



# Fast and Accurate Reaction Dynamics via Multiobjective Genetic Algorithm Optimization of Semiempirical Potentials

<sup>a</sup>Kumara Sastry, <sup>b</sup>Alexis L. Thompson,  
<sup>c</sup>Duane D. Johnson, <sup>a</sup>David. E. Goldberg, <sup>b</sup>Todd J. Martinez

<sup>a</sup>Industrial and Enterprise Systems Engineering,

<sup>b</sup>Chemistry and Beckman Institute

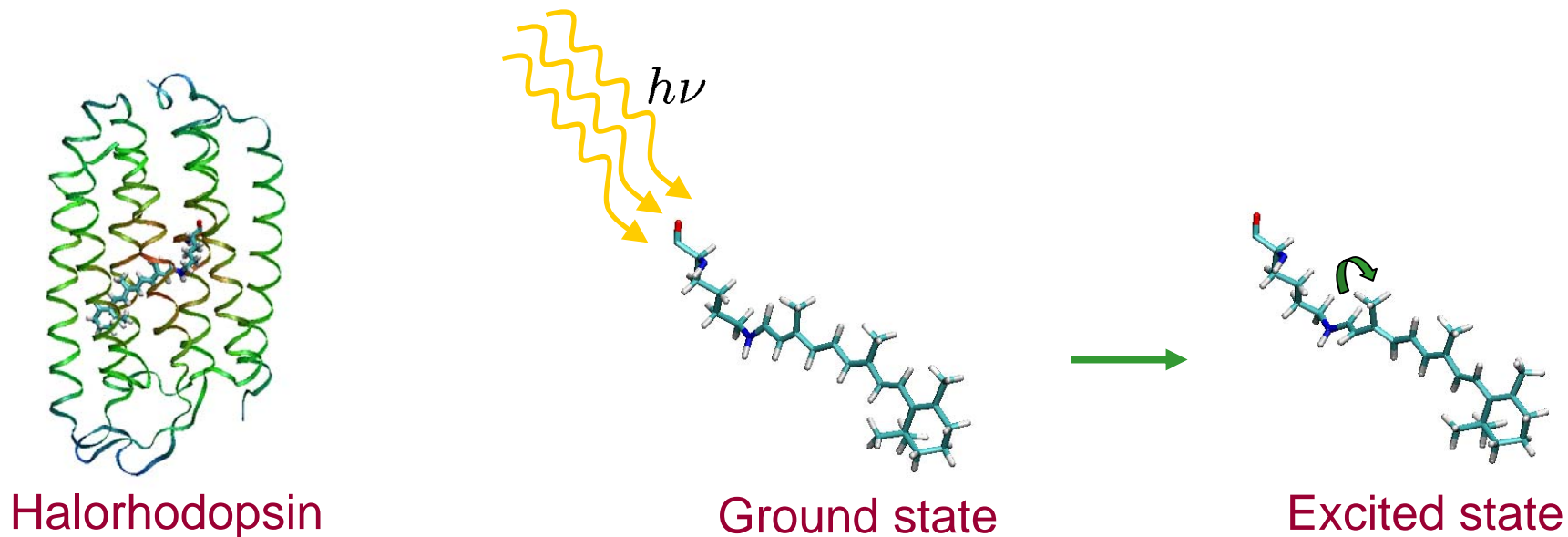
<sup>c</sup>Materials Science and Engineering

- ★ Silver Medal in Human Competitive Results at GECCO 2006 (ACM SIGEVO conference)
- ★ Best paper award, Real-world applications track, GECCO-2006 (ACM SIGEVO conference)



This work is supported by the Materials Computation Center (UIUC)  
NSF DMR 03-25939 and AFOSR FA9550-06-1-0096

# Modeling Photochemical Reactions



- ❖ Photochemistry important for photosynthesis, vision, solar energy

Need an accurate method that includes excited states

Dynamics requires many evaluations of the electronic energy

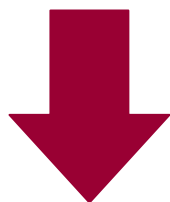
# Reaction Dynamics Over Multiple Timescales

*Ab Initio*  
Quantum Chemistry  
Methods

Accurate but slow (hours-days)  
Can calculate excited states

Semiempirical  
Methods

Fast (seconds-minutes)  
Calculate integrals from fit parameters  
Accuracy depends on parameters



Origin of our collaboration: Can we use optimization methods to do this reliably?

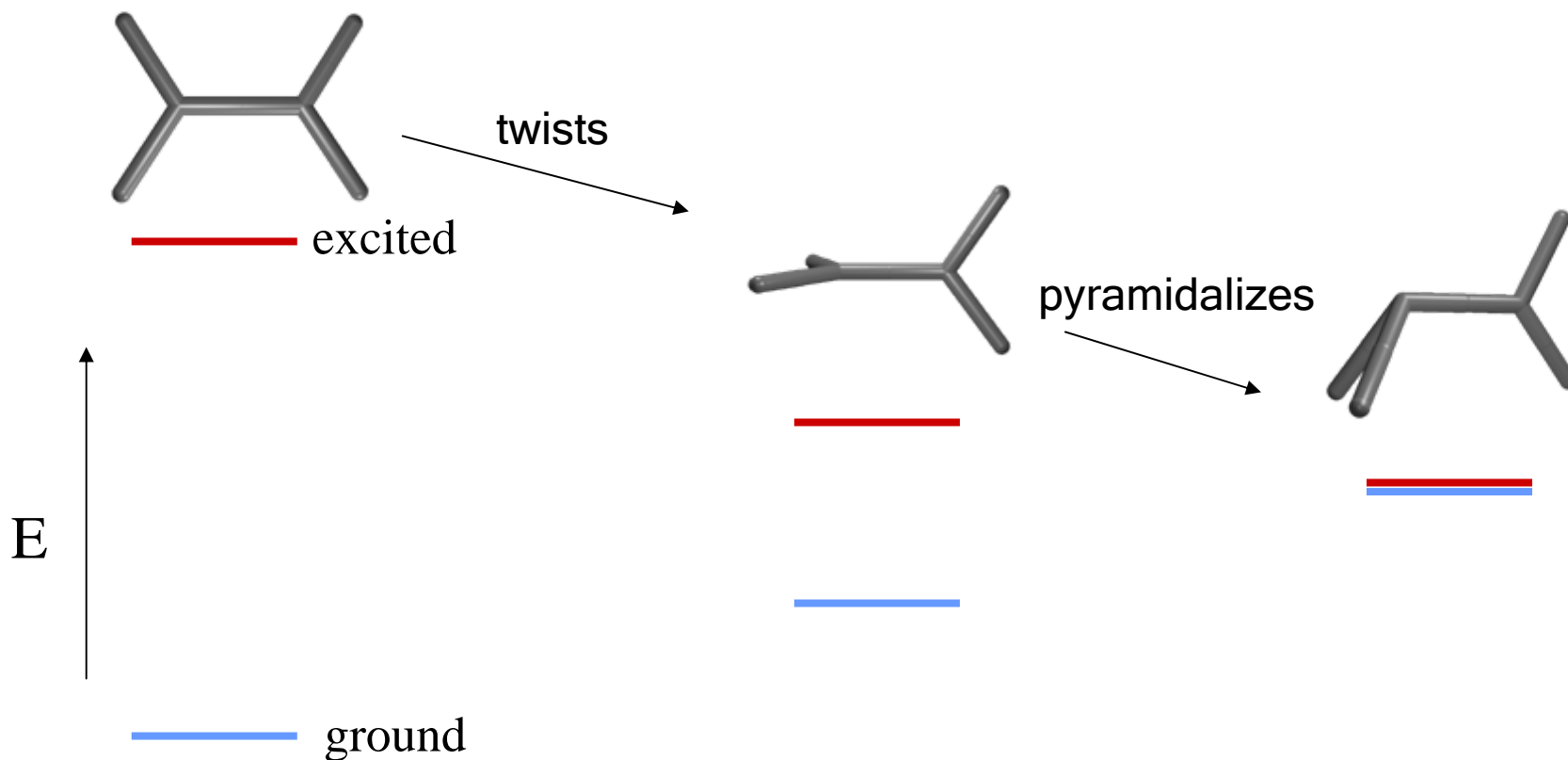
Can we get reasonable excited state surfaces with the speed of a semiempirical (SE) method?

This would allow us to study the dynamics of much larger systems: solvated chromophores, proteins, nanotubes...

# Methodology: Limited *Ab Initio* Results to Tune Semiempirical Parameters

- ❖ Standard parameter sets don't yield accurate potential energy surfaces (PESs)
  - ◆ Example: AM1, PM3, etc.
  - ◆ Parameterized for ground state properties
  - ◆ Calculate excited states using a CI technique within a small set of optimized orbitals
- ❖ **Parameter sets need to be reoptimized**
  - ◆ Optimize for a particular system, such as ethylene
  - ◆ Fit to only a few important geometries for the molecule
  - ◆ Maintain accurate description of ground-state properties
  - ◆ Yield globally accurate PES (including excited states)

# Ethylene Test Case for Reparameterization



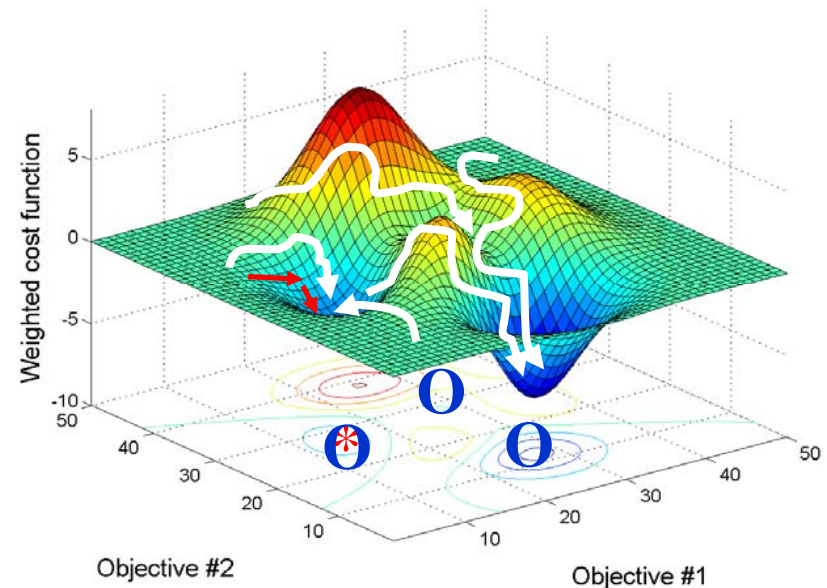
- ❖ Ethylene is the smallest molecule to show *cis-trans* isomerization
- ❖ Energetics of excited state involved pyramidalized intersection
- ❖ Semiempirical methods (AM1, PM3) do not reproduce this behavior
- ❖ Are they **transferable**? (Found for one system works for others.)

# Current Reparameterization Methods Fall Short; Need Multiobjective Optimization

- ❖ Current Method: Staged single-objective optimization
  - ◆ First minimize error in energies
  - ◆ Subsequently minimize weighted error in energy and gradient
- ❖ Multiple objectives and highly multimodal
  - ◆ Don't know the weights of different objectives
  - ◆ Local search gets stuck in low-quality optima

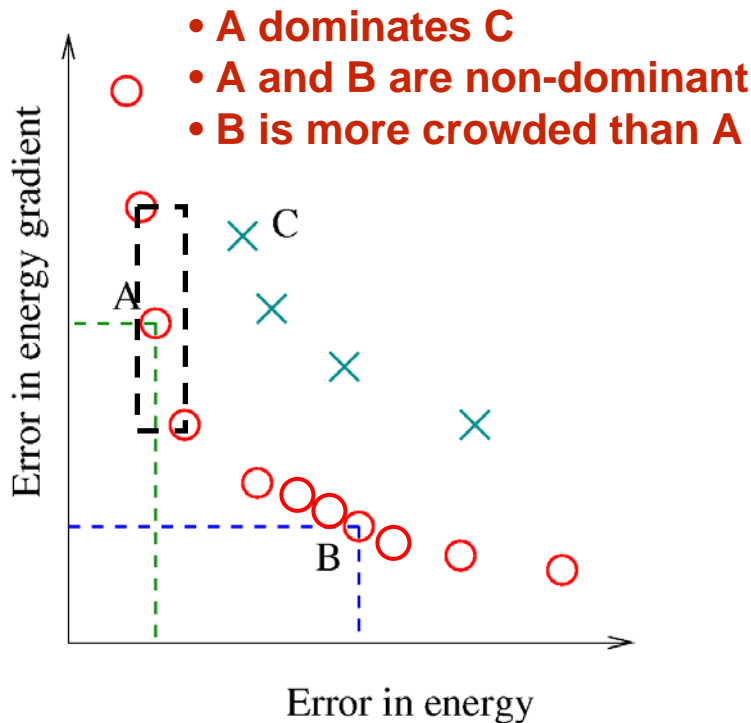
## Multiobjective optimization

- ❖ Simultaneously obtain “Pareto-optimal” solutions.
- ❖ *Avoid* potentially irrelevant and unphysical pathways.



# Multiobjective Optimization

- ❖ Unlike single-objective problems, multiobjective problems involve a set of optimal solutions often termed as **Pareto-optimal solutions**.
- ❖ **Notion of Non-Domination**



- ❖ **Solution A dominates C if:**
  - ◆ A is no worse than C in all objectives
  - ◆ A is better than C in at least one objective
- ❖ **Two goals:**
  - ◆ Converge onto the Pareto-optimal solutions (best non-dominated set )
  - ◆ Maintain as diverse a distribution as possible

# Multiobjective Genetic Algorithms (MOGAs)

- ❖ Search method based on “**evolutionary**” principles
- ❖ **Representation:** GAs operate on codes
  - ◆ Binary, gray, permutation, real, program
- ❖ **Fitness functions:** Relative quality measures of a solution
  - ◆ Objective, subjective, co-evolutionary
- ❖ **Population:** Candidate solutions (individuals)
- ❖ **Genetic operators:**
  - ◆ **Selection:** “Survival of the **non-dominated**”
  - ◆ **Recombination:** Combine parental traits to create offspring
  - ◆ **Mutation:** Modify an offspring slightly
  - ◆ **Replacement:** Replace parents with offspring

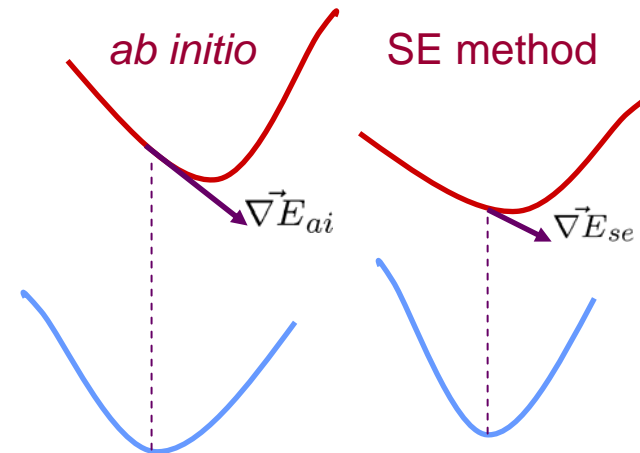
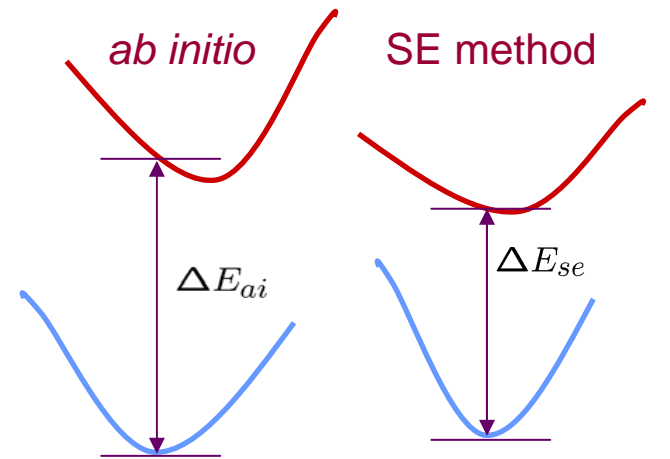
# Fitness: Errors in Energy and Energy Gradient

- ❖ Choose few ground- and excited-state configurations
- ❖ Fitness #1: Errors in energy
  - ◆ For each configuration, compute energy difference via *ab initio* and semiempirical methods

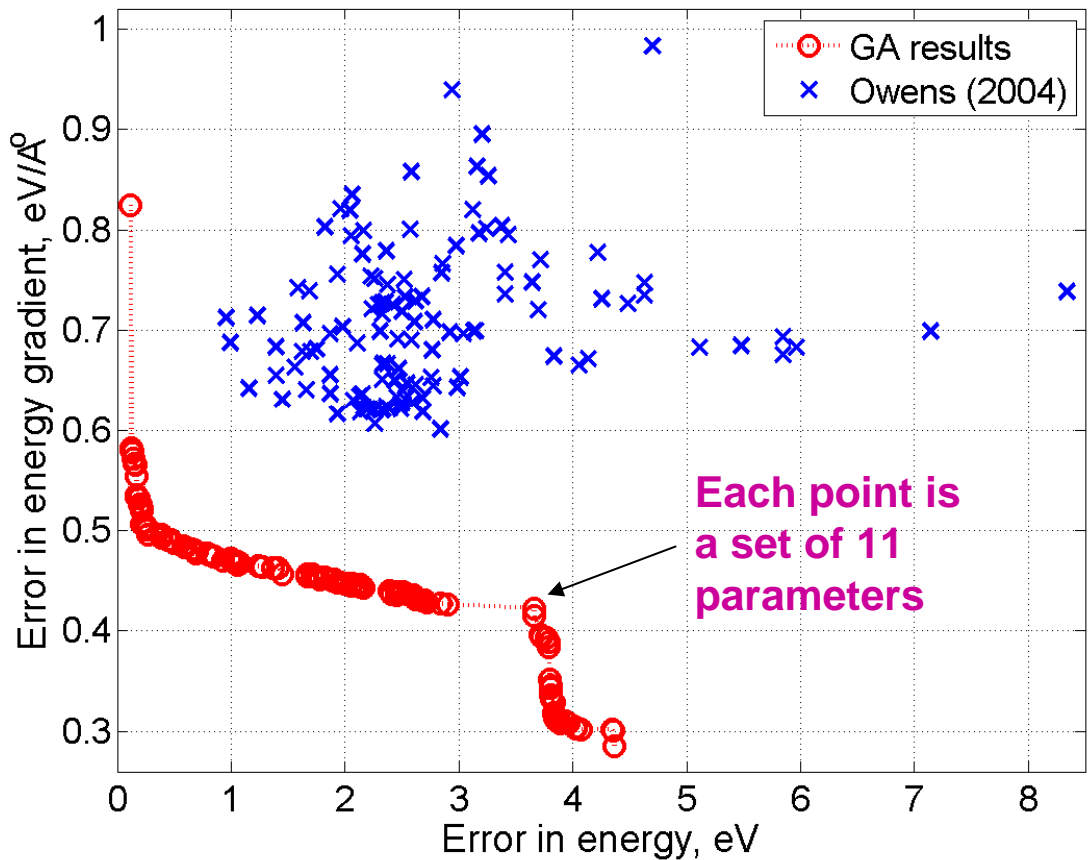
$$f_1 = \sum_1^{n_c} |\Delta E_{ai} - \Delta E_{se}|$$

- ❖ Fitness #2: Errors in energy gradient
  - ◆ For each configuration, compute energy gradient via *ab initio* and SE methods

$$f_2 = \sum_1^{n_c} |\vec{\nabla} E_{ai} - \vec{\nabla} E_{se}|$$



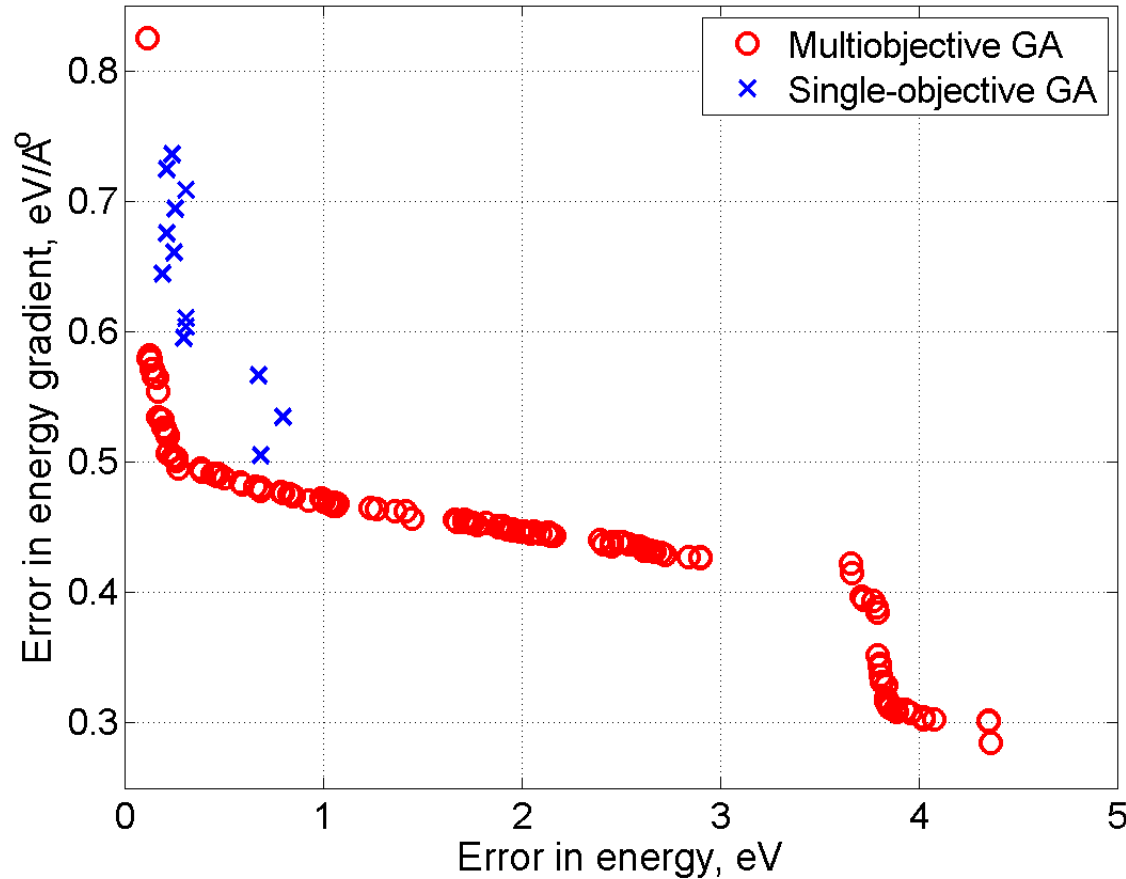
# MOGA Finds Physical and Accurate PES



MOGA  
VS  
Previously published  
results

- ❖ MOGA results have significantly lower errors than current results.
- ❖ Globally accurate PES yielding **physical** reaction dynamics

# Multiobjective Optimization Is Efficient



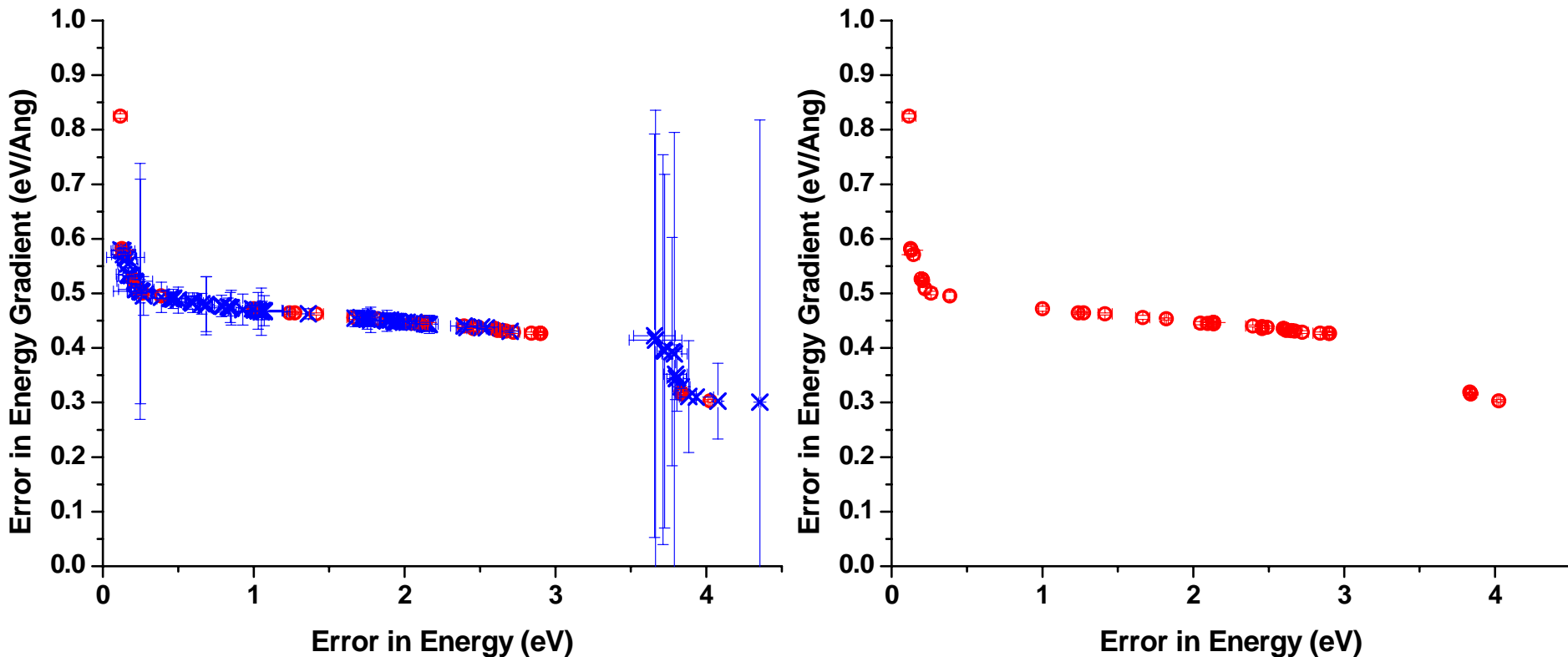
MOGA  
vs  
Multiple  
weighted  
single-objective  
GA runs

- ❖ **Not even one** Pareto-optimal solution obtained with multiple weighted single-objective optimizations

# Qualifying the Parameter Set

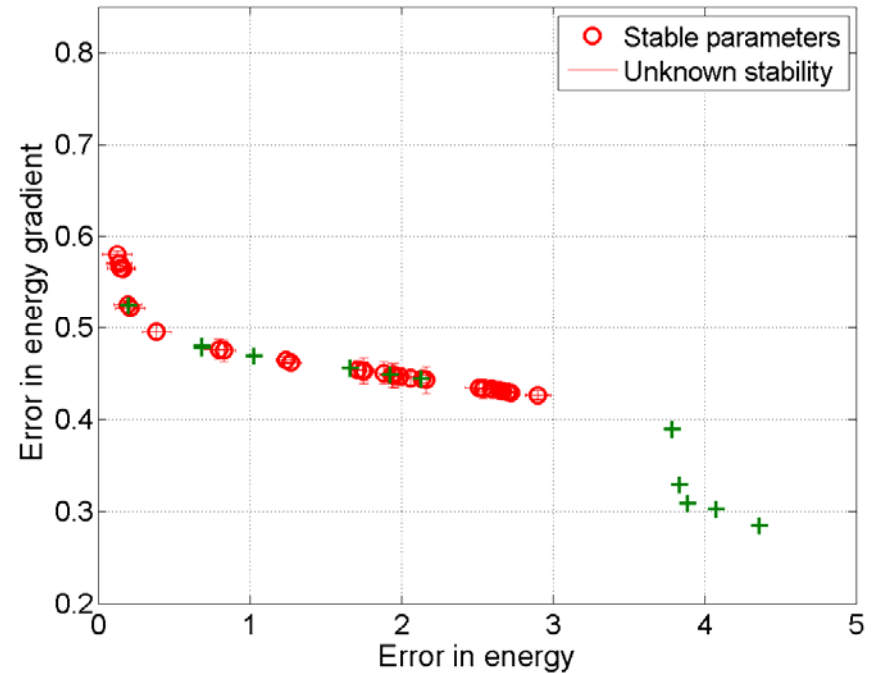
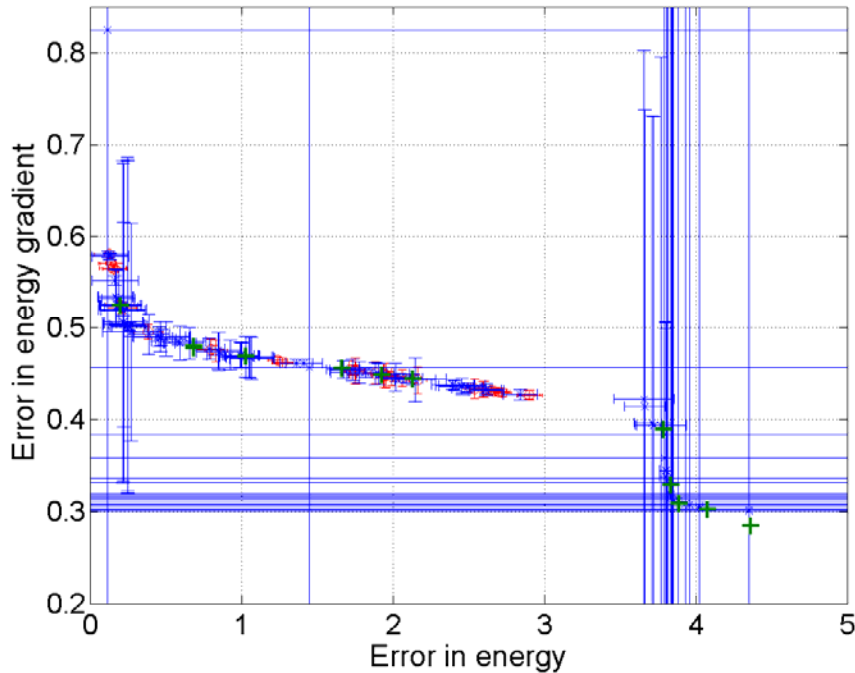
- ❖ Analyze the robustness of the parameter set
- ❖ Small changes in the parameters (0.1% or less) should have small effects on the error in energy and error in gradient
- ❖ Select ten parameters sets around each optimized parameter set and calculate error in energy and error in gradient for each set

# Eliminate Sensitive Parameter Sets (Manually)



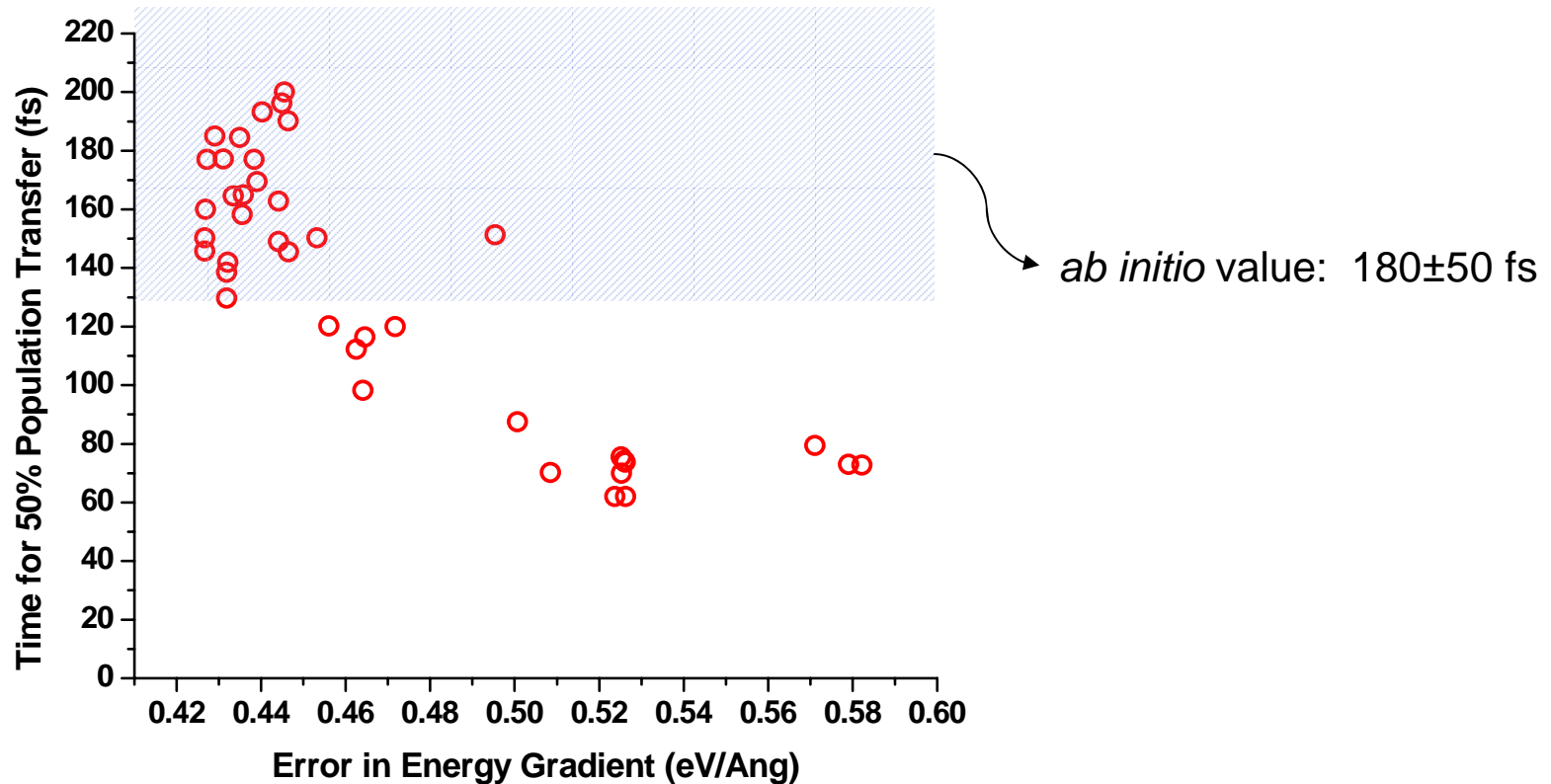
- ❖ Remove points that have large RMSD values
  - ◆ above 0.05 eV for error in energy
  - ◆ above 0.008 eV/Ang for error in energy gradient

# On-Line Parameter Stability Analysis in MOGA



- ❖ GA population contains data that can be mined, e.g., sensitivity of parameter sets
  - ◆ Analysis of quality of solutions around Pareto-optimal set yields a good measure of the SE parameter stability.
  - ◆ More perturb points in the MOGA analysis – higher reliability!

# Check potential energy surface with dynamics



- ❖ Population transfer determined using 20 initial conditions for each parameter set
- ❖ Parameter sets with lower error in energy gradient values have lifetimes close to *ab initio* value

# Key Results

- ❖ Yields multiple parameter sets that are **significantly higher quality than current published results.**
- ❖ Significantly outperforms single-objective optimization
- ❖ Enables  **$10^2$ - $10^5$**  increase in simulation time
- ❖  **$10$ - $10^3$**  times faster than current reparameterization method
- ❖ Population based search enables **on-line sensitivity analysis**
- ❖ **Currently investigating transferability of the SE parameters**
  - ◆ Enables accurate simulation of complex reactions without complete reoptimization
- ❖ **Pareto analysis using symbolic regression via GP**
  - ◆ Interpretable semiempirical potentials

# Outlook

- ❖ Broadly applicable in chemistry and materials science
  - ◆ Analogous applicability in modeling multiscaling phenomena.
- ❖ Facilitates fast and accurate materials modeling
  - ◆ Alloys: Kinetics simulations with *ab initio* accuracy.  $10^4$ - $10^7$  times faster than current methods [Sastry, et al (2006), *Phys. Rev. B*\*].
  - ◆ Chemistry: Reaction-dynamics simulations with *ab initio* accuracy.  $10^2$ - $10^5$  times faster than current methods.
- ❖ Lead potentially to new drugs, new materials, fundamental understanding of complex chemical phenomena
  - ◆ Science: Biophysical basis of vision, and photosynthesis
  - ◆ Synthesis: Pharmaceuticals, functional materials

★ chosen by the AIP editors as focused article of frontier research in *Virtual Journal of Nanoscale Science & Technology*, 12(9), 2005

# Software Contribution

## ❖ Genetic algorithms suite in C++

- ◆ Single & multiple objectives with and without constraints
- ◆ Array of selection, crossover, mutation, and other operators
- ◆ Optional interface with Matlab®
- ◆ Currently working on the documentation

## ❖ Other competent and efficient GAs (which acknowledge MCC) are available at <http://www-illigal.ge.uiuc.edu> and <http://medal.cs.umsl.edu/software.php>.

- ◆ Extended compact GA in C++, version 1.1
- ◆  $\chi$ -ary extended compact GA in C++
- ◆  $\chi$ -ary extended compact GA for Matlab in C++
- ◆ Generator of random additively decomposable problems

# Connections to MCC's Goals

## ❖ Cross-disciplinary collaborations

- ◆ Brought together a disparate, unique, and highly-qualified group to leverage expertise.

## ❖ Development of forefront algorithms

- ◆ Developed a novel method using MOGA that permits faster and reliable search of high-quality quantum-chemistry potentials.

## ❖ Development of extensible, user-friendly, open-source code

- ◆ Created a generic optimization and analysis code, which will be made available via Software Archive and IlliGAL sites.

## ❖ Enables analysis, understanding and prediction of properties

- ◆ Rapid and accurate simulation permits new science, synthesis, and facilitates bridging the gap between experiment and theory