Electronic-Spatial Rewiring and Topological Decoupling in Small Organic Molecules

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

While molecular properties are fundamentally coupled, the extent to which these relationships reorganize under varying electronic and topological constraints remains poorly understood. To investigate this dynamic behavior, we analyze the joint structure of geometric and electronic properties within the QM9 dataset. We isolate purely electronic effects by extracting non-linear atom-count baseline residuals using quantile regression and controlling for molecular shape via Ray's asymmetry parameter. By partitioning the dataset into HOMO-LUMO gap quantiles, we evaluate the coupling between electronic spatial extent and the third rotational constant using partial correlations. Additionally, we employ partial distance correlations to test the dependence of the highest occupied molecular orbital (HOMO) energy on spatial extent while conditioning on topological descriptors such as aromaticity and saturation. Our analysis reveals a strong global inverse coupling between spatial extent and rotational inertia that catastrophically breaks down in specific HOMO-LUMO gap regimes, indicating a localized decoupling of electron density from nuclear rotational dynamics. Furthermore, we demonstrate that the apparent correlation between HOMO energy and spatial extent is reduced by 81.73% when controlling for molecular topology. These findings establish that the relationship between orbital energy and spatial distribution in small organic molecules is primarily an architectural artifact of the underlying structural skeleton rather than a direct physical coupling, highlighting how electronic environments and topological constraints dynamically reconfigure molecular property relationships.

1 Introduction

Small organic molecules exhibit a complex interplay of geometric, electronic, and thermodynamic properties. In large-scale computational databases such as the QM9 dataset, these molecular properties are frequently treated as independent targets for predictive modeling. However, the physical reality is that these observables are fundamentally coupled. While the baseline relationships between molecular geometry and electronic structure are generally well-established, the extent to which these joint structures reorganize under varying electronic and topological constraints remains poorly understood. Understanding this dynamic behavior is crucial for uncovering the physical rules that govern molecular property spaces.

One fundamental relationship is the coupling between a molecule's electron density distribution, quantified by its electronic spatial extent (r^2), and its moment of inertia, represented by the third rotational constant (C). A baseline inverse relationship is physically expected: as the spatial extent of the electron cloud increases, the moment of inertia increases, leading to a decrease in the rotational constant. Yet, it remains unclear whether this rotational-spatial coupling is a static feature of the molecular architecture or if it reorganizes under varying degrees of electronic "tightness." Specifically, highly confined or conjugated electronic states—characterized by different HOMO-LUMO gap regimes—may force a measurable breakdown and structural rewiring of this spatial-rotational axis.

Furthermore, the relationship between the energy of the highest occupied molecular orbital (HOMO) and the molecule's spatial extent presents another layer of complexity. While orbital energy is often assumed to have direct spatial implications, molecules are heavily constrained by their underlying topological scaffolds, such as ring fusion and saturation levels. It is necessary to determine whether the apparent correlation between HOMO energy and spatial extent is a direct physical coupling or merely an architectural artifact dictated by the underlying molecular skeleton.

To investigate this dynamic behavior, we analyze the joint structure of geometric and electronic properties within the QM9 dataset. To isolate purely electronic and spatial effects from stoichiometric determinism, we extract non-linear atom-count baseline residuals for these properties using heteroscedasticity-aware quantile regression. We then bypass sparse topological scaffolds by using the HOMO-LUMO gap as a continuous conditioning variable, partitioning the dataset into discrete energetic windows. Within these regimes, we evaluate the coupling between the spatial extent and the third rotational constant using partial correlations, explicitly controlling for molecular shape via Ray's asymmetry parameter (κ).

Additionally, we employ partial distance correlations to rigorously test the dependence of HOMO energy on spatial extent while conditioning on topological descriptors, specifically aromaticity and the fraction of sp^3 hybridized carbons. By tracking the trajectory of these couplings across different chemical regimes, this work aims to establish how electronic environments and topological constraints dynamically reconfigure molecular property relationships, revealing localized decoupling phenomena and isolating true physical couplings from structural artifacts.

2 Methods

2.1 Data preparation and topological feature extraction

To investigate the joint structure of molecular properties, we utilized the QM9 dataset, which provides a comprehensive space of small organic molecules. Standard data ingestion and parsing were performed using RDKit, with loggers silenced to ensure clean processing. Following deduplication, the dataset was strictly filtered to isolate non-linear molecules, defined by a non-zero first rotational constant ($A \neq 0$). This filtering step yielded a final working set of 133,789 molecules and was necessary to bypass division-by-zero singularities in subsequent geometric calculations. For each molecule, we extracted the stoichiometric atom counts (C, N, O, F, and H) alongside key topological descriptors, specifically the number of aromatic rings and the fraction of sp^3 hybridized carbons (f_{csp3}), to quantify molecular saturation and ring fusion.

2.2 Baseline removal via quantile regression

Because molecular properties inherently scale with molecular size, it was necessary to isolate purely electronic and spatial effects from stoichiometric determinism. To achieve this, we extracted non-linear atom-count baseline residuals for the properties of interest. We employed GPU-accelerated LightGBM quantile regression, utilizing a 5-fold cross-validation strategy to generate heteroscedasticity-aware residuals. The predictors for this regression were the extracted atom counts, and the targets were the electronic spatial extent (r^2), the third rotational constant (C), the HOMO-LUMO gap, and the HOMO energy. This procedure yielded the size-independent residual variables used for all downstream analyses: res_r^2, res_C, res_gap, and res_homo.

2.3 Geometric controls and gap stratification

To ensure that any observed rewiring in the rotational-spatial axis was not an artifact of macroscopic molecular shape, we computed Ray's asymmetry parameter (κ) for the filtered non-linear subset. The parameter is defined as:

$$\kappa = \frac{2B - A - C}{A - C} \tag{1}$$

where A, B, and C are the rotational constants. This parameter provides a continuous measure of molecular shape, ranging from the prolate limit (-1) to the oblate limit ($+1$).

To bypass sparse topological scaffolds and evaluate the dynamic nature of property couplings, we used the HOMO-LUMO gap residual (res_gap) as a continuous conditioning variable. The dataset was partitioned into 20 discrete res_gap quantiles. This stratification maps the transition from high-strain/low-gap to low-strain/high-gap systems, providing discrete energetic windows within which the structural rewiring could be analyzed.

2.4 Statistical framework for coupling analysis

Within each of the 20 *res_gap* quantiles, we computed the partial correlation between the spatial extent residual (res_r^2) and the third rotational constant residual (res_C), explicitly controlling for Ray's asymmetry parameter (κ). This approach isolates the purely electronic rewiring effect from geometric shape artifacts. To map the physical mechanism underlying any observed breakdown in this coupling, we also analyzed the distribution shifts of κ (mean, median, and variance) across the 20 quantiles and correlated the partial correlation trajectory with the κ variance.

Finally, to rigorously test whether the electronic-spatial coupling exists independently of non-linear topological dependencies, we evaluated the relationship between the HOMO energy residual (*res_homo*) and the spatial extent residual (res_r^2). Replacing standard linear ordinary least squares (OLS) regression, we computed the partial distance correlation between these two variables, conditioning directly on the topological descriptors (number of aromatic rings and f_{csp3}). This non-linear statistical measure allowed us to determine if the apparent spatial dependence of the HOMO energy is a direct physical coupling or merely an architectural artifact of the underlying bonding patterns.

3 Results

3.1 Rewiring of the rotational-spatial axis

We first examine the fundamental coupling between the electronic spatial extent residual (res_r^2) and the third rotational constant residual (res_C). Across the entire dataset, we observe a strong global baseline coupling, characterized by a median partial correlation of approximately -0.8173 when explicitly controlling for shape asymmetry (κ). This inverse relationship aligns with physical expectations: an increase in the spatial extent of the electron cloud corresponds to an increased moment of inertia, thereby decreasing the rotational constant, independent of the overall molecular shape.

However, stratifying the dataset into 20 HOMO-LUMO gap residual (*res_gap*) quantiles reveals that this joint structure is not uniform across all electronic regimes. As illustrated in Figure 1, the coupling experiences an abrupt breakdown in specific energetic windows.

Specifically, in quantile 11 and quantile 19, the partial correlation effectively vanishes, dropping to -0.0562 and -0.0565, respectively. Outside of these localized breakdown points, the coupling remains robust, typically fluctuating between -0.70 and -0.84.

To ensure these breakdowns are not artifacts of macroscopic geometric shifts, we analyzed the trajectory of Ray's asymmetry parameter across the gap spectrum. The mean κ values remain highly stable, ranging only from -0.586 in quantile 1 to -0.654 in quantile 11 (Figure 1, bottom panel). Furthermore, the correlation between the partial correlation strength and the κ variance across the bins is weak (0.3108). This stability confirms that the observed decoupling is an electronic phenomenon—representing localized regimes where the spatial distribution of electron density decouples from the rotational dynamics of the nuclear framework—rather than a consequence of drastic changes in average molecular shape.

3.2 Topological decoupling of electronic and spatial properties

Next, we evaluate the relationship between the HOMO energy residual (*res_homo*) and the spatial extent residual (res_r^2) to determine if orbital energies possess intrinsic spatial dependencies beyond structural constraints. Using distance correlation to capture non-linear dependencies, we find an unconditioned global distance correlation of 0.2181 ± 0.0070 between the two residuals.

To test whether this apparent coupling is a direct physical relationship or an architectural artifact, we computed the partial distance correlation conditioning on key topological descriptors. As shown in Figure 2, when jointly controlling for the number of aromatic rings and the fraction of sp^3 hybridized carbons (f_{csp3}), the correlation collapses to 0.0399 ± 0.0033. This represents an 81.73% reduction in coupling strength.

Redundancy analysis reveals that conditioning on either topological feature individually yields similar reductions: controlling solely for aromaticity reduces the coupling to 0.0457, while controlling solely for saturation (f_{csp3}) reduces it to 0.0415. These results demonstrate that the apparent correlation between a molecule's HOMO energy and its electronic spatial extent is almost entirely a byproduct of its underlying

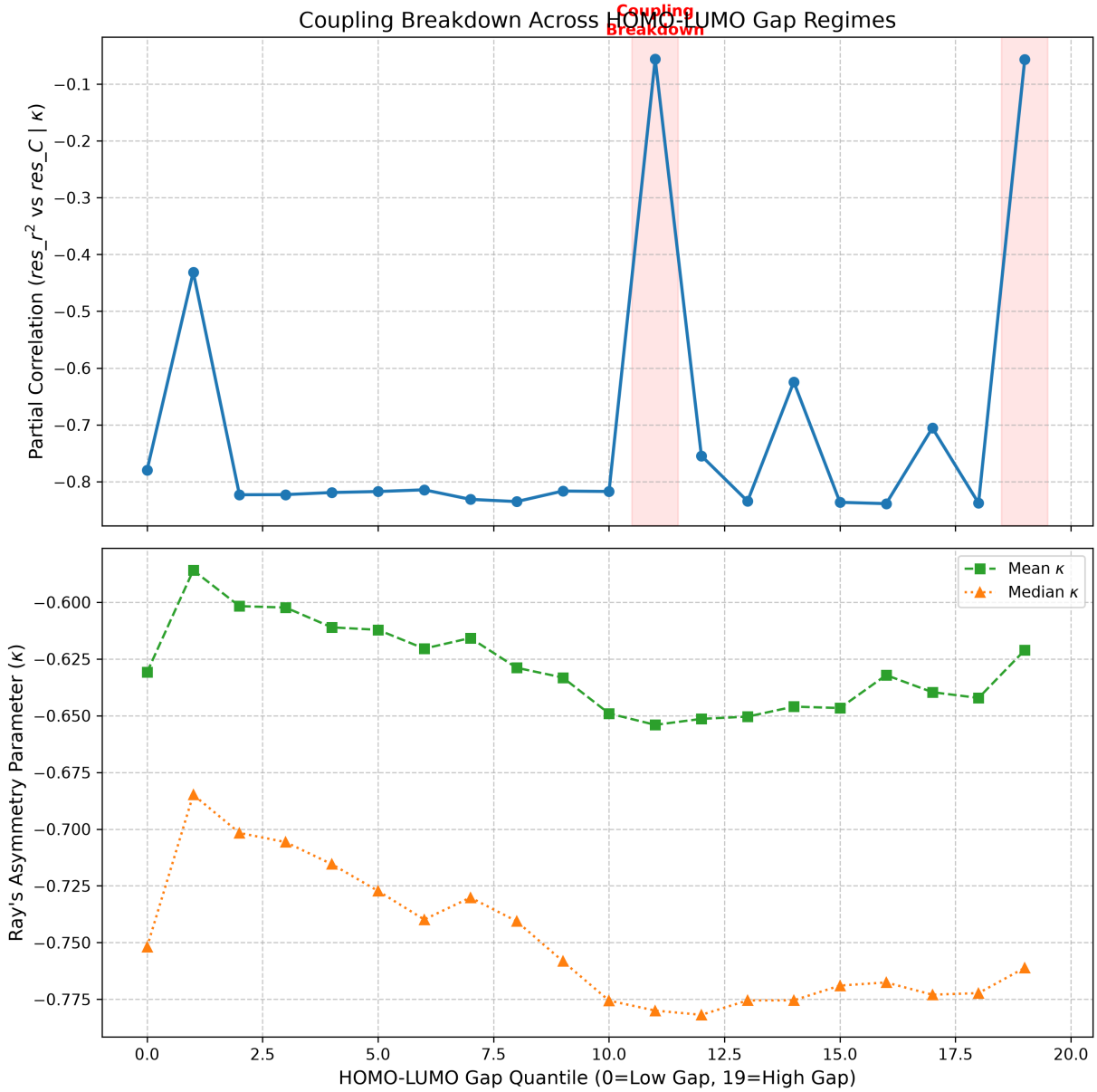

Figure 1: The top panel illustrates the partial correlation between the electronic spatial extent residual (res_r^2) and the third rotational constant residual (res_C), controlling for Ray's asymmetry parameter (κ), across 20 HOMO-LUMO gap quantiles. While a strong inverse coupling is maintained across most regimes, abrupt breakdowns occur at quantiles 11 and 19 (highlighted in red), where the correlation effectively vanishes. The bottom panel shows that the mean and median κ trajectories remain relatively stable across the entire gap spectrum. This stability demonstrates that the observed decoupling is not an artifact of drastic changes in average molecular shape, but rather represents electronic transitions where the spatial distribution of electron density decouples from the rotational dynamics of the nuclear framework.

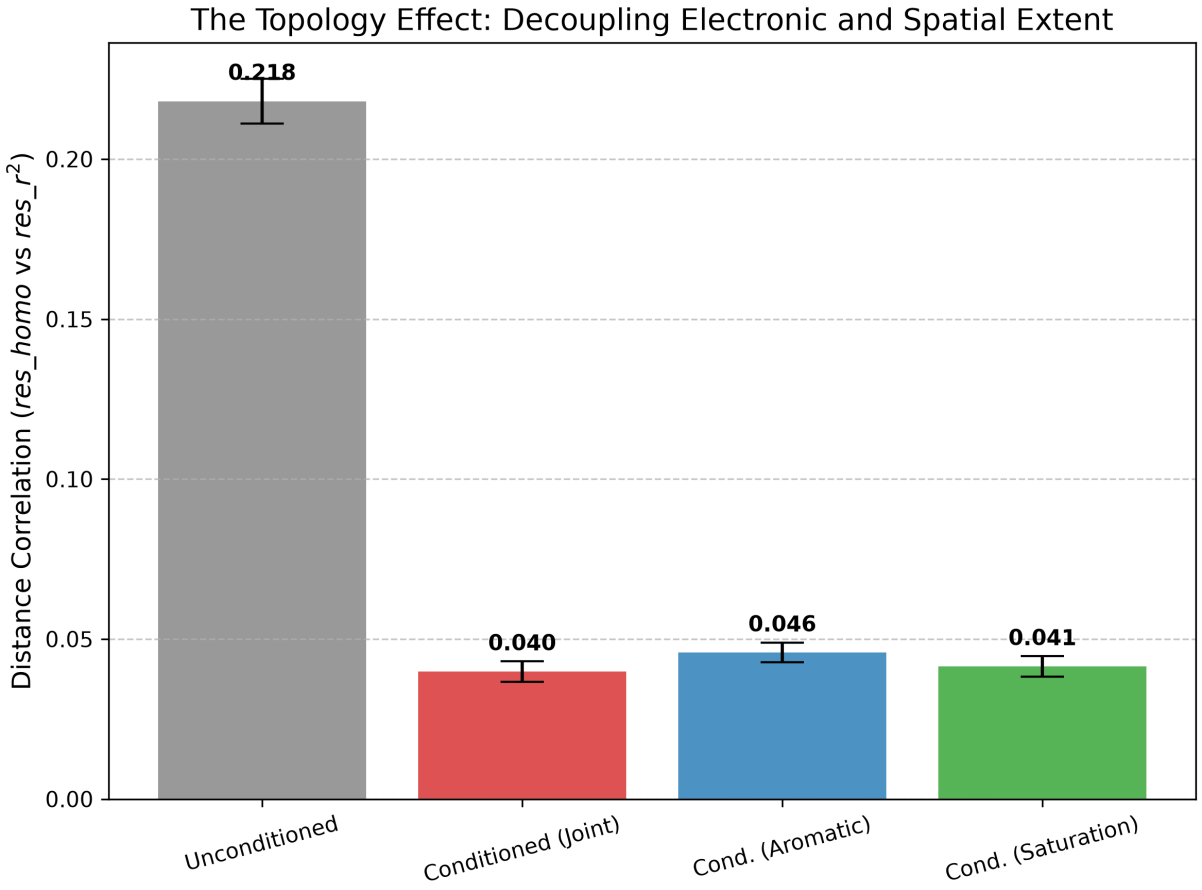

Figure 2: Distance correlation between the HOMO energy residual (*res_homo*) and the electronic spatial extent residual (res_r^2) under various topological conditioning schemes. The unconditioned coupling of 0.218 collapses to 0.040 when jointly controlling for the number of aromatic rings and the fraction of sp^3 carbons (f_{csp3}). Individual conditioning on aromaticity or saturation yields similar reductions to 0.046 and 0.041, respectively. This 81.73% decrease demonstrates that the apparent relationship between a molecule's HOMO energy and its spatial extent is an architectural artifact of its topological skeleton, meaning these properties are intrinsically decoupled once underlying bonding patterns are accounted for.

topological skeleton. Once the bonding patterns and saturation levels are accounted for, the electronic and spatial properties of the HOMO are intrinsically decoupled.

4 Conclusions

This study demonstrates that the joint structure of molecular properties in small organic molecules is not static, but rather highly sensitive to underlying electronic environments and topological constraints. By moving beyond the conventional treatment of QM9 properties as independent regression targets, we isolated the fundamental couplings between geometric, spatial, and electronic features using heteroscedasticity-aware quantile regression and rigorous structural controls.

Our analysis of the rotational-spatial axis confirms a strong global inverse coupling between the electronic spatial extent (res_r^2) and the third rotational constant (res_C). However, stratifying the dataset by HOMO-LUMO gap regimes revealed that this relationship is subject to catastrophic breakdowns. Specifically, in distinct energetic windows (quantiles 11 and 19), the partial correlation effectively vanishes. Because Ray's asymmetry parameter (κ) remains stable across these regimes, these breakdowns cannot be attributed to macroscopic shifts in molecular shape. Instead, they represent localized electronic phase transitions where the spatial distribution of electron density completely decouples from the rotational dynamics of the nuclear framework.

Furthermore, we resolved the nature of the relationship between HOMO energy and electronic spatial extent. While an unconditioned analysis suggests a meaningful correlation, applying partial distance correlations to control for molecular topology—specifically aromaticity and the fraction of sp^3 hybridized carbons (f_{csp3})—resulted in an 81.73% collapse in coupling strength. This reveals that the apparent spatial dependence of HOMO energy is predominantly an architectural artifact. Once the underlying structural skeleton is accounted for, the electronic energy and spatial extent of the highest occupied molecular orbital are intrinsically decoupled.

Ultimately, these findings highlight that molecular structure acts as a dynamic reconfiguring agent. It does not merely predict individual physical constants, but actively enforces or dissolves the joint organization between them. Recognizing these localized decoupling phenomena and separating true physical couplings from topological artifacts is essential for developing a more accurate, physically grounded understanding of molecular property spaces.