

**Multiobjective Genetic Algorithms for Multiscaling  
Excited State Direct Dynamics in Photochemistry**

**Kumara Sastry, D.D. Johnson,  
Alexis L. Thompson, David E. Goldberg,  
Todd J. Martinez, Jeff Leiding,  
and Jane Owens**

**IlliGAL Report No. 2006005  
February, 2006**

Illinois Genetic Algorithms Laboratory  
University of Illinois at Urbana-Champaign  
117 Transportation Building  
104 S. Mathews Avenue Urbana, IL 61801  
Office: (217) 333-2346  
Fax: (217) 244-5705

# Multiobjective Genetic Algorithms for Multiscaling Excited State Direct Dynamics in Photochemistry

Kumara Sastry<sup>1</sup>, D.D. Johnson<sup>2</sup>, Alexis L. Thompson<sup>3</sup>,  
David E. Goldberg<sup>1</sup>, Todd J. Martinez<sup>3</sup>, Jeff Leiding<sup>4</sup>, and Jane Owens<sup>4</sup>

<sup>1</sup>Illinois Genetic Algorithms Laboratory (IlliGAL),  
Department of Industrial and Enterprise Systems Engineering

<sup>2</sup>Department of Materials Science and Engineering

<sup>3</sup>Department of Chemistry, Beckman Institute, and  
Fredrick Seitz Materials Research Laboratory

<sup>4</sup>Department of Chemistry

University of Illinois at Urbana-Champaign, Urbana IL 61801

ksastry@uiuc.edu, duanej@uiuc.edu, alexis@spawn.scs.uiuc.edu,  
deg@uiuc.edu, tjm@spawn.scs.uiuc.edu, jeff@spawn.scs.uiuc.edu  
jane@spawn.scs.uiuc.edu

February 12, 2006

## Abstract

This paper studies the effectiveness of multiobjective genetic and evolutionary algorithms in multiscaling excited state direct dynamics in photochemistry via rapid reparameterization of semiempirical methods. Using a very limited set of *ab initio* and experimental data, semiempirical parameters are reoptimized to provide globally accurate potential energy surfaces, thereby eliminating the need for full-fledged *ab initio* dynamics simulations, which are very expensive. Through reoptimization of the semiempirical methods, excited-state energetics are predicted accurately, while retaining accurate ground-state predictions. The results show that the multiobjective evolutionary algorithm consistently yields solutions that are significantly better—up to 230% lower error in the energy and 86.5% lower error in the energy-gradient—than those reported in the literature. Multiple high-quality parameter sets are obtained that are verified with quantum dynamical calculations, which show near-ideal behavior on critical and untested excited state geometries. The results demonstrate that the reparameterization strategy via evolutionary algorithms is a promising way to extend direct dynamics simulations of photochemistry to multi-picosecond time scales.

## 1 Introduction

Many phenomena in science and engineering are inherently multiscale and in the recent years there has been growing interest in developing effective modeling and simulation methods to explain

or predict their behavior. In essence, there is a significant premium on cost-effective modeling techniques that can simulate physical, chemical or biological phenomena across multiple scales in both time and space, even at the price of losing information at intermediate scales. One multiscale approach is to apply modeling methods of a single scale and couple them by transferring key information from the finer scale to a coarser scale. An important and often daunting task in this multiscale approach is the development of proper coupling methods and evolutionary algorithms can potentially play an important role (Sastry, Johnson, Goldberg, & Bellon, 2004; Sastry, Johnson, Goldberg, & Bellon, 2005).

One such area of multiscale where evolutionary algorithms are very useful is in modeling excited-state dynamics in photochemistry. Photochemical reactions, as well as many spectroscopic measurements, involve electronic excited states of molecules and their concomitant structural changes. Such excited-state reactions are fundamental in many biological (for example, photosynthesis and vision) and technological (for example, solar cells and LED displays) settings. These reactions and the associated dynamics are energetically subtle and require highly accurate descriptions of the relevant interatomic forces. Thus, reliable predictions are costly even for small molecular reactions but rapidly become near impossible for reactions in complex environments, such as in solvents (for example, water), in solid cages (for example, zeolites), or with proteins.

The *ab initio* multiple spawning (AIMS) methods, which simultaneously solve both the electronic and nuclear Schrödinger equations (Ben-Nun & Martinez, 2002; Ben-Nun, Quenneville, & Martinez, 2000), while very flexible and accurate, can be computationally very expensive, especially for large molecules. Hence, having substantially faster semiempirical potentials that accurately reproduce higher-level quantum chemistry results would make it possible to address critical biological processes and technologically useful chemical reactions. However, the semiempirical methods (Dewar & Thiel, 1977; Dewar, Zoebisch, Healy, & Stewart, 1985; Stewart, 1989)—which neglect many of the two-electron integrals of *ab initio* methods—while significantly less expensive than AIMS, often yield erroneous energetics resulting in unphysical dynamics.

Therefore, in order to obtain globally accurate energetics, the parameter sets—which replace the non-neglected two-electron integrals of *ab initio* methods—have to be reoptimized for different classes of molecules using a very limited set of *ab initio* and experimental data (Owens, 2004; Toniolo, Thompson, & Martinez, 2004). The reparameterization strategy is a promising way to extend direct dynamics simulations of photochemistry to multi-picosecond time scales. However, the reoptimization problem is massively multimodal and involves multiple objectives such as minimizing the difference between calculated and predicted energies, gradients of energies, and stationary point geometries. Current methods, mostly based on a staged fixed-weight single-objective optimization, fall quite short of yielding globally correct PESs and thus can produce unphysical dynamics. Evolutionary algorithms, on the other hand, are robust search methods that simultaneously optimize multiple objectives, and hence, are particularly suited for rapid reparameterization of semiempirical parameters. Therefore, the purpose of this paper is to use multiobjective evolutionary algorithms for rapid reparameterization of semiempirical methods to obtain globally-correct excited-state dynamics.

This paper is organized as follows. In the next section, we provide a brief introduction to reparameterization of semiempirical methods. We then describe the multiobjective evolutionary algorithms used for reparameterization in section 3 followed by key results and discussion in section 4. We finally outline future work followed by summary and conclusions.

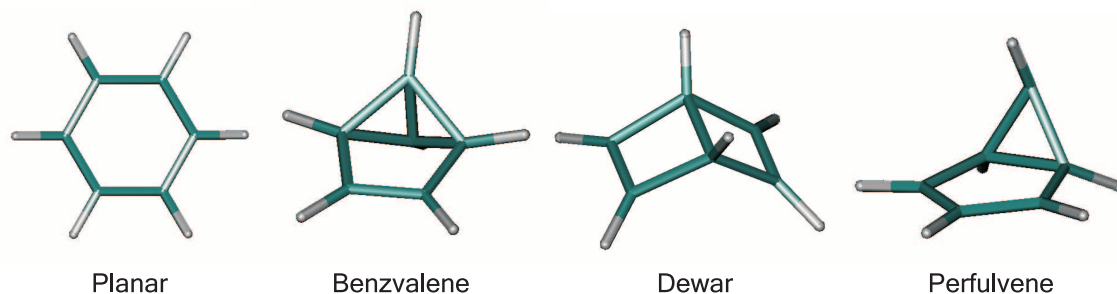


Figure 1: Ground state optimized geometries and important minimal energy conical intersections (MECIs) for benzene.

## 2 Reparameterization of Semiempirical Methods

In this section we provide a brief background of current computational methods for performing excited state dynamics in photochemistry and a more detailed overview is given elsewhere (Toniolo, Levine, Thompson, Quenneville, Ben-Nun, Owens, Olsen, Manohar, & Martinez, 2005; Owens, 2004; Toniolo, Thompson, & Martinez, 2004) and the references therein. As mentioned earlier, a comprehensive understanding of the photochemistry of molecules requires bridging the gap between molecular dynamics and quantum chemistry, and quantum dynamics simulations are required to simultaneously solve both the nuclear and electronic Schrödinger equations (Toniolo, Levine, Thompson, Quenneville, Ben-Nun, Owens, Olsen, Manohar, & Martinez, 2005). Additionally, the potential energy surfaces (PESs) must be of high quality and very robust because the portions of the PES that are critical to the behavior of the molecule may be far removed from the Franck-Condon region.

The *ab initio* multiple spawning (AIMS) method has been developed in order to address such problems (Ben-Nun & Martinez, 2002; Ben-Nun, Quenneville, & Martinez, 2000). While the AIMS method is extremely flexible and can describe quantum mechanical phenomena such as tunneling and non-adiabatic transitions, it is computationally very expensive because of a large number of *ab initio* electronic structure calculations involved, making long time dynamics simulations highly improbable, if not impossible.

In order to retain the flexibility of *ab initio* electronic structure methods with less computational cost, semiempirical methods—which ignore some two-electron integrals and use parameters for others—were developed (Dewar & Thiel, 1977; Dewar, Zoebisch, Healy, & Stewart, 1985; Stewart, 1989). The semiempirical parameters which are different for each element, have been optimized using ground state properties for a set of molecules. Standard parameter sets, such as MNDO (Dewar & Thiel, 1977), AM1 (Dewar, Zoebisch, Healy, & Stewart, 1985), and PM3 (Stewart, 1989), yield useful information concerning the locations of the minimal energy conical intersections (MECIs) which often dominate photochemical reactions. However, they often yield erroneous energetics, resulting in unphysical dynamics. Therefore, the parameter sets must be reoptimized using a very limited set of *ab initio* and experimental data to obtain an acceptable and an accurate description of the photodynamics. Also, the reparameterization strategy is a promising way to extend direct dynamics simulations of photochemistry to multi-picosecond time scales. It is also reasonable to expect *transferability* of the parameter sets optimized on simple molecules such as ethylene and benzene to other complex molecules such as stilbene and phenylacetylene dendrimers. Furthermore, the reparameterization approach opens up the possibility of accurate simulations of photochemistry in complex environments such as proteins and condensed phases.

It should be noted that while the reparameterization procedure only fits energetics of a few important stationary molecular geometries, much larger portions of the PESs will be accessed during dynamics simulations. Therefore, the semiempirical methods have to incorporate enough of the fundamental chemical physics to generate at least qualitatively correct global PESs. While it is possible to include geometries and energetics of the MECIs in the reparameterization, the strategy of using relatively little *ab initio* data is mandatory if reparameterization is to be applicable for larger molecules, where *ab initio* data is extremely expensive to obtain. Therefore, we intentionally use a minimal set of energies and gradients at ground state optimized geometries in our reparameterization.

In this paper we will concentrate on reparameterizing two simple molecules, which are fundamental building blocks of organic molecules: Ethylene and benzene. The small size of ethylene has many advantages: First, semi-empirical calculations can be run very quickly so a large number of reparameterization runs can be conducted. Second, the small number of atoms, basis functions, and possible geometries imply that the results may be less complex and more easily interpretable. Lastly, the size and simplicity enables the reoptimized parameter sets to be amenable for further analysis of ethylene dynamics and for transferability to stilbene or conjugated polyenes. Furthermore, despite its simplicity, ethylene has an associated set of ethylidene geometries than can be used to evaluate performance of the reoptimized parameter sets in calculations for which they were not optimized. Benzene plays an important role in photochemistry and photophysics of aromatic systems and has been extensively studied both experimentally and theoretically (Toniolo, Thompson, & Martinez, 2004).

Similar to Owens (2004), for ethylene reparameterization, we use energetics for the ground state planar and ethylidene geometries, twisted geometry on the excited state as well as the gradients on the excited and ground states. The *ab initio* results used for reparameterization are taken from previously reported calculations (Ben-Nun & Martinez, 2000). As in (Toniolo, Thompson, & Martinez, 2004), for reparameterization of benzene, we use four important local minima on  $S_0$ : planar, Dewar benzene, prefulvene and benzvalene (see figure 1) and use *ab initio* calculations and experimental results reported in and used by Toniolo, Thompson, and Martinez (2004). The semiempirical calculations are performed with a developmental version of MOPAC2000 (Stewart, 1999), while the *ab initio* results are performed with MOLPRO (Werner et al., ) and MolCas (Andersson et al., 2001), details of which are beyond the scope of this paper. For both ethylene and benzene, 11 parameters for carbon— $U_{ss}$ ,  $U_{pp}$ ,  $\beta_s$ ,  $\beta_p$ ,  $\zeta_s$ ,  $\zeta_p$ ,  $G_{ss}$ ,  $G_{sp}$ ,  $G_{pp}$ ,  $G_{p_2}$ , and  $H_{sp}$ —are reoptimized. Following earlier studies (Owens, 2004; Toniolo, Thompson, & Martinez, 2004), the core-core repulsion parameters— $\alpha$ ,  $a_i$ ,  $b_i$ , and  $c_i$  are not reoptimized.

With this general overview of reparameterization of semiempirical methods, we describe the multiobjective genetic algorithm used in reparameterization in the next section.

### 3 Multiobjective Genetic Algorithms

Many practical problems are inherently multiobjective in nature and evolutionary algorithms are particularly suited to handle multiple objectives as they can process a number of solutions in parallel and find all or majority of the solutions in the Pareto-optimal front. Based on Goldberg’s (Goldberg, 1989) suggestion of implementing a selection procedure that uses the non-domination principle, many multiobjective evolutionary algorithms have been proposed (Deb, 2001; Coello Coello, Van Veldhuizen, & Lamont, 2002). In this study, we used NSGA-II (Deb, Pratap, Agrawal, & Meyarivan, 2002) and provide the details of the algorithm in the following paragraphs.

As mentioned earlier, reparameterization of semiempirical methods involves optimizing the semi-

empirical parameters based on a very limited set of *ab initio* and/or experimental data. We use a real-valued encoding to represent the 11 parameters— $U_{ss}$ ,  $U_{pp}$ ,  $\beta_s$ ,  $\beta_p$ ,  $\zeta_s$ ,  $\zeta_p$ ,  $G_{ss}$ ,  $G_{sp}$ ,  $G_{pp}$ ,  $G_{p_2}$ , and  $H_{sp}$ —of the semi-empirical potentials. The two fitness functions involve minimizing the absolute error in energies and energy-gradients for a very limited set of excited-state and ground-state configurations either calculated by *ab initio* methods or obtained by experiments and those predicted by semiempirical methods. That is,

$$f_1(\mathbf{x}) = \sum_{i=1}^{n_c} [|\Delta E_{0,i} - \Delta E_{SE,i}(\mathbf{x})| + \Delta G_{0,SE,i}(\mathbf{x})] \quad (1)$$

$$f_2(\mathbf{x}) = \sum_{i=1}^{n_g} |\Delta(\nabla E_{0,i}) - \Delta(\nabla E_{SE,i}(\mathbf{x}))| \quad (2)$$

where  $x$  represents the semiempirical parameters to be optimized,  $n_c$  is the number of configurations, and  $n_g$  is the number of gradient-energy data used in reparameterization.  $\Delta E_{0,i}$  and  $\Delta E_{SE,i}$  are the differences in energy between the geometry  $i$  and the reference structure (planar ethylene and benzene) calculated by *ab initio* and semiempirical methods respectively. It should be noted that in the first objective we also include geometry difference between the reparameterized semiempirical geometries and the *ab initio* geometries,  $\Delta G_{0,SE,i}$ , by calculating the sum-squared differences between the corresponding atoms after the molecules have been rotated and translated such that they are in maximum coincidence.  $\Delta(\nabla E_{0,i})$ , and  $\Delta(\nabla E_{SE,i})$  represent the excited-state energy gradients using *ab initio* and semiempirical methods respectively.

We use a population size of 800 in accordance with population-sizing models (Goldberg, Deb, & Clark, 1992; Harik, Cantú-Paz, Goldberg, & Miller, 1999; Mahfoud, 1994), the verification of which is provided in section 4. The initial population is randomly generated within a certain percentage (20–50%) of the PM3 parameter values (Stewart, 1989). We restrict the parameter bounds around the PM3 set so as to maintain a reasonable representation of the ground-state potential energy surface. In our implementation of NSGA-II for reparameterization of semiempirical methods, we use a binary ( $s = 2$ ) tournament selection without replacement (Goldberg, Korb, & Deb, 1989; Sastry & Goldberg, 2001), simulated binary crossover (SBX) (Deb & Agarwal, 1995; Deb & Kumar, 1995)—which models the behavior of single-point crossover in binary genetic algorithms—with  $\eta_c = 5$ , and crossover probability  $p_c = 0.9$ , and a polynomial mutation (Deb, 2001) with  $\eta_n = 10$  and mutation probability  $p_m = 0.1$ .

## 4 Results

We demonstrate the effectiveness of using multiobjective genetic algorithm in rapid reparameterization of semiempirical methods for ethylene and benzene. We begin with estimating population-sizing and run-duration requirements and then compare the performance of the evolutionary approach in predicting globally accurate PESs—specifically on critical and untested excited states—with previously published results.

Since the fitness calculations for ethylene are reasonably fast—about 2 seconds per evaluation on a 1.7 GHz AMD Athlon XP workstation—we first verify the population-sizing and run-duration requirements using limited number of NSGA-II runs. In order to verify population-sizing requirements, we ran 5 independent runs of NSGA-II with a population size of 2000 for 200 generations and used the best non-dominated set out of those 5 runs as an approximation of the true Pareto-optimal front, which contains 61 distinct solutions. Using the population-size model for niching (Mahfoud, 1994), we compute that the population size required to maintain at least 1 copy of each

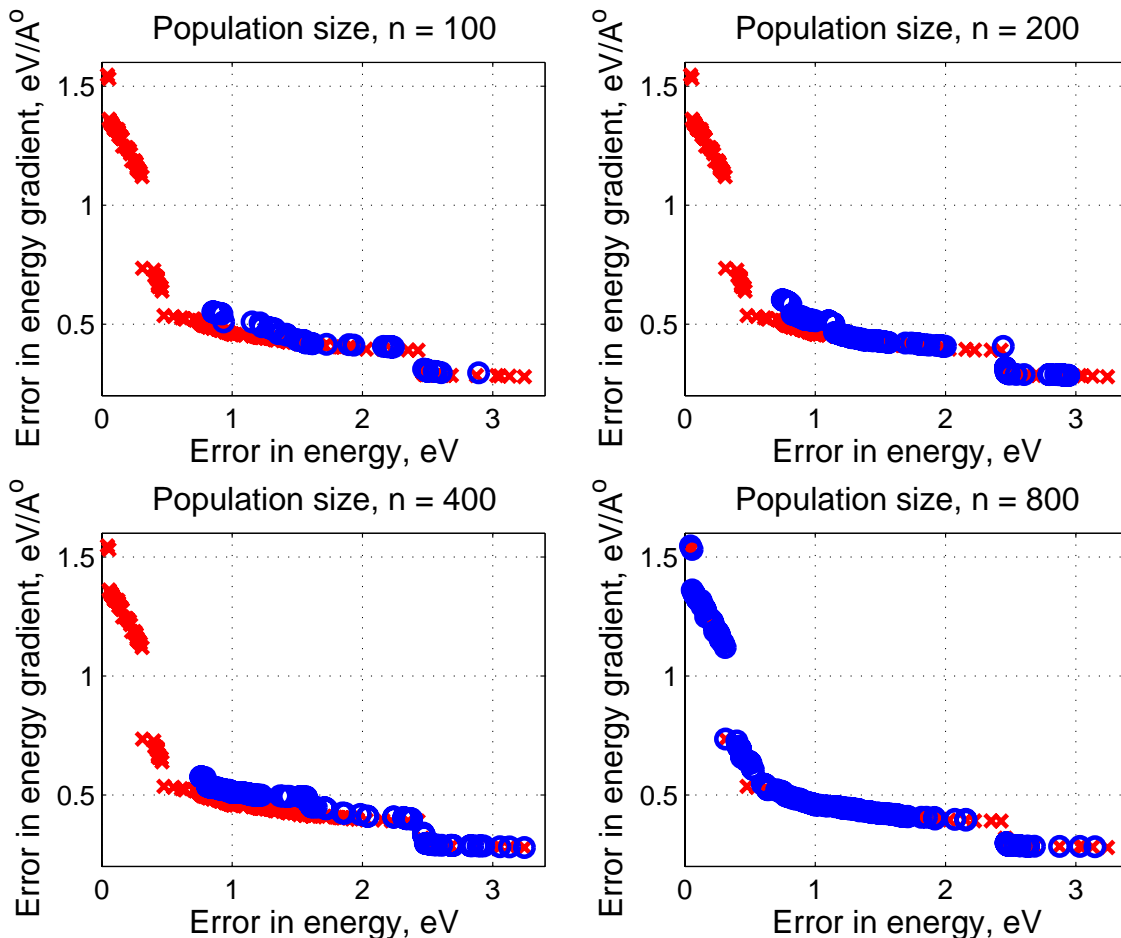


Figure 2: Effect of different population sizes on the convergence and coverage of the multi-objective GA. The results are shown for ethylene and are averaged over 10 independent runs. The results show that population sizes below 800 are not capable of converging onto the entire Pareto-front. The empirical results agree with the population size estimate of 750 predicted by the population-sizing model of Mahfoud (1994). Points denoted by crosses are obtained with a population of 2000 run for 200 generations and the points represented by circles are the best non-dominated solutions at population sizes of 100, 200, 400, and 800.

of the Pareto-optimal points with a probability of 0.98 to be 750. To verify this estimate we ran 10 independent runs of NSGA-II with population sizes between 50–800 with a fixed number of function evaluations of 80,000 for each run. The performance of NSGA-II with different population sizes are shown in figure 2. As shown in figure 2, while NSGA-II with population sizes below 800 are unable to converge to the approximate Pareto-optimal front, NSGA-II with a population size of 800 discovers almost all the Pareto-optimal points.

We now look at the convergence rate of NSGA-II and the run-duration requirements for reparameterization. Specifically, we considered 10 independent runs of NSGA-II with a population size of 800 and looked at the evolution of the best non-dominated front at different generations of the evolutionary process as shown in figure 3. The results show that reasonably good quality solutions start appearing as early as 10<sup>th</sup> generation and the solution quality improves at a steady pace till about 25 generations and gradually up to about 100 generations. We found that after about 100

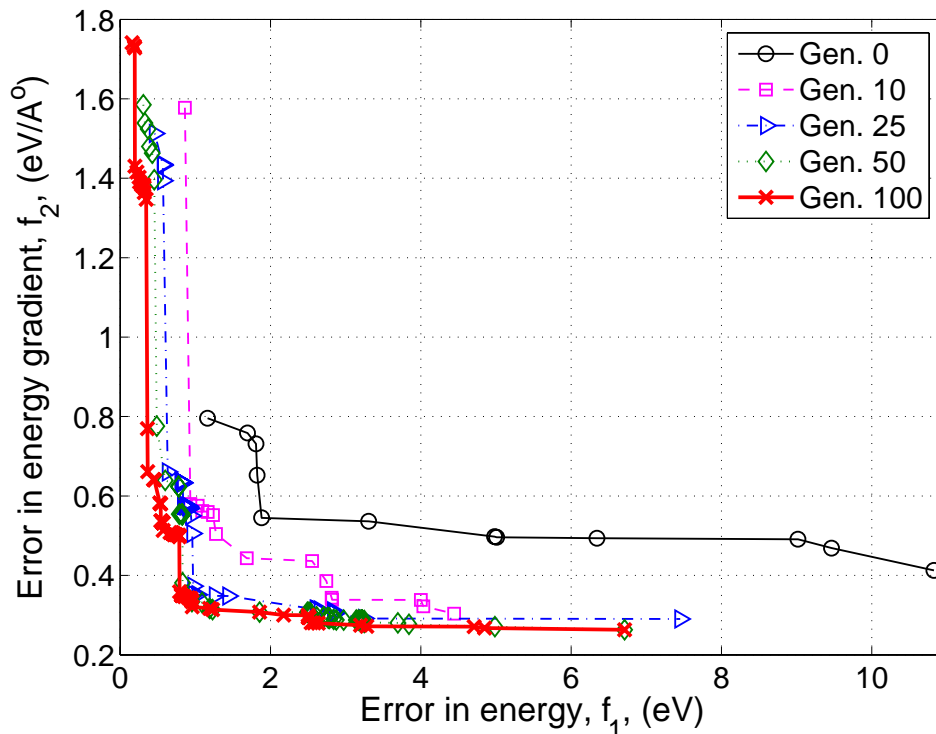


Figure 3: Convergence of NSGA-II for reparameterization of semiempirical parameters for ethylene. The best non-dominated front out of 10 independent runs are shown at five different generations. Solutions of reasonable quality start appearing in about 25 generations and high-quality solutions are discovered somewhere between 50–100 generations.

generations the improvement in solution quality was minimal.

Based on population-sizing and run-duration requirements in the remainder of the results we used a population size of 800 and ran NSGA-II for 100 generations. Moreover, the number of decision variables (semiempirical parameters) remain the same with different molecules involving carbon and hydrogen and the population-sizing and run-duration estimates should hold for reparameterization for those molecules as well. However, we note that the evaluation time increases with the complexity of the molecule under consideration.

We begin with comparing the solution qualities provided by the best non-dominated front of NSGA-II over the current published results of Owens (2004) for ethylene in figure 4. As shown in the figure, the solutions obtained through the genetic algorithm is significantly superior, both in terms of error in energy and energy-gradient, than those previously reported. Specifically, we obtain solutions that are 226% lower error in the energy and 32.5% lower error in the energy gradient. Additionally, one of the best points reported by Owens (2004) actually yields an inaccurate potential energy surface. In contrast, all 45 distinct solutions in the best non-dominated set with error in energy lower than 1.2 eV yield globally accurate PESs. All the unphysical points obtained through evolutionary approach have an error in energy greater than 1.23 eV.

We now consider solutions obtained through the GA and that of Owens (2004) with error in energy less than 2 eV, and evaluate their results on energetic calculations for a set of ethylidene geometries for which they were not reoptimized. Before comparing the results of GA with those of Owens, we provide certain salient properties of *cis-trans* isomerization of ethylene. The ground

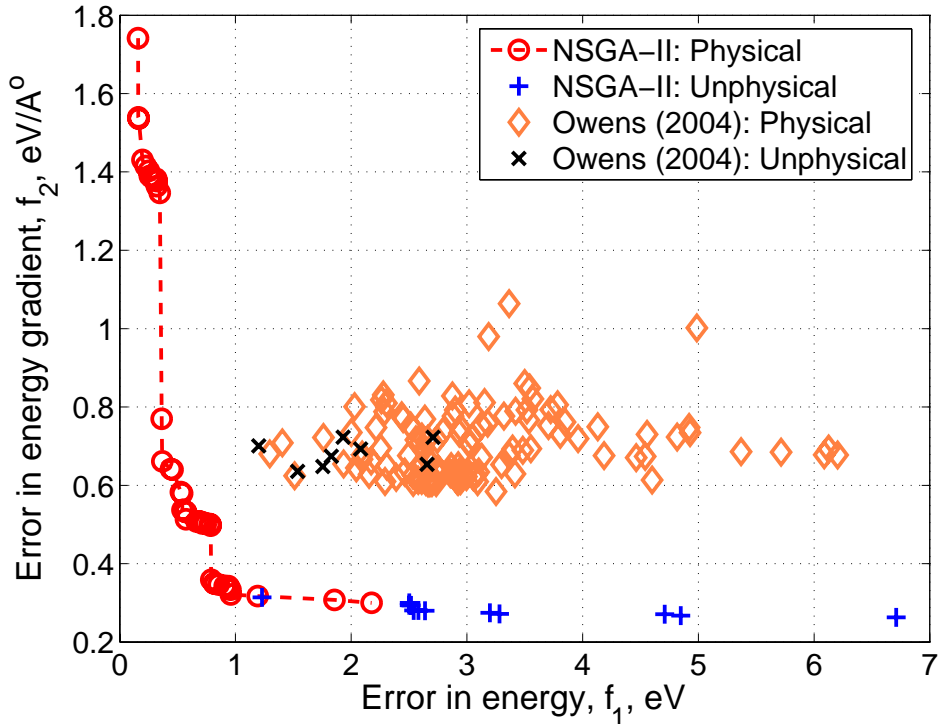


Figure 4: The best non-dominated front after 100 generations for ethylene compared to the published results (Owens, 2004). The GA results are for population size  $n = 800$ , and are averaged over 30 independent runs. The results obtained through GAs are significantly better—226% lower error in the energy, and 32.5% lower error in the energy gradient—than existing reparameterized sets.

state for ethylene is a planar structure as shown in figure 5. When it is excited, the carbon-carbon bond twists  $90^\circ$  and decreases in the energy gap from 7.8 eV to 2.5 eV. The twisted geometry, however is not an excited state minimum but a saddle point with respect to pyramidalization of one of the carbon atoms. However, as shown in figure 5, the PM3 and AM1 parameter sets incorrectly indicate that the twisted geometry is excited state minimum.

Therefore, we consider the above two important energetics, the results of which are shown in figure 6:

- Energy differences between planar ethylene (ground state,  $S_0$  minimized  $D_{2h}$ ) and twisted geometry ( $S_1$  minimized  $D_{2d}$ ), ideal value for which is 2.28 eV as calculated by *ab initio* methods (Ben-Nun & Martinez, 2000). If the energy difference between the planar and twisted geometry is less than zero, than the excited state minimum would be the planar structure, which is erroneous. In other words, for good parameter sets, the energy difference between the planar and twisted geometry should be greater than zero, preferably around 2.28 eV.
- Energy differences between the twisted geometry ( $S_1$  minimized  $D_{2d}$ ) and pyramidalized structure, ideal value for which is 0.88 eV as calculated by *ab initio* methods (Ben-Nun & Martinez, 2000). As shown in figure 5 the standard semiempirical parameter sets do not capture this feature, and therefore, this energetics is one of the critical phenomena in determining the quality of the reoptimized parameter sets. If the energy difference between

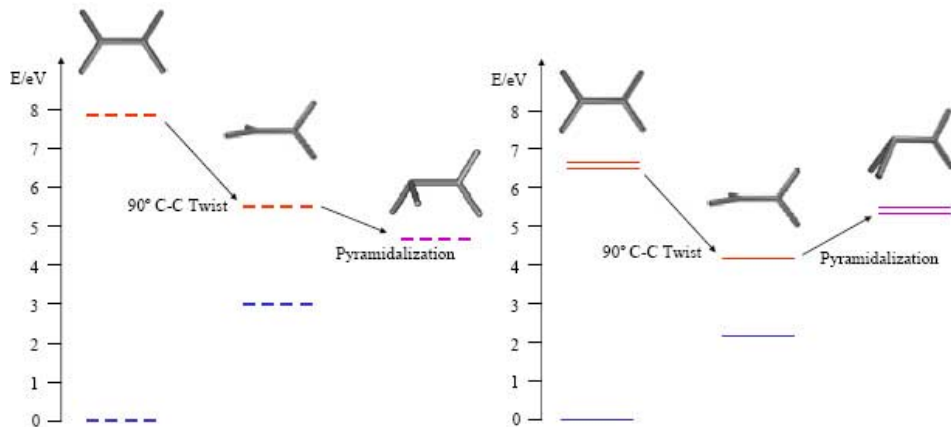


Figure 5: Energy levels of ethylene at various geometries: Planar, twisted, and pyramidalized (left) from *ab initio* calculations and (right) from PM3 and AM1 calculations.

the twisted geometry and the pyramidalized structure is less than zero, then the excited state minimum would be the twisted geometry (as predicted by standard parameter sets) which is inconsistent with *ab initio* and experimental results. Therefore, for good parameter sets, the energy difference between the twisted and pyramidalized geometries should be greater than zero, preferably around 0.88 eV.

The energy differences between planar and twisted geometry, and twisted geometry and pyramidalized structure, for both the best non-dominated set are shown in figure 6 along with the corresponding solutions. As shown in figure 6, the best non-dominated solutions with error in energies lower than 0.8 eV yield near ideal energies for both excited-state transitions. Indeed, during the evolutionary process, we find about 1,247 distinct parameter sets other than the best non-dominated ones that demonstrate near-ideal energetics. In essence, the genetic-algorithm re-optimized parameter sets correctly identify the lowest-energy excited state as the pyramidalized structure as opposed to standard semiempirical parameter sets and some of the previously reported reparameterized sets. In contrast, the energetics of the solutions reported by Owens (2004) deviate significantly from the ideal values. It should be noted that for the unphysical points in both solutions sets, the energy difference between the twisted and pyramidalized geometries was greater than -0.034 eV, which is well within the error bars of the *ab initio* methods.

To verify the effectiveness of NSGA-II, we tested reparameterization on benzene which is a more complex than ethylene. The results for benzene reoptimization are shown in figure 7. Similar to the results obtained for ethylene, we observe that the GA provides significant improvement—46% lower error in the energy and 86.5% lower error in the energy gradient—over previously reported results (Toniolo, Thompson, & Martinez, 2004). Furthermore, 75 out of 82 distinct best non-dominated solutions with error in energy less than 8 eV yield physically accurate dynamics. Similar to ethylene, while the standard semiempirical parameter sets yield unphysical dynamics, the genetic-algorithm reoptimized parameter sets yield results consistent with experiments and *ab initio* computations. For example, the newly optimized parameter sets predict an  $S_2$  lifetime of 100 fs, in agreement with experiment (Radloff, Stert, Freudenberg, Hertel, Jouvét, Dedonder-Lardeux, & Solgadi, 1997).

We have also conducted dynamics simulations using a limited set of most distinct and best non-dominated parameter sets for both ethylene and benzene, results of which are beyond the scope of this paper. In essence, the dynamics simulations indicate that for each of the reoptimized parameter sets, we obtained results that are in accord with *ab initio* and experimental results. The

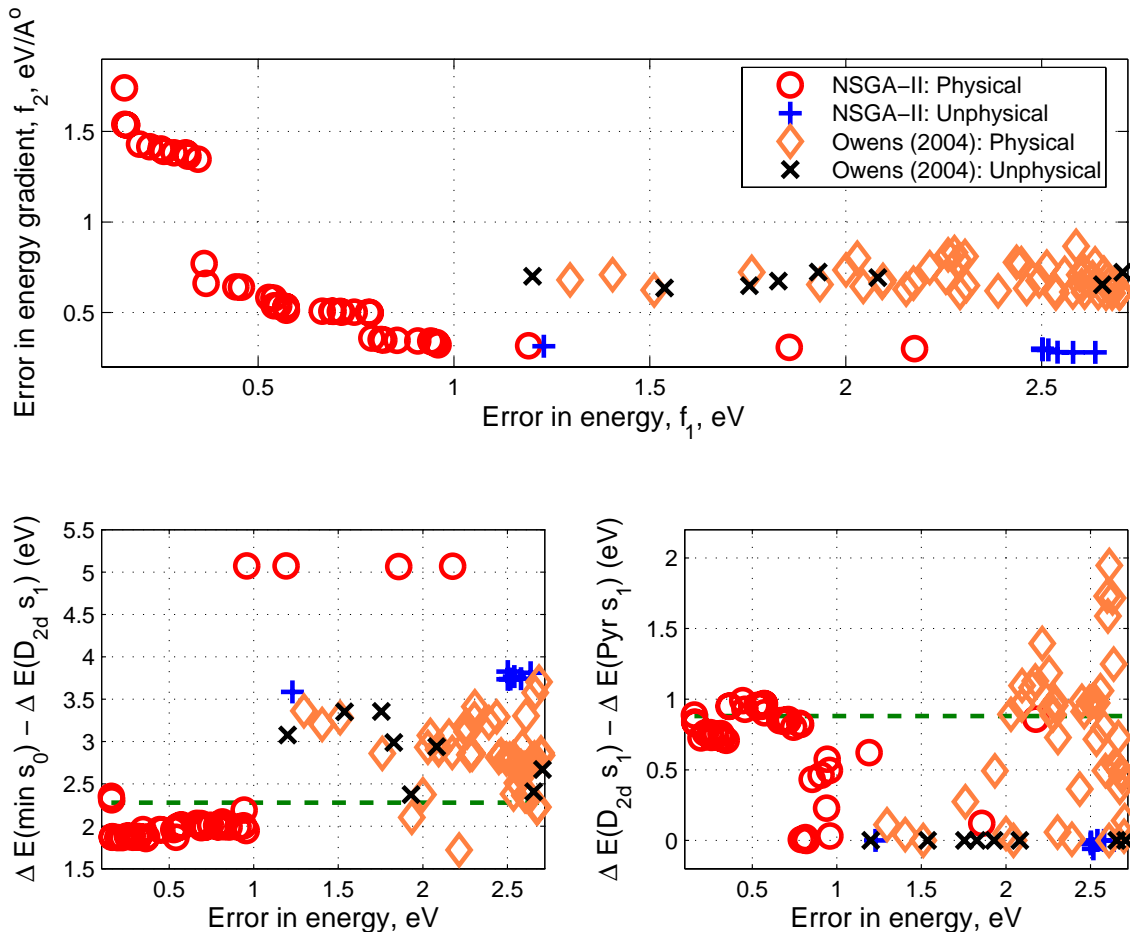


Figure 6: (Top) The best non-dominated sets and solution set reported by Owens (2004) with error in energy less than 2 eV reproduced from figure 4. (Bottom left) the energy difference between planar and twisted geometries. (Bottom right) The energy difference between twisted and pyramidalized geometry. The points on the best non-dominated set—especially those with low error in energy and energy-gradient—are near ideal value (shown by dashed lines) for energy difference between  $D_{2d} S_1$  and pyramidalized  $S_1$  configuration of 0.9 eV. Furthermore, the unphysical points with error in energy less than 2.0 eV, the energy difference between the  $D_{2d} S_1$  and pyramidalized  $S_1$  configurations is greater than -0.034 eV, which is well within the error bars of the *ab initio* methods.

results demonstrate the effectiveness of evolutionary search in providing multiple parameter sets for semiempirical methods that yield globally correct PESs even on untested and potentially critical reaction paths.

## 5 Future Work

We have taken the first step towards multiscaling excited state direct dynamics of photochemistry. Our main goal is to obtain high-quality reoptimized semiempirical parameters that are *transferable* to other complex molecules containing similar properties as ethylene and benzene and our limited tests so far have shown that this is indeed the case. We are now in the process of systematically

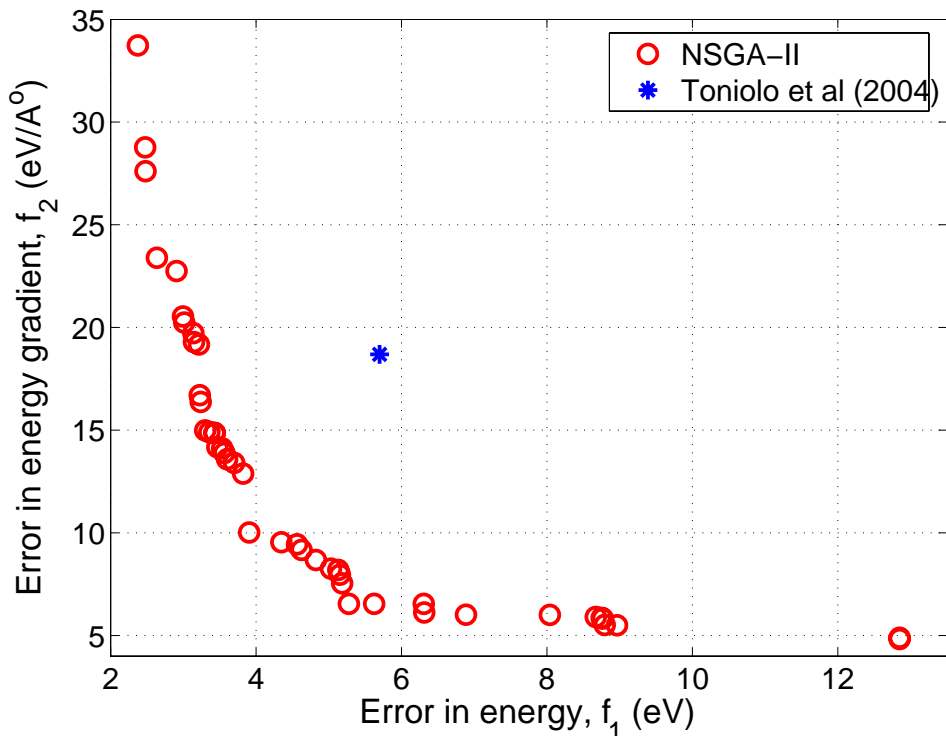


Figure 7: The best non-dominated front after 100 generations for benzene compared to the published results (Toniolo, Thompson, & Martinez, 2004). NSGA-II results are for population size  $n = 800$ , and are averaged over 10 independent runs. The results obtained through GAs are significantly better—226% lower error in the energy, and 32.5% lower error in the energy gradient—than existing reparameterized sets.

testing the transferability of the reoptimized parameter sets of ethylene and benzene to other molecules.

In this study we used an fixed evolutionary operators that do not adapt linkage, and the evolutionary algorithm might not be scalable in reparameterizing semiempirical methods for other complex molecules. One class of evolutionary algorithms that might be particularly effective—especially for complex molecules—might be the so called *competent* genetic algorithms (Goldberg, 2002; Pelikan, 2005).

The multiple reoptimized parameter sets of semiempirical methods all yield globally correct potential energy surfaces. Therefore, the interactions between the parameters contain important information about the molecules. Understanding these key interactions are central to developing a robust multiscale method. We have used traditional clustering techniques, the model-building process of Bayesian optimization algorithm (Pelikan, 2005; Ahn, 2005), and symbolic regression via genetic programming to understand the parameter interactions.

## 6 Summary and Conclusions

In this study, we investigated the use of genetic and evolutionary algorithms in multiscale simulations of excited state dynamics in photochemistry. Specifically, we used multiobjective genetic algorithms to rapidly reparameterize semiempirical methods to produce globally accurate potential

energy surfaces based on very limited *ab initio* and/or experimental data. Such reparameterization not only eliminates the need for a full-fledged *ab initio* dynamics simulation, which is prohibitively expensive for large molecules, but also eliminates drawbacks of semiempirical methods that use standard parameter sets and can yield unphysical dynamics. The results show that the evolutionary approach provides significantly better results—with up to 230% lower error in the energy and 86.5% lower error in the energy gradient—than those reported in literature. Furthermore, it also provides a large number of parameter sets, all of which yield globally accurate PESs and physical dynamics.

In the spirit of semiempirical methods, it may be reasonable to expect transferability to ethylene- and benzene-containing molecules, such as stilbene and phenylacetylene dendrimers. Our initial tests on transferability shows that this is indeed the case and that parameter sets optimized for ethylene are applicable to benzene, and vice versa. This is an ultimate goal of this work, since it would allow direct simulation of photoinduced cis-trans isomerization in molecules such as stilbene and azobenzene, as well as energy transfer in dendrimeric molecules. Furthermore, this opens up the possibility of accurate simulations of photochemistry in complex environments such as proteins and condensed phases.

## Acknowledgments

This work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-03-1-0129, the National Science Foundation under grant ITR grant DMR-03-25939 at the Materials Computation Center. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the National Science Foundation, or the U.S. Government.

## References

- Ahn, C. W. (2005). *Theory, design, and application of efficient genetic and evolutionary algorithms*. Doctoral dissertation, Gwangju Institute of Science and Technology, Korea.
- Andersson, K., et al. (2001). MOLCAS version 5.2. Lund University, Sweden.
- Ben-Nun, M., & Martinez, T. J. (2000). Photodynamics of ethylene: *Ab initio* studies of conical intersections. *Chemical Physics*, 259, 237–248.
- Ben-Nun, M., & Martinez, T. J. (2002). *Ab Initio* quantum molecular dynamics. *Advances in Chemical Physics*, 121, 439–512.
- Ben-Nun, M., Quenneville, J., & Martinez, T. J. (2000). *Ab Initio* multiple spawning: Photochemistry from first principles quantum molecular dynamics. *Journal of Physical Chemistry A*, 104(22), 5161–5175.
- Coello Coello, C. A., Van Veldhuizen, D. A., & Lamont, G. B. (2002). *Evolutionary algorithms for solving multi-objective problems*. Boston, MA: Kluwer Academic Publishers.
- Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Chichester, UK: John Wiley and Sons.

- Deb, K., & Agarwal, R. (1995). Simulated binary crossover for continuous search space. *Complex Systems*, 9, 115–148.
- Deb, K., & Kumar, A. (1995). Real-coded genetic algorithms with simulated binary crossover: Studies on multimodal and multiobjective problems. *Complex Systems*, 9, 431–454.
- Deb, K., Pratap, A., Agrawal, S., & Meyarivan, T. (2002). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. (Also KanGAL Report No. 2000001).
- Dewar, M. J. S., & Thiel, W. (1977). The MNDO method. Approximations and parameters. *Journal of The American Chemical Society*, 99(15), 4899–4907.
- Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., & Stewart, J. J. P. (1985). AM1: A new general purpose quantum mechanical molecular model. *Journal of The American Chemical Society*, 107(13), 3902–3909.
- Goldberg, D. E. (1989). *Genetic algorithms in search optimization and machine learning*. Reading, MA: Addison-Wesley.
- Goldberg, D. E. (2002). *Design of innovation: Lessons from and for competent genetic algorithms*. Boston, MA: Kluwer Academic Publishers.
- Goldberg, D. E., Deb, K., & Clark, J. H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, 6, 333–362. (Also IlliGAL Report No. 91010).
- Goldberg, D. E., Korb, B., & Deb, K. (1989). Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, 3(5), 493–530. (Also IlliGAL Report No. 89003).
- Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1999). The gambler’s ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation*, 7(3), 231–253. (Also IlliGAL Report No. 96004).
- Mahfoud, S. W. (1994). Population size and genetic drift in fitness sharing. *Foundations of Genetic Algorithms*, 3, 185–224. (Also IlliGAL Report No. 94005).
- Owens, J. M. (2004). *Theoretical studies of the solvation, dynamics, and photochemistry of ethylene, retinal protonated Schiff base, oligocellulose, and Gd(III) clusters*. Doctoral dissertation, University of Illinois at Urbana-Champaign, Department of Chemistry, Urbana, IL.
- Pelikan, M. (2005). *Hierarchical bayesian optimization algorithm: Toward a new generation of evolutionary algorithm*. Berlin: Springer Verlag.
- Radloff, W., Stert, V., Freudenberg, T., Hertel, I. V., Jouvét, C., Dedonder-Lardeux, C., & Solgadi, D. (1997). Internal conversion in highly excited benzene and benzene dimer: Femtosecond time-resolved photoelectron spectroscopy. *Chemical Physics Letters*, 21(1–3), 20–26.
- Sastry, K., & Goldberg, D. E. (2001). Modeling tournament selection with replacement using apparent added noise. *Intelligent Engineering Systems Through Artificial Neural Networks*, 11, 129–134. (Also IlliGAL Report No. 2001014).
- Sastry, K., Johnson, D., Goldberg, D. E., & Bellon, P. (2004). Genetic programming for multiscale modeling. *Int. J. of MultiScale Comput. Eng.*, 2(2), 239–256.
- Sastry, K., Johnson, D., Goldberg, D. E., & Bellon, P. (2005). Genetic programming for multi-timescale modeling. *Physical Review B*, 72, 085438.
- Stewart, J. J. P. (1989). Optimization of parameters for semiempirical methods 1. Method. *Journal of Computational Chemistry*, 10(2), 209–220.

- Stewart, J. J. P. (1999). MOPAC 2000. Fujitsu Limited, Tokyo, Japan.
- Toniolo, A., Levine, B., Thompson, A. L., Quenneville, J., Ben-Nun, M., Owens, J., Olsen, S., Manohar, L., & Martinez, T. J. (2005). Photochemistry from first principles and direct dynamics. New York, NY: Marcel-Dekker.
- Toniolo, A., Thompson, A. L., & Martinez, T. J. (2004). Excited state direct dynamics of benzene with reparameterized multireference semiempirical configuration interaction methods. *Chemical Physics*, *304*, 133–145.
- Werner, H.-J., et al. MOLPRO, version 2002.2, a package of *ab initio* programs. <http://www.molpro.net>.