Breaking the HOMO-Refusal Phenomenon via Targeted Functional Group Partitioning

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The "HOMO-refusal" phenomenon describes the unique decoupling of frontier molecular orbital energies from the global thermodynamic and spatial property manifold of a molecule. To determine whether this orthogonality is a universal baseline or a conditional state, we investigate the structural integration of frontier orbitals in the presence of specific electronic sinks and sources. Using the QM9 dataset of over 133,000 molecules, we isolate chemical architecture from stoichiometric determinism by extracting gradient-boosted atom-count residuals for non-redundant physical properties. We then partition the dataset based on exact functional group counts and compute the multivariate distance correlation between the frontier orbital residuals and the remaining nine-dimensional property space. Additionally, we employ Local Intrinsic Dimensionality and Optimal Transport to quantify manifold reorganization. Our analysis reveals that while frontier orbitals remain largely orthogonal to the physical manifold in baseline aliphatic systems, this independence breaks down under extreme electronic densities. Specifically, the accumulation of electron-withdrawing or electron-donating groups forces a sharp reintegration of the frontier orbitals into the global physical state, significantly increasing their distance correlation. Furthermore, highly conjugated systems exhibit a measurable contraction in intrinsic dimensionality and a significant shift in their global distribution compared to purely aliphatic molecules. These findings map the boundary conditions of electronic decoupling, demonstrating that the isolation of frontier orbitals is dynamically rewired by targeted bonding architectures.

1 Introduction

The mapping of molecular property spaces has revealed complex interdependencies between thermodynamic, geometric, and electronic features. However, a persistent anomaly within these spaces is the "HOMO-refusal" phenomenon. This describes the unique decoupling of frontier molecular orbital energies—specifically the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO)—from the global thermodynamic and spatial property manifold of a molecule. While extensive properties often form tightly bound, highly correlated manifolds driven by molecular size and stoichiometry, frontier electronic states frequently appear orthogonal to this underlying structure.

A critical unresolved question is whether this observed orthogonality represents a universal baseline of the underlying quantum mechanical theory level, such as the B3LYP/6-31G(2df,p) level of theory used in standard benchmark datasets, or if it is a conditional physical state. If conditional, the isolation of frontier orbitals should be dynamically rewired by specific chemical architectures, such as the introduction of targeted electronic sinks or sources. Understanding these boundary conditions is essential for predictive molecular design, where tuning electronic properties independently of thermodynamic stability or molecular geometry is often a primary objective.

To determine the boundary conditions of electronic decoupling, we investigate the structural integration of frontier orbitals across the QM9 dataset of over 133,000 organic molecules. Because molecular property spaces are overwhelmingly dominated by stoichiometric determinism—where simple atom counts account for nearly all variance in extensive properties—we first isolate the purely chemical architecture. This is achieved by extracting gradient-boosted atom-count residuals for a non-redundant set of physical properties. By transforming the data into this residual space, we expose the underlying joint structure of a nine-dimensional

physical manifold comprising geometric, response, thermodynamic, and vibrational properties, distinct from the frontier orbital manifold.

We then partition the dataset based on exact functional group counts, specifically targeting electron-withdrawing groups, electron-donating groups, and conjugated systems. To quantify the structural integration of the frontier orbitals, we compute the multivariate distance correlation ($dCor$) between the frontier orbital residuals and the nine-dimensional physical manifold across these discrete structural populations. Furthermore, we employ Local Intrinsic Dimensionality (LID) and Optimal Transport metrics, specifically the Wasserstein distance, to measure how the global manifold reorganizes and compresses under different chemical constraints.

Our analysis demonstrates that the "HOMO-refusal" phenomenon is indeed a conditional state rather than a universal constant. While frontier orbitals remain largely orthogonal to the physical manifold in baseline aliphatic systems, this independence breaks down under extreme electronic densities. The targeted accumulation of electron-withdrawing or electron-donating groups forces a sharp reintegration of the frontier orbitals into the global physical state, marked by a significant increase in distance correlation. Additionally, highly conjugated systems exhibit a measurable contraction in intrinsic dimensionality and a distinct shift in their global distribution compared to purely aliphatic molecules. These findings map the exact chemical thresholds that govern frontier electronic decoupling, revealing how targeted bonding architectures dynamically rewire the molecular state.

2 Methods

2.1 Data preparation and preprocessing

The analysis was conducted using the QM9 dataset, which provides quantum chemical properties for small organic molecules computed at the B3LYP/6-31G(2df,p) level of theory. To ensure data integrity, the initial dataset was deduplicated based on canonical SMILES strings. Molecules with linear geometries, identified by a rotational constant of zero ($A = 0$) or undefined values, were immediately removed prior to further processing, yielding a final filtered dataset of 133,793 molecules.

To stabilize the extreme variance associated with the rotational constant A, a logarithmic transformation, $\log(A + 1)$, was applied. The molecular properties were then partitioned into two distinct sets. The global physical manifold was defined as a nine-dimensional space comprising geometric, response, thermodynamic, and vibrational properties: the transformed rotational constant A, rotational constants B and C, dipole moment μ, isotropic polarizability α, electronic spatial extent $\langle R^2 \rangle$, zero-point vibrational energy (ZPVE), internal energy at 0 K (U_0), and heat capacity (C_v). The frontier orbital set was defined independently, comprising the highest occupied molecular orbital (HOMO) energy, the lowest unoccupied molecular orbital (LUMO) energy, and the HOMO-LUMO gap.

2.2 Non-linear stoichiometric residualization

Because extensive molecular properties are overwhelmingly dominated by stoichiometric determinism, it was necessary to isolate the variance attributable strictly to chemical architecture. We trained gradient boosting regressors to predict each of the twelve non-redundant properties using only exact atomic counts (C, N, O, F, H) as input features. The residuals from these non-linear stoichiometric baselines were extracted and standardized. This transformation projects the dataset into a residual space where the trivial scaling effects of molecular size and atomic composition are removed, exposing the underlying joint structural dependencies between the physical manifold and the frontier orbitals.

2.3 Structural motif extraction and partitioning

To investigate the conditional nature of the HOMO-refusal phenomenon, molecules were partitioned into statistically robust populations based on their specific chemical architectures. SMILES strings were parsed to extract exact counts of targeted structural and electronic motifs. The primary motifs of interest included electron-withdrawing groups (EWGs, such as nitriles, carbonyls, and fluorines), electron-donating groups (EDGs), conjugated heavy atoms, and aromatic rings. Molecules were subsequently grouped into discrete

structural bins based on the cumulative count of these specific motifs. Additionally, the total heavy-atom count (N_{heavy}) was computed for every molecule to enable rigorous size-matching in downstream comparative analyses.

2.4 Quantifying frontier orbital coupling

To measure the structural integration of the frontier orbitals into the global physical state, we computed the multivariate distance correlation ($dCor$) between the frontier orbital residuals and the nine-dimensional physical manifold. For each discrete structural bin, the joint distance correlation was calculated as $dCor((\text{HOMO}, \text{LUMO}), \text{Manifol}$

To map the trajectory of this coupling capacity as a function of targeted functional group perturbation, the joint $dCor$ was evaluated across the progressive accumulation of electron-withdrawing and electron-donating motifs. This approach allowed us to identify the exact chemical thresholds at which the frontier orbitals reintegrate into the global physical state. To establish robust confidence intervals for the joint coupling capacity, the distance correlation for each bin was computed using 100 bootstrap resamples.

2.5 Manifold reorganization and intrinsic dimensionality

Beyond direct correlation metrics, the global reorganization of the property space was quantified using Local Intrinsic Dimensionality (LID) and Optimal Transport.

To assess whether specific chemical architectures compress the physical property space, we implemented a k-Nearest Neighbors Maximum Likelihood Estimator for LID on the standardized nine-dimensional residual space, evaluating neighborhoods of size $k \in [10, 50]$. The LID of highly conjugated molecules was compared against that of purely aliphatic molecules. To eliminate size confounding, these two populations were strictly matched by their heavy-atom counts.

Furthermore, to measure the global displacement of the physical state distributions between these distinct structural families, we computed the Wasserstein distance. Using optimal transport methodologies, the Sliced 2-Wasserstein distance was calculated between the nine-dimensional residual distributions of the size-matched conjugated and aliphatic populations, quantifying the total geometric divergence induced by the targeted bonding architectures.

2.6 Robustness and validation

To ensure that the observed reintegration thresholds were driven by electronic motif density rather than size-dependent artifacts, structural claims were subjected to rigorous split-half stability testing. The dataset was partitioned by molecular size into "Small" ($N_{heavy} \leq 7$) and "Large" ($N_{heavy} > 7$) regimes. The joint $dCor$ trajectories and their corresponding bootstrap error bars were re-evaluated within these isolated size regimes to confirm the consistency of the functional group perturbation effects across the dataset.

3 Results

3.1 The global residual manifold and the baseline of stoichiometry

The property space of the QM9 dataset is dominated by molecular size and stoichiometric determinism. Our initial gradient-boosted regression on exact atom counts (C, N, O, F, H) confirms that stoichiometry accounts for nearly all variance in extensive thermodynamic properties. Specifically, the stoichiometric baseline achieves an R^2 of 0.9994 for internal energy at 0 K (U_0) and 0.9974 for zero-point vibrational energy (ZPVE). In stark contrast, frontier orbital energies are poorly predicted by stoichiometry alone, with the highest occupied molecular orbital (HOMO) yielding an R^2 of only 0.1990.

By extracting the residuals from these non-linear stoichiometric baselines, we successfully isolated the variance attributable strictly to chemical architecture. Within this residual space, the nine-dimensional physical manifold—comprising geometric, response, thermodynamic, and vibrational properties—exhibits a tightly bound joint structure. However, for baseline aliphatic molecules lacking targeted functional motifs

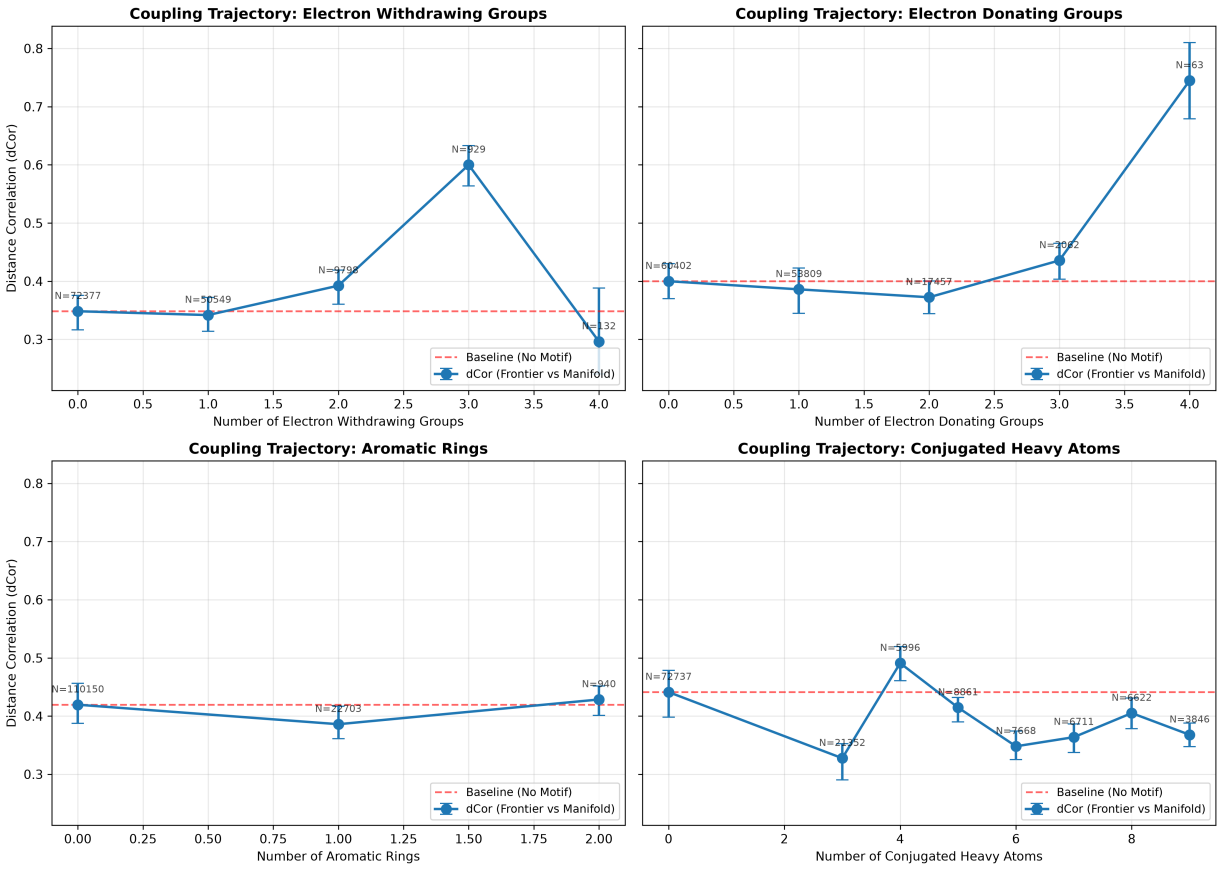

Figure 1: Distance correlation ($dCor$) trajectories illustrating the coupling strength between frontier orbital residuals and the 9-dimensional physical manifold as a function of specific chemical motifs. The top panels reveal the reintegration thresholds that break the baseline "HOMO-refusal" phenomenon; the accumulation of extreme electronic densities, specifically 3 electron-withdrawing groups or 4 electron-donating groups, triggers a sharp increase in $dCor$ (reaching 0.7448 for 4 EDGs). This demonstrates that frontier orbitals are forced to reintegrate into the global physical state under high electronic motif density. Conversely, the bottom panels show that purely topological features, such as aromatic rings and conjugated heavy atoms, fail to induce this strong reintegration, maintaining a weak coupling (e.g., 0.4285 for 2 aromatic rings) near the baseline. These trends confirm that the joint structure of the molecular property space is dynamically rewired by electronic density rather than structural topology.

(i.e., zero electron-withdrawing groups), the frontier orbital residuals remain largely orthogonal to this physical manifold, exhibiting a weak joint distance correlation ($dCor$) of approximately 0.350. This establishes the "HOMO-refusal" phenomenon as the baseline state for unperturbed small organic molecules.

3.2 The HOMO-refusal phenomenon and reintegration thresholds

To determine whether the isolation of frontier electronic states is a universal constant of the B3LYP/6-31G(2df,p) theory level or a conditional physical state, we systematically evaluated the coupling strength between the frontier orbitals and the physical manifold as a function of targeted electronic motifs. The progressive accumulation of electron-withdrawing groups (EWGs, such as nitriles, carbonyls, and fluorines) and electron-donating groups (EDGs) reveals that the HOMO-refusal phenomenon is highly sensitive to extreme electronic densities.

As illustrated in , the joint $dCor$ remains stable for molecules with few EWGs but undergoes a sharp reintegration threshold at counts ≥ 3. An even more pronounced effect is observed with the accumulation

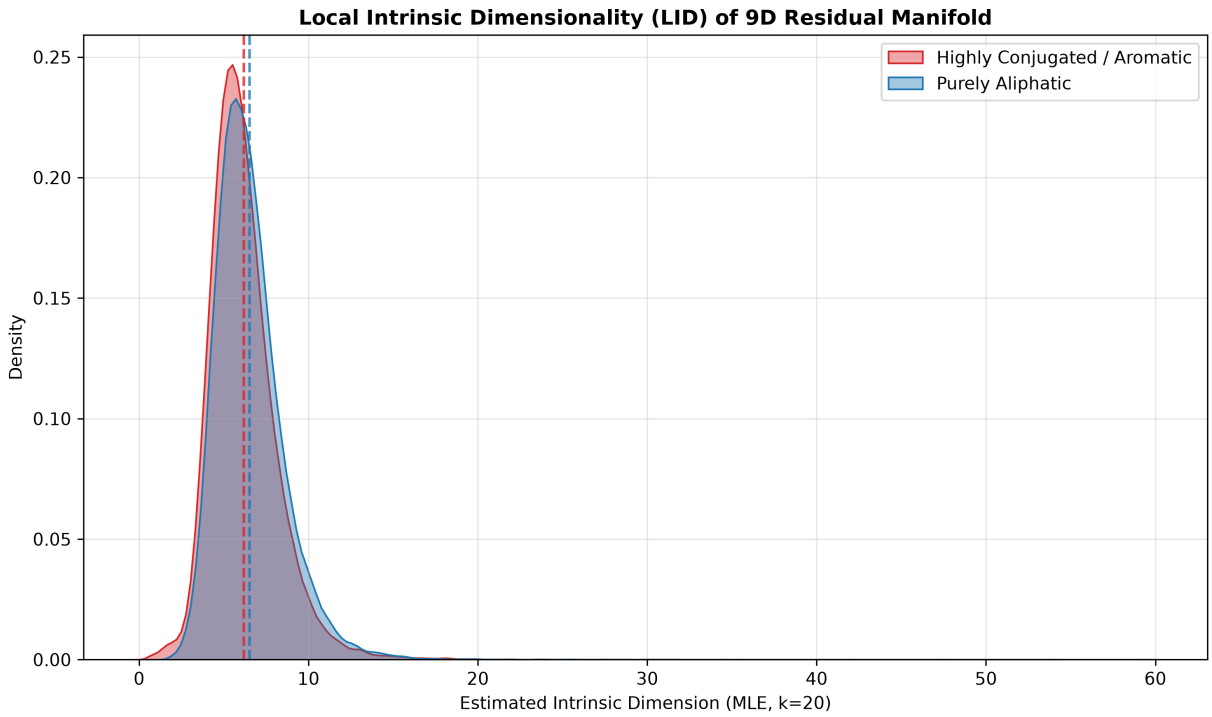

Figure 2: Density distributions of the Local Intrinsic Dimensionality (LID) for the 9-dimensional residual manifold, comparing highly conjugated or aromatic molecules (red) against purely aliphatic molecules (blue). The dashed vertical lines indicate the mean LID for each structural family. The plot illustrates a measurable dimensionality compression, where aliphatic molecules exhibit a mean LID of 6.526 compared to 6.174 for conjugated systems. This leftward shift reveals that highly conjugated molecules are more physically constrained, causing their geometric, thermodynamic, and response properties to collapse onto a lower-dimensional subspace within the global physical manifold.

of EDGs; molecules possessing 4 EDGs exhibit a *dCor* of 0.7448 (95% CI: 0.6789, 0.8100). This dramatic increase indicates that extreme electronic densities force the frontier orbitals to track the global spatial and thermodynamic reorganization of the molecule, effectively breaking their orthogonality.

Conversely, purely topological motifs fail to trigger this strong reintegration. The accumulation of aromatic rings or conjugated heavy atoms maintains a relatively weak coupling with the physical manifold. For instance, molecules with 2 aromatic rings exhibit a *dCor* of 0.4285 (95% CI: 0.4009, 0.4519), which is only marginally higher than the baseline. This contrast confirms that the reintegration of frontier orbitals is specifically driven by localized electronic sinks and sources rather than extended structural topology.

3.3 Manifold reorganization and intrinsic dimensionality

Beyond the direct coupling of frontier orbitals, we quantified the global reorganization of the physical property space using Optimal Transport and Local Intrinsic Dimensionality (LID). By comparing strictly size-matched populations of highly conjugated molecules against purely aliphatic molecules ($N = 39,704$ each), we observed a fundamental shift in the manifold's topology.

First, the global displacement between the two structural families is substantial. The Sliced 2-Wasserstein distance between the nine-dimensional residual distributions of the conjugated and aliphatic populations is 0.5143 (summarized in , Panel C). This metric represents the total geometric "work" required to transform the physical state distribution of one family into the other, confirming that conjugation significantly reorganizes the joint density of the property space.

Second, this reorganization is accompanied by a measurable compression in intrinsic dimensionality.

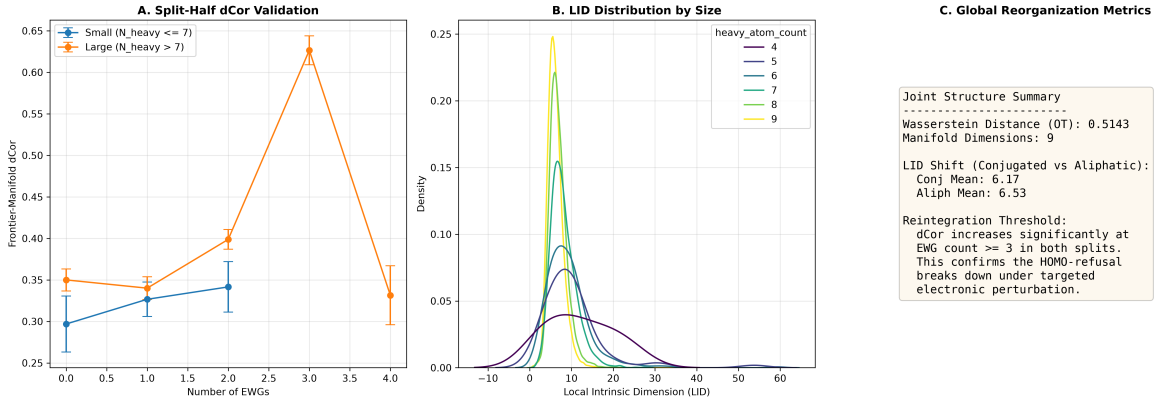

Figure 3: Validation of the HOMO-refusal breakdown and manifold reorganization in the QM9 dataset. Panel A displays the split-half validation of the distance correlation ($dCor$) between frontier orbitals and the 9-dimensional physical manifold as a function of electron-withdrawing groups (EWGs). A sharp reintegration threshold occurs at 3 EWGs for large molecules ($N_{heavy} > 7$), where $dCor$ spikes to 0.6265, demonstrating that extreme electronic densities break frontier orbital orthogonality and force them to couple with the global physical state. Panel B shows density distributions of Local Intrinsic Dimensionality (LID) partitioned by `heavy_atom_count`, illustrating the topological complexity of the property space across molecular sizes. Panel C summarizes global reorganization metrics, confirming a Sliced 2-Wasserstein distance of 0.5143 between conjugated and aliphatic populations and a dimensionality compression from a mean LID of 6.53 (aliphatic) to 6.17 (conjugated). These results collectively reveal that targeted chemical architectures, such as EWG accumulation and conjugation, rewire the molecular state and constrain physical properties onto a lower-dimensional subspace.

As shown in , aliphatic molecules occupy a manifold with a mean LID of 6.526 ($\pm$2.041). In contrast, conjugation reduces this topological complexity, shifting the distribution to a mean LID of 6.174 ($\pm$2.066). This contraction indicates that conjugated systems are subject to stricter physical constraints, causing their geometric, thermodynamic, and response properties to collapse onto a lower-dimensional subspace within the global physical manifold.

3.4 Robustness and size-dependent validation

To ensure that the observed reintegration thresholds and manifold reorganizations are fundamental consequences of electronic motif density rather than size-dependent artifacts, we validated these findings using a split-half stability test partitioned by heavy-atom count (N_{heavy}).

summarizes these validation metrics. Panel A demonstrates that the reintegration phenomenon remains consistent across different size regimes. For "Small" molecules ($N_{heavy} \leq 7$), the $dCor$ rises steadily from a baseline of 0.2969 at 0 EWGs to 0.3416 at 2 EWGs. For "Large" molecules ($N_{heavy} > 7$), the baseline coupling of 0.3500 at 0 EWGs is followed by a significant spike to 0.6265 ($\pm$0.0174) at the critical threshold of 3 EWGs. The persistence of this threshold across size partitions confirms that the breakdown of the HOMO-refusal phenomenon is a robust chemical behavior governed by functional group density.

Furthermore, the topological complexity of the property space exhibits consistent trends across molecular sizes, as illustrated by the LID density distributions in Panel B. Finally, Panel C corroborates the global reorganization metrics discussed previously, reinforcing the conclusion that targeted chemical architectures, such as EWG accumulation and conjugation, rewire the molecular state and constrain physical properties onto a lower-dimensional subspace.

4 Conclusions

The investigation of the QM9 dataset reveals that the joint structure of molecular property spaces is a dynamic state, intricately rewired by chemical architecture. We have demonstrated that the "HOMO-refusal" phenomenon—the striking independence of frontier molecular orbitals from the global thermodynamic, geometric, and response manifold—is not a universal constant of the underlying quantum mechanical theory. Instead, it represents the baseline physical state of small, unperturbed organic molecules.

This baseline orthogonality is highly brittle under the introduction of targeted electronic perturbations. By isolating chemical variance from stoichiometric determinism, our analysis identifies the exact boundary conditions where electronic decoupling breaks down. The accumulation of extreme electronic densities, specifically three or more electron-withdrawing groups or four electron-donating groups, forces a sharp reintegration of the frontier orbitals into the global physical state. In contrast, purely topological motifs such as aromaticity and extended conjugation fail to trigger this reintegration, maintaining the decoupled state of the electronic frontier.

Furthermore, while conjugation does not strongly couple frontier orbitals to the physical manifold, it fundamentally reorganizes the manifold itself. Optimal Transport and Local Intrinsic Dimensionality metrics confirm that conjugated architectures induce a significant global displacement of the property distribution and compress the physical state onto a lower-dimensional subspace. This indicates that conjugated systems are subject to stricter physical constraints than their aliphatic counterparts, even as their frontier orbitals remain relatively isolated.

Crucially, split-half validation confirms that these reintegration thresholds and manifold reorganizations are robust chemical behaviors driven by functional group density, independent of molecular size. These findings provide a critical framework for predictive molecular design. They establish that while frontier electronic properties can be tuned independently of thermodynamic and spatial constraints in baseline aliphatic systems, this design freedom is fundamentally lost when extreme electronic densities rewire the molecular state.