts. We note that our results extend to other discrete-event formalisms, such as generalized semi-Markov processes (Glynn 1989; Glynn and Haas 2006) and event graphs (Schruben 1983).

2 SPN PRELIMINARIES

We briefly review our non-Markovian SPN model, and describe the "hazard rate" representation of an SPN. Then we define a useful set of irreducibility and "positive density" conditions.

2.1 SPNs with General Clock-Setting Distributions

Following Haas and Shedler (1989, 1991), let $D = \{d_1, d_2, \ldots, d_L\}$ be a finite set of *places*, $E = \{e_1, e_2, \ldots, e_M\}$ be a finite set of *transitions*, and $E' \subset E$ a (possibly empty) set of *immediate transitions*. The transitions in $E \setminus E'$ are called *timed transitions*. Also let $I(e), L(e), J(e) \subseteq D$ be the sets of *normal input places*, *inhibitor input places*, and *output places*, respectively, for transition $e \in E$. Denote by G the finite or countably infinite set of *markings*. For $s \in G$ we write $s = (s_1, s_2, \ldots, s_L)$, where s_j is the number of *tokens* in place $d_j \in D$. Set $E(s) = \{e \in E : s_j \geq 1 \text{ for } d_j \in I(e) \text{ and } s_j = 0 \text{ for } d_j \in L(e)\}$, so that $E(s)$ (assumed nonempty) is the set of transitions that are *enabled* when the marking is s. A transition $e \in E \setminus E(s)$ is *disabled* when the marking is s. Define the set S' of *immediate markings* by $S' = \{s \in G : E(s) \cap E' \neq \varnothing\}$ and the set S of *timed markings* by $S = G \setminus S'$, so that an element of the marking set is an immediate marking if and only if at least one immediate transition is enabled.

The marking changes when one or more enabled transitions *fire*. For $E^* \subseteq E(s)$, denote by $p(s'; s, E^*)$ the probability that the new marking is s' given that the marking is s and the transitions in the set E^* fire simultaneously. The new-marking probabilities are constrained by the requirement that a transition remove at most one token from each normal input place and deposit at most one token in each output place when it fires. The token count of a place may increase or decrease by more than 1 when transitions fire simultaneously. A transition is *deterministic* if it always removes exactly one token from each input place and deposits exactly one token in each output place whenever it fires. A variety of mechanisms proposed in the literature for resolving "conflicts" between transitions whose clocks run down to zero simultaneously are special cases of our general marking-change mechanism. These former mechanisms include the use of weights or priorities on the transitions to determine a unique transition that fires, and our results therefore apply to SPN's that incorporate such conflict-resolution mechanisms.

As mentioned previously, the clock reading for an enabled transition indicates the remaining time until the transition is scheduled to fire. These clocks, along with the speeds at which the clocks run down, determine which of the enabled transitions trigger the next marking change. Denote by $r(s, e)$ (≥ 0) the *speed* (deterministic rate) at which the clock associated with transition e runs down when the marking is s. To ensure unambiguous behavior and a clean separation between timed and immediate transitions, we require all speeds to be finite and do not allow zero speeds for immediate transitions; such transitions always fire the instant that they become enabled. Moreover, by convention, $r(s, e) = 1$ if $e \in E' \cap E(s)$, and we assume that $r(s, e) > 0$ for some $e \in E(s)$. Let $C(s)$ be the set of possible *clock-reading vectors* when the marking is s: $C(s) = \{c = (c_1, \ldots, c_M) : c_i \geq 0 \text{ and } c_i > 0 \text{ if and only if } e_i \in E(s)\}$. (The ith component of a clock-reading vector $c = (c_1, \ldots, c_M)$ is the clock reading associated with transition e_i.) Beginning in marking s with clock-reading vector $c = (c_1, \ldots, c_M) \in C(s)$, the time $t^*(s, c)$ to the next marking change—also called the *holding time* in s—is given by

$$t^*(s, c) = \min_{\{i:\, e_i \in E(s)\}} c_i / r(s, e_i), \tag{1}$$

where $c_i / r(s, e_i)$ is taken to be $+\infty$ when $r(s, e_i) = 0$. The set of transitions $E^*(s, c)$ that fire simultaneously and trigger the next marking change is given by $E^*(s, c) = \{e_i \in E(s) : c_i - t^*(s, c) r(s, e_i) = 0\}$. Observe that $E^*(s, c) = E(s) \cap E'$ whenever the marking s is immediate.

At a marking change from s to s' triggered by the simultaneous firing of the transitions in the set E^*, a finite clock reading is generated for each *new transition* $e' \in N(s'; s, E^*) = E(s') \setminus (E(s) \setminus E^*)$. Denote the *clock-setting distribution function* (that is, the distribution function of such a new clock reading) by $F(\cdot; s', e', s, E^*)$. We require that $F(0; s', e', s, E^*) = 1$ if $e \in E'$ and $F(0; s', e', s, E^*) = 0$ if $e \in E \setminus E'$ so that immediate transitions always fire instantaneously and timed transitions never fire instantaneously. For

each *old transition* $e' \in O(s';s,E^*) = E(s') \cap \big(E(s) \setminus E^*\big)$, the old clock reading is kept after the marking change. For $e' \in \big(E(s) \setminus E^*\big) \setminus E(s')$, transition e' (that was enabled before the transitions in E^* fired) becomes disabled and the clock reading is discarded. When E^* is a singleton set of the form $E^* = \{e^*\}$, we write $p(s';s,e^*) = p(s';s,\{e^*\})$, $O(s';s,e^*) = O(s';s,\{e^*\})$, and so on.

2.2 The Marking Process

Denote by $X(t)$ the marking of the SPN at time t. Formal definition of the *marking process* $\{X(t): t \geq 0\}$ of an SPN with general firing times is in terms of a general state space Markov chain $\{(S_n,C_n): n \geq 0\}$ that describes the net at successive marking changes. Heuristically, $S_n = (S_{n,1}, \ldots, S_{n,L})$ represents the marking and $C_n = (C_{n,1}, \ldots, C_{n,M})$ represents the clock-reading vector just after the nth marking change; see Section 3.1 in Haas (2002) for a formal definition of this chain. The chain takes values in the set $\Sigma = \bigcup_{s \in G}\big(\{s\} \times C(s)\big)$. Denote by μ the *initial distribution* of the chain; for a subset $B \subseteq \Sigma$, the quantity $\mu(B)$ represents the probability that $(S_0,C_0) \in B$. We use the notations P_μ and E_μ to denote probabilities and expected values associated with the chain, the idea being to emphasize the dependence on the initial distribution μ; when the initial marking of the underlying chain is equal to some $(s,c) \in \Sigma$ with probability 1, we write $P_{(s,c)}$ and $E_{(s,c)}$. Typically, the SPN is initialized by selecting an initial marking s_0 according to a discrete distribution ν over G, and then generating a clock reading for each $e \in E(s_0)$ from an initial clock-setting distribution function $F_0(\cdot;e,s_0)$. We assume such an initialization procedure throughout.

We construct a continuous time process $\{X(t): t \geq 0\}$ from the chain $\{(S_n,C_n): n \geq 0\}$ in the following manner. Let ζ_n $(n \geq 0)$ be the non-negative, real-valued time of the nth marking change: $\zeta_n = \sum_{j=0}^{n-1} t^*(S_j,C_j)$. We focus throughout on SPNs whose marking set G is finite, in which case an argument as in the proof of Theorem 3.3.13 in Haas (2002) shows that

$$P_\mu\{\sup_{n \geq 0} \zeta_n = \infty\} = 1. \tag{2}$$

The marking process is then defined by setting $X(t) = S_{N(t)}$, where $N(t) = \sup\{n \geq 0: \zeta_n \leq t\}$ is the number of marking changes that occur in the interval $(0,t]$. By construction, the marking process takes values in the set S and has piecewise constant, right-continuous sample paths.

2.3 Modeling Power

A marking process $\{X(t): t \geq 0\}$ with timed marking set S is said to *mimic* another marking process $\{\tilde{X}(t): t \geq 0\}$ with timed marking set $\tilde{S}$ if there exists a mapping $\lambda: S \mapsto \tilde{S}$ such that the processes $\{\lambda X(t): t \geq 0\}$ and $\{\tilde{X}(t): t \geq 0\}$ have the same finite-dimensional distributions. Since a marking process has piecewise-constant sample paths, this congruence implies that $\{\lambda X(t): t \geq 0\}$ and $\{\tilde{X}(t): t \geq 0\}$ are identically distributed in that they have the same distributions with respect to probability measures over sets of sample paths.

2.4 Simple Transitions and Heavy-Tailed Transitions

To simplify the exposition, we focus throughout on SPNs in which each transition e' is *simple* in that there exists a distribution function $F(\cdot;e')$ such that $F(\cdot;s',e',s,E^*) \equiv F(\cdot;e')$ and $F_0(\cdot;e',s) \equiv F(\cdot;e')$ for all s', s, and E^*.

For an SPN with simple transitions, let H be the (possibly empty) subset of transitions in E such that $e \in H$ if and only if $\int_0^\infty t\, dF(t;e') = \infty$. Thus H is the set of "heavy-tailed" transitions. Set $\eta(H) = \max_{s \in G} |E(s) \cap H|$, so that $\eta(H)$ is the maximum number of heavy-tailed transitions that can be enabled simultaneously. In the following sections, we consider three possible scenarios: $\eta(H) = 0$, $\eta(H) = 1$, and $\eta(H) \geq 2$.

In applications, a heavy-tailed clock reading might correspond to the time required to process or transmit a computer file (Resnick and Rootzén 2000), the time between extreme geophysical events (Benson, Schumer, and Meerschaert 2007), or the time to deal with a financial loss (Moscadelli 2004), insurance claim (Powers

2010), extreme event in a highly optimized physical system (Carlson and Doyle 1999), or natural disaster such as a forest fire (Holmes, R.J. Huggett, and Westerling 2008)—in the latter examples we assume that these times are proportional to the loss amount, claim amount, magnitude of physical-system deviation, number of acres burned, and so on. In general, infinite-mean distributions are used to model situations in which there is a relatively large chance of seeing values much greater than any seen previously.

2.5 Hazard-Rate Representation

It is sometimes convenient to use an alternative construction of an SPN that centers around "hazard rates"—see, for example, Chapter 6 in Glasserman (1991) or Glynn (1989). Consider an SPN for which each clock-setting distribution $F(x;e)$ has a density function $f(x;e)$, and define the corresponding *hazard rate* by $h(x;e) = f(x;e)/\bar{F}(x;e)$, where $\bar{F} = 1 - F$ and we take $0/0 = 0$. For such an SPN, transitions never fire simultaneously (with probability 1). Also observe that if $\inf_x h(x;e) \geq \underline{h}$ for some $\underline{h} > 0$, then $f(x;e) > 0$ for all $x \in (0,\infty)$ and $F(\cdot;e)$ has finite moments of all orders. For $n \geq 1$, let $e_n^* = e^*(S_{n-1},C_{n-1})$ be the nth transition to fire, and for $n \geq 0$ let $t_n^* = t^*(S_n,C_n)$ be the holding time in marking S_n. Set $U = (S_0, t_0^*, e_1^*, S_1, t_1^*, e_2^*, \ldots)$, so that U completely specifies the continuous-time process $\{X(t): t \geq 0\}$, along with the sequence of trigger transitions.

Now consider a sequence $\tilde{U} = (\tilde{S}_0, \tilde{t}_0^*, \tilde{e}_1^*, \tilde{S}_1, \tilde{t}_1^*, \tilde{e}_2^*, \ldots)$ defined by the following alternative simulation algorithm. In the algorithm, the quantity $\alpha(n,i)$ is the random index of the most recent transition-firing time at or prior to $\tilde{\zeta}_n$ at which the clock for $e_i \in E(\tilde{S}_n)$ was set; by convention, $\alpha(n,i) = 0$ for $e_i \notin E(\tilde{S}_n)$.

0. (Initialization) Set $n = 0$, $\tilde{\zeta}_0 = 0$, and $\alpha(0,i) = 0$ for $e_i \in E$. Select $\tilde{S}_0$ according to ν. For $e_i \in E(\tilde{S}_0)$, generate $\tilde{C}_{n,i}$ according to a unit exponential distributon $F(x) = (1 - \exp(-x))I[x \geq 0]$.

1. For $e_i \in E(\tilde{S}_n)$, set $\tau_{n,i} = \inf\{t \geq 0: \int_{\tilde{\zeta}_n}^t h(x - \tilde{\zeta}_{\alpha(n,i)}; e_i)\, dx = \tilde{C}_{n,i}\}$. Then set $\tilde{\zeta}_{n+1} = \min_{i:e_i \in E(\tilde{S}_n)} \tau_{n,i}$, $\tilde{t}_n^* = \tilde{\zeta}_{n+1} - \tilde{\zeta}_n$, and $\tilde{e}_n^* = e_i$ such that $\tau_{n,i} = \tilde{\zeta}_{n+1}$.

2. Generate $\tilde{S}_{n+1}$ according to $p(\cdot; \tilde{S}_n, \tilde{e}_{n+1}^*)$.

3. For each $e_i \in N(\tilde{S}_{n+1}; \tilde{S}_n, \tilde{e}_{n+1}^*)$, set $\alpha(n+1,i) = n+1$ and generate $\tilde{C}_{n+1,i}$ according to a unit exponential distribution.

4. For each $e_i \in O(\tilde{S}_{n+1}; \tilde{S}_n, \tilde{e}_{n+1}^*)$, set $\alpha(n+1,i) = \alpha(n,i)$ and $\tilde{C}_{n+1,i} = \tilde{C}_{n,i} - \int_{\tilde{\zeta}_n}^{\tilde{\zeta}_{n+1}} h(x - \tilde{\zeta}_{\alpha(n,i)}; e_i)\, dx$.

5. For each $e_i \in (E(\tilde{S}_n) \setminus \{\tilde{e}_n^*\}) \setminus E(\tilde{S}_{n+1})$, set $\tilde{C}_{n+1,i} = \alpha(n+1,i) = 0$.

6. Set $n \leftarrow n+1$ and go to Step 1.

Finally, define a continuous-time process $\{\tilde{X}(t): t \geq 0\}$ in a manner analogous to the standard definition of an SPN by setting $\tilde{X}(t) = \tilde{S}_{\tilde{N}(t)}$, where $\tilde{N}(t) = \sup\{n \geq 0: \tilde{\zeta}_n \leq t\}$.

As can be seen, the key difference between the above construction and the usual SPN definition is that new clocks are set according to a unit exponential distribution and run down at a variable rate specified by the hazard function. The random variable $\tau_{n,i}$ represents the time at which a transition e_i that is enabled at time $\tilde{\zeta}_n$ is scheduled to fire.

Proposition 1 U and $\tilde{U}$ are identically distributed, as are $\{X(t): t \geq 0\}$ and $\{\tilde{X}(t): t \geq 0\}$.

The proof of this result rests on the following fact. Let F be a distribution function having a density f and hazard rate h, and let X be distributed according to a unit exponential distribution function. The random variable τ defined by $\tau = \inf\{t \geq 0: \int_0^t h(x)\, dx = X\}$ has distribution function F. That is, we can simulate a clock-setting distribution F by taking a unit exponential clock-setting distribution and running the corresponding new clock reading down to 0 at a time-varying rate given by h. Indeed, observe that $h(t) = f(t)/\bar{F}(t) = -d/dt(\ln \bar{F}(t))$, so that $\int_0^x h(t)\, dt = -\ln \bar{F}(x)$ and

$$P\{\tau \geq x\} = P\{\int_0^x h(t)\, dt \leq X\} = \exp\left(-\int_0^x h(t)\, dt\right) = \bar{F}(x).$$

In a similar manner, if we set $\tau = \inf\left\{t \geq 0 : \int_a^t h(x)\,dx = X\right\}$, then $P\{\tau \geq x\} = \bar{F}(x)/\bar{F}(a) = P\{X > x \mid X > a\}$, so that τ has the distribution of a residual clock reading from a clock-setting distribution F in a conventionally defined SPN, given that the current "age" of the clock—i.e., the amount of time that the clock has been running down—in such an SPN is a. Conditional on the past history of SPN markings, holding times, and trigger transitions, the conditional clock readings $\{\tilde{C}_{n,i} : e_i \in E(\tilde{S}_n)\}$ at any time point $\tilde{\zeta}_n$ are i.i.d unit exponential. This observation follows from the memoryless property of the exponential distribution along with a representation result for conditional clock readings; see Lemma 3.4.10 in Haas (2002). The overall argument is very similar to the proof of Theorem 3.4.21 in Haas (2002). When using the hazard-rate construction in the sequel, we suppress the tilde notation.

2.6 Irreducibility and Positive Density Conditions

In the following sections, we focus on "irreducible" SPNs having a finite marking set and simple transitions whose clock-setting distribution functions satisfy a "positive density" condition. For an SPN with marking set G and transition set E and for $s, s' \in G$ and $e \in E$, write $s \xrightarrow{e} s'$ if $p(s'; s, e) r(s, e) > 0$ and write $s \to s'$ if $s \xrightarrow{e} s'$ for some $e \in E(s)$. Also write $s \rightsquigarrow s'$ if either $s \to s'$ or there exist markings $s_1, s_2, \ldots, s_n \in G$ ($n \geq 1$) such that $s \to s_1 \to \cdots \to s_n \to s'$.

Definition 1 An SPN is said to be *irreducible* if $s \rightsquigarrow s'$ for each $s, s' \in G$.

Recall that a non-negative function G is a *component* of a distribution function F if G is not identically equal to 0 and $G \leq F$. If a component G of F is absolutely continuous, so that G has a density function g, then we say that g is a *density component* of F.

Definition 2 *Assumption PD(q) holds for a specified SPN and real number $q \geq 0$ if*

 (i)   the marking set G of the SPN is finite;
 (ii)   the SPN is irreducible;
 (iii) all speeds of the SPN are positive; and
 (iv)   there exists $\bar{x} \in (0, \infty)$ such that each clock-setting distribution function $F(\cdot; , e')$ of the SPN has finite qth moment and a density component that is positive and continuous on $(0, \bar{x})$.

Observe that Assumption PD(q) implies Assumption PD(r) whenever $q \geq r$. Moreover, Assumption PD(0) imposes finite-marking-set, positive-speed, irreducibility, and positive density conditions, but does not impose any conditions on the means of the clock-setting distributions.

3 NO HEAVY-TAILED TRANSITIONS

For an SPN with marking set G and underlying chain $\{(S_n, C_n) : n \geq 0\}$, we say that a marking $s \in G$ is *recurrent* if and only if $P_\mu\{S_n \in s \text{ i.o.}\} = 1$. Denote by $J_n(s)$ and $T_n(s)$ the nth hitting time of marking s (in discrete and continuous time, respectively) by the processes $\{S_n : n \geq 0\}$ and $\{X(t) : t \geq 0\}$.

Proposition 2 Suppose that Assumption PD(1) holds for an SPN, so that all clock-setting distributions have finite mean. Then, for any initial distribution μ and $n \geq 0$, (i) every marking of the SPN is recurrent, (ii) $E_\mu[J_n(s)] < \infty$ for $s \in G$, and (iii) $E_\mu[T_n(s)] < \infty$ for $s \in S$.

The above result establishes recurrence when Assumption PD(0) holds and every clock-setting distribution has finite mean; the final two assertions of the proposition can be viewed as a form of "positive" recurrence. The proposition follows from results in Chapter 5 of Haas (2002). The idea is to show that the "embedded chain" of the SPN—that is, the underlying Markov chain observed at marking changes where the new marking is timed—is positive Harris recurrent. (Roughly speaking, Harris recurrence means that any "dense enough" set of states is hit infinitely often with probability 1; positive Harris recurrence means that the expected time between successive hits is finite.) Applying this result to "dense" sets of the form $\{s\} \times C(s)$ for $s \in S$ proves the assertions of the theorem for timed markings, and geometric trials arguments then extend the assertions to immediate markings.

4 ONE HEAVY-TAILED TRANSITION

In this section we focus on SPNs with simple transitions for which $\eta(H) = 1$. We first show by means of a couple of examples that the hitting times (in continuous time) for a specified marking may or may not have finite means. We then show that, in any case, each marking in an SPN with $\eta(H) = 1$ is recurrent.

First consider an SPN with two places and two simple, deterministic timed transitions; see Figure 1. The marking set is $G = S = \{(1,0),(0,1)\}$. Suppose that at least one of the clock-setting distributions for e_1 and e_2 has infinite mean. Set $T_0 = 0$ and $T_n = \inf\{t > T_{n-1} : X(t) = (1,0)\}$, so that T_n is the nth hitting time for marking $s = (1,0)$. Then clearly each T_n is a.s. finite, but $E_\mu[T_n] = \infty$ for $n > 0$. Each hitting time for marking $(0,1)$ is also a.s. finite with infinite mean.
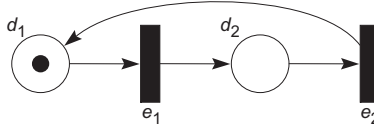

Figure 1: SPN with infinite hitting times.

The foregoing example shows that the hitting times to a marking may be infinite when $\eta(H) = 1$. To see that hitting times need not be infinite, define a transition e to be *uninfluential* if $e \in E(s)$ and $p(s;s,e) = 1$ for $s \in G$. The firing of an uninfluential transition e does not change the marking of the SPN and does not cause the scheduling or cancellation of any transition other than e. Observe that an SPN having one or more uninfluential transitions behaves essentially identically to an SPN in which the uninfluential transitions are not present. More precisely, the marking process of the original SPN can be mimicked—in the sense of Section 2.3—by the marking process of an SPN in which the uninfluential transitions have been dropped. If the original SPN satisfies Assumption PD(0) and only uninfluential transitions have infinite means, then the mimicking SPN satisfies Assumption PD(1) and the conclusion of Proposition 2 holds. Moreover, the transience and recurrence behavior of markings in the two SPNs is identical. Less trivial examples can be constructed along the same lines.

We now consider the question of recurrence in an SPN with simple transitions for which $\eta(H) = 1$. For definiteness and ease of exposition, we assume that $e_1 \in H$ and $e_2, e_3, \ldots, e_m \notin H$. We strengthen Assumption PD(0) by requiring that the clock-setting distributions for transitions $e_2, e_3, \ldots, e_m$ have hazard rates that are bounded from above and below. (Recall that such clock-setting distributions have density functions that are positive on $(0,\infty)$ and have finite moments of all orders.)

Theorem 3 Suppose that Assumption PD(0) holds for an SPN. Also suppose that all transitions are simple, that $e_1 \in H$, and that, for $2 \leq i \leq m$, the clock-setting distribution function $F(\cdot;e_i)$ has a hazard rate $h(\cdot;e_i)$ with $0 < \underline{h}_i \leq h(t;e_i) \leq \overline{h}_i < \infty$ for $t \geq 0$. Then each marking s of the SPN is recurrent.

To prove this result, we use the hazard-rate construction throughout, with the slight modification that the heavy-tailed transition e_1 is set according to $F(\cdot;e_1)$ and runs down at unit rate. We assume without loss of generality that all speeds are equal to 1. We also assume throughout that there are no "single states" s^* such that $E(s^*) = \{e_1\}$. This latter assumption also entails no loss of generality. If the SPN contains such a marking s^*, we can construct a modified SPN $\{X^*(t) : t \geq 0\}$ such that (1) s^* does not belong to the marking set of the modified SPN and (2) each marking in $\{X(t) : t \geq 0\}$ is recurrent if each marking in $\{X^*(t) : t \geq 0\}$ is recurrent. In the modified SPN, the new-marking probabilities are altered so that $p^*(s';s,e) = p(s';s,e) + p(s^*;s,e)p(s';s^*,e_1)$ for all s', s, and e. It is as if we simply take sample paths of the original SPN and delete all intervals in which the SPN is in marking s^*. Finally, we give the proof under the assumption that all transitions are timed; extension to the general case is straightforward.

The proof (merely sketched here) makes repeated use of a geometric trials recurrence criterion: if $\{\mathscr{F}_n\}$ is an increasing sequence of σ-fields and $\{A_n\}$ a sequence of events such that $A_n \in \mathscr{F}_n$, then $P\{A_n \text{ i.o.}\} = 1$ if $\inf_n P\{A_n \mid \mathscr{F}_{n-1}\} \geq \delta$ for some $\delta > 0$; see p. 88 in Haas (2002) or Corollary 2.3 in Hall and Heyde

(1980). The idea is to choose the $\{A_n\}$ sequence so that the recurrence of these events implies the recurrence of the SPN markings. Specifically, a "trial" begins at a time at which the clock for e_1 is set to a value in an interval $[\underline{u}, \overline{u}]$. The trial is a "success"—i.e., an event A_n occurs—if (a) while the e_1 clock runs down, the firing of transitions in $E \setminus \{e_1\}$ causes the SPN to visit all markings reachable from the initial marking, say, s, and end up in a specified marking s', all within $\underline{u}$ time units, and (b) upon arriving in s' the clock for each transition $e \in E(s') \setminus \{e_1\}$ takes at least $\overline{u}$ time units to run down, guaranteeing that e_1 fires in marking s'. The reason for the requirement in (b) is that not all markings in G may be reachable from s via firings of transitions in $E \setminus \{e_1\}$; when this is the case, the marking s' is chosen such that the firing of e_1 in s' will cause the marking process to jump to a previously inaccessible part of the marking set (with positive probability). An additional complication is that e_1 may be canceled and rescheduled one or more times during the excursion described above. For a trial to be a "success" we therefore also require that (c) the clock for e_1 be set to a value in $[\underline{u}, \overline{u}]$ at each such rescheduling; this requirement guarantees that there will be no firings of e_1 to interrupt the excursion through the markings driven by the transitions in $E \setminus \{e_1\}$ and that, when the SPN arrives in s', transition e_1 is still guaranteed to trigger the next marking change. The boundedness of the hazard rates for the transitions in $E \setminus \{e_1\}$ allows us to bound the probability of a "success" away from 0 uniformly in n, so that there are infinitely many successes with probability 1 by the geometric trials lemma.

Space limitations prohibit a detailed description of the bounding arguments, so we content ourselves with giving a small example of how to bound the probability of several simple events. The boundedness of the hazard rates, the memoryless property of the exponential distribution, and Lemma 3.4.10 in Haas (2002) imply that

$$P\{\tau_{n,i} > x \mid \mathscr{F}_n\} = P\left\{\int_{\zeta_n}^{x} h_i(y - \zeta_{\alpha(n,i)})\,dy < C_{n,i} \mid \mathscr{F}_n\right\} \geq P\{(x - \zeta_n)\overline{h}_i < C_{n,i} \mid \mathscr{F}_n\}$$
$$= \exp\left(-\overline{h}_i(x - \zeta_n)\right) \text{ a.s..}$$

Similarly,

$$P\{\tau_{n,i} \leq x \mid \mathscr{F}_n\} = P\left\{\int_{\zeta_n}^{x} h_i(y - \zeta_{\alpha(n,i)})\,dy > C_{n,i} \mid \mathscr{F}_n\right\} \geq P\{(x - \zeta_n)\underline{h}_i > C_{n,i} \mid \mathscr{F}_n\}$$
$$= 1 - \exp\left(-\underline{h}_i(x - \zeta_n)\right) \text{ a.s..}$$

Continuing in this fashion, we can eventually bound the probabilities of the complex events $\{A_n : n \geq 0\}$ uniformly away from 0, so that we can apply the geometric trials criterion.

Even though each marking is recurrent under the conditions of Theorem 3, it is easy to construct examples where the expected number of transition firings between successive visits to a given marking is infinite. (E.g., consider the SPN presented in the next section, but where $e_1 \in H$, transitions e_2 and e_3 have clock-setting distributions with bounded hazard rates, and $s = (2, 1, 1)$ is the specified marking.) Thus when $\eta(H) = 1$ one can only establish a form of "null" recurrence.

5   TWO OR MORE HEAVY-TAILED TRANSITIONS

We continue to assume that all transitions are simple, but now suppose that $\eta(H) \geq 2$. Arguments almost identical to those in the previous section show that the hitting times for a specified marking s may or may not have finite means. The key result of the current section, which we establish by means of an example, is that an SPN can have transient markings when $\eta(H) \geq 2$. It may seem surprising that, in the presence of a finite marking set and irreducibility, some sort of additional moment condition appears necessary to ensure recurrence.

Consider the SPN with three places and three timed transitions shown in Figure 2. The marking set for this SPN is $G = S = \{(2, 1, 1), (1, 2, 1), (1, 1, 2)\}$, so that the set of enabled transitions is given by $E(s) = E$ for $s \in G$. The new-marking probabilities are given by $p((2, 1, 1); s, e_1) = 1$ for $s \in G$,

$p\big((1,2,1);(2,1,1),e_2\big) = p\big((1,1,2);(1,2,1),e_2\big) = p\big((1,1,2);(1,1,2),e_2\big) = 1$, and $p\big((1,1,2);s,e_3\big) = 1$ for $s \in G$.

The marking-change mechanism for this SPN is "hidden" inside the definition of the new-marking probabilities. We can, however, construct an SPN having deterministic timed and immediate transitions—as defined in Section 2.1—that mimics the SPN in Figure 2 and for which the marking-change mechanism is completely apparent from the SPN graph; such a mimicking SPN is displayed in Figure 3. Our results therefore apply to SPN formalisms in which only deterministic transitions are allowed.
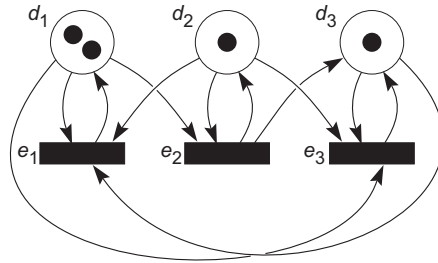

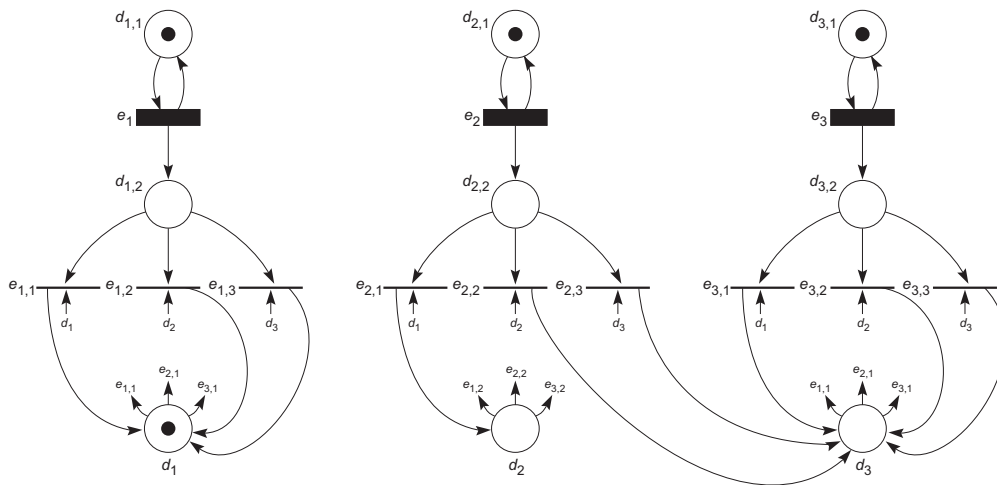Figure 2: SPN with a transient marking.

Figure 3: Alternative representation of SPN with a transient marking.

We use the SPN of Figure 2 in the current paper because of the simpler graph structure. To ease notation, we denote the markings $(2,1,1)$, $(1,2,1)$, and $(1,1,2)$ as markings 1, 2, and 3, respectively. With this notation, the state-transition diagram for the SPN can be represented as in Figure 4.

For each transition e_i, the clock-setting distribution function is of the form $F(\,\cdot\,;s',e_i,s,E^*) \equiv F_i(\,\cdot\,)$. In particular, $F_1(t) = 1 - (1+t)^{-\alpha}$ and $F_2(t) = 1 - (1+t)^{-\beta}$ for $t \geq 0$, and $F_3(t) = t/a$ for $t \in [0,a]$, where $a \in (0,\infty)$ and $\alpha, \beta \in (0,1)$ with $\beta > 1/2$ and $\alpha + \beta < 1$. Denote by f_i the density function of F_i and observe that f_1, f_2, and f_3 are positive and continuous on $(0,a)$. If each clock-setting distribution had a finite mean, then Assumption PD(1) would hold and each marking of the SPN would be recurrent by Proposition 2. The clock-setting distributions F_1 and F_2 have infinite mean, however, and we show in the following that marking 2 is transient.

Observe that the only way in which the SPN can hit marking 2 is if transition e_1 fires and then transition e_2 fires without an intervening firing of transition e_3. That is, if T_n denotes the nth time at which transition e_1 fires and E_n denotes the first transition to fire after T_n, then marking 2 is recurrent if and only if
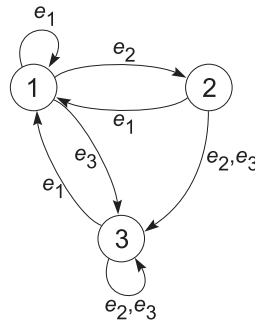
Figure 4: State-transition diagram for SPN with a transient marking.

$P_\mu \{ E_n = e_2 \text{ i.o.} \} = 1$. Roughly speaking, the clock readings for e_1 and e_2 must both be simultaneously close to 0 infinitely often with probability 1. The following result shows that this condition does not hold.

Theorem 4 Under the above assumptions on the clock-setting distribution functions, $P_\mu \{ E_n = e_2 \text{ i.o.} \} = 0$.

In applications, this result (or a slight variation of it) might correspond to a starvation scenario in which, e.g., resources are used by two processes with heavy-tailed durations. The resources are available for use by a third process—which periodically submits a resource request—only during short intervals starting at successive times when both heavy-tailed processes become simultaneously idle. The theorem implies that, with probability 1, the third process will only be granted the resource a finite number of times, since there will only be a finite number of times at which the heavy-tailed processes complete within a short time span of each other. Alternatively, the result might correspond to over-provisioning. For example, suppose that a resource is permanently dedicated to handling the simultaneous or near-simultaneous occurrence of two events, each of which is recurring but with a heavy-tailed distribution for the inter-event times. (One event might be a major forest fire and another event a magnitude 6.0 earthquake.) The theorem implies that with probability 1 such events will occur jointly at most a finite number of times.

Proof (sketch).   Observe that, for each transition e_i, the successive firing times for e_i form a renewal process; the three renewal processes thus defined are mutually independent. For $i = 1, 2, 3$, denote by $C_i(t)$ the clock reading for transition e_i at time $t \geq 0$. The random variable $C_i(t)$ is the residual life at time t for the renewal process associated with the successive firings of e_i. Setting $B_n = \{ C_2(T_n) \leq C_3(T_n) \}$, we have

$$P_\mu \{ E_n = e_2 \text{ i.o.} \} = P_\mu \{ C_2(T_n) \leq \min\big(C_1(T_n), C_3(T_n)\big) \text{ i.o.} \} \leq P_\mu \{ B_n \text{ i.o.} \}. \tag{3}$$

Conditioning on T_n and $C_3(T_n)$ and exploiting the independence of the various renewal processes yields the representation

$$P_\mu \{ B_n \} = \int_0^\infty \left(\int_0^a P_\mu \{ C_2(t) \leq y \} \, dG_t(y) \right) f_1^{*n}(t) \, dt \leq \int_0^\infty P_\mu \{ C_2(t) \leq a \} f_1^{*n}(t) \, dt, \tag{4}$$

where G_t is the distribution function of $C_3(t)$ and f_1^{*n} is the n-fold convolution of f_1 (and hence the density function of T_n). Denote by u_1 the renewal density function that corresponds to F_1; see, e.g., p. 147–148 in Asmussen (2003). Using the well known representation $u_1(t) = \sum_{n=1}^\infty f_1^{*n}(t)$, we find that

$$\sum_{n=1}^\infty P_\mu \{ B_n \} \leq \sum_{n=1}^\infty \int_0^\infty h(t) f_1^{*n}(t) \, dt = \int_0^\infty h(t) u_1(t) \, dt, \tag{5}$$

where $h(t) = P_\mu \{ C_2(t) \leq a \}$. The non-negativity of the terms in (5) justifies the interchange of summation and integration. A rather lengthy argument based on renewal theory for distributions with regularly-varying tails shows that $h(t) = O(t^{\beta - 1})$ and that $u_1(t) < t^{\alpha + \varepsilon - 1}$ for sufficiently large t, where $\varepsilon > 0$ is any constant such that $\alpha + \beta + \varepsilon < 1$. Thus $h(t) u_1(t) = O(t^\gamma)$, where $\gamma = \alpha + \beta + \varepsilon - 2 < -1$, and the rightmost integral in (5) is finite. The theorem now follows from (3), (5), and the Borel-Cantelli lemma. $\qquad\square$

Table 1: Summary of recurrence results for SPNs with simple events.

$\eta(H)$	Recurrent?	Positive Recurrent?	Effective Dimension
0	Yes	Yes	0
1	Yes?	Maybe	1 or 2
≥ 2	Maybe	Maybe	≥ 2

Note that an SPN with $\eta(H) \geq 2$ need not have transient markings. For example, if only uninfluential transitions have infinite means, then Proposition 2 implies that every marking s is recurrent; see Section 4.

6 SUMMARY AND CONCLUSION

Table 1 summarizes our results. The finite-mean requirement in Proposition 2 ensures that each marking s is recurrent, and indeed "positive recurrent" in the sense of the proposition. Theorem 3 asserts that recurrence (though not positive recurrence) is still ensured if at most one heavy-tailed transition can be enabled at any time point. (We imposed a bounded-hazard-rate assumption to facilitate the proof, but we conjecture that the conclusion holds in the absence of this assumption, hence the "?" in the second row of the table.) The results in Section 5 show that the requirement of at most one enabled heavy-tailed transition is "almost necessary" in that an SPN can have transient markings when the requirement is relaxed. A simple necessary moment condition for recurrence appears elusive, as does a simple sufficient condition weaker than those in Proposition 2 and Theorem 3. We expect that obtaining weaker conditions for recurrence would involve analysis of the detailed structure of the SPN under consideration. Indeed, SPNs contain networks of queues as special cases, and recurrence theory for such networks is quite intricate (Bramson 2008).

The rightmost column of the table illustrates an analogy to the theory of random walks on the d-dimensional integer lattice. As is well known (Chung and Fuchs 1951), the origin is positive recurrent (trivially) when $d = 0$, null recurrent when $d = 1$ or $d = 2$, and transient when $d \geq 3$. Our transience example centers around a requirement that multiple clock readings be "small" simultaneously, which roughly corresponds to a multi-dimensional random walk being close to the origin. A heavy-tailed clock-setting distribution can be viewed as analogous to the distribution of the time required for a 1- or 2-dimensional random walk to return to the origin. Thus the case $\eta(H) = 1$ roughly corresponds to a random walk situation in 1- or 2-dimensional space. The corresponding dimensions for the other cases are displayed in the rows above and below.

Ultimately, the results in this paper serve to illustrate the rich and complex behavior that can occur in non-Markovian SPN's, especially those with the sort of heavy-tailed clock-setting distributions that arise in financial, insurance, internet-traffic, reliability, and geophysical modeling. Much work remains to be done in gaining a fundamental understanding of this class of stochastic models. The behavioral complexity of SPN models described here also highlights the importance of simulation as a tool for studying such models in the context of practical system design and decision-making.

REFERENCES

Asmussen, S. 2003. *Applied Probability and Queues*. Second ed.

Balbo, G., and G. Chiola. 1989. "Stochastic Petri Net Simulation". In *1989*, 266–276. New York, NY: ACM Press.

Benson, D. A., R. Schumer, and M. M. Meerschaert. 2007. "Recurrence of extreme events with power-law interarrival times". *Geophys. Res. Lett.* 34:L16404.

Bramson, M. 2008. *Stability of queueing networks*, Volume 1950 of *Lecture Notes in Mathematics*. Springer-Verlag.

Carlson, J. M., and J. Doyle. 1999. "Highly optimized tolerance: A mechanism for power laws in designed systems". *Phys. Rev. E* 60 (2): 1412–1427.

Chiola, G. 1991. "Simulation framework for timed and stochastic Petri nets". 1:153–168.

Chiola, G., and A. Ferscha. 1993. "Distributed Simulation of Petri Nets". *Parallel Distributed Techn.* 1 (3): 33–50.

Chiola, G., G. Franceschinis, R. Gaeta, and M. Ribaudo. 1995. "GreatSPN 1.7: Graphical Editor and Analyzer for Timed and Stochastic Petri Nets". *Perform. Eval.* 24 (1-2): 47–68.

Chung, K. L., and W. H. J. Fuchs. 1951. "On the distribution of values of sums of random variables". *Mem. Amer. Math. Soc.* 6:1–12.

Gaeta, R., and G. Chiola. 1995. "Efficient Simulation of SWN Models". In *Sixth*, 137–146.

Glasserman, P. 1991. *Gradient Estimation via Perturbation Analysis*.

Glynn, P. W. 1989. "A GSMP Formalism for Discrete Event Systems". 77:14–23.

Glynn, P. W., and P. J. Haas. 2006. "A law of large numbers and functional central limit theorem for generalized semi-Markov Processes". 22:201–231.

Glynn, P. W., and D. L. Iglehart. 1990. "Simulation Output Analysis using Standardized Time Series". 15:1–16.

Haas, P. J. 2002. *Stochastic Petri Nets: Modelling, Stability, Simulation*.

Haas, P. J., and G. S. Shedler. 1985. "Regenerative Simulation of Stochastic Petri Nets". 14–23.

Haas, P. J., and G. S. Shedler. 1989. "Stochastic Petri Nets with Timed and Immediate Transitions". 5:563–600.

Haas, P. J., and G. S. Shedler. 1991. "Stochastic Petri Nets: Modeling Power and Limit Theorems". 5:477–498.

Hall, P., and C. C. Heyde. 1980. *Martingale Limit Theory and its Application*.

Holmes, T., J. R.J. Huggett, and A. Westerling. 2008. "Statistical Analysis of Large Wildfires". In *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species*, edited by T. P. Holmes, J. P. Prestemon, and K. L. Abt. Springer.

Moscadelli, M. 2004. "The modelling of operational risk: experience with the analysis of the data collected by the Basel committee". Technical Report 517, Banca d'Italia.

Powers, M. R. 2010. "Infinite-Mean Losses: Insurance's "Dread Disease"". *J. Risk Finance* 11 (2): 125–128.

Prisgrove, L. A., and G. S. Shedler. 1986. "Symmetric Stochastic Petri Nets". *IBM Journal of Research and Development* 30:278–293.

Resnick, S., and H. Rootzén. 2000. "Self-similar communication models and very heavy tails". *Ann. Appl. Probab.* 10:753–778.

Schruben, L. 1983. "Simulation Modeling with Event Graphs". *Commun. ACM* 26 (11): 957–963.

Shedler, G. S. 1993. *Regenerative Stochastic Simulation*.

AUTHOR BIOGRAPHIES

PETER W. GLYNN is the Thomas Ford Professor of Engineering in the Department of Management Science and Engineering at Stanford University. His email address is glynn@stanford.edu and his web page is http://www.stanford.edu/~glynn.

PETER J. HAAS is a Research Staff Member at the IBM Almaden Research Center. His email address is phaas@us.ibm.com and his web page is http://www.almaden.ibm.com/cs/people/peterh.