Spectral Graph Clustering and Permutation Testing of Distance Correlation Networks to Map Property Rewiring

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Understanding how molecular physical properties reorganize in response to topological complexity remains a significant challenge, as traditional property partitions often assume static dependencies across chemical space. To systematically quantify this architectural rewiring, we model the twelve-dimensional non-redundant property space of the QM9 dataset as scaffold-specific dependency networks. By isolating stoichiometric determinism through the extraction of atom-count regression residuals, we compute distance correlation matrices for statistically populated Murcko scaffold families. We then measure the macroscopic deformation of these networks by calculating the weighted mean absolute difference between cyclic scaffolds and a size-matched acyclic baseline, evaluating the rewiring as a continuous gradient driven by molecular rigidity and rotatable bond counts. Our analysis reveals that small, rigid, and aromatic systems, such as benzene and pyranone derivatives, exhibit the most significant global property rewiring. Furthermore, rigorous permutation testing highlights a fundamental fracture in the vibrational-spatial axis: the strong coupling between zero-point vibrational energy and heat capacity observed in rigid molecules systematically decays as flexibility increases. In contrast, rotational-spatial couplings remain remarkably stable across the structural gradient. Supported by sensitivity analyses confirming these transitions are genuine architectural effects rather than variance artifacts, these findings demonstrate that vibrational-spatial coupling serves as the central axis of physical reorganization, providing a more fluid and accurate representation of molecular property dependencies.

1 Introduction

1.1 The dynamic nature of molecular property space

Understanding how molecular physical properties reorganize in response to topological complexity remains a significant challenge in computational chemistry. Traditional approaches to mapping chemical space often partition molecular properties into discrete, static categories—such as thermodynamic, geometric, and electronic blocks. These partitions implicitly assume that the underlying dependencies between physical properties remain constant across different molecular architectures. However, as molecules transition from highly flexible, saturated chains to constrained, rigid polycyclic systems, their internal physical state spaces fundamentally reorganize. Quantifying this architectural "rewiring" is essential for developing a more fluid and accurate representation of molecular property dependencies.

1.2 Isolating architectural rewiring from stoichiometric determinism

A primary obstacle in observing genuine topological effects is the overwhelming influence of stoichiometric determinism. Raw thermodynamic properties, for instance, are almost entirely dictated by molecular size and atom counts, which can mask the higher-dimensional organization driven by molecular connectivity. To systematically isolate structural effects, we model the twelve-dimensional non-redundant property space of the QM9 dataset as scaffold-specific dependency networks. By extracting atom-count regression residuals, we remove the baseline stoichiometric variance, ensuring that the resulting topologies are driven strictly by the strength of physical couplings rather than absolute magnitude or compositional scaling.

These isolated chemical variances are then used to compute distance correlation ($dCor$) matrices for statistically populated Murcko scaffold families. Treating these matrices as weighted property-dependency

Scaffold	Heavy Atoms	wMAD	Size (N)
‘O=c1cccco1‘ (2-Pyranone derivative)	9	0.1828	108
‘N=c1cccco1‘	9	0.1679	134
‘O=c1cccc[nH]1‘ (2-Pyridone derivative)	9	0.1661	117
‘c1ccccc1‘ (Benzene)	9	0.1626	225
‘C1C2NC3C1C23‘ (Tricyclic aziridine)	9	0.1625	118

networks allows us to measure the macroscopic deformation of the physical state space. We quantify this deformation by calculating the weighted mean absolute difference (wMAD) between the dependency networks of cyclic scaffolds and a size-matched, fully acyclic baseline.

1.3 Macroscopic deformation and the gradient of flexibility

By evaluating property rewiring as a continuous gradient driven by molecular rigidity and the number of rotatable bonds, our spectral-statistical framework reveals precisely which non-linear physical couplings fracture under topological constraints. We find that small, rigid, and aromatic systems exhibit the most profound global property rewiring. As detailed in Table 1.3, scaffolds such as benzene, pyranone, and pyridone derivatives demonstrate the highest divergence from flexible acyclic behavior, indicating that ring constraints fundamentally alter their internal physical state spaces.

Furthermore, rigorous permutation testing across this structural gradient highlights a fundamental fracture in the vibrational-spatial axis. Specifically, the strong coupling between zero-point vibrational energy and heat capacity observed in rigid molecules systematically decays as molecular flexibility increases. In contrast, rotational-spatial couplings—such as the relationship between electronic spatial extent and the rotational constants—remain remarkably stable, acting as a preserved geometric identity across the structural gradient.

Supported by sensitivity analyses that confirm these transitions are genuine architectural effects rather than artifacts of variance collapse or geometric noise, this work demonstrates that vibrational-spatial coupling serves as the central axis of physical reorganization. By mapping these discrete topological thresholds and continuous physical transitions, we provide a comprehensive framework for understanding how molecular physics adapts to structural complexity.

2 Methods

2.1 Data curation and redundancy handling

To systematically map the property space of the QM9 dataset, we first deduplicated 83 molecules with identical SMILES strings to prevent data leakage during downstream network construction. We isolated a core twelve-dimensional space of non-redundant physical properties: rotational constants (A, B, C), dipole moment (μ), isotropic polarizability (α), frontier orbital energies (HOMO, LUMO), electronic spatial extent (r^2), zero-point vibrational energy (ZPVE), internal energy at 0 K (U_0), heat capacity (C_v), and atomization energy at 0 K (U_0^{atom}). Redundant thermodynamic properties and the HOMO-LUMO gap were explicitly excluded from the primary network to avoid rank-deficient covariance matrices and artificially tight couplings, though the gap was retained as a separate series for diagnostic validation.

Linear molecules (where $A = 0$ or $B = C$) were flagged for representative baseline calculations but excluded from specific scaffold network topologies to prevent their extreme rotational constants from biasing the graphs. Finally, a $\log(A + 1)$ transformation was applied to the rotational constant A to manage its extreme heavy tail. Pre- and post-transformation distance correlations ($dCor$) were evaluated to ensure that underlying physical dependencies were not artificially altered by this scaling.

2.2 Stoichiometric baseline validation and residual extraction

As raw thermodynamic properties are heavily dictated by molecular size, we isolated genuine structural effects from compositional determinism by extracting exact counts of carbon, nitrogen, oxygen, fluorine, and

hydrogen atoms from the RDKit-parsed SMILES strings. A linear Ridge regression model was trained for each of the twelve properties using only these atom counts as predictors across the entire dataset, including the flagged linear molecules to accurately capture the global stoichiometric effect.

The residuals from these linear models were extracted to represent the isolated chemical variance. These twelve residual columns were then globally standardized (z-scored) using the dataset-wide mean and standard deviation. This global standardization ensures that subsequent graph topologies are driven by the strength of physical dependencies rather than the absolute magnitude of variance within individual scaffold families.

2.3 Structural descriptors and network construction

To capture molecular symmetry and rigidity, we extracted the Murcko scaffold for each molecule and calculated the number of rotatable bonds alongside a continuous Rigidity Score, defined as the ratio of ring atoms to total heavy atoms. To address potential geometric inconsistencies in the uncleaned QM9 dataset, we identified and flagged the top 5% of extreme outliers within each scaffold family based on their r^2 and μ residuals.

The dataset was then stratified by heavy-atom count. Within each size stratum, acyclic molecules (lacking a Murcko scaffold) were isolated to serve as a size-matched, flexible baseline. For the remaining cyclic molecules, we retained scaffold families containing at least 100 molecules per size stratum. We computed the 12×12 pairwise distance correlation matrix for each size-stratified acyclic baseline and each retained cyclic scaffold family. These matrices were partitioned into physically meaningful sub-blocks: thermodynamic/vibrational (ZPVE, U_0, C_v, U_0^{atom}), geometric/rotational (A, B, C, r^2, α), and electronic (HOMO, LUMO, μ).

2.4 Quantifying macroscopic deformation and variance artifacts

To measure the global macroscopic deformation of the property networks, we calculated the weighted mean absolute difference (wMAD) between the *dCor* matrix of each cyclic scaffold and its corresponding size-matched acyclic baseline. The differences were weighted by the baseline *dCor* strength, ensuring that the fracturing of previously strong couplings was penalized more heavily than fluctuations in weak, noisy edges.

To verify that the observed fracturing of couplings was not merely an artifact of reduced dynamic range in rigid molecules, we computed the variance of the residuals for each property within each scaffold family. The edge strength was then correlated with the product of the variances of the two connected nodes, confirming that the rewiring signal remained independent of variance collapse.

2.5 Gradient analysis and permutation testing

Rather than clustering scaffolds into discrete groups, the rewiring was analyzed as a continuous gradient driven by molecular rigidity. Molecules were grouped by their number of rotatable bonds and their continuous Rigidity Score. Within these groups, we computed the absolute *dCor* for critical edges, specifically targeting the tightly bound vibrational-spatial (residual ZPVE versus residual C_v) and rotational-spatial (residual r^2 versus residual C) axes to evaluate monotonic physical transitions.

To pinpoint the chemical drivers of this rewiring, we performed rigorous permutation tests on the difference in specific edge weights between the extremes of the rigidity gradient (e.g., 0 rotatable bonds versus the maximum number), dynamically downsampling the larger group to match sizes during permutation. Finally, two sensitivity analyses were conducted: the first repeated the *dCor* calculations and permutation tests excluding the previously flagged top 5% of r^2 and μ outliers, and the second included the linear molecules to determine if extreme rigidity amplified the observed trends. An additional diagnostic check regressed the gap variable from both HOMO and LUMO targets, computing the *dCor* of their resulting residuals to ensure compatibility with the non-linear framework.

3 Results

3.1 Persistence of structural effects post-stoichiometry

To isolate genuine architectural effects from compositional determinism, all twelve non-redundant properties were conditioned on a five-feature atom-count baseline (C, N, O, F, H). While raw thermodynamic properties are almost entirely determined by stoichiometry (e.g., $R^2 \approx 1.000$ for internal energy at 0 K, U_0), the joint structure of the resulting residuals reveals significant persistent couplings. A principal component analysis on these residuals demonstrates that the effective dimensionality of the property space increases upon removing size effects. Specifically, eight principal components are required to explain 90% of the variance in the residual space, compared to only seven for the raw data. This expansion confirms that stoichiometric determinism masks a higher-dimensional, topology-driven organization of physical properties that persists after baseline extraction.

3.2 Quantified rewiring and scaffold-specific deformation

Using the weighted mean absolute difference (wMAD) of distance correlation ($dCor$) matrices, we quantified the macroscopic deformation of property networks for 15,989 unique Murcko scaffolds relative to a 13,998-molecule acyclic baseline. The most significant rewiring is observed in small, rigid, and aromatic cyclic systems. As detailed in Table 1.3, the top five scaffolds exhibiting the greatest divergence from flexible acyclic behavior are predominantly constrained ring systems, such as benzene and pyranone derivatives. These high wMAD values represent a systematic fracture of global property couplings, indicating that aromatic and rigid ring constraints fundamentally rewire the internal physical state space compared to size-matched open chains.

3.3 The continuous gradient of rewiring

The transition from flexible to rigid systems was mapped continuously using the number of rotatable bonds and a structural rigidity score (ratio of ring atoms to total heavy atoms). This gradient analysis reveals a profound divergence in how different physical axes respond to topological constraints.

The most dramatic architectural effect is the fracture of the vibrational-spatial axis, specifically the coupling between zero-point vibrational energy residuals (res_{zpve}) and heat capacity residuals (res_{cv}). In the rigid limit (0 rotatable bonds), these properties are tightly coupled ($dCor = 0.8350$). However, as molecular flexibility increases to the flexible limit (6 rotatable bonds), this coupling fractures significantly, dropping to $dCor = 0.4891$. This monotonic decline exhibits a strong negative trend correlation ($r = -0.8948$). Permutation testing across 1,000 iterations confirms that this observed difference ($\Delta dCor = 0.332$) is highly significant ($p < 0.001$), as illustrated in Figures 1 and 2.

In stark contrast, the rotational-spatial coupling between electronic spatial extent (res_{r^2}) and the molecular moment of inertia (res_C) remains remarkably stable across the same flexibility gradient. While this relationship exhibits a slight trend with the rigidity score ($r = 0.8253$), the difference in coupling strength between the 0-bond and 6-bond extremes ($\Delta dCor = 0.0244$) fails to reach statistical significance in permutation tests ($p = 0.487$). This stability implies that the relationship between spatial extent and rotational inertia constitutes a fundamental geometric identity that resists architectural rewiring.

3.4 Sensitivity and artifact validation

To ensure that the observed network deformations are genuine physical transitions rather than statistical artifacts, we conducted rigorous sensitivity analyses. First, we verified that the fracturing of couplings is not a side effect of reduced dynamic range in rigid molecules. The correlation between the change in $dCor$ and the product of property variances within scaffold families is negligible ($r = -0.0086$, $p = 0.389$), confirming that the rewiring signal is independent of variance collapse.

Second, the results remain highly stable when excluding the top 5% of geometric outliers (identified via res_{r^2} and res_μ). In this cleaned subset, the vibrational-spatial fracture remains robust at $\Delta dCor = 0.335$. Finally, a diagnostic check on the frontier orbital energies confirms that the fundamental electronic identity

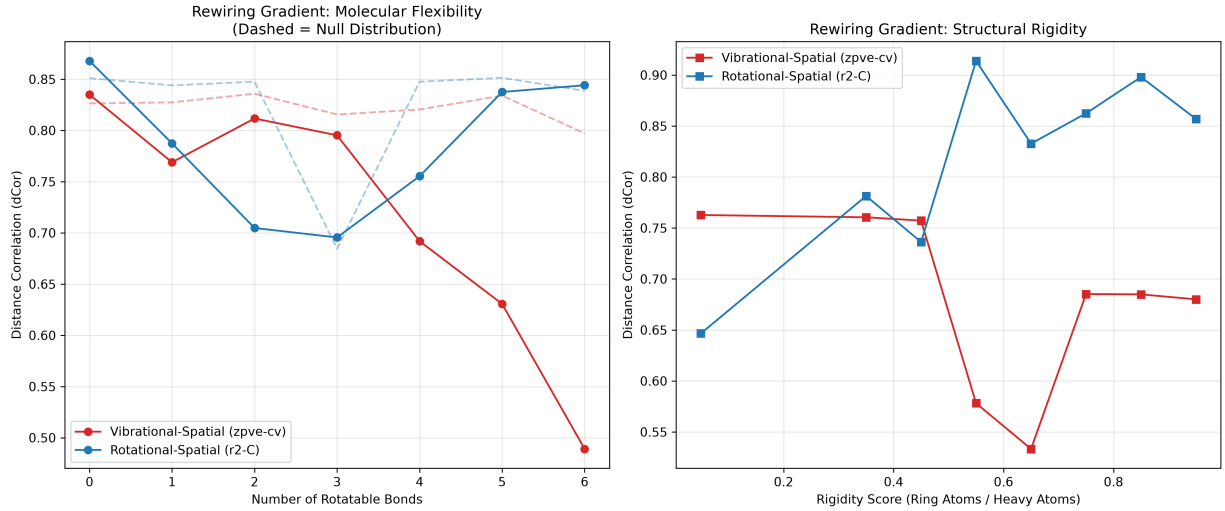

Figure 1: The figure displays the continuous gradient of molecular property rewiring across structural transitions, measured by distance correlation ($dCor$). The left panel maps $dCor$ against molecular flexibility (number of rotatable bonds), while the right panel maps it against structural rigidity (ratio of ring atoms to total heavy atoms). Two primary couplings are compared: the vibrational-spatial axis (zero-point vibrational energy and heat capacity residuals, res_{zpve} vs. res_{cv}) and the rotational-spatial axis (res_{r^2} vs. res_C). As demonstrated in the left panel, increasing flexibility induces a highly significant fracture in the vibrational-spatial coupling, which collapses from $dCor = 0.8350$ in the rigid limit (0 rotatable bonds) to $dCor = 0.4891$ in the flexible limit (6 rotatable bonds). Dashed lines represent the null distribution, confirming this monotonic decline is a genuine architectural effect rather than a statistical artifact. Conversely, the rotational-spatial coupling remains remarkably stable across the flexibility gradient, indicating a fundamental geometric identity that resists architectural rewiring.

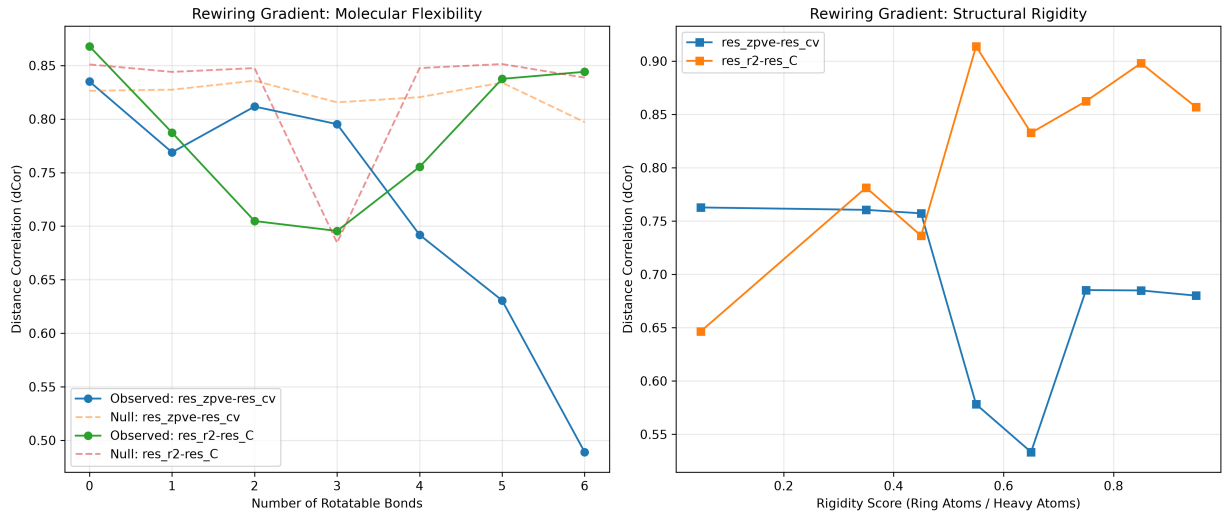

Figure 2: Distance correlation ($dCor$) of property residuals across gradients of molecular flexibility (left) and structural rigidity (right). The left panel illustrates the significant fracture of the vibrational-spatial coupling, showing a monotonic decline in the correlation between zero-point vibrational energy (res_{zpve}) and heat capacity (res_{cv}) as the number of rotatable bonds increases from the rigid to the flexible limit. In contrast, the rotational-spatial coupling between electronic spatial extent (res_{r^2}) and the molecular moment of inertia (res_C) remains stable across the flexibility gradient, reflecting a preserved geometric identity. The right panel maps these dependencies against a rigidity score (ratio of ring atoms to total heavy atoms), confirming that while the res_{r^2} and res_C relationship exhibits only a slight, non-significant trend, the res_{zpve} and res_{cv} coupling is highly sensitive to molecular architecture. These trends demonstrate that molecular property space is a dynamic network, where the vibrational-spatial axis serves as the primary site of physical rewiring as molecules transition from flexible acyclic chains to rigid polycyclic systems.

($LUMO - HOMO = gap$) is perfectly preserved in the residual space ($dCor = 0.999995$), validating the integrity of the linear residual extraction process.

3.5 Refinement of property partitions

The joint structure of the isolated chemical variance challenges traditional, static partitions of molecular properties. As visualized in the heatmap of distance correlation differences (Figure 3), the distinction between "Thermodynamic" (e.g., ZPVE, C_v) and "Geometric" (e.g., r^2) blocks is highly porous. The strongest couplings in the dataset ($dCor > 0.8$) bridge these conventional categories, specifically along the ZPVE-C_v and r^2-C_v axes.

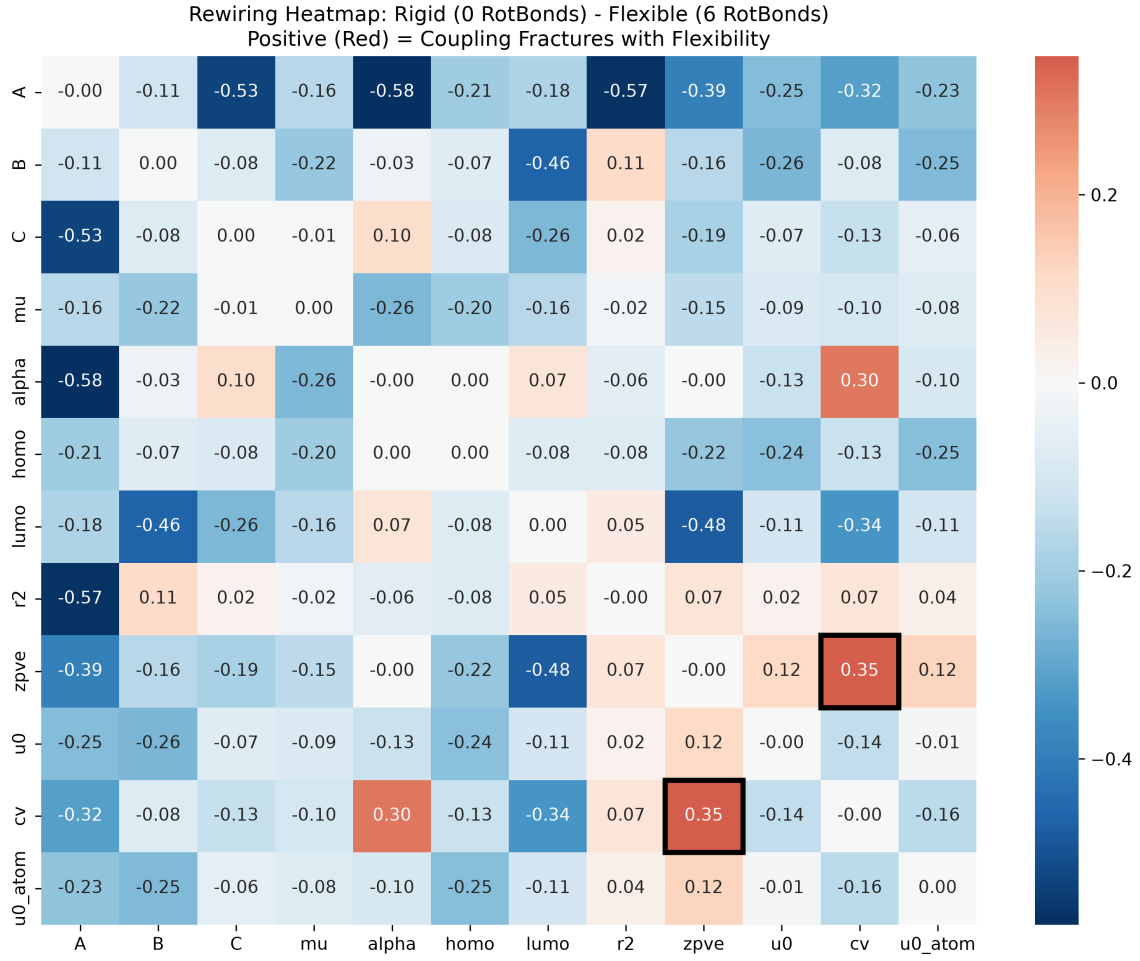

Figure 3: Heatmap of the difference in distance correlation ($\Delta dCor$) between rigid (0 rotatable bonds) and flexible (6 rotatable bonds) molecules for stoichiometric-conditioned property residuals. Positive values (red) indicate couplings that fracture as molecular flexibility increases. As highlighted by the boxed cell, the most significant rewiring occurs in the vibrational-spatial axis, specifically the collapse of the coupling between zero-point vibrational energy ($zpve$) and heat capacity (cv) residuals. This visualizes the central axis of reorganization in the chemical space, demonstrating that traditional property partitions are porous and that the internal physical state space dynamically rewires as a function of molecular architecture.

Conversely, the "Frontier-Electronic" block (HOMO, LUMO) is confirmed to be a distinct and remarkably isolated unit. While it maintains its internal identity, its coupling to the thermodynamic and geometric-response blocks remains consistently weak ($dCor < 0.3$) across all scaffold architectures. Ultimately, the

"Geometric-Response" block proves to be the most sensitive to molecular architecture. While r^2 and rotational constants are coupled by geometric necessity, their joint relationship with the thermodynamic state (via C_v) serves as the primary site of rewiring as molecules transition from acyclic chains to constrained polycyclic systems.

4 Conclusions

In this study, we demonstrated that molecular property space is not a static collection of independent variables, but a dynamic network of physical dependencies that systematically rewires in response to topological complexity. By isolating stoichiometric determinism through the extraction of atom-count regression residuals, we successfully unmasked the higher-dimensional organization of chemical space that is driven strictly by molecular architecture rather than compositional scaling.

Through the construction of distance correlation networks and rigorous permutation testing across structural gradients, we identified the vibrational-spatial axis as the primary site of physical reorganization. The strong coupling between zero-point vibrational energy and heat capacity, which is tightly bound in rigid molecules, systematically fractures as molecular flexibility increases. Conversely, the rotational-spatial coupling between electronic spatial extent and the molecular moment of inertia remains remarkably stable across the same structural gradient. This stability indicates that the relationship between spatial extent and rotational inertia acts as a preserved geometric identity that resists architectural rewiring.

Our quantification of macroscopic network deformation further highlights that small, rigid, and aromatic systems drive the most profound global property rewiring. As detailed in Table 1.3, scaffolds such as benzene, pyranone, and pyridone derivatives exhibit the highest divergence from flexible acyclic baselines. This confirms that aromatic and rigid ring constraints fundamentally alter the internal physical state space compared to size-matched open chains, a phenomenon that is independent of statistical variance collapse or geometric noise.

These findings challenge the conventional, static partitioning of molecular properties. We have shown that the boundary between thermodynamic and geometric properties is highly porous and dynamically rewired by molecular topology. Meanwhile, the frontier-electronic block remains a distinct and isolated unit, largely decoupled from the thermodynamic and geometric-response blocks regardless of the underlying scaffold.

Ultimately, this spectral-statistical framework provides a more fluid and accurate representation of molecular property dependencies. Recognizing that physical couplings fracture and reorganize across discrete topological thresholds is essential for developing more robust computational models and deepening our fundamental understanding of how molecular physics adapts to structural complexity.