ACCELERATED NEURAL RIEMANNIAN FLOWS FOR KOHN-SHAM DENSITY FUNCTIONAL THEORY

Anonymous authors
Paper under double-blind review

ABSTRACT

Density functional theory (DFT) underlies many electronic structure calculations. Typically, a key step in this framework is the iterative self-consistent field (SCF) procedure, which remains computationally expensive. Recent learning-based approaches accelerate parts of this pipeline by improving Hamiltonian prediction or initialization, but still retain the bottleneck at its root. We propose a *gauge-symmetric neural Riemannian flows* framework for SCF-free electronic structure computation. Our approach directly propagates electronic states on the Grassmann manifold via gauge-equivariant dynamics, rather than relying on a conventional SCF solver. This construction removes the structural bottleneck imposed by the SCF solver and makes the entire propagation trajectory differentiable, suitable for end-to-end learning of the electronic solver. Furthermore, we recast the propagation process at the operator level into a prefix-scan compatible form for efficient parallel execution and acceleration on modern GPUs. Empirically, our method reaches the electronic ground state without SCF, consistently improves energy and orbital related metrics, and achieves speedups of up to **30**$\times$ over prior approaches on MD17 molecules. These results point to a new SCF-free direction for electronic structure computation and highlight the potential of learned electronic solvers that are both accelerator-friendly and aligned with end-to-end differentiable learning.

1 INTRODUCTION

Density functional theory (DFT) has emerged as a standard framework for electronic structure calculations across a broad range of modern scientific disciplines, including computational chemistry (Becke, 2014; Mardirossian & Head-Gordon, 2017), condensed-matter physics (Jones & Gunnarsson, 1989; Baroni et al., 2001), and materials science (Hafner, 2008; Neugebauer & Hickel, 2013). A central advantage of DFT lies in the ability to predict energies, forces, and other electronic properties of molecular and condensed systems at a tractable computational cost. In practice, these quantities are typically computed through a numerical procedure known as the *self-consistent field (SCF)* method, which repeatedly updates the electron density and the effective Hamiltonian until self-consistency is achieved. The SCF procedure therefore constitutes the computational core of DFT, and the success of a DFT calculation depends critically on the efficiency and stability of SCF.

The iterative nature of SCF is thus a double-edged sword: while it is central to the practical success of DFT, it also constitutes one of its major computational bottlenecks. More specifically, attaining self-consistency entails repeated diagonalization and nonlinear updates of both the electron density and the effective Hamiltonian, each of which is computationally expensive (Cancès & Friesecke, 2023). Recent machine learning-based methods have sought to ease this burden by predicting the Hamiltonian or related electronic quantities to improve SCF initialization (Li et al., 2022; Yu et al., 2023; Wang et al., 2024; Tang et al., 2024). However, these methods still function primarily as auxiliary components to the SCF loop, leaving the dominant source of computational cost and convergence difficulty essentially unchanged. This suggests that addressing these challenges may require a more fundamental reformulation of the computational paradigm, rather than further improvements around the SCF loop.

Building on this insight, we introduce a new learning-compatible framework for electronic energy minimization, designed to align with the computational paradigm of modern deep learning rather than with classical SCF iteration. Instead of using learning merely to accelerate individual SCF

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

steps, our approach reformulates electronic structure computation itself as a smooth optimization process over the underlying geometry of electronic states. This reformulation removes the repeated solver iterations that dominate the computational cost of existing approaches and replaces them with a stable, fixed-length computational trajectory. As a result, the proposed framework operates as an end-to-end differentiable optimization pipeline, in contrast to previous methods that learn only partial components while remaining dependent on the SCF scheme. Alongside eliminating the SCF bottleneck, we reformulate the sequential optimization propagation as an operator sequence that admits a prefix-scan algorithm, which enables the logarithmic time complexity with parallel computation. Consequently, our method provides a SCF-free alternative and establishes a framework that accounts for the computational efficiency of the actual execution.

In summary, the main contribution of this paper is two-fold:

1. **Geometric Framework for Kohn-Sham DFT.** We introduce a new theoretical formulation that treats the electronic state update as gradient descent evolution directly on the geometry of electronic states, which eliminates the redundancy caused by the non-uniqueness of orbital representations. This perspective leads to an intrinsic and gauge-equivariant update expressed directly in the electronic space itself. As a result, our formulation provides a principled alternative to conventional SCF solver iteration, while enabling smooth, differentiable quantum trajectory.
2. **Acceleration of Quantum Dynamical System.** We design a scan-compatible operator framework for electronic propagation, which enables multiple time steps to be split into independently computed segments and recombined exactly. Leveraging this property, we develop a prefix-scan based electronic propagation algorithm that executes the composed propagation maps in parallel, which accelerates the optimization process in GPU environments. Overall, our framework achieves substantial runtime acceleration, with speedups of up to **30**×, even comparable to baseline inference time without any SCF iterations.

2 Kohn-Sham Density Functional Theory

In this section, we briefly recall the principle of variational Kohn-Sham formulation. Let $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_{N_\mathbf{r}}) \in \mathbb{R}^{3N_\mathbf{r}}$ denote the nuclear coordinates, where $N_\mathbf{r}$ is the number of atoms. The electronic ground-state energy is defined as the minimum value of the electronic energy functional $\mathcal{E}$ over the occupied orbital coefficient matrix $\mathbf{C}$ under the overlap-orthonormality constraint:

$$\mathbf{E}^\star(\mathbf{r}) \coloneqq \min_{\mathbf{C}} \mathcal{E}(\mathbf{r}, \mathbf{C}), \quad \mathbf{C}^\dagger \mathrm{S}(\mathbf{r}) \mathbf{C} = \mathrm{I}_N, \quad \mathrm{S}_{ij}(\mathbf{r}) = \int_{\mathbb{R}^3} \chi_i(\mathbf{x}; \mathbf{r})^\dagger \chi_j(\mathbf{x}; \mathbf{r}) d\mathbf{x}, \tag{1}$$

where $\{\chi_i\}_{i=1}^M$ are atom-centered basis functions, and M and N denote, respectively, the number of atomic basis functions and occupied orbitals. The matrix $\mathrm{S}(\mathbf{r})$ records the geometry-dependent overlap induced by the non-orthogonal basis, while the constraint $\mathbf{C}^\dagger \mathrm{S}(\mathbf{r}) \mathbf{C} = \mathrm{I}_N$ ensures that the occupied Kohn–Sham subspace remains orthonormal in the overlap metric.

To obtain a tractable representation of $\mathcal{E}$, we adopt the standard *Kohn-Sham* (Cancès & Friesecke, 2023) assumption, which decomposes the total energy into a one-electron contribution together with the Hartree and exchange-correlation terms:

$$\mathcal{E}(\mathbf{r}, \mathbf{C}) = \mathrm{Tr}\big(\mathbf{H}(\mathbf{r}) \mathbf{P}\big) + \frac{1}{2} \iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{\rho_\mathbf{P}(\mathbf{x}) \rho_\mathbf{P}(\mathbf{y})}{\|\mathbf{x} - \mathbf{y}\|} d\mathbf{x} d\mathbf{y} + \mathbf{E}_{\mathrm{xc}}\big[\rho_\mathbf{P}\big], \quad \mathbf{P} = \mathbf{C}\mathbf{C}^\dagger \mathrm{S}(\mathbf{r}), \tag{2}$$

where $\mathbf{C}$ is the orbital matrix and $\mathbf{P}$ is the corresponding density matrix. The double integral represents the classical Coulomb self-interaction of the induced electron density, whereas $\mathbf{E}_{\mathrm{xc}}[\rho_\mathbf{P}]$ accounts for the remaining many-body effects that are not captured by the Hartree term alone. The density associated with $\mathbf{P}$ and the single electron Hamiltonian matrix $\mathbf{H}(\mathbf{r}) \in \mathbb{C}^{M \times M}$ are both expressed with respect to a finite basis $\{\chi_i\}_{i=1}^M$, with the Hamiltonian obtained by projecting the kinetic operator and the external nuclear potential onto the chosen basis:

$$\rho_\mathbf{P}(\mathbf{x}) = \sum_{i,j=1}^{M} \mathbf{P}_{ij} \chi_i(\mathbf{x}) \chi_j(\mathbf{x})^\dagger, \quad \mathbf{H}_{ij}(\mathbf{r}) = -\int_{\mathbb{R}^3} \chi_i(\mathbf{x})^\dagger \left[\frac{1}{2} \nabla^2 + \sum_{I=1}^{N_\mathbf{r}} \frac{\mathrm{Z}_I}{\|\mathbf{x} - \mathbf{r}_I\|} \right] \chi_j(\mathbf{x}) d\mathbf{x},$$

where Z_I denotes the charge of the I-th nucleus. This matrix-level formulation has motivated a number of recent learning-based approaches that aim to accelerate the electronic solve by predicting

Hamiltonian matrices, or closely related operators, directly from molecular structure. In conventional KS-DFT, however, the constrained minimization in equation 1 remains nonlinear because the effective potential depends on the density itself, and the resulting problem is therefore solved in practice through an iterative self-consistent procedure.

Self-Consistent Field. Taking the first-order optimality conditions for minimizing $\mathcal{E}$ over $\mathbf{C}$ under the overlap-orthonormality constraint yields the classical *Roothaan–Hall equations* (Helgaker et al., 2013). At a formal level, the Euler–Lagrange condition is obtained by differentiating the energy with respect to the density matrix, which defines the effective one-electron operator, namely the *Fock matrix*, associated with the current electronic state:

$$\mathbf{F}_{ij}(\mathbf{D}) := \frac{\partial \mathcal{E}}{\partial \mathbf{D}_{ji}} = \mathbf{H}_{ij}(\mathbf{r}) + \sum_{k,l} \left(\iint_{\mathbb{R}^3 \times \mathbb{R}^3} \frac{\chi_i(\mathbf{x})^\dagger \chi_j(\mathbf{x}) \chi_k(\mathbf{y})^\dagger \chi_l(\mathbf{y})}{\|\mathbf{x} - \mathbf{y}\|} d\mathbf{x} d\mathbf{y} \right) \mathbf{D}_{lk} + \frac{\partial \mathbf{E}_{\mathrm{xc}}}{\partial \mathbf{D}_{ji}}.$$

Here, the first term is the one-electron Hamiltonian, the second term is the Hartree contribution written in the atomic-orbital basis, and the last term is the exchange-correlation contribution. The Kohn–Sham problem is nonlinear precisely because these two steps are coupled: the operator depends on the density, but the density is itself constructed from the orbitals obtained by solving the eigenvalue problem for that operator.

Because the effective Hamiltonian must be reconstructed at every iteration and the two-electron contribution is expensive to evaluate, the overall runtime is often dominated by repeated Fock matrix construction, which in standard implementations scales as $\mathcal{O}(M^4)$. Moreover, the SCF procedure is a fixed point iteration that induces a discrete nonlinear dynamical system whose stability depends sensitively on the molecular geometry and the local conditioning of the electronic problem. Even for moderately challenging configurations, the iterates can converge slowly or exhibit persistent oscillations unless stabilized by carefully tuned mixing strategies and solver specific heuristics. This sensitivity, together with variable length and data dependent control flow, is poorly aligned with modern learning pipelines and accelerator hardware, which favor smooth state evolution and fixed, uniform computation across geometries and time steps. We therefore replace the inner SCF iteration with a smooth propagation of electronic state variables that advances an occupied subspace representative $\mathbf{C}$ together with its density projector $\mathbf{P}$ through residual driven dynamics. The resulting update is fixed and accelerator friendly, and it recovers self consistency when the residual vanishes.

3 Riemannian Geometry of Electronic States

Before introducing our method, we formalize the *gauge symmetry* (Martin, 2004) of Kohn–Sham theory. In practice, the electron density $\mathbf{P}$ and total energy $\mathcal{E}$ depend on the occupied subspace $[\mathbf{C}]$, not on a particular choice of occupied orbitals $\mathbf{C}$. In particular, the gauge symmetry is given by

$$\mathbf{P}(\mathbf{C}) \coloneqq \mathbf{C}\mathbf{C}^\dagger \mathrm{S}(\mathbf{r}), \quad \mathbf{P}(\mathbf{C}\mathbf{U}) = \mathbf{P}(\mathbf{C}), \quad [\mathbf{C}] \coloneqq \{\mathbf{C}\mathbf{U} : \mathbf{U} \in \mathrm{U_S}(N)\}. \tag{3}$$

Physically, $\mathbf{P}(\mathbf{C})$ determines the electron density and hence fixes the effective potential and the Kohn–Sham total energy. Therefore, any gauge transformation $\mathbf{C} \mapsto \mathbf{C}\mathbf{U}$ within the occupied space leaves $\mathbf{P}$ unchanged and thus preserves energy $\mathcal{E}$. This symmetry motivates a geometric reformulation in which electronic states are represented as points on a *Riemannian manifold* (Absil et al., 2008), and gauge transformations correspond to motion along fibers that leaves the underlying state invariant.

Gauge-symmetric Electronic Spaces. As suggested by the gauge invariance in equation 3 and illustrated in Figure 1, a matrix $\mathbf{C}$ should be regarded as a convenient *orbital representative* of the physical electronic state, rather than the state itself. To formalize this distinction between orbital representatives and gauge-invariant states, we introduce the generalized *Stiefel* and *Grassmann* manifolds as follows:

$$\textbf{(Electronic Geometry)} \quad \mathrm{St_S} \coloneqq \left\{ \mathbf{C} \in \mathbb{C}^{M \times N} \;\middle|\; \mathbf{C}^\dagger \mathrm{S}(\mathbf{r}) \mathbf{C} = \mathbf{I}_N \right\}, \;\; [\mathbf{C}] \in \mathrm{Gr_S} \coloneqq \mathrm{St_S}/\mathrm{U_S}(N),$$

where the S-unitary group and its associated Lie algebra are defined as follows, respectively:

$$\mathrm{U_S}(M) \coloneqq \left\{ \mathbf{U} \in \mathrm{GL}(M, \mathbb{C}) \middle| \mathbf{U}^\dagger \mathrm{S} \mathbf{U} = \mathrm{S} \right\}, \quad \mathfrak{u}_\mathrm{S}(M) \coloneqq \left\{ \Omega \in \mathbb{C}^{M \times M} \middle| \Omega^\dagger \mathrm{S} + \mathrm{S}\Omega = 0 \right\}. \tag{4}$$

Intuitively, $\mathrm{St_S}$ serves as an *orbital-representative space* where each point $\mathbf{C} \in \mathrm{St_S}$ is a concrete choice of N occupied orbitals that satisfies the S-orthonormality constraint. This representation is

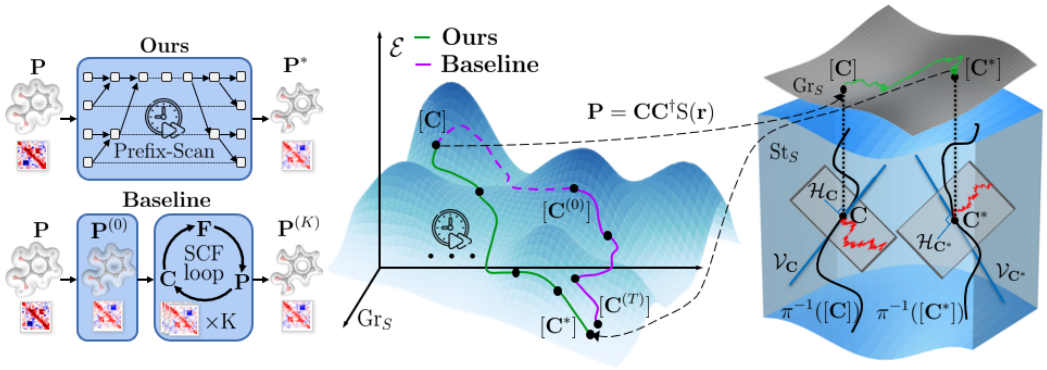

Figure 1: **Overall pipeline.** *(left)* Baselines predict an initialization $\mathbf{P}^{(0)}$ and apply sequential SCF iterations to reach $\mathbf{P}^{(K)}$, while our method directly propagates the electronic state with associative steps evaluated with prefix-scan. *(middle)* The dashed segment denotes the predicted path to the initialization point, while the solid segment shows the subsequent SCF convergence path. Our method instead follows smooth propagation dynamics on Grassmann manifold. *(right)* The electronic state corresponds to an equivalence class $[\mathbf{C}] \in \mathrm{Gr}_S$ represented on Stiefel manifold. The proposed learned dynamics follows horizontal directions to preserve gauge symmetry.

indispensable for computation, since quantities such as orbital velocities, tangent projections, and structure-preserving updates are naturally expressed in terms of $\mathbf{C}$. The Grassmann manifold Gr_S formalizes the identification in equation 3 by collapsing all gauge-equivalent representatives into a single equivalence class. In other words, Gr_S is the minimal configuration space on which the Kohn–Sham energy is well-defined without gauge redundancy.

Now, we view the projection $\pi : \mathrm{St}_S \to \mathrm{Gr}_S$ as a *Riemannian submersion* (Lee, 2018) under the S-metric, *i.e.*, $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle_{\mathbf{C}} := \mathrm{ReTr}\big(\mathbf{v}_1^\dagger \mathrm{S}(\mathbf{r}) \mathbf{v}_2\big)$ on the tangent space. This geometry induces the standard decomposition of the tangent space into vertical gauge and horizontal components, given by

$$T_{\mathbf{C}}\mathrm{St}_S = \Big\{ \mathbf{v} \;\Big|\; \mathbf{C}^\dagger \mathrm{S}\mathbf{v} + \mathbf{v}^\dagger \mathrm{S}\mathbf{C} = \mathbf{0} \Big\}, \;\; \mathcal{V}_{\mathbf{C}}(\mathbf{r}) := \big\{ \mathbf{C}\boldsymbol{\Omega} \;\big|\; \boldsymbol{\Omega}^\dagger = -\boldsymbol{\Omega} \big\}, \;\; \mathcal{H}_{\mathbf{C}} := \big\{ \mathbf{v} \;\big|\; \mathbf{C}^\dagger \mathrm{S}\mathbf{v} = \mathbf{0} \big\}.$$

where $\boldsymbol{\Omega} \in \mathbb{C}^{N \times N}$ denotes a skew-adjoint matrix and the direct-sum decomposition $T_{\mathbf{C}}\mathrm{St}_S = \mathcal{V}_{\mathbf{C}} \oplus \mathcal{H}_{\mathbf{C}}$ follows from orthogonality with respect to the S-metric. For a smooth curve $\mathbf{C}(t) \in \mathrm{St}_S$, let us introduce the *horizontal gauge-covariant derivative* ∇_t induced by the standard $\mathrm{U}_S(N)$-connection, and define the horizontal gradient flow with respect to the S-metric:

$$\underbrace{\nabla_t \mathbf{C}(t) := \dot{\mathbf{C}}(t) - \mathbf{C}(t)\big[\mathbf{C}(t)^\dagger \mathrm{S}(\mathbf{r}) \dot{\mathbf{C}}(t)\big]}_{\text{Horizontal dynamics in Figure 1 } \textit{(right)}}, \quad \underbrace{\dot{\mathcal{E}}(\mathbf{r}, \mathbf{C}(t)) = \big\langle (\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E})^\flat, \; \nabla_t \mathbf{C}(t) \big\rangle}_{\text{Energy minimizing dynamics in Figure 1 } \textit{(right)}}. \tag{5}$$

Here, $(\cdot)^\flat$ and $\mathrm{grad}_{\mathbf{C}}^{\mathrm{Hor}} \mathcal{E}$ denote musical isomorphism and their corresponding gradient flows of energy, respectively. Following by the Kohn-Sham assumption, the horizontal gradient admits a closed-form representation in terms of the self-consistent Hamiltonian, leading to the following result.

Theorem 3.1 (Local Convexity). *For the self-consistent Hamiltonian matrix $\mathbf{H}^\star$ solving Roothaan-Hall equation, the horizontal gradient flows satisfying energy minimizing flows in equation 5 can be written as follows:*

$$\dot{\mathbf{C}}(t) = -\mathrm{grad}_{\mathbf{C}(t)}^{\mathrm{Hor}} \mathcal{E} = -(\mathbf{I} - \mathbf{P})\mathbf{S}(\mathbf{r})^{-1} \mathbf{H}^\star \mathbf{C}(t) \tag{6}$$

Then, the energy $\mathcal{E}(\mathrm{r}, \mathbf{C}(t))$ in equation 5 exhibits local geodesic convexity at the stationary point $\mathbf{C}^\star$.

Theorem 3.1 identifies a smooth gauge-invariant relaxation directly on the quotient manifold Gr_S, so the state variable is the Born–Oppenheimer occupied subspace itself rather than an arbitrary orbital representative. In contrast to SCF iterations, which repeatedly rebuild a density-dependent Hamiltonian and solve a generalized eigenvalue problem, the proposed flows in equation 6 yield a continuous dissipative dynamics with a well-defined descent direction and local convexity near the target subspace. For learning, this is especially attractive because the resulting update is **continuously smooth** and **gauge-symmetric**, making it a natural trainable geometric layer for stable subspace propagation rather than imitation of SCF outputs.

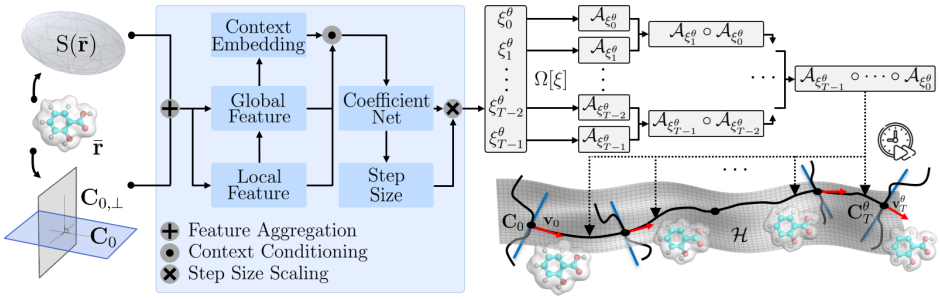

Figure 2: **Gauge-symmetric quantum prefix-scan.** *(left)* The molecular geometry $\mathbf{r}$ provides the overlap matrix, the initial representative $\mathbf{C}_0$ and complementary frame $\mathbf{C}_{0,\perp}$ as input. *(middle)* The model predicts a sequence of horizontal vectors $\{\xi_t^\theta\}_{t=0}^{T-1}$. *(right)* The predicted updates are lifted to quantum bundle actions $\mathcal{A}_{\xi_t}$ and accumulated with prefix-scan, which yields efficient propagation while preserving the gauge symmetry.

Objective Function. Based on the developed geometric setting above, we compare the induced tangent dynamics after projection onto the projector manifold. Specifically, we introduce *gauge-symmetric horizontal flow matching*[1] to quantify the discrepancy between the target and learned quantum horizontal dynamics rather than emulating exact SCF solvers:

$$\mathcal{L}(\theta) = \int_0^T \big(\Pi_{\mathbf{r}}^* g_{\mathcal{P}}\big)_{\mathbf{C}(t)}\big(\Delta_\theta(t), \Delta_\theta(t)\big)dt = \int_0^T \Big\|\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}(t)}\mathcal{E} - \mathbf{v}(\mathbf{C}(t), t; \theta)\Big\|^2_{\Pi_{\mathbf{r}}^* g_{\mathcal{P}}} dt. \quad (7)$$

Here, the vector fields $\mathbf{v}_t(\theta) \coloneqq \mathbf{v}(\mathbf{C}, t; \theta) : \mathcal{H}_{\mathcal{C}} \times [0, T] \times \Theta \to \mathcal{H}_{\mathcal{C}}$ are a parameterized family referred to as *neural Riemannian flows* with the pull-back metric Π^* computed as follows:

$$\Delta_\theta(t) \coloneqq \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}(t)}\mathcal{E} - \mathbf{v}_t(\theta), \quad \big(\Pi_{\mathbf{r}}^* g_{\mathcal{P}}\big)_{\mathbf{C}}(\Delta_\theta, \Delta_\theta) \coloneqq g_{\mathcal{P},\Pi_{\mathbf{r}}(\mathbf{C})}\Big((\Pi_{\mathbf{r}})_{*\mathbf{C}}\Delta_\theta, (\Pi_{\mathbf{r}})_{*\mathbf{C}}\Delta_\theta\Big). \quad (8)$$

Detailed derivations of equation 7 and equation 8 are provided in Section B.1. Intuitively, the proposed objective in equation 7 matches how the physical occupied subspace moves, not which orbital gauge is used to represent it. The target field specifies the correct local descent direction on the quotient geometry, and the neural Riemannian flow is trained to reproduce this direction under the pull-back metric induced by the projector map. As a result, the proposed loss penalizes only physically meaningful errors in subspace dynamics and ignores gauge-dependent discrepancies.

Despite these geometric advantages, a direct realization of the horizontal gradient flow can still incur substantial cost at inference time. As in many standard autoregressive models, the neural Riemannian field is reevaluated at every newly reached point along the sequential reconstruction of the trajectory. Moreover, each update entails geometry-preserving operations needed to preserve horizontality and gauge admissibility. As a result, inference over long horizons becomes increasingly slow and computationally inefficient. In light of these observations, a natural question arises: *How can this computational burden be reduced without sacrificing the underlying geometry?* In what follows, we answer this question through a new algorithmic construction referred to as *quantum prefix-scan*, a key mechanism for inference acceleration.

Quantum Prefix-Scan. A standard strategy in modern machine learning for reducing the sequential depth of long compositions from $\mathcal{O}(T)$ to $\mathcal{O}(\log T)$ is to recast the computation as a **prefix-scan** (Gu & Dao, 2024). The key ingredient is an *associative operation*, which allows partial accumulations to be evaluated in parallel rather than strictly step by step. In our setting, this means that our neural flow $\mathbf{v}(\cdot; \theta)$ in equation 7 should be reformulated in a scan-compatible form so that the induced quantum state updates can be composed and accumulated efficiently with logarithmic sequential depth. Unfortunately, the horizontal gradient flow in equation 5 is defined as a state-dependent evolution, and its native stepwise formulation therefore does not expose the associative algebraic structure required for such parallelization. To endow the propagation with a scan-compatible associative structure, we lift the dynamics to the horizontal bundle and define a bundle action

$$\mathcal{A} : \mathbb{C}^{(M-N)\times N} \cong \mathcal{H}_{\mathbf{C}} \times \mathcal{H} \to \mathcal{H}, \qquad (\xi, (\mathbf{C}, \mathbf{v})) \mapsto \mathcal{A}_\xi(\mathbf{C}, \mathbf{v}), \quad (9)$$

whose successive applications compose associatively, as formalized in the following proposition:

[1]For the mathematical preliminary, omitted notations and definitions, please refer to Appendix B.

Proposition 3.2 (Gauge-Symmetric Quantum Bundle Action)**.** *For each localized horizontal vector $\xi_t \in \mathbb{C}^{(M-N)\times N} \simeq \mathcal{H}_{\mathbf{C}_t}$, let us define the following operation on horizontal space $\mathcal{H}_{\mathbf{C}_t}$:*

$$\Omega[\xi_t] := \mathbf{C}_{t,\perp}\xi_t\mathbf{C}_t^\dagger \mathrm{S} - \mathbf{C}_t\xi_t^\dagger\mathbf{C}_{t,\perp}^\dagger \mathrm{S} = \mathbf{Q}_t \begin{bmatrix} 0 & -\xi_t^\dagger \\ \xi_t & 0 \end{bmatrix} \mathbf{Q}_t^{-1}, \quad \begin{cases} \Omega[\xi_t] \in \mathfrak{u}_{\mathrm{S}}(M), \\ \Omega[\xi_t]\mathbf{C}_t = \mathbf{C}_{t,\perp}\xi_t \in \mathcal{H}_{\mathbf{C}_t}. \end{cases}$$

*Then, the bundle automorphism $\mathcal{A}_\xi$ called a **quantum bundle action** preserves both horizontality and gauge symmetry:*

$$\mathcal{A}_{\xi_t}(\mathbf{C}, \mathbf{v}) := \exp(\Omega[\xi_t])(\mathbf{C}, \mathbf{v}), \quad \begin{cases} \mathbf{C}^\dagger \mathrm{S}\mathbf{v} = 0 \Longleftrightarrow \big(\exp(\Omega[\xi_t])\mathbf{C}\big)^\dagger \mathrm{S}\big(\exp(\Omega[\xi_t])\mathbf{v}\big) = 0, \ (\mathbf{A}), \\ \mathcal{A}_{\xi_t}\big((\mathbf{C}, \mathbf{v}) \cdot \mathbf{U}\big) = \mathcal{A}_{\xi_t}(\mathbf{C}, \mathbf{v}) \cdot \mathbf{U}, \quad \forall\, \mathbf{U} \in \mathrm{U}_{\mathrm{S}}(N), \ (\mathbf{B}). \end{cases}$$

Furthermore, the induced actions ensure a scan-compatible associative composition law:

$$\mathcal{A}_{\xi_{T-1}} \circ \cdots \circ \Big((\mathcal{A}_{\xi_{i+2}} \circ \mathcal{A}_{\xi_{i+1}}) \circ \mathcal{A}_{\xi_i}\Big) \circ \cdots \circ \mathcal{A}_{\xi_0} = \mathcal{A}_{\xi_{T-1}} \circ \cdots \circ \Big(\mathcal{A}_{\xi_{i+2}} \circ (\mathcal{A}_{\xi_{i+1}} \circ \mathcal{A}_{\xi_i})\Big) \circ \cdots \circ \mathcal{A}_{\xi_0}.$$

Proposition B.8 establishes two key properties of the proposed construction. First, it shows that the quantum bundle action $\mathcal{A}$ preserves both (A) *horizontality* and (B) *gauge symmetry*, thereby ensuring that the resulting evolution remains geometrically consistent and physically admissible. Second, it shows that these bundle actions compose associatively, thereby providing the algebraic foundation required for prefix-scan acceleration.

Figure 2 further illustrates how prefix-scan is incorporated into the propagation pipeline and ultimately reconstructs $\mathbf{C}_t$. Given $(t, \mathbf{C}(t))$ as input, the proposed neural network predicts local horizontal displacements, lifts them to the corresponding quantum bundle actions $\{\mathcal{A}_{\xi_t^\theta}\}$, and accumulates their compositions through prefix-scan to generate the propagated state efficiently. Accordingly, the generated horizontal vectors ξ_t^θ provide a structured parameterization of the vector field $\mathbf{v}_t(\theta) := \mathbf{v}(\mathbf{C}(t), t; \theta)$ in equation 7 as follows:

Figure 2 *(right)* $\quad (\mathbf{C}, \mathbf{v}_t(\theta)) = \mathcal{A}_{\xi_t^\theta}(\mathbf{C}_{t-1}, \mathbf{v}_{t-1}(\theta)), \quad (\mathbf{C}_t^\theta, \mathbf{v}_t^\theta) = (\mathcal{A}_{\xi_t^\theta} \circ \cdots \circ \mathcal{A}_{\xi_0^\theta})(\mathbf{C}_0, \mathbf{v}_0).$

This formulation turns long-horizon inference itself into a scan-compatible structure. Since the bundle actions induced by the predicted horizontal vectors $\{\xi_\cdot\}$ satisfy associativity, long trajectories can be reconstructed by first predicting the full sequence $\{\xi_t^\theta\}_{t=0}^{T-1}$ and then accumulating the corresponding actions in parallel through prefix-scan. Consequently, *the sequential depth of a T-step propagation is reduced from $\mathcal{O}(T)$ to $\mathcal{O}(\log T)$*, while each intermediate accumulation remains a physically admissible action on the horizontal bundle. This acceleration is not an external approximation layered on top of the geometry, but a direct consequence of recasting the learned evolution into an associative and constraint-preserving bundle action. In Section 5, we show that the proposed construction delivers substantial inference speedups while retaining strong predictive performance in the experiment.

4 Related Work

Electronic Structure Prediction for SCF Initialization. Existing learning-based approaches to electronic structure prediction focus on predicting electronic quantities from molecular structures. PhiSNet (Unke et al., 2021) predicts electronic wavefunctions and densities using an $\mathrm{SE}(3)$-equivariant architecture that represents the electronic state through a learned model. Subsequent work has focused on predicting the Kohn–Sham Hamiltonian or related operators to support the integration of learned models with existing electronic structure pipelines. For instance, QHNet (Yu et al., 2023) and DEQHNet (Wang et al., 2024) learn Hamiltonian representations from molecular structures. WANet (Li et al., 2025) and SPHNet (Luo et al., 2025a) further develop Hamiltonian prediction models with improved scalability and computational efficiency, while QHFlow (Kim et al., 2025a) models the distribution of Hamiltonian operators using flow-based generative models. These methods improve the accuracy of the electronic structure prediction, but are typically designed and used as initializations or auxiliary components for SCF calculations. As a result, the final electronic state is still determined through iterative SCF solvers.

Manifold Optimization for Electronic Structure. A complementary line of work explores solving electronic structure problems by casting them as direct energy minimization under orthogonality constraints instead of relying on SCF solver iterations. The geometric framework of optimization

Table 1: **Quantitative results on salicylic acid in MD17.** We report quantum physical metrics with inference time and SCF iteration counts. Best and second best results are highlighted in **bold** and **blue**, respectively.

Model	SCF	$\mathcal{E}$ (E_h) ↓	$\mathfrak{S}_c$ (%) ↑	ϵ_{occ} (E_h) ↓	$\mathbf{v}\epsilon$ (E_h) ↓	$\|\epsilon^{\text{DIIS}}\|$ (E_h) ↓	Inference (s) ↓	Speed Up ↑
PhiSNet (Unke et al., 2021)	0	-2.20363×10^2	20.98	3.84×10^0	1.34×10^{-1}	3.97×10^0	**2.56×10^0**	**5.69×**
	9	-4.94039×10^2	98.57	6.94×10^{-2}	7.49×10^{-2}			
QHNet (Yu et al., 2023)	0	-4.48601×10^2	28.05	1.42×10^1	1.23×10^{-1}	**1.07×10^0**	8.82×10^0	1.66×
	9	**-4.95030×10^2**	**99.87**	**1.81×10^{-2}**	**7.70×10^{-3}**			
DEQHNet (Wang et al., 2024)	0	-3.80573×10^2	36.89	2.82×10^0	1.09×10^{-1}	1.81×10^0	1.12×10^1	1.31×
	9	-4.94705×10^2	99.76	1.12×10^{-1}	1.11×10^{-1}			
WANet (Li et al., 2025)	0	-4.31822×10^2	66.98	2.67×10^0	7.16×10^{-2}	6.50×10^0	1.46×10^1	1.00×
	9	-4.92113×10^2	95.16	1.51×10^{-1}	8.18×10^{-2}			
SPHNet (Luo et al., 2025a)	0	-4.54412×10^2	43.58	2.06×10^0	6.17×10^{-2}	6.05×10^0	1.26×10^1	1.16×
	9	-4.91936×10^2	94.88	1.30×10^{-1}	9.57×10^{-2}			
QHFlow (Kim et al., 2025a)	0	-4.53624×10^2	65.07	9.37×10^{-1}	7.09×10^{-1}	5.18×10^0	5.66×10^0	2.58×
	9	-4.93197×10^2	96.66	1.31×10^{-1}	8.67×10^{-2}			
Ours	0	**-4.95109×10^2**	**99.99**	**3.55×10^{-4}**	**4.54×10^{-4}**	**3.63×10^{-1}**	**4.92×10^{-1}**	**29.67×**

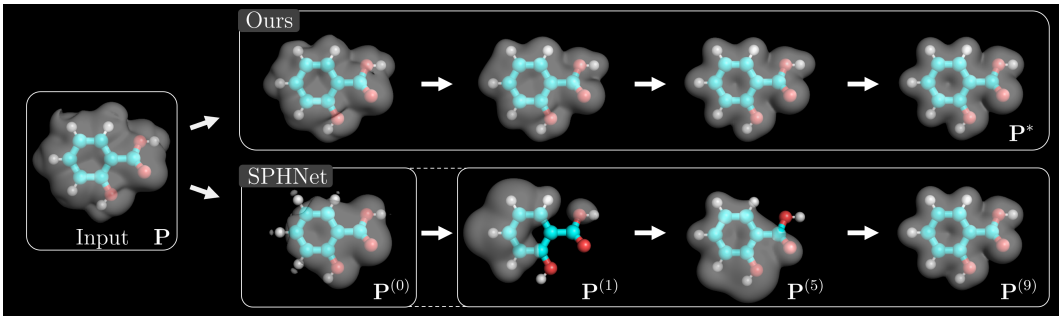

Figure 3: **Qualitative results on salicylic acid.** We visualize the electronic state dynamics of salicylic acid ($C_7H_6O_3$), with our framework shown on top and SPHNet on the bottom for comparison.

on Riemannian manifolds was established in earlier work (Edelman et al., 1998). Building on this geometric perspective, several approaches have been proposed for minimizing the Kohn–Sham energy, including orbital transformation based methodologies (VandeVondele & Hutter, 2003). More recent work has explored Riemannian optimization techniques such as gradient and Newton approaches on the Stiefel manifold, as well as direct minimization on complex Stiefel manifolds for Kohn–Sham DFT (Altmann et al., 2022; 2024; Luo et al., 2025b). However, these approaches are typically designed as hand-crafted numerical solvers with fixed update rules, which makes them difficult to parameterize in a learnable form or integrate with learning-based electronic structure models.

5 Experiments

In this section, we benchmark the proposed method against competitive baselines, including SPHNet and QHFlow, and report results on representative MD17 (Chmiela et al., 2017) and MD22 (Chmiela et al., 2023) molecular systems. We use quantum physical metrics, such as energy and orbital errors, to measure how closely the predicted electronic state matches converged references, and efficiency metrics to quantify computational cost through SCF iterations and inference time. All reference generation, post-hoc electronic structure evaluation, and SCF refinement procedures are performed in a unified PySCF environment (Sun et al., 2018), using def2-SVP, a split valence double-ζ basis with polarization functions. This setting yields atomic orbital dimensions of $M = 170$ to 222 on MD17 and $M = 390$ to 496 on MD22, defining Hamiltonian and overlap matrices of size $M \times M$.

MD17/MD22 Molecular Benchmarks. For the main experiments, we evaluate Salicylic Acid ($C_7H_6O_3$) from MD17 and the alanine tetrapeptide Ac-Ala3-NHMe ($C_{12}H_{22}N_4O_4$) from MD22. Salicylic Acid contains multiple functional groups, nonuniform bonding patterns, and polarization effects, making its electronic structure sensitive to molecular geometry. Ac-Ala3-NHMe provides a larger peptide system with an expanded atomic orbital space, allowing us to test the proposed SCF-free electronic propagation under increased molecular size and electronic degrees of freedom. The main quantitative results are reported in Tables 1 and 2. Additional experiments on Paracetamol ($C_8H_9NO_2$), Aspirin ($C_9H_8O_4$), and DHA ($C_{22}H_{32}O_2$) are provided in Section A.2.

Table 2: **Quantitative results on Ac-Ala3-NHMe in MD22.** We report quantum physical metrics with inference time and SCF iteration counts. Best and second best results are highlighted in **bold** and **blue**, respectively.

Model	SCF	$\mathcal{E}$ (E_h) ↓	$\mathfrak{S}_c$ (%) ↑	$\boldsymbol{\epsilon}_{\text{occ}}$ (E_h) ↓	$\mathbf{v}\boldsymbol{\epsilon}$ (E_h) ↓	$\|\boldsymbol{\epsilon}^{\text{DIIS}}\|$ (E_h) ↓	Inference (s) ↓	Speed Up ↑
PhiSNet (Unke et al., 2021)	0	-3.70551×10^2	21.00	3.60×10^0	1.62×10^{-1}	4.34×10^1	5.88×10^1	1.36×
	13	3.00311×10^1	25.93	1.03×10^0	1.45×10^{-1}			
QHNet (Yu et al., 2023)	0	1.12490×10^2	18.83	N/A	N/A	4.16×10^1	$\mathbf{5.60 \times 10^1}$	**1.42×**
	13	-5.98483×10^2	39.66	1.45×10^0	1.57×10^{-1}			
DEQHNet (Wang et al., 2024)	0	-7.37929×10^2	38.86	2.15×10^0	1.49×10^{-1}	$\mathbf{2.78 \times 10^0}$	7.98×10^1	1.00×
	13	$\mathbf{-9.87746 \times 10^2}$	**99.79**	9.13×10^{-2}	1.48×10^{-1}			
WANet (Li et al., 2025)	0	-8.48269×10^2	59.63	3.13×10^0	1.25×10^{-1}	8.08×10^0	7.06×10^1	1.13×
	13	-9.79237×10^2	91.51	7.64×10^{-2}	1.50×10^{-1}			
SPHNet (Luo et al., 2025a)	0	-9.09231×10^2	45.57	1.99×10^0	$\mathbf{7.97 \times 10^{-2}}$	4.70×10^0	7.00×10^1	1.14×
	13	-9.86387×10^2	97.93	$\mathbf{3.59 \times 10^{-2}}$	1.55×10^{-1}			
QHFlow (Kim et al., 2025a)	0	-8.96689×10^2	59.75	1.39×10^0	1.12×10^{-1}	1.92×10^1	6.81×10^1	1.17×
	13	-9.22758×10^2	68.40	1.82×10^{-1}	1.56×10^{-1}			
Ours	0	$\mathbf{-9.88139 \times 10^2}$	**99.09**	$\mathbf{1.12 \times 10^{-2}}$	$\mathbf{1.82 \times 10^{-2}}$	$\mathbf{1.46 \times 10^0}$	$\mathbf{1.60 \times 10^0}$	**49.88×**

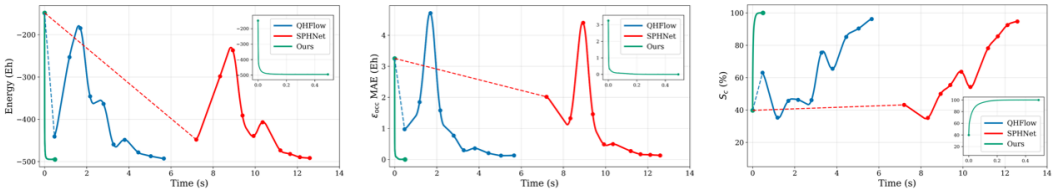

Figure 4: **Time efficiency and convergence comparison on salicylic acid.** We compare our method against the baselines over throughput for the following metrics. *(left)* Energy functional $\mathcal{E}$. *(middle)* Occupied orbital energies $\boldsymbol{\epsilon}_{\text{occ}}$. *(right)* Occupied orbital coefficient similarity $\mathfrak{S}_c$.

Comparison Results. We evaluate the proposed framework using physical fidelity and computational efficiency metrics. For physical fidelity, we consider the electronic energy functional $\mathcal{E}$, the occupied orbital coefficient similarity $\mathfrak{S}_c$, the mean absolute errors of the occupied orbital energies $\boldsymbol{\epsilon}_{\text{occ}}$, and the HOMO–LUMO gap error $\mathbf{v}\boldsymbol{\epsilon}$. We also monitor self-consistency through the DIIS residual norm, $\|\boldsymbol{\epsilon}^{\text{DIIS}}\|$, and report entries with large initial errors as N/A. For efficiency, we report inference time and speedup computed relative to the baseline with the largest inference cost.

Across all molecular systems, the proposed framework shows consistently strong physical fidelity. It attains a lower electronic energy functional relative to the strongest baseline and reduces $\|\boldsymbol{\epsilon}^{\text{DIIS}}\|$ by approximately **2.95×**, with similar improvements in $\boldsymbol{\epsilon}_{\text{occ}}$ and $\mathbf{v}\boldsymbol{\epsilon}$. These results suggest that gauge-symmetric horizontal flows on the Grassmann manifold help produce electronically stable states aligned with self-consistent solutions.

Figure 3 and Figure 4 illustrate how different initializations evolve during SCF refinement. Baseline refinement paths often fluctuate before reaching the refined solution, while energy, occupied orbital error, and coefficient similarity evolve at different rates. Thus, comparing only initial predictions or fully converged SCF solutions is less informative, since the former misses downstream solver behavior and the latter obscures learned initialization quality. For a controlled comparison, we apply a fixed SCF refinement budget to each model initialization. The budget is chosen as the number of iterations required to reduce the energy spread among baselines by 90% from its initial value, resulting in 9 iterations for MD17 and 13 iterations for MD22.

Under this fixed budget setting, the ranking of baselines differs from their ranking at initialization, suggesting that lower initial energy alone is insufficient to characterize downstream SCF behavior. Even after the baseline predictions are refined with the fixed SCF budget, the proposed framework remains competitive across the evaluated properties while using zero SCF iterations. The smoother and more monotonic prediction behavior is consistent with the local convexity induced by the formulation. Together, these results support the proposed framework as a route to accurate and physically consistent electronic states without relying on iterative SCF refinement.

We further evaluate computational efficiency. The gain comes from avoiding iterative SCF solvers, a dominant bottleneck in the baselines, and from reducing the cost of long propagation steps through prefix-scan acceleration. The proposed framework achieves the lowest inference time and additionally provides a conservative **35.00×** speedup relative to the fastest competing baseline. Overall, it combines stable near self-consistent electronic predictions with substantial runtime savings. Further implementation and experimental details are provided in Section A.1.

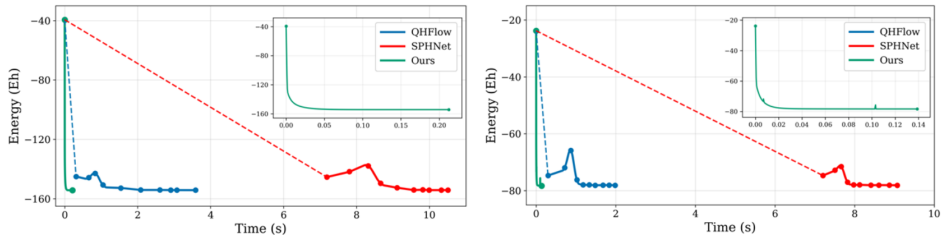

Figure 5: **Convergence comparison on degenerate molecular.** We compare the energy functional $\mathcal{E}$ over inference time between our method and the baselines on *(left)* Cyclobutadiene and *(right)* Twisted Ethylene.

Efficient and Consistently Convergent SCF-Free Inference. To evaluate the time efficiency and convergence performance of the proposed framework, we compare the time-resolved convergence behavior of the energy, ϵ_{occ} and $\mathfrak{S}_c$ for each model. As shown in Figure 4, existing methods only partially reduce SCF iterations through improved initialization, but the bottleneck induced by the iterative SCF solver and the instability of the convergence trajectory remain, resulting in substantial inference time. In contrast, our model exhibits stable and predominantly monotonic improvements consistent with the local geodesic convexity in Theorem 3.1, while achieving an $\underline{11.50\times}$ speedup, as shown in the inset subplots. This speedup stems from a prefix-scan based propagation algorithm, which further reduces the overall computational latency. Consequently, the combination of the SCF-free design of gauge-symmetric Riemannian flow for electronic state propagation and the prefix-scan algorithm leads to an advantage in terms of time efficiency and stable convergence behavior.

Degenerate Molecular Systems. We further broaden the evaluation to systems that are known to pose severe challenges for SCF. The main difficulty of these systems lies not merely in the final energy level, but in the stability of convergence under a sensitive electronic landscape. Near-degenerate systems form a class of molecules that have been reported to exhibit these instabilities as a result of electronic near-degeneracy, symmetry breaking, and competition among multiple solutions, with representative examples including Cyclobutadiene (C_4H_4) and Twisted Ethylene (C_2H_4) (Monino et al., 2022; Saade & Burton, 2024).

In this study, we adopt the challenging geometries established in the literature for these two systems. While the baselines remain affected by these challenging conditions and converge to less favorable solutions, our model exhibits a monotonic decrease in energy consistent with the behavior observed for structurally stable molecules as shown in Table 3, Figure 5, and Figure 6. This result strongly suggests that the local convexity induced by the proposed gauge-symmetric dynamics supports stable convergence behavior even in challenging cases, where small errors often lead to destabilized electronic solvers.

Table 3: **Energy comparison for challenging molecular systems.** For each molecular system, we report the resulting energy, with baseline methods evaluated after 9 SCF iterations.

Models	Cyclobutadiene	Twisted Ethylene
SPHNet	-1.54216×10^2	-7.80923×10^1
QHFlow	-1.54251×10^2	-7.80955×10^1
Ours	$\mathbf{-1.54316 \times 10^2}$	$\mathbf{-7.82468 \times 10^1}$

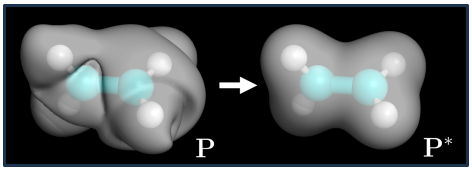

Figure 6: **Visualization on twisted ethylene.** We visualize the predicted electronic state on twisted ethylene, where $\mathbf{P}$ denotes the initial state and $\mathbf{P}^*$ denotes the state predicted by our model.

6 Conclusion

In this work, we develop an SCF-free framework for electronic state computation based on gauge-symmetric neural Riemannian flows on the Grassmann manifold. Rather than relying on a conventional SCF solver, the method propagates electronic states along horizontal directions consistent with gauge symmetry of the manifold. The propagation operators are expressed in a Lie group form that admits prefix-scan execution, which allows efficient parallel evaluation on GPU environments. Experiments on molecular benchmarks demonstrate that the framework produces accurate electronic states without SCF, substantially reduces inference time, and preserves stable energy relaxation in challenging electronic regimes. These results suggest that SCF-free electronic-structure computation can be realized by reformulating the problem as a smooth and parallelizable geometric optimization.

AI USE STATEMENT

We used generative AI tools to improve the clarity of the writing and support the development of mathematical proofs. We have reviewed all content produced with AI assistance and checked the mathematical arguments for correctness. We take full responsibility for the final content of this work.

REFERENCES

P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

Robert Altmann, Daniel Peterseim, and Tatjana Stykel. Energy-adaptive riemannian optimization on the stiefel manifold. *ESAIM: Mathematical Modelling and Numerical Analysis*, 56(5):1629–1653, 2022.

Robert Altmann, Daniel Peterseim, and Tatjana Stykel. Riemannian newton methods for energy minimization problems of kohn–sham type. *Journal of Scientific Computing*, 101(1):6, 2024.

Stefano Baroni, Stefano de Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Reviews of Modern Physics*, 73(2): 515–562, 2001.

Axel D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics*, 140(18):18A301, 2014.

Eric Cancès and Gero Friesecke. *Density Functional Theory*. Springer, 2023.

Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science Advances*, 3(5):e1603015, 2017.

Stefan Chmiela, Valentin Vassilev-Galindo, Oliver T. Unke, Adil Kabylda, Huziel E. Sauceda, Alexandre Tkatchenko, and Klaus-Robert Müller. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances*, 9(2):eadf0873, 2023. doi: 10.1126/sciadv.adf0873. URL https://www.science.org/doi/abs/10.1126/sciadv.adf0873.

Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998. doi: 10.1137/S0895479895290954.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *The First Conference on Language Modeling*, 2024.

Jürgen Hafner. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of Computational Chemistry*, 29(13):2044–2078, 2008. doi: 10.1002/jcc.21057.

Trygve Helgaker, Poul Jorgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, 2013. doi: 10.1002/9781119019572.

Robert O Jones and Olle Gunnarsson. The density functional formalism, its applications and prospects. *Reviews of Modern Physics*, 61(3):689–746, 1989. doi: 10.1103/RevModPhys.61.689.

Seongsu Kim, Nayoung Kim, Dongwoo Kim, and Sungsoo Ahn. High-order equivariant flow matching for density functional theory hamiltonian prediction. In *Advances in Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=iFIjNXb0Y5.

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. Pubchem 2025 update. *Nucleic Acids Research*, 53 (D1):D1516–D1525, 2025b. doi: 10.1093/nar/gkae1059.

John M. Lee. *Introduction to Riemannian Manifolds*. Springer, 2 edition, 2018. doi: 10.1007/978-3-319-91755-9.

He Li, Zun Wang, Nianlong Zou, Meng Ye, Runzhang Xu, Xiaoxun Gong, Wenhui Duan, and Yong Xu. Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nature Computational Science*, 2(6):367–377, 2022. doi: 10.1038/s43588-022-00265-6.

Yunyang Li, Zaishuo Xia, Lin Huang, Xinran Wei, Han Yang, Sam Harshe, Zun Wang, Chang Liu, Jia Zhang, Bin Shao, and Mark B. Gerstein. Enhancing the scalability and applicability of kohn-sham hamiltonians for molecular systems. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=twEvvkQqPS.

Erpai Luo, Xinran Wei, Lin Huang, Yunyang Li, Han Yang, Zaishuo Xia, Zun Wang, Chang Liu, Bin Shao, and Jia Zhang. Efficient and scalable density functional theory hamiltonian prediction through adaptive sparsity. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 41368–41390. PMLR, 13–19 Jul 2025a. URL https://proceedings.mlr.press/v267/luo25l.html.

Kai Luo, Tingguang Wang, and Xinguo Ren. Direct minimization on the complex stiefel manifold in kohn-sham density functional theory for finite and extended systems. *Computer Physics Communications*, 312, 2025b. doi: 10.1016/j.cpc.2025.109596.

Narbe Mardirossian and Martin Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Molecular Physics*, 115(19):2315–2372, 2017. doi: 10.1080/00268976.2017.1333644.

Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004.

Enzo Monino, Martial Boggio-Pasqua, Anthony Scemama, Denis Jacquemin, and Pierre-Francois Loos. Reference energies for cyclobutadiene: Automerization and excited states. *The Journal of Physical Chemistry A*, 126(28):4664–4679, 2022. doi: 10.1021/acs.jpca.2c02480.

Jörg Neugebauer and Tilmann Hickel. Density functional theory in materials science. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(5):438–448, 2013. doi: 10.1002/wcms.1125.

Sandra Saade and Hugh G. A. Burton. Excited state-specific casscf theory for the torsion of ethylene. *Journal of Chemical Theory and Computation*, 20(12):5105–5114, 2024. doi: 10.1021/acs.jctc.4c00212.

Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet K. L. Chan. Pyscf: the python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1340, 2018. doi: 10.1002/wcms.1340.

Zechen Tang, He Li, Peize Lin, Xiaoxun Gong, Gan Jin, Lixin He, Hong Jiang, Xinguo Ren, Wenhui Duan, and Yong Xu. A deep equivariant neural network approach for efficient hybrid density functional calculations. *Nature Communications*, 15(1):8815, 2024. doi: 10.1038/s41467-024-53028-4.

Oliver Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess Smidt, and Klaus-Robert Müller. Se(3)-equivariant prediction of molecular wavefunctions and electronic densities. *Advances in Neural Information Processing Systems*, 34:14434–14447, 2021.

Joost VandeVondele and Jürg Hutter. An efficient orbital transformation method for electronic structure calculations. *The Journal of Chemical Physics*, 118(10):4365–4369, 2003. doi: 10.1063/1.1543154.

Zun Wang, Chang Liu, Nianlong Zou, He Zhang, Xinran Wei, Lin Huang, Lijun Wu, and Bin Shao. Infusing self-consistency into density functional theory hamiltonian prediction via deep equilibrium models. *Advances in Neural Information Processing Systems*, 37:89652–89681, 2024.

Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Efficient and equivariant graph networks for predicting quantum hamiltonian. In *International Conference on Machine Learning*, pp. 40412–40424. PMLR, 2023.

A Appendix

A.1 Experimental Details

Data Generation. For reference generation, we use molecular geometries sampled from the MD17 and MD22 benchmark trajectories and compute the corresponding electronic reference quantities with a PySCF-based data generation pipeline. All calculations use the def2-SVP basis set, grid level 3, a convergence tolerance of 10^{-10}, and at most 80 SCF iterations for RKS with the PBE exchange-correlation functional. For each molecular geometry, we store the converged overlap matrix $S(r) \in \mathbb{R}^{M\times M}$, Fock matrix $F(r) \in \mathbb{R}^{M\times M}$, molecular orbital energies $\epsilon \in \mathbb{R}^{M}$, occupied orbital coefficients $C \in \mathbb{R}^{M\times N}$, and occupied orbital energies $\epsilon_{\mathrm{occ}} \in \mathbb{R}^{N}$. For analyses involving SCF refinement, we also store intermediate electronic quantities across SCF iterations. PubChem structures (Kim et al., 2025b) are used to initialize standalone inference simulations and qualitative visualizations when an external molecular structure is required.

Dataset Statistics. We summarize the dataset statistics used in our experiments in Table 4, including the molecular systems and their corresponding shapes.

Table 4: **Matrix and orbital dimensions for each molecular system.** We report the dimensions of the overlap matrix $\mathbf{S}$, Hamiltonian $\mathbf{H}$, and occupied orbital coefficient matrix $\mathbf{C}$, together with the number of atomic basis functions M and occupied orbitals N.

Molecule	**Atoms**	**Formula**	S	**H**	**C**	M	N
Salicylic acid	16	$C_7H_6O_3$	170×170	170×170	170×36	170	36
Aspirin	21	$C_9H_8O_4$	222×222	222×222	222×47	222	47
Paracetamol	20	$C_8H_9NO_2$	199×199	199×199	199×40	199	40
Ac-Ala3-NHMe	42	$C_{12}H_{22}N_4O_4$	390×390	390×390	390×77	390	77
DHA	56	$C_{22}H_{32}O_2$	496×496	496×496	496×90	496	90
Cyclobutadiene	8	C_4H_4	76×76	76×76	76×14	76	14
Twisted ethylene	6	C_2H_4	48×48	48×48	48×8	48	8

Training Setup. All experiments were implemented in Python 3.10.18 using PyTorch 2.5.1 with CUDA 12.4 and conducted on a single NVIDIA RTX 5000 Ada GPU. The predictor was optimized with AdamW using learning rate 10^{-4} and weight decay 10^{-6}, and trained using the objective defined in the methodology section. For runtime evaluation, CUDA synchronization was enabled to obtain consistent timing measurements, and reported inference times exclude initial warm up iterations. Across repeated runs, all methods exhibited less than 5% performance variation, and the qualitative conclusions were unchanged. We therefore omit standard deviations for clarity.

Baseline Implementations. We compare against the following neural electronic structure baselines, using the official public implementations of the respective authors whenever available. For fair comparison, all methods are evaluated on the same set of fixed molecular geometries under a common PySCF-based electronic structure pipeline with the same settings used for data generation, which includes the basis set, exchange-correlation functional, and grid level.

- **DEQHNet**: `https://github.com/Zun-Wang/DEQHNet`
- **QHFlow**: `https://github.com/seongsukim-ml/QHFlow`
- **QHNet (AIRS)**: `https://github.com/divelab/AIRS`
- **SPHNet / WANet**: `https://github.com/microsoft/SPHNet`
- **PhiSNet**: `https://github.com/AIRI-Institute/nablaDFT/tree/main`

Evaluation Protocol. Let $\mathbf{C}$ denote the predicted occupied orbital coefficient matrix and let $\mathbf{C}^\star$ be the converged reference. We report the electronic energy functional $\mathcal{E} \coloneqq \mathcal{E}(\mathbf{r}, \mathbf{C})$ directly, since lower energy indicates a state closer to the ground-state solution under the same Kohn–Sham functional. We also use the occupied orbital coefficient similarity $\mathfrak{S}_c$, which compares the predicted and reference coefficient matrices at the orbital level after orthogonal Procrustes alignment. Specifically, we

compute

$$\arg\min_{\mathbf{Q}^\top\mathbf{Q}=\mathbf{I}} \|\mathbf{C}\mathbf{Q}-\mathbf{C}^\star\|_{\mathrm{F}}, \quad \mathfrak{S}_c(\mathbf{C},\mathbf{C}^\star)=\frac{\langle \mathbf{C}\mathbf{Q}^\star,\mathbf{C}^\star\rangle_{\mathrm{F}}}{\|\mathbf{C}\mathbf{Q}^\star\|_{\mathrm{F}}\|\mathbf{C}^\star\|_{\mathrm{F}}}. \tag{10}$$

For the occupied orbital energies and the HOMO–LUMO gap, let $\boldsymbol{\epsilon}_{\mathrm{occ}}$ and $\boldsymbol{\epsilon}^\star_{\mathrm{occ}}$ denote the predicted and reference occupied orbital energies, respectively. We report the mean absolute error over occupied orbital energies, together with the absolute error in the HOMO–LUMO gap:

$$\boldsymbol{\epsilon}_{\mathrm{occ}}=\frac{1}{N}\sum_{i=1}^{N}|\epsilon_i-\epsilon_i^\star|, \quad \mathbf{v}\boldsymbol{\epsilon}=\left|\left(\epsilon_{\mathrm{LUMO}}-\epsilon_{\mathrm{HOMO}}\right)-\left(\epsilon^\star_{\mathrm{LUMO}}-\epsilon^\star_{\mathrm{HOMO}}\right)\right|. \tag{11}$$

These two metrics measure whether the predicted state reproduces not only the occupied spectrum itself but also the separation between frontier orbitals relevant to electronic structure.

As an auxiliary diagnostic, the norm of the DIIS error is monitored to quantify self-consistency. Using the Fock matrix $\mathbf{F}$, density matrix $\mathbf{P}$, and overlap matrix S, the DIIS error is defined as the commutator residual

$$\boldsymbol{\epsilon}^{\mathrm{DIIS}}=\mathbf{F}\mathbf{P}\mathrm{S}-\mathrm{S}\mathbf{P}\mathbf{F}, \quad \|\boldsymbol{\epsilon}^{\mathrm{DIIS}}\|=\|\mathbf{F}\mathbf{P}\mathrm{S}-\mathrm{S}\mathbf{P}\mathbf{F}\|_{\mathrm{F}}. \tag{12}$$

Smaller values indicate that the predicted state is closer to satisfying the self-consistent condition.

Table 5: **Quantitative results on aspirin in MD17.** We report quantum physical metrics with inference time and SCF iteration counts. Best and second best results are highlighted in **bold** and **blue**, respectively.

Model	SCF	$\mathcal{E}$ $(E_h)\downarrow$	$\mathfrak{S}_c$ (%) $\uparrow$	ϵ_{occ} $(E_h)\downarrow$	$\mathbf{v}\epsilon$ $(E_h)\downarrow$	$\|\epsilon^{\text{DIIS}}\|$ $(E_h)\downarrow$	Inference (s) $\downarrow$	Speed Up $\uparrow$
PhiSNet (Unke et al., 2021)	0	-2.86933×10^2	19.19	3.85×10^0	1.40×10^{-1}	8.64×10^0	1.81×10^1	1.34×
	9	-6.39879×10^2	89.37	1.02×10^{-1}	1.33×10^{-1}			
QHNet (Yu et al., 2023)	0	-5.90516×10^2	32.19	1.27×10^0	1.32×10^{-1}	1.49×10^1	$\mathbf{1.60 \times 10^1}$	**1.52**×
	9	-6.18090×10^2	74.67	1.61×10^{-1}	1.16×10^{-1}			
DEQHNet (Wang et al., 2024)	0	-4.86117×10^2	34.02	3.27×10^0	1.28×10^{-1}	5.72×10^0	1.99×10^1	1.23×
	9	-6.43662×10^2	98.65	3.31×10^{-1}	1.41×10^{-1}			
WANet (Li et al., 2025)	0	-5.60708×10^2	65.58	2.90×10^0	7.26×10^{-2}	2.99×10^0	2.44×10^1	1.00×
	9	-6.46805×10^2	99.14	3.57×10^{-2}	3.43×10^{-2}			
SPHNet (Luo et al., 2025a)	0	-5.87576×10^2	41.57	2.39×10^0	1.22×10^{-1}	4.05×10^0	1.97×10^1	1.24×
	9	-6.46336×10^2	98.55	7.57×10^{-2}	1.13×10^{-1}			
QHFlow (Kim et al., 2025a)	0	-5.90628×10^2	59.32	8.65×10^{-1}	7.55×10^{-2}	$\mathbf{1.95 \times 10^0}$	1.97×10^1	1.25×
	9	$\mathbf{-6.47220 \times 10^2}$	**99.70**	$\mathbf{3.07 \times 10^{-2}}$	$\mathbf{9.94 \times 10^{-3}}$			
Ours	0	$\mathbf{-6.47407 \times 10^2}$	**99.94**	$\mathbf{2.22 \times 10^{-3}}$	$\mathbf{1.93 \times 10^{-3}}$	$\mathbf{1.68 \times 10^0}$	$\mathbf{6.34 \times 10^{-1}}$	**38.47**×

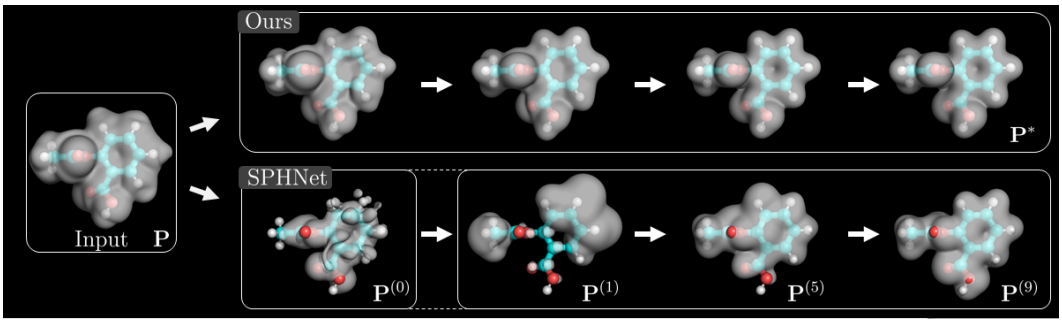

Figure 7: **Qualitative results on aspirin.** We visualize the electronic state dynamics of Aspirin ($C_9H_8O_4$), with our framework shown on top and SPHNet on the bottom.

A.2 Additional Experiments

Additional molecular systems. We extend the evaluation beyond the representative systems in the main text to examine whether the same SCF-free electronic propagation behavior persists across different molecular compositions and system sizes. The additional MD17 systems are Aspirin ($C_9H_8O_4$) and Paracetamol ($C_8H_9NO_2$), which complement Salicylic Acid by introducing different functional groups and local bonding patterns. Aspirin contains aromatic, carboxyl, and ester groups, whereas Paracetamol contains hydroxyl, amide, and aromatic components. These structures create heterogeneous local electronic environments and make the electronic state sensitive to changes in molecular geometry. We also evaluate DHA ($C_{22}H_{32}O_2$) from MD22 as a larger system with an expanded atomic orbital space. Compared with the MD17 molecules, the DHA system involves larger Hamiltonian and overlap matrices and therefore tests the proposed update with a larger number of electronic degrees of freedom.

Comparison results. Additional MD17 results indicate that the behavior observed in the main text extends beyond Salicylic Acid. On both Aspirin and Paracetamol, the proposed method maintains accurate electronic quantities with zero SCF iterations while keeping inference time substantially lower than baselines after SCF refinement. Comparing the zero SCF and SCF refined rows also clarifies the role of the iterative solver. For baselines that normally use SCF, raw electronic predictions are generally less consistent before refinement, especially in occupied orbital error, overlap consistency, and DIIS residual. SCF refinement improves these metrics but the improvement comes from repeated electronic solves at every MD step. In contrast, our method performs the electronic update directly through the propagated occupied subspace and density projector avoiding this iterative correction loop.

Qualitatively, the Aspirin result gives a complementary view of this behavior. Under the same simulation protocol, our method yields a smoother electronic state evolution along the trajectory, while the baseline shows less consistent dynamics. This visual trend is consistent with the quantitative results in Table 5, where the proposed update keeps low electronic errors without SCF correction. With DHA, the comparison further extends to a larger MD22 system. This adds a higher dimensional

Table 6: **Quantitative results on paracetamol in MD17.** We report quantum physical metrics with inference time and SCF iteration counts. Best and second best results are highlighted in **bold** and **blue**, respectively.

Model	SCF	$\mathcal{E}$ $(E_h)\downarrow$	$\mathfrak{S}_c$ (%) $\uparrow$	ϵ_{occ} $(E_h)\downarrow$	$\mathbf{v}\epsilon$ $(E_h)\downarrow$	$\|\epsilon^{\text{DIIS}}\|$ $(E_h)\downarrow$	Inference (s) $\downarrow$	Speed Up $\uparrow$
PhiSNet (Unke et al., 2021)	0	-2.65707×10^2	25.39	3.62×10^0	1.36×10^{-1}	3.89×10^0	1.45×10^1	1.07×
	9	-5.13465×10^2	98.71	7.24×10^{-2}	3.25×10^{-2}			
QHNet (Yu et al., 2023)	0	-4.62861×10^2	22.38	1.11×10^1	1.22×10^{-1}	2.73×10^0	6.62×10^0	2.34×
	9	-5.14014×10^2	99.29	4.49×10^{-2}	8.50×10^{-2}			
DEQHNet (Wang et al., 2024)	0	-3.79502×10^2	34.84	2.78×10^0	1.29×10^{-1}	2.66×10^0	1.54×10^1	1.00×
	9	-5.13509×10^2	99.41	1.91×10^{-1}	1.03×10^{-1}			
WANet (Li et al., 2025)	0	-4.59199×10^2	65.82	2.56×10^0	1.15×10^{-1}	2.30×10^0	1.24×10^1	1.25×
	9	-5.14083×10^2	99.06	5.68×10^{-2}	9.64×10^{-2}			
SPHNet (Luo et al., 2025a)	0	-4.72305×10^2	41.55	2.05×10^0	5.34×10^{-2}	$\mathbf{1.27 \times 10^0}$	1.26×10^1	1.22×
	9	$\mathbf{-5.14394 \times 10^2}$	**99.83**	$\mathbf{1.92 \times 10^{-2}}$	$\mathbf{2.05 \times 10^{-3}}$			
QHFlow (Kim et al., 2025a)	0	-4.76382×10^2	51.90	1.29×10^0	1.12×10^{-1}	4.28×10^0	$\mathbf{6.18 \times 10^0}$	**2.50**×
	9	-5.13204×10^2	97.68	7.43×10^{-2}	1.13×10^{-1}			
Ours	0	$\mathbf{-5.14476 \times 10^2}$	**99.93**	$\mathbf{3.80 \times 10^{-3}}$	$\mathbf{7.89 \times 10^{-4}}$	$\mathbf{4.82 \times 10^{-1}}$	$\mathbf{5.76 \times 10^{-1}}$	**26.87**×

Table 7: **Quantitative results on DHA in MD22.** We report quantum physical metrics with inference time and SCF iteration counts. Best and second best results are highlighted in **bold** and **blue**, respectively.

Model	SCF	$\mathcal{E}$ $(E_h)\downarrow$	$\mathfrak{S}_c$ (%) $\uparrow$	ϵ_{occ} $(E_h)\downarrow$	$\mathbf{v}\epsilon$ $(E_h)\downarrow$	$\|\epsilon^{\text{DIIS}}\|$ $(E_h)\downarrow$	Inference (s) $\downarrow$	Speed Up $\uparrow$
PhiSNet (Unke et al., 2021)	0	-5.47499×10^2	18.30	3.16×10^0	1.71×10^{-1}	3.88×10^1	$\mathbf{7.35 \times 10^1}$	**2.91**×
	13	1.20128×10^3	21.36	8.87×10^0	1.40×10^{-1}			
QHNet (Yu et al., 2023)	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	13	N/A	N/A	N/A	N/A			
DEQHNet (Wang et al., 2024)	0	-6.58837×10^2	30.23	3.32×10^0	1.71×10^{-1}	$\mathbf{3.60 \times 10^0}$	2.14×10^2	1.00×
	13	$\mathbf{-1.00281 \times 10^3}$	**98.71**	1.90×10^{-1}	$\mathbf{9.97 \times 10^{-3}}$			
WANet (Li et al., 2025)	0	-8.13970×10^2	66.75	2.68×10^0	1.64×10^{-1}	1.74×10^1	9.34×10^1	2.30×
	13	-9.56341×10^2	74.92	$\mathbf{1.22 \times 10^{-1}}$	1.60×10^{-1}			
SPHNet (Luo et al., 2025a)	0	-8.58110×10^2	53.36	1.18×10^0	1.62×10^{-1}	2.27×10^1	1.20×10^2	1.79×
	13	-9.38956×10^2	65.05	1.27×10^{-1}	1.57×10^{-1}			
QHFlow (Kim et al., 2025a)	0	-8.32087×10^2	37.18	1.73×10^0	1.68×10^{-1}	3.20×10^1	1.18×10^2	1.81×
	13	-8.02060×10^2	50.36	2.89×10^{-1}	1.64×10^{-1}			
Ours	0	$\mathbf{-1.00563 \times 10^3}$	**99.71**	$\mathbf{4.66 \times 10^{-3}}$	$\mathbf{7.95 \times 10^{-3}}$	$\mathbf{1.77 \times 10^0}$	$\mathbf{2.21 \times 10^0}$	**97.07**×

electronic setting to the comparison. Despite the increased Hamiltonian and overlap matrix sizes, our method remains competitive across the reported electronic metrics under an SCF-free setting. These observations suggest that the observed stability is tied to how the electronic state is propagated rather than to the absence of SCF iterations alone. Taken together, the additional results support the proposed framework as an SCF-free alternative for BOMD propagation beyond the molecular systems emphasized in the main text, covering varying molecular scales and compositions.

B MATHEMATICAL PRELIMINARY

We summarize the geometric setting used to describe occupied electronic states at a fixed nuclear configuration. Throughout this section, let $\mathbf{r} \in \Omega \subset \mathbb{R}^{3N_r}$ be fixed, and let $\mathrm{S} \coloneqq \mathrm{S}(\mathbf{r}) \in \mathbb{C}^{M\times M}$ be the overlap matrix associated with the atom-centered basis. The ambient space $\mathbb{C}^{M\times N}$ is regarded as a real inner product space with

$$\langle X, Y\rangle_{\mathrm{S}} \coloneqq \mathrm{Re}\,\mathrm{Tr}(X^\dagger \mathrm{S} Y), \quad X, Y \in \mathbb{C}^{M\times N}. \tag{13}$$

Physical State as an Occupied Subspace. The electronic state is determined by the occupied N-dimensional subspace, not by a particular choice of orbital basis spanning it. A convenient gauge-invariant description is therefore given by rank-N S-orthogonal projectors.

Definition B.1 (S-orthogonal projector manifold)**.** *Define*

$$\mathcal{P}_{\mathrm{S}}(M, N) \coloneqq \Big\{\mathbf{P} \in \mathbb{C}^{M\times M} : \mathbf{P}^2 = \mathbf{P},\ \mathrm{rank}(\mathbf{P}) = N,\ \mathbf{P}^\dagger \mathrm{S} = \mathrm{S}\mathbf{P}\Big\}. \tag{14}$$

Each $\mathbf{P} \in \mathcal{P}_{\mathrm{S}}(M, N)$ *represents an occupied subspace through its image* $\mathrm{Im}(\mathbf{P})$.

This viewpoint is consistent with the Kohn–Sham formulation in the main text, where physical observables depend on the occupied state through the induced projector.

Frame Representation of the State. For computation, it is convenient to work with orbital frames whose columns span the occupied subspace. These frames form the generalized Stiefel manifold.

Definition B.2 (Generalized Stiefel manifold)**.** *For fixed* $\mathrm{S} \succ 0$*, define*

$$\begin{aligned} \mathrm{St}_{\mathrm{S}}(M, N) &\coloneqq \big\{\mathbf{C} \in \mathbb{C}^{M\times N} : \mathbf{C}^\dagger \mathrm{S}\mathbf{C} = \mathrm{I}_N\big\}, \\ \mathbf{P}(\mathbf{C}) &\coloneqq \mathbf{C}\mathbf{C}^\dagger \mathrm{S}. \end{aligned} \tag{15}$$

Thus each $\mathbf{C} \in \mathrm{St}_{\mathrm{S}}(M, N)$ *is an* S*-orthonormal frame, while* $\mathbf{P}(\mathbf{C})$ *is the associated gauge-invariant projector.*

Different frames may generate the same projector and hence the same physical state. This redundancy is induced by right multiplication with unitary matrices, so the same occupied subspace may be represented by many different orbital frames. More precisely, the right action of $\mathrm{U}_{\mathrm{S}}(N)$ on $\mathrm{St}_{\mathrm{S}}(M, N)$ is given by

$$\mathcal{R}_{\mathbf{U}}(\mathbf{C}) \coloneqq \mathbf{C}\mathbf{U}, \quad \mathbf{U} \in \mathrm{U}_{\mathrm{S}}(N), \tag{16}$$

and it preserves both the Stiefel constraint and the associated projector:

$$(\mathbf{C}\mathbf{U})^\dagger \mathrm{S}(\mathbf{C}\mathbf{U}) = \mathrm{I}_N, \quad \mathbf{P}(\mathbf{C}\mathbf{U}) = \mathbf{P}(\mathbf{C}). \tag{17}$$

Hence two frames represent the same physical state precisely when they differ by a right unitary transformation. This leads to the quotient description of the manifold of occupied states.

Definition B.3 (Generalized Grassmann manifold)**.** *The generalized Grassmann manifold is the quotient*

$$\mathrm{Gr}_{\mathrm{S}}(M, N) \coloneqq \mathrm{St}_{\mathrm{S}}(M, N)/\mathrm{U}_{\mathrm{S}}(N), \tag{18}$$

whose points are the equivalence classes

$$[\mathbf{C}] = \{\mathbf{C}\mathbf{U} : \mathbf{U} \in \mathrm{U}_{\mathrm{S}}(N)\}, \quad [\mathbf{C}] \longmapsto \mathbf{P}(\mathbf{C}).$$

This correspondence identifies $\mathrm{Gr}_{\mathrm{S}}(M, N)$ *with* $\mathcal{P}_{\mathrm{S}}(M, N)$.

Tangent Directions of Frame Space. Although the physical state lives on the quotient, most calculations are carried out on $\mathrm{St}_{\mathrm{S}}(M, N)$. Accordingly, we record the tangent characterization induced by the orthonormality constraint together with the Riemannian structure induced by the ambient metric. A variation Δ belongs to the tangent space at $\mathbf{C} \in \mathrm{St}_{\mathrm{S}}(M, N)$ exactly when it preserves the constraint to first order, which gives

$$T_{\mathbf{C}}\mathrm{St}_{\mathrm{S}}(M, N) = \Big\{\Delta \in \mathbb{C}^{M\times N} : \mathbf{C}^\dagger \mathrm{S}\Delta + \Delta^\dagger \mathrm{S}\mathbf{C} = 0\Big\}. \tag{19}$$

Equivalently, $\mathbf{C}^\dagger \mathrm{S}\Delta$ is skew Hermitian. The resulting Riemannian metric is

$$g^{\mathrm{St}}_{\mathbf{C}}(\Delta_1, \Delta_2) \coloneqq \mathrm{Re}\,\mathrm{Tr}(\Delta_1^\dagger \mathrm{S}\Delta_2), \quad \Delta_1, \Delta_2 \in T_{\mathbf{C}}\mathrm{St}_{\mathrm{S}}(M, N). \tag{20}$$

Gauge and Intrinsic Directions. Let $\pi : \mathrm{St}_{\mathrm{S}}(M, N) \to \mathrm{Gr}_{\mathrm{S}}(M, N), \quad \pi(\mathbf{C}) = [\mathbf{C}]$, be the quotient projection. Tangent vectors along the fiber correspond to infinitesimal changes of gauge only, while the complementary directions describe genuine motion of the occupied subspace. With $\mathfrak{u}_{\mathrm{S}}(N) \coloneqq \{\Omega \in \mathbb{C}^{N\times N} : \Omega^\dagger = -\Omega\}$, the vertical space, the connection form induced by the metric, and the corresponding horizontal space are given by

$$\begin{aligned} \mathcal{V}_{\mathbf{C}} &\coloneqq \ker(d\pi_{\mathbf{C}}) = \{\mathbf{C}\Omega : \Omega \in \mathfrak{u}_{\mathrm{S}}(N)\}, \quad \omega_{\mathbf{C}}(\Delta)\mathbf{C}^\dagger \mathrm{S}\Delta, \\ \mathcal{H}_{\mathbf{C}} &\coloneqq \ker(\omega_{\mathbf{C}}) = \{\Delta \in T_{\mathbf{C}}\mathrm{St}_{\mathrm{S}}(M, N) : \mathbf{C}^\dagger \mathrm{S}\Delta = 0\}. \end{aligned} \tag{21}$$

Here $\mathcal{V}_{\mathbf{C}}$ contains pure gauge directions, whereas $\mathcal{H}_{\mathbf{C}}$ captures intrinsic subspace motion.

Motion Along a Curve. Consider now a smooth curve $t \mapsto \mathbf{C}_t \in \mathrm{St}_{\mathrm{S}}(M, N)$. Its velocity decomposes naturally into vertical and horizontal components.

Definition B.4 (Gauge Generator and Covariant Velocity)**.** *Let $t \mapsto \mathbf{C}_t \in \mathrm{St}_{\mathrm{S}}(M, N)$ be smooth. We define the gauge generator $\Xi(t)$ and the covariant velocity $\nabla_t \mathbf{C}_t$ by*

$$\Xi(t) \coloneqq \mathbf{C}_t^\dagger \mathrm{S}\dot{\mathbf{C}}_t \in \mathfrak{u}_{\mathrm{S}}(N), \quad \nabla_t \mathbf{C}_t \coloneqq \big(I - \mathbf{P}(\mathbf{C}_t)\big)\dot{\mathbf{C}}_t, \quad \dot{\mathbf{C}}_t = \mathbf{C}_t \Xi(t) + \nabla_t \mathbf{C}_t. \tag{22}$$

Here $\mathbf{C}_t\Xi(t) \in \mathcal{V}_{\mathbf{C}_t}$ and $\nabla_t \mathbf{C}_t \in \mathcal{H}_{\mathbf{C}_t}$.

Thus $\Xi(t)$ records the instantaneous rotation of the chosen frame inside the occupied subspace, while $\nabla_t \mathbf{C}_t$ captures the intrinsic transport of the state. Finally, the S-metric induces the standard flat and sharp maps

$$(\cdot)^\flat : X \mapsto \mathrm{S}X, \quad (\cdot)^\sharp : Z \mapsto \mathrm{S}^{-1}Z, \tag{23}$$

which will be used to pass between vector and covector representations under the same metric structure.

Lifted Lie-Group Geometry. To expose the precise algebraic carrier of the prefix-scan, we distinguish local horizontal coordinates, their infinitesimal Lie generators, and the resulting finite bundle actions. Concretely, fixing a reference frame identifies the horizontal space with a linear coordinate chart, while the actual scan is performed over the associated bundle automorphisms, whose composition is exactly associative. To make the scan structure explicit at the combinatorial level, we further represent admissible evaluation schedules by rooted planar full binary trees, equivalently by full binary parenthesizations of the ordered action sequence.

Definition B.5 (S-unitary group, isotropy, and horizontal generator space)**.** *Fix a nuclear configuration $\mathbf{r}$ and write $\mathrm{S} := \mathrm{S}(\mathbf{r})$. Define the S-unitary group and its Lie algebra by*

$$\mathrm{U}_{\mathrm{S}}(M) := \Big\{\mathbf{U} \in \mathrm{GL}(M, \mathbb{C}) \Big| \mathbf{U}^\dagger \mathrm{S}\mathbf{U} = \mathrm{S}\Big\}, \quad \mathfrak{u}_{\mathrm{S}}(M) := \Big\{\Omega \in \mathbb{C}^{M\times M} \Big| \Omega^\dagger \mathrm{S} + \mathrm{S}\Omega = 0\Big\}.$$

Fix a reference representative $\mathbf{C}_0 \in \mathrm{St}_{\mathrm{S}}(M, N)$ and choose $\mathbf{C}_{0,\perp} \in \mathbb{C}^{M\times(M-N)}$ to make $\mathbf{Q}_0 := [\mathbf{C}_0, \mathbf{C}_{0,\perp}]$ where $\mathbf{Q}_0^\dagger \mathrm{S}\mathbf{Q}_0 = I_M$ and $\mathbf{Q}_0^{-1} = \mathbf{Q}_0^\dagger \mathrm{S}$. Then, let us define the isotropy subgroup and isotropy Lie algebra of $\mathbf{P}_0$ defined by

$$\mathcal{K} := \left\{ \mathbf{Q}_0 \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \mathbf{Q}_0^{-1} \middle| U \in \mathrm{U}_{\mathrm{S}}(N), V \in \mathrm{U}(M-N) \right\}, \tag{24}$$

$$\mathfrak{k} := \left\{ \mathbf{Q}_0 \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \mathbf{Q}_0^{-1} \middle| A \in \mathfrak{u}_{\mathrm{S}}(N), B \in \mathfrak{u}_{\mathrm{S}}(M-N) \right\}, \tag{25}$$

$$\mathfrak{m} := \left\{ \mathbf{Q}_0 \begin{bmatrix} 0 & -Z^\dagger \\ Z & 0 \end{bmatrix} \mathbf{Q}_0^{-1} \middle| Z \in \mathbb{C}^{(M-N)\times N} \right\}. \tag{26}$$

Then, the direct sum gives decomposition as $\mathfrak{u}_{\mathrm{S}}(M) = \mathfrak{k}_{\mathbf{C}_0} \oplus \mathfrak{m}_{\mathbf{C}_0}$, where the Cartan-type decomposition is associated with the homogeneous geometry $\mathrm{Gr}_{\mathrm{S}}(M, N) \simeq \mathrm{U}_{\mathrm{S}}(M)/\mathcal{K}$, and followings are true:

$$[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k}, \quad [\mathfrak{k}, \mathfrak{m}] \subset \mathfrak{m}, \quad [\mathfrak{m}, \mathfrak{m}] \subset \mathfrak{k}. \tag{27}$$

B.1 Details on Gauge-Symmetric Horizontal Flow Matching

This section is devoted to explicate details of the objective function in equation 7.

Canonical Projection. We first define the canonical projection from an orbital representative to the corresponding physical occupied subspace is

$$\Pi_{\mathbf{r}} : \mathrm{St_S}(M,N) \longrightarrow \mathcal{P}_{\mathrm{S}}, \qquad \Pi_{\mathbf{r}}(\mathbf{C}) = \mathbf{C}\mathbf{C}^{\dagger}\mathrm{S}.$$

This map is invariant under the right action of the unitary gauge group: $\Pi_{\mathbf{r}}(\mathbf{C}\mathbf{U}) = \Pi_{\mathbf{r}}(\mathbf{C})$. Thus, the same electronic state is represented by the entire equivalence class $[\mathbf{C}] = \{\mathbf{C}\mathbf{U} : \mathbf{U} \in \mathrm{U}(N)\}$. For a horizontal tangent vector $\mathbf{\Delta} \in \mathcal{H}_{\mathbf{C}}$, the differential of the projector map is

$$(\Pi_{\mathbf{r}})_{*\mathbf{C}}\mathbf{\Delta} = \mathbf{\Delta}\mathbf{C}^{\dagger}\mathrm{S} + \mathbf{C}\mathbf{\Delta}^{\dagger}\mathrm{S}. \tag{28}$$

This expression is well-defined because the nuclear configuration is fixed at $\mathbf{r}$ in the definition of the objective, so no variation of $\mathrm{S}(\mathbf{r})$ appears. Moreover, $(\Pi_{\mathbf{r}})_{*\mathbf{C}}$ annihilates vertical directions:

$$(\Pi_{\mathbf{r}})_{*\mathbf{C}}(\mathbf{C}\mathbf{\Omega}) = \mathbf{C}\mathbf{\Omega}\mathbf{C}^{\dagger}\mathrm{S} + \mathbf{C}(\mathbf{C}\mathbf{\Omega})^{\dagger}\mathrm{S} = \mathbf{C}(\mathbf{\Omega} + \mathbf{\Omega}^{\dagger})\mathbf{C}^{\dagger}\mathrm{S} = 0.$$

Hence, the projector dynamics are determined only by the horizontal component of the Stiefel velocity. Restricted to $\mathcal{H}_{\mathbf{C}}$, the map

$$(\Pi_{\mathbf{r}})_{*\mathbf{C}} : \mathcal{H}_{\mathbf{C}} \longrightarrow T_{\Pi_{\mathbf{r}}(\mathbf{C})}\mathcal{P}_{\mathrm{S}}$$

is the horizontal lift identification induced by the quotient geometry. Let $g_{\mathcal{P}}$ denote the Riemannian metric on the projector manifold. We measure tangent errors on the physical projector manifold by pulling this metric back to the horizontal bundle:

$$\left(\Pi_{\mathbf{r}}^{*} g_{\mathcal{P}}\right)_{\mathbf{C}}(\mathbf{\Delta}_1, \mathbf{\Delta}_2) := g_{\mathcal{P},\Pi_{\mathbf{r}}(\mathbf{C})}\left((\Pi_{\mathbf{r}})_{*\mathbf{C}}\mathbf{\Delta}_1, (\Pi_{\mathbf{r}})_{*\mathbf{C}}\mathbf{\Delta}_2\right), \qquad \mathbf{\Delta}_1, \mathbf{\Delta}_2 \in \mathcal{H}_{\mathbf{C}}. \tag{29}$$

On the full Stiefel tangent space, this pull-back metric is degenerate because vertical directions correspond to pure gauge motion. However, on the horizontal subspace $\mathcal{H}_{\mathbf{C}}$ it is non-degenerate and therefore defines a valid way to compare physical tangent dynamics. This is precisely the reason we formulate the objective on horizontal vector fields.

Horizontal Discepancy as Objective Function. The target vector field is the horizontal Riemannian gradient of the Kohn–Sham energy $\mathbf{u}^{\star}(\mathbf{C}) := \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}$. Using the self-consistent Hamiltonian $\mathbf{H}^{\star}$, this field admits the closed-form expression

$$\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E} = (\mathbf{I} - \mathbf{P})\mathrm{S}^{-1}\mathbf{H}^{\star}\mathbf{C}, \qquad \mathbf{P} = \mathbf{C}\mathbf{C}^{\dagger}\mathrm{S}.$$

It is horizontal and gauge-equivariant because one can show that

$$\mathbf{C}^{\dagger}\mathrm{S}(\mathbf{I} - \mathbf{P}) = 0, \quad \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}\mathbf{U}}\mathcal{E} = \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}}\mathcal{E}\,\mathbf{U}, \quad \mathbf{U} \in \mathrm{U}(N)$$

since the projector $\mathbf{P}$ and the energy $\mathcal{E}$ are gauge-invariant. We parameterize the learned dynamics by a neural Riemannian flow

$$\mathbf{v}_{\theta} : \mathrm{St_S}(M,N) \times [0,T] \longrightarrow T\mathrm{St_S}(M,N), \qquad \mathbf{v}_{\theta}(\mathbf{C},t) \in \mathcal{H}_{\mathbf{C}}.$$

The horizontality condition ensures that the learned vector field represents a physical infinitesimal motion of the occupied subspace rather than a gauge rotation of the orbital frame. In practice, this condition can be enforced either by the architecture itself or by applying the horizontal projection

$$\mathrm{Hor}_{\mathbf{C}}(\mathbf{Z}) = (\mathbf{I} - \mathbf{P})\mathbf{Z}, \qquad \mathbf{P} = \mathbf{C}\mathbf{C}^{\dagger}\mathrm{S},$$

to an unconstrained raw output. For a trajectory $\mathbf{C}(t) \in \mathrm{St_S}(M,N)$, define the horizontal vector-field error

$$\mathbf{\Delta}_{\theta}(t) := \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}(t)}\mathcal{E} - \mathbf{v}_{\theta}(\mathbf{C}(t),t).$$

Since both terms lie in $\mathcal{H}_{\mathbf{C}(t)}$, the difference $\mathbf{\Delta}_{\theta}(t)$ is also horizontal. Therefore its norm under the pull-back projector metric is well-defined: To make the pull-back metric explicit, let us abbreviate

$$\mathbf{S} := \mathbf{S}(\mathbf{r}), \qquad \mathbf{C} := \mathbf{C}(t), \qquad \mathbf{\Delta} := \mathbf{\Delta}_{\theta}(t), \qquad \mathbf{P} := \Pi_{\mathbf{r}}(\mathbf{C}) = \mathbf{C}\mathbf{C}^{\dagger}\mathbf{S}.$$

Since both the target field and the learned field are horizontal, we have

$$\mathbf{\Delta} \in \mathcal{H}_{\mathbf{C}}, \qquad \mathbf{C}^{\dagger}\mathbf{S}\mathbf{\Delta} = 0.$$

The differential of the projector map is

$$(\Pi_{\mathbf{r}})_{*\mathbf{C}}\boldsymbol{\Delta} = \boldsymbol{\Delta}\mathbf{C}^{\dagger}\mathbf{S} + \mathbf{C}\boldsymbol{\Delta}^{\dagger}\mathbf{S}.$$

We use the quotient-induced metric on the projector manifold,

$$g_{\mathcal{P},\mathbf{P}}(\dot{\mathbf{P}}_1, \dot{\mathbf{P}}_2) := \frac{1}{2}\,\mathrm{ReTr}\left[\left(\mathbf{S}^{1/2}\dot{\mathbf{P}}_1\mathbf{S}^{-1/2}\right)^{\dagger}\left(\mathbf{S}^{1/2}\dot{\mathbf{P}}_2\mathbf{S}^{-1/2}\right)\right].$$

Therefore,

$$\begin{aligned}
\|\boldsymbol{\Delta}_\theta(t)\|^2_{\Pi_{\mathbf{r}}^* g_{\mathcal{P}}} &= \left(\Pi_{\mathbf{r}}^* g_{\mathcal{P}}\right)_{\mathbf{C}}(\boldsymbol{\Delta}, \boldsymbol{\Delta}) \\
&= g_{\mathcal{P},\Pi_{\mathbf{r}}(\mathbf{C})}\left((\Pi_{\mathbf{r}})_{*\mathbf{C}}\boldsymbol{\Delta}, (\Pi_{\mathbf{r}})_{*\mathbf{C}}\boldsymbol{\Delta}\right) \\
&= \frac{1}{2}\left\|\mathbf{S}^{1/2}\left(\boldsymbol{\Delta}\mathbf{C}^{\dagger}\mathbf{S} + \mathbf{C}\boldsymbol{\Delta}^{\dagger}\mathbf{S}\right)\mathbf{S}^{-1/2}\right\|_F^2 \\
&= \frac{1}{2}\left\|\left(\mathbf{S}^{1/2}\boldsymbol{\Delta}\right)\left(\mathbf{S}^{1/2}\mathbf{C}\right)^{\dagger} + \left(\mathbf{S}^{1/2}\mathbf{C}\right)\left(\mathbf{S}^{1/2}\boldsymbol{\Delta}\right)^{\dagger}\right\|_F^2.
\end{aligned} \tag{30}$$

Now define the orthonormalized variables

$$\widetilde{\mathbf{C}} := \mathbf{S}^{1/2}\mathbf{C}, \qquad \widetilde{\boldsymbol{\Delta}} := \mathbf{S}^{1/2}\boldsymbol{\Delta}.$$

Then, the following properties naturally hold:

$$\widetilde{\mathbf{C}}^{\dagger}\widetilde{\mathbf{C}} = \mathbf{C}^{\dagger}\mathbf{S}\mathbf{C} = \mathbf{I}_N, \qquad \widetilde{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\Delta}} = \mathbf{C}^{\dagger}\mathbf{S}\boldsymbol{\Delta} = 0.$$

Using these identities, the previous expression becomes

$$\begin{aligned}
\|\boldsymbol{\Delta}_\theta(t)\|^2_{\Pi_{\mathbf{r}}^* g_{\mathcal{P}}} &= \frac{1}{2}\left\|\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger} + \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\right\|_F^2 \\
&= \frac{1}{2}\,\mathrm{ReTr}\left[\left(\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger} + \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\right)^{\dagger}\left(\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger} + \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\right)\right] \\
&= \frac{1}{2}\,\mathrm{ReTr}\left[\left(\widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger} + \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger}\right)\left(\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger} + \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\right)\right] \\
&= \frac{1}{2}\,\mathrm{ReTr}\left[\widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger} + \widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger} + \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger} + \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger}\widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\right] \\
&= \frac{1}{2}\,\mathrm{ReTr}\left[\widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger} + \widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger}\widetilde{\mathbf{C}}\widetilde{\boldsymbol{\Delta}}^{\dagger}\right] \\
&= \frac{1}{2}\,\mathrm{ReTr}\left[\widetilde{\boldsymbol{\Delta}}^{\dagger}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger}\widetilde{\mathbf{C}} + \widetilde{\boldsymbol{\Delta}}^{\dagger}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{C}}^{\dagger}\widetilde{\mathbf{C}}\right] \\
&= \mathrm{ReTr}\left[\widetilde{\boldsymbol{\Delta}}^{\dagger}\widetilde{\boldsymbol{\Delta}}\right] \\
&= \mathrm{ReTr}\left[\boldsymbol{\Delta}^{\dagger}\mathbf{S}\boldsymbol{\Delta}\right] \\
&= \|\boldsymbol{\Delta}\|^2_{\mathbf{S}}.
\end{aligned} \tag{31}$$

In the fourth equality, the mixed terms vanish because

$$\widetilde{\boldsymbol{\Delta}}^{\dagger}\widetilde{\mathbf{C}} = \boldsymbol{\Delta}^{\dagger}\mathbf{S}\mathbf{C} = 0, \qquad \widetilde{\mathbf{C}}^{\dagger}\widetilde{\boldsymbol{\Delta}} = \mathbf{C}^{\dagger}\mathbf{S}\boldsymbol{\Delta} = 0,$$

which follows from horizontality. Hence, with the quotient-induced normalization of $g_{\mathcal{P}}$, the pull-back metric on horizontal tangent vectors coincides with the natural $\mathbf{S}$-metric on the horizontal bundle:

$$\|\boldsymbol{\Delta}_\theta(t)\|^2_{\Pi_{\mathbf{r}}^* g_{\mathcal{P}}} = \|\boldsymbol{\Delta}_\theta(t)\|^2_{\mathbf{S}(\mathbf{r})} = \mathrm{ReTr}\left[\boldsymbol{\Delta}_\theta(t)^{\dagger}\mathbf{S}(\mathbf{r})\boldsymbol{\Delta}_\theta(t)\right].$$

Equivalently, the loss compares the induced tangent vectors on the physical projector manifold:

$$\|\boldsymbol{\Delta}_\theta(t)\|^2_{\Pi_{\mathbf{r}}^* g_{\mathcal{P}}} = g_{\mathcal{P},\Pi_{\mathbf{r}}(\mathbf{C}(t))}\left((\Pi_{\mathbf{r}})_{*\mathbf{C}(t)}\boldsymbol{\Delta}_\theta(t), (\Pi_{\mathbf{r}})_{*\mathbf{C}(t)}\boldsymbol{\Delta}_\theta(t)\right).$$

We therefore define the gauge-symmetric horizontal flow matching objective by

$$\mathcal{L}(\theta) = \int_0^T \left(\Pi_{\mathbf{r}}^* g_{\mathcal{P}}\right)_{\mathbf{C}(t)}\left(\boldsymbol{\Delta}_\theta(t), \boldsymbol{\Delta}_\theta(t)\right) dt = \int_0^T \left\|\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}(t)}\mathcal{E} - \mathbf{v}_\theta(\mathbf{C}(t), t)\right\|^2_{\Pi_{\mathbf{r}}^* g_{\mathcal{P}}} dt. \tag{32}$$

This objective is gauge-symmetric in the following sense. If the representative is changed from $\mathbf{C}(t)$ to $\mathbf{C}(t)\mathbf{U}(t)$, and the learned vector field transforms equivariantly as

$$\mathbf{v}_\theta(\mathbf{C}(t)\mathbf{U}(t), t) = \mathbf{v}_\theta(\mathbf{C}(t), t)\mathbf{U}(t),$$

then

$$(\Pi_{\mathbf{r}})_{*\mathbf{CU}}\big(\boldsymbol{\Delta}_\theta \mathbf{U}\big) = (\Pi_{\mathbf{r}})_{*\mathbf{C}}\boldsymbol{\Delta}_\theta.$$

Consequently, the integrand in equation 7 is unchanged by the choice of orbital gauge. The loss therefore penalizes only errors in the physical occupied-subspace dynamics and ignores discrepancies that arise solely from different orbital representatives. In this sense, gauge-symmetric horizontal flow matching provides a quotient-level training criterion for learning SCF-free electronic relaxation dynamics on the Grassmann geometry.

B.2 PROOF

Lemma B.6 (Horizontal Gradient as Projected Ambient Gradient)**.** *On the Stiefel manifold* $\mathrm{St_S}(M,N) = \{\mathbf{C} \in \mathbb{C}^{M\times N} : \mathbf{C}^\dagger \mathrm{S}\mathbf{C} = I_N\}$*, define the gauge-invariant projector* $\mathbf{P} := \mathbf{C}\mathbf{C}^\dagger \mathrm{S}$ *and assume the energy has the form* $\mathcal{E}(\mathbf{r},\mathbf{C}) = \widetilde{\mathcal{E}}(\mathbf{r},\mathbf{P})$ *with Fréchet derivative* $\delta\widetilde{\mathcal{E}} = \langle \mathbf{H}^\star, \delta\mathbf{P}\rangle$ *for some Hermitian* $\mathbf{H}^\star = \mathbf{H}^\star(\mathbf{r},\mathbf{P})$*. Then the horizontal Riemannian gradient of* $\mathcal{E}$ *at* $\mathbf{C}$*, with respect to the* S*-metric* $\langle X, Y\rangle_{\mathrm{S}} := \mathrm{ReTr}(X^\dagger \mathrm{S} Y)$ *and the standard principal* $\mathrm{U_S}(N)$*-gauge, is given by*

$$\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}} \mathcal{E}(\mathbf{r},\mathbf{C}) = (\mathbf{I}-\mathbf{P})\mathrm{S}(\mathbf{r})^{-1}\mathbf{H}^\star\mathbf{C}. \tag{33}$$

Proof. To derive the expression for the horizontal gradient, we consider an arbitrary smooth path $\mathbf{C}_\varepsilon$ passing through $\mathbf{C}$ at $\varepsilon = 0$. Let us explicitly define the associated variations for the gauge-invariant projector and the wavefunction coefficients as follows:

$$\mathbf{P}_\varepsilon := \mathbf{P}(\mathbf{C}_\varepsilon) = \mathbf{C}_\varepsilon \mathbf{C}_\varepsilon^\dagger \mathrm{S}, \quad \delta\mathbf{P} := \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \mathbf{P}_\varepsilon, \quad \delta\mathbf{C} := \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \mathbf{C}_\varepsilon.$$

We first compute the variation of the projector $\delta\mathbf{P}$ in terms of $\delta\mathbf{C}$. By applying the chain rule directly to the definition of $\mathbf{P}_\varepsilon$, we obtain:

$$\begin{aligned}
\delta\mathbf{P} &= \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \big(\mathbf{C}_\varepsilon \mathbf{C}_\varepsilon^\dagger \mathrm{S}\big) \\
&= \left(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\mathbf{C}_\varepsilon^\dagger \mathrm{S} + \mathbf{C}_\varepsilon \frac{\partial \mathbf{C}_\varepsilon^\dagger}{\partial\varepsilon}\mathrm{S} + \mathbf{C}_\varepsilon \mathbf{C}_\varepsilon^\dagger \frac{\partial \mathrm{S}}{\partial\varepsilon}\right)\bigg|_{\varepsilon=0} \quad \text{(Leibniz rule)} \\
&= \left(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\mathbf{C}_\varepsilon^\dagger \mathrm{S} + \mathbf{C}_\varepsilon \frac{\partial \mathbf{C}_\varepsilon^\dagger}{\partial\varepsilon}\mathrm{S}\right)\bigg|_{\varepsilon=0} \quad (\delta\mathrm{S} = 0) \\
&= \left(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\mathbf{C}_\varepsilon^\dagger \mathrm{S} + \mathbf{C}_\varepsilon \Big(\frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\Big)^\dagger \mathrm{S}\right)\bigg|_{\varepsilon=0} \quad \Big(\frac{\partial}{\partial\varepsilon}(\mathbf{C}_\varepsilon^\dagger) = \frac{\partial \mathbf{C}_\varepsilon}{\partial\varepsilon}\Big)^\dagger \\
&= \delta\mathbf{C}\mathbf{C}^\dagger \mathrm{S} + \mathbf{C}(\delta\mathbf{C})^\dagger \mathrm{S}.
\end{aligned}$$

This result relates the variation of the physical state to the variation of the representation. Next, we proceed to compute the variation of the energy functional. Substituting the expression for $\delta\mathbf{P}$ into the definition of the Fréchet derivative $\delta\widetilde{\mathcal{E}} = \langle \mathbf{H}^\star, \delta\mathbf{P}\rangle$, and exploiting the Hermiticity of $\mathbf{H}^\star$ alongside the cyclic property of the trace, we deduce:

$$\begin{aligned}
\delta\mathcal{E}(\mathbf{r},\mathbf{C}) &= \langle \mathbf{H}^\star, \delta\mathbf{P}\rangle \\
&= \mathrm{ReTr}\big(\mathbf{H}^\star(\delta\mathbf{C}\mathbf{C}^\dagger \mathrm{S} + \mathbf{C}(\delta\mathbf{C})^\dagger \mathrm{S})\big) \\
&= \mathrm{ReTr}\big(\mathbf{H}^\star \delta\mathbf{C}\mathbf{C}^\dagger \mathrm{S}\big) + \mathrm{ReTr}\big(\mathbf{H}^\star \mathbf{C}(\delta\mathbf{C})^\dagger \mathrm{S}\big) \\
&= \mathrm{ReTr}\big(\mathbf{C}^\dagger \mathrm{S}\mathbf{H}^\star \delta\mathbf{C}\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \quad \text{(cyclicity of trace)} \\
&= \mathrm{ReTr}\big((\mathbf{C}^\dagger \mathrm{S}\mathbf{H}^\star \delta\mathbf{C})^\dagger\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \quad (\text{since } \mathrm{ReTr}(X) = \mathrm{ReTr}(X^\dagger)) \\
&= \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger (\mathbf{H}^\star)^\dagger \mathrm{S}^\dagger \mathbf{C}\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \\
&= \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathbf{H}^\star \mathrm{S}\mathbf{C}\big) + \mathrm{ReTr}\big((\delta\mathbf{C})^\dagger \mathrm{S}\mathbf{H}^\star\mathbf{C}\big) \quad (\mathbf{H}^\star = \mathbf{H}^{\star\dagger},\ \mathrm{S} = \mathrm{S}^\dagger) \\
&= \mathrm{ReTr}\Big((\delta\mathbf{C})^\dagger \mathrm{S}\Big(\mathrm{S}^{-1}\mathbf{H}^\star \mathrm{S} + \mathbf{H}^\star\Big)\mathbf{C}\Big) \\
&= 2\mathrm{ReTr}\Big((\delta\mathbf{C})^\dagger \mathrm{S}\mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}\Big) \quad \text{(identifying } \mathbf{H}^\star \text{ with the symmetrized representative)} \\
&= 2\langle \mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C},\ \delta\mathbf{C}\rangle_{\mathrm{S}}.
\end{aligned}$$

From this relation, we can identify the unconstrained gradient in the ambient Euclidean space. By definition of the gradient with respect to the S-metric, $\langle X, Y\rangle_{\mathrm{S}} := \mathrm{ReTr}(X^\dagger \mathrm{S} Y)$, it follows that the ambient gradient is characterized by

$$\delta\mathcal{E}(\mathbf{r},\mathbf{C}) = \langle \mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}}\mathcal{E},\ \delta\mathbf{C}\rangle_{\mathrm{S}} = \mathrm{ReTr}\big((\mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}}\mathcal{E})^\dagger \mathrm{S}\delta\mathbf{C}\big) \quad (\forall \delta\mathbf{C} \in \mathbb{C}^{M\times N}),$$

and comparing with $\delta\mathcal{E} = 2\langle \mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}, \delta\mathbf{C}\rangle_{\mathrm{S}}$ gives

$$\mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}}\mathcal{E} = 2\mathrm{S}^{-1}\mathbf{H}^\star\mathbf{C}.$$

To isolate the physical gradient, we employ the projection operator onto the horizontal space. We assert that the operator $\Pi_{\mathrm{Hor}}(\mathbf{Z}) = (\mathbf{I} - \mathbf{P})\mathbf{Z}$ acts as the S-orthogonal projector onto $H_{\mathbf{C}}$ for any $\mathbf{Z} \in \mathbb{C}^{M \times N}$. This is verified by checking the orthogonality condition:

$$\begin{aligned}\mathbf{C}^\dagger \mathrm{S}(\mathbf{I} - \mathbf{P}) &= \mathbf{C}^\dagger \mathrm{S} - \mathbf{C}^\dagger \mathrm{S}\mathbf{C}\mathbf{C}^\dagger \mathrm{S} \\ &= \mathbf{C}^\dagger \mathrm{S} - (\mathbf{C}^\dagger \mathrm{S}\mathbf{C})\mathbf{C}^\dagger \mathrm{S} \\ &= \mathbf{C}^\dagger \mathrm{S} - \mathrm{I}_N \mathbf{C}^\dagger \mathrm{S} \\ &= 0,\end{aligned}$$

where we used the fact that $\mathbf{C}^\dagger \mathrm{S}\mathbf{C} = I_N$ and that $\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger \mathrm{S}$ is self-adjoint with respect to S. This means that

$$\langle \mathbf{P}X, Y \rangle_{\mathrm{S}} = \mathrm{ReTr}(X^\dagger \mathbf{P}^\dagger \mathrm{S} Y) = \mathrm{ReTr}(X^\dagger \mathrm{S}\mathbf{P}Y) = \langle X, \mathbf{P}Y \rangle_{\mathrm{S}}.$$

Finally, the horizontal Riemannian gradient is uniquely characterized by the property that it represents the energy variation restricted to horizontal directions, i.e., $\delta\mathcal{E} = \langle \mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}} \mathcal{E}, \delta\mathbf{C} \rangle_{\mathrm{S}}$ for all $\delta\mathbf{C} \in H_{\mathbf{C}}$. Applying the projection to the ambient gradient derived earlier yields:

$$\begin{aligned}\mathrm{grad}^{\mathrm{Hor}}_{\mathbf{C}} \mathcal{E} &= \Pi_{\mathrm{Hor}}\big(\mathrm{grad}^{\mathrm{Amb}}_{\mathbf{C}} \mathcal{E}\big) = (\mathbf{I} - \mathbf{P})\big(2\mathrm{S}^{-1}\mathbf{H}^\star \mathbf{C}\big) \\ &= 2(\mathbf{I} - \mathbf{P})\mathrm{S}^{-1}\mathbf{H}^\star \mathbf{C}.\end{aligned}$$

Absorbing the conventional factor of 2 into the definition of $\mathbf{H}^\star$, this gives equation 33, as claimed. □

Proposition B.7 (Local geodesic convexity of the KS energy)**.** *Let $\mathbf{P}_\star$ be a stationary occupied projector of the Kohn–Sham energy on the Grassmann manifold. The horizontal Riemannian gradient flow converges exponentially in energy in that local neighborhood:*

$$\mathcal{E}(\mathbf{r}, \mathbf{P}(\tau)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star) \leq e^{-2m_{\mathrm{loc}}\tau} \left(\mathcal{E}(\mathbf{r}, \mathbf{P}(0)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star)\right) \tag{34}$$

Proof. We give the calculation in an $\mathbf{S}(\mathbf{r})$-orthonormalized basis. Equivalently, after the change of variables induced by $\mathbf{S}(\mathbf{r})^{1/2}$, we may assume

$$\mathbf{S}(\mathbf{r}) = \mathbf{I}, \qquad \mathbf{P} = \mathbf{C}\mathbf{C}^\dagger, \qquad \mathbf{P}^2 = \mathbf{P}, \qquad \mathbf{P}^\dagger = \mathbf{P}.$$

The generalized $\mathbf{S}$-Grassmann case follows by the same calculation after conjugation by $\mathbf{S}(\mathbf{r})^{1/2}$.

We write the Kohn–Sham energy as a function of the occupied projector:

$$\mathcal{E}(\mathbf{r}, \mathbf{P}) = \mathrm{Tr}\big(\mathbf{H}(\mathbf{r})\mathbf{P}\big) + \frac{1}{2}\iint \frac{\rho_{\mathbf{P}}(\mathbf{x})\rho_{\mathbf{P}}(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|} d\mathbf{x}d\mathbf{y} + \mathcal{E}_{\mathrm{xc}}[\rho_{\mathbf{P}}]. \tag{35}$$

Assume throughout the proof that the functional is twice Fréchet differentiable in a neighborhood of $\mathbf{P}_\star$. Define the effective Kohn–Sham Hamiltonian

$$\mathbf{F}(\mathbf{P}) \coloneqq \frac{\delta \mathcal{E}}{\delta \mathbf{P}} = \mathbf{H}(\mathbf{r}) + \mathbf{V}_{\mathrm{H}}[\rho_{\mathbf{P}}] + \mathbf{V}_{\mathrm{xc}}[\rho_{\mathbf{P}}]. \tag{36}$$

Thus, for any admissible projector variation $\delta\mathbf{P}$,

$$\delta\mathcal{E}(\mathbf{r}, \mathbf{P})[\delta\mathbf{P}] = \langle \mathbf{F}(\mathbf{P}), \delta\mathbf{P}\rangle = \mathrm{ReTr}\left[\mathbf{F}(\mathbf{P})^\dagger \delta\mathbf{P}\right]. \tag{37}$$

Since $\mathbf{F}(\mathbf{P})$ and $\delta\mathbf{P}$ are Hermitian, this becomes

$$\delta\mathcal{E}(\mathbf{r}, \mathbf{P})[\delta\mathbf{P}] = \mathrm{Tr}\left[\mathbf{F}(\mathbf{P})\delta\mathbf{P}\right].$$

We first record the stationarity condition. A tangent direction on the projector manifold can be written as

$$\delta\mathbf{P} = [\mathbf{A}, \mathbf{P}], \qquad \mathbf{A}^\dagger = -\mathbf{A}.$$

At a stationary point $\mathbf{P}_\star$, we therefore have

$$\begin{aligned} 0 &= \delta\mathcal{E}(\mathbf{r}, \mathbf{P}_\star)\big[[\mathbf{A}, \mathbf{P}_\star]\big] \\ &= \mathrm{Tr}\left[\mathbf{F}(\mathbf{P}_\star)(\mathbf{A}\mathbf{P}_\star - \mathbf{P}_\star\mathbf{A})\right] \\ &= \mathrm{Tr}\left[\mathbf{F}(\mathbf{P}_\star)\mathbf{A}\mathbf{P}_\star\right] - \mathrm{Tr}\left[\mathbf{F}(\mathbf{P}_\star)\mathbf{P}_\star\mathbf{A}\right] \\ &= \mathrm{Tr}\left[\mathbf{P}_\star\mathbf{F}(\mathbf{P}_\star)\mathbf{A}\right] - \mathrm{Tr}\left[\mathbf{F}(\mathbf{P}_\star)\mathbf{P}_\star\mathbf{A}\right] \\ &= \mathrm{Tr}\left[\big(\mathbf{P}_\star\mathbf{F}(\mathbf{P}_\star) - \mathbf{F}(\mathbf{P}_\star)\mathbf{P}_\star\big)\mathbf{A}\right] \\ &= \mathrm{Tr}\left[[\mathbf{P}_\star, \mathbf{F}(\mathbf{P}_\star)]\mathbf{A}\right]. \end{aligned} \tag{38}$$

Since this holds for every skew-Hermitian $\mathbf{A}$, stationarity implies

$$[\mathbf{F}_\star, \mathbf{P}_\star] = \mathbf{0}, \qquad \mathbf{F}_\star \coloneqq \mathbf{F}(\mathbf{P}_\star). \tag{39}$$

Hence $\mathbf{F}_\star$ and $\mathbf{P}_\star$ are simultaneously block diagonal with respect to the occupied–virtual splitting. We choose this eigenbasis so that

$$\mathbf{P}_\star = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{F}_\star = \begin{bmatrix} \boldsymbol{\varepsilon}_{\mathrm{occ}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varepsilon}_{\mathrm{vir}} \end{bmatrix}, \tag{40}$$

where

$$\boldsymbol{\varepsilon}_{\mathrm{occ}} = \mathrm{diag}(\varepsilon_i)_{i\in\mathrm{occ}}, \qquad \boldsymbol{\varepsilon}_{\mathrm{vir}} = \mathrm{diag}(\varepsilon_a)_{a\in\mathrm{vir}}.$$

Now consider an arbitrary occupied–virtual tangent direction. It is generated by a skew-Hermitian matrix of the form

$$\mathbf{K} = \begin{bmatrix} \mathbf{0} & -\boldsymbol{\kappa}^\dagger \\ \boldsymbol{\kappa} & \mathbf{0} \end{bmatrix}, \qquad \boldsymbol{\kappa} \in \mathbb{C}^{(M-N)\times N}. \tag{41}$$

The corresponding Grassmann geodesic through $\mathbf{P}_\star$ is

$$\mathbf{P}(t) = e^{t\mathbf{K}}\mathbf{P}_\star e^{-t\mathbf{K}}. \tag{42}$$

Differentiating this path gives

$$\begin{aligned}\dot{\mathbf{P}}(t) &= \frac{d}{dt}\left(e^{t\mathbf{K}}\mathbf{P}_\star e^{-t\mathbf{K}}\right)\\ &= \mathbf{K}e^{t\mathbf{K}}\mathbf{P}_\star e^{-t\mathbf{K}} - e^{t\mathbf{K}}\mathbf{P}_\star e^{-t\mathbf{K}}\mathbf{K}\\ &= \mathbf{K}\mathbf{P}(t) - \mathbf{P}(t)\mathbf{K}\\ &= [\mathbf{K}, \mathbf{P}(t)].\end{aligned} \tag{43}$$

Therefore,

$$\boldsymbol{\xi} \coloneqq \dot{\mathbf{P}}(0) = [\mathbf{K}, \mathbf{P}_\star]. \tag{44}$$

Using the block forms of $\mathbf{K}$ and $\mathbf{P}_\star$,

$$\mathbf{K}\mathbf{P}_\star = \begin{bmatrix}\mathbf{0} & -\boldsymbol{\kappa}^\dagger\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix}, \tag{45}$$

$$\mathbf{P}_\star\mathbf{K} = \begin{bmatrix}\mathbf{I}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{0} & -\boldsymbol{\kappa}^\dagger\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & -\boldsymbol{\kappa}^\dagger\\ \mathbf{0} & \mathbf{0}\end{bmatrix}. \tag{46}$$

Hence

$$\boldsymbol{\xi} = \dot{\mathbf{P}}(0) = \begin{bmatrix}\mathbf{0} & \boldsymbol{\kappa}^\dagger\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix}. \tag{47}$$

In particular,

$$\|\boldsymbol{\xi}\|_{\mathrm{F}}^2 = \mathrm{Tr}\left[\boldsymbol{\xi}^\dagger\boldsymbol{\xi}\right] = 2\,\mathrm{Tr}\left[\boldsymbol{\kappa}^\dagger\boldsymbol{\kappa}\right] = 2\sum_{i\in\mathrm{occ}}\sum_{a\in\mathrm{vir}}|\kappa_{ai}|^2. \tag{48}$$

We also need the second derivative of the geodesic. Since

$$\dot{\mathbf{P}}(t) = [\mathbf{K}, \mathbf{P}(t)],$$

we have

$$\begin{aligned}\ddot{\mathbf{P}}(t) &= \frac{d}{dt}[\mathbf{K}, \mathbf{P}(t)]\\ &= [\mathbf{K}, \dot{\mathbf{P}}(t)]\\ &= [\mathbf{K}, [\mathbf{K}, \mathbf{P}(t)]].\end{aligned} \tag{49}$$

Thus,

$$\ddot{\mathbf{P}}(0) = [\mathbf{K}, [\mathbf{K}, \mathbf{P}_\star]] = [\mathbf{K}, \boldsymbol{\xi}]. \tag{50}$$

Using the block form of $\boldsymbol{\xi}$,

$$\mathbf{K}\boldsymbol{\xi} = \begin{bmatrix}\mathbf{0} & -\boldsymbol{\kappa}^\dagger\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{0} & \boldsymbol{\kappa}^\dagger\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix} = \begin{bmatrix}-\boldsymbol{\kappa}^\dagger\boldsymbol{\kappa} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\kappa}\boldsymbol{\kappa}^\dagger\end{bmatrix}, \tag{51}$$

$$\boldsymbol{\xi}\mathbf{K} = \begin{bmatrix}\mathbf{0} & \boldsymbol{\kappa}^\dagger\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{0} & -\boldsymbol{\kappa}^\dagger\\ \boldsymbol{\kappa} & \mathbf{0}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\kappa}^\dagger\boldsymbol{\kappa} & \mathbf{0}\\ \mathbf{0} & -\boldsymbol{\kappa}\boldsymbol{\kappa}^\dagger\end{bmatrix}. \tag{52}$$

Consequently,

$$\ddot{\mathbf{P}}(0) = \mathbf{K}\boldsymbol{\xi} - \boldsymbol{\xi}\mathbf{K} = \begin{bmatrix}-2\boldsymbol{\kappa}^\dagger\boldsymbol{\kappa} & \mathbf{0}\\ \mathbf{0} & 2\boldsymbol{\kappa}\boldsymbol{\kappa}^\dagger\end{bmatrix}. \tag{53}$$

We now differentiate the energy along the geodesic. Let

$$E(t) \coloneqq \mathcal{E}(\mathbf{r}, \mathbf{P}(t)).$$

Using the first variation formula,

$$\begin{aligned}\frac{d}{dt}E(t) &= \left\langle \mathbf{F}(\mathbf{P}(t)), \dot{\mathbf{P}}(t)\right\rangle\\ &= \mathrm{Tr}\left[\mathbf{F}(\mathbf{P}(t))\dot{\mathbf{P}}(t)\right].\end{aligned} \tag{54}$$

Differentiating again,

$$\begin{aligned}\frac{d^2}{dt^2}E(t) &= \frac{d}{dt}\operatorname{Tr}\left[\mathbf{F}(\mathbf{P}(t))\dot{\mathbf{P}}(t)\right] \\ &= \operatorname{Tr}\left[\frac{d}{dt}\mathbf{F}(\mathbf{P}(t))\dot{\mathbf{P}}(t)\right] + \operatorname{Tr}\left[\mathbf{F}(\mathbf{P}(t))\ddot{\mathbf{P}}(t)\right] \\ &= \left\langle D\mathbf{F}(\mathbf{P}(t))[\dot{\mathbf{P}}(t)], \dot{\mathbf{P}}(t)\right\rangle + \left\langle \mathbf{F}(\mathbf{P}(t)), \ddot{\mathbf{P}}(t)\right\rangle. \end{aligned} \tag{55}$$

Evaluating at $t = 0$ gives

$$\left.\frac{d^2}{dt^2}\mathcal{E}(\mathbf{r}, \mathbf{P}(t))\right|_{t=0} = \underbrace{\langle D\mathbf{F}(\mathbf{P}_\star)[\boldsymbol{\xi}], \boldsymbol{\xi}\rangle}_{\text{density-response term}} + \underbrace{\left\langle \mathbf{F}_\star, \ddot{\mathbf{P}}(0)\right\rangle}_{\text{one-particle curvature term}}. \tag{56}$$

We first compute the one-particle curvature term. Using equation 40 and equation 53,

$$\begin{aligned}\left\langle \mathbf{F}_\star, \ddot{\mathbf{P}}(0)\right\rangle &= \operatorname{Tr}\left[\begin{bmatrix}\boldsymbol{\varepsilon}_{\text{occ}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varepsilon}_{\text{vir}}\end{bmatrix}\begin{bmatrix}-2\boldsymbol{\kappa}^\dagger\boldsymbol{\kappa} & \mathbf{0} \\ \mathbf{0} & 2\boldsymbol{\kappa}\boldsymbol{\kappa}^\dagger\end{bmatrix}\right] \\ &= -2\operatorname{Tr}\left[\boldsymbol{\varepsilon}_{\text{occ}}\boldsymbol{\kappa}^\dagger\boldsymbol{\kappa}\right] + 2\operatorname{Tr}\left[\boldsymbol{\varepsilon}_{\text{vir}}\boldsymbol{\kappa}\boldsymbol{\kappa}^\dagger\right] \\ &= -2\sum_{i\in\text{occ}}\varepsilon_i\sum_{a\in\text{vir}}|\kappa_{ai}|^2 + 2\sum_{a\in\text{vir}}\varepsilon_a\sum_{i\in\text{occ}}|\kappa_{ai}|^2 \\ &= 2\sum_{i\in\text{occ}}\sum_{a\in\text{vir}}(\varepsilon_a - \varepsilon_i)|\kappa_{ai}|^2. \end{aligned} \tag{57}$$

We next compute the density-response term. Let the density induced by the projector be written as

$$\rho_{\mathbf{P}}(\mathbf{x}) = \sum_{\mu,\nu} P_{\mu\nu}\chi_\mu(\mathbf{x})\overline{\chi_\nu(\mathbf{x})}.$$

The density variation induced by $\boldsymbol{\xi}$ is

$$\delta\rho(\mathbf{x}) \coloneqq \left.\frac{d}{dt}\rho_{\mathbf{P}(t)}(\mathbf{x})\right|_{t=0} = \sum_{\mu,\nu}\xi_{\mu\nu}\chi_\mu(\mathbf{x})\overline{\chi_\nu(\mathbf{x})}. \tag{58}$$

The linearized Hartree potential is

$$\delta v_{\text{H}}(\mathbf{x}) = \int \frac{\delta\rho(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{y}. \tag{59}$$

Similarly, the linearized exchange-correlation potential is

$$\delta v_{\text{xc}}(\mathbf{x}) = \int f_{\text{xc}}(\mathbf{x},\mathbf{y})\delta\rho(\mathbf{y})d\mathbf{y}, \qquad f_{\text{xc}}(\mathbf{x},\mathbf{y}) \coloneqq \frac{\delta^2\mathcal{E}_{\text{xc}}}{\delta\rho(\mathbf{x})\delta\rho(\mathbf{y})}. \tag{60}$$

Therefore,

$$\begin{aligned}\langle D\mathbf{F}(\mathbf{P}_\star)[\boldsymbol{\xi}], \boldsymbol{\xi}\rangle &= \int \delta\rho(\mathbf{x})\delta v_{\text{H}}(\mathbf{x})d\mathbf{x} + \int \delta\rho(\mathbf{x})\delta v_{\text{xc}}(\mathbf{x})d\mathbf{x} \\ &= \int \delta\rho(\mathbf{x})\left[\int \frac{\delta\rho(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{y}\right]d\mathbf{x} \\ &\quad + \int \delta\rho(\mathbf{x})\left[\int f_{\text{xc}}(\mathbf{x},\mathbf{y})\delta\rho(\mathbf{y})d\mathbf{y}\right]d\mathbf{x} \\ &= \iint \delta\rho(\mathbf{x})\frac{1}{\|\mathbf{x}-\mathbf{y}\|}\delta\rho(\mathbf{y})d\mathbf{x}d\mathbf{y} \\ &\quad + \iint \delta\rho(\mathbf{x})f_{\text{xc}}(\mathbf{x},\mathbf{y})\delta\rho(\mathbf{y})d\mathbf{x}d\mathbf{y} \\ &= \iint \delta\rho(\mathbf{x})\left(\frac{1}{\|\mathbf{x}-\mathbf{y}\|} + f_{\text{xc}}(\mathbf{x},\mathbf{y})\right)\delta\rho(\mathbf{y})d\mathbf{x}d\mathbf{y}. \end{aligned} \tag{61}$$

Combining equation 56, equation 57, and equation 61, we obtain

$$\frac{d^2}{dt^2}\mathcal{E}(\mathbf{r},\mathbf{P}(t))\Big|_{t=0} = 2\sum_{i\in\mathrm{occ}}\sum_{a\in\mathrm{vir}}(\varepsilon_a-\varepsilon_i)|\kappa_{ai}|^2 + \iint \delta\rho(\mathbf{x})\left(\frac{1}{\|\mathbf{x}-\mathbf{y}\|}+f_{\mathrm{xc}}(\mathbf{x},\mathbf{y})\right)\delta\rho(\mathbf{y})d\mathbf{x}d\mathbf{y}. \tag{62}$$

This is the desired spectral-gap plus density-response decomposition. We now derive a lower bound from this formula. At this point we impose the local stability assumptions needed for a convexity conclusion. First, assume that the occupied–virtual gap at $\mathbf{P}_\star$ is positive:

$$\Delta_\star \coloneqq \min_{i\in\mathrm{occ},a\in\mathrm{vir}}(\varepsilon_a-\varepsilon_i) > 0. \tag{63}$$

Then the spectral term satisfies

$$\begin{aligned}2\sum_{i\in\mathrm{occ}}\sum_{a\in\mathrm{vir}}(\varepsilon_a-\varepsilon_i)|\kappa_{ai}|^2 &\geq 2\Delta_\star\sum_{i\in\mathrm{occ}}\sum_{a\in\mathrm{vir}}|\kappa_{ai}|^2\\ &= \Delta_\star\left(2\sum_{i\in\mathrm{occ}}\sum_{a\in\mathrm{vir}}|\kappa_{ai}|^2\right)\\ &= \Delta_\star\|\boldsymbol{\xi}\|_{\mathrm{F}}^2.\end{aligned} \tag{64}$$

Second, assume that the exchange-correlation kernel is bounded below:

$$\iint u(\mathbf{x})f_{\mathrm{xc}}(\mathbf{x},\mathbf{y})u(\mathbf{y})d\mathbf{x}d\mathbf{y} \geq -\lambda_{\mathrm{xc}}\|u\|_{L^2}^2, \qquad \forall u\in L^2(\mathbb{R}^3). \tag{65}$$

Third, assume that the density map is locally bounded on tangent directions:

$$\|\delta\rho\|_{L^2} \leq L_\rho\|\boldsymbol{\xi}\|_{\mathrm{F}}. \tag{66}$$

The Hartree contribution is positive semidefinite:

$$\iint \delta\rho(\mathbf{x})\frac{1}{\|\mathbf{x}-\mathbf{y}\|}\delta\rho(\mathbf{y})d\mathbf{x}d\mathbf{y} \geq 0. \tag{67}$$

For instance, this follows from the positivity of the Coulomb kernel. In Fourier variables,

$$\iint \frac{\delta\rho(\mathbf{x})\delta\rho(\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|}d\mathbf{x}d\mathbf{y} = 4\pi\int\frac{|\widehat{\delta\rho}(\mathbf{k})|^2}{\|\mathbf{k}\|^2}d\mathbf{k} \geq 0.$$

The exchange-correlation contribution is bounded as

$$\begin{aligned}\iint \delta\rho(\mathbf{x})f_{\mathrm{xc}}(\mathbf{x},\mathbf{y})\delta\rho(\mathbf{y})d\mathbf{x}d\mathbf{y} &\geq -\lambda_{\mathrm{xc}}\|\delta\rho\|_{L^2}^2\\ &\geq -\lambda_{\mathrm{xc}}L_\rho^2\|\boldsymbol{\xi}\|_{\mathrm{F}}^2.\end{aligned} \tag{68}$$

Combining equation 62, equation 64, equation 67, and equation 68, we get

$$\begin{aligned}\frac{d^2}{dt^2}\mathcal{E}(\mathbf{r},\mathbf{P}(t))\Big|_{t=0} &\geq \Delta_\star\|\boldsymbol{\xi}\|_{\mathrm{F}}^2 - \lambda_{\mathrm{xc}}L_\rho^2\|\boldsymbol{\xi}\|_{\mathrm{F}}^2\\ &= \left(\Delta_\star-\lambda_{\mathrm{xc}}L_\rho^2\right)\|\boldsymbol{\xi}\|_{\mathrm{F}}^2.\end{aligned} \tag{69}$$

Let us define the local stability modulus $m_\star \coloneqq \Delta_\star - \lambda_{\mathrm{xc}}L_\rho^2$. Then, at an optimal point $\mathbf{P}_\star$, one can show that

$$\langle \mathrm{Hess}\,\mathcal{E}(\mathbf{r},\mathbf{P}_\star)[\boldsymbol{\xi}],\boldsymbol{\xi}\rangle \geq m_\star\|\boldsymbol{\xi}\|_{\mathrm{F}}^2. \tag{70}$$

If $m_\star > 0$, then the Hessian is positive definite on the Grassmann tangent space at $\mathbf{P}_\star$. By continuity of the Hessian in the finite-dimensional discretized problem, for every $m_{\mathrm{loc}} \in (0, m_\star)$ there exists a geodesic neighborhood $\mathcal{U}$ of $\mathbf{P}_\star$ such that

$$\langle \mathrm{Hess}\,\mathcal{E}(\mathbf{r},\mathbf{P})[\boldsymbol{\eta}],\boldsymbol{\eta}\rangle \geq m_{\mathrm{loc}}\|\boldsymbol{\eta}\|_{\mathrm{F}}^2, \qquad \forall\mathbf{P}\in\mathcal{U}, \quad \forall\boldsymbol{\eta}\in T_{\mathbf{P}}\mathrm{Gr}. \tag{71}$$

This is local geodesic strong convexity around $\mathbf{P}_\star$. It remains to show the stated exponential decay for the Riemannian gradient flow. Let $\mathbf{P}(\tau)$ be a gradient-flow trajectory contained in $\mathcal{U}$:

$$\dot{\mathbf{P}}(\tau) = -\operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P}(\tau)). \tag{72}$$

Define the local energy gap $\Phi(\tau) \coloneqq \mathcal{E}(\mathbf{r}, \mathbf{P}(\tau)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star)$. Along the flow in equation 72, infinitesimal changes of the energy gap over temporal states can be derived

$$\begin{aligned}\frac{d}{d\tau}\Phi(\tau) &= \left\langle \operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P}(\tau)), \dot{\mathbf{P}}(\tau) \right\rangle \\ &= -\left\|\operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P}(\tau))\right\|^2. \end{aligned} \tag{73}$$

Local geodesic strong convexity implies a local Polyak–Lojasiewicz inequality. Indeed, let

$$\boldsymbol{\eta} = \operatorname{Log}_{\mathbf{P}}(\mathbf{P}_\star) \in T_{\mathbf{P}}\mathrm{Gr}$$

for $\mathbf{P} \in \mathcal{U}$. Strong geodesic convexity gives

$$\mathcal{E}(\mathbf{r}, \mathbf{P}_\star) \geq \mathcal{E}(\mathbf{r}, \mathbf{P}) + \langle \operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P}), \boldsymbol{\eta}\rangle + \frac{m_{\mathrm{loc}}}{2}\|\boldsymbol{\eta}\|^2. \tag{74}$$

Rearranging the calculated terms one can have,

$$\begin{aligned}\mathcal{E}(\mathbf{r}, \mathbf{P}) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star) &\leq -\langle \operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P}), \boldsymbol{\eta}\rangle - \frac{m_{\mathrm{loc}}}{2}\|\boldsymbol{\eta}\|^2 \\ &\leq \|\operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P})\|\,\|\boldsymbol{\eta}\| - \frac{m_{\mathrm{loc}}}{2}\|\boldsymbol{\eta}\|^2 \\ &\leq \frac{1}{2m_{\mathrm{loc}}}\left\|\operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P})\right\|^2. \end{aligned} \tag{75}$$

Therefore, we show that

$$\left\|\operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P})\right\|^2 \geq 2m_{\mathrm{loc}}\left(\mathcal{E}(\mathbf{r}, \mathbf{P}) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star)\right). \tag{76}$$

Combining equation 73 and equation 76,

$$\begin{aligned}\frac{d}{d\tau}\Phi(\tau) &= -\left\|\operatorname{grad}\mathcal{E}(\mathbf{r}, \mathbf{P}(\tau))\right\|^2 \\ &\leq -2m_{\mathrm{loc}}\left(\mathcal{E}(\mathbf{r}, \mathbf{P}(\tau)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star)\right) \\ &= -2m_{\mathrm{loc}}\Phi(\tau). \end{aligned} \tag{77}$$

By Gronwall's inequality,

$$\mathcal{E}(\mathbf{r}, \mathbf{P}(\tau)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star) \leq e^{-2m_{\mathrm{loc}}\tau}\left(\mathcal{E}(\mathbf{r}, \mathbf{P}(0)) - \mathcal{E}(\mathbf{r}, \mathbf{P}_\star)\right), \tag{78}$$

as long as the trajectory remains in the local geodesically strongly convex neighborhood $\mathcal{U}$. This proves the claimed local convexity and the exponential energy decay of the horizontal Riemannian gradient flow. □

Proposition B.8 (Gauge-Symmetric Quantum Bundle Action)**.** *For each localized horizontal vectors $\xi_t \in \mathbb{C}^{(M-N)\times N} \simeq \mathcal{H}_{\mathbf{C}_t}$, let us define the following operation on horizontal space $\mathcal{H}_{\mathbf{C}_t}$:*

$$\Omega[\xi_t] := \mathbf{C}_{t,\perp}\xi_t\mathbf{C}_t^\dagger \mathrm{S} - \mathbf{C}_t\xi_t^\dagger\mathbf{C}_{t,\perp}^\dagger \mathrm{S} = \mathbf{Q}_t \begin{bmatrix} 0 & -\xi_t^\dagger \\ \xi_t & 0 \end{bmatrix} \mathbf{Q}_t^{-1}, \quad \begin{cases} \Omega[\xi_t] \in \mathfrak{m} \subset \mathfrak{u}_{\mathrm{S}}(M), \\ \Omega[\xi_t]\mathbf{C}_t = \mathbf{C}_{t,\perp}\xi_t \in \mathcal{H}_{\mathbf{C}_t}. \end{cases}$$

Then, the bundle automorphism $\mathcal{A}_\xi$ called a ***quantum bundle action*** *preserves both horizontality and gauge symmetry:*

$$\mathcal{A}_{\xi_t}(\mathbf{C},\mathbf{v}) := \exp(\Omega[\xi_t])(\mathbf{C},\mathbf{v}), \ \begin{cases} \mathbf{C}^\dagger\mathrm{S}\mathbf{v} = 0 \Longleftrightarrow \big(\exp(\Omega[\xi_t])\mathbf{C}\big)^\dagger \mathrm{S}\big(\exp(\Omega[\xi_t])\mathbf{v}\big) = 0, \ (\mathbf{A}), \\ \mathcal{A}_{\xi_t}\big((\mathbf{C},\mathbf{v})\cdot\mathbf{U}\big) = \mathcal{A}_{\xi_t}(\mathbf{C},\mathbf{v})\cdot\mathbf{U}, \quad \forall\, \mathbf{U} \in \mathrm{U}_{\mathrm{S}}(N), \ (\mathbf{B}). \end{cases}$$

Furthermore, this action ensures scan-compatible composition law carried by the induced actions:

$$\mathcal{A}_{\xi_{T-1}} \circ \cdots \circ \Big((\mathcal{A}_{\xi_{i+2}} \circ \mathcal{A}_{\xi_{i+1}}) \circ \mathcal{A}_{\xi_i}\Big) \circ \cdots \circ \mathcal{A}_{\xi_0} = \mathcal{A}_{\xi_{T-1}} \circ \cdots \circ \Big(\mathcal{A}_{\xi_{i+2}} \circ (\mathcal{A}_{\xi_{i+1}} \circ \mathcal{A}_{\xi_i})\Big) \circ \cdots \circ \mathcal{A}_{\xi_0}.$$

Proof. Fix a time index t and suppress the subscript t for readability. Write

$$\mathbf{C} := \mathbf{C}_t, \qquad \mathbf{C}_\perp := \mathbf{C}_{t,\perp}, \qquad \xi := \xi_t, \qquad \mathbf{Q} := \mathbf{Q}_t = [\mathbf{C}, \mathbf{C}_\perp].$$

By construction, $\mathbf{C}$ and $\mathbf{C}_\perp$ form an S-orthonormal frame:

$$\mathbf{C}^\dagger\mathrm{S}\mathbf{C} = \mathbf{I}_N, \qquad \mathbf{C}_\perp^\dagger\mathrm{S}\mathbf{C}_\perp = \mathbf{I}_{M-N}, \qquad \mathbf{C}^\dagger\mathrm{S}\mathbf{C}_\perp = \mathbf{0}, \qquad \mathbf{C}_\perp^\dagger\mathrm{S}\mathbf{C} = \mathbf{0}. \tag{79}$$

Equivalently,

$$\mathbf{Q}^\dagger\mathrm{S}\mathbf{Q} = \begin{bmatrix} \mathbf{C}^\dagger\mathrm{S}\mathbf{C} & \mathbf{C}^\dagger\mathrm{S}\mathbf{C}_\perp \\ \mathbf{C}_\perp^\dagger\mathrm{S}\mathbf{C} & \mathbf{C}_\perp^\dagger\mathrm{S}\mathbf{C}_\perp \end{bmatrix} = \begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-N} \end{bmatrix} = \mathbf{I}_M. \tag{80}$$

Hence

$$\mathbf{Q}^{-1} = \mathbf{Q}^\dagger\mathrm{S} = \begin{bmatrix} \mathbf{C}^\dagger\mathrm{S} \\ \mathbf{C}_\perp^\dagger\mathrm{S} \end{bmatrix}. \tag{81}$$

We first verify the block representation of the infinitesimal generator. Define

$$\mathbf{B}[\xi] := \begin{bmatrix} \mathbf{0} & -\xi^\dagger \\ \xi & \mathbf{0} \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{Q}\mathbf{B}[\xi] &= [\mathbf{C}, \mathbf{C}_\perp] \begin{bmatrix} \mathbf{0} & -\xi^\dagger \\ \xi & \mathbf{0} \end{bmatrix} \\ &= \Big[\mathbf{C}\mathbf{0} + \mathbf{C}_\perp\xi, \ -\mathbf{C}\xi^\dagger + \mathbf{C}_\perp\mathbf{0}\Big] \\ &= \Big[\mathbf{C}_\perp\xi, \ -\mathbf{C}\xi^\dagger\Big]. \end{aligned} \tag{82}$$

Therefore, using equation 81,

$$\begin{aligned} \mathbf{Q}\mathbf{B}[\xi]\mathbf{Q}^{-1} &= \Big[\mathbf{C}_\perp\xi, \ -\mathbf{C}\xi^\dagger\Big] \begin{bmatrix} \mathbf{C}^\dagger\mathrm{S} \\ \mathbf{C}_\perp^\dagger\mathrm{S} \end{bmatrix} \\ &= \mathbf{C}_\perp\xi\mathbf{C}^\dagger\mathrm{S} - \mathbf{C}\xi^\dagger\mathbf{C}_\perp^\dagger\mathrm{S} \\ &= \Omega[\xi]. \end{aligned} \tag{83}$$

This proves

$$\Omega[\xi] = \mathbf{Q} \begin{bmatrix} \mathbf{0} & -\xi^\dagger \\ \xi & \mathbf{0} \end{bmatrix} \mathbf{Q}^{-1}.$$

Since $\mathbf{B}[\xi]$ is purely off-diagonal with respect to the occupied–virtual splitting, this is precisely the horizontal Lie component. Hence

$$\Omega[\xi] \in \mathfrak{m}.$$

Next, we show that $\Omega[\xi]$ belongs to the S-skew-Hermitian Lie algebra $\mathfrak{u}_{\mathrm{S}}(M)$. Since

$$\mathbf{B}[\xi]^\dagger = \begin{bmatrix} \mathbf{0} & \xi^\dagger \\ -\xi & \mathbf{0} \end{bmatrix} = -\begin{bmatrix} \mathbf{0} & -\xi^\dagger \\ \xi & \mathbf{0} \end{bmatrix} = -\mathbf{B}[\xi],$$

and since $\Omega[\xi] = \mathbf{Q}\mathbf{B}[\xi]\mathbf{Q}^\dagger \mathrm{S}$, we have

$$\begin{aligned}\Omega[\xi]^\dagger \mathrm{S} &= \left(\mathbf{Q}\mathbf{B}[\xi]\mathbf{Q}^\dagger \mathrm{S}\right)^\dagger \mathrm{S} \\ &= \mathrm{S}^\dagger \mathbf{Q}\mathbf{B}[\xi]^\dagger \mathbf{Q}^\dagger \mathrm{S} \\ &= \mathrm{S}\mathbf{Q}\big(-\mathbf{B}[\xi]\big)\mathbf{Q}^\dagger \mathrm{S} \\ &= -\mathrm{S}\mathbf{Q}\mathbf{B}[\xi]\mathbf{Q}^\dagger \mathrm{S} \\ &= -\mathrm{S}\Omega[\xi]. \end{aligned} \tag{84}$$

Therefore, one obtains $\Omega[\xi]^\dagger \mathrm{S} + \mathrm{S}\Omega[\xi] = \mathbf{0}$, which proves $\Omega[\xi] \in \mathfrak{u}_{\mathrm{S}}(M)$. Next, we aim to verify that $\Omega[\xi]$ realizes the horizontal vector represented by ξ. Using the definition of $\Omega[\xi]$,

$$\begin{aligned}\Omega[\xi]\mathbf{C} &= \left(\mathbf{C}_\perp \xi \mathbf{C}^\dagger \mathrm{S} - \mathbf{C}\xi^\dagger \mathbf{C}_\perp^\dagger \mathrm{S}\right)\mathbf{C} \\ &= \mathbf{C}_\perp \xi \left(\mathbf{C}^\dagger \mathrm{S}\mathbf{C}\right) - \mathbf{C}\xi^\dagger \left(\mathbf{C}_\perp^\dagger \mathrm{S}\mathbf{C}\right) \\ &= \mathbf{C}_\perp \xi \mathbf{I}_N - \mathbf{C}\xi^\dagger \mathbf{0} \\ &= \mathbf{C}_\perp \xi. \end{aligned} \tag{85}$$

Then, let us consider the following identity:

$$\begin{aligned}\mathbf{C}^\dagger \mathrm{S}(\Omega[\xi]\mathbf{C}) &= \mathbf{C}^\dagger \mathrm{S}\mathbf{C}_\perp \xi \\ &= \mathbf{0} \cdot \xi \\ &= \mathbf{0}. \end{aligned} \tag{86}$$

Applying the result, one can show that

$$\Omega[\xi]\mathbf{C} = \mathbf{C}_\perp \xi \in \mathcal{H}_{\mathbf{C}}. \tag{87}$$

We also record that the construction is independent of the auxiliary complement once the same horizontal vector is represented. Indeed, if another complementary frame is chosen as

$$\mathbf{C}'_\perp = \mathbf{C}_\perp \mathbf{V}, \qquad \mathbf{V}^\dagger \mathbf{V} = \mathbf{I}_{M-N},$$

then the same horizontal vector is represented by

$$\xi' = \mathbf{V}^\dagger \xi, \qquad \mathbf{C}'_\perp \xi' = \mathbf{C}_\perp \mathbf{V}\mathbf{V}^\dagger \xi = \mathbf{C}_\perp \xi.$$

The corresponding generator is

$$\begin{aligned}\Omega'[\xi'] &= \mathbf{C}'_\perp \xi' \mathbf{C}^\dagger \mathrm{S} - \mathbf{C}(\xi')^\dagger (\mathbf{C}'_\perp)^\dagger \mathrm{S} \\ &= \mathbf{C}_\perp \mathbf{V}\mathbf{V}^\dagger \xi \mathbf{C}^\dagger \mathrm{S} - \mathbf{C}\left(\mathbf{V}^\dagger \xi\right)^\dagger \left(\mathbf{C}_\perp \mathbf{V}\right)^\dagger \mathrm{S} \\ &= \mathbf{C}_\perp \xi \mathbf{C}^\dagger \mathrm{S} - \mathbf{C}\xi^\dagger \mathbf{V}\mathbf{V}^\dagger \mathbf{C}_\perp^\dagger \mathrm{S} \\ &= \mathbf{C}_\perp \xi \mathbf{C}^\dagger \mathrm{S} - \mathbf{C}\xi^\dagger \mathbf{C}_\perp^\dagger \mathrm{S} \\ &= \Omega[\xi]. \end{aligned} \tag{88}$$

Thus the generator depends on the horizontal direction, not on the particular choice of complementary basis. Now let us define the finite transformation $\mathbf{U}_\xi := \exp(\Omega[\xi])$. While $\Omega[\xi] \in \mathfrak{u}_{\mathrm{S}}(M)$, the exponential lies in the S-unitary group. To see this explicitly, consider

$$\mathbf{G}(s) := \exp(s\Omega[\xi])^\dagger \mathrm{S} \exp(s\Omega[\xi]).$$

By differentiating $\mathbf{G}$, one obtains

$$\begin{aligned}\frac{d}{ds}\mathbf{G}(s) &= \frac{d}{ds}\left[\exp(s\Omega[\xi])^\dagger \mathrm{S} \exp(s\Omega[\xi])\right] \\ &= \exp(s\Omega[\xi])^\dagger \Omega[\xi]^\dagger \mathrm{S} \exp(s\Omega[\xi]) + \exp(s\Omega[\xi])^\dagger \mathrm{S}\Omega[\xi] \exp(s\Omega[\xi]) \\ &= \exp(s\Omega[\xi])^\dagger \left(\Omega[\xi]^\dagger \mathrm{S} + \mathrm{S}\Omega[\xi]\right) \exp(s\Omega[\xi]) \\ &= \exp(s\Omega[\xi])^\dagger \mathbf{0} \exp(s\Omega[\xi]) \\ &= \mathbf{0}, \end{aligned} \tag{89}$$

where we used vanishing identity above. Therefore $\mathbf{G}(s)$ is constant in s. Next, we observe that

$$\mathbf{G}(0) = \exp(\mathbf{0})^\dagger \mathrm{S} \exp(\mathbf{0}) = \mathrm{S}.$$

From the above fact and the definition, we obtain

$$\mathbf{G}(1) = \exp(\Omega[\xi])^\dagger \mathrm{S} \exp(\Omega[\xi]) = \mathrm{S}.$$

Equivalently,

$$\mathbf{U}_\xi^\dagger \mathrm{S} \mathbf{U}_\xi = \mathrm{S}. \tag{90}$$

Hence

$$\mathbf{U}_\xi \in \mathrm{U}_\mathrm{S}(M).$$

The quantum bundle action is therefore

$$\mathcal{A}_\xi(\mathbf{C}, \mathbf{v}) = (\mathbf{U}_\xi \mathbf{C}, \mathbf{U}_\xi \mathbf{v}).$$

We first prove that it preserves the Stiefel constraint. If $\mathbf{C}^\dagger \mathrm{S} \mathbf{C} = \mathbf{I}_N$, then

$$\begin{aligned}(\mathbf{U}_\xi \mathbf{C})^\dagger \mathrm{S} (\mathbf{U}_\xi \mathbf{C}) &= \mathbf{C}^\dagger \mathbf{U}_\xi^\dagger \mathrm{S} \mathbf{U}_\xi \mathbf{C} \\ &= \mathbf{C}^\dagger \mathrm{S} \mathbf{C} \\ &= \mathbf{I}_N. \end{aligned} \tag{91}$$

Thus $\mathbf{U}_\xi \mathbf{C} \in \mathrm{St}_\mathrm{S}(M, N)$.

We next prove preservation of tangent and horizontal conditions. For any pair $(\mathbf{C}, \mathbf{v})$,

$$\begin{aligned}(\mathbf{U}_\xi \mathbf{C})^\dagger \mathrm{S} (\mathbf{U}_\xi \mathbf{v}) &= \mathbf{C}^\dagger \mathbf{U}_\xi^\dagger \mathrm{S} \mathbf{U}_\xi \mathbf{v} \\ &= \mathbf{C}^\dagger \mathrm{S} \mathbf{v}. \end{aligned} \tag{92}$$

Similarly,

$$\begin{aligned}(\mathbf{U}_\xi \mathbf{v})^\dagger \mathrm{S} (\mathbf{U}_\xi \mathbf{C}) &= \mathbf{v}^\dagger \mathbf{U}_\xi^\dagger \mathrm{S} \mathbf{U}_\xi \mathbf{C} \\ &= \mathbf{v}^\dagger \mathrm{S} \mathbf{C}. \end{aligned} \tag{93}$$

Consequently,

$$\begin{aligned}&(\mathbf{U}_\xi \mathbf{C})^\dagger \mathrm{S} (\mathbf{U}_\xi \mathbf{v}) + (\mathbf{U}_\xi \mathbf{v})^\dagger \mathrm{S} (\mathbf{U}_\xi \mathbf{C}) \\ &\quad = \mathbf{C}^\dagger \mathrm{S} \mathbf{v} + \mathbf{v}^\dagger \mathrm{S} \mathbf{C}. \end{aligned} \tag{94}$$

Therefore, if $\mathbf{v} \in T_\mathbf{C} \mathrm{St}_\mathrm{S}(M, N)$, then $\mathbf{U}_\xi \mathbf{v} \in T_{\mathbf{U}_\xi \mathbf{C}} \mathrm{St}_\mathrm{S}(M, N)$. In particular, the horizontal condition is preserved exactly. From equation 92,

$$(\mathbf{U}_\xi \mathbf{C})^\dagger \mathrm{S} (\mathbf{U}_\xi \mathbf{v}) = \mathbf{C}^\dagger \mathrm{S} \mathbf{v}. \tag{95}$$

Hence

$$\mathbf{C}^\dagger \mathrm{S} \mathbf{v} = \mathbf{0} \quad \Longleftrightarrow \quad (\mathbf{U}_\xi \mathbf{C})^\dagger \mathrm{S} (\mathbf{U}_\xi \mathbf{v}) = \mathbf{0}. \tag{96}$$

This proves property $(\mathbf{A})$.

We now verify gauge symmetry. Let $\mathbf{R}$ be an admissible right gauge transformation, i.e., $\mathbf{R}^\dagger \mathbf{R} = \mathbf{I}_N$. The right action on the horizontal bundle is

$$(\mathbf{C}, \mathbf{v}) \cdot \mathbf{R} = (\mathbf{C}\mathbf{R}, \mathbf{v}\mathbf{R}).$$

First, this right action preserves the bundle constraints:

$$\begin{aligned}(\mathbf{C}\mathbf{R})^\dagger \mathrm{S} (\mathbf{C}\mathbf{R}) &= \mathbf{R}^\dagger \mathbf{C}^\dagger \mathrm{S} \mathbf{C} \mathbf{R} \\ &= \mathbf{R}^\dagger \mathbf{I}_N \mathbf{R} \\ &= \mathbf{I}_N, \end{aligned} \tag{97}$$

and if $\mathbf{C}^\dagger \mathrm{S} \mathbf{v} = \mathbf{0}$, the result can be improved as

$$\begin{aligned}(\mathbf{C}\mathbf{R})^\dagger \mathrm{S} (\mathbf{v}\mathbf{R}) &= \mathbf{R}^\dagger \mathbf{C}^\dagger \mathrm{S} \mathbf{v} \mathbf{R} \\ &= \mathbf{R}^\dagger \mathbf{0} \mathbf{R} \\ &= \mathbf{0}. \end{aligned} \tag{98}$$

Next, we compute the action after applying the gauge transformation:

$$\begin{aligned}\mathcal{A}_\xi\big((\mathbf{C},\mathbf{v})\cdot\mathbf{R}\big) &= \mathcal{A}_\xi(\mathbf{C}\mathbf{R},\mathbf{v}\mathbf{R})\\ &= (\mathbf{U}_\xi\mathbf{C}\mathbf{R},\mathbf{U}_\xi\mathbf{v}\mathbf{R})\\ &= (\mathbf{U}_\xi\mathbf{C},\mathbf{U}_\xi\mathbf{v})\cdot\mathbf{R}\\ &= \mathcal{A}_\xi(\mathbf{C},\mathbf{v})\cdot\mathbf{R}. \end{aligned} \tag{99}$$

This proves the second property $(\mathbf{B})$. It remains to prove the scan-compatible composition law. For any $\mathbf{G}\in \mathrm{U_S}(M)$, define the induced bundle action $\mathcal{A}_{\mathbf{G}}(\mathbf{C},\mathbf{v}) := (\mathbf{G}\mathbf{C},\mathbf{G}\mathbf{v})$. Then $\mathcal{A}_\xi = \mathcal{A}_{\mathbf{U}_\xi}$ with $\mathbf{U}_\xi = \exp(\Omega[\xi])$. Let $\mathbf{G}_1,\mathbf{G}_2\in\mathrm{U_S}(M)$. Their product is again S-unitary because

$$\begin{aligned}(\mathbf{G}_2\mathbf{G}_1)^\dagger \mathrm{S}(\mathbf{G}_2\mathbf{G}_1) &= \mathbf{G}_1^\dagger\mathbf{G}_2^\dagger\mathrm{S}\mathbf{G}_2\mathbf{G}_1\\ &= \mathbf{G}_1^\dagger\mathrm{S}\mathbf{G}_1\\ &= \mathrm{S}. \end{aligned} \tag{100}$$

Thus the finite actions are closed under composition. Moreover,

$$\begin{aligned}(\mathcal{A}_{\mathbf{G}_2}\circ\mathcal{A}_{\mathbf{G}_1})(\mathbf{C},\mathbf{v}) &= \mathcal{A}_{\mathbf{G}_2}(\mathbf{G}_1\mathbf{C},\mathbf{G}_1\mathbf{v})\\ &= (\mathbf{G}_2\mathbf{G}_1\mathbf{C},\mathbf{G}_2\mathbf{G}_1\mathbf{v})\\ &= \mathcal{A}_{\mathbf{G}_2\mathbf{G}_1}(\mathbf{C},\mathbf{v}). \end{aligned} \tag{101}$$

Hence composition of bundle actions is exactly multiplication of their corresponding finite S-unitary matrices, in the same order as function composition.

Now take

$$\mathbf{U}_i := \exp(\Omega[\xi_i]),\qquad \mathcal{A}_{\xi_i} = \mathcal{A}_{\mathbf{U}_i}.$$

For three consecutive actions, the left-parenthesized composition gives

$$\begin{aligned}\Big((\mathcal{A}_{\xi_{i+2}}\circ\mathcal{A}_{\xi_{i+1}})\circ\mathcal{A}_{\xi_i}\Big)(\mathbf{C},\mathbf{v}) &= (\mathcal{A}_{\mathbf{U}_{i+2}\mathbf{U}_{i+1}}\circ\mathcal{A}_{\mathbf{U}_i})(\mathbf{C},\mathbf{v})\\ &= \mathcal{A}_{(\mathbf{U}_{i+2}\mathbf{U}_{i+1})\mathbf{U}_i}(\mathbf{C},\mathbf{v})\\ &= ((\mathbf{U}_{i+2}\mathbf{U}_{i+1})\mathbf{U}_i\mathbf{C},(\mathbf{U}_{i+2}\mathbf{U}_{i+1})\mathbf{U}_i\mathbf{v})\,. \end{aligned} \tag{102}$$

The right-parenthesized composition gives

$$\begin{aligned}\Big(\mathcal{A}_{\xi_{i+2}}\circ(\mathcal{A}_{\xi_{i+1}}\circ\mathcal{A}_{\xi_i})\Big)(\mathbf{C},\mathbf{v}) &= (\mathcal{A}_{\mathbf{U}_{i+2}}\circ\mathcal{A}_{\mathbf{U}_{i+1}\mathbf{U}_i})(\mathbf{C},\mathbf{v})\\ &= \mathcal{A}_{\mathbf{U}_{i+2}(\mathbf{U}_{i+1}\mathbf{U}_i)}(\mathbf{C},\mathbf{v})\\ &= (\mathbf{U}_{i+2}(\mathbf{U}_{i+1}\mathbf{U}_i)\mathbf{C},\mathbf{U}_{i+2}(\mathbf{U}_{i+1}\mathbf{U}_i)\mathbf{v})\,. \end{aligned} \tag{103}$$

Since matrix multiplication is associative,

$$(\mathbf{U}_{i+2}\mathbf{U}_{i+1})\mathbf{U}_i = \mathbf{U}_{i+2}(\mathbf{U}_{i+1}\mathbf{U}_i),$$

the expressions in equation 102 and equation 103 are identical for every $(\mathbf{C},\mathbf{v})$. Therefore,

$$(\mathcal{A}_{\xi_{i+2}}\circ\mathcal{A}_{\xi_{i+1}})\circ\mathcal{A}_{\xi_i} = \mathcal{A}_{\xi_{i+2}}\circ(\mathcal{A}_{\xi_{i+1}}\circ\mathcal{A}_{\xi_i}). \tag{104}$$

The same argument extends to any finite sequence. Indeed, for any ordered block $\xi_0,\dots,\xi_k$,

$$(\mathcal{A}_{\xi_k}\circ\cdots\circ\mathcal{A}_{\xi_0})(\mathbf{C},\mathbf{v}) = (\mathbf{U}_k\cdots\mathbf{U}_0\mathbf{C},\mathbf{U}_k\cdots\mathbf{U}_0\mathbf{v})\,, \tag{105}$$

and the product $\mathbf{U}_k\cdots\mathbf{U}_0$ is independent of the binary parenthesization. Therefore any two scan trees that preserve the left-to-right order of the local quantum bundle actions produce the same accumulated action. In particular,

$$\begin{aligned}&\mathcal{A}_{\xi_{T-1}}\circ\cdots\circ\Big((\mathcal{A}_{\xi_{i+2}}\circ\mathcal{A}_{\xi_{i+1}})\circ\mathcal{A}_{\xi_i}\Big)\circ\cdots\circ\mathcal{A}_{\xi_0}\\ &\quad= \mathcal{A}_{\xi_{T-1}}\circ\cdots\circ\Big(\mathcal{A}_{\xi_{i+2}}\circ(\mathcal{A}_{\xi_{i+1}}\circ\mathcal{A}_{\xi_i})\Big)\circ\cdots\circ\mathcal{A}_{\xi_0}. \end{aligned} \tag{106}$$

This proves the scan-compatible composition law.

Finally, note that although the local coordinates ξ_i live in the horizontal chart, the scan is not performed by adding these coordinates. It is performed by composing the finite bundle actions $\mathcal{A}_{\xi_i}$, equivalently by multiplying the finite S-unitary matrices $\mathbf{U}_i = \exp(\Omega[\xi_i])$. The latter operation is closed and associative, which is exactly the algebraic structure required for prefix-scan propagation. □

B.3 Model Architecture

Neural Network Overview. We use a lightweight sequence-to-sequence network that predicts, at each step, a horizontal Lie generator parameterization together with a bounded step size. Given a context occupied subspace and a normalized time grid, the model performs feature aggregation, context embedding, and coefficient/step-size prediction. As illustrated in the middle box of Figure 2, the remainder of this section follows the block structure of feature aggregation, context embedding, coefficient network and step-size prediction.

Feature Aggregation. Let M denote the basis size and N the number of occupied orbitals, and let the network receive a normalized time grid $\mathbf{t} = (t_0, \dots, t_{T-1})^\top \in [0,1]^T$ together with either a single context occupied subspace $\mathbf{U}_{\mathrm{ctx}} \in \mathbb{R}^{M\times N}$ or a time-indexed context sequence $\mathbf{U}_{\mathrm{ctx}} \in \mathbb{R}^{T\times M\times N}$, where the former is broadcast across time. For each time step k, we form

$$\mathbf{P}_k = \mathbf{U}_k \mathbf{U}_k^\top \in \mathbb{R}^{M\times M}. \tag{107}$$

From $\mathbf{P}_k$, we extract local channels from the diagonal and b upper and lower near-diagonal bands, augment them with four channels based on off-diagonal magnitudes and first/second finite differences of the diagonal band, and compute global statistics including diagonal mean and standard deviation, off-diagonal magnitude summaries, and an orthogonality residual. Adaptive average pooling of the local features, concatenated with the global scalars, gives the aggregated feature vector $\mathbf{z}_k$.

Context Embedding. Time is encoded with Fourier features, and the aggregated feature vector and time embedding are projected to the model width and summed:

$$\mathbf{c}_k^{(0)} = \mathbf{W}_{\mathrm{ctx}}\mathbf{z}_k + \mathbf{b}_{\mathrm{ctx}} + \mathbf{W}_{\mathrm{t}}\phi(t_k) + \mathbf{b}_{\mathrm{t}}. \tag{108}$$

A Transformer encoder maps this sequence to a temporal conditioning vector $\mathbf{c}_k$, shared by the coefficient and step-size branches.

Coefficient and Step-Size. The local tensor $\mathbf{x}_k$ is lifted by a pointwise linear projection and processed by residual dilated 1D convolution blocks with exponentially increasing dilations. Each block applies GroupNorm, channel-wise FiLM modulation driven by $\mathbf{c}_k$, a dilated convolution, GLU gating, and a residual connection, with

$$\tilde{\mathbf{h}}_k^{(\ell)} = \big(\mathbf{1} + \boldsymbol{\gamma}_k^{(\ell)}\big) \odot \mathrm{GN}(\mathbf{h}_k^{(\ell)}) + \boldsymbol{\beta}_k^{(\ell)}. \tag{109}$$

Edge and global summaries are concatenated with the temporal context and mapped to generator coefficients $\tilde{\mathbf{a}}_k \in \mathbb{R}^D$, followed by a scaled $\tanh$ saturation $\mathbf{a}_k = s\tanh\big(\tilde{\mathbf{a}}_k/s\big)$. We set $D = (M-N)N$, so that $\mathbf{a}_k$ reshapes into $\mathbf{K}_k \in \mathbb{R}^{(M-N)\times N}$. A second branch applied to the same representation predicts the step size, which modulates the effective update magnitude:

$$\eta_k = \eta_{\min} + (\eta_{\max} - \eta_{\min})\sigma(\tilde{\eta}_k). \tag{110}$$

Summary of Hyperparameters. Unless otherwise stated, we use model width d, Fourier feature count F, band width b, pooling length p, B dilated convolution blocks, two temporal layers, eight attention heads, and step-size bounds $\eta_{\min}$ and $\eta_{\max}$. The default values are

$$d = 256, \quad F = 32, \quad b = 4, \quad p = 32, \quad \text{temporal layers} = 2, \quad \text{heads} = 8, \tag{111}$$

$$B = 6 \ \text{(dilations } 1,2,4,8,16,32), \quad \eta_{\min} = 0.02, \quad \eta_{\max} = 0.08, \quad \texttt{dropout} = 0. \tag{112}$$

B.4 Algorithm

Algorithms 1 and 2 describe the construction of the local finite actions and the subsequent state recovery. At each step k, Algorithm 1 predicts a horizontal coordinate $\mathbf{a}_k$ and a step size η_k, reshapes $\mathbf{a}_k$ into $\mathbf{K}_k \in \mathbb{R}^{(M-N)\times N}$, forms the associated skew-symmetric generator $\Omega[\mathbf{K}_k] = \mathbf{U}_{0,\perp}\mathbf{K}_k\mathbf{U}_0^\top - \mathbf{U}_0\mathbf{K}_k^\top\mathbf{U}_{0,\perp}^\top$, and constructs the finite orthogonal action $\mathbf{R}_k$ through a Cayley map. This yields a sequence of local bundle actions $\{\mathbf{R}_k\}_{k=0}^{T-1}$. In this way, each local update is represented directly as a geometry-preserving transport relative to the reference occupied subspace $(\mathbf{U}_0, \mathbf{U}_{0,\perp})$, rather than as an unconstrained Euclidean increment.

Algorithm 2 accumulates these local actions by an associative scan, producing the prefix products $\mathbf{G}_k = \mathbf{R}_k\mathbf{R}_{k-1}\cdots\mathbf{R}_0$ required for state reconstruction. Each propagated occupied frame

is then recovered as $\mathbf{U}_{k+1} = \mathbf{G}_k \mathbf{U}_0$, and the corresponding projector sequence is obtained as $\mathbf{P}_{k+1} = \mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top$. Accordingly, the full evolution is represented as a parallel scan over finite bundle actions rather than a sequential application of infinitesimal updates. This makes the state propagation compatible with parallel prefix-scan evaluation while retaining the gauge-invariant projector representation throughout.

Algorithm 1 Scan-compatible quantum bundle action

Require: $\mathbf{U}_0 \in \mathbb{R}^{M\times N}$, $\mathbf{t} = (t_0, \ldots, t_{T-1})^\top$, $\mathbf{U}_{\mathrm{ctx}}$, θ
Ensure: $\{\mathbf{R}_k\}_{k=0}^{T-1}$
function SKEW($\mathbf{X}$)
  return $\frac{1}{2}(\mathbf{X} - \mathbf{X}^\top)$
end function
function CAYLEY(Ω, η)
  return $(I + \frac{\eta}{2}\Omega)^{-1}(I - \frac{\eta}{2}\Omega)$
end function
$\mathbf{Q} \leftarrow \mathrm{qr}(\mathbf{U}_0, \mathrm{complete})$
$\mathbf{U}_{0,\perp} \leftarrow \mathbf{Q}[:, N:M]$
for $k = 0, \ldots, T-1$ **do**
  $\mathbf{U}_k \leftarrow \mathbf{U}_{\mathrm{ctx}}$ or $\mathbf{U}_{\mathrm{ctx},k}$
  $\mathbf{P}_k \leftarrow \mathbf{U}_k\mathbf{U}_k^\top$
  $\mathbf{z}_k \leftarrow \mathrm{BandFeat}(\mathbf{P}_k)$
  $\mathbf{c}_k \leftarrow \mathrm{TemporalEncoder}(\mathbf{W}_{\mathrm{ctx}}\mathbf{z}_k + \mathbf{W}_{\mathrm{t}}\phi(t_k))$
  $(\mathbf{a}_k, \tilde{\eta}_k) \leftarrow f_\theta(t_k, \mathbf{U}_k)$
  $\eta_k \leftarrow \eta_{\min} + (\eta_{\max} - \eta_{\min})\sigma(\tilde{\eta}_k)$
  $\mathbf{K}_k \leftarrow \mathrm{reshape}(\mathbf{a}_k, (M-N), N)$
  $\Omega[\mathbf{K}_k] \leftarrow \mathbf{U}_{0,\perp}\mathbf{K}_k\mathbf{U}_0^\top - \mathbf{U}_0\mathbf{K}_k^\top\mathbf{U}_{0,\perp}^\top$
  $\Omega[\mathbf{K}_k] \leftarrow$ SKEW($\Omega[\mathbf{K}_k]$)
  $\mathbf{R}_k \leftarrow$ CAYLEY($\Omega[\mathbf{K}_k], \eta_k$)
end for
return $\{\mathbf{R}_k\}_{k=0}^{T-1}$

Algorithm 2 Parallel prefix-scan of quantum bundle actions

Require: $\mathbf{U}_0 \in \mathbb{R}^{M\times N}$, $\{\mathbf{R}_k\}_{k=0}^{T-1}$
Ensure: $\{\mathbf{U}_k\}_{k=0}^{T}$, $\{\mathbf{P}_k\}_{k=0}^{T}$
function COMBINE($\mathbf{A}, \mathbf{B}$)
  return $\mathbf{BA}$
end function
$\mathbf{U}_0$ given
$\mathbf{P}_0 \leftarrow \mathbf{U}_0\mathbf{U}_0^\top$
$\{\mathbf{G}_k\}_{k=0}^{T-1} \leftarrow \mathrm{AssociativeScan}(\textsc{Combine}(\cdot,\cdot), \{\mathbf{R}_k\}_{k=0}^{T-1})$
for $k = 0, \ldots, T-1$ **do**
  $\mathbf{U}_{k+1} \leftarrow \mathbf{G}_k\mathbf{U}_0$
  $\mathbf{P}_{k+1} \leftarrow \mathbf{U}_{k+1}\mathbf{U}_{k+1}^\top$
end for
return $\{\mathbf{U}_k\}_{k=0}^{T}$, $\{\mathbf{P}_k\}_{k=0}^{T}$