JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

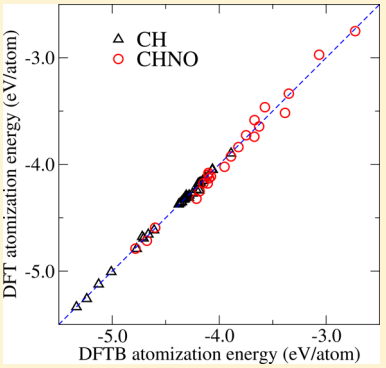
Numerical Optimization of Density Functional Tight Binding Models: Application to Molecules Containing Carbon, Hydrogen, Nitrogen, and Oxygen

A. Krishnapriyan,[†,‡] P. Yang,[‡] A. M. N. Niklasson,[‡] and M. J. Cawkwell*[,‡]

[†]Department of Materials Science and Engineering, Stanford University, Stanford, California 94305, United States

[‡]Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

ABSTRACT: New parametrizations for semiempirical density functional tight binding (DFTB) theory have been developed by the numerical optimization of adjustable parameters to minimize errors in the atomization energy and interatomic forces with respect to *ab initio* calculated data. Initial guesses for the radial dependences of the Slater-Koster bond integrals and overlap integrals were obtained from minimum basis density functional theory calculations. The radial dependences of the pair potentials and the bond and overlap integrals were represented by simple analytic functions. The adjustable parameters in these functions were optimized by simulated annealing and steepest descent algorithms to minimize the value of an objective function that quantifies the error between the DFTB model and *ab initio* calculated data. The accuracy and transferability of the resulting DFTB models for the C, H, N, and O system were assessed by comparing the predicted atomization energies and equilibrium molecular geometries of small molecules that were not included in the training data from DFTB to *ab initio* data. The DFTB models provide accurate predictions of the properties of hydrocarbons and more complex molecules containing C, H, N, and O.



1. INTRODUCTION

Self-consistent charge transfer density functional tight binding (DFTB) theory provides a fast and physically intuitive description of interatomic bonding in molecules and solids.[1−3] Bonding arising from the overlap of atom-centered orbitals is described by a Slater-Koster two-center tight binding Hamiltonian.[4,5] This traditional tight binding picture of interatomic bonding is extended in self-consistent tight binding theory by the introduction of charge transfer arising from a set of self-consistently calculated atom-centered point Mulliken charges.[1,6−8] The theoretical framework for the introduction of self-consistent charge transfer can be derived from the second-order expansion of the Hohenberg−Kohn−Sham functional in charge fluctuations about neutral, atom-centered charge densities.[7]

Density functional tight binding theory is semiempirical. Hence, while it is based rigorously on a quantum mechanical description of the underlying electronic structure of a system of atom-centered charge densities, many terms are approximated and parametrized rather than computed exactly. This step provides significant performance gains at the expense of (i) a loss of accuracy and transferability with respect to *ab initio* models and (ii) the requirement that parametrizations are developed for interactions between every pair of elements under study. The loss of accuracy with respect to *ab initio* electronic structure theory may be acceptable provided that the parametrization is of sufficient quality. We present a simple computational method for optimizing DFTB parametrizations to improve their accuracy and hence extend the number of potential applications of the method.

Unfortunately, there exists neither a direct correspondence between self-consistent DFTB theory and regular density functional theory (DFT) nor a unique, unambiguous method for the development of parametrizations. As a result, numerous schemes for the parametrization of semiempirical electronic structure models have been proposed. The most straightforward methods involve the reparametrization of the pairwise, mainly repulsive interactions in TB theory to accurate *ab initio* data, while keeping the TB Hamiltonian and overlap matrices fixed.[9−12] However, owing to the central force character of the pairwise repulsion, it is unlikely that these approaches can address any deficiencies in the angular character of interatomic bonding. Two groups have recently reported the parametrization of DFTB models in a two-step procedure whereby the electronic parameters, i.e., matrix elements, were first optimized to reproduce *ab initio* calculated data. The repulsive pairwise interactions were subsequently parametrized while keeping fixed the matrix elements defined in the first step.[13−15]

Here we present a new approach to the development of DFTB models whereby the matrix elements and repulsive potentials are optimized together, rather than sequentially, through the global optimization of a single objective function. Analogous schemes have been developed for the parametrization of empirical potentials.[16−19] The matrix elements

Received: July 25, 2017
Published: October 17, 2017

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

and pair potentials are optimized numerically such that errors in the atomization energies and interatomic forces of small molecules with randomly displaced atoms are reduced with respect to a database of high quality, *ab initio*-calculated reference data. The use of simple, analytic forms to represent the radial dependence of the Slater-Koster bond integrals and pair potentials is a crucial step since it enables a straightforward numerical exploration of a relatively small parameter space and reduces the risk of terms in the model developing unphysical features. This step would be more difficult if tabulated parameter sets were employed.

2. DENSITY FUNCTIONAL TIGHT BINDING THEORY

The derivation of DFTB theory has been described in detail in many outstanding articles, see for example, refs 2, 3, and 20−22. Here we outline only the key results. The wave functions are expanded in a linear combination of atomic-centered orbitals

$$|\Psi\rangle = \sum_{i,\alpha} c_{i,\alpha} |i\alpha\rangle \tag{1}$$

where i labels atoms, α orbitals, and $c_{i,\alpha}$ are expansion coefficients. The set of atomic orbitals is nonorthogonal with overlap matrix elements

$$S_{i\alpha,j\beta} = \langle j\beta | i\alpha \rangle \tag{2}$$

The DFTB Hamiltonian is

$$\mathbf{H} = \mathbf{H}^0 + \mathbf{H}^1 \tag{3}$$

where $\mathbf{H}^0$ is the charge-independent Slater-Koster Hamiltonian that represents the overlap of atom-centered orbitals. The elements on the main diagonal of the Slater-Koster Hamiltonian, $H^0_{i\alpha,i\alpha}$, or on-site energies, ε, are equal to the energy of the corresponding orbitals on noninteracting, neutral atoms. $\mathbf{H}^1$ is an electrostatic potential arising from a set of atom-centered charges, $\{q\}$.

$$H^1_{i\alpha,j\beta} = \frac{1}{2} S_{i\alpha,j\beta} (V_i + V_j) \tag{4}$$

V_i is the electrostatic potential plus the Hubbard U term that quantifies the change in energy upon charge transfer, that is

$$V_i = U_{(i)} q_i + \sum_{k \neq i=1}^{N} \gamma_{ik} q_k \tag{5}$$

where q_i is the Mulliken partial charge of atom i, $U_{(i)}$ the value of the Hubbard U corresponding to the chemical species of atom i, N the total number of atoms, and γ_{ik} a screened Coulomb potential. We use the same form for the screened Coulomb potential, γ_{ik}, as Elstner et al.[1] The Mulliken partial charges

$$q_i = \frac{1}{2} \sum_{\alpha \in i} (P_{j\beta,i\alpha} S_{i\alpha,j\beta} + P_{i\alpha,j\beta} S_{j\beta,i\alpha}) - n_i^e \tag{6}$$

where n_i^e is the number of valence electrons on neutral atom, i, depend on the density matrix, $\mathbf{P}$. The set of Mulliken charges in DFTB theory must be solved self-consistently because the DFTB Hamiltonian, eq 3, depends explicitly on charge transfer.

The potential energy in DFTB theory is

$$u = 2\mathrm{Tr}[(\mathbf{P} - \mathbf{P}_0)\mathbf{H}^0] + \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \gamma_{ij} q_i q_j + E_{\mathrm{pair}} \tag{7}$$

where $\mathrm{Tr}[\mathbf{X}]$ denotes the trace of matrix $\mathbf{X}$, and $\mathbf{P}_0$ is the density matrix for neutral, noninteracting atoms. E_{pair} is a sum of atom-centered pair potentials, Φ, that mainly provide short-range repulsion

$$E_{\mathrm{pair}} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \Phi(R_{ij}) \tag{8}$$

where R_{ij} is the distance between atoms i and j. The interatomic forces, $\mathbf{f}_i = -\partial u / \partial \mathbf{R}_i$, where $\mathbf{R}_i$ is the position of atom i, are computed from eq 7 once self-consistency in the partial charges is achieved.

2.1. Representation of Radial Dependences. The off-diagonal elements of the Slater-Koster Hamiltonian, $\mathbf{H}^0$, and overlap, $\mathbf{S}$, matrices are angularly dependent combinations of special bond and overlap integrals, respectively, that depend only on interatomic distance. The angular dependences of the matrix elements depend on the angular character of the orbitals forming the bond and are prescribed.[4] Hence, we are required only to parametrize the radial dependences of the bond and overlap integrals, $h_{ll'\tau}$ and $s_{ll'\tau}$, respectively, where l and l' label the angular momenta of the overlapping orbitals ($l = s, p, d, \ldots$), and τ labels the symmetry of bond ($\tau = \sigma, \pi, \delta, \ldots$).

Parameterizations for DFTB have traditionally been distributed in tabular form. We instead represent the bond integrals, overlap integrals, and pair potentials with simple analytic functions, which greatly simplifies their numerical optimization. The radial dependences of the bond and overlap matrix elements are

$$h_{ll',\tau}(R) = h_{ll',\tau}(R_0) \zeta^h_{ll',\tau}(R) \tag{9}$$

and

$$s_{ll',\tau}(R) = s_{ll',\tau}(R_0) \zeta^s_{ll',\tau}(R) \tag{10}$$

where R is the interatomic distance, R_0 is a reference interatomic distance, and

$$\zeta(R) = \prod_{i=1}^{p} \exp(A_i (R - R_0)^i) \tag{11}$$

where A_i are adjustable parameters.[23] We limit the product in eq 11 to $p = 2$ for the bond integrals and to $p = 4$ for the overlap integrals. Alternatives to the scaling functions we propose include those of Goodwin, Skinner, and Pettifor.[24] The pair potentials between each pair of elements are described using a similar functional form whereby

$$\Phi(R) = \Phi_0 \zeta(R) \tag{12}$$

where $R_0 = 0$ in the scaling function, ζ, and we extend the product to $p = 4$. The scaling function, eq 11, is replaced by a cut-off tail

$$t(R) = B_0 + (R - R_1)(B_1 + (R - R_1)(B_2 + (R - R_1) \times (B_3 + (R - R_1)(B_4 + (R - R_1)B_5)))) \tag{13}$$

at a specified distance $R = R_1$ that ensures that the functions decay smoothly to zero at $R = R_{\mathrm{cut}}$.

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

3. NUMERICAL OPTIMIZATION OF ADJUSTABLE PARAMETERS

The radial dependences of the overlap integrals and the on-site energies, ε_l, were held constant during the numerical optimization procedure. We optimized only the Hubbard U parameters for each element and the radial dependences of the bond integrals and pair potentials. The overlap integrals were derived from atom-centered numerical orbitals generated by minimum basis density functional theory calculations with the PLATO code.[25−27] We found that our numerical optimization procedure rapidly encounters severe numerical problems, for instance during Löwdin orthogonalization,[28] if the radial dependences of the overlap integrals are allowed to evolve during the parametrization process.

3.1. Generation of Initial Parameter Set. The on-site energies, ε_l, which parametrize the main diagonal of the Slater-Koster Hamiltonian, were obtained from plane wave density functional theory calculations on atoms in a large supercell. We used a plane wave cut-off energy of 500 eV, the exchange correlation functional of Perdew, Burke, and Ernzerhof (PBE),[29] and projector augmented wave (PAW) pseudopotentials[30,31] within the VASP code. A small, finite electronic temperature was used to achieve fractional occupation of the states. The energies of the valence orbitals are provided in Table 1. Initial guesses for the Hubbard U for each element were derived from the experimental values of its electron affinity, A, and ionization potential, I, via $U = I - A$.

Table 1. On-Site Energies, ε_l for C, H, N, and O

	ε_s (eV)	ε_p (eV)
C	−13.7199	−5.2541
H	−6.4835	
N	−18.5565	−7.0625
O	−23.9377	−9.0035

Overlap integrals and initial guesses for the radial dependences of the bond integrals were obtained from minimum basis density functional theory calculations using the PLATO package.[27] We generated short-ranged, atom-centered minimal sets of numerical orbitals for pseudoatoms by solving the modified Kohn−Sham equation

$$(T + V_{\text{eff}} + V_{\text{conf}})|\alpha\rangle = \varepsilon|\alpha\rangle \tag{14}$$

where T is the kinetic energy operator, V_{eff} the effective Kohn−Sham potential, and V_{conf} an additional confining potential that reduces the spatial range of the orbitals. A harmonic potential of the form

$$V_{\text{conf}}(R) = \left(\frac{R}{aR_{\text{cov}}}\right)^2 \tag{15}$$

where R_{cov} is the covalent radius of the atom and $1.8 < a < 2$ has been traditionally used to generate DFTB models.[2] Several alternative forms for the confining potential have been proposed.[25,32,33] We have opted to use, without modification, the confining potentials tabulated by Wahiduzzadin et al.[13] that are of the form

$$V_{\text{conf}}(R) = \left(\frac{R}{r_0}\right)^{\sigma} \tag{16}$$

The parameters for r_0 and σ for carbon, hydrogen, nitrogen, and oxygen were taken from ref 13.

The pseudoatoms were generated using the pseudopotentials of Goedecker, Teter, and Hutter,[34] and the PBE exchange-correlation functional. The radial dependences of the overlap and bond integrals were obtained from density functional theory calculations using the numerical orbitals on dimers as a function of bond distance. We parametrized the radial scaling functions, eqs 9, 10, and 11, to these data by least-squares fits. The parametrizations of the radial dependences of the overlap matrix elements for the C, H, N, and O system are provided in Table 2. The cut-off tail, eq 13, was applied between $R_1 = 3.5$ and $R_{\text{cut}} = 4.0$ Å for every overlap integral.

It is crucial that the analytic functions used to represent the radial dependences of the overlap integrals, eqs 10 and 11, reproduce the values obtained from minimal basis density functional theory with high accuracy. Any discrepancies between the parametrized analytic form and the exact overlap integrals from density functional theory might give rise to numerical problems during orthogonalization. We have tested the parametrized overlap integrals presented in Table 2 extensively in condensed-phase molecular dynamics simulations of organic materials without encountering any problems during orthogonalization. Hence, we are confident that the empirical scaling functions that we have chosen to represent the overlap integrals are adequate for their foreseeable applications.

Initial guesses for the radial dependences of the bond integrals were computed in the same way as the overlap integrals. The pairwise term in tight binding theory is typically used to capture all interactions that are not included explicitly in the other terms in the total energy. It can be parametrized to account for any differences between the band and Coulomb energies and a reference, high quality density functional theory calculation, E_{ref}, that is

$$E_{\text{pair}} = E_{\text{ref}} - \left(2\text{Tr}[(\mathbf{P} - \mathbf{P}_0)\mathbf{H}^0] + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i=1}^{N}\gamma_{ij}q_iq_j - \sum_{i=1}^{N}E_i^0\right) \tag{17}$$

where E_i^0 are the energies of spin-polarized, noninteracting neutral atoms in DFTB theory. We generated reference data, E_{ref}, specifically for the initial parametrization of the pair potentials using gas-phase density functional theory calculations of bond stretches in simple molecules. These calculations were performed with the cc-pVTZ basis set and B3LYP hybrid exchange correlation functional as implemented in the NWChem package.[35] Eq 17 was used to obtain the pairwise energy as a function of bond length, R, and we made a least-squares fit of eqs 12 and 11 to these data.

The reference density functional theory data, E_{ref}, in eq 17 is in terms of atomization energy, that is, the difference between the total energy of the molecule and sum of the energies of its spin polarized, isolated atoms. These calculations are straightforward in quantum chemical codes. In order to make a meaningful comparison between the reference atomization energies and those obtained from DFTB, we implemented the DFTB Mulliken spin density formalism presented in ref 2. The spin densities, m, are equal to the difference between the number of electrons in the spin up and spin down populations and are computed self-consistently from the density matrices for both spin populations. The self-consistent computation of spins differs somewhat from quantum chemical codes where

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

Table 2. Parametrization of the Radial Dependences of the Overlap Matrix Elements

	$ll'\tau$	$s(R_0)$	A_1 (Å^{-1})	A_2 (Å^{-2})	A_3 (Å^{-3})	A_4 (Å^{-4})	R_0 (Å)
N O	$ss\sigma$	0.340064	−1.703613	−0.622348	0.036738	−0.040158	1.2
N O	$sp\sigma$	−0.370946	−1.040947	−0.931097	0.252441	−0.115450	1.2
O N	$sp\sigma$	−0.420014	−1.107918	−0.905594	0.188424	−0.088365	1.2
N O	$pp\sigma$	−0.314073	0.499050	−2.914288	2.067657	−0.738439	1.2
N O	$pp\pi$	0.223937	−1.991867	−0.537630	−0.081270	−0.004130	1.2
C N	$ss\sigma$	0.263438	−1.754525	−0.584215	−0.007801	−0.021729	1.5
C N	$sp\sigma$	−0.326609	−1.197485	−0.807786	0.134891	−0.084373	1.5
N C	$sp\sigma$	−0.337943	−1.335442	−0.769693	0.119373	−0.079493	1.5
C N	$pp\sigma$	−0.350240	−0.467439	−1.849316	1.854403	−0.988471	1.5
C N	$pp\pi$	0.158424	−2.114409	−0.582346	−0.051076	−0.006183	1.5
C O	$ss\sigma$	0.375339	−1.547372	−0.642492	0.020614	−0.026699	1.2
C O	$sp\sigma$	−0.373027	−0.776043	−1.019920	0.257539	−0.102838	1.2
O C	$sp\sigma$	−0.458068	−1.035067	−0.937868	0.190562	−0.077841	1.2
C O	$pp\sigma$	−0.322293	0.795473	−3.476601	2.589965	−0.897800	1.2
C O	$pp\pi$	0.244570	−1.922717	−0.573671	−0.057280	−0.004108	1.2
N N	$ss\sigma$	0.231654	−1.879002	−0.572765	−0.004579	−0.031106	1.5
N N	$sp\sigma$	−0.305271	−1.385158	−0.751032	0.114531	−0.090839	1.5
N N	$pp\sigma$	−0.324668	−0.547805	−1.638658	1.495168	−0.827868	1.5
N N	$pp\pi$	0.142909	−2.162036	−0.571942	−0.071640	−0.004682	1.5
O O	$ss\sigma$	0.296445	−1.911896	−0.663451	0.038054	−0.046608	1.2
O O	$sp\sigma$	−0.362143	−1.285274	−0.939591	0.204641	−0.106438	1.2
O O	$pp\sigma$	−0.312044	0.121814	−2.519352	1.681266	−0.644566	1.2
O O	$pp\pi$	0.193010	−2.168462	−0.580629	−0.105104	0.004891	1.2
H O	$ss\sigma$	0.404725	−1.702546	−0.707938	0.074904	−0.039922	1.0
H O	$sp\sigma$	−0.447660	−0.952979	−1.163537	0.400616	−0.156965	1.0
H N	$ss\sigma$	0.446693	−1.500463	−0.657448	0.065741	−0.037004	1.0
H N	$sp\sigma$	−0.501530	−0.785734	−1.123232	0.394878	−0.148501	1.0
C C	$ss\sigma$	0.346977	−1.519820	−0.570812	−0.013518	−0.015829	1.4
C C	$sp\sigma$	−0.400467	−0.984048	−0.853949	0.157178	−0.073381	1.4
C C	$pp\sigma$	−0.382417	0.102889	−2.786680	2.646356	−1.134320	1.4
C C	$pp\pi$	0.214357	−1.948923	−0.578323	−0.034356	−0.007257	1.4
H C	$ss\sigma$	0.416003	−1.459596	−0.654874	0.009140	−0.012658	1.1
H C	$sp\sigma$	−0.495695	−0.901626	−1.007214	0.189808	−0.057087	1.1
H H	$ss\sigma$	0.575007	−1.391261	−0.778831	0.080209	−0.017759	0.75

the spin multiplicity is specified and remains constant during a self-consistent field procedure.

The change in energy upon the spin polarization of an atom in DFTB is

$$\Delta E = \frac{1}{2} W_l m_l^2 \tag{18}$$

where m_l is the number of unpaired electrons in orbitals with angular momentum l, and W_l is an adjustable parameter. We computed values for W_l for C, H, N, and O by evaluating the change in energy from nonspin polarized isolated atoms with fractional occupation to fully spin polarized isolated atom using the VASP code. The corresponding values are provided in Table 3. The free atom energies, E^0, used to compute atomization energies in DFTB, can be computed from eq 18 directly with Hund's rules.

Table 3. Parameters for the Mulliken Spin Density Model in DFTB Theory

	W_s (eV)	W_p (eV)
C		−0.61810
H	−2.2340	
N		−0.69342
O		−0.75765

3.2. Generation of Training and Validation Data.

Reference data for the atomization energy and interatomic forces in small, distorted organic molecules were obtained from gas-phase density functional theory calculations with the cc-pVTZ basis set and B3LYP exchange-correlation functional as implemented in the NWChem package. The distorted molecules were generated by adding to the equilibrium position of every atom, $\mathbf{x}_i$, in each molecule a small displacement, $\mathbf{D}_i$

$$\mathbf{x}_i' = \mathbf{x}_i + \mathbf{D}_i \tag{19}$$

where

$$\mathbf{D}_i = \eta_i d_{\max} \mathbf{n}_i \tag{20}$$

and $\mathbf{n}$ is a vector oriented randomly on the unit sphere, η a random number in the interval $[0, 1]$, and $d_{\max}$ a maximum displacement. We used $d_{\max}$ = 0.2 Å in this study.

We developed a database of 24 hydrocarbons for fitting the interactions between C−C and C−H. Smaller data sets containing 5−6 small molecules were constructed to parametrize the remaining interactions in the C, H, N, and O system. In total, 51 molecules were used to generate the parameter set for the CHNO system. The training data set included 20 replicas of each of the 51 molecules that were distorted using eq 20. The molecules used to generate the training data are listed in Tables 4 and 5. We did not

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

numerically optimize the parameters for H−H interactions (the $ss\sigma$ bond integral and pair potential) since the H_2 molecule was used to build the initial guess.

Table 4. Molecules Used To Parametrize C−C and C−H Interactions

formula	name
CH_4	methane
C_2H_6	ethane
C_2H_4	ethene
C_2H_2	ethyne
C_3H_4	propyne
C_3H_4	cyclopropene
C_3H_6	propene
C_3H_6	cyclopropane
C_3H_8	propane
C_4H_6	1-butyne
C_4H_6	2-butyne
C_4H_8	2-butene-*Z*
C_4H_8	2-butene-*E*
C_4H_8	1-propane 2-methyl
C_4H_{10}	butane
C_5H_8	spiropentane
C_5H_{10}	cyclopentane
C_6H_6	benzene
C_7H_8	toluene
C_8H_{10}	ethylbenzene
C_8H_{10}	*m*-xylene
$C_{10}H_8$	naphthalene
$C_{14}H_{10}$	anthracene
$C_{14}H_{10}$	phenanthrene

3.3. Numerical Optimization. The Hubbard *U*s and the adjustable parameters in the functions that describe the bond integrals and pair potentials were updated iteratively to minimize the discrepancies between the interatomic forces and atomization energies computed from our DFTB model and the training set of *ab initio* data. The errors in the forces and atomization energies were quantified by the objective functions

$$\chi_E^2 = \frac{1}{N_{\mathrm{mol}}N_{\mathrm{dist}}}\sum_{i=1}^{N_{\mathrm{mol}}}\sum_{j=1}^{N_{\mathrm{dist}}}\frac{(E_j - E_j^{\mathrm{ref}})^2}{\sigma(E)_i^2} \tag{21}$$

and

$$\chi_f^2 = \frac{1}{N_{\mathrm{mol}}}\sum_{i=1}^{N_{\mathrm{mol}}}\frac{\Delta\mathbf{f}_i}{\sigma(f)_i^2} \tag{22}$$

for the atomization energy and forces, respectively. Here, N_{mol} is the number of unique molecules in the training data, N_{dist} is the number of distortions per molecule, E_j and E_j^{ref} are the atomization energy of molecule j computed from DFTB and the reference data, respectively, and

$$\Delta\mathbf{f}_i = \frac{1}{N_a N_{\mathrm{dist}}}\sum_{j=1}^{N_{\mathrm{dist}}}\sum_{k=1}^{N_a}(\mathbf{f}_k - \mathbf{f}_k^{\mathrm{ref}})\cdot(\mathbf{f}_k - \mathbf{f}_k^{\mathrm{ref}}) \tag{23}$$

where N_a is the number of atoms in molecule i, and $\mathbf{f}_k$ and $\mathbf{f}_k^{\mathrm{ref}}$ are the forces acting on atom k from DFTB and the reference data, respectively. $\sigma(E)_i$ and $\sigma(f)_i$ are the standard deviations of the reference atomization energy and forces, respectively, which are computed over all N_{dist} distortions of molecule i. We

Table 5. Additional Molecules Used To Parametrize Interactions in the C, H, N, and O System

formula	name
N_2	nitrogen dimer
NH_3	ammonia
N_2H_2	*trans*-diazine
N_2O	nitrous oxide
HONO	nitrous acid
N_2H_4	hydrazine
H_2N_2O	nitrosamide
CO	carbon monoxide
CO_2	carbon dioxide
CH_2O	formaldehyde
CH_3OH	methanol
CH_3OCH_3	dimethyl ether
HCOH	hydroxycarbene
O_3	ozone
H_2O	water
H_2O_2	hydrogen peroxide
NH_2OH	hydroxylamine
CH_3NO_3	methylnitrate
HNO_3	nitric acid
HCN	hydrogen cyanide
CH_3NH_2	methylamine
$(CH_3)_2NH$	dimethylamine
CH_3NO_2	nitromethane
HCNO	isocyanic acid
NCOH	cyanic acid
ONCH	fulminic acid
$C_2H_6N_2O_2$	dimethylnitroamine

normalize the errors in the atomization energies and forces by the variances of those properties so that both objective functions are dimensionless. However, these forms of the objective functions are not valid if the reference data do not take the form we have adopted of multiple distortions of a number of unique molecules. These objective functions could also be ill-behaved if the variance of certain properties is very small. There exist numerous other similarity indices which could be substituted for those that we have adopted if required by the training data.

The dimensionless objective functions can be added to form a single objective function, χ^2, that we minimize, that is

$$\chi^2 = \chi_E^2 + \chi_f^2 \tag{24}$$

One of the advantages of this approach is that the objective function can be easily extended to include other properties of interest such as the HOMO−LUMO gap or molecular dipole moment.

We used simulated annealing and steepest descent algorithms to optimize the parametrizations. Simulated annealing has been used to parametrize interatomic potentials by a number of authors.[36−38] It is a robust algorithm that can deal with shallow local minima through the use of an appropriate effective temperature.[39] Numerous alternatives to simulated annealing exist which may exhibit greater efficiency or the ability to deal with poor initial guesses. Nevertheless, in this work we found simulated annealing to be sufficient. We typically started the optimization with a relatively high effective temperature that we reduced incrementally during the optimization. Most of the decrease in the objective function

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

was obtained over the first few hundred iterations after which its convergence slows.

We initially optimized the parameters bond integrals and pair potentials between each pair of elements using simulated annealing to subsets of the training data set while keeping the Hubbard *U*s constant. For instance, we parametrized the interactions between C−C and C−H to the hydrocarbon data set. Following the optimization of the interactions between each pair of elements, we performed another set of optimizations using the whole 51 molecule data set and allowed all of the adjustable parameters to evolve, including the Hubbard *U*s, for C, H, N, and O. A final steepest descent optimization was performed on all of the parameters following rounds of simulated annealing optimization at decreasing effective temperatures. The resulting parametrizations of the bond integrals, pair potentials, and Hubbard *U*s are provided in Tables 6, 7, and 8, respectively.

Table 6. Optimized Parametrization of the Radial Dependences of the Bond Integrals Elements

	$ll'\tau$	$h(R_0)$	A_1 (Å^{-1})	A_2 (Å^{-2})	R_0 (Å)
N O	$ss\sigma$	−9.360078	−1.293118	−0.379415	1.2
N O	$sp\sigma$	10.309052	−0.981652	−0.828497	1.2
O N	$sp\sigma$	10.723048	−0.454312	−0.916563	1.2
N O	$pp\sigma$	9.259131	−0.734112	−1.023762	1.2
N O	$pp\pi$	−4.532623	−1.999631	−0.286275	1.2
C N	$ss\sigma$	−7.409712	−1.940942	−0.219762	1.5
C N	$sp\sigma$	7.501761	−1.211169	−0.373905	1.5
N C	$sp\sigma$	8.697591	−1.267240	−0.178484	1.5
C N	$pp\sigma$	6.954600	−1.188456	−0.808043	1.5
C N	$pp\pi$	−2.921605	−2.203548	−0.409424	1.5
C O	$ss\sigma$	−13.986685	−1.931973	−0.432011	1.2
C O	$sp\sigma$	10.718738	−1.389459	−0.182128	1.2
O C	$sp\sigma$	14.194791	−1.371650	−0.248285	1.2
C O	$pp\sigma$	8.622023	−0.557144	−0.938551	1.2
C O	$pp\pi$	−5.327397	−2.190160	−0.089303	1.2
N N	$ss\sigma$	−7.165811	−2.348869	−0.541905	1.5
N N	$sp\sigma$	8.212268	−1.499123	−0.526440	1.5
N N	$pp\sigma$	7.102331	−1.252366	−0.552533	1.5
N N	$pp\pi$	−2.828938	−2.376886	−0.560898	1.5
O O	$ss\sigma$	−14.387756	−2.244278	−1.645605	1.2
O O	$sp\sigma$	13.699127	−1.602358	−0.114474	1.2
O O	$pp\sigma$	9.235469	−1.131474	−0.924535	1.2
O O	$pp\pi$	−4.526526	−2.487174	−0.201464	1.2
H O	$ss\sigma$	−12.189103	−1.800097	−0.325933	1.0
H O	$sp\sigma$	9.518733	−1.333235	−0.393710	1.0
H N	$ss\sigma$	−12.631030	−1.585597	−0.250969	1.0
H N	$sp\sigma$	9.837852	−1.234850	−0.324283	1.0
C C	$ss\sigma$	−9.197237	−1.607050	−0.535057	1.4
C C	$sp\sigma$	8.562436	−0.980182	−0.646929	1.4
C C	$pp\sigma$	6.614756	−0.528591	−0.951460	1.4
C C	$pp\pi$	−3.678302	−1.881668	−0.255951	1.4
H C	$ss\sigma$	−9.235812	−1.372683	−0.408433	1.1
H C	$sp\sigma$	8.104851	−0.936099	−0.626219	1.1
H H	$ss\sigma$	- 9.400000	−1.145903	−0.391777	0.75

The numerical optimization of the bond integrals, pair potentials, and Hubbard *U*s was performed without constraints on any parameter. In principle, these terms in the model could change significantly from their initial values. However, since the starting guesses for bond integrals were obtained from minimum basis density functional theory and the pair potentials from high quality density functional theory calculations, these parameters change relatively little during the numerical optimization procedure. Indeed, since DFTB theory is strongly connected to Kohn−Sham density functional theory, it would be both surprising and concerning if the optimized parameters differed significantly from their initial values. We present in Figure 1 the radial dependences of the C−C and C−H bond integrals to illustrate the effects of numerical optimization on their properties. The optimized bond integrals increase in magnitude by 10−20% at nearest neighbor distances and decay slightly faster with distance with respect to those obtained from minimal basis density functional theory. The relative values of the bond integrals are essentially unchanged after optimization, and, owing to the simple functional forms chosen to represent their radial dependences, the bond integrals do not evolve unphysical bumps. Hence, the adjustable parameters in the model retain their physical meaning. Moreover, recent work has clearly demonstrated that the confining potentials used to generate bond and overlap integrals for traditional DFTB models can be varied systematically in order to improve the accuracy of the resulting model. Our approach falls within the scope of these schemes.

4. EVALUATION OF NUMERICALLY OPTIMIZED DFTB PARAMETRIZATIONS

We have assessed the accuracy of our DFTB parametrizations by evaluating the errors in the atomization energies and bond lengths for a set of molecules that had not be used to train the model. The test molecules, 39 hydrocarbons and 23 CHNO-based, are listed in Tables 9 and 10. We have also evaluated the transferability of the parametrizations by comparing the mean errors in the atomization energies and bond lengths for the molecules used to train the model to those computed for the test molecules. If the errors computed for the training and test molecules are both small and of similar size, we can be confident in the transferability of the parametrization, that is, its ability to predict the properties of new molecules. The total energies and equilibrium geometries of the training and test molecules at the B3LYP/cc-pVTZ level of theory were either obtained from the Computational Chemistry and Comparison Benchmark Database[40] or computed using the NWChem package. The *ab initio* total energies were converted to atomization energies so that a meaningful comparison with the results of the DFTB calculations could be made. The atomization energies and equilibrium geometries from our DFTB parametrization were obtained following geometry optimization until the maximum force acting on any atom was less than 0.001 eV/Å.

The atomization energies normalized by the number of atoms in each molecule from density functional theory and DFTB for the hydrocarbon and CHNO-based molecules are presented in Figures 2 and 3, respectively. These figures show clearly the outstanding accord between the atomization energies predicted by DFTB and those obtained from density functional theory over a diverse set of chemical moieties.

The equilibrium covalent bond distances in molecules are controlled by the interatomic forces. Hence, by comparing the optimized, gas-phase structures of molecules obtained from density functional theory and DFTB we can readily assess the accuracy of the interatomic forces predicted by the latter. We plot in Figures 4 and 5 direct comparisons of the equilibrium bond distances in the 39 molecule hydrocarbon and 23 CHNO-based molecule test sets, respectively. Figure 4 clearly

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

Table 7. Optimized Parametrization of the Radial Dependence of Pair Potentials between All Pairs of Elements in the C, H, N, and O System[a]

	Φ_0 (eV)	A_1 (Å^{-1})	A_2 (Å^{-2})	A_3 (Å^{-3})	A_4 (Å^{-4})	R_1 (Å)	R_{cut} (Å)
N O	14.005908	19.769009	−46.607006	38.399015	−12.656658	1.6	1.7
C N	98.283078	10.289077	−27.709052	22.099235	−6.796462	1.6	1.7
N N	40.335850	14.958977	−36.644093	29.219613	−8.918783	1.6	1.7
C O	0.916287	30.115416	−59.612502	45.114207	−13.200384	1.5	1.6
N H	0.664002	28.086622	−63.415978	53.301425	−17.343446	1.3	1.4
O O	11.833452	19.281518	−45.763767	37.924165	−12.006535	1.5	1.6
O H	0.484351	33.176296	−81.154354	74.931992	−26.796460	1.2	1.3
C H	1.094168	28.606497	−71.558353	65.967464	−23.372892	1.2	1.3
C C	3.927770	24.439989	−51.156433	39.032536	−11.321277	1.6	1.7
H H	8.194700	16.371100	−75.246500	106.703000	−59.105700	0.8	0.9

[a]Cut-off tails are added between $R = R_1$ and R_{cut}.

Table 8. Optimized Values of the Hubbard *U*s for Carbon, Hydrogen, Nitrogen, and Oxygen

	U (eV)
C	14.240811
H	12.054683
N	17.372900
O	11.876141

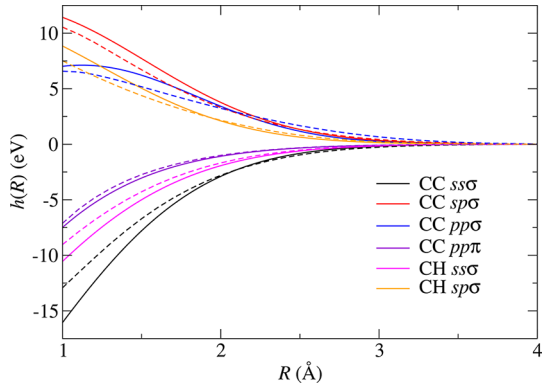


Figure 1. Radial dependences of the bond integrals between C−C and C−H from minimal basis density functional theory (broken lines) and after numerical optimization (solid lines).

depicts the high accuracy with which the optimized DFTB model can predict the gas-phase structure of hydrocarbons. The errors in predicted bond distances for the CHNO-based data set are larger than those seen for the hydrocarbons. Nevertheless, Figure 5 shows that beyond a small number of outliers, the optimized model still predicts accurately the gas-phase structures of the molecules in the test set.

We present in Table 11 the root-mean-square errors in the atomization energies and bond lengths obtained from our DFTB model with respect to density functional theory at the B3LYP/cc-pVTZ level for both the training and test sets of molecules. The results for hydrocarbons highlight the outstanding transferability of this parametrization since the errors in the atomization energies and bond lengths for the training and test molecules are almost identical. Hence, our parametrization for hydrocarbons is able to predict the energy and geometry of new molecules with, on average, the same accuracy as those to which it was parametrized. The results for the CHNO-based molecules indicate that this parametrization is neither as accurate nor as transferable as the parametrization for hydrocarbons. The errors in the atomization energy and bond lengths for the training molecules are larger than those of the hydrocarbons. Furthermore, the errors for the test molecules

Table 9. Molecules Used To Test the Accuracy and Transferability of the DFTB Parametrization for Hydrocarbons

formula	name
C_2H_2	vinylidene
C_3H_4	allene
C_4H_6	bicyclo[1.1.0]butane
C_4H_6	1,2-butadiene
C_4H_6	cyclobutene
C_4H_6	1-methylcyclopropene
C_4H_6	methylenecyclopropane
C_4H_8	cyclobutane
C_4H_8	methylcyclopropane
C_5H_8	1,3-butadiene, 2-methyl-
C_5H_8	cyclopentene
C_5H_8	bicyclo[2.1.0]pentane
C_5H_8	1,4-pentadiene
C_5H_8	1,2-pentadiene
C_5H_8	2,3-pentadiene
C_5H_8	1,2-butadiene, 3-methyl-
C_5H_8	1-pentyne
C_5H_8	ethenylcyclopropane
C_5H_8	cyclobutane, methylene-
C_5H_8	1,3-pentadiene, (*Z*)-
C_5H_8	1,3-pentadiene, (*E*)-
C_5H_{10}	1-pentene
C_5H_{10}	2-butene, 2-methyl-
C_5H_{10}	1-butene, 2-methyl-
C_5H_{10}	2-pentene, (*Z*)-
C_5H_{10}	2-pentene, (*E*)-
C_5H_{10}	cyclopropane, 1,1-dimethyl-
C_5H_{12}	butane, 2-methyl-
C_5H_{12}	pentane
C_5H_{12}	propane, 2,2-dimethyl-
C_6H_6	1,2,4,5-hexatetraene
C_6H_6	trimethylenecycopropane
C_6H_6	2,4-hexadiyne
C_6H_6	hexa-1,5-diene-3-yne
C_6H_6	fulvene
$C_{12}H_8$	acenaphthylene
$C_{12}H_8$	biphenylene
$C_{12}H_{10}$	biphenyl
$C_{12}H_{10}$	heptalene

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

Table 10. Molecules Used To Test the Accuracy and Transferability of the DFTB Parametrization for Molecules Containing CHNO

formula	name
HNO	nitrosyl hydride
N_2O_4	dinitrogen tetraoxide
N_2O_5	dinitrogen pentoxide
CH_3NO	formamide
CH_3NO	nitroso methane
CH_3NO	hydroxymethylimine
CH_3NO	1,2-oxaziridine
CH_3NO	formaldoxime
$CO(NH_2)_2$	urea
C_2H_5NO	acetaldoxime
C_2H_7NO	monoethanol amine
$C_2H_6N_2O$	methyl urea
C_2H_5NO	*N*-methyl-formamide
C_2H_5NO	acetamide
C_3H_3NO	oxazole
C_3H_3NO	isoxazole
CH_3NO_2	carbamic acid
$C_2H_6N_2O_2$	azodioxymethane
H_2NCH_2COOH	glycine
$C_6H_5NO_2$	nitrobenzene
H_2O_3	hydrogen trioxide
HCOOH	formic acid
HCOOH	dioxirane

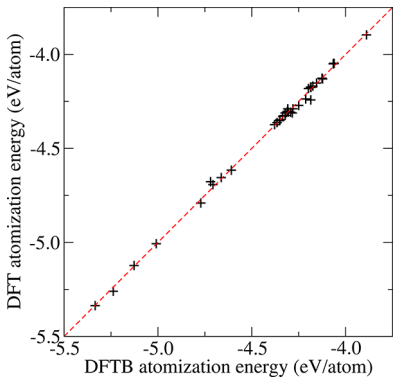


Figure 2. Comparison of atomization energies from optimized DFTB and density functional theory for the 39 molecule set of hydrocarbon molecules.

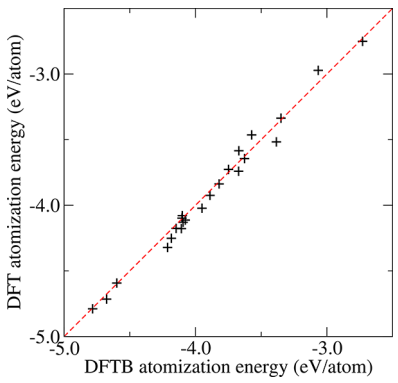


Figure 3. Comparison of atomization energies from optimized DFTB and density functional theory for the 23 molecule set of CHNO-based molecules.

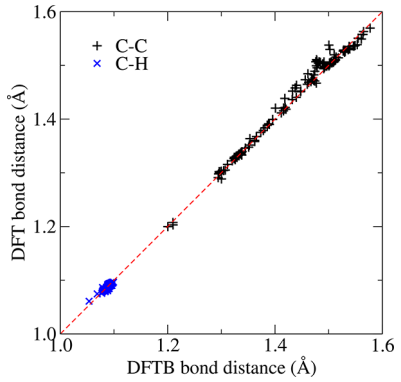


Figure 4. Comparison of equilibrium bond distances for the 39 hydrocarbon molecule set computed from density functional theory and optimized DFTB.

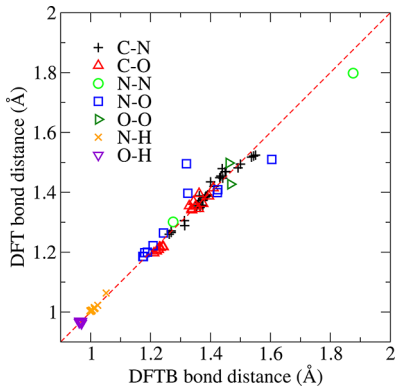


Figure 5. Comparison of equilibrium bond distances for the 23 CHNO molecule set computed from density functional theory and optimized DFTB.

are larger than those obtained for the training molecules by about 50%. We attribute the reduced accuracy and transferability of the CHNO parametrization with respect to that of the hydrocarbons to several factors. First, the interactions between each pair of elements was parametrized to 5−6 molecules rather than the 24 used to parametrize the hydrocarbons. Improved parametrizations might be achievable through the use of larger training sets, although resources such as the Computational Chemistry Comparison and Benchmark Database contain fewer CHNO-based molecules than hydrocarbons. Second, charge transfer is more significant in CHNO-based molecules owing to the greater differences in electronegativity. Finally, the underlying DFTB theory applied in this work, that is, a minimal basis with point charge monopoles, may be insufficient for molecules containing O and N. Self-consistent TB models including higher order moments of the charge density such as dipoles and quadrupoles[6] and/or extra valence orbitals[41] may be required for improved accuracy, albeit at greater computational expense. Nevertheless, the errors in both the atomization energies and geometries obtained with the full CHNO parametrization are acceptably small for the vast majority of real world applications.

5. SUMMARY AND CONCLUSIONS

We have presented a robust and extendible method for generating accurate and transferable parametrizations for DFTB theory that has been applied to molecules containing carbon, hydrogen, nitrogen, and oxygen. The use of simple analytic

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

Table 11. Root-Mean-Square Error in the Per-Atom Atomization Energies, ΔE, and Bond Lengths, Δd, for the Testing and Training Sets of Hydrocarbon and CHNO-Based Molecules[a]

	training molecules			test molecules		
	ΔE (eV/atom)	Δd_{XX} (Å)	Δd_{XH} (Å)	ΔE (eV/atom)	Δd_{XX} (Å)	Δd_{XH} (Å)
hydrocarbons	0.0159	0.011	0.0024	0.0163	0.010	0.0023
CHNO	0.0408	0.017	0.0038	0.0608	0.032	0.0043

[a]X represents C for the hydrocarbon data and C, N, and O for the CHNO data.

functions to represent the radial dependences of the bond integrals and pair potentials is a crucial step since it provides a small number of adjustable parameters that can be optimized numerically to minimize errors with respect to databases of *ab initio* calculated reference data. We have employed relatively simple yet effective algorithms for the numerical optimization of the adjustable parameters. The rate of convergence of the optimization could in principle be improved through the use of different algorithms and/or different objective functions.

The adjustable parameters in the DFTB model were updated iteratively to minimize the value of an objective function that measures errors in the atomization energy and interatomic forces with respect to *ab initio* data. The total objective function is a sum of dimensionless objective functions for the forces and atomization energies. This framework can be readily extended to include dimensionless objective functions written in terms of other properties. Hence, using this approach the parametrizations and DFTB models can be tailored for specific applications. The parametrizations provided in Tables 1−3 and Tables 6−8 can be easily tabulated for use in the variety of electronic structure codes in which DFTB theory has been implemented.

The optimized DFTB parametrizations, which we provide, were tested by comparing the atomization energies and equilibrium gas-phase geometries of molecules not included in the training data to the results of *ab initio* calculations. The root-mean-square errors in the atomization energies and bond lengths were only 0.0163 eV/atom and 0.010 Å for hydrocarbons and 0.0608 eV/atom and 0.032 Å for more complex molecules containing C, H, N, and O, respectively, that were not included in the training data. These assessments show that highly optimized DFTB models are capable of closely reproducing the results of density functional theory at a small fraction of the computational expense.

AUTHOR INFORMATION

Corresponding Author

*E-mail: cawkwell@lanl.gov.

ORCID

P. Yang: 0000-0003-4726-2860

M. J. Cawkwell: 0000-0002-8919-3368

Funding

This work was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory (P.Y., A.M.N.N., and M.J.C.) and a U.S. Department of Energy Computational Science Graduate Fellowship (A.K.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Nestor Aguirre, Enrique Batista, Thomas Lenosky, Matous Mrovec, Romain Perriot, Michael Wall, and Art Voter for numerous helpful discussions.

REFERENCES

(1) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260.

(2) Frauenheim, T.; Seifert, G.; Elsterner, M.; Hajnal, Z.; Jungnickel, G.; Porezag, D.; Suhai, S.; Scholz, R. A self-consistent charge density functional based tight binding method for predictive materials simulations in physics, chemistry, and biology. *Phys. Status Solidi B* **2000**, *217*, 41.

(3) Koskinen, P.; Mäkinen, V. Density-functional tight-binding for beginners. *Comput. Mater. Sci.* **2009**, *47*, 237.

(4) Slater, J. C.; Koster, G. F. Simplified LCAO Method for the Periodic Potential Problem. *Phys. Rev.* **1954**, *94*, 1498.

(5) Goringe, C. M.; Bowler, D. R.; Hernandez, E. Tight-binding modelling of materials. *Rep. Prog. Phys.* **1997**, *60*, 1447−1512.

(6) Finnis, M. W.; Paxton, A. T.; Methfessel, M.; van Schilfgaarde, M. Crystal Structures of Zirconia from First Principles and Self-Consistent Tight Binding. *Phys. Rev. Lett.* **1998**, *81*, 5149.

(7) Finnis, M. *Interatomic Forces in Condensed Matter*; Oxford University Press: 2003; DOI: 10.1093/acprof:oso/9780198509776.001.0001.

(8) Esfarjani, K.; Kawazoe, Y. Self-consistent tight-binding formalism for charged systems. *J. Phys.: Condens. Matter* **1998**, *10*, 8257−8267.

(9) Gaus, M.; Chou, C. P.; Witek, H.; Elstner, M. Automatized Parametrization of SCC-DFTB Repulsive Potentials: Application to Hydrocarbons. *J. Phys. Chem. A* **2009**, *113*, 11866−11881.

(10) Goldman, N.; Fried, L. E. *J. Phys. Chem. C* **2012**, *116*, 2198−2204.

(11) Goldman, N.; Fried, L. E.; Koziol, L. Using Force-Matched Potentials To Improve the Accuracy of Density Functional Tight Binding for Reactive Conditions. *J. Chem. Theory Comput.* **2015**, *11*, 4530−4535.

(12) Doemer, M.; Liberatore, E.; Knaup, J. M.; Tavernelli, I.; Rothlisberger, U. In situ parameterisation of SCC-DFTB repulsive potentials by iterative Boltzmann inversion. *Mol. Phys.* **2013**, *111*, 3595−3607.

(13) Wahiduzzaman, M.; Oliveira, A. F.; Philipsen, P.; Zhechkov, L.; van Lenthe, E.; Witek, H. A.; Heine, T. DFTB parameters for the periodic table: part 1, electronic structure. *J. Chem. Theory Comput.* **2013**, *9*, 4006.

(14) Oliveira, A. F.; Philipsen, P.; Heine, T. DFTB Parameters for the Periodic Table, Part 2: Energies and Energy Gradients from Hydrogen to Calcium. *J. Chem. Theory Comput.* **2015**, *11*, 5209.

(15) Chou, C.-P.; Nishimura, Y.; Fan, C.-C.; Mazur, G.; Irle, S.; Witek, H. A. Automatized Parameterization of DFTB Using Particle Swarm Optimization. *J. Chem. Theory Comput.* **2016**, *12*, 53−64.

(16) Baskes, M. I.; Srinivasan, S. G.; Valone, S. M.; Hoagland, R. G. Multistate modified embedded atom method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *75*, 094113.

(17) Larentzos, J. P.; Rice, B. M.; Byrd, E. F. C.; Weingarten, N. S.; Lill, J. V. Parameterizing Complex Reactive Force Fields Using Multiple Objective Evolutionary Strategies (MOES). Part 1: ReaxFF Models for Cyclotrimethylene Trinitramine (RDX) and 1,1-Diamino-2,2-dinitroethene (FOX-7). *J. Chem. Theory Comput.* **2015**, *11*, 381−391.

(18) Rice, B. M.; Larentzos, J. P.; Byrd, E. F. C.; Weingarten, N. S. Parameterizing Complex Reactive Force Fields Using Multiple Objective Evolutionary Strategies (MOES): Part 2: Transferability of

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200

ReaxFF Models to C-H-N-O Energetic Materials. *J. Chem. Theory Comput.* **2015**, *11*, 392−405.
(19) Stukowski, A.; Fransson, E.; Mock, M.; Erhart, P. Atomicrex - a general purpose tool for the construction of atomic interaction models. *Modell. Simul. Mater. Sci. Eng.* **2017**, *25*, 055003.
(20) Seifert, G.; Joswig, J.-O. Density-functional tight binding-an approximate density-functional theory method. *WIREs Comput. Mol. Sci.* **2012**, *2*, 456−465.
(21) Gaus, M.; Cui, Q.; Elstner, M. Density functional tight binding: application to organic and biological molecules. *WIREs Comput. Mol. Sci.* **2014**, *4*, 49−61.
(22) Elstner, M.; Seifert, G. Density Functional Tight Binding. *Philos. Trans. R. Soc., A* **2014**, *372*, 20120483.
(23) Cawkwell, M. J.; Niklasson, A. M. N.; Dattelbaum, D. M. Extended Lagrangian Born-Oppenheimer molecular dynamics simulations of the shock-induced chemistry of phenylacetylene. *J. Chem. Phys.* **2015**, *142*, 064512.
(24) Goodwin, L.; Skinner, A. J.; Pettifor, D. G. Generating transferable tight-binding parameters - application to silicon. *Europhys. Lett.* **1989**, *9*, 701−706.
(25) Horsfield, A. P. Efficient Ab Initio Tight Binding. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 6594.
(26) Kenny, S. D.; Horsfield, A. P.; Fujitani, H. Transferable atomic-type orbital basis sets for solids. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2000**, *62*, 4899.
(27) Kenny, S. D.; Horsfield, A. P. PLATO: a localised orbital based density functional theory code. *Comput. Phys. Commun.* **2009**, *180*, 2616.
(28) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Dover: 1989.
(29) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(30) Blöchl, G. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953.
(31) Kresse, G.; Joubert, J. From ultrasoft pseudopotentials to the projector augmented wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758.
(32) Sankey, O. F.; Niklewski, D. J. Ab initio multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1989**, *40*, 3979.
(33) Horsfield, A. P. Where does tight binding go from here? *Phys. Status Solidi B* **2012**, *249*, 231−236.
(34) Goedecker, S.; Teter, M.; Hutter, J. Seperable dual-space Gaussian pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703.
(35) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477.
(36) Lenosky, T. J.; Sadigh, B.; Alonso, E.; Bulatov, V. V.; de la Rubia, T. D.; Kim, J.; Voter, A. F.; Kress, J. D. Highly optimized empirical potential model of silicon. *Modell. Simul. Mater. Sci. Eng.* **2000**, *8*, 825−841.
(37) Tangney, P.; Scandolo, S. A many-body interatomic potential for ionic systems: Application to MgO. *J. Chem. Phys.* **2003**, *119*, 9673.
(38) Liu, S.; Grinberg, I.; Rappe, A. M. Development of a bond-valence based interatomic potential for $BiFeO_3$ for accurate molecular dynamics simulations. *J. Phys.: Condens. Matter* **2013**, *25*, 102202.
(39) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220*, 671.
(40) http://cccbdb.nist.gov (accessed Oct 25, 2017).
(41) Srinivasan, S. G.; Goldman, N.; Tamblyn, I.; Hamel, S.; Gaus, M. A density functional tight binding model with an extended basis set and three-body repulsion for hydrogen under extreme conditions. *J. Phys. Chem. A* **2014**, *118*, 5520−5528.

DOI: 10.1021/acs.jctc.7b00762
J. Chem. Theory Comput. 2017, 13, 6191−6200