

$$E_{tot} = E_{TB} + E_{disp} + E_{charge} + E_{rep}$$

$$E_{TB} =$$

$$E_{disp} =$$

$$E_{charge} =$$

## 1 Total Energy for GFN2-xTB

$$\begin{aligned} E_{GFN2-xTB} &= E_{rep}^{(0)} + E_{disp}^{(0,1,2)} + E_{EHT}^{(1)} + E_{IES+IXC}^{(2)} + E_{AES+AXC}^{(2)} + E_{IES+IXC}^{(3)} \\ &= E_{rep} + E_{disp}^{DA'} + E_{EHT} + E_{\gamma} + E_{AES} + E_{AXC} + E_{\Gamma}^{GFN2} \end{aligned} \quad (1)$$

### 1.1 Repulsion Energy

$$\begin{aligned} E_{rep} &= \frac{1}{2} \sum_{A,B} \frac{Z_A^{eff} Z_B^{eff}}{R_{AB}} e^{-\sqrt{a_A a_B} (R_{AB})^{(k_f)}} \\ k_f &= \begin{cases} 1 & \text{if } A, B \in \{\text{H, He}\} \\ \frac{3}{2} & \text{otherwise} \end{cases} \end{aligned} \quad (2)$$

$Z^{eff}$  and  $a$  are variables fitted for each element. A,B are the labels of atoms. Since we only have C and H in our systems we can simplify this quite a bit in code.  $R_{AB}$  is the distance between the A and B atoms.

### 1.2 Extended Hückel Theory Energy

$$E_{EHT} = \sum_{\mu\nu} P_{\mu\nu} + H_{\mu\nu}^{EHT} \quad (3)$$

where  $\mu$  and  $\nu$  are AO indecies,  $l$  and  $l'$  index shells. Both AO's are associated with an atom labled A and B.

$$\begin{aligned} P_{\mu\nu} &= P_{\mu\nu}^{(0)} + \delta P_{\mu\nu} \\ P_{\mu\nu}^{(0)} &= ?? \\ \delta P_{\mu\nu} &= ?? \\ H_{\mu\nu}^{EHT} &= \frac{1}{2} K_{AB}^{ll'} S_{\mu\nu} (H_{\mu\mu} + H_{\nu\nu}) \\ &\quad \cdot X(EN_A, EN_B) \\ &\quad \cdot \Pi(R_{AB}, l, l') \\ &\quad \cdot Y(\zeta_l^A, \zeta_{l'}^B), \forall \mu \in l(A), \nu \in l'(B) \end{aligned} \quad (4)$$

$K_{AB}^{ll'}$  is a element and shell specific fitted constant however, in GFN2 it only depends on the shells.  $S_{\mu\nu} = \langle \phi_\mu | \phi_\nu \rangle$  is just the overlap of the orbitals. In GFN2

$H_{\kappa\kappa} = h_A^l - \delta h_{CN'_A}^l CN'_A$  where  $CN'_A$  is the modified GFN2-type Coordinate Number for the element of atom A.  $h_A^l$  and  $\delta h_{CN'_A}^l$  are both fitted constants.  $EN_A$  is the electronegativity of the element of atom A.

$$\begin{aligned} X(EN_A, EN_B) &= 1 + k_{EN} \Delta EN_{AB}^2 \\ k_{EN} &= 0.02 \text{ in GFN2} \\ \Delta EN_{AB}^2 &= (EN_A - EN_B)^2 \end{aligned} \quad (5)$$

The electronegativity for C and H are 2.55 and 2.20 according to wikipedia. Thus here is a table for the combinations we will be working with:

A	B	$X(EN_A, EN_B)$
C	C	1
C	H	$1 + 0.02 \cdot (0.35^2)$
H	C	$1 + 0.02 \cdot (0.35^2)$
H	H	1

$$\Pi(R_{AB}, l, l') = \left( 1 + k_{A,l}^{\text{poly}} \left( \frac{R_{AB}}{R_{\text{cov},AB}} \right)^{\frac{1}{2}} \right) \left( 1 + k_{B,l'}^{\text{poly}} \left( \frac{R_{AB}}{R_{\text{cov},AB}} \right)^{\frac{1}{2}} \right) \quad (6)$$

$R_{\text{cov},AB}$  are the summed covalent radii and taken from Reference 61<sup>1</sup>, the covalent radii is the second number in the table for each element,  $H_{\text{cov}} = 0.32$ ,  $C_{\text{cov}} = 0.75$  (can be found on wiki as 0.75).  $k_{A,l}^{\text{poly}}$  and  $k_{B,l'}^{\text{poly}}$  are element and shell specific constants.

$$Y(\zeta_l^A, \zeta_{l'}^B) = \left( \frac{2\sqrt{\zeta_l^A \zeta_{l'}^B}}{\zeta_l^A + \zeta_{l'}^B} \right)^{\frac{1}{2}} \quad (7)$$

Here,  $\zeta_l^A$  are the STO exponents of the GFN2-xTB AO basis. Slater Type Orbitals are defined as such:

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = N Y_{l,m}(\theta, \varphi) r^{n-1} e^{-\zeta r}$$

N is a normalisation constant, Y are spherical harmonic functions, n, l, m are the quantum numbers for the AO.  $r, \theta, \varphi$  are polar 3D coordinates.  $\zeta$  determines the radial extent of the STO, a large value gives rise to a function that is "tight" around the nucleus and a small value gives a more "diffuse" function. This  $\zeta$  is the one mentioned in the Y term of  $E_{EHT}$  and is a value fitted when constructing the basis set, thus it is given to us.

<sup>1</sup>p.2109, <https://www.taylorfrancis.com/books/mono/10.1201/b12286/crc-handbook-chemistry-physics-william-haynes>

### 1.3 Isotropic electrostatic and XC energy

#### 1.3.1 Second order

$$\begin{aligned}
E_\gamma &= \frac{1}{2} \sum_{A,B}^{N_{atoms}} \sum_{l \in A} \sum_{l' \in B} q_l q_{l'} \gamma_{AB, ll'} \\
\gamma_{AB, ll'} &= \frac{1}{\sqrt{R_{AB}^2 + \eta_{AB, ll'}^{-2}}} \\
\eta_{AB, ll'} &= \frac{1}{2} \left[ \eta_A (1 + k_A^l) + \eta_B (1 + k_B^{l'}) \right]
\end{aligned} \tag{8}$$

$q_l$  is a partial muliken charge.  $\eta_A$  and  $\eta_B$  are element-specific fit parameters, while  $k_A^l$  and  $k_B^{l'}$  are element-specific scaling factors for the individual shells ( $k_A^l = 0$  when  $l = 0$ ).

#### 1.3.2 Third order

$$E_\Gamma^{GFN2} = \frac{1}{3} \sum_A^{N_{atoms}} \sum_{l \in A} (q_l)^3 K_l^\Gamma \Gamma_A \tag{9}$$

$K_l^\Gamma$  is a shell specific constant common for all elements and  $\Gamma_A$  is an element specific constant.

### 1.4 Anisotropic electrostatic energy

$$\begin{aligned}
E_{AES} &= E_{q\mu} + E_{q\Theta} + E_{\mu\mu} \\
&= \frac{1}{2} \sum_{A,B} \{ f_3(R_{AB}) [q_A (\boldsymbol{\mu}_B^T \mathbf{R}_{BA}) + q_B (\boldsymbol{\mu}_A^T \mathbf{R}_{AB})] \\
&\quad + f_5(R_{AB}) [q_A \mathbf{R}_{AB}^T \boldsymbol{\Theta}_B \mathbf{R}_{AB} + q_B \mathbf{R}_{AB}^T \boldsymbol{\Theta}_A \mathbf{R}_{AB} \\
&\quad - 3(\boldsymbol{\mu}_A^T \mathbf{R}_{AB})(\boldsymbol{\mu}_B^T \mathbf{R}_{AB}) + (\boldsymbol{\mu}_A^T \boldsymbol{\mu}_B) R_{AB}^2] \}
\end{aligned} \tag{10}$$

What is  $\mathbf{R}$ ? If  $^T$  is transposition then it has to be a tensor of some kind, we believe it to be a column vector due to the sandwiching of  $\boldsymbol{\Theta}$ .  $\boldsymbol{\mu}_A$  is the cumulative atomic dipole moment of atom A (probably a column vector) and  $\boldsymbol{\Theta}_A$  is the corresponding traceless quadrupole moment (probably a 2 dimensional matrix). The curly braces and brackets are used in the same way as normal

parenthesis for showing order of operations.  $q_A$  is the atomic charge of atom A.

$$\Theta_A^{\alpha\beta} = \frac{3}{2}\theta_A^{\alpha\beta} - \frac{\delta_{\alpha\beta}}{2}(\theta_A^{xx} + \theta_A^{yy} + \theta_A^{zz}) \quad (11)$$

$$\theta_A^{\alpha\beta} = \sum_{\kappa \in A} \sum_{\lambda} P_{\lambda} \left( \alpha_A D_{\lambda\kappa}^{\beta} + \beta_A D_{\lambda\kappa}^{\alpha} - \alpha_A \beta_A S_{\lambda\kappa} - Q_{\lambda\kappa}^{\alpha\beta} \right) \quad (12)$$

$$q_A = Z_A - \sum_{\kappa \in A} \sum_{\lambda} P_{\kappa\lambda} S_{\lambda\kappa} \quad (13)$$

$$\mu_A^{\alpha} = \sum_{\kappa \in A} \sum_{\lambda} P_{\kappa\lambda} (\alpha_A S_{\kappa\lambda} - D_{\kappa\lambda}^{\alpha}) \quad (14)$$

$$D_{\lambda\kappa}^{\alpha} = \langle \phi_{\lambda} | \alpha_i | \phi_{\kappa} \rangle \quad (15)$$

$$Q_{\lambda\kappa}^{\alpha\beta} = \langle \phi_{\lambda} | \alpha_i \beta_i | \phi_{\kappa} \rangle \quad (16)$$

$\alpha$  and  $\beta$  are Cartesian components (wtf is this? why do they disappear when we use  $\Theta$  and so on?? Is it because they are used to index in the matrix and vector? is  $\delta_{\alpha\beta}$  just the delta function then, that would make some sense. But what is  $\alpha_i$  then!?)

$$f_n(R_{AB}) = \frac{f_{damp}(a_n, R_{AB})}{R_{AB}^n} = \frac{1}{R_{AB}^n} \frac{1}{1 + 6 \left( \frac{R_0^{AB}}{R_{AB}} \right)^{a_n}} \quad (17)$$

$$R_0^{AB} = 0.5(R_0^{A'} + R_0^{B'}) \quad (18)$$

$$R_0^{A'} = \begin{cases} R_0^A + \frac{R_{max} - R_0^A}{1 + \exp[-4(CN_A' - N_{val} - \Delta_{val})]} & \text{if } N_{val} \text{ is given} \\ 5.0 \text{ bohrs} & \text{otherwise} \end{cases} \quad (19)$$

$$R_{max} = 5.0 \text{ bohrs} \quad (20)$$

$$\Delta_{val} = 1.2 \quad (21)$$

$R_0^A$  is a fitted value for 12 elements and 5.0 for the rest.  $a_n$  are adjusted global parameters.

## 1.5 Anisotropic XC energy

$$E_{AXC} = \sum_A (f_{XC}^{\mu_A} |\mu_A|^2 + f_{XC}^{\Theta_A} |\Theta_A|^2) \quad (22)$$

Where  $f_{XC}^{\mu_A}$  and  $f_{XC}^{\Theta_A}$  are fitted values. What norms are these?