

System Analysis, Modelling and Simulation, vol. 18-19, 1995, pp. 717-720

AN UPPER BOUND FOR A TIME STEP IN PARALLEL SPATIALLY EXPLICIT BIOLOGICAL SIMULATIONS*

BOLESŁAW K. SZYMANSKI

Rensselaer Polytechnic Institute, Troy, New York, USA

Abstract A new paradigm is proposed for parallel discrete-event simulation of lattice-based biological models. The paradigm, called periodic synchronization simulation, leads to highly parallel implementations with low communication-to-computation ratios. The periodic synchronization without rollback may introduce some errors in the results, however by proper selection of the synchronization time step such errors can be kept at an acceptable level. The bounds for an optimal synchronization time step are derived in the paper.

INTRODUCTION

Spatial aspects of biological processes are of prime importance for population biology, evolution theory, ecology and epidemiology. Spatially explicit biological simulations are an increasingly important component of high-performance computing¹. Many biological models of multi-species habitats are based on a lattice representation of space. The environment is partitioned into a two-dimensional grid of sites; each site so small that it can support at most one organism. An ecological stencil of a site consists of all the neighboring sites which affect biological processes at this site². In each experiment the size of the stencil and the location of the affected site relative to it are defined by biotic and abiotic factors. Such a space model matches well the plant dynamics and can be also used to simulate territorial animals.

*This material is based upon work supported by the National Science Foundation under Grants ASC-9318184 and BIR-9320264 BIR-9320264. The author is grateful to William Maniatty for implementing different versions of the spatially explicit ecological models which empirically confirmed the presented findings.

B.K. SZYMANSKI

Each site is in one of a few possible states, such as: unoccupied or occupied by an organism in a specific condition (i.e., healthy, infected with a virus or infested by a parasite). Usually, it is assumed that transitions between the site's states are governed by Poisson processes. Intensity of each process is defined by the states of the sites in the corresponding ecological stencil.

A parallel discrete-event simulation of such a model on distributed-memory machines relies on the partitioning of the space among the processors. Hence, stencils of sites near the partition boundaries (referred to as boundary sites) span across two to four processors. The computation of transition times for such sites involves communication. Moreover, such partitioning requires that any event associated with a site near the boundary of a partition must be executed on all processors that hold a part of this site's stencil. This requirement limits the simulation parallelism by tightly synchronizing processing of events on neighboring processors and imposes high communication overhead (time of each event must be communicated to all neighboring processors). We will refer to such an implementation as a continuous synchronization simulation (abbreviated as CSS).

To overcome this limitation, we propose to communicate the events happening to the boundary sites with some prescribed frequency; i.e., every t_s time units. As shown below, there is a well defined upper bound for such a time step that limits the error of the simulation and yet maintains high parallelism and small communication-to-computation ratio. We will refer to such an implementation as a periodic synchronization simulation, abbreviated as PSS.

TIME STEP ANALYSIS

Typically, the time unit in the model description is selected in such a way that it corresponds to the frequency with which experimental data are collected in nature. Therefore, the natural criterion for accuracy of the simulation is to require that at each time unit the average number of sites with states generated differently by PSS and CSS is less than one. This criterion imposes the upper bound on the value of the time step.

TIME STEP IN PARALLEL SIMULATIONS

Let N^2 denote the total number of sites in the environment arranged as two-dimensional lattice of size N . Consider a site undergoing a transition S according to a Poisson process with intensity i . Let the time of this transition be t . For a given coefficient α , the event S will be processed at such simulation step m that $(m-1+\alpha)t_s \leq t \leq (m+\alpha)t_s$. Note that if $\alpha = 1/2$ then there will be a trend to overestimate the time of events in the simulation. We want to find such a value of α that the expected difference D of simulated and actual times of the events is equal to zero. By definition:

$$D = \int_0^{\alpha t_s} x i e^{-ix} dx + \sum_{m=1}^{\infty} \int_{(m-1+\alpha)t_s}^{(m+\alpha)t_s} i(x - mt_s) e^{-ix} dx$$

so $D = 1/i - t_s \frac{e^{-it_s \alpha}}{1 - e^{-it_s}}$ and therefore $\alpha = (\ln(it_s) - \ln(1 - e^{-it_s})) / (it_s)$. For large intensities (which are typical in biological systems because the sampling is infrequent), α can be approximated as $\ln(it_s) / (it_s)$.

Consider now an experiment with sampling times $s_1, s_2 \dots$ and a sequence of transitions $S_1, S_2 \dots, S_n$ happening at actual times $t_1, t_2 \dots, t_n$. Assume that these transitions are simulated at steps $m_1, m_2 \dots, m_n$. Therefore, a time of occurrence error D_j is associated with each transition j . The simulation is exact if for every sampling time s_k and a transition S_j if $t_j < s_k$ then $m_j t_s < s_k$ and conversely if $t_j > s_k$ then $(m_j t_s > s_k$. Under the above conditions the same states are reported by the sampling and the simulation for all sites. To analyze the upper bound for a time step t_s , we consider a weaker condition, namely that the simulated and actual times are in accordance at the last step, i.e. that $m_n t_s - 1 < t_n < m_n t_s + 1$, or, equivalently that $-1 < \sum_{j=0}^n D_j < 1$.

Let T_s denote the total simulation time and i be the average intensity of the Poisson processes governing all state transitions. Then, the expected number of transitions that a site will undergo during the simulation is iT_s . It is easy to verify that for D_j the following holds:

$$E[D_j] = 0 \text{ and } E^2[D_j] = \frac{2}{i^2} - \frac{t_s e^{-iat_s} (2 - it_s(1 - 2a))}{i(1 - e^{-it_s})}$$

where $E[\]$ denotes an average value of its argument. For $it_s \ll 1$ the standard deviation σ can be approximated as $\sigma^2 = t_s/i$. Using central limit theorem and Gaussian function approximation for large arguments, the probability that the absolute value of the sum of $n = iT_s$

B.K. SZYMANSKI

independent variables D_j is at least 1 can be approximated as $P_m = \sigma\sqrt{n/(2\pi)}e^{-1/(2n\sigma^2)}$. Hence, from the requirement that $N^2P_m < 1$, we get (approximately):

$$t_s < \frac{1}{4T_s \ln(N)} \quad (1)$$

Equation (1) imposes a restriction on the synchronization frequency of a discrete-time simulation of Poisson processes. From equation (1), the average fraction of sites undergoing transitions in the environment during a single time step t_s is approximately $it_s < i/(4T_s \ln(N))$. In other words, the fraction of sites undergoing transitions in a single simulation time step is decreasing only logarithmically with the increase in the environment size.

Let p denote the number of available processors and g^2 be the number of sites in a stencil. Each processor is allocated N^2/p sites of which $2g(2N/\sqrt{p} - g) \approx 4gN/\sqrt{p}$ are boundary sites. On average each processor executes N^2it_s/p transitions and communicates to its neighbors $4gNit_s/\sqrt{p}$ events that happened at the boundary sites in a single time step t_s . Each transition involves g^2 evaluations of states of the stencil sites. Hence, the communication-to-computation ratio c and synchronization-to-computation ratio s for such a simulation are:

$$c < \frac{4g^3Nit_s/\sqrt{p}}{g^2N^2it_s/p} = \frac{4g\sqrt{p}}{N} \quad s = \frac{p}{g^2N^2it_s} = \frac{4pT_s \ln(N)}{ig^2N^2}. \quad (2)$$

In contrast, in CSS paradigm $s = p/g^2$ because processors synchronize after each transition. Thus, unlike CSS, the periodic synchronization simulation is scalable and efficiently parallelizable on distributed-memory machines.

REFERENCES

1. W. Maniatty, B.K. Szymanski and T. Caraco, in High Performance Computing, edited by A. Tentner (The Society For Computer Simulation, San Diego, 1994), pp. 114–119.
2. B.K. Szymanski and T. Caraco, Evolutionary Ecology, **8**, 299–314 (1994).