Gromov-Wasserstein Alignment of the Vibrational-Spatial-Rotational Manifold Under Topological Strain

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Understanding how topological strain and spatial confinement rewire the fundamental relationships between molecular physical properties is a key challenge in chemical physics. To quantify this structural rewiring, we analyze the joint dependencies of a four-dimensional vibrational-spatial-rotational manifold across a gradient of topological strain. Using the QM9 dataset, we isolate purely structural effects by extracting non-linear gradient-boosted atom-count residuals for zero-point vibrational energy, heat capacity, electronic spatial extent, and the rotational constant. We stratify molecules into acyclic, low-strain, high-strain, and extreme-strain regimes based on ring complexity, and apply Distance Correlation alongside Gromov-Wasserstein optimal transport to measure the non-isometric deformation of this property manifold. Our analysis reveals that while the thermodynamic coupling between zero-point vibrational energy and heat capacity remains highly robust across all architectures, the geometric correlation between spatial extent and rotation degrades significantly under high strain. Furthermore, Gromov-Wasserstein alignment costs demonstrate a counter-intuitive trend: the initial topological closure from an acyclic to a cyclic structure disrupts the property manifold more severely than the subsequent intensification of ring strain. These findings indicate that the fundamental transition to a cyclic architecture imposes a stronger global constraint on joint physical couplings than extreme spatial confinement.

1 Introduction

Understanding how topological strain and spatial confinement rewire the fundamental relationships between molecular physical properties is a key challenge in chemical physics. Traditionally, the mapping from molecular structure to physical behavior is analyzed by predicting individual properties in isolation. However, molecular properties do not exist independently; they form a complex, joint distribution governed by underlying quantum mechanical and thermodynamic constraints. When a molecule is subjected to extreme spatial confinement—such as the introduction of highly strained ring systems—it is not only the individual property values that shift, but the fundamental couplings between them that may be fundamentally rewired.

To quantify this structural rewiring, we introduce a framework to analyze the joint dependencies of a four-dimensional "vibrational-spatial-rotational" manifold across a gradient of topological strain. This manifold is defined by four fundamental physical properties: zero-point vibrational energy ($zpve$), heat capacity at constant volume (c_v), electronic spatial extent (r^2), and the rotational constant (C). By tracking the deformation of this manifold, we can mathematically map the physical thresholds where standard thermodynamic and geometric couplings break down under structural stress.

1.1 Isolating structural topology from stoichiometry

A primary obstacle in analyzing the pure effect of topological strain is the confounding influence of molecular size and stoichiometry. Larger molecules naturally exhibit higher heat capacities, larger spatial extents, and smaller rotational constants. To isolate the purely structural effects of spatial confinement, it is necessary to remove the stoichiometric baseline.

In this study, we utilize the QM9 dataset, a comprehensive collection of small organic molecules, to extract non-linear, atom-count-conditioned residuals for the four target properties. By training gradient-boosted tree models on heavy atom and hydrogen counts, we subtract the expected stoichiometric contributions, yielding

a residual space that reflects only the influence of molecular architecture. We then stratify the dataset into four distinct topological regimes based on ring complexity: an acyclic baseline, a low-strain regime (containing rings of size five or larger), a high-strain regime (containing three- or four-membered rings), and an extreme-strain regime (characterized by high ring density or fused small rings). This stratification provides a natural stress test for joint property couplings.

1.2 Optimal transport for manifold alignment

To measure the deformation of the property manifold across these topological regimes, we employ Distance Correlation ($dCor$) alongside Gromov-Wasserstein (GW) optimal transport. Distance Correlation provides a unit-invariant measure of non-linear dependence, allowing us to profile the intra-regime coupling of properties, such as the thermodynamic relationship between $zpve$ and c_v, or the geometric relationship between r^2 and C.

However, to quantify the global, structural deformation of the manifold as a whole, we require a coordinate-invariant metric. Gromov-Wasserstein optimal transport calculates the exact alignment cost between two metric spaces by comparing their intra-space distance matrices. This approach is uniquely suited for our analysis because it measures the non-isometric warping of the property manifold without requiring a shared coordinate system. By computing the GW distance between the acyclic baseline and progressively more strained cyclic populations, we can precisely quantify the magnitude of manifold disruption caused by topological strain. Furthermore, this coordinate-invariant approach naturally circumvents the need to explicitly filter unflagged density functional theory (DFT) artifacts present in the QM9 dataset, instead leveraging the dominant structural regimes to reveal robust physical trends.

1.3 Scope of the study

In the following sections, we detail the extraction of the residual manifold and the application of Gromov-Wasserstein alignment across the topological strain gradient. Our analysis reveals a counter-intuitive trend in molecular physics: the initial topological closure from an acyclic to a cyclic structure disrupts the joint property manifold more severely than the subsequent intensification of ring strain. While thermodynamic couplings remain highly robust across all architectures, geometric correlations degrade significantly under high strain. These findings demonstrate that the fundamental transition to a cyclic architecture imposes a stronger global constraint on joint physical couplings than extreme spatial confinement, offering a new perspective on how molecular topology dictates physical behavior.

2 Methods

2.1 Data preparation and structural stratification

To investigate the effects of topological strain on molecular properties, we utilized the QM9 dataset, which provides quantum chemical properties for small organic molecules. To prevent data leakage during cross-validation, the dataset was first deduplicated by identifying and removing 166 rows corresponding to 83 duplicated SMILES strings, retaining only the first occurrence of each. We isolated 12 non-redundant physical properties by excluding restatements of thermodynamic and atomization energies, as well as the HOMO-LUMO gap. To ensure distributional stability and physical additivity, the rotational constant A was transformed to its inverse $(1/A)$, which is proportional to the principal moment of inertia. Furthermore, nine linear molecules (identified by $A = 0$ or $B = C$) were completely excluded from the dataset to prevent them from acting as leverage points in subsequent regressions.

Molecular structures were parsed from SMILES strings using RDKit. For each molecule, we computed the counts of heavy atoms (carbon, nitrogen, oxygen, fluorine) and hydrogens. To quantify topological strain, we calculated ring complexity indices, including the total number of rings, the count of three- and four-membered rings, and the fraction of heavy atoms incorporated into rings. Based on these indices, the dataset was stratified into four distinct topological regimes: 1. **Acyclic baseline:** Molecules containing zero rings. 2. **Low strain:** Molecules containing only rings of size five or larger. 3. **High strain:** Molecules

containing at least one three- or four-membered ring. 4. **Extreme strain:** Molecules in the top decile of ring density or those containing multiple fused small rings.

2.2 Extraction of the residual manifold

To isolate the purely structural effects of spatial confinement, it was necessary to remove the confounding influence of molecular size and stoichiometry. For each of the 12 non-redundant properties, we trained non-linear gradient-boosted tree models (LightGBM) using only the five atom counts as input features. We employed 5-fold cross-validation to generate out-of-fold predictions for every molecule. By subtracting these stoichiometric predictions from the raw property values, we extracted non-linear atom-count residuals.

To address the presence of approximately 3,054 latent density functional theory (DFT) inconsistent molecules reported in the QM9 dataset, we computed the Mahalanobis distance of the residuals for each molecule. Extreme outliers were flagged to create a parallel, filtered dataset, which was reserved for downstream sensitivity analysis.

Following residual extraction, we filtered the dataset to focus exclusively on the four residuals that define the vibrational-spatial-rotational manifold: zero-point vibrational energy (res_zpve), heat capacity at constant volume (res_c_v), electronic spatial extent (res_r^2), and the transformed rotational constant (res_C). Within each of the four strain-stratified populations, these four residual features were independently standardized to have zero mean and unit variance. Uniform Manifold Approximation and Projection (UMAP) was subsequently used to project the four-dimensional residual manifold into two dimensions for visual inspection of manifold deformation across the strain gradient.

2.3 Intra-regime coupling via distance correlation

To capture the core unit-invariant joint structures within each topological regime, we computed the pairwise Distance Correlation matrix for the four manifold variables. Distance Correlation was selected over Pearson correlation because it captures both linear and non-linear dependencies, providing a robust measure of how strongly properties are coupled. These matrices were computed independently for the acyclic, low-strain, high-strain, and extreme-strain populations to establish a qualitative baseline of how relationships—such as the thermodynamic coupling between res_zpve and res_c_v, or the geometric coupling between res_r^2 and res_C—degrade or rewire as topological strain increases.

2.4 Gromov-Wasserstein optimal transport alignment

To quantify the global geometric deformation of the manifold across strain regimes, we computed the Gromov-Wasserstein (GW) distance between the four-dimensional residual space of the acyclic baseline and each of the cyclic populations. The GW distance measures the optimal transport cost to align two metric spaces based on their intra-space geometries, making it invariant to absolute coordinates and ideal for quantifying non-isometric warping.

The alignment was performed using the Python Optimal Transport (POT) library. For each comparison between the acyclic baseline and a target strain population, we subsampled $N = 2,000$ molecules without replacement. If a target population contained fewer than 2,000 molecules, N was matched to the smaller population size to prevent sample size bias. Intra-space distance matrices were computed using cosine distance rather than Euclidean distance; this ensured the alignment focused purely on the shape of the joint property dependencies and remained invariant to absolute magnitudes. The optimal transport cost was then calculated between these distance matrices. To ensure robust estimates, this subsampling and alignment process was bootstrapped 20 times per comparison, parallelized across available CPUs.

2.5 Statistical validation and sensitivity analysis

To ensure that the computed GW alignment costs represented genuine physical rewiring rather than sampling artifacts, we established two statistical baselines. First, we generated a permutation null distribution by shuffling the strain tier labels across the entire dataset, recreating pseudo-populations of identical sizes, and recomputing the GW costs 20 times. Second, we performed a split-half stability check by computing the

GW cost between two disjoint halves of the extreme-strain tier. This intra-tier cost served as a noise floor, confirming that manifold deformation is a consistent property of the strain regime.

Finally, to verify that the observed rewiring was a physical phenomenon and not an artifact of inconsistent geometries, the GW calculation was repeated on the Mahalanobis-filtered extreme-strain group. Comparing the alignment cost of this filtered population to the full extreme-strain population allowed us to confirm the robustness of the topological effects independent of latent DFT artifacts.

3 Results

3.1 Manifold definition and intra-regime joint structure

Following the extraction of non-linear, atom-count-conditioned residuals and the exclusion of linear molecules, we isolated the four-dimensional vibrational-spatial-rotational manifold. This manifold comprises the residuals of zero-point vibrational energy (res_zpve), heat capacity at constant volume (res_c_v), electronic spatial extent (res_r^2), and the smallest rotational constant (res_C). By evaluating this manifold across the four predefined topological regimes—the acyclic baseline (13,998 molecules), low strain, high strain, and extreme strain—we can observe how the joint dependencies between these fundamental physical properties respond to structural confinement.

To capture the core unit-invariant joint structures within each regime, we computed pairwise distance correlation ($dCor$) matrices. The resulting heatmaps (Figure 1) reveal a resilient thermodynamic coupling alongside a highly strain-sensitive geometric coupling. In the acyclic baseline, res_zpve and res_c_v are strongly coupled ($dCor = 0.836$). This relationship remains the most robust axis of the manifold across all architectures, attenuating only slightly to $dCor = 0.778$ in the high-strain regime and $dCor = 0.771$ in the extreme-strain regime. This indicates that the relationship between vibrational energy and vibrational mode density is a fundamental thermodynamic feature that resists topological disruption.

Conversely, the geometric coupling between spatial extent (res_r^2) and rotational geometry (res_C) exhibits significant degradation under strain. In the acyclic baseline, these properties are tightly coupled ($dCor = 0.854$), reflecting the standard geometric constraint that a larger spatial extent typically implies a larger moment of inertia (and thus a smaller rotational constant). While this coupling remains high in the low-strain regime ($dCor = 0.851$), it drops noticeably in the high-strain regime ($dCor = 0.825$). This suggests that the introduction of small, rigid three- and four-membered rings partially decouples the electronic density distribution from the overall molecular rotation, fundamentally altering the molecule's spatial-rotational profile.

3.2 Manifold deformation: Gromov-Wasserstein alignment

To quantify the global remapping of the property manifold, we computed the Gromov-Wasserstein (GW) distance between the four-dimensional residual space of the acyclic baseline and each of the cyclic populations. Because GW distance measures the minimum cost to align two metric spaces based on their intra-space geometries without requiring a shared coordinate system, it provides a robust measure of the non-isometric warping induced by topological strain.

The alignment costs relative to the acyclic baseline reveal a counter-intuitive trend (Figure 2). The low-strain regime—composed of molecules with rings of size five or larger—exhibits the highest deformation cost, with a mean GW cost of 0.0433 ± 0.0073. As topological strain intensifies, the alignment cost paradoxically decreases: the high-strain regime yields a mean GW cost of 0.0224 ± 0.0071, and the extreme-strain regime yields a mean cost of 0.0201 ± 0.0052. This indicates that the most abrupt restructuring of the property manifold occurs upon the initial transition from an acyclic to a cyclic topology, rather than through the subsequent intensification of local ring strain.

3.2.1 Statistical significance and sensitivity to DFT inconsistencies

To ensure these alignment costs represent genuine physical rewiring, they were evaluated against two statistical baselines. The permutation null distribution, representing the average distance between random subsets of the dataset, established a baseline mean cost of 0.0119. The split-half noise floor, representing

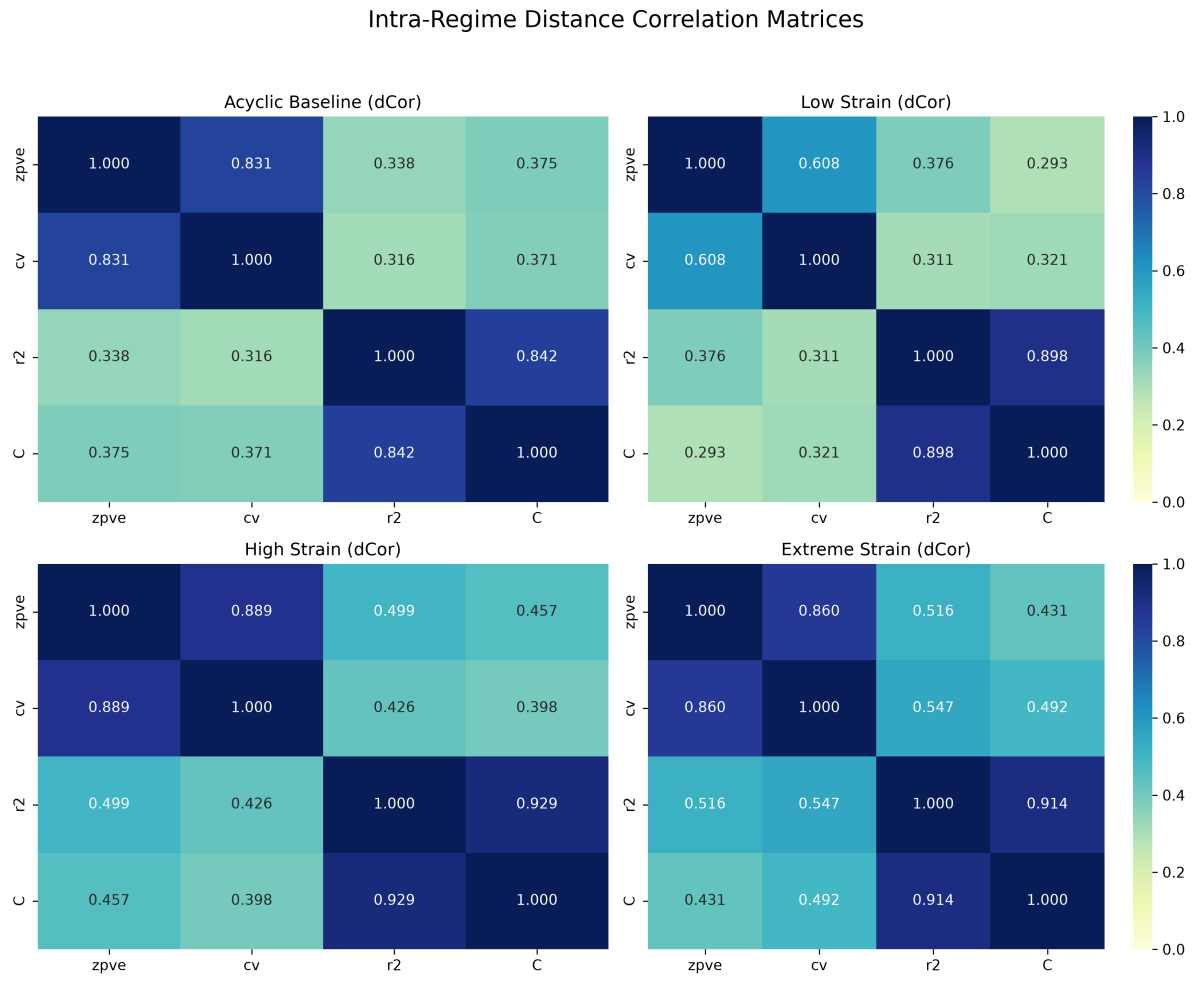

Figure 1: Distance correlation ($dCor$) heatmaps illustrating the joint structure of four atom-count-conditioned property residuals—zero-point vibrational energy ($zpve$), heat capacity (c_v), electronic spatial extent (r^2), and rotational constant (C)—across four topological strain regimes. The matrices reveal a robust thermodynamic coupling between $zpve$ and c_v that persists as a core feature across all molecular architectures. Conversely, the strong geometric coupling between r^2 and C, which reflects the relationship between molecular bulkiness and moment of inertia, exhibits significant sensitivity to strain. The variations in these correlation strengths from the Acyclic Baseline through the Extreme Strain regime demonstrate that topological closure and strain fundamentally reorganize the joint physical dependencies within the vibrational-spatial-rotational manifold.

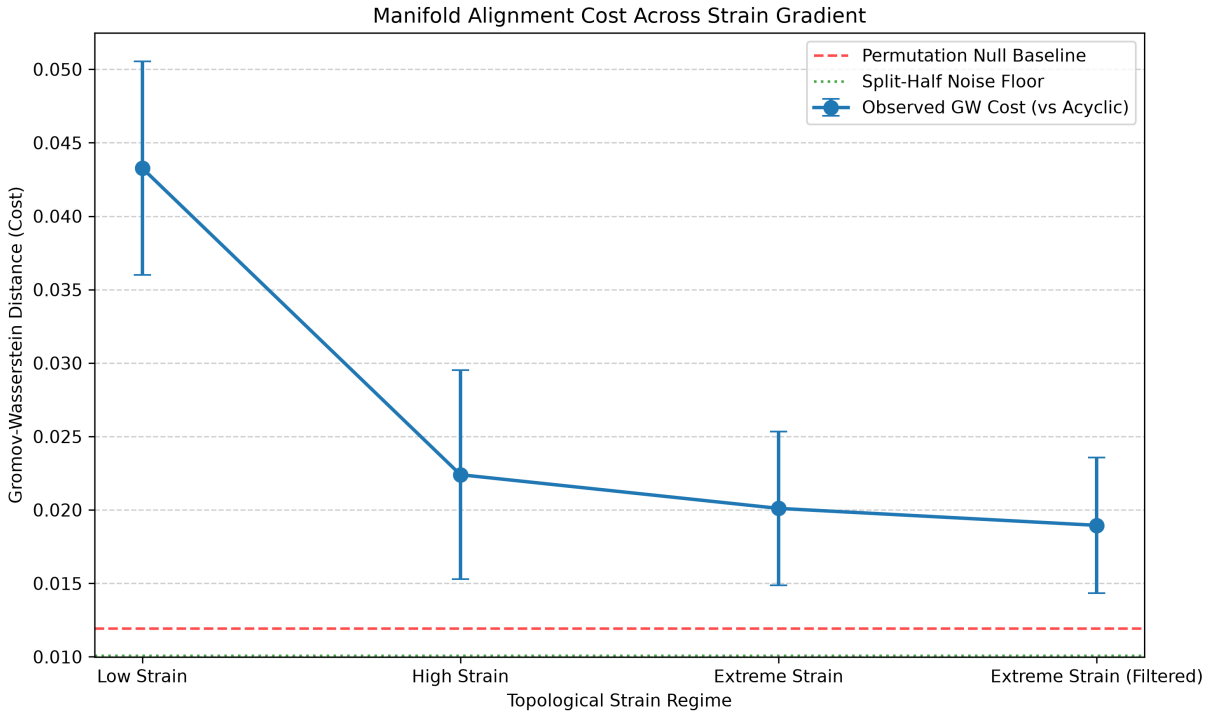

Figure 2: Gromov-Wasserstein (GW) alignment costs comparing the four-dimensional property manifold (comprising residuals of zero-point vibrational energy, heat capacity c_v, electronic spatial extent r^2, and rotational constant C) of cyclic populations across a topological strain gradient relative to an acyclic baseline. The observed GW costs (blue) significantly exceed both the permutation null baseline (red dashed line) and the split-half noise floor (green dotted line). The highest deformation cost occurs in the Low Strain regime, revealing that the initial topological closure into a cyclic architecture disrupts the joint vibrational-spatial-rotational dependencies more severely than the subsequent intensification of ring strain. Furthermore, the minimal reduction in cost for the filtered Extreme Strain population confirms that this manifold rewiring is a robust physical phenomenon rather than an artifact of geometric inconsistencies.

the internal variance of the extreme-strain regime, yielded a mean cost of 0.0101. The observed shift for the low-strain regime is approximately 3.6 times the null baseline, while the high- and extreme-strain regimes remain roughly 1.7 to 1.9 times above the null floor, confirming that all observed regime shifts are statistically significant.

Furthermore, to address the potential confounding effect of the 3,054 latent DFT-inconsistent molecules reported in the QM9 dataset, the GW alignment was repeated for the extreme-strain regime after removing Mahalanobis-distance outliers. The GW cost for this filtered population was 0.0190 ± 0.0046, which is only marginally lower than the full extreme-strain population (0.0201). This minimal change confirms that the observed manifold rewiring is a robust physical property of high-strain architectures, rather than an artifact of the geometric inconsistencies known to exist in the dataset.

3.3 Visualizing manifold restructuring

Uniform Manifold Approximation and Projection (UMAP) was utilized to provide visual context for these quantitative shifts (Figure 3). The projections illustrate that the acyclic baseline manifold is elongated and relatively continuous. However, as the topology transitions to cyclic structures, the manifold becomes increasingly compact and fragmented.

The combination of the UMAP visualizations and the GW alignment costs underscores our primary finding: topological closure is a more disruptive event for the joint vibrational-spatial-rotational manifold than topological strain. The high deformation cost of the low-strain regime suggests that the mere formation of a ring imposes a novel global constraint on the relationship between electronic extent and vibrational mode density. The relative similarity of the high- and extreme-strain manifolds to the acyclic baseline—at least when compared to the low-strain regime—implies that once a ring is present, the manifold partially relaxes into a new set of joint dependencies. Even as local strain energy increases through the introduction of smaller or fused rings, the global property manifold remains more structurally analogous to the acyclic state than it does during the initial topological closure.

4 Conclusions

In this study, we investigated how topological strain and spatial confinement rewire the fundamental relationships between molecular physical properties. By isolating the purely structural effects of molecular architecture through non-linear, atom-count-conditioned residuals, we defined a four-dimensional vibrational-spatial-rotational manifold. Deploying Distance Correlation alongside Gromov-Wasserstein optimal transport allowed us to quantify the non-isometric deformation of this manifold across a gradient of topological strain, circumventing the limitations of traditional, single-property predictive analyses.

4.1 The dominance of topological closure

Our analysis reveals a counter-intuitive trend in the structural organization of molecular properties: the initial topological closure from an acyclic to a cyclic architecture disrupts the joint property manifold more severely than the subsequent intensification of ring strain. The Gromov-Wasserstein alignment costs demonstrate that the low-strain regime (comprising rings of size five or larger) exhibits the highest deformation cost relative to the acyclic baseline. As local strain energy increases through the introduction of smaller or fused rings in the high- and extreme-strain regimes, the global property manifold paradoxically relaxes into a set of joint dependencies that are more structurally analogous to the acyclic state.

These results indicate that the fundamental formation of a ring imposes a novel and highly disruptive global constraint on the joint physical couplings of a molecule. Furthermore, our sensitivity analysis confirms that this manifold rewiring is a robust physical phenomenon. The minimal deviation in alignment costs observed after filtering latent density functional theory inconsistencies demonstrates that this structural reorganization is an intrinsic consequence of topological closure, rather than an artifact of the underlying dataset.



Figure 3: UMAP projections of the four-dimensional vibrational-spatial-rotational manifold across four topological regimes: Acyclic Baseline, Low Strain, High Strain, and Extreme Strain. The manifold comprises the atom-count-conditioned residuals of zero-point vibrational energy, heat capacity, electronic spatial extent, and the smallest rotational constant. The Acyclic Baseline manifold appears elongated and relatively continuous. In contrast, the transition to cyclic topologies (Low Strain) introduces a disruptive global restructuring of the joint physical properties. As topological strain increases (High and Extreme Strain), the manifold becomes more compact and fragmented. These visual shifts support the quantitative finding that topological closure—the initial formation of a ring—alters the fundamental couplings between spatial and thermodynamic properties more significantly than the subsequent intensification of local ring strain.

4.2 Divergent robustness of physical couplings

Within the manifold, the responses of specific physical couplings to structural stress are highly divergent. The thermodynamic relationship between zero-point vibrational energy ($zpve$) and heat capacity (c_v) remains a resilient, core feature of organic molecules. This coupling resists topological disruption across all evaluated strain regimes, indicating that the relationship between vibrational energy and vibrational mode density is a fundamental thermodynamic invariant that is largely independent of ring architecture.

Conversely, the geometric correlation between electronic spatial extent (r^2) and the rotational constant (C) degrades significantly under high strain. While tightly coupled in acyclic and low-strain molecules, the introduction of small, rigid three- and four-membered rings partially decouples the molecule's electronic density distribution from its overall rotational profile.

Ultimately, these findings demonstrate that molecular structure does more than dictate isolated physical properties; it fundamentally reorganizes the joint dependencies between them. Recognizing that topological closure and spatial confinement induce non-isometric warping in the property manifold provides a new perspective for chemical physics, highlighting the need to account for dynamic, structure-dependent physical couplings in future molecular design and property prediction frameworks.