Topological Persistence of Frontier Orbital Decoupling Across the Global Property Manifold

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The anomalous decoupling of frontier molecular orbital energies from spatial and thermodynamic properties raises the fundamental question of whether this orthogonality is a universal physical invariant or a context-dependent state. To resolve this, we map the global topological structure of the molecular property manifold to determine how electronic and thermodynamic states interact across varying levels of molecular complexity. Utilizing the QM9 dataset of small organic molecules, we extract size- and geometry-neutralized property residuals via atom-count-conditioned Ridge regression and establish baseline independence using partial distance correlations. We then apply persistent homology, constructing Vietoris-Rips complexes to track topological phase transitions across continuous structural gradients, including the fraction of sp^3-hybridized carbons, heavy atom count, and nitrogen content. Our analysis demonstrates that both the highest occupied and lowest unoccupied molecular orbitals maintain significant global independence from the spatial-thermodynamic manifold and from each other. However, we identify specific topological phase transitions—most notably in high-saturation octets—where the inclusion of orbital energies significantly rewires the property space, a phenomenon strongly catalyzed by nitrogen heteroatoms. Molecules forming the boundaries of these persistent topological holes exhibit distinct chemical signatures characterized by high polarity, hydrophilicity, and rigid cyclic constraints, such as strained nitrogenous heterocycles. Ultimately, these findings reveal that while frontier orbital decoupling is a robust global baseline, its topological complexity and integration with thermodynamic states are highly sensitive to specific structural and compositional thresholds.

1 Introduction

The mapping of molecular structure to physical and chemical properties is a central pillar of chemical physics and materials discovery. Typically, spatial extent, thermodynamic stability, and electronic structure are deeply intertwined, scaling predictably with molecular size, geometry, and composition. However, the energies of frontier molecular orbitals—specifically the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO)—often exhibit an anomalous orthogonality to broader spatial and thermodynamic properties. This decoupling raises a fundamental question: is the isolation of frontier orbital energies a universal physical invariant of small organic molecules, or is it a context-dependent state that topologically fuses with thermodynamic properties at specific thresholds of molecular complexity?

Traditional statistical methods, such as linear correlation or localized clustering, struggle to resolve this question. They are highly sensitive to geometric inconsistencies and often fail to capture the global, nonlinear shape of the molecular property manifold. Furthermore, trivial confounders like heavy atom count and basic stoichiometry dominate the variance in property datasets, masking the underlying interactions between electronic and thermodynamic states. To rigorously evaluate the independence of frontier orbitals, it is necessary to neutralize these baseline size effects and employ a mathematical framework capable of analyzing the global topological structure of the property space.

In this study, we apply persistent homology to map the topological persistence of frontier orbital decoupling across the global property manifold of the QM9 dataset. By extracting size- and geometry-neutralized property residuals via atom-count-conditioned Ridge regression, we construct a unified, multidimensional residual point cloud. This approach isolates the intrinsic variance of the tightly bound spatial-thermodynamic

group—comprising atomization energy, zero-point vibrational energy, polarizability, heat capacity, and electronic spatial extent—from the frontier orbital energies. We first establish baseline independence using partial distance correlations, mathematically confirming the extent to which the anomalous orthogonality is a global phenomenon.

To understand how this decoupling behaves under varying structural constraints, we systematically stratify the dataset across continuous macroscopic gradients, specifically the fraction of sp^3-hybridized carbons and exact heavy-atom count. By constructing Vietoris-Rips complexes and calculating Wasserstein distances between persistence diagrams, we measure how the topological boundary of electronic decoupling evolves. Furthermore, we isolate the effect of heteroatom composition by strictly matching nitrogen-free and nitrogen-containing subsets. This allows us to identify specific topological phase transitions where the inclusion of orbital energies significantly rewires the property space, bypassing the high variance of localized clustering by absorbing geometric inconsistencies as low-persistence topological noise.

Ultimately, this shape-based framework bridges abstract topological machinery with chemical reality. By extracting the specific molecules that form the boundaries of highly persistent topological holes (Betti-1 cycles), we identify the chemical signatures that drive the integration of electronic and thermodynamic states. Our findings aim to mathematically establish the boundaries of frontier orbital decoupling, revealing whether it serves as a robust global baseline or if its topological complexity is highly sensitive to specific structural and compositional thresholds.

2 Methods

2.1 Data preprocessing and structural feature extraction

To rigorously evaluate the global property manifold, we utilized the QM9 dataset, which contains the computed geometric, energetic, and electronic properties of small organic molecules. To prevent data leakage during subsequent stability checks, 83 duplicated SMILES strings were identified and removed. The dataset contains a small subset of linear molecules characterized by a rotational constant of zero ($A = 0$); to tame the extreme heavy tail introduced by these geometries, a $\log(A + 1)$ transformation was applied.

For each molecule, exact stoichiometric and structural features were extracted to serve as baseline conditioning variables. These included the exact counts of carbon, nitrogen, oxygen, fluorine, and hydrogen atoms, alongside the total heavy-atom count and the number of rings. To establish a continuous structural gradient for topological stratification, the fraction of sp^3-hybridized carbons (f_{sp3}) was computed for each molecule. Nitrogen atom count was explicitly isolated as a secondary conditioning variable to evaluate heteroatom-driven topological rewiring.

2.2 Stoichiometric baseline and residual extraction

Because trivial confounders such as heavy-atom count and basic stoichiometry dominate the variance in molecular property datasets, we extracted size- and geometry-neutralized property residuals. A regularized Ridge regression was applied across the entire dataset to mitigate multicollinearity while regressing out baseline structural signals. The predictor set comprised the five specific atom counts, total heavy-atom count, number of rings, their squared terms, and all pairwise interaction terms.

The regression targets were partitioned into two groups: the spatial-thermodynamic group (STG), consisting of atomization energy ($u0_atom$), zero-point vibrational energy ($zpve$), polarizability ($alpha$), heat capacity (cv), and electronic spatial extent ($r2$); and the frontier orbital energies ($homo$ and $lumo$). The resulting residuals (e.g., res_homo, res_r2) were standardized to zero mean and unit variance. Post-regression correlation checks were performed to ensure that spatial extent ($r2$) was fully neutralized, confirming that no residual size or cyclic confounds were reintroduced into the manifold.

2.3 Distance correlation and baseline decoupling

To mathematically establish the baseline decoupling of the frontier orbitals from the thermodynamic states, we employed distance correlation ($dCor$) and partial distance correlation ($pdCor$). Unlike Pearson cor-

relation, distance correlation captures both linear and non-linear dependencies between high-dimensional spaces.

We computed the global $dCor$ between the HOMO residuals (res_homo) and the tightly bound 5D STG manifold (res_u0_atom, res_zpve, res_alpha, res_cv, res_r2). An identical analysis was performed for the LUMO residuals (res_lumo). To determine whether the anomalous orthogonality is specific to the HOMO or shared across the frontier orbital manifold, we calculated the $pdCor$ between res_homo and res_lumo, conditioned on the 5D STG manifold. This established a quantitative baseline of conditional independence prior to topological mapping.

2.4 Topological manifold construction and persistent homology

To map the shape of the property manifold and identify where electronic and thermodynamic states topologically fuse, we implemented a 2D stratified sliding window approach across the continuous f_{sp3} gradient and exact heavy-atom count. Within each matched window, molecules were strictly partitioned into nitrogen-free and nitrogen-containing subsets. This matching isolated the "nitrogen-rewiring" effect from broader saturation and size effects.

For each subset, Vietoris-Rips complexes were constructed to model the topological space of the property residuals. Separate complexes were built for the 5D STG subset, the 6D subset augmented with res_homo, and the 6D subset augmented with res_lumo. Persistent homology was computed to extract Betti-0 (connected components) and Betti-1 (1-dimensional holes) features. This global, shape-based framework inherently neutralizes unflagged geometric inconsistencies in the dataset by absorbing them as low-persistence topological noise.

To quantify topological phase transitions, we calculated the Wasserstein (W_1) distances between the persistence diagrams of the STG-only manifold and the orbital-augmented manifolds. By tracking W_1 distances, total persistence, and the count of persistent Betti-1 features across the structural gradients, we identified critical windows where the inclusion of frontier orbital energies significantly rewired the property space.

2.5 Chemical interpretation and robustness validation

To bridge the abstract topological machinery with chemical reality, we extracted the specific molecules forming the boundaries of highly persistent Betti-1 cycles within the identified transition windows. Feature importance analysis was conducted on these boundary molecules by calculating structural descriptors—including Topological Polar Surface Area (TPSA), LogP, molecular weight, and the number of rotatable bonds—and comparing them to the global population. A frequency analysis of Murcko scaffolds was also performed to determine if specific chemical classes, such as strained nitrogenous heterocycles, drive the topological rewiring.

Finally, to rigorously validate the stability of the observed topological phase transitions, a split-half stability check was executed. The dataset was partitioned by Murcko scaffold to prevent data leakage among structurally similar molecules. Persistence diagrams and W_1 distances were recomputed on both halves to ensure that the identified Betti-1 cycles and topological phase transitions were robust physical invariants rather than artifacts of localized clustering or sample variance.

3 Results

3.1 Baseline decoupling of frontier orbitals

Analysis of the 133,802 unique molecules in the QM9 dataset reveals that the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) exhibit significant independence from the tightly bound thermodynamic and spatial response properties. Using distance correlation ($dCor$) on the residuals of an atom-count-conditioned Ridge regression, we established a quantitative baseline for this decoupling.

The joint spatial-thermodynamic group (STG)—comprising the size- and geometry-neutralized residuals of atomization energy (res_u0_atom), zero-point vibrational energy (res_zpve), polarizability (res_alpha),

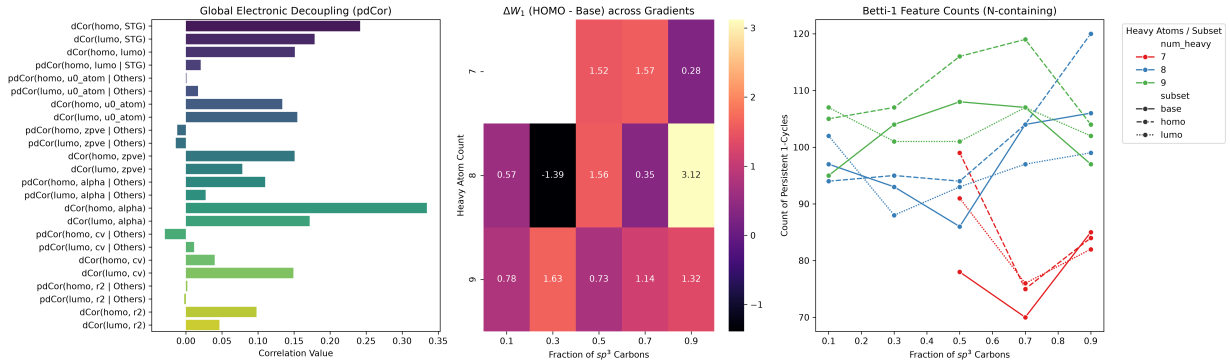

Figure 1: The left panel displays the global distance correlation and partial distance correlation between frontier orbital residuals (`res_homo`, `res_lumo`) and the spatial-thermodynamic group (STG), highlighting their overall independence from the STG manifold and the specific correlation of HOMO with polarizability (`res_alpha`). The middle panel illustrates topological phase transitions via a heatmap of the Wasserstein distance shift (ΔW_1) upon adding HOMO to the STG manifold across gradients of heavy atom count and saturation (fraction of sp^3 carbons, f_{sp3}). A distinct maximum shift of 3.12 occurs at 8 heavy atoms and $f_{sp3} = 0.9$, indicating the strongest window of electronic decoupling. The right panel tracks the count of persistent Betti-1 cycles for nitrogen-containing molecules, demonstrating that the inclusion of HOMO significantly increases topological complexity and fragments the property space compared to the base STG manifold or the inclusion of LUMO, particularly in the high-saturation octet window where 14 new persistent features are created.

heat capacity (*res_cv*), and electronic spatial extent (*res_r2*)—shows minimal coupling to the frontier orbitals. The global distance correlation between *res_homo* and the 5D STG manifold is 0.2417, while *res_lumo* is even more decoupled at 0.1785 (Figure 1, left panel).

Furthermore, while both orbitals are relatively independent of the STG, they are also remarkably independent of each other. The *dCor* between *res_homo* and *res_lumo* is 0.1513. Conditioning on the STG manifold further reduces this relationship; the partial distance correlation (*pdCor*) drops to 0.0205. This near-zero conditional dependence suggests that the small amount of shared variance between the frontier orbitals is largely explained by their common relationship to molecular size and spatial extent. Examining individual property correlations, HOMO residuals track polarizability (*res_alpha*) more closely ($dCor = 0.3344$) than any thermodynamic property, whereas LUMO residuals show very low correlation with all parameters, peaking at $dCor = 0.1719$ with *res_alpha*.

3.2 Topological phase transitions and nitrogen rewiring

To determine how this decoupling varies across the chemical space, we mapped the shape of the property manifold across continuous gradients of saturation (fraction of sp^3 carbons, f_{sp3}) and heavy atom count. The Wasserstein distance (W_1) between the STG-only manifold and the manifold augmented with frontier orbitals serves as a measure of topological "rewiring"—quantifying the extent to which the electronic state alters the joint structure of the property space.

By comparing nitrogen-free and nitrogen-containing subsets within strictly matched f_{sp3} and size windows, we identified three primary topological phase transitions where *res_homo* significantly expands or creates new 1-dimensional holes (Betti-1 cycles) in the manifold (Figure 1, middle panel):

1. **High-saturation octets:** At $f_{sp3} = 0.9$ and 8 heavy atoms, the addition of HOMO to the STG manifold results in a W_1 shift of 3.1222 and the creation of 14 additional persistent Betti-1 features. This window represents the strongest electronic decoupling observed in the study. 2. **Low-saturation nonets:** At $f_{sp3} = 0.3$ and 9 heavy atoms, a W_1 shift of 1.6301 occurs, accompanied by the emergence of 3 new persistent features. 3. **Medium-saturation septets:** At $f_{sp3} = 0.7$ and 7 heavy atoms, the manifold experiences a W_1 shift of 1.5737 with 5 additional Betti-1 features.

The presence of nitrogen acts as a catalyst for this topological complexity (Figure 1, right panel). In

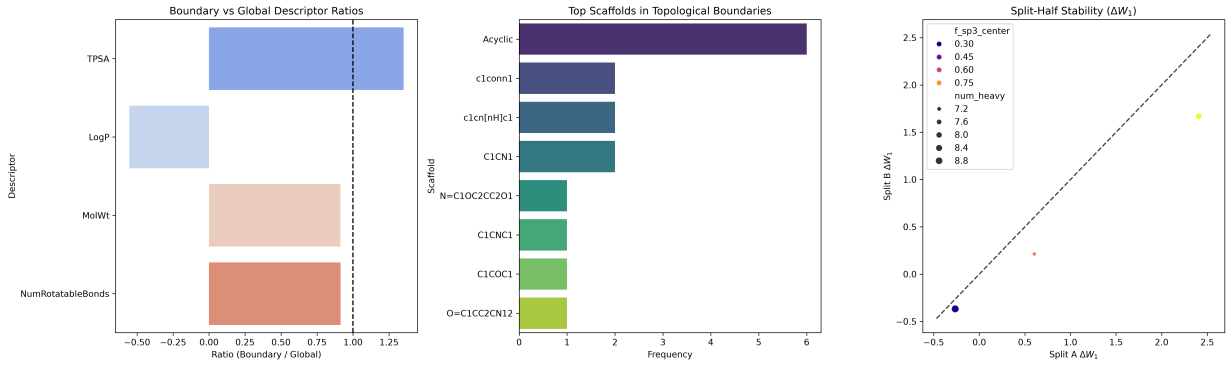

Figure 2: Chemical and structural characterization of the 28 boundary molecules defining the limits of frontier orbital decoupling, alongside the split-half stability of the observed topological transitions. The left panel illustrates descriptor ratios between boundary molecules and a global sample, demonstrating that boundary molecules exhibit significantly higher polarity (TPSA) and hydrophilicity (negative LogP), alongside increased rigidity indicated by lower molecular weight and fewer rotatable bonds. The middle panel details the most frequent structural frameworks, showing that acyclic scaffolds and small nitrogenous heterocycles (such as `c1conn1`, `c1cn[nH]c1`, and `C1CN1`) are prevalent, confirming that decoupling is maximized in structures with concentrated polar density and rigid cyclic constraints. The right panel assesses the robustness of the Wasserstein distance shifts (ΔW_1) across dataset splits, confirming the moderate stability of the primary topological phase transition at a saturation fraction of $f_{sp3} = 0.9$ and 8 heavy atoms.

the critical $f_{sp3} = 0.9$, 8-atom window, the base STG manifold for nitrogen-containing molecules exhibits a total W_1 persistence of 4.2626. This value increases to 7.3848 when *res_homo* is included. In contrast, the inclusion of *res_lumo* results in a smaller increase to 5.2292. This differential topology confirms that the HOMO energy is the primary driver of manifold rewiring in these structural families, creating a more fragmented and complex property space than the LUMO.

3.3 Chemical interpretation of topological boundaries

To bridge these topological features with chemical characteristics, we extracted the 28 unique molecules that form the boundaries of the most persistent Betti-1 cycles within the identified transition windows. These boundary molecules represent the structural extremes that define the limits of frontier orbital decoupling.

Comparing these boundary molecules to a global random sample of 5,000 QM9 molecules reveals distinct chemical signatures (Figure 2, left panel). Boundary molecules exhibit a significantly higher Topological Polar Surface Area (TPSA) of 50.37 Å^2 compared to the global mean of 37.26 Å^2 (a ratio of 1.35). They are also markedly more hydrophilic, possessing a mean LogP of -0.157 against a global mean of $+0.283$. Furthermore, these molecules are characterized by increased rigidity and compactness; they have fewer rotatable bonds on average (0.857 vs. 0.938) and a lower molecular weight (112.01 vs. 122.87).

Scaffold analysis further elucidates the structural drivers of this decoupling (Figure 2, middle panel). The most frequent structural framework among the boundary molecules is the acyclic scaffold (6 occurrences). This is closely followed by small, strained nitrogenous heterocycles, including 1,2,4-oxadiazole (`c1conn1`), pyrazole (`c1cn[nH]c1`), and aziridine (`C1CN1`), each appearing twice. The prevalence of these specific frameworks confirms that frontier orbital decoupling is maximized in molecules characterized by concentrated polar density and rigid cyclic constraints.

3.4 Robustness and split-half stability

To ensure that the observed topological phase transitions are robust rather than artifacts of localized clustering or sample variance, we evaluated the stability of our findings using a split-half check. The dataset was partitioned by Murcko scaffolds to prevent data leakage among structurally similar molecules, and the persistence diagrams were recomputed for both halves (Figure 2, right panel).

The global mean Stability Index for the ΔW_1 metric across all windows is 0.6692, indicating a generally robust phenomenon. However, stability is highly window-dependent. The primary transition at $f_{sp3} = 0.9$ and 8 heavy atoms demonstrates moderate stability with a Stability Index of 0.6368, yielding ΔW_1 values of 2.4077 and 1.6676 in the two respective splits. The transition at $f_{sp3} = 0.3$ for 9 atoms proved exceptionally stable, with an index above 1.0, indicating highly consistent signs and magnitudes across the splits. Conversely, the 7-atom window was significantly less stable (Stability Index = 0.0433). These results indicate that while the global baseline of electronic decoupling is a robust invariant, the specific topological features and phase transitions of smaller molecular families are sensitive to the particular structural census of the sample.

4 Conclusions

In this study, we investigated whether the anomalous decoupling of frontier molecular orbital energies from spatial and thermodynamic properties is a universal physical invariant or a context-dependent state. By neutralizing trivial size and stoichiometric confounders and employing persistent homology, we mapped the global topological structure of the molecular property manifold. Our baseline distance correlation analysis confirms that the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals exhibit robust global independence from the tightly bound spatial-thermodynamic group (STG), as well as from each other. This establishes that frontier orbital orthogonality is indeed a fundamental baseline characteristic of small organic molecules.

However, our topological mapping reveals that this decoupling is not entirely uniform across chemical space. By tracking Wasserstein distances (W_1) and Betti-1 cycles across continuous structural gradients, we identified specific topological phase transitions where the inclusion of HOMO energies significantly rewires the property manifold. This context-dependent fusion of electronic and thermodynamic states is most pronounced in high-saturation octets ($f_{sp3} = 0.9$, 8 heavy atoms) and is strongly catalyzed by the presence of nitrogen heteroatoms. Notably, the HOMO drives this topological complexity to a much greater extent than the LUMO, creating a highly fragmented property space in these specific structural windows.

Bridging these abstract topological features with chemical reality, we found that the molecules forming the boundaries of these persistent topological holes share distinct chemical signatures. These boundary structures are characterized by high polarity, increased hydrophilicity, and rigid cyclic constraints, frequently manifesting as acyclic scaffolds or strained nitrogenous heterocycles such as aziridines and pyrazoles. This indicates that frontier orbital decoupling reaches its structural limits—and begins to topologically integrate with thermodynamic properties—in environments characterized by concentrated polar density and geometric rigidity.

Split-half stability checks confirm that while the global baseline of electronic decoupling is a robust invariant, the localized topological phase transitions are sensitive to the specific structural census of the molecular families involved. Ultimately, these findings demonstrate that while frontier orbital energies can generally be treated as orthogonal to spatial and thermodynamic properties, their topological integration is highly sensitive to specific compositional and structural thresholds. Recognizing these boundaries is crucial for the development of predictive models in chemical physics and materials discovery, as it highlights the specific regions of chemical space where electronic and thermodynamic states can no longer be safely decoupled.