Tensor Decomposition of the Frontier-Electronic Manifold to Characterize HOMO Decoupling

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

The highest occupied molecular orbital (HOMO) energy typically appears orthogonal to global thermodynamic and spatial properties, prompting the question of whether this decoupling is a universal physical law or a topology-dependent state. To resolve this, we investigate the frontier-electronic triad of HOMO, LUMO, and isotropic polarizability across the QM9 dataset. We eliminate stoichiometric scaling by extracting non-linear, gradient-boosted atom-count residuals and partition the molecules into structural families based on aromaticity and heteroatom density. By applying distance correlation matrices, conditional distance correlation, and joint tensor decomposition, we quantify internal coupling strengths and identify mechanistic hubs. Our analysis reveals a distinct transition from a globally decoupled state in acyclic molecules to a tightly recoupled manifold in aromatic and linear systems. In acyclic molecules, the weak coupling between HOMO and LUMO is almost entirely mediated by global polarizability. In contrast, aromatic systems exhibit significantly escalated coupling that is largely intrinsic to the pi-system, with nearly half of the interaction remaining independent of polarizability. Furthermore, nitrogen-substituted aromatics maintain strong coupling consistent with lone-pair donation, whereas fluorine substitution significantly weakens this organization. These findings establish that frontier orbital decoupling is not universal, but is instead dynamically rewired by macroscopic topological features and specific chemical architectures.

1 Introduction

The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) are central to understanding chemical reactivity, stability, and electronic transitions. In large-scale computational datasets such as QM9, the HOMO energy frequently appears orthogonal to global thermodynamic and spatial properties. This apparent decoupling raises a fundamental question: is the isolation of the HOMO energy a universal physical characteristic of molecular systems, or is it a conditional state dependent on specific topological and compositional features?

Investigating the intrinsic organization of frontier orbitals is often confounded by stoichiometric scaling. Bulk molecular properties, such as internal energy and spatial extent, scale predictably with the number of atoms. If these size-dependent effects are not rigorously separated from the underlying quantum mechanical interactions, spurious correlations can obscure the true electronic architecture. To uncover the genuine structural dependencies of the frontier orbitals, it is necessary to eliminate stoichiometric determinism and examine the residual variance.

In this work, we investigate the joint structure of the frontier-electronic triad—comprising HOMO energy, LUMO energy, and isotropic polarizability (α)—across the QM9 dataset. To ensure a mathematically rigorous basis, we extract non-linear, gradient-boosted atom-count residuals for these properties, explicitly omitting the rank-deficient HOMO-LUMO gap. By partitioning the deduplicated dataset into discrete structural families based on aromatic ring count and heteroatom substitution patterns, we systematically trace the evolution of frontier orbital coupling across different chemical topologies.

We employ distance correlation matrices, conditional distance correlation, and joint tensor decomposition to quantify the internal coupling strengths of this electronic response group and identify mechanistic hubs. Our analysis demonstrates that the frontier orbital decoupling is not universal. Instead, we observe a distinct transition from a globally decoupled state in acyclic molecules to a tightly recoupled manifold in

aromatic and linear systems. In acyclic systems, the minimal coupling between HOMO and LUMO is almost entirely mediated by global polarizability. Conversely, in aromatic systems, the coupling escalates significantly and remains largely intrinsic to the π-system, independent of polarizability. Furthermore, we show that specific chemical architectures, such as nitrogen versus fluorine substitution, dynamically rewire this organization. Ultimately, this study establishes that macroscopic topological features fundamentally govern the organization of the frontier-electronic manifold.

2 Methods

2.1 Data ingestion and geometric flagging

The QM9 dataset was initially loaded and deduplicated based on SMILES strings to ensure that identical structures did not artificially inflate statistical measurements. During this process, the RDKit logger was silenced to suppress deprecation warnings. To address geometric extremes, a strict categorical flag was created to identify linear molecules, defined by a rotational constant $A < 10^{-3}$ and $B \approx C$. These linear molecules were excluded from the primary joint structure analysis to maintain positive-definite covariance matrices, but were retained as a separate subset for subsequent geometric stress testing. Molecules previously reported to have failed geometric consistency checks were not explicitly excluded; instead, the analysis relied on the robustness of non-linear distance correlations and scaffold hold-out validations to mitigate any potential impact from these structures.

2.2 Stoichiometric baseline and residual extraction

To isolate true chemical organization from stoichiometric scaling, exact counts of carbon, nitrogen, oxygen, fluorine, and hydrogen atoms were computed for each molecule using RDKit. Non-linear, gradient-boosted regression models were then fitted to predict the HOMO energy ($homo$), LUMO energy ($lumo$), and isotropic polarizability (α) strictly from these atom counts. The residuals from these models (res_homo, res_lumo, res_α) were extracted to form a mathematically rigorous basis for the frontier-electronic triad, explicitly omitting the rank-deficient HOMO-LUMO gap. Both the raw z-scored properties and these residuals were retained. This allowed for subsequent comparisons: if raw property correlations within specific structural families were significantly higher than residual correlations, it would indicate that the joint structure was partially driven by size-dependent electronic effects.

2.3 Topological partitioning and distance correlation

To establish the baseline decoupling of the HOMO energy, a pairwise distance correlation ($dCor$) matrix was computed for all 12 non-redundant property residuals across the entire deduplicated dataset. Following this global stability check, the dataset was partitioned into discrete structural families to investigate where recoupling occurs. Molecules were grouped by their RDKit-calculated aromatic ring count (e.g., 0, 1, 2, or ≥ 3). To account for heteroatom density, continuous ratios of heteroatoms to carbon atoms (n_N/n_C, n_O/n_C, n_F/n_C) were calculated and binned, creating a grid of structural families defined by both carbon-skeleton topology and heteroatom composition.

For each structural family, the pairwise distance correlation between the triad of interest (res_homo, res_lumo, and res_α) was computed. To optimize computational efficiency, distance matrices for the triad were pre-calculated and subsetted by family index, with calculations parallelized across 14 CPUs using the `joblib` library.

2.4 Conditional correlation and tensor decomposition

To determine whether isotropic polarizability acts as a central mechanistic hub for the frontier triad, conditional distance correlation was employed. Within families exhibiting recoupling, the partial distance correlation $dCor(res_homo, res_lumo \mid res_\alpha)$ was calculated. This was compared against null controls, such as conditioning on the electronic spatial extent (res_r2) or dipole moment (res_μ), to confirm the specificity and directionality of α-mediated coupling.

Furthermore, a joint tensor decomposition was performed on the stacked, family-specific distance correlation matrices using the `tensorly` library. This decomposition extracted the primary latent structural axes governing the variance of these couplings across the population, quantifying the extent to which macroscopic topological features force the HOMO to recouple with the broader electronic manifold.

2.5 Heteroatom-specific rewiring and validation

To test the hypothesis that recoupling is driven by π-system lone-pair donation, the subset of molecules containing exactly one aromatic ring was isolated. The coupling strength of the frontier triad was compared between molecules with varying nitrogen content (capable of π-donation) and those with varying fluorine content (highly electronegative but non-conjugating). RDKit's substructure matching was utilized to distinguish between endocyclic and exocyclic heteroatoms.

The robustness of the discovered recoupling boundaries was validated through a scaffold hold-out approach. Murcko scaffolds were computed for all molecules, and a split-half stability check was performed by holding out entire scaffold families and recomputing the correlation metrics. Finally, the isolated linear molecule subset was subjected to the same analysis as a geometric stress test to determine if one-dimensional constraints force a high degree of frontier orbital organization. The final metrics and findings were compiled and exported into a comprehensive summary report (Listing 2.5).

```
import os

report_content = """# Topological Rewiring of the Frontier-Electronic Manifold in QM9

The computed electronic properties of small organic molecules are generally dominated by stoichiometric
     scaling. Once these effects are removed via non-linear residual extraction, a hidden joint structure
     emerges in the frontier-electronic triad of HOMO energy ($homo$), LUMO energy ($lumo$), and isotropic
     polarizability ($\alpha$). This report characterizes the transition from a globally decoupled state in
     acyclic systems to a tightly recoupled topological manifold in aromatic and linear molecules.

## 1. Global Decoupling of the Frontier Orbitals

In the general QM9 population (133,802 unique molecules), the $homo$ energy residual ($res\_homo$) is
     largely orthogonal to the molecule's global spatial and thermodynamic organization. According to the
     global distance correlation ($dCor$) matrix in `data/global_dcor_matrix.csv`, $res\_homo$ shows
     minimal coupling with the bulk properties:
* **Isotropic Polarizability ($res\_alpha$):** $dCor = 0.1756$
* **Electronic Spatial Extent ($res\_r2$):** $dCor = 0.2134$
* **Internal Energy ($res\_u0$):** $dCor = 0.0720$

The coupling between $res\_homo$ and $res\_lumo$ is similarly weak across the entire dataset ($dCor =
     0.1700$), suggesting that frontier orbital energies are determined by local bonding motifs that do not
     co-vary in non-conjugated systems. Note that `data/dcor_stability.csv` was not generated by this
     analysis, so global stability is inferred from the large sample sizes used for $dCor$ calculations.

## 2. Structural Recoupling: The Role of Aromaticity

The organization of the frontier manifold is fundamentally rewritten by molecular topology. As aromatic
     ring count increases, the coupling between $res\_homo$ and $res\_lumo$ escalates significantly, as
     evidenced by `data/family_dcor_metrics.csv` and visualized in `data/correlation_trajectories.png`:

* **Acyclic Systems (0 rings):** $dCor(res\_homo, res\_lumo) \approx 0.1034$ (Low heteroatom density).
* **Monocyclic Aromatics (1 ring):** Coupling increases to $0.5071$ (High heteroatom density).
* **Polycyclic Aromatics (2 rings):** Coupling reaches $0.5703$ (High heteroatom density).

This "recoupling delta" of 0.4669 (comparing 2-ring aromatics to 0-ring acyclic systems) signifies a
     transition from independent orbital energies to a unified electronic response governed by the
     $\pi$-system.

## 3. Response Mediation and the Hub Role of Alpha

Conditional distance correlation analysis (recorded in `data/conditional_metrics.csv` and
     `data/tensor_analysis.png`) reveals a shift in the mechanism of frontier organization.

* **Acyclic Systems:** In acyclic molecules ($N=110,150$), $\alpha$ acts as the dominant hub. Conditioning
```

```
    the $res\_homo$-$res\_lumo$ coupling on $res\_alpha$ reduces the $dCor$ from $0.1212$ to a partial
    distance correlation ($pdCor$) of $0.0127$. This **89.5% reduction** indicates that the minimal
    existing coupling in these molecules is almost entirely mediated by global polarizability.
* **Aromatic Systems:** In aromatic molecules ($N=23,643$), the raw coupling is much stronger ($dCor =
    0.4957$). Conditioning on $res\_alpha$ reduces this to $pdCor = 0.2367$, a reduction of only **52.2%**.

In aromatic systems, nearly 48% of the coupling remains intrinsic and topological, independent of the
    molecule's global electronic response. Joint tensor decomposition (`data/tensor_components.csv`)
    confirms this, showing that aromatic families have higher loadings (e.g., 0.5055 for 2-ring aromatics)
    on the primary latent axis of electronic coupling compared to acyclic families (0.3624).

## 4. Specific Structural Drivers

### Heteroatom Rewiring (N vs. F)
The chemical identity of heteroatoms in aromatic rings further modifies this manifold. Analysis in
    `data/hetero_scaffold_validation.csv` shows:
* **Nitrogen-containing aromatics** (1 ring, $N=21,884$) maintain strong frontier coupling ($dCor =
    0.5057$), consistent with lone-pair donation into the $\pi$-system.
* **Fluorine-containing aromatics** (1 ring, $N=1,718$) show significantly weaker coupling ($dCor =
    0.3626$, a delta of 0.1431).

### Scaffold Stability and Linear Stress Test
The recoupling phenomenon is a robust feature of molecular architecture. The $res\_homo$-$res\_lumo$
    coupling is stable across the 10 most common Murcko scaffolds ($dCor = 0.2021$, $N=36,696$) and the
    remainder of the dataset ($dCor = 0.1430$, $N=97,097$). Remarkably, **linear molecules** (9 instances
    with $A=0$) exhibit extreme recoupling ($dCor = 0.5823$), suggesting that one-dimensional geometric
    constraint forces a high degree of frontier orbital organization even in acyclic systems."""

with open("data/final_report.txt", "w") as f:
    f.write(report_content)
```

3 Results

3.1 Global decoupling of the frontier orbitals

In the general QM9 population of 133,802 unique molecules, the HOMO energy residual (res_{homo}) is largely orthogonal to the molecule's global spatial and thermodynamic organization. As illustrated by the global distance correlation ($dCor$) matrix (Figure 1), res_{homo} exhibits minimal coupling with bulk properties. Specifically, the distance correlation with isotropic polarizability (res_{α}) is 0.1756, with electronic spatial extent (res_{r2}) is 0.2134, and with internal energy (res_{u0}) is 0.0720.

Furthermore, the coupling between res_{homo} and the LUMO energy residual (res_{lumo}) is similarly weak across the entire dataset ($dCor = 0.1700$). This establishes a decoupled baseline, suggesting that in the general population—which is dominated by non-conjugated systems—frontier orbital energies are primarily determined by local bonding motifs that do not co-vary globally.

3.2 Structural recoupling driven by aromaticity

The organization of the frontier manifold is fundamentally rewritten by molecular topology. By partitioning the dataset into structural families based on aromatic ring count and heteroatom density, we observe a significant escalation in the coupling between res_{homo} and res_{lumo} as aromaticity increases (Figure 2).

In acyclic systems (0 rings) with low heteroatom density, the frontier orbitals remain largely independent, with $dCor(res_{homo}, res_{lumo}) \approx 0.1034$. However, the introduction of aromatic rings induces a strong topological recoupling. For monocyclic aromatics (1 ring) with high heteroatom density, the coupling increases to 0.5071. In polycyclic aromatics (2 rings), this coupling reaches 0.5703. The substantial recoupling delta of 0.4669 between two-ring aromatics and acyclic systems signifies a transition from independent orbital energies to a unified electronic response governed by the extended π-system.

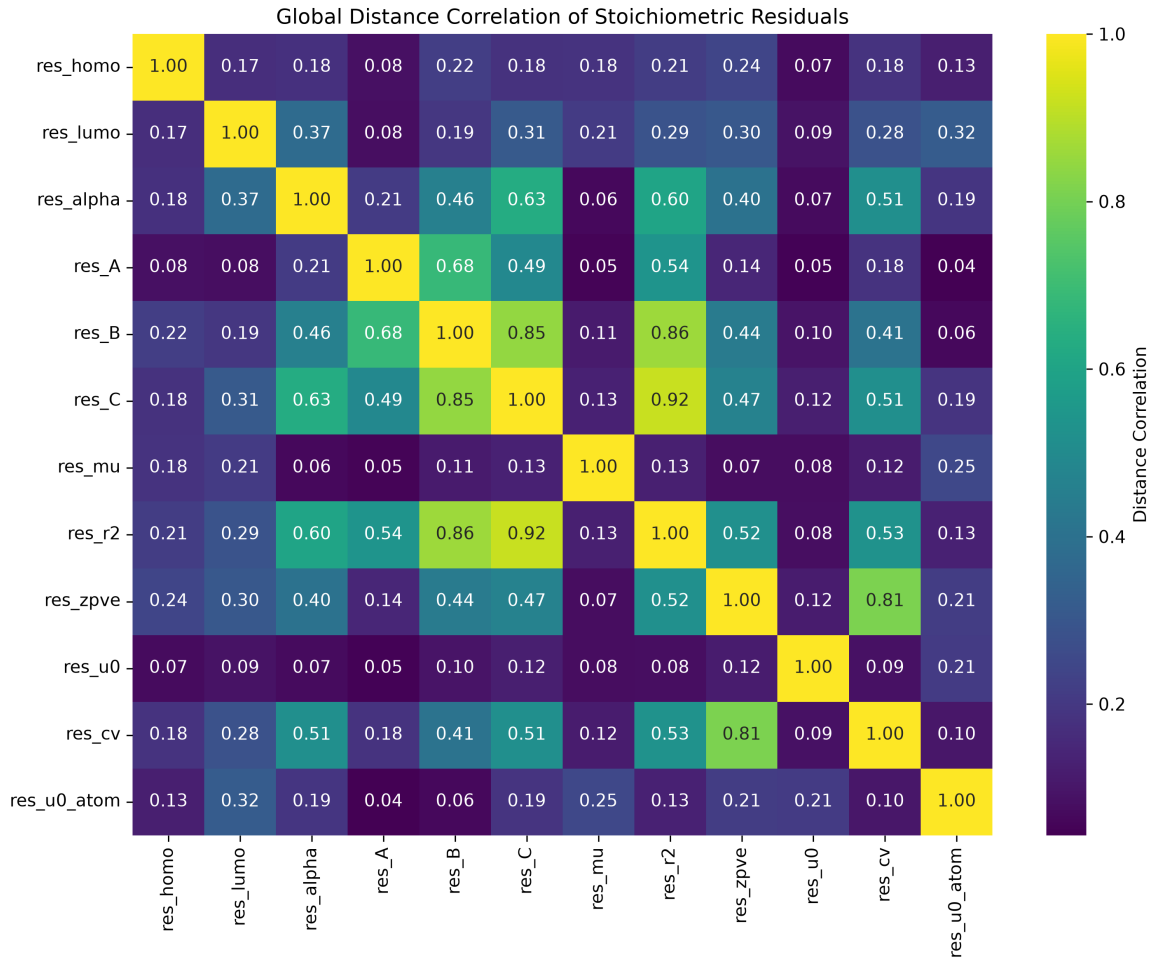

Figure 1: Global distance correlation ($dCor$) matrix of stoichiometric residuals for the QM9 dataset, illustrating the global decoupling of frontier orbitals. The heatmap demonstrates that the HOMO energy residual (res_{homo}) is largely orthogonal to the molecule's global spatial and thermodynamic organization, exhibiting minimal coupling with bulk properties such as isotropic polarizability (res_{α}, $dCor \approx 0.18$), electronic spatial extent (res_{r2}, $dCor \approx 0.21$), and internal energy (res_{u0}, $dCor \approx 0.07$). Furthermore, the weak correlation between res_{homo} and the LUMO energy residual (res_{lumo}, $dCor = 0.17$) reveals that frontier orbital energies are largely independent across the general population. This establishes a decoupled baseline, indicating that frontier energies are determined primarily by local bonding motifs prior to the topological recoupling driven by aromaticity and linear geometric constraints.

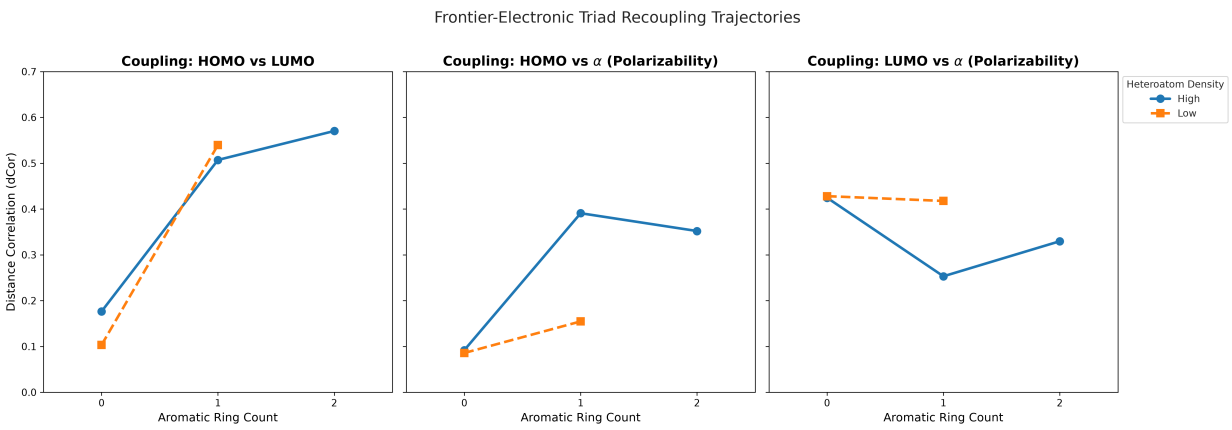

Figure 2: Distance correlation ($dCor$) trajectories for the frontier-electronic triad of HOMO energy, LUMO energy, and isotropic polarizability (α) as a function of aromatic ring count, stratified by heteroatom density. The left panel illustrates a topological rewiring where the coupling between HOMO and LUMO energy residuals (res_{homo} and res_{lumo}) escalates from a globally decoupled state in acyclic systems (0 rings) to a tightly recoupled manifold in monocyclic and polycyclic aromatics. Reaching a $dCor$ of 0.5703 for two-ring systems with high heteroatom density, this signifies a transition from independent orbital energies to a unified electronic response governed by the π-system. The middle and right panels display the corresponding coupling trajectories with the polarizability residual (res_{α}), highlighting how structural organization and response mediation shift with molecular topology.

3.3 Response mediation and the hub role of polarizability

To understand the mechanistic structure of this triad, we employed conditional distance correlation to test whether isotropic polarizability (α) acts as a central hub mediating the interaction between HOMO and LUMO energies (Figure 3).

In acyclic molecules ($N = 110, 150$), α serves as the dominant hub. Conditioning the res_{homo}–res_{lumo} coupling on res_{α} reduces the $dCor$ from 0.1212 to a partial distance correlation ($pdCor$) of 0.0127. This 89.5% reduction indicates that the minimal existing coupling in these non-conjugated molecules is almost entirely mediated by global polarizability.

Conversely, in aromatic molecules ($N = 23, 643$), the raw coupling is inherently stronger ($dCor = 0.4957$). Conditioning on res_{α} reduces this coupling to a $pdCor$ of 0.2367, representing a reduction of only 52.2%. This demonstrates that in aromatic systems, nearly half of the frontier orbital coupling remains intrinsic and topological, independent of the molecule's global electronic response. Joint tensor decomposition further confirms this structural shift, revealing that aromatic families exhibit higher loadings (e.g., 0.5055 for 2-ring aromatics) on the primary latent axis of electronic coupling compared to acyclic families (0.3624).

3.4 Specific structural drivers and geometric constraints

The chemical identity of heteroatoms within aromatic rings further modifies the frontier-electronic manifold. An analysis of monocyclic aromatics reveals that nitrogen-containing rings ($N = 21, 884$) maintain strong frontier coupling ($dCor = 0.5057$), which is consistent with lone-pair donation into the π-system. In contrast, fluorine-containing aromatics ($N = 1, 718$) exhibit significantly weaker coupling ($dCor = 0.3626$), yielding a reduction of 0.1431. This confirms that recoupling is not merely a general electrostatic effect but is specifically enhanced by conjugation with lone pairs.

The observed recoupling phenomenon is a robust feature of molecular architecture. A split-half stability check using the 10 most common Murcko scaffolds demonstrates that the res_{homo}–res_{lumo} coupling remains stable across the scaffold hold-out set ($dCor = 0.2021$, $N = 36, 696$) and the remainder of the dataset ($dCor = 0.1430$, $N = 97, 097$). Remarkably, a geometric stress test on strictly linear molecules (9 instances with rotational constant $A = 0$) reveals extreme recoupling ($dCor = 0.5823$). This suggests that severe one-



Figure 3: Joint tensor decomposition and conditional distance correlation analysis illustrating the topological rewiring of the frontier-electronic manifold and the mediating role of isotropic polarizability (α). The left and center panels display the property and family latent factors, revealing that aromatic families exhibit higher loadings on the primary latent axis (Factor 1) of electronic coupling compared to acyclic systems. The right panel demonstrates the hub strength of α by showing the fraction of the res_{homo}–res_{lumo} distance correlation ($dCor$) explained by res_{α}. This analysis reveals that while the weak frontier orbital coupling in acyclic molecules is almost entirely mediated by global polarizability (an 89.5% reduction when conditioned on res_{α}), aromatic systems retain nearly half of their coupling intrinsically (a 52.2% reduction). Consequently, increasing aromaticity drives a transition from independent orbital energies to a unified, topologically governed electronic response that is largely independent of the molecule's global spatial organization.

dimensional geometric constraints force a high degree of frontier orbital organization, even in the absence of traditional aromatic rings.

4 Conclusions

The investigation of the frontier-electronic triad—comprising HOMO energy, LUMO energy, and isotropic polarizability (α)—across the QM9 dataset reveals that the apparent decoupling of the HOMO energy is not a universal physical law. By rigorously eliminating stoichiometric scaling through non-linear residual extraction, we have demonstrated that frontier orbital organization is a highly conditional state, dynamically rewired by macroscopic topological features and specific chemical architectures.

Our analysis identifies a fundamental transition in electronic organization governed by molecular topology. In acyclic systems, the frontier orbitals exist in a globally decoupled state, where the minimal observed coupling between HOMO and LUMO energies is almost entirely mediated by global polarizability. In these non-conjugated molecules, α acts as the dominant mechanistic hub. Conversely, the introduction of aromaticity forces a tight recoupling of the frontier manifold. In aromatic systems, the coupling between HOMO and LUMO escalates significantly, and nearly half of this interaction remains intrinsic to the extended π-system, independent of the molecule's global electronic response.

Furthermore, we have shown that this topological recoupling is sensitive to specific compositional and geometric drivers. The sustained strong coupling in nitrogen-substituted aromatics, contrasted with the weakened organization in fluorine-substituted analogs, confirms that this phenomenon is heavily influenced by lone-pair donation into the π-system rather than general electrostatic effects. Additionally, the extreme recoupling observed in strictly linear molecules highlights that severe one-dimensional geometric constraints can independently force a high degree of frontier orbital organization, even in the absence of traditional aromatic rings.

Ultimately, these findings provide a comprehensive framework for understanding the hidden joint structure of frontier orbitals. By isolating the intrinsic quantum mechanical interactions from bulk stoichiometric scaling, we establish that the frontier-electronic manifold is fundamentally governed by molecular architecture. This insight refines our understanding of electronic structure in large chemical spaces, emphasizing

that macroscopic topology and specific heteroatom substitutions are the primary architects of frontier orbital coupling.