

1 Total Energy for GFN2-xTB

$$\begin{aligned} E_{GFN2-xTB} &= E_{rep}^{(0)} + E_{disp}^{(0,1,2)} + E_{EHT}^{(1)} + E_{IES+IXC}^{(2)} + E_{AES+AXC}^{(2)} + E_{IES+IXC}^{(3)} \\ &= E_{rep} + E_{disp}^{D4'} + E_{EHT} + E_{\gamma} + E_{AES} + E_{AXC} + E_{\Gamma}^{GFN2} \end{aligned} \quad (1)$$

1.1 Repulsion Energy

$$E_{rep} = \frac{1}{2} \sum_{A,B} \frac{Z_A^{eff} Z_B^{eff}}{R_{AB}} e^{-\sqrt{a_A a_B} (R_{AB})^{(k_f)}} \quad (2)$$

$$k_f = \begin{cases} 1 & \text{if } A, B \in \{\text{H, He}\} \\ \frac{3}{2} & \text{otherwise} \end{cases} \quad (3)$$

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

$$E_{EHT} = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu}^{EHT} \quad (4)$$

where μ and ν are AO indecies, l and l' index shells. Both AO's are associated with an atom labeled A and B.

$$P_{\mu\nu} = P_{\mu\nu}^{(0)} + \delta P_{\mu\nu} \quad (5)$$

$$P_{\mu\nu}^{(0)} = \sum_j^{occ.MO} c_{\mu j} c_{j\nu} \quad (6)$$

$$\delta P_{\mu\nu} = ?? \quad \text{comes from the iteration, can be skipped for now} \quad (7)$$

$$\begin{aligned} H_{\mu\nu}^{EHT} &= \frac{1}{2} K_{AB}^{ll'} S_{\mu\nu} (H_{\mu\mu} + H_{\nu\nu}) \\ &\cdot X(EN_A, EN_B) \\ &\cdot \Pi(R_{AB}, l, l') \\ &\cdot Y(\zeta_l^A, \zeta_{l'}^B), \forall \mu \in l(A), \nu \in l'(B) \end{aligned} \quad (8)$$

$K_{AB}^{ll'}$ 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.

$$\begin{aligned} CN'_A &= \sum_{B \neq A}^{N_{atoms}} (1 + e^{-10(4(R_{A,cov} + R_{B,cov})/3R_{AB} - 1)})^{-1} \\ &\times (1 + e^{-20(4(R_{A,cov} + R_{B,cov} + 2)/3R_{AB} - 1)})^{-1} \end{aligned} \quad (9)$$

h_A^l and $\delta h_{CN_A}^l$ are both fitted constants. EN_A is the electronegativity of the element of atom A.

$$X(EN_A, EN_B) = 1 + k_{EN} \Delta EN_{AB}^2 \quad (10)$$

$$k_{EN} = 0.02 \text{ in GFN2} \quad (11)$$

$$\Delta EN_{AB}^2 = (EN_A - EN_B)^2 \quad (12)$$

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

$$\Pi(R_{AB}, l, l') = \left(1 + k_{A,l}^{\text{poly}} \left(\frac{R_{AB}}{R_{\text{cov},AB}} \right)^{\frac{1}{2}} \right) \left(1 + k_{B,l'}^{\text{poly}} \left(\frac{R_{AB}}{R_{\text{cov},AB}} \right)^{\frac{1}{2}} \right) \quad (13)$$

$R_{\text{cov},AB}$ are the summed covalent radii ($R_{\text{cov},A} + R_{\text{cov},B}$), e.g. $R_{\text{cov},H} = 0.32$, $R_{\text{cov},C} = 0.75$. $k_{A,l}^{\text{poly}}$ and $k_{B,l'}^{\text{poly}}$ are element and shell specific constants.

$$Y(\zeta_l^A, \zeta_{l'}^B) = \left(\frac{2\sqrt{\zeta_l^A \zeta_{l'}^B}}{\zeta_l^A + \zeta_{l'}^B} \right)^{\frac{1}{2}} \quad (14)$$

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

$$\chi_{\zeta,n,l,m}(r, \theta, \varphi) = N Y_{l,m}(\theta, \varphi) r^{n-1} e^{-\zeta r} \quad (15)$$

N is a normalisation constant, Y are spherical harmonic funtions, 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.

1.3 Isotropic electrostatic and XC energy

1.3.1 Second order

$$E_\gamma = \frac{1}{2} \sum_{A,B} \sum_{l \in A} \sum_{l' \in B} q_l q_{l'} \gamma_{AB, ll'} \quad (16)$$

$$\gamma_{AB, ll'} = \frac{1}{\sqrt{R_{AB}^2 + \eta_{AB, ll'}^{-2}}} \quad (17)$$

$$\eta_{AB, ll'} = \frac{1}{2} \left[\eta_A (1 + k_A^l) + \eta_B (1 + k_B^{l'}) \right] \quad (18)$$

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 $l = 0$).

$$GAP_A = \sum_{l \in A} q_l \quad (19)$$

$$q_l = \sum_{l' \in B} P_{ll'} S_{ll'} = GOP_l \quad (20)$$

1.3.2 Third order

$$E_\Gamma^{GFN2} = \frac{1}{3} \sum_A \sum_{l \in A} (q_l)^3 K_l^\Gamma \Gamma_A \quad (21)$$

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

$$\begin{aligned} E_{AES} &= E_{q\mu} + E_{q\Theta} + E_{\mu\mu} \\ &= \frac{1}{2} \sum_{A,B} \{ f_3(R_{AB}) [q_A (\boldsymbol{\mu}_B^T \mathbf{R}_{BA}) + q_B (\boldsymbol{\mu}_A^T \mathbf{R}_{AB})] \\ &\quad + f_5(R_{AB}) [q_A \mathbf{R}_{AB}^T \boldsymbol{\Theta}_B \mathbf{R}_{AB} + q_B \mathbf{R}_{AB}^T \boldsymbol{\Theta}_A \mathbf{R}_{AB} \\ &\quad - 3(\boldsymbol{\mu}_A^T \mathbf{R}_{AB})(\boldsymbol{\mu}_B^T \mathbf{R}_{AB}) + (\boldsymbol{\mu}_A^T \boldsymbol{\mu}_B) R_{AB}^2] \} \end{aligned} \quad (22)$$

$\boldsymbol{\mu}_A$ is the cumulative atomic dipole moment of atom A and $\boldsymbol{\Theta}_A$ is the corresponding traceless quadrupole moment. Traceless simply means that the sum of the diagonal elements is 0. 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.

$$\Theta_A^{\alpha\beta} = \frac{3}{2}\theta_A^{\alpha\beta} - \frac{\delta_{\alpha\beta}}{2}(\theta_A^{xx} + \theta_A^{yy} + \theta_A^{zz}) \quad (23)$$

$$\theta_A^{\alpha\beta} = \sum_{l' \in A} \sum_l P_l \left(\alpha_A D_{ll'}^\beta + \beta_A D_{ll'}^\alpha - \alpha_A \beta_A S_{ll'} - Q_{ll'}^{\alpha\beta} \right) \quad (24)$$

$$q_A = Z_A - GAP_A \quad (25)$$

$$\mu_A^\alpha = \sum_{l' \in A} \sum_l P_{l'l} (\alpha_A S_{l'l} - D_{l'l}^\alpha) \quad (26)$$

$$D_{ll'}^\alpha = \langle \phi_l | \alpha_i | \phi_{l'} \rangle \quad (27)$$

$$Q_{ll'}^{\alpha\beta} = \langle \phi_l | \alpha_i \beta_i | \phi_{l'} \rangle \quad (28)$$

α and β are Cartesian components labeled $(x, y, z)^T$ with atom A being centered in $v_A = (x_i, y_i, z_i)^T$ where i is a form of pointer/label dereferencing. $\delta_{\alpha\beta}$ is just the delta function, i.e. is 1 if α and β are the same label and 0 otherwise, this serves to include the term only for the diagonal.

$$\Theta_A = \begin{pmatrix} \Theta_A^{xx} & \Theta_A^{xy} & \Theta_A^{xz} \\ \Theta_A^{yx} & \Theta_A^{yy} & \Theta_A^{yz} \\ \Theta_A^{zx} & \Theta_A^{zy} & \Theta_A^{zz} \end{pmatrix} \quad (29)$$

$$\mu_A = \begin{pmatrix} \mu_A^x \\ \mu_A^y \\ \mu_A^z \end{pmatrix} \quad (30)$$

$$\mathbf{R}_{AB} = v_A - v_B \quad (31)$$

$$f_n(R_{AB}) = \frac{f_{damp}(a_n, R_{AB})}{R_{AB}^n} = \frac{1}{R_{AB}^n} \frac{1}{1 + 6 \left(\frac{R_0^{AB}}{R_{AB}} \right)^{a_n}} \quad (32)$$

$$R_0^{AB} = 0.5(R_0^{A'} + R_0^{B'}) \quad (33)$$

$$R_0^{A'} = \begin{cases} R_0^A + \frac{R_{max} - R_0^A}{1 + \exp[-4(CN_A' - N_{val} - \Delta_{val})]} & \text{if } N_{val} \text{ is given} \\ 5.0 \text{ bohrs} & \text{otherwise} \end{cases} \quad (34)$$

$$R_{max} = 5.0 \text{ bohrs} \quad (35)$$

$$\Delta_{val} = 1.2 \quad (36)$$

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

$$E_{AXC} = \sum_A (f_{XC}^{\mu_A} |\mu_A|^2 + f_{XC}^{\Theta_A} \|\Theta_A\|^2) \quad (37)$$

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

1.6 Dispersion Energy

$$\begin{aligned}
E_{disp}^{D4'} = & - \sum_{A>B} \sum_{n=6,8} s_n \frac{C_n^{AB}(q_A, CN_{cov}^A, q_B, CN_{cov}^B)}{R_{AB}^n} f_{damp,BJ}^{(n)}(R_{AB}) \\
& - s_9 \sum_{A>B>C} \frac{(3\cos(\theta_{ABC})\cos(\theta_{BCA})\cos(\theta_{CAB}) + 1)C_9^{ABC}(CN_{cov}^A, CN_{cov}^B, CN_{cov}^C)}{(R_{AB}R_{AC}R_{BC})^3} \\
& \times f_{damp,zero}^{(9)}(R_{AB}, R_{AC}, R_{BC}).
\end{aligned} \tag{38}$$

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.

$$C_6^{AB} = \frac{3}{\pi} \sum_j w_j \bar{\alpha}_A(i\omega_j, q_A, CN_{cov}^A) \bar{\alpha}_B(i\omega_j, q_B, CN_{cov}^B) \tag{39}$$

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.

$$\bar{\alpha}_A(i\omega_j, q_A, CN_{cov}^A) = \sum_r^{N_{A,ref}} \xi_A^r(q_A, q_{A,r}) \bar{\alpha}_{A,r}(i\omega_j, q_{A,r}, CN_{cov}^{A,r}) W_A^r(CN_{cov}^A, CN_{cov}^{A,r}) \tag{40}$$

The Gaussian weighting for each reference system is given by:

$$W_A^r(CN_{cov}^A, CN_{cov}^{A,r}) = \sum_{j=1}^{N_{gauss}} \frac{1}{\mathcal{N}} \exp[-6j \cdot (CN_{cov}^A - CN_{cov}^{A,r})^2] \tag{41}$$

with

$$\sum_r^{N_{A,ref}} W_A^r(CN_{cov}^A, CN_{cov}^{A,r}) = 1 \tag{42}$$

\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 .

$$\xi_A^r(q_A, q_{A,r}) = \exp \left[3 \left\{ 1 - \exp \left[4\eta_A \left(1 - \frac{Z_A^{eff} + q_{A,r}}{Z_A^{eff} + q_A} \right) \right] \right\} \right] \quad (43)$$

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.

$$C_8^{AB} = 3C_6^{AB} \sqrt{Q^A Q^B} \quad (44)$$

$$Q^A = s_{42} \sqrt{Z^A} \frac{\langle r^4 \rangle^A}{\langle r^2 \rangle^A} \quad (45)$$

$\sqrt{Z^A}$ is the ad hoc nuclear charge dependent factor.

From the original xTB program we can see that s_{42} is 0.5, and Z^A is the atomic number of A.

$$\sqrt{0.5 * \left(\frac{r^4}{r^2} * \sqrt{Z^A} \right)} \quad (46)$$

$\langle r^4 \rangle$ 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} ???)

CN_{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 constant¹:

$$C_9^{ABC} = \frac{3}{\pi} \int_0^\infty \alpha^A(i\omega) \alpha^B(i\omega) \alpha^C(i\omega) d\omega \quad (47)$$

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 as¹:

¹https://www.researchgate.net/publication/43347348_A-Consistent_and_Accurate_Ab-Initio_Parametrization_of_Density_Functi
D_for_the_94_Elements_H-Pu

$$C_9^{ABC} \approx -\sqrt{C_6^{AB}C_6^{AC}C_6^{BC}} \quad (48)$$

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.

BJ = Becke-Johnson

$$f_n^{damp,BJ}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + (a_1 \cdot R_{AB}^{crit} + a_2)^6} \quad (49)$$

$$R_{AB}^{crit} = \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} \quad (50)$$

$$f_9^{damp,zero}(R_{AB}, R_{AC}, R_{BC}) = \left(1 + 6 \left(\sqrt{\frac{R_{AB}^{crit} R_{BC}^{crit} R_{CA}^{crit}}{R_{AB} R_{BC} R_{CA}}} \right)^{16} \right)^{-1} \quad (51)$$

1.7 SAD - Superposition of Atomic Densities

The superposition of atomic densities(SAD) is an approach to obtain a good approximation of a collection of atoms, to be used as an initial guess for solving the self-consistent field(SCF) equation.

As originally implemented in DISCO, the molecular electron density can be obtained by adding the densities of all the constituting atoms.

This is how we get the density matrix for an isolated atom? equation 15 from: (<https://sci-hub.box/10.1002/jcc.540030314>)

$$D_{ij} = \sum_a^{occ} c_{ia} c_{ja} \quad (52)$$

To get the coefficients we need to solve SCF for each atom? this is supposedly cheap, but idk how to do it. (<https://sci-hub.box/10.1002/jcc.20393>) Though the math for Direct SCF Approach is given in this paper at equation 10: (<https://sci-hub.box/10.1002/jcc.540030314>). This is probably how.

The SAD method is then the sum of all of these?

Equation 3 in the GFN2 paper talks about "superposition of (neutral) atomic reference densities". Is this relevant?

Direct SCF Approach

$$\begin{aligned}
\Delta F_{ab} = & (c_{ia}c_{jb} + c_{ja}c_{ib}) \\
& \Delta F_{ij} + (c_{ia}c_{kb} + c_{ka}c_{ib}) \\
& \Delta F_{ik} + (c_{ia}c_{lb} + c_{la}c_{ib}) \\
& \Delta F_{il} + (c_{ja}c_{kb} + c_{ka}c_{jb}) \\
& \Delta F_{jk} + (c_{ja}c_{lb} + c_{la}c_{jb}) \\
& \Delta F_{jl} + (c_{ka}c_{lb} + c_{la}c_{kb}) \Delta F_{kl} \\
& = l_{ijkl}(4E_{ij}^{ab}D_{kl} + 4D_{ij}E_{kl}^{ab} - E_{ik}^{ab}D_{jl} - D_{ik}E_{jl}^{ab} - E_{il}^{ab}D_{jk} - D_{il}E_{jk}^{ab})
\end{aligned} \tag{53}$$

where

$$E_{ij}^{ab} = c_{ia}c_{jb} + c_{ja}c_{ib} \tag{54}$$