

# QuantumCanvas: A Multimodal Benchmark for Visual Learning of Atomic Interactions

Can Polat  
Texas A&M University  
can.polat@tamu.edu

Mustafa Kurban  
Ankara University  
Texas A&M University at Qatar  
kurbanm@ankara.edu.tr

Erchin Serpedin  
Texas A&M University  
eserpedin@tamu.edu

Hasan Kurban  
Hamad Bin Khalifa University  
hkurban@hbku.edu.qa

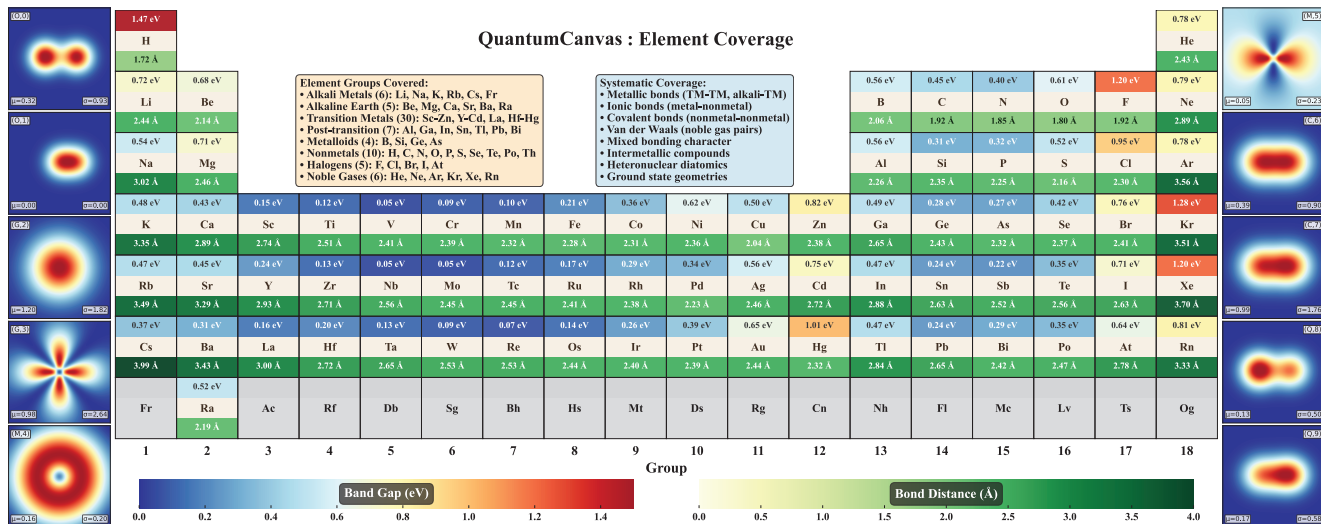


Figure 1. Periodic-table summary for the 75 elements included in 2,850 diatomic calculations. Each cell shows the element symbol, the mean Kohn–Sham energy gap, and the mean equilibrium bond distance. The energy-gap distribution is skewed toward metallic systems (median 0.00 eV, IQR 0.00–0.36 eV), while bond distances cluster near 2–3 Å (median 2.48 Å, IQR 2.16–2.91 Å). The left column displays channels 0–4 and the right column channels 5–9; badges report per-channel mean ( $\mu$ ) and standard deviation ( $\sigma$ ) over all 2,850 systems.

## Abstract

Despite rapid advances in molecular and materials machine learning, most models lack physical transferability: they fit correlations across whole molecules or crystals rather than learning the quantum interactions between atomic pairs. Yet bonding, charge redistribution, orbital hybridization, and electronic coupling all emerge from these two-body interactions that define local quantum fields in many-body systems. We introduce Quantum-

Canvas, a large-scale multimodal benchmark that treats two-body quantum systems as foundational units of matter. The dataset spans 2,850 element–element pairs, each annotated with 18 electronic, thermodynamic, and geometric properties and paired with ten-channel image representations derived from  $l$ - and  $m$ -resolved orbital densities, angular field transforms, co-occupancy maps, and charge-density projections. These physically grounded images encode spatial, angular, and electrostatic symmetries without explicit coordinates, providing an interpretable visual

modality for quantum learning. Benchmarking eight architectures across 18 targets, we report MAEs of **0.201 eV** on energy gap with GATv2, **0.265 eV** on HOMO, and **0.274 eV** on LUMO with EGNN; for energy-related quantities DimeNet attains **2.27 eV** total-energy MAE and **0.132 eV** repulsive-energy MAE, while a multimodal fusion model achieves a **2.15 eV** Mermin free-energy MAE. Pretraining on QuantumCanvas further improves convergence stability and generalization when fine-tuned on larger dataset such as QM9, MD17, and CrysMTM. By unifying orbital physics with vision-based representation learning, QuantumCanvas provides a principled and interpretable basis for learning transferable quantum interactions through coupled visual and numerical modalities. Dataset and model implementations can be found at <https://github.com/KurbanIntelligenceLab/QuantumCanvas>.

## 1. Introduction

Materials science underpins modern advances in semiconductors, energy conversion, catalysis, and quantum technologies [1]. A central challenge lies in understanding how atomic arrangements translate into measurable electronic, thermodynamic, and structural properties [2]. While data-driven models trained on large molecular and crystalline repositories have accelerated discovery, most remain statistical surrogates rather than physically grounded learners of interaction [3]. State-of-the-art architectures such as SchNet [4], DimeNet++ [5], PaiNN [6], GemNet [7], Equiformer [8], NequIP [9], Pure2DopeNet [10], Allegro [11], and GotenNet [12] achieve impressive accuracy through geometric message passing and equivariant tensor updates. However, they encode atomic interactions implicitly, capturing correlations across full structures rather than the fundamental quantum mechanisms through which atoms exchange charge, form bonds, and couple orbitals [13]. As a result, their generalization often deteriorates for unseen elements, charge states, or coordination motifs, revealing a limited understanding of transferable physical principles [14]. At the most fundamental level, bonding, charge redistribution, energy exchange, and orbital hybridization emerge from *two-body quantum interactions* [15]. These interactions define the local potential landscape on which all many-body effects are built. Explicitly learning them offers a direct route to physics-aware representations that are not tied to any specific geometry or composition [16]. Despite their foundational importance, large-scale datasets systematically targeting two-body quantum systems remain scarce, leaving this essential regime largely unexplored in machine learning.

To address this gap, we introduce *QuantumCanvas*, a large-scale multimodal benchmark that redefines two-body quantum systems as the elemental building blocks of matter.

The dataset enumerates all 2,850 element–element combinations across 75 elements and provides eighteen complementary descriptors covering energetic, electronic, thermodynamic, and geometric properties. Each pair is also represented by a ten-channel image tensor derived from *l*- and *m*-resolved orbital densities [17], angular field transforms such as GAF and MTF [18, 19], co-occupancy maps [20], and charge-density projections [21–23]. These physically grounded images encode spatial, angular, and electrostatic symmetries without explicit coordinates, creating a visually interpretable modality that complements graph-based representations. All diatomics are computed under a consistent Kohn–Sham framework [24, 25] at fixed electronic temperature using Mermin’s finite-temperature formalism [26], with identical grids, smearing, and dispersion settings, ensuring that energies, orbital levels, and charge distributions remain statistically comparable across pairs. The benchmark evaluates eight representative architectures across eighteen targets using mean absolute error under both random and composition-held-out splits, averaged over three seeds with uniform hyperparameter budgets. Pretraining on *QuantumCanvas* yields transferable embeddings that improve accuracy, convergence, and robustness when fine-tuned on established molecular and crystalline benchmarks such as QM9 [27], MD17 [28], and CrysMTM [29]. These results demonstrate that representations learned from two-body quantum interactions provide a physically meaningful foundation for generalizing across molecular, dynamical, and solid-state regimes.

The main contributions of this work are: (1) a comprehensive dataset of 2,850 diatomic systems covering 75 elements with 18 quantum-mechanical descriptors; (2) a ten-channel coordinate-free image representation derived from orbital populations and charge features that encodes symmetry for vision and multimodal learning; (3) a unified benchmark of eight architectures across eighteen targets under matched budgets, revealing modality-specific inductive biases and a well-defined difficulty gradient; and (4) experimental evidence that two-body pretraining improves accuracy, convergence, and generalization on QM9, MD17, and CrysMTM. Figure 1 summarizes the elemental coverage and per-channel statistics that motivate this multimodal design.

## 2. Related Work

**Modeling atomic interactions.** In quantum mechanics, all molecular and material phenomena originate from interactions between atoms [30]. The two-body problem forms the foundation of quantum chemistry, governing the potential energy and charge distribution between a pair of nuclei and their electrons [31]. Classical models such as the Lennard–Jones and Morse potentials approximate these interactions parametrically [32, 33], while first-principles approaches like density functional theory (DFT)

resolve them directly from the electronic structure [34]. In condensed-matter physics, pair potentials remain indispensable for modeling cohesive energies, lattice stability, and defect formation [35]. Despite their universality, explicit representations of two-body quantum behavior are largely absent from modern machine learning frameworks [36]. Most architectures encode atomic neighborhoods as high-dimensional embeddings with limited physical interpretability [37]. Capturing two-body interactions explicitly is critical for achieving transferability, as these interactions form a universal basis that generalizes across molecular, crystalline, and amorphous systems [38, 39].

**Existing datasets in molecular and materials machine learning.** The rapid progress of materials informatics has been enabled by large-scale quantum datasets. Benchmarks such as QM7-X [40] and MD22 [41] capture molecular geometries and DFT-level electronic properties for diverse organic systems, while the Materials Project, OQMD [42], NOMAD [43], and Matbench [44] provide millions of crystalline and surface calculations. These resources have catalyzed the development of graph-based and equivariant models that enable data-driven discovery of materials and molecules [45]. However, they primarily describe fully realized molecular or bulk systems, where pairwise interactions are entangled with many-body effects, polarization, and geometry-dependent coupling [46]. As a result, they do not isolate the primitive two-body quantum observables that underpin all higher-order material behavior [47].

**Two-body datasets and their limitations.** Only a few specialized efforts have targeted atomic pair interactions, often in the context of potential fitting and spectroscopy [48]. Frameworks such as ANI [49] and MACE [50] incorporate pairwise contributions through local energy decomposition but lack explicit two-body reference data. Experimental and *ab initio* diatomic datasets (e.g., NIST diatomic constants) provide limited elemental coverage and omit uniform descriptors beyond energy–distance curves [51]. Other benchmarks such as OC20 [52] and OC22 [53] emphasize many-body reaction energetics and surface adsorption, focusing on chemical processes rather than isolated atomic physics. Consequently, no existing benchmark systematically enumerates element–element combinations while providing both scalar quantum properties and spatially resolved orbital representations [54].

**Positioning of QuantumCanvas.** *QuantumCanvas* addresses this gap by modeling two-body quantum systems as the fundamental representational units of matter [55]. It unifies numerical descriptors with image-based orbital modalities that encode spatial, angular, and electrostatic symmetries [56]. This multimodal design provides an interpretable and physically consistent foundation for learning interatomic physics through both numerical and visual modalities, bridging quantum chemistry, computer vision,

and materials informatics [57]. By doing so, *QuantumCanvas* establishes the first large-scale benchmark that connects orbital-level quantum behavior to modern representation learning.

### 3. Dataset

#### 3.1. Two-body Quantum Simulations

Each sample in *QuantumCanvas* corresponds to an isolated diatomic (A–B) system computed through a consistent quantum-chemistry workflow based on an approximate finite-temperature Kohn–Sham density functional framework. Within the Mermin formalism, the electronic free energy is written as

$$\mathcal{F}[\rho] = E_{\text{KS}}[\rho] - T_e S[\{f_i\}], \quad (1)$$

where

$$E_{\text{KS}}[\rho] = T_s[\rho] + E_{\text{H}}[\rho] + E_{\text{xc}}[\rho] + \int \rho(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r}, \quad (2)$$

and  $S$  is the Fermi–Dirac entropy of the occupations  $\{f_i\}$ . Reported energy components include the band (one-electron) energy

$$E_{\text{band}} = \sum_i f_i \varepsilon_i, \quad (3)$$

the short-range repulsion  $E_{\text{rep}}$ , and the Mermin free energy evaluated at the self-consistent density  $\rho^*$  [58]. We record

$$E_{\text{tot}} = E_{\text{band}} + E_{\text{rep}} + \Delta E, \quad F = E_{\text{tot}} - T_e S, \quad (4)$$

where  $\Delta E$  collects exchange–correlation and double-counting corrections [59, 60]. Geometry optimization proceeds via SCF cycles to tolerance  $\varepsilon_{\text{SCF}}$  and line-search relaxation of the separation  $r = \|\mathbf{R}_B - \mathbf{R}_A\|$  until  $|\partial E_{\text{tot}}/\partial r| < \varepsilon_{\text{geom}}$  [61]. We also record the Fermi level  $E_F$ , the dipole moment

$$\boldsymbol{\mu} = \sum_{\alpha \in \{A,B\}} q_{\alpha} \mathbf{R}_{\alpha} - \int \mathbf{r} \rho(\mathbf{r}) d\mathbf{r}, \quad (5)$$

and Mulliken-like gross atomic charges  $q_{\alpha}$ , together with orbital populations resolved by angular momentum ( $\ell, m$ ) from projected density matrices.

All quantum-chemical calculations are carried out within the self-consistent charge density-functional tight-binding (SCC-DFTB) framework [62], which can be viewed as an approximate realization of finite-temperature Kohn–Sham density functional theory, as implemented in the DFTB+ code [60]. The Slater–Koster matrix elements were taken from the periodic table baseline parameters (PTBP) set, which provides robust density functional tight-binding parameters for solids across the periodic table and covers all

elements considered in this work [63]. This level of theory offers a good compromise between accuracy and computational cost, enabling the large-scale simulations.

### 3.2. Image Representations from Orbital Populations

Beyond scalars, each diatomic system is encoded as a fixed-size 10-channel image tensor  $\mathbf{I} \in \mathbb{R}^{10 \times 32 \times 32}$  derived from  $(\ell, m)$ -resolved orbital populations. Let  $n_{\ell m}^{(\alpha)}$  denote the population on atom  $\alpha \in \{A, B\}$  in the basis of real spherical harmonics  $Y_{\ell m}$ . We construct the following modalities.

**Orientation-aware orbital map (O-Map).** For each atom, we rasterize orbital populations over indices  $(\ell, m)$  up to  $\ell_{\max} = 2$  (covering  $s, p$ , and  $d$  shells):

$$\text{OMap}^{(\alpha)}[\ell, m] = n_{\ell m}^{(\alpha)}. \quad (6)$$

Each map is padded to a  $3 \times 5$  tile on a  $32 \times 32$  canvas; the two atomic maps form channels  $\mathbf{I}_{0:2}$ .

**Rotationally invariant power over shells (RIP) with GAF and MTF.** Shell populations  $P_{\ell}^{(\alpha)} = \sum_{m=-\ell}^{\ell} n_{\ell m}^{(\alpha)}$  are normalized to  $[0, 1]$  to obtain  $\tilde{P}_{\ell}^{(\alpha)}$ . The Gramian Angular Field (GAF) maps  $\tilde{\mathbf{P}}^{(\alpha)} = [\tilde{P}_0, \tilde{P}_1, \tilde{P}_2]$  to

$$\text{GAF}^{(\alpha)}[i, j] = \cos(\phi_i + \phi_j), \quad (7)$$

with angular encoding

$$\phi_i = \arccos \tilde{P}_i^{(\alpha)}. \quad (8)$$

The Markov Transition Field (MTF) quantizes  $\tilde{\mathbf{P}}^{(\alpha)}$  into  $Q = 3$  states by rank and encodes co-occurrence via

$$\text{MTF}^{(\alpha)}[u, v] = \mathbb{P}(s_{t+1} = v \mid s_t = u), \quad (9)$$

for  $u, v \in \{1, 2, 3\}$ . Both fields are upsampled to  $32 \times 32$  and assigned to channels  $\mathbf{I}_{2:4}$  (GAF) and  $\mathbf{I}_{4:6}$  (MTF).

**Co-occupancy map (COM).** Inter-atomic shell coupling is captured by the outer product of shell sums:

$$\text{COM}[\ell_1, \ell_2] = P_{\ell_1}^{(A)} P_{\ell_2}^{(B)}, \quad \ell_1, \ell_2 \in \{0, 1, 2\}. \quad (10)$$

This map is upsampled to  $32 \times 32$  and stored as channel  $\mathbf{I}_6$ . An optional second COM channel  $\mathbf{I}_7$  encodes normalized shell distributions or population-weighted  $|m|$  moments.

**Charge images (Q-Image).** Charge features are summarized by three  $2 \times 2$  matrices:  $\mathbf{Q}_{\text{diag}} = \begin{bmatrix} q_A & 0 \\ 0 & q_B \end{bmatrix}$ ,  $\mathbf{Q}_{|\Delta|} =$

$$\begin{bmatrix} 0 & |q_A - q_B| \\ |q_A - q_B| & 0 \end{bmatrix}, \mathbf{Q}_{\Pi} = \begin{bmatrix} 0 & q_A q_B \\ q_A q_B & 0 \end{bmatrix}, \text{ which}$$

are bilinearly upsampled to  $32 \times 32$  and mapped to channels  $\mathbf{I}_{8:10}$ . The final tensor  $\mathbf{I}$  integrates orientation-aware, rotationally invariant, and electrostatic information for multimodal learning.

### 3.3. Targets and Labels

Each pair is annotated with electronic, energetic, geometric, and charge-derived descriptors. Electronic observables include  $E_{\text{band}}$ ,  $E_{\text{rep}}$ ,  $E_{\text{tot}}$ , and the Mermin free energy  $F$ . Frontier levels  $\varepsilon_{\text{HOMO}}$  and  $\varepsilon_{\text{LUMO}}$  define the gap  $E_g = \varepsilon_{\text{LUMO}} - \varepsilon_{\text{HOMO}}$  and the Fermi level  $E_F$ . Thermodynamic quantities comprise the ionization potential  $I$  and electron affinity  $A$ , from which the following conceptual-DFT descriptors are derived:

$$\chi = \frac{I + A}{2}, \quad \eta = \frac{I - A}{2}, \quad S = \eta^{-1}, \quad (11)$$

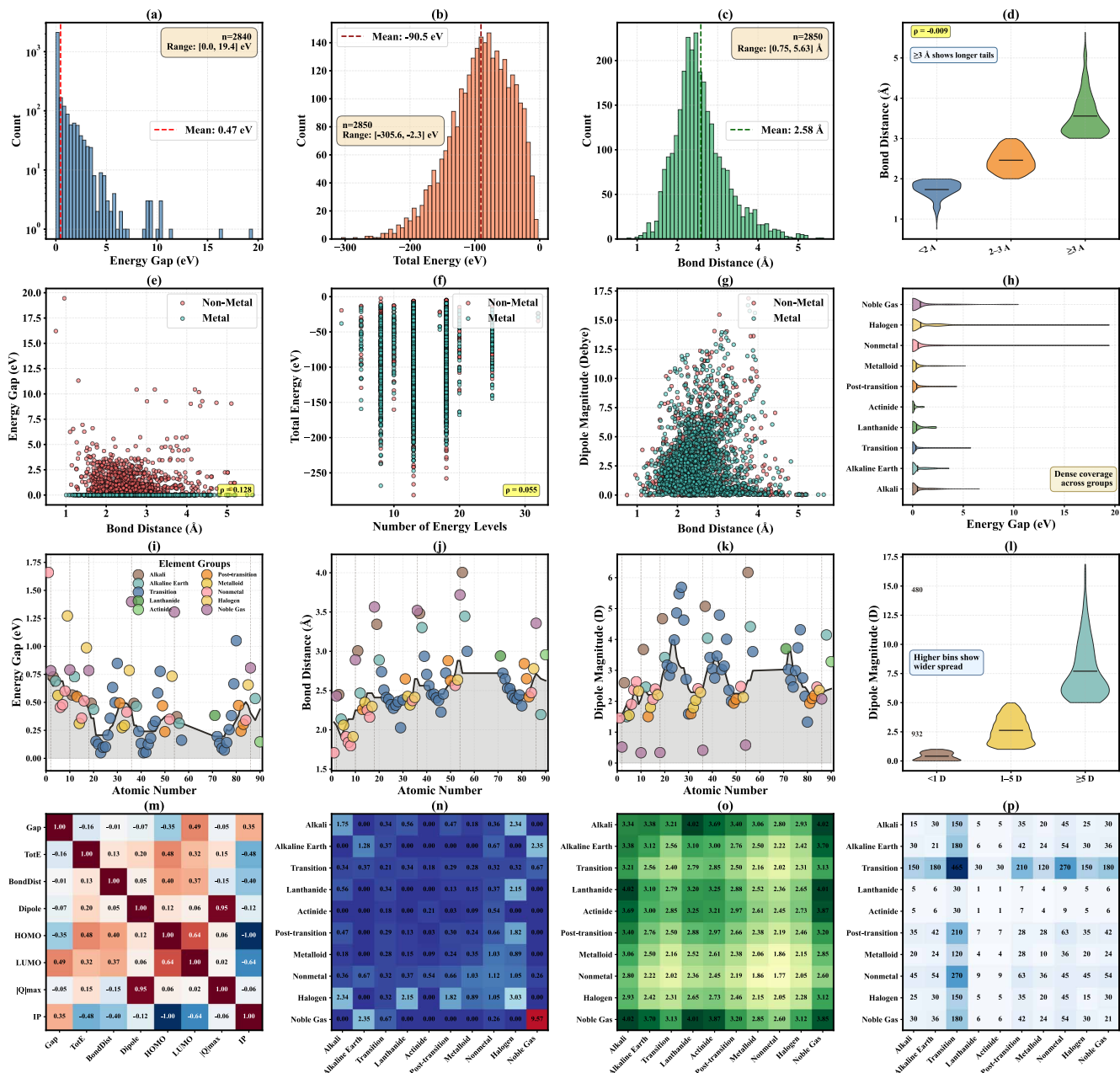
$$\mu_{\text{chem}} = -\chi, \quad \omega = \frac{\mu^2}{2\eta}.$$

where  $\chi$  denotes electronegativity,  $\eta$  hardness,  $S$  softness,  $\mu$  chemical potential, and  $\omega$  electrophilicity (all in eV) [64]. Dipole properties include the Cartesian components  $\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$  and the magnitude  $\|\boldsymbol{\mu}\|$ . The equilibrium bond length  $r$  (in Å) is extracted from optimized geometries. Charge statistics [65] summarize redistribution as follows:  $q_{\text{maxabs}} = \max(|q_A|, |q_B|)$ ,  $q_{\text{absmean}} = \frac{1}{2}(|q_A| + |q_B|)$ , and  $q_{\text{std}} = \sqrt{\frac{1}{2} \sum_{\alpha \in \{A, B\}} (q_{\alpha} - \bar{q})^2}$  with  $\bar{q} = \frac{1}{2}(q_A + q_B)$ .

**Label computation and units.** Energies and orbital levels are reported in eV after conversion from Hartree (1 Ha = 27.2114 eV). Dipoles are in Debye, charges in elementary charge  $e$ , and distances in Å. All values are obtained under a unified SCF and geometry protocol to ensure statistical consistency across pairs. Each record includes SCF diagnostics (iteration counts and residuals) to identify unconverged cases.

### 3.4. Analysis

**Periodic coverage and chemical diversity.** Figure 1 visualizes elemental and pairwise coverage for *QuantumCanvas*, summarizing 2,850 diatomics across 75 elements. For each element, the periodic table reports the mean Kohn–Sham band gap and equilibrium bond distance, yielding a compact map of physical diversity. Coverage includes nearly the entire s-, p-, and d-blocks: alkali (5/6), alkaline earth (6/6), transition (30/30), post-transition (7/7), metalloid (4/4), nonmetal (10/10), halogen (5/5), and noble gas (6/6). The overall band-gap distribution ( $\mu = 0.47$  eV,  $\sigma = 1.19$  eV) is sharply skewed toward metallic and semi-metallic behavior, while bond distances ( $\mu = 2.58$  Å,  $\sigma = 0.66$  Å) concentrate in the 2–3 Å range. *Implication: the dataset supports evaluation across chemically distinct*



regimes while maintaining realistic class balance for downstream generalization.

Channel-wise statistics in Figure 1 show complementary signal profiles. O-Maps (**ch0–ch1**) have low means/vari-

ances, reflecting sparsity in  $(\ell, m)$ -resolved populations and localized occupations. Gramian Angular Fields and Markov Transition Fields (**ch2–ch5**) exhibit the highest mean intensities and dispersions, capturing angular correlations and

transition structure among *s*-, *p*-, and *d*-shell populations. Co-occupancy maps (ch6–ch7) display moderate-to-high dispersion driven by cross-atomic shell coupling that varies with the chemical pair. Charge-derived channels (ch8–ch9) show low-to-moderate means with broad variance, encoding charge asymmetry and multiplicative charge interactions.

**Pairwise trends and structure–property patterns.** Pair counts range from 10 metalloid–metalloid examples to 465 transition–transition dimers, reflecting natural chemical prevalence and emphasis on metallic bonding. Average band gaps vary from 0.00 eV in alkali–alkaline-earth or noble-gas/post-transition pairs to  $\sim 9.6$  eV in noble-gas homodimers, spanning localized to delocalized regimes. Mean bond distances range from 1.9 Å in nonmetal–nonmetal pairs to 4.0 Å for weakly bound alkali–noble-gas pairs. *Implication: the dataset covers covalent, metallic, and van der Waals regimes, enabling stress-tests of extrapolation across coordination and bonding types.*

Figure 2 provides a comprehensive characterization of the *QuantumCanvas* diatomic corpus. Panels (a)–(c) summarize the global distributions of three core scalar properties. The energy-gap distribution in (a) is heavily right-skewed (mean 0.47 eV, median 0 eV, max 19.45 eV), reflecting the broad range of excitation behaviors across the periodic table. Total energies in (b) span over 300 eV, from weakly bound species to strongly attractive dimers, while bond distances in (c) cluster around 2.58 Å but still extend from extremely short (0.75 Å) to long-range (5.6 Å) separations. Panel (d) further partitions bond lengths into short ( $< 2$  Å, 16%), medium (2–3 Å, 62%), and long-range ( $> 3$  Å, 22%) regimes, showing that the dataset emphasizes chemically meaningful equilibrium-length dimers but preserves substantial coverage of compressed and stretched configurations.

Panels (e)–(g) explore structure–property relationships. The energy gap shows essentially no dependence on bond distance (e;  $\rho = -0.009$ ), indicating that orbital-level excitations are governed primarily by chemical identity rather than geometry alone. Total energy increases mildly with system size proxy (f;  $\rho = 0.128$ ), consistent with scaling of electron count, and dipole magnitude exhibits only a weak geometric trend (g;  $\rho = 0.055$ ), confirming that polarity is mostly driven by electronegativity contrast rather than internuclear separation.

Panel (h) organizes energy-gap distributions by periodic-table groups. Transition-metal dimers—the largest subset—show relatively small average gaps (0.32 eV), whereas halogen and noble-gas combinations yield broad, high-gap distributions (up to 19 eV), revealing strong periodic trends. Panels (i)–(k) provide element-level “constellation” sum-

maries of average gap, bond length, and dipole magnitude across 75 elements. Hydrogen forms the shortest and smallest-gap dimers, heavy alkali metals form the longest bonds, and dipole magnitudes peak for highly asymmetric pairs involving Cs, illustrating chemically interpretable patterns that a model must generalize across.

Panel (l) stratifies dipoles into physically meaningful categories: nearly one-third of dimers are weakly polar ( $< 1$  D), half lie in the moderate-polarity regime (1–5 D), and a significant tail (17%) exceeds 5 D. This variation is crucial for evaluating whether models capture subtle charge-transfer behavior across the periodic table.

The correlation matrix in panel (m) shows that most *QuantumCanvas* properties are only weakly correlated, confirming that the benchmark spans diverse and largely independent physical regimes. HOMO, LUMO, and gap form a coherent electronic cluster (e.g.,  $\rho_{\text{HOMO,LUMO}} = 0.642$ ), while strong dipole–charge coupling ( $\rho = 0.950$  between dipole and  $|Q|_{\text{max}}$ ) reflects consistent charge-asymmetry physics. Conversely, geometry (bond distance) correlates only moderately with frontier energies ( $\rho \sim 0.36$ – $0.40$ ) and very weakly with the gap, underscoring that the dataset does not collapse into a single geometric or energetic trend. *Implication: benchmarks should report per-target robustness and prioritize architectures that can disentangle electronic from geometric cues (e.g., equivariant graphs plus electronic-context heads).*

Finally, panels (n)–(p) present group-pair statistics. Average energy gaps vary dramatically across element-group combinations—from nearly metallic (alkali–alkaline-earth pairs, 0 eV) to highly insulating (noble-gas–noble-gas pairs, 9.57 eV). Group-level bond lengths in (o) range from compact (nonmetal–nonmetal, 1.77 Å) to extremely weakly bound (alkali–noble-gas, 4.02 Å). Pair-count distributions in (p) highlight chemical diversity: transition–transition pairs dominate with up to 465 samples, while rare combinations (e.g., lanthanide–lanthanide) remain minimally represented but non-zero.

Together, subfigures (a)–(p) reveal that *QuantumCanvas* is broad, chemically diverse, and structurally rich, with low redundancy across targets and strong periodic trends—providing an extensive and challenging benchmark for learning robust, transferable two-body quantum interactions.

## 4. Experiments

**Setup.** We evaluate learning of two-body quantum interactions on *QuantumCanvas*, which spans 18 scalar targets across electronic, energetic, thermodynamic, and geometric categories (Table 1). Each sample provides two modalities: a graph branch with atomic numbers and 3D coordinates, and a vision branch with ten-channel images derived from  $(\ell, m)$ -resolved orbital populations, angular transforms, and

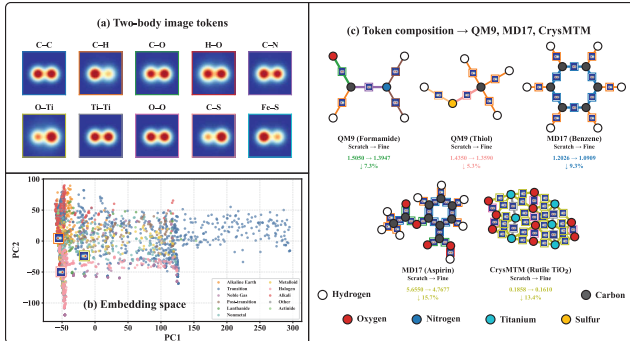


Figure 3. Panels (a–c) summarize the construction, organization, and downstream use of the two-body interaction tokens. (a) Representative two-body image tokens for selected element pairs; each inset is bordered by the color assigned to that pair type. (b) Two-dimensional PCA projection of all token embeddings, colored by the periodic-table group of one constituent element, showing that chemically related pairs cluster in the embedding space. (c) Assembly of two-body tokens into full molecular and crystalline structures for selected examples from QM9, MD17, and CrysMTM. For each bond, the corresponding token thumbnail is placed at the bond midpoint, illustrating how pairwise interactions compose into larger systems. Atoms are drawn with simplified CPK-like colors: carbon (dark gray), hydrogen (white), oxygen (red), nitrogen (blue), sulfur (yellow), and titanium (teal). Reported mean absolute errors show the improvement obtained when models are initialized from the two-body representation rather than trained from scratch.

charge maps. Models are trained with mean absolute error (MAE), early stopping on a held-out validation split, and three random seeds; we report mean  $\pm$  standard deviation. Splits are *element-pair disjoint* to prevent information leakage (i.e., a given A–B pair never appears across train/val/test), and preserve the class balance reported in Figure 2. Unless stated otherwise, methods receive identical inputs and differ only by encoder and fusion strategy. Implementation details are provided in the supplementary material.

**Baselines and modalities.** We compare geometric/equivariant graph networks (SchNet, DimeNet [66], EGNN [67], FAENet [68], GATv2 [69]) to vision encoders trained on images alone (ViT [70], QuantumShellNet [71]) and a late-fusion model that concatenates graph and image features before prediction. Hyperparameters (depth, width, cutoffs, radial/angle bases) are tuned uniformly per family to avoid target-specific bias.

**Difficulty landscape across targets.** Targets exhibit a clear complexity gradient. Charge statistics ( $Q_{\max}$ ,  $Q_{\text{mean}}$ ,  $Q_{\text{std}}$ ) and equilibrium bond distance achieve the lowest MAE, reflecting dependence on atomic identity and pair geometry. Band and total energies show larger

errors, consistent with many-electron and delocalized behavior that is only partially resolvable from two-body data. Frontier levels (HOMO, LUMO) and the gap are intermediate. Conceptual DFT quantities (electronegativity, chemical potential, hardness, softness, electrophilicity) are most sensitive to numerical stability due to non-linear transforms.

**Modality analysis: graph vs. image vs. fusion.** Graph encoders excel on geometry- and distance-driven properties (e.g.,  $r$ ,  $\|\mu\|$ ). Image encoders capture complementary electronic patterns from orbital symmetry, shell populations, and charge anisotropy. QuantumShellNet (image-only) approaches graph performance on several electronic targets, showing that orbital-field images contain rich quantum information even without explicit geometry. Late fusion improves selected properties such as HOMO, LUMO, and  $\|\mu\|$ , indicating synergistic rather than redundant cues. *Effect sizes:* for the energy gap  $E_g$ , GATv2 ( $0.201 \pm 0.020$ ) improves MAE by **11%** over EGNN ( $0.226 \pm 0.015$ ) and **19%** over DimeNet ( $0.248 \pm 0.020$ ); for  $\|\mu\|$ , EGNN ( $0.588 \pm 0.039$ ) reduces MAE by **7.8%** vs. DimeNet ( $0.638 \pm 0.035$ ).

**Architectural trends.** Among graph models, DimeNet achieves the lowest errors on geometry- and energy-sensitive tasks, benefiting from explicit radial/angle bases (e.g.,  $r$   $0.008 \pm 0.001$  Å). EGNN is competitive with lower capacity, highlighting the efficiency of equivariant design (best on  $\|\mu\|$   $0.588 \pm 0.039$  and tied-best/second on several electronic targets). FAENet favors directional observables (e.g.,  $\mu_z$   $0.888 \pm 0.091$ ). On the vision side, QuantumShellNet outperforms ViT on orbital/image-driven targets, supporting the value of physics-aware encodings. *Effect sizes:* HOMO MAE improves by **11.9%** from DimeNet ( $0.301 \pm 0.013$ ) to EGNN ( $0.265 \pm 0.014$ ); LUMO improves by **17.2%** from DimeNet ( $0.331 \pm 0.027$ ) to EGNN ( $0.274 \pm 0.032$ ). DimeNet reduces TOTAL ENERGY MAE by **28.8%** vs. EGNN (2.27 vs. 3.19 eV) and **11.7%** vs. FAENet (2.27 vs. 2.57 eV). For REPULSIVE ENERGY, DimeNet (0.132 eV) improves **41.6%** over the second-best GATv2 (0.226 eV).

**Variance, stability, and robustness across properties.** Different quantum targets exhibit markedly different levels of seed sensitivity. Among the conceptual DFT descriptors, *Softness* ( $S$ ) and *Electrophilicity* ( $\omega$ ) show the largest variance across runs, reflecting their nonlinear dependence on ionization potential and electron affinity; small perturbations in these upstream quantities propagate multiplicatively, amplifying fluctuations. Mermin free energy also displays elevated variability due to its explicit entropic term, which is harder to regress from static two-body fields.

In contrast, geometric observables such as the equilibrium bond length  $r$  and charge statistics ( $q_{\text{maxabs}}$ ,  $q_{\text{absmean}}$ ,  $q_{\text{std}}$ ) are substantially more stable, with standard deviations typically below  $10^{-2}$ .

Electronic-level predictions exhibit moderate but consistent variance patterns across GNNs and ViT-like models, while DimeNet and EGNN remain comparatively stable owing to their inductive biases. These metrics highlight which properties are robustly learnable from two-body interactions and which are intrinsically noisy due to higher-order coupling or thermodynamic contributions.

#### 4.1. Transfer Learning Across Molecular and Crystalline Domains

We pretrain on *QuantumCanvas* and fine-tune on QM9, MD17, and CrysMTM to test transfer across complementary regimes. QM9 probes molecular orbital energies, MD17 evaluates force-consistent potential energy surfaces along dynamics trajectories, and CrysMTM extends to periodic solids with composition and temperature effects. Pre-training on two-body interactions accelerates convergence and improves validation stability relative to training from scratch. Figure 3 visualizes the two-body interaction tokens, their embedding structure, and their assembly into downstream molecular and crystalline systems, highlighting how these representations bootstrap learning across datasets. On QM9, initialization reduces errors for LUMO and the gap; on MD17 it improves Aspirin, Benzene, and Ethanol targets; on CrysMTM it facilitates learning composition- and temperature-dependent band-gap variations with reduced overfitting. These results indicate that two-body pretraining internalizes low-order electronic principles—orbital coupling, charge asymmetry, short-range curvature—that transfer to many-body and periodic settings, supporting two-body pretraining as a scalable foundation for molecular and materials learning.

#### 5. Limitations

While *QuantumCanvas* establishes a rigorous foundation for modeling two-body quantum interactions, several limitations remain. First, the dataset is confined to diatomic systems and therefore cannot capture higher-order quantum effects such as collective electron correlation, polarization chains, or cooperative many-body phenomena that arise in extended materials. Second, although the ten-channel image representation encodes rich spatial and orbital structure, it remains static: time evolution, excited-state dynamics, and spin-resolved effects are not explicitly represented. Third, despite covering 75 elements and 2,850 unique dimers, the dataset does not include relativistic or high-pressure regimes where heavy-element and multi-reference effects become dominant. Finally, while pretraining on two-body interactions enhances transfer to

Table 1. **Comprehensive Benchmark Results on *QuantumCanvas***. MAE  $\pm$  standard deviation across 18 target properties and eight architectures. Units follow the dataset conventions: energies and orbital levels in eV, dipoles in Debye, charges in elementary charge  $e$ , and distances in Å. Green cells with bold values indicate the best performance, while blue underlined values denote the second-best. Graph-based models use atomic coordinates and chemical features, image-based models operate on 10-channel orbital-derived images, and the MultiModal model fuses both representations. Lower MAE indicates better predictive accuracy.

| Target                                             | DimeNet                           | EGNN                              | FAENet                           | GATv2                             | MultiModal                      | QShellNet         | SchNet            | ViT                             |
|----------------------------------------------------|-----------------------------------|-----------------------------------|----------------------------------|-----------------------------------|---------------------------------|-------------------|-------------------|---------------------------------|
| <i>Electronic Properties (eV)</i>                  |                                   |                                   |                                  |                                   |                                 |                   |                   |                                 |
| $E_g$                                              | 0.248 $\pm$ 0.020                 | 0.226 $\pm$ 0.015                 | 0.420 $\pm$ 0.070                | <b>0.201<math>\pm</math>0.020</b> | 0.639 $\pm$ 0.041               | 1.06 $\pm$ 0.30   | 0.662 $\pm$ 0.11  | 0.474 $\pm$ 0.05                |
| $E_{\text{HOMO}}$                                  | 0.301 $\pm$ 0.013                 | <b>0.265<math>\pm</math>0.014</b> | 0.363 $\pm$ 0.013                | 0.306 $\pm$ 0.026                 | 0.398 $\pm$ 0.019               | 0.883 $\pm$ 0.20  | 0.496 $\pm$ 0.02  | 0.868 $\pm$ 0.03                |
| $E_{\text{LUMO}}$                                  | 0.331 $\pm$ 0.027                 | <b>0.274<math>\pm</math>0.032</b> | 0.507 $\pm$ 0.028                | 0.322 $\pm$ 0.050                 | 0.551 $\pm$ 0.059               | 0.894 $\pm$ 0.06  | 0.666 $\pm$ 0.08  | 0.882 $\pm$ 0.02                |
| Energy gap                                         | <b>3.34<math>\pm</math>0.10</b>   | 3.74 $\pm$ 0.19                   | 3.83 $\pm$ 0.40                  | 8.56 $\pm$ 1.45                   | 3.34 $\pm$ 0.42                 | 19.4 $\pm$ 3.1    | 4.34 $\pm$ 0.46   | 17.5 $\pm$ 1.3                  |
| <i>Energy Properties (eV)</i>                      |                                   |                                   |                                  |                                   |                                 |                   |                   |                                 |
| $E_{\text{tot}}$                                   | <b>2.27<math>\pm</math>0.59</b>   | 3.19 $\pm$ 0.52                   | 2.57 $\pm$ 0.17                  | 9.23 $\pm$ 1.50                   | 3.00 $\pm$ 0.77                 | 19.5 $\pm$ 1.29   | 3.03 $\pm$ 0.47   | 16.5 $\pm$ 0.39                 |
| $E_{\text{rep}}$                                   | <b>0.132<math>\pm</math>0.020</b> | 0.244 $\pm$ 0.019                 | 0.245 $\pm$ 0.017                | 0.226 $\pm$ 0.001                 | 0.247 $\pm$ 0.081               | 0.769 $\pm$ 0.19  | 0.293 $\pm$ 0.02  | 0.397 $\pm$ 0.04                |
| Mermin $E$                                         | 3.22 $\pm$ 2.05                   | 2.68 $\pm$ 0.24                   | 2.73 $\pm$ 0.53                  | 8.17 $\pm$ 0.88                   | <b>2.15<math>\pm</math>0.38</b> | 18.9 $\pm$ 3.8    | 3.36 $\pm$ 0.17   | 16.7 $\pm$ 0.48                 |
| <i>Conceptual DFT Descriptors</i>                  |                                   |                                   |                                  |                                   |                                 |                   |                   |                                 |
| IP                                                 | <b>0.276<math>\pm</math>0.008</b> | 0.276 $\pm$ 0.007                 | 0.335 $\pm$ 0.024                | 0.323 $\pm$ 0.035                 | 0.395 $\pm$ 0.018               | 0.731 $\pm$ 0.04  | 0.534 $\pm$ 0.07  | 0.856 $\pm$ 0.02                |
| EA                                                 | 0.297 $\pm$ 0.017                 | <b>0.291<math>\pm</math>0.027</b> | 0.464 $\pm$ 0.014                | 0.339 $\pm$ 0.064                 | 0.513 $\pm$ 0.023               | 0.867 $\pm$ 0.07  | 0.715 $\pm$ 0.09  | 0.858 $\pm$ 0.06                |
| $\chi$                                             | 0.258 $\pm$ 0.027                 | <b>0.235<math>\pm</math>0.025</b> | 0.309 $\pm$ 0.005                | 0.256 $\pm$ 0.019                 | 0.323 $\pm$ 0.039               | 0.848 $\pm$ 0.15  | 0.405 $\pm$ 0.02  | 0.784 $\pm$ 0.02                |
| $\mu_{\text{chem}}$                                | 0.240 $\pm$ 0.011                 | <b>0.228<math>\pm</math>0.007</b> | 0.314 $\pm$ 0.017                | 0.234 $\pm$ 0.012                 | 0.271 $\pm$ 0.018               | 0.963 $\pm$ 0.14  | 0.419 $\pm$ 0.03  | 0.775 $\pm$ 0.03                |
| $\eta$                                             | 0.124 $\pm$ 0.010                 | 0.113 $\pm$ 0.008                 | 0.208 $\pm$ 0.038                | <b>0.101<math>\pm</math>0.010</b> | 0.299 $\pm$ 0.054               | 0.433 $\pm$ 0.04  | 0.331 $\pm$ 0.06  | 0.239 $\pm$ 0.03                |
| $S$                                                | 56.1 $\pm$ 38.9                   | 90.0 $\pm$ 44.5                   | 122.4 $\pm$ 41.5                 | 53.3 $\pm$ 40.0                   | 86.7 $\pm$ 23.4                 | 438.9 $\pm$ 29.8  | 122.5 $\pm$ 27.8  | <b>51.7<math>\pm</math>42.2</b> |
| $\omega$                                           | 844 $\pm$ 611                     | 2173 $\pm$ 1503                   | 1935 $\pm$ 846                   | 784.9 $\pm$ 696                   | 1888 $\pm$ 720                  | 8997 $\pm$ 1994   | 2264 $\pm$ 747    | <b>757.5<math>\pm</math>739</b> |
| <i>Molecular, Geometric, and Charge Properties</i> |                                   |                                   |                                  |                                   |                                 |                   |                   |                                 |
| $\ \mu\ $                                          | 0.638 $\pm$ 0.035                 | <b>0.588<math>\pm</math>0.039</b> | 0.737 $\pm$ 0.043                | 0.618 $\pm$ 0.035                 | 0.836 $\pm$ 0.13                | 0.857 $\pm$ 0.13  | 0.878 $\pm$ 0.03  | 1.66 $\pm$ 0.13                 |
| $\mu_z$                                            | 1.43 $\pm$ 0.03                   | 1.22 $\pm$ 0.19                   | <b>0.888<math>\pm</math>0.09</b> | 1.17 $\pm$ 0.05                   | 1.32 $\pm$ 0.10                 | 1.54 $\pm$ 0.14   | 2.21 $\pm$ 0.12   | 2.14 $\pm$ 0.08                 |
| $r$                                                | <b>0.008<math>\pm</math>0.001</b> | 0.105 $\pm$ 0.016                 | 0.070 $\pm$ 0.012                | 0.141 $\pm$ 0.023                 | 0.031 $\pm$ 0.006               | 0.292 $\pm$ 0.06  | 0.075 $\pm$ 0.01  | 0.425 $\pm$ 0.03                |
| $q_{\text{maxabs}}$                                | 0.055 $\pm$ 0.005                 | <b>0.052<math>\pm</math>0.006</b> | 0.059 $\pm$ 0.003                | 0.054 $\pm$ 0.003                 | 0.071 $\pm$ 0.010               | 0.067 $\pm$ 0.022 | 0.072 $\pm$ 0.005 | 0.128 $\pm$ 0.003               |
| $q_{\text{absmean}}$                               | 0.055 $\pm$ 0.005                 | <b>0.052<math>\pm</math>0.006</b> | 0.059 $\pm$ 0.001                | 0.054 $\pm$ 0.003                 | 0.078 $\pm$ 0.008               | 0.052 $\pm$ 0.007 | 0.072 $\pm$ 0.005 | 0.129 $\pm$ 0.002               |
| $q_{\text{std}}$                                   | 0.055 $\pm$ 0.005                 | <b>0.052<math>\pm</math>0.006</b> | 0.059 $\pm$ 0.002                | 0.054 $\pm$ 0.003                 | 0.070 $\pm$ 0.004               | 0.083 $\pm$ 0.013 | 0.072 $\pm$ 0.005 | 0.127 $\pm$ 0.007               |

larger molecules and crystals, the scalability of this approach to complex condensed phases, for example liquids, alloys, or solid-to-solid transitions, remains an open question. We view *QuantumCanvas* as a first step toward a hierarchical and multimodal language of quantum interactions, motivating future extensions to triatomic and many-body systems that incorporate time-dependent, spin-orbit, and external-field effects into unified learning frameworks.

#### 6. Conclusion

We introduced *QuantumCanvas*, a large-scale multimodal benchmark that reimagines two-body quantum systems as the fundamental building blocks of matter. Across 75 elements and 2,850 unique dimers, the dataset provides 18 scalar quantum descriptors and ten physically grounded image channels derived from orbital populations and charge distributions. These representations encode spatial, angular, and electrostatic symmetries that can be directly exploited by modern geometric and vision architectures, bridging data-driven learning and first-principles physics. Empirical studies demonstrate that models pretrained on *QuantumCanvas* achieve faster convergence and lower error across larger datasets, with gains up to 25% prediction accuracy. Beyond benchmarking, *QuantumCanvas* establishes a new paradigm for quantum representation learning by representing interatomic physics through interpretable visual and numerical modalities that generalize across molecular, crystalline, and condensed-matter domains. As a natural next step, we aim to extend *QuantumCanvas* to tri-

atomic and time-dependent datasets using spin-resolved and relativistic channels, advancing toward universal, physically interpretable foundation models for quantum materials.

## References

- [1] William D Callister Jr and David G Rethwisch. *Materials science and engineering: an introduction*. John Wiley & sons, 2020. 2
- [2] Feliciano Giustino. *Materials modelling using density functional theory: properties and predictions*. Oxford University Press, 2014. 2
- [3] Keith T Butler, Felipe Oviedo, and Pieremanuele Canepa. *Machine learning in materials science*, volume 29. American Chemical Society, 2022. 2
- [4] Kristof T Schütt, Huziel E Saucedo, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet—a deep learning architecture for molecules and materials. *The Journal of chemical physics*, 148(24), 2018. 2
- [5] Johannes Gasteiger, Shankari Giri, Johannes T Margraf, and Stephan Günnemann. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020. 2
- [6] Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International conference on machine learning*, pages 9377–9388. PMLR, 2021. 2
- [7] Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34:6790–6802, 2021. 2
- [8] Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022. 2
- [9] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022. 2
- [10] Can Polat, Mustafa Kurban, and Hasan Kurban. Multimodal neural network-based predictive modeling of nanoparticle properties from pure compounds. *Machine Learning: Science and Technology*, 5(4):045062, 2024. 2
- [11] Ken-ichi Nomura, Shinnosuke Hattori, Satoshi Ohmura, Ikumi Kanemasu, Kohei Shimamura, Nabankur Dasgupta, Aiichiro Nakano, Rajiv K Kalia, and Priya Vashishta. Allegro-fm: Toward an equivariant foundation model for exascale molecular dynamics simulations. *The Journal of Physical Chemistry Letters*, 16:6637–6644, 2025. 2
- [12] Sarp Aykent and Tian Xia. Gotennet: Rethinking efficient 3d equivariant graph neural networks. In *The Thirteenth International Conference on Learning Representations*, 2025. 2
- [13] Oliver T Unke, Stefan Chmiela, Huziel E Saucedo, Michael Gastegger, Igor Poltavsky, Kristof T Schütt, Alexandre Tkatchenko, and Klaus-Robert Müller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021. 2
- [14] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters*, 108(5):058301, 2012. 2
- [15] Robert G Parr. Density functional theory of atoms and molecules. In *Horizons of Quantum Chemistry: Proceedings of the Third International Congress of Quantum Chemistry Held at Kyoto, Japan, October 29-November 3, 1979*, pages 5–15. Springer, 1989. 2
- [16] Jörg Behler. Perspective: Machine learning potentials for atomistic simulations. *The Journal of chemical physics*, 145(17), 2016. 2
- [17] Dmitrii Aleksandrovich Varshalovich, Anatoli Nikolaevitch Moskalev, and Valerij Kel'manovich Khersonskii. *Quantum theory of angular momentum*. World Scientific, 1988. 2
- [18] Zhifeng Wu, Yaobin Zhu, Yanbing Wang, and Feng Wang. Extracting static and dynamic features in joint gaf-mtf image for space target recognition. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 9840–9843. IEEE, 2024. 2
- [19] Mengjiao Wang, Wenjie Wang, Xinan Zhang, and Herbert Ho-Ching Iu. A new fault diagnosis of rolling bearing based on markov transition field and cnn. *Entropy*, 24(6):751, 2022. 2
- [20] Knud Taulbjerg and JS Briggs. Inner-shell excitation in heteronuclear collisions. i. orbitals and couplings in a molecular basis. *Journal of Physics B: Atomic and Molecular Physics*, 8(11):1895, 1975. 2
- [21] Robert S Mulliken. Electronic structures of polyatomic molecules and valence. iv. electronic states, quantum theory of the double bond. *Physical Review*, 43(4):279, 1933. 2
- [22] Richard FW Bader and TT Nguyen-Dang. Quantum theory of atoms in molecules—dalton revisited. In *Advances in quantum chemistry*, volume 14, pages 63–124. Elsevier, 1981.
- [23] Nicholas J Mayhall, Krishnan Raghavachari, and Hrant P Hratchian. Oniom-based qm: Qm electronic embedding method using löwdin atomic charges: Energies and analytic gradients. *The Journal of chemical physics*, 132(11), 2010. 2
- [24] F Matthias Bickelhaupt and Evert Jan Baerends. Kohnsham density functional theory: predicting and understanding chemistry. *Reviews in computational chemistry*, pages 1–86, 2000. 2
- [25] Reiner M Dreizler and Eberhard KU Gross. *Density functional theory: an approach to the quantum many-body problem*. Springer Science & Business Media, 2012. 2
- [26] Mario V Stoitsov and Ivan Zh Petkov. Density functional theory at finite temperatures. *Annals of Physics*, 184(1):121–147, 1988. 2
- [27] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014. 2
- [28] Stefan Chmiela, Alexandre Tkatchenko, Huziel E Saucedo, Igor Poltavsky, Kristof T Schütt, and Klaus-Robert Müller.

- Machine learning of accurate energy-conserving molecular force fields. *Science advances*, 3(5):e1603015, 2017. 2
- [29] Can Polat, Erchin Serpedin, Mustafa Kurban, and Hasan Kurban. Crysmtn: a multiphase, temperature-resolved, multimodal dataset for crystalline materials. *Machine Learning: Science and Technology*, 6(3):030603, 2025. 2
- [30] Ira N Levine, Daryle H Busch, and Harrison Shull. *Quantum chemistry*, volume 6. Pearson Prentice Hall Upper Saddle River, NJ, 2009. 2
- [31] Peter W Atkins and Ronald S Friedman. *Molecular quantum mechanics*. Oxford university press, 2011. 2
- [32] Jacob N Israelachvili. *Intermolecular and surface forces*. Academic press, 2011. 2
- [33] Yatendra Pal Varshni. Comparative study of potential energy functions for diatomic molecules. *Reviews of Modern Physics*, 29(4):664, 1957. 2
- [34] Richard M Martin. *Electronic structure: basic theory and practical methods*. Cambridge university press, 2020. 3
- [35] Gerald D Mahan. *Many-particle physics*. Springer Science & Business Media, 2013. 3
- [36] Raghunathan Ramakrishnan and O Anatole von Lilienfeld. Machine learning, quantum chemistry, and chemical space. *Reviews in computational chemistry*, 30:225–256, 2017. 3
- [37] Thea Denell, Lauri Himanen, Markus Scheidgen, and Claudia Draxl. Automated identification of bulk structures, two-dimensional materials, and interfaces using symmetry-based clustering. *npj Computational Materials*, 11(1):25, 2025. 3
- [38] Justin Airas, Xinqiang Ding, and Bin Zhang. Transferable implicit solvation via contrastive learning of graph neural networks. *ACS Central Science*, 9(12):2286–2297, 2023. 3
- [39] Hironao Yamada, Chang Liu, Stephen Wu, Yukinori Koyama, Shenghong Ju, Junichiro Shiomi, Junko Morikawa, and Ryo Yoshida. Predicting materials properties with little data using shotgun transfer learning. *ACS central science*, 5(10):1717–1730, 2019. 3
- [40] Johannes Hoja, Leonardo Medrano Sandonas, Brian G Ernst, Alvaro Vazquez-Mayagoitia, Robert A DiStasio Jr, and Alexandre Tkatchenko. Qm7-x, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules. *Scientific data*, 8(1):43, 2021. 3
- [41] Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T Unke, Adil Kabylda, Huziel E Saucedo, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances*, 9(2):eadf0873, 2023. 3
- [42] James E Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and Christopher Wolverton. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd). *Jom*, 65(11):1501–1509, 2013. 3
- [43] Claudia Draxl and Matthias Scheffler. The nomad laboratory: from data sharing to artificial intelligence. *Journal of Physics: Materials*, 2(3):036001, 2019. 3
- [44] Alexander Dunn, Qi Wang, Alex Ganose, Daniel Dopp, and Anubhav Jain. Benchmarking materials property prediction methods: the matbench test set and automatminer reference algorithm. *npj Computational Materials*, 6(1):138, 2020. 3
- [45] Maarten De Jong, Wei Chen, Henry Geerlings, Mark Asta, and Kristin Aslaug Persson. A database to enable discovery and design of piezoelectric materials. *Scientific data*, 2(1):1–13, 2015. 3
- [46] Rina Ibragimova, Mikhail S Kuklin, Tigany Zarrouk, and Miguel A Caro. Unifying the description of hydrocarbons and hydrogenated carbon materials with a chemically reactive machine learning interatomic potential. *Chemistry of Materials*, 37(3):1094–1110, 2025. 3
- [47] Luigi Amico, Rosario Fazio, Andreas Osterloh, and Vlatko Vedral. Entanglement in many-body systems. *Reviews of modern physics*, 80(2):517–576, 2008. 3
- [48] Konrad Patkowski. Recent developments in symmetry-adapted perturbation theory. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 10(3):e1452, 2020. 3
- [49] Justin S Smith, Olexandr Isayev, and Adrian E Roitberg. Ani-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific data*, 4(1):1–8, 2017. 3
- [50] Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gábor Csányi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in neural information processing systems*, 35:11423–11436, 2022. 3
- [51] Hans-Joachim Werner, Peter J Knowles, Frederick R Manby, Joshua A Black, Klaus Doll, Andreas Heßelmann, Daniel Kats, Andreas Köhn, Tatiana Korona, David A Kreplin, et al. The molpro quantum chemistry package. *The Journal of chemical physics*, 152(14), 2020. 3
- [52] Lowik Chanasot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021. 3
- [53] Richard Tran, Janice Lan, Muhammed Shuaibi, Brandon M Wood, Siddharth Goyal, Abhishek Das, Javier Heras-Domingo, Adeesh Kolluru, Ammar Rizvi, Nima Shoghi, et al. The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *ACS Catalysis*, 13(5):3066–3084, 2023. 3
- [54] Anton Charkin-Gorbunin, Artem Kokorin, Huziel E. Saucedo, Stefan Chmiela, Claudio Quarti, David Beljonne, Alexandre Tkatchenko, and Igor Poltavsky. Atomic orbits in molecules and materials for improving machine learning force fields. *Machine Learning: Science and Technology*, 2025. 3
- [55] Bailey Christensen and Rusi Taleyharhan. Machine-learning-based real-time, high-fidelity neutron source identification via cavitation acoustic shock spectra in centrifugally tensioned metastable fluid detectors. *IEEE Access*, 2025. 3
- [56] Yaolong Zhang and Bin Jiang. Universal machine learning for the response of atomistic systems to external fields. *Nature Communications*, 14(1):6424, 2023. 3
- [57] Can Polat, Mehmet Tuncel, Mustafa Kurban, Erchin Serpedin, and Hasan Kurban. xchemagents: Agentic ai for explainable quantum chemistry. *arXiv preprint arXiv:2505.20574*, 2025. 3

- [58] John M Ziman. *Principles of the Theory of Solids*. Cambridge university press, 1979. [3](#)
- [59] G. Kresse and J. Furthmüller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169–11186, 1996. [3](#)
- [60] Balint Aradi, Ben Hourahine, and Th Frauenheim. Dftb+, a sparse matrix-based implementation of the dftb method. *The Journal of Physical Chemistry A*, 111(26):5678–5684, 2007. [3](#), [1](#)
- [61] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006. [3](#)
- [62] Michael Gaus, Qiang Cui, and Marcus Elstner. Dftb3: Extension of the self-consistent-charge density-functional tight-binding method (scc-dftb). *Journal of chemical theory and computation*, 7(4):931–948, 2011. [3](#), [1](#)
- [63] Mengnan Cui, Karsten Reuter, and Johannes T. Margraf. Obtaining robust density functional tight-binding parameters for solids across the periodic table. *Journal of Chemical Theory and Computation*, 20(12):5276–5290, 2024. [4](#), [1](#)
- [64] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012. [4](#)
- [65] F Richard and R Bader. Atoms in molecules: a quantum theory, 1990. [4](#)
- [66] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020. [7](#)
- [67] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021. [7](#)
- [68] Alexandre Agm Duval, Victor Schmidt, Alex Hernández-García, Santiago Miret, Fragkiskos D Malliaros, Yoshua Bengio, and David Rolnick. Faenet: Frame averaging equivariant gnn for materials modeling. In *International Conference on Machine Learning*, pages 9013–9033. PMLR, 2023. [7](#)
- [69] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021. [7](#)
- [70] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. [7](#)
- [71] Can Polat, Hasan Kurban, and Mustafa Kurban. Quantumshellnet: ground-state eigenvalue prediction of materials using electronic shell structures and fermionic properties via convolutions. *Computational Materials Science*, 246:113366, 2025. [7](#)
- [72] Sagar Imambi, Kolla Bhanu Prakash, and GR Kanagachidambaresan. Pytorch. In *Programming with TensorFlow: solution for edge computing applications*, pages 87–104. Springer, 2021. [2](#)
- [73] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019. [2](#)

# QuantumCanvas: A Multimodal Benchmark for Visual Learning of Atomic Interactions

## Supplementary Material

### 7. Detailed Materials Simulations

Following the finite-temperature Kohn–Sham formulation used in the main paper, each dimer is treated as an isolated two-body system within the self-consistent-charge density-functional tight-binding (SCC–DFTB) approximation at electronic temperature  $T_e$  [60, 62, 63]. For completeness, we summarize the finite-temperature occupations, entropy, and the second-order DFTB energy functional that underlies all reported scalar targets.

The single-particle orbitals  $\{\psi_i\}$  obey the generalized eigenvalue problem

$$\sum_{\nu} (H_{\mu\nu} - \varepsilon_i S_{\mu\nu}) c_{\nu i} = 0, \quad (12)$$

where  $H_{\mu\nu}$  and  $S_{\mu\nu}$  are the Hamiltonian and overlap matrix elements in a localized atomic basis and  $c_{\nu i}$  are molecular-orbital coefficients. At finite electronic temperature  $T_e$  the orbital occupations are given by Fermi–Dirac statistics

$$f_i = \frac{1}{1 + \exp\left(\frac{\varepsilon_i - \mu_F}{T_e}\right)}, \quad \sum_i f_i = N_e, \quad (13)$$

with Fermi level  $\mu_F$  chosen to conserve the total number of electrons  $N_e$ .

The corresponding electronic entropy is

$$S[\{f_i\}] = - \sum_i \left[ f_i \ln f_i + (1 - f_i) \ln(1 - f_i) \right], \quad (14)$$

so that the Mermin free energy at fixed nuclear geometry  $\mathbf{R} = (\mathbf{R}_A, \mathbf{R}_B)$  reads

$$F[\rho] = E_{\text{KS}}[\rho] - T_e S[\{f_i\}], \quad (15)$$

consistent with the expression used in Sec. 3.1 of the main text.

Within SCC–DFTB, the Kohn–Sham energy is approximated by a second-order expansion around a reference density built from neutral atoms. Introducing charge fluctuations  $\Delta q_A = q_A - q_A^{(0)}$  relative to the neutral reference, the total energy can be written as

$$E_{\text{DFTB}}(\mathbf{R}, \{\Delta q_A\}) = E_{\text{band}} + \Delta E_{\text{Coul}}(\{\Delta q_A\}) + E_{\text{rep}}(\mathbf{R}), \quad (16)$$

where the band energy is

$$E_{\text{band}} = \sum_i f_i \varepsilon_i, \quad (17)$$

the short-range repulsive contribution  $E_{\text{rep}}$  is given by a sum of tabulated pair potentials, and the charge-fluctuation term is approximated as

$$\Delta E_{\text{Coul}}(\{\Delta q_A\}) \simeq \frac{1}{2} \sum_{A,B} \gamma_{AB} \Delta q_A \Delta q_B. \quad (18)$$

Here  $\gamma_{AB}$  is the chemical hardness matrix derived from the underlying DFT reference and tabulated in the PTBP parameter set.

Mulliken-like gross atomic charges are computed from the density matrix

$$P_{\mu\nu} = \sum_i f_i c_{\mu i} c_{\nu i} \quad (19)$$

and the overlap matrix according to

$$q_A = \sum_{\mu \in A} \sum_{\nu} P_{\mu\nu} S_{\mu\nu}, \quad (20)$$

where the sum over  $\mu$  runs over basis functions centered on atom  $A$ . The SCC correction enters the Hamiltonian matrix elements as

$$H_{\mu\nu} = H_{\mu\nu}^{(0)} + \frac{1}{2} S_{\mu\nu} \sum_C (\gamma_{AC} + \gamma_{BC}) \Delta q_C, \quad \mu \in A, \nu \in B, \quad (21)$$

so that  $H_{\mu\nu}$ , the eigenvalues  $\varepsilon_i$ , the occupations  $\{f_i\}$ , and the charges  $\{q_A\}$  have to be determined self-consistently.

In practice, each diatomic calculation proceeds by iterating the coupled SCC and SCF equations until both the charge fluctuations and the total-energy change fall below prescribed tolerances,

$$\max_A |\Delta q_A^{(n)} - \Delta q_A^{(n-1)}| < \varepsilon_{\text{SCC}}, \quad |E_{\text{tot}}^{(n)} - E_{\text{tot}}^{(n-1)}| < \varepsilon_{\text{SCF}}, \quad (22)$$

after which a one-dimensional geometry relaxation along the internuclear separation  $r = \|\mathbf{R}_B - \mathbf{R}_A\|$  is performed until

$$\left| \frac{\partial E_{\text{tot}}}{\partial r} \right| < \varepsilon_{\text{geom}}. \quad (23)$$

All diatomic energies, orbital levels, charges, and dipole moments included in the QuantumCanvas benchmark are evaluated at these self-consistent, geometry-relaxed conditions. The Slater–Koster integrals  $H_{\mu\nu}^{(0)}(R_{AB})$ ,  $S_{\mu\nu}(R_{AB})$ , the hardness matrix  $\gamma_{AB}$ , and the repulsive potentials  $E_{\text{rep}}$  are all taken from the periodic-table baseline parameters (PTBP) and evaluated using the DFTB+ code [60, 62, 63].

## 8. Experiment Details

### 8.1. Training on QuantumCanvas

For the two-body experiments on *QuantumCanvas*, we implement a family of graph-based, vision-based, and multi-modal regressors that all operate on the same diatomic input representation. Each data point is a two-atom system  $A-B$  with equilibrium positions  $\mathbf{R} \in \mathbb{R}^{2 \times 3}$ , atomic numbers  $Z \in \mathbb{N}^2$ , a 10-channel  $32 \times 32$  orbital image tensor, and a scalar target  $y$  (one of the 18 quantum properties). All models are implemented in PyTorch [72] and PyTorch Geometric [73] and are trained as scalar regressors.

**Graph-based encoders.** **SchNet** follows the reference PyTorch Geometric implementation, with a lightweight configuration tailored to the two-body setting. We use hidden channels and filter size of 16, two interaction blocks, eight Gaussian radial basis functions, and a cutoff radius of 5.0 Å. The readout is a simple atom-wise sum, yielding a single scalar prediction per dimer.

**DimeNet++** is instantiated with hidden dimension 128, four interaction blocks, intermediate embedding sizes of 64 (interaction) and 8 (basis), out-embedding dimension 256, six radial basis functions, and seven spherical harmonics. We use a cutoff of 5.0 Å with an envelope exponent of 5, one pre-skip and two post-skip layers in each block, and three output MLP layers, closely following standard DimeNet++ hyperparameters but in a moderately sized regime suitable for a small-atom system.

**FAENet** is used in its message-passing configuration with a cutoff of 5.0 Å, eight Gaussian basis functions, 32 hidden channels and filters, and two interaction layers. Physics-aware embeddings are disabled (`phys_embeds=False`) so that the model focuses on learned representations rather than engineered descriptors. To accommodate heavy elements (up to Th,  $Z=90$ ), we extend FAENet’s internal embedding tables (main atomic embedding and optional period/group embeddings) from their default maximum atomic number to support up to  $Z=119$  by copying existing weights and randomly initializing the additional entries with small variance.

**EGNN-style GCN** is instantiated as a simple isotropic message-passing baseline. We embed atomic numbers with a 100-way embedding table to a feature dimension of 128. A  $k$ -nearest-neighbor graph with  $k=5$  is constructed from the 3D coordinates, and three stacked GCNConv layers (all 128 channels) are applied with ReLU activations. Global graph features are obtained via mean pooling over nodes, followed by a three-layer MLP (two hidden layers with ReLU and dropout, and a final linear layer) that maps from 128 to a scalar.

**GATv2** uses the same  $k$ -NN graph construction ( $k=5$ ) and a 100-way atomic embedding table to 64 dimensions.

We apply three GATv2Conv layers with 4 attention heads, no concatenation (head outputs averaged), and dropout 0.1 in both attention and intermediate activations. Node features are pooled with global mean pooling and passed through a linear layer to produce the scalar output. This architecture explicitly learns attention weights over pairwise interactions conditioned on local geometry.

**Vision-based encoders.** **ViTRegressor** is a lightweight Vision Transformer operating on  $32 \times 32$  images. For models that only consume images, we collapse the 10-channel tensor to 3 channels by selecting the first three channels as a pseudo-RGB representation. Patches of size  $4 \times 4$  yield 64 patch tokens, each embedded via a convolutional patch embedding layer into a 128-dimensional space. A learnable class token and learnable positional embeddings for 65 tokens (1 class + 64 patches) are added before feeding the sequence to a Transformer encoder with 4 encoder blocks, 4 self-attention heads, and a feed-forward dimension of  $4 \times 128$ . The class token output is passed through a layer-normalization and a final linear layer to produce the scalar prediction.

**QuantumShellNet** is a compact CNN specialized for the full 10-channel orbital image tensor. Five convolutional layers with stride 2 and kernel size 3 progressively down-sample the  $32 \times 32$  input to a  $1 \times 1$  spatial representation:  $10 \rightarrow 80 \rightarrow 160 \rightarrow 96 \rightarrow 48 \rightarrow 24$  channels. Each convolution is followed by ReLU and dropout (0.3). The resulting 24-dimensional latent vector is fed into a stack of fully connected layers interleaved with coarse atomic descriptors. In particular, we approximate per-dimer scalar features: total proton number (sum of  $Z$ ), approximate mass number (using  $A \approx 2Z$  for each atom), and total neutron number ( $A-Z$ ). These three scalars are concatenated at multiple stages of the MLP (after the first and second hidden layers) to explicitly inject basic nuclear information. The final MLP has hidden sizes 200, 100, and 50, with ReLU and dropout, culminating in a scalar regression head. All convolutions and linear layers operate in single-precision (float32).

**MultiModal** Combines SchNet-derived geometric features with a ResNet-18 backbone acting on the orbital images. On the graph side, we use the same SchNet configuration as above (16 hidden channels, two interactions, cutoff 5.0 Å). Rather than using only the final scalar output, we extract atom-wise hidden features after the last interaction block, and apply global mean pooling over atoms to obtain a 16-dimensional graph-level descriptor. On the vision side, we use torchvision’s ResNet-18 pretrained on ImageNet, remove the final classification fully-connected layer, and treat the penultimate 512-dimensional global average pooling vector as the image descriptor. The orbital tensor is

again reduced to 3 channels by using the first three channels for compatibility with the pretrained weights.

The two modality-specific embeddings are concatenated into a  $(16+512) = 528$ -dimensional joint representation and passed through a three-layer MLP with hidden widths  $256 \rightarrow 128 \rightarrow 64$ , ReLU activations, and dropout 0.1 after each hidden layer. A final linear layer maps the fused representation to a scalar prediction. This architecture corresponds to a late-fusion baseline in which the network must learn how to weight geometric and image-based cues without explicit cross-attention.

**Dataset interface and label normalization.** All models consume data through a unified `QuantumCanvas` wrapper built on `torch_geometric.data.Data`. Each sample stores atomic numbers  $z$ , Cartesian coordinates  $\mathbf{r}$ , the 10-channel image tensor (if available), and a scalar label  $y$ . For training stability, target values are min-max normalized to the range  $[-1, 1]$  using statistics computed on the training split only; the same statistics are reused for validation and test splits, and all metrics are reported after denormalizing predictions back to the original physical units.

All models are trained under a unified and fully reproducible benchmarking pipeline implemented in PyTorch and PyTorch Geometric. Each experiment is specified by a target property  $y$  and a random seed  $s$ , and all runs follow the same data handling, optimization scheme, evaluation protocol, and logging infrastructure.

**Dataset construction and splits.** For every experiment, we first instantiate the `QuantumCanvas` to read the full `QuantumCanvas` dataset and sample count  $N$ . Indices are randomly shuffled using a seed-controlled `torch.Generator`, and the dataset is partitioned into an 80/10/10 split:

$$\mathcal{D}_{\text{train}}, \quad \mathcal{D}_{\text{val}}, \quad \mathcal{D}_{\text{test}}.$$

Normalization statistics (min/max for the supervised target) are computed *only* from  $\mathcal{D}_{\text{train}}$ , and reused for validation and test. Each `Data` object contains atomic numbers  $z \in \mathbb{N}^2$ , positions  $\mathbf{r} \in \mathbb{R}^{2 \times 3}$ , the 10-channel orbital image tensor  $\mathbf{I}$ , and the normalized label  $\hat{y}$ .

**Input handling across model classes.** The training loop dispatches data to each architecture according to its expected interface:

- **SchNet**, **DimeNet++**, **GATv2**, **EGNN** consume  $(z, \mathbf{r}, \text{batch})$ .
- **FAENet** receives a full PyG batch object and may internally return either a tensor or a dictionary; the regression head is extracted accordingly.

- **ViTRegressor** and **MultiModal** use the first three channels of  $\mathbf{I}$  as a pseudo-RGB image (for compatibility with ImageNet-pretrained ResNet-18 in the multimodal setting).
- **QuantumShellNet** processes the full 10-channel  $32 \times 32$  tensor and also incorporates  $(z, \mathbf{r}, \text{batch})$  information in its fully connected layers.

**Training loop.** All models are optimized using the loss function specified in the configuration (MAE or MSE). For a mini-batch  $\{(\mathbf{x}_i, \hat{y}_i)\}$ , the model produces predictions  $\hat{y}_i^{\text{pred}}$ , the loss is backpropagated, and parameters are updated using one of the configured optimizers (Adam, AdamW, or SGD):

$$\mathcal{L} = \begin{cases} \frac{1}{B} \sum_i |\hat{y}_i - \hat{y}_i^{\text{pred}}| & \text{MAE,} \\ \frac{1}{B} \sum_i (\hat{y}_i - \hat{y}_i^{\text{pred}})^2 & \text{MSE.} \end{cases}$$

Learning rate schedules follow the configuration: ReduceLROnPlateau (validation-loss monitored), cosine annealing, step-decay, or no scheduling. Training continues for up to 50 epochs with early stopping (patience = 10).

**Evaluation and denormalization.** During validation and testing, predictions are first computed in the normalized space and then mapped back to physical units via the training statistics:

$$y = y_{\min} + \frac{\hat{y} + 1}{2}(y_{\max} - y_{\min}).$$

We report MAE and RMSE in the original physical units (eV, Debye, Å, etc.). The checkpoint with the lowest validation loss is used for final test evaluation. Random seeds are applied across PyTorch, NumPy, and dataset index permutation to ensure exact reproducibility.

**Benchmark coverage.** The full benchmark spans *all* models listed in Sec. 8.1, and 22 supervised targets drawn from electronic, energetic, geometric, dipole, and charge-related properties. For each target we run experiments across three independent seeds, totaling  $22 \times 3$  supervised tasks and training each of the 8 architectures under identical conditions.

## 8.2. Fine-tuning on downstream benchmarks

To assess how two-body representations learned on *QuantumCanvas* transfer to larger and more complex systems, we fine-tune a subset of our architectures on three standard downstream benchmarks: *QM9*, *MD17*, and *CrysMTM*. Throughout, we focus on two strong graph-based encoders, **SchNet** and **GotenNet**, which are widely used in molecular and materials machine learning and provide clean, scalable baselines for studying cross-domain transfer. In all cases we reuse exactly the same architectural configurations as in

the two-body setting so that *QuantumCanvas* checkpoints can be loaded without any structural changes.

Across all benchmarks we adopt a unified training protocol wherever possible: mini-batches of size 32, Adam optimization with weight decay 0.0, a ReduceLROnPlateau scheduler (factor 0.8, patience 10, minimum learning rate  $10^{-6}$ ), and early stopping based on validation MAE. For each dataset we run three seeds (42, 123, 456), and for every (target, seed) pair we train two variants per architecture: a *from-scratch* baseline initialized with random weights (learning rate  $10^{-4}$ ), and a *fine-tuned* model initialized from the corresponding two-body *QuantumCanvas* checkpoint and updated with a smaller learning rate  $10^{-5}$  to mitigate catastrophic forgetting.

**QM9: frontier orbital properties.** On *QM9* we target three central electronic quantities: the HOMO energy, the LUMO energy, and the HOMO–LUMO gap. We use the standard PyG split with 110,000 molecules for training, 10,000 for validation, and 10,000 for testing. For each target and seed we train both **SchNet** and **GotenNet** under the two regimes described above. Fine-tuned runs are initialized from two-body checkpoints trained on the matched *QuantumCanvas* targets (`e_homo_ev`, `e_lumo_ev`, `e_g_ev`). The QM9 models use the same SchNet and GotenNet hyperparameters as in Sec. 8.1 (96 hidden channels and six interaction blocks for SchNet; 64-dimensional atom features and three interaction blocks for GotenNet), ensuring that any performance gains can be attributed to two-body pretraining rather than changes in model capacity.

Training uses an L1 loss on the relevant component of the QM9 label vector. All optimization settings follow the shared protocol above, with the only difference between scratch and fine-tuning runs being the learning rate ( $10^{-4}$  vs.  $10^{-5}$ ). For each (target, seed, regime) we select the checkpoint with the lowest validation MAE and report MAE and RMSE on the test set in the original QM9 units.

**MD17: molecular dynamics energies.** On *MD17* we probe transfer to atomistic molecular dynamics by training on three canonical molecules: benzene, aspirin, and ethanol. For each system we construct a compact yet representative split of 950 geometries for training, 50 for validation, and a held-out test subset of 1,000 conformations. Both **SchNet** and **GotenNet** are trained to predict the scalar potential energy of each frame, again using the two-body architectures without modification (SchNet with 96 hidden channels, six interaction blocks, 50 radial basis functions and cutoff 5.0 Å; GotenNet with 64-dimensional atom features, three interactions, 10 radial basis functions, and 5.0 Å cutoff). Fine-tuned runs are initialized from the two-body `total_energy_ev` checkpoint.

For each molecule, targets are normalized using training-set statistics: we compute the mean and standard deviation of the energy over the 950 training frames and optimize an L1 loss in this normalized space. During evaluation, predictions are mapped back to the original MD17 energy units and we report MAE and RMSE on the denormalized values. Optimization hyperparameters follow the QM9 configuration (batch size 32, up to 50 epochs, Adam, scheduling, and early stopping as above), with the same distinction between scratch and fine-tuning learning rates. The best checkpoint per (molecule, seed, regime) is selected by validation MAE and used for final test reporting.

**CrysMTM: crystalline frontier-orbital energies.** On the *CrysMTM*  $\text{TiO}_2$  benchmark we examine transfer from diatomic pretraining to periodic oxide materials. The dataset consists of  $\text{TiO}_2$  crystals sampled across 21 temperatures (0–1000 K in 50 K increments) and multiple rotational configurations. We use all available configurations at these temperatures and focus on two frontier-orbital quantities, **HOMO** and **LUMO**, which are most directly aligned with the electronic targets in *QuantumCanvas* and thus form a clean testbed for cross-domain transfer. As in QM9 and MD17, we use **SchNet** and **GotenNet** with architectures fixed to their two-body configurations, enabling weight reuse without any structural modification.

For each property we randomly partition the dataset into an 80/20 split using a seed-controlled permutation at the crystal level, train on the 80% subset, and hold out the remaining 20% for evaluation. Targets are standardized per property using the mean and standard deviation computed on the training subset, and models are optimized with an MSE loss in this normalized space. All reported metrics (MAE and RMSE) are computed after denormalizing predictions back to the original physical units (eV). Optimization settings mirror the other benchmarks, with the exception that we allow up to 100 epochs and use a single base learning rate of  $10^{-4}$  (and  $10^{-5}$  for fine-tuning), and early stopping patience is reduced to 20 validation steps.

For each target (HOMO, LUMO) and each seed we train from-scratch and fine-tuned variants of both architectures, where fine-tuned runs are initialized from the corresponding two-body *QuantumCanvas* checkpoints (`e_homo_ev` for HOMO and `e_lumo_ev` for LUMO). This yields  $2 \times 3 \times 2 = 12$  CrysMTM experiments per model, carried out under the same data handling, normalization, and optimization protocol as the molecular benchmarks, enabling a controlled comparison of two-body pretraining across molecular, dynamical, and crystalline regimes.