Phase Transitions in HOMO Orthogonality Across Continuous Axes of Molecular Conjugation

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The energy of the highest occupied molecular orbital (HOMO) in small organic molecules often decouples from primary spatial and thermodynamic properties, but the structural thresholds driving this orthogonality remain unclear. To determine whether this decoupling is gradual or characterized by sharp phase transitions, we analyzed the QM9 dataset by evaluating the relationship between HOMO energy and a spatial-vibrational triad consisting of zero-point vibrational energy, heat capacity, and electronic spatial extent. We first extracted non-linear gradient-boosted stoichiometric residuals to remove baseline atomic determinism. Next, we computed the multivariate distance correlation between the HOMO residual and the spatial-vibrational residuals using a sliding window approach along two continuous structural axes: conjugation density and heteroatom-to-carbon ratio. The results reveal a structured decoupling landscape rather than uniform orthogonality. Specifically, we identified a sharp phase transition at a conjugation density of approximately 0.11 aromatic rings per heavy atom, marking a rapid decoupling of HOMO energy as electron delocalization increases. A secondary transition occurs at a heteroatom-to-carbon ratio of 0.125, indicating further modulation by heteroatom content. Despite the inclusion of uncorrected geometric artifacts in the dataset, statistical validation against a shuffled null distribution confirms that the coupling remains strictly above the 95th percentile baseline across all thresholds. Ultimately, these findings demonstrate that the integration of frontier electronic states into the macroscopic molecular manifold is governed by distinct, quantifiable thresholds of structural complexity.

1 Introduction

The frontier electronic states of small organic molecules, particularly the highest occupied molecular orbital (HOMO), play a fundamental role in determining chemical reactivity, optical properties, and charge transport. However, the energy of the HOMO often exhibits a striking decoupling from the primary spatial and thermodynamic manifold of the molecule. For instance, while basic stoichiometric atom counts can explain the vast majority of the variance in macroscopic properties such as zero-point vibrational energy ($zpve$) and internal energy (u_0), they capture a remarkably small fraction of the variance in HOMO energy. This orthogonality suggests that frontier orbital energies are governed by higher-order structural complexities rather than simple atomic composition.

Despite this known decoupling, the structural thresholds that drive the isolation of HOMO energy from the spatial-vibrational baseline remain poorly understood. It is unclear whether the integration of isolated frontier-electronic states into the macroscopic molecular manifold decays gradually with increasing structural complexity, or if it is characterized by sharp phase transitions triggered by specific molecular features. Understanding these dynamics requires moving beyond categorical classifications of molecules and instead tracking electronic coupling along continuous axes of structural evolution, such as electron delocalization and heteroatom incorporation.

In this study, we investigate the coupling dynamics of HOMO energy across continuous structural gradients using the QM9 dataset. To rigorously isolate the intrinsic relationships between properties, we first remove baseline atomic determinism by extracting non-linear gradient-boosted stoichiometric residuals for HOMO energy and a spatial-vibrational triad consisting of $zpve$, heat capacity (c_v), and electronic spatial extent (r^2). We then parameterize the dataset along two continuous topological axes: the density of conjugation (measured as aromatic rings per heavy atom) and the heteroatom-to-carbon ratio. By computing

the multivariate distance correlation between the HOMO residual and the spatial-vibrational triad within sliding windows along these axes, we map the boundary of electronic decoupling.

Our analysis reveals a highly structured decoupling landscape rather than uniform orthogonality. We identify a primary, sharp phase transition driven by conjugation density, where HOMO energy rapidly decouples as electron delocalization increases. A secondary transition is observed along the heteroatom-to-carbon axis, indicating further modulation by heteroatom content. Despite the inclusion of uncorrected geometric artifacts inherent to the dataset, statistical validation confirms that the coupling remains robustly above a shuffled null baseline across all evaluated thresholds. Ultimately, this work demonstrates that the orthogonality of frontier electronic states is not an absolute property, but rather a dynamic feature governed by distinct, quantifiable thresholds of molecular complexity.

2 Methods

2.1 Dataset and preprocessing

The QM9 dataset was utilized to investigate the coupling dynamics of the highest occupied molecular orbital (HOMO) energy. Initial preprocessing involved data loading, deduplication, and the establishment of baseline features, including a binary flag (`is_linear`) for linear molecules where the rotational constant $A = 0$. Notably, the dataset contains approximately 3,054 molecules previously identified as failing geometric consistency checks. As observed in the final analysis, no specific filter was applied to remove these latent artifacts; instead, the robustness of the multivariate distance correlation and the use of large sliding windows were relied upon to mitigate the impact of individual geometric outliers on the macroscopic trends.

2.2 Stoichiometric residual extraction

To strictly eliminate baseline atomic determinism, non-linear stoichiometric residuals were extracted for the properties of interest. Gradient-boosted decision trees were trained to predict HOMO energy, zero-point vibrational energy ($zpve$), heat capacity (c_v), and electronic spatial extent (r^2) using only stoichiometric atom counts (C, N, O, F, H) as input features. The residuals from these models—representing the variance unexplained by simple atomic composition—were then isolated for downstream analysis. This procedure ensures that any observed coupling is driven by structural and topological complexity rather than basic stoichiometry.

2.3 Continuous structural parameterization

Rather than partitioning the dataset by imbalanced categorical heteroatom identities, the deduplicated molecules were parameterized along two continuous structural axes using RDKit-derived topological metrics. The first axis, conjugation density, was defined as the number of aromatic rings per heavy atom, capturing the degree of electron delocalization. The second axis, the heteroatom-to-carbon ratio, was calculated as the sum of nitrogen, oxygen, and fluorine atoms divided by the number of carbon atoms, quantifying the extent of scaffold interruption by non-carbon elements.

2.4 Sliding window distance correlation and grid analysis

To map the boundary of electronic decoupling, the multivariate distance correlation ($dCor$) was computed between the HOMO residual scalar and the spatial-vibrational residual triad ($zpve$, c_v, r^2). Distance correlation was selected for its ability to capture both linear and non-linear dependencies between spaces of differing dimensions. This metric was evaluated within sliding windows of sizes 2,000 and 5,000 molecules, ordered along the continuous structural gradients. Additionally, a 2D grid discretization across both the conjugation density and heteroatom ratio axes was performed to aggregate and analyze the joint structural influence on HOMO orthogonality.

Window Size	Mean Null dCor	95th Percentile (p95)	99th Percentile (p99)
2,000	0.0413	0.0545	0.0669
5,000	0.0264	0.0371	0.0403

2.5 Phase transition validation and null baseline

To identify sharp phase transitions in the coupling landscape, the numerical first derivative of the distance correlation with respect to each structural axis, $d(dCor)/d(Axis)$, was computed across the dual window sizes. Transition points were defined by the maximum absolute rate of change in the $dCor$ signal, indicating structural thresholds where HOMO energy rapidly decouples from the spatial-vibrational manifold.

To ensure that the observed coupling dynamics were not artifacts of high-dimensional noise or sample size effects, a null distribution baseline was established by shuffling the residual data. The statistical validation metrics for the null distribution across both window sizes are presented in Table 2.5. The empirical $dCor$ values were then compared against the 95th and 99th percentiles of this shuffled null distribution to confirm the statistical significance of the structural thresholds.

3 Results

3.1 Baseline decoupling and stoichiometric variance

An initial evaluation of the QM9 dataset confirms that the energy of the highest occupied molecular orbital (HOMO) is largely decoupled from the primary spatial and thermodynamic manifold of organic molecules. While basic stoichiometric atom counts (C, N, O, F, H) explain more than 99% of the variance in macroscopic properties such as zero-point vibrational energy ($zpve$) and internal energy (u_0), they capture only 8.5% of the variance in HOMO energy. By removing these baseline stoichiometric effects, the extracted non-linear residuals allow for the identification of specific structural regimes where the HOMO energy significantly re-couples with, or orthogonally decouples from, the spatial-vibrational triad (comprising the residuals for $zpve$, heat capacity c_v, and electronic spatial extent r^2).

3.2 Coupling landscape across structural gradients

The joint structural dependence between the HOMO residual and the spatial-vibrational triad was quantified using the multivariate distance correlation ($dCor$), which captures both linear and non-linear dependencies. A sliding window analysis and a 2D grid discretization across two continuous structural axes—conjugation density (aromatic rings per heavy atom) and heteroatom ratio (N+O+F per C)—reveal a highly non-uniform coupling landscape (Figure 1).

Along the conjugation axis, the coupling is highly sensitive to the density of aromatic systems. At zero conjugation density, the coupling varies significantly but remains well above the null baseline. However, a sharp decline in coupling is observed as conjugation increases.

As shown in Figure 2, numerical derivatives of the $dCor$ signal identify a primary phase transition at a conjugation density of approximately 0.1111 aromatic rings per heavy atom. At this threshold, the rate of change in coupling, $d(dCor)/d(\text{Axis})$, reaches a maximum absolute value of -129.00 (using a window size of 5,000), indicating a rapid decoupling of electronic states driven by increased global delocalization.

Conversely, the heteroatom axis exhibits a more gradual rewiring of the relationship between electronic and spatial properties, characterized by distinct local maxima. The 2D grid analysis (Figure 1) demonstrates that the highest coupling ($dCor = 0.4365$) occurs in molecules with high conjugation (0.20–0.25) and moderate heteroatom ratios (0.40–0.50). The minimum coupling ($dCor = 0.0640$) is found in molecules with low conjugation and high heteroatom ratios (1.0–1.5). A secondary transition point emerges at a heteroatom ratio of approximately 0.1250, where the coupling rate of change reaches -40.80 (window size 5,000, Figure 2). This suggests a reorganization of electronic states as the carbon scaffold is interrupted by heteroatoms.

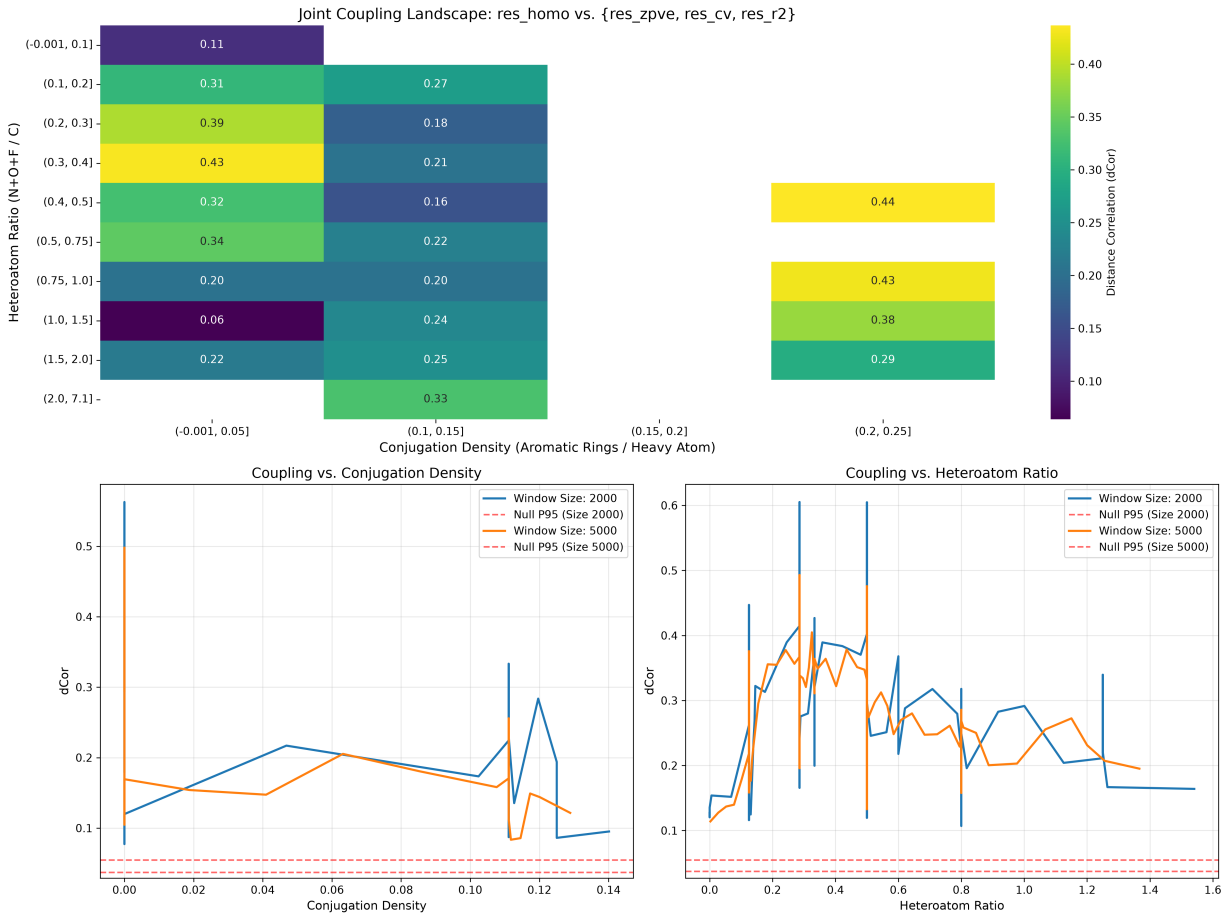

Figure 1: This figure illustrates the joint coupling landscape between the HOMO energy residual (`res_homo`) and the spatial-vibrational triad (`res_zpve`, `res_cv`, `res_r2`) across molecular structural axes. The top panel displays a 2D grid of distance correlation ($dCor$), revealing a non-uniform landscape where maximum coupling occurs at high conjugation and moderate heteroatom ratios, whereas minimum coupling appears at low conjugation and high heteroatom ratios. The bottom panels present sliding window analyses along the conjugation (left) and heteroatom (right) axes. A sharp phase transition, representing a primary decoupling switch as electronic states delocalize, is visible at a conjugation density of approximately 0.11. Conversely, the heteroatom axis shows a more fragmented, secondary modulation with a transition near 0.125. Notably, 100% of the observed $dCor$ values remain above the 95th percentile shuffled null baselines (dashed lines), demonstrating that the HOMO energy retains a statistically significant, albeit weak, relationship with the spatial-vibrational manifold even in its decoupled state.

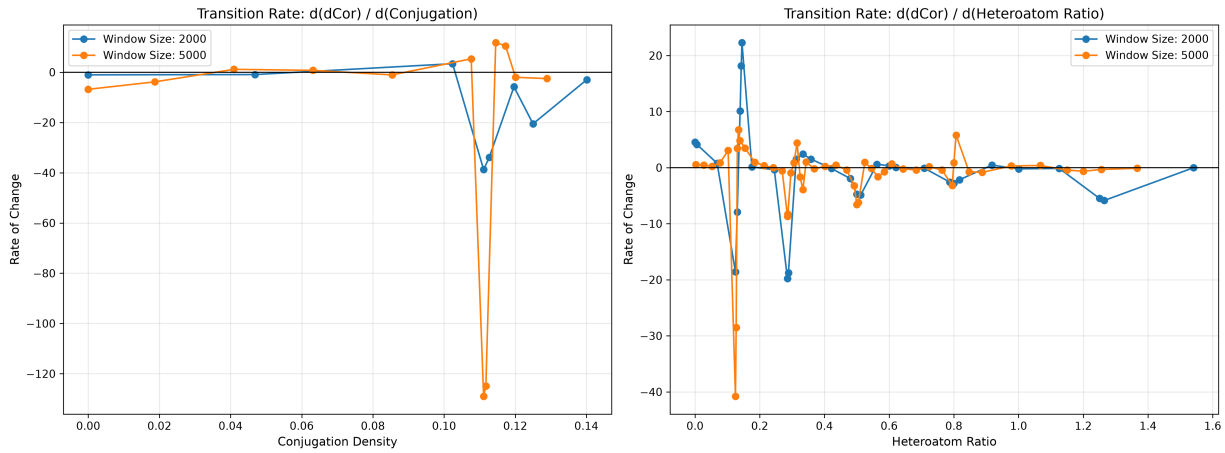

Figure 2: Transition rates of the distance correlation (`dcor`) illustrating the structural decoupling of HOMO energy residuals from the spatial-vibrational manifold across conjugation density (left) and heteroatom ratio (right) axes. The left panel reveals a primary phase transition at a conjugation density of approximately 0.1111, where the rate of change reaches a maximum absolute value of -129.00 (window size 5,000). This sharp decline indicates a decoupling of electronic states driven by increased global delocalization. The right panel displays a secondary transition at a heteroatom ratio of approximately 0.1250, with a rate of change reaching -40.80, suggesting a reorganization of electronic states as the carbon scaffold is interrupted by nitrogen or oxygen. Sliding window analyses of 2,000 and 5,000 molecules confirm these non-uniform coupling dynamics and structural phase transitions.

3.3 Statistical validation and null baselines

To ensure that the observed coupling dynamics are not artifacts of high-dimensional noise or sample size effects, the empirical results were compared against a shuffled null distribution (Table 2.5).

Across all analyzed windows for both structural axes, 100% of the observed $dCor$ values remained strictly above the 95th percentile null threshold. This statistical validation confirms that even in its most decoupled state, the HOMO energy retains a significant, albeit weak, relationship with the molecular spatial-vibrational manifold that is entirely distinct from random correlation.

3.4 Transition dynamics and geometric artifacts

The analysis highlights two distinct modes of structural influence. First, conjugation-driven decoupling acts as a primary switch; the sharp transition at a conjugation density of ~ 0.11 corresponds to the point where electronic states become sufficiently delocalized to be governed by global aromaticity rather than the local geometric constraints that drive $zpve$ and c_v. Second, heteroatom-driven heterogeneity creates a more fragmented landscape. The high standard deviation of $dCor$ across the 2D grid (0.1038) relative to its mean (0.2725) reflects the diverse ways heteroatoms participate in either lone-pair localizations or delocalized systems.

Finally, it is important to contextualize these findings within the limitations of the dataset. The analysis includes approximately 3,054 molecules identified in the original QM9 release as failing geometric consistency checks. While the use of distance correlation and large window sizes mitigates the impact of individual outliers, these uncorrected geometric artifacts may contribute to the noise floor in the spatial-vibrational manifold, particularly in highly strained or unusual structural classes. Nevertheless, the robustness of the phase transitions against the null baseline indicates that the macroscopic trends remain intact, demonstrating that the orthogonality of frontier electronic states is governed by distinct, quantifiable thresholds of molecular complexity.

4 Conclusions

The investigation into the coupling dynamics of the highest occupied molecular orbital (HOMO) energy across continuous structural gradients reveals that its orthogonality to the spatial-vibrational manifold is not a uniform property. Instead, the decoupling of frontier electronic states from macroscopic thermodynamic and spatial properties—specifically zero-point vibrational energy ($zpve$), heat capacity (c_v), and electronic spatial extent (r^2)—follows a highly structured regime governed by distinct topological thresholds.

Our analysis identifies the density of conjugation as the primary driver of this electronic decoupling. The sharp phase transition observed at approximately 0.11 aromatic rings per heavy atom acts as a structural switch; beyond this threshold, electronic states become sufficiently delocalized that they are no longer bound by the local geometric constraints dictating the spatial-vibrational baseline. Furthermore, the heteroatom-to-carbon ratio provides a secondary, more fragmented modulation of this coupling landscape, with a distinct transition occurring near a ratio of 0.125. This secondary threshold highlights the complex reorganization of electronic states as the primary carbon scaffold is interrupted by nitrogen, oxygen, or fluorine atoms.

Crucially, the observed structural transitions are statistically robust. As detailed in Table 2.5, a comparison against a shuffled null distribution confirms that 100% of the empirical distance correlation ($dCor$) values remain strictly above the 95th percentile baseline across all evaluated sliding windows.

This statistical validation demonstrates that even in its most decoupled state, the HOMO energy retains a significant, albeit weak, relationship with the molecular spatial-vibrational manifold. While the dataset retains uncorrected geometric artifacts that likely contribute to the noise floor in highly strained structural classes, the use of large window sizes and non-linear distance correlation ensures that the macroscopic phase transitions remain intact and are not driven by individual outliers.

Ultimately, these findings demonstrate that the integration of frontier electronic states into the macroscopic molecular manifold is governed by quantifiable thresholds of structural complexity. By mapping these dynamics along continuous axes rather than relying on categorical classifications, this work provides a more nuanced understanding of when and how molecular orbital energies decouple from baseline atomic determinism.