Neutral-atom quantum features as complementary structural encodings for graph learning

Mathieu Garrigues *†
mathieu.garrigues@pasqal.com

Igor O. Sokolov *†
i.sokolov@pasqal.com

Matthias Hecker †
matthias.hecker@pasqal.com

Lorenzo Moro †
lorenzo.moro@pasqal.com

Shaheen Acheche †
shaheen.acheche@pasqal.com

Abstract

Structural descriptors provide graph neural networks (GNNs) with positional and topological inductive biases. However, widely used classical encodings based on random walks or Laplacian spectra emphasize specific structural statistics of a graph, leaving room for alternative descriptors. Neutral-atom quantum processing units (QPUs) can realize graph-dependent Rydberg Hamiltonians, allowing graph topology to be probed through dynamical correlation measurements. In this work, we use these measurements as *quantum graph features* (QGFs), providing a complementary, hardware-derived family of structural encoding that — as we demonstrate empirically — is not accurately described by classical features derived from Laplacian and adjacency matrices, even when supplemented with an extended set of random-walk, geometric descriptors and classical physical models. By appropriately selecting the quench parameters of the Hamiltonian, we access two characteristic correlation regimes with distinct structural signatures: antiferromagnetic-like and ferromagnetic-like. On an intrinsic information benchmark of 15 topological and spectral probing tasks, quantum and hybrid quantum-classical features match or surpass classical descriptors on local topological properties.

When supplied to message-passing, attention, and graph-transformer backbones, hybrid quantum-classical features yield small improvements that are within cross-fold noise on most cells but sign-consistent across architectures. When comparing noiseless, shot-sampled emulation-derived quantum features with an end-to-end neutral-atom QPU implementation, the downstream hybrid results with compressed hardware features match their emulated counterparts within 0.01 F_1 on all six backbones. On a synthetic family of ladder graphs, QGFs distinguish two specific classes, outperforming the strongest classical structural descriptors by a margin exceeding cross-fold variability.

1 Introduction

Graph neural networks (GNNs) and graph transformers are the default tools for learning on relational data such as molecules, materials, and biological networks [1–4]. Their performance depends strongly on how graph structure is exposed to the model. Standard choices include shortest-path

*Equal contribution.

†All authors are with Pasqal, 24 avenue Émile Baudot, 91120 Palaiseau, France.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

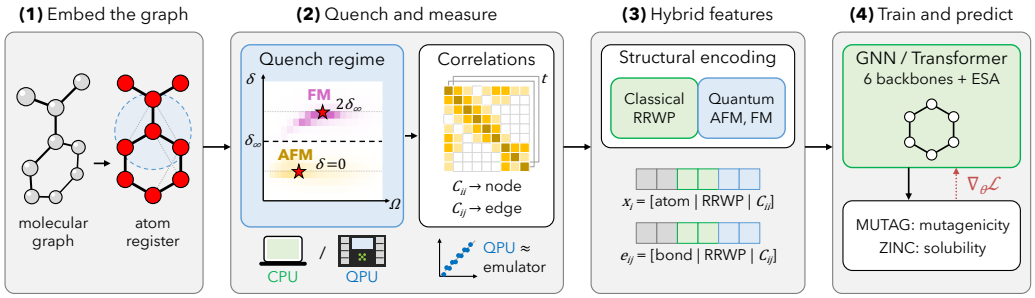


Figure 1: **Pipeline.** (1) Embed the graph in an atom register; (2) Run a quench in the antiferromagnetic (AFM) or ferromagnetic (FM) regime and measure the correlations $C_{i,j}$; (3) Diagonal/off-diagonal elements $C_{i,j}$ serve as node/edge features and are combined with classical RRWP; (4) Train GNNs and ESA on the combined encodings to predict molecular properties.

distances, Laplacian eigenvectors, heat kernels, and random-walk profiles such as relative random walk probabilities (RRWP) or random walk structural encodings (RWSE) or even application-specific like ring-membership, which are concatenated to node or edge features or injected into attention mechanisms [5–8]. These classical descriptors provide useful inductive biases and are efficiently computable, nonetheless, they encode specific structural statistics only. They inherit known limitations: for example, even expressive relative encodings such as RRWP fail on certain adversarial graph families, including non-isomorphic strongly regular graphs [5, 9–11]. This motivates the search for complementary graph descriptors. Physics-inspired architectures, which cast message passing as diffusion [12], magnetic-Laplacian propagation [13], or coupled oscillators [14], show that dynamics on graphs provide useful inductive biases; here, the dynamics is instead executed on a physical quantum system. Programmable neutral-atom quantum processors offer a concrete route towards such graph descriptors. In such a device, atoms can be arranged in user-defined registers. Atomic energy levels are used as qubit states which are time-evolved under a particular many-body Hamiltonian [15, 16].

Embedding a graph into such a register imprints its topology onto the quantum state, which can be extracted via observable measurement. Prior work has exploited this principle to define quantum evolution kernels and quantum positional encodings for graph learning [15–17, 10]. In particular, Thabet et al. [10] showed that quantum correlation-based encodings can be strictly more expressive than RRWP on selected graph families, while recent kernel studies have used neutral-atom observables to construct graph kernels competitive with classical baselines [18]. These results suggest that QPU-derived observables may capture information not contained in standard random-walk and spectral encodings.

In this work, using quantum correlations as *quantum graph features* (QGFs) we address three open questions. First, what intrinsic topological and spectral information do QPU-derived correlations carry, independently of any downstream neural architecture? Second, do they consistently improve GNNs across various architectures, rather than a single favorable backbone? Third, do they remain useful when produced by actual hardware involving noise sources from shot noise, over decoherence, to state-preparation-and-measurement (SPAM) errors? We focus on MUTAG and ZINC [19, 20], two standard molecular graph benchmarks spanning classification and regression, whose small molecular graphs largely admit two-dimensional unit-disk embeddings compatible with neutral-atom registers. In our datasets this covers 100% of MUTAG and 95% of $\mathrm{ZINC}_{\mathrm{subset}}$.

Our contributions. First, using two physically motivated operating regimes of the graph-encoded Rydberg Hamiltonian, antiferromagnetic (AFM) and ferromagnetic (FM) [Sec. 2.3], we introduce an intrinsic benchmark of 15 topological and spectral prediction tasks evaluated with permutation-invariant probes, showing that hybrid quantum–classical descriptors match or surpass purely classical baselines based on RRWP, RWSE and Laplacian eigenvectors, with AFM and FM correlations emphasizing different structural aspects of the graph [Sec. 3.1]. Second, regression analyses indicate that QGFs are not accurately recoverable from the classical descriptors considered here [Sec. 3.2] , even after adding ring, distance, register-geometry and restricted physical descriptors. Third, we show that concatenating QGFs with molecular features yields small but consistent gains across six message-passing, attention-based, and graph-transformer backbones, including GCN [1], GINE [21], GeneralConv [22], GAT [3], Transformer [23, 24], and our customized graph transformer (GT)

[Sec. 4.1]; we further validate the trend on ESA [25], a graph architecture outside the standard transformer family [Sec. 4.2]. Fourth, we execute the full feature-extraction pipeline on a neutral-atom QPU for all 188 MUTAG graphs at four measurement time steps. Despite the unavoidable noise in the hardware, the QPU-derived compressed quantum features match their emulated counterparts within 0.01 F_1 on all six backbones, and the full time-resolved features within 0.014 F_1 on five of six [Sec. 4.1]. Lastly, we introduce geometry-augmented baselines, atom-placement perturbations [App. I.3] and a synthetic family of ladder graphs where QGF bring a performance boost over RRWP outside of cross-fold noise [Sec. 5]. Together, these results establish QPU-derived correlations as hardware-validated structural graph descriptors that complement, rather than replace, classical positional encodings. The complete workflow is presented in Fig. 1.

2 Quantum graph feature extraction

To obtain QGFs, we encode each graph into a graph-dependent Rydberg Hamiltonian, let a fixed initial product state evolve under this Hamiltonian for a chosen time, after which we measure the desired observable: the occupation-based correlation.

2.1 Rydberg Hamiltonian

A neutral-atom register of N qubits encoded in a ground and a Rydberg state evolves under the analog Rydberg Hamiltonian [26–30]

$$\hat{\mathcal{H}}(\mathbb{J}, \Omega, \delta) = \sum_{i<j} J_{ij}\, \hat{n}_i \hat{n}_j + \hbar \sum_i \Big[(\Omega/2)\, \hat{\sigma}_i^x - \delta\, \hat{n}_i \Big], \tag{1}$$

with the Rabi frequency Ω, the global detuning δ, and the positive van-der-Waals coupling strength $J_{ij} = C_6/r_{ij}^6$ set by the inter-atomic distance r_{ij}, and the Rydberg coefficient C_6, see App. A for details. We write $(\mathbb{J})_{ij} = J_{ij}$ for the coupling matrix. By design, the QPU is initialized in the all-zero product state $|\mathbf{0}\rangle \equiv |0\rangle^{\otimes N}$, i.e. $|\psi\,(0)\rangle = |\mathbf{0}\rangle$. The implemented quantum-feature extraction scheme focuses on a specific observable: the occupation-based connected correlation matrix $C_{i,j}$ defined by

$$C_{i,j}(t) = \langle \hat{n}_i(t)\, \hat{n}_j(t) \rangle - \langle \hat{n}_i(t) \rangle \langle \hat{n}_j(t) \rangle \qquad \in [-1/4,\, 1/4]. \tag{2}$$

Due to symmetry, it contains $M = N(N+1)/2$ real-valued and independent entries. Importantly, the operator $\hat{n}_i$ is diagonal in the computational basis such that $C_{i,j}$ can be estimated efficiently from projective readout. Diagonal entries $C_{i,i}(t)$ effectively encode the per-site occupations $n_i(t) = \langle \hat{n}_i(t) \rangle$. The QGFs are defined by the entries of the correlation matrices: diagonal entries provide node-level features, whereas off-diagonal entries provide pairwise edge-level features for the models studied here. To obtain these features, we embed graphs into the QPU register, as described next.

2.2 Graph embedding

We focus on molecular graphs drawn from two established benchmark datasets: MUTAG (188 graphs, binary mutagenicity) and $\mathrm{ZINC}_{\mathrm{subset}}$(12'000 graphs, constrained solubility regression). Each molecule is represented as an attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{L}_{\mathcal{V}}, \mathcal{L}_{\mathcal{E}})$ with nodes (atoms) $\mathcal{V}$, edges (chemical bonds) $\mathcal{E}$, and attributes $\mathcal{L}_{\mathcal{V}}$ and $\mathcal{L}_{\mathcal{E}}$ for nodes and edges, respectively [18]. To encode a graph into the Rydberg Hamiltonian of Eq. (1), we assign each graph node to a qubit in the QPU register plane. The atomic positions are chosen such that pairs of atoms corresponding to graph edges interact strongly, while non-edge pairs are separated far enough that their residual van der Waals interactions are weak. The resulting graph-dependent coupling matrix is

$$\mathbb{J}_{ij}^{\mathcal{G}} = \begin{cases} J_{\mathrm{NN}} & (i,j) \in \mathcal{E}, \\ J_{ij} \ll J_{\mathrm{NN}} & (i,j) \notin \mathcal{E}, \end{cases} \tag{3}$$

where $J_{\mathrm{NN}} = C_6/r_{\mathrm{NN}}^6$ is the target nearest-neighbor coupling strength set by the edge distance r_{NN}. Unless stated otherwise, we fix $r_{\mathrm{NN}} = 5.01\,\mu\mathrm{m}$ in the emulations, corresponding to $J_{\mathrm{NN}}/\hbar \approx 54.75\,\mathrm{rad}/\mu\mathrm{s}$.

This embedding strategy is naturally related to unit-disk graphs [31]. Recognizing whether an arbitrary graph admits a unit-disk representation is NP-hard [32, 33]. Molecular graphs, however, originate from physical constraints and thus, they often admit a geometric realization consistent

with the scheme of Eq. (3); the primary limitation in current neutral-atom devices is that the programmable register is effectively two-dimensional. We therefore restrict the datasets to molecules that can be embedded in two dimensions: all 188 MUTAG graphs and 11,368 of $\mathrm{ZINC}_{\mathrm{subset}}$(95%; train/val/test 9,475/946/947), denoted $\mathrm{ZINC}^{*}_{\mathrm{subset}}$. The atomic layouts are computed using a force-based placement method following standard graph-drawing principles, `BLaDE` [34]. Here, we use a purely structural embedding: neither node attributes $\mathcal{L}_{\mathcal{V}}$ nor edge attributes $\mathcal{L}_{\mathcal{E}}$ are encoded, e.g., molecular bond types are not reflected in the coupling strengths J_{ij}. QGFs are therefore descriptors of a specified graph-to-register protocol: small positional noise leaves them only moderately perturbed (Spearman $\rho = 0.82$ at 2% noise), whereas alternative valid layouts yield measurably different features (App. I.3).

2.3 Quench-parameter selection: AFM and FM regimes

Having embedded a graph into the Rydberg Hamiltonian $\hat{\mathcal{H}}^{\mathcal{G}} = \hat{\mathcal{H}}\left(\mathbb{J}^{\mathcal{G}}, \Omega, \delta\right)$, we select two constant quench settings (Ω, δ) that generate distinct correlation patterns. The first setting produces antiferromagnetic (AFM)-like correlations, while the second produces ferromagnetic (FM)-like correlations. In both cases, the parameters are chosen to yield pronounced connected density correlations $C_{i,j}$ [Eq. (2)], which are used as quantum features in the downstream graph-learning task. Our parameter-selection strategy is guided by thermalization arguments [35], detailed in App. B and applied to the Rydberg Hamiltonian in App. C. The two quench settings used in this work define the AFM-like and FM-like Hamiltonians [cf. Eqs. (1), (3)]

$$\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{AFM}} = \hat{\mathcal{H}}\left(\mathbb{J}^{\mathcal{G}},\, \Omega = 0.3\, J_{\mathrm{NN}}/\hbar,\, \delta = 0\right),$$
$$\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{FM}} = \hat{\mathcal{H}}\left(\mathbb{J}^{\mathcal{G}},\, \Omega = 0.5\, J_{\mathrm{NN}}/\hbar,\, \delta = 2\delta^{\mathcal{G}}_{\infty}\right), \tag{4}$$

where $\delta^{\mathcal{G}}_{\infty} = \sum_{i \neq j} J^{\mathcal{G}}_{ij}/(4N\hbar)$. The value $\delta^{\mathcal{G}}_{\infty}$ corresponds to the detuning for which the all-zero initial state lies at infinite effective temperature. The choice $\delta = 2\delta^{\mathcal{G}}_{\infty}$ cancels the spatially averaged longitudinal field and therefore places the FM setting near a field-balanced transverse-field Ising regime, in the vicinity of a finite-size thermal crossover. Figure 2 confirms the expected qualitative behavior on a small molecular graph. In the FM setting, neighboring atoms develop positive off-diagonal correlations, whereas in the AFM setting the corresponding correlations are negative. This distinction follows from the sign of the effective interaction, set by δ relative to $\delta^{\mathcal{G}}_{\infty}$, rather than from finely tuned pulse values: both working points, AFM and FM, are chosen without task-level tuning, inside a broad region of pronounced correlations (App. C).

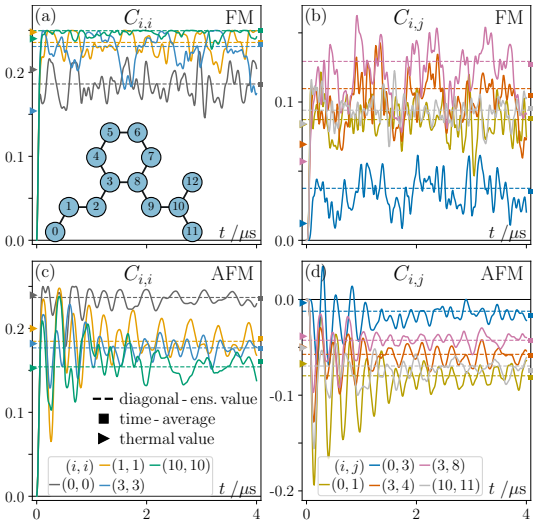


Figure 2: Time evolution of the correlation matrix $C_{i,j}$ (2) for the indicated molecular graph from the $\mathrm{ZINC}_{\mathrm{subset}}$dataset, under the two characteristic Hamiltonians of Eq. (4). Panels (a)–(b) show the FM setting $\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{FM}}$, while panels (c)–(d) show the AFM setting $\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{AFM}}$. The left panels display diagonal elements $C_{i,i}$, whereas the right panels show off-diagonal elements $C_{i,j}$ with $i \neq j$. The dashed lines indicate the diagonal-ensemble values $C^{\mathrm{diag}}_{i,j}$, square markers on the right indicate finite-time averages over $[0, 4\,\mu\mathrm{s}]$, and triangular markers on the left indicate thermal expectation values $C^{\mathrm{th}}_{i,j}$ (see App. C for details).

2.4 Quantum graph feature pipeline: emulation and hardware

For both datasets we emulate the quench dynamics under $\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{AFM}}$ and $\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{FM}}$ [Eq. (4)] with Pasqal's open-source emulators [36], interfaced with Pulser [37], which is also used for the hardware runs, ensuring consistency between emulated and hardware pulses. For the emulated benchmark, we record the correlation elements $C_{i,j}(t)$ [Eq. (2)] at $N_t = 20$ equally spaced time steps $t_\nu \in [0.1, 4.0]\,\mu\mathrm{s}$. At each time step, the observables are sampled $N_{\mathrm{shots}} = 1000$ times, which—based on a central-limit-theorem-derived budget (App. D.3)—is sufficient to resolve all correlation elements at 95% simultaneous confidence for all MUTAG graphs and for $\mathrm{ZINC}^{*}_{\mathrm{subset}}$ graphs with $N \leq 30$ (App. D.3).

The derived quantum descriptors are reused as fixed inputs across all downstream models. For each graph and each quench-pulse setting, AFM and FM, we concatenate the sampled correlation elements to form distinct descriptor sets

$$\boldsymbol{\mathcal{C}} = \left[C_{i,j}(t_\nu)\right]_{\nu=1}^{N_t} \in \mathbb{R}^{MN_t}, \quad \boldsymbol{\mathcal{C}}_i^{(1)} = \left[C_{i,i}(t_\nu)\right]_{\nu=1}^{N_t} \in \mathbb{R}^{N_t}, \quad \boldsymbol{\mathcal{C}}_{ij}^{(2)} = \left[C_{i,j\neq i}(t_\nu)\right]_{\nu=1}^{N_t} \in \mathbb{R}^{N_t}, \tag{5}$$

of indicated dimensions, which are interchangeably used in the ML protocols. For example, if a model permits, $\boldsymbol{\mathcal{C}}_i^{(1)}$ can be used for node-feature injection while $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ can be injected in edge features. Depending on which quench-pulse setting the downstream models are trained on, their respective method is labeled by AFM or FM. Additionally, assuming that much of the graph information is encoded in the time average [see App. B], we also train the model on the four-dimensional summary statistics $(\mathrm{mean}, \mathrm{variance}, \mathrm{minimal\ value}, \mathrm{maximal\ value})$, indicated later via $\mathrm{AFM_{stat}}$ and $\mathrm{FM_{stat}}$, respectively. A summary of all feature sets and their dimensionalities is given in App. E.1 (Table 6). Emulation details are given in App. D.1. For the hardware run, the full pipeline is executed on a Pasqal Fresnel neutral-atom QPU [38] for all 188 MUTAG graphs at four selected time steps due to budget and time restrictions. The required QPU wall time for the chosen setting is ~ 7.5 QPU-days. The four selected time steps are determined based on a principal component analysis, see App. D.2. Despite SPAM, decoherence and shot noise, the per-graph Spearman concordance between hardware and emulator features is high, with medians $\rho_s = 0.924$ for diagonal elements $\boldsymbol{\mathcal{C}}_i^{(1)}$ and $\rho_s = 0.870$ for off-diagonal elements $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$, see App. D.4. Models trained with hardware correlations are indicated via prefix HW-, for example HW-AFM.

2.5 Classical baselines for graph structural features

We mainly use two classical baselines: (i) RRWP [5] involving transition-matrix powers up to order $k_{\max} = 20$ to match $N_t = 20$ and (ii) Laplacian [39] with four leading eigenvectors to match dimension of the summary statistics, e.g. $\mathrm{AFM_{stat}}$. We choose RRWP over other positional encoding methods such as RWPE [40], SignNet [41], GRIT [5], because RRWP underpins GRIT on the one hand, and on the other hand, it matches or exceeds the alternatives on the studied molecular benchmarks. RWSE with ring-membership features are used only for the ESA model, since this feature set gives the best performance for that particular architecture [25]. Finally, we use two control baselines: (i) Random (Gaussian pseudo-features of matched dimension, fixed per graph) controls for input-dimensionality gains, and (ii) *No data* (atom/bond attributes only) gives each backbone's bare performance and is restricted to the downstream setting. To control for information beyond random walks, including the register geometry seen by the quantum dynamics, we also use a strengthened baseline RRWP+Laplacian+Rings+Geometry, which adds ring-membership descriptors, the interatomic distances r_{ij} and the couplings J_{ij} (Sec. 5). All features are z-scored using training-split statistics.

3 Intrinsic Information Benchmark

We first measure how much graph-theoretic information each descriptor family encodes intrinsically [Sec. 3.1], then test whether QGFs can be reconstructed from classical descriptors [Sec. 3.2].

3.1 Predicting graph properties

We design a lightweight probing benchmark to measure graph-theoretic information encoded in each descriptor while minimizing model-specific inductive biases. It spans 15 tasks, formally defined in App. F.1: node-level tasks (degree, betweenness, eccentricity, bridge-node detection, smallest cycle through node), edge-level tasks (link prediction, edge betweenness, bridge-edge detection, local connectivity, smallest cycle through edge), and graph-level tasks (number of cycles, diameter, density, Wiener index, algebraic connectivity).

The probe is a DeepSet [42] with a shared encoder ϕ and predictor ρ. Node- and edge-level targets are processed independently through $\rho \circ \phi$. For graph-level targets, node features are enriched with $\boldsymbol{\mathcal{C}}_i^{(1)}$ [Eq. (5)] and the row-wise mean of $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ over j, encoded, sum-pooled, and passed to ρ. The probe has no message passing, attention, or learned graph traversal; any recovered signal must already be present in the input encoding. It is diagnostic rather than competitive, revealing which structural targets are naturally accessible from each feature family. Training details are deferred to App. F.1.

Table 1: Intrinsic information benchmark on $\text{ZINC}^*_{\text{subset}}$. Means over 5-fold cross-validation are reported; standard deviations over folds are omitted for readability and provided in Table 13 (Appendix F.2). Metrics are MAE (↓) for regression tasks and accuracy (↑) for classification tasks. **Best** and 2nd best indicate the best and runner-up feature sets among the columns shown; entries marked † are tied with the best. R: RRWP ($d = 20$); L: Laplacian ($d = 4$); F: FM; A: AFM. Quantum columns FM_{stat} and AFM_{stat} use the four-statistic compression ($d = 4$), whereas FM and AFM use the full time series ($d = 20$).

	Classical			Quantum				Hybrid		
Task	R	L	R+L	F_{stat}	F	A_{stat}	A	R+F	R+A	R+F+A
Node Prediction										
Bridge Node↑	0.93	0.76	0.95	0.74	0.83	0.85	0.91	0.97	**0.98**†	**0.98**†
Degree↓ ($\times 10^2$)	14.4	46.7	6.98	30.5	18.7	9.07	1.26†	4.71	1.18†	1.25†
Betweenness↓ ($\times 10^2$)	7.77	12.5	4.76	9.10	7.17	10.1	7.11	4.33	3.90	**3.31**
Eccentricity↓	1.88	1.51	**1.08**	1.90	1.65	2.02	1.91	1.48	1.77	1.47
Min Cycle (Node)↓	0.73	1.68	0.60	1.80	1.42	1.40	0.88	0.48	0.36	**0.28**
Edge Prediction										
Link Pred.↑	1.00†	0.94	1.00†	0.99	0.99	1.00†	1.00†	1.00†	1.00†	1.00†
Bridge Edge↑	0.99†	0.98	0.99†	0.97	0.98	0.97	0.99	0.99†	0.99†	1.00†
Edge Betw.↓ ($\times 10^3$)	11.6	20.6	6.09	13.0	9.89	9.76	8.61	6.69	6.13	**5.36**
Local Conn.↓ ($\times 10^2$)	10.0	13.9	7.39	13.9	12.6	15.2	11.3	6.83	6.07	**5.14**
Min Cycle (Edge)↓ ($\times 10^2$)	13.3	42.4	10.2	23.0	19.4	23.1	13.2	10.0	**7.62**†	**7.17**†
Graph Prediction										
Num Cycles↓ ($\times 10^2$)	6.68	27.4	5.90	18.4	15.3	16.4	9.85	4.83†	4.48†	4.67†
Diameter↓	0.49	0.69	0.47	0.71	0.65	0.75	0.60	0.36	0.37	**0.32**
Density↓ ($\times 10^3$)	1.03†	2.48	1.14†	1.86	1.70	1.72	1.86	1.21†	1.10†	1.25†
Wiener Index↓ ($\times 10^2$)	10.8	19.2	9.61	16.4	14.1	19.5	15.0	**7.13**†	8.81	**7.24**†
Alg. Connect.↓ ($\times 10^3$)	2.63	5.59	**2.22**†	5.46	5.21	6.30	5.32	2.70	**2.34**†	2.38

Table 1 reports results on $\text{ZINC}^*_{\text{subset}}$ for ten feature sets: classical descriptors (RRWP, Laplacian, RRWP+Laplacian), quantum descriptors (FM, AFM, FM_{stat}, AFM_{stat}; see Sec. 2.4), and three RRWP-based hybrids. A random-feature control, per-Hamiltonian mean variants, and MUTAG results are given in App. F.2 and App. F.3. The random control is beaten by every structural feature by a large margin on non-trivial tasks, while link prediction and bridge-edge detection are close to saturation and retained only as sanity checks.

Quantum correlations carry strong local information and complement RRWP; RRWP alone is a strong classical baseline, consistent with its use as a positional encoding [5]. Quantum correlations provide a strong signal on selected local tasks: on degree prediction, AFM alone reaches 1.26 ($\times 10^{-2}$ MAE), clearly ahead of RRWP+Laplacian (6.98) and far ahead of RRWP (14.4). Combined with RRWP, this signal yields clear hybrid gains: degree prediction improves from 14.4 to 1.18 and 1.25 ($\times 10^{-2}$ MAE, statistically tied) when AFM and then FM+AFM are added. A monotone pattern holds for local connectivity ($10.0 \to 6.07 \to 5.14$), smallest cycle through a node ($0.73 \to 0.36 \to 0.28$), and Wiener index ($10.8 \to 8.81 \to 7.24$, $\times 10^{-2}$). These gains span node-, edge-, and graph-level probes, showing that QGFs provide information compatible with, but not redundant with, random-walk transition probabilities.

AFM and FM correlations probe different structural regimes. RRWP+AFM is stronger than RRWP+FM on local neighbourhood and cycle-related tasks, including degree, betweenness, and smallest cycle through a node or edge (e.g., node cycle: 0.36 vs. 0.48; edge cycle: 7.62 vs. 10.0, $\times 10^{-2}$). Conversely, RRWP+FM is stronger on global-distance targets, reaching 1.48 MAE on eccentricity against 1.77 for RRWP+AFM, and close to RRWP+AFM on diameter. This suggests, without formally establishing, that AFM-like dynamics are more sensitive to local frustration and cycle structure, whereas FM-like dynamics capture more coherent graph-scale correlation propagation. RRWP+FM+AFM inherits both strengths and is best, tied with the best, or second-best on most non-saturating structural tasks. The few remaining cases correspond to global or spectral targets better captured by RRWP+Laplacian, or by the simpler RRWP+AFM hybrid. Statistic-compressed variants impact and hardware results for MUTAG are discussed in resp. App. F.2.1 and App. F.3.1.

3.2 A ceiling on quantum feature reconstruction with classical descriptors

We empirically assess whether quantum features can be reconstructed from classical graph descriptors. For this purpose we construct a regression target $y(t)$ that, at a given time step t, combines the off-diagonal correlation elements $C_{i,j}(t)$ from all graphs in the dataset via a specific pooling scheme, see App. H. We test whether $y(t)$ is reducible to two classical pair-feature families up to a chosen order K: (i) walk-powers, i.e., powers of the adjacency matrix A, and (ii) RRWP, i.e., powers of the degree-normalized transition matrix $D^{-1}A$. For both physical regimes (AFM and FM), Figure 3 illustrates on a toy dataset ($N_{\mathrm{data}} = 2$) that quantum features at early times are well approximated by linear combination of powers of RRWP or adjacency ($R^2 \approx 1$), whereas at later times ($t \gtrsim 200\,\mathrm{ns}$ for AFM and $t \gtrsim 100\,\mathrm{ns}$ for FM) this is not the case anymore.

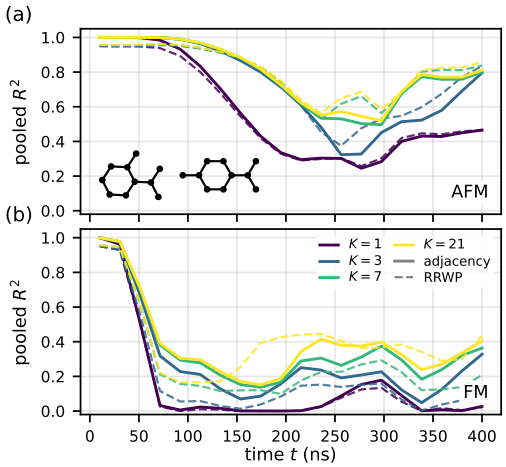


Figure 3: Linear recovery of quantum features from adjacency powers and RRWP features including matrix powers up to order K, for the two indicated graphs, and the two quench-parameter settings (a) AFM and (b) FM [Eq. (4)]. After an early-time period, quantum features are no longer fully recovered.

The results on the full MUTAG and $\mathrm{ZINC}^*_{\mathrm{subset}}$ datasets are provided in App. H, and summarized in the following. Under a pooled linear probe, the explained variance of $y(t)$ saturates at the measurement times: R^2 reaches ~ 0.83 on MUTAG and ~ 0.78 on $\mathrm{ZINC}^*_{\mathrm{subset}}$ at $K = 21$, improving by at most 2% up to $K = 50$ (see Table 15 and Figs. 9,8). To rule out a linearity limitation, we fit kernel ridge (RBF) and random forest on the same features and score them by 5-fold held-out R^2; both close at most 7–8 percent points of the gap for walk powers on MUTAG at $K = 21$, $t = 1536\,\mathrm{ns}$ (graph-level folds), leaving a residual of $\sim 9\%$ on AFM and $\sim 22\%$ on FM that is not recovered by any nonlinear function of adjacency or RRWP (see Table 16). Augmenting the design with the first K Laplacian eigenpairs additionally lifts the FM result to 0.85, identifying the FM beyond-walk content as partly spectral, while the AFM residual is nearly unchanged ($\leq 1\%$) and remains beyond any walk- or spectrum-derived feature we tested (see Table 17). We further expand the basis with ring, heat-kernel, geodesic and effective-resistance descriptors, the register distances r_{ij} and couplings J_{ij}, and restricted physical approximations of the same dynamics (classical-spin dTWA [43, 44], exact two-atom and mean-field correlations). The unexplained variance then remains 8.3% for AFM and 14.0% for FM (App. I.1). We therefore describe it as information not captured by the tested classical, geometric and restricted physical descriptors, rather than as purely quantum information.

4 Downstream evaluation across GNN backbones

4.1 Architectures, feature injection and results

To test robustness across downstream models, we evaluate six backbones: 5 from PyTorch Geometric [45]—GCN [1], GAT [3], GINE [21], Transformer [24], GeneralConv [22]—plus a custom dense graph transformer (GT). Diagonal features $\boldsymbol{\mathcal{C}}_i^{(1)}$ [Eq. (5)] are concatenated to the node embedding (or, in GT, projected and added residually); off-diagonal features $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ enter through whichever pairwise interface each backbone exposes (edge weights for GCN, edge attributes for the four MPNN/attention layers, dense per-head attention biases for GT). Models details are given in App. E.2. Architectural details of the GT, which consumes the dense tensor $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ for all node pairs, are given in App. E.3.

To control for model capacity, we approximately match the number of trainable parameters across backbones separately for each dataset. We keep the training protocol fixed across all feature configurations and do not perform per-feature hyperparameter tuning, thereby prioritizing a clean comparison of features over peak predictive performance. Hyperparameters are reported in App. E.2. The choice of number of trainable parameters is justified by the model-size ablation in App. E.4. It should

Table 2: Comparison of GNN architectures and feature sets on MUTAG. Mean macro F_1 (↑) ± standard deviation over 10 stratified cross-validation folds × 5 seeds is reported. Best and 2nd best indicate the best and runner-up feature sets in each column.

Method	MUTAG (F_1 macro ↑)					
	GT	GCN	GAT	GINE	Transformer	General
Baseline						
No Data	$0.690_{\pm 0.104}$	$0.652_{\pm 0.134}$	$0.734_{\pm 0.103}$	$0.785_{\pm 0.098}$	$0.781_{\pm 0.084}$	$0.823_{\pm 0.065}$
Classical						
Laplacian	$0.807_{\pm 0.065}$	$0.800_{\pm 0.099}$	$0.787_{\pm 0.090}$	$0.818_{\pm 0.106}$	$0.802_{\pm 0.072}$	$0.821_{\pm 0.079}$
RRWP	$0.854_{\pm 0.060}$	$0.849_{\pm 0.081}$	$0.862_{\pm 0.078}$	$0.861_{\pm 0.071}$	$0.869_{\pm 0.066}$	$0.868_{\pm 0.080}$
RRWP + Laplacian	$\underline{0.868}_{\pm 0.071}$	$0.851_{\pm 0.073}$	$0.850_{\pm 0.078}$	$0.843_{\pm 0.073}$	$0.852_{\pm 0.078}$	$0.854_{\pm 0.073}$
Quantum						
AFM	$0.830_{\pm 0.054}$	$0.817_{\pm 0.098}$	$0.791_{\pm 0.108}$	$0.851_{\pm 0.093}$	$0.848_{\pm 0.086}$	$0.849_{\pm 0.090}$
AFM_{stat}	$0.806_{\pm 0.047}$	$0.778_{\pm 0.075}$	$0.782_{\pm 0.102}$	$0.809_{\pm 0.084}$	$0.811_{\pm 0.081}$	$0.816_{\pm 0.094}$
FM	$0.851_{\pm 0.059}$	$0.832_{\pm 0.079}$	$0.821_{\pm 0.082}$	$0.844_{\pm 0.086}$	$0.860_{\pm 0.070}$	$0.814_{\pm 0.083}$
FM_{stat}	$0.841_{\pm 0.069}$	$0.792_{\pm 0.103}$	$0.828_{\pm 0.077}$	$0.834_{\pm 0.083}$	$0.829_{\pm 0.089}$	$0.831_{\pm 0.078}$
AFM + FM	$0.835_{\pm 0.077}$	$0.845_{\pm 0.074}$	$0.836_{\pm 0.092}$	$0.843_{\pm 0.083}$	$0.853_{\pm 0.075}$	$0.815_{\pm 0.064}$
Hybrid						
RRWP + AFM	$0.855_{\pm 0.078}$	$\mathbf{0.878}_{\pm 0.078}$	$0.862_{\pm 0.070}$	$0.848_{\pm 0.069}$	$0.868_{\pm 0.079}$	$0.867_{\pm 0.074}$
RRWP + FM	$0.862_{\pm 0.060}$	$0.856_{\pm 0.083}$	$0.866_{\pm 0.082}$	$0.847_{\pm 0.080}$	$0.869_{\pm 0.089}$	$0.848_{\pm 0.065}$
RRWP + AFM_{stat}	$0.854_{\pm 0.077}$	$0.852_{\pm 0.075}$	$0.860_{\pm 0.078}$	$0.846_{\pm 0.064}$	$0.852_{\pm 0.078}$	$0.873_{\pm 0.077}$
RRWP + FM_{stat}	$0.851_{\pm 0.070}$	$\underline{0.867}_{\pm 0.065}$	$\mathbf{0.876}_{\pm 0.075}$	$\underline{0.866}_{\pm 0.070}$	$0.870_{\pm 0.072}$	$\underline{0.873}_{\pm 0.082}$
RRWP + AFM + FM	$\mathbf{0.874}_{\pm 0.081}$	$0.850_{\pm 0.090}$	$\underline{0.872}_{\pm 0.077}$	$\mathbf{0.867}_{\pm 0.075}$	$\mathbf{0.878}_{\pm 0.080}$	$0.859_{\pm 0.079}$
RRWP + AFM_{stat} + FM_{stat}	$0.862_{\pm 0.053}$	$0.860_{\pm 0.076}$	$0.868_{\pm 0.074}$	$0.862_{\pm 0.078}$	$\underline{0.871}_{\pm 0.071}$	$0.870_{\pm 0.078}$
Quantum hardware						
HW-AFM	$0.784_{\pm 0.098}$	$0.793_{\pm 0.096}$	$0.796_{\pm 0.052}$	$0.826_{\pm 0.073}$	$0.805_{\pm 0.075}$	$0.823_{\pm 0.097}$
HW-AFM_{stat}	$0.778_{\pm 0.086}$	$0.767_{\pm 0.086}$	$0.801_{\pm 0.059}$	$0.833_{\pm 0.055}$	$0.810_{\pm 0.053}$	$0.809_{\pm 0.075}$
RRWP + HW-AFM	$0.854_{\pm 0.074}$	$0.837_{\pm 0.076}$	$0.852_{\pm 0.059}$	$0.841_{\pm 0.059}$	$0.854_{\pm 0.060}$	$0.859_{\pm 0.079}$
RRWP + HW-AFM_{stat}	$0.849_{\pm 0.072}$	$0.852_{\pm 0.055}$	$0.853_{\pm 0.062}$	$0.852_{\pm 0.059}$	$0.861_{\pm 0.060}$	$\mathbf{0.877}_{\pm 0.066}$

Table 3: Comparison of GNN architectures and feature sets on $ZINC^*_{subset}$. Mean MAE (↓) ± standard deviation over 10 seeds (GT) or 5 seeds (other backbones) is reported. Best and 2nd best indicate the best and runner-up feature sets in each column, † indicates a tie.

Method	$ZINC^*_{subset}$ (MAE ↓)					
	GT	GCN	GAT	GINE	Transformer	General
Baseline						
No Data	$0.317_{\pm 0.018}$	$0.591_{\pm 0.066}$	$0.434_{\pm 0.008}$	$0.236_{\pm 0.041}$	$0.339_{\pm 0.009}$	$0.331_{\pm 0.006}$
Classical						
Laplacian	$0.232_{\pm 0.020}$	$0.478_{\pm 0.006}$	$0.434_{\pm 0.005}$	$0.195_{\pm 0.012}$	$0.243_{\pm 0.034}$	$0.270_{\pm 0.031}$
RRWP	$0.113_{\pm 0.003}$	$0.252_{\pm 0.006}$	$0.223_{\pm 0.002}$	$0.117^{\dagger}_{\pm 0.007}$	$0.126_{\pm 0.006}$	$0.148_{\pm 0.004}$
RRWP + Laplacian	$0.112_{\pm 0.004}$	$0.278_{\pm 0.017}$	$0.229_{\pm 0.007}$	$0.117^{\dagger}_{\pm 0.007}$	$\underline{0.123}_{\pm 0.008}$	$\mathbf{0.145}_{\pm 0.005}$
Quantum						
AFM	$0.163_{\pm 0.004}$	$0.339_{\pm 0.040}$	$0.288_{\pm 0.010}$	$0.188_{\pm 0.002}$	$0.191_{\pm 0.007}$	$0.214_{\pm 0.006}$
AFM_{stat}	$0.168_{\pm 0.008}$	$0.419_{\pm 0.071}$	$0.360_{\pm 0.024}$	$0.192_{\pm 0.007}$	$0.195_{\pm 0.006}$	$0.218_{\pm 0.009}$
FM	$0.183_{\pm 0.010}$	$0.479_{\pm 0.017}$	$0.430_{\pm 0.005}$	$0.248_{\pm 0.040}$	$0.283_{\pm 0.036}$	$0.325_{\pm 0.005}$
FM_{stat}	$0.183_{\pm 0.010}$	$0.465_{\pm 0.016}$	$0.413_{\pm 0.004}$	$0.208_{\pm 0.012}$	$0.292_{\pm 0.013}$	$0.309_{\pm 0.005}$
AFM + FM	$0.154_{\pm 0.006}$	$0.314_{\pm 0.020}$	$0.304_{\pm 0.009}$	$0.189_{\pm 0.003}$	$0.195_{\pm 0.012}$	$0.215_{\pm 0.012}$
Hybrid						
RRWP + AFM	$0.111^{\dagger}_{\pm 0.004}$	$0.255_{\pm 0.011}$	$0.215_{\pm 0.008}$	$0.127_{\pm 0.008}$	$0.128_{\pm 0.006}$	$0.157_{\pm 0.007}$
RRWP + FM	$0.120_{\pm 0.004}$	$0.285_{\pm 0.016}$	$0.242_{\pm 0.008}$	$0.130_{\pm 0.004}$	$0.146_{\pm 0.006}$	$0.166_{\pm 0.005}$
RRWP + AFM_{stat}	$\mathbf{0.108}_{\pm 0.007}$	$\mathbf{0.244}_{\pm 0.004}$	$\underline{0.210}_{\pm 0.011}$	$0.124_{\pm 0.009}$	$0.125_{\pm 0.006}$	$0.150_{\pm 0.006}$
RRWP + FM_{stat}	$0.111^{\dagger}_{\pm 0.004}$	$0.257_{\pm 0.010}$	$0.228_{\pm 0.004}$	$0.121_{\pm 0.009}$	$\mathbf{0.120}_{\pm 0.010}$	$\mathbf{0.145}_{\pm 0.006}$
RRWP + AFM + FM	$0.114_{\pm 0.005}$	$0.262_{\pm 0.011}$	$0.235_{\pm 0.006}$	$0.135_{\pm 0.011}$	$0.147_{\pm 0.004}$	$0.165_{\pm 0.010}$
RRWP + AFM_{stat} + FM_{stat}	$0.111^{\dagger}_{\pm 0.005}$	$\underline{0.247}_{\pm 0.008}$	$\mathbf{0.209}_{\pm 0.004}$	$\mathbf{0.116}_{\pm 0.009}$	$0.128_{\pm 0.002}$	$0.149_{\pm 0.008}$

be noted that increasing model size after a certain point does not increase performance for RRWP, motivating the addition of supplemental features as a source of improvement.

Results. Tables 2 and 3 report resp. 19 and 15 feature sets on MUTAG and $ZINC^*_{subset}$ across six backbones. We emphasize that $ZINC^*_{subset}$ is a non-standard subset, so our results should not be directly compared with published results on $ZINC_{subset}$. The feature sets fall into three blocks: classical structural encodings (Random, Laplacian, RRWP, RRWP+Laplacian) [Sec. 2.5], standalone quantum features (AFM,FM, AFM_{stat} and FM_{stat}) [Sec. 2.4], and hybrids RRWP+{quantum pool}. As demonstrated in the tables, standalone quantum features do not substitute for RRWP in this setup. Yet, the column-best entry is always attained by a quantum–RRWP hybrid, with the GeneralConv column on $ZINC^*_{subset}$ tied between RRWP+Laplacian and RRWP+FM_{stat} at 0.145 MAE. RRWP+AFM_{stat}+FM_{stat} stays within one standard deviation of RRWP+AFM+FM and attains the column best in two cells (GAT and GINE on $ZINC^*_{subset}$).

Because per-cell recipe selection is a free parameter, we test this observation with fixed champions selected by best average performance across the sweep: RRWP for the classical block and RRWP+AFM_{stat}+FM_{stat} for the hybrid block. The hybrid champion ties or improves on the classical champion in ten of twelve (architecture, dataset) cells; the two losses (Transformer and GeneralConv on ZINC) sit within 0.5σ of equality. A one-sided sign test [46] gives $p = 0.019$, sup-

porting a sign-consistent but small-amplitude signal: most per-cell gaps lie inside the cross-fold noise band, with only GAT on ZINC reaching a 3σ effect (0.209 vs 0.223 MAE). Bayesian decompositions in App. G [47] support the same calibration: the effect is directionally asymmetric in favor of the hybrid champion, but mostly compatible with practical equivalence once cell- or architecture-level uncertainty is accounted for. Altogether, no single feature set dominates uniformly, but pooling with both quench settings combined, RRWP+AFM+FM and RRWP+AFM_{stat}+FM_{stat}, ranks in the top two in eight of the twelve cells, mirroring downstream the AFM/FM complementarity observed in the information benchmark.

Hardware. We repeat the MUTAG benchmark with AFM features obtained by executing the corresponding pulse on a neutral-atom device; the FM pulse and the $\text{ZINC}^{*}_{\text{subset}}$ sweep are deferred to future work. Rows HW-AFM and HW-AFM_{stat} of Table 2 show that the standalone hardware features remain close to their emulated counterparts despite SPAM errors, decoherence, and finite shots; readout errors are corrected in post-processing (App. D.3). HW-AFM differs from AFM by at most 0.046 F_1 across the six backbones, while HW-AFM_{stat} stays within 0.028 F_1 and even exceeds the emulator on GAT and GINE. The downstream hybrid picture is strongest for the compressed features. RRWP+HW-AFM is within 0.014 F_1 of RRWP+AFM on five of six backbones, with GCN the only larger deviation (0.837 vs 0.878). In contrast, RRWP+HW-AFM_{stat} matches RRWP+AFM_{stat} within 0.01 F_1 on all six backbones and reaches the MUTAG column best on GeneralConv (0.877); per-backbone differences are listed in Table 12 (App. E.5). The feature-level ranking of the compressed hybrid benchmark therefore transfers to a real Rydberg device under realistic finite-shot noise and residual coherent errors.

4.2 Classical–quantum feature fusion

The preceding benchmark shows that compressed quantum statistics are often on par with, or better than, the full time-resolved correlation vectors. This suggests that full QGFs are informative but not always easy to exploit downstream. Although the intrinsic benchmark (Sec. 3) shows strong hybrid information content, fixed task-agnostic time samples may include redundant or weakly informative entries, making it difficult for standard GNN concatenation to isolate the useful quantum signal and combine it effectively with RRWP. We therefore replace unconditional concatenation by gated fusion, allowing the model to regulate how much information is passed from each feature stream [48, 49].

Table 4: ESA model performance on MUTAG with concatenation or gated fusion strategy. RWSE* denotes RWSE augmented with ring-membership features. Mean binary F_1 (↑) ± standard deviation over 10 stratified cross-validation folds × 5 seeds is reported. Best and 2nd best indicate the best and runner-up methods overall.

Method	MUTAG (F_1 binary ↑)	
	Concat.	Gated
Baseline		
RWSE*	$0.905_{\pm 0.045}$	—
Hybrid (classical + quantum emulation)		
RWSE*+AFM_{stat}	$0.899_{\pm 0.044}$	$0.915_{\pm 0.042}$
RWSE*+FM_{stat}	$0.888_{\pm 0.076}$	$0.918_{\pm 0.046}$
RWSE*+AFM	$0.893_{\pm 0.049}$	$0.915_{\pm 0.048}$
RWSE*+FM	$0.913_{\pm 0.042}$	$0.911_{\pm 0.043}$
RWSE*+AFM+FM	$0.903_{\pm 0.050}$	**$0.922_{\pm 0.042}$**
Hybrid (classical + quantum hardware)		
RWSE*+HW-AFM	$0.907_{\pm 0.047}$	$\underline{0.920}_{\pm 0.044}$
RWSE*+HW-AFM_{stat}	$0.899_{\pm 0.046}$	$0.909_{\pm 0.046}$

We use Edge Set Attention (ESA) as a stringent test case: it is a strong molecular graph architecture based on attention over edge-level representations, and is therefore both competitive with modern graph-learning baselines and naturally suited to incorporating pairwise structural information [25]. Starting from the original ESA setup with RWSE* features, we augment the model with complete QGF feature sets, either by direct concatenation or by gated fusion. Table 4 shows that direct concatenation is not consistently beneficial, with most hybrids remaining close to the RWSE* baseline. Gated fusion is more effective: it improves nearly all hybrid variants over their concatenated counterparts and reaches the best score with RWSE*+AFM+FM, at 0.922 ± 0.042 binary F_1. The best hardware hybrid, RWSE*+HW-AFM with gating, reaches a comparable 0.920 ± 0.044. The gains remain modest relative to the cross-validation variance. Still, the learned gates are informative: from an initial value of 0.88, they stay close to initialization for compressed statistics (0.887 for AFM_{stat}, 0.855 for FM_{stat}) and decrease more for FM-containing full features (0.816 for FM, 0.830 for AFM+FM), whereas full AFM stays close to initialization (0.862). This is consistent with the model selectively attenuating the higher-dimensional QGFs rather than ignoring them.

Table 5: Ladder graphs: pinned versus frustrated ground state (394 graphs, 98 RRWP-2-blind groups, 18–26 atoms). Classical-All = RRWP+Laplacian+Ring+Geometry. Logistic regression, 5 folds grouped by blind group × 3 repeats; mean ± standard deviation over the 15 test folds. Label-permutation test (100 shuffles): $p = 0.010$ (the minimum) for all sets except RRWP ($p = 0.020$) and Geometry ($p = 0.446$). Best and 2nd best per metric.

Features	Acc. ↑	F_1 ↑	AUC ↑
Classical			
RRWP	$0.557_{\pm 0.048}$	$0.559_{\pm 0.047}$	$0.612_{\pm 0.065}$
Geom	$0.503_{\pm 0.053}$	$0.495_{\pm 0.080}$	$0.501_{\pm 0.063}$
Classical-All	$0.641_{\pm 0.077}$	$0.643_{\pm 0.091}$	$0.680_{\pm 0.065}$
Quantum emulation			
FM	$0.622_{\pm 0.063}$	$0.620_{\pm 0.071}$	$0.670_{\pm 0.065}$
AFM	$0.847_{\pm 0.045}$	$0.849_{\pm 0.045}$	$\underline{0.914}_{\pm 0.026}$
AFM+FM	$\underline{0.875}_{\pm 0.043}$	$\underline{0.875}_{\pm 0.044}$	$\mathbf{0.941}_{\pm 0.022}$
Hybrid (classical + quantum emulation)			
Classical-All+AFM+FM	$\mathbf{0.882}_{\pm 0.043}$	$\mathbf{0.884}_{\pm 0.043}$	$\mathbf{0.941}_{\pm 0.023}$

5 Quantum graph features beyond classical structural encodings

RRWP-based encodings cannot separate certain non-isomorphic graphs [10, 11]. We construct ladder graphs whose label depends only on the relative parity of two diagonal crossings—a global combinatorial property—and keep only graphs whose opposite-label partners share identical RRWP and ring features (394 graphs, 18–26 atoms, exact unit-disk registers; App. I.4). With a logistic-regression probe, AFM+FM reach 0.875 ± 0.043 accuracy (ROC-AUC 0.941), versus 0.641 ± 0.077 (ROC-AUC 0.680) for the combined classical baseline (Tab. 5); the gap of 0.23 is roughly three times the larger cross-fold standard deviation. This is a result on a synthetic family constructed to defeat these encodings, not evidence of a generic practical advantage for graph learning.

6 Conclusion

In this work we use the occupation-based connected correlation matrix $C_{i,j}$ [Eq. (2)] as *quantum graph features* (QGFs), whereby the observable results from two specific quench-parameter settings of a graph-encoded Rydberg Hamiltonian: FM-like and AFM-like [Eq. (4)]. Diagonal and off-diagonal elements $C_{i,j}$ serve as node-level and edge-level injections, respectively. Across all presented analyses (intrinsic information benchmark [Sec. 3], downstream GNNs [Sec. 4], including hardware-derived quantum features), the main conclusion is consistent: QGFs do not replace classical structural encodings, but provide complementary information that is most useful when concatenated with RRWP. In the intrinsic information benchmark of Sec. 3, the two quench settings, AFM and FM, exhibit distinct local and global biases. About 8% (AFM) and 14% (FM) of QGFs are not captured by the tested classical, geometric features, and on a ladder family QGFs separate two structural classes well beyond cross-fold noise [Sec. 5]. In the downstream benchmark of Sec. 4, a fixed compact RRWP+QGF champion ties or improves over RRWP in ten of twelve cells (one-sided sign test $p = 0.019$), with Bayesian analyses indicating a small effect size.

Crucially, we demonstrate an end-to-end QPU implementation: all 188 MUTAG graphs are embedded with quantum dynamics executed on a neutral-atom device, and the measured features drive the full machine-learning pipeline. The hybrid AFM_{stat} result matches emulation within $\Delta F_1 \sim 0.01$ across all six MUTAG backbones, demonstrating that QGFs are viable structural features for graph learning on molecules on present-day Rydberg hardware. The current limitations are clear. The encoding is restricted to two-dimensional unit-disk embeddings: although our protocol retains about 95% of $\text{ZINC}^{*}_{\text{subset}}$, it still limits graph coverage and introduces a geometry-dependent preprocessing step. Future neutral-atom processors with three-dimensional registers should reduce this bottleneck. The features also depend on the chosen placement: they are stable under small positional noise but not invariant under re-embedding, and canonical or placement-invariant quantum encodings remain an open problem. A further, more practical limitation is cost: hardware extraction remains much slower than classical preprocessing, and the current compute and QPU sampling time of the full pipeline is too high for practical use (see App. D.3). Much higher repetition rates will be required before such features can be generated at graph-learning scale. Finally, the downstream gains are modest, and the feature-extraction protocol is not yet optimized for learning. Fixed quenches, task-agnostic readout times, and low-order summaries may miss informative dynamics or introduce redundant samples. Future work should improve quantum-feature construction and hybrid fusion through optimized readout schedules, learned dynamical summaries, and gating mechanisms, as suggested by the ESA experiment in Sec. 4.2.

Code and data: github.com/pasqal-io/Quantum-Graph-Features, doi:10.5281/zenodo.23099924.

Acknowledgments and Disclosure of Funding

This work was funded by Pasqal SAS. All authors are employees of Pasqal, which develops and commercialises the neutral-atom quantum computer used in this work.

References

[1] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

[2] 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 (ICML)*, volume 70, pages 1263–1272. PMLR, 2017. URL https://arxiv.org/abs/1704.01212.

[3] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

[4] Ladislav Rampášek, Michael 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, pages 14501–14515, 2022. URL https://arxiv.org/abs/2205.12454.

[5] Liheng Ma, Chen Lin, Derek Lim, Adriana Romero-Soriano, Puneet K. Dokania, Mark Coates, Philip H.S. Torr, and Ser-Nam Lim. Graph inductive biases in transformers without message passing. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *PMLR*, 2023. URL https://arxiv.org/abs/2305.17589.

[6] Vijay Prakash Dwivedi, Ladislav Rampášek, Mikhail Galkin, Ali Parviz, Guy Wolf, Anh Tuan Luu, and Dominique Beaini. Long range graph benchmark. In *Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*, 2022. URL https://arxiv.org/abs/2206.08164.

[7] 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 (NeurIPS)*, volume 34, 2021. URL https://arxiv.org/abs/2106.03893.

[8] Romain Menegaux, Emmanuel Jehanno, Margot Selosse, and Julien Mairal. Self-attention in colors: Another take on encoding graph structure in transformers. *arXiv preprint arXiv:2304.10933*, 2023. URL https://arxiv.org/abs/2304.10933.

[9] Bohang Zhang, Shengjie Luo, Liwei Wang, and Di He. Rethinking the expressive power of GNNs via graph biconnectivity. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2301.09505.

[10] Slimane Thabet, Mehdi Djellabi, Igor O. Sokolov, Sachin Kasture, Louis-Paul Henry, and Loïc Henriet. Quantum positional encodings for graph neural networks. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *PMLR*, pages 47965–47996, 2024. URL https://proceedings.mlr.press/v235/thabet24a.html.

[11] Sachin Kasture, Shaheen Acheche, Loic Henriet, and Louis-Paul Henry. Multiparticle quantum walks for distinguishing hard graphs, 2025. URL https://arxiv.org/abs/2501.03683.

[12] Ben Chamberlain, James Rowbottom, Maria I. Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. GRAND: Graph neural diffusion. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1407–1418. PMLR, 2021. URL https://proceedings.mlr.press/v139/chamberlain21a.html.

[13] Xitong Zhang, Yixuan He, Nathan Brugnone, Michael Perlmutter, and Matthew Hirn. MagNet: A neural network for directed graphs. In *Advances in Neural Information Processing Systems*, volume 34, pages 27003–27015, 2021. URL https://arxiv.org/abs/2102.11391.

[14] T. Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 18888–18909. PMLR, 2022. URL https://proceedings.mlr.press/v162/rusch22a.html.

[15] Louis-Paul Henry, Slimane Thabet, Constantin Dalyac, and Loïc Henriet. Quantum evolution kernel: Machine learning on graphs with programmable arrays of qubits. *Physical Review A*, 104(3):032416, 2021. doi: 10.1103/PhysRevA.104.032416. URL https://doi.org/10.1103/PhysRevA.104.032416.

[16] Boris Albrecht, Constantin Dalyac, Lucas Leclerc, Luis Ortiz-Gutiérrez, Slimane Thabet, Mauro D'Arcangelo, Julia R. K. Cline, Vincent E. Elfving, Lucas Lassablière, Henrique Silvério, Bruno Ximenez, Louis-Paul Henry, Adrien Signoles, and Loïc Henriet. Quantum feature maps for graph machine learning on a neutral atom quantum processor. *Phys. Rev. A*, 107:042615, Apr 2023. doi: 10.1103/PhysRevA.107.042615. URL https://doi.org/10.1103/PhysRevA.107.042615.

[17] Thomas Hubregtsen, David Wierichs, Elies Gil-Fuster, Peter-Jan H. S. Derks, Paul K. Faehrmann, and Johannes Jakob Meyer. Training quantum embedding kernels on near-term quantum computers. *Physical Review A*, 106(4):042431, 2022. doi: 10.1103/PhysRevA.106.042431. URL https://doi.org/10.1103/PhysRevA.106.042431.

[18] Mehdi Djellabi, Matthias Hecker, and Shaheen Acheche. Attributed-graph kernel implementation using local detuning of a neutral-atom Rydberg hamiltonian. *Phys. Rev. A*, Apr 2026. doi: 10.1103/xybw-6sch. URL https://link.aps.org/doi/10.1103/xybw-6sch.

[19] Asim Kumar Debnath, Rosa L. Lopez de Compadre, Gargi Debnath, Alan J. Shusterman, and Corwin Hansch. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *Journal of Medicinal Chemistry*, 34(2):786–797, 02 1991. doi: 10.1021/jm00106a046. URL https://doi.org/10.1021/jm00106a046.

[20] Teague Sterling and John J. Irwin. ZINC 15 – Ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015. doi: 10.1021/acs.jcim.5b00559. URL https://doi.org/10.1021/acs.jcim.5b00559.

[21] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HJlWWJSFDH.

[22] Jiaxuan You, Rex Ying, and Jure Leskovec. Design space for graph neural networks, 2021. URL https://arxiv.org/abs/2011.08843.

[23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://arxiv.org/abs/1706.03762.

[24] Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjing Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 1548–1554, 2021. doi: 10.24963/ijcai.2021/214. URL https://www.ijcai.org/proceedings/2021/214.

[25] David Buterez, Jon Paul Janet, Dino Oglic, and Pietro Liò. An end-to-end attention-based approach for learning on graphs. *Nat. Comm.*, 16(1):5244, 2025. doi: 10.1038/s41467-025-60252-z. URL https://doi.org/10.1038/s41467-025-60252-z.

[26] Thomas F. Gallagher. *Rydberg Atoms*. Cambridge Monographs on Atomic, Molecular and Chemical Physics. Cambridge University Press, 1994. URL https://www.cambridge.org/core/books/rydberg-atoms/B610BDE54694936F496F59F326C1A81B.

[27] Antoine Browaeys and Thierry Lahaye. Many-body physics with individually controlled Rydberg atoms. *Nature Physics*, 16(2):132–142, 2020. doi: 10.1038/s41567-019-0733-z. URL https://doi.org/10.1038/s41567-019-0733-z.

[28] M Saffman. Quantum computing with atomic qubits and Rydberg interactions: progress and challenges. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 49(20): 202001, oct 2016. doi: 10.1088/0953-4075/49/20/202001. URL https://doi.org/10.1088/0953-4075/49/20/202001.

[29] Henning Labuhn, Daniel Barredo, Sylvain Ravets, Sylvain de Léséleuc, Tommaso Macrì, Thierry Lahaye, and Antoine Browaeys. Tunable two-dimensional arrays of single Rydberg atoms for realizing quantum Ising models. *Nature*, 534(7609):667–670, 2016. doi: 10.1038/nature18274. URL https://doi.org/10.1038/nature18274.

[30] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S. Zibrov, Manuel Endres, Markus Greiner, Vladan Vuletić, and Mikhail D. Lukin. Probing many-body dynamics on a 51-atom quantum simulator. *Nature*, 551(7682):579–584, 2017. doi: 10.1038/nature24622. URL https://doi.org/10.1038/nature24622.

[31] Brent N. Clark, Charles J. Colbourn, and David S. Johnson. Unit disk graphs. *Discrete Mathematics*, 86(1):165–177, 1990. ISSN 0012-365X. doi: https://doi.org/10.1016/0012-365X(90)90358-O. URL https://www.sciencedirect.com/science/article/pii/0012365X90903580.

[32] Fabian Kuhn, Thomas Moscibroda, and Rogert Wattenhofer. Unit disk graph approximation. In *Proceedings of the 2004 Joint Workshop on Foundations of Mobile Computing*, DIALM-POMC '04, pages 17–23, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581139217. doi: 10.1145/1022630.1022634. URL https://doi.org/10.1145/1022630.1022634.

[33] Heinz Breu and David G. Kirkpatrick. Unit disk graph recognition is NP-hard. *Computational Geometry*, 9(1):3–24, 1998. ISSN 0925-7721. doi: https://doi.org/10.1016/S0925-7721(97)00014-X. URL https://www.sciencedirect.com/science/article/pii/S092577219700014X. Special Issue on Geometric Representations of Graphs.

[34] Clément de Terrasson de Montleau, Victor Drouin-Touchette, and Wesley Coelho. Interaction-aware embedding optimization for neutral-atom quantum processors, 2026. URL https://arxiv.org/abs/2610.04068.

[35] Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics*, 65(3):239–362, 2016. doi: 10.1080/00018732.2016.1198134. URL https://doi.org/10.1080/00018732.2016.1198134.

[36] Anton Quelle, Mauro Mendizabal Pico, Kemal Bidzhiev, Stefano Grava, Elie Merhej, and Pablo le Henaff. Efficient emulation of neutral atom quantum hardware. 2025. URL https://arxiv.org/abs/2510.09813.

[37] Henrique Silvério, Sebastián Grijalva, Constantin Dalyac, Lucas Leclerc, Peter J. Karalekas, Nathan Shammah, Mourad Beji, Louis-Paul Henry, and Loïc Henriet. Pulser: An open-source package for the design of pulse sequences in programmable neutral-atom arrays. *Quantum*, 6:629, 2022. doi: 10.22331/q-2022-01-24-629. URL https://doi.org/10.22331/q-2022-01-24-629.

[38] Fresnel: A programmable neutral-atom quantum processor. portal.pasqal.cloud/devices/FRESNEL_CAN1, 2025. Fresnel Canada QPU, 100+ qubit register.

[39] Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. In *AAAI Workshop on Deep Learning on Graphs: Methods and Applications*, 2021. URL https://arxiv.org/abs/2012.09699.

[40] Vijay Prakash Dwivedi et al. Graph neural networks with learnable structural and positional representations. In *ICLR*, 2022. URL https://arxiv.org/abs/2110.07875.

[41] Derek Lim et al. Sign and basis invariant networks for spectral graph representation learning. In *ICLR*, 2023. URL https://arxiv.org/abs/2202.13013.

[42] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, pages 3391–3401. Curran Associates, Inc., 2017. URL https://arxiv.org/abs/1703.06114.

[43] William K Wootters. A wigner-function formulation of finite-state quantum mechanics. *Annals of Physics*, 176(1):1–21, 1987. URL https://www.sciencedirect.com/science/article/abs/pii/000349168790176X.

[44] Masaya Kunimi, Kazuma Nagao, Shimpei Goto, and Ippei Danshita. Performance evaluation of the discrete truncated wigner approximation for quench dynamics of quantum spin systems with long-range interactions. *Physical Review Research*, 3(1):013060, 2021. URL https://link.aps.org/doi/10.1103/PhysRevResearch.3.013060.

[45] Matthias Fey and Jan E. Lenssen. Fast graph representation learning with pytorch geometric, 2019. URL https://arxiv.org/abs/1903.02428.

[46] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006. URL https://jmlr.org/papers/v7/demsar06a.html.

[47] Alessio Benavoli, Giorgio Corani, Janez Demšar, and Marco Zaffalon. Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis. *Journal of Machine Learning Research*, 18(77):1–36, 2017. URL https://jmlr.org/papers/v18/16-305.html.

[48] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015. URL https://arxiv.org/abs/1505.00387.

[49] John Arevalo, Thamar Solorio, Manuel Montes-y Gómez, and Fabio A. González. Gated multimodal units for information fusion. In *Proceedings of the International Conference on Learning Representations Workshop*, 2017. URL https://arxiv.org/abs/1702.01992.

[50] N. Šibalić, J.D. Pritchard, C.S. Adams, and K.J. Weatherill. ARC: An open-source library for calculating properties of alkali Rydberg atoms. *Computer Physics Communications*, 220: 319–331, 2017. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2017.06.015. URL https://www.sciencedirect.com/science/article/pii/S0010465517301972.

[51] Marcos Rigol, Vanja Dunjko, and Maxim Olshanii. Thermalization and its mechanism for generic isolated quantum systems. *Nature*, 452(7189):854–858, 2008. doi: 10.1038/nature06838. URL https://doi.org/10.1038/nature06838.

[52] Rahul Nandkishore and David A. Huse. Many-body localization and thermalization in quantum statistical mechanics. *Annual Review of Condensed Matter Physics*, 6 (Volume 6, 2015):15–38, 2015. ISSN 1947-5462. doi: https://doi.org/10.1146/annurev-conmatphys-031214-014726. URL https://www.annualreviews.org/content/journals/10.1146/annurev-conmatphys-031214-014726.

[53] Dmitry A. Abanin, Ehud Altman, Immanuel Bloch, and Maksym Serbyn. Colloquium: Many-body localization, thermalization, and entanglement. *Rev. Mod. Phys.*, 91:021001, May 2019. doi: 10.1103/RevModPhys.91.021001. URL https://link.aps.org/doi/10.1103/RevModPhys.91.021001.

[54] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić. Weak ergodicity breaking from quantum many-body scars. *Nature Physics*, 14(7):745–749, 2018. doi: 10.1038/s41567-018-0137-5. URL https://doi.org/10.1038/s41567-018-0137-5.

[55] Takashi Mori, Tatsuhiko N Ikeda, Eriko Kaminishi, and Masahito Ueda. Thermalization and prethermalization in isolated quantum systems: a theoretical overview. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 51(11):112001, may 2018. doi: 10.1088/1361-6455/aabcdf. URL https://doi.org/10.1088/1361-6455/aabcdf.

[56] Sanjay Moudgalya, B Andrei Bernevig, and Nicolas Regnault. Quantum many-body scars and hilbert space fragmentation: a review of exact results. *Reports on Progress in Physics*, 85 (8):086501, jul 2022. doi: 10.1088/1361-6633/ac73a0. URL https://doi.org/10.1088/1361-6633/ac73a0.

[57] Maksym Serbyn, Dmitry A. Abanin, and Zlatko Papić. Quantum many-body scars and weak breaking of ergodicity. *Nature Physics*, 17(6):675–685, 2021. doi: 10.1038/s41567-021-01230-2. URL https://doi.org/10.1038/s41567-021-01230-2.

[58] J. M. Deutsch. Quantum statistical mechanics in a closed system. *Phys. Rev. A*, 43:2046–2049, Feb 1991. doi: 10.1103/PhysRevA.43.2046. URL https://link.aps.org/doi/10.1103/PhysRevA.43.2046.

[59] Mark Srednicki. Chaos and quantum thermalization. *Phys. Rev. E*, 50:888–901, Aug 1994. doi: 10.1103/PhysRevE.50.888. URL https://link.aps.org/doi/10.1103/PhysRevE.50.888.

[60] Norman F. Ramsey. Thermodynamics and statistical mechanics at negative absolute temperatures. *Phys. Rev.*, 103:20–28, Jul 1956. doi: 10.1103/PhysRev.103.20. URL https://link.aps.org/doi/10.1103/PhysRev.103.20.

[61] S. Braun, J. P. Ronzheimer, M. Schreiber, S. S. Hodgman, T. Rom, I. Bloch, and U. Schneider. Negative absolute temperature for motional degrees of freedom. *Science*, 339(6115): 52–55, 2013. doi: 10.1126/science.1227831. URL https://www.science.org/doi/abs/10.1126/science.1227831.

[62] Lars Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65:117–149, Feb 1944. doi: 10.1103/PhysRev.65.117. URL https://link.aps.org/doi/10.1103/PhysRev.65.117.

[63] G. H. Wannier. Antiferromagnetism. the triangular ising net. *Phys. Rev.*, 79:357–364, Jul 1950. doi: 10.1103/PhysRev.79.357. URL https://link.aps.org/doi/10.1103/PhysRev.79.357.

[64] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

[65] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019. URL https://arxiv.org/abs/1711.05101.

[66] John K. Kruschke. Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2):270–280, 2018. doi: 10.1177/2515245918771304. URL https://doi.org/10.1177/2515245918771304.

[67] Reinhard Diestel. *Graph Theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, 5 edition, 2017. URL https://link.springer.com/book/10.1007/978-3-662-70107-2.

[68] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 4 edition, 2013. URL https://www.press.jhu.edu/books/title/10678/matrix-computations.

[69] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, 2002. URL https://dl.acm.org/doi/book/10.5555/559923.

[70] Christopher K. I. Williams and Matthias Seeger. Using the Nyström method to speed up kernel machines. In *Advances in Neural Information Processing Systems 13 (NeurIPS)*, pages 682–688, 2001. URL https://proceedings.neurips.cc/paper/2000/hash/19de10adbaa1b2ee13f77f679fa1483a-Abstract.html.

[71] Damien Garreau, Wittawat Jitkrittum, and Motonobu Kanagawa. Large sample analysis of the median heuristic. *arXiv preprint arXiv:1707.07269*, 2017. URL https://arxiv.org/abs/1707.07269.

[72] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

[73] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. URL https://jmlr.org/papers/v12/pedregosa11a.html.

[74] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. doi: 10.1007/s11222-007-9033-z. URL https://doi.org/10.1007/s11222-007-9033-z.

[75] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2 edition, 2013. URL https://link.springer.com/book/10.1007/978-1-4612-0653-8.

A Equivalence between Rydberg and Ising Hamiltonian

In this appendix, we provide supporting information on the relevant neutral-atom Hamiltonian introduced in Sec. 2.1. In analog neutral-atom quantum simulators, each atom realizes an effective spin-1/2 degree of freedom spanned by a stable ground state $|0\rangle$ and a highly excited Rydberg state $|1\rangle$. In typical $^{87}\mathrm{Rb}$ implementations, these states may be chosen as $|0\rangle = \left|5S_{1/2}, F=2, m_F=2\right\rangle$ and $|1\rangle = \left|60S_{1/2}, m_J=1/2\right\rangle$. The Rydberg excitation gives rise to strong interactions between atoms occupying the $|1\rangle$ state. For the Rydberg levels and interatomic distances relevant here, these interactions are well approximated by a repulsive van der Waals potential, $V(\boldsymbol{r}) \sim |\boldsymbol{r}|^{-6}$, acting over micrometer-scale distances. The resulting dynamics of a register of N atoms are described by the Rydberg Hamiltonian (1), cf. Refs. [26–30]. The Hamiltonian acts on a 2^N-dimensional Hilbert space spanned by computational basis states $|\boldsymbol{b}\rangle = |b_1 b_2 \cdots b_N\rangle$ with $b_i \in \{0,1\}$. The operators satisfy $\hat{n}_i |b_i\rangle = b_i |b_i\rangle$, $\hat{\sigma}_i^x |b_i\rangle = |1-b_i\rangle$ and $\hat{\sigma}_i^z |b_i\rangle = (-1)^{b_i} |b_i\rangle$, with $\hat{n}_i = \left(1-\hat{\sigma}_i^z\right)/2$. Naturally, Eq. (1) can be rewritten in a conventional Ising form,

$$\hat{\mathcal{H}}(\mathbb{J}, \Omega, \delta) = \frac{1}{4}\sum_{i<j} J_{ij}\,\hat{\sigma}_i^z\hat{\sigma}_j^z + \hbar\sum_i \left[(\Omega/2)\,\hat{\sigma}_i^x + h_i^z\hat{\sigma}_i^z\right] + \bar{E}, \tag{6}$$

where the local longitudinal field h_i^z is given by $h_i^z = (\delta/2) - \bar{J}_i/(4\hbar)$ with $\bar{J}_i = \sum_{j=1}^N \bar{\delta}_{i,j} J_{ij}$ denoting the net interaction strength on atom i. The constant energy shift $\bar{E}$, given by

$$\bar{E} = \mathrm{tr}\left\{\hat{\mathcal{H}}\right\}/2^N = (N\bar{J}/8) - (\hbar N\delta/2), \tag{7}$$

coincides with the infinite-temperature average. The average interaction strength is $\bar{J} = \sum_{i=1}^N \bar{J}_i/N$. The positive van-der-Waals coupling strength $J_{ij} = C_6/r_{ij}^6$ is set by the inter-atomic distance r_{ij} between atoms i and j, and the Rydberg coefficient for the n=60 manifold used in this work evaluates to $C_6/\hbar = 8.66\times10^5\ \mu\mathrm{m}^6\,\mathrm{rad}/\mu\mathrm{s}$ [50]. The transverse driving Ω (Rabi frequency) and global detuning δ are set by the laser fields; their values are discussed in Sec. 2.3 and in App. C. Since $\hat{\mathcal{H}}(\mathbb{J}, \Omega, \delta)$ is time-independent during each experiment, the evolution corresponds to a quench,

leading to the time-evolved state $|\psi\left(t\right)\rangle = \exp\left[-\mathrm{i}(t/\hbar)\,\hat{\mathcal{H}}\left(\mathbb{J},\Omega,\delta\right)\right]\,|\psi\left(0\right)\rangle$, leading to the desired correlation matrix

$$C_{i,j}\left(t\right) = A^{nn}_{i,j}\left(t\right) - n_i\left(t\right)\,n_j\left(t\right), \tag{8}$$

where

$$A^{nn}_{i,j}\left(t\right) = \left\langle\psi\left(t\right)\right|\hat{n}_i\,\hat{n}_j\left|\psi\left(t\right)\right\rangle, \qquad n_i\left(t\right) = \left\langle\psi\left(t\right)\right|\hat{n}_i\left|\psi\left(t\right)\right\rangle, \tag{9}$$

see Eq. (2) where we used the Heisenberg operator $\hat{n}_i(t)$. The respective value ranges are $C_{i,j} \in [-1/4, 1/4]$ and $n_i \in [0,1]$. Since $\hat{n}_i^2 = \hat{n}_i$, it follows for the diagonal elements $C_{i,i}\left(t\right) = n_i\left(t\right)\left[1 - n_i\left(t\right)\right]$, i.e. the diagonal element $C_{i,i}\left(t\right) \in [0, 1/4]$ effectively encodes the site occupations $n_i\left(t\right)$. Importantly, both $\hat{n}_i$ and $\hat{n}_i\hat{n}_j$ are diagonal in the measurement basis, which allows $n_i\left(t\right)$ and $C_{i,j}\left(t\right)$ to be estimated efficiently by repeated projective readout of the atomic register.

B Thermalization

In this appendix, we collect the thermalization arguments which are implicitly used in Sec. 2.3, and explicitly used in App. C. We first recall the diagonal-ensemble description of long-time observables after a quench under a time-independent Hamiltonian. We then explain how the eigenstate thermalization hypothesis (ETH) motivates a canonical thermal approximation to the diagonal ensemble. Finally, we derive the simple criterion to determine the sign of the effective inverse temperature. The scheme is then applied to the Rydberg Hamiltonian (1) [or (6)] in App. C.

Let $\hat{A}$ be a generic observable and let the system evolve under a time-independent Hamiltonian $\hat{\mathcal{H}}$, $|\psi(t)\rangle = e^{-\mathrm{i}\hat{\mathcal{H}}t/\hbar}|\psi(0)\rangle$. We denote the eigenstates and eigenenergies of $\hat{\mathcal{H}}$ by $\hat{\mathcal{H}}|E_\alpha\rangle = E_\alpha|E_\alpha\rangle$, and expand the initial state as $|\psi(0)\rangle = \sum_\alpha c_\alpha|E_\alpha\rangle$ with $c_\alpha = \langle E_\alpha|\psi(0)\rangle$.

The expectation value of $\hat{A}$ can then be written as

$$A(t) = \langle\psi(t)|\hat{A}|\psi(t)\rangle \qquad \equiv A^{\mathrm{diag}} + A^{\text{off-diag}}(t), \tag{10}$$

where

$$A^{\mathrm{diag}} = \sum_\alpha |c_\alpha|^2 A_{\alpha\alpha}, \qquad A^{\text{off-diag}}(t) = \sum_{\alpha\neq\alpha'} c^*_\alpha c_{\alpha'} e^{i\omega_{\alpha,\alpha'}t} A_{\alpha\alpha'}, \tag{11}$$

with $A_{\alpha\alpha'} = \langle E_\alpha|\hat{A}|E_{\alpha'}\rangle$ and $\omega_{\alpha,\alpha'} = (E_\alpha - E_{\alpha'})/\hbar$. The first term is the time-independent diagonal-ensemble value. The second term contains oscillatory off-diagonal contributions.

For a finite closed quantum system, $A^{\text{off-diag}}(t)$ does not generally converge pointwise to zero: finite-size revivals and degeneracies can occur. What is relevant for equilibration is instead dephasing. Under generic non-resonance assumptions, the off-diagonal contribution vanishes in the infinite-time average. Experimentally, we approximate the long-time value by the finite-time average

$$\overline{A}(t_1,t_0) = \frac{1}{t_1 - t_0}\int_{t_0}^{t_1} dt'\, A(t'). \tag{12}$$

In the absence of long-lived nonergodic effects, one expects $\overline{A}(t_1,t_0)$ to approach A^{diag} as the averaging window becomes large. In practice, this convergence can be delayed or prevented by mechanisms that obstruct or weakly violate ergodicity, such as integrability, many-body localization, Hilbert-space fragmentation, many-body scars, or prethermalization [51–53, 30, 54–57]. Note that $C_{ij}(t)$ [Eq. (8)] is quadratic in the expectation values $n_i(t)$, and time averaging does not commute with taking products. Hence, the time average $\overline{C_{ij}}(t_1,t_0)$ does not in general converge to the diagonal-ensemble value $C^{\mathrm{diag}}_{ij} = A^{\mathrm{nn,diag}}_{ij} - n^{\mathrm{diag}}_i n^{\mathrm{diag}}_j$; the two differ by the temporal covariance $\overline{n_i n_j} - \overline{n_i}\,\overline{n_j}$. Instead, we use

$$C^{\infty}_{ij}(t_1,t_0) = \overline{A^{\mathrm{nn}}_{ij}}(t_1,t_0) - \overline{n_i}(t_1,t_0)\,\overline{n_j}(t_1,t_0), \tag{13}$$

which converges to C^{diag}_{ij}.

The diagonal ensemble still requires knowledge of the full eigensystem of $\hat{\mathcal{H}}$, which is generally unavailable for large many-body systems. The eigenstate thermalization hypothesis provides a

practical interpretation. If the dynamics are sufficiently ergodic for ETH to apply, the diagonal matrix elements of a few-body observable are smooth functions of energy, up to finite-size fluctuations,

$$A_{\alpha\alpha} = \langle E_\alpha|\hat{A}|E_\alpha\rangle \simeq \langle\hat{A}\rangle_{\rm micro}(E_\alpha), \tag{14}$$

with the microcanonical expectation value $\langle\hat{A}\rangle_{\rm micro}(E)$. The off-diagonal matrix elements are suppressed by the many-body entropy [58, 59, 51, 35]. If, in addition, the weights $|c_\alpha|^2$ have support in a sufficiently narrow energy window around the initial energy $E_{\rm init} = \langle\psi(0)|\hat{\mathcal{H}}|\psi(0)\rangle$, then the diagonal-ensemble value can be approximated by the microcanonical expectation value at $E_{\rm init}$, $A^{\rm diag} \simeq \langle\hat{A}\rangle_{\rm micro}(E_{\rm init})$. For sufficiently large systems, the microcanonical and canonical ensembles are equivalent for local observables. We therefore use the canonical thermal approximation at an effective inverse temperature $\beta(E_{\rm init})$,

$$A^{\rm diag} \simeq \langle\hat{A}\rangle_{\rm th}\left[\beta(E_{\rm init})\right], \tag{15}$$

where

$$A^{\rm th} = \langle\hat{A}\rangle_{\rm th}[\beta] = {\rm tr}\left\{\hat{A}e^{-\beta\hat{\mathcal{H}}}\right\}/Z[\beta], \qquad Z[\beta] = {\rm tr}\left\{e^{-\beta\hat{\mathcal{H}}}\right\}. \tag{16}$$

The inverse temperature is not a free parameter. It is fixed by energy conservation,

$$\langle\hat{\mathcal{H}}\rangle_{\rm th}[\beta] = E_{\rm init}, \qquad E_{\rm init} = \langle\psi(0)|\hat{\mathcal{H}}|\psi(0)\rangle. \tag{17}$$

This approximation is used in Sec. 2.3 and App. C as a diagnostic tool for selecting quench parameters (Ω, δ). It is not assumed to hold throughout the full parameter space. In particular, finite-size effects, graph symmetries, proximity to integrable limits, many-body localization, Hilbert-space fragmentation, many-body scars, or prethermalization can obstruct or delay convergence to the thermal value [52, 53, 30, 54–57].

Since the Hilbert space is finite, the spectrum of $\hat{\mathcal{H}}$ is bounded. Consequently, the solution of Eq. (17) may have either positive or negative inverse temperature [60, 61]. A negative-temperature ensemble of $\hat{\mathcal{H}}$ can equivalently be written as a positive-temperature ensemble of the negated Hamiltonian,

$$e^{-\beta\hat{\mathcal{H}}} = e^{-|\beta|(-\hat{\mathcal{H}})}, \qquad \beta < 0. \tag{18}$$

Thus, for $\beta < 0$, the effective thermal description reverses the sign of all terms in the Hamiltonian. In particular, the sign of the Ising interaction is reversed in this effective thermal description: turning the repulsive AFM model into an attractive FM model.

Conveniently, the sign of β can be easily determined, knowing the initial energy $E_{\rm init}$ (17) and the infinite-temperature energy

$$\bar{E} = \langle\hat{\mathcal{H}}\rangle_{\rm th}[\beta = 0] = {\rm tr}\left\{\hat{\mathcal{H}}\right\}/\dim(\hat{\mathcal{H}}), \tag{19}$$

see also Eq. (7). For many Hamiltonians both quantities can be evaluated analytically. The thermal energy $\langle\hat{\mathcal{H}}\rangle_{\rm th}[\beta]$ is a monotonically decreasing function of β. Indeed,

$$\frac{\partial\langle\hat{\mathcal{H}}\rangle_{\rm th}}{\partial\beta} = -\left[\langle\hat{\mathcal{H}}^2\rangle_{\rm th} - \langle\hat{\mathcal{H}}\rangle_{\rm th}^2\right] = -\left\langle\left(\hat{\mathcal{H}} - \langle\hat{\mathcal{H}}\rangle_{\rm th}\right)^2\right\rangle_{\rm th} \leq 0. \tag{20}$$

The inequality is strict unless the thermal state has zero energy variance. Clearly, an initial energy $E_{\rm init} = \bar{E}$ implies a thermal state at infinite temperature with $\beta = 0$. Given the monotonicity (20), it also follows that $E_{\rm init} < \bar{E}$ implies a positive solution β, whereas $E_{\rm init} > \bar{E}$ yields a negative solution β:

$$\begin{aligned} E_{\rm init} < \bar{E} \quad &\Rightarrow \quad \beta(E_{\rm init}) > 0, \\ E_{\rm init} = \bar{E} \quad &\Rightarrow \quad \beta(E_{\rm init}) = 0, \\ E_{\rm init} > \bar{E} \quad &\Rightarrow \quad \beta(E_{\rm init}) < 0. \end{aligned} \tag{21}$$

Applying this criterion to the Rydberg Hamiltonian (1) where $\bar{E}$ evaluates to Eq. (7), and $E_{\rm init} = 0$, one obtains the classification shown in App. C, Eq. (23).

C Quench-parameter selection

In this appendix, we motivate the quench-parameter selection that leads to the AFM-like and FM-like Hamiltonians, $\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{AFM}}$ and $\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{FM}}$, used in this work, and defined in Eq. (4) in Sec. 2.3. The quench settings (Ω, δ) are chosen to systematically yield pronounced connected density correlations $C_{i,j}$ [Eq. (2)]. Hereby, the selection strategy is guided by thermalization arguments which are detailed in App. B. In particular, we exploit that the diagonal-ensemble value A^{diag} (11), experimentally approximated by the time average $\overline{A}(t_1, t_0)$ (12), can be related to the canonical thermal expectation value A^{th} (16) under the assumption of thermalization. The main idea is illustrated on a small molecular graph in Fig. 2 and Fig. 4.

First, we apply the criterion (21) which determines the sign of the effective temperature on the Rydberg Hamiltonian [Eq. (1) or (6)] with an initial all-zero product state $|\psi(0)\rangle = |\mathbf{0}\rangle$ where $|\mathbf{0}\rangle \equiv |0\rangle^{\otimes N}$. The initial energy is $E_{\mathrm{init}} = \langle \mathbf{0}|\hat{\mathcal{H}}|\mathbf{0}\rangle = 0$, and the infinite-temperature energy $\bar{E}$ is evaluated in Eq. (7). Both energies equal each other, i.e. $E_{\mathrm{init}} = \bar{E}$, for the detuning value

$$\delta_\infty = \bar{J}/(4\hbar) = \sum_{i \neq j} J_{ij}/(4N\hbar), \tag{22}$$

such that the criterion (21) becomes

$$\begin{aligned} \delta < \delta_\infty \quad &\Rightarrow \quad \beta > 0 \quad &&\leftrightarrow \text{ AFM-like}, \\ \delta = \delta_\infty \quad &\Rightarrow \quad \beta = 0, \\ \delta > \delta_\infty \quad &\Rightarrow \quad \beta < 0 \quad &&\leftrightarrow \text{ FM-like}. \end{aligned} \tag{23}$$

Thus, increasing the detuning moves the thermal interpretation from an AFM-like positive-temperature regime through infinite temperature into an FM-like negative-temperature regime. This classification does not imply that the system thermalizes for all values of (Ω, δ); rather, it provides a useful organizing principle in the parameter regimes relevant for our feature-generation protocol for identifying regimes with robust AFM- or FM-like correlations.

For convenience, we repeat the two introduced Hamiltonians from Eq. (4) in Sec. 2.3

$$\hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{AFM}} = \hat{\mathcal{H}}\left(\mathbb{J}^{\mathcal{G}},\, \Omega = 0.3\, J_{\mathrm{NN}}/\hbar,\, \delta = 0\right), \quad \hat{\mathcal{H}}^{\mathcal{G}}_{\mathrm{FM}} = \hat{\mathcal{H}}\left(\mathbb{J}^{\mathcal{G}},\, \Omega = 0.5\, J_{\mathrm{NN}}/\hbar,\, \delta = 2\delta^{\mathcal{G}}_\infty\right), \tag{24}$$

see Eqs. (1), (6), (3) for definitions. To emphasize the graph dependence of $\delta_\infty = \delta_\infty\left(\mathbb{J}^{\mathcal{G}}\right)$ (22) we added the superscript $\mathcal{G}$ in Eq. (4) and (24).

The choice $\delta = 0$ places the AFM setting below δ_∞ and hence in the positive-temperature regime of Eq. (23). The choice $\delta = 2\delta_\infty$ places the FM setting in the negative-temperature regime and it has an additional useful interpretation: it cancels the spatially averaged longitudinal field. Indeed, using $h^z_i = \delta/2 - \bar{J}_i/(4\hbar)$ from Eq. (6), one finds $\sum_i h^z_i = 0$ is solved by $\delta = 2\delta_\infty$. Thus, the FM pulse setting approximately realizes a field-balanced transverse-field Ising regime, while the AFM setting probes the opposite side of the infinite-temperature line.

Figure 2 in the main text confirms the expected qualitative behavior on a small molecular graph. In the FM setting, neighboring atoms develop positive off-diagonal correlations, while in the AFM setting the corresponding correlations are negative. The molecule is small enough for exact diagonalization, which allows us to compute both the diagonal-ensemble values $C^{\mathrm{diag}}_{i,j}$ [Eq. (11)] and the thermal estimates $C^{\mathrm{th}}_{i,j}$ [Eq. (16)]. The finite-time averages $C^{\infty}_{i,j}(t_1, t_0)$ [Eq. (13)] agree quantitatively well with the diagonal-ensemble values, indicating that the chosen measurement window captures the dominant long-time correlation signal. The agreement with the thermal values is qualitative rather than quantitative, which is not surprising for such a small system, where finite-size effects, and the proximity to the $\Omega = 0$ integrable limit can prevent fully ergodic dynamics. Nevertheless, the thermal picture correctly predicts the AFM/FM character of the correlations and is therefore useful as a parameter-selection guide.

We next show that the chosen pulse settings not only produce the desired AFM- and FM-like correlation signs, but also lie in regimes with pronounced correlation magnitude. To quantify this, we scan the parameter space $\Omega/(J_{\mathrm{NN}}/\hbar) \in [0, 1]$ and $\delta/\delta_\infty \in [-1, 3]$, and compute the off-diagonal correlation weight per node, $\mathcal{C}_{\mathrm{off}} = \sum_{i \neq j} |C_{i,j}|/N$. This quantity serves as a simple proxy for the amount of nonlocal correlation information available in the correlation matrix. We evaluate this proxy

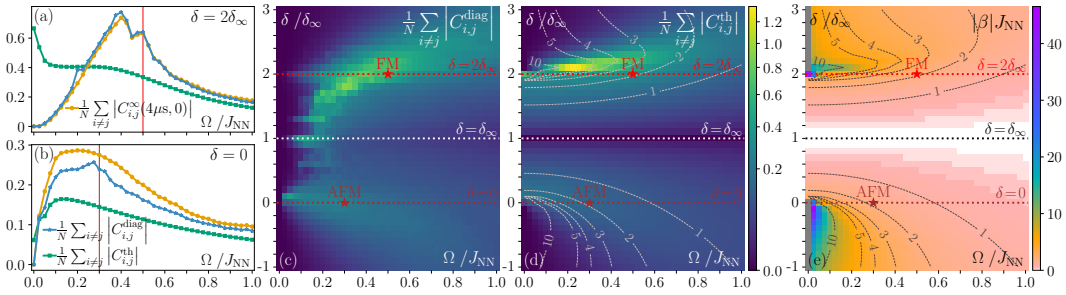


Figure 4: Exploration of the (Ω, δ) parameter space for the same molecular graph as in Fig. 2, initialized in $|\psi(0)\rangle = |\mathbf{0}\rangle$. The plotted quantity is the off-diagonal correlation weight per node, $\mathcal{C}_{\rm off} = \sum_{i\neq j}|C_{i,j}|/N$, used as a proxy for the amount of correlation information contained in the feature matrix. Panels (a)–(b) show the proxy as a function of Ω at the two relevant detunings: $\delta = 2\delta_\infty$ for the FM setting and $\delta = 0$ for the AFM setting, cf. Eq. (4) or (24), with the selected Ω values indicated by vertical lines. Panels (c)–(d) compare the diagonal-ensemble [Eq. (11)] and thermal estimates [Eq. (16)] over the full parameter window. Panel (e) shows the absolute value of the effective inverse temperature $|\beta|$ obtained from Eq. (17). The proxy for the time average based on $C^{\infty}_{i,j}(t_1, t_0)$ [Eq. (13)] is only shown in panels (a)–(b). The line $\delta = \delta_\infty$ separates the positive-temperature AFM-like regime from the negative-temperature FM-like regime. For this graph, $\delta_\infty = 0.53\, J_{\rm NN}/\hbar$ [cf. Eq. (23)]. Gray clipping in panel (e) indicates $|\beta| = \infty$.

using the finite-time average (12), the diagonal-ensemble value (11), and the thermal expectation value (16). The results are shown in Fig. 4. We omit the full phase diagram of the finite-time average, since it is qualitatively very similar to the diagonal-ensemble result in Fig. 4(c).

We first interpret the thermal phase diagram in Fig. 4(d), together with the corresponding effective inverse temperature shown in Fig. 4(e). As predicted by Eq. (23), the line $\delta = \delta_\infty$ separates the AFM-like regime at positive temperature from the FM-like regime at negative temperature. This separation is largely reproduced by the diagonal-ensemble estimate in Fig. 4(c), except at very small Ω, where the dynamics are more strongly affected by constrained subspaces and thermalization becomes less reliable.

On the FM side, the strongest correlations occur close to the regime where the average longitudinal field vanishes. In this regime, the model approaches a field-balanced transverse-field Ising model with effectively ferromagnetic thermal correlations. For the graph considered in Fig. 4, the thermal correlation maximum appears near $\Omega/(J_{\rm NN}/\hbar) \approx 0.23$ and $\delta/\delta_\infty \approx 2.1$, which we interpret as a finite-size thermal crossover. The existence of an enhanced-correlation regime is not tied to this particular molecular geometry: for a broad class of embedded graphs, one expects a finite-size thermal crossover at a graph-dependent effective temperature, provided the thermal description remains applicable. We therefore choose $\delta = 2\delta_\infty$ to place each graph near this crossover regime. The thermal maximum, however, occurs at relatively small Ω, where the finite-time dynamics and the diagonal ensemble show stronger deviations from the thermal estimate. We therefore select the slightly larger value $\Omega = 0.5\, J_{\rm NN}/\hbar$ [cf. Eq. (4) or (24)], which remains in a regime of enhanced correlations while producing more robust dynamical relaxation.

The crossover scale depends on the graph geometry through $\mathbb{J}^{\mathcal{G}}$ (3). For orientation, the effective crossover temperature observed in Fig. 4(e), $|\beta_{\rm cross}|J_{\rm NN} \approx 7$, lies between familiar classical Ising limits. In the nearest-neighbor field-free Ising model, the one-dimensional chain has no finite-temperature transition ($\beta_{\rm crit} = \infty$), whereas the square and triangular lattices have $\beta_{\rm crit} J_{\rm NN} \approx 1.76$ and $\beta_{\rm crit} J_{\rm NN} \approx 1.10$, respectively, in the present convention [62, 63]. These values include the factor-of-four conversion between the Rydberg interaction coefficient $J_{\rm NN}$ and the conventional Ising coupling in Eq. (6). The presence of the transverse field further shifts the crossover scale. These values should be understood only as reference scales, since the molecular graph is finite and irregular and the present model includes a transverse field.

The AFM setting is chosen more empirically. We fix $\delta = 0$, which places the system on the positive-temperature side at a detuning distance from δ_∞ comparable to the FM choice $\delta = 2\delta_\infty$. We then choose $\Omega = 0.3\, J_{\rm NN}/\hbar$, which yields a similar effective temperature scale and lies in a region with pronounced negative off-diagonal correlations, as shown in Fig. 4(c). Together, the two settings in

Eq. (4) or (24) provide complementary quantum feature maps: one dominated by FM-like positive correlations and one dominated by AFM-like negative correlations.

Sensitivity to the AFM working point and to the measurement window. The AFM point was selected without task-level tuning. The scans of Fig. 4 show that $\Omega = 0.3\, J_{\mathrm{NN}}/\hbar$ lies inside a broad AFM-like enhanced-correlation region rather than at an isolated optimum: for $\hbar\Omega/J_{\mathrm{NN}} \lesssim 0.4$, this region extends over a finite neighborhood of $\delta = 0$ on the positive-temperature side of the separatrix $\delta = \delta_\infty$. Within it, the diagonal-ensemble correlations vary smoothly with (Ω, δ) and remain AFM-like, with predominantly short-range negative correlations. Small changes of (Ω, δ) thus change the correlation magnitudes but not the regime, which is set by the sign of the effective interaction [Eq. (23)]. The AFM/FM complementarity is not specific to molecules: the working points are computed per graph from $\delta_\infty^{\mathcal{G}}$, and chemistry enters only through which graphs admit a unit-disk embedding. For denser graphs, $\delta_\infty^{\mathcal{G}}$ grows and the FM point $\delta = 2\delta_\infty^{\mathcal{G}}$ moves to larger detuning, which rescales the working points without changing the physical picture; the ladder graphs of Sec. 5 are a non-molecular example. These scans do not establish that every intrinsic or downstream score is stable across Ω; this would require regenerating the features and retraining all models. The exact sampling times are not expected to be critical either. For a thermalizing system, the information in $A(t)$ is mainly carried by its diagonal-ensemble value A^{diag} [Eq. (11)], which the window $[0.1, 4]\,\mu$s approximates well (Fig. 2). A suitable window must extend beyond the relaxation time, so that the time average approaches A^{diag}, and must contain enough samples to average out the oscillatory off-diagonal contributions; consistently, a static component dominates the PCA of App. D.2, and the four-statistic summaries retain most of the downstream performance.

D Emulation and QPU execution

This appendix describes the settings used for emulation, neutral-atom hardware execution, and provides analyses of QGFs produced by our pipeline.

D.1 Emulation budget and measurement-time window

The emulator (Sec. 2.4) uses a state-vector backend for graphs with $N < 25$ atoms (numerically exact expectation values from the full 2^N-dimensional state) and the matrix-product-state backend for larger graphs (controlled bond-dimension truncation). Both interface with a common library so emulated and hardware pulse sequences follow the same protocol. Total cost: $\sim$800 GPU-days on an NVIDIA A100 for $\mathrm{ZINC}^*_{\mathrm{subset}}$ (two pulses) and $\sim$14 GPU-days for MUTAG, which is roughly $60\times$ smaller.

We record observables at $N_t = 20$ equally spaced times t_ν with $\nu \in \{1, \ldots, N_t\}$ where $t_1 = 100\,\mathrm{ns}$ and $t_{N_t} = 4\,\mu$s. The lower bound is set by the onset of non-trivial many-body dynamics after the initial transverse-field-dominated regime; the upper bound is chosen long enough that the time-averaged correlations are well approximated.

D.2 PCA analysis of quantum correlation features

To better understand the structure of the quantum feature space, we perform a principal component analysis (PCA) on the normalised AFM and FM correlation features. Fig. 5 presents, for each pulse type and feature type diagonal and off-diagonal, the collection features across all graphs in the dataset, yielding matrices of shape $[N_{\mathrm{total}}, N_t]$ for diagonal (one row per node) and $[E_{\mathrm{total}}, N_t]$ for off-diagonal (one row per directed edge), where $N_t = 20$ is the number of simulation time steps.

One-body features $\mathcal{C}_i^{(1)}$ [Eq. (5)]. For both AFM and FM pulses, and on both datasets, the first principal component captures a dominant share of the total variance: approximately 80–83% for AFM and a comparable fraction for FM. This component corresponds to the average correlation level across timesteps, reflecting the time-averaged structural signature of each node. Reaching 90% explained variance requires 4–5 components, and 95% requires 9–10 components.

Two-body features $\mathcal{C}_{ij}^{(2)}$ [Eq. (5)]. The variance structure of two-body (pairwise) features is substantially more distributed. The first principal component accounts for 59–65% of the variance,

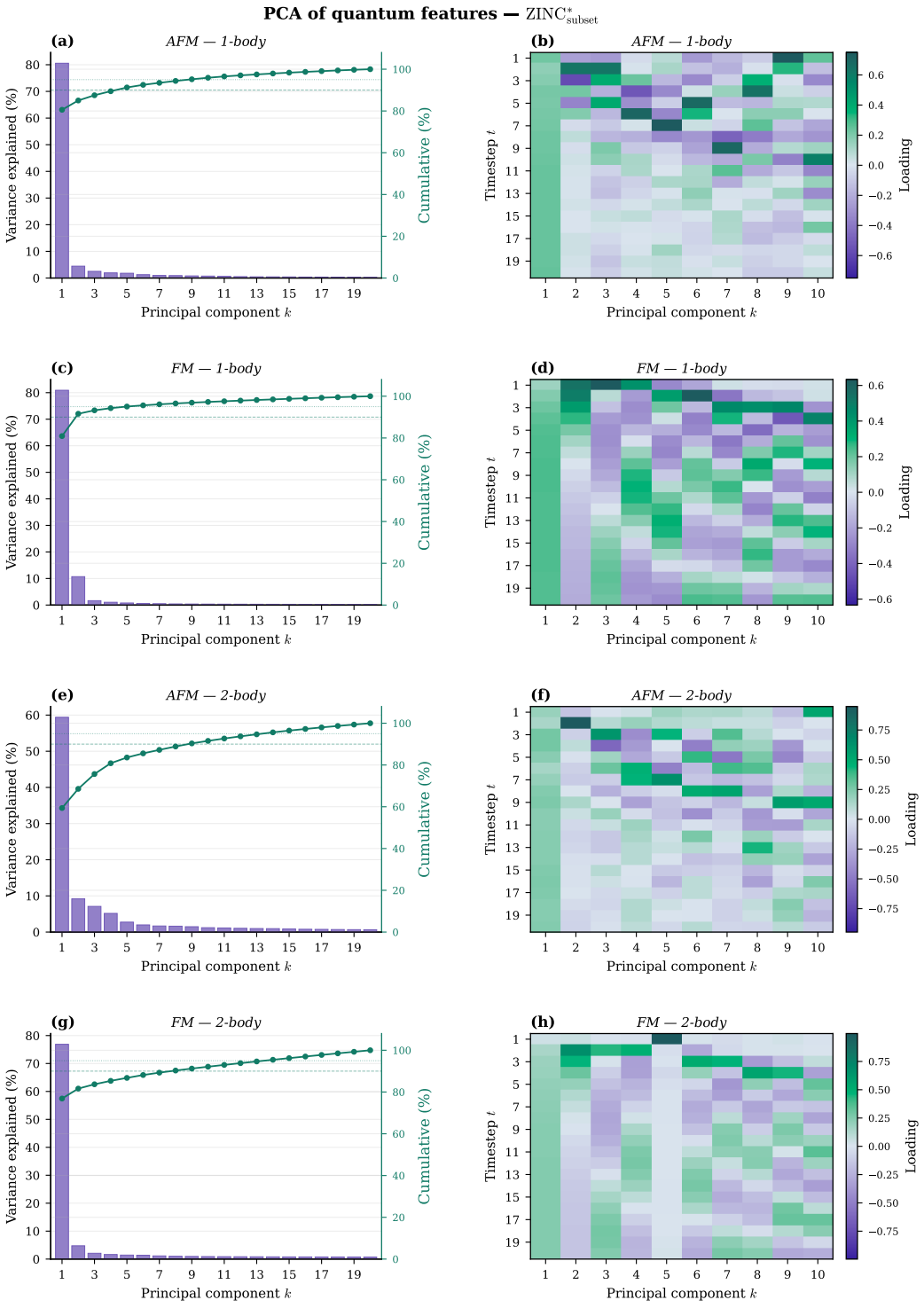


Figure 5: **PCA of quantum correlation features on** $\mathrm{ZINC}^{*}_{\mathrm{subset}}$**.** Each row shows a scree plot (left) and a loading heatmap (right) for a given pulse type and feature level. **(a,b)** AFM 1-body correlations $\boldsymbol{\mathcal{C}}_i^{(1)}$ [Eq. (5)]; **(c,d)** FM 1-body correlations; **(e,f)** AFM 2-body correlations $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ (edge-level); **(g,h)** FM 2-body correlations. In every case the first principal component captures the bulk of the variance (80%+ for 1-body, 60%+ for 2-body), indicating that time-averaged activation dominates the representation. The loading heatmaps reveal a structured temporal pattern: PC1 loads uniformly across all timesteps (static component), while PC2 and PC3 exhibit oscillatory sign alternations (dynamic components), consistently across both pulse types and datasets.

with the second and third components contributing 7–9% and 5–7% respectively. Reaching 90% explained variance requires 9–10 components, and 95% requires 14 components.

Interpretation. These results suggest that the quantum correlation features carry two distinct types of information:

1. A dominant *static* component (PC1), capturing the time-averaged correlation strength of each node or edge. This component is well approximated by the mean over timesteps and explains why the stat reduction $(\mu, \sigma, \min, \max)$ retains most of the performance in the PE benchmark.
2. A richer *dynamic* component (PC2 onwards), encoding temporal modulations of the correlations. This component is more prominent in two-body features than in one-body features, consistent with the fact that pairwise correlations encode distance-dependent oscillatory behaviour that varies across the time evolution.

Notably, this variance structure is nearly identical across the two datasets ($\text{ZINC}^*_{\text{subset}}$ and MUTAG), suggesting that the informational structure of the quantum features is intrinsic to the Hamiltonian simulation rather than dataset-dependent. This consistency supports the potential transferability of quantum PE across different molecular graph tasks.

D.3 Details on hardware runs

Readout-error correction. The readout-error associated with the QPU runs is corrected in a post-processing step. Hereby, one exploits the pre-calibrated values of false Rydberg detection $\epsilon = P_r\left(\text{observed} = |1\rangle \,\middle|\, \text{true} = |0\rangle\right)$ and false ground-state detection $\epsilon' = P_r\left(\text{observed} = |0\rangle \,\middle|\, \text{true} = |1\rangle\right)$; for the employed QPU the values are $\epsilon = 0.025$ and $\epsilon' = 0.1$. The observed and true qubit states are related through the 1-qubit confusion matrix A_{1q}, via

$$\begin{pmatrix} P\left(\text{true} = |0\rangle\right) \\ P\left(\text{true} = |1\rangle\right) \end{pmatrix} = A_{1q}^{-1} \begin{pmatrix} P\left(\text{obs} = |0\rangle\right) \\ P\left(\text{obs} = |1\rangle\right) \end{pmatrix}, \qquad A_{1q} = \begin{pmatrix} 1-\epsilon & \epsilon' \\ \epsilon & 1-\epsilon' \end{pmatrix}, \tag{25}$$

where $P\left(\text{obs} = |b\rangle\right)$ denotes the probability for observing the qubit in state $|b\rangle$. The true—readout-error corrected—occupation $\langle \hat{n}_i \rangle^{\text{true}} = P\left(\text{true} = |1_i\rangle\right)$ [cf. Eq. (8)] follows from the observed occupation $\langle \hat{n}_i \rangle^{\text{obs}} = P\left(\text{obs} = |1_i\rangle\right)$ by the simple relation $\langle \hat{n}_i \rangle^{\text{true}} = \sum_{k=1}^{2} \left(A_{1q}^{-1}\right)_{2,k} \left(v_i^{\text{n,obs}}\right)_k$ where $\boldsymbol{v}_i^{\text{n,obs}} = \left(1 - \langle \hat{n}_i \rangle^{\text{obs}}, \langle \hat{n}_i \rangle^{\text{obs}}\right)^T$. The two-qubit correlation observable $\langle \hat{n}_i \hat{n}_j \rangle^{\text{true}} = P\left(\text{true} = |1_i 1_j\rangle\right)$ [cf. Eq. (8)] is treated analogously. In extension to Eq. (25) we introduce the two-qubit confusion matrix $A_{2q} = A_{1q} \otimes A_{1q}$. The extended observation vector becomes

$$\boldsymbol{v}_{i,j}^{\text{nn,obs}} = \left(\langle (1-\hat{n}_i)(1-\hat{n}_j) \rangle^{\text{obs}}, \langle (1-\hat{n}_i)\hat{n}_j \rangle^{\text{obs}}, \langle \hat{n}_i (1-\hat{n}_j) \rangle^{\text{obs}}, \langle \hat{n}_i \hat{n}_j \rangle^{\text{obs}} \right)^T, \tag{26}$$

associated with observing the qubits in the states $(|0_i 0_j\rangle, |0_i 1_j\rangle, |1_i 0_j\rangle, |1_i 1_j\rangle)$. Finally, the readout-error corrected correlation follows from the relation $\langle \hat{n}_i \hat{n}_j \rangle^{\text{true}} = \sum_{k=1}^{4} \left(A_{2q}^{-1}\right)_{4,k} \left(v_{i,j}^{\text{nn,obs}}\right)_k$.

Shot budget estimate for MUTAG connected correlations. Using the Central Limit Theorem (CLT) for the simultaneous estimation of all one-body and two-body connected correlations at global confidence level $1 - \delta_\epsilon = 95\%$, we count $M = N(N+1)/2$ that is, N one-body observables together with $M - N = N(N-1)/2$ distinct off-diagonal correlations. Here Φ denotes the cumulative distribution function (CDF) of a standard normal random variable $Z \sim \mathcal{N}(0,1)$, i.e., $\Phi(x) = \mathbb{P}(Z \leq x)$, and Φ^{-1} is its quantile function. The corresponding Gaussian quantile is

$$k_N = \Phi^{-1}[1 - \delta_\epsilon/(2M)], \tag{27}$$

and we use $\delta_\epsilon = 0.05$. Fixing an absolute error tolerance ε and taking $v_{\max} = \sigma_{\max}^2$ as the maximal estimator variance (empirically estimated from the emulations performed), the required shot count per graph and per measurement time is

$$N_{\text{shots}}^{\text{CLT}}(N) = \left\lceil \frac{v_{\max}}{\varepsilon^2} k_N^2 \right\rceil. \tag{28}$$

With $v_{\max} = 0.06$ and $\varepsilon = 0.03$, this becomes $N_{\text{shots}}^{\text{CLT}}(N) = \lceil 66.67\, k_N^2 \rceil$. The choice $\varepsilon = 0.03$ is consistent with the natural scale of connected occupation correlations, which satisfy $|\mathcal{C}| \leq 1/4$; it therefore corresponds to an absolute resolution of about 12% of the full correlation range, sufficient to resolve medium-to-large local antiferromagnetic correlations.

For a standard MUTAG graph, taking $N \simeq 18$ gives $M = 171$ and $k_{18} \approx 3.62$ such that the shot count becomes $N_{\text{shots}}^{\text{CLT}}(18) = \lceil \frac{0.06}{0.03^2} \times k_{18}^2 \rceil = 874$.

QPU cost scaling. The acquisition time scales as $T_{\text{QPU}} \simeq G\, R\, N_t\, \bar{N}_{\text{shots}} / f_{\text{rep}}$, with G graphs, R quench regimes, N_t measurement times, $\bar{N}_{\text{shots}}$ the mean shot count per graph and time, and f_{rep} the effective repetition rate, including register loading, rearrangement and readout. The shot count grows only logarithmically with graph size ($N_{\text{shots}}^{\text{CLT}} = 734$ at $N = 10$, 901 at $N = 20$, 983 at $N = 28$), so the cost is dominated by f_{rep}. For the MUTAG run ($G = 188$, $R = 1$, $N_t = 4$, $\approx 6.5 \times 10^5$ shots in total), the ~ 7.5 QPU-days correspond to $f_{\text{rep}} \approx 1$ Hz. At this rate, a two-regime $\text{ZINC}_{\text{subset}}^{*}$ run ($G = 11{,}368$, $R = 2$, $N_t = 4$) would take ~ 950 days, against ~ 95 days at 10 Hz and ~ 10 days at 100 Hz; using all $N_t = 20$ times multiplies these figures by five. Three-dimensional registers would additionally improve the embedding quality, with fewer unwanted interactions between non-adjacent atoms.

D.4 Emulator–QPU concordance at t_4

We assess the agreement between idealized emulation and end-to-end QPU execution on the MUTAG dataset at the AFM operating point (Sec. 2.3), at the sampled measurement time t_4. The reported times, $t_{\text{QPU}} = 940$ ns and $t_{\text{emu}} = 715$ ns are different because, as the available max amplitude was lower in the QPU than in emulation, sampling times have to be rescaled accordingly (along with r_{NN}). All 188 MUTAG graphs admitting a 2D unit-disk embedding (Sec. 2.2) were measured under the shot budget of App. D.3 ($N_{\text{shots}} = 874$).

For each graph and each correlation type we evaluate the Spearman rank correlation ρ_s between QPU and emulation entries, and aggregate the result over graphs (per-graph distribution) and over (graph, entry) pairs (pooled scatter).

Figure 6 summarizes the comparison. Panel (a) shows one timestep of the raw $\mathcal{C}$matrices for four representative graphs of increasing size; the diagonal and the dominant off-diagonal patterns are preserved on hardware. Panels (b,c) report the per-graph distribution of ρ_s: the one-body channel concentrates above $\rho_s = 0.9$ (median 0.924, mean 0.912). The two-body channel concentrates lower (median 0.870, mean 0.861), as expected since two-body entries have lower magnitudes and are more sensitive to shot noise and SPAM. Another explanation is the expected low magnitude of long-range correlations, which are concentrated around 0 and induce a lower rank-correlation as they are in the noise. Panels (d,e) pool every correlation entry across the 188 graphs into a single scatter: the one-body channel ($\rho_s = 0.928$) lies tightly around the diagonal with a mild contraction towards zero on the high-magnitude side, consistent with finite-time decoherence; the off-diagonal correlations ($\rho_s = 0.863$) show a broader scatter and a small negative bias for entries with $|\mathcal{C}_{ij}^{(2)}| \gtrsim 0.5$, again compatible with depolarizing noise on the most strongly correlated pairs.

Two points support the use of QPU features as drop-in replacements for emulation features in Sec. 4. First, the rank ordering of correlation entries is preserved within each graph at both the one- and two-body level (panels b,c), so GNN backbones using these features as node or edge attributes see the same relative structure on hardware as in simulation. Second, the worst-case graphs in panels (b,c) are systematically among the smallest of MUTAG, where the absolute number of distinct entries is so small that a few miscalibrated shots can shift the rank substantially; this regime does not extend to the larger graphs of the dataset.

We emphasize that this concordance study is restricted to a single operating point (AFM, t_4) on a single dataset (MUTAG); a full QPU run on $\text{ZINC}_{\text{subset}}^{*}$ remains beyond the budget of this study and is left for future work.

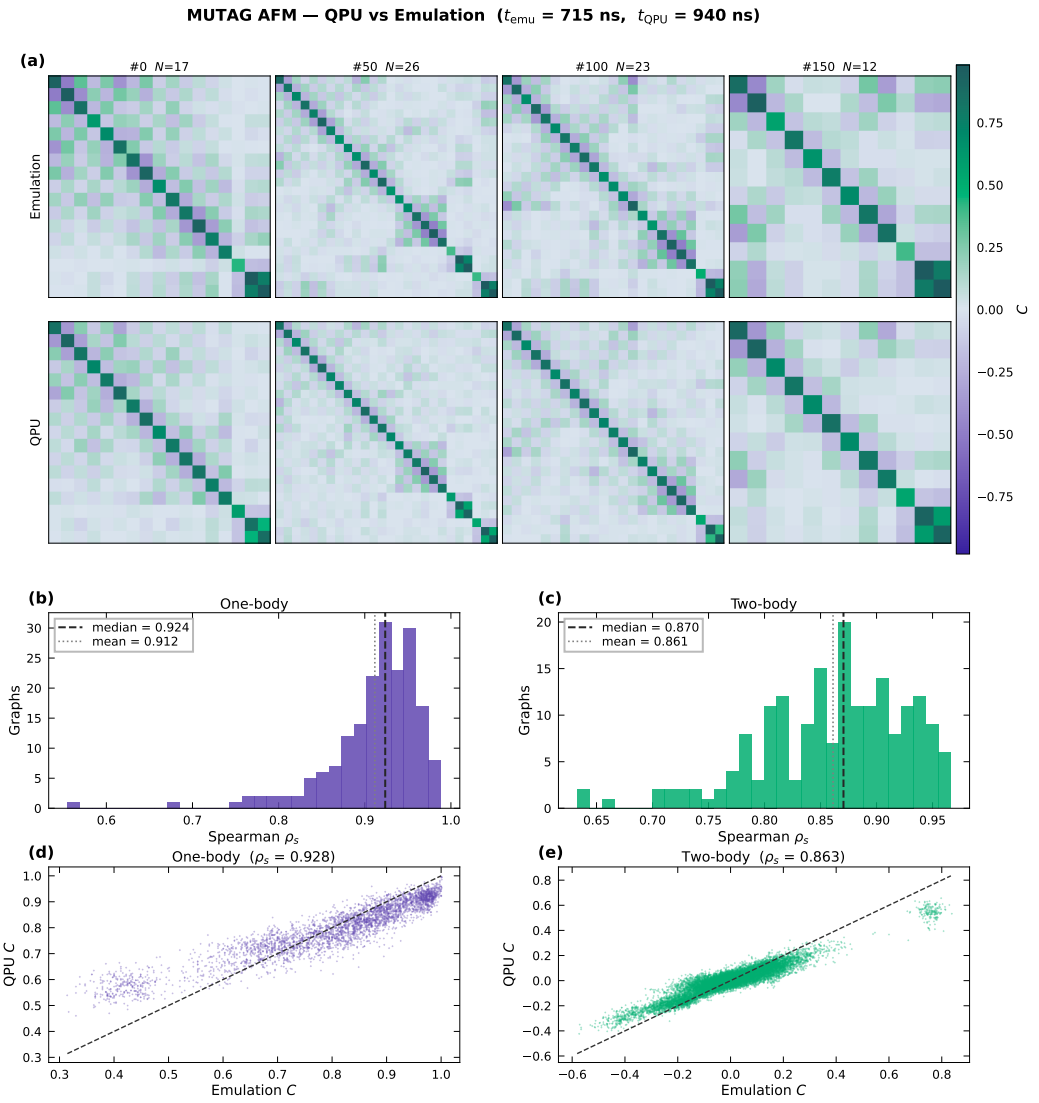


Figure 6: **Emulator–QPU concordance on MUTAG at** t_4 ($t_{\mathrm{emu}} = 715$ ns, $t_{\mathrm{QPU}} = 940$ ns; AFM operating point). **(a)** Normalized connected correlation matrices $C_{i,j}(t)$ [Eq. (2)] for the four representative graphs as functions of i and j; top row: emulation, bottom row: QPU. The presented correlations are normalized across the dataset, showing what is actually supplied to the models. **(b,c)** Per-graph distribution of the Spearman rank correlation ρ_s between QPU and emulation entries, over the 188 graphs for the (b) one-body $[C_{i,i}(t)]$ and (c) two-body $[C_{i,j\neq i}(t)]$ channels; dashed and dotted lines mark the median and mean. **(d,e)** Pooled scatter of every QPU entry against its emulation counterpart, for one-body (d, $\rho_s = 0.928$) and two-body (e, $\rho_s = 0.863$); the dashed line marks $y = x$. Hardware preserves the rank ordering of correlation entries within each graph; the residual contraction towards zero at large $|C_{i,j}(t)|$ is consistent with finite-time decoherence and finite-shot SPAM.

E GNN Methods

In this appendix, we provide various methodological details about the GNN benchmark methodology and model details.

Table 6: Feature sets used throughout the paper. Here $N_t = 20$ denotes the number of readout times in the full time-resolved quantum feature trajectories.

Feature set	**Description**	d
RRWP	$\mathbf{M}^k$, k=1...20	20
Laplacian	$\phi_2 \ldots \phi_5$ of $\mathbf{L}$	4
AFM / FM	$\boldsymbol{\mathcal{C}}_i^{(1)}$, $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ at N_t times	20
AFM+FM	concat. AFM & FM	40
AFM_{stat} / FM_{stat}	$\text{stat}(\cdot)$ over N_t	4
RRWP + X	concat. RRWP & X	$20{+}d_X$
HW-AFM	QPU measurements	4
$\text{HW-AFM}_{\text{stat}}$	$\text{stat}(\cdot)$ of HW-AFM	4

E.1 Features notations

Table 6 summarizes all feature sets, their origin, and dimensionality.

E.2 GNN details

This appendix complements Sec. 4.1 with the per-backbone descriptions, the GT block structure, and the training protocol omitted from the main text.

Feature injection. In every backbone the original dataset attributes (atom and bond types) are retained; structural descriptors are concatenated to the matching input channel. In the MPNN backbones, one-body features $\boldsymbol{\mathcal{C}}_i^{(1)}$ [Eq. (5)] are concatenated to the node feature vector before the first layer. In GT, dataset attributes and one-body correlations are projected independently to the hidden dimension and summed (see below). Two-body features $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ [Eq. (5)] are injected through whichever pairwise interface the architecture exposes: edge weights for GCN, edge attributes for GAT/GINE/Transformer/GeneralConv, and dense per-head attention biases for GT. This convention ensures that every architecture receives the same raw feature content; the architectural interface is the only differentiator.

Message-passing backbones. GCN [1] performs isotropic neighborhood aggregation through a symmetric degree-normalized adjacency operator. GINE [21] extends the Graph Isomorphism Network (GIN) [64] with explicit edge-feature injection into the message function, inheriting GIN's 1-WL expressiveness while accepting edge attributes. GeneralConv [22] provides a more flexible aggregation rule with learnable message transformations.

Attention-based backbones. GAT [3] computes learned attention coefficients over local neighborhoods. The Transformer layer [24] extends self-attention to the full node set with edge-aware biases, serving as a standard global-attention reference distinct from our custom GT.

Training protocol. The downstream benchmark compares *feature sets*, not architecture–feature combinations. We fix one set of training hyperparameters per architecture–dataset pair (Table 7) and reuse it across every feature configuration; no per-configuration hyperparameter search is performed. All downstream models are trained with AdamW [65] using a linear warm-up phase followed by ReduceLROnPlateau (factor 0.5); early stopping monitors the validation loss (MAE for ZINC, cross-entropy for MUTAG). Folds for MUTAG are stratified by graph size and class, and folds are consistent across all experiments.

E.3 Graph Transformer architecture

We describe the Graph Transformer (GT) architecture used in the GNN downstream benchmark experiments of Section 4. The key distinction from the message-passing backbones is that the GT can consume the dense off-diagonal correlation tensor $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ [Eq. (5)] for all node pairs, rather than only using pairwise information on molecular bonds.

Table 7: Main GT hyperparameter configurations used for each dataset. The corresponding parameter budgets are used to approximately match model capacity across backbones.

Parameter	**ZINC$_{\text{subset}}$***	**MUTAG**
Architecture		
Hidden dimension D	256	32
Number of heads H	16	4
Number of layers L	4	3
FFN dimension D_{ffn}	512	64
Dropout	0.1	0.3
Pooling	Mean	
Training		
Optimiser	AdamW	
Learning rate	5×10^{-4}	
Weight decay	10^{-5}	10^{-3}
Batch size	128	32
Max epochs	1000	300
Linear warmup epochs	50	20
Early stopping patience	200	75
LR scheduler	ReduceLROnPlateau (factor 0.5)	
Evaluation		
Loss	MAE	Cross-entropy
Metric	MAE	F_1 macro
Seeds	10	5 per fold
Folds	–	10

Input representation. Each molecular graph is represented by a set of atoms $\mathcal{V}_\mathcal{G}$ and a set of bonds $\mathcal{E}_\mathcal{G}$. Each atom type is mapped to a learned node embedding

$$\mathbf{h}_i^{(0)} \in \mathbb{R}^D, \qquad i \in \mathcal{V}_\mathcal{G}, \tag{29}$$

where D is the hidden dimension. Bond attributes are embedded separately and used to construct additive attention biases, as described below.

Diagonal correlation injection. The diagonal correlations $\boldsymbol{\mathcal{C}}_i^{(1)}$ are injected at the node level before the first Transformer block. When these correlations are used, they are mapped to the hidden dimension by a residual MLP encoder:

$$\mathbf{p}_i = f_\theta(\boldsymbol{\mathcal{C}}_i^{(1)}), \qquad f_\theta(\mathbf{x}) = \mathbf{W}_{\text{skip}}\mathbf{x} + \mathbf{W}_{\text{out}}\sigma\big(\mathbf{W}_{\text{in}}\mathbf{x}\big), \tag{30}$$

where σ is the ReLU nonlinearity. The encoded vector is added to the learned atom embedding:

$$\mathbf{h}_i^{(0)} \leftarrow \mathbf{h}_i^{(0)} + \mathbf{p}_i. \tag{31}$$

When diagonal correlations are not used, this encoder is omitted and the initial node representation is the learned atom embedding alone.

Off-diagonal correlation biases. The off-diagonal correlations $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ are injected as additive attention biases. Unlike message-passing backbones, which can only use pairwise information on molecular edges, the GT can consume $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ for all node pairs (i, j) in the dense graph.

When off-diagonal correlations are used, they are mapped to one scalar bias per attention head by a two-layer MLP:

$$b_{ij}^{\text{corr},(h)} = \Big[\mathbf{W}_4\,\sigma\big(\mathbf{W}_3\boldsymbol{\mathcal{C}}_{ij}^{(2)}\big)\Big]_h, \qquad h = 1, \ldots, H, \tag{32}$$

where H is the number of attention heads. When off-diagonal correlations are not used, this correlation-bias encoder is omitted and $b_{ij}^{\text{corr},(h)}$ is absent.

Edge biases. Molecular bond attributes are incorporated through a separate additive edge bias. For each attention head h, the bond representation associated with a node pair (i, j) is projected to a scalar

$$b_{ij}^{\mathrm{edge},(h)}. \tag{33}$$

This term encodes molecular bond information, whereas $b_{ij}^{\mathrm{corr},(h)}$ encodes the dense off-diagonal correlations. The two sources of pair information are kept separate and combined only at the level of the attention logits.

Multi-head attention with additive pair biases. At Transformer layer ℓ, node representations are projected to queries, keys, and values. For each head h, let

$$\mathbf{q}_i^{(h)}, \mathbf{k}_j^{(h)}, \mathbf{v}_j^{(h)} \in \mathbb{R}^{d_{\mathrm{head}}}, \qquad d_{\mathrm{head}} = D/H. \tag{34}$$

The attention logit from node i to node j is

$$\alpha_{ij}^{(h)} = \frac{\mathbf{q}_i^{(h)} \cdot \mathbf{k}_j^{(h)}}{\sqrt{d_{\mathrm{head}}}} + b_{ij}^{\mathrm{edge},(h)} + b_{ij}^{\mathrm{corr},(h)} + m_{ij}, \tag{35}$$

where m_{ij} is the padding mask, equal to 0 for valid node pairs and $-\infty$ when node j is a padding node. The attention weights are obtained by normalising over the key index:

$$a_{ij}^{(h)} = \frac{\exp\big(\alpha_{ij}^{(h)}\big)}{\sum_{k \in V_G} \exp\big(\alpha_{ik}^{(h)}\big)}. \tag{36}$$

The output of head h at node i is

$$\mathbf{z}_i^{(h)} = \sum_{j \in V_G} a_{ij}^{(h)} \mathbf{v}_j^{(h)}. \tag{37}$$

The head outputs are concatenated and linearly projected back to $\mathbb{R}^D$.

Transformer block. Each of the L Transformer blocks consists of biased multi-head attention followed by a feed-forward network, with residual connections, dropout, and layer normalisation:

$$\tilde{\mathbf{h}}^{(\ell)} = \mathrm{LayerNorm}\big(\mathbf{h}^{(\ell)} + \mathrm{Dropout}\big(\mathrm{MHA}\big(\mathbf{h}^{(\ell)}, \mathbf{b}^{\mathrm{edge}}, \mathbf{b}^{\mathrm{corr}}\big)\big)\big), \tag{38}$$

$$\mathbf{h}^{(\ell+1)} = \mathrm{LayerNorm}\big(\tilde{\mathbf{h}}^{(\ell)} + \mathrm{Dropout}\big(\mathrm{FFN}\big(\tilde{\mathbf{h}}^{(\ell)}\big)\big)\big). \tag{39}$$

Here MHA denotes multi-head attention with the additive edge and correlation biases defined above. The feed-forward network is

$$\mathrm{FFN}(\mathbf{x}) = \mathbf{W}_6\ \mathrm{Dropout}\big(\sigma\big(\mathbf{W}_5 \mathbf{x}\big)\big), \tag{40}$$

with hidden dimension D_{ffn}.

Batching and padding. Graphs of different sizes are batched by densifying the node representations with `to_dense_batch` from PyTorch Geometric [45]. This produces node tensors of shape

$$B \times N_{\max} \times D, \tag{41}$$

where B is the batch size and $N_{\max}$ is the maximum number of nodes in the batch. Off-diagonal correlation tensors and attention-bias tensors are padded consistently to shape

$$B \times H \times N_{\max} \times N_{\max}. \tag{42}$$

Padding nodes are excluded from the attention softmax by adding $-\infty$ to the corresponding logits through the mask term m_{ij} in Eq. (35).

Graph-level output. After the final Transformer block, node representations are pooled over valid, non-padding nodes:

$$\mathbf{h}_{\mathcal{G}} = \frac{1}{|\mathcal{V}_{\mathcal{G}}|} \sum_{i \in \mathcal{V}_{\mathcal{G}}} \mathbf{h}_i^{(L)}. \tag{43}$$

The graph representation $\mathbf{h}_{\mathcal{G}}$ is passed through a two-layer MLP with ReLU activation and dropout. The output is a scalar for regression tasks and a vector of logits for classification tasks.

This architecture therefore separates three sources of information: atom identities enter through learned node embeddings, diagonal correlations $\boldsymbol{\mathcal{C}}_i^{(1)}$are added to the node representations, and off-diagonal correlations $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$enter as dense additive biases in the attention logits. This is the mechanism that allows the GT to use pairwise correlations beyond the molecular edge set $\mathcal{E}_{\mathcal{G}}$.

Table 8: Model size configurations for GCN and GAT architectures. N_l: number of message-passing (backbone) or feedforward (prediction head) layers. N_d: hidden feature dimension per layer.

Model	Backbone		Prediction head	
	N_l	N_d	N_l	N_d
Tiny	2	16	2	16
Small	2	32	2	32
Medium	3	32	3	32
Large	4	64	4	64
Huge	4	128	4	8
V. Huge	5	256	3	256
Massive	4	512	3	512
Giant	5	512	3	512

Table 9: Quantum feature ablation on $\text{ZINC}_{\text{tiny}}$ ($\sim$1.5k graphs, emulation). All node and edge quantum features are used unless stated otherwise. Mean MAE ($\downarrow$) $\pm$ standard deviation over 10 seeds is reported. Best and 2nd best indicate the best and runner-up feature sets in each column.

Method	$\text{ZINC}^{*}_{\text{test}}$ (MAE $\downarrow$)	
	GCN	GAT
Baseline		
No PE	$0.520_{\pm 0.050}$	$0.480_{\pm 0.032}$
Quantum (FM)		
FM_{1b}	$0.512_{\pm 0.031}$	$0.473_{\pm 0.045}$
FM_{2b}	$0.478_{\pm 0.030}$	$0.456_{\pm 0.031}$
FM	$0.493_{\pm 0.035}$	$0.482_{\pm 0.055}$
Quantum (AFM)		
AFM_{1b}	$\mathbf{0.443_{\pm 0.031}}$	$0.454_{\pm 0.066}$
AFM_{2b}	$0.532_{\pm 0.025}$	$\mathbf{0.449_{\pm 0.041}}$
AFM	$\underline{0.446}_{\pm 0.020}$	$\underline{0.450}_{\pm 0.067}$

E.4 Ablation: diagonal and off-diagonal quantum features

We assess the impact of diagonal $\boldsymbol{\mathcal{C}}_i^{(1)}$ and off-diagonal $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ [Eq. (5)] quantum features on model performance. We use a subset $\text{ZINC}_{\text{tiny}}$ of 1.5×10^3 graphs, with property distributions matched to the $\text{ZINC}^{*}_{\text{small}}$ subset of $\approx 12 \times 10^3$ graphs, and evaluate two representative architectures: a message-passing GCN and an attention-based GAT. All experiments use the same fixed backbone as in the benchmark experiments. All models use the training and evaluation settings, including dropout and pooling, reported in Table 7. The model sizes are given in Table 8. We compare classical ZINC features (atom and bond types) against their augmentation with one- and two-body quantum features. Results in Table 9 show that quantum features consistently improve over classical features when averaged across six model sizes and five seeds. Tables 10 and 11 show that the large configuration has better performance overall, while larger models provide no systematic further improvement. Hence, we select the large backbone configuration for the emulated benchmark and QPU validation (Sec. 4).

In Fig. 7, we show the error metrics used for selection of the graph transformer (GT) model with the best capacity for our benchmarking experiments.

E.5 Hardware versus emulation per backbone

Table 12 lists, per backbone, the MUTAG F_1 of the RRWP hybrids with emulated and hardware AFM features. For the compressed features, agreement holds on all six backbones, with a maximum absolute deviation of 0.009. For the full time-resolved features, five backbones are within 0.014, while GCN deviates by 0.041.

Table 10: Model-size ablation for GCN on $\text{ZINC}_{\text{tiny}}$ (∼1.5k graphs, emulation). Mean MAE (↓) ± standard deviation over 10 seeds is reported. Best and 2nd best indicate the best and runner-up model sizes in each column.

Method	GCN on $\text{ZINC}^{*}_{\text{test}}$ (MAE ↓)							
	Tiny	Small	Medium	Large	Huge	V. Huge	Massive	Giant
Baseline								
ZINC (Baseline)	$0.530_{\pm 0.044}$	$0.529_{\pm 0.025}$	$0.490_{\pm 0.049}$	$0.513_{\pm 0.062}$	$0.515_{\pm 0.055}$	$0.527_{\pm 0.048}$	$0.528_{\pm 0.042}$	$0.546_{\pm 0.042}$
Quantum (FM)								
FM_{1b}	$0.518_{\pm 0.031}$	$0.520_{\pm 0.014}$	$0.514_{\pm 0.033}$	$0.520_{\pm 0.043}$	$0.503_{\pm 0.028}$	$0.506_{\pm 0.029}$	$0.508_{\pm 0.021}$	$0.516_{\pm 0.026}$
FM_{2b}	$0.478_{\pm 0.026}$	$0.479_{\pm 0.025}$	$0.461_{\pm 0.018}$	$0.473_{\pm 0.026}$	$0.468_{\pm 0.028}$	$0.478_{\pm 0.030}$	$0.494_{\pm 0.028}$	$0.511_{\pm 0.036}$
FM	$0.524_{\pm 0.013}$	$0.496_{\pm 0.024}$	$0.488_{\pm 0.011}$	$0.484_{\pm 0.035}$	$0.481_{\pm 0.042}$	$0.492_{\pm 0.036}$	$0.496_{\pm 0.033}$	$0.526_{\pm 0.038}$
Quantum (AFM)								
AFM_{1b}	$0.469_{\pm 0.023}$	**$0.450_{\pm 0.023}$**	**$0.441_{\pm 0.036}$**	**$0.434_{\pm 0.034}$**	**$0.434_{\pm 0.031}$**	$0.433_{\pm 0.026}$	$0.451_{\pm 0.033}$	$0.458_{\pm 0.021}$
AFM_{2b}	$0.554_{\pm 0.016}$	$0.532_{\pm 0.026}$	$0.538_{\pm 0.005}$	$0.522_{\pm 0.030}$	$0.534_{\pm 0.023}$	$0.527_{\pm 0.025}$	$0.533_{\pm 0.024}$	$0.546_{\pm 0.026}$
AFM	**$0.464_{\pm 0.009}$**	$0.459_{\pm 0.010}$	$0.455_{\pm 0.014}$	$0.448_{\pm 0.021}$	$0.438_{\pm 0.017}$	**$0.433_{\pm 0.024}$**	**$0.442_{\pm 0.016}$**	**$0.452_{\pm 0.016}$**

Table 11: Model-size ablation for GAT on $\text{ZINC}_{\text{tiny}}$ (∼1.5k graphs, emulation). Mean MAE (↓) ± standard deviation over 10 seeds is reported. Best and 2nd best indicate the best and runner-up model sizes in each column.

Method	GAT on $\text{ZINC}^{*}_{\text{test}}$ (MAE ↓)							
	Tiny	Small	Medium	Large	Huge	V. Huge	Massive	Giant
Baseline								
ZINC (Baseline)	$0.471_{\pm 0.024}$	$0.462_{\pm 0.036}$	$0.458_{\pm 0.019}$	$0.483_{\pm 0.020}$	$0.474_{\pm 0.024}$	$0.503_{\pm 0.030}$	$0.505_{\pm 0.041}$	$0.518_{\pm 0.020}$
Quantum (FM)								
FM_{1b}	$0.469_{\pm 0.058}$	$0.486_{\pm 0.059}$	$0.457_{\pm 0.042}$	$0.459_{\pm 0.039}$	$0.468_{\pm 0.045}$	$0.488_{\pm 0.029}$	$0.485_{\pm 0.029}$	$0.511_{\pm 0.018}$
FM_{2b}	**$0.447_{\pm 0.025}$**	**$0.459_{\pm 0.037}$**	$0.446_{\pm 0.021}$	$0.465_{\pm 0.040}$	$0.438_{\pm 0.016}$	$0.457_{\pm 0.021}$	$0.476_{\pm 0.030}$	$0.497_{\pm 0.015}$
FM	$0.486_{\pm 0.070}$	$0.495_{\pm 0.069}$	$0.497_{\pm 0.076}$	$0.454_{\pm 0.041}$	$0.469_{\pm 0.045}$	$0.484_{\pm 0.027}$	$0.504_{\pm 0.031}$	$0.510_{\pm 0.029}$
Quantum (AFM)								
AFM_{1b}	$0.501_{\pm 0.175}$	$0.460_{\pm 0.049}$	$0.444_{\pm 0.039}$	$0.450_{\pm 0.041}$	$0.438_{\pm 0.032}$	**$0.444_{\pm 0.022}$**	$0.459_{\pm 0.025}$	$0.465_{\pm 0.023}$
AFM_{2b}	$0.450_{\pm 0.041}$	$0.468_{\pm 0.058}$	**$0.443_{\pm 0.031}$**	$0.447_{\pm 0.046}$	**$0.430_{\pm 0.020}$**	$0.447_{\pm 0.032}$	$0.453_{\pm 0.040}$	$0.487_{\pm 0.032}$
AFM	$0.504_{\pm 0.155}$	$0.469_{\pm 0.075}$	$0.457_{\pm 0.064}$	**$0.421_{\pm 0.026}$**	$0.434_{\pm 0.023}$	$0.452_{\pm 0.027}$	**$0.445_{\pm 0.025}$**	**$0.461_{\pm 0.029}$**

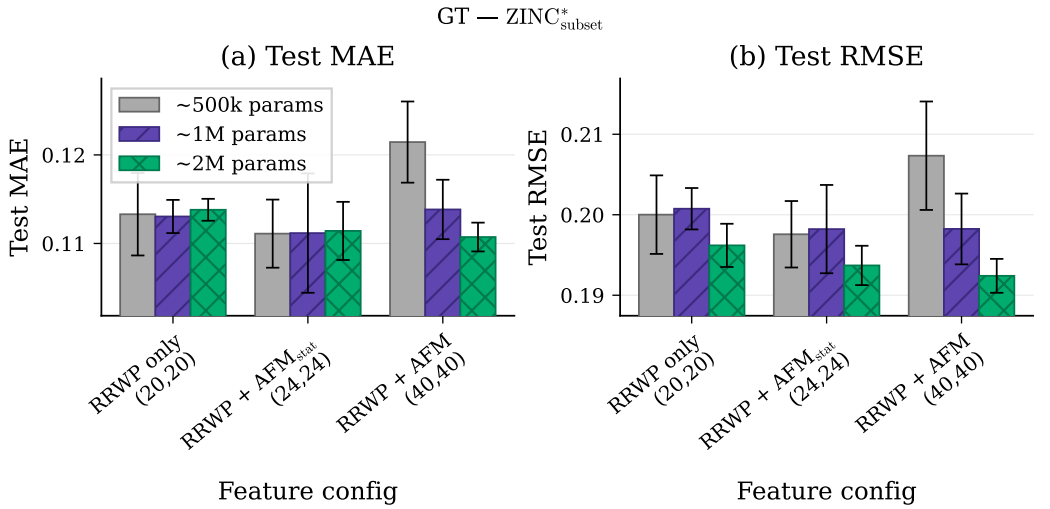


Figure 7: Results for model GT on $\text{ZINC}^{*}_{\text{subset}}$. Grouped bar charts show **(a)** test MAE and **(b)** test RMSE for three feature configurations: RRWP alone ($d_{\text{abs}} = 20,\ d_{\text{rel}} = 20$), RRWP + AFM_{stat} ($d_{\text{abs}} = 24,\ d_{\text{rel}} = 24$), and RRWP + AFM ($d_{\text{abs}} = 40,\ d_{\text{rel}} = 40$), each evaluated at three model capacities (Number of trainable parameters $\in \{500k, 1M, 2M\}$). Error bars report the standard deviation over 5 independent random seeds; in this study the reported epoch was selected on the test split. Increasing model capacity consistently reduces the gap between feature configurations, suggesting that the relative underperformance of RRWP + AFM at low capacity is a capacity bottleneck rather than a fundamental incompatibility between the two positional encodings.

Table 12: MUTAG, RRWP hybrids with emulated versus hardware (HW) AFM features, per backbone. Mean macro F_1 (↑) over 10 stratified folds × 5 seeds (values of Table 2); Δ = hardware − emulation.

Backbone	RRWP+AFM	RRWP+HW-AFM	Δ (full)	RRWP+AFM$_{\text{stat}}$	RRWP+HW-AFM$_{\text{stat}}$	Δ (stat)
GT	0.855	0.854	−0.001	0.854	0.849	−0.005
GCN	0.878	0.837	−0.041	0.852	0.852	+0.000
GAT	0.862	0.852	−0.010	0.860	0.853	−0.007
GINE	0.848	0.841	−0.007	0.846	0.852	+0.006
Transformer	0.868	0.854	−0.014	0.852	0.861	+0.009
General	0.867	0.859	−0.008	0.873	0.877	+0.004

F Intrinsic benchmark

In this appendix, we present the targets, training details and full results for our intrinsic information benchmark.

F.1 Intrinsic benchmark: targets and probe

This appendix complements Sec. 3 with the training details and formal definition of each of the 15 probed targets and with the technical details of the probe omitted from the main text.

Feature sets, dimensionalities and the z-score normalisation with training-split statistics are defined once in (see App. E.1) and reused unchanged here.

Target definitions. All structural quantities are computed on the undirected simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ underlying each molecule, i.e. after stripping bond multiplicities and atom-type information.

Node-level (5).

- **Degree**: $d_i = |\{j : (i,j) \in \mathcal{E}\}|$.
- **Betweenness centrality**: $b_i = \sum_{s \neq t \neq i} \sigma_{st}(i)/\sigma_{st}$, where σ_{st} is the number of shortest s-t paths and $\sigma_{st}(i)$ the number of those passing through i.
- **Eccentricity**: $\varepsilon_i = \max_{j \in \mathcal{V}} d(i,j)$, the longest shortest-path distance from i.
- **Bridge-node detection** (binary): 1 iff i is incident to at least one bridge edge.
- **Smallest cycle through node**: $g_i = \min\{|C| : C \text{ simple cycle}, i \in C\}$, set to the graph girth when i lies on no cycle.

Edge-level (5).

- **Link prediction** (binary): 1 iff $(i,j) \in \mathcal{E}$. Candidates are all pairs within two hops, so that negative examples lie in a realistic neighbourhood rather than being trivially far apart.
- **Edge betweenness**: $b_e = \sum_{s \neq t} \sigma_{st}(e)/\sigma_{st}$.
- **Bridge-edge detection** (binary): 1 iff removing e disconnects the graph.
- **Local edge connectivity**: κ_{ij}, the maximum number of edge-disjoint i-j paths (Menger).
- **Smallest cycle through edge**: $g_{ij} = \min\{|C| : C \text{ simple cycle}, e \in C\}$.

Graph-level (5).

- **Number of cycles**: circuit rank $|\mathcal{E}| - |\mathcal{V}| + c$ with c connected components.
- **Diameter**: $\max_{i,j} d(i,j)$.
- **Density**: $2|\mathcal{E}|/\big(|\mathcal{V}|(|\mathcal{V}|-1)\big)$.
- **Wiener index**: $W = \frac{1}{2}\sum_{i,j \in \mathcal{V}} d(i,j)$.
- **Algebraic connectivity**: $\lambda_2(\mathbf{L})$, the second-smallest eigenvalue of the graph Laplacian.

Probe architecture. The shared encoder ϕ is a two-layer MLP mapping the input feature vector to $\mathbb{R}^{64}$ with batch normalisation and ReLU activations between layers. The predictor ρ has one hidden layer of width 64 with batch normalisation and ReLU, and a linear output. For graph-level probes, ρ additionally applies a dropout layer with $p = 0.1$ before the output; node- and edge-level probes use

no dropout. Neither ϕ nor ρ has access to graph structure beyond what the input already encodes: there is no message passing, no attention, and no learned aggregation over neighbours, so any signal extracted reflects information already present in the encoding.

Entity-level input construction.

- **Node probes** use $\mathbf{x}_i = \boldsymbol{\mathcal{C}}_i^{(1)}$[Eq. (5)]
- **Edge probes** use $\mathbf{x}_{ij} = \boldsymbol{\mathcal{C}}_{ij}^{(2)}$ for $(i, j) \in \mathcal{E}$. For link prediction we additionally sample, per graph, negative candidates from the two-hop complement $\{(i, j) \notin \mathcal{E} : d(i, j) \leq 2\}$ in a $1:1$ ratio with positive edges.
- **Graph probes** enrich each node with its node-averaged pairwise feature: $\mathbf{x}_i = [\,\boldsymbol{\mathcal{C}}_{\boldsymbol{i}}^{(\mathbf{1})} \,\|\, \bar{\mathbf{c}}_i\,]$ with $\bar{\mathbf{c}}_i = \frac{1}{N}\sum_j \boldsymbol{\mathcal{C}}_{\boldsymbol{ij}}^{(\mathbf{2})}$, yielding a $2d$-dimensional per-node vector before sum pooling.

Training. Probes are trained on $\text{ZINC}^*_{\text{subset}}$ (11,368 graphs) and on MUTAG separately with 5-fold cross-validation means and standard deviations are computed across the five folds.

All probes use AdamW (learning rate 10^{-3}, default weight decay), MSE loss for regression, and binary cross-entropy with logits for classification. 5-fold cross-validation is performed at the *graph* level: all nodes and edges of a given graph belong to the same fold, which prevents leakage between train and test when probing node- or edge-level quantities. Folds are stratified by label for classification tasks. Node- and edge-level probes run for at most 50 epochs with patience 10 and batch size 512; graph-level probes for at most 100 epochs with patience 20 and batch size 32. The best checkpoint under validation loss is restored before computing the test metric, and the five test metrics are reported as mean $\pm$ standard deviation across folds.

F.2 Intrinsic benchmark: full results

Table 13 reports the complete $\text{ZINC}^*_{\text{subset}}$ intrinsic-benchmark results summarised in Table 1 of Sec. 3. It includes the random-feature control, the classical descriptors, the simulated FM and AFM descriptors, their compressed variants, the bare FM+AFM union, and the RRWP-based hybrid combinations. The corresponding MUTAG results are reported in Table 14. Compression effects are analysed in App. F.2.1, and hardware AFM results in App. F.3.1.

Random-feature control and saturating tasks. The random-feature control is separated from the structural descriptors on the non-saturating regression tasks. On $\text{ZINC}^*_{\text{subset}}$, for instance, random features give 0.564 MAE on degree prediction, whereas every non-random structural descriptor is substantially better. The same holds for cycle, betweenness, and graph-level distance targets, confirming that the probe is not solving the tasks from dataset-level shortcuts alone.

Link prediction, bridge-edge detection, and, to a lesser extent, bridge-node detection are close to saturation for many non-random feature sets. They are retained as sanity checks, but they should not be used to rank descriptor families.

F.2.1 Compression variants

The columns denoted FM_{stat} and AFM_{stat} use the four temporal summary statistics defined in Sec. 2: mean, variance, minimum, and maximum over the sampled evolution times. The columns denoted FM_{mean} and AFM_{mean} are a more aggressive compression in which only the temporal mean is retained. Note that our compression produces four descriptors per pair (mean, ...), of which the mean $\overline{C_{ij}}$ [Eq. (12)] is the leading one. Alternatively, the data could be compressed into the asymptotically exact long-time correlation C_{ij}^{∞} [Eq. (13)], at the cost of retaining a single descriptor per pair. In an ablation, adding C_{ij}^{∞} to the compressed features yielded no statistically significant change in performance.

The statistic-compressed descriptors preserve a substantial fraction of the signal despite reducing the temporal dimension from $N_t = 20$ to four values. On $\text{ZINC}^*_{\text{subset}}$, AFM_{stat} remains competitive with, and often better than, Laplacian features on several non-spectral topological tasks, including degree, betweenness, edge betweenness, smallest cycle through a node or edge, number of cycles, and density.

Table 13: Intrinsic information benchmark on $\text{ZINC}^*_{\text{subset}}$ graph tasks. Mean ± standard deviation over 5-fold cross-validation is reported. Metrics are MAE (↓) for regression tasks and accuracy (↑) for classification tasks. Best and 2nd best indicate the unique best and runner-up feature sets across the complete feature set, combining both halves of the table; entries marked † are tied with the best.

	Baseline	Classical			Quantum FM		
Task	Random	RRWP	Laplacian	RRWP+Lap.	FM_{mean}	FM_{stat}	FM
Node Prediction							
Bridge Node (Acc ↑)	$0.58_{\pm 0.00}$	$0.93_{\pm 0.01}$	$0.76_{\pm 0.00}$	$0.95_{\pm 0.00}$	$0.68_{\pm 0.00}$	$0.74_{\pm 0.00}$	$0.83_{\pm 0.00}$
Degree (MAE ↓) ($\times 10^2$)	$56.43_{\pm 0.63}$	$14.36_{\pm 0.99}$	$46.70_{\pm 0.36}$	$7.0_{\pm 2.4}$	$56.58_{\pm 0.45}$	$30.51_{\pm 0.27}$	$18.74_{\pm 0.69}$
Betweenness (MAE ↓) ($\times 10^2$)	$16.48_{\pm 0.05}$	$7.77_{\pm 0.72}$	$12.46_{\pm 0.11}$	$4.76_{\pm 0.29}$	$16.17_{\pm 0.07}$	$9.10_{\pm 0.13}$	$7.17_{\pm 0.05}$
Eccentricity (MAE ↓)	$2.14_{\pm 0.01}$	$1.88_{\pm 0.02}$	$1.51_{\pm 0.02}$	$\mathbf{1.08}_{\pm 0.04}$	$2.10_{\pm 0.01}$	$1.90_{\pm 0.01}$	$1.65_{\pm 0.01}$
Min Cycle (Node) (MAE ↓)	$2.66_{\pm 0.01}$	$0.73_{\pm 0.02}$	$1.68_{\pm 0.02}$	$0.60_{\pm 0.03}$	$2.36_{\pm 0.02}$	$1.80_{\pm 0.01}$	$1.42_{\pm 0.04}$
Edge Prediction							
Link Pred. (Acc ↑)	$0.91_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$0.94_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$0.97_{\pm 0.00}$	$0.99_{\pm 0.00}$	$0.99_{\pm 0.00}$
Bridge Edge (Acc ↑)	$0.96_{\pm 0.00}$	$0.99^{\dagger}_{\pm 0.00}$	$0.98_{\pm 0.00}$	$0.99^{\dagger}_{\pm 0.00}$	$0.96_{\pm 0.00}$	$0.97_{\pm 0.00}$	$0.98_{\pm 0.00}$
Edge Betw. (MAE ↓) ($\times 10^3$)	$35.20_{\pm 0.14}$	$11.6_{\pm 1.7}$	$20.57_{\pm 0.35}$	$\underline{6.09}_{\pm 0.95}$	$16.01_{\pm 0.14}$	$13.04_{\pm 0.28}$	$9.89_{\pm 0.08}$
Local Conn. (MAE ↓) ($\times 10^2$)	$26.96_{\pm 0.13}$	$10.01_{\pm 0.25}$	$13.92_{\pm 0.16}$	$7.39_{\pm 0.38}$	$15.24_{\pm 0.11}$	$13.89_{\pm 0.16}$	$12.60_{\pm 0.12}$
Min Cycle (Edge) (MAE ↓) ($\times 10^2$)	$60.92_{\pm 0.31}$	$13.3_{\pm 1.3}$	$42.41_{\pm 0.36}$	$10.2_{\pm 1.0}$	$25.67_{\pm 0.28}$	$23.02_{\pm 0.24}$	$19.36_{\pm 0.41}$
Graph Prediction							
Num Cycles (MAE ↓) ($\times 10^2$)	$60.24_{\pm 0.94}$	$6.68_{\pm 0.93}$	$27.4_{\pm 1.5}$	$5.90_{\pm 0.56}$	$22.92_{\pm 0.63}$	$18.38_{\pm 0.65}$	$15.33_{\pm 0.34}$
Diameter (MAE ↓)	$1.47_{\pm 0.03}$	$0.49_{\pm 0.04}$	$0.69_{\pm 0.01}$	$0.47_{\pm 0.05}$	$0.82_{\pm 0.02}$	$0.71_{\pm 0.01}$	$0.65_{\pm 0.01}$
Density (MAE ↓) ($\times 10^3$)	$3.72_{\pm 0.09}$	$1.03^{\dagger}_{\pm 0.22}$	$2.48_{\pm 0.22}$	$1.14^{\dagger}_{\pm 0.18}$	$1.95_{\pm 0.21}$	$1.86_{\pm 0.22}$	$1.70_{\pm 0.18}$
Wiener Index (MAE ↓) ($\times 10^2$)	$39.40_{\pm 0.36}$	$10.8_{\pm 2.1}$	$19.24_{\pm 0.24}$	$9.61_{\pm 0.92}$	$19.25_{\pm 0.56}$	$16.41_{\pm 0.14}$	$14.15_{\pm 0.19}$
Alg. Connect. (MAE ↓) ($\times 10^3$)	$9.86_{\pm 0.32}$	$2.63_{\pm 0.28}$	$5.59_{\pm 0.42}$	$\mathbf{2.22}_{\pm 0.12}$	$5.91_{\pm 0.32}$	$5.46_{\pm 0.17}$	$5.21_{\pm 0.30}$

	Quantum AFM			Hybrid			
Task	AFM_{mean}	AFM_{stat}	AFM	FM+AFM	RRWP+FM	RRWP+AFM	RRWP+FM+AFM
Node Prediction							
Bridge Node (Acc ↑)	$0.82_{\pm 0.00}$	$0.85_{\pm 0.00}$	$0.91_{\pm 0.00}$	$0.94_{\pm 0.00}$	$0.97_{\pm 0.00}$	$0.98^{\dagger}_{\pm 0.00}$	$0.98^{\dagger}_{\pm 0.00}$
Degree (MAE ↓) ($\times 10^2$)	$13.50_{\pm 0.21}$	$9.07_{\pm 0.09}$	$1.26_{\pm 0.16}$	$1.09^{\dagger}_{\pm 0.09}$	$4.71_{\pm 0.48}$	$1.18^{\dagger}_{\pm 0.34}$	$1.25_{\pm 0.39}$
Betweenness (MAE ↓) ($\times 10^2$)	$11.81_{\pm 0.15}$	$10.13_{\pm 0.10}$	$7.11_{\pm 0.19}$	$5.19_{\pm 0.06}$	$4.33_{\pm 0.26}$	$\underline{3.90}_{\pm 0.06}$	$\mathbf{3.31}_{\pm 0.10}$
Eccentricity (MAE ↓)	$2.10_{\pm 0.02}$	$2.02_{\pm 0.02}$	$1.91_{\pm 0.02}$	$1.58_{\pm 0.02}$	$1.48_{\pm 0.03}$	$1.77_{\pm 0.02}$	$\underline{1.47}_{\pm 0.02}$
Min Cycle (Node) (MAE ↓)	$1.61_{\pm 0.00}$	$1.40_{\pm 0.01}$	$0.88_{\pm 0.01}$	$0.66_{\pm 0.02}$	$0.48_{\pm 0.03}$	$\underline{0.36}_{\pm 0.01}$	$\mathbf{0.28}_{\pm 0.01}$
Edge Prediction							
Link Pred. (Acc ↑)	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$
Bridge Edge (Acc ↑)	$0.97_{\pm 0.00}$	$0.97_{\pm 0.00}$	$0.99_{\pm 0.00}$	$0.99_{\pm 0.00}$	$0.99^{\dagger}_{\pm 0.00}$	$0.99^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$
Edge Betw. (MAE ↓) ($\times 10^3$)	$11.81_{\pm 0.60}$	$9.76_{\pm 0.27}$	$8.61_{\pm 0.36}$	$6.87_{\pm 0.16}$	$6.69_{\pm 0.55}$	$6.13_{\pm 0.43}$	$\mathbf{5.36}_{\pm 0.35}$
Local Conn. (MAE ↓) ($\times 10^2$)	$16.36_{\pm 0.23}$	$15.16_{\pm 0.16}$	$11.30_{\pm 0.20}$	$8.79_{\pm 0.18}$	$6.83_{\pm 0.15}$	$\underline{6.07}_{\pm 0.15}$	$\mathbf{5.14}_{\pm 0.16}$
Min Cycle (Edge) (MAE ↓) ($\times 10^2$)	$26.15_{\pm 0.42}$	$23.07_{\pm 0.50}$	$13.21_{\pm 0.49}$	$11.75_{\pm 0.63}$	$10.03_{\pm 0.92}$	$7.62^{\dagger}_{\pm 0.44}$	$7.17^{\dagger}_{\pm 0.73}$
Graph Prediction							
Num Cycles (MAE ↓) ($\times 10^2$)	$21.41_{\pm 0.57}$	$16.36_{\pm 0.50}$	$9.85_{\pm 0.85}$	$7.20_{\pm 0.29}$	$4.83^{\dagger}_{\pm 0.38}$	$4.48^{\dagger}_{\pm 0.71}$	$4.67^{\dagger}_{\pm 0.42}$
Diameter (MAE ↓)	$0.89_{\pm 0.01}$	$0.75_{\pm 0.02}$	$0.60_{\pm 0.02}$	$0.51_{\pm 0.02}$	$\underline{0.36}_{\pm 0.04}$	$0.37_{\pm 0.03}$	$\mathbf{0.32}_{\pm 0.02}$
Density (MAE ↓) ($\times 10^3$)	$1.84_{\pm 0.26}$	$1.72_{\pm 0.23}$	$1.86_{\pm 0.40}$	$1.56_{\pm 0.23}$	$1.21^{\dagger}_{\pm 0.33}$	$1.10^{\dagger}_{\pm 0.32}$	$1.25^{\dagger}_{\pm 0.26}$
Wiener Index (MAE ↓) ($\times 10^2$)	$23.22_{\pm 0.29}$	$19.46_{\pm 0.32}$	$14.99_{\pm 0.84}$	$11.7_{\pm 1.3}$	$7.13^{\dagger}_{\pm 0.68}$	$8.81_{\pm 0.70}$	$7.24^{\dagger}_{\pm 0.71}$
Alg. Connect. (MAE ↓) ($\times 10^3$)	$7.34_{\pm 0.23}$	$6.30_{\pm 0.35}$	$5.32_{\pm 0.27}$	$4.29_{\pm 0.56}$	$2.70_{\pm 0.15}$	$\underline{2.34}_{\pm 0.22}$	$2.38_{\pm 0.20}$

The degradation from AFM to AFM_{stat} is nevertheless systematic on informative regression tasks except density, typically increasing MAE by a factor of roughly 1.1–1.7, and by 7.2 on degree.

The mean-only variants lose more information. For example, AFM_{mean} gives 13.50×10^{-2} MAE on degree prediction, compared with 1.26×10^{-2} for full AFM, and FM_{mean} gives 2.36 MAE on smallest-cycle-through-node prediction, compared with 1.42 for full FM. Thus, the full descriptors do not owe their advantage only to average correlation amplitude; the temporal profile of the dynamics carries additional structural information.

F.3 MUTAG corroborating results

Table 14 reports the same intrinsic-benchmark protocol on MUTAG. Because MUTAG is much smaller than $\text{ZINC}^*_{\text{subset}}$, fold-to-fold variability is larger and close numerical rankings should be interpreted cautiously. The table is therefore used as a qualitative robustness check rather than as a second precise ranking of all descriptors.

The main qualitative patterns are consistent with the $\text{ZINC}^*_{\text{subset}}$ results. AFM-based hybrids remain strong on local structural targets, while Laplacian and RRWP-based descriptors remain particularly effective on several spectral or distance-like quantities. As in the $\text{ZINC}^*_{\text{subset}}$ table, link prediction and bridge-edge detection are nearly saturated for the non-random feature sets and mainly serve as sanity checks.

Table 14: Intrinsic information benchmark on MUTAG graph tasks. Mean ± standard deviation over 5-fold cross-validation is reported. Metrics are MAE (↓) for regression tasks and accuracy (↑) for classification tasks. Best and 2nd best indicate the unique best and runner-up feature sets across the complete feature set, combining both halves of the table; entries marked tied† are tied with the best.

	Baseline	Classical			Quantum FM		
Task	Random	RRWP	Laplacian	RRWP+Lap.	FM_{mean}	FM_{stat}	FM
Node Prediction							
Bridge Node (Acc ↑)	$0.58_{\pm 0.01}$	$0.96_{\pm 0.01}$	$0.84_{\pm 0.03}$	$0.97_{\pm 0.01}$	$0.72_{\pm 0.02}$	$0.76_{\pm 0.03}$	$0.83_{\pm 0.02}$
Degree (MAE ↓) ($\times 10^2$)	$70.3_{\pm 1.7}$	$14.2_{\pm 1.0}$	$40.7_{\pm 1.6}$	$15.24_{\pm 0.88}$	$47.6_{\pm 1.3}$	$33.2_{\pm 2.7}$	$28.4_{\pm 1.2}$
Betweenness (MAE ↓) ($\times 10^2$)	$12.79_{\pm 0.57}$	**$3.89_{\pm 0.28}$**	$8.21_{\pm 0.24}$	$4.53_{\pm 0.40}$	$9.95_{\pm 0.94}$	$6.65_{\pm 0.44}$	$6.34_{\pm 0.18}$
Eccentricity (MAE ↓)	$1.82_{\pm 0.05}$	$\underline{1.01}_{\pm 0.10}$	$1.09_{\pm 0.07}$	**$0.92_{\pm 0.08}$**	$1.35_{\pm 0.08}$	$1.16_{\pm 0.07}$	$1.12_{\pm 0.04}$
Min Cycle (Node) (MAE ↓)	$3.09_{\pm 0.09}$	$0.73_{\pm 0.02}$	$1.61_{\pm 0.06}$	$0.81_{\pm 0.08}$	$1.70_{\pm 0.12}$	$1.62_{\pm 0.10}$	$1.62_{\pm 0.07}$
Edge Prediction							
Link Pred. (Acc ↑)	$0.88_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$0.95_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$0.94_{\pm 0.01}$	$0.98_{\pm 0.00}$	$0.99_{\pm 0.00}$
Bridge Edge (Acc ↑)	$0.97_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$0.99_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$0.97_{\pm 0.00}$	$0.97_{\pm 0.00}$	$0.98_{\pm 0.00}$
Edge Betw. (MAE ↓) ($\times 10^3$)	$45.0_{\pm 2.0}$	$9.00^{\dagger}_{\pm 0.40}$	$23.83_{\pm 0.83}$	$8.76^{\dagger}_{\pm 0.43}$	$20.0_{\pm 1.9}$	$17.1_{\pm 1.6}$	$13.34_{\pm 0.66}$
Local Conn. (MAE ↓)	$0.53_{\pm 0.00}$	$0.27_{\pm 0.02}$	$0.33_{\pm 0.01}$	$0.21_{\pm 0.01}$	$0.33_{\pm 0.01}$	$0.30_{\pm 0.01}$	$0.26_{\pm 0.00}$
Min Cycle (Edge) (MAE ↓)	$1.06_{\pm 0.03}$	$0.21_{\pm 0.02}$	$0.56_{\pm 0.01}$	**$0.17_{\pm 0.01}$**	$0.43_{\pm 0.03}$	$0.32_{\pm 0.01}$	$0.29_{\pm 0.02}$
Graph Prediction							
Num Cycles (MAE ↓)	$0.81_{\pm 0.05}$	$0.19_{\pm 0.05}$	$0.30_{\pm 0.04}$	**$0.14_{\pm 0.03}$**	$0.21_{\pm 0.04}$	$0.21_{\pm 0.04}$	$0.24_{\pm 0.02}$
Diameter (MAE ↓)	$1.66_{\pm 0.46}$	$0.54^{\dagger}_{\pm 0.04}$	$0.84_{\pm 0.10}$	$0.51^{\dagger}_{\pm 0.03}$	$0.70_{\pm 0.06}$	$0.81_{\pm 0.09}$	$0.82_{\pm 0.14}$
Density (MAE ↓) ($\times 10^2$)	$4.30_{\pm 0.79}$	$1.40^{\dagger}_{\pm 0.49}$	$4.5_{\pm 2.1}$	$2.8_{\pm 1.1}$	$1.89^{\dagger}_{\pm 0.52}$	$4.1_{\pm 1.7}$	$3.8_{\pm 1.1}$
Wiener Index (MAE ↓)	$0.62_{\pm 0.04}$	**$0.11_{\pm 0.01}$**	$0.31_{\pm 0.03}$	$0.17_{\pm 0.02}$	$0.18_{\pm 0.02}$	$0.20_{\pm 0.03}$	$0.22_{\pm 0.02}$
Alg. Connect. (MAE ↓) ($\times 10^2$)	$7.5_{\pm 2.4}$	**$1.71_{\pm 0.22}$**	$3.22_{\pm 0.82}$	$2.4_{\pm 1.5}$	$2.87_{\pm 0.73}$	$2.95_{\pm 0.44}$	$3.4_{\pm 1.1}$

	Quantum AFM			Hybrid				Hardware	
Task	AFM_{mean}	AFM_{stat}	AFM	FM+AFM	RRWP+FM	RRWP+AFM	RRWP+FM+AFM	HW-AFM	RRWP+HW-AFM
Node Prediction									
Bridge Node (Acc ↑)	$0.87_{\pm 0.00}$	$0.93_{\pm 0.01}$	$0.97_{\pm 0.01}$	$0.97_{\pm 0.01}$	$0.98^{\dagger}_{\pm 0.01}$	$0.98^{\dagger}_{\pm 0.01}$	$0.99^{\dagger}_{\pm 0.01}$	$0.92_{\pm 0.01}$	$0.98^{\dagger}_{\pm 0.00}$
Degree (MAE ↓) ($\times 10^2$)	$5.59_{\pm 0.62}$	$5.72_{\pm 0.45}$	$4.13^{\dagger}_{\pm 0.32}$	$4.51_{\pm 0.37}$	$15.8_{\pm 1.3}$	$4.31^{\dagger}_{\pm 0.64}$	$4.83_{\pm 0.39}$	$7.79_{\pm 0.26}$	$7.91_{\pm 0.79}$
Betweenness (MAE ↓) ($\times 10^2$)	$5.96_{\pm 0.36}$	$5.56_{\pm 0.27}$	$5.53_{\pm 0.27}$	$5.30_{\pm 0.26}$	$4.71_{\pm 0.20}$	$4.25_{\pm 0.15}$	$4.31_{\pm 0.22}$	$5.73_{\pm 0.18}$	$\underline{4.24}_{\pm 0.12}$
Eccentricity (MAE ↓)	$1.33_{\pm 0.04}$	$1.26_{\pm 0.03}$	$1.15_{\pm 0.04}$	$1.13_{\pm 0.06}$	$1.02_{\pm 0.06}$	$1.06_{\pm 0.06}$	$1.06_{\pm 0.08}$	$1.32_{\pm 0.06}$	$1.04_{\pm 0.06}$
Min Cycle (Node) (MAE ↓)	$1.12_{\pm 0.05}$	$1.03_{\pm 0.05}$	$0.82_{\pm 0.07}$	$0.87_{\pm 0.06}$	$0.80_{\pm 0.08}$	$0.72^{\dagger}_{\pm 0.04}$	$0.79_{\pm 0.06}$	$0.92_{\pm 0.06}$	$0.71^{\dagger}_{\pm 0.07}$
Edge Prediction									
Link Pred. (Acc ↑)	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$
Bridge Edge (Acc ↑)	$0.97_{\pm 0.00}$	$0.98_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$	$0.99_{\pm 0.00}$	$1.00^{\dagger}_{\pm 0.00}$
Edge Betw. (MAE ↓) ($\times 10^3$)	$9.65_{\pm 0.87}$	$11.0_{\pm 1.0}$	$9.46^{\dagger}_{\pm 0.87}$	$10.1_{\pm 1.0}$	$10.4_{\pm 1.2}$	$8.62^{\dagger}_{\pm 0.98}$	$8.76^{\dagger}_{\pm 0.86}$	$11.82_{\pm 0.88}$	$9.3^{\dagger}_{\pm 1.2}$
Local Conn. (MAE ↓)	$0.43_{\pm 0.00}$	$0.40_{\pm 0.00}$	$0.34_{\pm 0.01}$	$0.22_{\pm 0.01}$	$\underline{0.19}_{\pm 0.01}$	$0.22_{\pm 0.01}$	**$0.17_{\pm 0.01}$**	$0.40_{\pm 0.01}$	$0.25_{\pm 0.01}$
Min Cycle (Edge) (MAE ↓)	$0.30_{\pm 0.02}$	$0.24_{\pm 0.04}$	$0.20_{\pm 0.01}$	$0.19_{\pm 0.01}$	$0.19_{\pm 0.02}$	$0.17^{\dagger}_{\pm 0.02}$	$0.18^{\dagger}_{\pm 0.02}$	$0.22_{\pm 0.01}$	$0.18^{\dagger}_{\pm 0.02}$
Graph Prediction									
Num Cycles (MAE ↓)	$0.34_{\pm 0.01}$	$0.29_{\pm 0.02}$	$0.40_{\pm 0.06}$	$0.22_{\pm 0.01}$	$0.15^{\dagger}_{\pm 0.03}$	$0.16^{\dagger}_{\pm 0.02}$	$0.14^{\dagger}_{\pm 0.03}$	$0.49_{\pm 0.06}$	$0.16^{\dagger}_{\pm 0.04}$
Diameter (MAE ↓)	$0.99_{\pm 0.10}$	$0.79_{\pm 0.04}$	$1.28_{\pm 0.14}$	$0.82_{\pm 0.15}$	$0.64_{\pm 0.05}$	$0.55_{\pm 0.02}$	$0.67_{\pm 0.11}$	$1.34_{\pm 0.17}$	$0.62_{\pm 0.08}$
Density (MAE ↓) ($\times 10^2$)	$2.24_{\pm 0.57}$	$2.29_{\pm 0.85}$	$4.6_{\pm 1.5}$	$3.7_{\pm 1.7}$	$2.5_{\pm 1.2}$	$2.4_{\pm 1.1}$	$2.8_{\pm 1.2}$	$7.2_{\pm 2.1}$	$2.4_{\pm 1.1}$
Wiener Index (MAE ↓)	$0.41_{\pm 0.07}$	$0.23_{\pm 0.01}$	$0.51_{\pm 0.10}$	$0.27_{\pm 0.06}$	$0.16_{\pm 0.03}$	$0.20_{\pm 0.04}$	$\underline{0.12}_{\pm 0.02}$	$0.53_{\pm 0.07}$	$0.17_{\pm 0.04}$
Alg. Connect. (MAE ↓) ($\times 10^2$)	$3.2_{\pm 1.5}$	$2.23_{\pm 0.19}$	$4.7_{\pm 1.7}$	$4.1_{\pm 1.6}$	$\underline{1.79}_{\pm 0.12}$	$2.28_{\pm 0.66}$	$2.67_{\pm 0.60}$	$5.7_{\pm 2.3}$	$2.32_{\pm 0.63}$

F.3.1 Hardware AFM analysis

The MUTAG table also includes hardware AFM descriptors, denoted HW-AFM, and their RRWP hybrid. HW-AFM alone is generally weaker than the simulated AFM descriptor, which is expected given sampling noise, calibration effects, and the small size of the evaluation. However, RRWP+HW-AFM is consistently competitive with the simulated hybrids, is second on betweenness (4.24×10^{-2} MAE vs. 3.89×10^{-2} for RRWP) and ties the best on smallest cycle through a node (0.71 MAE). It also ties or nearly ties the best feature sets on bridge-node detection, link prediction and bridge-edge detection, but trails on diameter (0.62 vs. 0.51) and Wiener index (0.17 vs. 0.11).

These hardware results should not be interpreted as a stable ranking against noiseless simulated descriptors, since several MUTAG differences are within fold-level variability. They nevertheless provide a useful consistency check: experimentally obtained AFM correlations still carry graph-theoretic information, and this information is most useful when combined with RRWP.

G Statistical analysis of the GNN benchmark

The purpose of this appendix is to separate two questions that are easy to conflate: whether the hybrid champion is directionally favoured across benchmark cells, and whether the associated effect size is large enough to be practically resolved. The main text reports a one-sided sign test on the twelve paired (architecture, dataset) cell-level mean differences between the classical champion (RRWP) and the hybrid champion (RRWP+AFM_{stat}+FM_{stat}), giving $k = 10$ wins out of $n = 12$

and $p = 0.019$ under a binomial null with $p_0 = 0.5$ [46]. The headline directional claim rests on this fixed-champion sign test. We complement them with three Bayesian decompositions of the same paired data, evaluated at increasingly fine sample-size resolutions, whose role is to characterise the *amplitude* of the gap rather than its direction: a cell-level analysis on the twelve mean differences, a pooled signed-rank analysis on the $n = 300$ trial-level MUTAG observations, and a hierarchical model that partial-pools the trial-level evidence across architectures. We do not apply a multiple-comparison correction because the analyses answer distinct questions; their results are reported jointly for transparency.

Bayesian sign test (cell-level). We run a Bayesian sign test [47] on the twelve paired cell-level mean differences using the `baycomp` library. To put MUTAG (F_1) and $\text{ZINC}^*_{\text{subset}}$ (MAE) on a common scale, each paired mean difference is normalised by the pooled standard deviation $\sqrt{\sigma^2_{\text{cl}} + \sigma^2_{\text{hy}}}$ of its cell, and the region of practical equivalence (ROPE) is set to ± 0.5 pooled standard deviations, following the half-pooled-SD convention discussed by Kruschke [66] and corresponding to the medium-effect threshold of Cohen's d. Since differences are normalised by $\sqrt{\sigma^2_{\text{cl}} + \sigma^2_{\text{hy}}}$, this corresponds to $d \approx 0.71$ in Cohen's convention. With the default Dirichlet-process prior ($s = 0.5$, $z_0 = 0$), the posterior reads

$$\left(P_{\text{hybrid}},\ P_{\text{rope}},\ P_{\text{classical}}\right) = \left(0.018,\ 0.982,\ 0.000\right),$$

with 95% bootstrap intervals $[0.000,\ 0.389]$, $[0.611,\ 1.000]$, and $[0.000,\ 0.000]$ respectively, and where the first entry corresponds to the hybrid champion. The bulk of the posterior mass lies in the ROPE region, faithfully reflecting that per-cell mean gaps are individually below half the pooled noise; the decomposition is stable over the ROPE range ± 0.2 to ± 0.8 pooled standard deviations (P_{rope} from 0.97 at the tightest ROPE to 0.997 at the widest), and the corresponding posterior over the population-level Cohen's d has median 0.37 with 95% HDI $[-0.21,\ +0.95]$, consistent with anywhere from a small negative to a large positive normalised effect size. The small but nonzero P_{hybrid} paired with a vanishing $P_{\text{classical}}$ is consistent with the directional sign-test conclusion of the main text without overstating the per-cell amplitude.

Bayesian signed-rank test (MUTAG, trial-level). We additionally run a Bayesian signed-rank test [47] on the trial-level MUTAG observations: 50 paired (fold, seed) F_1 scores per architecture, pooled across the six architectures for a total of $n = 300$ paired trials between the classical and hybrid champions. We focus on MUTAG at this resolution because it is the dataset where the quantum features appear to bring slightly more uplift across our sweep, and where the trial-level posterior is therefore most informative. The ROPE is set to ± 0.01 F_1 points, the value recommended by Benavoli et al. [47] for classifier comparisons; the median within-cell trial-level standard deviation across the MUTAG champion cells is 0.077 F_1, so the ROPE captures only differences smaller than one-eighth of the cross-trial noise. The posterior reads

$$\left(P_{\text{hybrid}},\ P_{\text{rope}},\ P_{\text{classical}}\right) = \left(0.907,\ 0.000,\ 0.093\right),$$

with 95% bootstrap intervals $[0.281,\ 1.000]$, $[0.000,\ 0.000]$, and $[0.000,\ 0.720]$ respectively. Augmenting the signed-rank decomposition with a Bayesian estimate of the pooled mean F_1 difference between the two champions yields a posterior median of $+0.005$ F_1 with 95% HDI $[-0.003,\ +0.013]$, an interval that lies almost entirely on the positive side of zero but whose lower edge crosses zero; the signed-rank $91/9$ asymmetry should therefore be read as a directional rather than an amplitude statement. This pooled analysis treats the $n = 300$ paired trials as exchangeable, which over-counts evidence because the 50 trials within an architecture share architecture-level effects; the wide bootstrap interval on $P_{\text{classical}}$ reflects this fragility, and the hierarchical model below corrects for the clustering directly.

Hierarchical Bayesian test. We complement the pooled signed-rank with the hierarchical Bayesian test of Benavoli et al. [47], which models a per-architecture mean difference drawn from a shared prior and partial-pools the trial-level evidence across architectures rather than treating all $n = 300$ pairs as exchangeable. With six architectures, 50 paired trials each, and the same ± 0.01 F_1 ROPE, the posterior reads

$$\left(P_{\text{hybrid}},\ P_{\text{rope}},\ P_{\text{classical}}\right) = \left(0.241,\ 0.725,\ 0.034\right),$$

with 95% bootstrap intervals $[0.000,\ 0.457]$, $[0.057,\ 1.000]$, and $[0.000,\ 0.942]$ respectively. The bulk of the posterior returns to the practical-equivalence region, in agreement with the cell-level

decomposition; the residual directional mass nevertheless leans toward the hybrid side, in agreement with the sign tests. The associated posterior over the global mean F_1 difference has median $+0.005$ with 95% HDI $[-0.009, +0.019]$, the between-architecture standard deviation τ has median 0.006 F_1 with 95% HDI $[0.001, 0.026]$, suggesting non-negligible between-architecture heterogeneity, and the per-architecture posteriors all assign 95% HDIs to mean F_1 differences that include zero, with point medians from $+0.002$ (GeneralConv) to $+0.011$ (GCN). The qualitative decomposition is stable across the ROPE range ± 0.005 to ± 0.02 F_1: the bulk of the mass shifts from the directional region at the tightest ROPE ($P_{\text{hybrid}} = 0.56$) to overwhelmingly within the equivalence region at the widest ($P_{\text{rope}} = 0.97$), while the residual directional mass remains more concentrated on the hybrid side than on the classical side. We read the hierarchical posterior as the most defensible amplitude statement available: the per-architecture mean gap between the classical and hybrid champions on MUTAG is small with non-trivial probability of being practically zero, no individual architecture's mean is statistically resolvable from zero on its own, and the directional signal lives in the consistency of the per-architecture point estimates and in the residual asymmetry between P_{hybrid} and $P_{\text{classical}}$.

Independence of cells. The cell-level sign tests treat the twelve (architecture, dataset) cells as exchangeable observations. Because the six architectures of each dataset are evaluated on the same train/validation/test splits (the standard PyG $\text{ZINC}^*_{\text{subset}}$ partition and a fixed stratified 10-fold partition for MUTAG with random state 42), the per-cell mean differences may be correlated within each dataset. We therefore treat the sign-test p-values as descriptive evidence of directional consistency rather than as exact type-I error guarantees; this is also why the Bayesian analyses above explicitly separate the cell-level, pooled trial-level, and architecture-hierarchical resolutions.

H Linear and nonlinear regression of quantum features

In this appendix, we explain how we assess the capacity of combinations of classical features (powers of adjacency matrix or RRWP) to represent QGFs through linear and nonlinear regression experiments discussed in Sec. 3.2.

Per-graph target. Each molecular graph $\mathcal{G}_\alpha$ of the dataset, with $\alpha \in \{1, \dots, N_{\text{data}}\}$, is associated with a number of atoms N_α, a classical adjacency matrix A_α and a row-normalized adjacency matrix $M_\alpha = D_\alpha^{-1} A_\alpha$, as well as—after the quantum feature extraction pipeline of Sec. 2—with a sampled correlation matrix $C^\alpha_{i,j}(t)$ [cf. Eq. (2)]. We use the $M^{\text{off}}_\alpha = M^{\text{tot}}_\alpha - N_\alpha$ off-diagonal (and independent) correlation entries to construct the target vector $y_\alpha(t) \in \mathbb{R}^{M^{\text{off}}_\alpha}$, where $M^{\text{tot}}_\alpha = N_\alpha(N_\alpha + 1)/2$.

Pooling. For each graph $\mathcal{G}_\alpha$ with node-degree matrix D_α, we construct two pair-feature families indexed by $k = 0, \dots, K$,

$$\phi^{\text{walk}}_k(A_\alpha)_{ij} = (A^k_\alpha)_{ij}, \qquad \phi^{\text{rrwp}}_k(A_\alpha)_{ij} = ((D^{-1}_\alpha A_\alpha)^k)_{ij}, \tag{44}$$

i.e. the number of length-k walks from i to j [67], and the k-step random-walk transition probability respectively (RRWP, as in GRIT [5]). Each pair (i, j) contributes one row $[\phi_0, \dots, \phi_K]$ to a per-graph design $X^{\text{walk}}_\alpha, X^{\text{rrwp}}_\alpha \in \mathbb{R}^{M^{\text{off}}_\alpha \times (K+1)}$, which we then concatenate row-wise across all N graphs:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \in \mathbb{R}^{P \times (K+1)}, \qquad y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_N(t) \end{bmatrix} \in \mathbb{R}^P, \qquad P = \sum_{\alpha=1}^{N_{\text{data}}} M^{\text{off}}_\alpha. \tag{45}$$

Rows from different graphs share column indices $k = 0, \dots, K$, so a single coefficient vector $\beta \in \mathbb{R}^{K+1}$ is fit jointly to *all* pairs of *all* graphs. This is what we mean by "pooled": one regression, P data points, no per-graph parameters. The reported pooled R^2 is one number per (K, t, family) cell, computed against the global mean $\bar{y}(t) = P^{-1} \sum_p y_p(t)$ (MUTAG: $P = 30{,}505$ pairs over $N_{\text{data}} = 188$ graphs; ZINC: $P \approx 3{\times}10^6$ over $N_{\text{data}} = 11{,}368$).

Column normalization. Columns of X are rescaled by their max absolute value, $X^{\text{norm}}_{:,k} = X_{:,k} / \max(\max_p |X_{p,k}|, 1)$, to avoid $O(\lambda^k_{\max})$ overflow at high K (without this, $(A^k)_{ij}$ overflows `float64` around $K \approx 30$ on dense graphs), where $\lambda_{\max}$ is the maximal eigenvalue.

Probes. The linear probe solves ordinary least squares (OLS) regression, $\hat{\beta} = \arg\min_\beta \|X^{\text{norm}}\beta - y\|_2^2$, via the SVD-based Moore–Penrose pseudoinverse [68] (`numpy.linalg.lstsq`, `rcond=None`) and reports $R^2 = 1 - \text{SS}_{\text{res}}/\text{SS}_{\text{tot}}$ against the pooled mean. Nonlinear probes use kernel ridge with an RBF kernel [69] (Nyström-approximated [70] with $n_{\text{components}} = 500$, $\alpha = 10^{-3}$, bandwidth from the median pairwise-distance heuristic [71]; verified within 1 pp of exact KernelRidge) and random forest [72] (100 trees, `max_depth=20`, `min_samples_leaf=10`), scored by 5-fold held-out R^2 (`scikit-learn` implementations [73]).

Laplacian augmentation. We additionally augment each design X_α with the first K eigenpairs $\{u_k(i),\, u_k(j),\, \lambda_k\}_{k=1}^{K}$ (sorted by ascending eigenvalue λ, zero-padded when $n_\alpha < K$, eigenvector signs canonicalized by largest-magnitude entry) of the symmetric normalized Laplacian $L = I - D^{-1/2}AD^{-1/2}$ [74], chosen for its bounded spectrum $\lambda_k \in [0, 2]$ that is comparable across graphs of different size.

H.1 Regression results

We run the probes of App. H on MUTAG (188 graphs, $n \in [10, 28]$) and $\text{ZINC}^*_{\text{subset}}$ (11,368, $n \in [9, 37]$).

Linear probe (OLS). At $t \leq 30$ ns we observe $R^2 = 1.000$ on both pools. At measurement times the explained variance plateaus (Table 15, walk-power, AFM connected, time-averaged on $t \in \text{linspace}(100, 6000)$ ns):

Table 15: Pooled OLS R^2 of $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ on walk-power features. K-saturation is sharp ($\leq 1\%$ from K=7 to K=50).

dataset	$R^2_{K=1}$	$R^2_{K=7}$	$R^2_{K=21}$	$R^2_{K=50}$
MUTAG	0.59	0.78	0.83	0.83
ZINC	0.59	0.76	0.78	0.78

Per-graph ceiling. The plateau is also *per-graph*, not a pooling artifact: at the Cayley–Hamilton bound $K = n_\alpha - 1$ [75], beyond which $\{I, A_\alpha, \ldots, A_\alpha^K\}$ stops gaining rank, the per-graph saturated R^2 on MUTAG has median 0.876 (IQR $[0.855, 0.882]$) for AFM walk-power, with a worst-case graph (MUTAG #16) capping at $R^2 \approx 0.76$; RRWP gives essentially identical numbers (AFM median 0.879, worst case 0.75 for MUTAG #149). We additionally verify that pooled R^2 and the per-graph mean $R^2_{\text{pg}} = N_{\text{data}}^{-1} \sum_\alpha R^2_\alpha$ (regression refit independently per graph) agree to within a few percentage points, ruling out the possibility that the pooled headline number is dominated by between-graph correlator-magnitude variance.

Nonlinear probe (KR + RF). To show that the ceiling is not just the linearity of OLS, we fit the two nonlinear regressors above on the *same* column-normalized designs. MUTAG, $K = 21$, mid-plateau $t = 1536$ ns; the 5 folds are formed at the graph level, so that pairs of the same graph never appear in both training and test sets:

Table 16: Nonlinear probes close at most 7–8% of the linear gap for walk powers (OLS at the same time; graph-level folds, fold standard deviations ≤ 0.008 for AFM and ≤ 0.033 for FM). Two distinct ceilings separate cleanly.

target / family	OLS pool	KR (RBF)	RF	residual
AFM connected, walk	0.83	0.895	0.906	∼9%
AFM connected, RRWP	–	0.898	0.909	∼9%
FM connected, walk	0.70	0.753	0.779	∼22%
FM connected, RRWP	–	0.693	0.751	∼25%

The OLS-to-KR/RF gap is the *nonlinear* content of $\boldsymbol{\mathcal{C}}_{ij}^{(2)}$ that walk-powers carry — what an MPNN with a nonlinear readout could absorb. The KR/RF-to-1.0 gap is content that is **not a function of**

any walk-power / RRWP feature: it requires features outside the adjacency-power algebra (e.g. eigenvector geometry). Walk-powers and RRWP give numerically indistinguishable nonlinear-probe ceilings (within $\leq 1\%$), confirming the residual is a property of the feature family rather than of either specific parameterization.

Laplacian augmentation. Augmenting the design with normalized-Laplacian eigenpairs (App. H) localizes where the residual lives. MUTAG, RF 5-fold R^2, $K = 21$, $t = 1536$ ns, graph-level folds:

Table 17: Adding Laplacian eigenpairs lifts FM-conn from 0.75–0.78 to 0.84–0.85 — the FM "beyond-walk" residual is partly spectral. The AFM residual is mostly walk-content under nonlinearity (Laplacian augmentation adds $\leq 1\%$).

target	walk	walk+Lap	RRWP	RRWP+Lap
AFM connected	0.906	0.912	0.909	**0.913**
FM connected	0.779	**0.847**	0.751	0.840

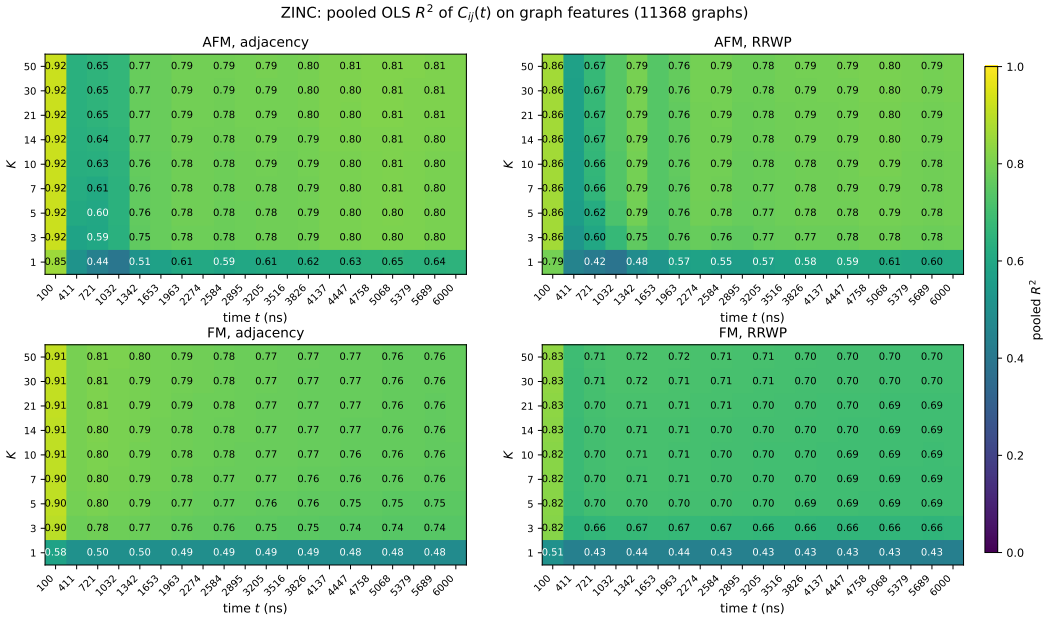


Figure 8: Pooled OLS results showing the R^2 score at different time steps for the AFM and FM regimes using K classical features (A^k for adjacency or $(D^{-1}A)^k$ for RRWP) on the $\text{ZINC}^*_{\text{subset}}$ dataset.

I Quantum graph features beyond classical structural encodings: details

This appendix details the controls summarized in Sec. 3.2 and Sec. 5.

I.1 Expanded reconstruction basis and residual

Table 18 extends the reconstruction of App. H with a cumulatively expanded basis. Beyond walk powers, RRWP and Laplacian eigenpairs, it adds ring-membership indicators; heat-kernel, geodesic-distance and effective-resistance pair features; the register distances r_{ij} and couplings $J_{ij} = C_6/r_{ij}^6$; and three restricted physical approximations of the same quench: classical-spin dynamics (discrete truncated Wigner approximation, dTWA), exact two-atom correlations under the same couplings, and mean-field correlations. All nonlinear scores use graph-level folds. With the full basis, the random forest leaves 8.3% (AFM) and 14.0% (FM) of the variance unexplained. The additive reconstruction process is shown in Fig. 10. The register geometry adds nothing once the graph descriptors are

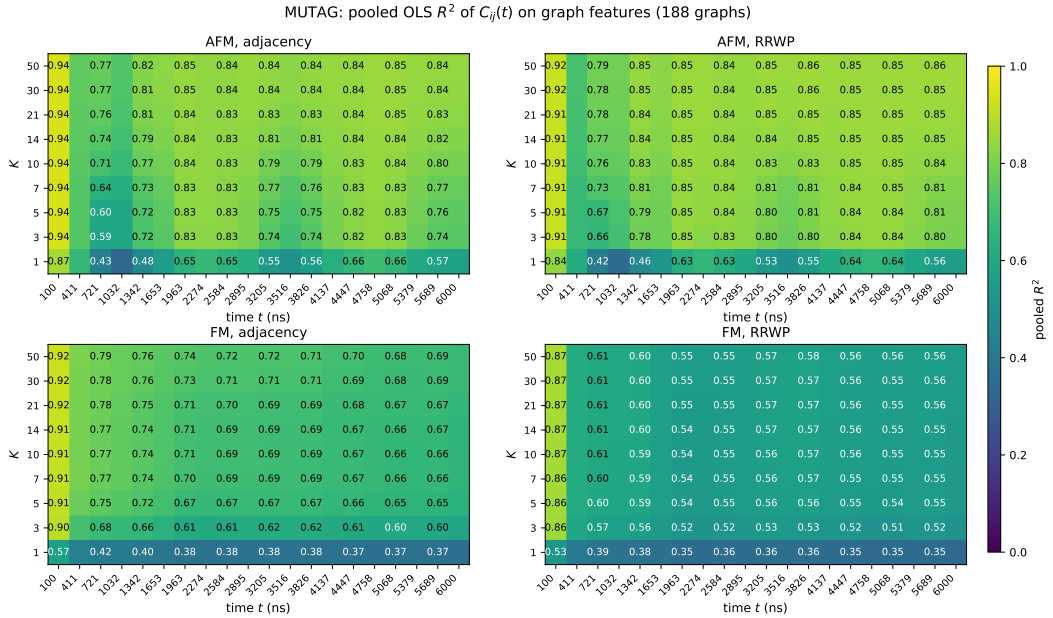


Figure 9: Pooled OLS results showing the R^2 score at different time steps for the AFM and FM regimes using K classical features (A^k for adjacency or $(D^{-1}A)^k$ for RRWP) on the MUTAG dataset.

present, and the restricted physical approximations add at most 0.5 percentage points (FM), so the residual is explained neither by the layout nor by these approximations.

Table 19 characterizes the residual on $\mathrm{ZINC}^*_{\mathrm{subset}}$. It is small in amplitude (0.03–0.05 at $d = 1$, against 0.25–0.43 for the initial correlations) but decays slowly with distance, so that it accounts for most of the long-range correlations (AFM: $d \gtrsim 4$; FM: $d \gtrsim 7$). In the normalized-Laplacian eigenbasis, the FM residual is concentrated in global, low-λ modes (49% of its energy below $\lambda = 0.04$), whereas the AFM residual is spread across the spectrum. Used as a standalone input to the intrinsic probe, the AFM+FM residual retains 97.4% of the performance of the full features on graph-level tasks and 82.5% on node-level tasks, against 42.4% on edge-level tasks, which mostly depend on short-range structure. No single graph invariant explains it: controlling for graph size, the mean absolute Spearman correlation between the residual and the benchmark targets is 0.46 for graph-level, 0.33 for edge-level and 0.17 for node-level targets, with the strongest correlations for the number of cycles (-0.57), density (-0.55) and algebraic connectivity ($+0.46$). A linear fit of the long-range residual ($d \geq 4$) on classical-spin (dTWA) correlations gives $R^2 < 10^{-5}$ for AFM and $R^2 = 0.004$ for FM.

The residual is a low-amplitude, longer-ranged component, concentrated in low-λ Laplacian modes for FM and spread across the spectrum for AFM; alone, it retains 97% (83%) of the full-feature probe performance on graph-level (node-level) tasks.

Table 18: Reconstruction of the off-diagonal correlations on MUTAG from an expanded classical basis ($K = 21$, reference time $t = 1536$ ns). Random forest, R^2 on graph-level 5-fold held-out data $\pm$ fold standard deviation; each row adds the indicated descriptors to all previous ones. Last row: unexplained variance $1 - R^2$ of the full basis.

Basis	AFM	FM
Walk powers, RRWP, Laplacian eigenpairs (Sec. 3.2)	0.915 ± 0.005	0.848 ± 0.041
+ ring, heat-kernel, geodesic and effective-resistance descriptors	0.917 ± 0.005	0.856 ± 0.036
+ register geometry (r_{ij}, J_{ij})	0.917 ± 0.005	0.855 ± 0.036
+ physical controls (dTWA, exact two-atom, mean-field)	0.917 ± 0.005	0.860 ± 0.031
Unexplained variance	8.3%	14.0%

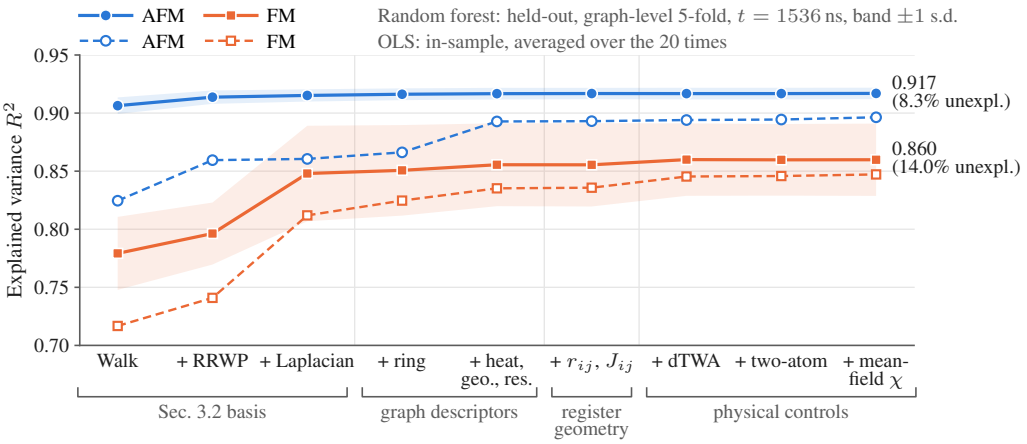


Figure 10: Explained variance R^2 of the off-diagonal correlations on MUTAG along the nine cumulative classical bases ($K = 21$); brackets group the steps into the four rows of Table 18, each row being the last step of its bracket. Solid: random forest, R^2 pooled over graph-level 5-fold held-out predictions at $t = 1536$ ns, band ± 1 standard deviation of the per-fold R^2; dashed: in-sample OLS, R^2 averaged over the 20 times. The band mainly reflects which graphs fall in each fold; the folds are identical at every step, so it is largely shared across steps and does not measure the uncertainty of the step-to-step changes. With the full basis the random forest leaves 8.3% (AFM) and 14.0% (FM) of the variance unexplained; register geometry adds nothing and the physical controls add at most 0.5 percentage points (FM). χ is the static mean-field susceptibility $(1 - \beta J)^{-1}$ of the register couplings.

I.2 Geometry-augmented baselines

The register geometry carries substantial information on its own (the couplings predict degree almost perfectly), so we compare against RRWP+Laplacian+Rings+Geometry (Sec. 2.5). On the intrinsic benchmark on $\text{ZINC}^*_{\text{subset}}$, adding either the AFM+FM correlations or only their unexplained residual (Sec. 3.2) to this baseline improves betweenness, bridge-node detection, edge betweenness and local connectivity beyond one standard error; eight tasks end in a statistical tie and the three cycle tasks are excluded, since the ring descriptors determine their targets almost by construction (Table 20).

The geometry-augmented baseline (Base) adds ring descriptors and the register distances r_{ij} and couplings J_{ij} to RRWP+Laplacian. For the GNN backbones, the geometry enters as five node-level and three pair-level register channels together with cycle-basis ring descriptors; for the GT, as the pair channels $(C_6/r_{ij}^6, r_{ij})$. Table 20 reports the intrinsic benchmark. Geometry alone is highly informative (the couplings predict degree almost exactly), and Base already improves over RRWP+Laplacian on most tasks. On top of it, both the AFM+FM correlations and their residuals improve betweenness, bridge-node detection, edge betweenness and local connectivity with disjoint standard-error bars, the residuals by a slightly larger margin. Eight tasks end in a statistical tie (for the residuals, density and algebraic connectivity are marginally worse than Base, at the edge of the standard-error bars), and the three cycle tasks are excluded. The geometry thus accounts for a substantial part of the signal, but not for all of it.

I.3 Placement sensitivity

We displace every atom of four $\text{ZINC}^*_{\text{subset}}$ graphs by Gaussian noise of standard deviation σ (relative to the minimal inter-atomic distance) and compare the AFM correlations with the reference ones (Table 21).

We distinguish small positional noise around a fixed layout from alternative valid layouts. For the first, every atom i with reference position $\boldsymbol{r}_i$ is displaced to $\boldsymbol{r}'_i = \boldsymbol{r}_i + \boldsymbol{\epsilon}_i$ with $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma^2 I_2)$, where the two Cartesian components are drawn independently and σ is given in units of the minimal inter-atomic distance. For the second, we rerun the force-based placement with different initializations

Table 19: Structure of the unexplained residual of the off-diagonal correlations on $\mathrm{ZINC}^{*}_{\mathrm{subset}}$ (time index 7), compared with the initial correlations. (a) Cumulative energy of the correlation matrix in the normalized-Laplacian eigenbasis up to eigenvalue λ. (b) Mean absolute value against shortest-path distance d (ZZ convention, $4C_{i,j}$). (c) Retention of probe performance when the residual replaces the full features, $(\text{random} - \text{residual})/(\text{random} - \text{full})$, averaged per task category.

(a) Cumulative spectral energy

λ	AFM initial	AFM residual	FM initial	FM residual
0.04	0.050	0.018	0.737	0.493
0.54	0.194	0.111	0.810	0.769
1.04	0.289	0.351	0.848	0.928
1.54	0.344	0.650	0.905	0.979
1.88	0.527	0.888	0.964	0.997

(b) Mean $|4C_{i,j}|$ versus distance

d	AFM initial	AFM residual	FM initial	FM residual
1	0.245	0.034	0.426	0.047
3	0.049	0.031	0.153	0.048
5	0.026	0.024	0.064	0.041
10	0.021	0.021	0.036	0.033

(c) Retention of probe performance

	Graph-level	Node-level	Edge-level
AFM+FM residual	97.4%	82.5%	42.4%

Table 20: Intrinsic information benchmark on $\mathrm{ZINC}^{*}_{\mathrm{subset}}$ with the geometry-augmented baseline. Mean $\pm$ standard error over 5 folds; MAE (↓) or accuracy (↑). Base: RRWP+Laplacian+Rings+Geometry (r_{ij}, J_{ij}); the last two columns add the AFM+FM correlations or only their unexplained residuals (Sec. 3.2) to it. Bold: improvement over Base with disjoint ± 1 standard-error bars. Cycle tasks (–) are excluded for the ring-augmented columns, since the ring descriptors determine their targets almost by construction.

Task	RRWP+Laplacian	Base	Base+AFM+FM	Base+residuals
Degree$^{\downarrow}$ ($\times 10^2$)	$6.98_{\pm 1.06}$	$0.84_{\pm 0.25}$	$0.86_{\pm 0.21}$	$0.83_{\pm 0.28}$
Betweenness$^{\downarrow}$ ($\times 10^2$)	$4.76_{\pm 0.13}$	$2.01_{\pm 0.08}$	$\mathbf{1.79}_{\pm 0.04}$	$\mathbf{1.74}_{\pm 0.06}$
Bridge Node$^{\uparrow}$	$0.950_{\pm 0.002}$	$0.992_{\pm 0.000}$	$\mathbf{0.995}_{\pm 0.000}$	$\mathbf{0.996}_{\pm 0.000}$
Eccentricity$^{\downarrow}$	$1.081_{\pm 0.018}$	$0.632_{\pm 0.006}$	$0.648_{\pm 0.015}$	$0.631_{\pm 0.003}$
Min Cycle (Node)$^{\downarrow}$	$0.603_{\pm 0.012}$	–	–	–
Link Pred.$^{\uparrow}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$
Edge Betw.$^{\downarrow}$ ($\times 10^3$)	$6.09_{\pm 0.43}$	$3.66_{\pm 0.08}$	$\mathbf{2.98}_{\pm 0.11}$	$\mathbf{2.67}_{\pm 0.09}$
Min Cycle (Edge)$^{\downarrow}$ ($\times 10^2$)	$10.23_{\pm 0.47}$	–	–	–
Bridge Edge$^{\uparrow}$	$0.994_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$
Local Conn.$^{\downarrow}$ ($\times 10^2$)	$7.39_{\pm 0.17}$	$2.79_{\pm 0.04}$	$\mathbf{2.43}_{\pm 0.08}$	$\mathbf{2.34}_{\pm 0.09}$
Num Cycles$^{\downarrow}$ ($\times 10^2$)	$5.90_{\pm 0.25}$	–	–	–
Diameter$^{\downarrow}$	$0.473_{\pm 0.022}$	$0.295_{\pm 0.010}$	$0.290_{\pm 0.010}$	$0.280_{\pm 0.010}$
Density$^{\downarrow}$ ($\times 10^3$)	$1.14_{\pm 0.08}$	$0.99_{\pm 0.05}$	$1.12_{\pm 0.10}$	$1.16_{\pm 0.12}$
Wiener Index$^{\downarrow}$ ($\times 10^2$)	$9.61_{\pm 0.41}$	$5.89_{\pm 0.06}$	$6.24_{\pm 0.47}$	$6.48_{\pm 0.70}$
Alg. Connect.$^{\downarrow}$ ($\times 10^3$)	$2.22_{\pm 0.05}$	$1.90_{\pm 0.12}$	$1.84_{\pm 0.09}$	$2.22_{\pm 0.19}$

and keep 10 valid layouts per graph. In both cases, we re-emulate the AFM dynamics and compare the correlation matrices with the reference ones (Table 21). Results are shown in Fig. 11.

At hardware-level noise ($\sigma = 2\%$) the Spearman correlation is 0.82 ± 0.07, and 0.61 ± 0.10 at $\sigma = 5\%$; above $\sigma = 20\%$ it falls to the level of a random node permutation. Ten alternative layouts obtained by re-running the force-based placement agree less (0.56 ± 0.25 after selecting embeddings of good quality, 0.23 ± 0.29 without selection). QGFs are thus robust to small positional noise around a fixed layout but not invariant under re-embedding: they describe a graph through a specified graph-to-register protocol.

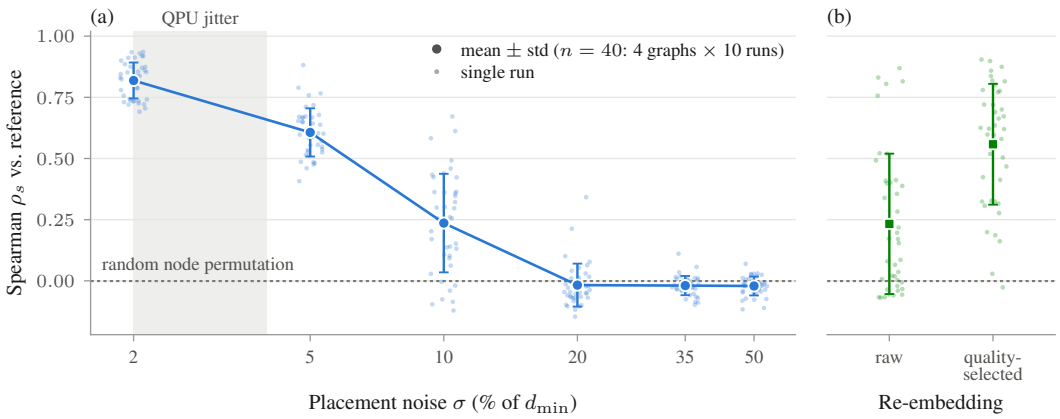


Figure 11: Sensitivity of the AFM correlations to atom placement, measured by the Spearman correlation ρ_s between the off-diagonal correlations (all 20 times pooled) of a perturbed layout and of the reference one. (a) Gaussian displacement of every atom with standard deviation σ, in units of the minimal inter-atomic distance $d_{\min} \approx 5\,\mu$m (log scale); the perturbed register is then rescaled so that its minimal distance is $5\,\mu$m. The gray band marks hardware-level jitter ($\sigma \approx 2$–4%). (b) Ten alternative layouts per graph from re-running the force-based placement, without and with selection on embedding quality. Markers: mean $\pm$ standard deviation over $n = 40$ cases (4 ZINC* subset graphs $\times$ 10 runs); dots: single runs; dotted line: random node permutation (Table 21).

Hardware-level jitter ($\sigma \approx 2$–4%, i.e., 0.1–0.2 μm) is already present in our QPU run, so the hardware results of Sec. 4 partly validate the robustness to it. The similarity decays smoothly with σ and reaches the permutation floor near $\sigma = 20\%$. Re-embeddings change the long-range couplings more strongly and agree less, particularly without selection on embedding quality. QGFs are therefore descriptors of a specified graph-to-register protocol; constructing canonical or placement-invariant quantum encodings is left for future work.

Table 21: Placement sensitivity of the AFM features on four ZINC$^*_{\text{subset}}$ graphs ($N = 16, 20, 12, 11$). Spearman correlation between the off-diagonal correlations (all 20 times pooled) of the perturbed and the reference layout; mean $\pm$ standard deviation over 4 graphs $\times$ 10 runs. Positional noise: Gaussian displacement of every coordinate with standard deviation σ, in units of the minimal inter-atomic distance; hardware-level jitter corresponds to $\sigma \approx 2$–4%.

Perturbation	Spearman ρ_s
$\sigma = 2\%$	0.819 ± 0.073
$\sigma = 5\%$	0.607 ± 0.099
$\sigma = 10\%$	0.236 ± 0.201
$\sigma = 20\%$	-0.017 ± 0.088
10 re-embeddings, raw	0.233 ± 0.287
10 re-embeddings, quality-selected	0.558 ± 0.247
Random node permutation	-0.001

I.4 Ladder graphs

In Fig. 12, a ladder with m rungs (T_i, B_i) has two alternating configurations of the blockade (maximum independent set) ground state, $T, B, T, B, \ldots$ and $B, T, B, T, \ldots$; a diagonal crossing added to a cell excludes one of them. If two crossings are separated by an even number of cells Δ, they exclude the same configuration and leave a unique ground state (pinned, $y = 1$); if they are separated by an odd number of cells, they exclude opposite configurations and leave a degenerate ground-state manifold (frustrated, $y = 0$). Partners of opposite label thus differ only by moving one crossing by one cell; labels are computed exactly by a transfer-matrix recursion over the rungs.

To exclude short random walks as an explanation, we enumerate all ladders with 9–13 rungs and group them by their pooled RRWP-2 signature (sorted entries of M and M^2, $M = D^{-1}A$), keeping

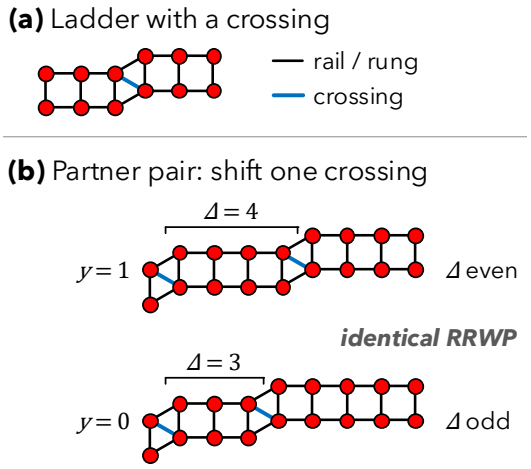


Figure 12: Ladder graphs (a) used to showcase the ability of QGFs to distinguish two classes y (b) characterised by even/odd shifts Δ of the crossing.

only groups that hold both labels, balanced per group. Within a group, every pooled RRWP-2 and ring feature is identical, so these features are at chance by construction. This yields 394 graphs in 98 groups, with 18–26 atoms. The registers use an exact unit-disk layout with square cells and rhombic crossing cells ($r_{\mathrm{NN}} = 5.01\,\mu$m; all non-edges at distance $\geq \sqrt{2}\, r_{\mathrm{NN}}$). The AFM and FM features are noiseless expectation values of the occupations and connected correlations at the 20 times of Sec. 2.4 (matrix-product-state emulation, bond dimension 16), pooled over nodes, pairs and edges into permutation-invariant graph-level vectors; the classical families are pooled in the same way. A standardized logistic regression is trained on 5 folds grouped by blind group, repeated 3 times.

Table 22 shows that random-walk and geometric descriptors, alone or combined with spectral and ring descriptors, remain far below the quantum features, whereas AFM alone reaches 0.914 ROC-AUC and AFM+FM 0.941; adding all classical features to AFM+FM changes little. The family is synthetic and its label is tied by construction to frustration and ground-state degeneracy; the result is therefore an existence statement, not evidence of a generic advantage on molecular benchmarks.

Table 22: Ladder graphs: pinned versus frustrated ground state (394 graphs, 98 RRWP-2-blind groups, 18–26 atoms). Logistic regression, 5 folds grouped by blind group × 3 repeats; mean ± standard deviation over the 15 test folds. Label-permutation test (100 shuffles): $p = 0.010$, the minimum, for all feature sets except RRWP ($p = 0.020$) and Geometry ($p = 0.446$).

Feature set	Accuracy	F_1	ROC-AUC
RRWP	0.557 ± 0.048	0.559 ± 0.047	0.612 ± 0.065
Geometry	0.503 ± 0.053	0.495 ± 0.080	0.501 ± 0.063
RRWP+Lap+Ring+Geom	0.641 ± 0.077	0.643 ± 0.091	0.680 ± 0.065
FM	0.622 ± 0.063	0.620 ± 0.071	0.670 ± 0.065
AFM	0.847 ± 0.045	0.849 ± 0.045	0.914 ± 0.026
AFM+FM	0.875 ± 0.043	0.875 ± 0.044	$\mathbf{0.941} \pm 0.022$
RRWP+Lap+Ring+Geom+AFM+FM	$\mathbf{0.882} \pm 0.043$	$\mathbf{0.884} \pm 0.043$	$\mathbf{0.941} \pm 0.023$

I.5 Non-unit-disk couplings

To probe couplings beyond what current two-dimensional registers implement, we compute exact state-vector dynamics for the 173 MUTAG graphs with $N \leq 24$, either with the register couplings C_6/r_{ij}^6 or with the ideal graph Hamiltonian $J_{\mathrm{NN}}A$ (cf. Eq. (3)), i.e., without residual non-edge couplings, as accessible with three-dimensional registers or digital implementations. Table 23 compares both, and the 1000-shot emulated features, on top of the geometry-augmented baseline for the five message-passing and attention backbones of Sec. 4. The exact features slightly improve over the baseline ($+0.011$ and $+0.009$ mean F_1, $p \approx 0.01$–0.02), and the ideal graph Hamiltonian

performs on par with the register couplings. This gain is not robust to the choice of backbones: including the two further backbones of the original run (GATv2 and GEN), none of the paired differences is significant ($p > 0.08$). The downstream signal therefore persists, but without a clear gain, when the couplings are replaced by those of the ideal graph Hamiltonian.

Table 23: MUTAG, the 173 graphs with $N \leq 24$: geometry-augmented baseline (RRWP+Laplacian+Rings+Geometry) versus its extension by AFM+FM features from the 1000-shot emulation, from exact state-vector dynamics with the register couplings C_6/r^6, and from exact dynamics of the graph Hamiltonian $J_{\mathrm{NN}}A$ (not realizable by a unit-disk register). Mean macro F_1 (↑) $\pm$ standard deviation over 10 stratified folds $\times$ 5 seeds; bold: best per backbone. Paired Wilcoxon tests over the 250 (backbone, fold, seed) trials, against the baseline: emulated $+0.004$ ($p = 0.31$); exact C_6/r^6 $+0.011$ ($p = 0.014$); exact $J_{\mathrm{NN}}A$ $+0.009$ ($p = 0.021$).

	Baseline	+AFM+FM emulated	+AFM+FM exact, C_6/r^6	+AFM+FM exact, $J_{\mathrm{NN}}A$
GCN	$0.853_{\pm 0.071}$	$0.851_{\pm 0.075}$	$0.856_{\pm 0.087}$	$\mathbf{0.857}_{\pm 0.085}$
GAT	$0.845_{\pm 0.071}$	$\mathbf{0.856}_{\pm 0.087}$	$0.852_{\pm 0.076}$	$0.852_{\pm 0.080}$
GINE	$0.845_{\pm 0.083}$	$0.854_{\pm 0.083}$	$\mathbf{0.860}_{\pm 0.061}$	$0.856_{\pm 0.076}$
Transformer	$0.854_{\pm 0.072}$	$0.859_{\pm 0.088}$	$\mathbf{0.872}_{\pm 0.071}$	$0.864_{\pm 0.071}$
General	$0.850_{\pm 0.083}$	$0.845_{\pm 0.099}$	$0.864_{\pm 0.090}$	$\mathbf{0.865}_{\pm 0.082}$
Mean	0.850	0.853	**0.861**	0.859

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: First, using two physically motivated operating regimes of the graph-encoded Rydberg Hamiltonian, antiferromagnetic (AFM) and ferromagnetic (FM) [Sec. 2.3], we introduce an intrinsic benchmark of 15 topological and spectral prediction tasks evaluated with permutation-invariant probes, showing that hybrid quantum–classical descriptors match or surpass purely classical baselines based on RRWP, RWSE and Laplacian eigenvectors, with AFM and FM correlations emphasizing different structural aspects of the graph [Sec. 3.1]. Second, regression analyses indicate that QGFs are not accurately recoverable from the classical descriptors considered here [Sec. 3.2], even after adding ring, distance, register-geometry and restricted physical descriptors. We also introduced geometry-augmented baselines, atom-placement perturbations and a synthetic family of ladder graphs to delimit where this residual information helps [App. I.2]. Third, we show that concatenating QGFs with molecular features yields small but consistent gains across six message-passing, attention-based, and graph-transformer backbones, including GCN [1], GINE [21], GeneralConv [22], GAT [3], Transformer [23, 24], and our customized graph transformer (GT) [Sec. 4.1]; we further validate the trend on ESA [25], a graph architecture outside the standard transformer family [Sec. 4.2]. Finally, we execute the full feature-extraction pipeline on a neutral-atom QPU for all 188 MUTAG graphs at four measurement time steps.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The following limitations are stated and discussed: limitations in embeddability (restricted to 2D unit-disk graphs on current hardware), compute time on current QPUs, and small magnitude of observed improvements over classical baselines.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A].

   Justification: The paper does not include theoretical results.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Methodological details are given throughout the paper. Given the numerous experiments and calculations performed, the code and data is available for full reproducibility.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: All code, datasets, features and raw QPU data are available, on github and zenodo cited at the end of main text.

   Code includes: QGF's emulation pipeline, intrinsic information benchmark pipeline, GNN downstream evaluation pipeline, predictability of quantum features from classical descriptors, statistical analysis.

   Everything is setup (seeds, dependencies, scripts) in a way that allows quick reproductability.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Details about splits, hyperparameters and types of optimizers are given. Not all of them are given in an extensive manner, as they are numerous, but the information required to understand the work is present.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Standard deviations across folds and seeds are reported throughout the results tables. An exception is table 1, which (explicitly) omits standard deviations for clarity, which are instead reported in full in the annex.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The computational bottleneck of the paper is the calculation time of QGFs, either on a real QPU hardware or in emulation. Both are reported as resp. QPU and GPU days. Training time on all described tasks is moderate in comparison and therefore omitted.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The research respects fully the NeurIPS Code of Ethics.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Datasets and libraries used in the paper are open-licensed and cited.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

Justification: All associated data is given in a Zenodo repository: 10.5281/zenodo.23099924

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: No LLMs are used or required in the paper or the methods.