Transition Matrix Methods

David Yevick (1) and Michael Reimer (2)

1) Department of Physics, University of Waterloo, Waterloo, ON N2L 3G1, Canada, yevick@uwaterloo.ca 2) Department of Physics, University of Waterloo, Waterloo, ON N2L 3G1, Canada, mareimer@sympatico.ca

Abstract: The transition matrix describes the probabilities of transitions of a system between different sets of states. In this presentation, we demonstrate that techniques similar to the multicanonical method can be employed to find the elements of the transition matrix between pairs of highly unlikely states in an optical communication system. This result is then applied to determine the probability distribution of outage times resulting from polarization mode delay (PMD).

Introduction
While biased sampling procedures, most notably the multicanonical or importance sampling techniques, are now frequently employed to investigate static system quantities such as the probability distribution functions (pdf) of the differential group delay (DGD) or of the eye opening penalty,[1-4], these methods have not been applied to dynamic system behavior such as the average time spent in an outage. Here we adapt the transition matrix methods of statistical mechanics[5-7] to dynamic communications problems (it should be noted that a transition matrix approach was earlier applied to analyze the error of static PMD calculations in [8]).

Multicanonical Method
Since our transition matrix discussion generalizes our previous work on biased sampling methods, we first review the foundations of the multicanonical method. A randomly fluctuating physical system can generally be described by a vector of $N_s$ system observables, $E$, each of which is affected by $N_a$ stochastically varying parameters $a$. For example, in this paper $E$ are suitably defined moments of a pulse that has passed through a series of polarization controllers to which fluctuating voltages specified by $a$ are applied. Typically, the average value of a function $p(E)$ in regions of sample space with an extremely small probability of occurrence must be determined without recourse to an unrealistically large number of samples, $a$.

Accordingly, the multicanonical method biases the sample space. In such a manner that, if $p(E)$ represents the pdf of $E$, the sampling probability after many iterations coincides with the inverse of the pdf and the sampling becomes uniform in the system observables $E$. In the standard implementation, the physically accessible region of $E$ is partitioned into $N$ histogram bins centered at $E_m$, with $m = 1, 2, \ldots, N$. Next, the elements of a histogram that will store the unnormalized pdf, $p(E_m)$, and a second histogram that holds temporary values, $H(E_m)$, are both initialized to unity. The starting value of $E$ is computed from a randomly chosen set of system variables $a_{\text{cur}}$. These system variables are then randomly perturbed according to $a_{\text{new}} = a_{\text{cur}} + \delta a$, leading to new observables $E_{\text{new}}$. To ensure that the computation spends more time in states with small values of $p(E)$, this transition is accepted with probability $\min(1, p(E_{\text{cur}})/p(E_{\text{new}}))$, in which case $a_{\text{cur}}$ is set to $a_{\text{new}}$, otherwise $a_{\text{cur}}$ is unchanged for the following iteration. In either case, the histogram element, $H(E_m)$, for the new value of $E_{\text{cur}}$ is incremented by unity. After this process is repeated $M$ times, an improved estimate, $p'(E)$, of $p(E)$ is obtained after the bias in the system variables during the iteration is removed according to $p'(E) = cp(E)/H(E)$. In the following iteration $p'(E)$ replaces $p'(E)$, $H(E)$ is reset to unity and the above process is repeated.

Transition Matrices
The multicanonical method can be employed to determine the relative probability for transitions between each pair of system variables even for highly unlikely pairs of states. In particular, for every accepted or rejected transition from a state in the $n$th to a state in the $m$th histogram bin obtained during a multicanonical calculation, the corresponding element, $T_{mn}$, of an unbiased but unnormalized transition matrix, $T$, is incremented by unity. At any point in the calculation, the rows of $T_{nm}$ can be normalized to unity, yielding an estimate of the standard transition matrix $T$. We can also construct a biased multicanonical transition matrix, $B$, such that $B_{nm}$ gives the likelihood of a transition from the $n$th to the $m$th state in the multicanonical procedure. Here $T_{nm}$, typically at the end of the last iteration, is multiplied by the multicanonical acceptance probability, $\min(1, p'(E_m)/p'(E_n))$ after which each row of this matrix is normalized to unity. A further matrix $R$, with $R_{nm} = T_{nm}/T_{mm}$, can be constructed that specifies the ratio of the probability of an unbiased transition from state $n$ to state $m$ to the probability of this transition transition probability in the multicanonical method.

An alternative method for biasing transitions is to formulate the acceptance rule directly from an estimate of $T$ that is dynamically updated as the calculation steps. Here, starting from initial values $T_{nm} = \text{1, transitions between two states $E_{nm}$ and $E_{cur}$ generated by $a_{new} = a_{cur} + \delta a$, are accepted with probability $\min(1, T_{nm}/T_{mm})$ while again incrementing
generate any desired rotation of the incoming polarization vector $\Omega_i$ on the Poincaré sphere. For certain mutual orientations of these polarization vectors, the ratio of the emulator DGD to the average DGD of the fiber segment will exceed a certain value associated with a system outage. The time interval over which the DGD exceeds this value is termed the outage time.

Identifying the vector of system variables $\mathbf{a}$ with the 2l-dimensional relative angles between the birefringence vectors of successive sections, and the system observables, $E_n$ with the DGD, the outage time distribution is normally determined by performing many time steps $a^{new} = a^{old} + \delta a$, and recording the length of time for which the system remains in the outage region for each outage. However, in our calculations, the Markov chain can be directly modeled by sampling transitions according to the probabilities stored in the unbiased or biased transition matrices (in the latter case, however, the histogram for the time duration of each system path must be incremented by an appropriate weighting factor).

Alternatively, we can take as the initial state of the system a distribution equal to $P(E_n)$ for $n$ in the non-outage region and zero elsewhere, corresponding to the average distribution of states that have not yet entered the outage region. The time evolution of this state is then modeled by repeatedly multiplying by the transition matrix. After the $k$th step, the sum of the histogram values that fall outside the outage region can be identified with the probability that the system returns to a non-outage state within the time interval $(k-1)\Delta t < t < k\Delta t$. The values of the system vector in the non-outage region are then set to zero before the $(k+1)$th step time. The time evolution can also be determined by noting that since the state vector of paths that are restricted to the outage region is repeatedly multiplied by the submatrix formed from the matrix entries for the outage states, the eigenvalues of this submatrix yield the decay rates corresponding to the distributions given by the associated eigenvectors. The desired outage time distributions can therefore be found through projection of the distribution after one time step onto these states.

Numerical Results
To illustrate the above methods, in Fig. 1 we show the pdf of the outage times, for both DGD>$2\tau_{avg}$ and DGD>$3\tau_{avg}$ outage conditions. The circles in this figure refer to an unbiased Markov chain calculation that recorded the outage times observed during $10^9$ time steps (fewer than 10,000 samples were recorded in the outage region for the $3\tau_{avg}$ outage condition). The solid line displays the result after constructing an unbiased transition matrix from the accepted and rejected transitions of three $5\times 10^6$ sample multican-
Onical iterations and then generating a biased Markov chain from the the multicanonical acceptance rule together with appropriate weight factors for the paths. Replacing the multicanonical acceptance rule by a rule derived from the ratio of transition matrix elements was found to decrease the sampling of the tail of the pdf for a fixed number of realizations. Finally, successive multiplications of the initial state vector by the unbiased transition matrix obtained after two $2 \times 10^7$ sample multicanonical iterations yielded the dashed line; as was also computed from the eigenvalues and eigenfunctions of the outage state submatrix. Although the biased transition matrix clearly samples a far wider range of states than the standard procedure, some accuracy is necessarily sacrificed since the standard method preserves all information about the system variable correlations in different regions of the system state space.

Finally, Fig. 2 displays the eigenvectors corresponding to the three largest eigenvalues of the transition matrix determined with three $5 \times 10^7$ multicanonical iterations. The inset contains the magnitude of the 50 largest eigenvalues. Note that the eigenvector with unit eigenvalue (solid line) yields the pdf of the GDD for infinitely long fiber samples while the other eigenvectors can be used to compute the DGD at finite fiber lengths.

**Fig. 1.** The pdf of the outage durations associated with an optical fiber emulator for thresholds of $2 \tau_{\text{avg}}$ and $3 \tau_{\text{avg}}$. The results of the standard method with $10^5$ samples, the multicanonical reweighing method with 3 iterations of $5 \times 10^7$ samples and repeated multiplication of the initial state vector by an unbiased transition matrix obtained after two $2 \times 10^7$ sample multicanonical iterations, are denoted by circles, solid and dashed lines, respectively.

**Fig. 2.** The first three (solid, dashed and dashed-dotted lines, respectively) transition matrix eigenvector computed after 3 multicanonical iterations of $5 \times 10^7$ samples. The inset contains the magnitude of the first 50 eigenvalues.

**Conclusions**

That time-dependent effects in communication and other physical systems can be rapidly and accurately modeled, even in very low probability regions by transition matrix methods can significantly impact future communication system simulations. We therefore hope to shortly extend our results to several challenging system calculations.

**Acknowledgments**

This work was supported in part by the National Sciences and Engineering Research Council of Canada (NSERC), the Canadian Institute for Photonic Innovation (CIPI), Nortel and the Ontario Research and Development Challenge Fund (ORDCF).

**References**