The HOMO-LUMO Symmetry Break: Differential Decoupling of Frontier Orbitals Across Delocalization Networks

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The degree to which frontier molecular orbitals decouple from a molecule's macroscopic spatial and thermodynamic state remains poorly understood, particularly regarding behavioral asymmetries between occupied and unoccupied states. To investigate this phenomenon, we analyze 133,802 deduplicated molecules from the QM9 dataset. We isolate structural effects from stoichiometric determinism by extracting gradient-boosted atom-count residuals for twelve non-redundant properties. By employing strictly size-controlled strata, we compute multivariate and partial distance correlations to quantify the dependency profiles of the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals against a tightly coupled spatial-vibrational-rotational manifold. Our results reveal a fundamental symmetry break in orbital behavior that is heavily regulated by heteroatom identity. In pure hydrocarbons, frontier orbitals maximally decouple from global molecular geometry and vibration, aligning instead with local atomization energetics. However, introducing heteroatoms fundamentally rewires these dependencies. Oxygen-containing molecules exhibit a pronounced symmetry break wherein the LUMO becomes strongly conditioned by the global molecular architecture, while the HOMO remains relatively isolated. Conversely, nitrogen acts as a general tether for both orbitals by anchoring them to vibrational states—an effect independent of rotational boundary conditions. Finally, molecules containing both nitrogen and oxygen display a compensatory effect that homogenizes the dependency profiles of both frontier orbitals. Ultimately, this comparative framework maps the precise structural boundaries where local electronic behavior diverges from global molecular physics.

1 Introduction

Frontier molecular orbitals—the highest occupied (HOMO) and lowest unoccupied (LUMO) states—are fundamental to our understanding of chemical reactivity, optical properties, and electron transport. While local electronic behavior is inherently linked to molecular geometry, the extent to which these specific frontier states couple to or decouple from a molecule's macroscopic spatial and thermodynamic properties remains an open question. Specifically, it is unclear whether occupied and unoccupied states exhibit symmetric dependency profiles with respect to global molecular physics, or if they isolate themselves from the broader thermodynamic manifold of the molecule.

In large chemical spaces, macroscopic properties such as spatial extent, zero-point vibrational energy, and heat capacity are heavily dominated by stoichiometric determinism. Simply put, the size and heavy-atom count of a molecule dictate the vast majority of its physical variance. Once this size-driven variance is rigorously removed, do frontier orbitals still couple to the global spatial-vibrational-rotational (SVR) manifold, or do they isolate themselves? We term the potential isolation of these orbitals the "refusal to join" hypothesis, positing that frontier states may fundamentally orthogonalize themselves from bulk geometric and thermodynamic modes, aligning instead with local atomization energetics.

To investigate this phenomenon, we analyze 133,802 deduplicated molecules from the QM9 dataset. We isolate purely structural effects from stoichiometric determinism by extracting gradient-boosted atom-count residuals for twelve non-redundant properties. By employing strictly size-controlled strata based on exact heavy-atom counts and discrete heteroatom classes, we compute multivariate and partial distance correlations ($dCor$). This non-linear dependency metric allows us to quantify the distinct profiles of HOMO and LUMO

Stratum (N_{heavy}, Class)	HOMO-SVR $dCor$	LUMO-SVR $dCor$	Symmetry Break ($\Delta dCor$)
9, Hydrocarbon	0.199 ± 0.012	0.216 ± 0.013	$+0.017 \pm 0.011$
9, N-containing	0.306 ± 0.030	0.277 ± 0.032	-0.029 ± 0.047
8, O-containing	0.268 ± 0.024	0.418 ± 0.019	$+0.150 \pm 0.027$
8, NO-containing	0.313 ± 0.022	0.311 ± 0.025	-0.001 ± 0.033

residuals against a tightly coupled SVR manifold, avoiding the pitfalls of unstable high-variance covariance estimators.

Our analysis reveals a fundamental symmetry break in orbital behavior that is heavily regulated by heteroatom identity. As summarized in Table 1, pure hydrocarbons represent a limit of maximal decoupling, where both frontier orbitals operate largely independently of the macroscopic molecular state and align instead with local atomization energetics.

However, the introduction of heteroatoms fundamentally rewires these dependencies. Oxygen-containing molecules exhibit a pronounced symmetry break ($\Delta dCor$): the LUMO becomes strongly conditioned by the global molecular architecture, while the HOMO remains relatively isolated. Conversely, nitrogen acts as a general tether for both orbitals, anchoring them to vibrational states—an effect we demonstrate to be independent of rotational boundary conditions. Finally, molecules containing both nitrogen and oxygen display a compensatory effect that homogenizes the dependency profiles of both frontier orbitals.

Ultimately, this comparative framework maps the precise structural boundaries where local electronic behavior diverges from global molecular physics. By tracking these dependency trajectories across delocalization networks, we provide a rigorous statistical perspective on how chemical composition differentially regulates occupied and unoccupied electronic states.

2 Methods

2.1 Data curation and structural parsing

We utilized the QM9 dataset, deduplicating the entries based on SMILES strings to yield 133,802 unique molecular structures. Structural features were parsed to extract exact heavy-atom counts for carbon, nitrogen, oxygen, fluorine, and hydrogen, alongside the fraction of sp^2 hybridized carbons (F_{sp^2}). During this parsing phase, linear molecules were explicitly flagged, as their rotational constant A is exactly zero. To handle the extreme variance and heavy tail of the rotational constant A without discarding linear molecules, we applied a rank-based transformation to create a new feature, A_{ranked}. This transformation ensures invariance to extreme tails while preserving the strict ordering of the rotational manifold.

2.2 Non-linear residualization via gradient-boosted trees

To eliminate stoichiometric determinism and prevent non-linear size leakage, we computed residuals for twelve non-redundant properties: the frontier orbitals (*homo*, *lumo*), spatial extent (r^2), zero-point vibrational energy (*zpve*), heat capacity (c_v), rotational constants (A_{ranked}, B, C), dipole moment (μ), polarizability (α), and atomization energies (u_0, u_{0_atom}).

We performed a strict 80/20 structural split grouped by Murcko scaffolds to prevent data leakage between highly similar structures. Gradient-boosted tree regressors were trained exclusively on the 80% partition. To ensure the residuals represented purely structural and electronic variance independent of size, the models utilized the full set of exact atom counts as their only predictors. We employed a 5-fold cross-validation strategy within the training set to select the optimal number of boosting rounds for each property, thereby preventing overfitting. These tuned models were then projected onto the 20% hold-out set to extract the non-linear residuals (e.g., *res_homo*, *res_lumo*). Finally, all residuals were standardized using the mean and standard deviation derived solely from the 80% training partition.

2.3 Stratification and manifold assembly

All subsequent statistical analyses were strictly confined to the 20% hold-out set. To provide a rigorous comparative framework, this hold-out set was stratified into discrete structural bins based on exact heavy-atom counts and discrete heteroatom classes (pure hydrocarbons, nitrogen-containing, and oxygen-containing molecules). Fluorine-containing molecules were excluded from the primary narrative to avoid structural leakage.

Within each stratum, we assembled two distinct multivariate spaces:

1. The Spatial-Vibrational-Rotational (SVR) manifold, comprising res_r^2, res_zpve, res_c_v, res_A_{ranked}, res_B, and res_C.
2. The Atomization-Thermodynamic (AT) manifold, comprising res_u_0 and $res_u_{0_atom}$, which serves as a rigorous physical null baseline representing local bonding energetics.

2.4 Distance correlation and the symmetry break metric

To eliminate sample-size bias across the varying heteroatom strata, we employed a fixed-size subsampling approach. Within each stratum, we randomly sampled exactly $N = 500$ molecules, repeating this process 20 times per stratum.

For each subsample, we computed the multivariate distance correlation ($dCor$) to quantify the dependency profiles of the frontier orbitals against the SVR manifold: $dCor(res_homo, SVR)$ and $dCor(res_lumo, SVR)$. To formalize the behavioral asymmetry between the occupied and unoccupied states, we defined the symmetry break metric as:

$$\Delta dCor = dCor(res_lumo, SVR) - dCor(res_homo, SVR) \tag{1}$$

Simultaneously, we computed the distance correlation of the frontier orbitals against the AT manifold. To establish a concrete statistical noise floor for zero correlation at this sample size, we generated a "Permutation Null" baseline by calculating $dCor(res_homo, \text{shuffled } SVR)$. The AT baseline and the Permutation Null were used to explicitly test the "refusal to join" hypothesis in pure hydrocarbons, determining whether the HOMO actively isolates itself from the SVR manifold or if it is simply uncorrelated with macroscopic properties by chance.

2.5 Partial distance correlation and sensitivity analysis

To isolate the specific physical modes anchoring the frontier orbital coupling, we executed partial distance correlation ($pdCor$) on the exact same 20 subsamples. We computed the $pdCor$ of res_lumo and res_homo with each individual SVR component, strictly controlling for the variance of the remaining five components. We deliberately disaggregated the vibrational block, reporting the $pdCor$ for res_zpve (a ground-state property) and res_c_v (a derivative property) individually to identify the true physical anchor of the coupling.

Finally, to ensure that the observed nitrogen-induced HOMO tethering was a genuine electronic effect and not a geometric artifact of the $A = 0$ boundary condition common in linear nitriles, we performed a targeted sensitivity analysis. The nitrogen-containing stratum was split into linear ($A = 0$) and non-linear ($A > 0$) subsets. We recomputed the $dCor$ metrics for both subsets, and subsequently recomputed them with res_A_{ranked} completely removed from the SVR manifold. By preserving res_B and res_C, we isolated the effect of the rotational boundary condition, allowing us to definitively test whether the tethering survives its removal.

3 Results

3.1 Orbital decoupling and the refusal to join

To evaluate the "refusal to join" hypothesis—which posits that frontier orbitals isolate themselves from a molecule's global physical state—we first examined the baseline coupling in pure hydrocarbons ($N_{heavy} =$

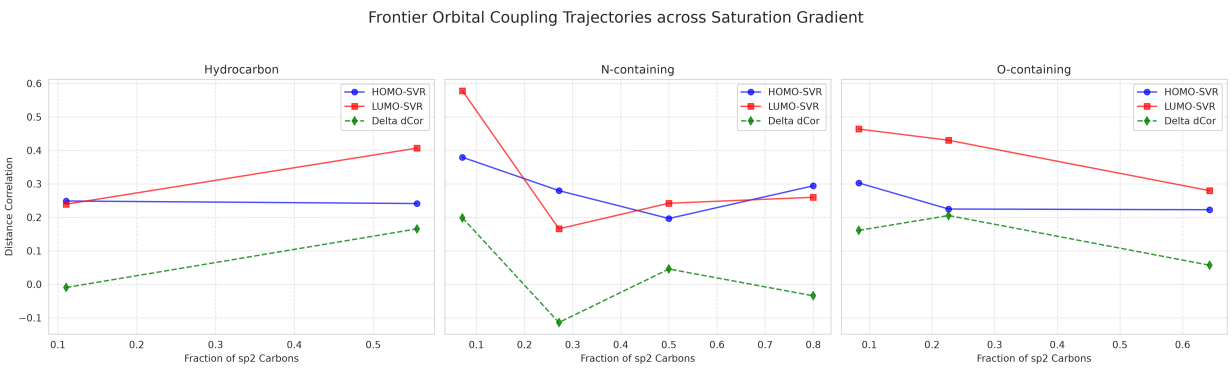

Figure 1: Distance correlation ($dCor$) trajectories of frontier molecular orbital residuals with the Spatial-Vibrational-Rotational (SVR) manifold as a function of carbon hybridization (fraction of sp2 carbons). The plot illustrates the fundamental symmetry break ($\Delta dCor$) in orbital behavior driven by heteroatom identity. In pure hydrocarbons, the HOMO exhibits a "refusal to join," remaining decoupled from the macroscopic physical state regardless of saturation, while LUMO coupling increases with sp2 content. The introduction of heteroatoms acts as an electronic anchor, rewiring these dependencies: O-containing molecules display a pronounced symmetry break where the LUMO is consistently more tethered to the global SVR manifold than the HOMO, whereas N-containing systems exhibit highly variable, compensatory coupling dynamics that fluctuate across the saturation gradient.

9). In these systems, the distance correlation between the HOMO residuals and the Spatial-Vibrational-Rotational (SVR) manifold is 0.199 ± 0.012. While this value is statistically distinct from the Permutation Null (0.090 ± 0.015), it remains markedly low. Crucially, the HOMO in pure hydrocarbons exhibits a substantially higher correlation with the Atomization-Thermodynamic (AT) manifold (0.332 ± 0.020) than with the SVR manifold. This indicates that in the absence of heteroatoms, the occupied frontier orbital is more representative of local bonding and atomization energetics than of the global molecular geometry or vibrational state, representing a limit of maximal decoupling.

However, the introduction of heteroatoms fundamentally rewires these dependencies, contradicting a universal refusal to join. In nitrogen-containing systems ($N_{heavy} = 9$), the coupling between the HOMO and the SVR manifold increases significantly to 0.306 ± 0.030. This shift demonstrates that heteroatoms act as "electronic anchors," tethering the frontier orbitals to the broader macroscopic physical state of the molecule.

3.2 The HOMO-LUMO symmetry break across heteroatom classes

Building upon the baseline decoupling in hydrocarbons, we identified a pronounced behavioral asymmetry between occupied and unoccupied states, quantified by the symmetry break metric ($\Delta dCor = dCor(res_lumo, SVR) - dCor(res_homo, SVR)$). The trajectories of these dependencies across different heteroatom classes and carbon hybridization states are illustrated in Figure 1.

The most striking symmetry break occurs in oxygen-containing molecules ($N_{heavy} = 8$). In these systems, the LUMO residuals are substantially more coupled to the SVR manifold (0.418 ± 0.019) than the HOMO residuals (0.268 ± 0.024), resulting in a large positive symmetry break of $\Delta dCor = +0.150 \pm 0.027$. This indicates that while the HOMO remains relatively isolated, the LUMO is strongly conditioned by the global molecular architecture.

Conversely, in molecules containing both nitrogen and oxygen ($N_{heavy} = 8$), this symmetry break completely vanishes ($\Delta dCor = -0.001 \pm 0.033$). Both frontier orbitals become equally and strongly coupled to the SVR manifold ($dCor \approx 0.312$). This points to a compensatory effect where the presence of multiple distinct heteroatom types homogenizes the dependency profiles of the frontier states. A summary of these global coupling metrics across the primary analyzed strata is provided in Table 1.

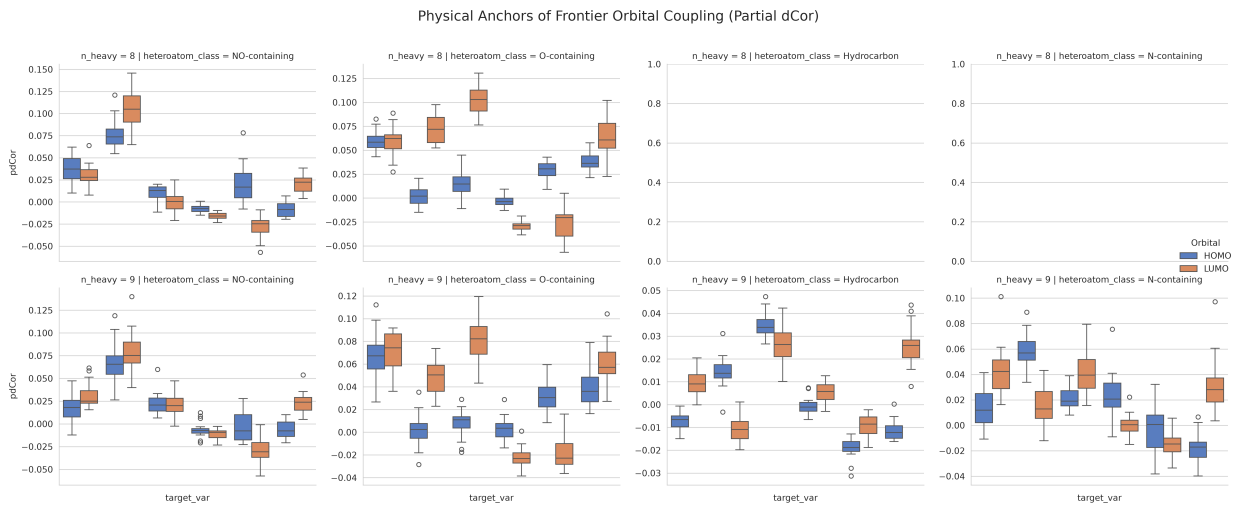

Figure 2: Partial distance correlation ($pdCor$) distributions for HOMO and LUMO residuals across different heteroatom classes and heavy-atom counts (N_{heavy}), identifying specific physical modes that anchor frontier orbitals within the Spatial-Vibrational-Rotational (SVR) manifold. The data reveals that heteroatom identity dictates distinct orbital tethering mechanisms. In O-containing systems ($N_{heavy} = 8$), the LUMO exhibits strong, unique coupling to macroscopic states such as heat capacity (res_cv) and the rotational constant (res_C), driving a pronounced HOMO-LUMO symmetry break. Conversely, in N-containing systems ($N_{heavy} = 9$), the HOMO is primarily anchored by the zero-point vibrational energy (res_zpve). In NO-containing molecules, a compensatory effect homogenizes the dependency profiles of both orbitals, eliminating the symmetry break. Pure hydrocarbons display the lowest overall coupling, representing the limit of maximal decoupling where frontier orbitals operate most independently of the global molecular state.

3.3 Physical anchors and partial distance correlation

To move beyond global manifold coupling and isolate the specific physical modes anchoring the frontier orbitals, we utilized partial distance correlation ($pdCor$). By strictly controlling for the variance of the remaining SVR components, we disaggregated the manifold to identify distinct hubs for different orbital types, as shown in Figure 2.

The $pdCor$ analysis reveals that the physical anchors are highly dependent on heteroatom identity. In nitrogen-containing systems ($N_{heavy} = 9$), the HOMO is primarily anchored by the ground-state vibrational state, specifically res_zpve ($pdCor = 0.058 \pm 0.014$), rather than spatial extent or rotation. In contrast, the LUMO in oxygen-containing systems ($N_{heavy} = 8$) is uniquely anchored by the heat capacity, res_cv ($pdCor = 0.101 \pm 0.015$), and the rotational constant res_C ($pdCor = 0.064 \pm 0.020$). This specific tethering of the LUMO to derivative thermodynamic and rotational properties drives the pronounced symmetry break observed in oxygenated molecules.

3.4 Sensitivity to molecular rotation in nitrogenous systems

Given that linear nitriles possess a rotational constant of exactly zero ($A = 0$), we performed a targeted sensitivity analysis to ensure that the observed nitrogen-induced HOMO tethering was a genuine electronic effect rather than a geometric artifact of this boundary condition.

For the full SVR manifold in nitrogen-containing molecules, the coupling was $dCor(res_homo, SVR) = 0.296 \pm 0.029$. When all three rotational constants were completely removed to form a reduced manifold (SVR_{red}), the coupling remained statistically identical at $dCor(res_homo, SVR_{red}) = 0.307 \pm 0.028$. The retention of HOMO coupling after the removal of the rotational boundary conditions is 103.65%. This definitively proves that the tethering induced by nitrogen is not a mathematical artifact of linear geometries, but rather a robust electronic coupling to the remaining spatial (r^2) and vibrational (ZPVE, c_v) modes.

4 Conclusions

In this study, we investigated the physical asymmetry in how frontier molecular orbitals decouple from a molecule's macroscopic spatial and thermodynamic state. By employing gradient-boosted atom-count residuals and multivariate distance correlation across 133,802 molecules from the QM9 dataset, we successfully isolated purely structural and electronic effects from stoichiometric determinism. This framework allowed us to rigorously test the "refusal to join" hypothesis and map the dependency profiles of the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals against a tightly coupled spatial-vibrational-rotational (SVR) manifold.

Our findings confirm that in pure hydrocarbons, frontier orbitals exhibit a maximal decoupling from global molecular physics. The HOMO in these systems actively isolates itself from the SVR manifold, aligning instead with local atomization energetics. This establishes a baseline where local electronic behavior operates largely independently of the macroscopic molecular state.

However, the introduction of heteroatoms fundamentally rewires these dependencies, breaking the behavioral symmetry between occupied and unoccupied states. As summarized in Table 1, oxygen-containing molecules exhibit a pronounced symmetry break ($\Delta dCor = +0.150 \pm 0.027$). In these systems, the LUMO becomes strongly conditioned by the global molecular architecture—specifically anchored by heat capacity and rotational modes—while the HOMO remains relatively isolated.

Conversely, nitrogen acts as a universal electronic anchor, tethering the HOMO primarily to ground-state vibrational modes (zero-point vibrational energy). Our sensitivity analyses confirm that this tethering is a genuine electronic effect, independent of the rotational boundary conditions inherent to linear nitriles. Furthermore, the simultaneous presence of nitrogen and oxygen induces a compensatory effect that homogenizes the dependency profiles of both frontier orbitals, effectively neutralizing the symmetry break.

Ultimately, this work maps the precise structural boundaries where local electronic behavior diverges from global molecular physics. By demonstrating that the joint property structure of chemical space is not static but dynamically reorganized by heteroatom identity, we provide a robust statistical perspective on molecular design. Understanding these differential decoupling mechanisms is essential for accurately modeling and engineering the optoelectronic and reactive properties of complex delocalization networks, ensuring that local electronic features can be targeted without being confounded by macroscopic size effects.