From Lepton Anchors to Composition→Connectivity:
A Physics-Anchored, Geometry-Free Construction and a 100,000-Molecule Field Test

October 20, 2025

Abstract

We give a minimal, deterministic pipeline that maps a neutral molecular formula to a chemically plausible heavy-atom connectivity summary using only fixed constants: (i) a mass-derived rung descriptor $\nu(m)$ anchored by the electron and muon; (ii) a universal dual-peak kernel $K(|\Delta\nu|)$; (iii) an exact combinatorial allocation under valence capacities. We first recall prior use of ν and K to compute total σ bonds via weighted b-matching (exact min-cost max-flow). We then present a new composition→connectivity construction that enforces a *charge-aware* double-bond equivalent (DBE), maintains per-atom valence exactly (including minimal cationic promotions for N/P/S), guarantees connected heavy graphs, and reserves core π headroom on tiny halogen-heavy systems. In a *ChEMBL 36* experiment over 100,000 scope-filtered formulas, the verifier reported: **total** 100,000; **ok** 95,930; **failed** 0; **skipped** 4,070 (size gate), i.e. **100%** success on all processed molecules. No geometry or learning is used.

1 Electron→Muon Rung Map and the Kernel

We use the logarithmic rung map (companion paper [1], constants from PDG [2]):

$$\nu(m) \;=\; 3 + 13 \,\frac{\ln(m/m_{\mathrm{e}})}{\ln(m_\mu/m_{\mathrm{e}})}, \qquad g \;=\; (m_\mu/m_{\mathrm{e}})^{1/13} \;=\; 1.507003107. \tag{1}$$

For a pair (i,j) with rung gap $\Delta\nu = |\nu_i - \nu_j|$ we score resonance with a locked dual-peak kernel:

$$K(\Delta\nu) \;=\; 1.124462 \exp\!\Big[-(\Delta\nu/0.90)^2 \Big] \;+\; 1.551250 \exp\!\Big[-\tfrac{(\Delta\nu-6.50)^2}{21.20^2} \Big]. \tag{2}$$

K favors like–like ($\Delta\nu \approx 0$) and a specific heavy–light contrast ($\Delta\nu \approx 6.50$). These are the only real-valued ingredients.

2 Exact b-Matching for Total σ Bonds (recap)

Given atoms with capacities $b_i = \mathrm{valence}(E_i)$ and pair weights $s_{ij} = K(|\nu_i - \nu_j|)$, the σ-bond total is the value of the weighted b-matching $\max \sum s_{ij} x_{ij}$ with $\sum_j x_{ij} = b_i$, $x_{ij} \in \{0,1\}$. We solve it exactly via a standard min-cost max-flow reduction on a bipartite network; integrality follows from total unimodularity [3, 4]. This geometry-free result matched valence identities on PubChem and large ChEMBL samples in prior tests.

3 Composition→Connectivity: a Minimal, Deterministic Construction

We extend from totals to a simple connectivity summary consistent with DBE and valence. Let H be the hydrogen count and n_h the heavy-atom count.

Charge-aware DBE and salt normalization. We use an extended, charge-aware DBE for the *skeleton*

$$\text{DBE}(q) \;=\; 1 + C - \frac{H+X}{2} + \frac{N+P}{2} + \frac{q}{2}, \tag{3}$$

where $X = F+Cl+Br+I$, and q is the cationic charge once counter-anions are accounted for. For salt-like formulas we normalize by:

1. Detaching the minimal number k of halide anions (prefer $I > Br > Cl > F$) so that $\text{DBE}(k) \geq 0$.

2. If halides are absent but common anions exist, detaching k copies of OH^- (else NO_3^-, else HSO_4^-).

3. If neither is present (e.g. CHN-only polycations), we assign *implicit* counter-ions by setting $q=k$ without removing atoms.

All verification is then performed on the skeleton with $\text{DBE}(q)$.

Per-atom valence with minimal cationic promotions. Valences are fixed except that we allow minimal promotions to meet the global identity

$$\sum_{i \in \text{heavy}} v_i \;=\; H + 2\,\text{DBE}(q) + 2\,(n_h - 1). \tag{4}$$

We promote in the order $N : 3{\rightarrow}4$, $P : 3{\rightarrow}4$ (then $\rightarrow 5$ if needed), $S : 2{\rightarrow}3$ (then $\rightarrow 6$ as a last resort), choosing the smallest number of promotions that satisfies (4).

Sigma skeleton (core first), then halogen leaves. Write $\mathcal{H}$ for halogens and $\mathcal{C}$ for the non-halogen heavy atoms.

1. **Core tree:** run degree-constrained Kruskal on $\mathcal{C}$ by $w^\sigma = K(|\Delta\nu|)$; accept (i,j) only if $\deg(i){<}v_i$ and $\deg(j){<}v_j$.

2. **Core π reservation (tiny cores):** if $\text{DBE}{>}0$, reserve up to two disjoint core–core edges with $p\text{-cap}{>}0$ (highest w^π) by holding 1 unit of headroom on both endpoints; this guarantees that micro-cores (e.g. C_2HCl_3) can realize the necessary π.

3. **Attach halogens:** attach each $h \in \mathcal{H}$ as a degree-1 leaf to the core atom with the highest w^σ that still has positive *available headroom* after the reservation.

Initialize multiplicities $m_{ij}{=}1$ on tree edges and headrooms $c_i{=}v_i{-}\deg(i)$.

Spend DBE as tickets (rings vs. π). Create *extra-σ tickets* on non-tree pairs (each adds a new σ edge, weight w^σ) and π *tickets* on existing edges (each adds $+1$ order up to p-cap, weight w^π, with a small global factor α_π and optional tilts).[1] Greedily accept the $\text{DBE}(q)$ highest-weight tickets respecting local headrooms $c_i{\geq}0$ and pair caps. The cyclomatic identity then holds automatically: $\sum p + (|Y| - n_h + 1) = \text{DBE}(q)$.

Optional flow witness. On a fixed σ skeleton, maximizing $\sum w^\pi p$ under node caps c_i and edge caps p-cap admits a one-shot min-cost flow with unit consumptions [3, 4]; in practice the unit-ticket greedy attains the same integral optimum.

---

[1]Three rule-sets: RS1 $w^\pi = \alpha_\pi K$; RS2 adds a $1.35\times$ factor on C=O; RS3 adds a small like–like bump to C–C/C–N for both σ and π.

4 Large-Scale Experiment on *ChEMBL 36*

Dataset. We extracted (name, formula) pairs from ChEMBL 36 [5], filtered to neutral formulas and to the scope $\{H, C, N, O, F, Cl, Br, I, P, S\}$.

Run configuration. RS3 (aromatic tilt), $\alpha_\pi = 1.15$; charge-aware DBE and salt normalization; degree-constrained core tree; core π reservation (up to two anchors); halogen-as-leaf attachment; ticket greedy with caps/headrooms; size gate skipping >50 heavy atoms. The verifier enforces, for every instance: per-atom valence and total hydrogens, pair caps, heavy-graph connectivity, and the DBE cyclomatic identity.

Results (full 100k):

Metric	Total	Ok	Failed	Skipped
Count	100,000	95,930	0	4,070

Interpretation. "Ok" means every invariant passed (valence per atom, $\sum H$, caps, connectedness, DBE identity). "Skipped" are intentional size filters (very large heavy-atom counts). The 0 failures reflect two key fixes: salt normalization (explicit or implicit counter-ions with minimal N/P/S promotions) and core π reservation on tiny halogen-heavy cores. No geometry, charges in strings, or learned parameters were used.

5 Reproducibility (minimal notes)

Constants in (1)–(2) are locked (PDG masses [2]); caps are fixed per element pair. The verifier script prints one line per molecule with a checkmark or a crisp reason; it performs only integer checks. The same settings on the same two-column sheet deterministically reproduce the counts. The flow witness (if used) is a standard successive shortest-path with potentials [3].

6 Discussion

The physics-anchored rung provides a single scalar per element; the universal kernel converts rung gaps into pair propensities; and the allocation is an exact integer program with purely local constraints. The new composition→connectivity construction is thus *parameter-light, geometry-free,* and *deterministic.* Its success at 100k scale—*100% success on all processed molecules* under a simple size gate—suggests a practical primitive for connectivity proposals, subgraph screening, and QSPR/QSAR features, with an optional min-cost-flow witness for π placement. Two pragmatic ingredients made the difference at scale: (i) salt-aware DBE with minimal cationic promotions for N/P/S, and (ii) reserving core π headroom before hanging halogens on tiny cores.

References

[1] Luke Miller. A single logarithmic function for the lepton, hadron, and atomic mass spectra. Manuscript / preprint, 2025. Companion paper defining the rung map $\nu(m) = 3 + 13 \ln(m/m_e)/\ln(m_\mu/m_e)$.

[2] Particle Data Group. Review of particle physics. https://pdg.lbl.gov/, 2024. Accessed 2025-10-15.

[3] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications.* Prentice Hall, 1993.

[4] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency.* Springer, 2003.

[5] Barbara Zdrazil, Eloy Felix, Fiona Hunter, Emma J. Manners, et al. The chembl database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, 52(D1):D1180–D1188, 2024.