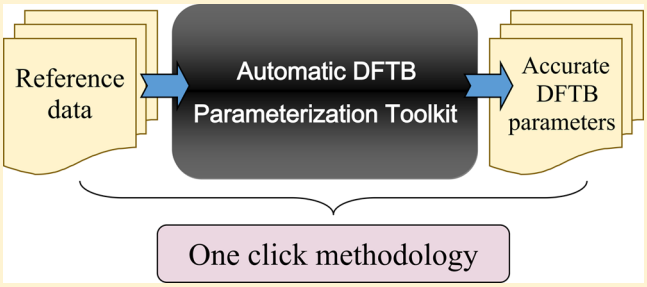
JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

Automatized Parameterization of DFTB Using Particle Swarm Optimization

Chien-Pin Chou,[†] Yoshifumi Nishimura,[†,‡,⊥] Chin-Chai Fan,[†] Grzegorz Mazur,[§] Stephan Irle,[‡,∥] and Henryk A. Witek*[,†]

[†]Department of Applied Chemistry and Institute of Molecular Science, National Chiao Tung University, Hsinchu 30010, Taiwan
[‡]Department of Chemistry, [∥]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan
[§]K. Gumiński Department of Theoretical Chemistry, Jagiellonian University, Ingardena 3, 30-060 Cracow, Poland

ABSTRACT: We present a novel density-functional tight-binding (DFTB) parametrization toolkit developed to optimize the parameters of various DFTB models in a fully automatized fashion. The main features of the algorithm, based on the particle swarm optimization technique, are discussed, and a number of initial pilot applications of the developed methodology to molecular and solid systems are presented.



1. INTRODUCTION

The density-functional tight-binding (DFTB) method is a collection of computational techniques developed for robust quantum mechanical calculations of properties of large molecular systems and solids. In addition to the standard, noniterative DFTB technique (usually abbreviated DFTB, DFTB1, or NCC-DFTB),[1,2] DFTB methods comprise more sophisticated, and consequently more accurate, methods based on low-order Tayler expansion of the DFT total energy with respect to the density fluctuations (SCC-DFTB[3] or DFTB2 and DFTB3[4,5]) or with respect to the spin density fluctuations (sDFTB).[6] Over the last 20 years since the DFTB model was introduced, DFTB methods have been applied for modeling numerous molecular and extended systems.[7−27]

Careful approximations were introduced in the DFTB methods in order to reduce the computational effort and to maintain reasonable accuracy. For example, tabulation of precomputed two-center overlap and Hamiltonian integrals ensures that DFTB is about 3 orders of magnitude faster than traditional DFT methods at this step of the computational process. The precomupted integrals, together with other necessary DFTB parameters, are usually stored in the form of so-called Slater−Koster (SK) files.[28] Several sets of SK files have been developed and published, for either general or specific purposes.[29−42]

Building a transferable and physically meaningful set of SK parameters is associated with a considerable parametrization effort. Until recently, the development of SK files has been performed manually in a somewhat *ad hoc* manner. Manual parametrization, despite providing good control over the physical relevance of the resulting parameters, has a number of considerable drawbacks. First, it is a very time-consuming and laborious process; the development of a single parameter set describing well a number of molecular properties can take easily a few months, if not longer. The second drawback concerns natural limitations in probing the parameter space associated with a limited number of fitness evaluations in the manual parametrization process. This limitation often leads to parameter sets that are not optimal and that reproduce the targeted molecular properties only in an approximate manner. This limitations of the manual parametrization process have been recently recognized, and a number of techniques aiming at semiautomatic or automatic development of SK files have been presented.[43−49]

One of the main challenges in DFTB parametrization methodology is developing a fully automatized approach that allows for dense and representative probing of the parameter space in a robust and efficient manner. An algorithm of this sort has not been reported yet. The main purpose of the current Article is to report a methodology that is capable of fulfilling such requirements. Namely, we employ the particle swarm optimization (PSO) algorithm[50] to find the optimal values of DFTB parameters. The resulting parametrization methodology is fully automatic and can be used for efficiently developing the DFTB parameter sets (i.e., the Slater−Koster files) optimized to reproduce an arbitrary physical property of an arbitrary molecular or extended system. This Article discusses the main features of such a PSO-based DFTB parametrization algorithm and gives a number of its applications. It is important to stress here that the proposed methodology can be used for developing both general and target-specific DFTB parametrizations. We believe that both directions of development

Received: July 15, 2015
Published: November 19, 2015

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

are important and have merits. General DFTB parametrizations, reproducing approximately a wide spectrum of various molecular properties or a specific solid property for a large group of systems, will be of primary importance for novice DFTB users. Experienced DFTB users may be interested in developing target-specific DFTB parametrizations, which will be particularly suitable for accurately modeling properties of a limited group of systems containing, in an extreme case, only a single molecule or solid. The framework presented here is suitable for both types of applications. It is clear that existing general parametrizations (e.g., *mio*,[3] *3ob*,[39] or *pbc*[51]) can be further fine-tuned using the methodology presented here. Our long-term goal is to provide the DFTB community with a fully automated DFTB parametrization toolkit for creating DFTB parameters for new elements and for specific, purpose-driven applications. Such a program will be presented in due course.

2. DFTB

This section briefly introduces the DFTB method and its self-consistent extensions. For more details, see refs 1−5.

The total DFTB energy can be written as

$$E_{\mathrm{DFTB1}} = E_{H_0} + E_{\mathrm{rep}} = \sum_{i}^{MO} n_i \langle \psi_i | \hat{H}_0 | \psi_i \rangle + \frac{1}{2} \sum_{\mathrm{A,B}} V_{\mathrm{AB}}^{\mathrm{rep}} \tag{1}$$

where n_i is the occupation number of a molecular orbital ψ_i. The molecular orbital ψ_i is expressed as a linear combination

$$\psi_i = \sum_{\mu} c_{\mu i} \phi_\mu \tag{2}$$

of pseudoatomic valence orbitals ϕ_μ, which are conventionally determined by solving a confined Kohn−Sham equation

$$[\hat{T} + V + V_{\mathrm{conf}}]\phi_\mu = \varepsilon_\mu \phi_\mu \tag{3}$$

where V comprises the standard Coulomb and centrifugal terms and V_{conf} is an additional confining potential. The actual form of the confining potential V_{conf} is a parameter of the DFTB model; it is customary to choose it in a parabolic $(\frac{r}{r_{\mathrm{orb}}})^2$ or quartic $(\frac{r}{r_{\mathrm{orb}}})^4$ representation, with r_{orb} being subject to optimization. In many applications,[3,40,52] r_{orb} was not optimized but was *ad hoc* selected to be twice the atomic covalent radius. Recently, Heine and collaborators discovered[44] that releasing the integer-like limitations imposed on the power n_{orb} of the confining potential is advantageous for reproducing the shape of the DFT band structure of solids; in many cases, it was necessary to select n_{orb} greater than 10 in order to achieve good correspondence with DFT, which, in practice, signifies that a particle is confined in a spherical 3D box of radius r_{orb}. In Heine's model, the optimal values of the orbital compression radius r_{orb} and the orbital compression exponent n_{orb} of each atom remain to be determined in the parametrization process.

In this work, we go a few steps further in the generalization process and utilize the Woods−Saxon form of the confining potential.[53] The Woods−Saxon potential V_{WS} has the following form

$$V_{\mathrm{WS}}(r) = \frac{W}{1 + \exp^{-a(r-r_0)}} \tag{4}$$

where W, a, and r_0 are the height, slope, and half-height radius, respectively, of the confining potential. The Woods−Saxon confinement has two important advantages over power confinement: (i) it does not affect the wave functions and orbital energies of core electrons and (ii) it guarantees correct asymptotic behavior of valence orbitals. The first property comes from the fact that the Woods−Saxon confinement potential is practically zero in the core region while power potentials take on nonzero values for small radial distances, particularly for small values of n_{orb}. The second property is linked to the constant finite value of the Woods−Saxon confinement potential at large nuclear separations, which does not alter the correct asymptotics of the wave function. This correct asymptotic behavior of V_{WS} and additional variational freedom associated with the third optimizable confinement parameter often proves useful for reproducing various molecular properties of solids and molecules.

The total energy in the self-consistent variants of DFTB is given by the following expressions

$$E_{\mathrm{DFTB2}} = E_{\mathrm{DFTB1}} + \frac{1}{2} \sum_{AB} \Delta q_A \Delta q_B \gamma_{AB}$$

$$E_{\mathrm{DFTB3}} = E_{\mathrm{DFTB2}} + \frac{1}{3} \sum_{AB} \Delta q_A^2 \Delta q_B \Gamma_{AB} \tag{5}$$

where Δq_{A} is the induced, self-consistently determined Mulliken charge on atom A and γ_{AB} and Γ_{AB} are distance-dependent functionals computed from atomic hardnesses and their derivatives; for details, see elsewhere.[3,4,5]

The symbol $V_{\mathrm{AB}}^{\mathrm{rep}}$ in eq 1 denotes the distance-dependent repulsive pair potential between atoms A and B, which physically corresponds mainly to the interaction between the chemical cores of both atoms. In principle, it is possible to derive the shape of $V_{\mathrm{AB}}^{\mathrm{rep}}$ directly from theory.[47] However, in practice, one usually represents a repulsive potential in the DFTB method as a set of spline functions. The spline coefficients are subsequently selected in a way to reproduce targeted molecular properties of selected reference systems. In our approach, the spline coefficients are obtained by solving the least-squares problem of an accompanying linear system constructed directly from energies and geometries of the reference molecules and/or solids.[43]

3. PARTICLE SWARM OPTIMIZATION

PSO, proposed by Eberhart and Kennedy in 1995,[50] is an evolutionary optimization algorithm inspired by the social behavior of flocking animals. The main objective of the PSO algorithm is to locate a point in a multidimensional search space that gives an optimal value of some functional, usually referred to as the fitness function, defined over this space. Very little limitations are imposed on the functional; it can be nonlinear, discontinuous, or even undefined at some points of the search space. The optimization process proceeds by selecting a set (swarm) of probing points, usually referred to as particles, in the parameter space and varying semistochastically their positions using the PSO algorithm until an optimal value of the functional (preferably, the global minimum) is located. Numerous applications have utilized PSO in various fields and for various purposes because of the relative simplicity of the algorithm.[54−62]

The PSO algorithm was developed to solve the following problem

$$\text{minimize } f(\vec{x})\colon \mathbb{R}^n \rightarrow \mathbb{R} \tag{6}$$

where $\vec{x} = [x_1, x_2, ..., x_n]^T$ is the vector of decision variables.

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

Let N be the number of particles in the swarm, and let $\vec{x}_i \in \mathbb{R}^n$ and $\vec{v}_i \in \mathbb{R}^n$ denote the current position and velocity of the ith particle, respectively. Furthermore, let $\vec{p}_i$ denote the optimal position (personal best) of the ith particle during all previous PSO iterations, and let $\vec{g}$ denote the optimal position (global best) within the whole swarm of particle during all previous PSO iterations. A basic PSO algorithm is presented in Algorithm 1. The PSO parameters ω, c_1, and c_2 are called inertia

Algorithm 1 Basic particle swarm optimization (PSO) algorithm

```
for i=1 to N do
    Randomize position x_i and velocity v_i of each particle
    Set person best to the initial position : p_i ← x_i
end for
for iIter=1 to maxIter do
    for i=1 to N do
        Evaluate the function value f(x_i)
        if f(x_i) < f(p_i) then
            p_i ← x_i
            if f(p_i) < f(g) then
                g ← p_i
            end if
        end if
        for each dimension k of the vector x do
            Update the velocity vector v_i by
            v_i(k) ← ω(iIter)·v_i(k) + c_1·[p_i(k) − x_i(k)] + c_2·[g(k) − x_i(k)]
            Update the position vector x_i ← x_i + v_i
        end for
    end for
end for
```

weight, cognitive acceleration constant, and social acceleration constant, respectively. Inertia weight controls the degree to which the velocity contributes to the position change of the particle in the next iteration. The constants c_1 and c_2 control the tendency of particles to move toward the personal or global best positions, respectively. The PSO parameters can be iteration-dependent. It is customary to represent ω by a linearly decreasing (e.g., from 0.9 to 0.4) function of the iteration number to provide greater diversity in the early stages and better convergence in the final stages of the optimization process. The values of c_1 and c_2 are commonly chosen to be between 1.5 and 2.

In most real-world scenarios, one does not usually want to minimize a single fitness function $f(\vec{x})$ but rather a set of fitness functions $\{f_i(\vec{x}),\ i = 1 \ldots m\}$. In the DFTB context, one may want, for example, to find a parameter set that would produce the best geometries, best frequencies, and best dissociation energies simultaneously. It is always possible to combine the fitness function components $f_i(\vec{x})$ into an overall fitness function using some appropriate weighting scheme

$$\text{minimize } f(\vec{x}) = \sum_i^m a_i f_i(\vec{x}) \tag{7}$$

A serious drawback of this approach is that an improper choice of the weight factors a_i can introduce bias in the final function value, especially when the component objective functions have different scales or units. Alternatively, one may use a multicomponent fitness function in the optimization process

$$\text{minimize } \vec{f}(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x}), \ldots, f_m(\vec{x})] \tag{8}$$

where m signifies the number of components in the multicomponent fitness function. The first multiobjective PSO (MOPSO) algorithm based on the Pareto-optimal approach was proposed by Moore and Chapman in 1999;[63] many modifications and improvements have been developed since then (for details, see a comprehensive review by Reyes-Sierra and Coello[64]).

The main practical difference between conventional single-objective PSO and the Pareto-optimal-based MOPSO algorithm is the form of the output. The PSO algorithm produces a single solution, which is possibly different in consecutive runs of the algorithm, whereas the MOPSO algorithm returns a set of Pareto-optimal solutions giving a discrete representation of the Pareto front. The decision of whether to choose only one of these solutions remains with the user. Usually, the selection of the final solution is performed by inspection of the Pareto-optimal solutions and choosing the one that is optimal in the opinion of the user.

4. DFTB PARAMETERIZATION USING THE PSO ALGORITHM

In this section, we briefly discuss how the PSO algorithm has been adapted for the purpose of DFTB parametrization. We have defined the following set of fitness functions for molecules

- bond length fitness function
- bond angle fitness function
- dihedral angle fitness function
- atomization energy fitness function
- reaction energy fitness function
- conformation change energy fitness function
- vibrational frequencies fitness function

and the following set of fitness functions for solids

- bond length fitness function
- lattice parameters fitness function
- band structure fitness function
- atomization energy fitness function
- phase change energy fitness function
- bulk modulus fitness function

These sets can be easily extended to other properties such as dipole moments, polarizabilities, IR and Raman spectra, elastic constants, and many others. Augmenting our program with a new property requires that a short script be written that extracts a given molecular or crystal property from the DFTB output file and ports it to the parametrization program. Each of these fitness functions can be used as an independent component of the vector fitness function in MOPSO, or, alternatively, one may define an overall scalar fitness function by a linear combination of various components according to eq 7.

Let $\{A_1, A_2, \ldots, A_p\}$ be the set of chemical elements defining the scope of a given DFTB parametrization. Let the parameter search space be spanned by the following set of variables

- for each element A_i, $i \in \{1, \ldots, p\}$
  - orbital confining parameters: r_i^{orb}, a_i^{orb}, and W_i^{orb}
  - (optional) shell-resolved orbital confining parameters:
    - $r_i^{\text{orb}}(s)$, $r_i^{\text{orb}}(p)$, ...
    - $a_i^{\text{orb}}(s)$, $a_i^{\text{orb}}(p)$, ...
    - $W_i^{\text{orb}}(s)$, $W_i^{\text{orb}}(p)$, ...
  - (optional) density confining parameters: r_i^{dens}, a_i^{dens}, and W_i^{dens}
  - orbital energies of unoccupied shells (e.g., $\varepsilon_i(d)$)
  - (optional) orbital energies of occupied shells (e.g., $\varepsilon_i(p)$)
  - (optional) Hubbard parameters H_i or shell-resolved Hubbard parameters $H_i(s)$, $H_i(p)$, and $H_i(d)$
  - (optional) derivatives of Hubbard parameters H_i' or shell-resolved Hubbard parameters $H_i'(s)$, $H_i'(p)$, and $H_i'(d)$

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

- for each pair of elements (A_i, A_j), $i, j \in \{1, ..., p\}$
  - the number $K(i,j)$ and the location $x_k(i,j)$, $k \in \{1, ..., K(i,j)\}$ of the spline knots defining the repulsive DFTB potential $V^{rep}(A_i, A_j)$
  - (optional) additional constraints for $V^{rep}(A_i, A_j)$ [Additional constraints of this form often provide a very convenient way of speeding-up the optimization process, e.g., constrained optimization of the value of the second derivative of $V^{rep}(A_i, A_j)$ at the distance corresponding to a typical equilibrium distance between A_i and A_j is a very effective method of obtaining correct value of the stretching frequency of this bond.]

Then, a PSO particle corresponds to a vector $\vec{x}$ denoting a point in the parameter space defined by these variables.

It is sometimes advantageous to restrict the value of a given parameter by providing a value for its upper and lower bounds. Such a restriction prevents the PSO algorithm from probing regions of the parameter space that are mathematically infeasible. However, in the DFTB parametrization process, some infeasible regions can also be located inside the predefined search interval. Usually, their presence cannot be anticipated prior to performing the optimization. PSO particles that enter an infeasible region are reset to random values in the next iteration.

The basic PSO and a specific MOPSO algorithm[65] have been implemented using the C++ programming language. Both implementations comprise the modified Erepfit code,[43] which has been extended to also treat periodic boundary conditions. Confined atomic orbitals and density are generated with the relativistic atomic code.[53] A patched DFTB+ code[66] is interfaced with the PSO optimizer to obtain the data necessary for the parametrization and to evaluate the properties necessary to define the objective functions. Two graphical user interface tools have been developed to improve the efficiency of preparing the input for the optimization process.

5. TEST APPLICATIONS

Initial applications of the PSO DFTB parametrization toolkit, meant to illustrate the capabilities and limitations of the new tool reported here, include developing optimal DFTB parameters suitable for reproducing the following physical properties of selected solids and molecules:

- band structures of various phases of crystalline silicon
- accurate band gaps for diamond-like silicon
- lattice constants, bulk moduli, and correct relative energies of basic crystalline silicon phases
- vibrational frequencies of the methanol molecule
- internal rotation energy curve for hydrogen peroxide

A detailed discussion of the results is presented below in separate sections.

5.1. Band Structures of Silicon. The objective of this test application is to develop a set of DFTB electronic parameters for silicon, referred to in the following as *psopbe*, capable of reproducing DFT band structures of basic Si crystalline phases. We consider explicitly the following four phases of solid silicon: diamond, lonsdaleite, body-centered cubic (BCC), and face-centered cubic (FCC). The lattice constants of diamond (a = 5.431 Å; ref 67) and lonsdaleite (a = 3.800 Å, c = 6.280 Å; ref 68) were taken from experiment. The lattice constants a = 3.081 Å for the BCC phase and a = 3.868 Å for the FCC phase were optimized with VASP[69−72] (PBE functional,[73,74] projector augmented-wave potentials,[75,76] plain-wave basis set with energy cutoff of 500 eV, 32 × 32 × 32 Monkhorst−Pack mesh). The four sets (diamond, lonsdaleite, FCC, and BCC) of reference electron dispersion curves were computed with the same methodology. The resulting band structures were shifted in such a way that the valence band maximum of each structure matches the origin of the energy axis. The definition of the PSO fitness function was based on the L^1 norm of the pointwise difference vector between the DFT/PBE reference and DFTB/*psopbe* bands for all of the valence and conductance DFT bands located below 15 eV for FCC and BCC and below 5 eV for diamond and lonsdaleite. With the minimal basis sets pertinent to the DFTB model, it is usually not legitimate to include more conduction bands in the definition of the fitting function. Intuitively, the PSO fitness function corresponds to the mean absolute deviation (MAD) between the reference DFT bands and the calculated DFTB ones.

Before reporting the results of our optimization, we briefly discuss the performance of existing DFTB parametrizations for silicon: *pbc* and *matsci*. Figure 1 shows the band structures of

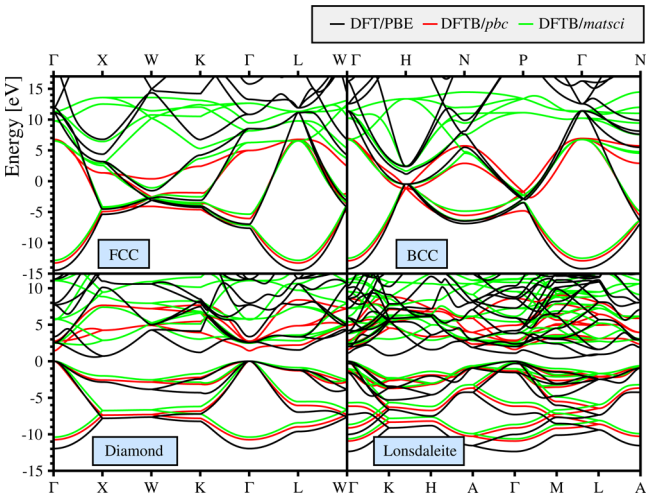


Figure 1. Comparison of band structures of four phases of solid silicon computed with DFT/PBE, DFTB/*pbc*, and DFTB/*matsci*.

the diamond, lonsdaleite, BCC, and FCC phases of solid silicon. Both parameter sets were fitted mainly to silicon clusters, so it should not be expected that the *pbc* and *matsci* band structures will superbly resemble the DFT reference. Indeed, the correspondence is rather mediocre. The average deviation from DFT/PBE per point for valence bands is 0.63 eV for *pbc* and 0.92 eV for *matsci*, and for conductance bands that are not located higher than 5 eV above the valence band maximum, it is 1.21 eV for *pbc* and 0.71 eV for *matsci*. Both sets show similar performance, with a slightly better description of the conduction bands produced by *matsci*. The reason for this behavior is simple. The *pbc* set utilizes only 3s and 3p atomic orbitals, whereas the *matsci* set includes 3d orbitals as well. The major problem with *pbc* and *matsci* is the incorrect description of the lowest conduction bands in the nonmetallic structures. In particular, both sets fail miserably in predicting the indirect band gap for diamond and lonsdaleite.

The DFTB band structures computed with the *psopbe* parameters are shown in Figure 2 together with the reference DFT/PBE curves. The DFT band structures are well-reproduced by the set of PSO-optimized parameters in the fitting regions. The average deviation from DFT/PBE per point for valence bands

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

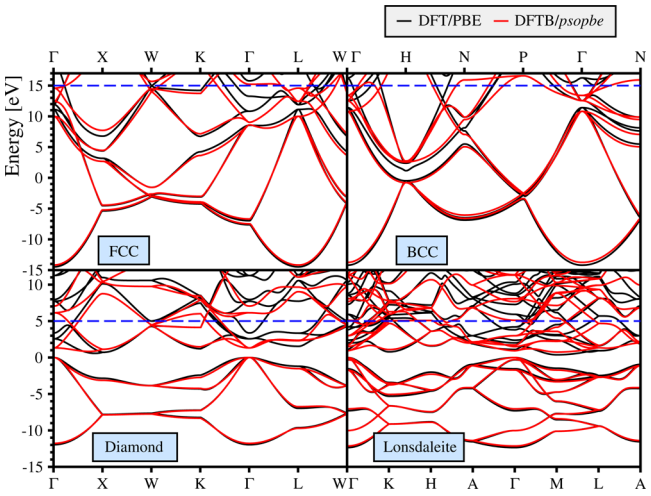


Figure 2. DFTB band structures of four phases of solid silicon computed with the *psopbe* parameters closely resemble the DFT/PBE reference. The DFT bands located below the blue dashed lines were included in the definition of the fitness function.

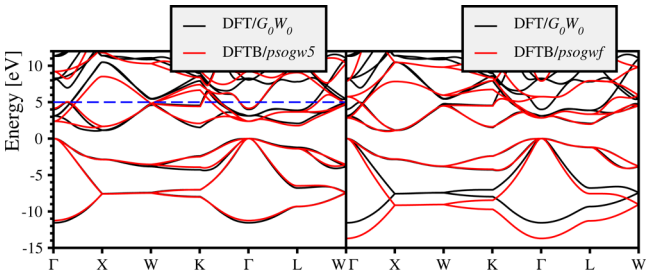


Figure 3. Comparison of Si diamond band structures computed using the DFTB/*psogw5* parameters (left) and DFTB/*psogwf* parameters (right) with the reference DFT/G_0W_0 band structure. The valence band maximum of each band structure determines the origin of the energy axis. The blue dashed line gives the upper offset of bands included in the definition of the fitness function in the DFTB/*psogw5* parametrization.

is only 0.14 eV, whereas for the conductance bands that are not located higher than 5 eV above the valence band maximum, it is 0.46 eV. The most distinct difference between the DFT/PBE and DFTB/*psopbe* band structure curves concerns the behavior of low-lying conductance bands around the Γ point. This behavior is responsible for a substantial error of 1.25 eV with respect to the DFT/PBE reference in the reproduced direct band gap for the diamond phase. A natural solution for this problem is to extend the minimal DFTB basis by including more virtual pseudoatomic orbitals. Another possible solution to this problem is discussed in the next section.

5.2. Accurate Band Gaps in Diamond Si. The objective of this test application is to develop a set of DFTB electronic parameters capable of reproducing accurate band gaps in diamond-like silicon. The DFT/PBE band gaps, used as a reference in the previous section, are quite inaccurate in comparison with experiment. Therefore, in the current section, we resort to band structures computed within the GW approximation.[77] The GW reference band structures were computed at the experimental lattice geometry of diamond (a = 5.431 Å; ref 67) using the VASP code[69−72] (PBE functional,[73,74] projector augmented-wave potentials,[75,76] plain-wave basis set with energy cutoff of 500 eV, followed by the quasi-particle single-shot approach, aka G_0W_0, and 12 × 12 × 12 Monkhorst−Pack k-point sampling). The final band structures were interpolated by maximally localized Wannier functions (MLWFs) using the wannier90[78] code. The resulting direct (3.12 eV) and indirect (1.05 eV) reference band gaps correspond to experimental[79,80] values of 3.40 and 1.17 eV, respectively, much better than PBE band gaps in the previous section of 2.57 and 0.57 eV, respectively.

Two different definitions of the fitness function were used. The first fitness function was defined using all valence and conductance bands that are not located higher than 5 eV above the valence band maximum; we denote the resulting DFTB parameters *psogw5*. The second fitness function was defined using only two bands, the highest valence band and the lowest conductance band; the resulting parametrization will be denoted *psogwf*, where the letter *f* stands for frontier bands.

The resulting DFTB band structures obtained with the *psogw5* and *psogwf* parameters are shown in Figure 3. The quality of both parameter sets is quite good; the average deviation from the DFT reference is only 0.17 eV per point for *psogw5* and 0.06 eV for *psogwf* in the fitted region. The computed band gaps are listed in Table 1 together with analogous results corresponding to previous DFTB parametrizations, DFT reference, and experiment. As expected, the best band gaps, differing from the reference by 0.27 eV for the direct gap and by 0.07 eV for the indirect gap, are produced by the *psogwf* parameters.

Note that in this test application only the band structures of the diamond phase were included in the optimization. We also performed transferability tests on the band structures of the three other studied phases. The *psogw5* parametrization shows better transferability than the *psogwf* parametrization on the BCC and FCC phases owing to its better description of the valence bands of diamond silicon. Nevertheless, the *psogwf* parametrization gives a better description of the band gaps of the diamond phases, which also improves the band gaps of the lonsdaleite phase (see Table 1).

5.3. Band Structures Parameterization of Solid Silicon with Double Valence Basis Set. The objective of this test application is to develop a set of DFTB electronic parameters including 4s, 4p, and 4d atomic orbitals of silicon to investigate whether the description of the low-lying conduction bands of the diamond and lonsdaleite phases of silicon can be improved by including the second shell of the atomic orbitals of silicon.

In this section, we considered two parametrizations, *psopbe_dv* optimized to the DFT/PBE reference band structures and *psogw8_dv* optimized to the DFT/G_0W_0 reference band structures. The reference band structures were computed using the same geometries and configurations as those in the previous two sections. The fitness definitions have been slightly adjusted by including all bands up to 8 eV above the Fermi energy for the diamond and lonsdaleite phases in order to obtain a better description of the low-lying conduction bands, whereas the fitness definitions for the BCC and FCC phases have been taken up to 15 eV above the Fermi energy as before.

Figures 4 and 5 show the band structures computed with the *psopbe_dv* and *psogw8_dv* parameters, respectively. As expected, including the 4s, 4p, and 4d atomic orbitals gives a better description of the low-lying conduction bands in both the diamond and lonsdaleite phases. The computed band gaps are also listed in Table 1. Generally, the PSO optimizer was able to optimize the SK parameter files for both reference band structures. However, we have observed that in the *psogw8_dv* parametrization the band structure correspondence is slightly

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

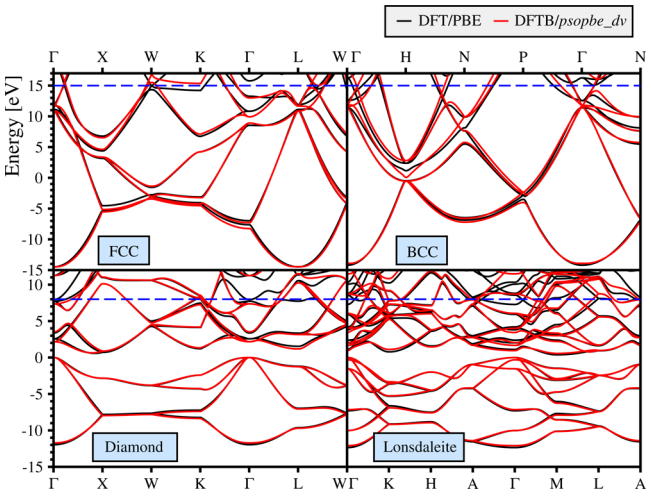


Figure 4. DFTB band structures of four phases of solid silicon computed with the *psopbe_dv* parameters closely resemble the DFT/PBE reference. The DFT bands located below the blue dashed lines were included in the definition of the fitness function.

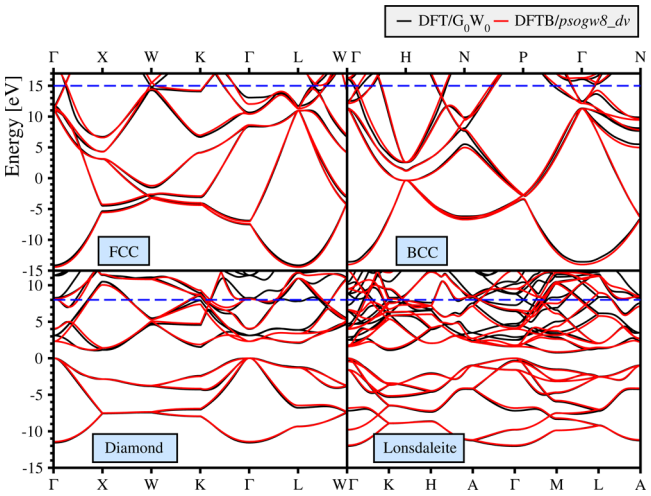


Figure 5. DFTB band structures of four phases of solid silicon computed with the *psopbe_dv* parameters closely resemble the DFT/G_0W_0 reference. The DFT bands located below the blue dashed lines were included in the definition of the fitness function.

worse than the PBE reference *psopbe_dv* parametrization. Since the PBE density functional was used to create the SK parameter files, it is possible that in order to obtain a better description of the GW band gaps the band structure shapes have been slightly spoiled. Nevertheless, the band gaps obtained from the *psogw8_dv* parametrization are still better than the DFT/PBE band gaps in these two studied nonmetallic phases.

The performance of the double valence parametrizations shows possibilities of creating a parametrization suitable for applications requiring an accurate description of the conduction bands (e.g., for TD-DFTB). We believe that the PSO optimizer will benefit those target-specific applications which require parameters that have a more accurate description of virtual orbitals or conduction bands.

5.4. Lattice Constants, Bulk Moduli, and Relative Phase Energies of Solid Silicon. The objective of this test application is to develop DFTB repulsive potentials capable of reproducing lattice parameters, bulk moduli, and correct relative energies of basic Si crystalline phases for the Si−Si pair.

Table 1. Comparison of Direct and Indirect Band Gaps (eV) in Si Diamond and Lonsdaleite Phases Computed Using Various PSO-Optimized DFTB Parameters with the DFT/G_0W_0 Reference Data[a]

	diamond		lonsdaleite	
	direct gap	indirect gap	direct gap	indirect gap
DFT/PBE	2.57	0.57	1.11	0.37
DFTB/*psopbe*	1.32	0.84	0.93	0.75
DFTB/*psopbe_dv*	2.19	0.63	0.95	0.56
DFT/G_0W_0	3.12	1.05	1.60	0.76
DFTB/*psogw5*	2.39	1.52	1.57[e]	1.47[e]
DFTB/*psogwf*	2.85	0.98	1.44[e]	0.66[e]
DFTB/*psogw8_dv*	2.66	1.03	1.34	0.77
DFTB/*pbc*	1.44	[d]	1.62	[d]
DFTB/*matsci*	2.48	2.26	2.21	[d]
experiment	3.40[b]	1.17[c]		

[a]Other theoretical and experimental values are given for reference. [b]Ref 79. [c]Ref 80. [d]Conductance band minimum is located at the Γ point. [e]Not included in the fitting set.

We consider explicitly three phases of silicon: diamond, BCC, and FCC. The reference data for all phases were obtained as follows. The lattice constants were optimized using the EXCITING code[81] (all-electron, full-potential DFT with PBEsol functional,[82] linearized augmented plane-wave (FP-LAPW) method, and 12 × 12 × 12 Monkhorst−Pack grid for diamond and 32 × 32 × 32 for BCC and FCC). Subsequently, we have compressed (5 points) and expanded (5 points) the lattice constants starting from their equilibrium values with a step of 1%, generating the DFT energy−volume ($E-V$) reference curves. These curves were subsequently fitted to the Birch−Murnaghan equation of state, yielding the values of equilibrium distance, bulk modulus, and its first derivative. The energy difference between the minima of the curves defined the relative FCC and BCC reference energies with respect to the diamond phase.

We produced a Si−Si repulsive potential for each set of electronic parameters developed in the previous sections: DFTB/*psopbe*, DFTB/*psogw5*, DFTB/*psogwf*, DFTB/*psopbe_dv*, and DFTB/*psogw8_dv*. The Si−Si repulsive potentials are represented as nine-segment fourth-order splines starting at r_{min} = 4 bohr and terminating at r_{max} = 12 bohr; the motivation for selecting this particular value of the cutoff distance is explained below. The optimal positions of the spline knots are determined by the PSO algorithm, and the spline coefficients are obtained by solving a set of linear equations as explained in ref 43. The fitness function, based on the difference between the reference DFT/PBEsol $E-V$ curve and the DFTB $E-V$ curve, is constructed as follows. Each of the $E-V$ curves is shifted such that the minimum has an energy of 0. Then, the pointwise energy difference is computed between the DFTB and the reference DFT $E-V$ curves for the three phases of silicon. The fitness function is defined as the L^1 norm of this difference vector. Again, intuitively, the fitness function corresponds to the mean absolute deviation (MAD) between the shifted reference DFT and DFTB $E-V$ curves.

The resulting $E-V$ curves for the three studied phases of crystalline silicon (BCC, FCC, and diamond) are shown in Figure 6. The $E-V$ curves obtained with the PSO-optimized parameters reproduce the DFT reference data with high fidelity. The other existing DFTB parametrizations, *pbc* and *matsci*, perform much worse. The relative phase energies, equilibrium

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

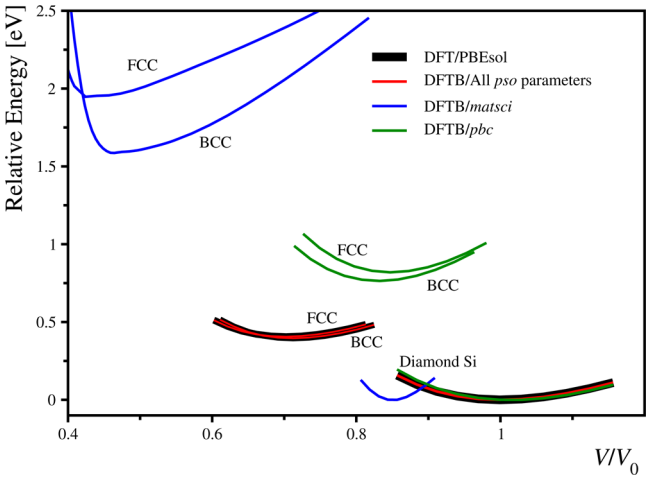


Figure 6. $E-V$ curves of three phases of silicon (FCC, BCC, and diamond) computed with DFT and DFTB (*pbc*, *matsci*, *psopbe*, *psogw5*, *psogwf*, *psopbe_dv*, and *psogw8_dv* parameters).

lattice constants, bulk moduli, and their derivatives computed using the DFTB/*psopbe*, DFTB/*psogw5*, DFTB/*psogwf*, DFTB/*psopbe_dv*, and DFTB/*psogw8_dv* Slater–Koster files are given in Table 2 together with the DFT/PBEsol reference data and analogous quantities determined with the existing DFTB parametrizations, *pbc* and *matsci*. The accuracy of the new PSO-optimized repulsive potentials is good for all five sets of electronic parameters. All of the reference data are reproduced simultaneously with high fidelity for all of the silicon phases included in the parametrization. The accuracy seems to be independent of the level of the underlying electronic parametrization. The new optimized repulsive potentials perform much better than any existing parametrizations, which, of course, is not surprising since *pbc* and *matsci* were not fitted to those properties.

The Si–Si repulsive potentials optimized for the DFTB/*psopbe*, DFTB/*psogw5*, DFTB/*psogwf*, DFTB/*psopbe_dv*, and DFTB/*psogw8_dv* electronic parameters are shown in Figure 7. These repulsive potentials are much longer than the usual choice for silicon; the cutoff distance for the Si–Si repulsive potentials in *pbc* and *matsci* are only 4.80 and 4.38 bohr, respectively. In our parametrization, we set the cutoff distance to 12 bohr in order to reproduce all of the parametrization objectives. Such a long cutoff distance signifies that not only do the first neighbor atoms in the analyzed solids have a nonzero repulsive interaction with the central atom but the next neighbors do as well. In particular, the cutoff of 12 bohr corresponds to including up to the fourth nearest neighbor interaction in diamond and up to the fifth nearest neighbor interaction in the FCC and BCC phases. The cutoffs of the repulsive potentials were initially selected to be 6.5 bohr, which effectively limited the repulsive interactions in the studied solids only to the first neighbor. With such a short cutoff, we were able to reproduce the diamond phase properties well, but no good fit for properties of the FCC and BCC phases was possible. Inspection of the $E-V$ curves clearly showed the incorrect shape of the total energy curves at large volume for these phases. Extending the repulsive potential cutoff to 8, 10, and, finally, 12 bohr systematically improved the shapes of the $E-V$ curves and the correct energy separation between them. The 12 bohr range of the repulsive potential required using nine spline segments to have enough flexibility to obtain good

Table 2. Relative Phase Energies ΔE, Equilibrium Lattice Constants a_0, Bulk Moduli B_0, and Their Derivatives B_0' of Three Phases of Solid Si Calculated from the Birch–Murnaghan Equation of State

	method	diamond	FCC	BCC
ΔE (eV)	DFT/PBEsol		0.406	0.395
	DFTB/*psopbe*		0.406	0.396
	DFTB/*psopbe_dv*		0.408	0.391
	DFTB/*psogw5*		0.406	0.367
	DFTB/*psogwf*		0.406	0.395
	DFTB/*psogw8_dv*		0.409	0.393
	DFTB/*pbc*		0.819	0.764
	DFTB/*matsci*		1.420	1.593
a_0 (Å)	DFT/PBEsol	5.433	3.833	3.057
	DFTB/*psopbe*	5.433	3.830	3.056
	DFTB/*psopbe_dv*	5.433	3.832	3.060
	DFTB/*psogw5*	5.433	3.833	3.055
	DFTB/*psogwf*	5.434	3.830	3.062
	DFTB/*psogw8_dv*	5.433	3.831	3.056
	DFTB/*pbc*	5.545	4.081	3.218
	DFTB/*matsci*	5.147	3.238	2.647
	experiment[a]	5.431		
B_0 (GPa)	DFT/PBEsol	93.2	91.8	100.9
	DFTB/*psopbe*	93.2	91.8	101.1
	DFTB/*psopbe_dv*	93.5	93.4	102.9
	DFTB/*psogw5*	93.2	92.4	101.4
	DFTB/*psogwf*	93.2	91.8	102.4
	DFTB/*psogw8_dv*	93.4	91.6	101.4
	DFTB/*pbc*	97.7	180.8	174.4
	DFTB/*matsci*	738.2	1630.0	867.7
	experiment[b]	99.2		
B_0'	DFT/PBEsol	4.23	4.07	4.34
	DFTB/*psopbe*	4.25	4.63	4.44
	DFTB/*psopbe_dv*	4.30	4.02	2.96
	DFTB/*psogw5*	4.24	3.94	4.77
	DFTB/*psogwf*	4.21	5.10	2.08
	DFTB/*psogw8_dv*	4.38	4.39	4.60
	DFTB/*pbc*	3.67	3.40	3.33
	DFTB/*matsci*	10.74	42.03	65.07
	experiment[c]	4.09		

[a]Ref 83. [b]Ref 84. [c]Ref 85.

agreement of the bulk moduli and their derivatives with the reference.

The similar accuracy of the solid properties computed with the new DFTB/*psopbe*, DFTB/*psogw5*, and DFTB/*psogwf* parameters suggests that it should be possible to obtain physically relevant repulsive potentials for a wide spectrum of electronic DFTB parameters. The presented PSO methodology seems to be particularly suited for fine-tuning and tailoring the DFTB parametrization for the desired set of objectives.

5.5. Vibrational Frequencies of Methanol. The objective of this test application is to develop a set of DFTB parameters for carbon, oxygen, and hydrogen capable of reproducing basic molecular properties (equilibrium geometry, atomization energy, set of vibrational harmonic frequencies, and dimer binding energy) of the methanol molecule. The reference atomization energy (511 kcal/mol) was taken from experimental data.[86] The binding energy for the methanol dimer was taken from the

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53–64

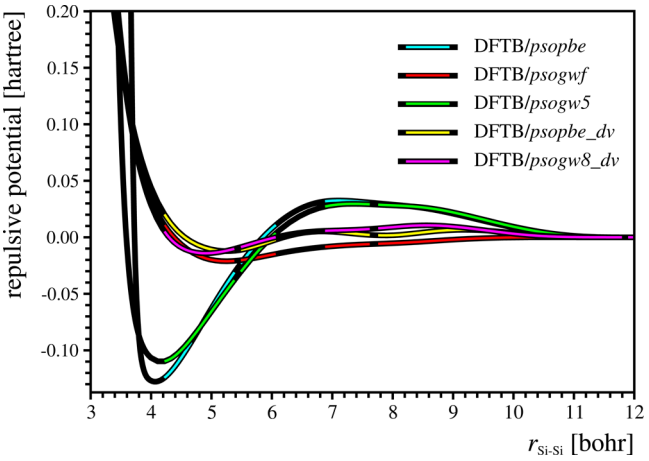


Figure 7. Si−Si repulsive potentials optimized with PSO for each DFTB electronic parametrization. Colored segments of each curve denote interatomic distances actively involved in the parametrization procedure. The black segments can, in principle, have an arbitrary form because they do not influence the shape of the $E-V$ curves of the FCC, BCC, and diamond phases.

S66 benchmark database.[87] For the equilibrium geometry reference, we use the CCSD(T)/aug-cc-pVTZ equilibrium structure of methanol. We use two reference sets of vibrational frequencies: (i) CCSD(T)/aug-cc-pVTZ harmonic vibrational frequencies and (ii) experimental fundamental frequencies.[88] The definition of the total fitness function is based on the weighted average of the atomization energy, dimer binding energy, geometrical parameters, and vibrational frequencies fitness function components. Two distinct DFTB parametrizations utilizing DFTB3 formalism are developed here, DFTB3/*psocc* and DFTB3/*psoexp*, differing by the set of reference vibrational frequencies used in the definition of the fitness function. The reference CCSD(T)/aug-cc-pVTZ equilibrium geometry is computed using the MOLPRO package,[89] and the set of CCSD(T)/aug-cc-pVTZ harmonic vibrational frequencies is computed using the CFOUR package.[90]

The geometrical parameters of methanol computed with the optimized DFTB3/*psocc* and DFTB3/*psoexp* parameters show very good agreement to the reference; for details, see Table 3. The atomization and binding energies are identical to the reference data. The resulting DFTB3/*psocc* and DFTB3/*psoexp* harmonic vibrational frequencies of methanol, shown in Table 4, reproduce the reference data accurately with a MAD of about 6 and 7 cm^{-1} with respect to the CCSD(T) and experimental reference values, respectively. The here reported DFTB3/*psocc* and DFTB3/*psoexp* parameters should provide a viable approach for modeling properties of methanol clusters,[91] methanol nanodroplets, and bulk systems and possibly for determining the anharmonic vibrational frequencies of small methanol clusters. This pilot application of the PSO optimizer demonstrates that it is possible to obtain a DFTB parametrization designed specifically to reproduce molecular properties of some specific molecule, which may, for example, prove to be useful in MD simulations of solids.

5.6. Internal Rotation Potential for Hydrogen Peroxide. One of well-known deficiencies[39] of existing DFTB parametrizations is the poor description of the internal rotation curve for H_2O_2, which results in an incorrect, planar equilibrium structure of this molecule. This failure has been recognized for other semiempirical techniques as well,[92] which led to speculation regarding whether this limitation is inherently associated with semiempirical models. Below, we explicitly demonstrate that a correct geometrical description of the H_2O_2 molecule is quite possible within the framework of the DFTB method.

The internal rotation energy curves of H_2O_2, computed with DFT (B3LYP/cc-pVTZ) and with applicable existing DFTB parameter sets (*mio*, *3ob*, *pbc*, and *matsci*), are shown in the left panel of Figure 8. The *mio*, *3ob*, and *pbc* parametrizations predict a flat equilibrium geometry of H_2O_2 in a *trans* conformation. The energy barrier to the energetically unfavorable *cis* conformation predicted with these parameters is 3.5−5.5 kcal/mol. Only the *matsci* parametrization correctly predicts that H_2O_2 is not planar with the dihedral angle underestimated by 8° and the energy barrier being too small by approximately a factor of 2. The quite satisfactory performance

Table 3. Atomization Energies, Dimer Binding Energies, and Geometrical Parameters of Methanol Computed with the Optimized DFTB/*psocc* and DFTB/*psoexp* Parameters Show Very Good Correspondence with the Reference Data

property		reference	DFTB3/*psocc*	DFTB3/*psoexp*
D_e (kcal/mol)	atomization	511.95	511.95	511.95
	binding in dimer	−5.757	−5.757	−5.757
r (Å)	C−O	1.426	1.425	1.425
	$C-H_1$[a]	1.089	1.087	1.087
	$C-H_2$[b]	1.095	1.096	1.095
	O−H	0.961	0.961	0.961
∠ (deg)	$O-C-H_1$	106.7	107.9	107.4
	$O-C-H_2$	111.8	110.9	111.2
	H_1-C-H_2	108.6	108.6	109.0
	C−O−H	108.0	106.1	106.8
dih ∠ (deg)	$H_1-C-O-H$	180.0	180.0	180.0
	$H_2-C-O-H$	61.4	60.5	60.8

[a]Anti with respect to the OH group. [b]Gauche with respect to the OH group.

Table 4. Vibrational Frequencies (cm^{-1}) of Methanol Computed with the Optimized DFTB Parameters and the Reference *Ab Initio* and Experimental Results

		CCSD(T)		experiment	
mode	symmetry	reference	DFTB3/*psocc*	reference	DFTB3/*psoexp*
torsion	A''	287	+10.4	295	+0.0
CO str	A'	1053	−1.9	1033	−0.0
CH_3 rock	A'	1082	+7.6	1060	+1.4
CH_3 rock	A''	1175	−0.0	1165	+0.4
OH bend	A'	1379	−0.6	1345	−0.9
CH_3 s-deform	A'	1484	−25.6	1455	−48.6
CH_3 d-deform	A''	1512	+3.7	1477	+0.1
CH_3 d-deform	A'	1522	−0.0	1477	+4.4
CH_3 s-str	A'	3010	−0.3	2844	+8.0
CH_3 d-str	A''	3069	−0.3	2960	−11.9
CH_3 d-str	A'	3128	+14.0	3000	−0.5
OH str	A'	3843	−0.1	3681	−1.4
MAD			5.5		6.5
max			25.6		48.6

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

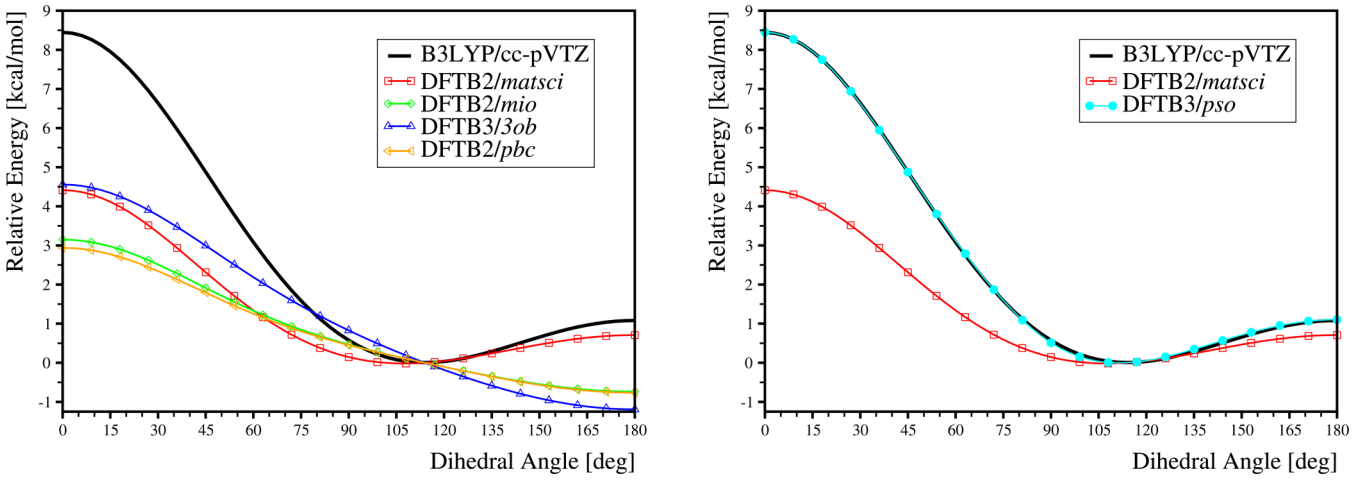


Figure 8. Left: Rigid energy scan for the internal rotation in H_2O_2 computed with DFT and existing DFTB parametrizations. Right: Comparison of an analogous curve computed with the PSO-optimized parameters (DFTB3/*pso*, cyan) with the reference DFT curve (B3LYP/cc-pVTZ, black) and the best existing DFTB curve (DFTB2/*matsci*, red).

Table 5. Deviation in Atomization Energies, Geometrical Parameters, and Vibrational Frequencies of H_2, O_2, H_2O, and H_2O_2 Computed with the Optimized DFTB3/*pso* and DFTB3/*pso2* Parameters Shows Very Good Correspondence with the Reference Data

property		B3LYP/cc-pVTZ	DFTB3/*3ob*	DFTB2/*matsci*	DFTB3/*pso*	DFTB3/*pso2*
D_e (kcal/mol)	H_2	110.26	−33.94[a]	+44.20	0.00	0.00
	O_2 (triplet)	123.06	−19.66[a,b]	+23.68	0.00	+11.67
	H_2O	227.62	+10.90[a]	+43.77	0.00	0.00
	H_2O_2	264.77	+5.94[a]	+88.07	−12.15	0.00
r (Å)	H−H in H_2	0.743	0.000	0.000	0.000	0.000
	O−O in O_2	1.206	+0.005	+0.020	0.000	0.000
	O−H in H_2O	0.962	−0.005	+0.007	0.000	0.000
	O−O in H_2O_2	1.452	+0.020	+0.001	0.000	−0.002
	O−H in H_2O_2	0.966	+0.014	+0.008	−0.005	−0.005
∠ (deg)	H−O−H in H_2O	104.5	+6.0	−3.1	0.0	0.0
	H−O−O in H_2O_2	100.4	+0.3	−0.1	+0.2	+0.2
dih ∠ (deg)	H−O−O−H	115.0	+64.0	−7.5	−0.1	−1.5
ω (cm^{-1})	H−H str. in H_2	4425	+1	+15	0	0
	O−O str. in O_2	1628	−201	−57	0	0
	H−O−H bend. in H_2O	1637	−256	+218	−3	−3
	H−O sym. str. in H_2O	3795	−144	−95	+14	+14
	H−O anti. str. in H_2O	3895	+24	−69	−1	−1
ω (cm^{-1})	H−O−O−H torsion in H_2O_2	375	−134	−96	−32	−21
	O−O str. in H_2O_2	953	−28	+108	+1	−0
	H−O a-bend. in H_2O_2	1324	−109	+190	−3	+3
	H−O s-bend. in H_2O_2	1439	−46	+128	+31	+32
	H−O anti. str. in H_2O_2	3763	+212	−131	−12	−14
	H−O sym. str. in H_2O_2	3765	+214	−133	+1	−2

[a]Note that the original paper on the *3ob* parameters uses a nonstandard way of defining the atomization energies. [b]Fitted to singlet O_2.

of *matsci* for the H_2O_2 molecule seems to be overlooked by the DFTB community, which is not surprising when taking into account the fact that the *matsci* parameters were created for quite different purposes. The performance of the *matsci* parametrization suggests that it should be possible to refine the DFTB parameters even further to describe H_2O_2 with better fidelity.

The objective of this test application is to create a set of DFTB parameters, referred to in the following as DFTB3/*pso* parameters, for H and O atoms that can correctly describe the internal rotation energy curve of H_2O_2 and at the same time reproduce basic chemical properties of H_2, O_2, and H_2O. The reference data for this parametrization was generated using

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

DFT with the B3LYP/cc-pVTZ model. The set of reference data comprises the internal rotation curve shown in Figure 8 along with atomization energies and equilibrium geometries of H_2, O_2, H_2O, and H_2O_2. The geometric information is specified in a 2-fold manner, once as the equilibrium value of each geometrical parameter (five interatomic distances, two bond angles, and one dihedral angle; for details, see Table 5) and the second time as an 11-point energy curves (rigid scan) around the equilibrium value of each parameter with a displacement of ±1.5%.

The resulting internal rotation energy curve of H_2O_2 computed with the new optimized parameters is shown in the right panel of Figure 8 together with analogous curves obtained with B3LYP/cc-pVTZ and DFTB2/*matsci*. The DFTB3/*pso* curve reproduces the DFT reference with very high fidelity, making these two curves almost indistinguishable in Figure 8. Consequently, the optimized dihedral angle differs only by 0.1° from the reference value of 115°, whereas the corresponding harmonic frequency shows a deviation of 32 cm^{-1} from the corresponding DFT value. The computed atomization energies, geometrical parameters, and vibrational frequencies of the H_2, O_2, H_2O, and H_2O_2 molecules are listed in Table 5. Virtually all of the computed molecular constants are reproduced better with the DFTB3/*pso* parameters than with the previous parametrizations. The DFTB3/*pso* and DFTB3/*pso2* parametrizations perform better than the *3ob* parameters for almost all of the molecular constants. The only problem with the *pso* and *pso2* parameters is in the simultaneous prediction of correct atomization energies for the H_2O_2 and O_2 molecules. This shortcoming originates from the problems that DFTB has in correctly describing the multireference character of the bonding situation in the $^3\Sigma_g^-$ ground state of the O_2 molecule. In the DFTB description, the electronic density of the O_2 molecule is described as a mixture of the density for the three lowest electronic states of O_2: $^3\Sigma_g^-$, $^1\Delta_g$, and $^1\Sigma_g^+$. In this sense, the *pso2* parametrization is more meaningful as it ignores the dubious atomization energy of O_2 while perfectly reproducing the remaining atomization energies. Note that the *3ob* parameters use a nonstandard way of defining the atomization energy (e.g., using fitted atom energy), which leads to substantial errors in atomization energies with the standard definition.

We feel that it is important to address here the question of why the *mio*, *3ob*, and *pbc* parametrizations failed to correctly described the nonplanar structure of H_2O_2 while the *matsci* and *pso* parametrizations were successful at this task. Note that within the DFTB model internal rotation is fully (or almost fully) controlled by the electronic part because the repulsive potentials either vanish or are negligible at the second and third neighbor distances that vary during such processes. The main difference between various DFTB parametrizations must stem, then, from different spatial extents and shapes of the valence orbitals of hydrogen and oxygen. The radial wave functions associated with the 1s orbital of H and the 2s and 2p orbitals of O are shown in Figure 9 for all of the parametrizations analyzed here. For oxygen, the confinement of the atomic tail region results in similar 2s and 2p orbitals for all parametrizations (*pso* oxygen is slightly larger). It is clear that the main difference between parametrizations comes from different size of the hydrogen atom, which is much larger (by approximately 1 bohr) in the *mio*, *3ob*, and *pbc* parametrizations. For the *matsci* parametrization, the size of the H atom is similar to that in *pso* but the shape of the tail part is quite different, with the *pso* hydrogen decaying much

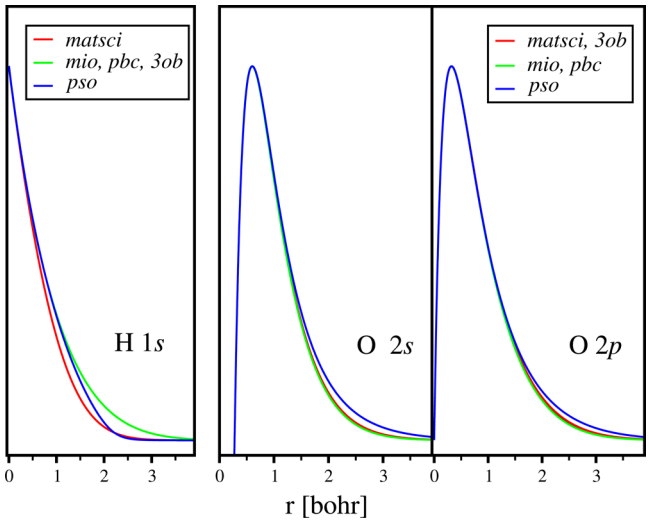


Figure 9. Radial wave function of the 1s orbital of hydrogen (left) and the 2s (middle) and 2p (right) orbitals of oxygen.

faster than the *matsci* hydrogen. The decay rate of the *matsci* hydrogen is governed by the harmonic potential (proportional to r^2), whereas for the *pso* hydrogen, the value of *a* in the Woods−Saxon potential corresponds effectively to a much higher power, close to r^{14}. This observation confirms the importance of using a high power in the confining DFTB potentials, as advocated earlier by Heine and collaborators in the context of fitting the band structures of elemental solids.[44]

6. CONCLUSIONS

We have demonstrated how to use the PSO algorithm to find optimal DFTB parameters for various molecular and solid systems in an automatized fashion. The approach discussed here was implemented and is available upon request from the authors as a stand-alone C++ code. We show that it is relatively straightforward to obtain a set of DFTB parameters using the program reported here for an arbitrary solid or molecular system. Initial pilot applications of the code to solid systems include the development of DFTB parameters for Si capable of reproducing DFT band structures of multiple crystalline silicon phases, accurate band gaps in Si diamond, and lattice parameters and bulk moduli of the FCC, BCC, and diamond silicon phases as well as a correct energy ordering of these phases. The application of the developed code to molecular systems is illustrated by developing a set of DFTB parameters for hydrogen, carbon, and oxygen that are optimized to reproduce accurate reference molecular properties of methanol: atomization energy, dimer binding energy, equilibrium geometry, and a set of vibrational frequencies. Another illustration of the capabilities of the program is the development of (O, H) parametrization that reproduces the internal rotation potential for the H_2O_2 molecule and simultaneously reproduces important molecular constants of hydrogen- and oxygen-containing molecules. In all cases reported here, DFTB with the PSO-optimized parameters is capable of reproducing the reference data with superb accuracy that is much better than the existing DFTB parametrization. We hope that the program developed and reported here becomes a standard tool for the process of developing DFTB parameters.

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

AUTHOR INFORMATION

Corresponding Author

*E-mail: hwitek@mail.nctu.edu.tw.

Present Address

⊥(Y.N.) Theoretical and Computational Chemistry Initiative, Institute for Molecular Science, Okazaki 444-8585, Japan.

Funding

The Ministry of Science and Technology of Taiwan (MOST 102-2113-M-009-015-MY3, 104-2914-I-009-035-A1) and the Ministry of Education (MOE-ATU project) are acknowledged for financial support.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1995**, *51*, 12947−12957.
(2) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185−192.
(3) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260−7268.
(4) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861−10873.
(5) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931−948.
(6) Köhler, C.; Seifert, G.; Gerstmann, U.; Elstner, M.; Overhof, H.; Frauenheim, T. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5109−5114.
(7) Witek, H. A.; Irle, S.; Morokuma, K. *J. Chem. Phys.* **2004**, *121*, 5163−5170.
(8) Witek, H. A.; Morokuma, K. *J. Comput. Chem.* **2004**, *25*, 1858−1864.
(9) Małolepsza, E.; Witek, H. A.; Morokuma, K. *Chem. Phys. Lett.* **2005**, *412*, 237−243.
(10) Witek, H. A.; Irle, S.; Zheng, G.; De Jong, W. A.; Morokuma, K. *J. Chem. Phys.* **2006**, *125*, 214706.
(11) Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Chem. Phys.* **2004**, *121*, 5171−5178.
(12) Kazachkin, D. V.; Nishimura, Y.; Witek, H. A.; Irle, S.; Borguet, E. *J. Am. Chem. Soc.* **2011**, *133*, 8191−8198.
(13) Małolepsza, E.; Witek, H. A.; Irle, S. *J. Phys. Chem. A* **2007**, *111*, 6649−6657.
(14) Witek, H. A.; Morokuma, K.; Stradomska, A. *J. Theor. Comput. Chem.* **2005**, *4*, 639−655.
(15) Li, W.; Irle, S.; Witek, H. A. *ACS Nano* **2010**, *4*, 4475−4486.
(16) Witek, H. A.; Trzaskowski, B.; Małolepsza, E.; Morokuma, K.; Adamowicz, L. *Chem. Phys. Lett.* **2007**, *446*, 87−91.
(17) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565−581.
(18) Yang, Y.; Yu, H.; Cui, Q. *J. Mol. Biol.* **2008**, *381*, 1407−1420.
(19) Klein, P.; Gottwald, B.; Frauenheim, T.; Köhler, C.; Gemmler, A. *Surf. Coat. Technol.* **2005**, *200*, 1600−1603.
(20) Choi, T. H.; Liang, R.; Maupin, C. M.; Voth, G. A. *J. Phys. Chem. B* **2013**, *117*, 5165−5179.
(21) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569−585.
(22) Frauenheim, T.; Seifert, G.; Elstner, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. *Phys. Status Solidi B* **2000**, *217*, 41−62.
(23) Woodcock, H. L.; Hodošcek, M.; Brooks, B. R. *J. Phys. Chem. A* **2007**, *111*, 5720−5728.
(24) Goyal, P.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2011**, *115*, 6790−6805.
(25) Lukose, B.; Supronowicz, B.; St. Petkov, P.; Frenzel, J.; Kuc, A. B.; Seifert, G.; Vayssilov, G. N.; Heine, T. *Phys. Status Solidi B* **2012**, *249*, 335−342.
(26) Hazebroucq, S.; Picard, G. S.; Adamo, C.; Heine, T.; Gemming, S.; Seifert, G. *J. Chem. Phys.* **2005**, *123*, 134510.
(27) Miró, P.; Cramer, C. J. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1837−1843.
(28) Slater, J. C.; Koster, G. F. *Phys. Rev.* **1954**, *94*, 1498−1524.
(29) Zheng, G.; Witek, H. A.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K.; Lundberg, M.; Elstner, M.; Köhler, C.; Frauenheim, T. *J. Chem. Theory Comput.* **2007**, *3*, 1349−1367.
(30) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067−2084.
(31) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2014**, *10*, 1518−1537.
(32) Lu, X.; Gaus, M.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2015**, *119*, 1062−1082.
(33) Kubar, T.; Bodrog, Z.; Gaus, M.; Köhler, C.; Aradi, B.; Frauenheim, T.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 2939−2949.
(34) Dolgonos, G.; Aradi, B.; Moreira, N. H.; Frauenheim, T. *J. Chem. Theory Comput.* **2010**, *6*, 266−278.
(35) Saha, S.; Pal, S.; Sarkar, P.; Rosa, A. L.; Frauenheim, T. *J. Comput. Chem.* **2012**, *33*, 1165−1178.
(36) Sarkar, S.; Pal, S.; Sarkar, P.; Rosa, A. L.; Frauenheim, T. *J. Chem. Theory Comput.* **2011**, *7*, 2262−2276.
(37) Grundkotter-Stock, B.; Bezugly, V.; Kunstmann, J.; Cuniberti, G.; Frauenheim, T.; Niehaus, T. A. *J. Chem. Theory Comput.* **2012**, *8*, 1153−1163.
(38) Hellstrom, M.; Jorner, K.; Bryngelsson, M.; Huber, S. E.; Kullgren, J.; Frauenheim, T.; Broqvist, P. *J. Phys. Chem. C* **2013**, *117*, 17004−17015.
(39) Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338−354.
(40) Niehaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct.: THEOCHEM* **2001**, *541*, 185−194.
(41) Jahangiri, S.; Dolgonos, G.; Frauenheim, T.; Peslherbe, G. H. *J. Chem. Theory Comput.* **2013**, *9*, 3321−3332.
(42) Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. *J. Chem. Theory Comput.* **2015**, *11*, 332−342.
(43) Gaus, M.; Chou, C.-P.; Witek, H.; Elstner, M. *J. Phys. Chem. A* **2009**, *113*, 11866−11881.
(44) Wahiduzzaman, M.; Oliveira, A. F.; Philipsen, P.; Zhechkov, L.; van Lenthe, E.; Witek, H. A.; Heine, T. *J. Chem. Theory Comput.* **2013**, *9*, 4006−4017.
(45) Knaup, J. M.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5637−5641.
(46) Trani, F.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*, 713−719.
(47) Bodrog, Z.; Aradi, B.; Frauenheim, T. *J. Chem. Theory Comput.* **2011**, *7*, 2654−2664.
(48) Zhu, X.; Koenig, P.; Hoffmann, M.; Yethiraj, A.; Cui, Q. *J. Comput. Chem.* **2010**, *31*, 2063−2077.
(49) Witek, H. A.; Chou, C.-P.; Mazur, G.; Nishimura, Y.; Irle, S.; Aradi, B.; Frauenheim, T.; Morokuma, K. *J. Chin. Chem. Soc.* **2015**, DOI: 10.1002/jccs.201500066.
(50) Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. *Proceedings of the Sixth International Symposium on Micro Machine and Human Science, MHS '95*; IEEE: Piscataway, NJ, 1995; pp 39−43.
(51) Sieck, A.; Frauenheim, T.; Jackson, K. A. *Phys. Status Solidi B* **2003**, *240*, 537−548.
(52) Eschrig, H.; Bergert, I. *Phys. Status Solidi B* **1978**, *90*, 621−628.
(53) Witek, H. A.; Köhler, C.; Frauenheim, T.; Morokuma, K.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5712−5719.
(54) Wang, Y.; Miao, M.; Lv, J.; Zhu, L.; Yin, K.; Liu, H.; Ma, Y. *J. Chem. Phys.* **2012**, *137*, 224108.
(55) Zendehboudi, S.; Ahmadi, M. A.; James, L.; Chatzis, I. *Energy Fuels* **2012**, *26*, 3432−3447.
(56) Tian, F.; Dong, X.; Zhao, Z.; He, J.; Wang, H.-T. *J. Phys.: Condens. Matter* **2012**, *24*, 165504.

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64

(57) Nguyen, T. T.; Li, Z.; Zhang, S.; Truong, T. K. *Expert Syst. Appl.* **2014**, *41*, 2134−2143.
(58) Wen, J.-H.; Zhong, K.-J.; Tang, L.-J.; Jiang, J.-H.; Wu, H.-L.; Shen, G.-L.; Yu, R.-Q. *Talanta* **2011**, *84*, 13−18.
(59) Liu, Y.; Zhao, L.; Li, W.; Zhao, D.; Song, M.; Yang, Y. *J. Comput. Chem.* **2013**, *34*, 67−75.
(60) Ibrahim, Z.; Khalid, N. K.; Buyamin, S.; Ibrahim, I.; Mukred, J. A. A.; Yusof, Z. M.; Mohamad, M. S.; Mokhtar, N.; Saaid, M. F. M.; Engelbrecht, A. *Int. J. Innov. Comput. I.* **2012**, *8*, 3441−3450.
(61) Shao, G.-F.; Wang, T.-N.; Liu, T.-D.; Chen, J.-R.; Zheng, J.-W.; Wen, Y.-H. *Comput. Phys. Commun.* **2015**, *186*, 11−18.
(62) Yu, X.; Li, L.; Xu, X.-W.; Tang, C.-C. *J. Phys. Chem. C* **2012**, *116*, 20075−20079.
(63) Moore, J.; Chapman, R. *Technol. Rep.: Application of particle swarm to multiobjective optimization*, 1999.
(64) Reyes-Sierra, M.; Coello, C. A. C. *Int. J. Comput. Intell. Res.* **2006**, *2*, 287−308.
(65) Leung, M.-F.; Ng, S.-C.; Cheung, C.-C.; Lui, A. K. A new strategy for finding good local guides in MOPSO. 2014 IEEE Congress on Evolutionary Computation (CEC), Beijing, July 6−11, 2014; pp 1990−1997.
(66) Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678−5684.
(67) Wang, S. Q.; Ye, H. Q. *J. Phys.: Condens. Matter* **2003**, *15*, L197.
(68) Wentorf, R. H., Jr.; Kasper, J. S. *Science* **1963**, *139*, 338−339.
(69) Kresse, G.; Hafner, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558−561.
(70) Kresse, G.; Hafner, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *49*, 14251−14269.
(71) Kresse, G.; Furthmüller, J. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(72) Kresse, G.; Furthmüller, J. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169−11186.
(73) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(74) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396−1396.
(75) Blöchl, P. E. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953−17979.
(76) Kresse, G.; Joubert, D. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758−1775.
(77) Jiang, H.; Gómez-Abal, R. I.; Li, X.-Z.; Meisenbichler, C.; Ambrosch-Draxl, C.; Scheffler, M. *Comput. Phys. Commun.* **2013**, *184*, 348−366.
(78) Mostofi, A. A.; Yates, J. R.; Lee, Y.-S.; Souza, I.; Vanderbilt, D.; Marzari, N. *Comput. Phys. Commun.* **2008**, *178*, 685−699.
(79) Zucca, R. R. L.; Shen, Y. R. *Phys. Rev. B* **1970**, *1*, 2668−2676.
(80) Kittel, C.; McEuen, P. *Introduction to Solid State Physics*, 8th ed.; Wiley: New York, 2004.
(81) Gulans, A.; Kontur, S.; Meisenbichler, C.; Nabok, D.; Pavone, P.; Rigamonti, S.; Sagmeister, S.; Werner, U.; Draxl, C. *J. Phys.: Condens. Matter* **2014**, *26*, 363202.
(82) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. *Phys. Rev. Lett.* **2008**, *100*, 136406.
(83) Hubbard, C. R.; Swanson, H. E.; Mauer, F. A. *J. Appl. Crystallogr.* **1975**, *8*, 45−48.
(84) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2004**, *69*, 075102.
(85) Beattie, A. G.; Schirber, J. E. *Phys. Rev. B* **1970**, *1*, 1548−1551.
(86) Adamo, C.; Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **2000**, *112*, 2643−2649.
(87) Rezac, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(88) Shimanouchi, T. *J. Phys. Chem. Ref. Data* **1972**, *1*, 189.
(89) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. J. *MOLPRO*, version 2012.1, a package of ab initio programs, 2012. http://www.molpro.net.
(90) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G. et al. *CFOUR*, a quantum chemical program package. http://www.cfour.de/.
(91) Nishimura, Y.; Lee, Y.-P.; Irle, S.; Witek, H. A. *J. Chem. Phys.* **2014**, *141*, 094303.
(92) Krüger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. *J. Chem. Phys.* **2005**, *122*, 114110.

DOI: 10.1021/acs.jctc.5b00673
J. Chem. Theory Comput. 2016, 12, 53−64