SYMMETRY-AWARE RESIDUAL LEARNING OF FEW-BODY MOLECULAR WAVE FUNCTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Ultracold polar molecules are a highly controllable platform for engineering quantum matter, but modeling their few-body properties remains computationally challenging. Their anisotropic long-range interactions and rich internal structure lead to high-dimensional eigenvalue problems for which conventional grid-based solvers are largely restricted to two-molecule systems. To address this limitation, we introduce a physics-informed neural variational method that exploits known physical structure of the molecular wave function. Our key idea is to factor the wave function into analytic and neural components. The analytic component encodes rotational symmetry and asymptotic decay, while the neural residual learns the remaining intermediate-range behavior. This decomposition reduces the effective dimensionality of the problem without relying on ad hoc approximations or problem-specific discretization grids. We benchmark our method against high-accuracy two-molecule solvers and recover both ground and excited bound states with relative errors at the percent level. We also validate the numerical building blocks required for extension to three-body systems. Specifically, we test the angular-channel decomposition and two-coordinate radial representation in settings where reliable benchmarks are available. This provides a systematic route to larger molecular systems and opens opportunities to study correlations and dynamics in interacting quantum matter.

1 INTRODUCTION

Neural networks (NNs) have become powerful computational tools in quantum physics (Carleo et al., 2019; Karniadakis et al., 2021; Raissi et al., 2019) and chemistry (Goh et al., 2017; Westermayr & Marquetand, 2021; Nomura, 2023; Mroz et al., 2025). For example, they have enabled accurate descriptions of correlated spin dynamics (Carleo & Troyer, 2017) and molecular electronic wave functions (Pfau et al., 2020; Hermann et al., 2020). A central challenge in these problems is the presence of many strongly coupled degrees of freedom, which renders conventional numerical approaches inefficient at high dimensions (Troyer & Wiese, 2005; Schuch & Verstraete, 2009). NNs provide general and flexible representations of candidate solutions that can capture nontrivial quantum structure through direct optimization of the underlying physical objective. While many-body physics has been extensively studied in this context, the full quantum dynamics of few-particle systems with realistic interactions likewise present a high-dimensional NP-hard challenge (Troyer & Wiese, 2005; Kempe et al., 2006) well suited to flexible neural representations.

Ultracold polar molecules with anisotropic long-range interactions coupling multiple internal degrees of freedom represent such a physical setting. They have emerged as a versatile experimental platform for engineered quantum matter due to their high controllability and rich internal structure (Carr et al., 2009; Moses et al., 2017), enabling a wide range of applications from exotic quantum phases (Baranov et al., 2012) to quantum simulation and computation (DeMille, 2002; Wall et al., 2015; Cornish et al., 2024). For fully exploiting control over such systems, it is first crucial to understand the few-body properties of molecules, as they encode the effective interactions and correlations that govern their collective behavior. Solving molecular few-body problems with conventional grid-based methods is primarily limited to two-molecule systems. This limitation motivates new computational approaches that can broaden the range of tractable molecular quantum problems.

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.

In this work, we introduce a NN representation of molecular wave functions to address this bottleneck. A general NN without an inductive bias would have to infer from optimization all physical properties that are characteristic of molecular few-body Hamiltonians, including the overall translational symmetry, angular structure, and asymptotic decay. Instead, our *central idea* is to explicitly incorporate this general physical structure into the wave function representation, leaving the network to learn only a symmetry-aware residual. We realize this idea by factoring the full wave function into an analytic component that encodes known structure, and a neural residual that captures the remaining correlations. The angular dependence is represented using spherical harmonics, reducing the problem to coupled radial channels, and an exponential envelope imposes the asymptotic behavior of each radial channel. The network is left to learn the residual intermediate-range structure. This construction follows the physics-informed neural-network (PINN) philosophy (Karniadakis et al., 2021; Raissi et al., 2019), while remaining closer to a structured variational wave function method than to a black-box PDE solver. Our main contributions are:

- We introduce a neural residual representation for molecular few-body wave functions that combines analytic symmetries and asymptotic structure with a learned residual
- We develop a stable physics-informed initialization and grid-sampling procedure
- We formulate the approach for both two- and three-body systems, and validate it for two-body field-linked bound states against mapped discrete variable representation (MDVR) calculations (Willner et al., 2004)

Extensibility to larger systems: In few-body physics, the particle count (N) indexes a family of related N-body problems, whose complexity scales exponentially with N. While each N requires a corresponding wave function formulation, the principles underlying our representation remain unchanged. Translational invariance is enforced using Jacobi coordinates, while the angular dependence of multiple relative coordinates is represented through tensor products of spherical harmonics and standard angular-momentum coupling (Varshalovich et al., 1988). These ingredients allow the representation to extend systematically to larger few-body systems, although the computational complexity still grows rapidly with N. We explicitly formulate the representation for both 2-body and 3-body systems, while performing full benchmarking on the 2-body case, where reliable reference solutions are available. On top of that, we validate the core numerical building blocks required by the 3-body formulation, including the angular decomposition and two-dimensional radial optimization pipeline. Three-body bound states of field-linked molecules remain substantially less established: recent calculations by Singh & Greene (2026) provide an initial study, but reliable independent benchmarks are not yet available. More broadly, our results illustrate the benefit of using analytical physical structure to reduce the learning burden of neural wave function representations.

2 Related Work

Neural few-body solvers. Neural approaches for continuous-coordinate few-body systems remain underexplored relative to their many-body counterparts. Early works used feedforward networks for few-body bound states of atoms, focusing on universal phenomena and small $^4\mathrm{He}_N$ clusters (Saito, 2018; Yokoi et al., 2026; Freitas & Vitiello, 2023), demonstrating neural representations for a few structureless particles interacting through isotropic short-range potentials. Other work (Li et al., 2025) extended NNs to few-neutron systems with spins, though still under simple short-range interactions. In contrast, we focus on dipolar molecular systems, where anisotropic long-range interactions couple internal degrees of freedom.

Physics-informed PDE solvers. Our work also connects to neural PDE solvers (Raissi et al., 2019), variational approaches, like Deep Ritz (E & Yu, 2018), and extensions to eigenvalue problems and operator learning (Jin et al., 2022; Li et al., 2021; Lu et al., 2021). However, standard neural PDE solvers rarely incorporate the physical symmetries, asymptotic behavior, or variational structure intrinsic to quantum few-body Hamiltonians, although some recent approaches explicitly encode Lie-group symmetries (Klausen et al., 2025).

Symmetry-aware architectures. Our approach aligns with symmetry-aware architectures that embed geometric priors via equivariant NNs to improve efficiency and generalization (Thomas et al., 2018; Kondor & Trivedi, 2018; Batzner et al., 2022). Our method follows this philosophy by combining variational learning with systematic physics-informed structure tailored to first-quantized few-body Schrödinger problems.

3 BACKGROUND

Wave function and Hamiltonian. The wave function $|\Psi\rangle \sim \Psi(\boldsymbol{\alpha})$, defined over generalized coordinates $\boldsymbol{\alpha}$, fully describes the quantum state of a physical system (Sakurai & Napolitano, 2020). However, $|\Psi\rangle$ itself is not directly observable. Instead, according to the Born rule (Born, 1926), the normalized wave function defines the probability density: $p(\boldsymbol{\alpha}) = |\Psi(\boldsymbol{\alpha})|^2$. Quantum observables are represented by Hermitian operators $\hat{O}$ acting on $\Psi(\boldsymbol{\alpha})$, whose expectation values are computed with respect to the probability density $p(\boldsymbol{\alpha})$. One such operator is the Hamiltonian $\hat{\mathcal{H}}(\boldsymbol{\alpha};\boldsymbol{\eta})$, which represents the total energy of the system $\mathcal{E}$ and may depend on extrinsic parameters $\boldsymbol{\eta}$. Expectation values are computed as: $\mathcal{E} = \langle\Psi|\,\hat{\mathcal{H}}\,|\Psi\rangle = \int \Psi^*(\boldsymbol{\alpha})\{\hat{\mathcal{H}}\Psi\}(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha}$, where $\Psi^*(\boldsymbol{\alpha})$ is a complex conjugate.

Schrödinger equation. Our goal is to numerically approximate *stationary states* of a physical system. These states are eigenstates of the Hamiltonian operator $\hat{\mathcal{H}}(\boldsymbol{\alpha};\boldsymbol{\eta})$, and are solutions to the time-independent Schrödinger equation (Sakurai & Napolitano, 2020; Schrödinger, 1926)

$$\hat{\mathcal{H}}(\boldsymbol{\alpha};\boldsymbol{\eta})\Psi_n(\boldsymbol{\alpha}) = \mathcal{E}_n\Psi_n(\boldsymbol{\alpha})\,, \tag{1}$$

where index n labels states $\Psi_n(\boldsymbol{\alpha})$ with eigenenergies $\mathcal{E}_n$, and $\boldsymbol{\eta}$ denotes extrinsic parameters.

Variational method. The conventional numerical approach to solving eigenvalue problems, such as Eq. (1), employs the variational principle (Yuan et al., 2019). One considers an arbitrary function $|\Psi_{\boldsymbol{w}}\rangle$, parametrized in a complete functional basis $|\psi_k^w\rangle$, such that $|\Psi_{\boldsymbol{w}}\rangle = \sum_{k=1}^{\infty} w_k\,|\psi_k^w\rangle$. The normalized expectation value of the Hamiltonian is

$$\mathcal{E}_{\boldsymbol{w}}[\hat{\mathcal{H}}] \equiv \langle\hat{\mathcal{H}}\rangle = \langle\Psi_{\boldsymbol{w}}|\,\hat{\mathcal{H}}\,|\Psi_{\boldsymbol{w}}\rangle\,/\,\langle\Psi_{\boldsymbol{w}}|\Psi_{\boldsymbol{w}}\rangle\ . \tag{2}$$

The parameters $\boldsymbol{w}_0$ that minimize $\mathcal{E}_{\boldsymbol{w}}[\hat{\mathcal{H}}]$ yield the lowest-energy eigenstate of $\hat{\mathcal{H}}(\boldsymbol{\alpha};\boldsymbol{\eta})$ within the chosen parametrization. Its energy, $\mathcal{E}_0 = \mathcal{E}_G = \mathcal{E}_{\boldsymbol{w}_0}[\hat{\mathcal{H}}]$, is called the *ground-state* energy.

Higher-energy (excited) states can be obtained via the Gram–Schmidt orthogonalization (MacDonald, 1933; Ding et al., 2024). Given n_0 lowest-energy wave functions $\{|\Psi^0\rangle, ..., |\Psi^{n_0-1}\rangle\}$, one can repeat the variational algorithm in the subspace orthogonal to these functions using the trial wave function

$$|\Psi_{\boldsymbol{w}}\rangle = |\Psi_{\boldsymbol{w}}^{n_0}\rangle - \sum\nolimits_{n=0}^{n_0-1}\big(\,\langle\Psi^n|\Psi_{\boldsymbol{w}}^{n_0}\rangle\,/\,\langle\Psi^n|\Psi^n\rangle\,\big)\,|\Psi^n\rangle\ . \tag{3}$$

Eq. (3) is not the only strategy for calculating excited states (González & Lindh, 2020). Alternatives include secular determinant or generalized Rayleigh–Ritz method (Coulson, 1979; Roothaan, 1951), used, for example, by Pfau et al. (2024). These methods construct overlap matrices from trial functions and solve a generalized eigenvalue problem, yielding several low-energy states simultaneously. While useful when orthogonalization becomes numerically unstable, these methods require more resources. For our purposes, a simpler method, based on Eq. (3), provides sufficient performance.

Training objective. Eqs. (2) and (3) define the training objective used throughout this work. Implementation details and pseudocode for the optimization procedure are provided in Apps. B.2 and E.

4 PROBLEM FORMULATION

4.1 PHYSICAL MODEL

Hamiltonian of a few-body system. The Hamiltonian operator $\hat{\mathcal{H}}(\boldsymbol{\alpha};\boldsymbol{\eta})$ (see Sec. 3) describes an N-body system with coordinates $\boldsymbol{\alpha} \equiv \{\boldsymbol{R}_i\}_{i=1}^N$ and physical parameters $\boldsymbol{\eta} \equiv \boldsymbol{\eta}_{\text{pot}} \cup \{m_i\}_{i=1}^N$. For particles with masses $\{m_i\}_{i=1}^N$, the Hamiltonian partitions into kinetic and potential terms:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{\text{kin}} + \hat{\mathcal{H}}_{\text{pot}}\ ,\ \text{where}\quad \hat{\mathcal{H}}_{\text{kin}} = \sum\nolimits_{i=1}^{N}\mathcal{E}_i^{\text{kin}}\boldsymbol{\nabla}_{\boldsymbol{R}_i}^2 \quad\text{and}\quad \hat{\mathcal{H}}_{\text{pot}} = \sum\nolimits_{i<j}^{N} V(\boldsymbol{\eta}_{\text{pot}}, \boldsymbol{R}_i - \boldsymbol{R}_j)\,. \tag{4}$$

Here $V(\boldsymbol{\eta}_{\text{pot}}, \boldsymbol{R}_i - \boldsymbol{R}_j)$ is the interaction between i-th and j-th particles, $\boldsymbol{\nabla}_{\boldsymbol{R}_i}$ is the gradient operator, and $\mathcal{E}_i^{\text{kin}} = -\hbar^2/(2m_i)$ are kinetic factors, where $\hbar$ is the reduced Planck constant. We assume that all particles have the same mass m, and that $V(\boldsymbol{\eta}_{\text{pot}}, \boldsymbol{R}_i - \boldsymbol{R}_j)$ depends only on the relative position vector $\boldsymbol{R}_i - \boldsymbol{R}_j$. Such an interaction implies overall *translational invariance* in the system, which is characteristic of systems without inhomogeneous external fields.

For translationally invariant systems, we replace laboratory coordinates $\{\boldsymbol{R}_i\}_{i=1}^{N}$ with Jacobi relative coordinates $\{\boldsymbol{r}_i\}_{i=1}^{N-1}$ (defined in App. A.1, see Suzuki & Varga (1998) and Jacobi (2009)) and the center-of-mass coordinate $\boldsymbol{R}_{\mathrm{CM}}$. $\boldsymbol{R}_{\mathrm{CM}}$ describes trivial global motion, separate from the intrinsic dynamics, and can be discarded. We therefore solve only for $\boldsymbol{\Psi}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N-1})$, reducing the dimensionality while explicitly enforcing translational invariance on the level of the Hamiltonian.

Field-linked molecules. We focus on ultracold polar molecules. When these molecules approach each other at short range, they undergo inelastic or reactive processes (Ospelkaus et al., 2010; Liu et al., 2020; Gregory et al., 2020; Bause et al., 2023), which leads to clustering and loss of correlations. To overcome this challenge, an external (microwave) field can be coupled to molecules to induce a potential barrier that strongly suppresses short-range losses, a mechanism known as *microwave shielding* (Büchler et al., 2007; Micheli et al., 2007; Cooper & Shlyapnikov, 2009; Karman & Hutson, 2018; Anderegg et al., 2021; Deng et al., 2023). The potential barrier manifests itself in a rapid increase of the interaction energy $V(\boldsymbol{\eta}_{\mathrm{pot}}, \boldsymbol{r})$, as $|\boldsymbol{r}| \to 0$ (Fig. 1(b)). The shielding potential between microwave-dressed molecules is well approximated by the following expression (Deng et al., 2023):

$$V(\boldsymbol{\eta}_{\mathrm{pot}}, \boldsymbol{r}) = C_6(\boldsymbol{\eta}_{\mathrm{pot}})[1 - \cos^4\theta]\, r^{-6} + C_3(\boldsymbol{\eta}_{\mathrm{pot}})[3\cos^2\theta - 1]\, r^{-3}\,, \tag{5}$$

where $\boldsymbol{r} = \{r, \theta, \phi\}$ denote spherical coordinates, and interaction strength coefficients C_3 and C_6 depend on molecular properties and can be tuned via field parameters $\boldsymbol{\eta}_{\mathrm{pot}} = \{\nu, \ldots\}$. Their explicit form is provided in App. A.2. The main extrinsic physical parameter is the Rabi coupling ν that can be adjusted to shift the barrier to shorter range (Fig. 1(b)), deepening the potential well. When depth of the well is sufficient to fit a matter-wave mode, the potential supports molecular binding. Such a molecular bond is stable only in the presence of the external electric field and is therefore referred to as a *field-linked* bound state. Recent observation of field-linked dimer states in dipolar molecules (Chen et al., 2024) demonstrates that these bonds can be reversibly created and destroyed by tuning the external field, enabling controllable assembly of larger multi-molecular bound states.

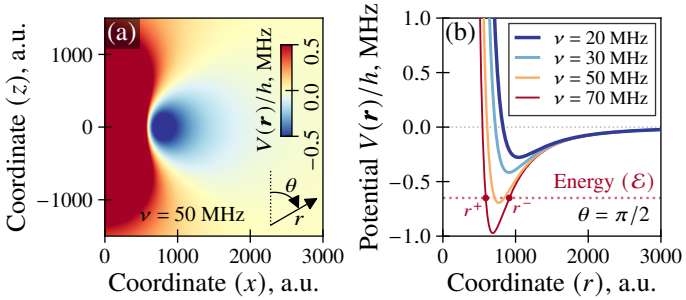

Figure 1: **Interaction potential of field-linked molecules.** (a) Potential $V(\boldsymbol{r})$ is strongly anisotropic with respect to the angle θ, yielding intricate correlation dynamics. (b) Potential $V(\boldsymbol{r})$ exhibits energy barrier for short-range ($|\boldsymbol{r}| \to 0$) interactions, whose position and strength depend on the Rabi parameter ν.

4.2 Challenges for numerical approaches

The physical model introduced in Sec. 4.1 poses several challenges for eigenvalue solvers. It requires simultaneously resolving continuous spatial coordinates and intricate interaction-induced correlations.

Quantum wave function representations are often tailored to two limiting regimes: many-body problems with established ansätze or analytically tractable few-body systems. In many-body physics (electronic or spin states), single-particle states can often be represented in truncated bases, enabling physically motivated multi-particle trial states, like Slater determinants (Pfau et al., 2020; Hermann et al., 2020). These approaches efficiently capture correlations, but become increasingly difficult to construct as the complexity of single-particle states grows. At the other extreme are simple few-body systems, such as harmonic confinement or isotropic interactions, where closed-form solutions are often available. These methods typically focus on the continuous nature of particle coordinates, but do not embed correlations into the trial state. Our problem lies between these regimes: although involving only a few particles, the interactions are long-range and anisotropic (see Fig. 1(a)), leading to intrinsically coupled continuous degrees of freedom. As a result, the wave function does not admit a simple factorized form or compact low-dimensional parametrization, and must instead be approximated as the solution of a high-dimensional PDE eigenvalue problem on a continuous manifold.

Interaction potential (5) is particularly difficult for numerical representation. On the one hand, the Hamiltonian (4) is invariant under global translations and rotations, implying that a physically meaningful solution to Eq. (1) needs to be symmetric with respect to the Euclidean group E(3). On the other hand, anisotropic potential $V(\boldsymbol{\eta}_{\mathrm{pot}}, \boldsymbol{r})$ results in single particle states that violate the

rotational symmetry of SO(3), a subgroup of E(3). Moreover, the competing length scales, arising from r^{-3} and r^{-6} terms in Eq. (5), induce both sharp and slowly varying radial structure in the wave function, particularly for excited states with an additional nodal structure.

These challenges call for a flexible representation of a wave function, which is capable of capturing the rich structure of solutions to Eq. (1) within a unified framework. Such a representation should avoid restrictive functional bases, as it has to adapt to the structural diversity of the eigenfunctions. The variational formulation introduced in Sec. 3 applies to any parametric model. In this work, we adopt a neural network parametrization of the wave function, augmented with physics-informed inductive biases to capture the underlying structure of the problem.

5 Physics-Informed Neural Variational Method

We introduce a neural wave function that embeds the geometric and physical structure of the problem, described in Sec. 4.1. The *key idea* is to encode relevant symmetries into the representation, reducing the problem to learning a low-dimensional residual within a unified variational formulation.

5.1 Spherical decomposition

The action of the Hamiltonian operator (4) on the wave function has an internal symmetry structure. While this structure can be learned by the network (see Sec. 6.2 and App. C.3), doing so is highly inefficient when this angular structure is analytically known and can be embedded into the representation (Klausen et al., 2025). We therefore encode the SO(3) angular-momentum coupling rules into the representation for $\mathbf{\Psi}(\boldsymbol{r}_1, ..., \boldsymbol{r}_{N-1})$, treating molecular rotations purely analytically. Full derivations for two- and three-molecule systems are presented respectively in Apps. A.3 and A.4. In the 2-body case, we expand the wave function in spherical harmonics (Varshalovich et al., 1988): $\Psi(\boldsymbol{r}) = \sum_{l,m}[\psi_{l,m}(r)/r]\, Y_{l,m}(\theta, \phi)$, where $\boldsymbol{r} = \{r, \theta, \phi\}$ are spherical coordinates and $Y_{l,m}(\theta, \phi)$ are spherical harmonics. Matrix elements of the Hamiltonian become: $\mathcal{H}_{l,m;l',m'}(r) = \langle Y_{l',m'}|\, \hat{\mathcal{H}}\, |Y_{l,m}\rangle$, reducing the Schrödinger equation (1) to a leaner system of coupled radial equations

$$\sum\nolimits_{l',m'} \mathcal{H}_{l,m;l',m'}(r)\, \psi_{l',m'}(r) = \mathcal{E}\, \psi_{l,m}(r)\,. \tag{6}$$

Spherical harmonics $Y_{l,m}(\theta, \phi)$ offer several advantages as an angular basis. First, they form irreducible representations of the SO(3) group and provide a natural orthonormal basis for angular functions. Unlike problem-specific polynomial bases, they arise directly from rotational symmetry and therefore remain applicable under changes in interactions or boundary conditions. Second, they analytically absorb singularities of spherical coordinates and diagonalize the kinetic energy operator (4) in the angular basis, yielding numerically stable differential operators and reducing computational cost. Third, the decomposition naturally extends to higher-body systems through tensor products of spherical harmonics and angular-momentum coupling rules, enabling scalable representations of multi-coordinate wave functions. Finally, the expansion admits a physically justified truncation at finite $l_{\max}$, since higher-l channels are suppressed by centrifugal barriers and therefore contribute negligibly to the dynamics. Further discussion is provided in App. B.1.

5.2 WKB asymptotic as an inductive bias

After incorporating the symmetry structure of the Hamiltonian into the wave function representation, we further improve efficiency by encoding the asymptotic behavior of radial functions $\psi_{l,m}(r)$. The asymptotic form is obtained through multiple-scale analysis (Bender & Orszag, 1999) of Eq. (6), yielding the Wentzel–Kramers–Brillouin (WKB) approximation (7). Specifically, in the "forbidden" region, where $V(r) > \mathcal{E}$ (see Fig. 1(b)), the wave function can be approximated by

$$\psi_{\text{WKB}}(r, \mathcal{E}) \sim \exp\left(\pm \int_{r_\pm}^{r} \kappa_{\mathcal{E}}(x)\mathrm{d}x \right), \text{where} \quad \kappa_{\mathcal{E}}(r) = \sqrt{(V(r) - \mathcal{E})/|\mathcal{E}_{\text{kin}}|}\,, \tag{7}$$

where $\mathcal{E}$ is the energy of the system, $V(r) \equiv V(r, \theta = \frac{\pi}{2}, \phi = 0)$ (see Eq. (5)) and the sign is "+" for $r \ll r_+$ and "−" for $r \gg r_-$, where $r_\pm$ are the turning points, defined as solutions to $\kappa_{\mathcal{E}}(r) = 0$ (see Fig. 1(b)). Approximation Eq. (7) is rather inaccurate around and between the turning points $r_\pm$, however it provides the correct asymptotics in $r \to 0$ and $r \to \infty$ regimes.

Since the exact system energy $\mathcal{E}$ is unknown before the training, we define the reference envelope using the proxy energy $\mathcal{E}_{\text{WKB}} = \min_r V(r)/2$. This proxy does not approximate the ground-state energy, in fact, for $\nu = 50\,\text{MHz}$: $\mathcal{E}_{\text{WKB}} \approx 10\mathcal{E}_{\text{G}}$. Instead, it tracks changes in interaction parameters, such as ν. Altogether, we use WKB-inspired asymptotic envelope $\psi_{\text{WKB}}(r, \mathcal{E}_{\text{WKB}})$ instead of employing the approximation itself.

Figure 2 provides comparison between typical behavior of target ground-state wave function $\psi_{\text{G}}(r)$ and the corresponding WKB approximations (7). A literal WKB envelope can be overly restrictive, limiting the neural network's freedom to explore nearby parameters. Hence, we intentionally weaken the asymptotic prior by broadening the envelope via $\kappa_{\mathcal{E}}(r) \to \beta\kappa_{\mathcal{E}}(r)$, where $\beta \in (0, 1)$. The details can be found in App. E.

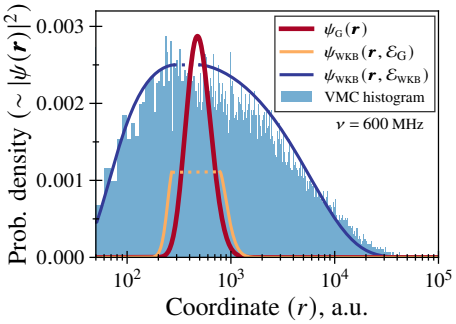

Figure 2: **WKB-inspired asymptotic envelope used for sampling.** The approximation $\psi_{\text{WKB}}(r, \mathcal{E}_{\text{G}})$ captures the asymptotic behavior of the wave function $\psi_{\text{G}}(r)$. To improve convergence, we use a broader reference envelope $\psi_{\text{WKB}}(r, \mathcal{E}_{\text{WKB}})$ as a prior and as a fixed sampling distribution for the VMC. The histogram shows VMC samples drawn from this envelope.

5.3 Neural Residual

Combining the angular decomposition of Sec. 5.1 with the asymptotic envelope of Sec. 5.2, the neural network only needs to represent the remaining radial structure. We parametrize these residual radial functions with a *shared* multi-layer perceptron (MLP), which maps the radial coordinate r jointly to all angular channels $\psi^{\text{NN}}_{l,m}(r)$. Sharing the network parameters across channels allows correlations between different angular components to be captured within a common low-dimensional representation. For the two-body system, the resulting wave function representation is

$$\Psi(\boldsymbol{r}) = \sum_{l,m} [\psi^{\text{NN}}_{l,m}(r)\psi_{\text{WKB}}(r, \mathcal{E}_{\text{WKB}})/r]\, Y_{l,m}(\theta, \phi)\,. \tag{8}$$

The network parameters are optimized variationally using the objective of Sec. 3. For the ground state, we minimize the normalized expectation value of the Hamiltonian (2), while excited states are obtained by applying the same optimization to the orthogonalized trial wave function (3). At each optimization step, the neural residual $\psi^{\text{NN}}_{l,m}(r)$ is evaluated on a minibatch of radial coordinates, combined with the fixed analytical components to construct the full wave function (8), and used to evaluate the variational energy. Gradients are then backpropagated through the neural residual. Thus, only the residual functions are learned, while the angular basis and asymptotic envelope remain fixed. The complete procedure and implementation details are provided in Apps. B.2 and E.

5.4 Sampling and Grid Construction

Accurate evaluation of the variational objective depends critically on the sampling strategy: regions that contribute substantially to the expectation value must be sufficiently resolved to avoid biased energy estimates. Ideally, the sampling grid should therefore follow the underlying wave function distribution. A straightforward approach uses a fixed discretization of coordinate space. However, the wave function exhibits strong multi-scale structure (Bender & Orszag, 1999), with sharp localization in some regions and decay over many orders of magnitude elsewhere. Uniform grids are therefore inefficient: they either fail to resolve narrow features or allocate excessive resolution to regions with negligible probability mass (see App. D.4). Fixed coordinate transformations improve resolution across scales, but remain problem-specific and nonadaptive.

This motivates sampling from a target probability distribution. In variational Monte Carlo (VMC), samples are drawn from the model-induced density $\rho_{\boldsymbol{\omega}}(\boldsymbol{r}) \sim |\Psi_{\boldsymbol{\omega}}(\boldsymbol{r})|^2$ using Markov chain Monte Carlo (MCMC) methods (Brooks et al., 2011; Chib & Greenberg, 1995; Robert & Casella, 2004). Because this distribution evolves during optimization, the Markov chain must be repeatedly re-equilibrated, increasing computational cost and destabilizing training. Moreover, the learned distribution may transiently become nearly uniform (Sec. 6.4), making it uninformative for sampling. We therefore decouple grid construction from the evolving model by initializing a fixed MCMC

sampling grid before training. Specifically, we sample points from a target density proportional to $|\psi_{\mathrm{WKB}}(r, \mathcal{E}_{\mathrm{WKB}})|^2$, introduced in Sec. 5.2, and draw minibatches from this fixed grid during optimization. As shown in Fig. 2, the resulting distribution captures the large-scale structure of the problem while remaining independent of the evolving model. Implementation details are provided in App. E.

6 Experiments

We evaluate our method on two-body field-linked molecular systems, where high-accuracy reference solutions are available. Our experiments address four questions: (i) can the method reproduce MDVR reference solutions, (ii) which components of the physics-informed design are essential, (iii) do the core numerical building blocks for higher-body extensions function reliably, (iv) how does optimization behave across interaction regimes? All experiments use a shared model configuration *without fine-tuning* and exhibit stable optimization behavior. Extended results are provided in App. D, with implementation details and pseudocode in Apps. E and B.2, respectively[1].

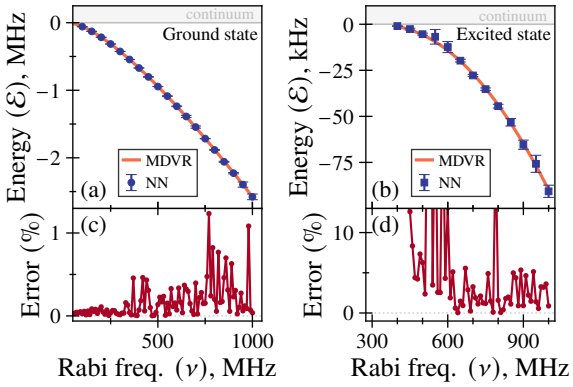

Figure 3: **Comparison with MDVR.** Predictions from our approach vs. MDVR reference for ground and excited states across ν. Top: predicted and reference energies. Bottom: relative absolute errors, $|\mathrm{pred} - \mathrm{ref}|/|\mathrm{ref}|$. The model accurately reproduces the energy spectrum across a broad range of interaction regimes.

6.1 Comparison with MDVR

We compare our approach against MDVR (App. A.3.3), a high-accuracy grid-based variational solver widely used for molecular bound-state calculations (which we independently validate against a Weinberg-expansion benchmark in App. D.6). We evaluate ground and excited states across Rabi frequencies $\nu \in [50, 1000]$ MHz ($\Delta\nu = 10$ MHz). Sweeping ν alters the depth and position of the field-linked potential well, probing physically distinct optimization regimes: low ν produces shallow wells with spatially extended states and weak gradients, whereas high ν induces strong confinement and tightly localized wave functions. Excited states appear only for $\nu \gtrsim 400$ MHz, as lower frequencies lack a sufficiently deep potential well to support bound excited states.

As shown in Fig. 3, our model closely matches the MDVR reference energies across all physical regimes, accurately reproducing both the ordering and energy spacing of bound states. The mean absolute relative error is 3×10^{-3} for the ground state and 1.14×10^{-1} for excited states. Higher excited-state errors concentrate near the emergence threshold ($\nu \approx 400$ MHz), where states are extremely weakly bound, making optimization particularly sensitive to nodal structure and minor energy shifts. As ν increases and the potential deepens, error drops significantly and optimization stabilizes. Multi-seed experiments demonstrate low variance and high stability across initialization seeds (App. D.5). Although MDVR remains computationally faster in this low-dimensional setting, our goal is not to outperform specialized grid solvers, but to demonstrate that our representation robustly captures distinct physical regimes while extending to higher-dimensional few-body systems.

6.2 Ablations of Physics-Informed Components

We isolate the contribution of the main design choices by modifying individual components while keeping the variational objective and optimizer fixed. Specifically, we compare the full model against (i) a variant without the WKB-inspired asymptotic envelope, (ii) a direct matrix parametrization without shared structure across channels, and (iii) a 2D representation without analytical spherical harmonics. Additional experiments are provided in App. D.3.

Effect of the WKB-induced inductive bias. The WKB envelope has the largest effect at low ν, where bound states are weakly confined and extend over a broad spatial region. In this regime, the

[1]Code will be released upon publication.

envelope acts as a physics-informed inductive bias by analytically enforcing the correct asymptotic decay, allowing the network to focus on the intermediate-range structure instead of learning long-range behavior from optimization alone. Incorporating this prior directly into the representation substantially accelerates convergence (Fig. 4(a)). As ν increases and the ground state becomes more localized, the difference between the two variants decreases, since optimization becomes dominated by the local structure of the potential well.

Effect of neural parametrization. We consider a direct matrix parametrization as a baseline, using 100k parameters compared to 84k in the proposed neural representation. Here, spherical channel coefficients are represented explicitly as an $N_r \times N_\ell$ matrix, where each entry is treated as an independent learnable parameter without shared structure across radial or angular dimensions. This corresponds to independently discretizing the radial dependence of each angular channel. To isolate the effect of the parametrization, both models are multiplied by the same WKB-inspired envelope.

Despite having more parameters, the matrix parametrization performs substantially worse than the shared MLP, particularly at low ν, where it often fails to converge within the optimization budget (Fig. 4(b)). This reflects the difficulty of learning spatially extended states with independent parameters and weak distributed gradients. As ν increases and states become more localized, the performance gap narrows. Increasing the parameter count from 100k to 200k does not recover convergence in the low-frequency regime, indicating that the gap is not explained by capacity alone. These results highlight the importance of shared parametrization across coupled angular channels.

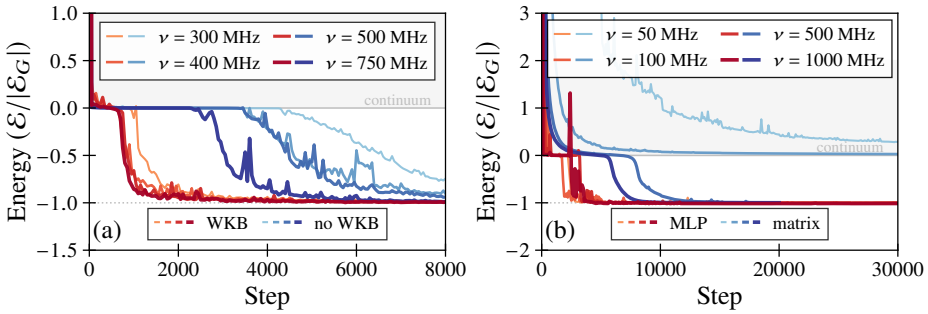

Figure 4: **Effects of the WKB-inspired asymptotics and neural parametrization.** (a) Ground-state training trajectories with and without the WKB-inspired envelope across different ν. Encoding the correct asymptotic decay directly into the wave function representation accelerates convergence. (b) Training dynamics of a direct matrix parametrization versus the shared MLP parametrization across ν. The matrix baseline exhibits slower convergence and instability, particularly at low ν, while the MLP model converges reliably and reaches lower energies. In both cases, the effect is more pronounced at low ν, where the wave function is spatially extended.

Effect of analytical symmetries. Without the analytical angular decomposition described in Sec. 5.1, the network must learn angular structure directly from a full 2D representation (see Fig. 5(a)) in reduced cylindrical coordinates $\boldsymbol{r} = \{r_\perp, z\}$. To evaluate this representation, we construct a 2D numerical solver that samples coordinate points randomly, integrates via Delaunay triangulation, evaluates kinetic energy operators on local interpolations to maintain stability, and extends the WKB-inspired prior to a product of 2D coordinate envelopes (see App. C). As shown in Fig. 5(b), this unconstrained 2D network eventually converges to the correct bound-state energy. Yet, it suffers from slower convergence ($\sim \times 100$ longer optimization time), requires $\times 15$ larger batch sizes, and exhibits reduced training stability under identical hardware configurations. This confirms that analytically factoring out rotational symmetry provides the efficiency and stability gains in our framework.

6.3 Extension to Three-Body Systems

Extending our framework to a three-body ($N = 3$) system requires two Jacobi coordinates $(\boldsymbol{r}, \boldsymbol{R})$, expanding the coordinate space to 6D. Following the same structural decomposition, the 3-body wave function factorizes into dual-angular channels and coupled 2D radial residuals (see App. A.4):

$$\Psi_{\text{3-body}}(\boldsymbol{r}, \boldsymbol{R}) = \tfrac{1}{rR} \sum\nolimits_{l,m;L,M} \psi^{\text{2D}}_{l,m;L,M}(r, R) \left[Y_{l,m}(\boldsymbol{r}) \otimes Y_{L,M}(\boldsymbol{R}) \right]. \tag{9}$$

While universally accepted 3-body reference energies for field-linked molecules remain an open question in the domain, we explicitly validate the core numerical machinery required for this higher-

dimensional formulation using reliable two-body baselines. Specifically, we validate the spherical harmonic decomposition in Sec. 6.1, and the 2D radial residual $\psi^{2D}(r, R)$ via the unconstrained baseline in Sec. 6.2. Validating both components independently in benchmarked regimes demonstrates that our variational framework systematically extends to higher-dimensional few-body problems.

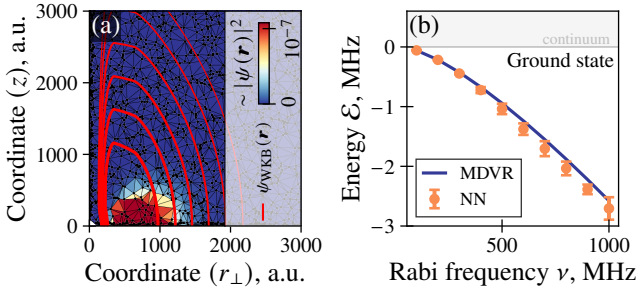

Figure 5: **Validation of the 2D representation.** (a) Ground-state probability density learned in reduced cylindrical coordinates $\{r_\perp, z\}$, with the sampling mesh and WKB-inspired envelope. (b) Predicted ground-state energies across ν compared with MDVR. The agreement validates the multi-coordinate optimization machinery required for higher-body extensions.

6.4 Optimization Dynamics

To better understand optimization, Fig. 6 shows the evolution of energies and wave functions during training for representative low- and high-ν ground-state regimes. For small ν, optimization exhibits a two-stage behavior: the model initially remains in a diffuse, near-zero-amplitude state (continuum) with weak gradients, followed by gradual accumulation of probability mass near the attractive region and rapid convergence to the bound state. Correspondingly, the wave function contracts progressively from a broad delocalized profile. In contrast, large ν induces strong gradients from the outset, leading to early localization and substantially faster convergence. In this regime, the wave function rapidly forms a localized bound state, with subsequent training only refining its structure. Overall, optimization difficulty is governed primarily by the strength of the confining potential: shallow potentials lead to slow discovery of bound states, while deep potentials enable rapid convergence.

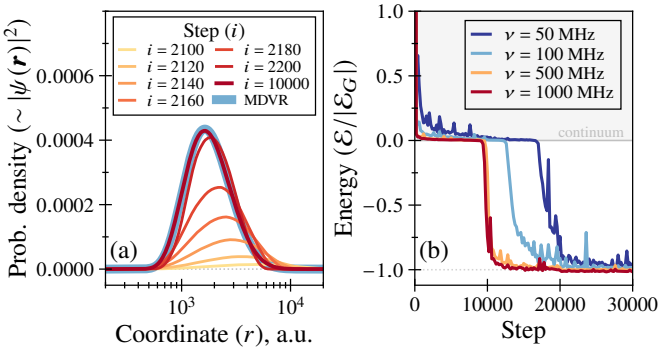

Figure 6: **Optimization dynamics across interaction regimes.** (a) Evolution of the ground-state probability density during training, compared with MDVR. A reduced learning rate is used to resolve the transient localization dynamics. (b) Normalized energy trajectories for representative ν. Low ν shows a plateau before localization, and high ν converges rapidly due to stronger confinement.

7 Discussion

In this work, we introduced a physics-informed neural framework for solving quantum few-body molecular problems with anisotropic long-range interactions. Rather than relying on handcrafted discretizations or problem-specific bases, the method combines a variational neural representation with symmetry-aware and asymptotic inductive biases derived directly from the Hamiltonian structure. Rotational symmetries are embedded analytically through spherical harmonic decomposition, while asymptotic behavior is incorporated through a tempered WKB-inspired envelope. Across interaction regimes, the approach accurately reproduces MDVR solutions for both ground and excited states while maintaining stable optimization behavior. While our primary benchmarks focus on two-body field-linked molecular systems to enable controlled evaluation against established baselines, we explicitly formulate the three-body extension and experimentally validate each of its core numerical components. More broadly, these results suggest that embedding physical structure into neural wave function representations is a promising direction for scalable first-principles simulation of correlated molecular quantum systems and engineered quantum matter.

AI USE STATEMENT

We have not used generative AI tools for tasks with required disclosure. Additionally, we used generative AI tools for editing text, unit testing and generating figures in code, and identifying the relevant literature. We have reviewed all AI-assisted work. We made sure the text is consistent with our research direction and accurate and the code runs correctly without any biases. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To ensure full reproducibility of our results, all algorithmic details, hyperparameter configurations, and experimental setups are provided in the Appendix. Specifically, the complete optimization procedure and model workflow are summarized as pseudocode in App. B.2, while training hyperparameters, network architectures, and technical details are documented in App. E. Furthermore, to verify optimization stability against stochastic initialization, we report mean and standard deviation metrics across multiple random seeds in App. D.5. Source code and data generation scripts will be made publicly available upon publication.

REFERENCES

Loïc Anderegg, Sean Burchesky, Yicheng Bao, Scarlett S. Yu, Tijs Karman, Eunmi Chae, Kang-Kuen Ni, Wolfgang Ketterle, and John M. Doyle. Observation of microwave shielding of ultracold molecules. *Science*, 373(6556):779–782, August 2021. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.abg9502. URL https://www.science.org/doi/10.1126/science.abg9502.

M. A. Baranov, M. Dalmonte, G. Pupillo, and P. Zoller. Condensed matter theory of dipolar quantum gases. *Chemical Reviews*, 112(9):5012–5061, August 2012. ISSN 1520-6890. doi: 10.1021/cr2003568. URL http://dx.doi.org/10.1021/cr2003568.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.

Roman Bause, Arthur Christianen, Andreas Schindewolf, Immanuel Bloch, and Xin-Yu Luo. Ultracold Sticky Collisions: Theoretical and Experimental Status. *The Journal of Physical Chemistry A*, 127(3):729–741, January 2023. ISSN 1089-5639, 1520-5215. doi: 10.1021/acs.jpca.2c08095. URL https://pubs.acs.org/doi/10.1021/acs.jpca.2c08095.

Carl M. Bender and Steven A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Springer New York, New York, NY, 1 edition, 1999. ISBN 978-0-387-98931-0. doi: 10.1007/978-1-4757-3069-2. Originally published by McGraw-Hill, 1978.

Max Born. Zur quantenmechanik der stoßvorgänge. *Zeitschrift für Physik*, 37:863–867, 1926. doi: 10.1007/BF01397477.

Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng (eds.). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011. ISBN 978-1-4200-7941-8. doi: 10.1201/b10905.

H. P. Büchler, E. Demler, M. Lukin, A. Micheli, N. Prokof'ev, G. Pupillo, and P. Zoller. Strongly correlated 2d quantum phases with cold polar molecules: Controlling the shape of the interaction potential. *Phys. Rev. Lett.*, 98:060404, Feb 2007. doi: 10.1103/PhysRevLett.98.060404. URL https://link.aps.org/doi/10.1103/PhysRevLett.98.060404.

Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017. doi: 10.1126/science.aag2302.

Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019. doi: 10.1103/RevModPhys.91.045002.

Lincoln D Carr, David DeMille, Roman V Krems, and Jun Ye. Cold and ultracold molecules: science, technology and applications. *New Journal of Physics*, 11(5):055049, may 2009. doi: 10.1088/1367-2630/11/5/055049. URL https://doi.org/10.1088/1367-2630/11/5/055049.

Xing-Yan Chen, Shrestha Biswas, Sebastian Eppelt, Andreas Schindewolf, Fulin Deng, Tao Shi, Su Yi, Timon A. Hilker, Immanuel Bloch, and Xin-Yu Luo. Ultracold field-linked tetratomic molecules. *Nature*, 626(7998):283–287, February 2024. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-023-06986-6. URL https://www.nature.com/articles/s41586-023-06986-6.

Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335, 1995. doi: 10.1080/00031305.1995.10476177.

N. R. Cooper and G. V. Shlyapnikov. Stable topological superfluid phase of ultracold polar fermionic molecules. *Phys. Rev. Lett.*, 103:155302, Oct 2009. doi: 10.1103/PhysRevLett.103.155302. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.155302.

Simon L. Cornish, Michael R. Tarbutt, and Kaden R. A. Hazzard. Quantum computation and quantum simulation with ultracold molecules. *Nature Physics*, 20(5):730–740, May 2024. ISSN 1745-2481. doi: 10.1038/s41567-024-02453-9. URL http://dx.doi.org/10.1038/s41567-024-02453-9.

Charles Alfred Coulson. *Coulson's Valence*. Oxford Chemistry Series. Oxford University Press, Oxford, 3 edition, 1979. ISBN 0198551452.

Boris Delaunay. Sur la sphère vide. à la mémoire de georges voronoï. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et naturelles*, (6):793–800, 1934.

D. DeMille. Quantum computation with trapped polar molecules. *Phys. Rev. Lett.*, 88:067901, Jan 2002. doi: 10.1103/PhysRevLett.88.067901.

Fulin Deng, Xing-Yan Chen, Xin-Yu Luo, Wenxian Zhang, Su Yi, and Tao Shi. Effective potential and superfluidity of microwave-shielded polar molecules. *Phys. Rev. Lett.*, 130:183001, May 2023. doi: 10.1103/PhysRevLett.130.183001. URL https://link.aps.org/doi/10.1103/PhysRevLett.130.183001.

Lexin Ding, Cheng-Lin Hong, and Christian Schilling. Ground and Excited States from Ensemble Variational Principles. *Quantum*, 8:1525, November 2024. ISSN 2521-327X. doi: 10.22331/q-2024-11-14-1525.

Weinan E and Bing Yu. The deep ritz method: a deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018.

William Freitas and SA Vitiello. Synergy between deep neural networks and the variational monte carlo method for small $^4\mathrm{He}_N$ clusters. *Quantum*, 7(arXiv: 2302.00599):1209, 2023.

Garrett B. Goh, Nathan O. Hodas, and Abhinav Vishnu. Deep learning for computational chemistry. *Journal of Computational Chemistry*, 38(16):1291–1307, 2017. doi: https://doi.org/10.1002/jcc.24764. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.24764.

Leticia González and Roland Lindh (eds.). *Quantum Chemistry and Dynamics of Excited States: Methods and Applications*. John Wiley & Sons, 2020. ISBN 9781119417750. doi: 10.1002/9781119417774.

Philip D. Gregory, Jacob A. Blackmore, Sarah L. Bromley, and Simon L. Cornish. Loss of Ultracold $^{87}\mathrm{Rb}^{133}\mathrm{Cs}$ Molecules via Optical Excitation of Long-Lived Two-Body Collision Complexes. *Physical Review Letters*, 124(16):163402, April 2020. ISSN 0031-9007. doi: 10.1103/PhysRevLett.124.163402. URL http://arxiv.org/abs/2002.04431.

Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic schrödinger equation. *Nature Chemistry*, 12(10):891–897, Oct 2020. ISSN 1755-4349. doi: 10.1038/s41557-020-0544-y.

Carl Gustav Jakob Jacobi. *Jacobi's Lectures on Dynamics: Delivered at the University of Königsberg in the Winter Semester 1842–1843 and According to the Notes Prepared by C. W. Borchardt*. Texts and Readings in Mathematics. Springer, 2 edition, 2009. ISBN 9789386279620.

Henry Jin, Marios Mattheakis, and Pavlos Protopapas. Physics-informed neural networks for quantum eigenvalue problems. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2022.

Tijs Karman and Jeremy M. Hutson. Microwave shielding of ultracold polar molecules. *Phys. Rev. Lett.*, 121:163401, Oct 2018. doi: 10.1103/PhysRevLett.121.163401. URL https://link.aps.org/doi/10.1103/PhysRevLett.121.163401.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, Jun 2021. ISSN 2522-5820. doi: 10.1038/s42254-021-00314-5. URL https://doi.org/10.1038/s42254-021-00314-5.

Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006. doi: 10.1137/S0097539704445226.

René P. Klausen, Ivan Timofeev, Johannes Frank, Jonas Naujoks, Thomas Wiegand, Sebastian Lapuschkin, and Wojciech Samek. Liesolver: A pde-constrained solver for ibvps using lie symmetries, 2025. URL https://arxiv.org/abs/2510.25731.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International conference on machine learning*, pp. 2747–2755. PMLR, 2018.

Ruitian Li, Xuan Luo, Hao Sun, and Pablo G. Ortega. Solving two and three-body systems with deep neural networks, 2025. URL https://arxiv.org/abs/2507.17559.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, Anima Anandkumar, et al. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Yu Liu, Ming-Guang Hu, Matthew A. Nichols, David D. Grimes, Tijs Karman, Hua Guo, and Kang-Kuen Ni. Photo-excitation of long-lived transient intermediates in ultracold reactions. *Nature Physics*, 16(11):1132–1136, November 2020. ISSN 1745-2473. doi: 10.1038/s41567-020-0968-8. URL http://doi.org/10.1038/s41567-020-0968-8.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

J. K. L. MacDonald. Successive approximations by the rayleigh-ritz variation method. *Phys. Rev.*, 43: 830–833, May 1933. doi: 10.1103/PhysRev.43.830. URL https://link.aps.org/doi/10.1103/PhysRev.43.830.

A. Micheli, G. Pupillo, H. P. Büchler, and P. Zoller. Cold polar molecules in two-dimensional traps: Tailoring interactions with external fields for novel quantum phases. *Phys. Rev. A*, 76:043604, Oct 2007. doi: 10.1103/PhysRevA.76.043604. URL https://link.aps.org/doi/10.1103/PhysRevA.76.043604.

Steven A. Moses, Jacob P. Covey, Matthew T. Miecnikowski, Deborah S. Jin, and Jun Ye. New frontiers for quantum gases of polar molecules. *Nature Physics*, 13(1):13–20, Jan 2017. ISSN 1745-2481. doi: 10.1038/nphys3985. URL https://doi.org/10.1038/nphys3985.

Austin M. Mroz, Annabel R. Basford, Friedrich Hastedt, Isuru Shavindra Jayasekera, Irea Mosquera-Lois, Ruby Sedgwick, Pedro J. Ballester, Joshua D. Bocarsly, Ehecatl Antonio del Río Chanona, Matthew L. Evans, Jarvist M. Frost, Alex M. Ganose, Rebecca L. Greenaway, King Kuok (Mimi) Hii, Yingzhen Li, Ruth Misener, Aron Walsh, Dandan Zhang, and Kim E. Jelfs. Cross-disciplinary perspectives on the potential for artificial intelligence across chemistry. *Chem. Soc. Rev.*, 54:5433–5469, 2025. doi: 10.1039/D5CS00146C. URL http://dx.doi.org/10.1039/D5CS00146C.

Yusuke Nomura. Boltzmann machines and quantum many-body problems. *Journal of Physics: Condensed Matter*, 36(7):073001, nov 2023. doi: 10.1088/1361-648X/ad0916. URL https://doi.org/10.1088/1361-648X/ad0916.

S. Ospelkaus, K.-K. Ni, D. Wang, M. H. G. de Miranda, B. Neyenhuis, G. Quéméner, P. S. Julienne, J. L. Bohn, D. S. Jin, and J. Ye. Quantum-State Controlled Chemical Reactions of Ultracold Potassium-Rubidium Molecules. *Science*, 327(5967):853–857, 2010. ISSN 0036-8075. doi: 10.1126/science.1184121. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1184121.

David Pfau, James S. Spencer, Alexander G. D. G. Matthews, and W. M. C. Foulkes. Ab initio solution of the many-electron schrödinger equation with deep neural networks. *Phys. Rev. Res.*, 2:033429, Sep 2020. doi: 10.1103/PhysRevResearch.2.033429. URL https://link.aps.org/doi/10.1103/PhysRevResearch.2.033429.

David Pfau, Simon Axelrod, Halvard Sutterud, Ingrid von Glehn, and James S. Spencer. Accurate computation of quantum excited states with neural networks. *Science*, 385(6711):eadn0137, 2024. doi: 10.1126/science.adn0137. URL https://www.science.org/doi/abs/10.1126/science.adn0137.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, New York, NY, 2 edition, 2004. ISBN 978-1-4757-4145-2. doi: 10.1007/978-1-4757-4145-2.

C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23:69–89, Apr 1951. doi: 10.1103/RevModPhys.23.69.

Hiroki Saito. Method to solve quantum few-body problems with artificial neural networks. *Journal of the Physical Society of Japan*, 87(7):074002, 2018. doi: 10.7566/JPSJ.87.074002.

J. J. Sakurai and Jim Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, 3 edition, 2020. ISBN 9781108473224.

E. Schrödinger. Quantisierung als eigenwertproblem. *Annalen der Physik*, 384(6):489–527, 1926. doi: https://doi.org/10.1002/andp.19263840602.

Norbert Schuch and Frank Verstraete. Computational complexity of interacting electrons and fundamental limitations of density functional theory. *Nature Physics*, 5(10):732–735, Oct 2009. ISSN 1745-2481. doi: 10.1038/nphys1370. URL https://doi.org/10.1038/nphys1370.

T. Secker, J.-L. Li, P. M. A. Mestrom, and S. J. J. M. F. Kokkelmans. Three-body recombination calculations with a two-body mapped grid method. *Phys. Rev. A*, 103:032817, Mar 2021. doi: 10.1103/PhysRevA.103.032817. URL https://link.aps.org/doi/10.1103/PhysRevA.103.032817.

Shayamal Singh and Chris H Greene. Efimov effect in ultracold microwave-shielded polar molecules. *Physical Review Letters*, 136(23):233402, 2026.

Yasuyuki Suzuki and Kálmán Varga. *Stochastic Variational Approach to Quantum-Mechanical Few-Body Problems*, volume 54 of *Lecture Notes in Physics Monographs*. Springer, Berlin, Heidelberg, 1998. ISBN 978-3-540-65152-9. doi: 10.1007/3-540-49541-X.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Matthias Troyer and Uwe-Jens Wiese. Computational complexity and fundamental limitations to fermionic quantum monte carlo simulations. *Phys. Rev. Lett.*, 94:170201, May 2005. doi: 10.1103/PhysRevLett.94.170201. URL https://link.aps.org/doi/10.1103/PhysRevLett.94.170201.

Dmitriĭ Aleksandrovich Varshalovich, Anatolij Nikolaevič Moskalev, and Valerii Kel'manovich Khersonskii. *Quantum Theory of Angular Momentum*. World Scientific, 1988. ISBN 9789971501075. doi: 10.1142/0270.

M. L. Wall, K. R. A. Hazzard, and A. M. Rey. *Quantum Magnetism with Ultracold Molecules*, pp. 3–37. WORLD SCIENTIFIC, June 2015. ISBN 9789814678704. doi: 10.1142/9789814678704_0001. URL http://dx.doi.org/10.1142/9789814678704_0001.

Steven Weinberg. Quasiparticles and the born series. *Phys. Rev.*, 131:440–451, Jul 1963. doi: 10.1103/PhysRev.131.440. URL https://link.aps.org/doi/10.1103/PhysRev.131.440.

Julia Westermayr and Philipp Marquetand. Machine learning for electronically excited states of molecules. *Chemical Reviews*, 121(16):9873–9926, Aug 2021. ISSN 0009-2665. doi: 10.1021/acs.chemrev.0c00749. URL https://doi.org/10.1021/acs.chemrev.0c00749.

K. Willner, O. Dulieu, and F. Masnou-Seeuws. Mapped grid methods for long-range molecules and cold collisions. *The Journal of Chemical Physics*, 120(2):548–561, 01 2004. ISSN 0021-9606. doi: 10.1063/1.1630031. URL https://doi.org/10.1063/1.1630031.

Sora Yokoi, Shimpei Endo, and Hiroki Saito. Neural-network quantum states for solving few-body problems: application to efimov physics, 2026. URL https://arxiv.org/abs/2604.04435.

Xiao Yuan, Suguru Endo, Qi Zhao, Ying Li, and Simon C. Benjamin. Theory of variational quantum simulation. *Quantum*, 3:191, October 2019. ISSN 2521-327X. doi: 10.22331/q-2019-10-07-191.

APPENDIX

A EXTENDED PHYSICAL MODEL

A.1 JACOBI COORDINATES

To simplify analytical expressions that describe few-body systems, it is useful to define a reference frame with origin at the system's *center of mass (CM)*, whose coordinate and gradient are defined as

$$\boldsymbol{R}_{\text{CM}} = \sum_{j=1}^{N} \frac{m_j}{M} \boldsymbol{R}_j \quad \Rightarrow \quad \boldsymbol{\nabla}_{\boldsymbol{R}_{\text{CM}}} = \sum_{j=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{R}_j} \,, \tag{10}$$

where $M = \sum_{j=1}^{N} m_j$ is the total mass of the system, and m_j, $\boldsymbol{R}_j$ and $\boldsymbol{\nabla}_{\boldsymbol{R}_j}$ are respectively mass, coordinate and gradient of the j-th particle. The naïve definition of particle coordinates in the CM reference frame is

$$\boldsymbol{\rho}_j = \boldsymbol{R}_j - \boldsymbol{R}_{\text{CM}} \,, \tag{11}$$

which provides a set of *non-independent* vectors, since $\sum_j m_j \boldsymbol{\rho}_j = 0$. The laboratory-frame ($\boldsymbol{\nabla}_{\boldsymbol{R}_j}$) and CM-frame ($\boldsymbol{\nabla}_{\boldsymbol{\rho}_j}$) gradients are related via

$$\boldsymbol{\nabla}_{\boldsymbol{R}_j} = \frac{m_j}{M} \boldsymbol{\nabla}_{\boldsymbol{R}_{\text{CM}}} + \boldsymbol{\nabla}_{\boldsymbol{\rho}_j} - \frac{m_j}{M} \left(\sum\nolimits_{k=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{\rho}_k} \right) . \tag{12}$$

The coordinate choice (11) results in a convoluted expression for the kinetic energy (4)

$$\begin{aligned} \hat{\mathcal{H}}_{\text{kin}} = \sum_{j=1}^{N} \mathcal{E}_j^{\text{kin}} \boldsymbol{\nabla}^2_{\boldsymbol{R}_j} &= \sum_{j=1}^{N} \mathcal{E}_j^{\text{kin}} \left[\frac{m_j}{M} \boldsymbol{\nabla}_{\boldsymbol{R}_{\text{CM}}} + \boldsymbol{\nabla}_{\boldsymbol{\rho}_j} - \frac{m_j}{M} \left(\sum\nolimits_{k=1}^{N} \boldsymbol{\nabla}_{\boldsymbol{\rho}_k} \right) \right]^2 \\ &= -\frac{\hbar^2}{2M} \boldsymbol{\nabla}^2_{\boldsymbol{R}_{\text{CM}}} + \sum_{j=1}^{N} \mathcal{E}_j^{\text{kin}} \boldsymbol{\nabla}^2_{\boldsymbol{\rho}_j} + \frac{\hbar^2}{2M} \sum_{j,k=1}^{N} \left(\boldsymbol{\nabla}_{\boldsymbol{\rho}_j} \cdot \boldsymbol{\nabla}_{\boldsymbol{\rho}_k} \right) , \end{aligned} \tag{13}$$

where kinetic factors $\mathcal{E}_j^{\text{kin}} = -\hbar^2/(2m_j)$, and the mass polarization term, containing the scalar product, arises, since vectors $\boldsymbol{\rho}_j$ are not independent.

The issue with the non-independence can be resolved by considering *Jacobi coordinates* $\boldsymbol{r}_j$ (Suzuki & Varga, 1998; Jacobi, 2009). These vectors are defined recursively for $j = 1, ..., N-1$, as

$$\boldsymbol{r}_j = \frac{1}{M_j} \left(\sum\nolimits_{k=1}^{j} m_k \boldsymbol{R}_k \right) - \boldsymbol{R}_{j+1} \,, \tag{14}$$

where the Jacobi masses $M_j = \sum_{k=1}^{j} m_k$, and the last vector $\boldsymbol{r}_N \equiv \boldsymbol{R}_{\text{CM}}$. Intuitively, the Jacobi method recursively replaces a pair of particles with a virtual particle placed at their center of mass. The process is repeated, until one reaches the overall center of mass of an N-particle system.

The kinetic energy (4) of the system can be expressed in Jacobi coordinates (14) the following way:

$$\hat{\mathcal{H}}_{\text{kin}} = -\frac{\hbar^2}{2M} \boldsymbol{\nabla}^2_{\boldsymbol{R}_{\text{CM}}} - \sum_{j=1}^{N-1} \frac{\hbar^2}{2\mu_j} \boldsymbol{\nabla}^2_{\boldsymbol{r}_j} \,, \tag{15}$$

where the reduced masses $\mu_j = m_{j+1} M_j / M_{j+1}$. Here, the first term is the kinetic energy of the center of mass, which is usually omitted, and the second term is the total kinetic energy of particles in the CM reference frame. Ultimately, the generic N-particle Hamiltonian can be approximated by

$$\hat{\mathcal{H}} = -\sum_{j=1}^{N-1} \frac{\hbar^2}{2\mu_j} \boldsymbol{\nabla}^2_{\boldsymbol{r}_j} + \sum_{i<j}^{N} V(\tilde{\boldsymbol{R}}_i - \tilde{\boldsymbol{R}}_j) \,, \tag{16}$$

where $\tilde{\boldsymbol{R}}_{i,j}$ need to be treated as functions of Jacobi coordinates $\boldsymbol{r}_i$.

Consequently, the wave function factorizes as

$$\Psi(\boldsymbol{R}_1, ..., \boldsymbol{R}_N) = \Psi_{\text{CM}}(\boldsymbol{R}_{\text{CM}}) \, \Psi_{\text{internal}}(\boldsymbol{r}_1, ..., \boldsymbol{r}_{N-1}) \,. \tag{17}$$

The center-of-mass factor describes free motion of the entire system and does not affect the intrinsic bound-state structure. Discarding $\boldsymbol{R}_{\text{CM}}$ is therefore equivalent to working in the intrinsic frame, or quotienting out the global translation symmetry. In our PINN formulation, the neural network represents only the internal wave function $\Psi_{\text{internal}}(\boldsymbol{r}_1, ..., \boldsymbol{r}_{N-1})$, reducing the input dimension from $3N$ to $3(N-1)$ while enforcing translational invariance exactly.

A.2 Microwave-shielded interaction potential

From the physics perspective, we are interested in ultracold diatomic molecules, hereafter referred to as monomers, that have a large permanent electric dipole moment d. The dipole moment is measured in the body-fixed frame, but it averages to zero in the laboratory frame. To get a net dipole moment in the laboratory frame, one can couple the transition (with frequency ω_0) from the ground to the first excited rotational state of the monomer, to a microwave field (with frequency ω). The coupling strength is characterized by the Rabi frequency $\Omega = 2\pi\nu$, defined as the product of the dipole moment d and the amplitude of the microwave electric field F_{mw}. For convenience, we also define here the rescaled detuning $\delta = (\omega_0 - \omega)/\Omega$ by the Rabi frequency. Not only does the microwave-field-to-molecule coupling induce a net dipole moment $d_{\text{eff}} = d/\sqrt{12(1+\delta^2)}$ in the dressed monomer state, but also—more importantly—a repulsive core in the dipole–dipole interaction potential between two monomers. This, so-called, microwave shielding (Karman & Hutson, 2018) is crucial for maintaining the stability of ultracold dipolar molecules, which would otherwise undergo significant losses when they encounter each other at short range. The adiabatic shielding potential between two dressed monomers can be given in a simple closed form (Deng et al., 2023)

$$V(\boldsymbol{r}) = C_6[1 - \cos^4\theta]\, r^{-6} + C_3[3\cos^2\theta - 1]\, r^{-3}\,, \tag{18}$$

where $C_3 = d_{\text{eff}}^2/(4\pi\epsilon_0) = d^2/[48\pi\epsilon_0(1+\delta^2)]$, $C_6 = d^4/[128\pi^2\epsilon_0^2\hbar\Omega(1+\delta^2)^{3/2}]$ and $\boldsymbol{r} = \{r, \theta, \phi\}$ is the relative interparticle in spherical coordinates. Note that the potential can also depend on ϕ, when one turns on the ellipticity $\xi \neq 0$ of the field (Deng et al., 2023). The sign of δ plays no role in Eq. (18), however, in practice, we do need a minus sign (blue detuning) for it. In our study, we can simply set the microwave field to be on resonance: $\delta = 0$.

For numerical calculations, it is instrumental to use the *atomic unit (a.u.) system*, in which $\hbar = 1$ and $4\pi\epsilon_0 = 1$. Using these units, the interaction potential reads

$$V_{\text{eff}}^{\text{a.u.}}(\boldsymbol{r}) = C_6^{\text{a.u.}}[1 - \cos^4\theta]\, r^{-6} + C_3^{\text{a.u.}}[3\cos^2\theta - 1]\, r^{-3}, \tag{19}$$

where $C_6^{\text{a.u.}} = (d^{\text{a.u.}})^4/(8\Omega^{a.u.})$ and $C_3^{\text{a.u.}} = (d^{\text{a.u.}})^2/12$. Note, that we already set $\delta = 0$. The actual physical parameters for ^{23}Na^{40}K in SI units:

$$m = 62.95376748\,\text{Da}\,, \qquad d = 2.72\,\text{D}\,, \qquad \Omega/(2\pi) = \nu \in [16, 50]\,\text{MHz}\,.$$

Conversion from SI units to a.u.:

$$[1\,\text{Da}]_{\text{a.u.}} = 1822.888486\,[m_e]_{\text{a.u.}}\,, \qquad [1\,\text{D}]_{\text{a.u.}} = 0.3934303\,[ea_0]_{\text{a.u.}}\,,$$
$$[1\,\text{MHz}]_{\text{a.u.}} = 1.5198298460570 \times 10^{-10}\,[E/h]_{\text{a.u.}}\,.$$

A.3 Two-body system

For a system with $N = 2$ molecules, there is a single Jacobi coordinate (see App. A.1). Namely, the relative interparticle position: $\boldsymbol{r} = \boldsymbol{R}_1 - \boldsymbol{R}_2$. In this setting, the Hamiltonian (4) reads

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2\mu}\boldsymbol{\nabla}_{\boldsymbol{r}}^2 + V(\boldsymbol{r})\,, \tag{20}$$

where $\mu = m_1 m_2/(m_1 + m_2) = m/2$ is the reduced mass.

A.3.1 Cylindrical coordinates

Since the interaction potential (19) does not depend on the spherical angle ϕ, the straightforward way to solve the Hamiltonian (20) is by using reduced cylindrical coordinates $\boldsymbol{r} = \{r_\perp, z\}$ with the angle ϕ omitted. In particular, one considers the following wave function

$$\Psi(\boldsymbol{r}) = \psi_{\text{2D}}(r_\perp, z)/\sqrt{r_\perp}\,, \tag{21}$$

for which the kinetic energy contribution to the Hamiltonian reads

$$\hat{\mathcal{H}}_{\text{kin}} = -\frac{\hbar^2}{2\mu}\boldsymbol{\nabla}_{\boldsymbol{r}}^2 = -\frac{\hbar^2}{2\mu}\left(\boldsymbol{\nabla}_{\boldsymbol{r}_\perp}^2 + \boldsymbol{\nabla}_{\boldsymbol{z}}^2 + \frac{1}{4r_\perp^2}\right), \tag{22}$$

and the potential can be obtained by substituting into Eq. (19)

$$r = \sqrt{r_\perp^2 + z^2} \quad \text{and} \quad \cos\theta = z/r\,. \tag{23}$$

The design of the optimization procedure for wave function (21) is thoroughly explained in App. C.

A.3.2 Spherical decomposition

To simplify the learning problem, one can analytically take care of the angular behavior of the wave function. In particular, wave function $\Psi(\boldsymbol{r})$ and the interaction potential $V(\boldsymbol{r})$ can be expanded over the complete basis of spherical harmonics

$$\Psi(\boldsymbol{r}) = \sum_{l,m} \psi_{l,m}(r) Y_{l,m}(\theta,\phi)/r \quad \text{and} \quad V(\boldsymbol{r}) = \sum_{l,m} V_{l,m}(r) Y_{l,m}(\theta,\phi)\,, \tag{24}$$

where $\boldsymbol{r} = \{r,\theta,\phi\}$ are spherical coordinates and $Y_{l,m}(\theta,\phi)$, where $l = 0,1,2,...$ and $m \in \{-l,l\}$, are spherical harmonics with the Condon-Shortley phase (Varshalovich et al., 1988). The spherical components $V_{l,m}(r)$ of the interaction potential (5) can be explicitly calculated

$$V_{0,0}(r) = \frac{8\sqrt{\pi}C_6}{5}\frac{1}{r^6}\,,\; V_{2,0}(r) = 4\sqrt{\frac{\pi}{5}}\left(C_3\frac{1}{r^3} - \frac{2C_6}{7}\frac{1}{r^6}\right),\; V_{4,0}(r) = -\frac{16\sqrt{\pi}C_6}{105}\frac{1}{r^6}\,. \tag{25}$$

Expansions (24) allow us to remove angles $\{\theta,\phi\}$ from the numerical optimization problem. After applying these expansions, the energy functional (2) can be written down as

$$\mathcal{E}_{\boldsymbol{w}}[\hat{\mathcal{H}}] \sim \int \Psi_{\boldsymbol{w}}^*(\boldsymbol{r})\hat{\mathcal{H}}\Psi_{\boldsymbol{w}}(\boldsymbol{r})\mathrm{d}\boldsymbol{r} = \int \sum_{l,m;l',m'} \psi_{\boldsymbol{w},l',m'}^*(r)\hat{\mathcal{H}}_{l,m;l',m'}(r)\psi_{\boldsymbol{w},l,m}(r)\mathrm{d}r\,, \tag{26}$$

where $\boldsymbol{w}$ are neural network parameters, and the cross-channel Hamiltonian

$$\hat{\mathcal{H}}_{l,m;l',m'}(r) \equiv \hat{\mathcal{H}}^{\text{kin}}_{l,m;l',m'}(r) + \mathcal{H}^{\text{pot}}_{l,m;l',m'}(r)\,. \tag{27}$$

The cross-channel potential Hamiltonian reads

$$\mathcal{H}^{\text{pot}}_{l,m;l',m'}(r) \equiv \sum_{L,M} V_{L,M}(r) G^{L,M}_{l,m;l',m'}\,, \tag{28}$$

where the Gaunt coefficients are defined as integrals

$$G^{L,M}_{l,m;l',m'} = \int Y^*_{l',m'}(\Omega) Y_{L,M}(\Omega) Y_{l,m}(\Omega)\,\mathrm{d}\Omega = \sqrt{\frac{(2l+1)(2L+1)}{4\pi(2l'+1)}} C^{l',m'}_{l,m;L,M} C^{l',0}_{l,0;L,0}\,, \tag{29}$$

where $C^{l',m'}_{l,m;L,M}$ are the Clebsch–Gordan coefficients that encapsulate angular-momentum algebra. The cross-channel kinetic Hamiltonian is defined as

$$\hat{\mathcal{H}}^{\text{kin}}_{l,m;l',m'}(r) = \hat{\mathcal{H}}^{\text{kin,rad}}_{l,m;l',m'} + \mathcal{H}^{\text{kin,cent}}_{l,m;l',m'}(r)\,, \tag{30}$$

with the diagonal radial derivative and centrifugal terms defined as

$$\hat{\mathcal{H}}^{\text{kin,rad}}_{l,m;l',m'} = -\delta_{l,l'}\delta_{m,m'}\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} \quad \text{and} \quad \mathcal{H}^{\text{kin,cent}}_{l,m;l',m'}(r) = \delta_{l,l'}\delta_{m,m'}\frac{\hbar^2}{2\mu}\frac{l(l+1)}{r^2}\,, \tag{31}$$

where $\delta_{l,l'}$ is the Kronecker delta. One has to be careful, when implementing the operator $\hat{\mathcal{H}}^{\text{kin,rad}}_{l,m;l',m'}$ within a neural network. Its expectation value needs to be integrated by parts, in order to change the overall sign and provide correct motivation within the loss function.

A.3.3 MAPPED DISCRETE VARIABLE REPRESENTATION (MDVR)

As a conventional numerical solver for two-molecule systems, used as benchmark for our PINN, we extend the mapped discrete variable representation (MDVR) (Willner et al., 2004; Secker et al., 2021) to include multiple partial-wave couplings. The MDVR approach defines a map function: $x(r) = \frac{\sqrt{m}}{p_{\max}} \int_{r_{\text{in}}}^{r} dr' \sqrt{E_{\max} - V_{\text{env}}(r')}$, such that a uniformly spaced grid in the x-representation corresponds to a denser r-grid, where the local de Broglie wavelength $(1/\sqrt{E - V(r)})$ is small, and a coarser r-grid where it is large. Here $E_{\max}$ denotes the maximum energy considered, and $p_{\max}$ is the corresponding maximum local momentum. The envelope potential $V_{\text{env}}(r)$ is set as the lowest ($l = 0$ channel) adiabatic potential after diagonalizing $V^{\text{eff}}_{ll'}(r) + \hbar^2 l(l+1)/(2\mu r^2)$. After the mapping, the remaining procedure in x-representation is standard:

1. Construct the Fourier DVR basis set $\{\tilde{s}_k(x)\}_{k=1}^{N-1}$ from uniformly spaced x-grid points $\{x_k\}_{k=0}^{N}$. Two endpoints are removed to impose a hard-wall boundary condition.
2. Build up the Hamiltonian matrix in this basis set: $H_{ij} = T_{ij}\delta_{ll'} + V^{\text{eff}}_{ll'}[r(x_i)]\delta_{ij}$.
3. Diagonalize the Hamiltonian matrix to get the wave function in x-representation: $\varphi_n(x) = \sum_{i,l} c_n^{i,l} \tilde{s}_i(x)/\sqrt{\Delta x}$, where $c_n^{i,l}$ is taken from the n-th eigenvector of the Hamiltonian matrix and $\Delta x = x_i - x_{i-1}$.
4. Map the x-representation wave function back to r-representation via: $\psi(r) = \varphi[x(r)]/\sqrt{J(x)}$, where $J(x) = \partial r(x)/\partial_x$ denotes the Jacobian of the $r \to x$ transformation.

We refer the reader to Willner et al. (2004) for the explicit expression of the Fourier DVR basis $\{\tilde{s}_k(x)\}$ and of the kinetic matrix element T_{ij}. The MDVR can significantly reduce the numerical cost compared to the DVR using uniformly spaced grids in r representation (Willner et al., 2004). We use $N = 512$ DVR grids and $l \leq 8$ for two-molecule calculations.

A.4 THREE-BODY SYSTEM

For a system of $N = 3$ particles of the same mass m at positions $\{\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3\}$, the Jacobi coordinates (see App. A.1) are defined as

$$\boldsymbol{r}_1 = \boldsymbol{R}_1 - \boldsymbol{R}_2\,, \qquad \boldsymbol{r}_2 = \frac{\boldsymbol{R}_1 + \boldsymbol{R}_2}{2} - \boldsymbol{R}_3\,, \qquad \boldsymbol{r}_3 = \boldsymbol{R}_{\text{CM}} = \frac{\boldsymbol{R}_1 + \boldsymbol{R}_2 + \boldsymbol{R}_3}{3}\,. \tag{32}$$

The pair-separation vectors can be expressed as

$$\boldsymbol{R}_1 - \boldsymbol{R}_2 = \boldsymbol{r}_1\,, \qquad \boldsymbol{R}_2 - \boldsymbol{R}_3 = \boldsymbol{r}_2 - \frac{\boldsymbol{r}_1}{2}\,, \qquad \boldsymbol{R}_3 - \boldsymbol{R}_1 = -\boldsymbol{r}_2 - \frac{\boldsymbol{r}_1}{2}\,. \tag{33}$$

In the absence of an external potential, the free center-of-mass motion that depends on $\boldsymbol{R}_{\text{CM}}$ can be trivially described by a plane wave, leaving only the relative degrees of freedom in the Hamiltonian:

$$\hat{\mathcal{H}}_{\text{3-body}} = -\frac{\hbar^2}{m}\nabla^2_{\boldsymbol{r}_1} - \frac{3\hbar^2}{4m}\nabla^2_{\boldsymbol{r}_2} + V(\boldsymbol{r}_1) + V(\boldsymbol{r}_2 - \boldsymbol{r}_1/2) + V(-\boldsymbol{r}_2 - \boldsymbol{r}_1/2)\,. \tag{34}$$

We implement the spherical harmonic expansion (see App. A.3.2) of the wave function $\Psi_{\text{3-body}}(\boldsymbol{r}_1, \boldsymbol{r}_2)$ on both $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$:

$$\Psi_{\text{3-body}}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{1}{r_1 r_2} \sum_{l,m_l;L,M_L} \psi^{\text{2D}}_{l,m_l;L,M_L}(r_1, r_2)\left[Y_{l,m_l}(\boldsymbol{r}_1) \otimes Y_{L,M_L}(\boldsymbol{r}_2)\right], \tag{35}$$

where $\psi^{\text{2D}}_{l,m_l;L,M_L}(r_1, r_2)$ is the 2D radial wave function, discussed in detail in App. C, and $[Y_{l,m_l}(\boldsymbol{r}_1) \otimes Y_{L,M_L}(\boldsymbol{r}_2)]$ is the tensor product of spherical harmonics (see Varshalovich et al. (1988)). The purpose of the expansion (35), just like in the two-body case, is to eliminate all angular coordinates analytically, leaving only the two radial coordinates $\{r_1, r_2\}$ for the numerical optimization. The overall logic is straightforward. The single-coordinate interaction $V(\boldsymbol{r}_1)$ is treated exactly as in App. A.3.2. The remaining pair interactions, $V(\boldsymbol{r}_2 - \boldsymbol{r}_1/2)$ and $V(-\boldsymbol{r}_2 - \boldsymbol{r}_1/2)$, depend on linear combinations of the two Jacobi coordinates and therefore couple their angular momenta. We evaluate these terms by re-expanding their angular dependence in products of spherical harmonics of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, while expanding the remaining dependence on their relative angle in Legendre polynomials. The

resulting angular integrals reduce to standard Clebsch–Gordan and Gaunt coefficients, yielding the matrix elements of the Hamiltonian that depend only on $\{r_1, r_2\}$:

$$\begin{aligned}\mathcal{H}_{l,m_l;L,M_L\to l',m_l';L',M_L'}(r_1,r_2) = \\ \mathcal{H}^{\text{kin}}_{l,L}(r_1,r_2)\,\delta_{l,l'}\delta_{m_l,m_l'}\delta_{L,L'}\delta_{M_L,M_L'} + V_{l,l';m}(r_1)\delta_{L,L'}\delta_{M_L,M_L'}\delta_{m,m'} \\ + V^{(23)}_{l,m_l;L,M_L\to l',m_l';L',M_L'}(r_1,r_2) + V^{(31)}_{l,m_l;L,M_L\to l',m_l';L',M_L'}(r_1,r_2)\,,\end{aligned} \tag{36}$$

where $V_{l,l';m}(r_1) \equiv \mathcal{H}^{\text{pot}}_{l,m;l',m'}(r_1)$ from Eq. (28), and the kinetic Hamiltonian is

$$\mathcal{H}^{\text{kin}}_{l,L}(r_1,r_2) = -\frac{\hbar^2}{m}\left(\frac{d^2}{dr_1^2} - \frac{l(l+1)}{r_1^2}\right) - \frac{3\hbar^2}{4m}\left(\frac{d^2}{dr_2^2} - \frac{L(L+1)}{r_2^2}\right). \tag{37}$$

To separate the angular dependence carried by the two Jacobi coordinates, we re-expand the spherical functions of $\boldsymbol{\rho} = \boldsymbol{r}_2 - \boldsymbol{r}_1/2$ into coupled products of spherical harmonics of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. This way, the interaction matrix associated with the pair-separation $\boldsymbol{\rho} = \boldsymbol{R}_2 - \boldsymbol{R}_3$ can be written down as

$$\begin{aligned}&V^{(23)}_{l,m_l;L,M_L\to l',m_l';L',M_L'}(r_1,r_2) = \langle l,m_l;L,M_L\,|V(\boldsymbol{\rho})|\,l',m_l';L',M_L'\rangle \\ &= \sum_{\lambda,\lambda',m_\lambda}\left\langle l,m_l;L,M_L\left|Y_{\lambda,\mu}\Big(\boldsymbol{r}_2-\frac{\boldsymbol{r}_1}{2}\Big)\,V_{\lambda,\lambda';\mu}(\rho)\,Y^*_{\lambda',\mu}\Big(\boldsymbol{r}_2-\frac{\boldsymbol{r}_1}{2}\Big)\right|l'm_l'L'm_L'\right\rangle \\ &= \sum_{\substack{\lambda,\lambda',\mu \\ \lambda_{r_1}+\lambda_{r_2}=\lambda;\lambda'_{r_1}+\lambda'_{r_2}=\lambda'}} (-1)^{\lambda_{r_1}+\lambda'_{r_1}} {r_2}^{\lambda_{r_2}+\lambda'_{r_2}}\left(\frac{r_1}{2}\right)^{\lambda_{r_1}+\lambda'_{r_1}}\times \\ &\sqrt{\tfrac{4\pi(2\lambda+1)!}{(2\lambda_{r_2}+1)!(2\lambda_{r_1}+1)!}}\sqrt{\tfrac{4\pi(2\lambda'+1)!}{(2\lambda'_{r_2}+1)!(2\lambda'_{r_1}+1)!}}\sum_{m_{r_2},m_{r_1};m'_{r_2},m'_{r_1}} C^{\lambda,\mu}_{\lambda_{r_2},m_{r_2};\lambda_{r_1},m_{r_1}} C^{\lambda,'\mu}_{\lambda'_{r_2},m'_{r_2};\lambda'_{r_1},m'_{r_1}}\times \\ &\left\langle l,m_l;L,M_L\left|Y_{\lambda_{r_2},m_{r_2}}(\boldsymbol{r}_2)Y_{\lambda_{r_1},m_{r_1}}(\boldsymbol{r}_1)\tfrac{V_{\lambda,\lambda';\mu}(\rho)}{\rho^{\lambda+\lambda'}}Y^*_{\lambda'_{r_2},m'_{r_2}}(\boldsymbol{r}_2)Y^*_{\lambda'_{r_1},m'_{r_1}}(\boldsymbol{r}_1)\right|l',m_l';L',M_L'\right\rangle,\end{aligned} \tag{38}$$

where $|l,m_l;L,M_L\rangle \sim \big[Y_{l,m_l}(\boldsymbol{r}_1)\otimes Y_{L,M_L}(\boldsymbol{r}_2)\big]$. After expansion (38), the remaining coupling between $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ enters only through their relative angle $\xi = (\boldsymbol{r}_1\cdot\boldsymbol{r}_2)/(r_1r_2)$, embedded in the definition of $\rho = \sqrt{r_2^2 + r_1^2/4 - r_1r_2\xi}$. We further expand the corresponding scalar radial factor

$$\frac{V_{\lambda,\lambda';\mu}(\rho(\xi))}{\rho(\xi)^{\lambda+\lambda'}} = \sum\nolimits_{\kappa,\nu}\mathcal{F}^{\kappa}_{\lambda,\lambda';\mu}(r_1,r_2)Y_{\kappa,\nu}(\boldsymbol{r}_2)Y^*_{\kappa,\nu}(\boldsymbol{r}_1), \tag{39}$$

where

$$\mathcal{F}^{\kappa}_{\lambda,\lambda';\mu}(r_1,r_2) = 2\pi\int_{-1}^{1}\frac{V_{\lambda,\lambda';\mu}\Big(\sqrt{r_2^2+r_1^2/4-r_1r_2\xi}\Big)}{(r_2^2+r_1^2/4-r_1r_2\xi)^{(\lambda+\lambda')/2}}P_\kappa(\xi)\,\mathrm{d}\xi\,, \tag{40}$$

where $P_\kappa(\xi)$ is the Legendre polynomial of degree κ. For the angular integral, we obtain

$$\begin{aligned}&\left\langle l,m_l;L,M_L\left|Y_{\lambda_{r_2},m_{r_2}}(\boldsymbol{r}_2)Y_{\lambda_{r_1},m_{r_1}}(\boldsymbol{r}_1)\frac{V_{\lambda,\lambda';\mu}(\rho)}{\rho^{\lambda+\lambda'}}Y^*_{\lambda'_{r_2},m'_{r_2}}(\boldsymbol{r}_2)Y^*_{\lambda'_{r_1},m'_{r_1}}(\boldsymbol{r}_1)\right|l',m_l';L',M_L'\right\rangle \\ &\qquad = \sum_{\kappa,\nu}\mathcal{F}^{\kappa}_{\lambda,\lambda';\mu}(r_1,r_2)\left\langle L,M_L\left|Y_{\lambda_{r_2},m_{r_2}}(\boldsymbol{r}_2)Y_{\kappa,\nu}(\boldsymbol{r}_2)Y^*_{\lambda'_{r_2},m'_{r_2}}(\boldsymbol{r}_2)\right|L',M_L'\right\rangle \\ &\qquad\qquad\times\left\langle l,m_l\left|Y_{\lambda_{r_1},m_{r_1}}(\boldsymbol{r}_1)Y^*_{\kappa,\nu}(\boldsymbol{r}_1)Y^*_{\lambda'_{r_1},m'_{r_1}}(\boldsymbol{r}_1)\right|l',m_l'\right\rangle,\end{aligned} \tag{41}$$

And finally, for the matrix elements of the potential

$$V^{(23)}_{l,m_l;L,M_L\to l',m'_l;L',M'_L}(r_1,r_2) = \sum_{\substack{\lambda,\lambda',\mu \\ \lambda_{r_1}+\lambda_{r_2}=\lambda;\lambda'_{r_1}+\lambda'_{r_2}=\lambda'}} (-1)^{\lambda_{r_1}+\lambda'_{r_1}} r_2^{\lambda_{r_2}+\lambda'_{r_2}} \left(\frac{r_1}{2}\right)^{\lambda_{r_1}+\lambda'_{r_1}}$$
$$\sqrt{\tfrac{4\pi(2\lambda+1)!}{(2\lambda_{r_2}+1)!(2\lambda_{r_1}+1)!}}\sqrt{\tfrac{4\pi(2\lambda'+1)!}{(2\lambda'_{r_2}+1)!(2\lambda'_{r_1}+1)!}} \sum_{\substack{m_{r_2},m_{r_1} \\ m'_{r_2},m'_{r_1}}} C^{\lambda,\mu}_{\lambda_{r_2},m_{r_2};\lambda_{r_1},m_{r_1}} C^{\lambda',\mu}_{\lambda'_{r_2},m'_{r_2};\lambda'_{r_1},m'_{r_1}}$$
$$\times \sum_{\kappa,\nu} \mathcal{F}^{\kappa}_{\lambda,\lambda';\mu}(r_1,r_2)(-1)^{m'_{r_2}+m'_{r_1}+\nu}$$
$$\times \sum_{j_{r_2},K_{r_2}} G^{j_{r_2},m_{r_2}+\nu}_{\lambda_{r_2},m_{r_2};\kappa,\nu} G^{K_{r_2},m_{r_2}+\nu-m'_{r_2}}_{j_{r_2},m_{r_2}+\nu;\lambda'_{r_2},-m'_{r_2}} G^{L,M_L}_{K_{r_2},m_{r_2}+\nu-m'_{r_2};L',M'_L}$$
$$\times \sum_{j_{r_1},K_{r_1}} G^{j_{r_1},m_{r_1}-\nu}_{\lambda_{r_1},m_{r_1};\kappa,-\nu} G^{K_{r_1},m_{r_1}-\nu-m'_{r_1}}_{j_{r_1},m_{r_1}-\nu;\lambda'_{r_1},-m'_{r_1}} G^{l,m_l}_{K_{r_1},m_{r_1}-\nu-m'_{r_1};l',m'_l}\,, \tag{42}$$

where $G^{L,M}_{l,m;l',m'}$ are Gaunt coefficients (29). Similarly, we obtain matrix elements of the other pair interaction

$$V^{(31)}_{l,m_l;L,M_L\to l',m'_l;L',M'_L}(r_1,r_2) = \left\langle l,m_l;L,M_L \middle| V\Big(-\boldsymbol{r}_2-\frac{\boldsymbol{r}_1}{2}\Big)\middle| l',m'_l;L',M'_L\right\rangle$$
$$= \sum_{\lambda,\lambda',\mu} \sum_{\substack{\lambda_{r_2}+\lambda_{r_1}=\lambda \\ \lambda'_{r_2}+\lambda'_{r_1}=\lambda'}} (-1)^{\lambda+\lambda'} r_2^{\lambda_{r_2}+\lambda'_{r_2}} \left(\frac{r_1}{2}\right)^{\lambda_{r_1}+\lambda'_{r_1}} \sqrt{\tfrac{4\pi(2\lambda+1)!}{(2\lambda_{r_2}+1)!(2\lambda_{r_1}+1)!}}\sqrt{\tfrac{4\pi(2\lambda'+1)!}{(2\lambda'_{r_2}+1)!(2\lambda'_{r_1}+1)!}}$$
$$\times \sum_{\substack{m_{r_2},m_{r_1} \\ m'_{r_2},m'_{r_1}}} C^{\lambda,\mu}_{\lambda_{r_2},m_{r_2};\lambda_{r_1},m_{r_1}} C^{\lambda',\mu}_{\lambda'_{r_2},m'_{r_2};\lambda'_{r_1},m'_{r_1}} \sum_{\kappa,\nu} \mathcal{F}^{\kappa,+}_{\lambda,\lambda';\mu}(r_1,r_2)(-1)^{m'_{r_2}+m'_{r_1}+\nu}$$
$$\times \sum_{j_{r_2},K_{r_2}} G^{j_{r_2},m_{r_2}+\nu}_{\lambda_{r_2},m_{r_2};\kappa,\nu} G^{K_{r_2},m_{r_2}+\nu-m'_{r_2}}_{j_{r_2},m_{r_2}+\nu;\lambda'_{r_2},-m'_{r_2}} G^{L,M_L}_{K_{r_2},m_{r_2}+\nu-m'_{r_2};L',M'_L}$$
$$\times \sum_{j_{r_1},K_{r_1}} G^{j_{r_1},m_{r_1}-\nu}_{\lambda_{r_1},m_{r_1};\kappa,-\nu} G^{K_{r_1},m_{r_1}-\nu-m'_{r_1}}_{j_{r_1},m_{r_1}-\nu;\lambda'_{r_1},-m'_{r_1}} G^{l,m_l}_{K_{r_1},m_{r_1}-\nu-m'_{r_1};l',m'_l}\,. \tag{43}$$

Note that in this case, the radial part $\mathcal{F}^{\kappa,+}_{\lambda,\lambda';\mu}(r_1,r_2)$ is now given by

$$\mathcal{F}^{\kappa,+}_{\lambda,\lambda';\mu}(r_1,r_2) = 2\pi \int_{-1}^{1} \frac{V_{\lambda,\lambda';\mu}\Big(\sqrt{r_2^2+r_1^2/4+r_1r_2\xi}\Big)}{(r_2^2+r_1^2/4+r_1r_2\xi)^{(\lambda+\lambda')/2}} P_\kappa(\xi)\,\mathrm{d}\xi\,. \tag{44}$$

In summary, expansion (35) eliminates the angular coordinates from the three-body optimization problem. All rotational structure and angular-momentum coupling between the two Jacobi coordinates is encoded analytically through Clebsch–Gordan and Gaunt coefficients, while the remaining Hamiltonian is a matrix-valued function (36) of the two radial coordinates $\{r_1, r_2\}$. The neural representation therefore only needs to learn the corresponding two-dimensional radial channel functions $\psi^{\mathrm{2D}}_{l,m_l;L,M_L}(r_1,r_2)$. The same construction generalizes to additional Jacobi coordinates through higher tensor products of spherical harmonics and standard angular-momentum coupling algebra.

B Extended method

B.1 Properties of the Spherical Harmonic Representation

This section summarizes several mathematical and physical properties of spherical harmonic representations that motivate their use in the proposed neural wave function parametrization. While not required for the formulation itself, these properties explain the numerical stability, scalability, and physical interpretability of the decomposition.

I. The natural basis. From the mathematical perspective, spherical harmonics $Y_{l,m}(\theta, \phi)$ form a complete orthonormal basis for L^2-integrable functions of angles $\{\theta, \phi\}$. Unlike other polynomial sets, like Hermite or Chebyshev polynomials, commonly used for numerical methods, the use of spherical harmonics is not motivated by the boundary conditions of the problem. Problem-specific functional bases are a common cause of the fragility of numerical methods: changing the interaction or boundary conditions would require redesigning the representation. In contrast, for every value of l, spherical harmonics form the standard *irreducible* representation of the SO(3) group and provide the natural, symmetry-inspired basis for functions of angles $\{\theta, \phi\}$.

II. Dealing with singularities. Spherical coordinates $\boldsymbol{r} = \{r, \theta, \phi\}$ are curvilinear. Because of that, they are singular at the origin ($r = 0$) and along the polar axes ($\theta \in \{0, \pi\}$). Differential operators, such as the Laplacian $\boldsymbol{\nabla}^2_{\boldsymbol{r}}$ that constitutes kinetic energy (4), include terms, like $1/\sin\theta$, that behave numerically unstable around singular points. This highly affects the performance of numerical methods that employ spherical coordinates. In contrast, spherical harmonics $Y_{l,m}(\theta, \phi)$ embed singularities of the coordinate system into their structure. Thus, they provide numerically stable expressions for differential operators, like

$$\boldsymbol{\nabla}^2_{\boldsymbol{r}} Y_{l,m}(\theta, \phi) \sim -l(l+1)\, Y_{l,m}(\theta, \phi)/r^2\,. \tag{45}$$

The expectation value of the kinetic energy (4) in terms of angular momentum pairs $\{l, m\}$ is diagonal: $\langle Y_{l',m'}|\hat{\mathcal{H}}_{\text{kin}}|Y_{l,m}\rangle \sim \delta_{l,l'}\delta_{m,m'}$, which substantially reduces the computational cost.

III. Scalability. For an N-body system, the angular dependence of $N-1$ relative coordinates can be represented by tensor products of spherical harmonics. Under rotations, these products transform according to tensor-product representations of SO(3), which can be decomposed into total-angular-momentum sectors using standard Clebsch–Gordan coupling. Consequently, angular-momentum coupling and transfer between different relative coordinates are handled analytically through Clebsch–Gordan and Wigner coefficients, rather than evaluated numerically. This construction extends recursively to an arbitrary number of coordinates, allowing the rotational structure of the full N-body wave function to be tracked algebraically. For example, a three-body wave function $\Psi_{\text{3-body}}(\boldsymbol{r_1}, \boldsymbol{r_2})$ can be expanded in coupled products of spherical harmonics (see Eq. (9) and App. A.4).

IV. Truncation. Truncating the expansion $\Psi(\boldsymbol{r}) = \sum_{l,m} \psi_{l,m}(r)\, Y_{l,m}(\theta, \phi)/r$, at a finite l_{max} is physically justified. Higher-l harmonics $Y_{l,m}(\theta, \phi)$ are associated with larger centrifugal terms (45) in the expression for the kinetic energy (4). Similar to the interaction potential (5), these terms represent barriers to short-range molecular collisions. A higher barrier decreases the likelihood of molecule scattering into the higher-l state, which renders these modes insignificant.

B.2 Training pseudo-code

Algorithm 1 summarizes the optimization procedure used throughout this work.

C Technical Details of the 2D Wave Function Representation

This Appendix details the two-dimensional formulation used for the two-body problem, including its numerical integration, differentiation, asymptotic envelope, and optimization.

C.1 Integration and spatial derivatives

We implement the two-coordinate wave function representation (21) in reduced cylindrical coordinates, introduced in App. A.3.1. These describe a single relative position vector $\boldsymbol{r} = \{r_\perp, z\}$, with $r = \sqrt{r_\perp^2 + z^2}$. For the azimuthally symmetric, reflection-even potential (5), we restrict numerical calculations to $r_\perp > 0$ and $z > 0$. Following App. A.3.1, we represent the rescaled wave function

$$\psi_{\text{2D}}(r_\perp, z) = \sqrt{r_\perp}\, \Psi(\boldsymbol{r})\,. \tag{46}$$

This transformation absorbs the cylindrical Jacobian into the wave function, giving

$$\langle \Psi|\Psi\rangle = 4\pi \int_0^\infty \int_0^\infty |\psi_{\text{2D}}(r_\perp, z)|^2\, \mathrm{d}r_\perp\, \mathrm{d}z\,. \tag{47}$$

The factor 4π accounts for azimuthal integration and reflection across $z = 0$.

Algorithm 1 Variational training with residual learning

Require: Hamiltonian $\hat{\mathcal{H}}$, WKB envelope $\psi_{\mathrm{WKB}}(r)$, angular cutoff $l_{\max}$, optimizer $\mathcal{O}$, reference energy $\mathcal{E}_{\mathrm{ref}}$

1: Initialize neural network parameters $\boldsymbol{\theta}$

2: Construct fixed MCMC sampling grid $\mathcal{G} = \{\boldsymbol{r}^{(i)}\}_{i=1}^{N}$ from the WKB-inspired density

3: **while** not converged **do**

4: Sample mini-batch $\mathcal{B} \subset \mathcal{G}$

5: Encode radial coordinates

$$z^{(i)} = \mathrm{PE}(r^{(i)})$$

6: Evaluate neural residual channels

$$\psi_{l,m}^{\mathrm{NN}}(r^{(i)}; \boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(z^{(i)})_{l,m}$$

7: Construct radial functions

$$\psi_{l,m}(r^{(i)}) = \psi_{l,m}^{\mathrm{NN}}(r^{(i)}; \boldsymbol{\theta})\psi_{\mathrm{WKB}}(r^{(i)})$$

8: Assemble wave function

$$\Psi(\boldsymbol{r}^{(i)}) = \sum_{l,m} \psi_{l,m}(r^{(i)}) Y_{l,m}(\theta^{(i)}, \phi^{(i)})/r^{(i)}$$

9: Evaluate normalized variational energy

$$\mathcal{L}_{\boldsymbol{\theta}} = \frac{1}{|\mathcal{E}_{\mathrm{ref}}|} \frac{\langle \Psi_{\boldsymbol{\theta}} | \hat{\mathcal{H}} | \Psi_{\boldsymbol{\theta}} \rangle}{\langle \Psi_{\boldsymbol{\theta}} | \Psi_{\boldsymbol{\theta}} \rangle}$$

10: Update parameters

$$\boldsymbol{\theta} \leftarrow \mathcal{O}\left(\boldsymbol{\theta}, \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}\right)$$

11: **end while**

Sampling. We construct the 2D sampling grid from the one-dimensional coordinate pool used in the radial formulation (Sec. 5.4). At each optimization step, coordinates are sampled independently from this pool and randomly paired to form points $\{r_{\perp,i}, z_i\}$, yielding an unstructured 2D point cloud. Points outside the prescribed radial domain $r_{\min} \leq r \leq r_{\max}$ are discarded.

Triangulation and integration. We construct a Delaunay triangulation $\mathcal{T}$ (Delaunay, 1934) of each retained point cloud. For a triangle $T \in \mathcal{T}$ with vertices $\{\boldsymbol{q}_1^T, \boldsymbol{q}_2^T, \boldsymbol{q}_3^T\}$, its area is

$$A_T = \frac{1}{2} \left| \det \left(\boldsymbol{q}_2^T - \boldsymbol{q}_1^T, \boldsymbol{q}_3^T - \boldsymbol{q}_1^T \right) \right| . \tag{48}$$

We approximate integrals using the vertex quadrature

$$\mathcal{I}_{\mathcal{T}}[f] = 4\pi \sum_{T \in \mathcal{T}} \frac{A_T}{3} \sum_{j=1}^{3} f\left(\boldsymbol{q}_j^T\right) . \tag{49}$$

This rule integrates the piecewise-linear interpolation of the sampled integrand (f) exactly. We use the same triangulation $\mathcal{T}$ for the norm, potential-energy density, and gradient-energy density. Sampling determines the spatial resolution, while triangle areas A_T determine the integration weights, no inverse sampling-density factor is used.

Spatial derivatives. While the Delaunay triangulation provides stable integration routine, differentiating the corresponding piecewise-linear interpolation is numerically unstable. We therefore decouple differentiation from the triangulation by evaluating the wave function at additional stencil points around each sampled coordinate $\boldsymbol{q}_i$. For each sampled point $\boldsymbol{q}_i \in \mathcal{T}$, the triangulation $\mathcal{T}$ is used to define a local length scale h_i as half the distance to its nearest neighbor in the Delaunay mesh:

$$h_i = \min_{j \in \mathcal{N}(i)} |\boldsymbol{q}_i - \boldsymbol{q}_j|/2 \tag{50}$$

where $\mathcal{N}(i)$ denotes the set of vertices connected to $\boldsymbol{q}_i$ by a triangulation edge.

Along each coordinate axis, we evaluate the wave function at five nearby points around each sampled coordinate. For the z-axis, these evaluations are: $\psi_{\mathrm{2D}}(r_{\perp,i}, z_i + \delta^z_{i,j})$, for $j = 1, \ldots, 5$, where the stencil size $\delta^z_{i,j}$ is determined by the local mesh spacing h_i. We approximate the local wave function by the unique fourth-order polynomial $p_i(z)$ passing through these five values and take its derivative at the central point:

$$\partial_z \psi_{\mathrm{2D}}(r_{\perp,i}, z_i) \approx p_i'(z = z_i)\,. \tag{51}$$

The $r_\perp$ derivative is computed analogously. Near the domain boundary, the stencil points are shifted inward while retaining the same polynomial differentiation procedure.

C.2 Extension of WKB to 2D

For the two-dimensional representation in reduced cylindrical coordinates $\{r_\perp, z\}$, we extend the one-dimensional WKB-inspired envelope of Sec. 5.2, so that the wave function representation is

$$\psi_{\mathrm{2D}}(r_\perp, z) = \psi^{\mathrm{NN}}_{\mathrm{2D}}(r_\perp, z)\, \mathcal{W}\!\left(\sqrt{r_\perp^2 + z^2}\right) \mathcal{W}(r_\perp)\,, \tag{52}$$

where $\mathcal{W}(r) \equiv \psi_{\mathrm{WKB}}(r, \mathcal{E}_{\mathrm{WKB}})$ is the same tempered WKB-inspired envelope used in the one-dimensional formulation (see Sec. 5.2). The radial factor $W(\sqrt{r_\perp^2 + z^2})$ imposes the physical short- and long-range asymptotics, while the additional factor $W(r_\perp)$ suppresses the wave function near the cylindrical axis $r_\perp \to 0$, where the transformed two-dimensional Hamiltonian contains a singular $1/r_\perp^2$ contribution. The product therefore serves as a regularizing asymptotic prior rather than as a direct WKB approximation to the full two-dimensional wave function.

C.3 Two-dimensional training versus radial-channel training

Both representations minimize the same physical energy functional, but differ in how much structure is supplied analytically. In the two-dimensional formulation, the network learns the joint dependence on $\{r_\perp, z\}$ through the representation (52), including the remaining angular structure. In the radial-channel formulation (8), the angular dependence is represented analytically by spherical harmonics. The network therefore learns only the radial channel functions $\psi^{\mathrm{NN}}_{l,m}(r)$, while angular kinetic and potential couplings are handled through the centrifugal terms and precomputed Gaunt coefficients.

Removing this decomposition substantially increases the numerical burden. The 2D samples must resolve both radial localization and angular variation, while evaluation of the energy additionally requires Delaunay integration and spatial derivatives along both $r_\perp$ and z. This leads to slower convergence and more expensive updates than in the radial-channel formulation. At the same time, the 2D calculation provides a controlled test in which the analytical angular decomposition is removed while retaining the same variational objective, and separately validates optimization with two continuous coordinates required by the three-body radial representation.

D EXTENDED RESULTS

D.1 WAVE FUNCTION INPUT SCALE

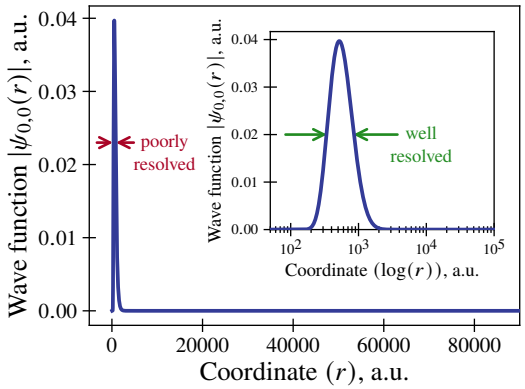

Figure 7: **Logarithmic versus linear discretization.** In linear coordinates (main), a localized peak is poorly resolved due to the large dynamic range, while logarithmic coordinates (inset) distribute resolution across scales and capture the feature more accurately.

D.2 QUALITATIVE WAVE FUNCTION RESULTS

Figures 8 and 9 compare the learned wave functions against MDVR reference solutions across interaction regimes for ground and excited states. The model accurately reproduces the overall spatial structure, asymptotic decay, and excited-state nodal behavior across the different spherical channels.

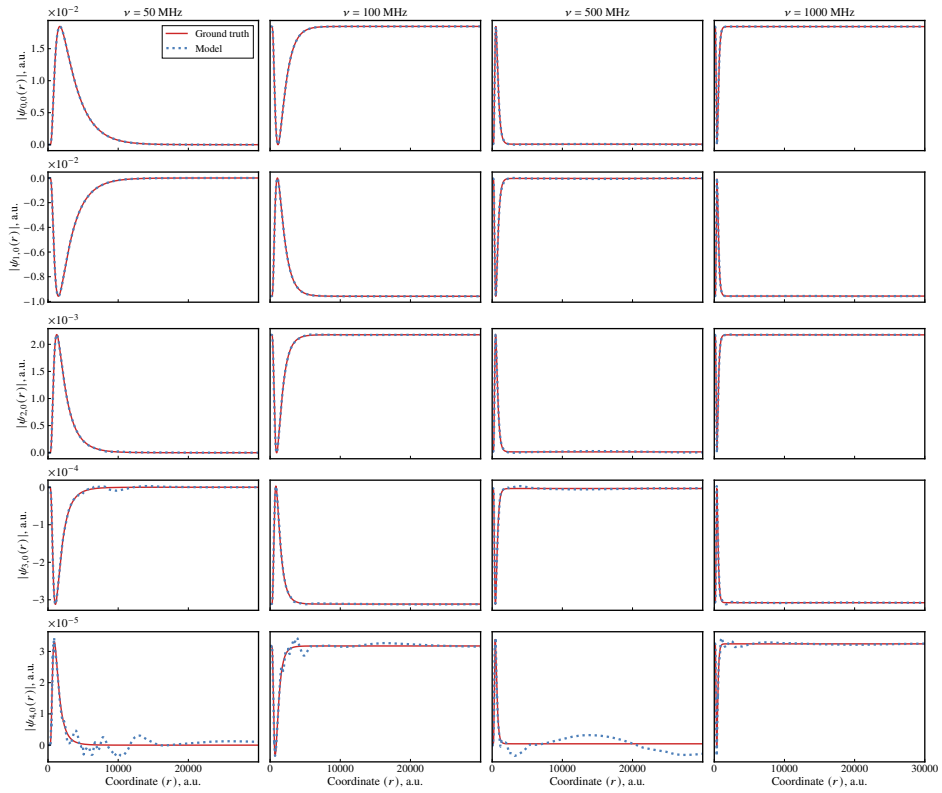

Figure 8: **Ground-state wave functions across interaction strengths ν.** Solid lines denote MDVR reference solutions and dashed lines denote the neural predictions. The model accurately captures both localized and spatially extended regimes across all spherical channels. Larger discrepancies appear primarily in higher-order spherical channels, whose amplitudes are substantially smaller and therefore contribute only weakly to the total energy.

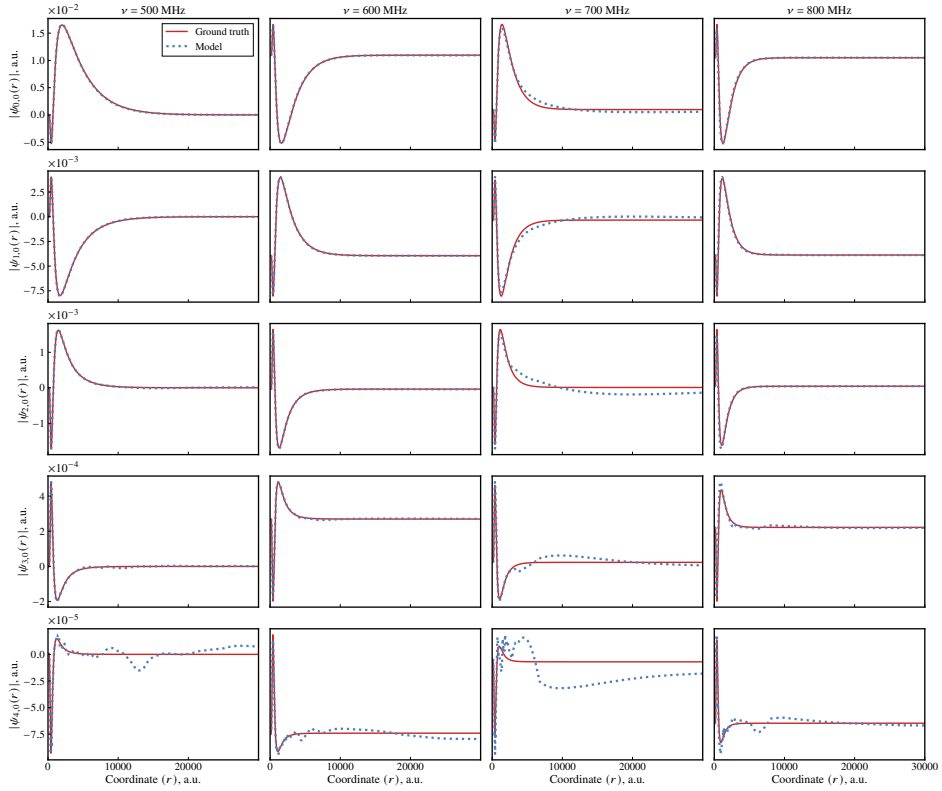

Figure 9: **Excited-state wave functions across interaction strengths ν.** Solid lines denote MDVR reference solutions and dashed lines denote the neural predictions. The model reproduces the nodal structure and asymptotic decay of the excited states across interaction regimes. Deviations are largest in higher-order spherical channels, which have comparatively small amplitudes and limited impact on the total energy.

D.3 Ablation

D.3.1 Effect of the WKB envelope on excited states

Figure 10 shows the effect of removing the WKB-inspired asymptotic envelope for excited-state optimization. Compared to the ground-state case, optimization is substantially more sensitive to the asymptotic prior, particularly at large ν. Although larger ν leads to stronger spatial confinement, excited states additionally require learning sign-changing wave functions with accurate zero crossings, making optimization considerably less stable without the envelope. Incorporating the asymptotic prior stabilizes training and consistently improves convergence across interaction regimes.

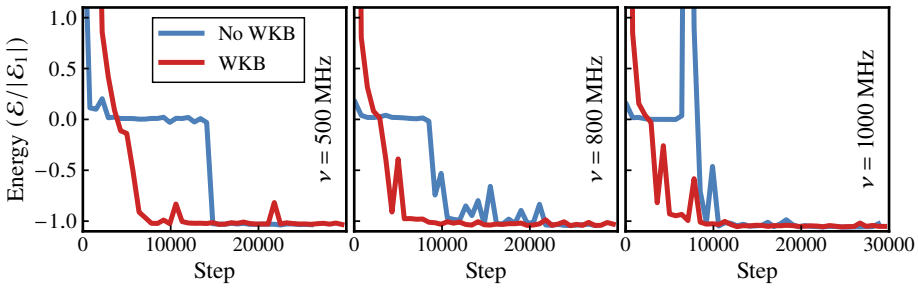

Figure 10: **Effect of the WKB envelope on excited-state optimization.** Training trajectories with and without the WKB-inspired asymptotic envelope for excited states across interaction regimes. Incorporating the envelope significantly stabilizes optimization and accelerates convergence, while training without the asymptotic prior exhibits large fluctuations and poor convergence behavior.

D.3.2 WKB ENVELOPE SPREAD

The WKB-inspired envelope acts as an asymptotic prior that guides optimization toward physically relevant regions of the solution space. To control the strength of this prior, we broaden the asymptotic decay by scaling the WKB exponent as $\kappa_{\mathcal{E}}(r) \to \beta\kappa_{\mathcal{E}}(r)$, where smaller β corresponds to a broader and weaker envelope. Figure 11 shows optimization trajectories for several choices of the broadening coefficient β. Across all tested values, optimization eventually converges to the correct solution, indicating that the method is not highly sensitive to the precise envelope shape. However, β strongly affects optimization dynamics. Large values of β impose overly sharp asymptotic decay, restricting exploration of nearby configurations and producing unstable gradients, particularly for the finite-difference kinetic-energy operator that depends on neighboring grid points. This leads to pronounced energy spikes during training. In contrast, excessively broad envelopes provide weaker asymptotic guidance and slower convergence. Intermediate values yield the most stable and efficient optimization behavior. In all experiments, we use $\beta = 0.01$, which provided a robust trade-off between stability and convergence speed.

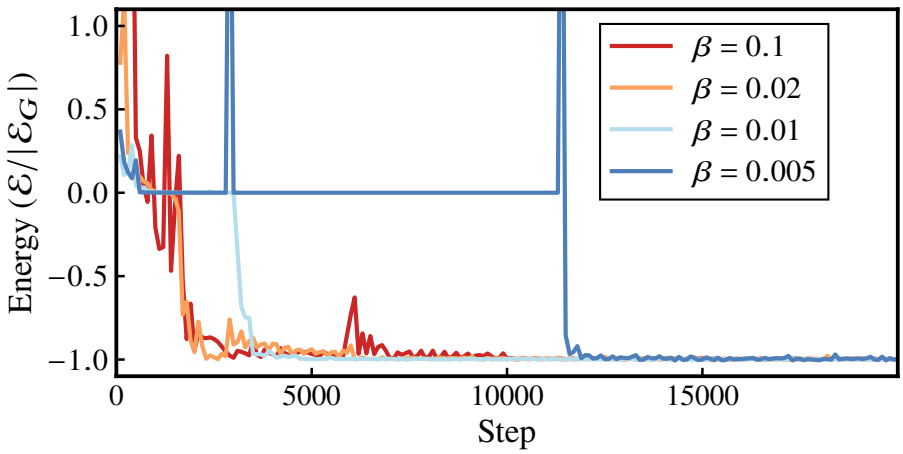

Figure 11: **Effect of the WKB envelope spread parameter.** Curves show training trajectories for different broadening coefficients β. Very narrow envelopes lead to unstable optimization and large energy spikes, while overly broad envelopes provide weaker asymptotic guidance. Nevertheless, all tested values eventually converge to the correct solution. An intermediate choice, $\beta = 0.01$, yields the most stable convergence and best overall performance.

D.4 Effect of sampling strategy

We compare three sampling strategies: uniform, logarithmic, and the proposed WKB-based MCMC sampling. Uniform sampling leads to slow convergence and higher final energy due to poor coverage of relevant regions (Fig. 12). Both logarithmic and WKB-based sampling improve optimization. While the logarithmic grid performs comparably in this setting, it relies on a fixed, problem-specific coordinate transformation. In contrast, the proposed approach adapts to the physical structure through the sampling distribution itself and does not require manual tuning. This distinction becomes critical in higher-dimensional settings, where suitable coordinate transformations are unavailable.

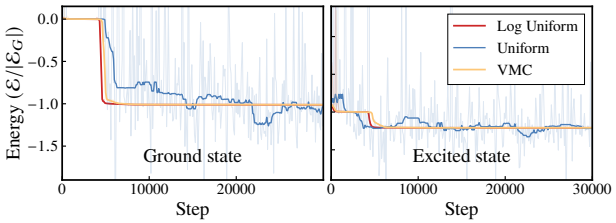

Figure 12: **Effect of sampling strategy.** Training trajectories for ground and excited states using uniform, logarithmic, and WKB-based sampling. Uniform sampling leads to noisy optimization and slow convergence, while logarithmic and WKB-based sampling improve stability and convergence.

D.5 Robustness across random seeds

To evaluate optimization stability and reproducibility, we repeated training using five independent random seeds across representative field strengths (ν, sampled in 100 MHz increments). Table 1 reports the resulting mean energies and standard deviations alongside baseline benchmark values computed via mapped discrete variable representation (MDVR). While absolute variability increases at larger ν, energy magnitudes scale proportionally, keeping the relative variance small across all regimes. Overall, the optimization remains stable across random initializations and all physical trends are preserved.

Table 1: Ground-state and excited-state energies across representative interaction regimes (ν), reported as mean $\pm$ standard deviation across 5 random seeds compared to MDVR benchmarks.

	Ground State (kHz)		**Excited State (kHz)**	
ν (MHz)	Ours	MDVR	Ours	MDVR
100	-57.963 ± 0.030	-57.948	–	–
200	-215.458 ± 0.101	-215.223	–	–
300	-425.085 ± 0.589	-425.145	–	–
400	-670.347 ± 1.169	-670.824	-1.005 ± 0.145	-1.214
500	-944.922 ± 1.340	-943.628	-5.397 ± 0.562	-5.757
600	-1237.390 ± 4.307	-1238.276	-12.386 ± 2.985	-14.252
700	-1545.055 ± 16.572	-1551.201	-27.791 ± 0.564	-26.914
800	-1883.576 ± 6.018	-1879.825	-44.475 ± 1.020	-43.781
900	-2227.304 ± 9.927	-2222.201	-65.399 ± 2.503	-64.806
1000	-2577.781 ± 41.851	-2576.802	-90.684 ± 3.324	-89.896

D.6 Additional classical baseline

The MDVR method is based on the discrete variable representation (DVR), also known as the pseudospectral method. For smooth wave functions, the discretization error of pseudospectral methods decreases exponentially as the number of grid points increases. MDVR further improves the efficiency of DVR by introducing an optimized coordinate mapping, allowing the same level of accuracy to be achieved with significantly fewer grid points. Consequently, for the present problem,

the numerical error introduced by the MDVR method can be made negligibly small by using a sufficiently dense grid.

To validate the accuracy of the MDVR approach, we compare its results with those obtained from an independent momentum-space calculation based on the Weinberg expansion (Weinberg, 1963) of the interaction potential. As shown in Table 2, the two methods agree to within 0.02% for the ground-state energy, demonstrating the high accuracy of the MDVR method. The relative error first decreases and then gradually increases with increasing energy, reflecting the competing effects of the increasing absolute deviation (numerator) and the increasing binding energy (denominator). The absolute deviation may arise from numerical errors in both methods.

Table 2: Comparison of MDVR with an independent Weinberg-expansion (WeinE) benchmark

ν (MHz)	WeinE energy (kHz)	MDVR energy (kHz)	Relative error
50	-11.071	-11.069	0.018%
100	-57.952	-57.948	0.007%
200	-215.227	-215.223	0.002%
300	-425.143	-425.145	$< 0.001\%$
400	-670.813	-670.824	0.002%
500	-943.603	-943.628	0.003%
600	-1238.235	-1238.276	0.003%
700	-1551.141	-1551.201	0.004%
800	-1879.743	-1879.825	0.004%
900	-2222.096	-2222.201	0.005%
1000	-2576.673	-2576.802	0.005%

E TECHNICAL DETAILS

Optimization. All experiments were implemented in PyTorch and optimized using Adam with learning rate 10^{-3}, gradient clipping threshold 10^{-3}, and a step learning-rate scheduler with decay factor $\gamma = 0.5$ every 10^4 training steps. Models were trained for 3×10^4 steps, although convergence typically occurred substantially earlier. All computations were performed in double precision. Experiments were run on a single NVIDIA H100 GPU. The radial-channel experiments used a batch size of 2×10^4, with training time below 10 minutes per run.

2D optimization. The optimization for the 2D wave function used the same hyperparameters as for the radial case. One noticeable exception is the batch size. In order to accurately represent integrals over 2D coordinate domain, batches of 3×10^5 coordinate points were used ($\times 15$ compared to the radial case). This substantially increases every step's runtime and GPU memory usage. Nevertheless, the training converges within the aforementioned 3×10^4 steps.

Wave function representation. The wave function was parameterized using a multilayer perceptron (MLP) with:

- 6 hidden layers,
- hidden dimension 128,
- sinusoidal positional encoding with $N = 3$ frequency bands.

These hyperparameters were selected by first training candidate architectures in a supervised setting against reference wave function data, and subsequently using the same architecture in the variational formulation proposed in this work.

The positional encoding was defined as

$$\mathrm{PE}(r) = \left[\sin\left(2^k 2\pi\tilde{r}\right), \cos\left(2^k 2\pi\tilde{r}\right)\right]_{k=0}^{N-1},$$

where $\tilde{r} = (r - r_{\min})/(r_{\max} - r_{\min})$ denotes the normalized radial coordinate.

The model outputs correspond to even spherical harmonic channels up to $l_{\max} = 8$, namely $l = 0, 2, 4, 6, 8$. The WKB-inspired asymptotic envelope described in Sec. 5.2 was tempered using $\beta = 0.01$ to avoid over-constraining the representation.

Sampling grids. For the sampling-grid comparison experiments, each grid contained 10^5 points over the radial domain $r \in [50, 10^6]$. The uniform grid used linear spacing with step size 10, while the logarithmic grid used logarithmically spaced points over the same interval. The WKB-based grid was constructed using the MCMC sampling procedure described in Sec. 5.4.

VMC sampling. The VMC sampler used 5000 walkers in a one-dimensional radial space with cutoff range $r \in [50, 10^6]$. Sampling was performed in log-space to efficiently resolve the multiple spatial scales present in the problem. A symmetric Gaussian proposal with initial step size 0.5, burn-in length 5000, and thinning factor 10 was used. After burn-in, 20 samples were retained per walker, resulting in a total of 10^5 sampled points.

During burn-in, the proposal step size was adaptively adjusted using a windowed acceptance-rate controller targeting an acceptance probability of 0.3. Let a_t denote the acceptance rate measured over an adaptation window of size 50, and let $a^\star$ denote the target acceptance rate. The proposal scale was updated according to

$$\log s_{t+1} = \log s_t + \eta \, \mathrm{clip}(a_t - a^\star, -c, c),$$

where s_t is the proposal step size, $\eta = 0.6$ is the controller learning rate, and $c = 0.2$ is a clipping threshold. The multiplicative update was additionally constrained using growth and shrinkage caps of 1.6 per adaptation step to prevent unstable changes in proposal variance.

Energy scaling. We work within the atomic unit system. In these units, typical energy values for our system are on the order of 10^{-12} (Hartrees). Such extremely small numerical values lead to poor numerical conditioning during optimization. To improve stability, we rescale the energies using the reference value $|\mathcal{E}_{\text{WKB}}|$, defined in Sec. 5.2. All reported energies were normalized by this reference scale during training and evaluation.

E.1 Physical Hyperparameters

Choice of ν. The first and second bound states appear when $\nu > 16\,\text{MHz}$ and $\nu > 300\,\text{MHz}$, respectively. We choose our benchmark dataset within $\nu \in [50, 1000]\,\text{MHz}$ for the ground state and $\nu \in [400, 1000]\,\text{MHz}$ for the excited state, though our approach applies across other ν ranges.

Choice of r. The interaction potential features a repulsive barrier around 700 a.u. (decreasing slightly with increasing ν), which exponentially suppresses the wave function amplitude at smaller distances. Consequently, at $r_{\min} = 50$ a.u., the wave function amplitude already approaches zero, allowing us to exclude the origin and small r values that would otherwise induce numerical instability. Similarly, setting $r_{\max} = 10^5$ a.u. ensures sufficient wave function decay at large distances.