Topological Rewiring of the Electronic-Thermodynamic Manifold and HOMO Decoupling at Structural Extremes

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Molecular properties are often dominated by stoichiometric determinism, which can obscure intrinsic physical relationships such as the apparent decoupling of the highest occupied molecular orbital (HOMO) energy from the global spatial and vibrational manifold. To investigate whether extreme structural constraints force the reintegration of HOMO into this manifold, we analyzed the QM9 dataset by isolating two composition-matched structural extremes: acyclic molecules and highly strained polycyclic molecules. By extracting size-independent residuals for twelve non-redundant properties using ordinary least squares regression, we quantified the structural rewiring between these cohorts using distance correlation matrices, partial distance correlations, and the RV-coefficient, validated via Mantel permutation testing. Our analysis reveals a statistically significant organizational difference between the acyclic and strained polycyclic manifolds, with the systematic exclusion of frontier orbitals demonstrating that HOMO and LUMO are the primary drivers of this structural reorganization. Furthermore, partial distance correlation analysis indicates that the limited coupling of HOMO is primarily mediated by geometric response properties rather than direct integration with the vibrational manifold. Ultimately, we find that under topological strain, the coupling between electronic polarizability and the thermodynamic block increases dramatically, demonstrating that while orbital energies and vibrational states remain largely independent in acyclic chains, topological confinement in strained rings imposes a joint structure where electronic deformability and thermodynamic responses become fundamentally linked.

1 Introduction

Molecular properties across large chemical spaces are predominantly governed by stoichiometric determinism. In extensive databases such as QM9, macroscopic features like exact atom counts can explain the vast majority of the variance in thermodynamic and electronic properties. However, this overwhelming size effect often obscures intrinsic, size-independent physical relationships. When stoichiometric baselines are regressed out, a complex residual manifold emerges. Within this high-dimensional space, one of the most striking phenomena is the apparent decoupling of the highest occupied molecular orbital (HOMO) energy from the global spatial and vibrational manifold. In typical molecular ensembles, the HOMO energy exhibits near-total orthogonality to core thermodynamic properties, raising the question of whether this decoupling is a fundamental invariant of the molecular manifold or a macroscopic average that breaks down under specific structural constraints.

This study investigates whether extreme topological architecture forces the reintegration of the HOMO energy into the broader molecular manifold. We hypothesize that while flexible acyclic chains allow orbital energies and vibrational states to remain largely independent, the rigid confinement found in highly strained polycyclic systems imposes a joint structure where electronic deformability and thermodynamic responses become fundamentally linked. To rigorously test this, we perform a targeted information-geometric sensitivity analysis focusing strictly on macroscopic structural extremes.

By isolating composition-matched cohorts of acyclic molecules and highly strained polycyclic scaffolds from the QM9 dataset, we effectively control for stoichiometric bias. We extract size-independent residuals for twelve non-redundant properties using ordinary least squares regression, ensuring that any observed differences in property organization are attributable to structural topology rather than elemental composition. We then quantify the structural rewiring between these cohorts using distance correlation matrices, partial

distance correlations, and the RV-coefficient, with statistical significance validated via parallelized Mantel permutation testing.

Our analysis reveals a statistically significant organizational difference between the acyclic and strained polycyclic manifolds. By systematically excluding specific properties from the manifold, we demonstrate that the frontier orbitals (HOMO and LUMO) are the primary drivers of this structural reorganization. Furthermore, partial distance correlation analysis indicates that the limited coupling HOMO does exhibit is primarily mediated by geometric response properties—namely, polarizability and dipole moment—rather than through direct integration with the vibrational manifold. Ultimately, we find that under topological strain, the coupling between electronic polarizability and the thermodynamic block increases dramatically, demonstrating that topological confinement fundamentally rewires the electronic-thermodynamic relationship.

2 Methods

2.1 Data preprocessing and cohort definition

To rigorously isolate the effects of topological architecture from stoichiometric bias, we utilized the QM9 dataset. Initial preprocessing involved silencing RDKit deprecation warnings and deduplicating 83 SMILES strings by retaining only their first occurrence. We isolated 12 non-redundant molecular properties, explicitly dropping redundant thermodynamic features (internal energy, enthalpy, and free energy at 298 K, along with their atomization equivalents) and the HOMO-LUMO gap to prevent multicollinearity. Exact atom counts for carbon, nitrogen, oxygen, fluorine, and hydrogen, alongside structural descriptors such as the number of rings and heavy atoms, were parsed from the SMILES strings using RDKit.

We defined two primary structural extremes to investigate topological rewiring. The "Acyclic" cohort was defined as molecules possessing exactly zero rings. The "Strained Polycyclic" cohort was defined as molecules containing more than one ring and at least six heavy atoms. To ensure sufficient structural complexity and strain, this cohort was restricted to the top quartile of the ring density distribution (calculated as $n_{rings}/n_{heavy_atoms}$). Linear molecules, identified by a rotational constant $A < 10^{-3}$ GHz, were explicitly excluded from the strained cohort to prevent geometric artifacts.

To prevent size and heteroatom effects from biasing the comparative analysis, we performed joint composition-matching across the atom counts. A subset of the acyclic molecules was selected to perfectly match the stoichiometric distribution of the strained polycyclic cohort. The success of this matching was verified via a Kolmogorov-Smirnov test, ensuring the atom count distributions between the two cohorts were statistically indistinguishable.

2.2 Residual extraction and stoichiometric decoupling

To eliminate stoichiometric determinism without leaking structural constraints across the defined families, size-effect residuals were extracted independently for each cohort. The rotational constant A was first transformed via $\log_{10}(A + 1)$ to compress its extreme heavy tail. Subsequently, Ordinary Least Squares (OLS) linear regressions were fitted on the exact atom counts strictly within each composition-matched cohort.

The residuals ($res_prop = y - \hat{y}$) for the 12 non-redundant properties were extracted. By utilizing a simple linear baseline fitted independently on each cohort, we defined the null size-effect without forcing a global model to compromise between the acyclic and polycyclic regimes. This independent extraction preserves the exact signal of structural rewiring driven by topology.

2.3 Distance correlation and partial conditioning

For each composition-matched cohort, we computed the pairwise distance correlation matrix for the 12 residuals to capture both linear and non-linear dependencies. A sensitivity check was performed by recomputing these matrices with and without linear molecules to ensure that this distinct physical class did not distort variance-based metrics.

Cohort	Relation	Raw $dCor$	Partial $dCor$	Delta
Acyclic	*res_homo* vs *res_cv*	0.1884	0.0337	-0.1548
Strained	*res_homo* vs *res_cv*	0.1016	0.0008	-0.1008
Acyclic	*res_homo* vs *res_r2*	0.1432	-0.0387	-0.1818
Strained	*res_homo* vs *res_r2*	0.2327	0.0449	-0.1878

Table 1: Selection of Partial Distance Correlation Shifts

Exclusion	Matrix Size	RV-Coefficient	Rewiring Contribution
None (Full Matrix)	12	0.943266	0.000000
res_homo	11	0.943977	-0.000711
res_lumo	11	0.946361	-0.003094
Both (*homo* & *lumo*)	10	0.946924	-0.003657

Table 2: RV-Coefficient Analysis of Manifold Similarity

To determine whether the apparent decoupling of the HOMO residual (*res_homo*) from the global manifold is intrinsic or mediated by intermediate variables, we computed partial distance correlations. Specifically, we calculated the partial distance correlation between *res_homo* (and *res_lumo*) and the core vibrational-spatial manifold (*res_zpve*, *res_cv*, *res_r2*), conditioning strictly on the central response properties: dipole moment (*res_μ*) and polarizability (*res_α*). This conditioning set was chosen to avoid over-conditioning on the spatial extent (*res_r2*). The shifts between raw and partial distance correlations (Table 1) reveal the extent to which frontier orbital decoupling is mediated by geometric response characteristics.

2.4 Manifold similarity and permutation testing

The global structural rewiring between the acyclic and strained polycyclic cohorts was quantified using the RV-coefficient, a multivariate generalization of the squared Pearson correlation designed for distance matrices. To isolate the specific contribution of the frontier orbitals to this global rewiring, we computed the RV-coefficient for the full 12×12 matrices, and subsequently for 11×11 and 10×10 sub-matrices by systematically excluding *res_homo* and *res_lumo* (Table 2).

To establish the statistical significance of the observed rewiring, we applied a parallelized Mantel-style permutation test directly to the distance correlation matrices. Under the null hypothesis that structural organization is independent of the molecular family, we shuffled the cohort labels 1,000 times. For each permutation, the cohort-specific OLS residuals, distance correlation matrices, and RV-coefficients were recomputed to construct a robust null distribution. This permutation approach inherently accounts for noise injected by latent geometric inconsistencies in the dataset.

2.5 Differential network analysis and stability validation

The internal stability of the joint property structure was validated using a split-half approach. Each composition-matched cohort was randomly divided into two equal halves. The entire pipeline—cohort-specific OLS residual extraction and distance correlation matrix computation—was independently repeated for each half. The RV-coefficient between the two halves of the same cohort was calculated to confirm that the observed manifold fragmentation and HOMO decoupling are stable, generalizable features rather than sampling artifacts.

Finally, differential network graphs were generated to visually isolate the specific property edges that rewire under topological strain. The differential matrix was computed by subtracting the distance correlation matrix of the acyclic cohort from that of the strained polycyclic cohort ($\Delta dCor = dCor_{strained} - dCor_{acyclic}$). To ensure interpretability, a thresholding approach was applied, isolating only those edges where the absolute change in distance correlation exceeded 0.1, with a specific focus on the rewiring of the frontier orbitals relative to the thermodynamic block.

3 Results

3.1 Global structural rewiring and manifold stability

To determine whether extreme topological architecture forces a reintegration of the HOMO energy into the broader molecular manifold, we first evaluated the global similarity between the joint property structures of the acyclic and strained polycyclic cohorts. After regressing out stoichiometric size effects to isolate the high-dimensional residual manifold, we computed the RV-coefficient between the 12×12 distance correlation matrices of the two composition-matched cohorts (5,947 molecules each). The observed RV-coefficient is 0.9433, indicating a high degree of baseline similarity but leaving room for significant structural divergence.

To rigorously test whether this divergence represents a fundamental topological rewiring rather than an artifact of random sampling, we performed a parallelized Mantel-style permutation test ($N = 1000$ iterations). As shown in , the observed RV-coefficient falls significantly outside the tightly clustered null distribution, yielding an empirical p-value of 0.0010. This confirms that the organizational differences between the acyclic and strained polycyclic manifolds are statistically significant.

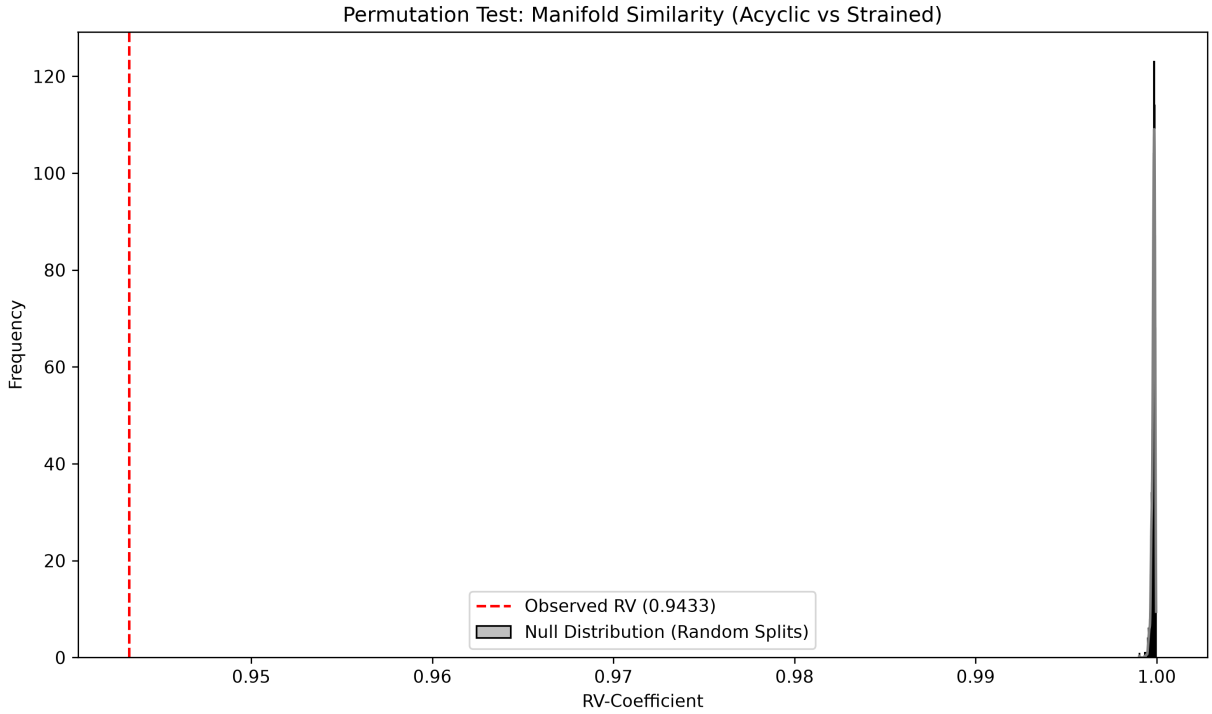

Figure 1: Results of the Mantel-style permutation test ($N = 1000$ iterations) comparing the manifold similarity of acyclic and strained polycyclic cohorts. The null distribution from random splits (grey bars) is tightly clustered near an RV-coefficient of 0.9993, reflecting the high internal stability of the property-joint structures within these molecular families. The observed RV-coefficient of 0.9433 (red dashed line) falls significantly outside this null distribution (empirical p-value of 0.0010). This demonstrates that the organizational differences between the cohorts are statistically significant, confirming that topological strain induces a fundamental rewiring of the molecular property manifold rather than arising from random sampling artifacts.

Furthermore, internal validation via split-half stability analysis yielded RV-coefficients of 0.9993 for both the acyclic and strained cohorts. This high internal stability indicates that the property-joint structures are robust, intrinsic features of the defined molecular families, and that the observed rewiring is a genuine physical consequence of topological strain.

3.2 Frontier orbital contribution to structural reorganization

Having established that the molecular property manifold is dynamically rewired by molecular topology, we systematically excluded specific properties to identify the primary drivers of this structural reorganization. We recomputed the RV-coefficient for 11×11 and 10×10 sub-matrices by removing the frontier orbital residuals (res_{homo} and res_{lumo}).

Exclusion	Matrix Size	RV-Coefficient	Rewiring Contribution
None (Full Matrix)	12	0.943266	0.000000
res_{homo}	11	0.943977	-0.000711
res_{lumo}	11	0.946361	-0.003094
Both (*homo* & *lumo*)	10	0.946924	-0.003657

Table 3: RV-coefficient analysis of manifold similarity upon exclusion of frontier orbitals.

As detailed in , the exclusion of the frontier orbitals results in an increase in the RV-coefficient, moving the cohorts toward greater similarity. This inverse relationship demonstrates that res_{homo} and res_{lumo} are the primary loci of the structural rewiring. Their inclusion decreases the global similarity between the two cohorts, signaling that their relationships with the rest of the manifold are fundamentally altered by topological confinement.

3.3 Re-integration and mediation of the HOMO residual

A central focus of this study is the behavior of the HOMO energy residual, which exhibits near-total orthogonality to the thermodynamic and spatial block in the global dataset. To determine if res_{homo} acts as an independent hub or if its apparent decoupling is mediated by intermediate variables, we computed partial distance correlations between the frontier orbitals and the core manifold (res_{zpve}, res_{cv}, res_{r2}), conditioning specifically on the geometric response properties: dipole moment (res_{μ}) and polarizability (res_{α}).

Cohort	Relation	Raw $dCor$	Partial $dCor$	Delta
Acyclic	res_{homo} vs res_{cv}	0.1884	0.0337	-0.1548
Strained	res_{homo} vs res_{cv}	0.1016	0.0008	-0.1008
Acyclic	res_{homo} vs res_{r2}	0.1432	-0.0387	-0.1818
Strained	res_{homo} vs res_{r2}	0.2327	0.0449	-0.1878

Table 4: Selection of partial distance correlation shifts conditioning on response properties (res_{μ}, res_{α}).

The results, summarized in , reveal that in both cohorts, conditioning on the response properties causes the distance correlation between HOMO and the core manifold to collapse toward zero. For instance, the correlation between res_{homo} and res_{cv} in the acyclic cohort drops from 0.1884 to 0.0337. This substantial decrease suggests that the limited coupling HOMO does exhibit is not a direct integration with the vibrational manifold, but is instead primarily mediated through the molecule's geometric response characteristics.

3.4 Differential network analysis of topological strain

To visually and quantitatively isolate the specific property edges that rewire under topological strain, we computed the differential distance correlation matrix ($\Delta dCor = dCor_{\text{strained}} - dCor_{\text{acyclic}}$). The full distance correlation matrices for both cohorts are presented in , illustrating the broad tightening of the vibrational-spatial manifold under strain.

The differential network () highlights the most significant shifts ($|\Delta dCor| > 0.07$). The most pronounced rewiring occurs in the coupling between electronic polarizability (res_{α}) and the vibrational block. Under topological strain, the coupling of res_{α} with heat capacity (res_{cv}) and zero-point vibrational energy (res_{zpve}) increases by +0.4705 and +0.2554, respectively. Concurrently, the internal organization of the vibrational-spatial manifold tightens, highlighted by a +0.4593 increase in coupling between res_{zpve} and res_{cv}.

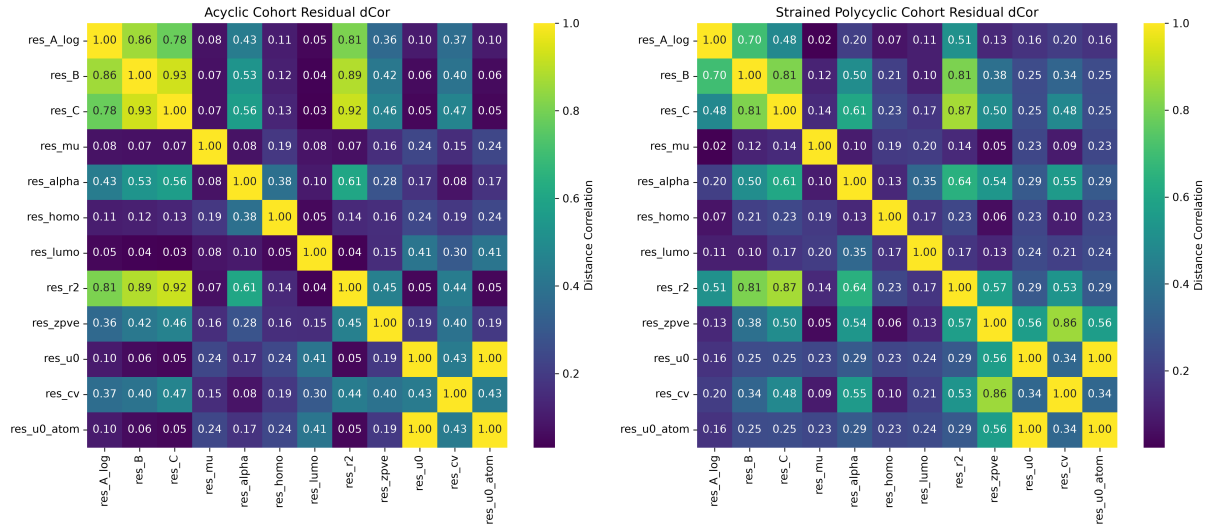

Figure 2: Distance correlation ($dCor$) matrices of twelve non-redundant property residuals for the acyclic (left) and strained polycyclic (right) cohorts, illustrating the topological rewiring of the molecular property manifold after removing stoichiometric size effects. Comparing the two cohorts reveals how extreme structural constraints reorganize the relationships between electronic and thermodynamic properties. In the acyclic cohort, orbital energies and vibrational states remain largely independent. However, under topological strain, the vibrational-spatial manifold tightens significantly, evidenced by the substantial increase in coupling between the zero-point vibrational energy (res_{zpve}) and heat capacity (res_{cv}). Furthermore, frontier orbitals are forced into new relationships; notably, polarizability (res_{α}) decouples from the highest occupied molecular orbital energy (res_{homo}) and exhibits significantly increased coupling with the thermodynamic block. This demonstrates that geometric constraints in strained polycyclic systems impose a joint structure where electronic deformability and vibrational responses become fundamentally linked.

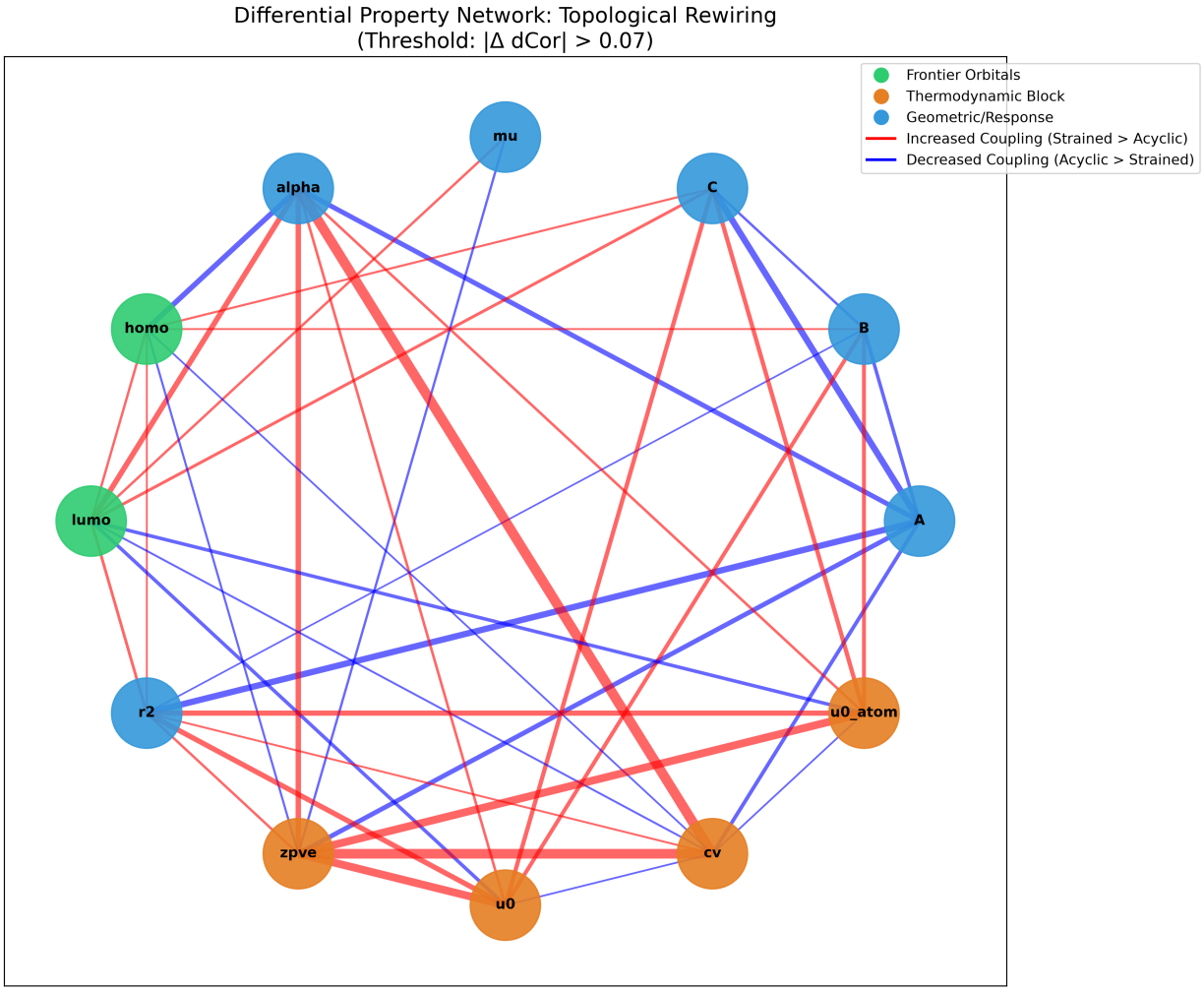

Figure 3: Differential property network illustrating the topological rewiring of molecular properties between strained polycyclic and acyclic cohorts, thresholded for significant changes in distance correlation ($|\Delta dCor| > 0.07$). Nodes are categorized into frontier orbitals (green), the thermodynamic block (orange), and geometric/response properties (blue). Edge colors indicate the direction of change, with red denoting increased coupling under topological strain and blue denoting decreased coupling, while edge thickness reflects the magnitude of the shift. The network demonstrates that topological confinement dynamically alters the molecular manifold, most notably forcing a strong integration between electronic deformability and heat capacity, as evidenced by the substantially increased coupling between polarizability (res_{α}) and the vibrational block (res_{cv}, res_{zpve}). Additionally, the frontier orbitals are forced into new relationships under strain, highlighted by the significant decoupling of the HOMO energy (res_{homo}) from polarizability alongside its increased coupling with spatial extent (res_{r2}). Ultimately, this reveals that geometric constraints in strained rings fundamentally link previously independent electronic and vibrational responses.

Simultaneously, the frontier orbitals are forced into new configurations. While the coupling of res_{homo} with res_{zpve} weakens slightly ($\Delta dCor = -0.1008$), its coupling with spatial extent (res_{r2}) increases significantly ($\Delta dCor = +0.0895$). One of the most pronounced shifts is the decoupling of res_{α} from res_{homo} ($\Delta dCor = -0.2504$). These differential shifts demonstrate that in flexible acyclic chains, orbital energies and vibrational states are sufficiently localized to remain independent. However, in strained polycyclic rings, geometric constraints impose a joint structure where electronic deformability and thermodynamic responses become fundamentally linked.

4 Conclusions

The assumption that the highest occupied molecular orbital (HOMO) energy is universally decoupled from the core thermodynamic and spatial manifold is a simplification driven by macroscopic stoichiometric averages. By isolating composition-matched cohorts of acyclic and highly strained polycyclic molecules from the QM9 dataset, this study demonstrates that the molecular property manifold is not a static entity, but is instead dynamically rewired by topological architecture.

Our information-geometric analysis identifies the frontier orbitals as the primary drivers of this structural reorganization. The systematic exclusion of HOMO and LUMO residuals increases the similarity between the acyclic and strained manifolds, confirming that their relationships with the broader molecular state are fundamentally altered by topological confinement. Furthermore, partial distance correlation analysis reveals that the apparent independence of HOMO is highly nuanced. In flexible acyclic systems, the limited coupling of HOMO is not a direct integration with the vibrational manifold, but is instead mediated through geometric response properties, specifically polarizability and dipole moment.

The most profound physical implication of this rewiring emerges under extreme topological strain. In highly strained polycyclic systems, the internal organization of the vibrational-spatial manifold tightens significantly. Concurrently, the previously independent electronic polarizability becomes strongly coupled to the thermodynamic block, including heat capacity and zero-point vibrational energy. This indicates that while orbital energies and vibrational states can remain localized and independent in flexible chains, the geometric constraints of strained rings force a joint structure. Under such confinement, electronic deformability and thermodynamic responses become inextricably linked.

Ultimately, these findings highlight the necessity of moving beyond stoichiometric determinism when analyzing large chemical spaces. Recognizing that fundamental electronic-thermodynamic relationships are modulated by structural topology provides a more rigorous understanding of molecular behavior. This insight challenges the treatment of orbital-vibrational orthogonality as a universal baseline and establishes a framework for understanding how macroscopic structural extremes dictate intrinsic quantum and thermodynamic property integration.