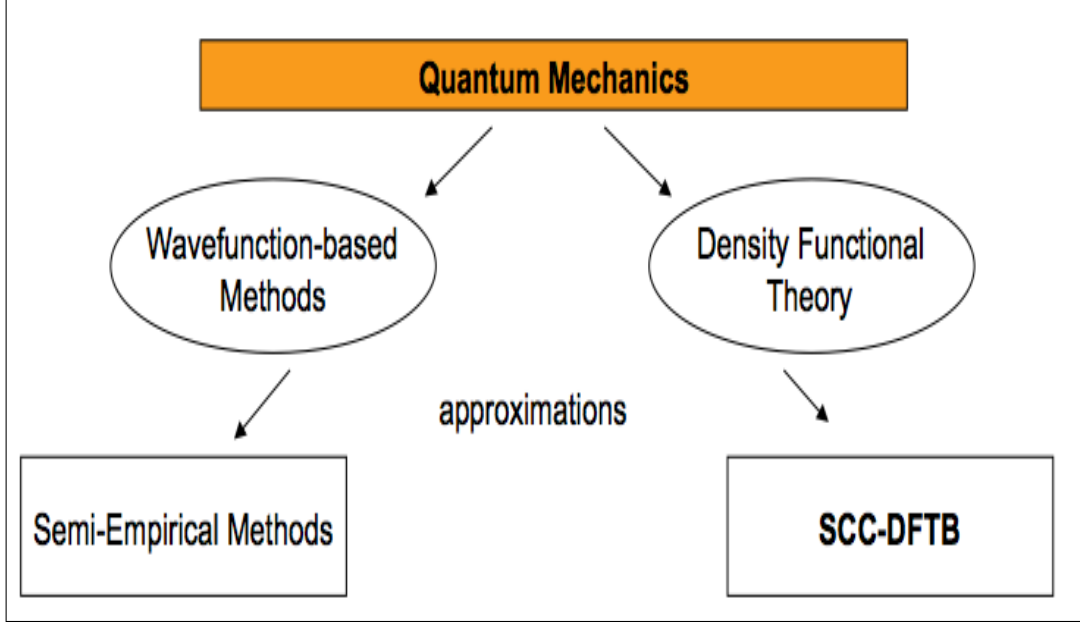


## I. SEMI-EMPIRICAL DFT: SCC-DFTB

DFT can treat up to 100 atoms in routine applications, sometimes even more and about several ps in MD simulations. Very often, one would like to go to larger systems, therefore approximations to DFT are required.



### A. Non-selfconsistent schemes

To get started, consider a case, where you know the ground state density  $\rho_0$  already to sufficient accuracy. In this case, one can omit the self consistent solution of the KS equations and get the orbitals immediately through:

$$\left[ -\frac{1}{2}\nabla^2 + v_{eff}[\rho_0] \right] \phi_i = \epsilon_i \phi_i$$

( $\rho_0$  stands for a proper chosen input density in the following). This saves a factor of 5 -10 already, however, it is the starting point for further approximations.

Consider a minimal basis set consisting of atomic orbitals, i.e.  $\eta_\mu = 2s, 2p_x, 2p_y, 2p_z$  for first row elements (we omit the core states in the following, since they are in a good

approximation chemically inactive) and  $\eta_\mu = 1s$  for H. With the basis set expansion

$$\phi_i = \sum_{\mu} c_{\mu}^i \eta_{\mu}$$

and the Hamiltonian

$$\hat{H}[\rho_0] = \hat{T} + v_{eff}[\rho_0]$$

we find:

$$\sum_{\mu} c_{\mu}^i \hat{H}[\rho_0] |\eta_{\mu}\rangle = \epsilon_i \sum_{\mu} c_{\mu}^i |\eta_{\mu}\rangle \quad (1)$$

Multiplication with  $\langle \eta_{\nu} |$

$$\sum_{\mu} c_{\mu}^i \langle \eta_{\nu} | \hat{H}[\rho_0] | \eta_{\mu} \rangle = \epsilon_i \sum_{\mu} c_{\mu}^i \langle \eta_{\nu} | \eta_{\mu} \rangle \quad (2)$$

or in matrix notation

$$\mathbf{H}\mathbf{C} = \mathbf{S}\mathbf{C}\epsilon \quad (3)$$

This means, we just have to solve the **eigenvalue equation** once, i.e. we have to diagonalize the **Hamilton matrix**  $H_{\mu\nu} = \langle \eta_{\nu} | \hat{H}[\rho_0] | \eta_{\mu} \rangle$ . Note, that our basis set is **non-orthogonal**, i.e. the **overlap matrix**  $S_{\mu\nu} = \langle \eta_{\nu} | \eta_{\mu} \rangle$  appears in the eigenvalue equations.

### 1. Empirical Tight-Binding: ETB or Hückel Theory

In empirical schemes, the basis functions are taken to be orthogonal, i.e.  $S_{\mu\nu} = \delta_{\mu\nu}$ .

Background is the so called Löwdin orthogonalization, where we get orthonormal orbitals through:

$$\eta' = \mathbf{S}^{1/2} \eta$$

Introducing orthonormal orbitals means multiplying with  $\mathbf{S}^{-1/2}$  and inserting a '1':

$$\mathbf{S}^{-1/2} \mathbf{H} \mathbf{S}^{-1/2} \mathbf{S}^{1/2} \mathbf{C} = \mathbf{S}^{-1/2} \mathbf{S}^{1/2} \mathbf{S}^{1/2} \mathbf{C} \epsilon$$

to get the orthonormal equations ( $\mathbf{C}' = \mathbf{S}^{1/2}\mathbf{C}$ ):

$$\mathbf{H}'\mathbf{C}' = \mathbf{C}'\epsilon$$

Introducing orthonormal orbitals means effectively changing the Hamiltonian. And this is convenient, since in empirical schemes the Hamiltonmatrix is completely fitted to empirical data, e.g. for Carbon to the solid state band-structures of several crystal structures (diamond, graphite, bcc etc.), or, in Hückel theory, to properties of Hydrocarbons.

Diagonalization leads to the one-particle energies  $\epsilon_i$ , i.e. to the so called electronic energy:

$$E_{elec} = \sum_i \epsilon_i$$

If we compare this to the total energy in DFT,

$$E[\rho] = \sum_i^{occ} \epsilon_i - \frac{1}{2} \int \frac{\rho_0(r)\rho_0(r')}{|r-r'|} dr dr' + E_{xc}[\rho_0] - \int v_{xc}(r)\rho_0(r)dr + \frac{1}{2} \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \quad (4)$$

it is obvious, that a big part of energy is missing, the so called **double-counting** and core-core repulsion terms in DFT. First of all, it is interesting to note, that the double counting terms depend on the input/reference density  $\rho_0$  only. The XC parts are hard to evaluate, however, in GGA we can say that they decay exponentially due to the exponential decay of the density-overlap. If we assume an atomic density decomposition,  $\rho = \sum_\alpha \rho_\alpha$ , the coulomb contributions

$$\frac{1}{2} \sum_{\alpha\beta} \left[ \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} - \int \frac{\rho_\alpha(r)\rho_\beta(r')}{|r-r'|} dr dr' \right]$$

show an exponential decay as well!

Therefore, first ETB models had the form:

$$E_{tot} = \sum_i \epsilon_i + \frac{1}{2} \sum_{\alpha\beta} U_{\alpha\beta}$$

with the two-body terms  $U_{\alpha\beta}$  being exponentials fitted to reproduce e.g. geometries, vibrational frequencies and reaction energies of suitable systems.

## 2. Density Functional Tight-Binding

The derivation of parameters via fitting is a quite complicated process. If one could derive the parameters from DFT calculations one would gain much more flexibility and a simplified parametrization scheme.

In a first step, one needs a basis set. In TB theory, basis functions are atomic orbitals, and these can be calculated from the atomic KS equations:

$$\left[ -\frac{1}{2}\nabla^2 + v_{eff}[\rho_{atom}] \right] \eta_\mu = \epsilon_\mu \eta_\mu$$

For a basis, we are free to choose whatever we want. Atomic orbitals have the disadvantage, that they are very diffuse, in solids, molecules or clusters the size of the orbitals is 'compressed' due to the interaction with the neighbors. A measure for the distance between neighbors is given by the so called 'covalent radius'  $r_0$  and is determined for all atoms empirically.

Therefore, it is wise to use orbitals, which incorporate this information somehow. One way to enforce this, is to add an additional (harmonic) potential to the atomic Kohn-Sham equations, which leads to compressed atomic orbitals, or optimized atomic orbitals (O-LCAO).

$$\left[ -\frac{1}{2}\nabla^2 + v_{eff}[\rho_{atom}] + \left( \frac{r}{r_0} \right)^2 \right] \eta_\mu = \epsilon_\mu \eta_\mu \quad (5)$$

As a result of the atomic calculations, we get the orbitals  $\eta_\mu$ , the electron density at atom  $\alpha$ ,  $\rho_\alpha = \sum |\eta_\mu|^2$  and the overlap matrix  $S_{\mu\nu} = \langle \eta_\nu | \eta_\mu \rangle$ . To solve the eigenvalue problem eq. 2 or eq. 3, we only need the Hamiltonian matrix. This leads to further approximations, since although we have the complete input density  $\rho_0 = \sum \rho_\alpha$  the Hamiltonian evaluation would be very complicated:

$$H_{\mu\nu} = \langle \eta_\nu | \hat{H}[\rho_0] | \eta_\mu \rangle = \langle \eta_\nu | \hat{H}[\sum \rho_\alpha] | \eta_\mu \rangle$$

We therefore usually make the so called 2-center approximation for  $\mu \neq \nu$ :

$$H_{\mu\nu} = \langle \eta_\nu | \hat{H}[\rho_0] | \eta_\mu \rangle = \langle \eta_\nu | \hat{H}[\rho_\alpha + \rho_\beta] | \eta_\mu \rangle \quad (6)$$

where orbital  $\nu$  is located on atom  $\alpha$  and orbital  $\mu$  is located on atom  $\beta$ . The diagonal Hamitonelements  $H_{\mu\mu} = \epsilon_\mu$  are taken from eq. 5.

$H_{\mu\nu}$  and  $S_{\mu\nu}$  are tabulated for various distances between atom pairs up to 10 , where they vanish (also due to compression!). For any molecular geometry, these matrix elements are read in based on the distance between to atoms and then oriented in space using the Slater-Koster sin/cos combination rules. Then the generalized eigenvalue problem 3 is solved and the first part of the energy can be calculated. It should be emphasized that this is a non-orthogonal TB scheme, which is more **transferable** due to the appearance of the overlap matrix.

The second part,

$$E_{rep}[\rho_0] = \frac{1}{2} \sum_{\alpha\beta} U_{\alpha\beta}$$

is calculated pointwise as follows: To get the repulsive potential e.g. for Carbon one could take the Carbon dimer  $C_2$ , stretch its bond and for each distance calculate the total energy with DFT and the electronic TB part  $\sum_i \epsilon_i$ .  $U_{CC}(R_{C-C})$  is the given point-wise for every  $R_{C-C}$  by:

$$U_{CC}(R_{C-C}) = E_{tot}^{DFT}(R_{C-C}) - \sum_i \epsilon_i$$

In practice, we include several structures in the fit. For the C-C bond, e.g. we take molecules with C-C single, double and triple bonds, as shown in Fig 1.

The resulting DFTB method works very well for homo-nuclear systems, where charge transfer between the atoms in the system does not occur or is very small. As soon as charge is flowing between atoms because of an electronegativity difference, the resulting density is no more well approximated by the superposition of the atomic densities  $\rho_0 = \sum \rho_\alpha$ . However, the formalism works very well, when charge flow is small, therefore an extension will try to start form the non-selfconsistent scheme.

The problem is, that the effective Kohn-Sham potentials contain only the neutral re-

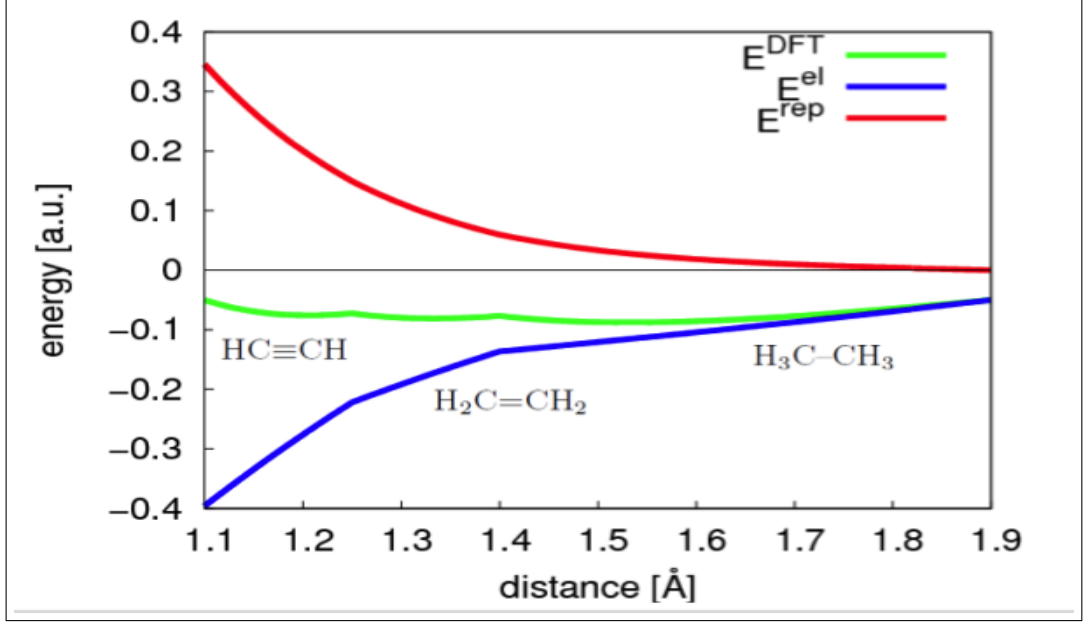


Abbildung 1:  $E_{rep}$ , as derived from the three molecules shown

ference density  $\rho_0$ , which does not account for charge transfer between atoms. Lets try a Taylor series expansion (functional expansion) of the potential with the ground state density  $\rho$  around the reference density  $\rho_0$ :

$$v_{eff}[\rho] = v_{eff}[\rho_0] + \int \frac{\delta v_{eff}[\rho]}{\delta \rho} \delta \rho dr$$

This is a very interesting procedure, since now the potential inserted into the KS equations will lead to the same matrix elements of  $H_{\mu\nu}[\rho_0]$  depending on the reference density as above, and corrections terms, which have to deal with the functional derivative.

## B. Self-Consistent Charge Density Functional Tight-Binding: SCC-DFTB

Since we need the total energy and not only the KS equations, it is better to start the functional expansion with the DFT total energy. The SCC-DFTB method is derived from density functional theory (DFT) by a second order expansion of the DFT total energy functional with respect to the charge density fluctuations  $\delta \rho$  around a given reference density  $\rho_0$  ( $\rho'_0 = \rho_0(\vec{r}')$ ,  $\int' = \int d\vec{r}'$ ):

$$\begin{aligned}
E = & \sum_i^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \iint' \left( \frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta \rho \delta \rho'} \bigg|_{n_0} \right) \Delta \rho \Delta \rho'. \\
& - \frac{1}{2} \iint' \frac{\rho'_0 \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0] n_0 + E_{cc}
\end{aligned} \tag{7}$$

After introducing the LCAO basis  $\Psi_i = \sum c_\mu^i \eta_\mu$ , the first term becomes:

$$\langle \Psi_i | \hat{H}^0 | \Psi_i \rangle = \sum c_\mu^i c_\nu^i H_{\mu\nu}$$

and can be evaluated as discussed above. The last four terms depend only on the reference density  $\rho_0$  and represent the repulsive energy contribution  $E_{rep}$  discussed above. Therefore, we only have to deal with the second order terms.

The second order term in the charge density fluctuations  $\Delta \rho$  (second term in Eq.7) is approximated by writing  $\Delta \rho$  as a superposition of atomic contributions

$$\Delta \rho = \sum_\alpha \Delta \rho_\alpha,$$

which decay quickly with increasing distance from the corresponding center. To further simplify  $E^{2nd}$ , we apply a monopole approximation

$$\Delta \rho_\alpha \approx \Delta q_\alpha F_{00}^\alpha Y_{00}, \tag{8}$$

Basically,  $\Delta \rho_\alpha$  is assumed to look like an 1s orbital.  $F_{00}^\alpha$  denotes the normalized radial dependence of the density fluctuation on atom  $\alpha$ , which is constrained (approximated) to be spherical ( $Y_{00}$ ); i.e., the angular deformation of the charge density change in second order is neglected.

$$E^{2nd} \approx \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \iint' \left( \frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta \rho \delta \rho'} \bigg|_{n_0} \right) F_{00}^\alpha F_{00}^\beta Y_{00}^2 dr dr' \tag{9}$$

This formular looks complicated, but has a quite simple curve shape:

- For large distances,  $R_{\alpha\beta} = |\vec{r} - \vec{r}'| \rightarrow \infty$  the XC terms vanish, and the integral describes the coulomb-interaction of two spherical normalized charge densities, which reduces basically to  $1/R_{\alpha\beta}$ , i.e. we get:

$$E^{2nd} \approx \frac{1}{2} \sum_{\alpha\beta} \frac{\Delta q_\alpha \Delta q_\beta}{R_{\alpha\beta}}$$

- For vanishing interatomic distance,  $R_{\alpha\beta} = |\vec{r} - \vec{r}'| \rightarrow 0$ , the integral describes the e-e interaction on atom  $\alpha$ . We can approximate the integral as:

$$E^{2nd} \approx 0.5 \frac{\partial^2 E_\alpha}{\partial^2 q_\alpha} = U_\alpha$$

$U_\alpha$  is known as the Hubbard parameter or the chemical hardness. It describes, how much the energy of a system changes upon adding or removing electrons.

Now we need a formula to interpolate between these two cases. A very similar situation appears in semi-empirical quantum chemical methods like MNDO, AM1 or PM3, where  $\gamma$  has a simple form, as given, for example, by the Klopman-Ohno approximation,

$$\gamma_{\alpha\beta} = \frac{1}{\sqrt{R_{\alpha\beta}^2 + 0.25(1/U_\alpha + 1/U_\beta)^2}}. \quad (10)$$

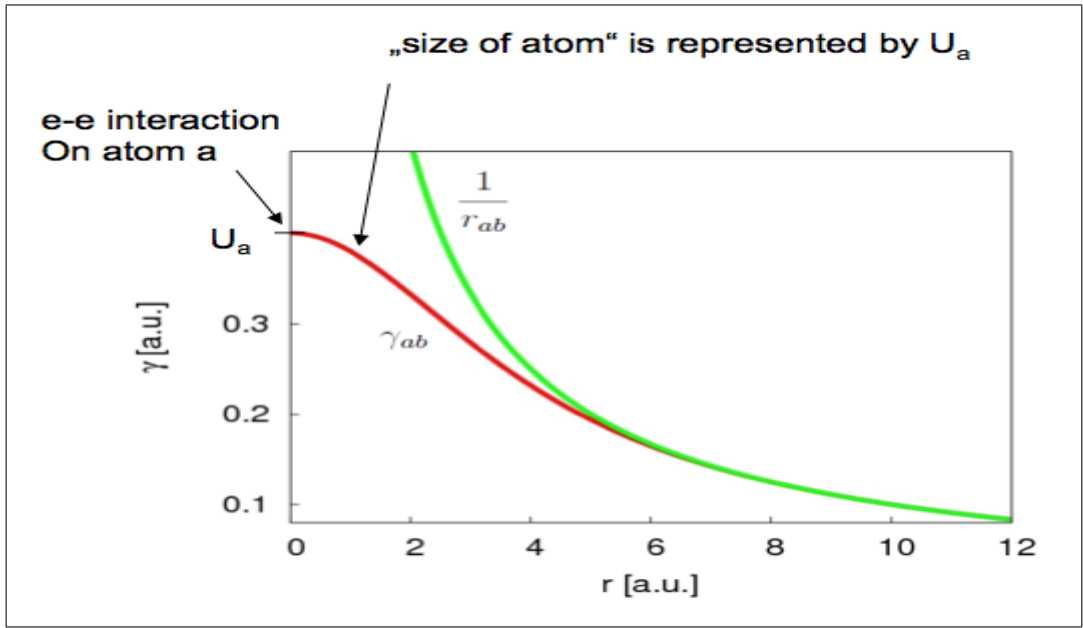


Abbildung 2: The 'gamma' function used to interpolate  $1/R$  into the bonding region

To derive analytically an expression, we approximate the charge density fluctuations with spherical charge densities. Slater like distributions

$$F_{00}^\alpha = \frac{\tau_\alpha}{8\pi} \exp(-\tau_\alpha |r - R_\alpha|) \quad (11)$$

located at  $R_\alpha$  allow for an analytical evaluation of the Hartree contribution of two spherical charge distributions. This leads to a function of  $\gamma_{\alpha\beta}$ , which depends on the parameters  $\tau_\alpha$



and  $\tau_\beta$ , determining the extension of the charge densities of the atoms  $\alpha$  and  $\beta$  as shown in Fig 9. This function has a  $1/R_{\alpha\beta}$  dependence for large  $R_{\alpha\beta}$  and approaches a finite value for  $R_{\alpha\beta} \rightarrow 0$ . For zero interatomic distances, i.e.  $\alpha = \beta$ , one finds that:

$$\tau_\alpha = \frac{16}{5}\gamma_{\alpha\alpha} \quad (12)$$

Eq. 12 implies, that the extension of the charge distribution is inversely proportional to the chemical hardness of the respective atom; i.e. the size of an atom is inversely related to its chemical hardness.

After integration,  $E^{2nd}$  becomes a simple two-body expression depending on atomic-like charges

$$E^{2nd} = \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \quad (13)$$

The diagonal terms  $\gamma_{\alpha\alpha}$  model the dependence of the total energy on charge density fluctuations (decomposed into atomic contributions) in second order. The monopole approximation restricts the change of the electron density considered and no spatial deformations are included; only the change of energy with respect to change of charge on the atom  $\alpha$  is considered. By neglecting the effect of the chemical environment on atom  $\alpha$ , the diagonal part of  $\gamma$  can be approximated by the chemical hardness  $\eta$  of the atom,

$$\gamma_{\alpha\alpha} = 2\eta_\alpha = U_\alpha = \frac{\partial^2 E_\alpha}{\partial^2 q_\alpha}, \quad (14)$$

$E_\alpha$  is the energy of the isolated atom  $\alpha$ .  $U_\alpha$  is known as the Hubbard parameter and is twice the chemical hardness of atom  $\alpha$ , which can be estimated from the difference of the ionization potential and the electron affinity of atom  $\alpha$ . For SCC-DFTB, it is calculated using Janak's theorem by taking the first derivative of the energy of the highest occupied molecular orbital with respect to occupation number.

## 1. Performance

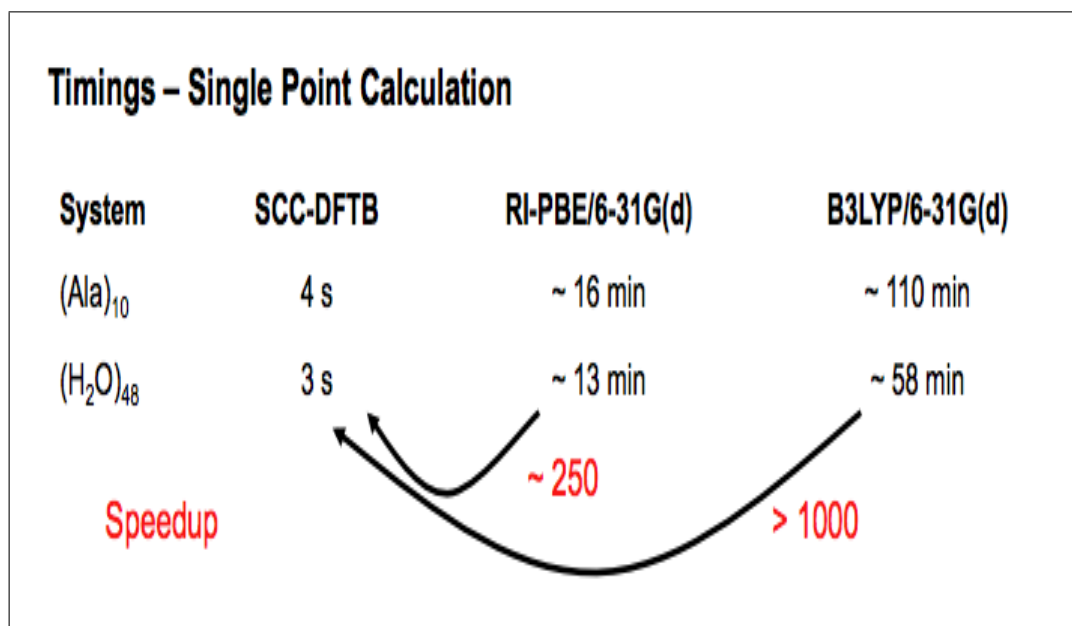


Abbildung 3: CPU-timings

*Performance for small organic molecules*

(mean absolut deviations)

Reaction energies<sup>a)</sup>: ~ 5 kcal/mole

Bond-lenghts<sup>a)</sup>: ~ 0.014 Å°

Bond angles<sup>b)</sup>: ~ 2°

Vib. Frequencies<sup>c)</sup>: ~6-7 %

a) J. Andzelm and E. Wimmer, J. Chem. Phys. **96**, 1280 1992.

b) J. S. Dewar, E. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Am. Chem. Soc. **107**, 3902 1985.

c) J. A. Pople, et al., Int. J. Quantum Chem., Quantum Chem. Symp. **15**, 269 1981.

Abbildung 4: DFTB Test

Mean Absolute Errors in Calculated Heats of Formation for Neutral Molecules Containing the Elements C, H, N and O (kcal/mol). *J. Phys. Chem. A* **2006**, *110*, 13551

	N	AM1	PM3	PDDG/PM3	SCC-DFTB
Hydrocarbons	254	5.6	3.6	2.6	4.8
All Molecules	622	6.7	4.4	3.2	5.9
Training Set	134	6.1	4.3	2.7	7.0
Test Set	488	6.8	4.4	3.3	5.6

Abbildung 5: DFTB Test

Absolute Errors for Additional Molecular Properties of CHNO-containing Species. *J. Phys. Chem. A* **2006**, *110*, 13551

	N	AM1	PM3	PDDG/PM3	SCC-DFTB
Bond lengths (Å)	218	0.017	0.012	0.013	0.012
Bond angles (deg.)	126	1.5	1.7	1.9	1.0
Dihedral angles (deg.)	30	2.8	3.2	3.7	2.9
Dipole moments (D)	47	0.23	0.25	0.23	0.39

Abbildung 6: DFTB Test

CCSD(T): 5.0 kcal/mole (Klopper et al PCCP 2000 2, 2227)

BLYP: 4.2 kcal/mole

PBE: 5.1 kcal/mole

B3LYP: 4.6 kcal/mole

HF: 3.7 kcal/mole

(from Xu&Goddard, JCPA 2004)

For larger systems:

DFTB: 3.3 kcal/mole

HF: 5.7 kcal/mole @ 6-31G\*

B3LYP: 6.8 kcal/mole @ 6-31G\* ~2 kcal/mole BSSE (BSIE)

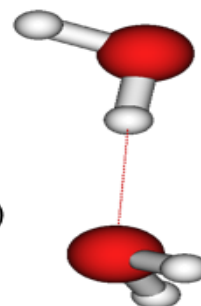


Abbildung 7: DFTB Test

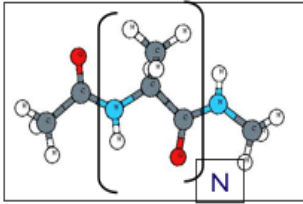
## Hydrogen bonds of 57 complexes

JPCA 111, 5753

	AM1	OM2	SCC-DFTB
E	2.8	1.5	2.7
R	0.25	0.20	0.08
$\theta$	33.7°	12.1°	6.2°

- DFTB scatters around B3LYP values
- AM1 0.12 Å too long
- OM2 0.14 Å too short

Abbildung 8: DFTB Test



Otte, Scholten & Thiel JPCA 111, 5753

**TABLE 5: Mean Absolute Deviations for the Peptide Test Set<sup>a</sup>**

	<i>N</i>	AM1	OM2	DFTB
relative energies (kcal/mol)	22	2.0	1.7	1.1
backbone H-bond lengths (Å)	67	0.22	0.34	0.26
backbone dihedral angles (deg)	190	17.0	12.0	9.0

Abbildung 9: DFTB Test