JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

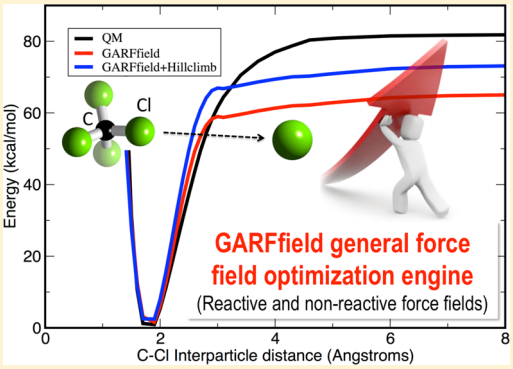
General Multiobjective Force Field Optimization Framework, with Application to Reactive Force Fields for Silicon Carbide

Andres Jaramillo-Botero,* Saber Naserifar, and William A. Goddard, III

Chemistry and Chemical Engineering Division, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, United States

Supporting Information

ABSTRACT: First-principles-based force fields prepared from large quantum mechanical data sets are now the norm in predictive molecular dynamics simulations for complex chemical processes, as opposed to force fields fitted solely from phenomenological data. In principle, the former allow improved accuracy and transferability over a wider range of molecular compositions, interactions, and environmental conditions unexplored by experiments. That is, assuming they have been optimally prepared from a diverse training set. The trade-off has been force field engines that are functionally complex, with a large number of nonbonded and bonded analytical forms that give rise to rather large parameter search spaces. To address this problem, we have developed GARFfield (genetic algorithm-based reactive force field optimizer method), a hybrid multiobjective Pareto-optimal parameter development scheme based on genetic algorithms, hill-climbing routines and conjugate-gradient minimization. To demonstrate the capabilities of GARFfield we use it to develop two very different force fields: (1) the ReaxFF reactive force field for modeling the adiabatic reactive dynamics of silicon carbide growth from an methyltrichlorosilane precursor and (2) the SiC electron force field with effective core pseudopotentials for modeling nonadiabatic dynamic phenomena with highly excited electronic states. The flexible and open architecture of GARFfield enables efficient and fast parallel optimization of parameters from quantum mechanical data sets for demanding applications like ReaxFF, electronic fast forward (or electron force field), and others including atomistic reactive charge-optimized many-body interatomic potentials, Morse, and coarse-grain force fields.



I. INTRODUCTION

The use of evolutionary algorithms (EA) in computational optimization of chemistry and biochemistry related problems has seen applications to ligand−protein docking,[1,2] molecular structure predictions,[3−5] and force field development,[6−10] among other applications.[11] In particular, genetic algorithms (GAs) enable searching of different regions of a solution space for problems involving nonconvex, discontinuous, and multimodal solutions. However, force field parameter solvers give the developer responsibility for establishing parameter search boundaries to avoid physically unreasonable or discontinuous potential energy surfaces (PES), while providing robust solutions suitable outside the training set.

Prior attempts to systematize this problem have been limited in scope and flexibility. Hunger and Huttner[6] combined neural networks and genetic algorithms to optimize specific force field parameters for tripod metal templates, tripodMO(CO)$_3$, using the root mean square deviation between the observed and force field predicted structures as the sole fitness function. Wang and Kollman[7] reported on an automatic engine for empirical force-field parametrization called Parmscan and applied it to reproduce the relative energies and conformations of small hydrocarbons by optimizing valence parameters in a conventional harmonic force field using systematic searches and a genetic algorithm. Angibaud et al.[8] reported on a hard-wired genetic algorithm (GA) for the parametrization of a reactive force field that included directional covalent bonds, dispersion terms, a charge transfer term, and atom coordinations; and used it to optimize a silicon−silicon force field. Handley and Deeth[9] describe the use and application of multiobjective evolutionary algorithms (MOEAs) by reparameterizing the ligand field molecular mechanics (LFMM) force field for modeling spin crossover in iron(II)−amine complexes. Recently, Larsson et al.[10] also reported on the use of an in-house code for unbiased global optimization using genetic algorithms to fit a ReaxFF[12] reactive force field for SiOH. For the most part, these efforts focus on applying evolutionary algorithms to optimizing parameters for specific (hard-wired) atomistic force field engines, do not enable Pareto-optimal solutions, rely on single or a limited number of objective fitness functions, and on energy minimization as means to establish an error norm with reference values.

Here, we introduce GARFfield, a genetic algorithm-based reactive force field optimizer framework. GARFfield is a hybrid evolutionary, gradient-based parameter optimization framework that allows multiobjective fitness function specification,

Received: February 7, 2014
Published: March 18, 2014

   dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

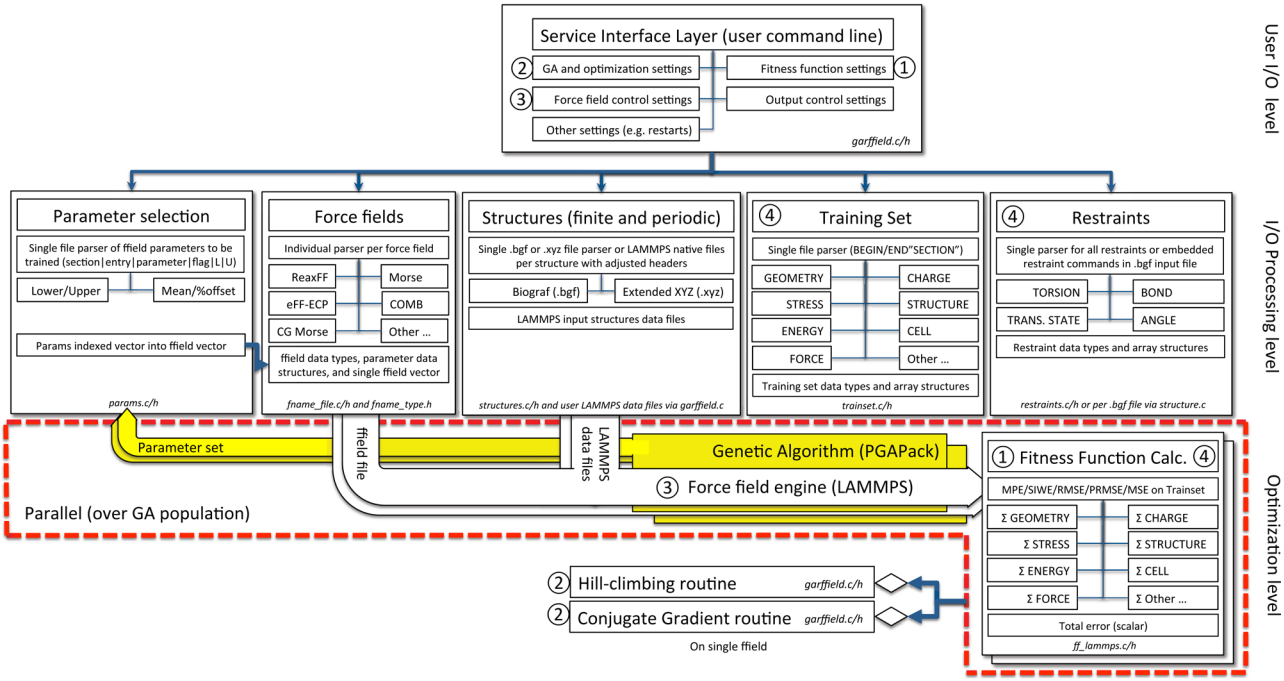


Figure 1. Software architecture of GARFfield. Circled numbers indicate connections 1−1, 2−2, 3−3, and 4−4 between functional blocks. Dotted block indicates all calculations performed in parallel, over the population size as pop/procs.

user-defined training sets and constraints, general parameter optimization for different force field engines, including conventional, complex reactive adiabatic[12,13] and nonadiabatic[14,15] force fields, and different optimization criteria (such as 2-norm force vector for all atoms, as opposed to just energy minimization on geometries). Furthermore, GARFfield provides unique capabilities to enable exploration of multiple (nondeterministic) Pareto-optimal solutions, via hill-climbing and weight randomization of objective functions.

The general multiobjective decision problem addressed by GARFfield is defined as the following: given an n-dimensional force field parameters decision vector $x = x + 1, x_2,..., x_n$ in the solution hyperspace space x, find a vector instance, x', that minimizes the k objective functions $z(x') = z_1(x'), z_2(x'),..., z_k(x')$ for a given set of molecular structure properties predicted from first-principles quantum mechanics, where x is constrained by the parameters bounds and not dominated by any particular objective z_i $(i = 1,..., k)$. A solution x is said to dominate another feasible solution y $(x < y)$ if and only if $z_i(x) \leq z_i(y)$ and $zj(x) < z_j(y)$ for at least one objective function j.

This paper describes in section II the GARFfield architecture, program flow, and the different objective functions currently supported for force field development; then in section III, we illustrate the application of GARFfield to the development of the ReaxFF reactive force field[12] for studying adiabatic reactions and dynamics of silicon carbide (SiC) growth from trichlorosilane. We also apply GARFfield to the development of effective core pseudopotentials (ECP)[16] for the electron fast forward (or electron force field, eFF[14,15]), a mixed quantum-classical mechanics method for describing nonadiabatic electron dynamics phenomena for highly electronically excited systems. These individual applications produce optimized parameter sets, but here we focus only on the validation against quantum mechanics (QM), eschewing detailed analysis of the applications to other publications. Because the focus here is on force field development using the GARFfield framework, we provide implementation details, a User's Manual (as part of the Supporting Information), and performance metrics in section IV, along with a final set of observations and recommendations for the development of force fields using GARFfield in section V.

II. GARFFIELD FORCE FIELD DEVELOPMENT FRAMEWORK

GARFfield uses an evolutionary algorithm to produce a random population of force field parameter sets (strings, or chromosomes in natural evolution terminology). The user has the option of selecting a subset of force field parameters for training (i.e., selecting particular string positions, or genes in the natural evolution terminology), or using the parameter entire set (Parameter Selection block in Figure 1). Each parameter value within a string (allele, in the natural evolution terminology) needs to be specified within a real space range that prevents discontinuities or nonphysical behavior in the potential energy landscape of the force field being optimized. Every string (force field) is then used by a compliant molecular mechanics/dynamics engine to evaluate a full set of user-defined training motifs (Training Set block in Figure 1). The result of evaluating each training motif using a particular string is compared against a corresponding reference value obtained from first-principles QM calculations. The training set prepared by the user may contain different objective functions, and multiple entries per objective function, both of which can be weighted. Reference labels are used to uniquely identify training set entries and serve as index keys to the corresponding atomic or coarse-grain structure definition files, which should also be provided by the user. A total fitness value, calculated using an error function selected by the user (Fitness Function Calc. block in Figure 1), is produced for every string (force field) in the random population evaluated over the entire training set. This fitness value is used to rank each string and to decide if it should be replaced (i.e., discarded) or used to evolve new and better fit offsprings. Evolution of new siblings is achieved via mutation (reproduction from a single fit string)

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

and/or crossover (reproduction from two different and fit strings) operations. The iterated population improves toward a higher-quality set of force fields, and upon convergence, the best force field performer is selected and written to an output file. The overall computational flow diagram for GARFfield is depicted in Figure 2.

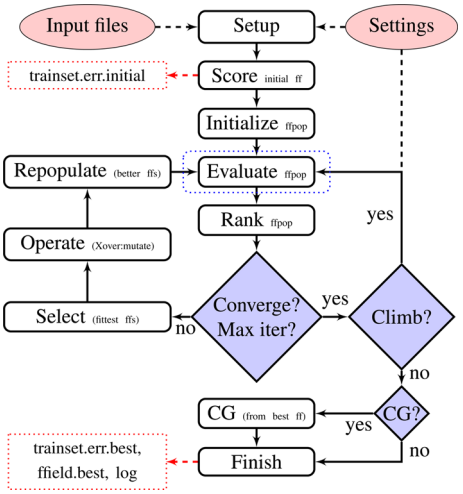


Figure 2. Flow diagram for GARFfield. Ovals represent user-provided input, shaded rectangles represent GARFfield calculations and library calls (to the evolutionary algorithm, LAMMPS, or the conjugate gradient solver CG), and dashed boxes represent primary outputs. Shaded boxes represent decision points or manual input requirements. The dotted evaluate box depicts parallelized computations over the GA force field population size.

GARFfield provides many unique features for training complex force fields, including improved convergence via hybrid global and local gradient-based optimization, local minima avoidance via hill-climbing, application of relative restraints (harmonic with variable start/end restraint coefficients) for handling nonequilibrium definitions, and support for multiobjective optimization, among others.

To improve convergence and speed of convergence, GARFfield includes an option to automatically switch from the evolutionary GA scheme to a conjugate gradient (Polak−Ribiere version) (CG) optimizer, when the GA is within the quadratic basin of a local minima in the PES. The CG operates on the single best accumulated force field string at the time of the GA−CG switch and proceeds by calculating the directional gradients from central finite differences of the force field parameters applied on the fitness function over the full training set. The CG option may be activated when the error change between iterations is within a threshold tolerance. GARFfield also provides a hill-climbing option that may be used to periodically test the existence of other minima (i.e., to climb out of local minima).

For training nonequilibrium cases in reactive systems, such as transitions states along reaction coordinates, cold-compression curves in periodic systems, or bond breaking or rotational barriers, among others, GARFfield enables the definition and application of relative structural restraints between interparticle lengths, angles, and torsions (Restraints block in Figure 1).

A critical step in preparing a force field from first-principles QM is defining an appropriate objective function or set of functions. GARFfield allows the force field developer to systematically define multiple objective functions in a training set, and to combine them into a single weighted sum composite function. Furthermore, it enables randomization of individual objective function weights that allow the force field developer to produce nondeterministic, multiobjective and Pareto-optimal solutions. The force field developer is at liberty to choose which objective functions to use for a particular force field development case (i.e., any combination from $k = 1,..., N_{of}$).

The following objective functions are currently supported: q_e, atomic charges; s_e, structural/geometric properties including 2-body bond lengths (s_b), 3-body angles (s_a), 4-body torsions or inversions (s_t, s_i), cell parameters (s_l) (for periodic systems only), or pairwise radial distributions (s_{RDF}). Equilibrium definitions are the following: f_e: zero atomic forces, from 2-norm (length) force vector for all atoms (to specify a minimum or transition state geometry) and σ_e: zero cell stresses (to specify an optimum periodic cell). E_e is the differential energies (and off-equilibrium definitions).

The charges (s_b) are computed from atomic coordinates using the Charge Equilibration (QEq) method from Rappe and Goddard.[17] This allows QEq parameters (electronegativity, hardness, and atomic radius) to be optimized, if required in the particular force being trained (e.g., ReaxFF).

2-body bond lengths are computed from the two particle coordinates

$$s_b = \sqrt{\sum_{p=x,y,z} (p_1 - p_2)^2} \tag{1}$$

3-body angles from two bond vectors (or the cosine of the angle), $\vec{\mathbf{a}} = (p_0 - p_1)$ and $\vec{\mathbf{b}} = (p_2 - p_1)$, as

$$s_a = \cos^{-1}\left(\frac{\vec{\mathbf{a}}\cdot\vec{\mathbf{b}}}{|a|\cdot|b|}\right) \tag{2}$$

4-body torsions from the dihedral angle defined by three bond vectors (or the cosine of the angle) as

$$s_t = a\tan 2(|b|\vec{\mathbf{a}}\cdot[\vec{\mathbf{b}}\times\vec{\mathbf{c}}],\ [\vec{\mathbf{a}}\times\vec{\mathbf{b}}]\cdot[\vec{\mathbf{b}}\times\vec{\mathbf{c}}]) \tag{3}$$

the cell lattice parameters directly from a energy minimized structure and the unit cell type definition (cell angles from the cosine of the angle), the radial distribution function for any particular pair of particles, from the number density (ρ) multiplied by the spherical shell, as

$$s_{RDF} = g(r) = 4\pi r^2 \rho dr \tag{4}$$

The cell stress trace is computed from the sum of the $i = j$ virial components of the pressure tensor, given by

$$\sigma_{ij} = \sum_{k}^{N} r_{k_i} f_{k_j} \tag{5}$$

and the 2-norm (length) of the global force vector for all atoms is used as an alternative to full structure energy minimization (finite and periodic)

$$|f| = \sqrt{f_1^2 + f_2^2 + \cdots + f_n^2} \tag{6}$$

Differential energy entries may be any combination of relative energies with respect to some reference state. For example, discretized energies in an equation of state or a bond-energy bond-distance relationship (single or multiple bonds) with respect to the corresponding equilibrium values; or relative reactant, transition state, and product energies; binding energy versus unbound energy; relative energies between different material phases at equilibrium; relative energies between clusters

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

of different sizes; surface energy versus bulk energy; interaction energy between two molecules as the difference between the complex and the individual molecular energies; relative energies between under and overcoordinated case; heats of formation at 0 K for molecular reactions; among many other options.

To minimize the k objective functions, GARFfield computes an error between the force field predicted values (using the random parameters produced by the evolutionary algorithm) and the corresponding data from quantum mechanics (or experiment). Un-normalized weight (MSE = mean square error and SIWE = squared inverse weighted error) and normalized weight (RMSE = root mean square error, MPE = mean percent error, and NRMSE = normalized RMSE) error functions are available in GARFfield, as defined, respectively, by the following functional forms

$$z_{\mathrm{MSE}}(x_i) = w_i[p_{\mathrm{QM}}(x_i) - p_{\mathrm{FF}}(x_i)]^2 \tag{7}$$

$$z_{\mathrm{SIWE}}(x_i) = \left[\frac{p_{\mathrm{QM}}(x_i) - p_{\mathrm{FF}}(x_i)}{w_i}\right]^2 \tag{8}$$

where, w_i is a corresponding entry case weight, p_{FF} is the property computed by the force field being optimized, and p_{QM} the corresponding value from a first-principles quantum mechanical (QM) calculation, and

$$z_{\mathrm{RMSE}}(x_i) = \sqrt{w_i'[p_{\mathrm{QM}}(x_i) - p_{\mathrm{FF}}(x_i)]^2} \tag{9}$$

$$z_{\mathrm{NRMSE}}(x_i) = \frac{z_{\mathrm{RMSE}}(x_i)}{p_{\mathrm{FF}_{\max}} - p_{\mathrm{FF}_{\min}}} \tag{10}$$

$$z_{\mathrm{MPE}}(x_i) = 100w_i'\frac{|p_{\mathrm{QM}}(x_i) - p_{\mathrm{FF}}(x_i)|}{p_{\mathrm{QM}}(x_i)} \tag{11}$$

where w' corresponds to the mean normalized weight (over the total number of entries per section, k, and entries per class, n_c) (i.e., $\sum w_i' = 1$) precalculated from

$$w_i' = \frac{1}{k}\frac{w_{\mathrm{of}_i}}{\sum_{j=1}^{k} w_{\mathrm{of}_j}}\frac{1}{n_c}\frac{w_{c_i}}{\sum_{l=1}^{n_c} w_{c_l}} \tag{12}$$

where w_{of_i} is the corresponding objective function section ($z = q_e$, S_e, E_e, σ_e, f_e, as described below) weight and w_{c_i} the corresponding section entry weight.

GARFfield will recognize the number of entries per objective function in the training set file and normalize depending on the selected error function (see the Supporting Information for file format details). By default, every $w_{\mathrm{of}_i} = 1$ (hence the first sum turns into an average over k, unless the user supplies weights per section).

This hierarchy of weights provides the necessary flexibility to the user. For example, to capture the correct dissociation barriers for a reactive force field, one may require higher weighting for points of a bond-energy bond-distance curve that are near equilibrium and at and away from the bond dissociation length. In the same manner, weights can be set to alter the relative importance of an entire objective function section in the training set (e.g., relative energies versus structural details).

A default a priori approach is implemented, which expects the user to provide the weights within the trainset file (otherwise weighs are set to 1). An alternative, and supported, approach is to let GARFfield randomly generate (using "-W" option) the normalized weight vector, $w_{1\ldots\mathrm{Nof}}$. This will enable a search in multiple directions, requiring the user's decision a posteriori on which solution satisfies his goals best.

GARFfield uses by default, explicit energy minimization on every structure, before calculating and comparing the particular objective function result to the literal reference value provided in the training set from QM or experiment. Normally, this requires multiple energy/force calculations, depending on how far the result comes to the reference value, and complete minimization convergence. Two minimization schemes are available, a conjugate gradient algorithm (default) and a fast inertial relaxation engine or FIRE;[18] the latter performs damped dynamics minimization.

Alternatively, one can compute the single-point energy for a specified geometry and the 2-norm force (length of the global force vector for all atoms) to use a geometry relaxation criterion (i.e., a fully relaxed structure would have zero 2-norm force vector). This involves a single energy/force calculation step, as opposed to a finite number of minimization steps, which would lead to a significant speedup in computing the training set. Nevertheless, it must be used with caution because the relative direction of the force is lost in the 2-norm and the reference value used by GARFfield to compute the error is still the systemssed by GARFfield to compute the error is still energy, albeit scaled by $(1 + f_{2\text{-norm}})^{-1}$. Alternatively, the user can directly use 2-norm force or cell stress as part of the training set (see the Supporting Information).

III. SILICON CARBIDE REACTIVE FORCE FIELDS FROM GARFFIELD

This section covers two different force field development examples for modeling silicon carbide (SiC) material systems. The first case is an adiabatic ReaxFF[12] reactive potential to study the dynamic synthesis of SiC from Methyltrichlorosilane and the second an effective core pseudopotential (ECP) in the nonadiabatic eFF formulation to study the electronic excitations during SiC shock loading.

SiC has a wide range of applications due to its many unique properties, low-density, such as high thermal conductivity,[19] thermal shock resistance,[20] biocompatibility,[21] resistance to acidic and alkali environments,[22] chemical inertness, and high mechanical strength.[23,24] It is also used in the design and synthesis of fuels and chemicals as separation and reactive media, such as adsorbents, membranes, and catalysts, as well as in various types of electronic sensors. Silicon carbide is also an important material in TRISO-coated fuel particles, the type of nuclear fuel found in high temperature gas cooled reactors (such as the pebble bed reactor). A layer of silicon carbide gives coated fuel particles structural support and is the main diffusion barrier to the release of fission products.[25] Silicon carbide composite material has also been investigated for use as a replacement for Zircaloy cladding in light water reactors. The composite consists of SiC fibers wrapped around a SiC inner layer and surrounded by an SiC outer layer.[26] Because SiC rarely happens in nature, optimizing its fabrication processes (e.g., using sintering or chemical vapor deposition (CVD) or infiltration (CVI) from methyltrichlorosilane (MTS) or other precursors) remains an active research topic.

Therefore, the development of reactive and nonadiabatic force fields capable of accurately capturing the complex phenomena found in the ample range of SiC applications, including synthesis,

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

Table 1. ReaxFF Parameters Optimized for SiC−MTS[a]

parameter	section	atompair/angle	initial value	final value	% |difference|
D_e^σ	bond	C−Cl	131.9833	135.7409	2.85
p_{be1}	bond	C−Cl	−1.0000	−0.8229	17.71
p_{be2}	bond	C−Cl	1.2060	1.8090	50.00
p_{bo1}	bond	C−Cl	−0.0824	−0.0884	7.28
p_{bo_1}	bond	C−Cl	5.7013	5.1701	9.32
D_e^σ	bond	H−Cl	96.3886	112.8570	17.09
p_{be1}	bond	H−Cl	−0.0572	0.0944	265.03
p_{be2}	bond	H−Cl	2.2822	1.5000	34.27
p_{bo1}	bond	H−Cl	−0.1093	−0.1027	6.04
p_{bo2}	bond	H−Cl	5.1686	5.7987	12.19
D_e^σ	bond	Si−Cl	117.3136	126.4334	7.77
p_{be1}	bond	Si−Cl	0.0148	0.0074	50.00
p_{be2}	bond	Si−Cl	0.9459	0.4729	50.01
p_{bo1}	bond	Si−Cl	−0.1181	−0.0590	50.04
p_{bo2}	bond	Si−Cl	5.4270	5.9410	9.47
D_{ij}	off-diagonal	C−Cl	0.2535	0.2651	4.58
α_{ij}	off-diagonal	C−Cl	12.0000	11.4529	4.56
D_{ij}	off-diagonal	H−Cl	0.0129	0.0628	386.82
α_{ij}	off-diagonal	H−Cl	12.0000	9.6285	19.76
D_{ij}	off-diagonal	Si−Cl	0.1014	0.1840	81.46
α_{ij}	off-diagonal	Si−Cl	12.0000	10.8573	9.52
p_{val1}	angle	C−C−Cl	75.9138	89.0053	17.25
p_{val2}	angle	C−C−Cl	21.2437	28.5172	34.24
p_{val1}	angle	C−Si−Cl	76.1248	76.2780	0.20
p_{val2}	angle	C−Si−Cl	29.9800	26.1530	12.77
p_{val1}	angle	H−C−Cl	78.3243	85.1985	8.78
p_{val2}	angle	H−C−Cl	30.0000	30.0000	0.00
p_{val1}	angle	H−Si−Cl	81.6322	77.2718	5.34
p_{val2}	angle	H−Si−Cl	30.0000	29.5037	1.65
p_{val1}	angle	Si−C−Cl	69.7801	69.5117	0.38
p_{val2}	angle	Si−C−Cl	12.4545	11.0362	11.39
p_{val1}	angle	Si−Si−Cl	73.4201	71.7728	2.24
p_{val2}	angle	Si−Si−Cl	16.9804	14.2543	16.05
p_{val1}	angle	Cl−C−Cl	74.2883	77.1943	3.91
p_{val2}	angle	Cl−C−Cl	30.0000	30.0000	0.00
p_{val1}	angle	Cl−Si−Cl	75.5382	75.1218	0.55
p_{val2}	angle	Cl−Si−Cl	29.9993	30.0000	0.00

[a]See ReaxFF file format for the location of these parameters in the Supporting Information.

is an important contribution and a complementary goal of this paper.

III.1. ReaxFF Reactive Force Field for SiC Synthesis from MTS. SiC can be fabricated through thermal degradation of various preceramic precursors via a series of tailored reactions leading to the final ceramic. Understanding the atomistic events and mechanisms involved in this process is paramount to optimizing it. Unfortunately, there is nearly no information from neither experiments on the reaction sequences nor on the local geometric configurations of this amorphous structure, both of which control the overall material performance properties. To provide input into how the chemical reaction processes can affect the final structural and physical properties, we use reactive molecular dynamics with ReaxFF force fields to model the thermal decomposition of the precursors and growth of the final SiC, under realistic experimental conditions. In refs 27 and 28, we studied the production of SiC nanoporous membranes via pyrolysis of hydridopolycarbosilane (HPCS) $[-SiH_2CH_2-]_n$, used for H_2 gas separation at high temperatures.[29,30] We developed the HPCS ReaxFF force field to study the ceramic formation using HPCS as a precursor and for modeling the process-based dynamics of growth. This led to a material product model characterization in close agreement with the experimental results.[27,28]

SiC growth using the CVI technique via microwave-enhanced heating of methyltrichlorosilane (MTS), CH_3SiCl_3, has recently received significant attention. Mainly, because the microwave-assisted heating makes it possible to heat the substrate internally to enable the deposition of SiC in an inside-out fashion. As a result, a spatially uniform composite with higher density is generated and formation of pores is avoided. Furthermore, the manufacturing time is reduced considerably.[31−33]

Here, we describe the use of GARFfield to develop an accurate ReaxFF force field capable of modeling the thermal decomposition of MTS, as a precursor for SiC formation. To facilitate the development process, we start from the ReaxFF force field parameters for HPCS[27] and extend it to include Cl self-interactions and its 2-, 3-, and 4-body interactions with the other elements involved (i.e., Si, C, and H). 37 parameters were chosen for optimization; from the bond, off-diagonal, and angle terms in the ReaxFF energy expression (see ref 34), as listed in Table 1. A corresponding parameter selection file was prepared, as

described in GARFfield's User Manual (see the Supporting Information), using the values from ref 27 as mean values for the subsequent training. A training set file was prepared from 235 unweighted QM cases that include atomic charges, bond lengths, angles, transition state geometries, reactions and dissociation energies for bonds and angles (relevant to MTS), and energies from all thermal decomposition products (see the Supporting Information for details on the training set).

The reference geometrical (bond, angle, etc.) and charge data in the training set were obtained by carrying out quantum mechanical (QM) calculations on every given molecule in the training set file. The QM data for nonperiodic systems was obtained from the density functional theory calculations and carried out using Jaguar (version 5.5),[35] the B3LYP functional,[36,37] and Pople's 6-31G** basis set.[38,39] The Mulliken charges were also computed using the 6-31G** basis set. Molecular geometry files for all the cases referenced in the training set were prepared in MSI BioGraf (.bgf) format. GARFfield was set to single-point mutations at a rate of 1/235, no crossover, and a mean percent error (MPE) fitness function.

The optimal strategy for optimizing this ReaxFF force field involved partitioning the GARFfield optimization process in incremental stages, based on the force field parameter types. A ReaxFF force field file contains 7 sections: a general parameters section, an atom specific parameters section, a 2-body (bond) interaction parameters section, an off-diagonal parameters section, a 3-body (valence angle) parameters section, a 4-body (torsion angle) parameters section, and a hydrogen bond parameters section. Note that in this case, we started from a known set of parameters, albeit not optimized for SiC.

If the force field under development involves elements for which there are no available parameters, the fundamental atom, bond, angle, and torsion parameters should be populated from running some of the same training motifs (finite and periodic) with the universal force field, UFF[40] (covers the periodic table of elements up to Lawrencium), from quantum mechanical calculations, or, as a last resort, from existing ones for an element with similar chemical properties. The sequence described in Figure 3 should then change to 1, 5, 2, 3, 4, i.e., optimizing the valence terms first, followed by the nonbond terms. The automated UFF step capability will be part of a future GARFfield release.

For the present case, atom, off-diagonal, bond, and angle parameters were optimized sequentially (see Table 1). No hydrogen bonds are involved. This builds upon direct knowledge of the interactions to be trained and parameter sensitivity analysis performed on the starting point, SiC HPCS ReaxFF force field.

200 GA iterations were performed for each case, and the training result errors were checked at each step to make sure the optimization proceeded as expected. The range for each parameter was selected to be within ±100% of its mean value, defined from the initial force field. Convergence to an acceptable minimum was tested by running another instance up to 500 iterations, using the hill-climbing option every 50 iterations. A larger parameter range (up to ±500%) was also examined to test the sensitivity of the optimization to the parameter range. The results were consistently similar to the previous ranges tested. The bond parameters were then optimized in a two-stage process, beginning with the first parameter for every bond pair, D_e^σ, which corresponds to the pre-exponential factor of the bond-energy expression in ReaxFF, and finishing with the rest of the section parameters. D_e^σ plays a critical role in the value of equilibrium bond energy. The resulting D_e^σ was obtained using ±100% offset from the mean reference value, whereas the other

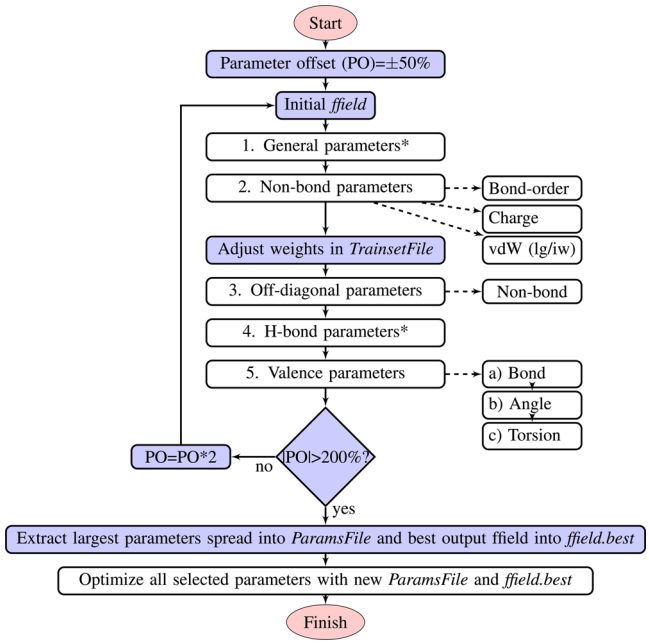


Figure 3. General strategy for training ReaxFF force fields. The initial forcefield should be chosen from an existing set, when available, or else from parameters for elements with similar properties. The best output ffield corresponds to the best out of the three runs (at PO = ± 50, 100, 200). White boxes represent force field parameter types, and shaded boxes correspond to decision points or manual input requirements. The asterisk implies these parameters should be optimized only when strictly necessary, vdW = van der Waals, lg = low-gradient vdW dispersion correction, iw = vdW inner-wall correction, and arrows indicate sequence/concurrence.

section parameters were obtained from ±5%−10% of their respective mean values. Finally, the valence angle parameters were optimized using a range of ±100% of the initial mean values.

The best force field obtained after this procedure was used to reoptimize all the parameters jointly using a small parameter range (e.g ± 5%). The final set of optimized parameters and the percent difference with the initial value are shown in Table 1. The training set and resulting force field for this case, and a description of parameters in a ReaxFF force field file can be found in the Supporting Information.

In general, the proposed strategy for optimizing a ReaxFF force field with GARFfield involves partial and incremental optimization of parameters in the sequence described in Figure 3.

Alternatively, and assuming the force field developer has significant experience with the ReaxFF energy expression, lower and upper range values for each parameter may be specified. An additional test of convergence is to use the "-s nochange" command-line option to decide convergence when the total error does not change over a range of iterations.

A sufficiently general training set should contain a diverse set of cases to describe elemental, binary. and *n*-ary interactions (when needed) through corresponding finite, surface or bulk models. Nonetheless, keep in mind that the cost of calculating the training set will be proportional to the size of the models used (i.e., finite systems are faster to compute than equivalent sized periodic systems). Also note that charge training cases must be included in the trainset file when optimizing charge parameters (i.e., electronegativity, hardness, and radius).

Figure 4 shows the total error as a function of iteration, during the sequential optimization procedure described above,

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

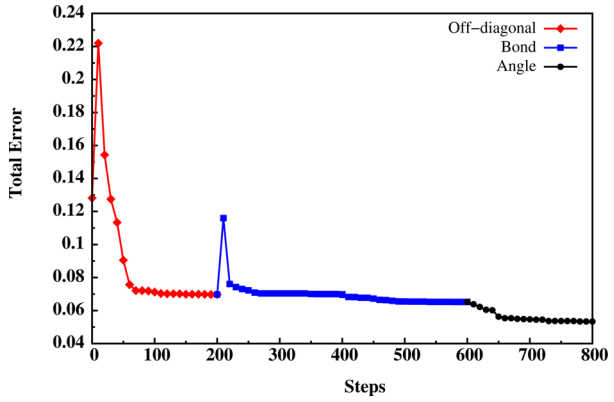


Figure 4. Change in the total error with the number of steps during one complete optimization.

Figure 5 compares the GARFfield−ReaxFF and QM H−Cl, C−Cl, and Si−Cl bond dissociation curves in HCl, H_3CCl, and H_3SiCl molecules (i.e., bond distance versus bond energy), and Figure 6 compares the GARFfield−ReaxFF and QM C−C−Cl, C−Si−Cl, H−Si−Cl, Cl−C−Cl, Cl−C−H, Cl−Si−Cl, and Si−Si−Cl angle curves (i.e., angle value versus energy).

In general, there is good agreement between the GARFfield−ReaxFF predicted values and the QM references, except for a significant discrepancy between the energies computed by ReaxFF and QM in the compressed regime of H−Cl and in the expanded regime of Si−Cl. Thus, ReaxFF uses a bond energy-to-bond order-to-bond distance relation that can capture double and single bonds, as well as lower-order bonds during reactions. As a result, the compression and expansion curves computed by ReaxFF are generally not as accurate as those from QM. For covalent systems, accurate second-order properties are best carried out by using ReaxFF to obtain the structure, and then using a nonreactive force field to analyze the distortions near equilibrium geometries.

GARFfield provides a variety of options that can be used for improving optimization results (see the Supporting Information). For example, the hill-climbing option enables upward moves to search for alternative and lower minima. Figure 7 shows how the total C−Cl bond dissociation curve is improved by using the hill-climbing option with a frequency of 100 (the hill-climb frequency depends on the size and complexity of the training set.)

To test the accuracy of the developed force field, we study thermal decomposition of the MTS in presence of H_2 gas. The goal is to determine if the reaction products obtained from reactive molecular dynamics (RMD) simulations using the prepared ReaxFF force field agrees with what is known theoretically. To do so, we built an atomistic model of the MTS with H_2 gas. A system of 480 H_2 and 40 MTS molecules (1280 atoms), at the room-temperature density of the mixture, was prepared in a periodic simulation box with dimensions of 29.5 × 29.5 × 29.5 Å. The system was then minimized using an energy criterion of 1e-6 and subsequently equilibrated using an NVT ensemble at a temperature of 300 K using ReaxFF (this took 8 ps). A cook-off simulation of the system was performed to study the thermal degradation of MTS, using a linear heating rate

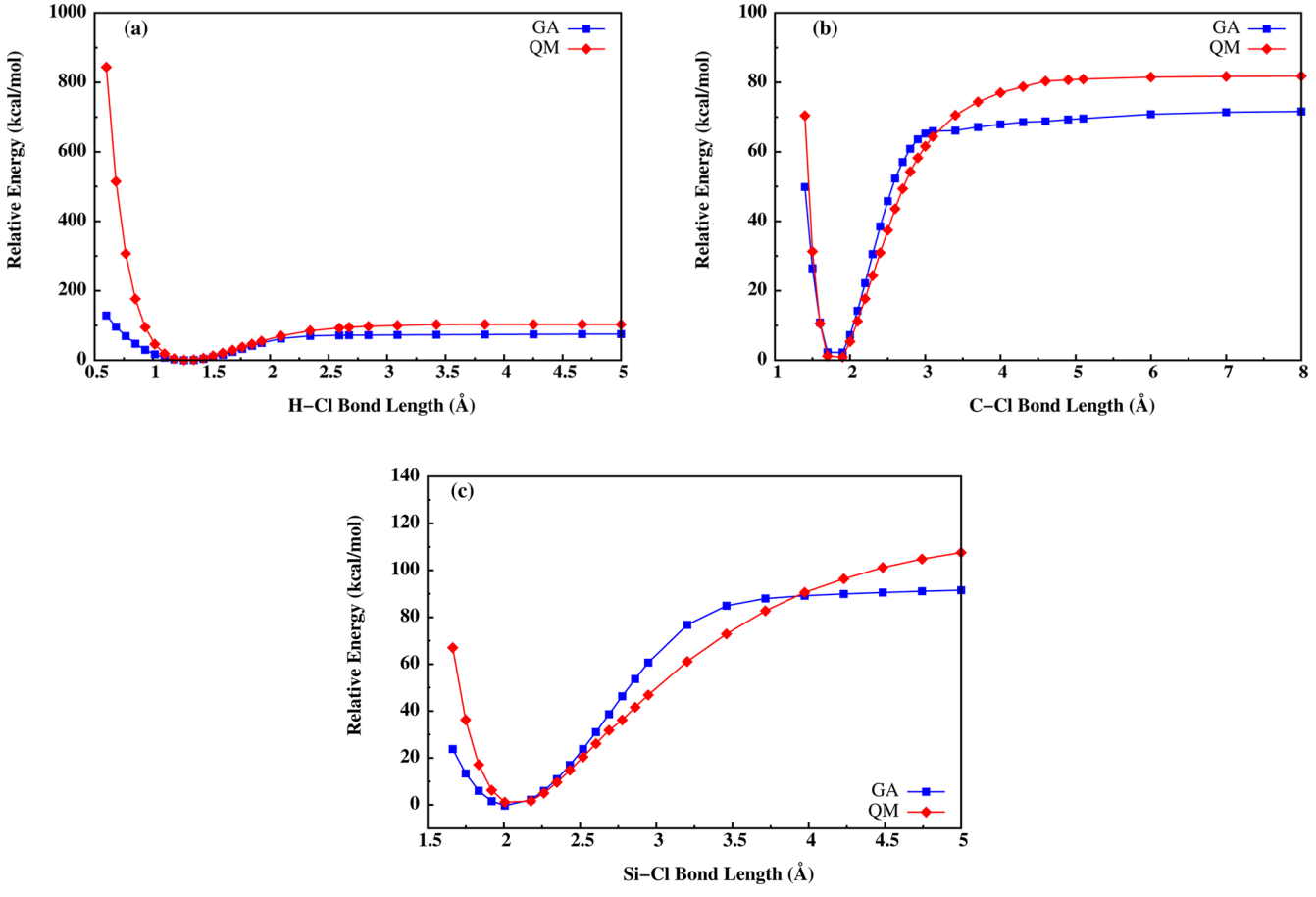


Figure 5. Comparison of the energies computed by QM and GA for (a) H−Cl, (b) C−Cl, and (c) Si−Cl bond dissociation.

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

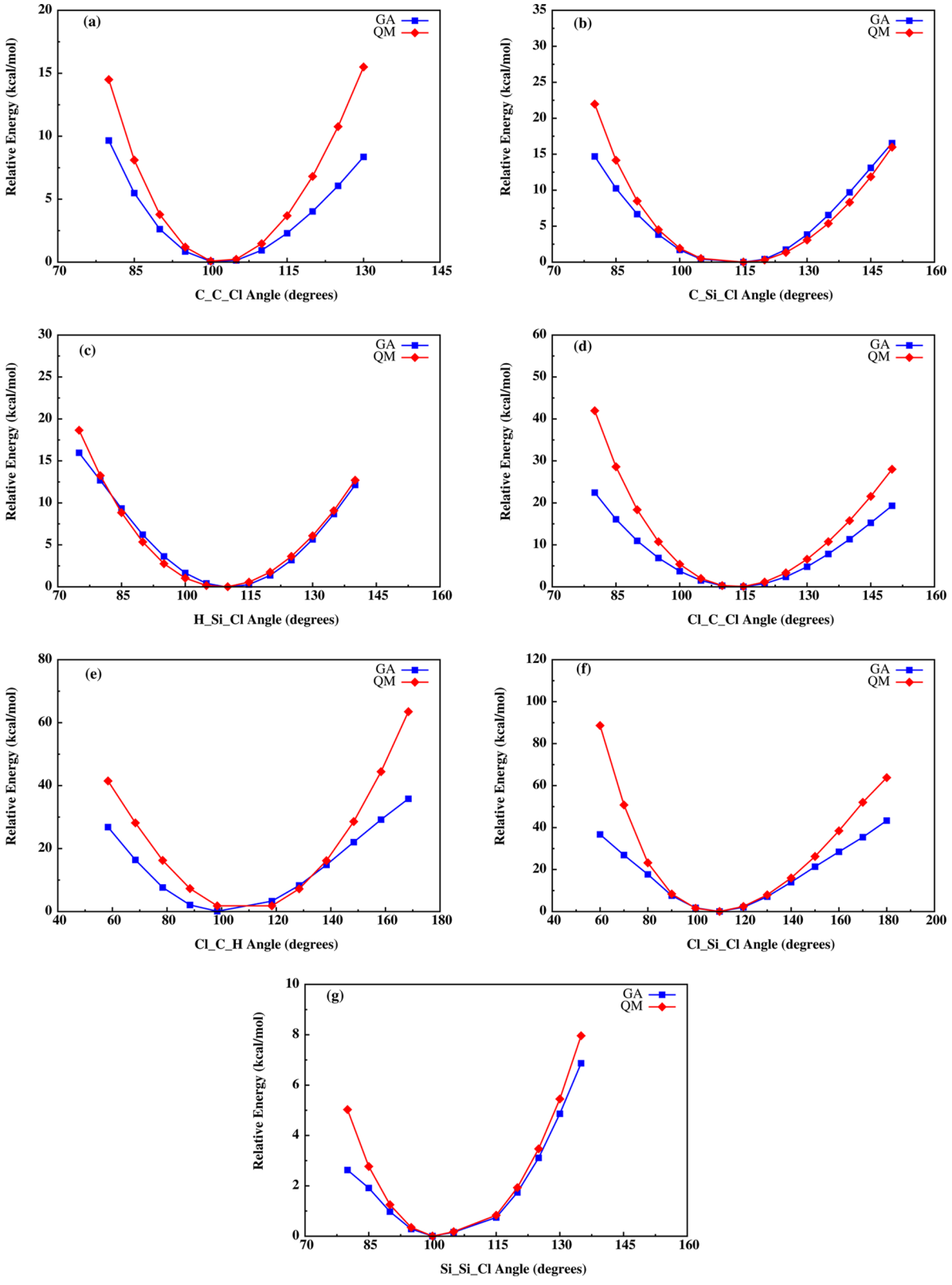


Figure 6. Comparison of distortion energies computed by GARFfield−ReaxFF and QM for the angles in (a) C−C−Cl, (b) C−Si−Cl, (c) H−Si−Cl, (d) Cl−C−Cl, (e) Cl−C−H, (f) Cl−Si−Cl, and (g) Si−Si−Cl.

of 0.1 K/fs, up to a final temperature of 3500 K. The system was subsequently maintained at 3500 K for 45 ps. The variations of the temperature and pressure during the simulation are shown in 8 (upper left and upper right), and the time evolution of the gaseous compounds is shown in 8 (bottom).

For temperatures less than but close to 2500 K, the Si−Cl bonds were very mobile (see Cl and MTS production), forming and breaking before actual MTS decomposition. The simulations indicate that the degradation of the MTS initiates at about 2500 K with Si−C, Si−Cl, and few C−H bonds breaking.

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

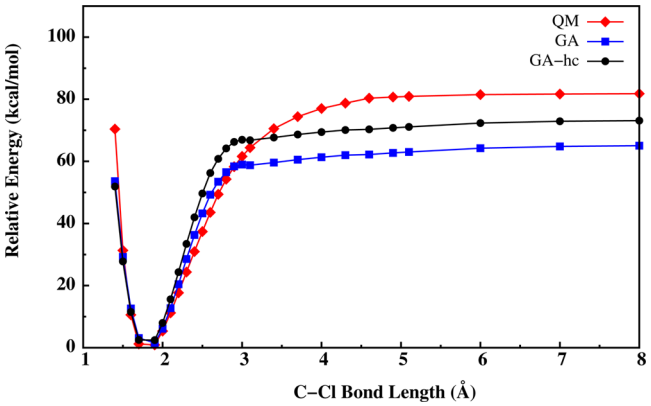


Figure 7. Comparison of the energies computed by QM and GA (with and without hill-climbing routine) for C–Cl bond.

At higher temperatures and up to 3500 K, the bond breaking process accelerates and free H, Cl, CH_n, and $SiCl_n$ radicals are produced. Secondary reactions involving bond formation between Cl and H radicals and CH_n and H radicals at 3500 K led to the formation of HCl and CH_4 gases, respectively.

Fragment analysis of the simulation products reveals a series of reactions that were found to be the most important and frequent (see Table 2). These reactions may be used as a guide in predicting the real mechanism of the MTS pyrolysis. We find that Si–C and Si–Cl bond breaking in MTS initiate its decomposition (reactions 1 to 20). In some cases (reactions 1 to 3), the Si–C bond breaks due to the thermal effect and in others due to the hydrogen gas and H radicals reacting with the CH_3 groups to

Table 2. Reactions during MTS Thermal Decomposition with ReaxFF

reaction no.	reaction	no. of reactions
1	$CH_3SiCl_3 \rightarrow CH_3 + SiCl_3$	2
2	$CH_3SiCl_2 \rightarrow CH_3 + SiCl_2$	2
3	$CH_3SiCl \rightarrow CH_3 + SiCl$	1
4	$H + CH_3SiCl_3 \rightarrow CH_4 + SiCl_3$	2
5	$H + CH_3SiCl_2 \rightarrow CH_4 + SiCl_2$	3
6	$H + CH_3SiCl \rightarrow CH_4 + SiCl_2$	8
7	$H + CH_3SiH \rightarrow CH_4 + SiH$	2
8	$H + CH_3Si \rightarrow CH_4 + Si$	2
9	$H + CH_3SiH_2 \rightarrow CH_4 + SiH_2$	4
10	$H + CH_3SiHCl \rightarrow CH_4 + SiHCl$	2
11	$H + CH_2SiHCl \rightarrow CH_3 + SiHCl$	1
12	$CH_3SiH \rightarrow CH_3 + HSi$	2
13	$CH_3SiH_2Cl \rightarrow CH_3 + SiHCl + H$	1
14	$CH_3SiH_2 \rightarrow CH_3 + SiH_2$	1
15	$CH_3Si \rightarrow CH_3 + SiH$	1
16	$CH_3SiHCl \rightarrow CH_3 + SiHCl$	1
17	$CH_2SiH_2 \rightarrow CH_2 + SiH_2$	1
18	$CH_3SiCl_3 \rightarrow CH_3SiCl_2 + Cl$	39
19	$CH_3SiCl_2 \rightarrow CH_3SiCl + Cl$	18
20	$CH_3SiCl \rightarrow CH_3Si + Cl$	16
21	$CH_3SiCl_2 + Cl \rightarrow CH_3SiCl_3$	18
22	$CH_3SiCl + Cl \rightarrow CH_3SiCl_2$	11
23	$CH_3Si + Cl \rightarrow CH_3SiCl$	14
24	$H_2 \rightarrow H+H$	92
25	$H + Cl \rightarrow HCl$	180
26	$HCl \rightarrow H + Cl$	128
27	$CH_3 + H \rightarrow CH_4$	5

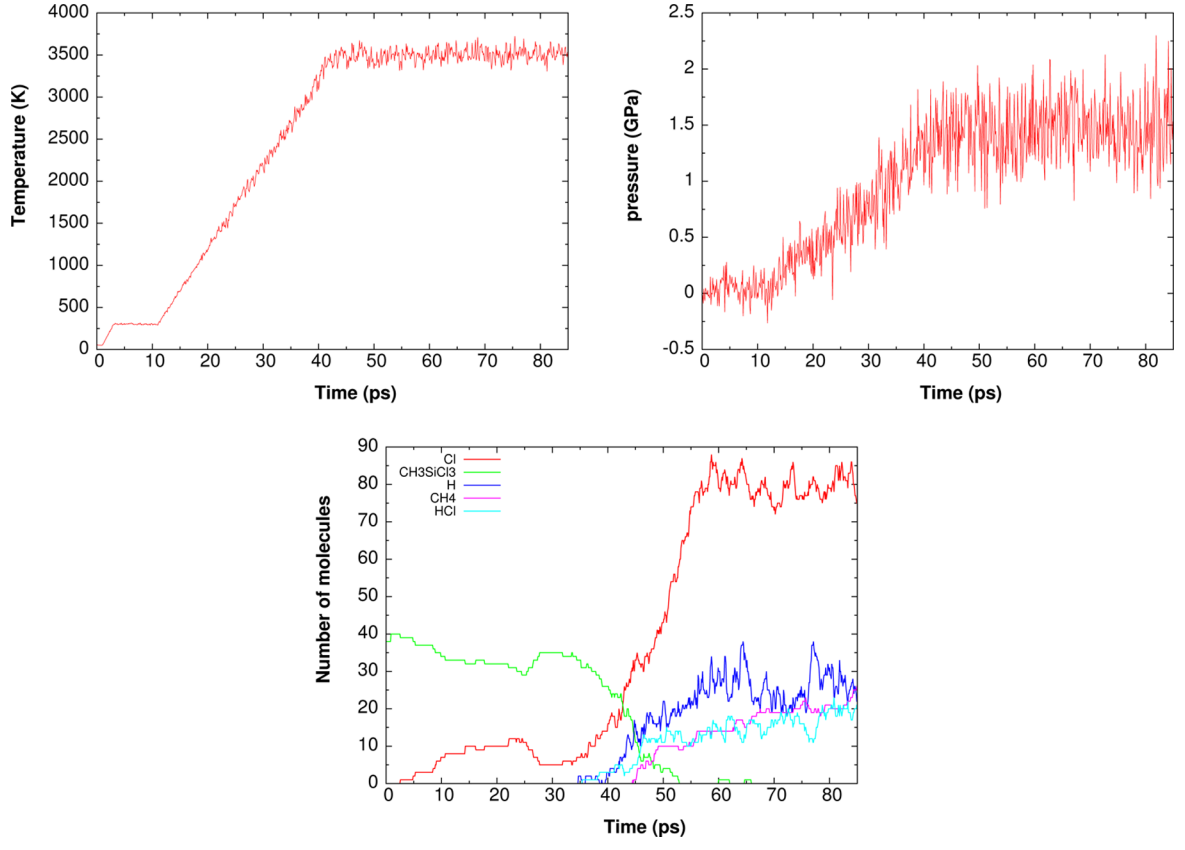


Figure 8. (Upper left) temperature and (upper right) pressure and fragment evolution (bottom) during MTS cook-off simulation.

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426–1439

produce CH_4 gas (reactions 4 to 10). Some of the H radicals react with $CH_3SiH_nCl_m$ (n, m = 0−2) to form more stable structures, whereas in subsequent steps, the Si−C bonds in these structures break and decomposition of MTS proceeds (reactions 11 to 20). As mentioned earlier, the Si−Cl bonds are very mobile and they form and break frequently (reactions 21 to 23). Decomposition of MTS and H_2 forms H radicals that react with Cl and CH_n radicals to produce other gases, including HCl and CH_4 (reactions 24 and 27). This sequence of reactions and products agrees very well with published results from other theoretical studies.[41]

III.2. Nonadiabatic eFF Effective Core Pseudopotential (ECP) for SiC. The Born−Oppenheimer (BO) approximation, which decouples the nuclear and electronic motions, underlies most of the atomistic modeling applications. However, the BO approximation breaks down for systems involving highly excited electronic states. These are common in warm dense and hot matter regimes, high velocity impact, high energy radiation, and other extreme material conditions for which the electronic portion of the wave function contains contributions from many stationary states.[15]

The electron fast forward (eFF)[14] (also called the electron force field) was developed to overcome this limitation by enabling very fast nonadiabatic dynamics of highly excited electronic states. eFF with recent improvements[15] has confirmed its scalability and applicability to challenging problems including explaining emissions of hot electrons during brittle fracture of silicon,[42] describing the dynamics of Auger induced chemical decomposition,[43] characterizing hydrostatic and dynamic shock Hugoniots for different materials,[14,15,44,45] and tracking the dynamics of Coulomb explosion in silicon and diamond nanoparticles,[46] among others.[47]

In the framework of eFF, nuclei are classical point charges and the total electronic wave function is represented by a Hartree product of one-electron floating spherical Gaussian (FSG) wave packets, eq 13, whose positions, $\vec{\mathbf{x}}_i$, and sizes, s_i, are both dynamic variables.

$$\Psi(\vec{\mathbf{r}}_i) \propto \prod_i \exp\left[-\left(\frac{1}{s_i^2} - \frac{2p_{s_i}}{s_i}\frac{i}{\hbar}\right)(\vec{\mathbf{r}}_i - \vec{\mathbf{x}}_i)^2\right]\exp\left[\frac{i}{\hbar}\vec{\mathbf{p}}_{\vec{\mathbf{x}}_i}\cdot\vec{\mathbf{r}}_i\right] \tag{13}$$

$$U(R, r, s, \sigma) = E_{NN}(R_{NN}) + E_{Ne}(R_{Ne}, s) + E_{ee}(r_{ee}, s) + E_{KE}(s) + E_{PR}(\sigma, s) \tag{14}$$

where R_{NN}, R_{Ne}, and r_{ee} correspond to the internucleus, nucleus-electron and interelectron distances, respectively, s to the electron radius, and σ to the electron spin.

This representation leads to a rather simple electronic energy expression shown in eq 14, $\langle\Psi|\hat{H}|\Psi\rangle$, constituted from the sum of single-particle QM-based kinetic energy (E_{KE}), pairwise Coulomb energies (for nucleus−nucleus, E_{NN}, electron−electron, E_{ee}, and nucleus-electron, E_{Ne}), and the Pauli pairwise spin-dependent (E_{PR}) Hamiltonian that provides the forces that would otherwise arise from antisymmetrization of the wave function product to account for the Pauli Principle. With eFF, we do not solve the Schrödinger equation to describe the ground or low-lying excited states. Rather, we propagate the wave function using the Hamiltonian to describe nonadiabatic dynamics. This leads to a trajectory describing both nuclear and electron motions. To obtain proper statics, we obtain 100−1000s of trajectories starting with various initial conditions and collect statistics on the product distributions. The total eFF energy expression shown in eq 14, has a standard description for electrostatic interactions between a set of zero-dimensional point and Gaussian charges which include nucleus−nucleus (E_{NN}), electron−electron (E_{ee}), and nucleus-electron (E_{Ne}). In addition to the electrostatics, eFF introduces quantum effects through an electron kinetic energy from the Gaussian (E_{KE}) and a spin-dependent Pauli repulsion potential term (E_{PR}) between Gaussians (further details can be found in previous work[14,15]).

An intrinsic limitation of the all-electron FSG-based eFF emanates from the spherical symmetry of the underlying basis functions. For atoms with valence electrons of higher angular momenta, such as p-block elements, the FSG representation misses part of the interaction between the core and valence electrons, due to the absence of nodal structures.

The the SiC electron force field with effective core pseudopotentials (eFF−ECP) scheme requires reformulating and parametrizing the Pauli energy term, E_{PR} in eq 14, for pseudoparticles with Gaussian charge replacing the core electrons and the nucleus, and adjusting the classical electrostatic energies between the pseudocore and valence electrons (core−elec), nuclei (core−nuc), and other pseudocore (core−core) particles as described in.[16]

Various Pauli potentials in the eFF−ECP representation have been tested, although the form of E_{PR} is based on the effect on the kinetic energy of orthogonalizing spherical Gaussians, we have tested formulations based instead on the relation $E_{PR} \propto S^2$, where S is the overlap between two Gaussians: one representing the core and the other an interacting valence electron. In this model representation, we developed the following two different types of ECP: (1) an $s-s$ overlap, for an s-type valence electron and (2) an $s-p$ overlap, for an p-type valence electron.

The corresponding functional forms are given (see refs 16 and 42 for additional details on their derivation) as the following:

$$E_{PR_{s-s}} = a\exp\left(-\frac{br^2}{c + s^2}\right) \tag{15}$$

$$E_{PR_{s-p}} = \mathbf{a}\left(\frac{2}{\mathbf{b}/s + s/\mathbf{b}}\right)^5 (r - \mathbf{c}s)^2\exp\left(-\frac{\mathbf{d}(r - \mathbf{c}s)^2}{\mathbf{e} + s^2}\right) \tag{16}$$

where r in eq 15 is the distance between the s-type pseudocore and an interacting s-type valence electron and, in eq 16, it corresponds to the distance between the s-type pseudocore and the s-type Gaussian representing one of the lobes of a p-type valence electron (see Figure 9). s is the size of the corresponding valence electron, **a** corresponds to the pseudocore wave function amplitude, **b** in eq 15 and **d** in eq 16 to the pseudocore wave function decay factor, **c** in eq 15 and **e** eq 16 to the square of effective pseudocore particle size). For the $s-p$ case **c** corresponds to an off-center measure and **b** to a second effective size that adjusts the overlap amplitude.

We used GARFfield to develop the eFF−ECP force field for SiC tabulated in Table 3. These were obtained by optimizing the parameters for a p-type carbon ECP expression (eq 16) along with our previously published s-type silicon parameters.[42] In this eFF−ECP scheme, both silicon and carbon atoms are represented by a core effective potential and four valence electrons. Because the electrons are represented by spherical Gaussians with a position, size and spin, the parameter optimization for geometries with complex bonding structures is more challenging than that of a conventional force field. We used lower and upper

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

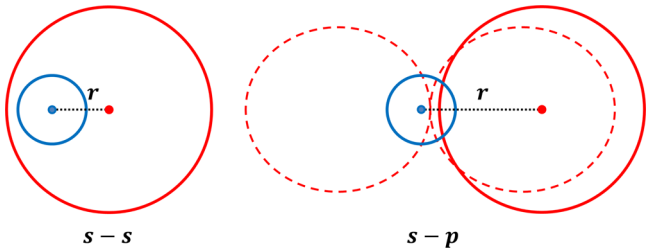


Figure 9. Illustration of distances used in the two functional types of eFF−ECP, where blue circles represent core electrons and red valence electrons. Note that the center of real *p*-type Gaussian is the nodal point. Figure courtesy of Hai Xiao (Caltech).

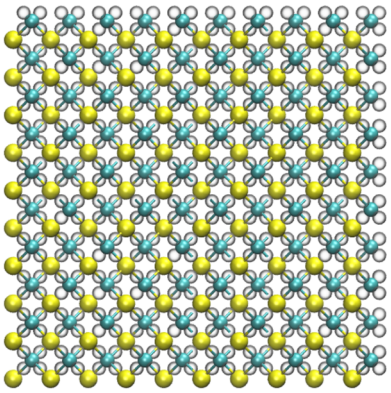


Figure 10. 3C−SiC molecular structure. 5000 atoms, with Si shown in yellow, C in green, and electrons as small circles (not to scale).

parameter range values between 0.5 and 3 for the ECP radius and 0−100.0 for *a*, *b*, *c*, *d*, and *e*.

We find that the eFF−ECP force field obtained with GARFfield accurately captures the proper lattice constant for the 3C−SiC Zinc blend, 4.3496 at 0 K and 4.3551 at 297 K. The latter was computed from the radial distribution function (RDF) between nuclear pairs and the corresponding coordination number. A 5000 atom 3C−SiC fully periodic cubic unit cell (see Figure 10) was minimized to a 1e-8 energy difference using conjugate gradient minimization. Then 100 fs of equilibrated NVT molecular dynamics at 297 K (controlled with a Nosé−Hoover thermostat) was run in LAMMPS with the eff/cut pair style[15] to determine the average RDF and coordination number. Experimental results report a value of 4.3596 at 297 K[48]).

The results for the Si−C bond, and Si−Si−C, Si−C−C, H−C−H, H−Si−H, C−C−Si, C−Si−C, C−Si−Si, Si−C−Si, Si−C−H, and C−Si−H angles from the different conformers in the training set are tabulated in Table 4.

For this case, we ran 80 GARFfield iterations without hill-climbing and a single restart step after 40 iterations. No restraints were applied.

The ECP functional forms used in the reported SiC case, and our previously published applications, are based on the potential interpretation that captures the orbital overlap between the core electrons and a single valence electron. For atoms beyond He, where *p*, *d*, or *f*, angular momentum states play an important role, the absence of nodal structures in the FSG representation misses part of the interaction between the core and valence electrons. Moreover, for elements beyond Neon (Ne), the multiple spatial scales for the core and valence electrons, complicate the form of the E_{PR} based on orbital overlaps. Consequently, we have recently introduced a new formalism into the eFF methodology to treat the core electrons with an effective core potential that replaces the core electrons for norm conserving pseudopotentials.[49] Further details on the functional form of the newer ECP, using angular momentum projections, can be found in ref 16. This new ECP form is also supported in GARFfield.

As mentioned, training of eFF force fields with effective core pseudopotentials (ECPs) involves parametrizing the spin-dependent Pauli repulsion interaction energies between a core pseudoparticle, with other cores and valence electrons, as a function of interparticle distance, electronic spin, and size. For this reason, the eFF−ECP forms involve optimization of the pseudocore size (as a parameter), while electronic wave function centers and sizes are allowed to change (as variables) during bond formation, bond braking, or ionization events. Due to the explicit electrons (and electron cores) in the eFF−ECP description, parameter optimization requires a larger number of geometry optimization iterations to minimize the energy for every core size, while valence electron positions and sizes vary. As opposed to an adiabatic approximation, the nuclear and electronic motions in eFF are coupled through the energy expression, except when the electrons become fully delocalized from nuclei (e.g., plasma state), in which case electrons become noninteracting particles. Excited states are captured from the kinetic energy of the electronic wave function term in the eFF energy expression (eq 14), which, in turn, is related to the electron size. During the dynamic interactions of electrons with other particles in a system, a high wave function kinetic energy is associated with localized electrons (small radius) and a low wave function kinetic energy with diffuse electrons (large radius). Because eFF formally derives the kinetic energy of the electronic wave function from first-principles quantum mechanics and this is not affected during the training of the Pauli repulsion energies (i.e., establishes ground state configuration), electronic excitations remain first-principles based.

IV. FRAMEWORK IMPLEMENTATION AND PERFORMANCE

The core source of GARFfield is written in ANSI C with a message passing interface for parallel computation support. GARFfield uses a modified PGAPack[50] library to perform the evolutionary functions described in a master−slave configuration and a modified interface to the LAMMPS parallel molecular dynamics simulation code[51] to perform the molecular mechanics/dynamics operations needed to rank strings (force fields) residing on each processor. The LAMMPS library used by GARFfield is compiled to include the Reax/c[52] and eFF[15] user packages, both required for training adiabatic ReaxFF and nonadiabatic eFF reactive force fields. The code is written in a modular fashion

Table 3. Carbon *s*−*p* Functional Form Parameters in the SiC−ECP Force Field and Silicon *s*−*s* Parameters from ref 42

atom	ECP radius	*a*	*b*	*c*	*d*	*e*
Si	1.691	0.320852	2.283269	0.814857		
C	0.630348	21.344508	0.715963	0.954384	14.261287	5.314525

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

Table 4. Silicon Carbide Training Set and Optimization Results Using the $s-p$ form ECP for Carbon[a]

structure	feature	QM(BY3LYP)	eFF−ECP	% error
SiH_3CH_3	Si−C bond	3.5527	3.8177	7.4594
SiH_2CH_2	Si−C bond	3.2314	4.0231	24.5002
$2SiH_22CH_2$	Si−C bond	3.6417	3.5643	2.1241
$2(SiH_2CH_2)$	Si−Si−C angle	78.500	80.1411	2.0906
$2(SiH_2CH_2)$	Si−C−C angle	101.5000	99.8589	1.6168
$2(SiH_2CH_2)$	H−C−H angle	106.5000	118.1523	10.9411
$2(SiH_2CH_2)$	H−Si−H angle	107.6000	113.7215	5.6891
$C_2H_5SiH_3$	C−C−Si angle	114.1000	114.8811	0.6846
$2(CH_3)SiH_2$	C−Si−C angle	112.1000	116.4574	3.8871
$Si_2H_5CH_3$	C−Si−Si angle	112.0000	112.298	0.2662
$2(SiH_3)CH_2$	Si−C−Si angle	115.7000	111.8900	3.2930
SiCH	Si−C−H angle	111.3000	116.6054	4.7667
CSiH	C−Si−H angle	109.3000	101.8441	6.8215

[a]All bond lengths in Bohrs and angles in degrees.

to enable incorporation of other force fields (code components are shown in Figure 1).

The speedup from using multiple processors will vary as a function of the GA population size (controlled using the p option in the command line) and the number of population strings replaced at each GA generation (i.e., evolution strategy). Of course, speedup also depend on the number of processors available for the computation and the communication/synchronization overheads associated with distributing and collecting information to and from processors. For example, for a population size of 100 one could use 101 processors effectively to run the (1) master and (100) slaves processes with a full population replacement strategy and a theoretical ×100 speedup (assuming no overheads). However, the default population replacement strategy (steady-state or SSGA) only replaces 10% of the population each GA iteration, hence only 10 processors can be effectively used under parallelism. The percent replacement can be modified using the option (-z, see the Supporting Information). Depending on the computer architecture used to run GARFfield, significant nonlinear speedup can also be achieved from the use of deep cache hierarchies (see ref 11).

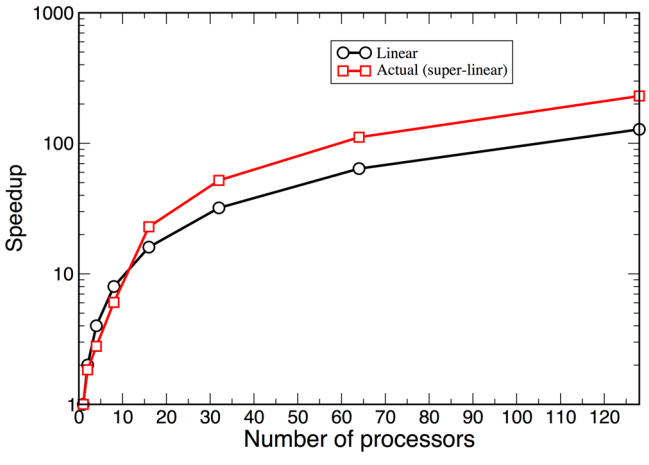


Figure 11. Speedup for ReaxFF SiC parameter set optimization under Los Alamos' National Laboratory supercomputer Mapache.

V. CONCLUDING REMARKS

The multiobjective randomized weighted sum approach used by GARFfield can, in principle, guarantee solutions uniformly distributed over a nonconvex trade-off surface. It is up to the force field developer to choose an optimal force field set among a set of nondeterministic solutions. On the other hand, GARFfield can produce relatively accurate force fields without the use of nondeterminism.

One of the single most critical factors in obtaining fast and reliable force fields from GARFfield is having precise understanding of the underlying energy expressions, and their sensitivity to the individual parameters. Other equally important factors are having a complete, accurate, and diverse (i.e., not exhaustive) training set, having a reasonable starting force field (i.e., complete set of parameters with no singularities), a well-bracketed range for each of the parameters being trained, and proper physical and chemical insight to narrow down the parameter search space.

In general, for preparing ReaxFF force fields, we suggest starting from existing parameters. If these are unavailable, start from parameters available for atoms of similar properties; there is an extensive database of existing ReaxFF parameters for variety of materials and interactions. We find that the use of damped dynamics-based minimization (e.g., the FIRE algorithm[18]) avoids incorrect PES traps that were otherwise common when using local gradient-based minimization (e.g., CG). The latter failed to converge to adequate minima, correct energetics, and geometries.

On the other hand, eFF−ECP parameter optimization requires sampling an increased geometrical and radial search space, given the need to describe explicit electrons and their interaction with the effective core electrons representations. This was solved by increasing the number of minimization steps (-m option in GARFfield's command line, see the Supporting Information) to guarantee geometry convergence and appropriate energetics in the eFF−ECP framework.

Thus, our experience using GARFfield has led us to suggest a general strategy that requires a stepwise parameter optimization process based on incremental training of general, nonbond, valence, and H-bond parameter parameters (or a subset of these, depending on the nature of the force field), followed by an all-inclusive training with small percent offsets from the mean parameter values found at the end of the incremental optimizations (as described in Figure 3). The physical/chemical reasoning for this sequential optimization strategy stems from the fact that nonbond interaction terms (e.g., atom, off-diagonal, and H-bond sections in a ReaxFF force field) play a major role in chemistry of any reactive process. Many reactions start and

dx.doi.org/10.1021/ct5001044 | J. Chem. Theory Comput. 2014, 10, 1426−1439

proceed due to intermolecular interactions. Therefore, it is of crucial importance to first obtain nonbond energy terms that can accurately capture these interactions. Simultaneous optimization of covalent term parameters (i.e., bond, angle, and torsion sections) would negatively affect the quality and partitioning of nonbond and valence energies. Consequently, in our sequential optimization strategy, we start with the nonbond energy parameters, followed by bond, angle, and torsion energy terms. Valence terms are also optimized in sequence by treating highest energy contributions first (i.e., bonds have larger contributions to the total energy of a system than angles or torsions, and angle terms have larger contributions to the total energy than torsions). This strategy can be adapted to the specific force field being trained, considering the training set definitions and other inputs to GARFfield are standardized (see GARFfield's User Manual in the Supporting Information).

As a general rule of thumb, we recommend performing an initial optimization with an unweighted training set, followed by a subsequent run with weights adjusted according to the application requirements (e.g., exact geometries versus energetics) and based on the initial unweighted training errors (trainset.err.best). A thorough revision of the initial training set results, including evaluating the errors reported in trainset.err.best, the physical significance of parameters in ffield.best, visualizing the minimized geometries, and confirming any applied restraints, is essential for adequately spotting and troubleshooting any problems early on. Last but not least, the user should experiment with the different error calculation forms depending on the target properties. For example, the use of squared error criteria (e.g., RMSE) emphasizes values away from equilibrium, whereas normalized differences (e.g., MPE) are less dependent on the rate of change of the fitted functions.

The fact that GARFfield generates multiple sets of solutions per iteration allows for the computation of an approximate of the entire Pareto front. Although this does not necessarily guarantee Pareto optimality, it is known that none of the generated solutions dominates the others. GARFfield, on the other hand, enables selection of a Pareto-optimal solution from the execution of multiple runs, with or without randomization of training set weights. Manual consolidation and inspection of the trainset.err.best file results for all runs is then required for the selection of the Pareto-optimal solution.

Future versions of the GARFfield force field optimization framework will include heuristics that allow finer control on the optimization procedure. For example, the user will be able to specify sequential optimization stages by groups of parameters (e.g., bond-order parameters, followed by atom and off-diagonal parameters associated to nonbond interactions, followed by bond parameters, followed by angle and torsion parameters, and finalizing with H-bond parameters (when required)). We also expect to add a parameter prepopulation step from an UFF pass of the training set. This will allow initialization of parameters for new elements in force fields like ReaxFF.

In summary, GARFfield (1) is a general force field optimization framework, with significant flexibility, as opposed to a hardwired single force field solution, (2) enables multiobjective optimization of complex force fields (including adiabatic and nonadiabatic reactive force fields), (3) supports alternative optimization criteria to improve computationally efficiency (i.e., 2-norm force versus standard geometry optimization) and convergence speed (automatic switching to gradient-based optimization), (4) enables nondeterministic solutions, via hill-climbing and weight randomization of objective functions, (5) scales linearly in the string population size (p_S) and population replacement (p_{rf}) fraction (i.e., with a speedup proportional to $p_S \times p_{rf}/p$) (for p processors), and has demonstrated superlinear speedups under multihierarchy cache cluster architectures, (6) is designed around a modular software architecture that is amenable to extensions of default base of force fields (i.e., ReaxFF, eFF−ECP, charge-optimized many-body interaction potentials −COMB and atomistic and coarse-grain Morse potentials), and objective functions and restraints, and (7) offers a unique opportunity for systematic reversible multiscale simulations, via up-scaling (e.g., from atomistic to coarse-grain force field) and down-scaling (e.g., coarse-grain tuning from atomistic).

An open source version of the GARFfield code will be made available after this publication, through the LAMMPS Web site (http://lammps.sandia.gov) or through http://www.wag.caltech.edu/home/ajaramil/.

■ ASSOCIATED CONTENT

Supporting Information

Four additional files associated with this paper are included: (1) GARFfield Users Manual, (2) a file specifying the ReaxFF force field file format, (3) the optimized SiC ReaxFF force field file, and (4) the QM training set used to obtain the optimized SiC ReaxFF force field reported. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*A. Jaramillo-Botero. E-mail: ajaramil@caltech.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was partially funded by the Defense Advanced Research Projects Agency (DARPA) (Grant No. N660011214037, Drs. Tyler McQuade and Anne Fischer) and by the US Department of Transportation (DOT), Federal Highway Administration (FHWA) (Award Number BAA No. DTFH61-09-R-00017, Dr. Kunik Lee). The authors thank Hai Xiao for his contributions on the eFF−ECP and Drs. Sergey Zybin and Qi An for insightful discussions.

■ REFERENCES

(1) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727−748.
(2) So, S. S.; Karplus, M. *J. Med. Chem.* **1996**, *39*, 1521−1530.
(3) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368−1372.
(4) Unger, R. *Struct. Bonding (Berlin, Ger.)* **2004**, *110*, 153−175.
(5) Ho, K. M.; Shvartsburg, A. A.; Pan, B. C.; Lu, Z. Y.; Wang, C. Z.; Wacker, J. G.; Fye, J. L.; Jarrold, M. F. *Nature* **1998**, *392*, 582−585.
(6) Hunger, J.; Huttner, G. *J. Comput. Chem.* **1999**, *20*, 455−471.
(7) Wang, J. M.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1219−1228.
(8) Angibaud, L.; Briquet, L.; Philipp, P.; Wirtz, T.; Kieffer, J. *Nucl. Instrum. Methods Phys. Res., Sect. B* **2011**, *269*, 1559−1563.
(9) Handley, C. M.; Deeth, R. J. *J. Chem. Theory Comput.* **2012**, *8*, 194−202.
(10) Larsson, H. R.; van Duin, A. C. T.; Hartke, B. *J. Comput. Chem.* **2013**, *34*, 2178−2189.
(11) Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182−197.
(12) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *J. Phys. Chem. A* **2001**, *105*, 9396−9409.

(13) Yu, J. G.; Sinnott, S. B.; Phillpot, S. R. *Phys. Rev. B* **2007**, *75*, 08531101−08531113.
(14) Su, J. T.; Goddard, W. A. *Phys. Rev. Lett.* **2007**, *99*, 1850031−185004.
(15) Jaramillo-Botero, A.; Su, J.; Qi, A.; Goddard, W. A. *J. Comput. Chem.* **2011**, *32*, 497−512.
(16) Xiao, H.; Jaramillo-Botero, A.; Theofanis, P.; Goddard, W. *Electron force field effective core potentials for high-Z elements of the periodic table*; PSAAP Final Report, Materials and Process Simulation Center: California, USA, 2013.
(17) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358−3363.
(18) Bitzek, E.; Koskinen, P.; Gahler, F.; Moseler, M.; Gumbsch, P. *Phys. Rev. Lett.* **2006**, *97*, 17020101−17020104.
(19) Takeda, Y.; Nakamura, K.; Maeda, K.; Matsushita, Y. *J. Am. Ceram. Soc.* **1987**, *70*, C266−C267.
(20) Schulz, K.; Durst, M. *Filtr. Sep.* **1994**, *31*, 25−28.
(21) Rosenbloom, A. J.; Sipe, D. M.; Shishkin, Y.; Ke, Y.; Devaty, R. P.; Choyke, W. J. *Biomed. Microdevices* **2004**, *6*, 261−267.
(22) Li, Z. Y.; Kusakabe, K.; Morooka, S. *J. Membr. Sci.* **1996**, *118*, 159−168.
(23) Zorman, C. A.; Fleischman, A. J.; Dewa, A. S.; Mehregany, M.; Jacob, C.; Nishino, S.; Pirouz, P. *J. Appl. Phys.* **1995**, *78*, 5136−5138.
(24) Kenawy, S. H.; Nour, W. M. N. *J. Mater. Sci.* **2005**, *40*, 3789−3793.
(25) Lopez-Honorato, E.; Tan, J.; Meadows, P. J.; Marsh, G.; Xiao, P. *J. Nucl. Mater.* **2009**, *392*, 219−224.
(26) Carpenter, D.; Ahn, K.; Kao, S.; Hejzlar, P.; Kazimi, M. *Assessment of Silicon Carbide Cladding for High Performance Light Water Reactors*; Nuclear Fuel Cycle Program (CANES Reports), 2007; Vol. MIT-NFC-TR-098.
(27) Naserifar, S.; Goddard, W. A.; Liu, L. C.; Tsotsis, T. T.; Sahimi, M. *J. Phys. Chem. C* **2013**, *117*, 3320−3329.
(28) Naserifar, S.; Liu, L. C.; Goddard, W. A.; Tsotsis, T. T.; Sahimi, M. *J. Phys. Chem. C* **2013**, *117*, 3308−3319.
(29) Elyassi, B.; Sahimi, M.; Tsotsis, T. T. *J. Membr. Sci.* **2007**, *288*, 290−297.
(30) Elyassi, B.; Sahimi, M.; Tsotsis, T. T. *J. Membr. Sci.* **2008**, *316*, 73−79.
(31) Devlin, D. J.; Currier, R. P.; Barbero, R. S.; Espinoza, B. F.; Elliott, N. In *Chemical Vapor Deposition of Refractory Metals and Ceramics II. Materials Research Society Symposium Proceedings*; Besmann, T., Gallois, B., Warren, J., Eds.; Materials Research Society: Pittsburgh, PA, 1992; Vol. *250*; pp 245−250.
(32) Morell, J. I.; Economou, D. J.; Amundson, N. R. *J. Mater. Res.* **1992**, *7*, 2447−2457.
(33) Gupta, D.; Evans, J. W. *J. Mater. Res.* **1991**, *6*, 810−818.
(34) Chenoweth, K.; van Duin, A. C. T.; Goddard, W. A. *J. Phys. Chem. A* **2008**, *112*, 1040−1053.
(35) Bachrach, S. M. *J. Am. Chem. Soc.* **2004**, *126*, 5018−5018.
(36) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(37) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(38) Hariharan, P.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.
(39) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(40) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(41) Mousavipour, S. H.; Saheb, V.; Ramezani, S. *J. Phys. Chem. A* **2004**, *108*, 1946−1952.
(42) Theofanis, P. L.; Jaramillo-Botero, A.; Goddard, W. A.; Mattsson, T. R.; Thompson, A. P. *Phys. Rev. B* **2012**, *85*, 941091−941097.
(43) Su, J. T.; Goddard, W. A. *J. Chem. Phys.* **2009**, *131*, 24450101−24450120.
(44) Kim, H.; Su, J. T.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 15101−15105.
(45) Theofanis, P. L.; Jaramillo-Botero, A.; Goddard, W. A.; Xiao, H. *Phys. Rev. Lett.* **2012**, *108*, 045501.
(46) Chenard-Lemire, C.; Lewis, L. J.; Meunier, M. *Appl. Surf. Sci.* **2012**, *258*, 9404−9407.
(47) Su, J. T.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 1001−1005.
(48) Taylor, A.; Jones, R. In *Silicon Carbide - A High Temperature Semiconductor*; O'Connor, J., Smiltens, J., Eds.; Pergamon Press: Oxford, London, New York, Paris, 1960; Vol. *1*; p 147.
(49) Kahn, L. R.; Goddard, W. A. *Chem. Phys. Lett.* **1968**, *2*, 667−670.
(50) Levine, D. *Users Guide to the PGAPack Parallel Genetic Algorithm Library*; Argonne National Laboratory, U.S. Department of Energy: Argonne, IL, 1996; pp 1−73.
(51) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1−19.
(52) Aktulga, H. M.; Fogarty, J. C.; Pandit, S. A.; Grama, A. Y. *Parallel Computing* **2012**, *38*, 245−259.