Applied Parallel Computing, Umea, Sweden, LNCS, Vol. 1541, Springer-Verlag, Berlin, 1998, pp. 322-331

High-Performance Simulation of Evolutionary Aspects of Epidemics*

William Maniatty[1], Boleslaw K. Szymanski[1], and Thomas Caraco[2]

[1] Department of Computer Science
Rensselaer Polytechnic Institute, Troy, NY 12180
WWW home Pages: `http://www.cs.rpi.edu/~maniattb` and
`http://www.cs.rpi.edu/~szymansk`
[2] Department of Biological Sciences
SUNY Albany, Albany, NY 12192
WWW home Page: `http://www.albany.edu/biology/caraco/tcaraco.htm`

Abstract. Local interactions between individual organisms influence the population dynamics of species and impact their evolution. We describe high-performance simulation of evolutionary aspects of epidemics in spatially explicit, individual based models of multi-species habitat. Evolution consists of two processes, selection between genotypes and mutations producing novel genotypes. In this paper we focus on the effects of selection between genotypes in a model with a single host species and two competing pathogens with fixed (i.e. non-evolving) genotypes. We present the foundations of a model that represents two competing host species, a parasite serving as a disease vector, and a vector borne pathogen. The model is implemented as cellular automaton that tracks individual organisms to account for heterogeneity of the habitat. The implementation targets parallel distributed memory machines (including IBM SP-2 and a network of workstations) and NUMA shared memory architectures (SGI Origin 2000). We demonstrate also that this model yields qualitatively new biological results.

1 Introduction

Certain parasites evolve to impair their host's survival and reproduction only minimally. The primate lentiviruses offer an interesting example [16]. In older coevolutionary associations, such as simian immunodeficiency virus and African green monkeys, infection is essentially non-pathogenic, while in the recent association between HIV and humans, infection leads to serious disease. However, a long-standing coevolutionary association does not always imply reduced parasite virulence [7]. In fact, virulence exhibits a great deal of variation among host-parasite associations, and often varies temporally within a particular association [11, 9, 19].

* This work was supported in part by NSF Grant CCR-9527151. The content of this paper does not necessarily reflect the position or policy of the U.S. Government — no official endorsements should be inferred or implied.

Current theory for the evolution of virulence equates virulence with extra host mortality due to parasite infection, and than assumes that selection should increase a parasite's rate of reproduction. As a parasite exploits host resources at a greater rate, it increases the rate of transmission to new hosts. But the host's mortality rate increases as a consequence, decreasing the length of the period during which the parasite can be transmitted. Depending on the functional relationship between parasite transmission rate and virulence (i.e., the trade-off between transmission rate and infectious period), selection may favor low, intermediate or increasing virulence [15, 16, 10].

A number of recent models address variations of the trade- off just described. Important extensions include analyses of (i) interactions between virulence and host recovery rate ("clearance" by the immune system [1, 2]); (ii)competition between different parasite strains infecting the same host individual ("coinfection" [3, 8, 19]); (iii) competition between parasites when a more virulent strain excludes a less virulent strain infecting the same host ("superinfection" [11, 12, 16]); and and (iv) effects of mutation and relatedness among parasites infecting the same and different hosts [10].

This paper presents a series of computational models of increased complexity, with the goal of creating a final model encompassing analysis of all described above interactions (separately or simultaneously). We introduce two important elements into coevolutionary analysis of virulence. Ecologically, our model assumes an individual-based, spatially explicit basis for birth, death and infection processes [6]. Evolutionally, our model employs genetic algorithms to simulate mutation, recombination [13], and selection processes.

2 A Model of Selection between Competing Strains

To verify that the spatial effects are important for evolution in epidemics, we simulated competition between two microparasite strains. The simplest case of such competition arise when there are just two competing strains transmitted via direct contact and a host is infected by at most one strain at a time [11].

For simplicity, we assume the selection on virulence. The reward function for increased virulence is defined by the primary effect of increased pathogen reproductive rate (as is manifested by infecting nearby hosts). The penalty function for increased virulence (i.e., a constraint) is the consequence of the increased mortality rate for hosts in presence of a more virulent pathogen.

Let s_0, and s_1 denote two pathogen strains that differ in virulence. *Superinfection* is a disease preemption during which a host infected with s_0 becomes infected by strain s_1 via exposure from a nearby host. *Coinfection* occurs when multiple strains infect a host and order of infection matters. To denote coinfection when s_0 arrives first, we adopt the notation for the "pseudo-strain" s_{01}, but when s_1 arrives first then we introduce a pseudo-strain s_{10} [1]. Typically, coin-

[1] To model coinfective systems in which order of infection does not matter one can treat states with infection by either strain as equivalent thereby "merging" these two states into one.

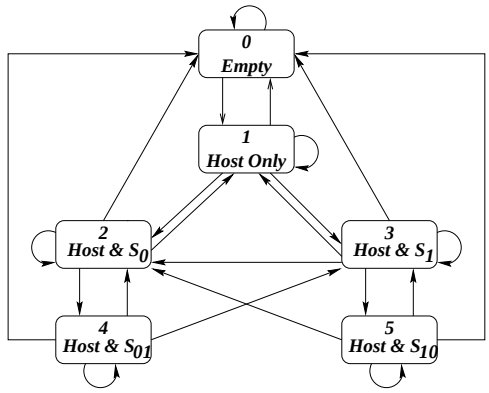


Fig. 1. State Diagram of Coinfection/Competition Model

fection and superinfection are modeled exclusively of each other, so our model provides a mechanism for disabling these events (although, as can be seen below, we permit the user to enable both as well). The state diagram of this simple model is shown in Figure 1.

The parameters for the system are given in Table 1. Let $p(x, y)$ be the probability of going from state x to state y in a single time step.

An empty site will remain unpopulated with probability $p(0,0) = \prod_{i=1}^{5} \left[(1-\rho_i)^{\sigma_i}\right]$, otherwise, it is colonized by an offspring from a nearby populated site. Since congenital immunity is disallowed: $p(0,1) = 1 - p(0,0)$.

Consider next a site occupied by a susceptible (i.e., an uninfected host). It may either die with probability $p(1,0) = \mu_1$ or be infected by a strain from their neighbors. The probability of a host in state 1 avoiding exposure to strain $s_i, i \in 0,1$ is denoted A_i with: $A_0 = (1-\alpha_0)^{\sigma_2}(1-\alpha_{01,0})^{\sigma_4}(1-\alpha_{10,0})^{\sigma_5}$ and $A_1 = (1-\alpha_1)^{\sigma_3}(1-\alpha_{01,1})^{\sigma_4}(1-\alpha_{10,1})^{\sigma_5}$ The Poisson single transition rule implies that if a susceptible is exposed to both strains, only one of them will be successful in infecting the host, so the probability of catching strain 0 is $p(1,2) = (1-\mu_1)(1-A_0)\left[A_1 + \gamma(1-A_1)\right]$ and the probability of catching strain 1 is then: $p(1,3) = (1-\mu_1)(1-A_1)\left[A_0 + (1-\gamma)(1-A_0)\right]$

A host infected with strain s_0 will die (i.e., be removed) with probability $p(2,0) = \mu_2$, and will recover (without immunity) with probability $p(2,1) = (1-\mu_2)\mu_{s_0}$. It will become coinfected with strain s_1 after being infected by $s_{0,1}$ with probability $p(2,4) = (1-\mu_2)(1-\mu_{s_0})(1-A_0)$ and will remain in the same state otherwise.

Finally, consider the case of a host infected with strain s_1. The probability that it dies is $p(3,0) = \mu_3$. It will recover with probability $p(3,1) = (1-\mu_3)\mu_{s_1}$, and superinfection by s_0 will occur with probability $p(3,2) = (1-\mu_3)(1-\mu_{s_1})p_{SI}(1-p_{01})(1-A_1)$. Coinfection by s_0 will occur with probability $p(3,5) = (1-\mu_3)(1-\mu_{s_1})(1-p_{SI})p_{01}(1-A_0)$.

Symbol	Meaning and values
$\alpha_i, i \in \{0,1\}$	Probability of exposure to s_i from an infective in state 2 or 3 within δ.
$A_i, i \in \{0,1\}$	The probability that a susceptible avoids exposure to s_i
$\alpha_{ij,k} i,j,k \in \{0,1\}, i \neq j$	Probability of exposure to s_i from a coinfected host infected by strain s_i then s_j
γ	The competitive advantage of s_0 over s_1 when susceptibles are exposed to both strains
δ_k	The interaction region about site k, $\delta_k \in \{3^2, 11^2, 33^2\}$
$\mu_i, i \in \{1,2,3\}$	Probability of host mortality, $\mu_1 = 0.5, \mu_3 = 0.25$ $\mu_2 \in \{\{0.251, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95\}$
$\mu_{s_i}, i \in \{0,1\}$	Probability of recovery from a single strain s_i
$\mu_{ij,k}, i,j,k \in \{0,1\}, i \neq j$	Probability of recovery of a coinfected host from s_k
p_{SI}	Post-Exposure Superinfection probability of s_0 $p_{SI} = 0.3$
p_{01}	The probability that a host infected with s_0 can be coinfected with s_{01} upon exposure to s_1
p_{10}	The probability that a host infected with s_1 can be coinfected with s_{10} upon exposure to s_0
$\rho_i, i \in 1,2,3,4$	Probability of host in state i placing
a propagule nearby (in δ),	$(1-\rho_i)^\delta \in \{0.24, 0.2, 0.15, 0.11\}$
$\sigma_i(k), i \in \{0,1,2,3\}$	Number of hosts in state i about site k

Table 1. Symbols Used in Direct Transmission Two Competing Strains Model

Consider a coinfected host, where s_0 infected the host first. The host can die with probability $p(4,0) = \mu_4$, or it can recover from strain s_0 with probability $p(4,3) = (1-\mu_4)(1-\mu_{01,1})\mu_{01,0}$. The host can recover from strain s_1 with probability $p(4,2) = (1-\mu_4)(1-\mu_{01,0})\mu_{01,1}$, otherwise, the host will remain in its current state [2].

Consider the complementary coinfected host, where s_1 infected the host first. The host can die with probability $p(5,0) = \mu_5$. It can recover from strain s_0 with probability $p(5,3) = (1-\mu_5)(1-\mu_{10,1})\mu_{10,0}$. The host could recover from strain s_1 with probability $p(5,2) = (1-\mu_5)(1-\mu_{10,0})\mu_{10,1}$ otherwise the host will remain in its current state [3].

Similar reasoning can be used to extend this model to vector-borne pathogenic strains.

A subset of the parameters presented in Table 1 was selected as a collection of control parameters and based on TEMPEST, we implemented a tool, called

[2] Due to the Poisson single transition rule, we treat the possibility of simultaneous recovery from both strains as an impossible event. However if this event were to be considered possible with non-negligible probability, the formulation might look like $p(4,1) = (1-\mu_4)\mu_{01,0}\mu_{01,1}$

[3] Again, due to the Poisson single transition rule, we treat the possibility of simultaneous recovery from both strains as an impossible event. However if this event were to be considered possible with non-negligible probability, the formulation might look like $p(5,1) = (1-\mu_5)\mu_{10,0}\mu_{10,1}$.

STORM, to simulate the simplified model. In the simulation runs, we varied the spatial parameters governing host fecundity, $\rho_i, i \in \{1, 2, 3\}$ and rate of disease exposure from local infectives $\alpha_j, j \in \{0, 1\}$ so that the intensity of these processes would not be unduly impacted by variations in the area of the ecological stencil of each site.

The space was toroidally wrapped to avoid impact of boundaries on the simulation results. Initially 25% of the sites were populated by susceptible hosts in a spatially uniform density. Small clusters of infective hosts (0.25% of the sites) were placed in the initial environment, with maximum spatial separation to avoid premature extinctions induced by local superinfection events. The experiments presented use a region of 100×100 sites, however STORM has capacity for much larger simulations. STORM was run using the fixed spatial configuration for 2000 generations for each parameter combination.

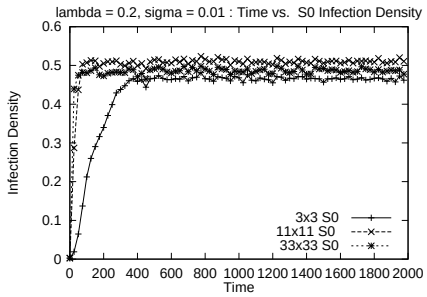
The results of the experiment have shown a remarkable richness in the range of outcomes generated. First consider a case with two competing strains which are equal except for (i) a small difference in their virulence induced mortality and (ii) the possibility of superinfection by the more virulent strain, s_0. We observe that the the more virulent strain, s_0 can drive the less virulent strain, s_1 to extinction via superinfection as seen in Figure 2(a) and 2(b). The rise and decline of s_1 is governed by the rate at which the epidemic can spread through the environment, which is in turn correlated to the stencil's area, δ. With larger δ values, both s_0 and s_1 infection disperse more quickly so that the superinfection induced extinction occurs earlier.

Now consider a similar case, except that the virulence induced mortality difference is large. In this case, even the competitive advantage of superinfection is not sufficient to prevent extinction of s_0, as shown in Figure 2(e) and 2(f). The rise and decline of s_0 in this case reflects expansion of s_0 infection from its original small cluster and later a denial of access to susceptibles occurs due to the prevalence of s_1 infection.

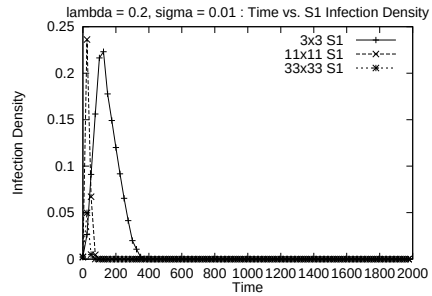
Finally consider an intermediate level of virulence induced mortality, as seen in Figure 2(c) and 2(d). For a small stencil (e.g., the 3×3 case), the more virulent strain, s_0, drives the less virulent strain, s_1 extinct by gradually excluding it from susceptible hosts. For intermediate sized interaction neighborhoods (e.g., the 11×11 case), the virulent less strain, s_1, drives the more virulent strain, s_0 extinct because of rapid death of hosts infected by s_0. For large stencils (e.g., 33×33) s_0 and s_1 coexist, the first thanks to superinfection and the second because of increased access to susceptibles. This range of results can only be found in spatially explicit models and cannot be generated by spatially homogeneous models.

3 Implementation and Performance of STORM

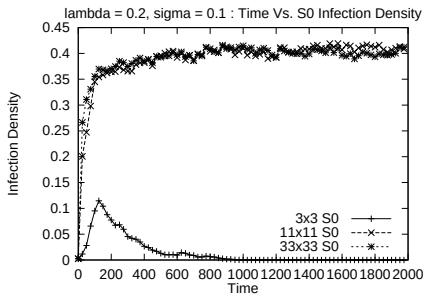
The STORM model was implemented using C++ and MPI much like the TEMPEST model described in [14]. Prior performance analysis of TEMPEST [17, 14] was done on a SIMD mesh architecture (a MasPar MP-1), a network of worksta-



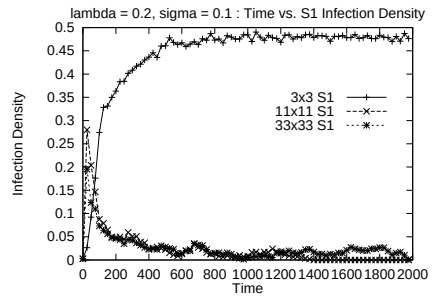
(a) Strain with low virulence, $\sigma = 0.01$



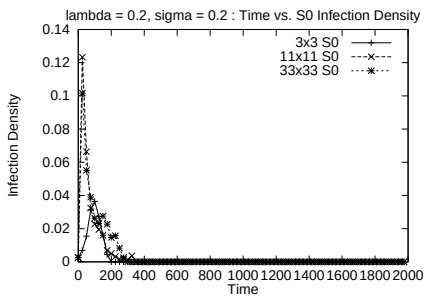
(b) Strain with high virulence, $\sigma = 0.01$



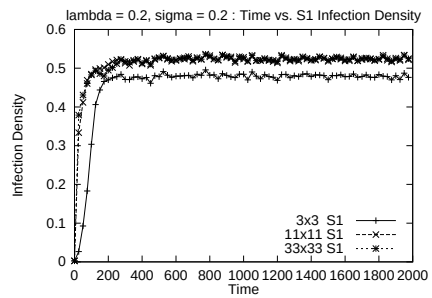
(c) [Strain with low virulence, $\sigma = 0.1$



(d) Strain with high virulence, $\sigma = 0.1$



(e) Strain with low virulence, $\sigma = 0.2$



(f) Strain with high virulence, $\sigma = 0.2$

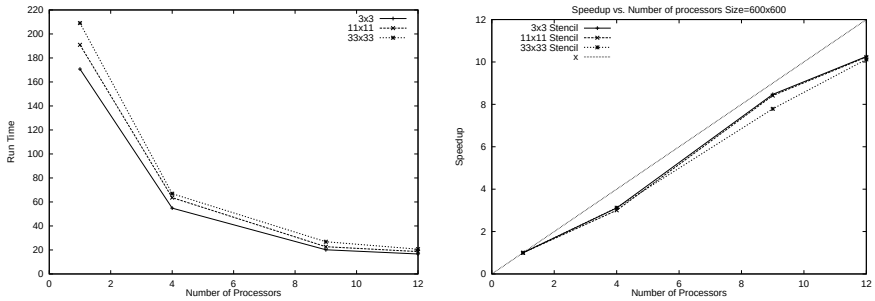
Fig. 2. The impact of stencil area, δ, and relative virulence levels, $\sigma = \mu_1 - \mu_2$, on Infection densities

tions and a coarse grained MIMD architecture(an IBM SP2) [14]. In this paper we introduce performance results from a tightly coupled cache coherent SMP architecture, an SGI Origin 2000 (our configuration has 12 processors). We briefly review implementation issues and then provide a performance comparison.

The underlying model of STORM has strong locality of interaction, so a static block data decomposition was selected (much like in TEMPEST [14]) At each time step, the state transition probabilities are computed, which involves counting the number of sites in each state within the interaction neighborhood [14]. The processor which "owns" the partition in which a particular site resides is responsible for computing its next state, and therefore requires information about the current state of boundary sites on neighboring processors (since some stencils can span partitions). We exchanged this information using MPI directly on the SGI, rather than allowing a direct read of data owned by neighboring processors. This allowed a short development time (with regards to porting the code) and avoided race conditions. We used the shared memory chameleon (mpich) implementation.

The stochastic nature of the model (state transitions are selected randomly according to predefined distribution) require many runs of the same model with different parameters, initial configurations and random number generator seeds to obtain meaningful results (this process resembles sampling the state space with Monte Carlo methods in numerical analysis). Hence, the speed of computation is of utmost importance. We ran simulations for 100 time steps for an environment of 600×600 sites using stencil sizes of 3×3, 11×11 and 33×33 on 1, 4, 9 and 12 processors. The run times were relatively insensitive to variation of the interaction neighborhood's area as shown in Figure 3. The associated speedup curves (see Figure 3) were within 85% of optimal, but approximately linear with a small knee at 9 processors. We believe that some of the slowdown observed at 12 processors might have to do with the systems software stealing cycles (since we utilized all 12 processors in our configuration).

Typically users, including those using our system, are often interested in the ability of machines to run large scale simulations, and will increase the problem size in response to increased capacity. In the following experiment we assigned a fixed number of sites to each processor, with per processor allocation being one of the following: $100^2, 200^2, 300^2, 400^2, 500^2$ or 600^2 sites for 100 time steps on $1, 4, 9$ or 12 processors on the SGI Origin 2000. The interaction neighborhood size was fixed at 11×11. The timings shown in Figure 4 demonstrate some slowdown when going from single processor to multiprocessor runs (due to copying of boundary information) and another slowdown when all 12 processors were utilized. The slowdown when 12 processors were used was most pronounced for large per processor allocations. This might reflect increased contention of the cache coherent memory, and the overhead of systems software stealing resources to run during the simulations. The 600×600 problem size had a repeatable increase in run time when 4 processors were used over the 9 processor runs of about 10%. We suspect that this increase is due to contention of processors for memory access in the interconnection hub, but more research is required to check this hy-



(a) Run Time Vs. Number of Processors

(b) Speedup Vs. Number of Processors

Fig. 3. Performance of Fixed Size Workload vs. Number of Processors on an SGI Origin 2000

pothesis. We noted that the scaled speedup curves looked remarkably similar to the shape of the speedup curves for 9 or less processors, but had a performance degradation when all 12 processors were used (achieving a speedup of about 8). For smaller problem sizes, the system finished within 85% of optimal speedup, but for larger problem sizes, the parallel efficiency was limited to 66%.

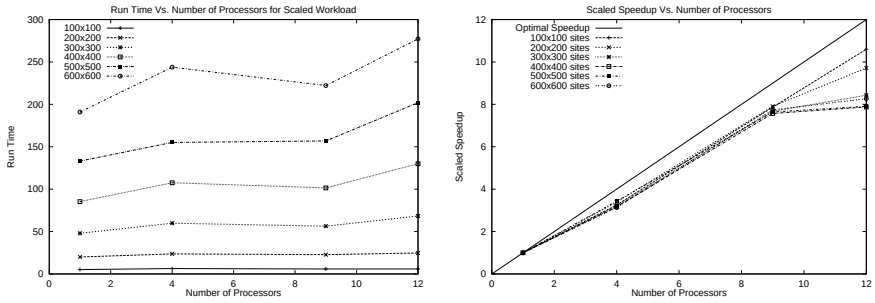
Measuring the per processor throughput of the SGI Origin 2000 (denoted T_{SGI}) for the largest problem size yields:

$$T_{SGI} = \frac{100\text{timesteps} \times 4.32 \times 10^6 \frac{\text{transitions}}{\text{timestep}}}{12\text{processors} \times 276.7\text{sec.}} \approx \frac{1.56 \times 10^6\text{transitions}}{\text{sec.}} \tag{1}$$

Similar computations for TEMPEST presented in [14] showed that the per processor throughput of a MasPar MP-1 was $T_{MP-1} \approx 44\frac{transitions}{sec.}$, while the per processor throughput of an SP2 was $T_{SP2} = 1.2 \times 10^5 \frac{transitions}{sec.}$. The departmental network of workstations achieved $T_{NOW} = 2.7 \times 10^4 \frac{transitions}{sec.}$. Assuming that the two simulations do roughly equivalent per site computation (which is reasonable), the SGI Origin 2000's processors can be thought of as capable of doing the work of over 35000 MasPar MP-1 processors, 13 SP2 processors and 57 NOW processors.

4 Conclusions and Future Directions

Evolution is the result of the combined effects of selection and mutation. In this paper we presented a model that analyzes only one these processes. By isolating and modeling the impact of selection on simple non-evolving systems, we provided a platform for validating and interpreting results when the complexity



(a) Run Time Vs. Number of Processors

(b) Speedup Vs. Number of Processors

Fig. 4. Performance of Scaled Size Workload vs. Number of Processors on an SGI Origin 2000

of evolution is added. We have observed good performance characteristics on a surprisingly wide range of architectures due to their localized interactions and algorithm selection [18, 14] for frequently performed operations.

Work is in progress to develop a cellular automaton model of genetic variation (and mutation) of two competing host species, one macro-parasite spread via direct contact transmission, and a vector-borne micro-parasite. The evolution of the species is accomplished using the genetic algorithms on the genetic information, copying, mutation, and (where applicable) crossover. Hosts are modeled individually and sites are small enough to contain at most one host, in the same fashion as the model presented here and the Szymanski-Caraco model [17]. The host species are assumed to have sexual reproduction (and hence crossover of genetic material) while the parasitic species are treated as asexual. Such a model poses novel computational challenges, including: computing the lineage of each organism, measuring the simulation trajectory given within site diversity, modeling intrahost competition of parasites. This work also addresses specific defenses against infection (i.e. acquired immunity).

Efficiency and fidelity concerns motivate the consideration of whether a synchronous cellular automata or discrete event simulation engine is better for the simulation kernel. Our experience with Lyme disease simulation (without evolutionary effects, cf. [5, 4]), indicates that in the case of complex species interactions or large diversity of time scales between species an optimistic approach with rollback for correcting potential causality errors may be very efficient. Hence, we plan to implement the presented system using PDES approach and compare efficiency of the resulting system with the current implementation.

References

1. R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford, 1991.
2. R. Antia, B. R. Levin, and R. M. May. Within host population dynamics and the evolution and maintenance of microparasite virulence. *American Naturalist*, 144:457–472, 1994.
3. H. J. Bremermann and J. Pickering. A game-theoretical model of parasite virulence. *J. Theoretical Biology*, 100:400–426, 1983.
4. E. Deelman, T. Caraco, and B. Szymanski. Breadth-first rollback in spatially explicit simulations. In *Proc. PADS97, 11th Workshop on Parallel and DistributedSimulation*, Los Alamitos, CA, 1997. IEEE Computer Society.
5. E Deelman, T. Caraco, and B. K. Szymanski. Parallel discrete event simulation of lyme disease. In *Proceedings of the Conference on Biocomputing*, Hawaii, 1996. World Scientific Publishing Co, Singapore.
6. M. Duryea, J. Gardner, T. Caraco, B. K. Szymanski, and W. A. Maniatty. Host spatial heterogeneity and extinction of an SIS epidemics. *Journal of Theoretical Biology*, 1998. To appear.
7. P. W. Ewald. *Evolution of Infectious Disease.* Oxford University Press, New York City, NY USA, 1994.
8. S. A. Frank. A kin selection model for the evolution of virulence. *Proc. Royal Society London*, B250:195–197, 1992.
9. S. A. Frank. Coevolutionary genetics of plants and pathogens. *Evolutionary Ecology*, 7(1):45–75, 1993.
10. S. A. Frank. Models of parasite virulence. *Quarterly Review of Biology*, 71:37–78, 1996.
11. S. Levin and D. Pimentel. Selection of intermediate rates of virulence in parasite-host systems. *American Naturalist*, 117:308–315, 1981.
12. S. A. Levin. Some approaches to the modelling of coevolutionary interactions. In M. H. Nitecki, editor, *Coevolution*, pages 21–65. University of Chicago Press, Chicago, 1983.
13. Hastings I. M. and B. Wedgwood-Oppenheim. Sex, strains and virulence. *Parasitology Today*, 13:375–383, 1997.
14. W. A. Maniatty, B. K. Szymanski, and T. Caraco. Parallel computing with generalized cellular automata. *Parallel and Distributed Programming Practices*, 1(1):85–104, January 1998. Also Technical Report 97-3 Department of Computer Science, Rensselaer Polytenchic Institute, Troy, NY 12180.
15. R. M. May and Anderson R. M. Parasite-host coevolution. *Parastiology*, 100:S89–S100, 1990.
16. M. A. Nowak and R. M. May. Superinfection and the evolution of parasite virulence. *Proc. Royal Soc. London*, B 255:81–89, 1994.
17. B. K. Szymanski and T. Caraco. Spatial analysis of vector-borne disease: A four species model. *Evolutionary Ecology*, 8:299–314, 1994. Online information available at ftp://ftp.cs.rpi.edu/pub/szymansk/eec.ps.
18. B.K. Szymanski, W. Maniatty, and B. Sinharoy. Simultaneous parallel reduction. *Parallel Processing Letters*, 5(1), 1995. RPI Dept. of Computer Science Tech Report CS 92-31.
19. M. Van Baalen and M. W. Sabelis. The dynamics of multiple infection and the evolution of virulence. *American Naturalist*, 146:881–910, December 1995.