Connectivity Verification: Mathematical Rules and Procedures

1 Notation

- Multiset of atoms: $A = \{a_1, \ldots, a_n\}$; element symbol of a_i is e_i.
- Heavy atoms: $H = \{i : e_i \neq \mathrm{H}\}$, $n_h = |H|$.
- Count of element E: $N_E = |\{i : e_i = E\}|$.
- Base valence $v_0(E)$ for $E \in \{\mathrm{H,C,N,O,F,Cl,Br,I,P,S,B,Si,Ge,As,Se,Te}\}$ as in Table 1.
- Net skeleton charge parameter (after salt handling): $Q \in \mathbb{Z}$.
- Degree/bond multiplicity on heavy pair (i, j): $m_{ij} \in \{0, 1, 2, 3\}$, symmetric, with $m_{ij} = 0$ if not used.
- Halogen set: $\mathcal{X} = \{\mathrm{F,Cl,Br,I}\}$.

E	H	C	N	O	F	Cl	Br	I	P	S	B	Si	Ge	As	Se	Te
$v_0(E)$	1	4	3	2	1	1	1	1	3	2	3	4	4	3	2	2

Table 1: Base valence used by the verifier.

2 DBE (Double Bond Equivalents)

Let

$$s \;=\; \sum_{E} N_E \,(v_0(E) - 2).$$

The verifier uses an integer, charge-aware DBE:

$$\mathrm{DBE} \;=\; 1 + \frac{s + Q}{2} \quad \text{interpreted as integer via } \; 1 + \left\lfloor \frac{s + Q}{2} \right\rfloor$$

(equivalently $1 + (s + Q)//2$).

3 Pair Capacities

Bond multiplicity cap $c(e_i, e_j) \in \{0, 1, 2, 3\}$:

$$c(e_i, e_j) = \begin{cases} 0, & e_i, e_j \in \mathcal{X} \\ 1, & |\{e_i, e_j\} \cap \mathcal{X}| = 1 \\ 1, & \{e_i, e_j\} = \{\mathrm{O,O}\} \\ 3, & \{e_i, e_j\} \in \{\{\mathrm{C,C}\}, \{\mathrm{C,N}\}, \{\mathrm{N,N}\}\} \\ 2, & \{e_i, e_j\} \in \{\{\mathrm{C,O}\}, \{\mathrm{N,O}\}, \{\mathrm{S,O}\}\} \\ 2, & \text{otherwise.} \end{cases}$$

The π capacity is $c_\pi(e_i, e_j) = \max\{0,\, c(e_i, e_j) - 1\}$.

4 Rung Weights and Rulesets

Define a rung value $\nu(E)$ from atomic mass $m(E)$ (in u):

$$\nu(E) = 3 + 13\frac{\ln\left(\frac{m(E)\cdot 931.49410242}{0.51099895}\right)}{\ln(105.6583755/0.51099895)}.$$

Gaussian kernel on $d = |\nu(e_i) - \nu(e_j)|$:

$$K(d) = W_0 e^{-(d/\sigma_0)^2} + W_1 e^{-(d-\Delta_1)^2/2\sigma_1^2}$$

with W_0=1.124462, W_1=1.551250, σ_0=0.9, σ_1=1.2, Δ_1=6.5.

For each pair (i, j) the σ and π weights are:

$$(w_{ij}^{\sigma}, w_{ij}^{\pi}) = \begin{cases} (K,\, \alpha_\pi K) & \text{RS1} \\ \left(K,\, \alpha_\pi\, \gamma_{ij} K\right) & \text{RS2, with } \gamma_{ij}\text{=1.35 if } \{e_i, e_j\}\text{=\{C,O\} else 1} \\ \left(K + \delta_{ij}^{\sigma},\, \alpha_\pi\,(K + \delta_{ij}^{\pi})\right) & \text{RS3} \end{cases}$$

where for RS3, letting $L(d) = e^{-(d/0.80)^2}$ and χ_{ij} true if $\{e_i, e_j\} \in \{\{\text{C,C}\}, \{\text{C,N}\}\}$,

$$\delta_{ij}^{\sigma} = \begin{cases} 0.20\, L(d) & \chi_{ij} \\ 0 & \text{else} \end{cases}, \qquad \delta_{ij}^{\pi} = \begin{cases} 0.10\, L(d) & \chi_{ij} \\ 0 & \text{else.} \end{cases}$$

5 Dynamic Valence Assignment (Heavy Atoms)

Let DBE be as above, and let $H = N_{\text{H}}$. Over heavy indices $i \in H$:

$$\sum_{i \in H} v_i \;=\; H + 2\,\text{DBE} + 2\,(n_h - 1).$$

Start with $v_i = v_0(e_i)$ and raise minimally by these promotions until the identity holds:

(a) N: $3 \to 4$ as needed,
(b) P: $3 \to 4$, then $4 \to 5$ if still needed,
(c) S: $2 \to 3$, then (only if large residual) up to 6 in unit steps.

6 Graph Construction

6.1 Phase 1: Core Tree + Halogens

1. Build candidate heavy–heavy pairs with caps $c(\cdot,\cdot)$ and weights w^{σ}, w^{π}.
2. Core set $H_{\text{core}} = \{i \in H : e_i \notin \mathcal{X}\}$. Run a maximum-weight spanning tree (Kruskal) on H_{core} using w^{σ}, under degree constraints $\deg(i) \le v_i$. If impossible, fail with `tree_disconnected`.
3. If DBE > 0, reserve ≤ 2 disjoint core–core edges with highest w^{π} (1 unit headroom on both ends).
4. Attach each halogen $h \in H \backslash H_{\text{core}}$ as a leaf to the best core neighbor by (available capacity, w^{σ}).

6.2 Phase 2: Spend DBE (Merged Greedy)

Let current multiplicities $m_{ij} = 1$ on tree edges and 0 elsewhere. Headroom per node i is $r_i = v_i - \deg(i) - \pi_i$ (with $\pi_i = \sum_j \max(0, m_{ij} - 1)$).

1. Maintain two streams:
   - extra-σ: non-tree pairs (i, j) with m_{ij}=0 (costs 1 headroom each end, adds one cycle),
   - π-tickets: on each used edge (i, j), up to $c_\pi(e_i, e_j)$ increments (costs 1 headroom each end, increases π).
2. While remaining DBE > 0: pick the feasible item with larger weight between next best extra-σ (w^σ) and next best π-ticket (w^π); apply it and update headrooms. If neither feasible exists, fail with `dbe_unspendable`.

7 Verification Conditions

Given multiplicities m_{ij} over heavy pairs:
(V1) (No H edges) $m_{ij} = 0$ if e_i=H or e_j=H.
(V2) (Caps) $m_{ij} \le c(e_i, e_j)$ for all i<j.
(V3) (Hydrogen slack identity) For each heavy i,

$$h_i = v_i - \deg(i) - \pi_i \ \ge 0, \quad \sum_{i \in H} h_i = N_{\mathrm{H}}.$$

(V4) (Connectivity) The heavy subgraph is connected if $n_h \ge 2$.
(V5) (DBE equality) Let $Y = \#\{(i, j) : m_{ij} \ge 1\}$ and $\Pi = \sum_{i<j} \max(0, m_{ij} - 1)$. Then

$$\Pi + (Y - n_h + 1) \ = \ \mathrm{DBE}.$$

8 Salt & Charge Handling (Deterministic Rules)

8.1 Parsing Charge

From a formula suffix (e.g. `+2`, `-`, `+++`), set initial Q accordingly.

8.2 Spectator Cations

Remove {Li,Na,K,Rb,Cs,Ca,Mg} from A and *add their positive charge* to Q:

$$Q \leftarrow Q + N_{\mathrm{Li}} + N_{\mathrm{Na}} + \cdots + 2N_{\mathrm{Ca}} + 2N_{\mathrm{Mg}}.$$

8.3 Transition/Rare Metals

Remove all TMs from A; estimate counter-anion units needed per metal (typ. 2 for Pt/Pd/Ni/Fe/Co/Ru/Rh/Os/Ir; 1 for Au/Ag/Cu), and *remove* that many halides (priority I>Br>Cl>F) or common anions (BF_4^-, NO_3^-, PF_6^-, AsF_6^-, SbF_6^-, ClO_4^-, ReO_4^-, $CF_3SO_3^-$, HPO_4^{2-}, SO_4^{2-}, $H_2PO_4^-$, HSO_4^-, OH^-), adding their charge to Q. If insufficient anions were found, carry the residual as charge in Q.

8.4 Hydrogen-Balance Fix

On the cleaned core (no spectator cations/TMs; halogens retained), recompute DBE with current Q and assign v_i. Compute

$$H_{\text{need}} = \sum_{i \in H} v_i - 2\,\text{DBE} - 2\,(n_h - 1),$$

then add/remove hydrogens to A so that $N_{\text{H}} = H_{\text{need}}$.
Parity nudge: If $H_{\text{need}} - N_{\text{H}}$ has nonzero parity (or small magnitude), adjust Q by ± 1 or $\pm 2k$ (bounded) to match parity before final H-fix.

8.5 Early Inorganic/Salt Accept (DBE=0)

Accept immediately (skip graphing) as a salt with DBE=0 if any of the following holds:

- no carbon; or non-halogen non-metal organic core size ≤ 2;
- small CHO alkali salt: only {C,H,O} plus alkali/alkaline, no halogens, no TMs;
- polyoxo-alkali: ≥ 1 alkali and many O with P/S present; or ≥ 2 alkali with moderately many O and P/S present;
- very halide-rich ion pair: ≥ 6 total halides and no O/P/S;
- aurate-like salts: Au present with alkali and at least 4 of O+S.

9 Determinism

All sorts are by weight with stable tie-breaking on (w, i, j) lexicographic order. Reservations consume headroom one unit at endpoints but do not add m_{ij}.

10 Parameters

Unless specified: RS3, α_π=1.15, reservation ≤ 2, kernel constants as above.