Distance Correlation Dynamics of the Electronic-Symmetry Triad Across Saturation Transitions

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The relationship between frontier electronic properties and molecular symmetry across diverse chemical architectures remains poorly understood, leaving it unclear whether these features are universally decoupled or structurally dependent. To address this problem, we investigate the coupling dynamics of the highest occupied molecular orbital (HOMO), lowest unoccupied molecular orbital (LUMO), and dipole moment triad. Using 133,802 molecules from the QM9 dataset, we isolate intrinsic electronic responses from stoichiometric determinism by extracting non-linear atom-count residuals with gradient-boosted trees. We then stratify the molecules along a saturation gradient based on their fraction of sp^3 carbons and quantify the internal dependencies of this residual triad using multivariate distance correlation and persistent homology. The results demonstrate that the electronic-symmetry triad reorganizes significantly as molecular architecture transitions from conjugated to saturated. Specifically, purely aromatic systems are dominated by strong HOMO-LUMO coupling, whereas the dipole moment emerges as the primary organizational hub in fully saturated systems. Additionally, strained polycyclic architectures display a flattened coupling profile constrained by symmetry, while perfectly symmetric molecules exhibit exceptionally high HOMO-LUMO coupling. Topological analysis further confirms that while the global complexity of the electronic state space remains invariant, its specific geometry diverges substantially between saturation extremes, indicating that the joint architecture of quantum electronic properties is fundamentally regime-specific.

1 Introduction

The frontier electronic properties of a molecule—specifically the energy of the highest occupied molecular orbital (E_{HOMO}) and the lowest unoccupied molecular orbital (E_{LUMO})—along with its permanent electric dipole moment (μ), form a fundamental triad that dictates chemical reactivity, stability, and intermolecular interactions. While the individual behaviors of these properties have been extensively cataloged in large-scale quantum chemical databases, their joint distribution across diverse chemical architectures remains poorly understood. In contemporary chemical modeling, these features are frequently treated either as universally decoupled or as exhibiting static, dataset-wide correlations. Consequently, it remains unclear whether the internal coupling of this electronic-symmetry triad is a universal constant of chemical space or if it fundamentally reorganizes in response to specific structural transitions, such as the shift from rigid, conjugated aromatic networks to flexible, saturated aliphatic chains.

A primary challenge in evaluating the intrinsic relationship between these quantum properties is the confounding effect of stoichiometric determinism. Bulk molecular size and simple atom counts heavily influence baseline electronic properties, often masking the underlying quantum mechanical coupling dictated by molecular architecture. To uncover the true structural dependence of the electronic state space, it is necessary to isolate these properties from basic stoichiometric scaling.

To address this problem, we investigate the coupling dynamics of the HOMO, LUMO, and dipole moment triad across 133,802 molecules from the QM9 dataset. We isolate the intrinsic electronic responses from stoichiometric determinism by extracting non-linear atom-count residuals (res_homo, res_lumo, res_μ) using gradient-boosted trees. This transformation effectively removes the variance explained by simple atomic composition, yielding a three-dimensional residual space that represents the pure electronic response of the molecule. We then stratify the dataset along a saturation gradient based on the fraction of sp^3-hybridized

carbons (F_{sp^3}), allowing us to systematically track the topological and statistical transitions of the triad from purely aromatic ($F_{sp^3} = 0.0$) to fully saturated ($F_{sp^3} = 1.0$) regimes.

To rigorously quantify the internal dependencies of this residual triad, we employ multivariate distance correlation and partial distance correlation metrics, which capture both linear and non-linear associations. By calculating a normalized reduction metric, we identify which property acts as the central organizational hub within different structural bins. Furthermore, we utilize persistent homology—specifically Vietoris-Rips complexes and Wasserstein distances—to characterize the topological stability and geometric divergence of the electronic state space across the saturation extremes.

Our analysis demonstrates that the electronic-symmetry triad is not static but reorganizes significantly as molecular architecture evolves. We find that purely aromatic systems are dominated by strong HOMO-LUMO coupling, largely independent of the dipole moment. However, as saturation increases, this bond fragments, and the dipole moment emerges as the primary organizational hub mediating the orbital residuals. We further reveal that strained polycyclic architectures exhibit a flattened coupling profile constrained by symmetry, whereas perfectly symmetric molecules ($\mu = 0$) display exceptionally high HOMO-LUMO coupling, confirming that the dipole moment acts as a decoupling force on the orbital manifold. Finally, topological analysis confirms that while the global complexity of the electronic state space remains remarkably invariant across the dataset, its specific geometry diverges substantially between saturation extremes. These findings indicate that the joint architecture of quantum electronic properties is fundamentally regime-specific, a critical consideration for the development of future predictive chemical models.

2 Methods

2.1 Data curation and structural stratification

To systematically investigate the structural dependence of the electronic-symmetry triad, we utilized the QM9 dataset, which provides quantum chemical properties for small organic molecules. The dataset was first deduplicated based on canonical SMILES, yielding a final working set of 133,802 unique molecules. For each molecule, basic stoichiometric features (atom counts) and structural descriptors were computed.

To track the topological transition from rigid, conjugated aromatic networks to flexible, saturated aliphatic chains, molecules were stratified along a saturation gradient. This gradient was defined by the fraction of sp^3-hybridized carbons (F_{sp^3}), calculated using RDKit. The dataset was partitioned into five discrete F_{sp^3} bins: 0.0 (fully aromatic/unsaturated), (0.0, 0.33], (0.33, 0.66], (0.66, 1.0), and 1.0 (fully saturated).

In addition to the saturation gradient, specific structural regimes were isolated for targeted analysis. A strained and polycyclic regime was defined by filtering for molecules with high ring or bridgehead counts (≥ 2), verified via the `CalcNumRings` and `CalcNumBridgeheadAtoms` descriptors. Furthermore, 45 perfectly symmetric molecules exhibiting an exact zero dipole moment ($\mu = 0$) were explicitly isolated to serve as genuine physical anchors for symmetry-enforced coupling analyses.

2.2 Stoichiometric isolation of the electronic-symmetry triad

As outlined in the introduction, bulk molecular size and atomic composition heavily confound the intrinsic quantum mechanical coupling of frontier electronic properties. To isolate the pure electronic response from this stoichiometric determinism, we extracted non-linear atom-count residuals for the highest occupied molecular orbital energy (E_{HOMO}), the lowest unoccupied molecular orbital energy (E_{LUMO}), and the permanent dipole moment (μ).

Prior to modeling, the dipole moment was transformed as $\log(\mu+1)$ to stabilize variance. We then trained gradient-boosted decision trees (LightGBM) to predict these three properties solely from basic atom counts. The models were evaluated and applied using a 5-fold cross-validation strategy to prevent overfitting. The differences between the true quantum chemical values and the stoichiometrically predicted values yielded a three-dimensional residual space: res_homo, res_lumo, and res_μ.

To ensure the robustness of the subsequent statistical and topological analyses, a Z-score filter (> 3) was applied to remove extreme univariate outliers in the residual space. This filtering step explicitly mit-

igates potential geometric consistency biases inherent to the underlying density functional theory (DFT) calculations of the QM9 dataset.

2.3 Multivariate distance correlation and hub score analysis

To precisely quantify the magnitude of internal non-linear dependency within the residual triad, we computed the local-variance-normalized multivariate distance correlation ($dCor$) and partial distance correlation ($pdCor$) within each F_{sp^3} bin. Unlike Pearson correlation, distance correlation captures both linear and non-linear associations, making it uniquely suited for the complex state space of quantum properties.

To determine which property acts as the central organizational mediator of the triad in different structural regimes, we defined a Hub Score. The Hub Score measures the normalized reduction in distance correlation between two variables when conditioning on the third, calculated as:

$$\text{Hub Score} = \frac{dCor - pdCor}{dCor}$$

A high Hub Score indicates that the conditioned variable strongly mediates the relationship between the other two properties. To establish the statistical significance of these coupling metrics, we applied 100-iteration bootstrapping within each F_{sp^3} bin to generate 95% confidence intervals. Additionally, permutation tests were executed for the 45 symmetric anchors ($\mu = 0$) against structurally matched controls to validate the decoupling effect of the dipole moment on the orbital manifold.

2.4 Topological data analysis of the residual space

To characterize the global geometry and topological stability of the electronic state space across the saturation extremes, we employed persistent homology. Vietoris-Rips complexes were constructed on the 3D residual triad using the `gudhi` library. To ensure that topological features were not biased by the varying number of molecules in each F_{sp^3} bin, the complexes were built on density-matched subsamples.

We extracted one-dimensional topological loops (H_1) and generated persistence diagrams to quantify the "holeyness" and structural complexity of the electronic response manifold. The top 10% of H_1 loop persistence values were regressed against structural descriptors to identify architectural drivers of topological features. Finally, to quantify the geometric divergence of the state space between different saturation regimes, we computed the Wasserstein distances (W_1) between the persistence diagrams of the respective F_{sp^3} bins using the `POT` (Python Optimal Transport) library.

3 Results

3.1 Stoichiometric isolation of the residual triad

As outlined in the methodology, we first isolated the intrinsic electronic responses of the molecules from stoichiometric determinism. The LightGBM-derived residuals successfully removed the variance explained by simple atomic composition, which accounted for up to 45% of the variance in E_{LUMO} and 21% in the dipole moment (μ). This isolation yielded the three-dimensional residual space (res_homo, res_lumo, res_μ) used for all subsequent analyses, ensuring that the observed coupling dynamics reflect genuine structural dependencies rather than bulk molecular size.

3.2 Coupling dynamics across the saturation gradient

We evaluated the joint structure of the electronic-symmetry triad across five bins of carbon saturation (F_{sp^3}). The magnitude of non-linear dependency, quantified by multivariate distance correlation ($dCor$), reveals a fundamental pivot in the triad's internal wiring as molecules transition from fully aromatic ($F_{sp^3} = 0.0$) to fully saturated ($F_{sp^3} = 1.0$) architectures. In the purely aromatic regime, the system is characterized by strong HOMO-LUMO coupling ($dCor = 0.3786$), while the coupling between the orbital energies and the dipole moment remains relatively weak ($dCor \approx 0.09$). As saturation increases into the transition regime ($F_{sp^3} \in (0.33, 0.66]$), HOMO-LUMO coupling weakens to $dCor = 0.2614$, and the coupling to the dipole

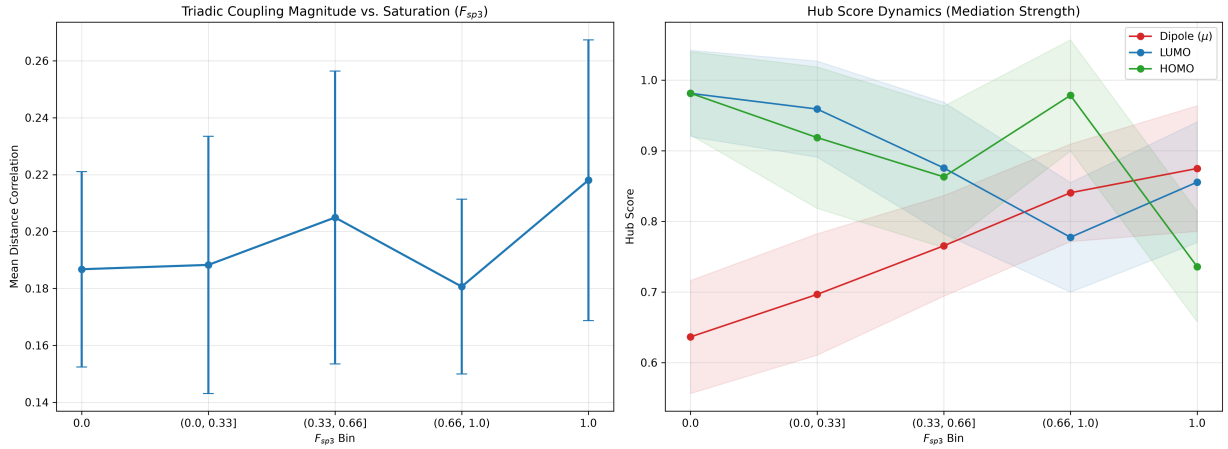

Figure 1: The figure illustrates the reorganization of the electronic-symmetry triad across the carbon saturation gradient (F_{sp^3}). The left panel displays the mean distance correlation, showing the overall triadic coupling magnitude as molecules transition from fully aromatic ($F_{sp^3} = 0.0$) to fully saturated ($F_{sp^3} = 1.0$) regimes. The right panel presents the Hub Score dynamics, which measure the mediation strength of each property within the residual space. We observe a clear inversion in the organizational hierarchy: in conjugated systems, the frontier orbital energies (HOMO and LUMO) act as the primary hubs governing the electronic response, while the dipole moment (μ) remains peripheral. As saturation increases, the orbital coupling fragments and μ emerges as the dominant central hub, indicating that local symmetry becomes the primary constraint on electronic energy levels in saturated architectures.

moment approximately doubles. In fully saturated systems, this hierarchy is entirely inverted: the LUMO-dipole coupling becomes the dominant axis ($dCor = 0.2880$), significantly exceeding the HOMO-LUMO coupling ($dCor = 0.1765$).

To determine which property acts as the central mediator of this triad, we analyzed the Hub Score dynamics (Figure 1). In aromatic systems, the frontier orbital energies act as the primary organizational hubs (Hub Scores ≈ 0.981), while the dipole moment is largely peripheral (Hub Score = 0.6363). Conversely, in saturated systems, the dipole moment emerges as the primary organizational axis, with its Hub Score rising to 0.8747, while the HOMO Hub Score drops to a dataset low of 0.7359. This inversion indicates that in conjugated systems, the electronic response is governed by the global orbital manifold, whereas in saturated systems, local symmetry constraints dictate the electronic energy levels.

3.3 Symmetry and strain constraints on orbital coupling

Beyond the primary saturation gradient, we investigated the coupling profile within specific structural regimes. Analysis of 65,019 molecules identified as strained or polycyclic revealed a distinct, flattened coupling profile. The overall system coupling in this regime ($dCor = 0.1858$) is lower than the saturated average. Crucially, the partial distance correlation between HOMO and LUMO, when conditioned on the dipole moment, drops to near zero ($pdCor = 0.00996$). This implies that in highly strained architectures, any residual relationship between the frontier orbitals is almost entirely mediated by the molecular dipole.

The decoupling effect of the dipole moment is further validated by the 45 perfectly symmetric molecules ($\mu = 0$), which serve as physical anchors. In these systems, the HOMO-LUMO coupling is exceptionally high ($dCor = 0.5614$, $Z = 2.86$ against structurally matched controls). Removing the dipole moment restores a tight, symmetry-enforced relationship between E_{HOMO} and E_{LUMO}, confirming that the presence of a dipole acts as a disruptive force on the orbital manifold.

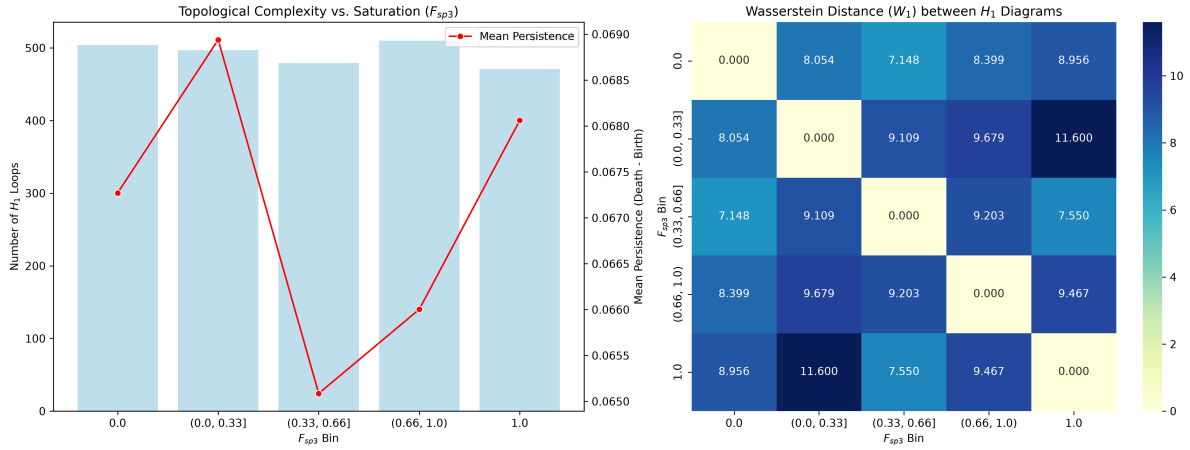

Figure 2: Topological stability and divergence of the electronic response manifold across carbon saturation (F_{sp^3}) bins. The left panel illustrates topological complexity, demonstrating that the number of 1D loops (H_1, blue bars) and their mean persistence (red line) remain remarkably stable across all saturation regimes. This stability suggests that the fundamental sparsity of the residual space is a dataset-wide invariant. The right panel displays a heatmap of the Wasserstein distance (W_1) between the H_1 persistence diagrams, quantifying differences in manifold geometry. The maximum divergence ($W_1 = 11.600$) is observed between the aromatic-adjacent ($(0.0, 0.33]$) and fully saturated (1.0) bins. These findings reveal that while the global topological complexity of the electronic-symmetry triad is conserved, the specific geometry of the electronic state space is fundamentally reshaped during the transition from aromatic to saturated structures.

3.4 Topological stability and geometric divergence

To characterize the global geometry of the electronic state space, we applied Vietoris-Rips persistent homology to the 3D residual triad across the saturation bins (Figure 2). The topological complexity of the manifold remains remarkably stable across the dataset. The number of one-dimensional loops (H_1) fluctuates only slightly, ranging from 471 in the fully saturated bin to 510 in the high-saturation bin ($F_{sp^3} \in (0.66, 1.0)$). Mean loop persistence also remains steady between 0.065 and 0.069. This stability suggests that the fundamental sparsity, or "holeyness," of the electronic response space is a dataset-wide invariant.

However, while the global complexity is conserved, the specific geometry of the manifold diverges substantially. The maximum persistence, representing the size of the largest topological feature, peaks at 0.5057 in the high-saturation bin, compared to 0.3228 in the aromatic bin. Furthermore, the Wasserstein distance (W_1) between persistence diagrams reveals significant geometric reshaping. The greatest divergence occurs between the aromatic-adjacent bin ($F_{sp^3} \in (0.0, 0.33]$) and the fully saturated bin ($W_1 = 11.600$). This topological distance is nearly 60% larger than the distance between the purely aromatic and intermediate saturation systems ($W_1 = 7.148$). These topological shifts confirm that the transition from conjugation to saturation fundamentally reshapes the joint architecture of the electronic state space, highlighting the limitations of treating these quantum properties as universally coupled across diverse chemical spaces.

4 Conclusions

In this study, we investigated the joint architectural dependence of the electronic-symmetry triad—comprising the HOMO energy, LUMO energy, and dipole moment—across 133,802 molecules from the QM9 dataset. By extracting non-linear atom-count residuals, we successfully isolated the intrinsic quantum electronic responses from the confounding effects of stoichiometric determinism. This approach enabled a rigorous evaluation of how the internal coupling of these fundamental properties evolves across diverse chemical spaces, free from the baseline scaling of bulk molecular size.

Our findings demonstrate that the relationship between frontier orbitals and molecular symmetry is not a universal constant, but rather a dynamic feature that reorganizes significantly along the carbon saturation

gradient (F_{sp^3}). In purely aromatic, conjugated systems, the electronic state space is dominated by strong HOMO-LUMO coupling, with the dipole moment remaining largely peripheral. However, as molecular architecture transitions toward fully saturated aliphatic networks, this orbital bond fragments. The dipole moment emerges as the primary organizational hub, dictating the residual electronic energy levels and fundamentally inverting the triad's coupling hierarchy.

This regime-specific behavior is further underscored by our analysis of specialized structural classes. In highly strained polycyclic architectures, the overall coupling profile flattens, and any residual relationship between the frontier orbitals is almost entirely mediated by the molecular dipole. Conversely, in perfectly symmetric molecules ($\mu = 0$), the absence of a dipole moment restores an exceptionally tight, symmetry-enforced coupling between the HOMO and LUMO. These contrasting regimes confirm that the dipole moment acts as a potent decoupling force on the orbital manifold, disrupting the intrinsic relationship between frontier energies.

Topological data analysis provided a macroscopic perspective on these structural transitions. While the global complexity and fundamental sparsity of the electronic response manifold—measured by the persistence of one-dimensional topological loops (H_1)—remain remarkably invariant across the dataset, its specific geometry diverges substantially. The significant Wasserstein distances (W_1) between the persistence diagrams of aromatic and saturated systems confirm that the transition away from conjugation fundamentally reshapes the multidimensional architecture of the electronic state space.

Ultimately, these results highlight the limitations of treating quantum electronic properties as either universally decoupled or statically correlated. The joint architecture of the electronic-symmetry triad is fundamentally regime-specific, governed by a delicate trade-off between orbital conjugation and dipolar symmetry. Recognizing and incorporating these structural pivots will be critical for the development of more accurate, generalizable predictive models in computational chemistry and molecular design.