GINO: GRAPH DINO THROUGH CANONICAL TOKEN SPACES

Anonymous authors
Paper under double-blind review

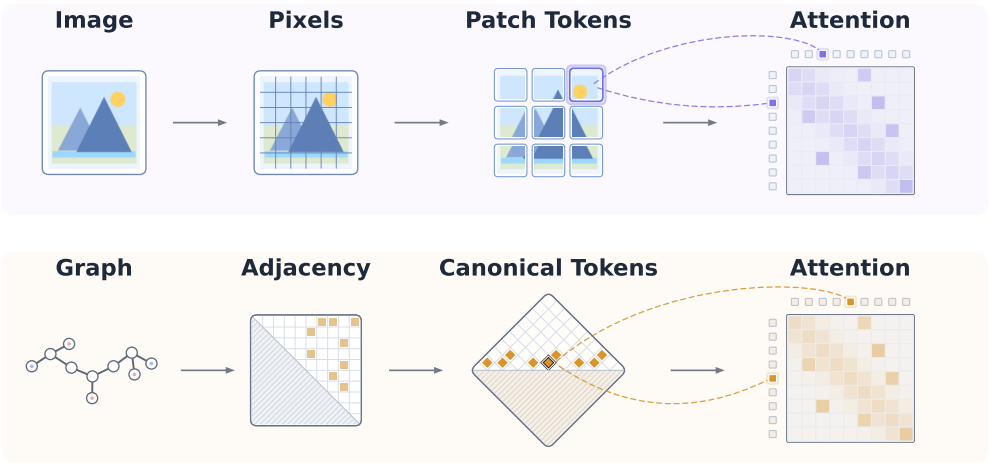

Figure 1: **Mapping arbitrary orderings to canonical tokens.** Vision Transformers (top) operate on patches with fixed spatial coordinates. GINO (bottom) uses spectral canonicalization to assign graph interactions topology-derived pair coordinates, yielding a consistent token space under node relabeling. This enables standard positional mechanisms and DINO/iBOT-style self-supervision to be applied consistently to graphs.

ABSTRACT

Modern self-supervised learning in language and vision leverages tokens with well-defined positional structure. Words follow an ordered sequence, while image patches occupy a spatial grid. Graphs, in contrast, have no intrinsic node order, causing any positional machinery tied to input indices to fail under different node labelings. We introduce GINO, a canonicalization-based framework that equips graph tokens with a consistent spatial layout. GINO extracts a canonical node ordering from the graph's spectral structure and uses it to assign deterministic, two-dimensional coordinates to graph edges. Embedded in this consistent coordinate system, positional encodings and DINO/iBOT-style objectives can operate seamlessly, independent of the initial node labeling. We provide theoretical backing for this framework by guaranteeing the ordering's invariance and bounding its proximity to the exact optimum under a normalized Dirichlet energy. Across molecular benchmarks, self-supervised GINO pretraining improves downstream performance and achieves competitive transfer using only 2D graph structure.

1 INTRODUCTION

Transformers represent inputs as collections of tokens and model their interactions through attention. For many forms of structured data, however, token content alone is not sufficient: word order

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

is essential to language, and spatial arrangement is essential to images. Transformers therefore incorporate positional information alongside token features, allowing attention to distinguish tokens according to where they occur (Vaswani et al., 2017; Shaw et al., 2018). While early approaches primarily encoded individual token positions, subsequent methods increasingly focused on relative position—the displacement between interacting tokens. Rotary position embeddings (RoPE) incorporate this relative information directly into attention by transforming queries and keys according to their positions, such that their interaction depends on the relative offset between them (Su et al., 2024). Two-dimensional RoPE extends the same principle to images, encoding relative offsets between patches on a spatial grid (Heo et al., 2024).

Extending this paradigm to graphs is less direct: unlike words or image patches, graph nodes have no intrinsic ordering, and their input labels can be arbitrarily permuted without changing the graph. This permutation symmetry creates a basic obstacle for positional encoding. If positions are derived directly from the input node labels, relabeling the same graph changes the positions assigned to its nodes and edges, even though the underlying structure is unchanged. Graph neural networks avoid this dependence through permutation-equivariant architectures (Gilmer et al., 2017), while Graph Transformers typically augment such architectures with topology-derived structural or positional information (Ying et al., 2021; Kreuzer et al., 2021; Rampášek et al., 2022; Kim et al., 2022). However, invariant encodings typically suffer from limited expressivity. We therefore consider an alternative approach: we first place the graph in a canonical coordinate system, and only then encode it using its canonical coordinates, as illustrated in Figure 1).

We propose **GINO**, which uses spectral canonicalization to construct a consistent coordinate system for graph tokens (Ma et al., 2024). From the graph topology, GINO builds an affinity matrix, computes its Fiedler eigenvector, and deterministically sorts its entries to assign a canonical index to each vertex. Since this procedure depends only on the graph structure, different node labelings of the same graph produce the same canonical form. GINO then represents each observed edge as a token and assigns it a two-dimensional coordinate derived from the canonical indices of its two endpoints. These coordinates embed the graph tokens in a consistent 2D layout, allowing machinery developed for spatially organized tokens—including 2D RoPE, structured cropping, and DINO/iBOT-style teacher–student objectives—to be applied directly to graphs (Caron et al., 2021; Zhou et al., 2021). Thus, rather than modifying the Transformer architecture or introducing a new self-supervised objective, GINO first resolves the arbitrary node labeling and then applies established Transformer mechanisms in the resulting canonical token space.

We validate the proposed representation by testing its consistency under extensive random node relabelings and by evaluating self-supervised molecular representation learning across eight MoleculeNet classification datasets (Wu et al., 2018). Our results show that canonical token spaces support competitive molecular transfer and that self-supervised pretraining improves downstream performance.

Our main contributions are:

- **Canonical token spaces for graphs.** We construct a canonical coordinate system for graph tokens, allowing standard 2D RoPE and DINO/iBOT-style objectives to be applied consistently across equivalent node labelings.
- **Spectral guarantees for canonical ordering.** We characterize the spectral ordering underlying the canonicalizer, show that sorting the Fiedler coordinate gives its optimal discrete ranking, and bound the resulting normalized Dirichlet energy in terms of the rounding error.
- **Large-scale validation and molecular transfer.** We verify canonicalization robustness under extensive relabeling tests, show that the spectral analysis describes the vast majority of evaluated molecular graphs, and demonstrate competitive molecular transfer using only 2D graph structure.

2 Canonical Token Coordinates

A graph canonicalizer selects a common representation for all vertex labelings of the same graph. Let $B \in \mathbb{R}^{n \times n}$ be the symmetric weighted adjacency matrix of an undirected graph. A canonicalizer

Algorithm 1 Spectral canonicalization (generic case).

Require: Connected weighted adjacency matrix $B \in \mathbb{R}^{n\times n}$, $n \geq 2$, and bandwidth $\sigma > 0$

1. Construct the graph affinity
Assign each edge $\{i, j\}$ the length $\ell_{ij} \leftarrow 1/B_{ij}$
Compute all-pairs shortest-path distances Δ
Set $A_{ij} \leftarrow \exp\big(-d_{ij}^2/(2\sigma^2)\big)$ for $i \neq j$, and $A_{ii} \leftarrow 0$

2. Extract the spectral embedding
$L \leftarrow I - D^{-1/2}AD^{-1/2}$
Compute a unit Fiedler vector u_2 of L
$i^\star \leftarrow \arg\max_i |u_2(i)|$
Orient the coordinate: $\widehat{u}_2 \leftarrow \mathrm{sgn}\big(u_2(i^\star)\big)\, u_2$

3. Obtain the canonical order
Sort $\widehat{u}_2$ increasingly to obtain τ_B and P_B
$C(B) \leftarrow P_B B P_B^\top$

Output: $\tau_B,\ P_B,\ C(B)$

assigns a permutation matrix P_B and returns

$$C(B) = P_B B P_B^\top.$$

The defining requirement is that this procedure should not depend on the ordering of the graph nodes, since this order is arbitrarily. Formally, for every permutation matrix Π, we require that

$$P_{\Pi B \Pi^\top} = P_B \Pi^\top, \text{ and }, C(\Pi B \Pi^\top) = C(B),$$

We construct P_B by sorting a spectral coordinate derived from graph structure. We first describe the construction and give sufficient conditions for its invariance. We then explain its relation to a discrete ordering objective. The production implementation supplements the spectral construction with the refinement and fallback procedures described in Appendix A

2.1 Constructing the canonical order

Our construction follows the spectral ordering approach of CanonNet and related work (Friedmann & Werman, 2025; Zhou et al., 2026), using graph distances to define the affinity and the symmetric normalized Laplacian as the spectral operator. For the main construction, assume that the graph is connected, $n \geq 2$, and $B_{ij} \geq 0$, with $B_{ij} > 0$ precisely on edges and $B_{ii} = 0$.

We assign each edge $\{i, j\}$ the length $\ell_{ij} = 1/B_{ij}$ and let Δ_{ij} be the resulting shortest-path distance. Nonedges are absent from the shortest-path computation; no reciprocal is taken when $B_{ij} = 0$. To obtain a dense affinity, we normalize distances by the weighted diameter and apply a Gaussian kernel:

$$d_{ij} = \frac{\Delta_{ij}}{\max_{k,\ell} \Delta_{k\ell}}, \qquad A_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right) \quad (i \neq j), \qquad A_{ii} = 0.$$

The experiments use $\sigma = 0.35$. Thus, vertices that are closer in the weighted graph receive larger affinity. We then define

$$D = \mathrm{diag}(A\mathbf{1}), \qquad L = I - D^{-1/2}AD^{-1/2}.$$

The matrix L is symmetric positive semidefinite, with a one-dimensional kernel. Write its eigenvalues as $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$, and let u_2 be a unit eigenvector associated with λ_2.

When λ_2 is simple, u_2 is determined up to sign. If it has a unique entry of largest magnitude, we fix the sign by making this entry positive and denote the resulting vector by $\widehat{u}_2$. If its entries are also pairwise distinct, sorting them in increasing order defines a unique rank assignment $\tau_B : V \to \{1, \ldots, n\}$. The corresponding permutation matrix satisfies $(P_B)_{\tau_B(i),i} = 1$, so that $P_B\widehat{u}_2$ is increasing.

Repeated eigenvalues, unresolved signs, and tied coordinates require additional treatment. Appendix A specifies the production sign rules, structural refinement, exact canonical-labeling fallback, and handling of disconnected graphs. If the bounded fallback search cannot complete, the graph is quarantined. Algorithm 1 describes the construction in exact arithmetic; the numerical tolerances and search budgets used in the implementation are given in the appendix.

2.2 INVARIANCE UNDER VERTEX RELABELING

Relabeling vertices preserves weighted path lengths and therefore transports Δ, A, D, and L by the same permutation. In particular,

$$L(\Pi B \Pi^\top) = \Pi L(B) \Pi^\top.$$

This equivariance, together with an unambiguous orientation and sorting step, yields the following guarantee. The result also applies to other equivariant symmetric positive semidefinite operators, including the combinatorial Laplacian.

Theorem 2.1 (Permutation invariance)**.** *Let $B \mapsto L(B)$ be a permutation-equivariant map into symmetric positive semidefinite matrices. Assume that $L(B)$ has a one-dimensional kernel and a simple second eigenvalue $\lambda_2 > 0$. Suppose that a corresponding eigenvector has a unique entry of largest magnitude and pairwise distinct entries. Orient and sort this vector as in Section 2.1. Then, for every permutation matrix Π,*

$$P_{\Pi B \Pi^\top} = P_B \Pi^\top, \qquad C(\Pi B \Pi^\top) = C(B).$$

Proof. Let $\widetilde{B} = \Pi B \Pi^\top$. Equivariance gives $L(\widetilde{B}) = \Pi L(B) \Pi^\top$, so $\Pi \widehat{u}_2$ is a unit eigenvector of $L(\widetilde{B})$ with eigenvalue λ_2. Its unique largest-magnitude entry is positive. Simplicity of λ_2 therefore implies that the oriented eigenvector for $\widetilde{B}$ is exactly $\Pi \widehat{u}_2$. Since its entries are distinct, its sorting permutation is uniquely determined and satisfies $P_{\widetilde{B}} = P_B \Pi^\top$. Consequently,

$$C(\widetilde{B}) = (P_B \Pi^\top)(\Pi B \Pi^\top)(\Pi P_B^\top) = P_B B P_B^\top = C(B).$$

□

The theorem guarantees a consistent vertex-to-rank assignment under its assumptions. More generally, invariance of the canonical graph form does not require a unique assignment of individual vertices to ranks. If the graph has nontrivial automorphisms, two permutations realizing the same canonical form may differ by an automorphism of B. While canonicalizing B alone yields equivalent structural positions for nodes with identical topological roles, we resolve these automorphisms by deterministically breaking ties using the node attributes. This ensures a strictly unique and invariant assignment of the fully attributed graph. Appendix A.2 details this attribute-based tie-breaking procedure, and Section 4.1 empirically validates its strict robustness.

2.3 FROM SPECTRAL COORDINATES TO VERTEX ORDERINGS

We have show that our canonicalization ensures consistency under relabeling. We now explain why it preserves the affinity structure of the graph. To do so, we introduce an energy for discrete orderings and analyze the ordering obtained by sorting the Fiedler vector. Throughout this subsection, L is the symmetric normalized Laplacian of a connected affinity graph with $n \geq 2$.

An ordering is a bijection $\tau : V \to \{1, \ldots, n\}$, where $\tau(i)$ is the position of vertex i. We represent it by the centered, unit-norm rank vector $r_\tau(i) = \frac{\tau(i) - (n+1)/2}{\sqrt{n(n^2-1)/12}}$, $\quad r_\tau^\top \mathbf{1} = 0, \quad \|r_\tau\|_2 = 1$. Every ordering thus assigns the same equally spaced values to the vertices. We evaluate an ordering by its normalized Dirichlet energy,

$$E(\tau) := r_\tau^\top L r_\tau = \frac{1}{2} \sum_{i,j} A_{ij} \left(\frac{r_\tau(i)}{\sqrt{D_{ii}}} - \frac{r_\tau(j)}{\sqrt{D_{jj}}} \right)^2, \qquad \tau_* \in \operatorname*{arg\,min}_{\tau \in S_n} E(\tau), \tag{1}$$

where S_n denotes the set of vertex orderings. High-affinity pairs incur a large penalty when their degree-normalized rank values are far apart. If the affinity degrees are uniform, this objective is proportional to the affinity-weighted sum of squared rank differences. For nonuniform degrees, it measures smoothness after degree normalization.

Optimizing over all orderings to attain τ_* is a difficult discrete optimization problem. A related continuous problem is obtained from the variational characterization of the Fiedler vector. Define

$u_1 = \frac{D^{1/2}\mathbf{1}}{\sqrt{\mathbf{1}^\top D\mathbf{1}}}$, $\quad Lu_1 = 0$. For a connected affinity graph, u_1 spans the kernel of L, and

$$\lambda_2 = \min_{\substack{\|x\|_2=1 \\ x^\top u_1=0}} x^\top Lx. \tag{2}$$

A minimizer is a unit Fiedler vector. Its orthogonality constraint excludes the zero-energy direction, along which $x_i/\sqrt{D_{ii}}$ is constant. Notice that this constraint differs from the rank constraint $r_\tau^\top \mathbf{1} = 0$ unless the affinity degrees are uniform. Thus, for general degrees, Eq. (2) is not an exact relaxation of Eq. (1).

Let $\widehat{u}_2$ denote the oriented unit Fiedler vector and let τ_{spec} be an ordering obtained by sorting its entries in nondecreasing order. The approximation statements below hold for any fixed choice of unit Fiedler vector, including when eigenvalues or coordinates are tied. Uniqueness and canonicality of that choice require the additional conditions discussed above.

While this sorted assignment strictly minimizes the ℓ_2 distance to $\widehat{u}_2$, this nearest-rank optimality does not imply that τ_{spec} minimizes the discrete energy E. To bound its energy gap, we quantify both its approximation error and the mismatch between the two orthogonality constraints. Let

$$\varepsilon = \|r_{\tau_{\mathrm{spec}}} - \widehat{u}_2\|_2, \qquad e = \frac{\mathbf{1}}{\sqrt{n}}, \qquad \gamma_D = (e^\top u_1)^2 = \frac{\left(\sum_i \sqrt{D_{ii}}\right)^2}{n\sum_i D_{ii}} \in (0, 1]. \tag{3}$$

The quantity γ_D is the squared cosine between the two constraint directions; it equals one exactly when the affinity degrees are uniform.

Theorem 2.2 (Energy gap of the spectral ordering)**.** *The spectral ordering satisfies*

$$E(\tau_{\mathrm{spec}}) - E(\tau_*) \le \lambda_2(1-\gamma_D) + (\lambda_n - \lambda_2)\left(\varepsilon^2 - \frac{\varepsilon^4}{4}\right). \tag{4}$$

Proof. We first lower-bound the energy $E(\tau_*)$. Denote $\eta = \tau_*$, the vectors e and r_η are orthonormal. Bessel's inequality gives

$$(e^\top u_1)^2 + (r_\eta^\top u_1)^2 \le 1, \qquad (r_\eta^\top u_1)^2 \le 1 - \gamma_D.$$

Expanding r_η in an orthonormal eigenbasis of L yields

$$E(\tau_*) = E(\eta) = \sum_{k=2}^{n} \lambda_k (r_\eta^\top u_k)^2 \ge \lambda_2 \sum_{k=2}^{n} (r_\eta^\top u_k)^2 = \lambda_2\big(1 - (r_\eta^\top u_1)^2\big) \ge \gamma_D \lambda_2.$$

Next denote $\tau = \tau_{\mathrm{spec}}$, write $r = r_\tau$, and set $c = r^\top \widehat{u}_2$. Since both vectors have unit norm,

$$c = 1 - \frac{\varepsilon^2}{2}, \qquad 1 - c^2 = \varepsilon^2 - \frac{\varepsilon^4}{4}.$$

Decompose $r = c\widehat{u}_2 + z$ with $z \perp \widehat{u}_2$. Then

$$E(\tau) = \lambda_2 c^2 + z^\top L z \le \lambda_2 c^2 + \|L\|\|z\|^2 = \lambda_2 c^2 + \lambda_n(1-c^2) = \lambda_2 + (\lambda_n - \lambda_2)\left(\varepsilon^2 - \frac{\varepsilon^4}{4}\right). \tag{5}$$

Subtracting the lower bound on $E(\tau_*)$ proves the statement. □

Since $\lambda_n \le 2$ for the symmetric normalized Laplacian, the bound above implies a simpler bound:

$$E(\tau_{\mathrm{spec}}) - E(\tau_*) \le \lambda_2(1-\gamma_D) + (2-\lambda_2)\varepsilon^2.$$

Thus, a small rank-approximation error and nearly aligned constraint directions imply a small additive gap for this energy objective. In Section 4.2 we use this bound to show that in practice, the ordering obtained from our algorithm is close to the actual minimum of the discrete optimum $E(\tau_*)$.

3 METHOD

We introduce **GINO (Graph DINO)**, a Transformer for graphs built around a canonical coordinate system for graph edges (see Figure 3). GINO first computes a canonical graph form—independent of the input labeling—and uses the resulting vertex order to assign canonical indices. Each observed edge is then represented as a token and assigned a two-dimensional coordinate derived from the canonical indices of its endpoints. This canonical pair layout serves two purposes: it establishes the relative geometry for 2D rotary position embeddings (RoPE) and provides a coordinate system for constructing structured crops during self-supervised learning.

GINO can be trained directly for downstream prediction or pretrained with DINO/iBOT-style self-supervision. We next describe the edge tokens and their canonical coordinates, relative positional encoding and canonical cropping, and the self-supervised pretraining objective.

3.1 EDGE TOKENS IN CANONICAL COORDINATES

Let $G = (V, E, X, F)$ be an attributed undirected graph with structural matrix B. To resolve topological symmetries, the canonical vertex permutation P_B deterministically breaks structural ties using the vertex attributes X. This induces a unique canonical index $c_B(v) \in \{0, \dots, n-1\}$ for each vertex v, ensuring the resulting spatial layout is strictly invariant to the initial input labeling. GINO represents each observed edge $e = \{u, v\} \in E$ as a Transformer token, with the initial representation $z_e^{(0)} = \phi_{\text{tok}}\big(\psi_{\text{sym}}(x_u, x_v), f_e\big)$, where ϕ_{tok} is the token embedding map and ψ_{sym} is a symmetric function ($\psi_{\text{sym}}(x_u, x_v) = \psi_{\text{sym}}(x_v, x_u)$) to guarantee endpoint order invariance. In our molecular experiments, x_u, x_v contain atom features and f_e contains bond features; complete feature and embedding definitions are given in Appendix E.

Each edge token is paired with a two-dimensional coordinate derived from its endpoints' canonical indices: $\rho(e) = (\min\{c_B(u), c_B(v)\}, |c_B(u) - c_B(v)|)$. This is an invertible representation of the canonically ordered endpoint pair, using the smaller endpoint index and the nonnegative offset to the larger.

3.2 RELATIVE POSITION AND CANONICAL CROPPING

Each edge receives a fixed coordinate $\rho(e)$ in the canonical pair layout. GINO leverages this shared coordinate system in two ways: RoPE converts canonical coordinates into relative offsets for attention, while canonical cropping uses the same layout to construct structured graph views for self-supervised learning.

Relative positional encoding. We encode $\rho(e) = (\rho_1(e), \rho_2(e))$ using two-dimensional rotary position embeddings (RoPE) (Su et al., 2024; Heo et al., 2024). Let $R(\rho(e))$ denote the corresponding rotation. For an edge-token state h_e, $\widetilde{q}_e = R(\rho(e))W_Q h_e, \qquad \widetilde{k}_e = R(\rho(e))W_K h_e$. The rotary representation satisfies $R(p)^\top R(q) = R(q-p)$, so the attention score between edge tokens e and f becomes $\widetilde{q}_e^\top \widetilde{k}_f = (W_Q h_e)^\top R\left(\rho(f) - \rho(e)\right) W_K h_f$. This row–offset parameterization ensures that the relative displacement $\rho(f) - \rho(e)$ is structurally meaningful. The difference in the first axis captures the displacement between the edges' starting vertices in the canonical sequence, while the difference in the second axis captures the relative variation in their canonical spans. Consequently, RoPE allows the attention mechanism to naturally distinguish structural roles—such as densely connected local neighborhoods versus long-range topological bridges—based entirely on relative canonical geometry, independent of arbitrary input vertex labels.

Canonical cropping. The canonical vertex indices define a canonicalized adjacency matrix for the graph. Because the graph is undirected, we restrict our attention to the strictly upper right triangle of this matrix. Crucially, canonicalization naturally organizes the vertices such that structurally similar nodes are positioned near one another. As a result, this adjacency grid acquires meaningful spatial locality, much like an image.

To construct a cropped view for the student, we leverage this spatial structure. Rather than dropping edges independently at random—which destroys local structure—canonical cropping removes a contiguous band of anti-diagonals from the upper triangle. Because the matrix is spatially organized, this band corresponds to a localized group of edges. Dropping it removes a continuous

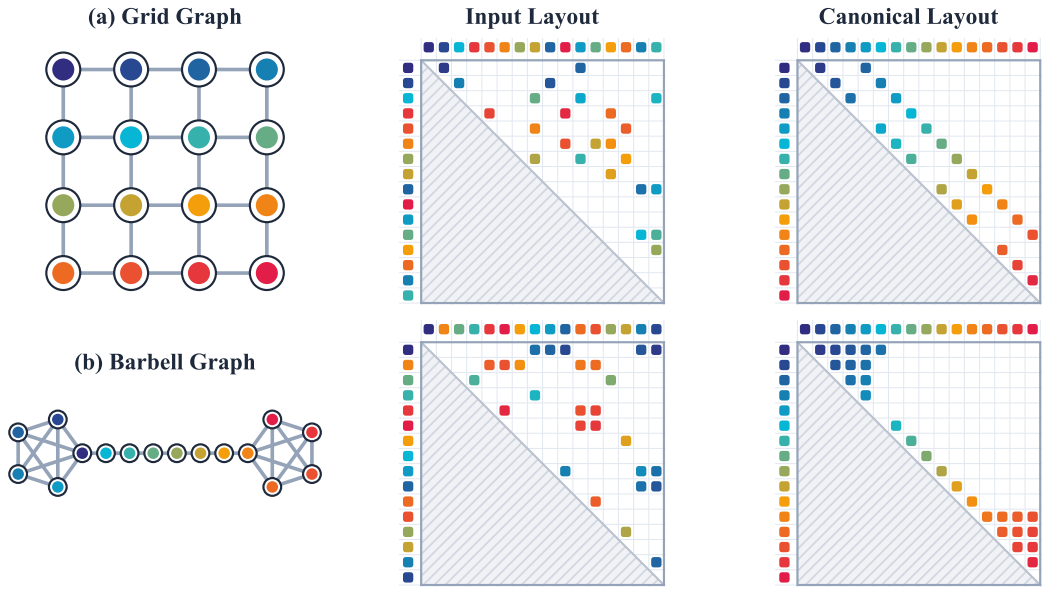

Figure 2: **Structured token spaces.** Spectral canonicalization (right) transforms scattered, arbitrary adjacency matrices (middle) into consistent 2D layouts. Grouping local neighborhoods together allows Transformers to natively process graphs like images.

chunk of the graph's topology—much like cropping a section out of an image—creating a structurally distinct view for the student model. All vertices are retained, and no unobserved vertex pair is introduced as a token.

The remaining graph G_{crop} is then recanonicalized based on its surviving topology and receives new coordinates determined solely by that topology. Consequently, a surviving edge's canonical position may shift between full and cropped views, ensuring that cropped coordinates depend only on the remaining topology. The exact deterministic crop schedule and precomputation procedure are given in Appendix E.6

3.3 GRAPH DINO

GINO processes the edge tokens using a standard Transformer architecture. Graph structure is encoded entirely through the observed edge tokens and their canonical 2D RoPE, processed via global self-attention.

We pretrain GINO using a standard DINO/iBOT-style student–teacher objective (Caron et al., 2021; Zhou et al., 2021). The teacher encodes the full graph G, while the student encodes the cropped view $G_S = G_{\text{crop}}$ derived in Section 3.2. Because G_S is recanonicalized independently, a surviving edge $e \in E_S$ receives a student coordinate $\rho_S(e)$ that may differ from its teacher coordinate $\rho_T(e)$. Consequently, teacher–student correspondence is anchored by the underlying edge identity rather than its spatial coordinate.

The pretraining objective combines two complementary perturbations: **Graph-level self-distillation (DINO)** aligns the global representations of the cropped student and full teacher graphs, while **Edge-level masked prediction (iBOT)** requires the student to predict a masked subset of its surviving edge tokens against corresponding unmasked teacher targets.

Cropping alters the graph topology and its resulting coordinate system, whereas masking leaves the cropped topology unchanged while substituting only token content. The final objective optimizes a weighted sum of losses, including DINO and iBOT. Complete architectural specifications, including exponential moving average parameter updates, projection head structures, and specific loss temperatures, are detailed in Appendix E.

4 EXPERIMENTS

We evaluate GINO along three axes: **1) Canonicalization robustness:** assessing consistency under vertex relabeling and certification coverage; **2) Spectral rounding in practice:** measuring adherence to the spectral analysis of Section 2.3; and **3) Molecular transfer:** evaluating downstream classification of self-supervised representations. We conduct these experiments using the GEOM pretraining corpus (Axelrod & Gomez-Bombarelli, 2022) and eight MoleculeNet classification datasets (Wu et al., 2018). Complete details are in Appendix F.

4.1 CANONICALIZATION ROBUSTNESS

To verify strict invariance to vertex relabeling, both random relabeling audits and cross-checks against BLISS (Junttila & Kaski, 2007) yielded zero discrepancies, successfully certifying canonical forms for 99.975% of the corpus ($\sim$304k molecules) at just 35 ms per molecule.

By deterministically breaking structural ties with atom features, GINO guarantees a unique, invariant assignment for the entire attributed graph. Consequently, our robustness audits confirm zero discrepancies not only in the topological coordinates but in the exact placement of all node and edge attributes. Additional stress tests, along with certification and quarantine details, are provided in Appendix .

4.2 SPECTRAL ROUNDING IN PRACTICE

To evaluate the practical significance of Theorem 2.2, we analyze $495{,}833$ graphs drawn from GEOM (Axelrod & Gomez-Bombarelli, 2022), MoleculeNet (Wu et al., 2018), and OGBG-MolHIV (Hu et al., 2020). The vast majority of this population ($474{,}144$ graphs, or 95.63%) fall into the simple-eigenvalue, sign-resolved regime. For these graphs, let $\widehat{u}_2$ denote the oriented Fiedler vector and τ_{spec} the ordering obtained by sorting its entries. We measure the continuous-to-discrete rounding error as $\varepsilon = \|r_{\tau_{\mathrm{spec}}} - \widehat{u}_2\|_2$, which yields a median error of 0.226.

Theorem 2.2 bounds the energy gap $E(\tau_{\mathrm{spec}}) - E(\tau_*)$ by $B_{\mathrm{full}} = B_D + B_R$, where B_D captures the degree-induced mismatch between the discrete centering constraint and the Laplacian's null direction, and B_R captures the continuous-to-discrete rounding error. We evaluate these quantities graph by graph over the full spectral regime. The median value of B_{full} is 0.0349, and $\mathrm{median}[B_{\mathrm{full}}/E(\tau_{\mathrm{spec}})] = 0.0632$. Thus, the theorem limits the potential improvement over the spectral ordering to 6.32% of its energy at the median. The bound is dominated by spectral rounding: the median fraction contributed by the degree-mismatch term, B_D/B_{full}, is just 0.134.

To directly assess the tightness of this bound, we exhaustively evaluate all discrete orderings for the $4{,}253$ graphs with $n \leq 10$, allowing the optimal ordering τ_* to be computed exactly. On this subset, the median realized optimality gap is $E(\tau_{\mathrm{spec}}) - E(\tau_*) = 0.0021$, while the median ratio of the realized gap to the theoretical bound is $(E(\tau_{\mathrm{spec}}) - E(\tau_*))/B_{\mathrm{full}} = 0.032$ Thus, the realized gap occupies only 3.21% of the graph-specific upper bound at the median, demonstrating that the full bound is substantially conservative on this exactly solvable small-graph subset. Moreover, the production ordering attains the discrete optimum on 32% of these instances.

While Theorem 2.2 analyzes the direct spectral ordering τ_{spec}, the production canonicalizer closely matches this theoretical baseline (median Kendall $\tau_b = 0.992$). Its deterministic refinement only reorders vertices whose Fiedler coordinates are indistinguishable within numerical tolerance.

4.3 SELF-SUPERVISED MOLECULAR PRETRAINING

To evaluate representation learning utility, we pretrain the model using our graph adaptation of the DINO/iBOT self-supervised objectives (Caron et al., 2021; Zhou et al., 2021) (Section 3.3).

We consider two evaluation settings. In the *frozen* setting, the encoder is fixed, and a linear predictor is trained on a graph-level embedding obtained by mean-pooling the final edge-token states. In the *fine-tuned* setting, the entire encoder is updated end-to-end for each task.

Comparison with 2D and Multimodal Baselines. Table 2 evaluates GINO against established 2D self-supervised baselines—including InfoGraph (Sun et al., 2020), GROVER (Rong et al.,

Table 1: **Empirical spectral bound.** Evaluated across $474k$ graphs in the sign-resolved regime. The exact-optimum (τ_*) subset contains $4{,}253$ small graphs ($n \leq 10$). All reported values are medians.

Quantity	Value
B_{full}	0.035
$B_{\text{full}}/E(\tau_{\text{final}})$	6.3%
$E(\tau_{\text{final}}) - E(\tau_*)$	0.002
Realized gap $\big(E(\tau_{\text{final}}) - E(\tau_*)\big)/B_{\text{full}}$	3.2%
$E(\tau_{\text{final}}) = E(\tau_*)$	32%

Table 2: **Molecular self-supervised transfer on MoleculeNet.** ROC-AUC (%) across eight classification datasets. **Bold** values denote the highest reported result per column in each setting. GINO achieves the highest average ROC-AUC in both the frozen and fine-tuned settings, outperforming all 2D baselines.

Method	BBBP	BACE	ClinTox	HIV	MUV	SIDER	Tox21	ToxCast	Avg. ↑
Frozen encoder									
InfoGraph	65.90	67.40	61.40	71.40	63.90	57.20	65.80	54.60	63.44
GROVER	67.00	69.00	65.00	67.80	62.70	**59.90**	63.90	53.60	63.62
GraphCL	64.70	**77.00**	60.80	**72.50**	60.60	59.50	69.10	56.20	65.04
C-FREE$_{\text{2D-MOL}}$	60.50	75.80	62.70	68.70	67.60	59.00	76.10	62.70	66.63
C-FREE$_{\text{2D-SUB}}$	64.20	73.90	71.40	65.50	64.60	58.00	**76.70**	**63.90**	67.27
GINO (ours)	**80.39**	76.77	**79.94**	65.07	**68.70**	56.24	69.53	58.76	**69.43**
Fine-tuned encoder									
GraphCL	69.70	75.40	76.00	78.50	69.80	60.50	73.90	62.40	70.78
GraphLog	72.50	83.50	76.70	77.80	76.00	61.20	75.70	63.50	73.40
GraphMAE	72.00	83.10	82.30	77.20	76.30	60.30	75.50	64.10	73.85
MGSSL	69.70	79.10	80.70	**78.80**	**78.70**	61.80	76.50	64.10	73.70
Mole-BERT	71.90	80.80	78.90	78.20	78.60	**62.80**	**76.80**	64.30	74.04
GINO (ours)	**87.87**	**85.57**	**91.97**	76.35	77.87	59.52	74.95	**66.27**	**77.55**

2020), GraphCL (You et al., 2020), GraphLoG (Xu et al., 2021), GraphMAE (Hou et al., 2022), MGSSL (Zhang et al., 2021), and Mole-BERT (Xia et al., 2023)—with baseline scores sourced primarily from C-FREE (Ariguib et al., 2026). GINO consistently outperforms these 2D frameworks. In the frozen setting, it achieves an average ROC-AUC of 69.43, surpassing the prior best of 67.27 (C-FREE$_{\text{2D-SUB}}$). When fine-tuned end-to-end, it reaches an average ROC-AUC of 77.55, a success driven explicitly by self-supervised pretraining (yielding a $+0.028$ mean validation ROC-AUC improvement over the untrained step-zero baseline). Beyond outperforming these 2D benchmarks, GINO remarkably rivals multimodal approaches: it outperforms C-FREE's 3D-aware SchNet baselines and remains highly competitive with their top PaiNN variant (extended comparisons in Appendix G.4).

5 Conclusion

GINO introduces a consistent coordinate system for graphs through spectral canonicalization. By assigning canonical 2D coordinates to edges, GINO naturally extends standard 2D RoPE and DINO/iBOT pretraining to permutation-symmetric graphs without modifying the core Transformer architecture. Theoretically, this spectral ordering yields the discrete ranking closest to the continuous Fiedler coordinate with a strictly bounded energy gap; empirically, it guarantees strict invariance to vertex relabeling and drives competitive downstream molecular transfer using only 2D graph topology. Ultimately, GINO demonstrates a powerful principle for representation learning: resolving arbitrary labeling symmetries prior to encoding empowers off-the-shelf Transformers to natively process graph-structured data.

AI USE STATEMENT

Generative AI tools were used during the development of this work as research assistants. Their use included refining mathematical statements and arguments, providing feedback on proofs, methodology, and experimental design, assisting with software implementation and debugging, and helping interpret experimental results. Generative AI tools were also used for brainstorming, literature search and synthesis, manuscript organization, drafting and editing text for clarity, and improving the presentation of figures, tables, and references.

All mathematical claims, proofs, experimental procedures, numerical results, citations, and AI-assisted text were subsequently reviewed and verified by the authors. Experimental results reported in the paper were obtained from the authors' implementations and computational experiments; generative AI was not used to fabricate or substitute experimental measurements. The authors take full responsibility for the final content of the paper, including all claims, results, text, code, and other artifacts produced with the assistance of generative AI.

ETHICS STATEMENT

This work develops methods for representation learning on molecular graphs and evaluates them using established molecular benchmark datasets. The study does not involve human participants, personally identifiable information, or sensitive personal data. The proposed method is methodological and is not intended, by itself, to support clinical, medical, or other safety-critical decisions.

We use existing datasets for research evaluation and follow their applicable usage conditions. As with other machine-learning methods for molecular representation learning, downstream use may require additional domain-specific validation, particularly in settings in which model predictions could affect health, safety, or other consequential decisions. The authors have considered the broader impacts of the work and are not aware of additional ethical concerns specific to the experiments presented in this paper.

REPRODUCIBILITY STATEMENT

We provide the information needed to reproduce the main theoretical and empirical results throughout the paper and supplementary material. Section 2 states the assumptions and main theoretical guarantees underlying the spectral ordering, with complete supporting derivations provided in the appendix. The appendix additionally documents the full canonicalization procedure, including numerical tolerances, structural refinement, fallback procedures, and handling of degenerate cases. It also specifies the graph preprocessing and feature construction, model architecture, self-supervised pretraining procedure, optimization and downstream evaluation protocols, dataset processing, hyperparameters, and the protocols used for the canonicalization and spectral-ordering experiments. These details are intended to make both the canonicalization pipeline and the reported experimental results independently reproducible.

REFERENCES

Boshra Ariguib, Mathias Niepert, and Andrei Manolache. Learning the neighborhood: Contrast-free multimodal self-supervised molecular graph pretraining. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, 2026.

Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9650–9660, 2021.

Benjy Friedmann and Michael Werman. Canonnet: Canonical ordering and curvature learning for point cloud analysis, 2025.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 2017.

Byeongho Heo, Song Park, Dongyoon Han, and Sangdoo Yun. Rotary position embedding for vision transformer. In *European Conference on Computer Vision*, pp. 289–305, 2024. doi: 10.1007/978-3-031-72684-2_17.

Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. Graphmae: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. In *Advances in Neural Information Processing Systems*, 2020.

Tommi Junttila and Petteri Kaski. Engineering an efficient canonical labeling tool for large and sparse graphs. In *Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments (ALENEX)*, pp. 135–149, 2007. doi: 10.1137/1.9781611972870.13.

Jinwoo Kim, Tien Dat Nguyen, Seonwoo Min, Sungjun Cho, Moontae Lee, Honglak Lee, and Seunghoon Hong. Pure transformers are powerful graph learners. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Devin Kreuzer, Dominique Beaini, William L. Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 60936–60979, 2024.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 14501–14515, 2022.

Isaac Reid, Arijit Sehanobish, Cederik Höfs, Bruno Mlodozeniec, Leonhard Vulpius, Federico Barbero, Adrian Weller, Krzysztof Choromanski, Richard E Turner, and Petar Veličković. Rotary position encodings for graphs. *arXiv preprint arXiv:2509.22259*, 2025.

Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. In *Advances in Neural Information Processing Systems*, volume 33, pp. 12559–12571, 2020.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 464–468, 2018. doi: 10.18653/v1/N18-2074.

Jianlin Su, Yu Lu, Shengfeng Pan, Ahmed Murtadha, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

Fan-Yun Sun, Jordan Hoffman, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. In *International Conference on Learning Representations*, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. Moleculenet: A benchmark for molecular machine learning. *Chemical Science*, 9(2):513–530, 2018.

Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, and Stan Z. Li. Mole-BERT: Rethinking pre-training graph neural networks for molecules. In *International Conference on Learning Representations*, 2023.

Minghao Xu, Hang Wang, Bingbing Ni, Hongyu Guo, and Jian Tang. Self-supervised graph-level representation learning with local and global structure. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11548–11558. PMLR, 2021.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do transformers really perform bad for graph representation? In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Cai Zhou, Zijie Chen, Zian Li, Jike Wang, Kaiyi Jiang, Pan Li, Rose Yu, Muhan Zhang, Stephen Bates, and Tommi Jaakkola. Rethinking diffusion models with symmetries through canonicalization with applications to molecular graph generation. In *The 2026 Workshop on Generative and Agentic AI for Biology*, 2026. URL https://openreview.net/forum?id=TqNEdNHfjQ.

Jinghao Zhou, Chen Wei, Huiyu Wang, Wei Shen, Cihang Xie, Alan Yuille, and Tao Kong. ibot: Image bert pre-training with online tokenizer. *arXiv preprint arXiv:2111.07832*, 2021.

Supplementary Material

Appendices

A **Production Canonicalizer and Scope of the Guarantee** 15

B **Proofs and Spectral Theory** 19

C **Canonicalization Audits and Spectral Rounding** 23

D **Datasets, Splits, and Evaluation Populations** 29

E **Exact GINO Architecture and Canonical Views** 31

F **Pretraining and Statistical Protocol** 34

G **Self-Supervised Molecular Transfer** 36

H **Positional and Ordering Controls** 37

I **Model-Design Ablations** 39

J **Efficiency, Scaling, and Compute** 41

K **Cross-Domain Portability Diagnostics** 42

L **Reproducibility, Result Lineage, and Claim Boundaries** 42

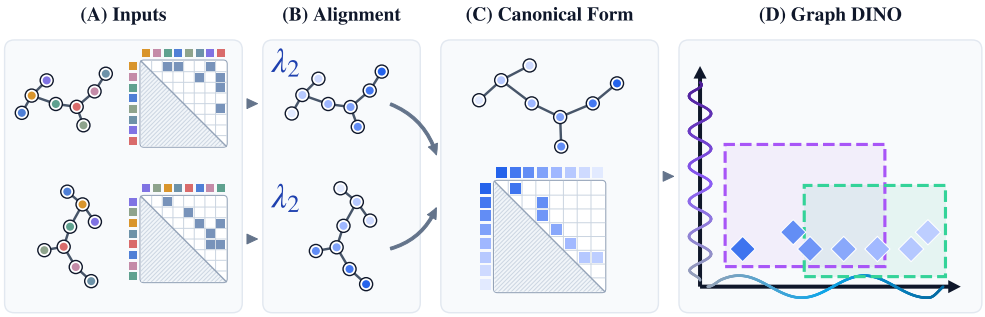

Figure 3: **Spectral canonicalization establishes a consistent 2D coordinate system for graphs. (A) Inputs:** Arbitrary permutations produce drastically different adjacency layouts. **(B-C) Alignment & Canonical Form:** Sorting nodes by the structural Fiedler vector (λ_2) maps isomorphic graphs into one canonical form. **(D) Graph DINO:** This canonical grid natively enables 2D relative positional encodings (Axial RoPE) and spatial crop masking for Teacher-Student learning.

A Production Canonicalizer and Scope of the Guarantee

The main paper introduces Fiedler sorting as the primary canonicalization mechanism. The production implementation uses this spectral order when it is uniquely resolved, refines residual spectral ambiguity using label-independent structural information, and invokes bounded exact canonical labeling only when ambiguity remains. The implementation stages are defined below: the production sign cascade (C6), repeated-eigenspace projector refinement (M2), and bounded exact fallback (M3). Their combined version identifier is `fiedler_v2_float64_c6_m2_m3`.

Implementation identifiers. The labels `C6`, `M2`, and `M3` are internal identifiers inherited from two canonicalization design studies; they are not successive steps numbered from one. The sign study compared a legacy rule, `C0`, with corrected candidate cascades `C1`–`C8`. All corrected candidates were permutation-stable in the final comparison, and `C6_fiedler_only` was selected by the prespecified preference for a Fiedler-only, cheaper, and simpler rule. The production cascade called `C6` is therefore the four-stage S1–S4 procedure defined in Appendix A.5; the S1–S4 labels are its internal decision stages and are unrelated to the C1–C8 candidate numbering. A separate repeated-eigenspace study compared methods `M0`–`M3`. Production uses the basis-invariant projector refinement denoted `M2` when the Fiedler eigenspace is repeated, and the bounded exact individualization–refinement procedure denoted `M3` whenever ambiguity remains. There is no `M4` stage in the production canonicalizer. We retain these identifiers only to match the executable configuration and experiment records; the algorithmic roles are described explicitly below.

Experiment status follows the lineage terminology defined in Appendix L; unless stated otherwise, molecular results refer to the current Family-C configuration, i.e., the observed-bond, recanonicalized-crop recipe specified in Appendices E and F.

A.1 Canonicalization contract

Let $B \in \mathbb{R}^{n\times n}$ denote the symmetric structural matrix supplied to the canonicalizer, and let Q_B denote a permutation matrix returned by a certified run. The corresponding canonicalized structural matrix is

$$\mathcal{C}(B) = Q_B B Q_B^\top. \tag{6}$$

If certification fails, the procedure returns Quarantine instead of a permutation. For any relabeling $\widetilde{B} = \Pi B \Pi^\top$, certified outputs satisfy

$$\mathcal{C}(\widetilde{B}) = \mathcal{C}(B). \tag{7}$$

Algorithm 2 Production structural canonicalization

Require: symmetric structural matrix B
Ensure: certified permutation Q_B or QUARANTINE

handle disconnected inputs componentwise if necessary
construct weighted shortest paths, dense affinity A, and normalized L
compute the first nonzero eigenspace of L in float64
if the eigenspace is one-dimensional **then**
    orient the Fiedler vector using the production sign cascade
    if the orientation resolves **then**
        form spectral tolerance cells and refine them structurally
        if all cells are singletons **then**
            return the resulting order
        end if
        run exact individualization–refinement seeded by the ordered cells
    else
        initialize a label-independent structural partition and run exact individualization–refinement without a spectral ordering
    end if
else
    initialize cells from the eigenspace projector and structurally refine
    if all cells are singletons **then**
        return the resulting order
    end if
    run exact individualization–refinement without a spectral ordering
end if
if the bounded exact search certifies its minimum **then**
    return the certified order
end if
return QUARANTINE

The realizing vertex permutations need not be unique. In the presence of graph automorphisms they are determined only up to

$$Q_{\widetilde{B}}\Pi = Q_B S, \qquad S \in \mathrm{Aut}(B). \tag{8}$$

Thus the guarantee concerns the canonicalized structural matrix, not a unique vertex ordering: vertices in the same automorphism orbit may occupy interchangeable positions.

A.2 SCOPE FOR ATTRIBUTED MOLECULAR GRAPHS

The canonicalizer supports both structure-only and attribute-aware canonicalization. In the attribute-aware configuration, feature information can participate in resolving residual structural ambiguities. All experiments reported in this paper, however, use the more information-restricted `topology_only=True` configuration: the canonical order is determined exclusively from the bond-order-weighted structural graph, and atom attributes are attached only after canonicalization. Consequently, the invariance guarantee for the reported experiments concerns the canonical structural form $\mathcal{C}(B)$ rather than a unique attributed vertex-to-position assignment. In particular, structurally automorphic vertices carrying different atom attributes may exchange canonical positions while leaving $\mathcal{C}(B)$ unchanged. The attribute-aware configuration (`topology_only=False`) can use feature information to refine such residual ambiguities, but was not used to produce the pretraining or downstream results reported in this work. The corresponding structure-only invariance audits are reported in Appendix C.

A.3 PRODUCTION DECISION FLOW

Algorithm 2 summarizes the production hierarchy: use the Fiedler order when it is uniquely resolved, refine spectral ties using invariant structural information, and invoke exact canonical labeling only when ambiguity remains. The following subsections define these stages and their numerical tolerances.

Stage	Decision rule	Production tolerance
S1	common sign among maximal-$\lvert u\rvert$ entries	relative tie 10^{-9}
S2	standardized third central moment m_3	$\lvert m_3\rvert > 10^{-6}$
S3	first resolving moment in m_3, m_5, m_7	$\lvert m_k\rvert > 10^{-6}$
S4	sorted-lexicographic comparison of u and $-u$	relative 10^{-9}

Table 3: Production sign-resolution cascade. If every stage abstains, no index-dependent sign rule is introduced; the graph proceeds to exact structural canonical labeling.

A.4 Structural signal and spectral operator

The structural edge values are

$$\text{single} \mapsto \tfrac{1}{3}, \qquad \text{aromatic} \mapsto \tfrac{1}{2}, \qquad \text{double} \mapsto \tfrac{2}{3}, \qquad \text{triple} \mapsto 1. \tag{9}$$

The matrix is symmetric with zero diagonal; if duplicate entries occur for one pair, the largest structural value is retained.

Positive structural weights are converted to path lengths

$$\ell_{ij} = B_{ij}^{-1} \tag{10}$$

on the support $|B_{ij}| > 10^{-12}$. Weighted all-pairs shortest paths are computed by Floyd–Warshall and divided by the largest finite positive distance, so the weighted diameter is one. Writing the normalized distances as d_{ij}, the dense affinity is

$$A_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right), \qquad A_{ii} = 0, \qquad \sigma = 0.35. \tag{11}$$

The affinity is symmetrized after construction. The production spectral operator is the symmetric normalized Laplacian

$$L = I - D^{-1/2} A D^{-1/2}, \qquad D = \operatorname{diag}(A\mathbf{1}), \tag{12}$$

with degrees clamped below by 10^{-12}. The spectral stage uses float64 and `torch.linalg.eigh`.

The implementation first skips the null block according to the number of connected components (one eigenpair for the connected graphs used by the main spectral branch), rather than applying a separate eigenvalue-zero threshold. Let λ_F denote the first eigenvalue after this block; for a connected graph, $\lambda_F = \lambda_2$ is the Fiedler eigenvalue. Eigenvalues are treated as belonging to the same first nonzero eigenspace when

$$|\lambda_j - \lambda_F| \le \max\bigl(10^{-10}, 10^{-8}\max(1, |\lambda_F|)\bigr). \tag{13}$$

The resulting branch is determined by this eigenspace multiplicity, as summarized in Algorithm 2.

A.5 Simple second eigenvalue: production sign orientation

For a connected graph, a one-dimensional first nonzero eigenspace yields the usual Fiedler vector, which is defined only up to a global sign. The production sign cascade (internal identifier `C6`) applies the following permutation-invariant tests in order and uses the first test that determines a unique orientation.

At S1, all entries satisfying $|u_i| \ge (1 - 10^{-9})\max_j |u_j|$ must share a sign; the orientation is chosen so that these maximal-magnitude entries are positive. S2 orients the vector so the standardized third central moment m_3 is positive whenever $|m_3| > 10^{-6}$. S3 is the implemented generic odd-moment stage: it checks m_3, m_5, m_7 in that order with the same 10^{-6} threshold. Thus its first m_3 check is redundant after S2 under the current equal tolerances, but it is retained here because that is the executed cascade. If all moment tests abstain, S4 compares the descending sorted coordinate lists for u and $-u$ and uses the first difference exceeding $10^{-9}\max_i |u_i|$. If S4 also abstains, no orientation is returned. Vertices are sorted by the resolved coordinate in ascending order. Empirical resolution rates are reported in Appendix C.1.

A.6 TOLERANCE CELLS AND STRUCTURAL REFINEMENT

After orientation, float64 Fiedler coordinates are stably sorted and grouped into tolerance cells using the first coordinate u_a in each cell as a fixed reference. A later sorted coordinate u_j remains in that cell exactly when

$$|u_j - u_a| \le 10^{-9} + 10^{-7} \max_i |u_i|. \tag{14}$$

Anchoring every comparison to u_a, rather than chaining successive near-equalities, prevents cell membership from depending on a sequence of pairwise-near coordinates. Distinct cells retain their spectral order.

Within unresolved cells, the simple-spectrum branch performs at most six rounds of structural refinement using the structural matrix B only. For each vertex i and each current cell C, the descriptor contains

$$\left(\sum_{j \in C} B_{ij},\ \sum_{j \in C} B_{ij}^2,\ \max_{j \in C} B_{ij},\ \min_{j \in C} B_{ij},\ \#\{j \in C : |B_{ij}| > 10^{-12}\} \right). \tag{15}$$

Descriptor columns are compared with absolute tolerance 10^{-9} plus relative tolerance 10^{-7} times the global scale of that column. The dense kernel is excluded from this refinement because roundoff-scale kernel asymmetries can be amplified into label-dependent splits.

A.7 REPEATED SECOND EIGENSPACES: PROJECTOR REFINEMENT

When the first nonzero eigenspace has multiplicity $m > 1$, individual eigenvectors are not canonical. Let $U \in \mathbb{R}^{n \times m}$ be any orthonormal basis and set

$$P = UU^\top. \tag{16}$$

The projector is invariant to the basis choice. The repeated-eigenspace branch (internal identifier `M2`) initializes cells from

$$[P_{ii}, \mathrm{sort}_\downarrow(P_{i,:})] \tag{17}$$

and applies at most 12 refinement rounds using descriptors computed from both B and P, with the same five cell aggregates as Eq. (15). Any residual ambiguity proceeds to the exact fallback. In the executed production path, however, an `M2` outcome that remains ambiguous does *not* seed M3 with the M2 partition: M3 restarts from a label-independent structural partition. Empirical multiplicity and routing statistics are reported in Appendix C.1.

A.8 EXACT INDIVIDUALIZATION–REFINEMENT FALLBACK

The exact fallback (internal identifier `M3`) performs deterministic individualization–refinement on the structural matrix. M3 has two entry modes. If spectral orientation succeeded, the remaining ordered tolerance cells initialize the search. Otherwise—after sign abstention or a repeated eigenspace—M3 initializes a label-independent structural partition and proceeds without a spectral ordering.

At each non-leaf state, the smallest non-singleton cell is selected, with ties broken by its position in the ordered partition. Each candidate vertex is then individualized, the resulting partition is refined to a fixed point, and the search recurses. At a leaf, the strict upper triangle of the permuted structural matrix is concatenated with its diagonal and rounded to nine decimal places; we call this sequence the leaf serialization. The lexicographically smallest serialization defines the canonicalized structural matrix. Equal minimum serializations are automorphic realizations of the same structural form.

The bounded-search settings are

If exhaustive certification exceeds the leaf-count, wall-clock, or input-vertex limit, the canonicalizer discards any best-so-far ordering and returns QUARANTINE. Certification coverage and failure routes are reported in Appendix C.3.

Setting	GEOM	MoleculeNet
Maximum leaves	512	2,048
Wall-clock safety valve	30 s	60 s
M3 input-vertex limit	128	128
Builder graph-size cap	$n \leq 91$	$n \leq 91$

Table 4: Bounded exact-search settings. The molecular builder limits the complete strict-upper-triangular pair grid to at most 4,096 possible vertex pairs; $\binom{91}{2} = 4,095$, which yields the effective $n \leq 91$ builder cap. Under this cap, the M3 input-vertex limit is inactive for the reported molecular datasets.

A.9 Disconnected and trivial graphs

Disconnected inputs are supported by recursively canonicalizing each connected component. Components are ordered first by size and then by their nine-decimal canonical structural serialization. Equal component keys identify isomorphic components and may be exchanged by an automorphism. If any component cannot be certified, the whole graph is quarantined. Inputs with $n \leq 2$ are handled by invariant structural refinement without an index-based tie-break. The reported GEOM and MoleculeNet builders exclude disconnected molecules before model evaluation, so the disconnected branch is not exercised by the principal molecular results.

B Proofs and Spectral Theory

This appendix has two parts. First, we prove that every certified branch of the production canonicalizer is invariant to vertex relabeling. Second, we analyze Fiedler sorting as a discrete rank approximation and derive both the discrete-optimum energy-gap theorem used by the main theoretical claim and the separate excess-energy bound used by the production-order diagnostic.

B.1 Relabeling covariance

Lemma B.1 (Covariance of the spectral construction)**.** *Let $\widetilde{B} = \Pi B \Pi^\top$. The weighted shortest-path matrix, dense affinity, degree matrix, and symmetric normalized Laplacian satisfy*

$$\widetilde{\Delta} = \Pi \Delta \Pi^\top, \qquad \widetilde{A} = \Pi A \Pi^\top, \qquad \widetilde{D} = \Pi D \Pi^\top, \qquad \widetilde{L} = \Pi L \Pi^\top.$$

Proof. Relabeling changes only vertex names and therefore preserves the weight of every path. Shortest-path distances transform by simultaneous row/column permutation. Diameter normalization and the Gaussian kernel commute with simultaneous permutation, giving the result for A. The identities for D and L then follow directly. □

Lemma B.2 (Simple-eigenspace covariance)**.** *If λ_2 is simple, a unit second eigenvector of the relabeled operator satisfies*

$$\widetilde{u}_2 = \pm \Pi u_2.$$

Proof. Since $\widetilde{L} = \Pi L \Pi^\top$, $\widetilde{L}(\Pi u_2) = \Pi L u_2 = \lambda_2 \Pi u_2$. Simplicity leaves only the global sign ambiguity. □

Lemma B.3 (Sign-cascade covariance)**.** *Whenever the production sign cascade resolves, the oriented vectors satisfy*

$$\widehat{\widetilde{u}}_2 = \Pi \widehat{u}_2.$$

Proof. Each stage is label independent and transforms consistently under the global sign flip $u \mapsto -u$. Relabeling preserves every multiset, moment, and sorted comparison used by the cascade, while changing u to $-u$ reverses the selected orientation whenever a stage resolves. Hence the first resolving stage is preserved by relabeling and the deterministic orientation commutes with the permutation. □

B.2 Repeated eigenspaces and exact fallback

Lemma B.4 (Projector covariance). *For an orthonormal basis U of a repeated eigenspace, $P = UU^\top$ is basis invariant. Under relabeling,*

$$\widetilde{P} = \Pi P \Pi^\top.$$

Proof. For any orthogonal R, $(UR)(UR)^\top = UU^\top$. A basis of the relabeled subspace can be written ΠUR, hence $(\Pi UR)(\Pi UR)^\top = \Pi UU^\top \Pi^\top$. □

Lemma B.5 (Equivariance of projector-cell refinement). *Let $P = UU^\top$ be the projector used by the repeated-eigenspace branch. The initial projector partition and every subsequent refinement step using B and P are permutation equivariant.*

Proof. Under relabeling, $B \mapsto \Pi B \Pi^\top$ and, by the previous lemma, $P \mapsto \Pi P \Pi^\top$. The initial descriptor for vertex i is $[P_{ii}, \mathrm{sort}_\downarrow(P_{i,:})]$. Relabeling only permutes the row entries before sorting, so the descriptor attached to vertex i is transported unchanged to vertex Πi. During refinement, each descriptor is formed from sums, sums of squares, extrema, and nonzero counts of entries from B and P between a vertex and the current cells. These aggregates depend only on the corresponding multisets and are therefore preserved under simultaneous relabeling. Consequently, the ordered partition at every refinement round is transported by Π. □

Lemma B.6 (Equivariance of tolerance-cell refinement). *The tolerance-cell construction and structural refinement of Appendix A.6 are permutation equivariant.*

Proof. Relabeling preserves the multiset of oriented spectral coordinates and hence transports every tolerance cell by the same permutation. Each structural descriptor in Eq. (15) depends only on the multiset of structural edge values between a vertex and a current cell. Those multisets are preserved under simultaneous relabeling. Ordering newly created subcells by descriptor content is label independent, so the refined partition is transported equivariantly. □

Lemma B.7 (Exact-fallback canonicality). *If the exact individualization–refinement search of Appendix A.8 completes, the selected structural serialization is canonical.*

Proof. Relabeling induces a bijection between complete search leaves. Corresponding leaves have identical structural serializations, so both searches encounter the same set of leaf keys and choose the same lexicographically minimal key. The order in which candidates are enumerated cannot change this minimum when the search is exhaustive. If the bounded search does not complete, no canonical output is returned. □

Theorem B.8 (Certified canonicality). *For every $\widetilde{B} = \Pi B \Pi^\top$, whenever the production canonicalizer returns certified outputs for both inputs,*

$$\mathcal{C}(\widetilde{B}) = \mathcal{C}(B).$$

Moreover, the realizing permutations satisfy

$$Q_{\widetilde{B}} \Pi = Q_B S \qquad \text{for some } S \in \mathrm{Aut}(B).$$

Proof. The preceding lemmas cover the simple-spectrum sign and tolerance-cell branch, the repeated-eigenspace projector branch including its B, P refinement, and the exact fallback. Thus every certified branch is transported by the same vertex relabeling, while the final exact structural serialization is invariant. Two realizing permutations that produce the same canonicalized structural matrix can differ only by an automorphism S of B. Hence

$$\begin{aligned}\mathcal{C}(\widetilde{B}) &= Q_{\widetilde{B}} \Pi B \Pi^\top Q_{\widetilde{B}}^\top \\ &= Q_B S B S^\top Q_B^\top \\ &= Q_B B Q_B^\top = \mathcal{C}(B).\end{aligned}$$

Uncertified branches return QUARANTINE rather than a best-so-far permutation. □

Having established that every certified production branch is canonical under vertex relabeling, we now turn to the second question: why the Fiedler coordinate provides a meaningful one-dimensional ordering.

B.3 Variational characterization of the implemented coordinate

For the dense affinity A, consider

$$\min_y \frac{1}{2}\sum_{i,j} A_{ij}(y_i - y_j)^2 \quad \text{s.t.} \quad y^\top Dy = 1, \qquad y^\top D\mathbf{1} = 0. \tag{18}$$

The objective equals $y^\top(D-A)y$, so stationary points satisfy $(D-A)y = \lambda Dy$. With $u = D^{1/2}y$, this becomes

$$Lu = \lambda u, \qquad L = I - D^{-1/2}AD^{-1/2}.$$

For a connected affinity graph, the normalized-Laplacian null vector is

$$u_1 = \frac{D^{1/2}\mathbf{1}}{\sqrt{\mathbf{1}^\top D\mathbf{1}}}, \tag{19}$$

and the first nonconstant direction satisfies

$$u_2 \in \arg\min_{\substack{\|u\|_2=1 \\ u^\top u_1 = 0}} u^\top Lu. \tag{20}$$

For any u,

$$u^\top Lu = \frac{1}{2}\sum_{i,j} A_{ij}\left(\frac{u_i}{\sqrt{D_{ii}}} - \frac{u_j}{\sqrt{D_{jj}}}\right)^2. \tag{21}$$

Equivalently, $y_2 = D^{-1/2}u_2$ minimizes the weighted Dirichlet energy in Eq. (18) subject to $y^\top Dy = 1$ and $y^\top D\mathbf{1} = 0$. The implementation uses the corresponding normalized-Laplacian eigenvector u_2 as its spectral coordinate.

B.4 Nearest discrete rank representation

For an ordering $\tau : V \to \{1, \dots, n\}$, define

$$r_\tau(i) = \frac{\tau(i) - (n+1)/2}{\sqrt{n(n^2-1)/12}}. \tag{22}$$

Then $r_\tau^\top \mathbf{1} = 0$ and $\|r_\tau\|_2 = 1$. We write S_n for the set of all permutations of the n vertices.

Proposition B.9 (Nearest rank representation)**.** *Let τ_{spec} sort the entries of the deterministically oriented Fiedler vector $\widehat{u}_2$ in nondecreasing order. Then*

$$\tau_{\text{spec}} \in \arg\min_{\tau \in S_n} \|r_\tau - \widehat{u}_2\|_2^2. \tag{23}$$

Proof. Because both vectors have unit norm,

$$\|r_\tau - \widehat{u}_2\|_2^2 = 2 - 2r_\tau^\top \widehat{u}_2.$$

Hence minimizing distance is equivalent to maximizing $r_\tau^\top \widehat{u}_2$. Every r_τ is a permutation of the same centered rank values. By the rearrangement inequality, the inner product is maximized by pairing increasing ranks with increasing entries of $\widehat{u}_2$, which is exactly the sorted Fiedler order. □

B.5 Energy gap to the optimal discrete ordering

Define the discrete normalized-Laplacian energy

$$E(\tau) := r_\tau^\top Lr_\tau, \qquad \tau_* \in \arg\min_{\tau \in S_n} E(\tau). \tag{24}$$

Let

$$e = \frac{\mathbf{1}}{\sqrt{n}}, \qquad \gamma_D = (e^\top u_1)^2 = \frac{\left(\sum_i \sqrt{D_{ii}}\right)^2}{n\sum_i D_{ii}} \in (0, 1]. \tag{25}$$

The quantity $1 - \gamma_D$ measures the mismatch between the uniform centering constraint satisfied by every rank vector and the degree-weighted null direction of the normalized Laplacian.

Lemma B.10 (Energy floor for rank vectors)**.** *For every ordering τ,*

$$E(\tau) \geq \gamma_D \lambda_2. \tag{26}$$

Proof. The vectors e and r_τ are orthonormal. Bessel's inequality applied to the unit vector u_1 gives

$$(e^\top u_1)^2 + (r_\tau^\top u_1)^2 \leq 1,$$

so

$$(r_\tau^\top u_1)^2 \leq 1 - \gamma_D.$$

Expanding r_τ in an orthonormal eigenbasis $u_1, \ldots, u_n$ of L yields

$$\begin{aligned} E(\tau) &= \sum_{k=2}^{n} \lambda_k (r_\tau^\top u_k)^2 \\ &\geq \lambda_2 \sum_{k=2}^{n} (r_\tau^\top u_k)^2 \\ &= \lambda_2 \big(1 - (r_\tau^\top u_1)^2\big) \\ &\geq \gamma_D \lambda_2. \end{aligned}$$

□

The bound below has two sources of looseness: the mismatch between rank centering and the normalized-Laplacian nullspace, and the approximation of the continuous Fiedler coordinate by discrete ranks.

Theorem B.11 (Energy gap of the spectral ordering)**.** *Let*

$$\varepsilon = \| r_{\tau_{\mathrm{spec}}} - \widehat{u}_2 \|_2.$$

Then the spectral ordering satisfies

$$E(\tau_{\mathrm{spec}}) - E(\tau_*) \leq \lambda_2(1 - \gamma_D) + (\lambda_n - \lambda_2) \left(\varepsilon^2 - \frac{\varepsilon^4}{4} \right). \tag{27}$$

Proof. The previous lemma gives $E(\tau_*) \geq \gamma_D \lambda_2$. Set

$$r = r_{\tau_{\mathrm{spec}}}, \qquad c = r^\top \widehat{u}_2.$$

Since both vectors have unit norm,

$$\varepsilon^2 = 2 - 2c, \qquad c = 1 - \frac{\varepsilon^2}{2},$$

and therefore

$$1 - c^2 = \varepsilon^2 - \frac{\varepsilon^4}{4}. \tag{28}$$

Because $\widehat{u}_2$ is a unit eigenvector with eigenvalue λ_2, expanding r in the same eigenbasis gives

$$\begin{aligned} E(\tau_{\mathrm{spec}}) &= \lambda_2 c^2 + \sum_{k=3}^{n} \lambda_k (r^\top u_k)^2 \\ &\leq \lambda_2 c^2 + \lambda_n \sum_{k=3}^{n} (r^\top u_k)^2 \\ &\leq \lambda_2 c^2 + \lambda_n (1 - c^2) \\ &= \lambda_2 + (\lambda_n - \lambda_2)(1 - c^2). \end{aligned}$$

Subtracting the lower bound on $E(\tau_*)$ and using Eq. (28) proves Eq. (27). □

The two contributions are

$$B_D := \lambda_2(1-\gamma_D), \qquad B_R := (\lambda_n - \lambda_2)\left(\varepsilon^2 - \frac{\varepsilon^4}{4}\right), \qquad B_{\text{full}} := B_D + B_R. \tag{29}$$

Here B_D captures the constraint mismatch and B_R the discrete rounding contribution; they are components of one upper bound rather than independent errors.

Corollary B.12 (Discrete-optimum bound for any rank ordering)**.** *Let τ be any ordering and define $\varepsilon_\tau = \|r_\tau - \widehat{u}_2\|_2$. Then*

$$E(\tau) - E(\tau_*) \le \lambda_2(1-\gamma_D) + (\lambda_n - \lambda_2)\left(\varepsilon_\tau^2 - \frac{\varepsilon_\tau^4}{4}\right). \tag{30}$$

Proof. The lower bound $E(\tau_*) \ge \gamma_D\lambda_2$ is independent of the chosen ordering. For $r = r_\tau$ and $c = r^\top\widehat{u}_2$, the same eigenbasis argument as in the preceding theorem gives

$$E(\tau) \le \lambda_2 + (\lambda_n - \lambda_2)\left(\varepsilon_\tau^2 - \frac{\varepsilon_\tau^4}{4}\right).$$

Subtracting the lower bound on $E(\tau_*)$ proves the claim. □

B.6 Excess energy relative to the continuous Fiedler level

The empirical analysis in Appendix C.4 evaluates the *final* production order, so we also require an inequality that applies to an arbitrary rank ordering.

Corollary B.13 (Fiedler-excess bound for any rank ordering)**.** *Let τ be any ordering, let $\varepsilon_\tau = \|r_\tau - \widehat{u}_2\|_2$, and define*

$$\Delta_E^+(\tau) := \max\{E(\tau) - \lambda_2, 0\}.$$

Then

$$\Delta_E^+(\tau) \le (\lambda_n - \lambda_2)\left(\varepsilon_\tau^2 - \frac{\varepsilon_\tau^4}{4}\right). \tag{31}$$

Proof. Set $r = r_\tau$ and $c = r^\top\widehat{u}_2$. Unit norms again give $1 - c^2 = \varepsilon_\tau^2 - \varepsilon_\tau^4/4$. Expanding r in the eigenbasis of L,

$$E(\tau) \le \lambda_2 c^2 + \lambda_n(1-c^2),$$

so

$$E(\tau) - \lambda_2 \le (\lambda_n - \lambda_2)(1-c^2).$$

The right-hand side is nonnegative, so replacing the left-hand side by its positive part preserves the inequality. □

Since the symmetric normalized Laplacian satisfies $\lambda_n \le 2$, a looser comparison is

$$\Delta_E^+(\tau) \le 2\left(\varepsilon_\tau^2 - \frac{\varepsilon_\tau^4}{4}\right) \le 2\varepsilon_\tau^2. \tag{32}$$

The two bounds address different quantities: Eq. (31) controls excess above the Fiedler energy λ_2, whereas Eq. (27) controls the gap to the optimal discrete rank ordering and additionally includes B_D.

C Canonicalization Audits and Spectral Rounding

We report the empirical checks supporting the production canonicalizer and the final spectral-rounding analysis. All statistics in this section use the final audited populations described below; the earlier singleton-only spectral analysis is superseded.

Table 5: **Canonicalization robustness, certification coverage, and preprocessing cost.** Relabeling audits test whether equivalent vertex labelings produce the same canonical topology, while the BLISS cross-check provides an independent reference implementation for canonical labeling. Molecules for which bounded exact search cannot certify a canonical form are quarantined before pretraining.

Check	Scope	Result
Canonical-form invariance		
Random relabeling audit	160 cases	0 failures
Built-dataset relabeling audit	1,536 cases	0 failures
BLISS cross-check	10,571 graphs $\times$20 relabelings	0 mismatches
GEOM certification coverage		
Certified by canonicalizer	303,796/303,873 molecules	99.975%
Quarantined by bounded search	77/303,873 molecules	0.025%
Preprocessing cost		
Runtime (accepted molecules)	GEOM corpus	35 ms median / 250 ms p_{99}

Production route	Accepted graphs
Sign-resolved spectral/refinement path without exact fallback	47,235
Sign-resolved path $\rightarrow$ accepted exact fallback	252,363
Sign-abstained path $\rightarrow$ accepted exact fallback	4,125
Repeated eigenspace $\rightarrow$ exact fallback	72
Single-vertex/trivial case	1
Total accepted	**303,796**

Table 6: Accepted GEOM molecules by final canonicalization route. The exact fallback is invoked for 256,560 accepted molecules in total: 252,363 after sign resolution, 4,125 after sign abstention, and 72 after a repeated Fiedler eigenspace.

C.1 Sign, multiplicity, and automorphism audits

A 100,000-molecule sign-resolution study (seed 20260723) used 5,173,695 permutation trials. The legacy float32 sign rule exhibited three explicit $u \leftrightarrow -u$ violations and 1,349 failing graphs. Each corrected candidate cascade, including the selected production cascade, had zero sign violations and zero relabeling failures among cases in which it returned an orientation. The selected cascade abstained on approximately 1.42% of simple-spectrum cases and routed them to exact structural refinement rather than introducing an index-dependent sign rule.

Across all 303,873 GEOM canonicalization attempts, the float64 multiplicity criterion identifies 75 repeated first nonzero eigenspaces: 72 of multiplicity two and 3 of multiplicity three. On the 100k study, the legacy float32 rule misclassified 14 of 23 genuinely repeated cases as simple.

The projector-based repeated-eigenspace branch was separately stress-tested on 120 graphs containing 2,908 vertices. It left 353 unresolved cells; all 353 were contained entirely within true automorphism orbits, and none mixed distinct orbits.

Separately, we test the attributed-graph scope of the guarantee rather than the projector branch itself. An audit of 400 connected BBBP molecules under three atom renumberings each found that the packed canonical structural matrix changed in 0/400 molecules, whereas the canonical-order node-feature matrix changed in 48/400 (12.0%). These results support the structural-canonicality scope stated in Appendix A.2.

Table 6 conditions on the 303,796 accepted GEOM molecules. Including the 77 quarantined attempts gives the full 303,873-row canonicalization diagnostic population.

Applying the production decision flow to the accepted molecules gives:

Across the full 303,873-attempt diagnostic population, the first sign test (maximal absolute coordinate) resolves 299,652 cases (98.61%). Of these, 47,235 are accepted without exact fallback,

Audit	Result
Legacy packed-structure relabeling test	63 / 160 failures
Production packed-structure relabeling test	**0 / 160 failures**
Production built-dataset relabeling audit	**0 / 1,536 failures**
BLISS audit, $10{,}571 \times 20$ relabelings	**0 mismatches**
Production quarantines reproduced	**77 / 77**
Quarantined graphs resolved by BLISS	**77 / 77**

Table 7: Canonicalization correctness audits. All comparisons concern canonical structural outcomes rather than arbitrary choices within automorphism orbits.

Quarantine route	GEOM graphs
Sign resolved $\rightarrow$ exact fallback, 512-leaf budget exhausted	54
Sign abstained $\rightarrow$ exact fallback, 512-leaf budget exhausted	20
Repeated eigenspace $\rightarrow$ exact fallback, 512-leaf budget exhausted	3
Total	**77**

Table 8: Production GEOM quarantines. All 77 are deterministic leaf-budget exhaustions; none is caused by the 30 s wall-clock safety valve.

252,363 are accepted after exact fallback, and 54 exhaust the fallback leaf budget and are quarantined. A further 4,145 simple-spectrum cases abstain: 4,125 are accepted after exact fallback and 20 are quarantined. The later sign tests remain part of the production cascade but resolve no additional simple-spectrum GEOM cases beyond the first test.

C.2 Relabeling and BLISS oracle audits

The production relabeling checks compare the packed canonical structural form after independent vertex renumberings. One audit uses 40 GEOM molecules under four presentations each (160 comparisons); a second built-dataset audit uses 512 graphs under three relabelings each (1,536 comparisons). The final canonicalizer has zero structural-canonicalization failures in these audits: 0/160 and 0/1,536, respectively.

We additionally cross-check against BLISS, an independent graph canonical-labeling tool used only as an oracle rather than as the production fallback.

The BLISS audit uses `python-igraph==0.11.6`. Bond orders are represented exactly through a colored incidence gadget in which each bond is replaced by a vertex whose color encodes the bond class. The audited population contains 10,400 stratified exact-fallback GEOM graphs, all 75 repeated-eigenspace graphs, all 77 production quarantines rebuilt from raw inputs, and 19 synthetic symmetry families, for 10,571 graphs total.

For each graph, 20 relabelings are tested. The protocol checks invariance of the production structural serialization, invariance of the BLISS certificate, agreement of the induced structural/isomorphism verdicts, and deterministic re-quarantine of bounded-search failures. It reports zero mismatches; all 77 quarantined graphs are independently resolved by BLISS.

C.3 Exact fallback behavior and certification coverage

Among the 256,560 accepted GEOM graphs that invoke the exact fallback, the number of explored leaves has median 4, 90th percentile 12, 99th percentile 48, and maximum 432. Exact-fallback time has median 15 ms, 90th percentile 65 ms, 99th percentile 230 ms, and maximum 4.33 s among accepted graphs. For 256,536 of 256,560 calls (99.99%), all explored leaves produce the same serialization. Only 24 accepted graphs produce multiple distinct leaf serializations, in which case the lexicographically smallest is selected.

Seventy-five of the 77 quarantined graphs have all explored leaves tied. The policy is deliberately conservative: a bounded search that does not finish is excluded instead of being accepted on a best-

Population	Graphs
All retained graphs	495,833
Primary spectral population	**474,144 (95.63%)**
Graphs with only singleton Fiedler tolerance cells	85,935 (17.33%)
Tied cells refined without exact fallback	1,360
Tied cells refined through exact fallback	386,849

Table 9: Population accounting for the corrected spectral-rounding analysis. The 17.33% subset is a stricter direct-sort subset, not the applicability rate of the analysis.

Diagnostic	Value
Median $\varepsilon_{\mathrm{prod}}$	0.2259
95th percentile $\varepsilon_{\mathrm{prod}}$	0.3844
Median $r_{\tau_{\mathrm{final}}}^{\top}\widehat{u}_2$	0.9745
Median δ_{round}	0
95th percentile δ_{round}	5.55×10^{-17}
$\lvert\delta_{\mathrm{round}}\rvert \le 10^{-12}$	99.9975%
Median sharp-bound utilization	0.8329
95th percentile sharp-bound utilization	0.9305
95th percentile positive excess energy	0.06518

Table 10: Spectral-rounding diagnostics over the 474,144-graph primary population. The inner-product median is the stored aggregate $0.974481\ldots$, not a value reconstructed from the rounded median error.

so-far order. The observed 0.025% quarantine rate is therefore a certification-coverage loss, not an observed correctness failure among accepted graphs.

Across accepted GEOM molecules, total production canonicalization time has median 35 ms, 90th percentile 95 ms, 99th percentile 250 ms, and maximum 4.38 s. These are production-pipeline measurements; hardware-normalized latency is not reported.

The preceding audits address canonicalization correctness and certification coverage. We next evaluate a separate question: on graphs for which the simple-spectrum Fiedler coordinate is defined and oriented, how closely does the final production order preserve direct spectral sorting?

C.4 Corrected spectral-rounding population

The corrected spectral analysis covers all retained graphs from GEOM, the eight MoleculeNet datasets, and OGBG-MolHIV, totaling 495,833 graphs. Its primary population contains connected graphs with $n \ge 2$, a simple production λ_2, and a resolved production sign. Graphs with tied Fiedler-coordinate cells remain in the primary population provided refinement does not reverse the ordering between distinct spectral cells. The singleton-only subset accounts for 17.33% of all retained graphs, whereas the corrected primary population covers 95.63%.

The fraction of retained graphs satisfying the primary spectral-regime criteria is shown for each source:

C.5 Excluded spectral cases

The 21,689 graphs outside the primary population are fully accounted for by the following mutually exclusive reasons:

C.6 Production refinement relative to direct Fiedler sorting

For each primary graph, let

$$\varepsilon_{\mathrm{prod}} = \|r_{\tau_{\mathrm{final}}} - \widehat{u}_2\|_2, \qquad \varepsilon_{\mathrm{sort}} = \|r_{\tau_{\mathrm{spec}}} - \widehat{u}_2\|_2, \tag{33}$$

Source	Primary (%)	Source	Primary (%)
GEOM	98.62	BBBP	94.97
BACE	99.60	ClinTox	90.45
HIV	86.71	MUV	98.70
SIDER	90.88	Tox21	79.66
ToxCast	78.22	OGBG-MolHIV	80.64

Table 11: Fraction of each retained source satisfying the corrected primary spectral-regime definition. These rates measure spectral-analysis scope rather than canonicalization success.

Exclusion reason	Graphs
Disconnected	3,006
Nonsimple second eigenspace	1,180
$n < 2$	42
Other production cases not evaluated by the C6 simple-spectrum diagnostic	30
Production sign unresolved	17,431
Total	**21,689**

Table 12: Cases excluded from the corrected spectral-rounding population. These are scope exclusions for the spectral diagnostic, not canonicalization failures. The finalized aggregate records the 30 rare non-C6 production cases as one grouped category rather than a finer breakdown.

where τ_{spec} is the stable direct sort of the oriented float64 Fiedler coordinates and τ_{final} is the final production permutation. Define

$$\delta_{\mathrm{round}} = \varepsilon^2_{\mathrm{prod}} - \varepsilon^2_{\mathrm{sort}}. \tag{34}$$

The median is exactly zero in the stored aggregate, the 95th percentile is 5.55×10^{-17}, and 99.9975% of primary graphs satisfy $|\delta_{\mathrm{round}}| \le 10^{-12}$. Thus deterministic refinement preserves the nearest-rank Fiedler-distance objective to numerical precision on essentially the entire primary population, although the final permutation may still change within tolerance-indistinguishable cells.

C.7 Production-order instantiation of the energy-gap bound

The main theorem is stated for direct Fiedler sorting. The finalized empirical artifact instantiates the same rank-vector inequality at the *final production order*, for which all quantities are stored explicitly. Define

$$B_D = \lambda_2(1-\gamma_D), \qquad B_R^{\mathrm{prod}} = (\lambda_n - \lambda_2)\left(\varepsilon^2_{\mathrm{prod}} - \frac{\varepsilon^4_{\mathrm{prod}}}{4}\right), \qquad B^{\mathrm{prod}}_{\mathrm{full}} = B_D + B_R^{\mathrm{prod}}. \tag{35}$$

By Corollary B.12,

$$E(\tau_{\mathrm{final}}) - E(\tau_*) \le B^{\mathrm{prod}}_{\mathrm{full}}. \tag{36}$$

Table 13 reports the stored graph-wise distribution over the 474,144-graph primary population. Ratios are formed per graph before aggregation. The merged result artifact stores the production-order energy $E(\tau_{\mathrm{final}})$, so the full-population normalization reported here is exactly $B^{\mathrm{prod}}_{\mathrm{full}}/E(\tau_{\mathrm{final}})$.

The 0.063246 full-population ratio is therefore measured with $E(\tau_{\mathrm{final}})$, the energy stored by the merged artifact. We independently recomputed $E(\tau_{\mathrm{spec}})$ on 6,913 graphs spanning all ten data sources: the recomputed production energy matched the stored value to 6×10^{-17}, while the median relative difference between spectral and production energies was 10^{-16} (95th percentile 3×10^{-4}; maximum 5.9%), with all differences confined to tolerance-tied cells. On this sample, the median bound-to-energy ratio is 0.06182 with either denominator to reported precision, and the 95th percentile changes from 0.17476 to 0.17477. Hence the 6.325% figure is the stored full-population production-energy normalization; the direct spectral-energy normalization is independently sample-validated, not stored over all 474,144 graphs.

Graph-wise quantity	Median	95th percentile
$B_{\text{full}}^{\text{prod}}$	0.03485	0.09333
$B_{\text{full}}^{\text{prod}}/E(\tau_{\text{final}})$	0.06325	0.16270
$B_D/B_{\text{full}}^{\text{prod}}$	0.13438	0.40771

Table 13: Empirical scale of the production-order discrete-optimum bound. The stored full-population median is 6.325% of the final production-order energy, and the degree-mismatch term contributes 13.44% of the bound at the median. Ratios are computed graph by graph before aggregation.

Ordering	Median gap	Median gap / bound	Exact optimum	Violations
Direct Fiedler τ_{spec}	3.44×10^{-4}	0.0057	46.7%	0 / 4,253
Production τ_{final}	2.103×10^{-3}	0.0321	32.0%	0 / 4,253

Table 14: Exhaustive discrete optimization on the same 4,253 eligible graphs with $n \leq 10$. The first row is the theorem-facing direct Fiedler ordering; the second is the final production ordering. Ratios use the corresponding graph-wise rank-vector bound before aggregation.

C.8 Exact discrete optimization on small graphs

For every eligible primary-regime graph with $n \leq 10$, the evaluation enumerates all $n!$ rank assignments and computes the exact discrete optimum $E(\tau_*)$, yielding 4,253 exactly solved graphs. The finalized v3 artifact stores the production-order gap. For the theorem-facing direct spectral order, we recompute $E(\tau_{\text{spec}})$ on the same 4,253 graphs by stable sorting of the oriented Fiedler coordinate and reuse the stored exhaustive $E(\tau_*)$; recomputed production energies agree with the stored values to within 1.1×10^{-16}.

Thus the theorem-facing direct-Fiedler values are 3.44×10^{-4} for the median realized gap, 0.0057 for the median realized-gap/bound ratio, and 46.7% exact optimality. The production-order values are reported separately because refinement inside tolerance-indistinguishable spectral cells can change the exact energy.

C.9 Production order versus direct Fiedler sorting

The theorem is stated for the direct spectral order τ_{spec}, whereas the production canonicalizer may refine vertices inside tolerance-indistinguishable Fiedler cells. By construction, refinement preserves the order between distinct spectral cells; the reported audit observes no reversal of a vertex pair whose oriented Fiedler coordinates lie in different production tolerance cells. Across the primary population, the median Kendall rank correlation between τ_{spec} and the final production order is

$$\operatorname{median}\big[\tau_b(\tau_{\text{spec}}, \tau_{\text{final}})\big] = 0.992095. \tag{37}$$

This statistic complements the nearest-rank diagnostic in Eq. (34): the median δ_{round} is zero, the 95th percentile is 5.55×10^{-17}, and 99.9975% of primary graphs satisfy $|\delta_{\text{round}}| \leq 10^{-12}$. Thus production refinement can change the vertex permutation inside numerically indistinguishable cells while leaving the Fiedler-distance objective unchanged to numerical precision on essentially the entire primary population.

C.10 Production-order excess above the continuous Fiedler level

For the final production rank vector, the separate inequality proved in Appendix B.6 gives

$$\Delta_E^+ \leq B_{\text{sharp}} := (\lambda_n - \lambda_2)\left(\varepsilon_{\text{prod}}^2 - \frac{\varepsilon_{\text{prod}}^4}{4}\right), \qquad \Delta_E^+ = \max\{E(\tau_{\text{final}}) - \lambda_2, 0\}. \tag{38}$$

The stored utilization ratio $\Delta_E^+/B_{\text{sharp}}$ has median 0.8329 and 95th percentile 0.9305; the 95th percentile of Δ_E^+ is 0.06518. This quantity should not be confused with the discrete-optimum gap $E(\tau_{\text{spec}}) - E(\tau_*)$: it measures how tightly the sharp Fiedler-level inequality describes the final production rank vector.

Quantity	Value
Discovered molecules	304,339
Canonicalization attempts	303,873
Accepted molecules	**303,796**
Disconnected exclusions	465
Empty exclusions	1
Canonicalization quarantines	77
Failed builds	0
Accepted atoms	7,553,370
Accepted molecular bonds	8,160,775
Mean atoms / molecule	24.86
Mean bonds / molecule	26.86

Table 15: GEOM-Drugs population used for the reported molecular self-supervised pretraining.

C.11 Scope of the operator choice

Permutation invariance is an exact same-graph guarantee; it is not a continuity guarantee under perturbations of the graph operator. The production bandwidth $\sigma = 0.35$, inverse-weight path lengths, dense normalized Laplacian, and single-resolution Fiedler construction were not independently ablated. We therefore do not attribute predictive performance to any one of these operator choices.

D Datasets, Splits, and Evaluation Populations

We describe the molecular populations used for pretraining and downstream evaluation, including identity handling, graph-level exclusions, split construction, and the VALID/TEST separation. GEOM pretraining and MoleculeNet are not guaranteed to be identity-disjoint; the available historical overlap audit is reported separately below.

D.1 GEOM pretraining population

Pretraining uses the GEOM-Drugs one-resolution corpus, which retains one conformer per molecule during data construction. The current molecular encoder, however, consumes only the resulting 2D molecular graph and atom features; 3D coordinates do not enter GINO. GEOM preprocessing discovers 304,339 molecules. The production builder accepts 303,796, excludes 465 disconnected molecules and one empty molecule, and quarantines 77 molecules whose canonical structural form cannot be certified within the bounded exact-search budget. The accepted corpus contains 7,553,370 atoms and 8,160,775 molecular bonds.

For each SSL seed, the accepted corpus is deterministically partitioned into 297,720 training molecules and 6,076 validation molecules. The partition is seed dependent: the corpus is permuted with the corresponding SSL seed before taking the 2% validation subset. These pretraining partitions are distinct from the downstream scaffold split seeds described below.

D.2 MoleculeNet identity handling and graph construction

We evaluate BBBP, BACE, ClinTox, HIV, MUV, SIDER, Tox21, and ToxCast. The current P1-v2 evaluation population is constructed as follows:

1. molecular identity is computed using the full InChIKey;
2. identity groups with conflicting label vectors are removed entirely;
3. agreeing duplicate rows are collapsed to a canonical representative;
4. non-isomeric RDKit Bemis–Murcko scaffolds are computed;
5. identity groups are assigned to TRAIN, VALID, and TEST by a seeded greedy scaffold split targeting approximately 80/10/10 proportions;
6. hydrogens are removed and molecular graphs/features are rebuilt through the centralized feature path compatible with RDKit and OGB; and

Dataset	Raw	Post-ID	Built	> 91	Disconn.	Quar.
BBBP	2,039	1,960	1,850	4	104	2
BACE	1,513	1,513	1,512	1	0	0
ClinTox	1,480	1,434	1,401	16	14	3
HIV	41,120	41,118	37,717	147	3,066	188
MUV	93,087	93,087	93,087	0	0	0
SIDER	1,427	1,427	1,206	56	165	0
Tox21	7,823	7,819	7,557	12	243	7
ToxCast	8,579	8,579	6,826	17	1,726	10

Table 16: Current P1-v2 MoleculeNet population accounting.

Dataset	Seed 42: TRAIN/VALID/TEST	Seed 43: TRAIN/VALID/TEST	Seed 44: TRAIN/VALID/TEST
BBBP	1,486/179/185	1,486/179/185	1,482/183/185
BACE	1,210/151/151	1,210/150/152	1,210/150/152
ClinTox	1,118/140/143	1,123/139/139	1,123/137/141
HIV	30,192/3,740/3,785	30,203/3,746/3,768	30,175/3,779/3,763
MUV	74,469/9,309/9,309	74,469/9,309/9,309	74,469/9,309/9,309
SIDER	964/120/122	968/113/125	983/113/110
Tox21	6,064/749/744	6,064/748/745	6,069/749/739
ToxCast	5,557/639/630	5,564/621/641	5,564/655/607

Table 17: Current P1-v2 scaffold-separated downstream splits for seeds 42–44.

7. graphs that are disconnected, exceed the builder size limit, or fail certified canonicalization are excluded.

An independent validator re-derives identity and scaffold assignments and finds zero identity crossings and zero scaffold crossings across all 24 manifests (8 datasets $\times$ 3 split seeds).

D.3 Historical GEOM–MoleculeNet overlap diagnostic

The following identity-exposure audit predates the P1-v2 rebuild and is reported only to document possible GEOM exposure. It was computed on the earlier P1-v1 seed-42 manifests; none of its percentages is used as a current P1-v2 split statistic.

Exposure varies substantially across benchmarks. We therefore do not treat raw GEOM-to-MoleculeNet transfer as an identity-disjoint generalization test, and we draw no overlap-stratified performance claims from this historical audit. This limitation does not affect the canonicalization or spectral-rounding analyses.

D.4 External-comparison scope

External-method comparability, including differences between the current P1-v2 evaluation populations and the published C-FREE protocol, is discussed with the corresponding results in Appendix G.4.

D.5 VALID/TEST discipline

The headline molecular results are VALIDATION aggregates. Frozen evaluation uses the full 3×3 grid of P1-v2 scaffold split seeds $\{42, 43, 44\}$ and SSL seeds $\{42, 43, 44\}$; full fine-tuning uses three paired split/SSL seeds 42, 43, and 44. TEST is untouched for the current results, and the fine-tuning pipeline does not construct a TEST loader.

A historical P2 TEST comparison was evaluated once after development and was not subsequently used for model selection, tuning, or current claims. It is retained only as provenance and is not pooled with the current VALID results.

Dataset	TRAIN overlap	VALID overlap	TEST overlap
BACE	2/1,210 (0.17%)	0/151	0/151
BBBP	412/1,486 (27.73%)	41/179 (22.91%)	44/185 (23.78%)
ClinTox	646/1,118 (57.78%)	86/140 (61.43%)	68/143 (47.55%)
HIV	874/30,192 (2.89%)	77/3,740 (2.06%)	83/3,785 (2.19%)
MUV	69,321/74,469 (93.09%)	8,579/9,309 (92.16%)	8,589/9,309 (92.27%)
SIDER	490/964 (50.83%)	54/120 (45.00%)	67/122 (54.92%)
Tox21	1,410/6,064 (23.25%)	275/749 (36.72%)	250/744 (33.60%)
ToxCast	1,309/5,557 (23.56%)	264/639 (41.31%)	243/630 (38.57%)

Table 18: Historical P1-v1 GEOM identity-exposure audit. The values document possible pretraining exposure but are not exact current P1-v2 overlap rates.

E Exact GINO Architecture and Canonical Views

We specify the molecular encoder and canonical student-view construction used by the current molecular results. The model uses one token per molecular bond, global Transformer self-attention, and canonical two-dimensional RoPE; sampled non-edge tokens are not used.

E.1 Edge-token construction and molecular features

For an undirected molecular graph $G = (V, E, X, F)$, where X collects the per-atom categorical features and F the per-bond structural values, GINO creates one token for each bond $e = \{u, v\} \in E$. Each endpoint is represented by nine OGB-style categorical atom fields: atomic number, chirality, degree, formal charge, number of hydrogens, number of radical electrons, hybridization, aromaticity, and ring membership. Their category cardinalities are: atomic number 119, chirality 5, degree 12, formal charge 12, number of hydrogens 10, number of radical electrons 6, hybridization 6, aromaticity 2, and ring membership 2. Each field has a trainable 128-dimensional embedding initialized from $\mathcal{N}(0, 0.02^2)$; the nine embeddings are summed to obtain $h_v \in \mathbb{R}^{128}$. Out-of-range category indices are clamped by the centralized categorical encoder.

Endpoint content is fused symmetrically:

$$s(h_u, h_v) = [h_u + h_v,\ |h_u - h_v|,\ h_u \odot h_v] \in \mathbb{R}^{384}. \tag{39}$$

For each bond, the selected stem additionally receives the structural bond-order scalar in $\{1/3, 1/2, 2/3, 1\}$ and a learned 64-dimensional edge-presence embedding. The shared tokenization interface retains this presence embedding; in the selected observed-bond-only configuration its indicator is always one, so it acts as a learned token-type bias rather than a varying edge attribute. The separate categorical OGB bond-type, stereo, and conjugation embedding is not used by the selected model.

The $1 + 384 + 64 = 449$ dimensional stem input is mapped to the Transformer width by

$$449 \xrightarrow{\text{Linear}} 256 \xrightarrow{\text{GELU}} 256 \xrightarrow{\text{Linear}} 512 \xrightarrow{\text{RMSNorm}} 512, \tag{40}$$

with RMSNorm $\epsilon = 10^{-6}$ and zero feature dropout.

E.2 Canonical pair coordinates

Let $c(v) \in \{0, \dots, n-1\}$ denote the canonical structural index. For an edge $e = \{u, v\}$, define

$$i_e = \min\{c(u), c(v)\}, \qquad j_e = \max\{c(u), c(v)\}.$$

The canonical edge coordinate records the lower endpoint and the offset to the upper endpoint:

$$\rho(e) = \big(i_e,\ j_e - i_e\big). \tag{41}$$

This representation is invertible: if $\rho(e) = (r, s)$, the ordered endpoint indices are $(r, r + s)$. The centered normalized rank vectors used in the theory are analysis objects and are not supplied to the encoder.

Configuration field	Selected value
Embedding dimension	512
Transformer blocks	12
Attention heads / head dimension	8 / 64
MLP ratio / hidden width	4.0 / 2,048
Transformer MLP	SwiGLU
Block normalization	pre-norm LayerNorm, $\epsilon = 10^{-5}$
QKV bias	no
Attention-output bias	yes
MLP linear biases	yes
Transformer dropout / attention dropout / drop path	0 / 0 / 0
CLS tokens / register tokens	1 / 4
RoPE base / reference nodes	10,000 / 64
Token set	molecular bonds only
Downstream encoder parameters	50,692,096

Table 19: Selected molecular GINO encoder. The parameter count includes the stem and Transformer backbone and excludes SSL projection heads.

E.3 Normalized axial 2D RoPE

The implementation supplies these same two coordinates to RoPE in offset-first order. Let

$$d_e = j_e - i_e, \qquad r_e = i_e, \tag{42}$$

where d_e is the endpoint offset and r_e the lower canonical index. Thus (d_e, r_e) contains the same pair information as $\rho(e) = (r_e, d_e)$; the ordering is only an implementation convention. For $n > 1$, both are scaled to a 64-node reference range:

$$\bar{d}_e = d_e \frac{63}{n-1}, \qquad \bar{r}_e = r_e \frac{63}{n-1}. \tag{43}$$

Thus Eq. (41) defines the graph's canonical pair geometry, while Eq. (43) gives the numerical coordinates supplied to RoPE.

The Transformer has head dimension 64 and rotary base 10,000. All 64 query/key dimensions participate in RoPE: 16 two-dimensional rotary pairs encode d_e, and the remaining 16 pairs encode r_e. Values are not rotated. The learned CLS token and four register tokens receive the identity rotary transform. Writing the rotary operator as $R(\bar{d}_e, \bar{r}_e)$,

$$q_e^{\mathrm{rot}} = R(\bar{d}_e, \bar{r}_e) W_Q h_e, \qquad k_e^{\mathrm{rot}} = R(\bar{d}_e, \bar{r}_e) W_K h_e.$$

The rotary composition law therefore makes the attention interaction depend on relative scaled canonical displacements rather than arbitrary input labels.

E.4 Transformer configuration

Each block is pre-norm:

$$x \leftarrow x + \mathrm{Attn}(\mathrm{LN}_1(x)), \qquad x \leftarrow x + \mathrm{MLP}(\mathrm{LN}_2(x)).$$

The MLP uses a 512-to-4096 linear map split into two 2,048-dimensional branches, a SiLU gate, and a 2,048-to-512 output map. A final LayerNorm is applied after the twelfth block. The selected backbone uses global variable-length attention; no graph-specific message-passing layer is inserted.

E.5 Graph-level readout

For downstream graph prediction, GINO uses the mean of the final post-norm edge states,

$$h_G = \frac{1}{|E|} \sum_{e \in E} h_e^{(L)}. \tag{44}$$

CLS and register states are excluded. Frozen probes standardize h_G using TRAIN-only statistics before fitting the classifier.

E.6 CANONICAL ANTI-DIAGONAL CROP

The student view uses a topology-changing crop followed by independent canonicalization. Viewing the canonical adjacency matrix in its strict upper triangle, entries with equal $i + j$ lie on one anti-diagonal. For an observed edge with teacher-canonical endpoint indices $i < j$, define

$$a(e) = i + j. \tag{45}$$

The valid anti-diagonal indices are $1, \ldots, 2n - 3$. Let

$$M = 2n - 3, \qquad \ell_{\text{crop}} = \max\left\{1, \left\lfloor 0.35M + \tfrac{1}{2} \right\rfloor\right\}. \tag{46}$$

For a valid start index $a_0 \in \{1, \ldots, M - \ell_{\text{crop}} + 1\}$, the crop removes

$$E_{\text{drop}} = \{e \in E : a_0 \le a(e) < a_0 + \ell_{\text{crop}}\}. \tag{47}$$

Thus the crop is one contiguous band of anti-diagonals in the teacher's canonical pair layout. All vertices are retained, and no unobserved pair is introduced as a token.

The surviving graph is reconstructed in raw vertex identifiers,

$$G_S = (V, E \setminus E_{\text{drop}}), \tag{48}$$

and the full-graph vertex order is discarded. The same production canonicalizer is then applied from scratch to G_S; student endpoint indices, token order, and RoPE coordinates are rebuilt from the resulting student permutation. Full-graph coordinates are used only to choose the crop. Because the surviving graph is canonicalized independently, its student coordinates need not agree with the teacher coordinates.

E.7 DETERMINISTIC CROP BANK AND EDGE CORRESPONDENCE

The current run precomputes $K = 16$ independently canonicalized student views for each of the 303,796 accepted GEOM molecules, yielding 4,860,736 stored views. The global crop-bank seed is 20260914. For graph identifier g, bank-view identifier v, and resampling attempt a, the candidate-view generator derives its pseudorandom seed from the SHA256 digest of

$$(\texttt{version}, 20260914, \texttt{dataset_id}, g, v, a), \tag{49}$$

and draws the valid start a_0 uniformly from $\{1, \ldots, M - \ell_{\text{crop}} + 1\}$. Thus every candidate crop is a pure function of the graph and recorded bank metadata rather than process-local RNG state.

At training time, the bank entry is selected by the project's stable `derive_seed` function applied to

$$(20260914, \texttt{graph_fingerprint}, \texttt{epoch}, \texttt{view_id}, \texttt{salt=canonical_antidiag_recanon_v2|view_choice}) \tag{50}$$

followed by a uniform integer draw from the available stored views. The structural view therefore changes across epochs while remaining exactly reproducible and independent of data-loader ordering.

A candidate is deterministically resampled if it is empty, duplicates an already stored survivor set while another distinct crop remains available, or cannot be certified by the production canonicalizer. Up to 64 attempts are permitted per view identifier, and bank construction continues until every accepted GEOM graph has exactly 16 stored views. Disconnected or isolated-node student graphs are allowed and are handled by the componentwise canonicalizer in Appendix A.9.

There is one exceptional GEOM graph with two vertices and one edge: every nontrivial anti-diagonal crop removes its only edge. For this graph the bank stores 16 independently canonicalized identity student views rather than copying teacher coordinates. This case is excluded from nontrivial crop augmentation.

The resulting bank confirms that recanonicalization changes the student coordinate system in practice: on average, 89.5% of surviving edges change canonical coordinates between teacher and student views (median 100%), and the mean L_1 displacement in unnormalized canonical coordinates is 15.8. These are implementation-audit statistics, not additional model inputs.

Every surviving bond carries an immutable teacher-edge identifier corresponding to its row in the teacher's lexicographically sorted real-edge list. The iBOT and Gram targets are gathered through this identifier, never through positional coordinate equality. All crop canonicalizations occur offline; the training loop performs zero runtime canonicalizer calls.

F PRETRAINING AND STATISTICAL PROTOCOL

We describe the current molecular self-supervised training and downstream evaluation protocol. The student receives two distinct perturbations: a structure-changing canonical crop, which changes the graph and its coordinate system, and an iBOT content mask, which changes token content without altering the already-cropped topology.

F.1 TEACHER AND STUDENT VIEWS

Each optimization step contains one full teacher view and one cropped student view. The teacher receives the complete molecular bond graph with its stored production canonicalization and is not content-masked. The student receives one of the independently recanonicalized views described in Appendix E.7. Consequently, a surviving bond generally has different teacher and student coordinates.

After selecting the structural crop, iBOT applies a fresh uniform content mask over the surviving student edge tokens. For a student with $T \geq 2$ edge tokens, the target number of masked tokens is

$$m(T) = \min\{\max(1, \lceil 0.35T \rceil), T - 1\}. \tag{51}$$

The mask itself is redrawn deterministically each epoch. Its seed is derived from the tuple

$$\begin{aligned}&(20260914,\ \texttt{graph_fingerprint},\ \texttt{epoch},\ \texttt{view_id},\\ &\ \texttt{salt=canonical_antidiag_recanon_v2|ibot_on_student})\end{aligned} \tag{52}$$

through the same stable `derive_seed` routine, then sampled over the surviving student-token set. Masked token embeddings are replaced by a learned mask vector; their student coordinates are unchanged. In the exceptional identity-view case, the parent mask draw is reused, matching the executed data path. Thus structural cropping and iBOT masking are complementary: the former removes complete edge tokens and induces a new canonical frame, while the latter hides content at surviving positions.

F.2 SELF-SUPERVISED OBJECTIVE

The graph-level DINO loss matches the cropped student's CLS prediction to the full teacher's CLS target. The edge-level iBOT loss is evaluated on masked surviving student bonds; corresponding teacher targets are gathered by immutable edge identity rather than by positional-coordinate equality. The complete objective used by the current molecular runs is

$$\mathcal{L} = \mathcal{L}_{\text{DINO}} + \mathcal{L}_{\text{iBOT}} + 0.05\,\mathcal{L}_{\text{KoLeo}} + 3.0\,\mathcal{L}_{\text{Gram}}. \tag{53}$$

KoLeo is applied to the normalized student global representation with $\epsilon = 10^{-6}$. The Gram anchor is computed over visible student edge-token representations. For graph g, let $Z_g^s, Z_g^t \in \mathbb{R}^{m_g \times d}$ denote the aligned visible student and frozen-teacher edge-token features, respectively, with teacher rows gathered by immutable edge identity. In float32, the implementation forms

$$H_g^a = \frac{(Z_g^a)^\top Z_g^a}{m_g}, \qquad \bar{H}_g^a = \frac{H_g^a}{\max(\|H_g^a\|_F, 10^{-8})}, \qquad a \in \{s, t\}, \tag{54}$$

then averages these normalized per-graph matrices, $\bar{H}^a = |\mathcal{G}|^{-1} \sum_{g \in \mathcal{G}} \bar{H}_g^a$, over graphs with at least one visible token. The Gram loss is

$$\mathcal{L}_{\text{Gram}} = \text{MSE}(\bar{H}^s, \bar{H}^t). \tag{55}$$

The teacher features come from a frozen full-teacher snapshot refreshed every 5,000 optimization steps. This graph-segmented construction never forms one Gram matrix across unrelated packed graphs.

Teacher parameters are updated by exponential moving average,

$$\theta_t \leftarrow m_t \theta_t + (1 - m_t)\theta_s, \tag{56}$$

with cosine-ramped momentum

$$m_t = 1 - (1 - 0.998)\frac{1 + \cos(\pi t / T)}{2}, \tag{57}$$

where $T = 348{,}900$ is the total number of optimization steps. Buffers are copied rather than exponentially averaged, and teacher parameters receive no gradients.

Hyperparameter	Selected value
Accepted GEOM molecules	303,796
SSL train / validation	297,720 / 6,076
Epochs	300
Batch size	256
Updates per epoch / total updates	1,163 / 348,900
Optimizer	AdamW
AdamW β / ϵ	$(0.9, 0.999)$ / 10^{-8}
Peak learning rate	4.362×10^{-4}
Minimum learning rate	10^{-6}
Warmup	5 epochs from 2×10^{-6}
Post-warmup schedule	cosine
Weight decay	0.02 on eligible matrix parameters; 0 otherwise
Gradient clipping	global norm 0.75
EMA base momentum	0.998, cosine-ramped to 1
Student / teacher temperature	0.10 / 0.04
KoLeo / Gram weight	0.05 / 3.0
Gram snapshot refresh	every 5,000 steps
Structural crop	35% anti-diagonal band, Eq. (46)
Precomputed student views / graph	16
Online iBOT mask ratio	0.35
Sampled non-edge ratio	0
Precision	CUDA bfloat16 autocast; no GradScaler
SSL seeds	42, 43, 44

Table 20: Current molecular pretraining recipe. The full-precision peak learning rate recorded in the executable configuration is $4.361852518229932 \times 10^{-4}$.

F.3 Projection heads and teacher targets

Both DINO and iBOT use 1,024 prototypes. The implemented projection heads are

$$\text{DINO: } 512 \to 4096 \to 4096 \to 256 \to 1024, \tag{58}$$

$$\text{iBOT: } 512 \to 4096 \to 256 \to 1024. \tag{59}$$

Hidden layers use GELU. The bottleneck is L2-normalized and the final prototype layer is weight-normalized and bias-free. Student and teacher temperatures are 0.10 and 0.04, respectively. Teacher targets are produced by three Sinkhorn–Knopp iterations in float32 with numerical $\epsilon = 10^{-12}$; no separate EMA center vector is maintained.

F.4 Current 300-epoch pretraining recipe

For each SSL seed, the accepted GEOM corpus is deterministically partitioned into 297,720 training and 6,076 validation molecules; these pretraining splits are distinct from the downstream scaffold splits in Appendix D. The epoch-300 EMA teacher is used downstream. No downstream metric is used for SSL checkpoint selection.

F.5 Frozen molecular evaluation

The primary downstream metric is ROC-AUC. Multi-task datasets are scored per task after masking missing labels and macro-averaged over task columns for which both classes are present. The encoder is fixed and the graph representation is the mean edge-token readout in Eq. (44). A `StandardScaler` is fit on TRAIN only. Each task is then fit with `LogisticRegression`($C = 1$, `max_iter`=2000).

The current frozen results use all nine combinations of scaffold split seed $\{42, 43, 44\}$ and SSL seed $\{42, 43, 44\}$ for each dataset. The reported dataset value is the mean of these nine VALID cells. For HIV only, frozen TRAIN is deterministically capped at 6,000 molecules with `np.random.default_rng(0)`. TEST is not evaluated.

Fine-tuning setting	Value
Trainable representation	complete encoder + supervised head
Encoder learning rate	3×10^{-5}
Head learning rate	10^{-3}
Optimizer	AdamW, PyTorch default β/ϵ
Weight decay	0.01
Batch size	32
Maximum epochs	50
Early stopping	patience 10 on VALID ROC-AUC
Gradient clipping	global norm 1.0
Prediction head	Dropout(0.1) $\rightarrow$ Linear(512, tasks)
Readout	mean final edge-token state
Loss	masked binary cross-entropy
Seeds	paired 42, 43, 44
Reported endpoint	best VALID ROC-AUC
TEST access	none

Table 21: Current full-fine-tuning protocol. The downstream hyperparameters were fixed before the current campaign using VALID performance on BBBP and Tox21; TEST was not used.

F.6 Full end-to-end fine-tuning

The current full-fine-tuning evaluation uses two seed-matched initialization arms: the epoch-300 EMA teacher after SSL and the exact pre-SSL student initialization from the corresponding run (step 0). The step-zero tensors are verified byte-for-byte against the recorded initialization, providing the matched control for the effect of SSL updates. For each dataset, seed $s \in \{42, 43, 44\}$ pairs SSL seed s with scaffold split seed s; the complete comparison therefore contains $8 \times 3 \times 2 = 48$ cells.

For multitask datasets, binary cross-entropy and ROC-AUC use only finite, measured labels; task AUCs are then averaged with equal weight. Full fine-tuning uses the complete HIV TRAIN split, without the frozen-probe cap. The test loader is never constructed. For each run, the prespecified early-stopping rule selects the checkpoint with the highest eligible VALID ROC-AUC; that score is reported. Because this is selection on the same validation split, the absolute value is a validation estimate rather than a held-out TEST estimate. The matched SSL-versus-step-0 comparison uses the same protocol in both arms.

F.7 Aggregation and uncertainty

Repeated cells are first averaged within each dataset. Overall means weight the eight datasets equally. For paired effects, the primary confidence interval is a dataset-clustered percentile bootstrap: the eight dataset-level effects are resampled 10,000 times with `np.random.default_rng(0)`, and the 2.5th and 97.5th percentiles form the reported 95% interval. Individual task columns and seed-level cells are not treated as independent benchmark replicates.

G Self-Supervised Molecular Transfer

We first report the current molecular results used in the main paper, then place them in context with the matched step-zero control and published C-FREE results. The historical sparse predecessor (internally S2/P2) is retained only as a diagnostic and is clearly separated from the current values.

G.1 Current molecular downstream results

The frozen results use the epoch-300 EMA teachers and the protocol in Appendix F.5. Each dataset averages the nine combinations of three P1-v2 scaffold split seeds and three SSL seeds. The fine-tuned results use the protocol in Appendix F.6 and average three paired split/SSL seeds. Both are VALIDATION aggregates; TEST is not evaluated for these current results.

The unrounded overall means are 0.694273 for frozen transfer and 0.775479 for full fine-tuning.

Setting	BBBP	BACE	ClinTox	HIV	MUV	SIDER	Tox21	ToxCast	Avg.
Frozen	80.39	76.77	79.94	65.07	68.70	56.24	69.53	58.76	69.43
Fine-tuned	87.87	85.57	91.97	76.35	77.87	59.52	74.95	66.27	77.55

Table 22: Current GINO VALID ROC-AUC (%) on the deduplicated, scaffold-separated P1-v2 MoleculeNet populations. Frozen values average three split seeds × three SSL seeds; fine-tuned values average three paired split/SSL seeds.

Dataset	SSL init	Step-0 init	Δ
BBBP	0.8787	0.8618	+0.0169
BACE	0.8557	0.7904	+0.0653
ClinTox	0.9197	0.7822	+0.1376
HIV	0.7635	0.7532	+0.0103
MUV	0.7787	0.7586	+0.0201
SIDER	0.5952	0.5970	-0.0018
Tox21	0.7495	0.7717	-0.0222
ToxCast	0.6627	0.6653	-0.0026
Mean	0.7755	0.7475	+0.0280

Table 23: Current full-fine-tuning SSL-versus-step-zero diagnostic on VALID. The mean effect is $+0.02796$ ROC-AUC; the dataset-clustered 95% confidence interval is $[-0.00009, +0.06384]$.

G.2 Matched full-fine-tuning step-zero control

The strongest available pretraining control compares each epoch-300 SSL teacher with the byte-identical initialization of the same training run before any SSL updates. Both arms use the same downstream data, seed, architecture, and fine-tuning procedure.

The point estimate is positive on five of eight datasets and on all three seed-level aggregate comparisons, but the primary dataset-clustered 95% confidence interval narrowly includes zero. The experiment therefore does not resolve a nonzero average initialization benefit across datasets, despite the positive point estimate. A matched current frozen step-zero probe was not run, so no such contrast is reported.

G.3 Historical S2 diagnostics

The historical sparse predecessor, internally referred to as S2/P2, used the same underlying sparse architecture but predated the current topology-changing crop. The name inherits the selected S2 observed-bond arm from the earlier model-design study in Table 29; it is a campaign-lineage label rather than a second, unrelated S2 method. Its frozen VALID mean was 0.7082 versus 0.7380 for its historical random-initialization control ($\Delta = -0.0298$, 95% CI $[-0.0434, -0.0141]$). Historical end-to-end fine-tuning yielded 0.7719 versus 0.7524 ($\Delta = +0.0195$, 95% CI $[+0.0024, +0.0417]$). These results are retained only as diagnostics of the earlier lineage and are not pooled with the current molecular results.

G.4 Extended C-FREE context

C-FREE (Ariguib et al., 2026) provides a recent GEOM-to-MoleculeNet reference point. GINO uses only the 2D molecular graph. C-FREE reports 2D graph/subgraph variants as well as multimodal variants using 3D conformers and dedicated 3D encoders. Moreover, the downstream preprocessing, split construction, and reported evaluation populations differ. The values below are therefore shown for context, not as a controlled cross-method ranking.

H Positional and Ordering Controls

These secondary controls test canonical addressing, pair RoPE, and the choice of spectral order. They are separate from the current molecular-transfer experiment in Appendix G; where a control

Method	Input	BBBP	BACE	ClinTox	HIV	MUV	SIDER	Tox21	ToxCast	Avg.
Frozen encoder — 2D only										
GINO (ours)	2D graph	80.39	76.77	79.94	65.07	68.70	56.24	69.53	58.76	69.43
C-FREE$_{\text{2D-MOL}}$	2D graph	60.5	75.8	62.7	68.7	67.6	59.0	76.1	62.7	66.63
C-FREE$_{\text{2D-SUB}}$	2D subgraph	64.2	73.9	71.4	65.5	64.6	58.0	76.7	63.9	67.27
Frozen encoder — 2D+3D										
C-FREE$_{\text{MM-MOL}}$	2D+3D graph	69.8	75.3	69.9	72.8	76.6	58.5	79.9	65.8	71.07
C-FREE$_{\text{MM-SUB}}$	2D+3D subgraph	73.8	75.5	75.7	71.9	70.6	56.4	76.7	66.8	70.92
Fine-tuned encoder — GINO 2D / C-FREE 2D+3D										
GINO (ours)	2D graph	87.87	85.57	91.97	76.35	77.87	59.52	74.95	66.27	77.55
C-FREE$_{\text{Sch-1C}}$	2D+3D (1 conformer)	88.6	78.8	73.1	78.9	75.9	61.8	84.7	71.3	76.63
C-FREE$_{\text{Sch-3C}}$	2D+3D (3 conformers)	78.9	78.6	83.7	77.9	82.5	62.5	84.2	71.7	77.50
C-FREE$_{\text{PaiNN-3C}}$	2D+3D (3 conformers)	88.9	82.3	87.6	81.4	75.7	64.8	86.2	71.6	79.81

Table 24: Extended molecular-transfer context. GINO values are the current VALID aggregates from this work. All C-FREE values are reproduced from the published C-FREE tables. Because preprocessing, downstream splits, evaluation populations, and input modalities differ, no controlled cross-method ranking is implied.

Contrast	Mean Δ ROC-AUC	95% CI
Per-epoch relabeling − canonical+RoPE	+0.0018	$[-0.0100, +0.0119]$
Per-epoch relabeling − fixed arbitrary	+0.0107	$[-0.0000, +0.0186]$
Canonical+RoPE − fixed arbitrary	+0.0089	$[+0.0015, +0.0189]$
Canonical+RoPE − canonical, no RoPE	-0.0014	$[-0.0150, +0.0152]$

Table 25: Corrected presentation/augmentation controls for the 512×12 molecular model. Confidence intervals are dataset-clustered.

belongs to an earlier valid lineage, it is reported as a historical diagnostic rather than mixed with the current headline results.

H.1 Canonical presentation versus true relabeling augmentation

An early random-relabel augmentation implementation accidentally kept the epoch fixed at zero and therefore reused one static presentation. That arm is invalid as an augmentation experiment and is excluded. The corrected experiment propagates epoch state and redraws the relabeling each epoch. Across eight MoleculeNet datasets and three scaffold split seeds, all remaining model and training factors are held fixed. The values below are VALIDATION aggregates.

The corresponding VALID means are 0.7557 for per-epoch relabeling, 0.7552 for canonical coordinates without RoPE, 0.7538 for canonical+RoPE, and 0.7450 for one fixed arbitrary presentation. Thus a single arbitrary presentation is weaker in this experiment. At the statistical resolution of this study, however, the data do not resolve a difference between deterministic canonicalization and true per-epoch relabeling augmentation. The isolated canonical+RoPE versus canonical-no-RoPE contrast likewise does not establish a measurable molecular-prediction gain from RoPE alone.

H.2 WIRE, canonical coordinates, random coordinates, and no position

A separate GraphGPS-style control compares four positional constructions after a fixed random relabeling of every graph. WIRE (Wavelet-Induced Rotary Encodings) (Reid et al., 2025) is the external graph-RoPE construction used in this comparison. MolHIV uses the official OGB split and official OGB evaluator, with eight paired seeds and TEST ROC-AUC as the reported metric. Within a seed, all MolHIV arms share an A40 host and bit-identical backbone initialization; backup reruns are not pooled with the primary comparison. PATTERN uses three paired seeds.

On PATTERN, the four TEST means are 0.86813 (WIRE), 0.86777 (canonical), 0.86737 (random), and 0.86772 (no position); the maximum difference is 0.00076. Across CPU hosts, recomputed WIRE coordinates differ by an equivalent sign or basis choice on approximately 35% of retained

MolHIV arm	Mean TEST AUC	SD	SE
No positional encoding	0.7585	0.0129	0.0046
Random-index RoPE	0.7626	0.0085	0.0030
Canonical-index RoPE	0.7659	0.0140	0.0050
WIRE	0.7675	0.0257	0.0091

Table 26: Paired positional-coordinate comparison on MolHIV.

Paired contrast	Mean Δ	95% bootstrap CI
Canonical − random	+0.0033	$[-0.0056, +0.0142]$
Canonical − no position	+0.0074	$[-0.0070, +0.0224]$
Canonical − WIRE	-0.0015	$[-0.0174, +0.0158]$
WIRE − no position	+0.0089	$[-0.0150, +0.0290]$

Table 27: Primary paired MolHIV contrasts. None of the reported intervals resolves a positional-method ranking.

MolHIV graphs, while aligned vectors or subspaces remain numerically equivalent. This is a reproducibility property of the coordinate representation, not evidence of a predictive difference.

H.3 Spectral order versus exact canonical order

A MolHIV VALID diagnostic compares the production spectral order with an exact M3 order. Starting from weights pretrained with the spectral coordinate system, we fine-tune three variants: spectral+RoPE, spectral without RoPE, and exact-M3+RoPE. Their mean AUCs are 0.7658, 0.7681, and 0.7318 over three seeds. The corresponding random-initialization spectral and exact-M3 means are 0.7433 and 0.7481. Pair RoPE is again unresolved, while the large pretrained spectral versus exact-M3 difference is absent under random initialization.

One possible explanation is that replacing the spectral coordinate system by an exact M3 order at fine-tuning creates a coordinate-system mismatch with pretraining. This diagnostic does not establish that mechanism causally, nor that M3 is intrinsically worse for prediction.

H.4 Ordering locality diagnostic

A separate diagnostic spans 14 graph sources (ten molecular and four non-molecular). Across molecular datasets, the Spearman correlation between canonical index separation and graph-hop separation lies between 0.58 and 0.92 for spectral order and between approximately -0.06 and $+0.07$ for exact M3. Moreover, 68–82% of molecular bonds under spectral ordering satisfy $|i-j| \leq 2$. These measurements establish that the spectral branch is strongly local in index space. In contrast, exact M3 exhibits near-zero correlation between index separation and graph-hop separation under these locality statistics. They do not establish that locality itself causes predictive gains.

I Model-Design Ablations

These studies document design decisions that preceded the current molecular configuration. They are reported as controlled or historical diagnostics and are not used to strengthen the canonicalization theorem or to infer causal mechanisms beyond the comparisons actually performed.

I.1 Graph readout

Using checkpoints from the historical sparse predecessor (internally S2/P2), we evaluated six frozen graph-level readouts without retraining the encoder.

The mean molecular-bond readout improves over CLS by approximately 0.033 ROC-AUC in this historical screen. This readout is defined identically whether or not the token set also contains sampled non-edges: it always averages only over molecular-bond tokens.

Readout	Mean Δ VALID ROC-AUC vs. CLS
Mean all pair tokens	+0.0314
Mean molecular-bond tokens	+0.0330
Max molecular-bond tokens	-0.0017
Min molecular-bond tokens	-0.0034
CLS + mean concatenation	+0.0185
Mean + max concatenation	+0.0148

Table 28: Historical frozen readout screen. Mean molecular-bond pooling gives the largest average improvement over CLS and is the readout used by the current molecular encoder.

Arm	Non-edge ratio	Non-edge chemistry	Mean VALID	Δ vs. S0	95% CI
S0	1.0	full	0.6828	–	–
S1	1.0	node chemistry zeroed	0.7134	+0.0305	$[+0.0110, +0.0526]$
S2	0.0	n/a	0.7148	+0.0320	$[+0.0145, +0.0497]$
S3	0.5	full	0.7057	+0.0228	$[+0.0075, +0.0421]$

Table 29: Historical sampled-non-edge ablation. The selected sparse architecture uses real molecular bonds only (S2).

I.2 Observed bonds versus sampled non-edges

A controlled four-arm study varies the sampled non-edge ratio and the treatment of chemistry on sampled non-edges while keeping the sparse architecture fixed.

When rescored on the later deduplicated, scaffold-separated downstream population, S2 remains positive relative to S0 by +0.0178 mean ROC-AUC (7/8 datasets; 95% CI $[+0.0050, +0.0327]$).

I.3 Endpoint chemistry and categorical bond features

With observed-bond tokenization fixed, adding symmetric endpoint atom chemistry to structural tokens yields mean $\Delta = +0.0536$, positive on all eight datasets, with dataset-clustered 95% CI $[+0.0280, +0.0991]$.

Adding a separate categorical OGB bond embedding on top of endpoint chemistry yields mean $\Delta = -0.0077$, with 95% CI $[-0.0311, +0.0099]$ and positive effects on 5/8 datasets. The supported conclusion is that this additional bond embedding has no resolved incremental benefit at the tested scale. The current stem still includes the structural bond-order scalar specified in Appendix E.

I.4 Masking choice

The current configuration uses uniform random iBOT content masking over the student edge tokens with target ratio 0.35, as defined in Eq. (51). This content mask is separate from the structural anti-diagonal crop. No clean final ablation isolates uniform random masking against alternative masking modes under the current Family-C recipe; we therefore make no claim that the selected masking mode is superior to those alternatives.

I.5 Sparse versus dense system reference

The historical dense design represents the complete strict upper triangle, whereas the selected sparse design represents molecular bonds only. On the P2 VALID grid, the tuned sparse system exceeds the tuned dense reference by +0.0790 mean ROC-AUC. In the single historical TEST comparison, sparse exceeds dense by +0.0654, positive on all eight datasets, with dataset-clustered 95% CI $[+0.0404, +0.0943]$.

This is a system-level comparison rather than a causal estimate of sparsity: the dense and sparse recipes were tuned separately, so the observed gap does not isolate sparsity as its cause.

Metric	Sparse	Dense reference	Relative value
Encoder parameters	50.69M	50.42M	$1.005\times$
Mean bond/pair tokens per MoleculeNet graph	25.7	302.2	$11.75\times$ fewer
Mean bond/pair tokens per GEOM graph	26.9	312.7	$11.64\times$ fewer
Historical minutes per pretraining epoch	11.1	36.6	$3.30\times$ faster
Historical GPU-h per 300-epoch seed	≈ 55.4	≈ 183.2	$3.31\times$ lower
Peak inference memory, A40 batch 64	0.39 GiB	1.41 GiB	$3.59\times$ lower
Probe-path inference throughput	269 mol/s	550 mol/s	sparse $0.49\times$

Table 30: Sparse observed-bond representation versus the historical dense full-triangle reference. Token count, training time, memory, and throughput are reported separately because their scaling is not interchangeable.

J Efficiency, Scaling, and Compute

Canonicalization-time distributions are reported with the production audit in Appendix C.3. Here we focus on asymptotic scaling, neural representation size, accelerator measurements, and the compute used by the current molecular SSL runs.

J.1 Canonicalization scaling and amortization

For n vertices, the current implementation forms weighted all-pairs shortest paths, a dense $n \times n$ affinity matrix, and a dense eigendecomposition. Its dominant asymptotic cost is therefore cubic in n. The present implementation is designed for the small-graph molecular regime studied in this paper and is not intended for large-scale graphs.

For a fixed topology, production canonicalization is performed once and cached. Its cost is therefore amortized across training epochs. Family-C student canonicalizations are likewise performed offline when constructing the 16-view crop bank; training itself performs no runtime canonicalizer calls.

J.2 Current Family-C pretraining compute

Each of the three current molecular SSL seeds uses one NVIDIA A40 GPU. All runs complete 300 epochs and 348,900 optimization steps. Their measured wall-clock runtimes are 57.9, 65.8, and 65.9 hours for seeds 42, 43, and 44, respectively; with one GPU per run these are also the corresponding GPU-hour totals. The runs use eight data-loading workers, 64 GB requested host memory, and bfloat16 CUDA autocast. Crop-bank generation is an offline CPU-side preprocessing stage; the final audit does not contain a single hardware-normalized aggregate for its construction time or storage cost, so those quantities are not reported here.

J.3 Sparse tokenization versus the historical dense reference

The following measurements compare the selected observed-bond architecture with the historical complete-upper-triangle representation. Training-time entries in this table come from the predecessor sparse S2 campaign and quantify the underlying tokenization/backbone rather than the additional current crop-bank preprocessing.

Although sparse tokenization substantially reduces sequence length and peak memory, the measured sparse probe path is slower in molecules per second on these small graphs. Python tokenization/collation and small-kernel overhead contribute materially to this measurement. Token reduction should therefore not be interpreted as a universal wall-clock inference speedup.

Dataset	Metric	SSL frozen	Step-0 frozen	Δ
MNIST	Accuracy	0.5476	0.5029	+0.0448
PROTEINS	Accuracy	0.7339	0.7393	-0.0054
IMDB-BINARY	Accuracy	0.6507	0.6147	+0.0360
COLORS-3	Accuracy	0.2776	0.2799	-0.0024

Table 31: Frozen cross-domain diagnostic for the corrected recanonicalized-crop study.

Dataset	Metric	SSL fine-tune	Step-0 fine-tune	Δ
MNIST	Accuracy	0.9315	0.9474	-0.0159
PROTEINS	Accuracy	0.7327	0.7303	+0.0024
IMDB-BINARY	Accuracy	0.6587	0.6713	-0.0127
COLORS-3	Accuracy	0.3157	0.3521	-0.0364

Table 32: Fine-tuned cross-domain diagnostic for the corrected recanonicalized-crop study.

K Cross-Domain Portability Diagnostics

These secondary experiments test whether the canonical observed-edge interface and recanonicalized crop can be instantiated outside molecular graphs. They are not used to claim universal positive transfer.

K.1 Small-scale corrected recanonicalized-crop study

A completed small-scale study evaluates MNIST superpixels, PROTEINS, IMDB-BINARY, and COLORS-3 with a 256-dimensional, three-block encoder. The structural augmentation follows the same corrected principle used by the current molecular system: crop in the teacher canonical frame, remove edges, and independently canonicalize the student graph before recomputing its coordinates. The tables report classification accuracy, averaged over five target-label budgets and three downstream seeds. These budget-averaged values summarize the study rather than a single operating point and should not be compared across datasets as if they shared the same task difficulty.

The effects vary in both sign and magnitude across domains and evaluation regimes. The experiment therefore supports portability of the implementation, but not a general positive-transfer claim.

K.2 Additional completed transfer diagnostics

A separate heterogeneous-source predecessor (internally Family B) evaluated transfer to MoleculeNet under full fine-tuning and yielded mean SSL/step-zero scores of 0.779697/0.782079, a difference of -0.002382 with 95% CI $[-0.015724, +0.012418]$. In the corresponding budget-matched PROTEINS study, the mean SSL-minus-step-zero effect is -0.022876; at target-label budgets 0.01, 0.1, and 1.0, the respective effects are -0.016340, -0.035948, and -0.016340. These completed diagnostics likewise provide no evidence of uniformly positive cross-domain transfer.

K.3 Excluded portability branches

Invalid and incomplete portability experiments are excluded from the results above; their failure modes and status are documented centrally in Appendix L.

L Reproducibility, Result Lineage, and Claim Boundaries

We use four status terms throughout the supplement. *Current* denotes the configuration or result used by the main paper. A *historical diagnostic* is a valid earlier experiment retained for context or mechanism analysis. *Superseded* denotes a result replaced by a corrected protocol, and *invalid—excluded from claims* denotes an experiment with a known defect that prevents scientific interpretation. Commit hashes and machine-specific paths are omitted from the supplementary material to preserve double-blind review.

Track	Lineage	Current role
Canonicalizer	legacy float32/index-sensitive path → float64 spectral/refinement/exact v2	v2 is the current production canonicalizer; legacy outputs are historical only
Downstream data	P1-v1 manifests → deduplicated P1-v2 manifests with no within-dataset identity or scaffold crossings	P1-v2 is used for current molecular evaluation; P1-v1 survives only in the historical overlap audit
Molecular SSL	historical sparse predecessor (internally S2/P2) → corrected recanonicalized-crop Family C	Family C supplies the current molecular headline results; the S2/P2 predecessor is a historical diagnostic
Canonical crop	full-position-leak crop v1 → independently recanonicalized crop v2	v1 is invalid; recanonicalized v2 is the current crop mechanism
Spectral audit	singleton-only v1 population → corrected tied-cell population	corrected 95.63% population is current; the 17.33% singleton subset is superseded as an applicability estimate

Table 33: Compact lineage of the principal components and experiments used in this paper.

L.1 Result lineage at a glance

The current molecular system therefore combines the production canonicalizer, P1-v2 downstream manifests, and Family-C SSL with independently recanonicalized anti-diagonal student crops. SSL seeds are 42, 43, and 44. Current frozen results use all nine split/SSL-seed combinations per dataset; current full-fine-tuning results use three paired split/SSL seeds. Both are VALIDATION aggregates.

For canonicalization nomenclature, `C0` denotes the legacy sign rule, `C1`–`C8` the corrected sign-cascade candidates from the sign study, and `C6` the selected production cascade. The independent repeated-eigenspace study used labels `M0`–`M3`; production uses `M2` projector refinement and `M3` bounded exact fallback. These identifiers describe experiment lineage rather than an ordered C1→C6 or M1→M3 execution pipeline.

L.2 Reproducibility controls

The independent canonicalization oracle uses `python-igraph==0.11.6`. Dependency versions not enumerated here are fixed by the executable project environment and its package lockfile; this appendix does not reconstruct versions absent from the validated project artifacts.

The principal positional-coordinate comparison keeps all MolHIV arms for a paired seed on the same A40 host with bit-identical backbone initialization; hardware-noise reruns are not pooled with the primary paired results.

The current TEST-access policy is specified in Appendix D.5. No additional TEST result enters the current Family-C claims.

L.3 Superseded and invalid experiments

L.4 Incomplete work

Incomplete experiments are excluded from all reported claims and aggregates. The scientifically relevant omissions are: no matched current frozen step-zero probe; no matched-hyperparameter causal dense-versus-sparse study; and several cross-domain follow-ups without a finalized complete evaluation. These absences are reflected directly in the narrower interpretations used in the corresponding sections.

L.5 Claim boundaries

The validated record supports the following statements:

Experiment	Issue	Correction	Status
Legacy canonicalizer	float32 and index-sensitive sign/tie behavior	float64 sign cascade, projector refinement, bounded exact fallback	Superseded
Early sparse sweep	stale epoch state, biased non-edge sampling, and masking-contract defects	corrected data loader, sampler, and regression tests	Invalid—excluded
Pre-fix feature experiments	molecular categorical-feature pipeline mismatch	centralized RDKit/OGB-compatible feature rebuild	Superseded
Early downstream manifests	molecular-identity leakage and unstable tie handling	dedup-first P1-v2 scaffold manifests with no within-dataset identity or scaffold crossings	Superseded
Static random-relabel arm	epoch remained zero, producing one fixed presentation	corrected true per-epoch relabeling experiment	Invalid—excluded
Anti-diagonal crop v1	student retained full-teacher positions after topology changed	independent student recanonicalization and fresh coordinates	Invalid—excluded
Spectral-rounding v1	singleton sorts mistaken for theorem applicability	tied-cell-preserving population logic	Superseded
Historical sparse S2/P2 molecular SSL	valid predecessor without the current structural crop	current Family-C molecular SSL experiment	Historical diagnostic

Table 34: Major experiment-lineage corrections. Invalid and superseded runs are not pooled with current results.

- certified outputs of the production canonicalizer define an invariant canonicalized structural matrix, with zero observed failures in the reported relabeling audits and independent BLISS cross-checks;
- the corrected spectral-rounding regime covers 95.63% of the evaluated graph population, and the final production rank remains very close to direct Fiedler sorting under the reported rounding metric;
- the current molecular system implements topology-changing anti-diagonal cropping, independent student recanonicalization, and DINO/iBOT-style self-supervision on molecular-bond tokens; and
- the 2D-only current system obtains the VALID molecular-transfer values reported in Appendix G and the main paper.

The validated record does *not* establish that:

1. canonicalization outperforms true per-epoch relabeling augmentation;
2. pair RoPE alone improves molecular prediction;
3. WIRE, canonical, random, and no-position coordinates have a resolved predictive ranking;
4. spectral locality itself causes predictive gains;
5. sparsity alone causes the sparse–dense system-level performance gap; or
6. self-supervised pretraining improves every downstream dataset or every evaluation regime.