

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

$$E_{TB} =$$

$$E_{disp} =$$

$$E_{charge} =$$

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 μ and ν 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 61¹, 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, ζ_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, θ, φ are polar 3D coordinates. ζ 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 ζ 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.

¹p.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. η_A and η_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^Γ is a shell specific constant common for all elements and Γ_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.

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

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

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

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

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

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

α and β are Cartesian components (wtf is this? why do they disappear when we use Θ 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 α_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?

1.6 Dispersion Energy

$$\begin{aligned}
E_{disp}^{D4'} = & - \sum_{A>B} \sum_{n=6,8} s_n \frac{C_n^{AB}(q_A, CN_{cov}^A, q_B, CN_{cov}^B)}{R_{AB}^n} f_{damp,BJ}^{(n)}(R_{AB}) \\
& - s_9 \sum_{A>B>C} \frac{(3\cos(\theta_{ABC})\cos(\theta_{BCA})\cos(\theta_{CAB}) + 1)C_9^{ABC}(CN_{cov}^A, CN_{cov}^B, CN_{cov}^C)}{(R_{AB}R_{AC}R_{BC})^3} \\
& \times f_{damp,zero}^{(9)}(R_{AB}, R_{AC}, R_{BC}).
\end{aligned} \tag{23}$$

The term in the second line is the three-body Axilrod–Teller–Muto (ATM) (What is this?????) term and the last line is the corresponding zero-damping function for this term.

C_6^{AB} is the pairwise dipole-dipole dispersion coefficients calculated by numerical integration via the Casimir-Polder relation.

$$C_6^{AB} = \frac{3}{\pi} \sum_j w_j \bar{\alpha}_A(i\omega_j, q_A, CN_{cov}^A) \bar{\alpha}_B(i\omega_j, q_B, CN_{cov}^B) \tag{24}$$

w_j are the integration weights, which are derived from a trapezoidal partitioning between the grid points $j (j \in [1, 23])$.

The isotropically averaged, dynamic dipole-dipole polarizabilities $\bar{\alpha}$ at the j th imaginary frequency $i\omega_j$ are obtained from the self-consistent D4 model; i.e., they are depending on the covalent coordination number and are also charge dependent.

$$\bar{\alpha}_A(i\omega_j, q_A, CN_{cov}^A) = \sum_r^{N_{A,ref}} \xi_A^r(q_A, q_{A,r}) \bar{\alpha}_{A,r}(i\omega_j, q_{A,r}, CN_{cov}^{A,r}) W_A^r(CN_{cov}^A, CN_{cov}^{A,r}) \tag{25}$$

The Gaussian weighting for each reference system is given by:

$$W_A^r(CN_{cov}^A, CN_{cov}^{A,r}) = \sum_{j=1}^{N_{gauss}} \frac{1}{\mathcal{N}} \exp[-6j \cdot (CN_{cov}^A - CN_{cov}^{A,r})^2] \tag{26}$$

with

$$\sum_r^{N_{A,ref}} W_A^r(CN_{cov}^A, CN_{cov}^{A,r}) = 1 \tag{27}$$

\mathcal{N} is a normalization constant.

The number of Gaussian function per reference system N_{gauss} is mostly one, but equal to three for $CN_{cov}^{A,r} = 0$ and reference systems with similar coordination number.

The charge-dependency is included via the empirical scaling function ξ_A^r .

$$\xi_A^r(q_A, q_{A,r}) = \exp \left[3 \left\{ 1 - \exp \left[4\eta_A \left(1 - \frac{Z_A^{eff} + q_{A,r}}{Z_A^{eff} + q_A} \right) \right] \right\} \right] \quad (28)$$

where η_A is the chemical hardness taken from ref 98.

Z_A^{eff} is the effective nuclear charge of atom A, which has been determined by subtracting the number of core electrons represented by the def2-ECPs in the time-dependent DFT reference calculations.

C_8^{AB} is calculated recursively from the lowest order C_6^{AB} coefficients.

$$C_8^{AB} = 3C_6^{AB} \sqrt{Q^A Q^B} \quad (29)$$

$$Q^A = s_{42} \sqrt{Z^A} \frac{\langle r^4 \rangle^A}{\langle r^2 \rangle^A} \quad (30)$$

$\sqrt{Z^A}$ is the ad hoc nuclear charge dependent factor. (How do we get $Z^{????}$)

$\langle r^4 \rangle$ and $\langle r^2 \rangle$ are simple multipole-type expectation values derived from atomic densities which are averaged geometrically to get the pair coefficients. (What is 'r', how we get $^{????}$) (what is $s_{42}^{??}$)

C_{cov}^A is the covalent coordination number for atom A.

q is the atomic charge, so q_A is the atomic charge for atom A.

The damping and scaling parameters in the dispersion model are:

$$a1 = 0.52 \quad | \quad a2 = 5.0 \quad | \quad s6 = 1.0 \quad | \quad s8 = 2.7 \quad | \quad s9 = 5.0$$

C_9^{ABC} is the triple-dipole constant²:

$$C_9^{ABC} = \frac{3}{\pi} \int_0^\infty \alpha^A(i\omega) \alpha^B(i\omega) \alpha^C(i\omega) d\omega \quad (31)$$

The three-body contribution is typically $< 5-10\%$ of E_{disp} , so it is small enough that we can reasonably approximate the coefficients by a geometric mean as²:

$$C_9^{ABC} \approx -\sqrt{C_6^{AB} C_6^{AC} C_6^{BC}} \quad (32)$$

θ_{ABC} is the angle between the two edges going from B to the other two atoms.

θ_{BCA} is the angle between the edges going from C to the other two and so on.

²https://www.researchgate.net/publication/43347348_A_Consistent_and_Accurate_Ab_Initio_Parametrization_of_Density_Functionals_for_the_94_Elements_H-Pu

BJ = Becke-Johnson

$$f_n^{damp,BJ}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + (a_1 \cdot R_{AB}^{crit} + a_2)^6} \quad (33)$$

$$R_{AB}^{crit} = \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} \quad (34)$$

$$f_9^{damp,zero}(R_{AB}, R_{AC}, R_{BC}) = \left(1 + 6 \left(\sqrt{\frac{R_{AB}^{crit} R_{BC}^{crit} R_{CA}^{crit}}{R_{AB} R_{BC} R_{CA}}} \right)^{16} \right)^{-1} \quad (35)$$