Quantifying Topological Rewiring of the Molecular Property Manifold under Cyclic Confinement via Gromov-Wasserstein Alignment

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

While molecular properties are frequently modeled as additive atomic contributions, this approach overlooks how topological features, such as cyclic confinement, fundamentally reorganize the joint covariance structures of these properties. To strictly quantify this topological rewiring, we track the intrinsic deformation of a 12-dimensional property manifold across discrete structural families using the QM9 dataset. After removing stoichiometric determinism by residualizing properties against atom counts and applying Isolation Forest anomaly filtering, we partition the molecules into cohorts based on ring counts. We then employ entropic Gromov-Wasserstein optimal transport to map the internal distance structures of acyclic molecules onto increasingly complex cyclic manifolds. By evaluating the transport-driven shifts in Fisher Z-transformed distance correlations, we pinpoint how spatial confinement alters strongly coupled physical regimes. Our findings reveal that cyclic confinement systematically decouples the rotational-spatial and polarizability axes; specifically, the strong acyclic dependencies between electronic spatial extent, isotropic polarizability, and the principal rotational constant are severely weakened by the presence of rings. Additionally, we observe a topological divergence in thermodynamic properties: while monocyclic structures decouple vibrational energy from heat capacity, extreme polycyclic confinement in molecules with four or more rings imposes rigid geometric constraints that recouple them. These results demonstrate that molecular topology is not merely a predictor of isolated values, but a fundamental modifier of the intrinsic covariance structure within chemical property spaces.

1 Introduction

The physical state of a molecule—encompassing its geometry, electronic configuration, and thermodynamic stability—is frequently modeled through the lens of additive atomic contributions. For many macroscopic and extensive properties, such as total internal energy or isotropic polarizability, simple stoichiometric determinism explains the vast majority of variance across chemical space. However, this additive paradigm fundamentally overlooks the joint covariance structures of molecular properties. Molecules are not merely collections of atoms; they are complex topological entities. Macroscopic architectural features, particularly cyclic confinement and ring structures, impose rigid geometric constraints that reorganize the global physical states of the molecule.

While the direct impact of ring structures on isolated properties like ring strain or steric hindrance is well documented, the extent to which cyclic confinement reorganizes the *joint* dependencies between diverse physical properties remains poorly quantified. In acyclic molecules, degrees of freedom are relatively unconstrained, leading to highly redundant relationships between spatial, rotational, and electronic properties. As molecules become cyclically confined, these redundancies are expected to collapse or rewire, fundamentally altering the intrinsic covariance structure of the chemical property space. To rigorously understand molecular behavior, we must transition from viewing topology as a simple predictor of isolated values to recognizing it as a fundamental modifier of the property manifold itself.

1.1 Quantifying topological rewiring

To strictly quantify this topological rewiring, we track the intrinsic deformation of a 12-dimensional property manifold across discrete structural families using the QM9 dataset. Because properties such as zero-point vibrational energy and atomization energy are overwhelmingly determined by simple atom counts, any direct analysis of the raw property space primarily measures molecular size. To expose the underlying joint structure, we eliminate stoichiometric determinism by residualizing the properties against atom counts. We further apply a $\log(A+1)$ transformation to the principal rotational constant to mitigate extreme outliers while preserving the continuity of linear molecules, and we employ robust Isolation Forest filtering to prevent latent geometric anomalies from distorting spatial metrics.

The resulting residual point clouds represent the stoichiometry-independent property manifolds of distinct topological cohorts, partitioned by their ring counts (ranging from acyclic structures to complex polycyclic systems with four or more rings). Comparing these manifolds requires a mathematically rigorous framework capable of aligning spaces that do not share a trivial coordinate mapping.

1.2 Manifold alignment via Gromov-Wasserstein transport

To measure the exact geometric cost of mapping the internal distance structures of acyclic molecules onto increasingly complex cyclic manifolds, we deploy entropic Gromov-Wasserstein optimal transport. Unlike standard optimal transport, which requires aligned feature spaces, Gromov-Wasserstein transport aligns metric measure spaces based on their intra-domain pairwise distances. By computing the transport plan between the acyclic baseline and each cyclic cohort, we can project the acyclic distance correlation matrix into the cyclic space.

Evaluating the transport-driven shifts in Fisher Z-transformed distance correlations allows us to pinpoint precisely how spatial confinement alters strongly coupled physical regimes. In this work, we demonstrate that cyclic confinement systematically decouples the rotational-spatial and polarizability axes. Specifically, the strong acyclic dependencies between electronic spatial extent, isotropic polarizability, and the principal rotational constant are severely weakened by the introduction of ring structures. Furthermore, we identify a topological divergence in thermodynamic properties: while monocyclic structures decouple vibrational energy from heat capacity, extreme polycyclic confinement imposes rigid geometric constraints that recouple them. Through this Gromov-Wasserstein alignment framework, we provide a mathematically rigorous demonstration that molecular topology fundamentally deforms the joint thermodynamic and geometric landscape of chemical space.

2 Methods

2.1 Data preparation and stoichiometric residualization

To isolate the topological signal from stoichiometric determinism, we first preprocessed the QM9 dataset by deduplicating the entries and removing redundant thermodynamic columns. The principal rotational constant, A, exhibits extreme outliers (up to 619,868 GHz) and includes linear molecules where $A = 0$. To preserve the continuity of this metric without introducing artifacts, we applied a $\log(A+1)$ transformation. Subsequently, the 12-dimensional property space was residualized against the counts of constituent atoms (C, N, O, F, and H). This step yielded a stoichiometry-independent residual manifold, ensuring that downstream analyses measured intrinsic structural relationships rather than simple molecular size. Finally, to prevent latent geometric anomalies from distorting spatial metrics, we applied an Isolation Forest anomaly detection filter directly to the residual point clouds, establishing a robust baseline manifold for each structural family.

2.2 Cohort partitioning and stratified bootstrapping

The filtered residual space was partitioned into discrete structural cohorts defined by their RDKit-calculated ring counts, specifically grouping molecules into sets containing 0 (acyclic), 1, 2, 3, and 4 or more rings. To ensure the statistical stability of the rewiring signal and to generate confidence intervals, we implemented a stratified bootstrapping framework. For each cohort, we extracted a random sample of exactly $N = 500$

molecules. This sampling process was repeated across at least three independent random seeds. The sampling strategy inherently captured scaffold diversity without enforcing strict whole-scaffold block requirements, thereby maintaining computational stability for downstream tensor operations.

2.3 Entropic Gromov-Wasserstein optimal transport

For each bootstrap iteration, we computed the pairwise Euclidean distance matrix for the 500-point residual cloud of the acyclic baseline and each cyclic cohort. To prevent numerical overflow or underflow during Sinkhorn iterations, these distance matrices were normalized by their respective maximum values. This normalization ensured the distance scales remained consistent with the chosen entropic regularization parameter, $\epsilon = 0.02$. We then computed the optimal transport plan, T, between the acyclic distance matrix and each cyclic distance matrix using entropic Gromov-Wasserstein optimal transport. To maximize computational efficiency, these independent cohort-to-acyclic alignment tasks were parallelized across 14 CPU cores.

2.4 Transport-driven correlation warping and rewiring extraction

To quantify the topological rewiring, we computed the local distance correlation matrices for the 12-dimensional residuals of each cohort. The acyclic distance correlation matrix was projected into the cyclic space using a barycentric mapping derived from the transport plan T. This point re-weighting ensures a mathematically valid warped matrix that represents the acyclic correlation structure as if it were deformed into the cyclic geometry.

We then applied a Fisher Z-transformation to both the warped acyclic and the empirical cyclic distance correlation matrices to yield symmetric, statistically meaningful values. The exact rewiring was extracted by subtracting the empirical cyclic matrix from the warped acyclic matrix, producing a Δ correlation matrix. These Δ matrices were averaged across the bootstrap runs to report a stable mean shift. This allowed us to pinpoint specific matrix elements corresponding to strongly coupled physical regimes, such as the vibrational-spatial and rotational-spatial axes, and rigorously evaluate how cyclic confinement reorganizes the property manifold.

3 Results

3.1 Stoichiometric baseline and manifold deformation

As established in the methodology, raw molecular properties are heavily dominated by simple atom counts. Univariate analyses of properties such as total internal energy (U_0), zero-point vibrational energy (ZPVE), atomization energy ($U_{0,atom}$), and isotropic polarizability (α) primarily measure molecular size, exhibiting near-perfect linear correlations with stoichiometry ($R^2 = 0.99999$, 0.9971, 0.9889, and 0.8591, respectively). By residualizing these properties against atom counts and applying a $\log(A+1)$ transformation to the principal rotational constant A to mitigate the influence of extreme outliers and accommodate linear molecules, we successfully isolated the intrinsic joint structure of the 12-dimensional property manifold.

To quantify how this manifold deforms under cyclic confinement, we computed the entropic Gromov-Wasserstein (GW) optimal transport plans between the acyclic baseline and each cyclic cohort. The resulting GW distances reflect the intrinsic geometric cost of mapping the acyclic property distribution onto cyclic topologies. For a representative bootstrap sample ($N = 500$, regularization $\epsilon = 0.02$), the GW distances were 0.0161 for the 1-ring cohort, 0.0172 for 2-rings, 0.0140 for 3-rings, and 0.0154 for 4+ rings. These distances confirm a substantial deformation of the property space. The corresponding transport plans, visualized in Figure 1, illustrate the complex, non-linear point-to-point correspondences required to align these manifolds, confirming that cyclic confinement induces a global structural reorganization rather than a simple linear shift.

3.2 Decoupling of the rotational-spatial and polarizability axes

The topological "rewiring" of the manifold is defined as the shift in distance correlation (Δ Fisher Z) between property pairs after accounting for the global manifold warp via the GW transport plan. As shown in Figure

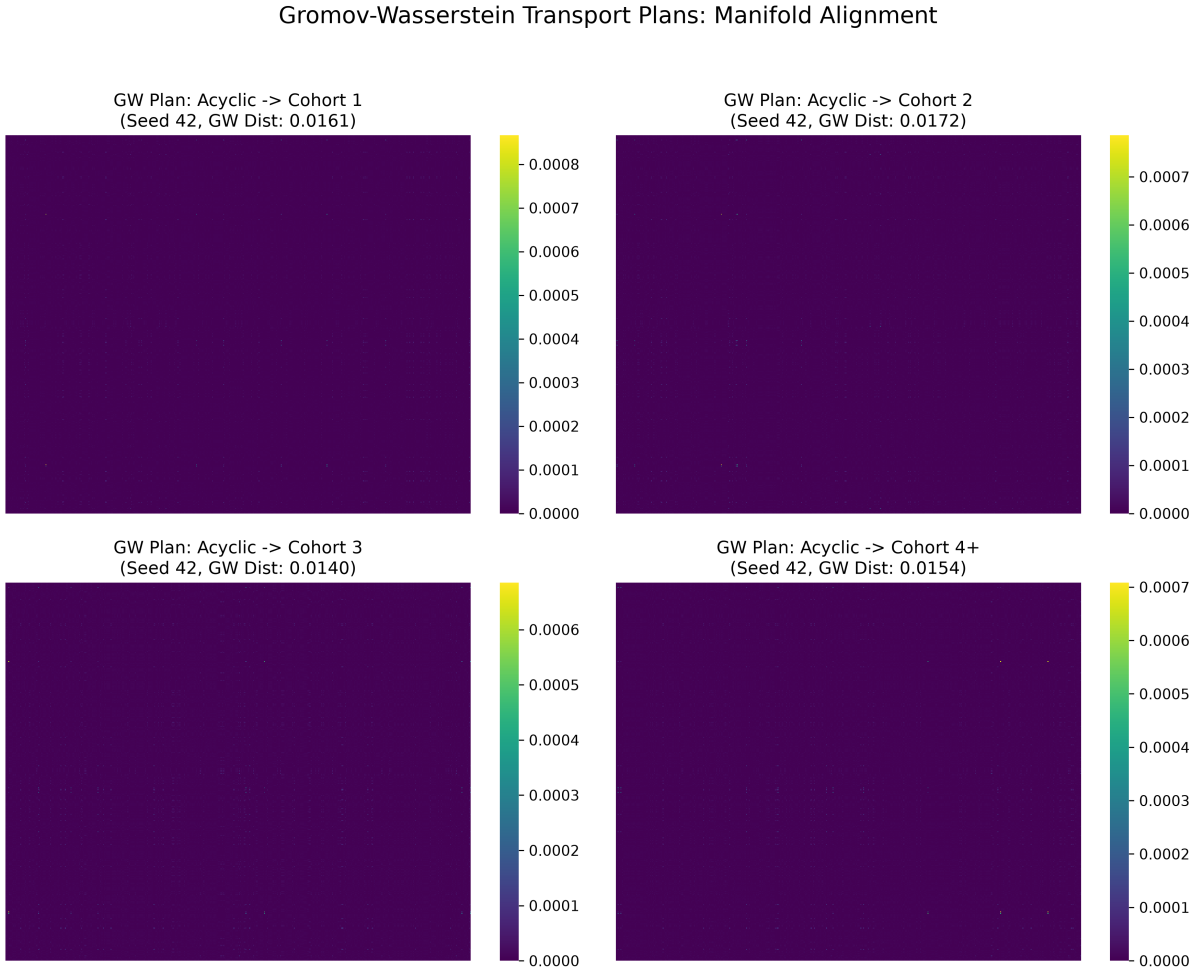

Figure 1: Heatmaps of entropic Gromov-Wasserstein transport plans ($\epsilon = 0.02$) aligning the 12-dimensional property manifold of acyclic molecules to cyclic cohorts (1 to 4+ rings) for a representative bootstrap sample (Seed 42, $N = 500$). The plots reveal the non-linear point-to-point correspondences required to map the internal distance structures of the acyclic baseline onto cyclic states. The associated Gromov-Wasserstein distances (0.0161, 0.0172, 0.0140, and 0.0154) quantify the intrinsic geometric cost and the degree of manifold deformation induced by cyclic confinement, illustrating how molecular topology fundamentally reorganizes joint property structures.

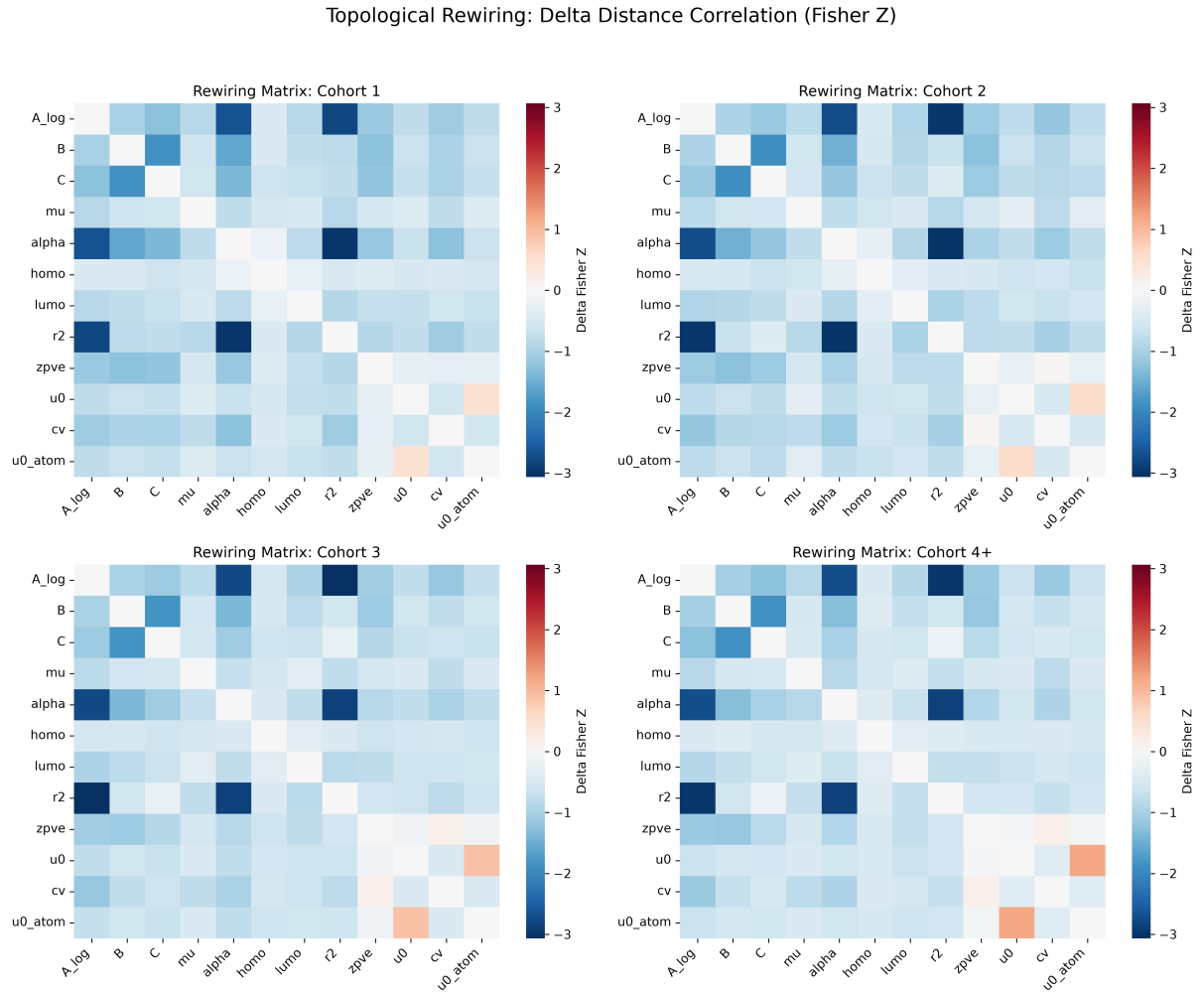

Figure 2: Heatmaps illustrating the topological rewiring of the QM9 property manifold across four cyclic cohorts (1 to 4+ rings). The matrices display the shift in distance correlation (Δ Fisher Z) between residualized molecular properties relative to an acyclic baseline following Gromov-Wasserstein manifold alignment. Negative values (dark blue) indicate a systematic decoupling of properties that are strongly interdependent in acyclic molecules. The most pronounced decoupling occurs among the primary rotational constant A, electronic spatial extent r^2, and isotropic polarizability α, demonstrating that cyclic geometric constraints break the redundancy between electron cloud distribution and principal axes of rotation. Conversely, the emergence of positive shifts in the 4+ ring cohort, notably between zero-point vibrational energy and heat capacity, reveals that extreme polycyclic confinement sufficiently reduces degrees of freedom to establish new, rigid joint thermodynamic structures distinct from those of acyclic chains.

2, the heatmaps of Δ Fisher Z values demonstrate a systematic decoupling of properties that are strongly interdependent in acyclic molecules.

Our analysis reveals that the most profound reorganization occurs among the principal rotational constant A, electronic spatial extent (r^2), and isotropic polarizability (α). Table 1 summarizes the largest shifts across the cyclic cohorts. The consistently large negative ΔZ values indicate that the strong joint dependencies observed in acyclic chains are severely weakened by the introduction of ring structures. In acyclic molecules, spatial extent and polarizability are tightly coupled to the primary rotational constant; once the atom count is fixed, the spatial extent largely dictates the moments of inertia. However, cyclic systems introduce rigid

Table 1: Largest shifts in distance correlation (Δ Fisher Z ± SD) between residualized property pairs across cyclic cohorts relative to the acyclic baseline. The most significant rewiring consistently involves the rotational constant A, electronic spatial extent r^2, and isotropic polarizability α.

Property Pair	Cohort 1 (1-Ring)	Cohort 2 (2-Rings)	Cohort 3 (3-Rings)	Cohort 4+ (4+ Rings)
`res_alpha` ↔ `res_r2`	-2.9910 ± 0.6451	-3.0018 ± 0.6670	-2.8665 ± 0.7161	-2.8541 ± 0.6296
`res_A_log` ↔ `res_r2`	-2.7916 ± 0.4325	-2.9855 ± 0.3065	-3.0668 ± 0.2651	-2.9868 ± 0.4496
`res_A_log` ↔ `res_alpha`	-2.6599 ± 0.3328	-2.7439 ± 0.2079	-2.7590 ± 0.3047	-2.7269 ± 0.2870
`res_B` ↔ `res_C`	-1.8500 ± 1.1383	-1.8853 ± 1.3124	-1.8333 ± 1.2066	-1.8479 ± 1.1908

geometric constraints that break this redundancy, decoupling the spatial distribution of the electron cloud from the principal axes of rotation.

3.3 Topological divergence in targeted coupled regimes

Beyond the primary rotational-spatial decoupling, we tracked specific physically motivated coupled axes to observe how rewiring scales with increasing topological complexity. The vibrational-spatial axis (residual ZPVE versus residual heat capacity, C_v) and the minor rotational-spatial axis (residual r^2 versus residual rotational constant C) exhibited notable divergences depending on the degree of cyclic confinement.

For the vibrational-spatial relationship, the introduction of a single ring slightly decouples the properties ($\Delta Z = -0.2939 \pm 0.5967$). However, in complex polycyclic systems with four or more rings, this shift reverses ($\Delta Z = +0.1575 \pm 0.5790$). This transition from a negative to a positive shift indicates that while simple rings disrupt acyclic redundancies, extreme polycyclic confinement sufficiently reduces the degrees of freedom to establish new, rigid joint thermodynamic structures.

Similarly, the decoupling of electronic spatial extent from the minor rotational constant C is most severe in monocyclic systems ($\Delta Z = -0.7663 \pm 0.7051$) but partially recovers as the system becomes more constrained by multiple rings ($\Delta Z = -0.1865 \pm 0.7005$). These results demonstrate that molecular topology does not merely degrade acyclic property correlations; rather, varying degrees of cyclic confinement uniquely sculpt the intrinsic covariance structure of the chemical space.

4 Conclusions

In this study, we demonstrated that the physical state of a molecule cannot be fully captured by additive atomic contributions alone. While stoichiometric determinism accounts for the vast majority of variance in macroscopic properties, it fundamentally obscures the intrinsic covariance structures that define chemical space. By residualizing the 12-dimensional QM9 property manifold against atom counts and deploying entropic Gromov-Wasserstein optimal transport, we provided a mathematically rigorous quantification of how cyclic confinement rewires these joint property dependencies.

Our manifold-alignment framework revealed that molecular topology acts as a fundamental modifier of the property covariance structure. The most profound topological rewiring occurs within the rotational-spatial and polarizability axes. In acyclic molecules, the electronic spatial extent (r^2), isotropic polarizability (α), and principal rotational constant (A) exhibit highly redundant dependencies. However, the introduction of even a single ring imposes rigid geometric constraints that systematically dismantle these redundancies, decoupling the spatial distribution of the electron cloud from the molecule's principal axes of rotation.

Furthermore, tracking targeted coupled regimes across varying degrees of cyclic complexity uncovered a distinct topological divergence in thermodynamic properties. While monocyclic structures disrupt the acyclic redundancies between zero-point vibrational energy and heat capacity, extreme polycyclic confinement (molecules with four or more rings) sufficiently reduces the available degrees of freedom to recouple these properties. This indicates that highly constrained polycyclic regimes do not merely degrade acyclic relationships, but rather establish new, rigid joint thermodynamic structures that are entirely distinct from those of acyclic chains.

Ultimately, these findings emphasize that the joint thermodynamic and geometric landscape of organic molecules is topologically contingent. Molecular topology is not merely a predictor of isolated property

values, but a global deformer of the intrinsic chemical property manifold. Future predictive models and representations of chemical space must account for these topologically driven shifts in internal correlation structures; otherwise, they will inevitably fail to capture the physical reality of complex cyclic and polycyclic systems.