QUIVER: QUantum-Informed Views for Enhanced Representations in Large Machine Learning Models

Aritra Bal [1 2] **Michael Binder** [2] **Markus Klute** [2] **Benedikt Maier** [3] **Michael Spannowsky** [1 4]

Abstract

Large machine learning models benefit substantially from multimodal inputs that provide a complementary view of the same example. We introduce QUIVER (**QU**antum-**I**nformed **V**iews for **E**nhanced **R**epresentations), a paradigm that enriches classical data-driven features with a *quantum Fisher view*: a geometrically motivated, basis-independent summary of higher-order correlations captured by a variational quantum circuit (VQC) trained to perform the same task. Unlike classical feature augmentation, the quantum Fisher information matrix encodes the intrinsic geometry of the learned quantum state manifold. This feature map, motivated by quantum information theory, is ordinarily non-trivial to model classically. However, it can reveal statistical structure that additional classical data or model capacity finds difficult to learn, making it a complementary modality. We demonstrate that QUIVER improves standard performance metrics on two benchmark datasets from very different fields: the JETCLASS dataset for predicting jet flavor at the Large Hadron Collider (LHC), and the QM9 dataset for predicting molecule properties. The core contribution, however, is domain-agnostic: the quantum Fisher view can be fused into a broad class of model architectures via targeted modifications to the base architecture, to incorporate information about the quantum geometry of the problem. These results demonstrate that quantum-geometric features, extracted from simulated variational circuits, can deliver measurable value for standard machine learning tasks, well before the advent of fault-tolerant quantum hardware.

[1]Institute of Theoretical Physics (ITP), KIT, Karlsruhe, DE [2]Institute of Experimental Particle Physics (ETP), Karlsruhe Institute of Technology (KIT), Karlsruhe, DE [3]Blackett Laboratory, Imperial College of Science and Technology, London, UK [4]Institute for Quantum Materials and Technologies, KIT, Karlsruhe, DE. Correspondence to: Aritra Bal <aritra.bal@kit.edu>.

Preprint. June 9, 2026.

1. Introduction

Multi-million parameter models play an increasingly important role in scientific analysis of extremely high-dimensional data. Architectures such as graph neural networks (GNNs) (Gilmer et al., 2017; Battaglia et al., 2018; Scarselli et al., 2009) and transformers (Vaswani et al., 2017) have become standard tools in this regard. Two domains where this paradigm has driven significant methodological progress are high-energy physics (HEP) and molecular chemistry. In both cases the inputs are high-dimensional structured objects, particles in a jet or atoms in a molecule, yet the models that process them are trained exclusively on classical representations of these systems. While these representations are highly effective, they are ultimately restricted to correlations that can be efficiently expressed within standard feature spaces and architectures.

Critical tasks in the context of major experiments such as the Large Hadron Collider (LHC), among others, are performed by such machine learning models, two prominent examples being *jet flavor classification*, and *anomaly detection* for new-physics searches. Jets are collimated sprays of hadrons produced when a high-energy quark or gluon fragments and hadronizes; identifying their parton-level origin (e.g. light quark vs. gluon, or boosted top quark vs. QCD background) is a central ingredient of essentially every analysis at the LHC. State-of-the-art jet taggers represent each jet as a *point cloud* of its constituent particles, each carrying kinematic features (Komiske et al., 2019; Qu & Gouskos, 2020; Qu et al., 2022b), in addition to other information. Similar principles are used to design model-agnostic anomaly-detection pipelines aimed at uncovering physics beyond the Standard Model (Nachman & Shih, 2020; Kasieczka et al., 2021).

While such architectures excel at exploiting kinematic correlations in the particle-level representation of a jet, they are limited by the structure of the feature space in which these correlations are expressed. In particular, higher-order and non-local correlations must be learned implicitly through model capacity rather than being directly exposed to the model. This motivates the exploration of alternative representations that can surface such structure more directly. Jets arise from a coherent branching process governed by quantum chromodynamics, and consequently exhibit rich correla-

tion patterns among their constituents. In practice, however, these objects are represented for machine learning through classical feature constructions, usually kinematic, which can obscure or compress precisely those multi-particle correlations that are most discriminative—for example, between a color-singlet W jet and a color-connected QCD jet. This motivates the exploration of alternative representations that make such structure more directly accessible. In this work, we operationalize this idea through the notion of *quantum views* of a jet: embeddings of classical data into a Hilbert space that expose its geometric and correlation structure in ways not readily captured by standard kinematic summaries.

This structural limitation extends equally to molecular chemistry, where a central task is the prediction of quantum-mechanical molecular properties from structure, with targets such as the HOMO-LUMO gap, dipole moment, isotropic polarizability and atomization energy available in benchmark datasets like QM9 (Ramakrishnan et al., 2014). GNNs and transformers tailored to molecular graphs and atomic point clouds, SchNet (Schütt et al., 2017), MPNN (Gilmer et al., 2017), DimeNet (Gasteiger et al., 2020b) and equivariant successors, now define the state of the art on these benchmarks. These targets often depend on complex, highly correlated interactions that are only indirectly reflected in classical structural descriptors. As a result, models must infer these relationships from data rather than accessing a representation in which such correlations are naturally organized.

We address this limitation by introducing a *quantum Fisher view* of the input in which we map classical data into a parameterized quantum state via a variational quantum circuit (VQC) (Cerezo et al., 2021), and extract the associated quantum Fisher information matrix (QFIM) (Liu et al., 2020; Meyer, 2021; Abbas et al., 2021). This construction does not require an assumption that the underlying system is quantum or that the encoding reflects a physical quantum state. Rather, it provides a principled mapping of classical data into a Hilbert space where geometric structure, in particular, sensitivity and higher-order correlations, can be probed through the induced metric.

On this basis, we propose QUIVER, a paradigm that fuses the quantum Fisher view with the classical view of the same input. QUIVER is deliberately architecture-agnostic: for transformer backbones we condition the model through cross-attention between the quantum and classical modalities, and for GNN backbones we modulate the learned graph messages by features derived from the QFIM. We show that QUIVER delivers consistent improvements over classical baselines of similar or greater complexity, including the Particle Transformer (Qu et al., 2022b), a > 2 M-parameter state-of-the-art jet tagger used in LHC analyses, and DimeNet++, a state-of-the-art model on QM9 property prediction.

2. Motivating QUIVER: A Mathematical Background

In this section, we provide a short mathematical overview of a VQC and the QFIM. Thereafter, we describe the quantum encoding used for the HEP task, and the novel quantum encoding we develop for the molecular chemistry application. All quantum circuit operations described in this paper were simulated classically using the quantum simulator library PENNYLANE (Bergholm et al., 2022).

2.1. Variational quantum circuits

A VQC is a parameterized unitary $U(\boldsymbol{\theta})$ acting on a fixed reference state of N qubits (Cerezo et al., 2021). With the standard initialization $|0\rangle^{\otimes N}$, the circuit prepares first an input encoding:

$$|\psi(\boldsymbol{\Theta})\rangle \;=\; U(\boldsymbol{\Theta})\,|0\rangle^{\otimes N}, \qquad \boldsymbol{\Theta} \in \mathbb{R}^P, \tag{1}$$

where the P angles $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_P)$ are functions of the inputs: $\boldsymbol{\Theta} = \boldsymbol{\Theta}(x)$, with x a jet or a molecule. This encoding is typically followed by a series of entanglement operations and trainable single-qubit rotations $R(\theta)$, which together constitute a variational ansatz. The output of the circuit is obtained by measuring one or more qubit observables, yielding an expectation value that serves as the prediction of the VQC.

2.2. The Quantum Fisher Information Matrix

Now, because the map of inputs $\boldsymbol{\theta} \mapsto |\psi(\boldsymbol{\Theta}, \boldsymbol{\theta})\rangle$ is smooth, the resultant quantum states form a submanifold of pure states whose canonical Riemannian structure is the Fubini–Study metric ((Provost & Vallee, 1980)). On pure states this metric coincides, up to an overall factor of four, with the QFIM ((Braunstein & Caves, 1994)):

$$F_{ij}(\boldsymbol{\theta}) \;=\; 4\,\mathrm{Re}[\,\langle \partial_i\psi \,|\, \partial_j\psi\rangle - \langle \partial_i\psi \,|\, \psi\rangle\langle \psi \,|\, \partial_j\psi\rangle\,], \tag{2}$$

where $\partial_i \equiv \frac{\partial}{\partial\theta_i}$, yielding the line element

$$\mathrm{d}s^2 \;=\; \tfrac{1}{4}\sum_{i,j} F_{ij}(\boldsymbol{\theta})\,\mathrm{d}\theta_i\,\mathrm{d}\theta_j, \tag{3}$$

which measures the statistical distinguishability of two infinitesimally separated parameter points $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + \mathrm{d}\boldsymbol{\theta}$ through the states they prepare. In our setting, the input enters through the data-dependent state preparation that precedes the trainable rotations, so $F_{ij}(\boldsymbol{\theta}; x)$ is an input-conditioned object: evaluated at a fixed reference $\boldsymbol{\theta}_0$, it characterizes how the encoded state of x shapes the local geometry of the trainable-parameter manifold. This computation is tractable on existing classical simulators, using standard implementations such as that in PENNYLANE.

The diagonal F_{ii} records how strongly the prepared state responds to a perturbation of θ_i alone, and so acts as a per-feature "dynamic" importance score. The off-diagonal F_{ij} couple distinct parameters and are non-zero precisely when the two corresponding directions in input space act *coherently* on overlapping qubit subsystems; they vanish whenever the two parameters drive factorized, independent parts of the state. Under an encoding of the form described in the two previous sections, this gives a direct relational reading: large off-diagonal entries between two qubits flag collective behavior of the corresponding input elements; while a nearly diagonal F signals effectively independent contributions. The result is a compact relational tensor whose entries are directly consumable by attention layers, or by message-passing networks, as will now be detailed in the following sections.

2.3. The 1P1Q particle embedding

For jets we adopt the one-particle–one-qubit (1P1Q) encoding of Bal et al. (2025), in which each reconstructed constituent is mapped to a dedicated qubit using its kinematic features, followed by two-qubit entanglement, and standard Pauli rotation operations. We represent each jet as an ordered set of its ten highest p_{T} constituents, using $(p_{\mathrm{T}}, \eta, \phi)$ as the set of kinematic input features. Here, p_{T} is the transverse momentum (magnitude of momentum in the plane perpendicular to the collider beam axis), $\eta = -\ln(\tan(\theta/2))$ is the pseudorapidity, and θ, ϕ are the zenith and azimuthal angles. We use the standard coordinate references in collider physics where the Z axis is defined as being along the collider beam direction. We omit further details of the circuit and embedding, this being amply described in (Bal et al., 2025).

2.4. The 2A2Q molecular embedding

For the QM9 molecular dataset, we design and use a novel two-atom–two-qubit embedding which we call 2A2Q. The objective is to regress $\Delta\epsilon = \epsilon_{\mathrm{HOMO}} - \epsilon_{\mathrm{LUMO}}$, defined as the energy difference between the highest occupied (HOMO) and lowest unoccupied molecular orbitals (LUMO), on the QM9 dataset. We represent each molecule as a 10-qubit system, with one qubit assigned to each heavy atom. The unused qubits are populated with randomly sampled hydrogen atoms, and all remaining explicit hydrogen information is otherwise discarded.

Starting from the initial state $|0\rangle^{\otimes N}$, we first learn a per-atom embedding by applying $R_Y(w^j_{\mathrm{atom}})\,|0\rangle$ on each qubit j, where R_Y denotes the Pauli rotation about the Y-axis and w^j_{atom} is a trainable parameter associated with the atomic species occupying qubit j. A naive one-atom–one-qubit encoding of Cartesian coordinates would introduce a dependence on the choice of reference frame, which is undesirable. To mitigate this, we combine the encoding and entanglement stages into a single pairwise operation, defined by the angles

$$\begin{aligned} \omega_1^{(ij)} &= e_{d_1} \cdot \left(1 - \frac{d_{ij}}{d_{\mathrm{CUTOFF}}}\right) \cdot \cos(\theta_{ij}), \\ \omega_2^{(ij)} &= e_{\mathrm{bond}}^{(ij)} \cdot \pi, \\ \omega_3^{(ij)} &= e_{d_2} \cdot \left(1 - \frac{d_{ij}}{d_{\mathrm{CUTOFF}}}\right) \cdot \cos(\phi_{ij}), \end{aligned} \tag{4}$$

followed by the two-qubit unitary

$$\begin{aligned} \mathcal{U}_{ij} = \big(&I_{YY}(\omega_3^{(ij)})\, I_{ZZ}(\omega_2^{(ij)})\, I_{XX}(\omega_1^{(ij)})\big) \\ &\big(R_Y(w^i_{\mathrm{atom}}) \otimes R_Y(w^j_{\mathrm{atom}})\big)\,|00\rangle, \end{aligned} \tag{5}$$

where e_{d_1} and e_{d_2} are learnable scaling parameters, $e_{\mathrm{bond}}^{(ij)}$ is a learnable bond-type entanglement parameter, $d_{\mathrm{CUTOFF}} = 1.7$ Å is fixed, and I_{XX}, I_{YY}, I_{ZZ} denote the Ising-type two-qubit interactions $\exp(-i\,\omega\,\sigma_X \otimes \sigma_X/2)$, $\exp(-i\,\omega\,\sigma_Y \otimes \sigma_Y/2)$, and $\exp(-i\,\omega\,\sigma_Z \otimes \sigma_Z/2)$, respectively. The pairwise distance d_{ij} is frame-invariant by construction, while the pairwise zenith and azimuthal angles θ_{ij} and ϕ_{ij} retain a residual frame dependence; we accept this trade-off, as the objective of the small variational quantum circuit is not to achieve state-of-the-art performance in isolation. The entanglement block $\mathcal{U}_{ij}$ is applied only for atom pairs (i, j) satisfying $d_{ij} < d_{\mathrm{CUTOFF}}$ and connected by a chemical bond. The pairwise stage is followed by a per-qubit trainable rotation sequence $R_Z \cdot R_Y \cdot R_Z$ with independent parameters on each qubit. Together, the atom embedding, conditional pairwise entanglement, and single-qubit rotations constitute one layer of the circuit, and we stack $N = 2$ such layers in the final architecture.

The prediction for the gap $\Delta\epsilon$ is extracted from the VQC via measurement of the observable:

$$\mathcal{H} = \sum_{i=1}^{N} c_i\, Z_i, \tag{6}$$

where the index i runs over all $N = 10$ qubits of the system, Z_i denotes the Pauli-Z operator acting on the i-th qubit, and $\{c_i\}$ are trainable coefficients. Since the HOMO–LUMO gap is strictly positive, the raw expectation value $\langle\mathcal{H}\rangle \in [-\sum_i |c_i|, \sum_i |c_i|]$ is shifted by $\sum_i |c_i|$ so that the predicted value lies in $[0, 2\sum_i |c_i|]$. The circuit parameters are optimized by minimizing the Huber loss ((Huber, 1964; PyTorch Contributors, 2024)) between the predicted gap and the target value in units of meV, which provides robustness to outliers by interpolating between an ℓ_2 behavior for small residuals and an ℓ_1 behavior for large ones.

3. Adding the Quantum Views

3.1. Jet Flavor Classification

We evaluate QUIVER on the JETCLASS dataset of (Qu et al., 2022a), focusing on the binary classification task of distinguishing hadronic top-quark jets ($t \to Wb \to q\bar{q}b$) against the QCD multijet background. As our classical baseline, we adopt the Particle Transformer (Qu et al., 2022b), a state-of-the-art model with approximately 2.14M parameters [1], and incorporate the QFIM into its attention mechanism via implicit cross-attention through sequence concatenation, as described in Section 3.1.2.

Table 1 summarizes the per-particle feature sets. The kinematic baseline uses the per-particle kinematic-only features relative to the jet axis; the full-feature baseline additionally includes calorimeter energy deposits, particle identification flags, and track impact-parameters.

Table 1. Per-particle input features for the jet tagging experiments.

Scope	Features
Kinematics	$\log(p_{T,\text{rel}})$, $\Delta\eta$, $\Delta\phi$
Full	all kinematic features above, **plus**: $\log p_T$, $\log E$, $\log E_{\text{rel}}$, ΔR, charge particle-ID flags ($e^{\pm}$, $\mu^{\pm}$, charged/neutral hadron, γ), impact-parameter features (d_0, d_z and their errors)

All models also receive four-vectors (p_x, p_y, p_z, E) for the Lorentz-vector pair embedding and $(\Delta\eta, \Delta\phi)$ as spatial point coordinates.

3.1.1. QFIM REPRESENTATION

Given the compute requirements of simulating large multi-qubit states, we restrict our implementation of the 1P1Q encoding to a maximum of $N = 10$ qubits, allowing us to use only the ten highest-p_{T} constituents of a given jet, with all subsequent per-particle information being discarded. For 10 particles encoded under the 1P1Q scheme with three local rotation-gate parameters per qubit, the QFIM is a 30×30 real symmetric matrix, stored in the data pipeline as 90 channels over 10 particle slots. Each particle slot i receives all 90 QFIM channels as its feature vector, which is embedded by a Particle-Transformer-style MLP into a token of dimension 128 and appended to the classical particle-token sequence. The transformer therefore receives 20 tokens in total: 10 classical particle tokens followed by 10 QFIM tokens.

[1] We use the official release (Qu & Li, 2022) of Particle Transformer on GitHub for the baseline

Algorithm 1 QUIVER: QFIM token injection into ParT.

Require: $\{x_i\}_{i=1}^{P}, \{v_i\}_{i=1}^{P}, \mathbf{Q}$
  for $i = 1, \ldots, P$ **do**
    $k_i \leftarrow \text{MLP}_{\text{tok}}(x_i)$
    $q_i \leftarrow \text{MLP}_{\text{qfim}}(\mathbf{Q}[:, i])$
  end for
  $b_{ij} \leftarrow \text{PairEmbed}(v_i, v_j)$
  Transformer Input $\leftarrow$ $[k_1, \ldots, k_P,\ q_1, \ldots, q_P]$

3.1.2. THE QUIVER PARADIGM: QFIM INJECTION

Let $\{x_i\}_{i=1}^{P}$ be the particle features, $\{v_i\}_{i=1}^{P}$ be the four-vectors used for pairwise embedding and $\mathbf{Q} \in \mathbb{R}^{90\times 10}$ be the reshaped QFIM with the second dimension representing each constituent particle of the jet. We incorporate the QFIM into the Particle Transformer by embedding the per-particle QFIM channels independently using a Particle-Transformer-style MLP and appending the resulting tokens to the classical particle sequence:

$$\text{transformer input} = \big[k_1, \ldots, k_P,\ q_1, \ldots, q_P\big], \quad (7)$$

where $k_i = \text{MLP}_{\text{tok}}(x_i) \in \mathbb{R}^{128}$ are the classical particle tokens and $q_i = \text{MLP}_{\text{QFIM}}(\mathbf{Q}[:, i]) \in \mathbb{R}^{128}$ are the embedded QFIM tokens, with $\mathbf{Q}[:, i]$ denoting all 90 QFIM channels associated with particle slot i. The Lorentz-vector pair bias is computed for the original particle sequence and zero-padded to the doubled sequence length. Algorithm 1 details the complete forward pass.

With the architectural modifications of Algorithm 1, the QUIVER-augmented Particle Transformer has a parameter count of 2.29M, a modest increase of 7% over the original.

3.2. Molecular Property Regression

The task, as before, remains the regression of $\Delta\epsilon$. The quantum encoding was previously described in Section 2. The QFIM is computed per molecule and stored as a 10×10 grid of 6×6 sub-blocks, corresponding to 2 circuit layers times 3 single-qubit rotations per qubit, resulting in a 60×60 matrix whose off-diagonal block Q_{ij} captures the coherent coupling between the rotation-gate parameter groups of qubits i and j, which by the conditions of the 2A2Q encoding necessarily corresponds to that between atoms i and j.

3.2.1. THE QUIVER PARADIGM: QUANTUM-INFORMED EDGE-STATE RESCALING

Rather than introducing an independent QFIM processing branch, which risks an improvement by generic parameter capacity rather than physically aligned information, we constrain the QFIM to act as a modulating factor on top of the existing baseline edge-state vectors.

We apply this mechanism to DimeNet++ (Gasteiger

Algorithm 2 QUIVER in DimeNet++: QFIM-gated edge-state rescaling.

Require: atomic numbers $\{z_i\}$, positions $\{\mathbf{r}_i\}$, QFIM sub-blocks $\{Q_{ij}\}$, learnable scalar α
CONSTRUCT: DimeNet++ radius graph and geometric basis functions: RBF_{ij} and SBF_{kji}.
$x_{ij}^{(0)} \leftarrow \textsc{Embed}^{\mathrm{DimeNet++}}(z_i, z_j, \mathrm{RBF}_{ij})$
$\widetilde{x}_{ij}^{(0)} \leftarrow \textsc{Rescale}(x_{ij}^{(0)}, Q_{ij}, \alpha)$
$o \leftarrow \textsc{Output}_{\mathrm{DimeNet++}}^{(0)}(\widetilde{x}_{ij}^{(0)}, \mathrm{RBF}_{ij})$
for $l = 1, \ldots, L$ **do**
  $x_{ij}^{(l)} \leftarrow \textsc{Interaction}_{\mathrm{DimeNet++}}^{(l)}(\widetilde{x}_{ij}^{(l-1)}, \mathrm{RBF}_{ij}, \mathrm{SBF}_{kji})$
  $\widetilde{x}_{ij}^{(l)} \leftarrow \textsc{Rescale}(x_{ij}^{(l)}, Q_{ij}, \alpha)$ {QFIM gate}
  $o \leftarrow o + \textsc{Output}_{\mathrm{DimeNet++}}^{(l)}(\widetilde{x}_{ij}^{(l)}, \mathrm{RBF}_{ij})$ {pooling}
end for
$\hat{y}_G \leftarrow \sum_{i \in G} o_i$ {graph-level sum readout}

et al., 2020a)[2], an improvement to the original DimeNet (Gasteiger et al., 2020b), which operates on directed-edge embeddings $x_{ij}^{(l)}$ updated by interaction blocks rather than explicit node messages. This rescaling is applied after the initial embedding block and after each interaction block, so the steps described in Algorithm 2 are realizable in DimeNet++'s native edge state. The QFIM-modulated rescaling results in a parameter size of $1.891\mathrm{M}$, a negligible increase over the original's $1.886\mathrm{M}$ parameters. We refer to this model as $\mathcal{Q}$DimeNet++.

The edge-state rescaling of Step 2, Algorithm 2 is implemented as a residual multiplicative gate on the baseline directed edge state $x_{ij}^{(l)}$

$$\widetilde{x}_{ij}^{(l)} = \textsc{Rescale}(x_{ij}^{(l)}, Q_{ij}, \alpha) = \Big(1 + \alpha \cdot \Theta(Q_{ij})\Big) x_{ij}^{(l)}. \tag{8}$$

Here, α is a global learnable scalar initialized to zero, ensuring that the two networks are identical in the beginning. The function $\Theta(Q_{ij})$ is a per-edge bounded scalar learnt using a convolutional neural network (CNN) applied to the 6×6 QFIM sub-block, followed by a scaling multilayer perceptron (MLP) (Rosenblatt, 1958) with a final $\tanh$ activation; ensuring $\Theta(Q_{ij}) \in [-1, 1]$. The exact details of the CNN and training setup are provided in the appendix.

The inputs and prediction targets are standardized using the statistics from the train sample, with subsequent de-standardization of the output before reporting the results. Ten independent seed initializations are run for each variant to obtain the error bars.

[2]The baseline uses the DimeNet++ implementation in the package PYTORCH Geometric (Fey & Lenssen, 2019; Contributors, 2024).

4. Results

4.1. Jet Flavor Classification

The evaluation criteria and classical benchmark was previously described in Section 3.1. We train both the Particle Transformer (ParT) and its QUIVER-augmented variant with five independent random seed initializations, and report our results in Table 2, using the AUC score and the QCD background rejection rate as our performance metrics. The latter quantity is a standard quantity used in HEP, defined as $1/\epsilon_B$. We evaluate this at a top-tagging efficiency of $\epsilon_S = 0.5$, (where $\epsilon_B = \mathrm{FPR}$, $\epsilon_S = \mathrm{TPR}$). Each run, comprising the aforementioned five initializations, is carried out twice, once for the kinematic-only set of features and then again for the full set of available inputs, both described in Table 1.

Table 2. Comparison of ParT and QUIVER on the top tagging task.

Features	Method	N	Results
Kin.	ParT	0.1 M	AUC $= 0.97140 \pm 0.00038$ $1/\epsilon_B = 107 \pm 2$
		0.5 M	AUC $= 0.97629 \pm 0.00015$ $1/\epsilon_B = 146 \pm 3$
		5 M	AUC $= 0.97832 \pm 0.00004$ $1/\epsilon_B = 176 \pm 1$
	QUIVER	0.1 M	AUC $= 0.97368 \pm 0.00013$ $1/\epsilon_B = 130 \pm 1$
		0.5 M	AUC $= 0.97848 \pm 0.00045$ $1/\epsilon_B = 191 \pm 9$
		5 M	AUC $= 0.98070 \pm 0.00003$ $1/\epsilon_B = 240 \pm 1$
Full	ParT	0.1 M	AUC $= 0.98875 \pm 0.00008$ $1/\epsilon_B = 570 \pm 13$
		0.5 M	AUC $= 0.99080 \pm 0.00017$ $1/\epsilon_B = 921 \pm 13$
		5 M	AUC $= 0.99235 \pm 0.00003$ $1/\epsilon_B = 1306 \pm 8$
	QUIVER	0.1 M	AUC $= 0.98893 \pm 0.00005$ $1/\epsilon_B = 590 \pm 7$
		0.5 M	AUC $= 0.99095 \pm 0.00003$ $1/\epsilon_B = 951 \pm 17$
		5 M	AUC $= 0.99244 \pm 0.00003$ $1/\epsilon_B = 1362 \pm 28$

These results demonstrate that the QUIVER paradigm contributes to improving the performance of even large classical state-of-the-art baseline model such as the Particle Transformer in terms of community-standard performance metrics.

4.2. Molecular Property Regression

We evaluate QUIVER on the benchmark described in Section 3.2.1, with the results reported as mean $\pm$ standard deviation over ten independent seed initializations.

Figure 1 shows the validation MAE curves over training for both DimeNet++ variants.

The upper panel shows that $\mathcal{Q}$DimeNet++ consistently achieves lower validation MAE than the baseline, with the gap opening early in training and persisting through convergence. The lower panel shows the per-epoch difference between DimeNet++ and $\mathcal{Q}$DimeNet++, which remains positive throughout training.

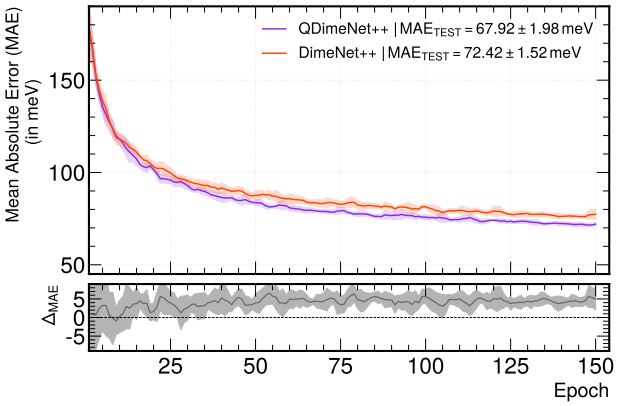

Figure 1. Validation MAE during training for the DimeNet++ baseline and the quantum-inspired $\mathcal{Q}$DimeNet++ model. Solid lines show the mean across $N = 10$ paired seeds (identical data splits and initialization RNG state across the two models; only the architecture differs), smoothed with a 3-epoch rolling window. Shaded bands denote ± 1 sample standard deviation across seeds (calculated using standard library functions with `ddof`=1). The lower panel shows the per-epoch paired difference $\Delta\mathrm{MAE} = \mathrm{MAE}_{\mathrm{DimeNet++}} - \mathrm{MAE}_{\mathcal{Q}\mathrm{DimeNet++}}$ with the corresponding $\pm 1\sigma$ band over the same paired seeds; positive values indicate lower MAE for $\mathcal{Q}$DimeNet++.

Our results show that $\mathcal{Q}$DimeNet++ achieves a mean test MAE of $67.92 \pm 1.98\,\mathrm{meV}$ against $72.42 \pm 1.52\,\mathrm{meV}$ for the classical baseline, a relative reduction of 6.21% achieved with a negligible parameter overhead of 0.27%.

The mean paired difference is $\Delta\mathrm{MAE} = 4.50 \pm 2.46\,\mathrm{meV}$. A paired t-test across the ten seeds yields $t_9 = 5.78$ ($p < 10^{-3}$), confirming that the observed reduction is statistically significant and is not derived from seed-level noise. This serves as a demonstration that the QFIM contributes genuine information rather than acting as a source of generic additional model capacity.

5. Conclusion

Our work introduced QUIVER, an architecture-agnostic paradigm in which the quantum Fisher information matrix of a variational quantum circuit, evaluated on classical inputs, furnishes a geometry-aware view that complements the standard classical representation of the same example. The construction is deliberately decoupled from any assumption that the underlying system is quantum: it treats the embedding solely as a mapping into a Hilbert space whose induced Fubini–Study metric (equivalent to the QFIM up to an overall factor of four) exposes higher-order correlations that are not naturally organized within standard kinematic or structural feature spaces. Fusing this quantum Fisher view with the classical view, via targeted architectural modifications, yields consistent improvements over two state-of-the-art baselines on tasks drawn from very different domains. On the JETCLASS top-quark tagging benchmark, the QUIVER-augmented Particle Transformer improves both the AUC and the QCD background rejection $1/\epsilon_B$ at $\epsilon_S = 0.5$ across all training sample sizes and feature sets considered, at a 7% parameter overhead. On the QM9 HOMO–LUMO gap regression task, $\mathcal{Q}$DimeNet++ reduces the test mean absolute error from $72.42 \pm 1.52\,\mathrm{meV}$ to $67.92 \pm 1.98\,\mathrm{meV}$, a 6.21% relative improvement obtained at a 0.27% parameter overhead, with a mean paired difference $\Delta\mathrm{MAE} = 4.50 \pm 2.46\,\mathrm{meV}$ that remains positive within $\pm 1\sigma$ across the ten paired seeds. The persistence of these gains under negligible parameter overhead, and across distinct architectures, input modalities and physical symmetries, supports the interpretation that the QFIM supplies genuine and complementary information rather than acting as a source of generic model capacity. Taken together, these results indicate that quantum-geometric features extracted from classically simulated variational quantum circuits can deliver measurable value to large classical models today, decoupling the practical utility of quantum-informed representations from progress toward fault-tolerant hardware.

6. Limitations

We provide a short discussion of the main limitations of this work, with a clear pathway for tackling these in future research.

First, the computational overhead of simulating extremely large quantum circuits constrains us to use at most 10 qubits for the VQC of the 1P1Q encoding, which contains the kinematic information of up to ten jet constituents, a limitation propagated therefore to the Particle Transformer benchmark. Even though most of the critical information required for jet flavor classification is often contained in these high-momentum constituents, this still results in a loss of performance, as compared to what would have been attained by using all 150 jet constituent particles available in the JETCLASS dataset.

This limitation also affects the task of molecular property prediction: the absolute performance of our benchmark DimeNet++, and QUIVER-augmented $\mathcal{Q}$DimeNet++ is marginally below that of the numbers reported in the original DimeNet++ paper (Gasteiger et al., 2020a), a necessary consequence of our setup operating on a restricted subset of up to 10 atoms (these being mapped to 10 qubits),

resulting in the information contained in the remaining (hydrogen) atoms being lost. The results are interpretable as a clear methodological gain under identical conditions for both large-parameter models. The goal therefore remains to scale up these systems to more qubits by the usage of, for example, **HPC resources with multi-GPU nodes** for large-qubit quantum system simulations.

Finally, a hybrid quantum-classical pipeline that simultaneously minimizes the parameters of both the precursor VQC and the subsequent large neural model could in principle, converge to a global minimum with performance better than what is observed in the current iteration of this work. This remains among our goals, with the main technical challenge lying in optimizing the quantum circuit based on a measurement of its QFIM rather than an observable.

References

Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., and Woerner, S. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021.

Bal, A., Klute, M., Maier, B., Oughton, M., Pezone, E., and Spannowsky, M. One particle - one qubit: Particle physics data encoding for quantum machine learning. *Phys. Rev. D*, 112:076004, Oct 2025. doi: 10.1103/l8y2-87vq. URL https://link.aps.org/doi/10.1103/l8y2-87vq.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., and Pascanu, R. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., Alam, M. S., Alonso-Linaje, G., Akash-Narayanan, B., Asadi, A., Arrazola, J. M., Azad, U., Banning, S., Blank, C., Bromley, T. R., Cordier, B. A., Ceroni, J., Delgado, A., Matteo, O. D., Dusko, A., Garg, T., Guala, D., Hayes, A., Hill, R., Ijaz, A., Isacsson, T., Ittah, D., Jahangiri, S., Jain, P., Jiang, E., Khandelwal, A., Kottmann, K., Lang, R. A., Lee, C., Loke, T., Lowe, A., McKiernan, K., Meyer, J. J., Montañez-Barrera, J. A., Moyard, R., Niu, Z., O'Riordan, L. J., Oud, S., Panigrahi, A., Park, C.-Y., Polatajko, D., Quesada, N., Roberts, C., Sá, N., Schoch, I., Shi, B., Shu, S., Sim, S., Singh, A., Strandberg, I., Soni, J., Száva, A., Thabet, S., Vargas-Hernández, R. A., Vincent, T., Vitucci, N., Weber, M., Wierichs, D., Wiersema, R., Willmann, M., Wong, V., Zhang, S., and Killoran, N. Pennylane: Automatic differentiation of hybrid quantum-classical computations, 2022. URL https://arxiv.org/abs/1811.04968.

Braunstein, S. L. and Caves, C. M. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439–3443, 1994.

Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S. C., Endo, S., Fujii, K., McClean, J. R., Mitarai, K., Yuan, X., Cincio, L., and Coles, P. J. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

Contributors, P. G. Dimenetplusplus. https://pytorch-geometric.readthedocs.io/en/latest/generated/torch_geometric.nn.models.DimeNetPlusPlus.html, 2024. PyTorch Geometric Documentation.

Fey, M. and Lenssen, J. E. Fast graph representation learning with pytorch geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. URL https://arxiv.org/abs/1903.02428.

Gasteiger, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *CoRR*, abs/2011.14115, 2020a. URL https://arxiv.org/abs/2011.14115.

Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020b. URL https://openreview.net/forum?id=B1eWbxStPH.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 1263–1272, 2017.

Huber, P. J. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. doi: 10.1214/aoms/1177703732.

Kasieczka, G. et al. The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics. *Rept. Prog. Phys.*, 84(12):124201, 2021. doi: 10.1088/1361-6633/ac36b9.

Komiske, P. T., Metodiev, E. M., and Thaler, J. Energy flow networks: deep sets for particle jets. *Journal of High Energy Physics*, 2019(1), January 2019. ISSN 1029-8479. doi: 10.1007/jhep01(2019)121. URL http://dx.doi.org/10.1007/JHEP01(2019)121.

Liu, J., Yuan, H., Lu, X.-M., and Wang, X. Quantum Fisher information matrix and multiparameter estimation.

Journal of Physics A: Mathematical and Theoretical, 53 (2):023001, 2020.

Meyer, J. J. Fisher Information in Noisy Intermediate-Scale Quantum Applications. *Quantum*, 5:539, 2021. doi: 10.22331/q-2021-09-09-539.

Nachman, B. and Shih, D. Anomaly detection with density estimation. *Physical Review D*, 101(7), April 2020. ISSN 2470-0029. doi: 10.1103/physrevd.101.075042. URL http://dx.doi.org/10.1103/PhysRevD.101.075042.

Provost, J. P. and Vallee, G. Riemannian structure on manifolds of quantum states. *Communications in Mathematical Physics*, 76(3):289–301, 1980.

PyTorch Contributors. HuberLoss – PyTorch Documentation. https://pytorch.org/docs/stable/generated/torch.nn.HuberLoss.html, 2024. Accessed: 2026.

Qu, H. and Gouskos, L. Jet tagging via particle clouds. *Physical Review D*, 101(5), March 2020. ISSN 2470-0029. doi: 10.1103/physrevd.101.056019. URL http://dx.doi.org/10.1103/PhysRevD.101.056019.

Qu, H. and Li, C. weaver-core: A streamlined deep-learning framework for high energy physics, 2022. URL https://github.com/hqucms/weaver-core. Accessed: May 2026.

Qu, H., Li, C., and Qian, S. Jetclass: A large-scale dataset for deep learning in jet physics, June 2022a. URL https://doi.org/10.5281/zenodo.6619768.

Qu, H., Li, C., and Qian, S. Particle transformer for jet tagging. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pp. 18281–18292, 2022b.

Ramakrishnan, R., Dral, P. O., Rupp, M., and von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, 2014.

Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958. doi: 10.1037/h0042519.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009. doi: 10.1109/TNN.2008.2005605.

Schütt, K. T., Kindermans, P.-J., Sauceda, H. E., Chmiela, S., Tkatchenko, A., and Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

A. Appendix: Jet Flavor Classification

A.1. Training setup and times

As mentioned in Table 2, we train the particle transformer per epoch on training sizes of $0.1\mathrm{M}$, $0.5\mathrm{M}$ and $5\mathrm{M}$ jets, equally divided between the two classes. We use a validation set of $1\mathrm{M}$ jets (equally balanced) and a held-out test set of $3.9\mathrm{M}$ jets. Training employs the Ranger optimizer (combining RAdam with LookAhead) with learning rate $\eta = 1 \times 10^{-3}$, batch size 512, and a flat+decay learning rate schedule that maintains the initial learning rate for the first 70% of training before exponentially decaying to 0.01η over the final 30%. The model is trained with cross-entropy loss and early stopping is disabled, allowing models to train for the full 60 epochs. For the largest training size of $5\mathrm{M}$ examples, the baseline Particle Transformer requires approximately $0.46\,\mathrm{h}$ per epoch, with its QUIVER-augmented variant requiring $1.26\,\mathrm{h}$ for the same.

B. Appendix: Molecular Property Regression

B.1. Training setup and times

The QM9 dataset is partitioned into training, validation, and test subsets containing $65{,}390$, $13{,}078$, and $52{,}313$ molecules respectively, summing to the full corpus of $130{,}781$ samples. The VQC is trained on a subset of $5{,}000$ molecules drawn from the training partition, with $1{,}000$ molecules taken from the validation partition for monitoring convergence. The classical DimeNet++ benchmark and the proposed $\mathcal{Q}\mathrm{DimeNet{+}{+}}$ architecture are both trained on the remaining $60{,}390$ molecules and validated on the remaining $12{,}078$ molecules, with final performance reported on the held-out test set of $52{,}313$ molecules. This protocol guarantees that the classical and quantum-enhanced models are evaluated on identical test data, while ensuring that no sample seen by the VQC during its pre-training stage is reused for validation or testing of the downstream graph network.

We train all DimeNet++ models with the ADAM optimizer using an initial learning rate of 10^{-3}, batch size 128, and zero weight decay. No learning-rate schedule or decay is used. Models are trained for at most 300 epochs with early stopping on validation MAE: training terminates when validation MAE fails to improve by at least $0.25\,\mathrm{meV}$ for 30 consecutive epochs. Since the HOMO–LUMO gap target is standardized during training, this corresponds to a normalized threshold of 1.95×10^{-4}.

The training objective is the L1 loss, equivalent to MAE, applied to the standardized HOMO–LUMO gap target. Validation and test MAEs are reported after converting back to physical units using the training-set target standard deviation.

Across the 10 seeds, DimeNet++ trained for 196.9 ± 30.9 epochs on average, while $\mathcal{Q}$DimeNet++ trained for 216.5 ± 54.6 epochs. Because runs were executed under parallel GPU scheduling, wall-clock timings should be interpreted as approximate runtime bounds rather than isolated architecture benchmarks. In this setup, epochs completed within approximately $80\,\mathrm{s}$ for DimeNet++ and $100\,\mathrm{s}$ for $\mathcal{Q}$DimeNet++.

For both cases, training was carried out on a single NVIDIA L40S GPU with $48\,\mathrm{GB}$ of VRAM, on a locally available university compute cluster with 128 CPU cores.

B.2. Technical Details

The edge-state rescaling in Equation 8, proposed under the QUIVER paradigm, operates on the 6×6 QFIM sub-matrix that encodes the pairwise interaction between qubits i and j. This sub-matrix is first processed by a two-dimensional convolutional layer with 16 output channels and a kernel of size 3, followed by a ReLU non-linearity. A global average pooling operation then collapses the spatial dimensions, and the resulting representation is flattened from 16 channels into an 8-dimensional vector, which is subsequently standardized by a LayerNorm operation to stabilize the learned embedding distribution. The normalized embedding is then passed through a scaling multilayer perceptron s_{ij} consisting of a linear projection from d_{QFIM} to $\max(4, d_{\mathrm{QFIM}})$ hidden units, a SiLU activation, a second linear projection to a single scalar, and a final $\tanh$ non-linearity that bounds the output to $(-1, 1)$. The resulting scalar acts as a learned, edge-specific multiplicative gate on the message exchanged between qubits i and j, allowing the model to attenuate or amplify each QFIM-derived interaction in a fully data-driven manner.