Local Sensitivity Mapping of HOMO Decoupling Across Continuous Topological Gradients

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The energy of the highest occupied molecular orbital (HOMO) is frequently observed to be orthogonal to macroscopic spatial and thermodynamic properties, yet it remains unknown whether this decoupling is a universal constant or a conditional state governed by molecular topology. To address this, we analyzed 133,802 molecules from the QM9 dataset to determine how electronic-thermodynamic integration shifts across continuous structural gradients. We employed gradient-boosted regressors to extract non-linear property residuals, eliminating stoichiometric determinism, and mapped the local structural neighborhood of each molecule using K-Nearest Neighbors across a topological space defined by sp3 carbon fraction and BertzCT complexity. By calculating partial distance correlations, we evaluated the local sensitivity between the frontier electronic state and the spatial-thermodynamic manifold while controlling for heteroatom confounding. Our results reveal a distinct topological phase transition where the electronic state suddenly recouples with macroscopic properties. Persistent homology analysis demonstrates that this recoupled regime exhibits a drastically simplified property manifold with a near-complete reduction in higher-order topological voids. Furthermore, motif analysis indicates that this recoupling is driven by increased carbonyl density and polar surface area, whereas rigid bicyclic strain favors decoupling. Ultimately, we demonstrate that the independence of the frontier orbital energy is a localized phenomenon in simple molecules that transitions into a highly integrated joint manifold as specific topological complexity increases.

1 Introduction

The frontier electronic structure of a molecule, particularly the energy of the highest occupied molecular orbital (HOMO), fundamentally dictates its chemical reactivity, stability, and interaction potential. However, in large-scale cheminformatic analyses, the HOMO energy is frequently observed to be orthogonal to macroscopic spatial and thermodynamic properties, such as zero-point vibrational energy, heat capacity, and spatial extent. This apparent decoupling raises a critical question: is the independence of the frontier orbital energy a universal constant of chemical space, or is it a conditional state governed by underlying molecular topology?

Investigating this relationship requires disentangling true structural effects from stoichiometric determinism. In comprehensive databases such as QM9, basic atom counts explain the vast majority of the variance in extensive thermodynamic properties. For instance, atomic composition alone can account for over 99% of the variance in total energies and zero-point vibrational energies. This overwhelming stoichiometric baseline masks the subtle, non-linear ways in which atomic arrangement and geometric strain influence electronic-thermodynamic integration. To expose the true joint structure of these properties, it is necessary to isolate the chemical variance—the property residuals—from the dominant atom-count effects.

In this work, we analyze 133,802 molecules from the QM9 dataset to determine how the coupling between the frontier electronic state and the spatial-thermodynamic manifold shifts across continuous structural gradients. By employing gradient-boosted regressors, we extract non-linear property residuals (including res_homo, res_zpve, res_c_v, and res_r^2) to eliminate stoichiometric confounding. We then map the local structural neighborhood of each molecule across a continuous topological space defined by the fraction of sp^3 hybridized carbons (F_{sp^3}) and BertzCT complexity.

To robustly quantify the integration between the HOMO energy and macroscopic properties, we compute the partial distance correlation within local K-Nearest Neighbor (KNN) neighborhoods, explicitly controlling for heteroatom confounding. This continuous formulation bypasses the instability of clustering discrete, low-population scaffolds, directly isolating the geometric conditions that govern electronic-thermodynamic integration. Furthermore, we apply persistent homology to mathematically characterize the shape and connectivity of the resulting 11-dimensional residual property manifold, tracking how its topological features evolve across different chemical regimes.

Our analysis reveals a distinct topological phase transition within chemical space. We demonstrate that the baseline decoupling of the HOMO energy is not universal, but rather a localized phenomenon characteristic of simple, non-polar molecules. As specific topological complexity increases—driven notably by carbonyl density and polar surface area—the electronic state suddenly recouples with the macroscopic property manifold. This recoupled regime exhibits a drastically simplified property manifold with a near-complete reduction in higher-order topological voids. Conversely, rigid bicyclic strain is shown to favor electronic decoupling. Ultimately, this work establishes that the joint structure of molecular properties is a highly conditional state dictated by continuous topological gradients, offering a new framework for understanding electronic-thermodynamic integration in chemical space.

2 Methods

2.1 Data curation and feature extraction

The analysis was conducted on the QM9 dataset, from which 133,802 unique molecules were retained following SMILES string deduplication. For each molecule, exact atom counts for carbon, nitrogen, oxygen, fluorine, and hydrogen were parsed. To map the chemical space along continuous axes of molecular topology, we computed two primary structural descriptors: the fraction of sp^3 hybridized carbons (F_{sp^3}) and the BertzCT complexity index. Additionally, perfectly linear molecules, defined by a rotational constant $A = 0$, were identified and isolated to serve as a boundary-condition contrast group in later stages of the analysis.

2.2 Non-linear residualization of the property manifold

As established previously, basic atom counts overwhelmingly dictate extensive thermodynamic properties, masking the subtle structural dependencies of the frontier orbitals. To eliminate this stoichiometric determinism, we isolated the chemical variance by training gradient-boosted regressors (LightGBM) on the exact atom counts to predict the dataset's non-redundant properties. Seven highly collinear thermodynamic properties (the HOMO-LUMO gap, u_{298}, h_{298}, g_{298}, u_{298_atom}, h_{298_atom}, and g_{298_atom}) were excluded to prevent artificial low-dimensionality. The residuals from these non-linear models—specifically res_homo, res_zpve, res_c_v, and res_r^2—were extracted. These robust residuals form an 11-dimensional manifold that represents true chemical and geometric organization independent of molecular size and composition.

2.3 Continuous topological mapping and local sensitivity analysis

To track the trajectory of electronic-thermodynamic coupling smoothly across the dataset, we mapped the local structural neighborhood of each molecule. The F_{sp^3} and BertzCT features were first standardized using Z-score normalization. We then defined the local structural neighborhood for each molecule using a K-Nearest Neighbors (KNN) query in this 2D topological space, setting $k = 2000$ to ensure statistical robustness.

Within each sliding analytical window, we evaluated the local sensitivity between the frontier electronic state and the spatial-thermodynamic manifold. This was quantified by computing the partial distance correlation ($dCor$) between the res_homo vector and the tightly bound macroscopic manifold (res_zpve, res_c_v, res_r^2). To ensure that the observed coupling was driven by underlying topology rather than simple heteroatom presence, we explicitly controlled for the exact counts of nitrogen and oxygen atoms within each neighborhood during the $dCor$ computation.

2.4 Artifact screening and geometric consistency

To guarantee that the identified recoupling zones were physical phase transitions and not artifacts of latent density functional theory (DFT) convergence failures, we evaluated the geometric strain of molecules in high-$dCor$ neighborhoods. Rapid 3D conformations were generated using RDKit, and the MMFF94 strain energy was computed. For each KNN neighborhood, we calculated the mean relative strain energy and the 3D embedding failure rate. These local artifact proxies were then correlated with the local $dCor$ values to verify that geometric inconsistencies were uniformly distributed and did not masquerade as structural tipping points.

2.5 Phase boundary extraction and split-half validation

The continuous $dCor$ map was analyzed to identify structural tipping points where the HOMO energy suddenly recouples with macroscopic properties. We defined strictly "decoupled" and "recoupled" regimes based on the bottom and top 15th percentiles of the $dCor$ distribution, respectively. To validate the stability of these phase boundaries, we performed a split-half validation by partitioning the dataset into two independent halves using a strict Murcko scaffold split. The KNN mapping and partial distance correlation computations were re-executed independently on both halves to verify that the continuous mapping yielded consistent high and low $dCor$ regions across disjoint chemical families.

2.6 Persistent homology and motif enrichment

To provide a rigorous mathematical description of the reorganization within the joint property structure, we performed a persistent homology analysis on the 11-dimensional residual property manifold using the Gudhi library. We subsampled 2,000 representative molecules from both the strictly recoupled and decoupled regimes using a stratified sampling strategy based on F_{sp^3} and BertzCT density. For each regime, a Vietoris-Rips complex was constructed from the residual space, and persistence barcodes (H0, H1, H2) were computed to quantify the topological shape and void structure of the data manifold.

Finally, we extracted the specific molecules located within the stable recoupled zones to identify shared chemical drivers. Kolmogorov-Smirnov (KS) statistics were employed to measure the enrichment of specific motifs—such as carbonyl density, polar surface area, and rigid bicyclic strain—that govern the recoupling independent of simple N/O counts. A formal contrast analysis was also performed on the isolated linear molecules, comparing their res_homo behavior against their nearest non-linear neighbors to establish boundary conditions for the formulated chemical rules.

3 Results

3.1 Isolation of chemical variance and local sensitivity mapping

As outlined in the methodology, basic atomic composition accounts for over 99.99% of the variance in total energies and 99.71% of the variance in zero-point vibrational energy (`zpve`) within the QM9 dataset. By extracting the residuals from gradient-boosted regressors trained on exact atom counts, we successfully stripped this stoichiometric determinism. The resulting 11-dimensional residual manifold isolates the true chemical variance—the component of each property dictated purely by the topological arrangement of atoms.

To determine whether the frontier electronic state is universally orthogonal to macroscopic properties, we evaluated the partial distance correlation ($dCor$) between the highest occupied molecular orbital residual (`res_homo`) and the spatial-thermodynamic manifold (`res_zpve`, `res_cv`, `res_r2`). By mapping this local sensitivity across a continuous topological space defined by the fraction of sp^3 carbons (F_{sp^3}) and BertzCT complexity, we observed a distinct topological phase transition.

As shown in Figures and , the coupling strength is not uniform. We identified two distinct regimes: a "Decoupled" regime (defined by the bottom 15th percentile, $dCor \leq 0.0097$) where the frontier electronic state is essentially orthogonal to the molecule's spatial and thermodynamic organization, and a "Recoupled" regime (above the 85th percentile, $dCor \geq 0.0924$) where `res_homo` strongly integrates with the macroscopic properties.

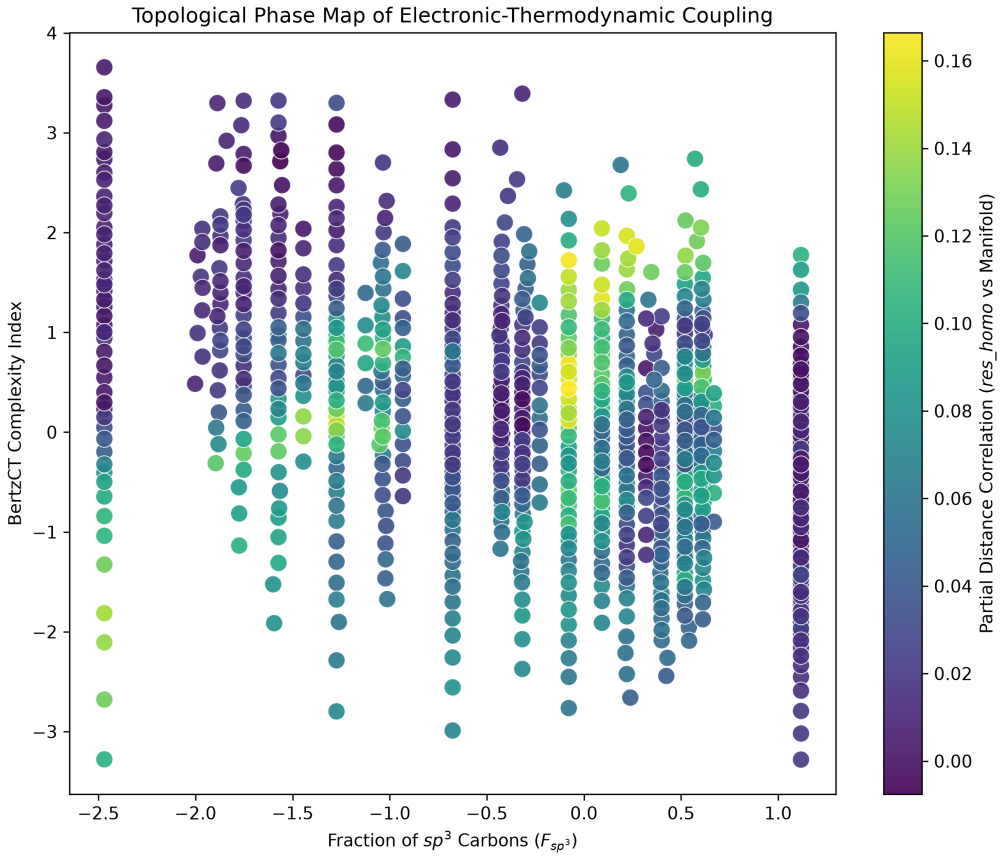

Figure 1: Topological phase map illustrating the electronic-thermodynamic coupling across the QM9 dataset. The plot displays the partial distance correlation ($dCor$) between the `res_homo` vector and the spatial-thermodynamic manifold as a function of the fraction of sp^3 carbons (F_{sp^3}) and the BertzCT complexity index. A sharp transition is visible between a Decoupled regime (dark purple, $dCor \leq 0.0097$), where the frontier electronic state is essentially orthogonal to the macroscopic property space, and a Recoupled regime (yellow/green, $dCor \geq 0.0924$), where `res_homo` strongly integrates with the manifold. This reveals that the independence of the HOMO energy is a local phenomenon; as molecular complexity and saturation vary, the electronic state becomes locked into a joint manifold with spatial and vibrational properties.

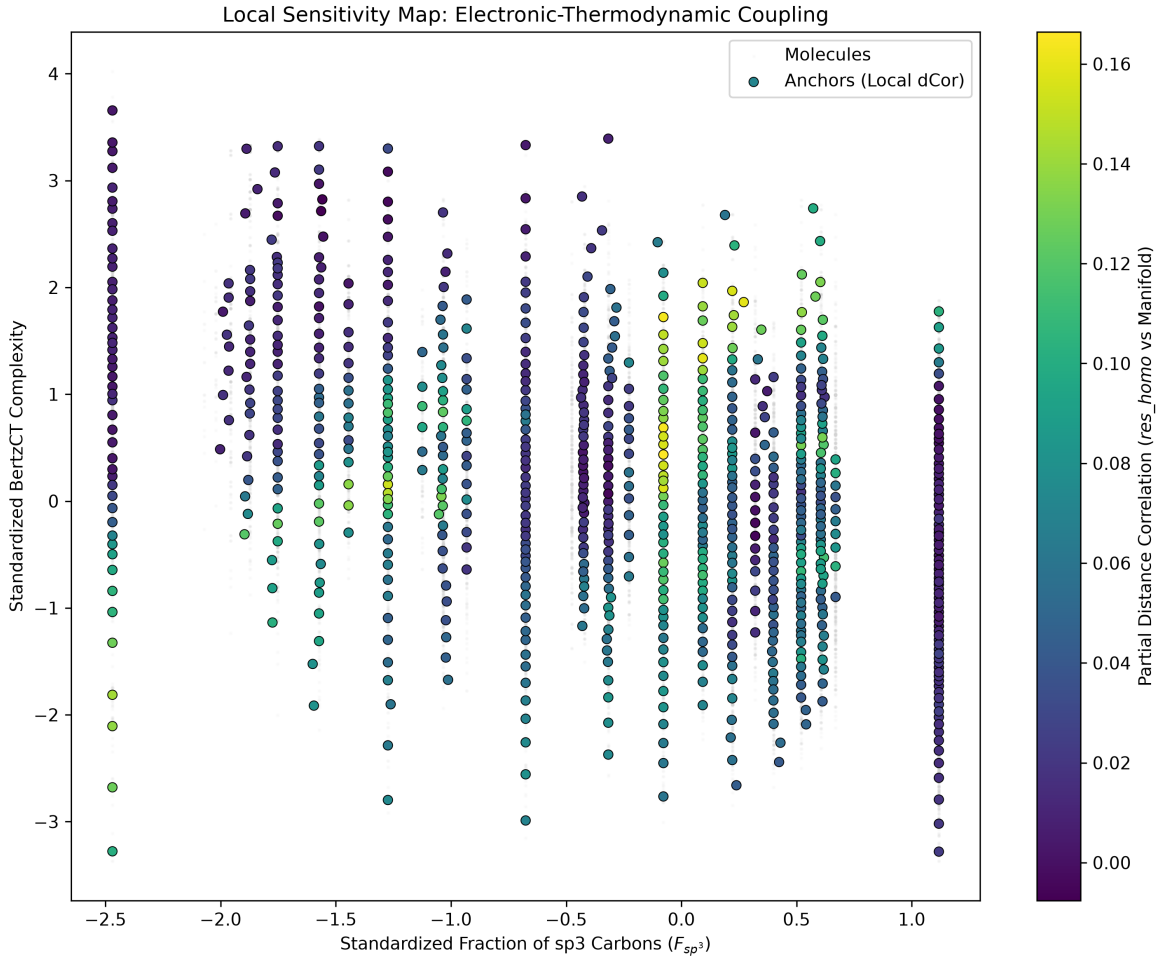

Figure 2: Local sensitivity map illustrating the topological phase transition in electronic-thermodynamic coupling across the QM9 dataset. The topological space is defined by the standardized fraction of sp^3 carbons (F_{sp^3}) and BertzCT complexity. Colored anchor points represent the local partial distance correlation ($dCor$) between the highest occupied molecular orbital residuals (`res_homo`) and the spatial-thermodynamic manifold (`res_zpve`, `res_cv`, `res_r2`), controlling for heteroatom counts. The map reveals a distinct transition from a "Decoupled" regime (dark purple, low $dCor$), where the frontier electronic state is orthogonal to spatial and thermodynamic organization, to a "Recoupled" regime (yellow and green, high $dCor$), where `res_homo` strongly integrates with the macroscopic property manifold. We learn that the independence of the HOMO energy is not a global characteristic, but a conditional state dictated by molecular topology and complexity.

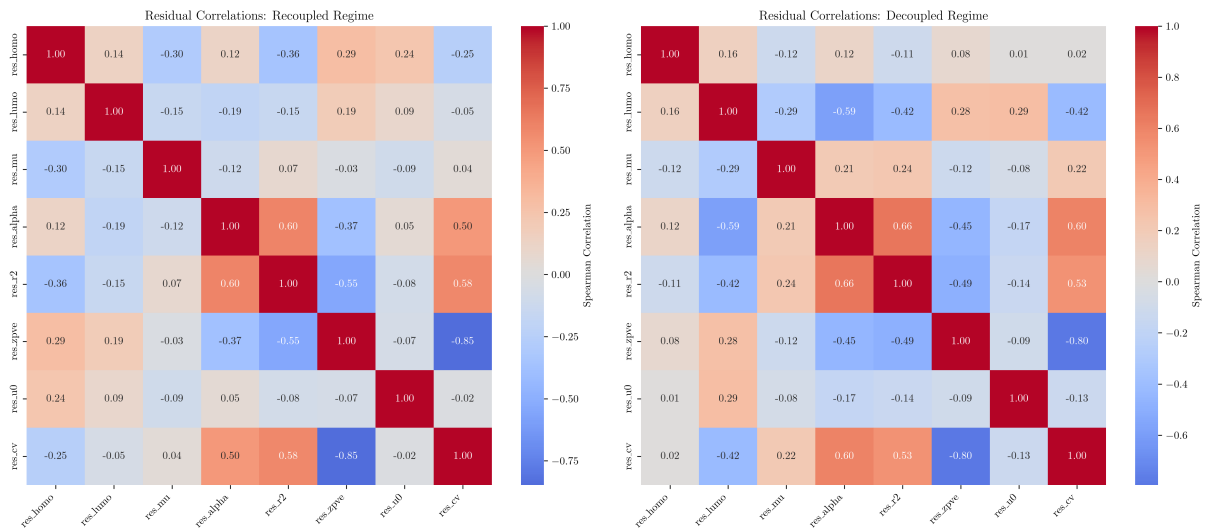

Figure 3: Spearman correlation heatmaps of property residuals for the Recoupled (left) and Decoupled (right) regimes, illustrating the topological phase transition in electronic-thermodynamic coupling. The residuals isolate structural variance from stoichiometric determinism. In the Decoupled regime, the frontier electronic state (`res_homo`) is nearly orthogonal to the spatial and thermodynamic manifold, showing near-zero correlations with spatial extent (`res_r2`), zero-point vibrational energy (`res_zpve`), and heat capacity (`res_cv`). Conversely, in the Recoupled regime, `res_homo` strongly integrates with these macroscopic properties, displaying significantly larger correlation magnitudes (e.g., -0.36 with `res_r2` and 0.29 with `res_zpve`). This visualizes the finding that under specific topological conditions, such as increased molecular complexity, the previously independent HOMO energy recouples into a highly structured, joint property manifold.

The stark contrast in the joint property structure between these two states is further evidenced by the Spearman correlation heatmaps of the property residuals (). In the Decoupled regime, `res_homo` exhibits near-zero correlation with spatial extent and thermodynamic variance. However, upon crossing the topological threshold into the Recoupled regime, the HOMO energy becomes locked into the joint manifold, displaying significantly larger correlation magnitudes.

3.2 Phase stability and scaffold validation

To ensure that the observed topological phase transition is a physical property of the chemical space rather than a stochastic artifact of the KNN mapping, we performed a rigorous split-half validation. The dataset was partitioned into two independent halves based on a strict Murcko scaffold split, ensuring that the training and validation sets comprised disjoint chemical families.

The local $dCor$ map remained highly consistent across these independent halves (). Across 800 topological anchors, the local sensitivity estimates yielded a Pearson correlation of $r = 0.3708$ and a Spearman rank correlation of $\rho = 0.3410$. Furthermore, the mean absolute error (MAE) in $dCor$ estimation between the two halves was only 0.0339. This stability confirms that the recoupling of electronic and thermodynamic properties is a robust, generalized feature of the molecular property space, validating the extracted phase boundaries.

3.3 Topological reorganization of the property manifold

To mathematically quantify the structural reorganization of the joint property space, we applied persistent homology to the 11-dimensional residual manifold. By constructing Vietoris-Rips complexes for representative subsamples of both regimes, we tracked the birth and death of topological features across dimensions 0 (connected components), 1 (loops), and 2 (voids).

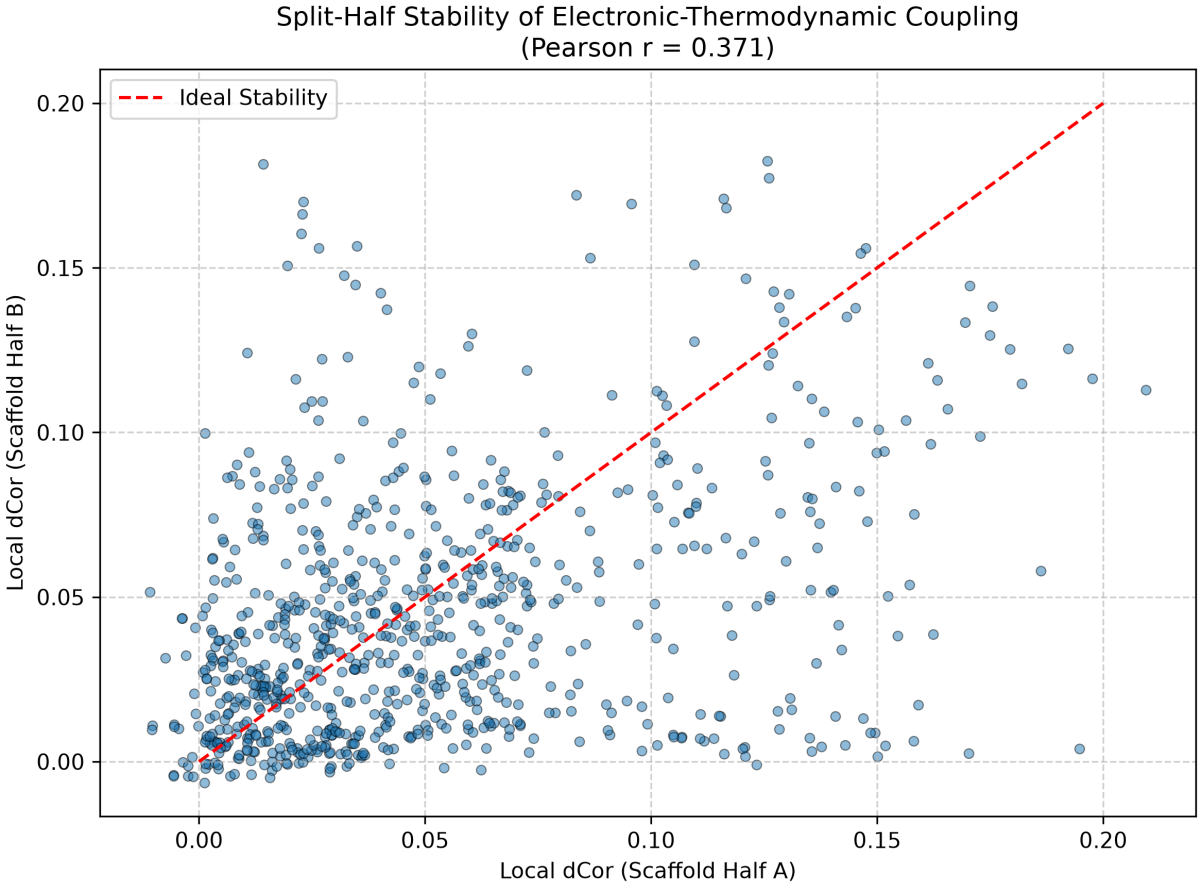

Figure 4: Split-half validation of the local distance correlation ($dCor$) measuring the coupling between the frontier electronic state (`res_homo`) and the spatial-thermodynamic manifold. To ensure the observed topological phase transitions are physical rather than stochastic, the dataset was partitioned by Murcko scaffold into two disjoint chemical families. The scatter plot compares the local $dCor$ values across 800 topological anchors for Scaffold Half A and Scaffold Half B, with the red dashed line indicating ideal stability. The positive correlation (Pearson $r = 0.371$) and low mean absolute error (0.0339) between the independent halves demonstrate that the recoupling of electronic and thermodynamic properties is a robust, stable feature of the molecular property space, confirming the validity of the phase boundaries.

Metric (600 molecules)	Decoupled Regime	Recoupled Regime
H0 Mean Lifetime	0.7950	0.8691
H1 Count	253	206
H1 Max Lifetime	0.2552	0.2764
H2 Count	97	7
H2 Max Lifetime	0.1239	0.0551

Table 1: Persistence metrics comparing the topological features of the property manifold in the Decoupled and Recoupled regimes.

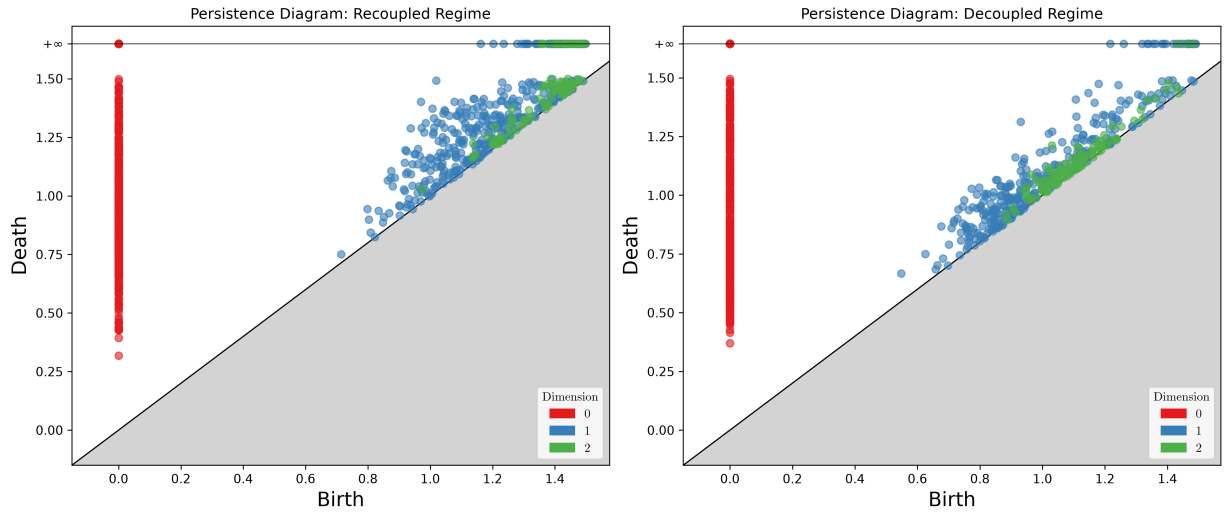

Figure 5: Persistence diagrams comparing the topological structure of the property manifold in the Recoupled (left) and Decoupled (right) regimes. The plots display the birth versus death of topological features across dimensions 0 (red), 1 (blue), and 2 (green). The transition to the Recoupled regime is characterized by a dramatic simplification of higher-order connectivity, visually evident as a 93% reduction in H2 features (green points). This demonstrates that as the HOMO energy recouples with the spatial-thermodynamic manifold, the joint property space collapses into a more structured, lower-dimensional surface with fewer voids. Additionally, the Recoupled regime exhibits an increased maximum lifetime for H1 features (blue points), indicating the formation of larger, more persistent loops in the coupled state.

The persistence metrics (Table 1) and corresponding diagrams (Figures and) reveal a dramatic simplification of higher-order connectivity upon entering the Recoupled regime. Most notably, there is a 93% reduction in H2 features, dropping from 97 voids in the Decoupled state to just 7 in the Recoupled state. This indicates that as the HOMO energy integrates with the spatial-thermodynamic manifold, the previously complex property space collapses into a more structured, lower-dimensional surface. Concurrently, the maximum lifetime of H1 features increases from 0.2552 to 0.2764, suggesting the formation of larger, more persistent loops within this newly coupled joint structure.

3.4 Structural drivers and geometric boundary conditions

Having established the existence and stability of the phase transition, we investigated the specific chemical motifs driving the recoupling. Kolmogorov-Smirnov (KS) enrichment analysis revealed that the Recoupled regime is heavily driven by the presence of carbonyl and ketone groups. Specifically, the `fr_C_O_noCOO` motif ($KS = 0.354$) and `fr_ketone` motif ($KS = 0.186$) act as primary structural drivers. This is accompanied by a significant increase in topological polar surface area (TPSA), which is enriched in the Recoupled regime ($KS = 0.245$) with a mean difference of $+5.07$ Å^2. Conversely, rigid bicyclic systems are significantly depleted in the Recoupled regime ($KS = 0.147$, mean difference -0.293), indicating that extreme rigid strain favors the decoupling of the frontier electronic state.

To rule out the possibility that these structural drivers were merely capturing latent DFT convergence failures, we evaluated geometric consistency across the regimes (Figures and). The mean 3D embedding failure rate, a proxy for geometric inconsistency, was found to be 1.80% across the dataset. Crucially, this failure rate remained uniform across both the Recoupled and Decoupled regimes, confirming that the observed phase transition is driven by physical topology rather than computational artifacts.

Finally, we examined perfectly linear molecules ($A = 0$) to test the boundary conditions of our topological rules. Linear molecules exhibited a mean `res_homo` of 0.0146 Hartree, compared to 0.0203 Hartree for their nearest non-linear topological neighbors. A KS test ($p = 0.989$) indicated that linear molecules do not significantly deviate from their local topological baseline. This demonstrates that the recoupling rules

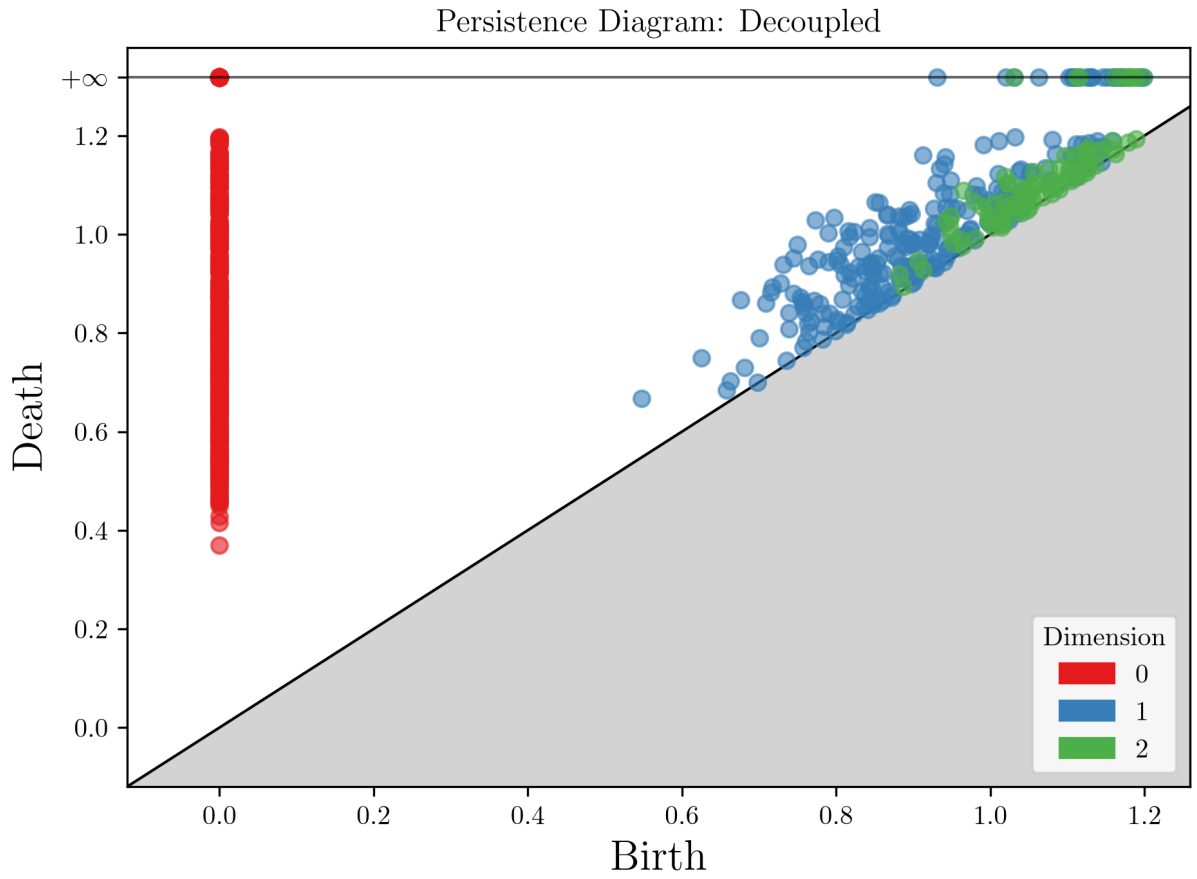

Figure 6: Persistence diagram for the Decoupled regime, illustrating the birth and death of topological features across dimensions 0 (red), 1 (blue), and 2 (green). The plot highlights the high topological complexity of the property manifold when the highest occupied molecular orbital energy is orthogonal to the spatial and thermodynamic manifold (`zpve`, `cv`, `r2`). Specifically, the dense cluster of H2 features (green) reflects the 97 higher-order voids present in this state. This demonstrates that the Decoupled regime possesses significantly greater higher-order connectivity compared to the Recoupled regime, which undergoes a 93% reduction in H2 features as the property space collapses into a simpler, lower-dimensional structure.

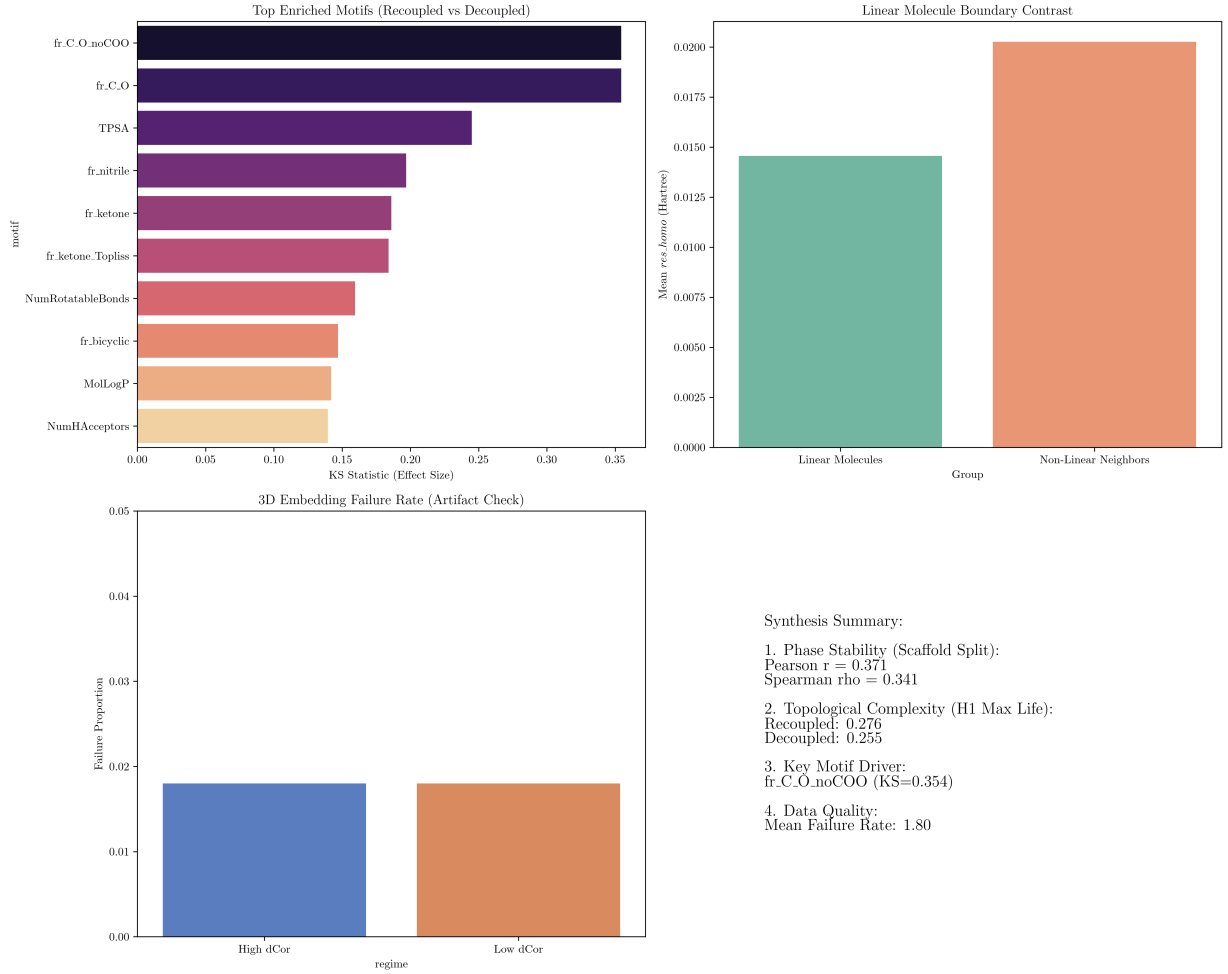

Figure 7: Structural drivers, boundary conditions, and validation metrics of the topological phase transition in electronic-thermodynamic coupling. The top-left panel displays the Kolmogorov-Smirnov (KS) statistics for motif enrichment, revealing that carbonyl groups (e.g., fr_C_O_noCOO, KS = 0.354) and polar surface area (TPSA) are the primary drivers of the recoupled regime. The top-right panel contrasts the mean `res_homo` residuals of linear molecules (0.0146 Hartree) with their non-linear topological neighbors (0.0203 Hartree), indicating that recoupling rules remain continuous across geometric singularities. The bottom-left panel confirms a uniform 3D embedding failure rate of 1.80% across both high and low distance correlation ($dCor$) regimes, demonstrating that the observed transition is not an artifact of DFT convergence failures. Finally, the synthesis summary validates the robustness of these regimes, highlighting phase stability across disjoint Murcko scaffolds (Pearson $r = 0.371$) and an increase in topological persistence (H1 max lifetime of 0.276 in the recoupled state). Together, these analyses demonstrate that the recoupling of frontier electronic states with the macroscopic property manifold is a robust, motif-driven phenomenon rather than a stochastic or geometric artifact.

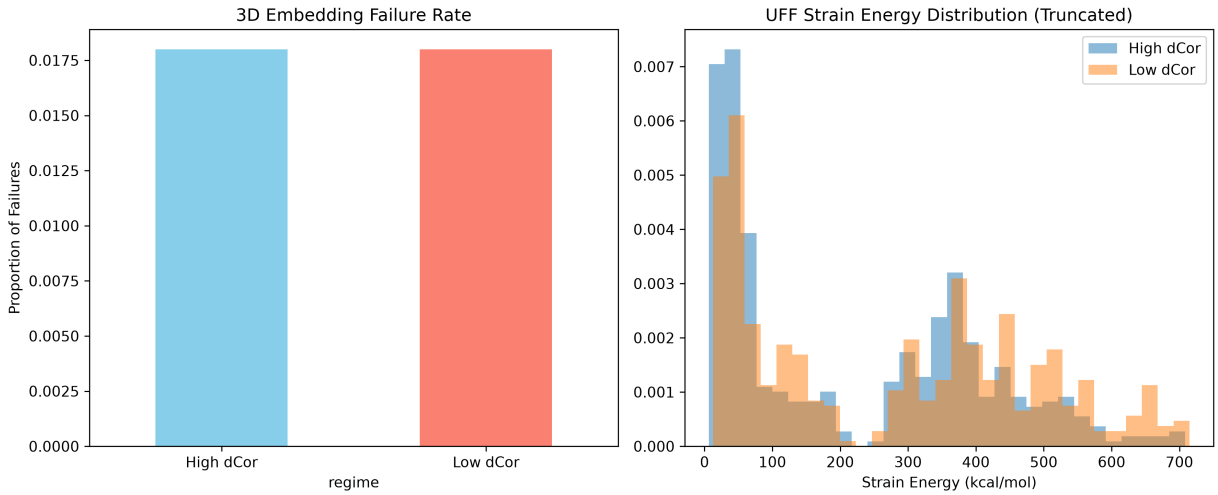

Figure 8: The figure evaluates potential structural and geometric artifacts across the Recoupled (High dCor) and Decoupled (Low dCor) regimes. The left panel demonstrates that the 3D embedding failure rate, a proxy for geometric inconsistency, remains uniform at approximately 1.80% across both regimes, confirming that the observed topological phase transition is not an artifact of DFT convergence failures. The right panel displays the truncated UFF strain energy distribution, revealing that the Decoupled regime (Low dCor) contains a higher proportion of high-strain molecules. This distribution supports the observation that extreme rigid strain, characteristic of bicyclic systems, favors the decoupling of the `res_homo` frontier electronic state from the macroscopic spatial and thermodynamic manifold.

governing the electronic-thermodynamic integration remain continuous even through geometric singularities, further validating the universality of the topological gradients mapped in this study.

4 Conclusions

4.1 Re-evaluating electronic-thermodynamic orthogonality

The fundamental assumption that the highest occupied molecular orbital (HOMO) energy is globally orthogonal to macroscopic spatial and thermodynamic properties has been shown to be incomplete. By systematically stripping away the overwhelming stoichiometric determinism present in the QM9 dataset using gradient-boosted regressors, we successfully isolated the true chemical variance of the molecular property manifold. Our continuous topological mapping across the fraction of sp^3 carbons (F_{sp^3}) and BertzCT complexity reveals that the independence of the frontier electronic state is a highly localized phenomenon, characteristic primarily of simple, non-polar molecules. As molecular complexity increases, the chemical space undergoes a distinct topological phase transition where the HOMO energy suddenly recouples with the spatial-thermodynamic manifold, integrating with properties such as zero-point vibrational energy, heat capacity, and spatial extent.

4.2 Topological reorganization of the property manifold

This recoupling is not merely a statistical correlation but represents a fundamental geometric reorganization of the property space. Persistent homology analysis demonstrates that the transition from the decoupled to the recoupled regime is defined by a dramatic collapse in higher-order topological complexity.

As detailed in Table 1, the recoupled state is characterized by a near-complete elimination of H2 topological voids—reducing from 97 to 7—and an increase in the maximum lifetime of H1 loops. This mathematical signature indicates that upon electronic-thermodynamic integration, the joint property manifold simplifies into a more structured, lower-dimensional surface with significantly fewer voids.

4.3 Structural drivers and broader implications

We have established that this phase transition is governed by specific structural drivers rather than stochastic or computational artifacts. Motif enrichment analysis identifies that the recoupling is heavily driven by increased carbonyl density and topological polar surface area. Conversely, rigid bicyclic systems exhibit extreme geometric strain that actively favors the decoupling of the frontier electronic state. Crucially, split-half validation across disjoint Murcko scaffolds and boundary-condition testing on linear molecules confirm that these topological rules are robust, continuous, and independent of latent density functional theory convergence failures.

Ultimately, this work demonstrates that the joint structure of molecular properties is a conditional state dictated by continuous topological gradients. Recognizing that electronic-thermodynamic integration is a localized, motif-driven phenomenon provides a new framework for understanding chemical space. These findings suggest that future cheminformatic models and machine learning architectures must account for these topological phase boundaries, as the fundamental relationships between electronic and macroscopic properties can fundamentally rewire across different regimes of molecular complexity.