Virtual Epidemics - Ecological Modeling on a Parallel Machine

Bill Maniatty [1], Boleslaw Szymanski [2], Tom Caraco [3]

1 Biological Issues

Ecologists have long recognized that the spatial context of ecological interactions is fundamentally important to understanding population dynamics, community stability, and biodiversity [3, 4]. However, the computational power necessary to deal with the volume of data and computations in spatially explicit models has only recently been available to ecologists ([7]). In this paper we discuss the implementation of such spatially explicit models of epidemics on massively parallel computers. We have designed and implemented two basic but ecologically significant models for epidemic: disease transmitted directly and vector-borne diseases.

Many natural systems, including epidemics, can be characterized by a global state that is defined as a vector of local states. Correspondingly, the global state transitions are determined by the composition of local transitions [5]. To simulate the spatial and temporal dynamics of a multi-species ecosystem, its habitat is partitioned into a grid of sites, each site made sufficiently small so that it can support at most one host. Each site's state indicates the presence or absence of species involved in a simulation. A probabilistic model can describe a site's state transition as a function of the states of its neighboring sites. The neighboring sites having an influence on an affected site forms a *stencil*. Typically, the stencil is defined as a rectangular collection of sites with the affected site located inside it (but not necessarily at the center). The size of the ecological stencil and the location of the affected site within it are based on the biotic and abiotic characteristics of the species and the habitat.

Computer simulation of population densities and associated spatial patterns in ecological models can be used to predict the future of an ecosystem. The implemented model described in this paper is highly data parallel and displays regular communication patterns, hence it is suitable for SIMD architectures. The implementation has been done on a Maspar MP-1. The implemented software is parameterized to enable an easy experimentation with the different properties of the species involved in the epidemics. The results produced by the simulation include the number of sites in each state at the end of simulation, the distribution of the species in the habitat, pathogen contagion, as well as size, frequency and fractal dimensions of species' spatial clusters.

Two dimensional spatial modeling of epidemics has seldom been attempted. A population distribution lattice has often been assumed where each individual occupies a separate location, and each individual has m neighbors. Most previous models restrict contact and direct transmission to direct neighbors. Kendall (reviewed in Bartholomew [1]) considered a deterministic general epidemic in this way. Recently, some stochastic models have modeled epidemics as percolations on a two dimensional lattice; transmission is again restricted to neighbors.

Compared to earlier works, the model considered in this paper allows far more complete analyses. We can examine both direct transmission and vector-borne disease. More importantly we can manipulate the size, shape and the probabilistic dynamics of the ecological stencil to simulate a variety of local effects, and then observe the changes in the ecology across a series of large scale environments.

2 Computational Issues

The simulation is performed on a Maspar MP-1 SIMD architecture. The array of processing elements (PEs) of Maspar is organized in a two-dimensional grid. Each processing element has eight nearest neighbors with whom it can communicate directly using *xnet* operations. There is also a global router which allows communication between any two arbitrary PEs but imposes much higher communication cost than xnet operations.

[1] Graduate Student: Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180
[2] Faculty Advisor: Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180
[3] Faculty Advisor: Biology Department, SUNY Albany, Albany, NY 12192

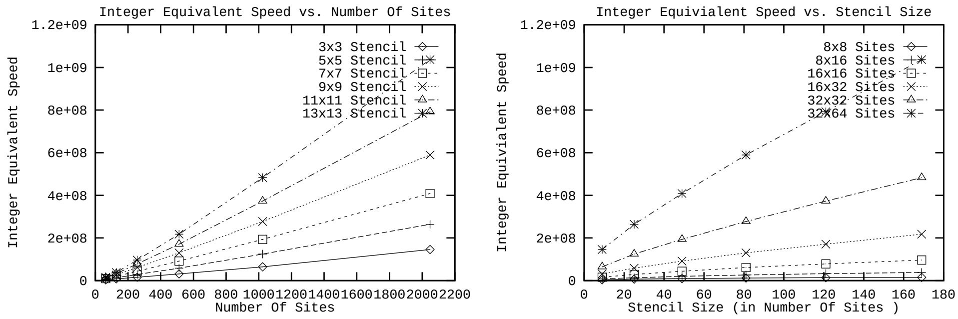

Figure 1: Integer-Equivalent Speed of MasPar MP-1

The architecture matches the grid-partitioned environment being simulated. The sites naturally map onto the PEs. Local changes of state at each site are programmed as operations on local variables whereas the global statistics about the temporal and spatial developments are stored in the front-end processor. The tight synchronization of SIMD architecture execution facilitates the update of the global state at each time step.

The main algorithmic challenges were:

- *Computation of Transition Probabilities* - These probabilities are used in selecting the next state of each site. Computing them involves simultaneous evaluation of some functions over ecological stencils. In [6], we have presented algorithms for such operations that take advantage of associativity and commutativity of used binary operators.

- *Fractal Dimension* - A measure of complexity of the spatial patterns arising in the habitat. Our implementation of this measurement is based on an algorithm developed by Cypher [2], for image labeling on mesh connected SIMD machines.

- *Relative Patchiness* - A measure of ecological diversity, i.e. the average rate of change of the landscape along some direction. This measure is implemented through parallel reduction algorithm.

3 Achieved Performance

The model was executed thirty six times using 32 bit integer and 64 bit floating point arithmetic on both the Maspar MP-1 (with 2048 PEs, each with a 1.56 MIPS and 0.0336 MFLOPS rating) and the DECstation 5000/240HX (42.9 MIPS and 6.0 MFLOPS). The simulations differ in both the number of sites in the environment (we allocate one PE per site on the MasPar) and the area of influence of each site (i.e., the sizes of the ecological stencil).

The speedup on Maspar relative to DECstation is computed as the ratio of the corresponding wall-clock computation times. The speedup is nearly linear both with respect to the number of sites (which is also the number of MasPar PE's used) and the stencil size. The biggest speedup obtained was about 24 for a 13×13 stencil and 2048 sites (i.e. 2048 MasPar PE's).

In comparing performance of Maspar vs. DECstation we need to consider processing speed disparity of floating point and integer arithmetics on both processors. On MasPar the floating point arithmetic is ≈ 46.4 slower than integer one, whereas slow-down is just ≈ 3.97 for the DECstation. Therefore we introduced the measure of the integer-equivalent speed of Maspar, the speed that would be achieved if each floating point operation were replaced by 46.4 integer operations on Maspar and 3.97 operations on DECstation and the DECstation (with sequantial implementation) would perform at its advertised speed. This measure is plotted in Figure 1 as a function of the number of sites and a function of the stencil size. The maximum integer-equivalent speed was obtained for the 13×13 stencil with 2048 processors, and it was equal $\approx 10^9$ instructions per second (about 1 GIPS). There are two reasons why the integer-equivalent speed curves are not strictly linear. Firstly, conditional flow of control

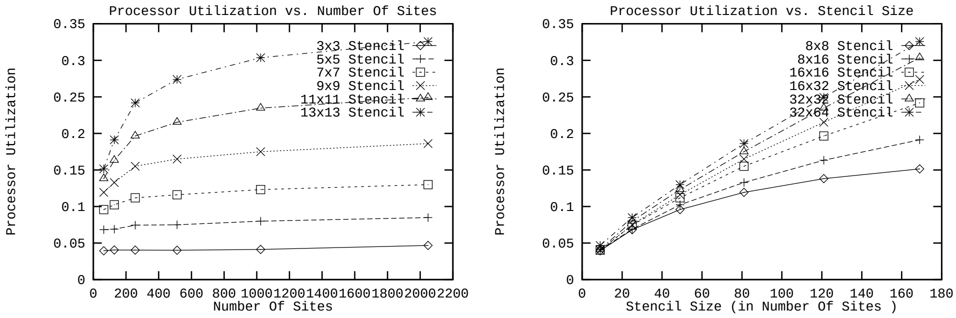

Figure 2: Processor Utilization of MasPar MP-1

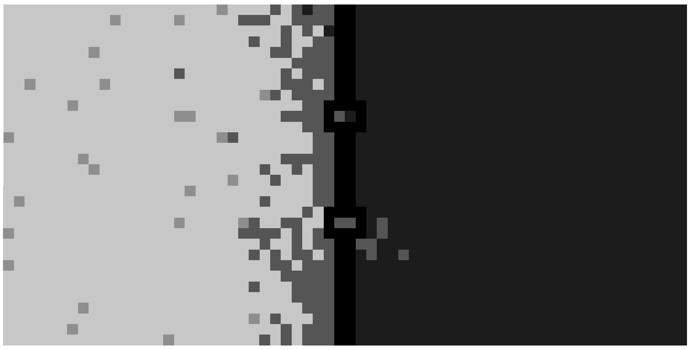

Figure 3: Tomography of epidemic about to cross the river

on SIMD architectures causes processors not taking a branch to wait for branch completion. We have a nine way branch when computing the state transition probability which is inherent in the nine state ecological model. Secondly, sublinear (logarithmic or square root), in term of stencil size, complexity of several algorithms on the Maspar differs from obviously linear complexity of comparable sequential algorithms for DECstation.

From MasPar's advertised speed and achieved integer-equivalent speed it is easy to obtain utilization of Maspar processors. It is plotted as a function of the number of sites and as a function of the stencil size in Figure 2. For the largest stencil and the maximum number of sites, simulation sustained high for SIMD architectures utilization of 32.6%.

4 Ecological Results

Each simulation starts with the same ecosystem. Hosts without parasite or pathogen, occupy most sites(90.2%). A thin barrier that cannot be occupied splits the environment into left and right halves. Two "islands" are ecological "stepping stones" [4] for the parasite and pathogen, seen in figure 3, A few hosts carrying the parasite, pathogen, or both (2% of each state) are at the far left. The epidemic spreads from left to right.

The simulation results we present are:

1. the overall frequency of diseased hosts at each sampling points,

2. the overall biodiversity of the environment at each sampling point.

The extent of infection is an epidemic's most fundamental ecological attribute. The frequency of diseased hosts describes the epidemic's temporal evolution.

[4]This is an abstraction of more realistic case of wild fox rabies moving from continental Europe to Great Britain across the English Channel via the tunnel.

Variable	Definition
α	Probability of *Parasite* attack
β	*Host* pathogen susceptibility
H	$H = -\sum_{i=1}^{S} p_i \ln(p_i)$
μ_d	*Parasite* mortality rate
p_i	the fraction of sites with state i
S	The number of distinct states in the environment

Table 1: Notations of model parameters and measurements

Diversity increases with the number of distinct ecological states, and strongly influences extinction trends. To calculate diversity we use the quantity H, in table 4.

The parameters controlling these runs are in table 4. The simulation was run for 100 generations, with sampling every 10^{th} generation. We group the simulations by the patterns which emerged:

1. $\beta = 0.001$ and $\alpha = 0.002$: Low susceptibility; neither parasite nor pathogen spread. Entries are means across μ_d values $0.001, 0.01, 0.1$ and 0.5.

2. $\beta = 0.0001$ and $\alpha = 0.015$: Low susceptibility; the parasite is stopped at the barrier. The pathogen does not spread. Entries are means for $\mu_d = 0.001$, 0.01 and 0.1. The parasite was extinct by time step 20 for $\mu_d = 0.5$.

3. $\beta = 0.001$ and $\alpha = 0.125$: Low susceptibility; the parasite spans the environment. Disease frequency remains low due to low susceptibility. Entries are means for $\mu_d = 0.001, 0.01$ and 0.1. The decline in biodiversity at the sampling time 4 shows a decline in hosts with the pathogen but lacking the parasite.

4. $\beta = 0.5$ and $\alpha = 0.002$: High susceptibility; spread of the disease is constrained by the parasite's extinction or slow growth. For $\mu_d = 0.1$ and 0.5, parasite extinction occurred before the simulation was through half the time steps. The numbers below are for $\mu_d = 0.001$ and 0.01.

5. $\beta = 0.5$ and $\alpha = 0.015$: High susceptibility; the parasite, and thus the disease, is stopped at the barrier. For $\mu_d = 0.5$, the parasite went extinct by 20 time steps. Entries are means for $\mu_d = 0.001, 0.01$ and 0.1.

6. $\beta = 0.5$ and $\alpha = 0.5$: High susceptibility; the disease occurs throughout the environment. Entries are means for $\mu_d = 0.001, 0.01, 0.1$ and 0.5. The decline in biodiversity after sampling time 2 occurs as hosts contract the parasite and pathogen.

We present graphs of average frequency of infected hosts over time, and the mean biodiversity over time (see figure 4) for the patterns listed above.

The epidemiological results confirm validity of the model. What is exciting is that we can produce the full range of epidemic conditions by manipulating only a few, easily understood, probabilistic parameters. Hence we can explore significant ecological questions in a computationally convenient manner.

5 Conclusion

The initial results have been both ecologically and computationally encouraging. We have developed novel parallel algorithm for simultaneous reduction operation. We have simulated ecosystems with larger ecological stencils that those considered to date. Future research directions include:

- Automated parallel code generation: Some speedups can be achieved by customizing the generated code to the required input parameters to avoid run-time condition execution that is expensive in SIMD architectures. Automated code generation also facilitates configuration and architecture independence, leading to ease of porting the software to other architectures.

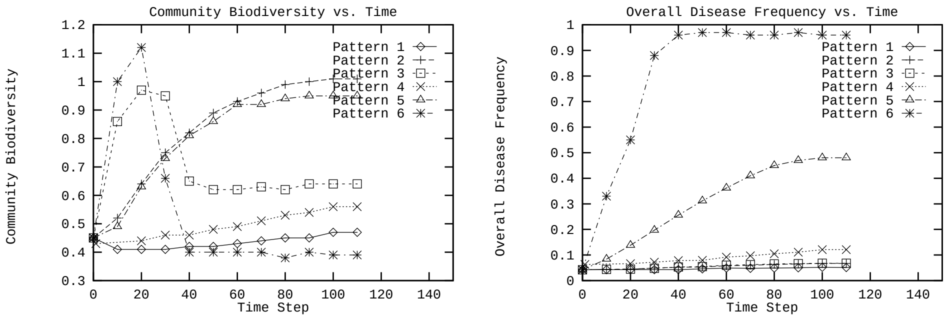

Figure 4: Ecological Results

- Virtualization of processors: Sites can be mapped onto logical processors, so that the number of sites is not limited by the number of processors physically present in the architecture.
- Interactive graphical user interface development: In particular visualization of data is critical when dealing with such large volumes of data.
- Tools to compare simulation results with experimental results: This will help in determining the accuracy of a model in predicting the temporal behavior of an ecology.

Future simulations will be based on more complex interspecies interactions than most analyses of spatial pattern in ecological dynamics offer.

References

[1] Bartholomew, D.J. 1983. Stochastic Models for Social Processes, 3rd Edn. Wiley: New York.

[2] Cypher, R. & Sanz, J. L. 1990. Algorithms for Image Component Labeling on SIMD Mesh-Connected Computers. IEEE Transactions on Computers, 39:277-281.

[3] Hastings, A. 1990. Spatial heterogeneity and ecological models. Ecology 71:426-428.

[4] Kareiva, P. 1990. Population dynamics in spatially complex environments: theory and data. Philosophical Transactions of the Royal Society, London, B, 30:175-190.

[5] Caraco, T., Szymanski, B. Spatial Analysis of Vector-Borne Disease: A Four Species Model, In review for: Evolutionary Ecology.

in review: Spatial Modeling of Epidemics on Parallel Machines, Fifth IEEE Symposium On Parallel and Distributed Processing.

Efficiency of Data Alignment on MasPar. SIGPLAN Notices, **28**, 1:48-51.

SIMD machines. IEEE Trans. Parallel and Distributing Programming, also Technical Report 91-9. Rensselaer Polytechnic Institute. Department of Computer Science, 1991.

[6] Szymanski B.K., B. Maniatty & B. Sinharoy. 1993. Simultaneous Parallel Reduction. Rensselaer Polytechnic Institute. Tech. Rep. CS 92-31, also submitted to Parallel Programming Letters.

[7] Turner, M.G. 1989. Landscape ecology: the effect of pattern on process. Annual Review of Ecology & Systematics 20:171-198.