Gromov-Wasserstein Measurement of Local-Global Physical Decoupling Across Conjugation Regimes

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The apparent decoupling of localized frontier orbital energies from global spatial and thermodynamic molecular properties remains poorly understood, raising the question of whether this phenomenon is a universal physical constant or depends on molecular structure. To investigate this, we reframe the anomaly as a dynamic boundary between distinct physical manifolds. Using the QM9 dataset, we extract size-independent property residuals via atom-count linear regression and partition them into a local electronic manifold and a global thermodynamic-spatial manifold. We then stratify the dataset based on carbon hybridization and aromaticity, employing the entropic Gromov-Wasserstein distance to rigorously quantify the topological divergence between these manifolds across structural classes. Our analysis reveals that the decoupling of frontier orbitals is heavily modulated by structural conjugation rather than being a universal artifact. While localized electronic states and global thermodynamics are highly divergent in saturated acyclic systems, they undergo significant topological reintegration in conjugated and aromatic regimes, demonstrating that extended conjugated systems fundamentally reorganize the geometric relationship between different domains of molecular physics.

1 Introduction

The prediction and interpretation of molecular properties often rely on the assumption that a molecule's structural architecture dictates its physical behavior in a unified, coherent manner. However, empirical observations frequently reveal a persistent decoupling between localized electronic states—most notably the energies of frontier orbitals such as the Highest Occupied Molecular Orbital (HOMO)—and global macroscopic features like thermodynamic stability and spatial extent. This phenomenon, often referred to as the HOMO anomaly, raises a fundamental question: is this decoupling a universal physical constant across all chemical space, or is it a dynamic feature modulated by specific structural motifs?

Historically, investigations into property decoupling have relied on simple linear correlation heuristics or sparse combinatorial binning. These methods struggle to separate intrinsic chemical organization from the overwhelming stoichiometric determinism driven by molecular size and heavy-atom count. Furthermore, standard correlation metrics fail to capture the multi-dimensional geometric relationships between distinct domains of molecular physics, treating properties as isolated scalars rather than interconnected manifolds. To rigorously investigate this local-global physical decoupling, we must move beyond pairwise correlations and evaluate the structural boundaries between entire physical regimes.

1.1 A topological manifold framework

In this work, we reframe the HOMO anomaly as a dynamic structural boundary between two distinct physical manifolds. Using the QM9 dataset, we isolate intrinsic chemical organization by extracting size-independent property residuals via exact atom-count linear regression. To manage heavy-tailed variance and tame extreme rotational sensitivity, we explicitly apply a $\log(A + 1)$ transformation to the rotational constants prior to residual extraction. The resulting residual space is partitioned into two distinct domains:

- A **local electronic manifold**, comprising the HOMO energy (res_homo), the HOMO-LUMO gap (res_gap), the dipole moment (res_μ), and the isotropic polarizability (res_α).

- A **global thermodynamic-spatial manifold**, comprising the zero-point vibrational energy (res_zpve), heat capacity (res_cv), spatial extent (res_r^2), and the transformed rotational constant (res_C).

To quantify the relationship between these manifolds, we stratify the deduplicated dataset into statistically dense bins based on RDKit-derived sp^2 carbon fractions and aromatic ring counts. We then employ the entropic Gromov-Wasserstein (GW) distance to measure the topological divergence between the local and global property manifolds within each structural class. Unlike standard distance metrics that require direct point-to-point alignment, the GW distance provides a mathematically rigorous measure of the formal transport cost between metric measure spaces. By tracking this topological divergence as a function of conjugation, and comparing it against a rotation-based null model that preserves covariance structure, we can isolate true chemical organization from marginal property distributions.

1.2 Conjugation as a physical re-wiring agent

Our analysis reveals that the decoupling of frontier orbitals is heavily modulated by structural conjugation rather than being a universal artifact of molecular physics. We demonstrate that while localized electronic states and global thermodynamics are highly divergent in saturated acyclic systems, extended π-systems force a fundamental reorganization of these physical domains. In acyclic unsaturated systems, the local and global manifolds reach a state of near-perfect topological alignment, leaving the HOMO as the primary source of any remaining divergence.

Crucially, in polycyclic aromatic regimes, we observe a topological reintegration: the inclusion of the HOMO energy actively decreases the divergence between the electronic and thermodynamic descriptions. This indicates that extended conjugated systems and aromatic rings act as physical re-wiring agents, forcing localized quantum states to reintegrate with the macroscopic spatial-thermodynamic architecture. By replacing simple correlation heuristics with a rigorous measure of geometric divergence, this framework provides a new perspective on how molecular structure dictates the interplay between distinct domains of physical behavior.

2 Methods

2.1 Data preprocessing and residual extraction

Building upon the theoretical framework introduced previously, we utilized the QM9 dataset, comprising 133,802 deduplicated molecules, to isolate intrinsic chemical organization from stoichiometric determinism. To remove the dominant effects of molecular size and heavy-atom count, we performed exact ordinary least squares (OLS) linear regression on atom counts (C, N, O, F, H) to extract size-independent property residuals.

To manage heavy-tailed variance and tame extreme rotational sensitivity, we applied a $\log(A+1)$ transformation to the rotational constant A prior to regression. This transformation ensures that the nine linear molecules in the dataset (where $A = 0$) are handled consistently without introducing numerical artifacts into the resulting residual (res_C). The redundant LUMO energy residual (res_lumo) was explicitly dropped from the analysis. The remaining property residuals were strictly partitioned into two distinct domains: a local electronic manifold consisting of res_homo, res_gap, res_μ, and res_α, and a global thermodynamic-spatial manifold consisting of res_zpve, res_cv, res_r^2, and res_C.

2.2 Structural stratification

To evaluate the topological divergence across different structural regimes, we partitioned the dataset into a two-dimensional quantile grid based on RDKit-derived sp^2 carbon fractions and aromatic ring counts. To eliminate size-proxy artifacts, we calculated the mean heavy-atom count for each bin. Bins that deviated significantly from the global mean or lacked a sufficient range of heavy-atom counts were merged or filtered out.

This stratification yielded five statistically dense bins, each containing over 900 molecules: saturated acyclic (Bin 0_0), partially unsaturated acyclic (Bin 0_1), highly unsaturated acyclic (Bin 0_2), mono-aromatic (Bin 1_2), and polycyclic aromatic (Bin 2_2).

2.3 Gromov-Wasserstein topological divergence

We quantified the geometric divergence between the local and global property manifolds within each structural class using the entropic Gromov-Wasserstein (GW) distance. For each bin, intra-manifold distance matrices were computed, and the GW distance was evaluated using the `sinkhorn` solver from the Python Optimal Transport (POT) library. To optimize memory and computational efficiency, the solver was executed with chunked processing distributed across 14 CPUs.

To rigorously isolate the contribution of the HOMO anomaly, the GW calculation was performed twice per bin: once including the HOMO residual (GW_{full}) and once excluding it (GW_{no_homo}). The relative contribution of the HOMO energy to the manifold's structural integrity was then extracted as a normalized percentage delta:

$$\Delta HOMO = \frac{GW_{full} - GW_{no_homo}}{GW_{full}} \times 100$$

2.4 Rotation-based null model

To distinguish regime-specific physical coupling from universal manifold alignment, we implemented an intra-bin rotation-based null model. Bypassing standard combinatorial scaffold shuffling, we applied a random orthogonal matrix to the data within each bin. This transformation strictly preserves the covariance structure—and thus the intrinsic shape of the manifolds—while destroying the specific point-to-point alignment between the local and global properties.

By comparing the true GW distance (GW_{true}) to this structural null distance (GW_{struct}), we defined an Organization Index (OI) to measure the degree of specific physical organization:

$$OI = \frac{GW_{struct} - GW_{true}}{GW_{struct}}$$

A positive OI indicates that the true physical coupling between the manifolds is stronger than a random pairing of manifolds with identical covariance structures.

2.5 Principal component analysis

Finally, to physically determine which specific property axes drive the manifold dimensionality collapse or expansion under extended π-systems, we extracted the local Principal Component Analysis (PCA) eigenvalue spectra for the bins exhibiting maximal and minimal relative GW distances against the null model. We computed and compared the eigenvectors (loadings) for the top two principal components, allowing us to track the shifting influence of individual properties across different conjugation regimes.

3 Results

3.1 Stoichiometric baseline and structural stratification

To rigorously investigate the decoupling of localized frontier orbitals from bulk molecular physics, it is first necessary to isolate intrinsic chemical organization from the dominant effects of molecular size and composition. As detailed in the methodology, we extracted size-independent property residuals via exact atom-count linear regression. The necessity of this step is highlighted by the stoichiometric baseline, which demonstrates the proportion of variance explained purely by atom counts prior to residual extraction (Figure 1). Thermodynamic properties such as zero-point vibrational energy and heat capacity exhibit strong stoichiometric determinism, whereas electronic and geometric properties like the HOMO energy and rotational constants are poorly predicted by simple atom counts.

Following residual extraction, the deduplicated QM9 dataset was stratified into a two-dimensional grid based on sp^2 carbon fractions and aromatic ring counts. This partitioning yielded five statistically dense structural regimes ($N > 900$), ranging from saturated acyclic molecules to polycyclic aromatic systems. As shown in Figure 2, the mean heavy atom count remains relatively uniform across these bins (ranging from 8.50 to 8.91), confirming that our structural stratification successfully avoids severe molecular size biases that could otherwise confound the topological analysis.

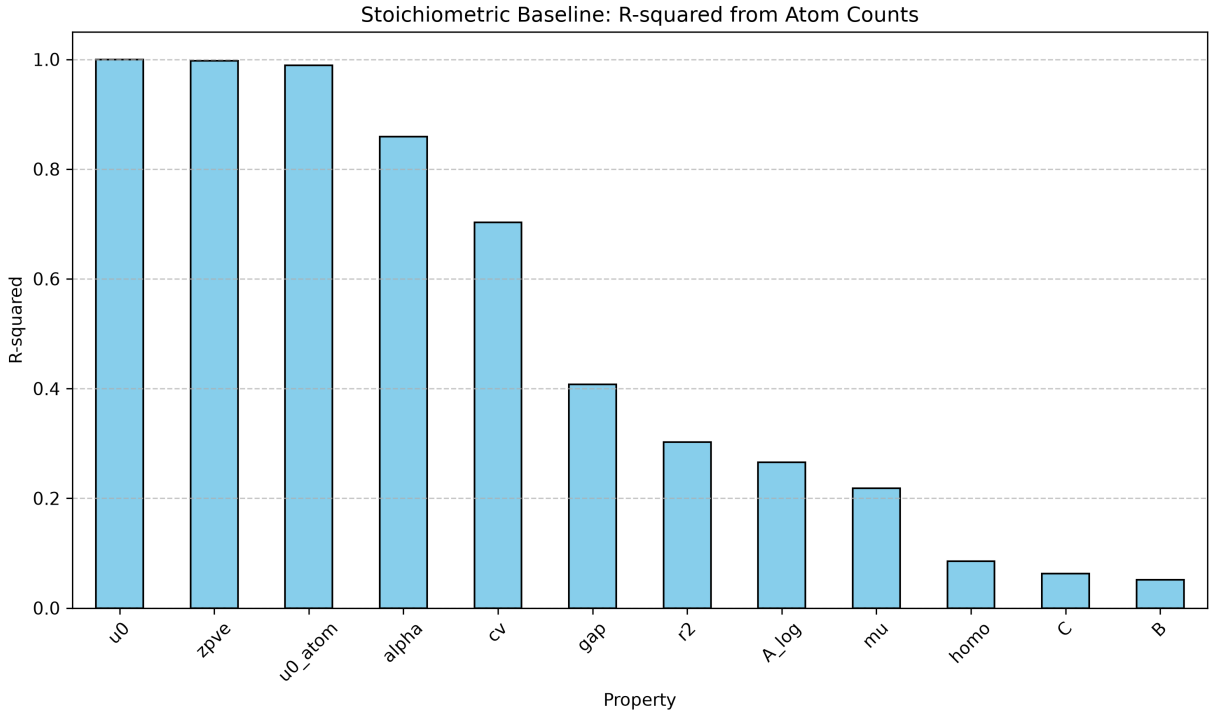

Figure 1: Bar chart displaying the R^2 values from the stoichiometric baseline linear regression on atom counts (C, N, O, F, H) for various molecular properties. The plot illustrates the proportion of variance explained purely by molecular size and composition prior to residual extraction. Thermodynamic properties such as zero-point vibrational energy (zpve) and heat capacity (cv) exhibit high R^2 values, indicating a strong dependence on basic stoichiometry. In contrast, electronic and geometric properties like the HOMO energy (homo), dipole moment (μ), and rotational constants (A_log, C, B) show significantly lower R^2 values. This demonstrates that frontier orbital energies and specific spatial configurations are poorly predicted by simple atom counts, justifying the extraction of regression residuals to isolate true chemical organization and accurately evaluate the topological divergence between the local electronic and global thermodynamic manifolds.

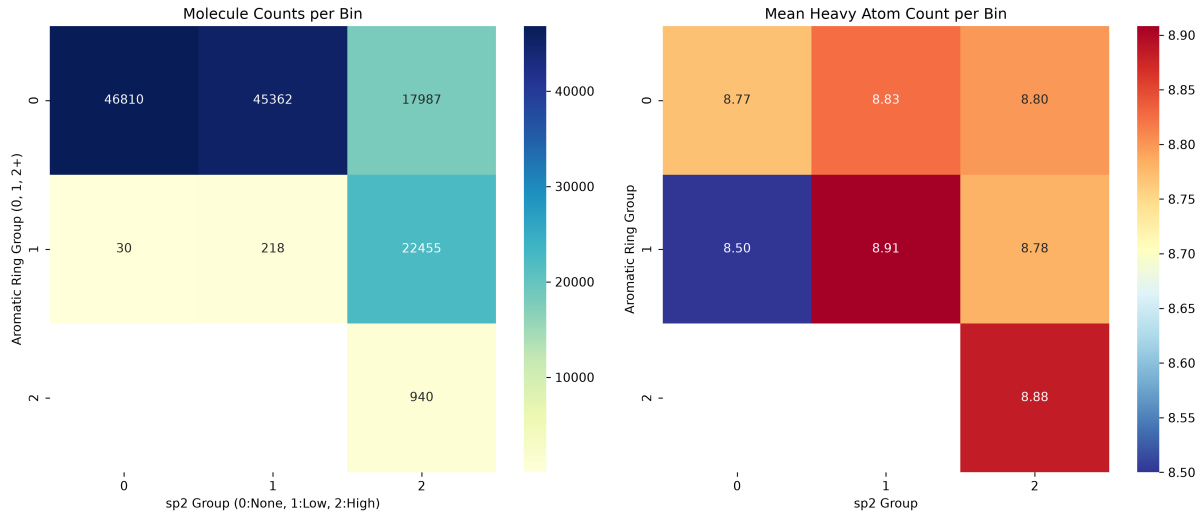

Figure 2: Heatmaps displaying the structural stratification of the QM9 dataset based on sp^2 carbon fraction and aromatic ring count. The left panel shows the molecule counts per bin, identifying the five statistically dense regimes ($N > 900$) used for topological analysis: saturated acyclic (Bin 0_0), partially unsaturated acyclic (Bin 0_1), highly unsaturated acyclic (Bin 0_2), mono-aromatic (Bin 1_2), and polycyclic aromatic (Bin 2_2). The right panel illustrates the mean heavy atom count across these bins. The relatively uniform heavy atom distribution (ranging from 8.50 to 8.91) demonstrates that the structural partitioning avoids severe molecular size biases. This supports the methodology of using atom-count residuals to isolate intrinsic chemical organization, allowing for an unbiased evaluation of how the HOMO energy topologically diverges or integrates across different structural and aromatic regimes.

Table 1: Entropic Gromov-Wasserstein distance (GW_{full}) and the percentage contribution of the HOMO energy ($\Delta HOMO$) across five structural regimes.

Bin ID	Rings	sp^2 Group	Count	GW_{full}	$\Delta HOMO$ (%)
0_0	0	0	46,810	3.79×10^{-1}	27.35%
0_1	0	1	45,362	3.95×10^{-31}	94.31%
0_2	0	2	17,987	5.16×10^{-30}	98.79%
1_2	1	2	22,455	2.07×10^{-25}	93.06%
2_2	2	2	940	1.69×10^{-1}	-25.61%

3.2 Topological divergence and the HOMO contribution

With the dataset stratified and size effects removed, we computed the entropic Gromov-Wasserstein distance (GW_{full}) between the local electronic and global thermodynamic-spatial manifolds for each structural bin. To isolate the specific influence of the HOMO anomaly, we also calculated the percentage contribution of the HOMO energy ($\Delta HOMO$) to the total topological misalignment. The results across the five structural regimes are summarized in Table 1 and visualized in Figure 3.

In saturated acyclic systems (Bin 0_0), we observe a significant base-level topological divergence ($GW_{full} = 0.379$), with the HOMO energy contributing 27.35% to this misalignment. However, as conjugation increases in acyclic and mono-aromatic systems (Bins 0_1, 0_2, and 1_2), the absolute GW distance collapses dramatically toward zero (on the order of 10^{-31} to 10^{-25}). This collapse indicates a near-perfect structural alignment between the electronic and thermodynamic manifolds. In these highly conjugated regimes, the HOMO energy becomes the near-sole driver of the remaining infinitesimal divergence, reaching a contribution of 98.79% in Bin 0_2.

Crucially, this trend does not hold indefinitely. In polycyclic aromatic systems (Bin 2_2), the overall manifold divergence rebounds to 0.169, but the HOMO contribution undergoes a topological reversal,

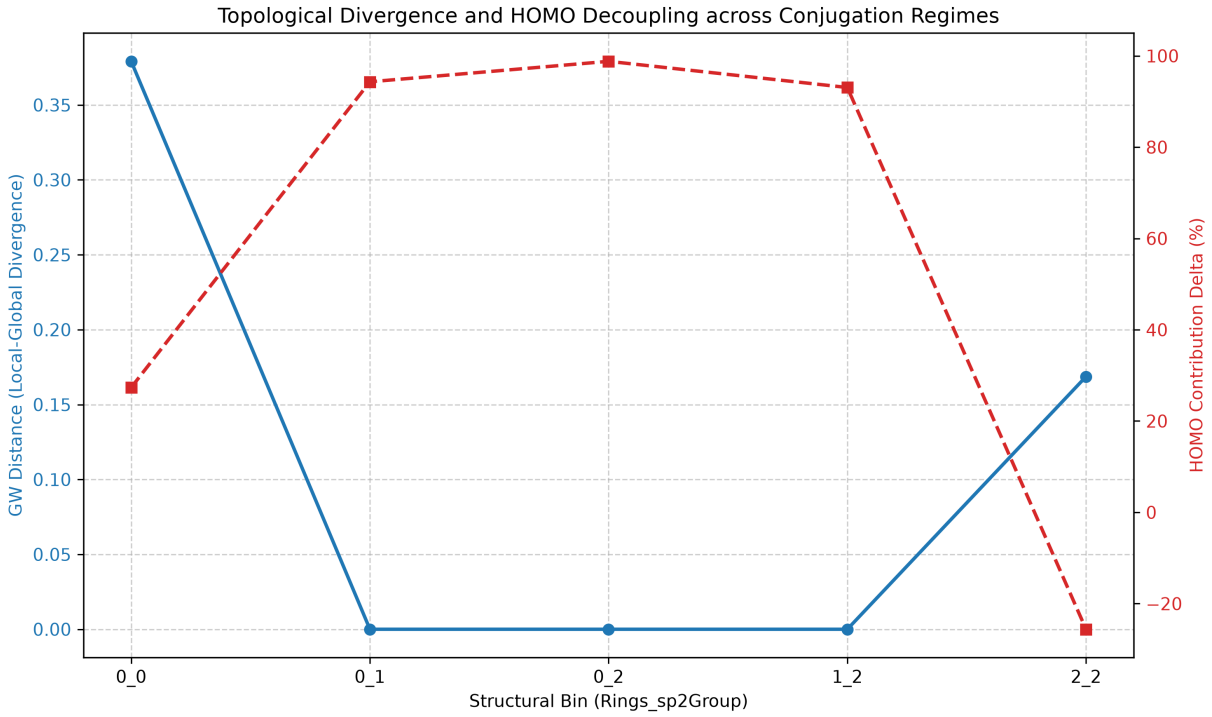

Figure 3: Entropic Gromov-Wasserstein distance (GW_{full}) between local electronic and global thermodynamic-spatial manifolds (blue solid line) and the percentage contribution of the HOMO energy ($\Delta HOMO$) to this divergence (red dashed line) across five structural bins. In saturated acyclic systems (bin 0_0), a significant base-level topological divergence is observed. As conjugation increases in acyclic and mono-aromatic systems (bins 0_1, 0_2, and 1_2), the absolute GW_{full} distance collapses toward zero, indicating near-perfect structural alignment, while the HOMO energy becomes the near-sole driver of the remaining infinitesimal divergence (reaching up to 98.79%). Conversely, in polycyclic aromatic systems (bin 2_2), the divergence rebounds, but the HOMO contribution undergoes a topological reversal to -25.61%. This negative $\Delta HOMO$ demonstrates that in extended π-systems, the inclusion of frontier orbital energies coherently integrates with and improves the alignment of the global thermodynamic architecture, rather than acting as a decoupled anomaly.

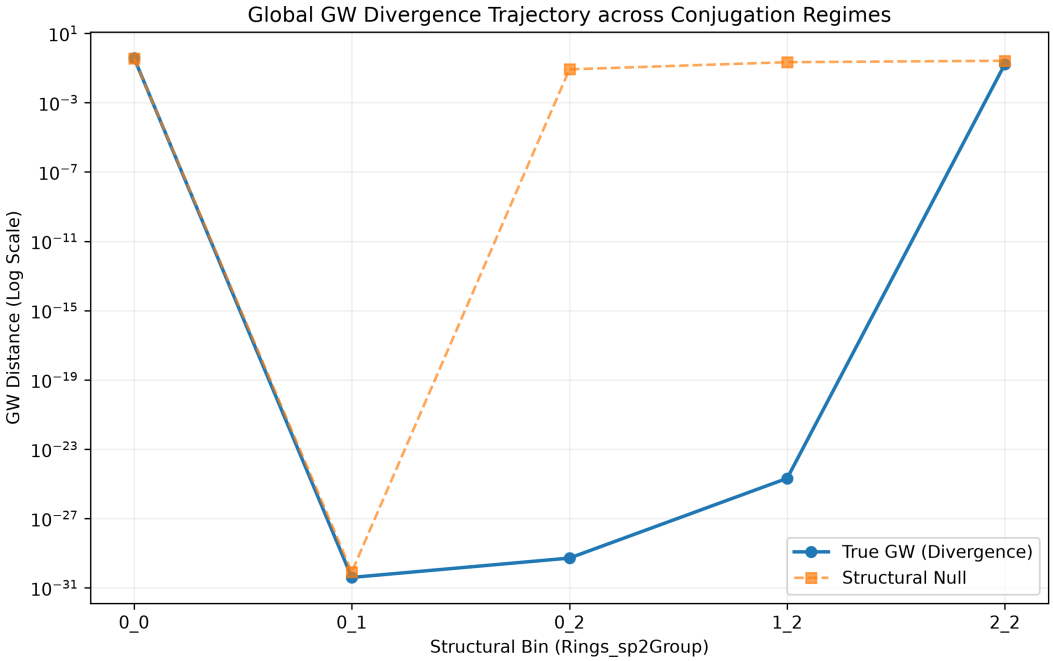

Figure 4: Comparison of the true Entropic Gromov-Wasserstein distance (GW_{true}) and a structural null model (GW_{struct}) across five molecular structural bins, illustrating the topological divergence between local electronic and global thermodynamic-spatial manifolds. The logarithmic y-axis highlights a massive collapse in GW_{true} for acyclic unsaturated and mono-aromatic regimes (Bins 0_1, 0_2, and 1_2), indicating near-perfect structural alignment. The significant deviation of GW_{true} from GW_{struct} in these highly conjugated regimes demonstrates a unique, regime-specific physical coupling rather than universal manifold alignment. Conversely, saturated acyclic (Bin 0_0) and polycyclic aromatic (Bin 2_2) systems exhibit higher baseline divergence, with Bin 0_0 showing no stronger coupling than random manifold pairings. Ultimately, the trajectory reveals that extended π-systems and aromatic rings fundamentally reorganize the geometric relationship between electronic and thermodynamic domains.

dropping to -25.61%. This negative $\Delta HOMO$ indicates that the inclusion of the HOMO energy actively decreases the divergence between the manifolds. In extended π-systems, frontier orbital energies are therefore more coherently integrated with the global thermodynamic architecture than other electronic properties, challenging the notion of a universal HOMO anomaly.

3.3 Validation via structural null model

To ensure that the observed manifold alignments are driven by regime-specific physical coupling rather than universal geometric similarities, we compared the true intra-bin GW distance (GW_{true}) against a rotation-based structural null model (GW_{struct}). The trajectory of this comparison across the structural bins is shown in Figure 4.

We quantified the strength of this coupling using the Organization Index (OI), detailed in Table 2 and visualized in Figure 5. The saturated bin (0_0) displays a negative OI (-0.054), implying that the coupling between local and global manifolds in saturated molecules is no stronger than a random pairing of manifolds from different structural classes. In stark contrast, the highly unsaturated (0_2) and polycyclic aromatic (2_2) bins exhibit high positive OI values (1.000 and 0.357, respectively). This confirms that extended π-systems and aromatic rings act as physical re-wiring agents, driving a unique organization that aligns electronic and thermodynamic properties significantly more than the baseline structural null.

Table 2: Null model statistics comparing the true intra-bin GW distance (GW_{true}) to a structural null model (GW_{struct}). The Organization Index (OI) quantifies the strength of regime-specific physical coupling.

Bin ID	GW_{true}	GW_{struct}	Organization Index (OI)
0_0	0.379	0.360	-0.054
0_2	5.16×10^{-30}	0.084	1.000
2_2	0.169	0.263	0.357

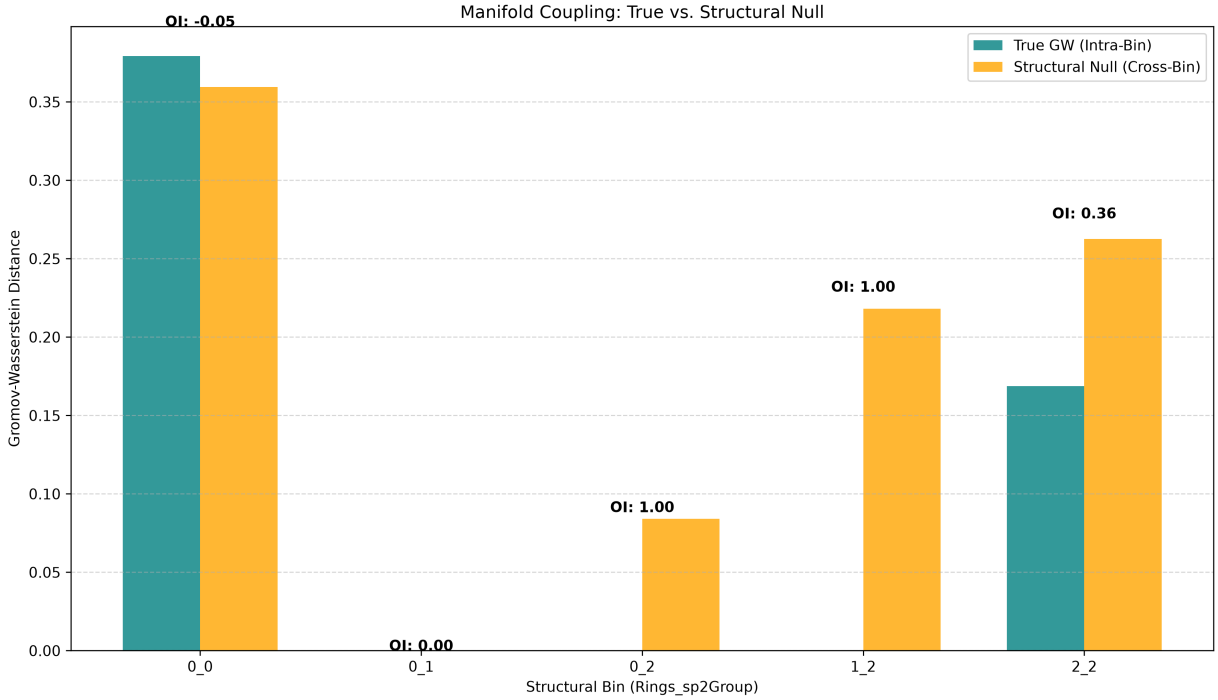

Figure 5: Comparison of the true intra-bin Gromov-Wasserstein distance (GW_{true}) and the cross-bin structural null model (GW_{struct}) across five molecular structural regimes, illustrating the topological divergence between local electronic and global thermodynamic-spatial manifolds. The annotated Organization Index (OI) quantifies the strength of regime-specific physical coupling. Saturated acyclic molecules (Bin 0_0) exhibit a negative OI (-0.05), implying that their manifold coupling is no stronger than random pairings of different structural classes. Conversely, highly unsaturated acyclic (Bin 0_2) and aromatic (Bins 1_2 and 2_2) systems display high positive OI values. This demonstrates that extended π-systems and aromatic rings act as re-wiring agents, driving a unique physical organization that fundamentally aligns electronic and thermodynamic properties significantly more than the baseline structural null.

Table 3: Principal Component Analysis (PCA) eigenvalue and loading analysis for the regimes exhibiting maximum (Bin 0_0) and minimum (Bin 0_2) relative divergence.

Regime	Bin	PC	Exp. Var.	*load_res_homo*	*load_res_α*	*load_res_C*
Max Divergence	0_0	1	42.94%	-0.181	0.436	-0.105
		2	19.45%	0.659	0.252	-0.003
Min Divergence	0_2	1	37.94%	-0.063	0.375	-0.364
		2	23.21%	0.637	0.315	0.008

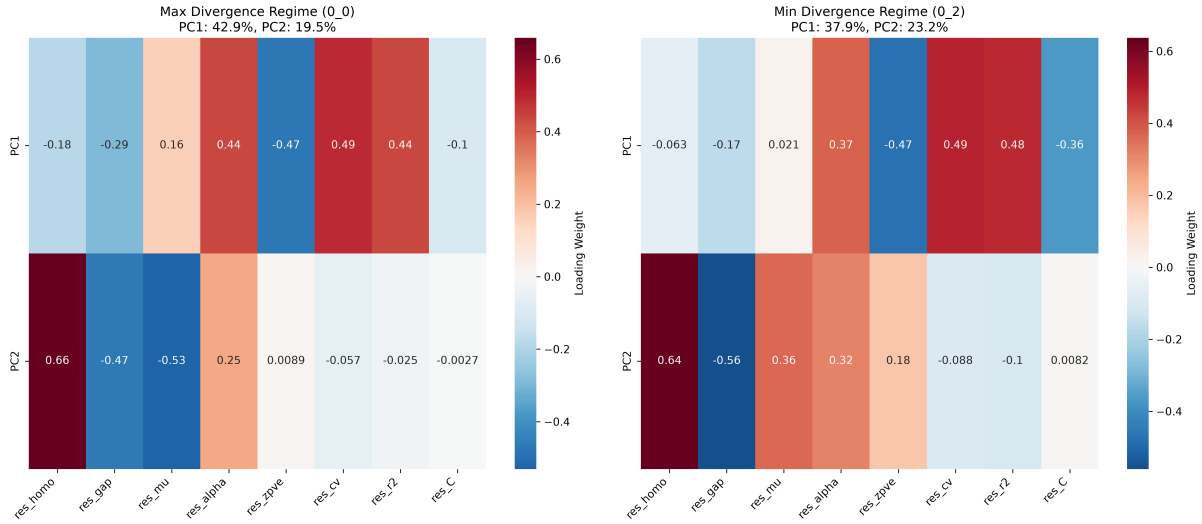

Figure 6: Principal Component Analysis (PCA) loading weights for the first two principal components (PC1 and PC2) across local electronic and global thermodynamic-spatial residual properties for the maximum divergence (saturated, Bin 0_0) and minimum divergence (highly unsaturated, Bin 0_2) regimes. In the saturated regime, PC1 is dominated by spatial and thermodynamic properties, while the HOMO energy (`res_homo`) is largely excluded from PC1 and instead dominates PC2, illustrating its decoupling from the broader manifold. In the highly unsaturated regime, the explained variance of PC1 decreases (from 42.9% to 37.9%), indicating an expansion of manifold dimensionality. Concurrently, the loading magnitude of the rotational constant (`res_C`) in PC1 substantially increases, demonstrating that increased conjugation forces a tighter alignment between geometric responses and the broader property manifold, even as the HOMO energy maintains its partially decoupled status in PC2.

3.4 Dimensionality expansion and manifold loadings

To physically interpret the axes driving the manifold dimensionality collapse and expansion, we performed Principal Component Analysis (PCA) on the combined local and global residuals for the regimes exhibiting maximum (Bin 0_0) and minimum (Bin 0_2) relative divergence. The eigenvalue spectra and corresponding loadings for the top two principal components are presented in Table 3 and Figure 6.

In the saturated regime, the first principal component (PC1) explains 42.94% of the variance and is dominated by spatial and thermodynamic properties. The HOMO energy (*res_homo*) is largely excluded from this primary axis (loading -0.181) and instead dominates PC2 (loading 0.659), illustrating its decoupling from the broader physical manifold.

Transitioning to the highly unsaturated regime, the explained variance of PC1 decreases to 37.94%, indicating an expansion of manifold dimensionality. This expansion is accompanied by a significant reorganization of property couplings. While *res_homo* remains primarily isolated in PC2 (loading 0.637), the loading magnitude of the rotational constant (*res_C*) in PC1 more than triples, shifting from -0.105 to -0.364. This demonstrates that increased conjugation forces a tighter alignment between geometric responses and the broader property manifold, fundamentally altering the structural boundaries between distinct domains

of molecular physics.

4 Conclusions

The apparent decoupling of localized frontier orbital energies from global spatial and thermodynamic properties has long been treated as a persistent anomaly in molecular physics. In this work, we moved beyond simple pairwise correlations and stoichiometric determinism to reframe this phenomenon as a dynamic structural boundary between distinct physical manifolds. By extracting size-independent property residuals and employing the entropic Gromov-Wasserstein distance, we rigorously quantified the topological divergence between local electronic and global thermodynamic-spatial domains across the QM9 dataset.

Our findings demonstrate that the decoupling of the HOMO energy is not a universal physical constant, but rather a highly variable feature modulated by structural conjugation. In saturated acyclic systems, the local and global manifolds exhibit significant baseline divergence, confirming a fundamental decoupling where the HOMO energy operates largely independently of the broader thermodynamic architecture. However, as molecules transition into highly unsaturated acyclic regimes, extended π-systems act as physical re-wiring agents. In these structures, the electronic and thermodynamic manifolds achieve near-perfect topological alignment, leaving the localized HOMO energy as the near-sole driver of any remaining geometric divergence. As supported by our principal component analysis, this dimensionality expansion highlights a regime where geometric responses and global properties become tightly coupled.

Most strikingly, this structural modulation culminates in a topological reversal within polycyclic aromatic systems. Rather than acting as a decoupled anomaly, the inclusion of the HOMO energy in these extended aromatic regimes actively decreases the topological divergence between the physical manifolds. This indicates that in polycyclic aromatics, frontier orbital energies are coherently integrated with the macroscopic spatial-thermodynamic structure, completely subverting the baseline anomaly observed in saturated systems.

Ultimately, this work establishes that molecular structure does more than merely dictate individual chemical properties; it fundamentally reorganizes the geometric and topological relationships between entire domains of physical behavior. By leveraging optimal transport metrics to evaluate these multi-dimensional property manifolds, we provide a mathematically rigorous framework for understanding when and why localized quantum states integrate with bulk molecular physics. These insights offer a new perspective for molecular design and property prediction, emphasizing that the coupling between electronic and thermodynamic features is a dynamic, structure-dependent property in its own right.