Evolving Molecular Force Field Parameters for Si and Ge

Al Globus
CSC
NASA Ames Research Center

Madhu Menon
University of Kentucky

Ecleamus Ricks
Morehouse College

Deepak Srivastava
CSC
NASA Ames Research Center

1 Abstract

A genetic algorithm (GA) procedure has been developed to fit parameters for multi-species reactive interatomic force field functions. Given an analytic form (of which several are available), fitting parameters to multi-species reactive force fields is extremely tedious and error prone because the parameter space is large and includes complex correlations. As a result, parameters are available for only a few reactive systems (Si, C, and a few others). By automating parameter fitting, we seek to significantly expand the reactive systems that may be investigated using molecular dynamics.

2 Introduction

The ability to model reactive solid systems with fast molecular dynamics, as opposed to much more compute-intensive quantum calculations, can enable a wide variety of crack propagation, thin-film deposition and etching, ion and cluster bombardment, surface diffusion, surface reactions, molecular machine manufacture, nanotube strength and dynamics, and many other studies. Unfortunately, even given good functional form for the molecular potential, parameterization is extremely tedious, error prone, and makes a risky thesis topic as many parameterization projects fail entirely. Worse, one rarely finds a good functional form immediately, so parameterization must be repeated as the functional form is developed. By automating parameterization with a genetic algorithm (GA) using massive quantities of cycle-scavenged CPU time we seek to change multi-species reactive molecular potential development from a multi-year to a multi-week affair.

3 Genetic Algorithm

GAs seek to mimic natural evolution's ability to produce highly functional objects. Natural evolution produces organisms, whereas GAs can produce sets of parameters, programs, molecular designs, and many other structures. Our GA, JavaGenes, employs the following algorithm (words in quotes are typical GA terminology):

1. Represent potential parameters as a set of floating point numbers; each set is called an "individual"
2. Generate a "population" of individuals with random parameters
3. Calculate the "fitness" of each individual
4. Repeat
   (a) Randomly select "parents" with a bias towards better fitness
   (b) Produce "children" from the parents with either:
      i. "crossover" that combines parts of two parents into a child
      ii. "mutation" that modifies a single parent
      iii. or a combination of the two
   (c) Calculate the fitness of the child
   (d) Randomly replace individuals of less fitness in the population with the child
5. Until satisfied

The current fitness function is a measure of the distance between the energies calculated for many Si or Ge clusters using an individual's parameters and reference values calculated using higher quality quantum codes and/or derived from experiment. Many additional fitness functions are possible (bond length, vibrational energies, minimal energy conformation, bulk properties, etc.), but have not yet been implemented. Multiple fitness functions may be combined by a weighted sum, product, or using multi-objective methods.

While GA has successfully parameterized force fields in the past [Hunger, Beyreuther, Huttner, Allinger, Radelof, Zsolnai (1998); Hunger, Huttner (1999); Cundari, Fu (2000)], until now GA has not been applied to parameterization of reactive force fields suitable for critical nanotechnology tasks.

4 Functional Form

For proof of concept, we chose the well-established Stillinger-Weber (SW) functional form for Si [Stillinger, Weber (1985)].

SW expresses the total energy as the sum of two- and three-body contributions:

$$E = \sum_{\substack{i,j \\ i<j}} v_2(i,j) + \sum_{\substack{i,j,k \\ i<j<k}} v_3(i,j,k) \quad (1)$$

where E is total interaction energy, i, j, k indicate individual atoms, and v is the interaction energy of n atoms. To reduce computation, SW has a cutoff function which forces each term to zero at large atomic separations. This converts the problem from $O(n^3)$ to $O(n)$ since only nearby atoms need be considered. The terms are:

$$v_2(i,j) = A(Br^{-p} - r^{-q})c_1 \quad (2)$$

$$c_1 = e^{\frac{C}{r-a}};\ r < a \quad (3)$$

$$c_1 = 0;\ r \geq a \quad (4)$$

where r is the i, j inter-atomic distance and all other values are adjustable parameters.

$$v_3(i,j,k) = \alpha + \lambda(\cos\Theta - \cos\Theta_0)^2 c_2 \quad (5)$$

$$c_2 = e^{\frac{\gamma}{r_{i,j}-a_1} + \frac{\gamma}{r_{j,k}-a_2}}; r_{i,j} < a_1 \wedge r_{j,k} < a_2 \quad (6)$$

$$c_2 = 0;\ r_{i,j} \geq a_1 \vee r_{j,k} \geq a_2 \quad (7)$$

where $r_{i,j}$ and $r_{j,k}$ are the two inter-atomic distances, θ is the angle and all other values are adjustable parameters. The parameters a, a_1, and a_2 defining the cut-off distances are not evolved because they optimize performance by degrading accuracy. The equilibrium angle, $\Theta_0 \approx 109$ degrees for the tetrahedral geometry of Si and Ge, is also not evolved.

5 Results

As a proof of concept, we demonstrate the procedure by (a) reproducing the published parameters for Si by using SW energetics in the fitness function, (b) evolving a new set of parameters, with a fitness function based on a non-orthogonal tight-binding method [Menon, M.; Subbaswamy, K. R. (1993)] better suited to Si cluster energetics than the published SW Si potential, and (c) evolving parameters for Ge clusters, for which SW parameters were previously unavailable. Evolution was driven by a fitness function based on the energies and forces calculated for Si_n and Ge_n clusters ($n < 7$) and was able to predict accurate energies for minimum energy and deformed configurations of Si_n ($n = 7, 8, 33$) clusters, which were not used in the fitness function.

Figure 1 shows that the evolved Si parameters match the tight-binding energies for Si_{33} clusters much more closely than the original published parameters, both quantitatively and qualitatively. Figure 2 shows that the Ge parameters match the tight-binding energies quite well. Examining the evolved parameters (table 1) we see that evolution has found parameters that would probably never be chosen intuitively or by deductive reasoning, although they match the overall data quite well. In particular, note that p and q have different signs and that for Ge α and λ have very large values. For the mixed hetero-atomic systems, the methods involving taking arithmatic or geometric mean of the parameters belonging to individual components may not work with the above set because of the relatively large differences in the magnitudes of some parameters. The heteroatomic systems with mixed species need to be fitted separately.

Parameter	Si	Ge
A	40.83	21184
B	784.12	1979
C	6.31	29.18
p	9.94	3.59
q	-0.8	-9.07
α	16.88	-9.75e17
λ	77.23	2.4e18
γ	3.91	282.3

Table 1: Best Parameters for eV/angstrom

6 Computational Issues

GAs typically require a great deal of processing, and we run 1000s of jobs with randomized GA parameters (population size, transmission operations, bounds for parameters in the initial population, etc.). Fortunately, GAs are embarrassingly parallel so the jobs can run simultaneously on many workstations. To provide these cycles inexpensively, we use the Condor cycle scavenger [Litzkow, Livny,Mutka (1988)] running on about 350 SGI and Sun machines at the NASA Advanced Supercomputing (NAS) Division. Each machine runs a daemon that watches user I/O and CPU load and runs jobs when the machines are otherwise idle.

7 Discussion

JavaGenes is able to generate a few dozen good parameterizations, out of a thousand jobs, in a few days using the NAS Condor pool. Introduction of a new functional form takes a few days to a week or so. If our initial results are confirmed by additional use, development of new classical potentials for reactive systems should become routine. Furthermore, the process involved generates many parameterizations, not just one. Thus, studies can use multiple parameterizations, or even multiple functional forms, to insure there is no dependence on the precise parameterization used.

8 References

Cundari, T. R.; Fu, W. T. (2000): Genetic algorithm optimization of a molecular mechanics force field for technetium. Inorganica Chimica Acta 300-302, pp. 113-124

Litzkow, M.; Livny, M.; Mutka, M. W. (1988): Condor - a hunter of idle workstations. Proceedings of the 8th International Conference of Distributed Computing Systems, pp. 104-111

Hunger, J.; Beyreuther, S.; Huttner, G.; Allinger, K.; Radelof , U.; Zsolnai, L. (1998): How to derive force field parameters by genetic algorithms: modeling tripod-Mo(CO)3 compounds as an example. European Journal of Inorganic Chemistry, pp. 693-702

Hunger, J.; Huttner, G. (1999): Optimization and analysis of force field parameters by a combination of genetic algorithms and neural networks. Journal of Computational Chemistry, vol. 20, pp. 455-471

Menon, M.; Subbaswamy, K. R. (1993): Non-orthogonal tight-binding molecular-dynamics study of silicon clusters. Physical Review B, vol. 47, no. 19, pp. 754-759

Stillinger, F. H.; Weber, T. A. (1985): Computer simulation of local order in condensed phases of silicon. Physical Review B, vol. 31, no. 8, pp. 5262-5271

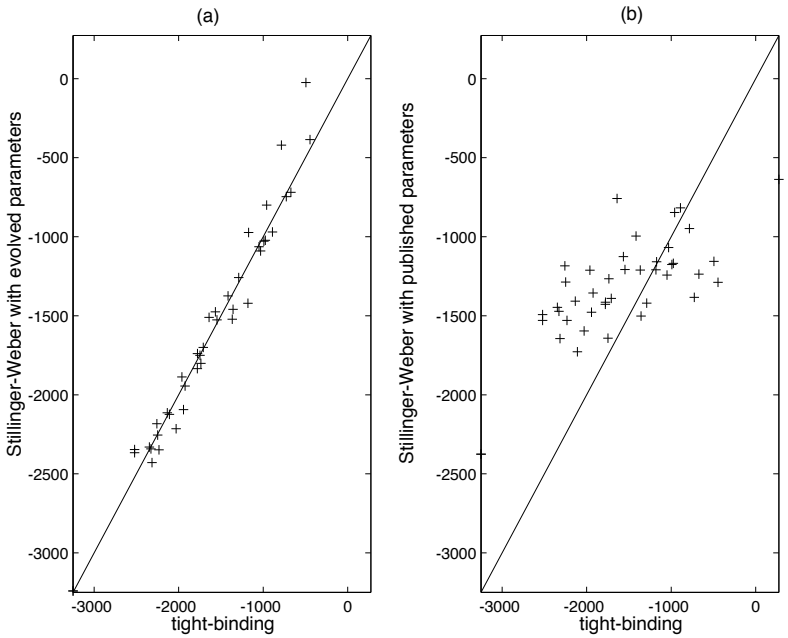

Figure 1: Comparison of energies calculated for Si_{33} clusters using tight-binding vs. evolved and published SW parameters. Each cross represents a cluster. Crosses on the diagonal line are a perfect fit. Evolved parameters were generated using a fitness function with Si_{2-6} cluster energies calculated by the tight-binding method. In (a) the vertical axis is the energy calculated using evolved parameters in the same units as [Stillinger, Weber (1985)]. In (b) the vertical axis is the energy calculated using the published parameters. Horizontal axes are the energies calculated by the tight-binding method. The figures clearly show that the evolved parameters match the tight-binding energies much more closely than the published parameters. All energies are kcal/mol.

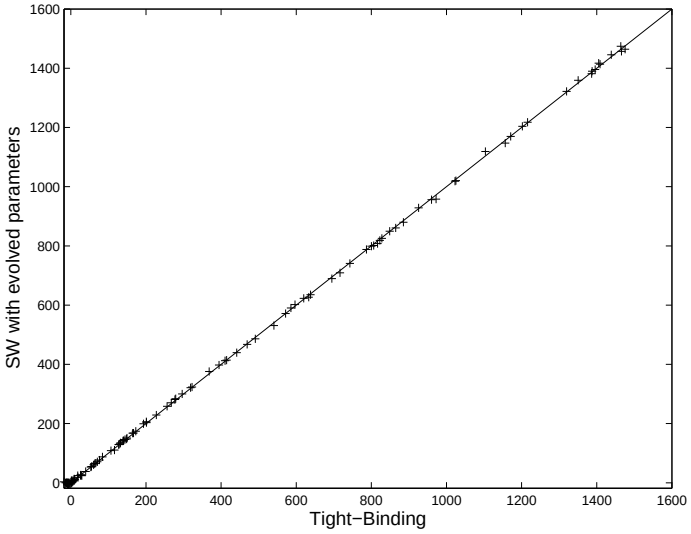

Figure 2: Comparison of energies calculated for Ge_{2-6} clusters of tight-binding vs. evolved. Each cross represents a cluster. Crosses on the diagonal line are a perfect fit. Evolved parameters were generated using a fitness function with Si_{2-6} cluster energies calculated by the tight-binding method. The vertical axis is the energy calculated using evolved parameters in eV/angstrom. The horizontal axis are the energies calculated by the tight-binding method. All energies are eV.