LAST COPY
NJG.

Information Processing Letters 32 (1989) 7–13
North-Holland

3 July 1989

PATH INTEGRAL METHODS FOR COMPUTER PERFORMANCE ANALYSIS

Neil GUNTHER

Computer Science Laboratory, Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, U.S.A.

Communicated by G.R. Andrews
Received 18 July 1988
Revised 23 February 1989

The Feynman path integral is presented for analyzing computer performance models which exhibit critical behavior due to the presence of unstable states. We examine previously ignored deterministic "jump" solutions associated with instability transients. Using these solutions, an explicit expression for the lifetime of an induced metastable state is derived for a queueing model of virtual memory.

Keywords: Bistability, computer networks, Feynman–Kac theorem, large transients, metastability, path integral, performance modeling, queueing theory, system control, virtual memory

1. Introduction

The role of instabilities in the performance degradation of a broad range of computer systems has been recognized for a long time [3,4,15,21] but has proven difficult to express in a mathematically tractable way [11,18,19,21,26]. A chief cause of such difficulties is accomodating the *global* change in system behavior which arises as a consequence of *local* state fluctuations. Under heavy demand, the system may become unstable to local fluctuations such that the performance collapses spontaneously. In multiprogrammed computers with virtual memory this dramatic degradation in performance is known affectionately as "thrashing" [4] and manifests itself as a sudden escalation in the average response time of the system.

In the Courtois model of virtual memory [2,3], the finite size of main memory is reflected in a finite-population multiprogram queue. Additional programs or jobs are kept outside the multiprogram queue in a stable arrival state. If, however, the programming load becomes inordinately large, the system can become unstable to small fluctuations and may spontaneously "jump" into a new stable regime (the thrashing state). The transition between these two regimes corresponds to large transients and is essentially intractable using conventional performance methods.

Similar transition effects are familiar in another discipline: statistical physics. We revisit the bistable Courtois model and examine it using a mathematical formalism borrowed from statistical physics; the imaginary-time ($t \to -it$) Feynman path integral [8] or statistical density matrix [7]. Since the path integral has proven successful in understanding physical systems which exhibit unstable states [20] it is therefore natural to ask whether this method might find similar application in the context of performance models which entail unstable queueing states. The purpose of this note is to outline how such a program can be carried out. Although we concentrate on just one example as an application of the path integral method, the fact that similar drift functions have been shown to belong to other computer systems, e.g., buffer flow-controlled networks [19,26], multi-access networks [19,24], is suggestive that the path integral method should be generally applicable to these and other computer performance models [13].

The technique described in the following sections corresponds to a small-variance continuum approximation and in that sense represents another asymptotic method [17] related to WKB [10] matching for queues [16] and the large deviations [5,25] extension of the Laplace method [5,9,20]. Each of these methods requires a large parameter, $N \to \infty$. In the method of large deviations [11,26] the probability of rare transients occurring after some long time-interval τ is distributed asymptotically as $\exp(-W\tau)$ with transition frequency, $W \approx A \exp(-nB)$. As we shall demonstrate using the path integral, determination of the constants A and B is associated with the lifetime of metastable states through the choice of system control parameters.

2. Path integral formalism

We seek to express the dynamics of arbitrary state transitions with mathematical consistency from the outset. To this end we introduce the definition of the path integral [7,8,20,21], and some relevant identities. Our interest is in the transient behavior of queueing systems, i.e., the system in a state x at time t undergoing a transition to a new state x' at time t'. There is an infinite number of intermediate states through which the final state can be reached. Consequently, the transition can be thought of as a set of possible trajectories in state-space subject to some function which influences them according to system characteristics: the "potential" function. Each trajectory or path has a certain probability weighting. We express this more formally by introducing the statistical density matrix [7] denoted here $\rho(x', t' \mid x, t)$. If Q is a self-adjoint time-development operator for a real-valued state space such that the stationary states $u_n(x)$ are determined by the eigenvalue equation

$$Qu_n = C_n u_n, \tag{1}$$

then the density matrix can be defined in spectral form [7,8,25] as

$$\rho(x', t' \mid x, t) = \sum_n u_n(x') u_n(x) \exp - C_n \tau \tag{2}$$

with $\tau = (t' - t)$. The integral $\int \rho(x \mid x)\, \mathrm{d}x \equiv \operatorname{Tr} \rho$ is the partition function from which appropriate statistical physics [8] or queueing theory [18] quantities can be derived. In accord with the Feynman–Kac theorem [20,22,25], the dominant contributions to the density matrix come from the lowest eigenvalue in the limit as $\tau \to \infty$.

An alternative definition of the density matrix can be given using a cost functional $I[x]$, defined in terms of a *deterministic* (rotated-time) Lagrangian function, $L(x, \dot{x}) = -(\frac{1}{2}\dot{x}^2 + V(x))$, where $V(x)$ is some potential function for a point-particle and $\dot{x}$ denotes differentiation with respect to continuous time t. The time-development operator $Q(x, p)$ is related to the Lagrangian function via a Legendre transform, $Q = p\dot{x} - L$ where $p = \partial L / \partial \dot{x}$. The objective or cost functional (also called the action [7,8] or (convergence) rate-function [25]) is

$$I[x] = \int^{\tau} L(x, \dot{x})\, \mathrm{d}t. \tag{3}$$

The density matrix can also be defined in terms of the path integral

$$\rho(x', t' \mid x, t) = \mathcal{N} \int \mathrm{d}[x] \exp\{-I[x]\} \tag{4}$$

over all continuous state transitions or paths between (x, t) and (x', t'). Here $\int \mathrm{d}[x]$ denotes functional integration over paths $x(t)$ obeying the boundary conditions $x(0) = x$ and $x(\tau) = x'$ and $\mathcal{N}$ is an appropriate normalization constant for the measure. The value of $I[x(t)]$ for some path $x(t)$ represents the "cost" for that trajectory. Paths corresponding to large excursions between the initial and final states incur a higher cost, and thus have a lower probability than paths which remain close to *the* deterministic trajectory: the path of *least cost*. Equation (4) is an alternative expression of the probability-theoretic Feynman–Kac formula

$$\rho(x', t' \mid x, t) = E_x\left\{\exp\left[-\int V(x(t))\, \mathrm{d}t\right] \cdot \delta(x(t') - x')\right\}, \tag{5}$$

where $E_x\{\ldots\}$ is the expectation with respect to the Wiener process $x(t)$ for Brownian motion on $\mathbb{R}^d$. The reader is referred to the literature [7,8,9,20,22] for further details regarding variants of the path integral formalism. The key feature, for our discussion, is the way in which the path integral connects *deterministic* time-development with *stochastic* time-development via the functional integral over paths.

In this version of the Feynman path integral (4), all quantities are naively real-valued. In general, a real-valued cost functional with bounded potential $V(x)$ implies stable solutions. We remind the reader that if an imaginary part should develop in the cost functional it would signal the presence of an unstable state since a complex eigenfunction cannot be a true eigenfunction of the self-adjoint operator Q.

3. Continuum stochastic equations

With the path integral defined, we now use it to derive the general solution for time-dependent state transitions in queueing systems. Consider a one-dimensional system described by a continuous-time, continuous-state Markov process with random scalar state variable $x(t) \in [0, N]$. In the continuum limit, Markov states evolve approximately as a conditional probability distribution function $P(x', t' | x, t)$, which satisfies the *forward* Kolmogorov continuity equation [1]

$$\partial_t P = Q(x)P \equiv \tfrac{1}{2}\partial_x^2[\sigma^2(x)P] - \partial_x[f(x)P]. \tag{6}$$

Here ∂_x denotes partial differentiation with respect to x. The probability satisfies the initial condition $P(x, t = t' | x', t) = \delta(x - x')$ and the Neumann boundary condition, $\frac{1}{2}\partial_x[\sigma^2(x)P] - f(x)P = 0$ at $x = 0$ and N. The infinitesimal drift field $f(x) = -\delta_x V(x)$ is defined in terms of the external potential $V(x)$ (Fig. 1). A stochastic process is then characterized by the drift function $f(x)$. The stability points of $V(x)$ correspond to the roots of $f(x)$.

For the purposes of the large transient analysis to be given in the next section, it will prove beneficial to convert (6) to self-adjoint form. This

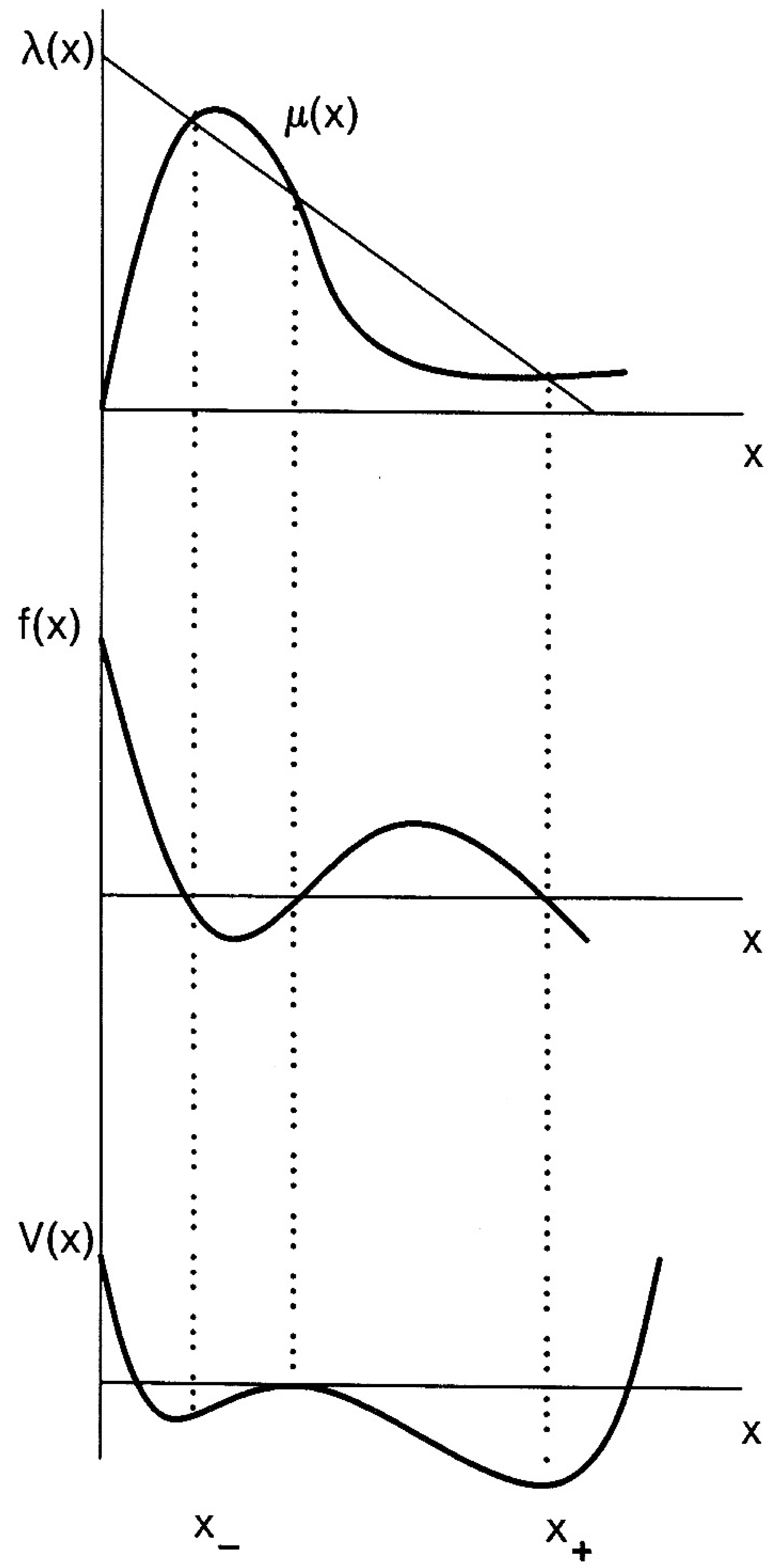

Fig. 1. Schematic representation of queueing rates, drift and potential for the Courtois model.

is conveniently accomplished by transforming to new variables; $dz = dx/\sigma(x)$ and $\Psi(z) = \sigma(x)P(x)$ so that (6) becomes

$$Q\Psi = \tfrac{1}{2}d_z^2\Psi - d_z[F(z)\Psi], \tag{7}$$

which is equivalent to a diffusion equation with unit diffusion coefficient and a transformed drift-field

$$F(z) = [f(z) - \tfrac{1}{4}\partial_z\sigma^2(z)]/\sigma(z). \tag{8}$$

A standard technique (see, e.g., [22]) enables the first-order drift term in (7) to be replaced by an effective potential via a gauge transformation Q^*

$= U^{-1}QU$ and $\rho = U\Psi$ where $U(z) = \exp[-\int F(z)\,dz]$. Then (7) becomes

$$Q^*\rho = \tfrac{1}{2}d_z^2\rho - V_{\text{eff}}(z)\rho \tag{9}$$

with the effective potential $V_{\text{eff}}(z) = \frac{1}{2}[F^2(z) + \partial_z F(z)]$. The Markov process described by the state variable $z(t)$ is equivalent to the dynamics of a single particle in a one-dimensional potential, $V_{\text{eff}}(z)$. This representation with drift-field and state-dependent variance contributions is to be contrasted with the simpler zero-variance or pure-drift approximation that underlies the methods described in [11,26].

Reverting to previous notation, the path integral defined in (1) is a solution of the transformed Kolmogorov equation with objective function

$$I[x(t)] = \int \left\{ \tfrac{1}{2}(\dot{x}/\sigma(x))^2 + V_{\text{eff}}(x) \right\} dt. \tag{10}$$

The form of the Lagrangian in (10) is identical to that described in Section 2 (cf. action functionals in [9,25]). The first term represents the "kinetic energy" of a single Brownian-like particle subject to the potential V_{eff} in (9) with $F(x)$ defined by (8). In the absence of an effective potential the path integral solution simplifies to a Gaussian functional with stationary mean, corresponding to the Wiener process [6,14,25].

4. Virtual memory model

We now turn to a path integral analysis of large transients in the M/M/1/N model of multiprogrammed computers due to Courtois [2]. This closed system comprises N terminals (or programs) connected to a computer subsystem. No terminal can issue more than one request at a time to the computer so that the interarrival time of these requests is exponentially distributed with mean $1/\lambda$ sec. The transient behavior of this multisource system is revealed by modeling it as a finite population, single-server queue [14]. If the subsystem has $x \in [0, N]$ requests enqueued, with $(N - x)$ requests remaining in an arrival state, then the effective arrival into the subsystem is a linear state-dependent function $\lambda(x) = (N - x)\lambda$, and the throughput is modelled as a nonlinear function $\mu(x)$ which subsumes page-faulting and process suspension effects. These rates are depicted schematically in Fig. 1 and we note, in passing, the similarity with rates for certain bistable networks [13,19,24]. Further details about the assumptions underlying the Courtois model of virtual memory can be found in [3,21].

Although queue occupancy is formally a discrete-valued state variable, because of our interest in the limit as $N \to \infty$, it is consistent with our earlier discussion to represent the queue-state distribution by a continuous real-valued variable $x(t)$. Accordingly, we approximate the Markov queueing model by the continuum Kolmogorov equation (6) with infinitesimal drift function defined by

$$f(x) = \lambda(x) - \mu(x), \tag{11}$$

and infinitesimal variance

$$\sigma^2(x) = \lambda(x) + \mu(x). \tag{12}$$

In Section 3 we noted that system stability is primarily determined by the roots of (11). Noting the similarities with WKB [10] to uncover the asymptotics of (4) we can safely assume, in the large-N limit, that both (11) and (12) have weak state-dependence. The variance (12) mainly effects time scales [13] and since we seek only asymptotic estimates of the mean transition rate or first-passage time (FPT) to the new stable point, we further assume that $\sigma^2(x)$ is essentially constant and denote it by μ (the asymptotic service rate). Recognizing that constant variance or coefficient of diffusion depends inversely on the characteristic system size N, $1/\mu$ plays the role of an intrinsic large parameter. Alternatively, one could retain (12) and simply impose an artificial parameter to develop a perturbation expansion about the deterministic drift-dynamics (see, e.g., [9]). Our small-variance approximation has the added benefit of simplifying the equations without severe loss of generality.

The infinitesimal drift function in Fig. 1 can be most simply approximated by a third-degree polynomial possessing, at most, three real positive roots corresponding to local equilibrium of the service and arrival rates. Upon integrating (11) with respect to x, a general polynomial expression for the drift-potential of the Courtois model is

$$V(x)=\tfrac{1}{4}x^4+\tfrac{1}{2}\beta^2x^2+\alpha x, \tag{13}$$

where $\beta^2<0$ implies that the system is in the critical regime exhibiting bistability due to the presence of two local minima in the potential function, and $\alpha=0$ means that the bistable states are degenerate. Only two control parameters are required for a complete description of system stability in agreement with potentials belonging to the class of "cusp" catastrophes [19]. In this potential model, the arbitrary parameter β is identified with the maximal programming level [2,3] or the potential-barrier height while the arbitrary parameter α can be identified with either the terminal-load N or the terminal "think time" $1/\lambda$. As a consequence of the small-variance approximation, the potential defined by (8) and (10) now reduces to

$$V_{\text{eff}}(x)=\left[f^2(x)+\mu\,\partial_x f(x)\right]/2\mu. \tag{14}$$

Since the choice of β, relative to μ in (14), appears arbitrary, $V_{\text{eff}}(x)$ could be expanded as a sixth-degree polynomial having up to three minima. It can be shown [13], however, that when $V(x)$ is expressed in terms of actual model parameters, only two minima are significant. Provided there are at least two local minima, the exact form of the potential is unimportant for our qualitative discussion of metastability. Performing a first variation $\delta I[x]/\delta x=0$, it can be shown that the objective functional has stationary solutions at the values of the local minima, and at the instantaneous "jump" or "instanton" [23] solution given by

$$t=\int^x\left(2V_{\text{eff}}(y)\right)^{-1/2}\,\mathrm{d}y. \tag{15}$$

As depicted in Fig. 2, the width of this step-like transient corresponds to a "hopping time" between wells while the separation between jumps corresponds to the overall transition rate. The value of the cost functional I_0 at the jump is proportional to the barrier height in agreement with the more familiar WKB result [10,20]. The

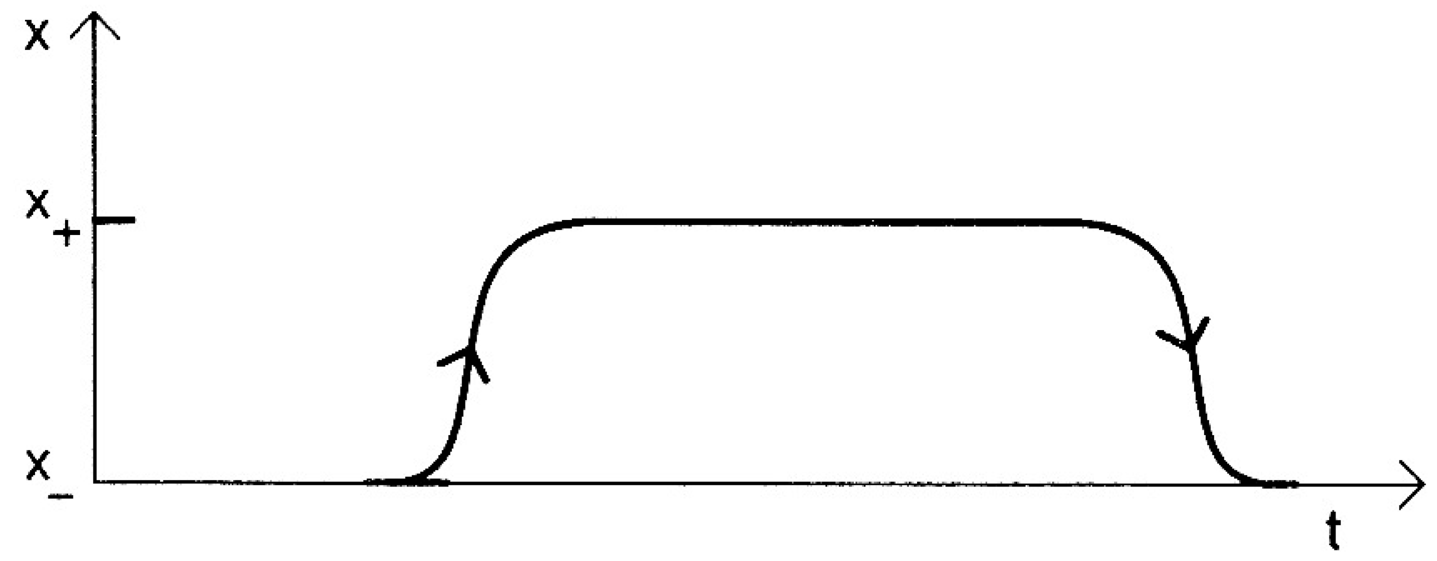

Fig. 2. A jump–anti-jump pair.

jump is the deterministic path of least cost that interpolates between the minima—the escape path by which the probability leaks into the global minimum. Here we see the advantage of working in the continuum approximation; it reveals the nature of the *large scale* fluctuations which are present at the onset of the transition between the bistable states. All that remains is to show how this "jump" solution can be applied to calculate the time before the system leaves the metastable state (i.e., FPT).

5. Lifetime of the metastable state

Permitting the control parameter α to take a small negative value in the potential function (9) lifts the degeneracy between the optimal queue congestion x_- and the degraded queue congestion x_+ as shown in Fig. 1. The previously optimal performance regime now becomes metastable (with respect to the lower minimum) and has a characteristic life expectancy before the system undergoes a complete transition into the degraded regime (new global minimum).

Stationary solutions of the density matrix dominate in the limit $\tau\to\infty$. If the variance corrections are identified with an expansion (up to second order) about the dominant jump solution $x_0(t)$, we may express an arbitrary path as $x(t)=x_0(t)+\Sigma_n a_n x_n$. The x_n are a complete set of real functions vanishing at the boundaries such that $x_n(\tau)=x_n(0)=0$, $\int^\tau \mathrm{d}t\,x_m(t)x_n(t)=\delta_{mn}$. They are also eigensolutions of the second variation of $I[x]$ and we write

$$Mx_n=\left(\delta^2I[x]/\delta x^2\right)x_n=\theta_n x_n, \tag{16}$$

where M is the small variance operator having eigenvalues θ_n. Taking the small-μ limit we find that the path integral (4) becomes a product of ordinary Gauss–Fresnel integrals [8,20],

$$\mathcal{N}\exp(-I[x_0]/\mu)\prod_n\left\{\int(\exp-a_nMa_n)\,\mathrm{d}a_n\right\}\times[1+\mathrm{O}(\mu)] \qquad (17)$$

which can be re-expressed more simply as

$$\mathcal{N}\exp(-I_0/\mu)(\det M)^{-1/2} \qquad (18)$$

where the $\mathrm{O}(\mu)$ notation is now suppressed. More generally, there can be a succession of j jump–anti-jump pairs (Fig. 2) that are well separated in time, and transitions can occur at any instant on the t-axis. These combinatoric considerations require that the correct contribution to the lowest cost C_0 is given by the grand partition function over j odd jump times

$$\exp-C_0\tau \simeq \sum_{j\,(\mathrm{odd})}\exp(-jF_0/\mu)\left\{K\tau(\det M)^{-1/2}\right\}^j/j! \equiv \exp\left\{-K\tau(\det M)^{-1/2}\exp(-F_0/\mu)\right\}, \qquad (19)$$

where the newly introduced constant K is chosen to provide normalization in agreement with $\mathcal{N}$.

The lowest eigenvalue of M is negative, $\theta_0<0$, due to the absence of any turning point on the interval $[x_-, x_+]$. It corresponds to an unstable expansion leading to an infinite separation between the jump and the next anti-jump. This negative eigenvalue gives rise to an imaginary prefactor in the exponent of (19). Recalling our remarks in Section 2, it is the signature of an unstable state. Finally then, we have an expression for the transition rate due to the separation between the jump and anti-jump pairs:

$$K\left(\prod_n\theta_n\right)^{-1/2}\exp\left\{-\int[2V_{\mathrm{eff}}(\alpha,\beta,x)]^{1/2}\,\mathrm{d}x/\mu\right\} \qquad (20).$$

It has the same form as the mean transition frequency or mean FPT estimate [7,21] mentioned in Section 1 with the constants now determined by the value of the control parameters α, β in the effective potential. The inverse of (20) characterizes the average lifetime of the metastable state x_- in Fig. 1. Since this rate is exponentially small in μ, it cannot be seen in a perturbative expansion of the original cost functional. In this way the path integral describes how the transient probability density escapes the metastable equilibrium state (essentially via a unique *deterministic* path) into the global equilibrium state accompanied by a sudden shift in probability mass of the equilibrium distribution.

6. Summary

Analogy with statistical physics has guided our application of the path integral and the path integral has guided our small-variance analysis of large transients that is otherwise unreachable by conventional steady-state performance methods. The analogy also provides another perspective on the complexity of solving such bistable computer systems. It is as difficult as one of the most subtle problems in physics; calculating the transmission probability for a "quantum particle" to tunnel through a potential barrier [12].

We have used a polynomial "mechanical potential" approximation in this paper because it is natural to the Lagrangian in the Feynman formulation of path integrals [7,8] and it also makes contact with aspects of the catastrophe theory approach discussed in [19]. The path integral method presented here also has formal connections with the generalized method of large deviations [9,23,25]. Defining Lagrangians directly in terms of (nonpolynomial) probability-theoretic distributions [11,13,26] avoids arbitrary coefficients like those introduced in Section 4.

Calculational methods, not reliant on the small-variance approximation, are known for the path integral [20] and should be developed for performance analysis. Numerical analysis based on the path integral has been investigated in [13]. The path integral formalism is extensible to higher-dimensional topologies having greater internal degrees of freedom and Bickerstaff has considered an application for a network of processors using process migration as a load-balancing mechanism: a 2-dimensional Courtois network. These and other ideas will be pursued in greater detail elsewhere.

Acknowledgment

It is a pleasure to acknowledge discussions with R. Nelson and helpful remarks from J. Nievergelt and P. Fleming. I also wish to thank the following colleagues at PARC for discussions and encouragement: G. Beretta, J. Gastinel, A. Perlis (Yale), B. Preas, J. Shaw, S. Shenker and M. Weiser, F. Yao. Special thanks to S. Shenker for *critically* reviewing earlier versions of this work and for bringing the papers by Nelson and Weiss to my attention.

I am grateful to the anonymous referees for suggested improvements and additional references [21] and [22].

References

[1] E.G. Coffman Jr., and M.I. Reiman, Diffusion approximation for computer and communication systems, In: G. Iazeolla, P.J. Courtois and A. Hordijk, eds., *Mathematical Computer Performance and Reliability* (North-Holland, Amsterdam, 1984) 33–53.

[2] P.J. Courtois, Decomposability, instabilities, and saturation in multiprogramming systems, *Comm. ACM* **18** (1975) 371.

[3] P.J. Courtois, *Decomposability: Queueing and Computer System Applications* (Academic Press, New York, 1977).

[4] P.J. Denning, Thrashing: Its causes and prevention, *Proc. AFIPS Fall JCC* **33** (1968) 915.

[5] R.S. Ellis, *Entropy, Large Deviations and Statistical Mechanics* (Springer, New York, 1985).

[6] W. Feller, *Introduction to Probability Theory and Its Applications, Vol. I* (Wiley, New York, 1971).

[7] R.P. Feynman, *Statistical Mechanics* (W.A. Benjamin Inc., New York, 1981).

[8] R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[9] M.I. Freidlin and A.D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1984).

[10] N. Froman and P.O. Froman, *JWKB Approximation* (North-Holland, Amsterdam, 1965).

[11] A. Greenberg and A. Weiss, *An Analysis of ALOHA Systems via Large Deviations*, AT&T Bell Labs. preprint, July 1986.

[12] N.J. Gunther, A new interpretation of instabilities in computer performance models, In: *Proc. ACM SIGMETRICS Conf. on Measurement and Modeling of Computer Systems*, Santa Fe, NM (May 1988) 276.

[13] N.J. Gunther and J.G. Shaw, Path integral evaluation of ALOHA network transients, In preparation.

[14] L. Kleinrock, *Queueing Systems, Vol. 1: Computer Applications* (Wiley, New York, 1976).

[15] L. Kleinrock and S. Lam, Packet switching in a multiaccess broadcast channel: Performance evaluation, *IEEE Trans. Comm.* **COM-23** (1975) 410.

[16] C. Knessl, B.J. Matkowsky, Z. Schuss, and C. Tier, Two repairmen problem: A finite source M/G/2 queue, *SIAM J. Appl. Math.* **47** (1987) 367.

[17] J. McKenna, D. Mitra, and K.G. Ramakrishnan, A class of closed Markovian queueing networks: Integral representations, asymptotic expansions, generalizations, *Bell Syst. Tech. J.* **60** (1981) 599.

[18] R. Nelson, The stochastic cusp, swallowtail, and hyperbolic umbilic catastrophes as manifest in a simple communications model, In: *Performance 84* (North-Holland, Amsterdam, 1984) 207–224.

[19] R. Nelson, Stochastic catastrophe theory in computer performance modeling, *J. ACM* **34** (1987) 661.

[20] L.S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).

[21] P. Semal, and P.J. Courtois, Stability analysis of large Markov chains, In: *Performance 87* (North-Holland, Amsterdam, 1988) 363–382.

[22] B. Simon, *Functional Integration and Quantum Physics* (Academic Press, New York, 1979).

[23] B. Simon, Instantons, double wells and large deviations, *Bull. Amer. Math. Soc.* **8** (1983) 323.

[24] A. Tanenbaum, *Computer Networks* (Prentice-Hall, Englewood Cliffs, NJ, 1981).

[25] S.R.S. Varadhan, *Large Deviations And Applications* (SIAM, Philadelphia, PA, 1984).

[26] A. Weiss, A new technique for analyzing large traffic systems, *Adv. Appl. Prob.* **18** (1986) 506.