Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна

Факультет (навчально-науковий інститут) «Фізико-технічний факультет»

Кафедра Фізики ядра та високих енергій імені О. І. Ахієзера

До захисту допущено
Кафедрою фізика ядра та високих енергій імені О.І. Ахієзера протокол № 10 від 15.05.2026 р.

завідувач кафедри _____ Юрій Слюсаренко

(підпис) *(ім'я, прізвище)*

«15» травня 2026 р.

Кваліфікаційна робота
здобувача другого (магістерського) рівня вищої освіти
(першого (бакалаврського) / другого (магістерського))

Застосування архітектури мереж Колмогорова-Арнольда для обчислення фізичних величин за допомогою нейронних квантових станів у неперервних системах багатьох тіл.
(назва роботи)

Спеціальність (спеціалізація) 105 «Прикладна фізика та наноматеріали»
(код та найменування спеціальності; спеціалізації спеціальності - за наявності)

Освітня програма "Експериментальна ядерна фізика та фізика плазми"
(назва освітньої програми)

Виконавець _____ Микола Луганько
(підпис) *(ім'я, прізвище)*

Науковий керівник _____ Андрій Сотніков
(підпис) *(ім'я, прізвище)*

Харків – 2026

Анотація

У роботі розроблено нейромережевий квантовий стан на основі мереж Колмогорова–Арнольда (KAN–NQS) для взаємодійних систем бозонів у неперервному просторі з періодичними граничними умовами. Запропонований анзац поєднує перестановочно-інваріантну архітектуру Deep Sets із KAN-шарами та явними парними ознаками, що дає змогу описувати нетривіальні кореляції у неперервному просторі зі збереженням фізично важливих симетрій задачі.

Розроблений підхід протестовано на моделях із гаусівською взаємодією в одно-, дво- та тривимірному випадках, на одновимірній моделі Калоджеро–Сазерленда із сингулярною нелокальною взаємодією, а також на додаткових моделях, призначених для аналізу обмежень парно-добуткових варіаційних форм. Для моделі Калоджеро–Сазерленда до анзацу включено явний cusp-множник, який стабілізує оптимізацію та відтворює правильну короткодіапазонну структуру хвильової функції. Чисельно отримано енергії основного стану, а також репрезентативні профілі густини й парні кореляційні функції для всіх розглянутих тестових задач.

Крім того, у межах єдиного варіаційного підходу проведено порівняння кількох типів KAN-шарів і показано, що параметрично редукована реалізація KAN забезпечує найкращий компроміс між виразністю, стабільністю оптимізації та кількістю параметрів. У прямих порівняннях із базовими анзацами ястрового типу KAN-посилений анзац Deep Sets забезпечує нижчі варіаційні енергії в режимах, де суттєвими є колективні або вищі багаточастинкові кореляції. Також проаналізовано внутрішню структуру вивчених KAN-шарів за допомогою діагностик рангу та сингулярних значень. Загалом отримані результати свідчать, що KAN-посилені нейронні квантові стани є перспективним і практично конкурентним варіаційним підходом для опису систем взаємодійних бозонів у неперервному просторі.

Ключові слова: нейронні квантові стани, мережі Колмогорова–Арнольда, Deep Sets, варіаційний метод Монте-Карло, бозонні системи, парні кореляції, модель Калоджеро–Сазерленда.

KAN Neural Quantum States for Bosonic Systems with Nonlocal Interactions

Abstract

We develop a Kolmogorov–Arnold network neural quantum state (KAN–NQS) for interacting bosonic systems in continuous space with periodic boundary conditions. The proposed ansatz combines the permutation-invariant Deep Sets architecture with KAN layers and explicit pairwise features, allowing the wave function to represent nontrivial correlations in continuous space while preserving the relevant symmetries of the problem.

We benchmark the approach on finite-range Gaussian interactions in one, two, and three spatial dimensions, on the one-dimensional Calogero–Sutherland (CS) model with singular nonlocal interactions, and on additional models designed to probe the limits of pair-product variational forms. For the CS model, we incorporate an explicit cusp term to stabilize optimization and reproduce the correct short-distance structure of the wave function. Numerically, we obtain ground-state energies together with representative density profiles and pair-correlation functions across all benchmark models.

We further compare several KAN layer constructions within the same variational pipeline and find that our parameter-reduced KAN implementation provides the best overall trade-off between expressivity, optimization stability, and parameter efficiency. In direct comparisons with Jastrow-type baselines, the KAN-enhanced Deep Sets ansatz yields lower variational energies in regimes where collective or higher-order correlations become important, indicating a clear advantage over simpler pair-product descriptions. We also analyze the internal structure of the learned KAN layers through rank and singular-value diagnostics. Taken together, these results support KAN-enhanced neural quantum states as a promising and practically competitive variational framework for interacting bosons in continuous space.

Keywords: neural quantum states, Kolmogorov–Arnold networks, Deep Sets, variational Monte Carlo, bosonic systems, pair correlations, Calogero–Sutherland model.

CONTENTS

1. Introduction

Neural Quantum States (NQS) have emerged as a powerful variational framework for approximating ground states and dynamics of quantum many-body systems, leveraging the universal approximation capabilities of neural networks to parameterize complex wave functions. By representing quantum states as neural networks and optimizing variational parameters through techniques such as variational Monte Carlo (VMC), NQS have demonstrated remarkable success in capturing highly entangled states across diverse platforms, from spin systems to quantum chemistry applications. However, despite their promise, current NQS implementations face significant computational and methodological challenges that limit their applicability to increasingly complex quantum systems.

Recent advances in extending NQS to continuum quantum field theories have revealed critical limitations in optimization stability and physical accuracy. Martyn et al. [1] demonstrated that Neural-Network Quantum Field States (NQFS) suffer from substantial optimization instabilities, particularly in inhomogeneous systems where direct optimization via the time-dependent variational principle is prone to numerical instabilities. Furthermore, the treatment of interaction potentials that diverge as particles approach each other requires precise handling of Kato cusp conditions[2], which traditional multilayer perceptron (MLP) architectures struggle to capture accurately, leading to large variances in local energy estimates that hinder efficient optimization. Additional challenges arise in systems that violate particle number conservation, where the local energy exhibits singular behavior that is difficult to regularize with conventional neural network ansätze.

Complementary limitations emerge in the treatment of periodic quantum systems, where spatial periodicity introduces additional architectural constraints. Pescia et al.[3] identified that while Deep Sets neural networks provide the necessary permutation invariance for bosonic systems, they face expressivity limitations and computational bottlenecks. The challenge of simultaneously capturing both local correlations and long-range periodic structure requires neural architectures that can efficiently represent multi-scale physics, particularly in two-dimensional systems where computational demands scale unfavorably. These limitations collectively highlight the need for more sophisticated neural network architectures that can address the fundamental mathematical challenges inherent in quantum many-body variational approaches while maintaining computational tractability.

One promising direction is to replace the standard multilayer perceptron blocks used in many NQS architectures by Kolmogorov–Arnold networks (KANs). In contrast to conventional MLPs, where the nonlinearities are fixed and only the linear weights are learned, KANs assign learnable univariate functions to the network edges. This change gives the model more direct control over the shape of the nonlinear transformations and therefore can improve both expressivity and trainability. In the context of many-body wave functions, this is particularly attractive because the target functions often combine smooth collective structure, sharp short-range features, and strong sensitivity to coordinate transformations; these properties place substantial demands on the chosen basis and on the optimization dynamics.

In this work, we use KANs not as a standalone black-box replacement, but as the functional building blocks inside a symmetry-aware NQS architecture for interacting bosonic systems in continuous space. Specifically, we combine KAN layers with a Deep Sets construction so that bosonic permutation symmetry is built into the ansatz by design, while explicit pairwise features allow the network to represent physically relevant two-body correlations in periodic geometries. This leads to a KAN-enhanced Deep Sets wave-function ansatz in which the expressive part of the model is concentrated in learnable one-dimensional edge functions, whereas the many-body symmetries are enforced through invariant aggregation and correlation-aware embeddings.

The main goal of this thesis is, therefore, not only to test whether KANs can be used in neural quantum states, but also to evaluate how they perform in a physically meaningful continuum setting. To this end, we study finite-range Gaussian interactions in one, two, and three spatial dimensions, the one-dimensional Calogero–Sutherland model with singular inverse-square interactions, and additional comparison problems designed to expose the limitations of simpler pair-product wave functions. We also compare several KAN layer constructions within the same variational pipeline, analyze optimization behavior and internal rank structure, and benchmark the resulting ansatz against Jastrow-type baselines. In this way, the present work aims to provide both an architectural contribution–a practical KAN-based NQS for bosons in continuous space–and a numerical assessment of when such added flexibility yields a clear advantage.

2. Literature Overview

Neural quantum states (NQS) are variational quantum-state ansätze in which the many-body wave function is parameterized by a neural network. In this framework, the network takes a physical configuration of the system and returns the corresponding wave-function amplitude, phase, or both, thereby replacing more traditional handcrafted trial states with a flexible data-driven representation. In practice, the parameters of the neural-network ansatz are usually optimized within the variational Monte Carlo (VMC) framework, where configurations are sampled from the probability distribution induced by the current wave function and the variational energy is minimized with respect to the network parameters. This combination of expressive neural parameterizations and stochastic variational optimization has made NQS a powerful tool for studying quantum many-body problems because it allows one to treat systems whose correlation structure is difficult to capture with conventional ansätze.

The modern NQS literature grew from the observation that neural networks can serve not only as generic function approximators, but also as physically meaningful wave-function models capable of representing nontrivial entanglement and correlation patterns. Since then, the field has expanded rapidly to include architectures for lattice and continuum systems, symmetry-preserving constructions, autoregressive models, excited-state methods, and applications to both bosonic and fermionic problems. A broad overview of these developments is given by Lange et al. [4], who review the main NQS architectures and their applications to ground states, excited states, finite-temperature problems, open-system dynamics, and quantum-state tomography.

2.1. Foundation of NQS

The modern foundation of the field was established by Carleo and Troyer [5], who showed that a restricted Boltzmann machine can be used as a variational representation of a quantum many-body wave function and optimized efficiently within the variational Monte Carlo framework. Their work demonstrated that neural-network ansätze are practical variational states capable of describing both ground-state properties and nonequilibrium dynamics in interacting spin systems. This paper is widely regarded as the starting point of the NQS program because it formulated the now-standard idea of representing the wave function directly by a trainable neural archi-

tecture and optimizing it through stochastic sampling.

Soon afterward, Deng, Li, and Das Sarma [6] clarified why such representations are physically interesting by analyzing the entanglement structure of neural-network states. In particular, they showed that short-range restricted Boltzmann machine states naturally obey area-law entanglement scaling, while long-range constructions can represent states with volume-law entanglement. This result was important because it connected neural-network wave functions to one of the central structural properties of quantum many-body states, namely their entanglement scaling, and therefore provided theoretical support for the claim that NQS can capture classes of states that are difficult for more traditional variational ansätze.

At the same time, Gao and Duan [7] strengthened the theoretical basis of the field by studying the expressive power of deep neural-network representations for quantum many-body systems. Their analysis showed that deep architectures can represent highly complex quantum states efficiently and, in some regimes, more compactly than shallow constructions. This line of work helped establish that the success of NQS is not only empirical, but also rooted in a genuine representational advantage: neural networks can encode nonlocal correlations and complicated sign or amplitude structures in a way that is systematically more flexible than many earlier handcrafted wave-function forms.

Taken together, these foundational papers defined the three main pillars of the NQS field: a practical variational training framework based on neural-network wave functions and VMC, a physical understanding of the entanglement properties that such states can realize, and a theoretical argument for the strong expressive power of deep neural architectures in quantum many-body settings.

2.2. Deep Architectures

As the field matured, attention shifted from the original shallow restricted-Boltzmann-machine constructions toward richer neural architectures and improved optimization procedures. One important methodological ingredient in this broader development is stochastic reconfiguration, which has become one of the standard optimization tools in variational Monte Carlo because it incorporates information about the geometry of the variational manifold and often yields more stable parameter updates than naive gradient descent. Within the NQS literature, the move toward deeper and more structured architectures has gone hand in hand with the need for optimization schemes that remain robust as expressive power increases.

Sharir et al. [8] provided one of the clearest demonstrations of this shift by introducing deep autoregressive neural quantum states for many-body systems. Their factorized construction gives a normalized wave-function representation together with direct sampling, thereby removing the need for an internal Markov chain to generate configurations. This is an important conceptual step because autoregressive models turn sampling, probability evaluation, and variational optimization into a single coherent framework, making them especially attractive for large systems and for architectures in which conventional Monte Carlo updates can become inefficient.

Hibat-Allah et al. [9] developed a closely related but distinct direction based on recurrent neural-network wave functions. By using recurrent architectures to sequentially generate many-body configurations, they showed that autoregressive ideas can be adapted naturally to quantum states while preserving exact likelihood evaluation and efficient independent sampling. Their work strengthened the case for sequence-model-inspired wave functions and demonstrated that neural architectures originally designed for language or time-series data can be repurposed successfully for strongly correlated quantum problems.

Zhang and Di Ventra [10] extended this trend further with the transformer quantum state, arguing that attention-based architectures can serve as flexible general-purpose models for quantum many-body problems. The relevance of this contribution lies not only in introducing a new neural architecture, but also in emphasizing that modern deep-learning components such as self-attention can capture long-range and global dependencies that are difficult to encode efficiently with more local or sequential models. In that sense, transformer-based NQS represent part of a broader transition from early proof-of-principle neural wave functions to a more mature design space that draws directly on advances from mainstream deep learning.

Taken together, these works show that the evolution of NQS is not limited to better physical applications; it also involves a continuous refinement of the underlying machine-learning architectures and optimization strategies. This broader architectural context is relevant for the present thesis because it motivates the search for alternatives to standard multilayer perceptrons, including KAN-based layers, as potentially more expressive and trainable building blocks for variational wave functions.

2.3. Physically-relevant Symmetries

An important stage in the development of neural quantum states was the realization that physically-relevant symmetries should not merely be learned approximately from data, but can often be incorporated directly into the variational construction. Choo, Carleo, Regnault, and Neupert [11] made a foundational contribution in this direction by extending NQS methods beyond ground states and showing how neural-network wave functions can be constrained to target eigenstates of Hamiltonian symmetries. In the same work, they also introduced a practical strategy for computing low-lying excited states within the neural variational framework. This was a major advance because it demonstrated that NQS can be adapted not only to reproduce the lowest-energy state, but also to access symmetry-resolved sectors and excitation spectra, which are central to the physical interpretation of many-body systems.

From a broader perspective, this paper clarified that symmetry information can serve as a structural principle for neural ansätze rather than as a secondary diagnostic applied after optimization. Embedding symmetry constraints directly into the wave function reduces the effective search space, improves physical interpretability, and allows one to address questions that are inaccessible to a purely unconstrained ground-state calculation. For this reason, symmetry-aware constructions have become one of the central design principles in modern NQS research.

Luo et al. [12] pushed this idea substantially further by considering symmetries and constraints that are more intricate than global quantum numbers, namely gauge invariance and anyonic symmetry in lattice models. They developed a general framework for autoregressive neural networks that explicitly satisfy these constraints while retaining efficient sampling properties. Their results are especially important because they show that neural quantum states can be built to respect highly nontrivial kinematic structures from the outset, making them suitable for lattice gauge theories, topological systems, and other constrained many-body problems. In this sense, the work of Luo et al. broadens the role of symmetry in NQS from conventional symmetry sectors to genuinely constrained Hilbert spaces.

Together, these papers establish that symmetry is not an optional refinement in neural quantum states, but one of the main mechanisms by which expressive neural ansätze can be made physically faithful and computationally efficient. This perspective is directly relevant for the present thesis, where bosonic permutation symmetry and periodic structure are likewise treated as built-in ingredients of the

variational architecture rather than as properties to be learned only indirectly.

2.4. Continuous Systems

Moving from spin systems to problems with continuous degrees of freedom required more than a simple change of input representation: it required neural wave functions that act directly on continuous coordinates while remaining trainable within variational Monte Carlo. An important early step in this direction was made by Stokes et al. [13], who formulated continuous-variable neural-network quantum states for the quantum rotor model. Their work showed that RBM-inspired trial states and the associated VMC machinery can be extended to continuous-variable settings in first quantization, providing a proof of principle that neural variational methods remain viable beyond discrete Hilbert spaces. At the same time, their results made it clear that realistic continuous systems would likely require richer architectures and more sophisticated optimization strategies than the simplest shallow constructions.

A more direct precursor to the present thesis is the work of Pescia et al. [14] on neural-network quantum states for periodic systems in continuous space. In that paper, the authors introduced a Deep Sets-based ansatz in which particle coordinates are transformed in a way that is compatible with periodic boundary conditions, while permutation invariance is built into the architecture for bosonic systems. This was a key conceptual advance because it addressed two central difficulties of continuum bosonic problems at once: the need to encode exchange symmetry exactly and the need to represent spatial periodicity in a form suitable for neural processing. Their benchmarks on one- and two-dimensional quantum gases with Gaussian interactions, as well as on confined helium, showed that symmetry-aware NQS can produce accurate ground-state energies and correlation observables in realistic continuous-space settings.

An important complementary development was provided by Kim et al. [15], who considered ultracold Fermi gases across the strongly correlated BCS–BEC crossover. Although their focus is fermionic rather than bosonic, the paper is highly relevant for the continuous-systems literature because it shows that neural quantum states can remain competitive even in the presence of strong pairing correlations and short-range interactions. Their Pfaffian–Jastrow neural-network ansatz, augmented by message-passing backflow features, demonstrates that physically informed architectures can outperform more conventional Slater–Jastrow baselines and can even com-

pete with state-of-the-art diffusion Monte Carlo benchmarks. Equally important, the authors show that transfer learning can stabilize optimization near the most challenging interaction regimes. For the present thesis, this work reinforces the broader lesson that successful NQS treatments of continuum matter often require architectures that encode the dominant correlation structure explicitly rather than relying on generic dense networks alone.

3. Models and Methods

The central problem addressed by neural quantum states is the ground-state problem for a given many-body Hamiltonian $\hat{H}$. One seeks the lowest eigenvalue E_0 and the corresponding wave function Ψ_0, which satisfy

$$\hat{H}\Psi_0 = E_0\Psi_0. \tag{3.1}$$

Within the variational framework, this problem is approximated by introducing a parameterized trial state Ψ_θ and minimizing the Rayleigh quotient

$$E[\Psi_\theta] = \frac{\langle\Psi_\theta|\hat{H}|\Psi_\theta\rangle}{\langle\Psi_\theta|\Psi_\theta\rangle} \geq E_0. \tag{3.2}$$

Neural quantum states arise when the variational ansatz Ψ_θ is represented by a neural network, so that the search for the ground state is recast as an optimization problem over network parameters. In practice, this optimization is usually carried out within variational Monte Carlo, where configurations are sampled from $|\Psi_\theta|^2$ and observables are estimated stochastically.

In the present work, we apply this variational strategy to bosonic Hamiltonians in continuous space with periodic boundary conditions. The class of systems considered here can be written schematically as

$$\hat{H} = -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\nabla_i^2 + \hat{V}, \tag{3.3}$$

where the first term is the kinetic energy and $\hat{V}$ encodes the interaction potential. The benchmark models considered in this thesis span several complementary interaction regimes, including the periodic Calogero–Sutherland model, finite-range Gaussian interactions, a Gaussian three-body interaction, and a regularized Lennard-Jones-type potential. Their explicit Hamiltonians are introduced in the dedicated subsections below.

This viewpoint was introduced by Carleo and Troyer, who showed that neural-network wave functions can accurately represent interacting quantum spin states and can be optimized variationally with high precision [5]. Since then, the same general strategy has been extended to a broad range of problems, includ-

ing frustrated spin systems such as the two-dimensional J_1–J_2 Heisenberg model [16], continuous-space quantum many-body systems [17], electronic-structure and quantum-chemistry problems [18], and periodic systems in continuous space described by symmetry-aware architectures [3]. These examples illustrate that the NQS paradigm is not tied to one specific physical platform: rather, it provides a flexible variational framework whose effectiveness depends on how well the chosen neural ansatz incorporates the relevant symmetries, correlations, and geometric structure of the target Hamiltonian.

3.1. Physical system and motivation

We study dilute bosonic many-body systems in continuous space with periodic boundary conditions. Periodic geometry is useful here for two reasons: it removes boundary effects that would otherwise obscure bulk correlations, and it provides a natural setting in which translational symmetry can be built directly into the variational ansatz. The benchmark suite is chosen to cover four complementary regimes within this common framework: the one-dimensional Calogero–Sutherland model on a ring, finite-range Gaussian interactions in one, two, and three spatial dimensions, a Gaussian three-body interaction, and a regularized Lennard-Jones-type soft-core potential. Taken together, these models allow us to separate several distinct challenges for neural quantum states (NQS), namely smooth versus singular interactions, pairwise versus genuinely many-body structure, and single-scale versus competing-scale correlations.

The motivation for this selection is methodological rather than tied to a single experimental platform. We want a benchmark set that spans analytically structured models, smooth reference problems, explicitly non-pairwise interactions, and pairwise interactions whose collective response is effectively many-body. This makes it possible to test not only variational accuracy but also how well the ansatz handles permutation symmetry, periodicity, short-distance structure, and optimization stability across qualitatively different physical regimes.

Throughout this work we use dimensionless variables. A characteristic length scale a is introduced to nondimensionalize distances, so that physical coordinates are rescaled as $x \to x/a$. Here m denotes the mass of a bosonic particle. In these units, the kinetic energy prefactor is absorbed by setting $\hbar^2/(2ma^2) = 1$, and all coordinates and distances are therefore measured in units of a.

Inverse-square Calogero–Sutherland interaction. We begin with the periodic Calogero–Sutherland (CS) model (also known as Calogero–Moser–Sutherland model with the trigonometric interaction potential) governed by the Hamiltonian

$$H_{\mathrm{CS}} \;=\; -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\nabla_i^2 \;+\; V_0 \sum_{1\le i<j\le N}\frac{g\,\pi^2/L^2}{\sin^2\!\big[\pi(x_i-x_j)/L\big]}, \tag{3.4}$$

with dimensionless coupling $g = 5$, $V_0 = 1$ as a result $\frac{\hbar^2}{2m} = 1$ and particle coordinates x_i on a ring of length L, $x_i \in [-L/2, L/2]$. The circumference sets the only external length scale, while the periodic denominator makes the long-range interaction compatible with the ring geometry.

Given in the form (3.4), the system Hamiltonian corresponds to an integrable one-dimensional model [19] with nonlocal repulsive interactions, power-law correlations, and a short-distance singularities in the potential. These properties make it a stringent and analytically structured test for NQS: successful learning requires the ansatz to capture algebraic order, global periodic organization, and the correct short-distance suppression of the wave function in order to keep the local energy under control. In practice, this is exactly the regime in which Jastrow-like priors or other physics-informed features are often most valuable. On the finite ring, the model is not strictly scale invariant because the circumference L sets an explicit length scale. Nevertheless, after introducing the dimensionless coordinates $x_i = Ly_i$, the Hamiltonian can be written as an overall factor L^{-2} multiplying a dimensionless operator that depends only on λ, so its finite-size dependence remains strongly constrained.

From the physical perspective, the CS model is a paradigmatic one-dimensional system with inverse-square repulsion. It is used here mainly as a benchmark rather than as the most literal microscopic description of a cold-atom experiment, although related effective inverse-square couplings can emerge in engineered one-dimensional settings with long-range interactions, constrained lattice realizations, or collective dipolar modes. For the present manuscript, the importance of the model is methodological: it tests whether the ansatz can simultaneously reproduce global periodic ordering, nonlocal correlations along the ring, and the correct short-distance behavior of the wave function.

The difficulty of this interaction lies precisely in that combination of features. The potential is singular when two particles approach one another, while the interac-

tion also remains long-ranged along the ring, so optimization must capture nonlocal correlations rather than only local pair structure. For this reason, the CS benchmark is the natural place to assess whether a KAN-enhanced ansatz can handle singular continuum physics more robustly than a smoother pair-product baseline.

Finite-range Gaussian interaction. We next turn to the Hamiltonian with the Gaussian type of nonlocal interaction,

$$H = -\frac{\hbar^2}{2m}\sum_i \nabla_i^2 + \epsilon \sum_{i<j} e^{-r_{ij}^2}, \tag{3.5}$$

where ϵ sets the interaction strength and σ characterizes the effective interaction range, and r_{ij} is the distance between two particles i and j. If it is not stated otherwise, the default value for the interaction strength is $\epsilon = 2$.

In contrast to the one-dimensional CS model (3.4) with singularities, the Hamiltonian (3.5) serves as a smoother reference problem with a finite-range two-body interaction. The Gaussian model is non-polynomial yet smooth, captures controlled boson-boson repulsion of finite range, and is often used as a benchmark for variational methods. A Gaussian interaction is rarely the exact microscopic potential of a real material or atomic gas. Instead, it is widely used as an effective finite-range repulsion and as a smooth surrogate for more complicated short-range forces. In the present work, it plays the role of the clean reference problem in one, two, and three spatial dimensions.

This model is useful when one wants to study the impact of interaction range without introducing a hard core or an ultraviolet divergence. In that sense, it is less a literal physical law than a controlled benchmark representation of finite-range bosonic correlations. For NQS, it stresses the ability to represent nonlocal but analytic correlations without enforcing cusps, and it enables systematic tests as functions of the range ratio σ/L and the density $n = N/L$. Moreover, many other physically relevant interactions can be represented or approximated as perturbations or linear combinations of Gaussian potentials, which makes this model both a natural starting point and a versatile benchmark. Within the ordered benchmark suite of this subsection, it is the interaction family that isolates finite-range correlations without the additional complications caused by singular potentials.

The main difficulty here is therefore subtler than in the CS case. Because the

potential is smooth, the challenge is not to reproduce a cusp, but to capture how a finite interaction range modifies the collective structure of the bosonic state across different dimensions and densities. The ansatz must learn when pairwise correlations remain local and when they become sufficiently extended that the aggregated many-body response matters. This makes the Gaussian benchmark an important baseline: if the model cannot describe this smooth case reliably, one should not expect it to remain accurate for more singular or more collective interactions.

Three-body Gaussian interaction. To move beyond pairwise physics, we also consider a Hamiltonian with three-body interactions of the Gaussian type,

$$H = -\frac{\hbar^2}{2m}\sum_i \nabla_i^2 + W \sum_{i<j<k} \exp\left[-\left(\frac{r_{ij}^2 + r_{jk}^2 + r_{ki}^2}{R^2}\right)^{\alpha}\right], \tag{3.6}$$

where r_{ij} denotes the minimum-image interparticle distance on the periodic box. The parameters with $W = 20$, $R = 0.6$, $\alpha = 2$ control the strength, spatial extent, and sharpness of the three-body term. Unlike the previous Hamiltonians, this interaction cannot be reduced to a sum of independent pair factors, which is precisely why it is useful for the present work.

Three-body interactions in bosonic systems are usually effective rather than fundamental. In ultracold gases they can emerge after eliminating higher bands, internal states, or other virtual degrees of freedom, and they have been discussed in optical lattices, double-well systems, Rydberg-type platforms, and continuum gases with engineered multibody couplings. Their physical significance is that they can alter the phase diagram qualitatively, stabilize states that are inaccessible in purely two-body models, and modify the structure of few-body bound states in low-dimensional systems.

The main theoretical difficulty is that three-body correlations are intrinsically more complex to represent and optimize. A Jastrow wave function can incorporate pair structure very naturally, but it does not encode irreducible triple correlations in a direct way. Moreover, the number of interacting triples grows rapidly with system size, so both the expressivity requirement and the computational burden increase. The short-range Gaussian form used in Eq. (3.6) is therefore a practical compromise: it is smooth, translationally invariant, and permutation symmetric, while still forcing the ansatz to learn genuinely many-body structure. Because the potential remains

finite when particles coincide, it avoids unnecessary numerical singularities and lets us focus on the representational challenge itself.

Lennard-Jones-type soft-core interaction. Finally, to study a model with competing length scales, we use a regularized attractive-repulsive interaction of Lennard-Jones type,

$$H_{\mathrm{DS}} \;=\; -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\nabla_i^2 \;+\; \sum_{1\le i<j\le N}\left[\frac{U_{\mathrm{rep}}}{1+(r_{ij}/R_{\mathrm{rep}})^{p_{\mathrm{rep}}}} - \frac{U_{\mathrm{att}}}{1+(r_{ij}/R_{\mathrm{att}})^{p_{\mathrm{att}}}}\right], \quad (3.7)$$

with minimum-image pair distances r_{ij}. The parameters $U_{\mathrm{rep}} = 1.0$, $R_{\mathrm{rep}} = 0.8$, and $p_{\mathrm{rep}} = 12.0$ control the soft repulsive core, while $U_{\mathrm{att}} = 0.6$, $R_{\mathrm{att}} = 1.8$, and $p_{\mathrm{att}} = 6.0$ define the attractive tail. This model is the Lennard-Jones-type interactions, retaining the competition between short-range repulsion and longer-range attraction while softening the core enough for stable optimization.

Lennard-Jones interactions are a standard effective model for neutral atoms and simple molecular systems. Their defining feature is the competition between a steep short-range repulsive core, which represents the effective Pauli exclusion-driven overlap repulsion, and a longer-range attractive part associated with dispersion forces. In few-body and molecular physics, Lennard-Jones-type potentials are widely used to study binding, cluster formation, and near-threshold scattering properties. In many-body settings, the same attraction–repulsion balance can generate nontrivial collective organization, so they provide a natural benchmark for testing whether a variational ansatz can capture structure that remains pairwise at the Hamiltonian level but becomes increasingly collective at the level of the wave function.

For the present manuscript, the important point is that such systems are challenging even though the interaction is formally pairwise. The wave function must respond not only to local avoidance at short distance but also to the longer-range attractive tail, and the balance between these two tendencies can produce collective structure that is effectively many-body in character. At the numerical level, the challenge is to retain this qualitative competition without introducing the severe short-distance divergence of the standard Lennard–Jones potential, which would increase the local-energy variance and destabilize optimization. The regularized double-soft-core form used in Eq. (3.7) is therefore a computationally convenient Lennard-Jones surrogate: for $p_{\mathrm{att}} = 6$ it preserves the characteristic long-range de-

cay, while the bounded core keeps the model smooth enough for stable continuous-space variational calculations.

Taken together, these four Hamiltonians probe complementary aspects of the same general problem. The Gaussian benchmark tests smooth finite-range correlations, the CS model stresses singular and long-range structure, the three-body Hamiltonian isolates genuinely non-pairwise correlations, and the double-soft-core model examines how competing attractive and repulsive scales generate collective organization. This combination gives a broad and physically meaningful test bed for the KAN-enhanced Deep Sets ansatz developed in the following sections.

Why these benchmarks matter for NQS. Together, Eqs. (3.4), (3.5), (3.6), and (3.7) span singular long-range, smooth finite-range, explicitly three-body, and regularized attractive-repulsive interaction regimes. This diversity allows us to: (i) validate NQS against analytic or quasi-exact results where available, (ii) quantify sample efficiency and variance control in continuous space, (iii) assess symmetry handling (bosonic exchange, translation on a ring) and inductive biases, (iv) study observables of immediate experimental relevance such as ground-state energy densities and pair correlations $g^{(2)}(r)$ under conditions accessible to ultracold-atom platforms.

These metrics directly reflect the suitability of NQS for analog-quantum benchmarks and for modeling metrologically useful quantum states.

3.2. Theoretical approaches

In continuous bosonic systems, the main theoretical approaches range from numerically exact methods to structured variational ansätze. In this work, we position our KAN-based neural quantum state relative to three important classes of methods: exact diagonalization, continuous matrix product states, and Jastrow-type wave functions. These approaches provide useful reference points because each of them realizes a different balance between accuracy, physical interpretability, and computational cost.

Exact diagonalization Exact diagonalization (ED) provides the most direct benchmark because the many-body Hamiltonian is represented explicitly in a trun-

cated basis and the eigenvalue problem

$$\hat{H}\,|\Psi_n\rangle = E_n\,|\Psi_n\rangle \tag{3.8}$$

is then solved numerically. Within the chosen basis, the method is exact up to the basis truncation, so it serves as a valuable reference for ground-state energies and other observables. In our implementation, the calculation is performed entirely in momentum space using a bosonic Fock basis of plane-wave modes.

For N_b bosons in a periodic box, the linear system size is fixed by the density as

$$L = \left(\frac{N_b}{\rho}\right)^{1/d}, \tag{3.9}$$

where d is the spatial dimension. Instead of discretizing real space directly, we truncate the one-particle momentum modes to integer vectors $\mathbf{k}$ whose components lie in the interval $[-N_{\mathrm{mode}}, N_{\mathrm{mode}}]$. The total number of one-particle modes is therefore

$$M = (2N_{\mathrm{mode}} + 1)^d. \tag{3.10}$$

The many-body Hilbert space is then the bosonic occupation-number basis over these M modes. Its dimension grows combinatorially as

$$\dim \mathcal{H} = \binom{M + N_b - 1}{N_b}, \tag{3.11}$$

which is the main practical bottleneck of the ED approach.

In this basis, the kinetic term is diagonal. For a mode labeled by momentum vector $\mathbf{k}$, the single-particle dispersion is

$$T_{\mathbf{k}} = 2\left(\frac{\pi}{L}\right)^2 |\mathbf{k}|^2, \tag{3.12}$$

which corresponds to units with $\hbar = 1$ and $m = 1$. The interaction part is built from the Gaussian pair potential

$$V(\mathbf{r}) = \varepsilon \exp\!\left(-\frac{|\mathbf{r}|^2}{\sigma^2}\right). \tag{3.13}$$

After transforming to second quantization in momentum space, the interacting

Hamiltonian takes the form

$$H_{\text{int}} = \sum_{\substack{\mathbf{m},\mathbf{k},\mathbf{l},\mathbf{n} \\ \mathbf{m}+\mathbf{k}=\mathbf{l}+\mathbf{n}}} \hat{V}(\mathbf{m}-\mathbf{l})\, a^\dagger_{\mathbf{m}} a^\dagger_{\mathbf{k}} a_{\mathbf{l}} a_{\mathbf{n}}, \tag{3.14}$$

where the momentum-conservation constraint is imposed explicitly when generating matrix elements. For the Gaussian interaction, the Fourier coefficient is known analytically,

$$\hat{V}(\mathbf{q}) = \frac{\varepsilon}{2}\left(\frac{\sqrt{\pi}\,\sigma}{L}\right)^d \exp\left[-\left(\frac{\pi\sigma}{L}\right)^2 |\mathbf{q}|^2\right]. \tag{3.15}$$

This analytical form is especially convenient because it avoids an additional real-space discretization step and generalizes directly to arbitrary spatial dimension.

In this work we have used QuSpin framework to do the exact diagonalization solutions [20].

The resulting Hamiltonian is assembled as a sparse matrix and only its lowest eigenpair is extracted, rather than computing the full spectrum. In practice, this is done with an iterative Lanczos/Arnoldi-type routine for the smallest algebraic eigenvalue. This makes ED tractable for moderate system sizes and allows it to serve as a benchmark for the neural ansatz, even though it still scales poorly with N_b and with the mode cutoff N_{mode}.

The main approximation parameter is the plane-wave cutoff N_{mode}, which controls basis convergence. It must be large enough that higher-momentum occupations are negligible in the target state. At the same time, increasing N_{mode} rapidly enlarges the Hilbert space, so in practice ED is restricted to relatively small boson numbers and is best viewed as a validation tool rather than a scalable method for the regimes targeted in this paper. Compared with an older one-dimensional implementation based on a discrete real-space Fourier transform, the momentum-space formulation with the analytical Gaussian transform avoids discretization error and extends naturally to d dimensions.

Continuous matrix product states. Continuous matrix product states (cMPS) provide the continuum analogue of matrix product states and are especially natural for one-dimensional quantum fields [21]. For a single bosonic species, a standard

cMPS ansatz can be written as

$$|\Psi[Q,R]\rangle = v_L^\dagger \mathcal{P} \exp\left(\int_0^L dx \left[Q(x) \otimes I + R(x) \otimes \hat{\psi}^\dagger(x) \right] \right) v_R \, |0\rangle \, . \tag{3.16}$$

Here Q and R are variational matrix-valued functions, and optimization is usually performed with gradient-based schemes or methods based on the time-dependent variational principle. cMPS are powerful when the target state has effectively one-dimensional structure and can be captured with a moderate bond dimension. However, they are less natural for higher-dimensional settings, for periodic problems where one wants a single architecture shared across several dimensions, or for benchmarks designed to compare the same ansatz across several qualitatively different interaction classes.

Jastrow wave functions. Jastrow-type ansätze remain one of the most physically transparent approaches for bosonic matter because they encode correlations directly through products of low-body factors [22–24]. In their simplest form, they capture the dominant effect of pairwise interactions and often provide excellent variational baselines for dilute systems. They are particularly appealing when the relevant physics is governed by two-body structure, cusp conditions, or known asymptotic behavior. However, a pure Jastrow form can become too restrictive when the true ground state contains more complicated many-body correlations, nonlocal collective structure, or features that cannot be reduced to a simple product of pair factors. Because of their conceptual importance, we discuss the Jastrow construction in more detail below and use it as a bridge to the Deep Sets–KAN representation.

Neural quantum states. Neural quantum states generalize traditional variational wave functions by replacing hand-designed correlation factors with flexible function approximators. In principle, this allows the ansatz to represent correlations that go well beyond low-body product forms and to adapt to a much wider class of Hamiltonians. The price for this increased expressivity is a more difficult optimization landscape, larger sampling noise, and a stronger dependence on architectural choices. This is precisely the motivation for the present work: we seek an NQS architecture that preserves the essential bosonic symmetries of the problem, remains interpretable, and improves the functional flexibility of standard Deep Sets models through KAN layers. In that sense, the KAN-enhanced ansatz studied here

is intended to occupy an intermediate position between highly structured physics-driven baselines such as Jastrow wave functions and more generic black-box neural parametrizations.

4. Methodology

4.1. Research objective

This section formulates the methodological framework used to study KAN-enhanced neural quantum states for continuous bosonic systems with periodic boundary conditions. Our objective is to construct a symmetry-preserving variational ansatz based on Deep Sets and Kolmogorov–Arnold network (KAN) layers, augmented by correlation-aware inputs and problem-specific corrections where required, and to evaluate it within a unified variational Monte Carlo pipeline. The methodology is therefore organized to make explicit the relation between standard MLP-based models, Jastrow wave functions, and the proposed KAN-enhanced construction; to specify the optimization, pretraining, and layer-design choices used in practice; and to define the diagnostics by which expressivity, optimization stability, and parameter efficiency are assessed across the benchmark Hamiltonians introduced above. In this way, the section serves as the methodological bridge between the physical models and the numerical results presented later.

4.2. Architectural motivation and Jastrow baseline

Most early neural quantum-state constructions relied on multilayer perceptrons (MLPs) or closely related dense neural networks. In an MLP, each layer applies an affine transformation followed by a fixed nonlinearity such as tanh or ReLU. This design is broadly applicable and often effective, but it also means that the functional form of the nonlinearity is chosen in advance rather than adapted to the structure of the physical problem. For continuum many-body wave functions, where one must often represent smooth collective behavior together with sharp short-distance features, this restriction can become limiting.

From the parameter-count perspective, a fully connected MLP layer with n_{in} inputs and n_{out} outputs contains $n_{\mathrm{in}}n_{\mathrm{out}}$ weights, ignoring biases. In a spline-based KAN layer, each scalar connection is replaced by a learnable univariate function, so for grid size G and spline order k each edge carries roughly $G + k$ coefficients. The corresponding KAN layer therefore scales as $n_{\mathrm{in}}n_{\mathrm{out}}(G + k)$ trainable edge parameters. This larger parameterization increases computational cost, but it also gives the model substantially more direct control over the shape of the nonlinear transformations on each connection.

Kolmogorov–Arnold networks (KANs) provide one natural way to relax the fixed-

nonlinearity restriction of standard MLPs[25, 26]. From the perspective of neural quantum states, they are attractive because they offer a compromise between flexibility and structure: the model becomes more expressive than a standard MLP, while still retaining an internal organization that is more interpretable than that of a completely generic dense network. Recent work in both continuous and discrete quantum systems has already suggested that KAN-based architectures can be competitive with, and in some cases superior to, more conventional choices. This motivates the central question of the present work: whether the same advantages can be realized for continuous bosonic systems with periodic geometry.

Before introducing the formal KAN construction used in this thesis, it is helpful to recall the Jastrow ansatz, since it provides the main physics-driven baseline and also clarifies what is generalized by the subsequent Deep Sets formulation. In its standard form, the wave function is written as

$$\psi_{\text{Jastrow}}(x_i) = \prod_{1\le i<j\le N} f(x_i - x_j), \tag{4.1}$$

where f is a variational pair function that encodes correlations directly in real space. This form is especially effective when the physics is dominated by two-body constraints: it can incorporate cusp conditions naturally, it reproduces the exact ground states of Calogero–Moser- and Calogero–Sutherland-type models, and it underlies classic constructions ranging from liquid-helium wave functions to Laughlin states.

At the same time, the Jastrow form makes its main limitation explicit. Because correlations enter only through a product of pair factors, genuinely collective structure can be represented only indirectly, if at all. This observation provides a natural bridge to the KAN-enhanced Deep Sets architecture. The Jastrow wave function can be rewritten in the logarithmic form

$$\begin{aligned} \log \psi_{\text{Jastrow}} &= \sum_{1\le i<j\le N} \log f(x_i - x_j) \\ &= \sum_{1\le i<j\le N} \varphi_{1,ij}(x_i - x_j) \\ &= \Phi_1\left(\sum_{1\le i<j\le N} \varphi_{1,ij}(x_i - x_j) \right). \end{aligned} \tag{4.2}$$

where the outer map is the trivial function $\Phi_1(u) = u$ and $\varphi_{1,ij}(u) = \log f(u)$.

Written in this way, the Jastrow ansatz can be interpreted as a highly restricted invariant architecture acting on the set of particle pairs: the feature map is hand-designed, the aggregation is a simple sum, and the readout is trivial. The Deep Sets ansatz introduced next preserves this invariant aggregation principle, but replaces the fixed pair map and the trivial post-pooling readout by learned transformations, thereby allowing the wave function to represent correlations beyond the standard pair-product level. In this sense, the KAN-enhanced architecture studied in this work can be viewed as an extension of the Jastrow idea: rather than optimizing a single hand-chosen pair factor, the model learns a latent pair representation before aggregation and then applies a learned readout after pooling. This additional embedding-and-readout structure is precisely what allows the ansatz to move beyond conventional pair-product forms and to encode more collective correlations when the physics requires them.

4.3. Formal KAN construction

We now turn from architectural motivation to the formal definition of the KAN layers used in this work. Kolmogorov–Arnold networks are grounded in the Kolmogorov–Arnold representation theorem. This theorem, proven independently by Kolmogorov and Arnold in the 1950s [25], establishes that any multivariate continuous function

$$f : [0, 1]^n \to \mathbb{R}$$

can be exactly represented as:

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right). \tag{4.3}$$

Here, $\phi_{q,p}$ and Φ_q are univariate continuous functions. This remarkable result demonstrates that complex multivariate functions can be decomposed into compositions of simpler univariate functions, providing the mathematical foundation for KAN architectures.

Unlike traditional multilayer perceptrons (MLPs) that place fixed activation functions on nodes and learnable parameters on edges, KANs invert this paradigm by placing learnable activation functions on the edges themselves. Each edge in a KAN is parameterized by a univariate function, typically implemented as a spline-based representation that can be trained end-to-end. This architectural choice transforms

the network from learning linear combinations followed by fixed nonlinearities to learning the nonlinear transformations directly. The resulting network can be written as

$$\mathrm{KAN}(x) = \mathrm{KAN}_{L-1} \circ \mathrm{KAN}_{L-2} \circ \cdots \circ \mathrm{KAN}_1 \circ \mathrm{KAN}_0(x), \tag{4.4}$$

where each layer KAN_l applies learnable univariate edge functions $\phi_{i,j}^{(l)}$ along the connections between successive neurons. This is fundamentally different from standard neural networks, which alternate matrix multiplications with fixed activation functions. This architectural innovation yields several key advantages particularly relevant to quantum many-body systems. Enhanced approximation capabilities emerge from the network's ability to learn optimal functional forms rather than being constrained to fixed activation functions, potentially requiring fewer parameters to represent complex relationships.

Improved interpretability results from the explicit representation of learned functions on each edge, allowing direct visualization and analysis of how the network processes different inputs—a crucial advantage for understanding physical correlations in quantum systems.

These properties make KANs particularly well-suited to address the fundamental challenges in Neural Quantum States, where the need to accurately represent complex many-body correlations, handle singular interactions, and maintain computational efficiency has proven challenging for conventional neural architectures. The ability of KANs to learn problem-specific functional forms suggests they may provide more efficient and accurate representations of quantum wave functions while offering insight into the underlying physics through their interpretable structure.

An additional advantage of KAN layers is the regularity of the represented functions. In particular, for the polynomial-based KAN layers considered in this work, each edge map is written as a finite expansion in smooth basis functions, so the resulting transformation is differentiable with respect to its inputs; in the Chebyshev-based case, it is simply a finite polynomial and is therefore smooth everywhere. This contrasts with common choices such as ReLU-based dense networks, whose activations are only piecewise linear and are not differentiable at the kink points. Such regularity is especially useful in neural quantum states because physically relevant quantities, including the kinetic-energy contribution to the local energy, depend explicitly on first and second derivatives of the wave function. Also KAN layers might be useful for analytical understanding of the functions in case of the weight analysis.

4.4. KAN-Enhanced Deep Sets

The variational architecture used in this work is based on the Deep Sets framework, which provides a natural permutation-invariant parametrization for bosonic systems. This is important at the methodological level because bosonic exchange symmetry should be built into the ansatz rather than recovered only approximately through optimization. Our modification of the standard Deep Sets construction is to replace the usual MLP-based embedding and readout networks by KAN layers, thereby retaining the invariant structure of Deep Sets while increasing the flexibility of the learned transformations[27, 28].

Theorem 4.1 (Deep Sets representation [27])**.** *Let F be a permutation-invariant function acting on a set of particle coordinates. Then, under suitable regularity assumptions, there exist functions ϕ and ρ such that*

$$F(\{x_1, \ldots, x_n\}) = \rho\left(\sum_{i=1}^{n} \phi(x_i)\right). \tag{4.5}$$

In other words, a permutation-invariant set function can be represented as a function of a summed elementwise embedding.

For bosonic many-body wave functions, this theorem provides the formal rationale for using Deep Sets: each particle is processed by a shared feature map ϕ, the resulting latent vectors are combined by a permutation-invariant sum, and the pooled representation is transformed by a readout map ρ. Figure 4.1 summarizes this structure schematically. In our setting, both ϕ and ρ are implemented with KAN blocks, so the symmetry structure of Deep Sets is preserved while the fixed nonlinearities of standard dense networks are replaced by learnable univariate edge functions.

The ansatz used here extends the basic one-body Deep Sets construction by adding a second invariant branch built from pair variables. More specifically, we use separate Deep Sets–KAN blocks for the one-body variables $\{x_i\}_{i=1}^{N}$ and for the pair variables $\{x_i - x_j\}_{i<j}$. Each branch therefore has its own embedding map, invariant aggregation, and post-pooling readout. Because the set of pair variables is itself permutation invariant, this extension preserves bosonic exchange symmetry while allowing the model to encode correlation structure that cannot be captured by purely additive single-particle embeddings.

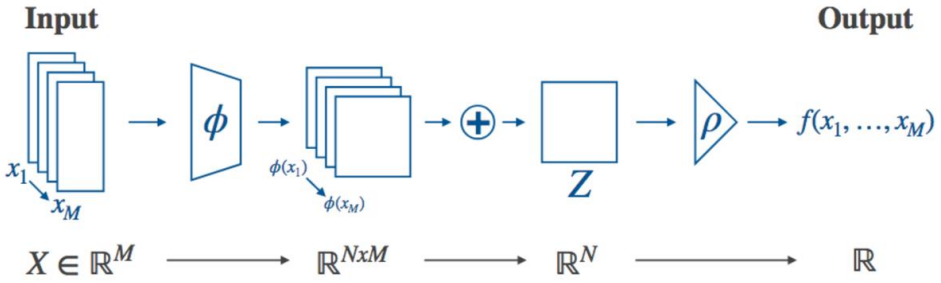


Figure 4.1: Schematic illustration of the Deep Sets architecture. Shared elementwise embeddings ϕ are applied to each particle, the resulting latent features are aggregated by a permutation-invariant sum, and the pooled representation is mapped by ρ to the final output. In the present work, the ϕ and ρ blocks are implemented with KAN layers.

Since the network parametrizes the logarithm of the wave function, the one-body and pair contributions enter additively rather than multiplicatively:

$$\log \Psi_\theta(X) = \rho^{(1)}_{\mathrm{KAN}}\left(\sum_{i=1}^{N} \phi^{(1)}_{\mathrm{KAN}}(x_i)\right) + \rho^{(2)}_{\mathrm{KAN}}\left(\sum_{1\le i<j\le N} \phi^{(2)}_{\mathrm{KAN}}(x_i - x_j)\right). \tag{4.6}$$

Here $\phi^{(1)}_{\mathrm{KAN}}$ and $\phi^{(2)}_{\mathrm{KAN}}$ map one-body and pair variables, respectively, to latent features, while $\rho^{(1)}_{\mathrm{KAN}}$ and $\rho^{(2)}_{\mathrm{KAN}}$ are the corresponding readout networks. In the translationally invariant systems considered in this work, both branches can be made physically meaningful. With an appropriate periodic embedding, the one-body branch is compatible with translational invariance and can capture physically relevant single-particle structure, while the pair branch encodes relative geometry explicitly and is therefore naturally suited to interaction-driven correlations. In this sense, the two branches should be viewed as complementary rather than mutually exclusive: the one-body branch provides a compact description of effective single-particle information, whereas the pair branch gives direct access to two-body structure.

The KAN components themselves are not restricted to the original spline construction. In the present work, the embedding networks $\phi^{(1)}_{\mathrm{KAN}}$, $\phi^{(2)}_{\mathrm{KAN}}$ and the readout networks $\rho^{(1)}_{\mathrm{KAN}}$, $\rho^{(2)}_{\mathrm{KAN}}$ are instantiated with several KAN layer families, including spline-based, Chebyshev-type, parameter-reduced radial-basis, and sinusoidal variants. At an abstract level, each such layer follows the same KAN principle that an

edge carries a learnable univariate map,

$$\mathrm{KAN}_{\mathrm{edge}}(x) = \sum_m c_m \, b_m(x), \tag{4.7}$$

where $\{b_m\}$ denotes the chosen basis family and c_m are learnable coefficients or effective edge parameters. The choice of basis determines the inductive bias of the layer: spline bases provide local resolution and explicit control over smoothness, polynomial bases are efficient for smooth global structure, radial-basis constructions support localized responses with reduced parameterization, and sinusoidal bases are natural for oscillatory or approximately periodic features. This viewpoint is central for our comparative study, since the main architectural differences considered in the numerical experiments arise from changing the edge basis while keeping the surrounding Deep Sets structure fixed.

4.5. Embeddings

The embedding is one of the main sources of inductive bias in the variational pipeline, because it determines which geometric relations are presented explicitly to the network and which must instead be reconstructed implicitly during optimization. In the present setting, the embedding must do more than encode particle coordinates: it must also remain compatible with bosonic exchange symmetry and with periodic boundary conditions.

Periodic geometry is handled through the minimum-image convention. Intuitively, the system is viewed as a periodic tiling of identical cells, and distances are measured using the shortest displacement between two particles across this periodic structure.

Permutation symmetry is enforced by constructing features either from single-particle coordinates or from pairwise coordinate differences, and then combining them through symmetric aggregation. Within this framework, we considered three embedding families, all designed to respect periodicity through sine–cosine encoding while differing in how much relative geometric information is made explicit to the network.

The simplest choice is a single-particle embedding, denoted by PeriodicEmbedding. In this case, each particle is represented independently

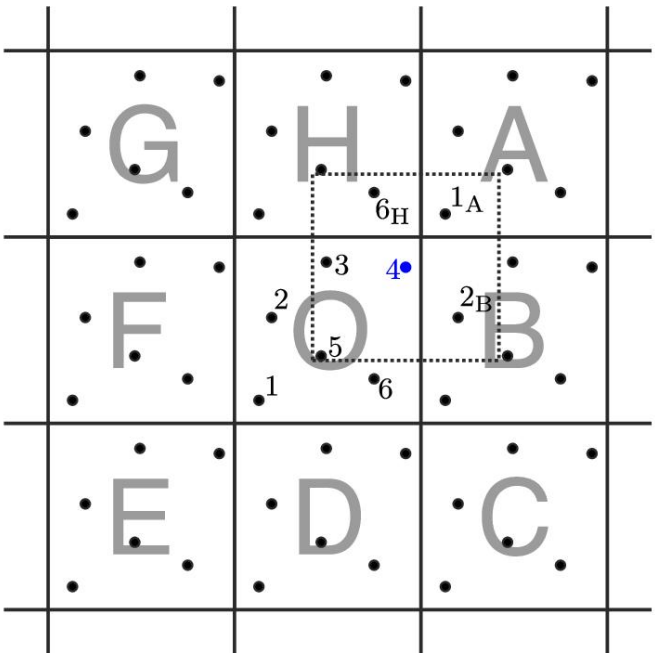


Figure 4.2: Illustration of the minimum-image convention used to encode periodic boundary conditions.

and the network must infer interparticle relations only in the subsequent layers:

$$\text{PeriodicEmbedding}(x_i) = \left[\sin\left(2\pi\frac{x_i}{L}\right), \cos\left(2\pi\frac{x_i}{L}\right)\right]. \tag{4.8}$$

For a d-dimensional system with N particles, this produces a feature space of size $2dN$. The normalization by L makes the trigonometric argument dimensionless and ensures that the embedding is exactly compatible with the periodic box geometry. This representation is simple and inexpensive, but it leaves the task of reconstructing relative distances entirely to the neural network.

A more informative alternative is a pair-based embedding built from minimum-image displacement vectors. We denote this representation by VectorPeriodicEmbedding:

$$\text{VectorPeriodicEmbedding}(x_i, x_j) = \left[\sin\left(2\pi\frac{\Delta x_{ij}^{\min}}{L}\right), \cos\left(2\pi\frac{\Delta x_{ij}^{\min}}{L}\right)\right], \tag{4.9}$$

where $\Delta x_{ij}^{\min}$ is the minimum-image displacement vector. In our implementation, its feature dimension scales as $dN(N-1)$ because we keep the full matrix of pairwise displacements rather than only its upper-triangular part. This redundancy is acceptable in practice because it is naturally compatible with permutation symmetry: exchanging $x_i \leftrightarrow x_j$ only permutes the pair-feature array and therefore leaves the

representation equivalent after symmetric aggregation. Although this embedding is more expensive, it provides a much stronger inductive bias for interaction-driven problems because relative geometry is encoded explicitly from the outset [29].

A third option is to compress the pair information further by keeping only the minimum-image pair distance and applying the trigonometric map to that scalar variable:

$$\text{DistancePeriodicEmbedding}(x_i, x_j) = \left[\sin\left(2\pi\frac{r_{ij}}{L_{\max}}\right), \cos\left(2\pi\frac{r_{ij}}{L_{\max}}\right)\right], \quad (4.10)$$

where $r_{ij} = \|\Delta x_{ij}^{\min}\|$ and $L_{\max} = L\sqrt{d}/2$ is the largest possible minimum-image distance in the box. The normalization by $L_{\max}$ rescales the admissible interval to a dimensionless argument in $[0, 1]$. In practice, however, this representation performed worse than both PeriodicEmbedding and VectorPeriodicEmbedding. One reason is that it discards directional information while still increasing the input size quadratically in the number of particles. Even for isotropic interactions, the network benefits from seeing the full displacement vectors because componentwise information helps reconstruct the many-body geometry. A second drawback is more specific to the trigonometric encoding itself: coordinates and displacement components are genuinely periodic variables, but the radial distance is not. Mapping $r_{ij} \in [0, L_{\max}]$ to a sine–cosine pair therefore introduces an artificial identification between $r_{ij} = 0$ and $r_{ij} = L_{\max}$, even though these correspond to maximally different physical situations. This aliasing weakens the inductive bias and makes optimization harder.

The comparison in Fig. 4.3 illustrates how these choices affect the final performance in the two-dimensional Gaussian problem. The main practical conclusion is that the vector-based periodic embedding provides the most useful balance between symmetry preservation and explicit geometric information. More broadly, this comparison shows that, for NQS in continuous space, the embedding is not merely a preprocessing step but an integral part of the variational design.

4.6. Optimization

The optimization of our neural quantum state (NQS) ansatz based on the Deep Set KAN architecture follows the variational Monte Carlo (VMC) framework[30, 31]. The many-body wave function $\Psi_\theta(\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_N)$ collects all particle coordinates, is parameterized by the KAN-enhanced Deep Sets network. The central

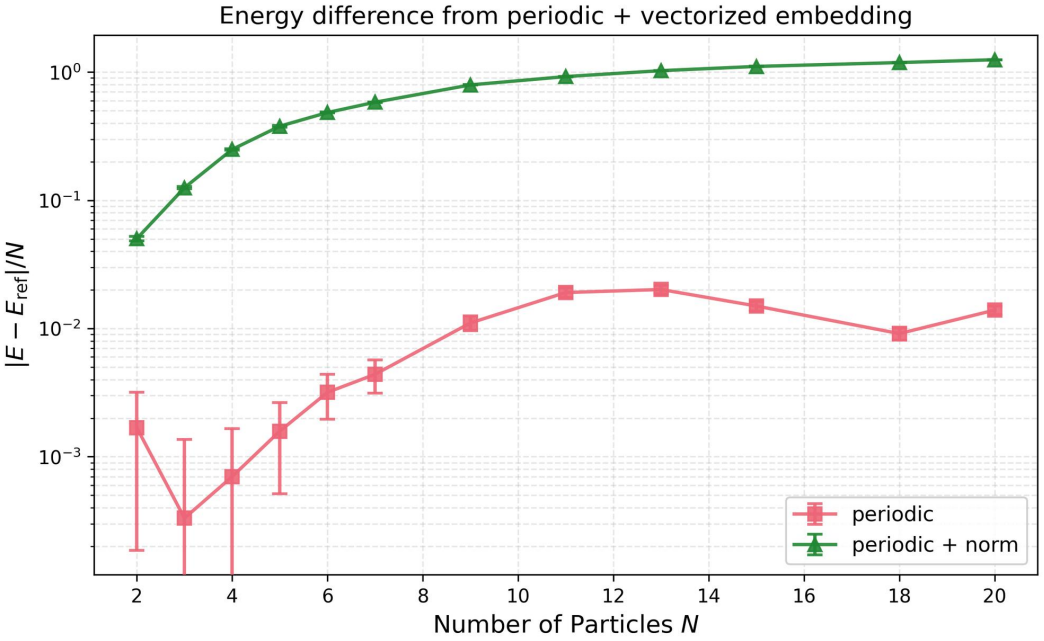


Figure 4.3: Comparison of embedding choices for the two-dimensional Gaussian problem.

objective is the variational energy

$$E(\theta) = \frac{\langle \Psi_\theta | H | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle}, \tag{4.11}$$

which is minimized with respect to the network parameters θ. Introducing the probability density induced by the ansatz,

$$p_\theta(\mathbf{x}) = \frac{|\Psi_\theta(\mathbf{x})|^2}{Z_\theta}, \qquad Z_\theta = \int d\mathbf{x}\, |\Psi_\theta(\mathbf{x})|^2, \tag{4.12}$$

the energy can be rewritten as an expectation value over sampled configurations,

$$E(\theta) = \int d\mathbf{x}\, p_\theta(\mathbf{x}) E_{\text{loc}}(\mathbf{x}), \qquad E_{\text{loc}}(\mathbf{x}) = \frac{(H\Psi_\theta)(\mathbf{x})}{\Psi_\theta(\mathbf{x})}. \tag{4.13}$$

In practice, the integral is replaced by a Monte Carlo estimator over configurations $\{\mathbf{x}^{(k)}\}_{k=1}^{M}$ drawn from p_θ,

$$\hat{E}(\theta) = \frac{1}{M} \sum_{k=1}^{M} E_{\text{loc}}\big(\mathbf{x}^{(k)}\big). \tag{4.14}$$

This converts the optimization of an exponentially complex wave function into the minimization of a stochastic objective that depends only on sampled particle configurations.

The quantities reported during optimization have a direct physical and numerical meaning. First, the **samples** $\mathbf{x}^{(k)}$ are particle configurations drawn from $p_\theta(\mathbf{x}) \propto |\Psi_\theta(\mathbf{x})|^2$. They are the basic data from which all observables are estimated. Good samples must explore the relevant regions of configuration space; otherwise, the estimated energy and gradients become biased or excessively noisy. For a generic observable $A(\mathbf{x})$, the Monte Carlo estimator is

$$\langle A \rangle \approx \frac{1}{M} \sum_{k=1}^{M} A\big(\mathbf{x}^{(k)}\big), \tag{4.15}$$

and its uncertainty decreases approximately as $1/\sqrt{M_{\text{eff}}}$, where $M_{\text{eff}} \leq M$ is the effective number of independent samples after accounting for autocorrelations. Thus, increasing the number and quality of samples directly improves the reliability of training.

Second, the **energy** $E(\theta)$ is the central optimization target because, by the variational principle, it provides an upper bound to the exact ground-state energy. A lower variational energy means that the ansatz describes the true ground state more accurately. During training, the energy therefore measures physical accuracy, while the change in energy from one iteration to the next is a practical indicator of whether optimization is still making progress.

Theorem 4.2 (Variational principle)**.** *Let H be a self-adjoint Hamiltonian with ground-state energy E_0. For any normalized trial state Ψ in the domain of H, satisfying $\langle \Psi \mid \Psi \rangle = 1$, the variational energy*

$$E[\Psi] = \langle \Psi \mid H \mid \Psi \rangle \tag{4.16}$$

satisfies

$$E[\Psi] \geq E_0. \tag{4.17}$$

This theorem provides the basic justification for the VMC approach used throughout this thesis: by minimizing $E[\Psi_\theta]$ over the parameters of the neural-network ansatz, we obtain a controlled upper bound on the true ground-state energy.

To compute parameter updates explicitly, it is convenient to introduce the logarithmic derivatives of the wave function,

$$O_i(\mathbf{x}) = \partial_{\theta_i} \ln \Psi_\theta(\mathbf{x}). \tag{4.18}$$

Differentiating the normalized energy functional yields the standard NQS gradient formula

$$g_i \equiv \partial_{\theta_i} E(\theta) = 2\,\mathrm{Re}\left[\langle O_i^* E_{\mathrm{loc}}\rangle_{p_\theta} - \langle O_i^*\rangle_{p_\theta}\langle E_{\mathrm{loc}}\rangle_{p_\theta}\right], \tag{4.19}$$

where $\langle\cdot\rangle_{p_\theta}$ denotes averaging over the Monte Carlo samples. For the real-valued wave functions considered here, this expression reduces to the covariance form used in stochastic-gradient optimization.

An equally important property for practical optimization is the zero-variance principle. Let

$$E_{\mathrm{loc}}(\mathbf{x}) = \frac{(H\Psi)(\mathbf{x})}{\Psi(\mathbf{x})} \tag{4.20}$$

denote the local energy associated with a trial wave function Ψ. Then the variance

$$\mathrm{Var}(E_{\mathrm{loc}}) = \left\langle \big(E_{\mathrm{loc}} - E[\Psi]\big)^2 \right\rangle \tag{4.21}$$

vanishes if and only if Ψ is an exact eigenstate of the Hamiltonian on the support sampled from $|\Psi|^2$. Therefore, the decay of the local-energy variance during training is not only a numerical indicator of optimization stability, but also a direct diagnostic of how close the ansatz is to an exact stationary state.

This makes the **variance** one of the most informative diagnostics in NQS optimization. Two trial states can have similar variational energies, yet the state with smaller $\mathrm{Var}(E_{\mathrm{loc}})$ is usually closer to a true eigenstate and yields more stable gradients. In practice, a high variance signals that the ansatz still misses important correlations or cusp structure, whereas a low variance indicates both better physical fidelity and better-conditioned optimization.

When sampling is performed with several independent Markov chains, an additional convergence diagnostic is the **potential scale reduction factor** $\hat{R}$. Suppose C chains are run, each producing n post-warmup measurements of some observable A. Denote by $\bar{A}_c$ the mean within chain c, by $\bar{A}$ the average over chains, and define

the within-chain and between-chain variances as

$$W = \frac{1}{C}\sum_{c=1}^{C} s_c^2, \qquad B = \frac{n}{C-1}\sum_{c=1}^{C}(\bar{A}_c - \bar{A})^2, \tag{4.22}$$

where s_c^2 is the sample variance inside chain c. An estimator of the target variance is then

$$\hat{V} = \frac{n-1}{n}W + \frac{1}{n}B, \qquad \hat{R} = \sqrt{\frac{\hat{V}}{W}}. \tag{4.23}$$

Values $\hat{R} \approx 1$ indicate that different chains have mixed well and are sampling the same distribution, while noticeably larger values signal insufficient equilibration or poor exploration of configuration space. In the context of NQS, monitoring $\hat{R}$ is important because inaccurate sampling propagates directly into the estimated energy, variance, and parameter gradients.

Another practical diagnostic is the **acceptance rate** of the Monte Carlo sampler. In a Metropolis–Hastings update, a proposed move from configuration $\mathbf{x}$ to $\mathbf{x}'$ is accepted with probability

$$a(\mathbf{x} \to \mathbf{x}') = \min\left(1, \frac{|\Psi_\theta(\mathbf{x}')|^2\, T(\mathbf{x} \mid \mathbf{x}')}{|\Psi_\theta(\mathbf{x})|^2\, T(\mathbf{x}' \mid \mathbf{x})}\right), \tag{4.24}$$

where $T(\mathbf{x}' \mid \mathbf{x})$ is the proposal distribution. For symmetric proposals this simplifies to

$$a(\mathbf{x} \to \mathbf{x}') = \min\left(1, \frac{|\Psi_\theta(\mathbf{x}')|^2}{|\Psi_\theta(\mathbf{x})|^2}\right). \tag{4.25}$$

If N_{prop} moves are proposed and N_{acc} are accepted, the empirical acceptance rate is

$$A_{\text{acc}} = \frac{N_{\text{acc}}}{N_{\text{prop}}}. \tag{4.26}$$

This quantity is important because it measures how efficiently the Markov chain explores configuration space during training. If the acceptance rate is too low, the chain becomes stuck and successive samples are strongly correlated, which reduces the effective sample size and increases the noise of energy and gradient estimates. If it is too high, the proposal steps are often too small, so the chain moves only locally and explores the distribution slowly. Monitoring the acceptance rate during training therefore helps tune the proposal scale and maintain a good balance between local

acceptance and global exploration.

Gradients are computed using the standard log-derivative trick,

$$\nabla_\theta E = 2 \left\langle \big(E_{\mathrm{loc}} - \langle E \rangle\big) \nabla_\theta \ln \Psi_\theta \right\rangle , \tag{4.27}$$

where E_{loc} is the local energy evaluated on each configuration. Thus, a simple stochastic-gradient step takes the form

$$\theta^{(t+1)} = \theta^{(t)} - \eta_t \, \widehat{\nabla_\theta E}, \tag{4.28}$$

with learning rate η_t and Monte Carlo estimate $\widehat{\nabla_\theta E}$. A common refinement in NQS optimization is stochastic reconfiguration (SR), or natural-gradient descent, in which the update direction is preconditioned by the covariance matrix of the log-derivatives,

$$S_{ij} = \langle O_i^* O_j \rangle_{p_\theta} - \langle O_i^* \rangle_{p_\theta} \langle O_j \rangle_{p_\theta} , \tag{4.29}$$

and the parameter increment is obtained from

$$\sum_j S_{ij} \, \Delta\theta_j = -\eta_t \, g_i. \tag{4.30}$$

This update follows the geometry of the variational manifold more closely than bare gradient descent and is widely used to stabilize NQS training. In our implementation, the actual parameter updates are performed with adaptive stochastic-gradient optimizers, while the SR form provides the natural preconditioning picture for the optimization landscape discussed below.

A critical ingredient in our optimization pipeline is a **pretraining strategy based on iterative particle growth**. Instead of training the network directly on the full N-particle system, we initialize training with a smaller particle number and progressively add particles one by one. At each stage, the network parameters from the previous training cycle are used to initialize the wave function of the larger system. This hierarchical pretraining scheme significantly accelerates convergence, as the network leverages previously learned correlations while gradually adapting to the increased system size.

All systems are studied at **unit density**, if not stated otherwise. It is defined such that one particle occupies one unit of spatial length. This constraint ensures

that comparisons across different particle numbers remain physically consistent and that the pretraining procedure respects the underlying density of the system.

Through the combination of large-batch stochastic optimization, iterative particle-growth pretraining, and the KAN-enhanced Deep Sets ansatz, our approach achieves stable and efficient training of neural quantum states even in regimes where traditional NQS methods face severe optimization bottlenecks.

Through the combination of large-batch stochastic optimization, iterative particle-growth pretraining, and the KAN-enhanced Deep Sets ansatz, our approach achieves stable and efficient training of neural quantum states even in regimes where traditional NQS methods face severe optimization bottlenecks.

In our work, we assess both the convergence and the quality of the NQS ansatz using a consistent set of diagnostics: the potential scale reduction factor $\hat{R}$, the Monte Carlo acceptance rate, the statistical errors on the estimated observables, the local-energy variance, and the convergence of the variational energy itself. In all reported results, we require $\hat{R} < 1.01$, which indicates good agreement between independent sampling chains. The acceptance rate is typically maintained in the range 0.4–0.7; in practice, it tends to be larger during the early stages of training and for smaller particle numbers, and then gradually decreases as the system size grows. The statistical uncertainties are controlled to remain below 0.1% of the reported energies, and in many cases they are substantially smaller. We also monitor the variance relative to the number of available samples to ensure that the estimators remain reliable throughout optimization. Finally, for all experiments we verify that the energy curves reach a stable plateau and that the main conclusions are reproducible across different random seeds, different numbers of layers, and different architectural choices.

4.7. Layer types

The behavior of a KAN model depends strongly on the basis used to parameterize its edge functions, so both the expressive power and the optimization stability of the network are layer-dependent. Different bases impose different inductive biases: spline bases provide local support and flexible resolution, polynomial bases favor smooth global approximations, and sinusoidal or radial bases are naturally suited to oscillatory or sharply localized structure. As a result, the basis choice affects not only which functions can be represented efficiently, but also how well the corresponding parameters can be trained with stochastic gradient methods.

In this work, we compare five layer types inside the same KAN-enhanced Deep Sets pipeline: the original spline layer, a modified Chebyshev layer, parameter-reduced KAN (PRKAN), a sine layer, and a lean Chebyshev layer. This comparison is important because all other components of the variational ansatz are kept fixed, so differences in final performance can be attributed primarily to the representation used for the learnable one-dimensional edge functions. In particular, we are interested in the trade-off between approximation quality, parameter efficiency, and robustness of optimization.

At a high level, these layers fall into three groups. The spline layer is the closest to the original KAN construction and offers strong local flexibility. The modified Chebyshev and lean layers use polynomial expansions and therefore provide a compact description for smooth functions, with different levels of parameterization. PRKAN replaces polynomial features by radial basis functions together with an explicit low-rank parameterization, while the sine layer uses a harmonic basis that is naturally compatible with periodic structure. The corresponding definitions are summarized below.

4.7.1 Spline layer

The spline layer is the canonical KAN layer introduced in the original KAN construction. For each input–output pair, the edge function is written as

$$\psi_{ij}(x) = w_{ij}^{\mathrm{res}}\,\mathrm{SiLU}(x) + w_{ij}^{\mathrm{spl}} \sum_{m} c_{ijm} B_m(x), \tag{4.31}$$

where $B_m(x)$ are B-spline basis functions on a predefined grid and c_{ijm} are learnable coefficients. In our implementation, the coefficient tensor has shape $n_{\mathrm{out}} \times n_{\mathrm{in}} \times (G + k)$, where G is the number of grid intervals and k is the spline order.

This layer can be viewed as an interpolation between a standard MLP and a full KAN: when the spline contribution is small, the residual SiLU term dominates, whereas large spline weights allow the network to learn a genuinely nonparametric edge function. The main advantage of the spline layer is its strong local adaptability. Its main drawback is the relatively large number of parameters and the sensitivity of the representation to the grid resolution and knot placement.

4.7.2 Modified Chebyshev layer

The modified Chebyshev layer replaces local spline bases by a global polynomial basis and augments it with an additional learnable edge-wise scaling factor. After applying a bounded transformation to the inputs,

$$\tilde{x}_j = \tanh(x_j), \tag{4.32}$$

the layer evaluates the Chebyshev basis up to a fixed maximal degree k and forms the output as

$$y_i = \sum_{j=1}^{n_{\text{in}}} s_{ij} \sum_{m=0}^{k} c_{ijm}\, T_m(\tilde{x}_j), \tag{4.33}$$

where T_m denotes the Chebyshev polynomial of degree m, c_{ijm} are learnable expansion coefficients, and s_{ij} is an additional learnable edge-wise scaling parameter. Keeping the maximum degree static preserves fixed parameter shapes and simplifies implementation, while the effective polynomial order can still be adjusted within this representation. In practice, this layer retains the efficiency of Chebyshev features but adds extra flexibility through the additional activation weights. Compared with spline layers, the Chebyshev basis is more compact for smooth target functions, but its global nature can make it less effective when the edge functions develop sharply localized structure.

4.7.3 PRKAN

In this work, the best-performing variant is the parameter-reduced KAN (PRKAN), which is based on radial basis functions rather than splines or polynomials [32]. We use a modified implementation that preserves the KAN idea of learnable edge functions while introducing an explicit low-rank factorization of the basis coefficients.

The radial centers are placed uniformly on a grid from $\mu_0 = -1$ to $\mu_1 = 1$,

$$\mu_g = \mu_0 + g\frac{\mu_1 - \mu_0}{G - 1}, \qquad g = 0, \ldots, G - 1, \tag{4.34}$$

and the initial bandwidth is chosen as

$$\gamma = \frac{\mu_1 - \mu_0}{1.5\,(G - 1)}. \tag{4.35}$$

The corresponding RBF features are

$$B_{blg} = \exp\left[-\left(\frac{x_{bl} - \mu_g}{\gamma}\right)^2\right]. \tag{4.36}$$

The basis contribution is then written as

$$y_{\text{basis}}[b, o] = \sum_{l,g} \Big(a[o, l]\ \text{softmax}\big(c_{\text{amp}}[o, g]\big) \Big) B[b, l, g], \tag{4.37}$$

which corresponds to the factorization

$$c[o, l, g]\ \rightarrow\ a[o, l]\ \text{softmax}\big(c_{\text{amp}}[o, g]\big). \tag{4.38}$$

The residual branch is

$$y_{\text{res}}[b, o] = \sum_{l} c_{\text{res}}[o, l]\, \text{SiLU}(x[b, l]), \tag{4.39}$$

and the final output is

$$y_{\text{layer}} = y_{\text{basis}} + y_{\text{res}}. \tag{4.40}$$

This construction separates the mixing across input channels from the distribution of basis amplitudes over the RBF grid, which substantially reduces the number of free parameters. Empirically, this layer provides the best overall balance between expressivity and optimization stability in our benchmarks, especially when the wave function must resolve both smooth collective behavior and localized correlation features.

4.7.4 Sine layer

The sine layer uses a harmonic expansion for each input channel,

$$y_{b,o} = \sum_{l=1}^{n_{\text{in}}} \sum_{k=1}^{K} A_{o,l,k} \sin(\omega_{o,l,k} x_{b,l} + \varphi_{o,l,k}) + \sum_{l} c_{o,l}^{\text{res}}\, \text{SiLU}(x_{b,l}), \tag{4.41}$$

with learnable amplitudes, frequencies, and phases

$$A, \omega, \varphi \in \mathbb{R}^{n_{\text{out}} \times n_{\text{in}} \times K}. \tag{4.42}$$

Unlike PRKAN, this parameterization does not use an explicit low-rank factorization. The main motivation for the sine layer is that periodic systems often contain oscillatory structure that may be represented naturally in a trigonometric basis. However, in our setting the learned edge functions are not purely periodic objects, and the simultaneous optimization of amplitudes, frequencies, and phases tends to be less stable than in PRKAN or Chebyshev-type layers.

4.7.5 Lean layer

The lean layer is a parameter-efficient Chebyshev-based KAN layer [33]. Instead of introducing additional activation scalers or auxiliary edge parameters, it keeps only the coefficients of the basis expansion as trainable parameters. For an input x, the layer first applies a bounded transformation such as $\tanh(x)$ for numerical stability and then builds the Chebyshev basis $T_0(x), \ldots, T_k(x)$ recursively up to a fixed degree k. The output is obtained by a linear combination of these basis functions with learned coefficients for every input–output pair,

$$y_i = \sum_{j=1}^{n_{\text{in}}} \sum_{m=0}^{k} c_{ijm}\, T_m(\tanh(x_j))\,. \tag{4.43}$$

In this way, the lean layer preserves the main approximation mechanism of polynomial KANs while reducing both the parameter count and the computational overhead, which makes it a useful lightweight baseline in our layer comparison. Relative to the modified Chebyshev layer, it sacrifices some flexibility in exchange for a simpler and more regular parameterization.

Figures 4.4 and 4.5 summarize the empirical comparison on the common one-dimensional Gaussian benchmark. The main qualitative conclusion is that the basis choice matters already at the level of a single hidden-layer building block: layers with a compact but stable parameterization consistently train better than either overly rigid or overly flexible alternatives. Among the tested options, PRKAN gives the most favorable overall trade-off, reaching the lowest energies together with competitive variance, which is why it is used as the default layer in the remainder of this work.

The figures below compare the performance of several KAN layer types on the same benchmark problem under identical training conditions. As a common refer-

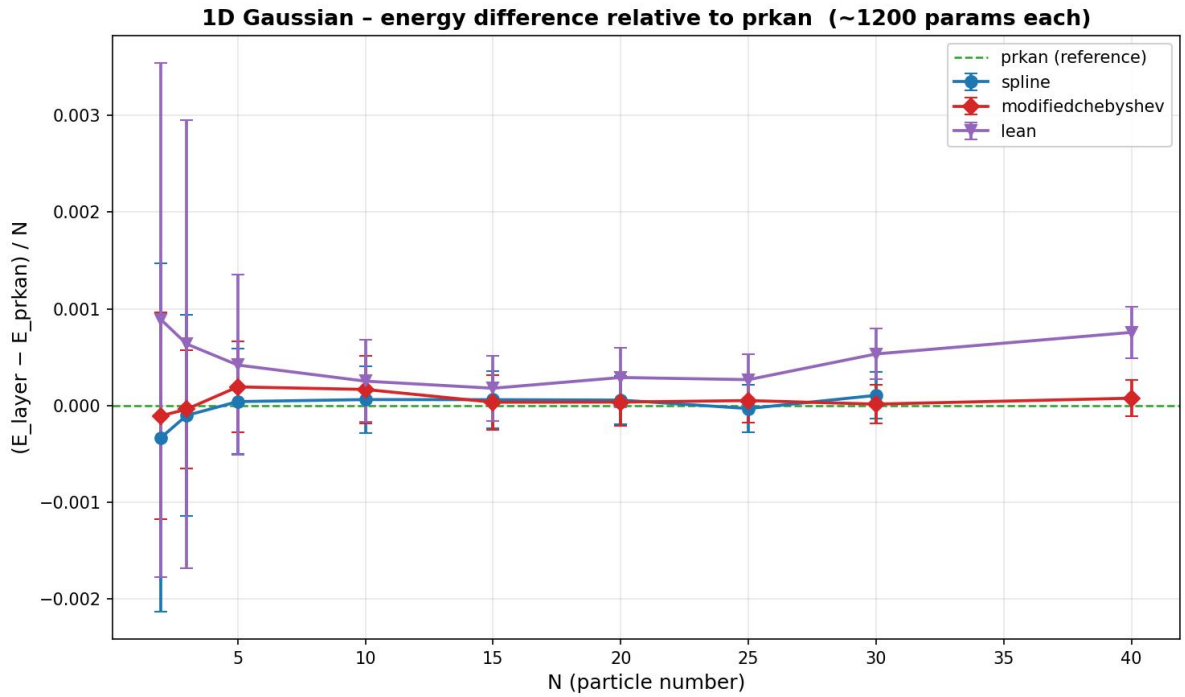


Figure 4.4: Energy difference relative to the PRKAN baseline for different layer types.

ence case, we use the one-dimensional Gaussian-interaction model with the default parameter settings.

4.8. Cusp

One of the main difficulties in applying neural quantum states to continuum many-body systems is the treatment of short-distance singularities. When the interaction diverges at particle coincidence, the kinetic and potential terms in the local energy separately become singular, and these singular pieces cancel only if the wave function has the correct non-analytic behavior at coalescence. This requirement is usually referred to as a cusp condition. If the ansatz does not reproduce it, the local energy develops large spikes, the variance increases dramatically, and variational optimization becomes unstable.

This problem is particularly important for the one-dimensional Calogero–Sutherland model[1], whose inverse-square interaction imposes a specific power-law suppression when two particles approach one another. Near a coincidence point, with relative distance $r_{ij} = |x_i - x_j|$, the exact many-body wave function behaves as

$$\Psi(X) \propto \prod_{i<j} r_{ij}^{\lambda} \qquad (r_{ij} \to 0), \tag{4.44}$$

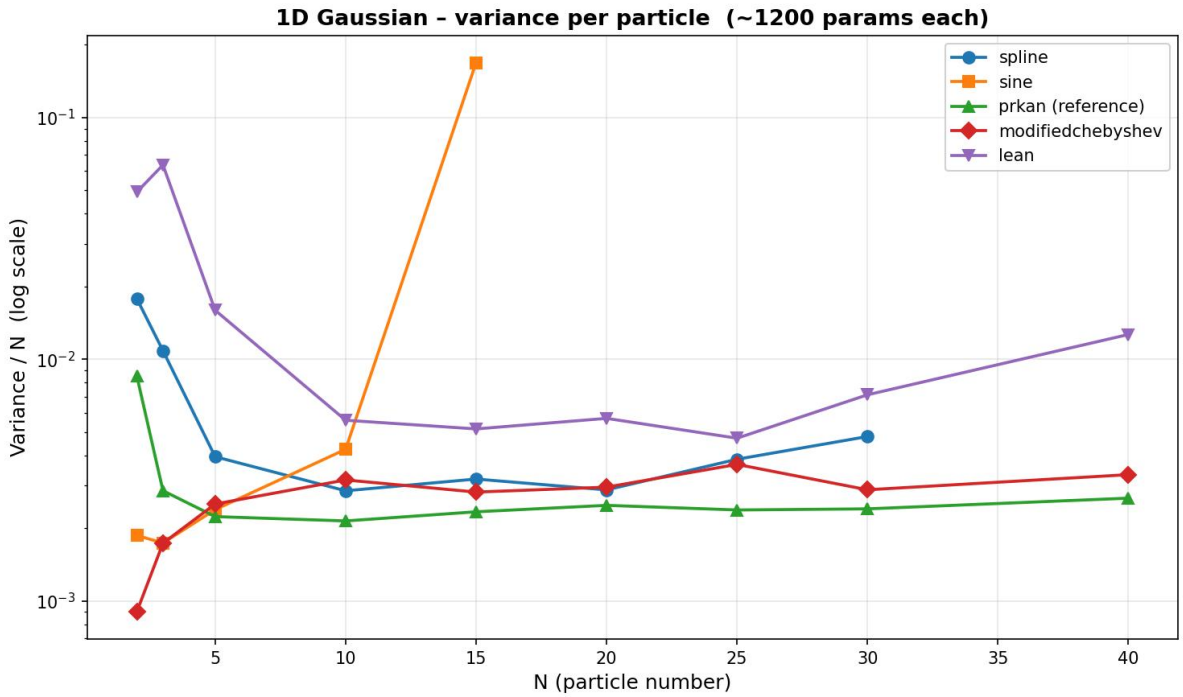


Figure 4.5: Energy variance for different layer types.

where the exponent is fixed by the coupling strength. A generic neural network with smooth activations does not reproduce this structure automatically, so asking the network to learn it from data alone is both inefficient and numerically fragile.

To address this issue, we use a cusp-aware KAN–Deep Sets ansatz in which the singular short-distance structure is built in explicitly and the neural network models only the remaining smooth part of the wave function. Concretely, we write

$$\log \Psi_\theta(X) = \log J_{\text{cusp}}(X) + f_\theta(X), \tag{4.45}$$

where J_{cusp} is a fixed physics-informed prefactor and f_θ is the learnable network output. This decomposition is advantageous for two reasons. First, it guarantees the correct leading-order behavior at particle contact. Second, it regularizes the learning problem, because the network no longer needs to approximate a non-analytic feature with smooth basis functions.

For the one-dimensional Calogero–Sutherland problem, we use a Jastrow cusp factor of the form

$$J_{\text{cusp}}(X) = \prod_{i<j} \left[\tanh\left(\pi \frac{|x_i - x_j|}{L} \right) \tanh\left(\pi \left(1 - \frac{|x_i - x_j|}{L} \right) \right) \right]^\lambda, \tag{4.46}$$

with

$$\lambda = \frac{1}{2}\left(1 + \sqrt{1 + 4mg}\right). \tag{4.47}$$

This periodic form is chosen so that the same short-distance suppression appears both when $|x_i - x_j| \to 0$ and when particles approach one another through the periodic image, $|x_i - x_j| \to L$. Indeed, for small separations,

$$\tanh\left(\pi \frac{|x_i - x_j|}{L}\right) \sim \pi \frac{|x_i - x_j|}{L}, \tag{4.48}$$

so the prefactor reproduces the expected power-law cusp locally. The second hyperbolic tangent plays the analogous role for the opposite side of the ring and therefore makes the construction compatible with periodic boundary conditions.

Since the network parametrizes the logarithm of the wave function, the contribution actually added to the ansatz is

$$\begin{aligned} \log J_{\text{cusp}}(X) = \lambda \sum_{i<j} \Bigg[& \log \tanh\left(\pi \frac{|x_i - x_j|}{L}\right) \\ & + \log \tanh\left(\pi \left(1 - \frac{|x_i - x_j|}{L}\right)\right) \Bigg]. \end{aligned} \tag{4.49}$$

In this representation, the non-analytic contribution is handled analytically, while the network is responsible for the regular remainder. As a result, the local energy is much better behaved during Monte Carlo sampling and optimization.

At the architectural level, we further support this construction by providing the model with explicit pairwise distances $|x_i - x_j|_{i<j}$ and by using KAN layers whose basis resolution is biased toward the short-distance regime. This is important because even after factoring out the exact leading cusp, the residual function still varies most rapidly when particles are close. Concentrating representational capacity in that region improves efficiency and reduces the burden on the optimizer.

Overall, the cusp-aware ansatz combines analytic prior knowledge with a flexible neural representation: the prefactor enforces the correct singular structure, while the KAN-enhanced Deep Sets network captures the remaining smooth many-body correlations. In practice, this hybrid design is essential for stable optimization of the Calogero–Sutherland problem and substantially reduces both the variance of the local energy and the difficulty of training in the singular regime.

4.9. Grid update

Grid updates are available only for spline-based KAN models, because the learned edge functions are represented on an explicit B-spline knot grid. In this setting, a layer represents a univariate map through spline basis functions defined on a grid of knots. The basic idea of a grid update is to rebuild this knot grid from the current sample distribution and then refit the spline coefficients so that the represented function changes as little as possible. In principle, this should allocate more resolution to regions of configuration space where the wave function varies rapidly[34].

Operationally, the procedure has two parts. First, a new knot vector is constructed from the currently sampled inputs, typically mixing a data-adaptive grid with a small uniform component for numerical stability. A compact way to write this update is

$$g_r^{\mathrm{new}} = \varepsilon\, g_r^{\mathrm{unif}} + (1-\varepsilon)\, g_r^{\mathrm{adapt}}, \qquad r = 0, \ldots, G,$$

where g_r^{adapt} is obtained from the empirical sample quantiles and g_r^{unif} is the corresponding uniform reference grid. Second, after replacing the old grid by the new one, the spline coefficients are re-estimated by a least-squares fit,

$$c' = \arg\min_{\tilde{c}} \|B'(X)\tilde{c} - B(X)c\|_2^2,$$

so that the function represented on the new grid approximates the old one on the current batch. In this way, the discretization is adapted without reinitializing the layer from scratch.

In the present study, however, this mechanism did not provide a clear advantage over the corresponding fixed-grid model. We compared training with and without grid updates for the two-dimensional Gaussian benchmark while keeping the remaining parts of the architecture unchanged. As shown in Fig. 4.6, the adaptive-grid version does not yield a systematic improvement in the final variational performance. For this reason, grid updates should be viewed here as an implementation option rather than as a key ingredient of the method.

The main practical conclusion is therefore negative but useful: the mechanism is conceptually reasonable and remains available in the implementation, but within the current framework the simpler no-grid-update setup is sufficient and was used

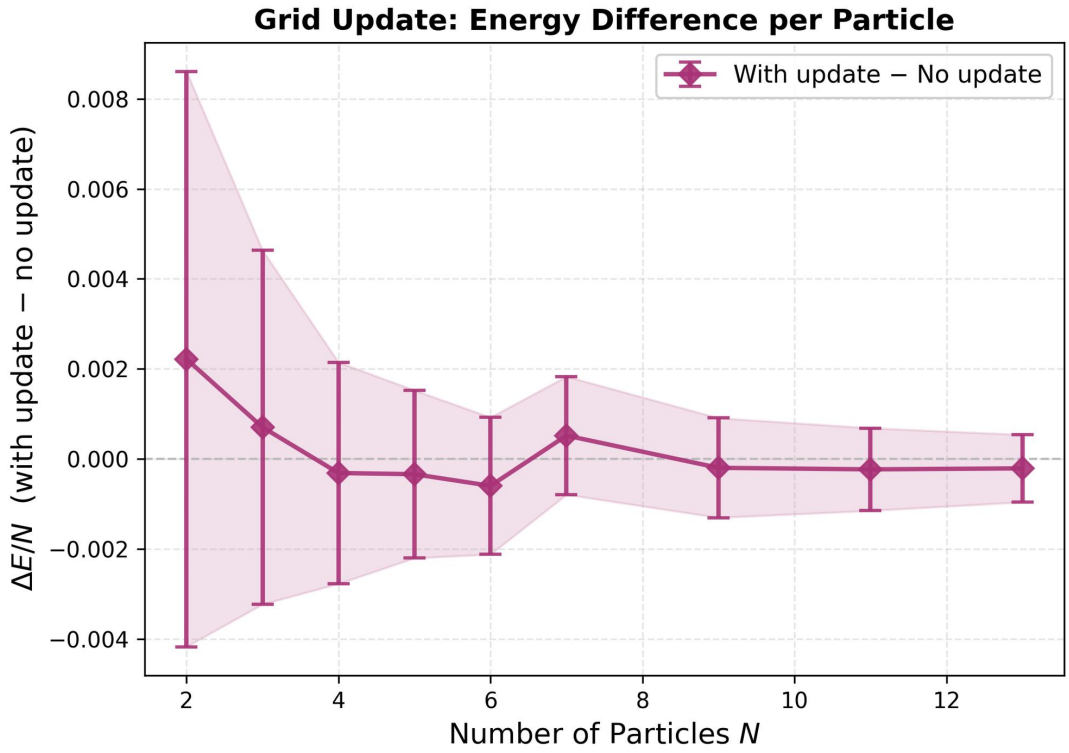


Figure 4.6: Comparison between training with and without spline-grid updates. For the benchmark considered here, the adaptive-grid variant does not produce a clear improvement over the fixed-grid baseline.

as the main reference in the later experiments.

4.10. Memory Requirements

We have got many times into a situation where there is not enough memmory for running the script. In our current implmentation we use Jastrow anzatz with MLP correelation functions. Number of paramters in the dnese network which have to be saved in the RAM to run the network. VecPeriodic embedding generates $N(N-1)/2$ particles from M and 2 dim, so input of the network transforms from batch,N,dim to batch,$N(N-1)/2$,2 dim. For each of the input the network is used so overall $N(N-1)/2$ copies of network is used. The dense network amount of parameters, where n_l is the width of layer l, is

$$P_{\mathrm{MLP}} = \sum_{l=1}^{L} \left(n_{l-1} n_l + n_l\right), \tag{4.50}$$

for the Jastrow network the $n_0 = 2\,\mathrm{dim}$ and the last layer has width 1. For a KAN network, the parameter count depends on how many coefficients are used to represent a single edge function. If the l-th layer has width $n_{l-1} \to n_l$ and each edge

in that layer is parameterized by m_l learnable coefficients, then the total number of KAN parameters can be written as

$$P_{\mathrm{KAN}} = \sum_{l=1}^{L} n_{l-1} n_l \, m_l, \tag{4.51}$$

up to additional bias or residual parameters that depend on the specific layer implementation. For the original spline-based KAN, one typically has $m_l \approx G_l + k_l$, where G_l is the number of grid intervals and k_l is the spline order, so that

$$P_{\mathrm{KAN}}^{\mathrm{spline}} \approx \sum_{l=1}^{L} n_{l-1} n_l \, (G_l + k_l). \tag{4.52}$$

4.11. Adiabatic evolution

For some Hamiltonians, direct optimization of a neural quantum state becomes unreliable when the target problem is too difficult from the outset. In our Gaussian benchmark family, this happens when the interaction parameter ϵ is large. For moderate values, such as $\epsilon \approx 0.5$, the variational Monte Carlo optimization converges without major difficulty. By contrast, when training starts directly at $\epsilon = 20$, the optimization often becomes trapped in poor local minima, the sampled configurations are not representative of the true ground state, and the final variational error remains large.

To improve robustness, we use an adiabatic-continuation strategy in Hamiltonian space. The idea is not to simulate real-time quantum dynamics, but to solve a sequence of nearby variational problems[35],

$$H\big(\epsilon_0\big), H\big(\epsilon_1\big), \ldots, H\big(\epsilon_M\big), \tag{4.53}$$

with gradually increasing interaction strength. After optimizing the parameters θ_k for $H(\epsilon_k)$, we use them to initialize the next optimization at $H(\epsilon_{k+1})$. In this way, the learned wave function is transported through parameter space from an easy regime to a hard one.

In the experiments reported here, we start from $\epsilon = 0.5$ and increase the interaction strength uniformly up to $\epsilon = 20$ using 10 intermediate values. This continuation makes the overall procedure slower than a single direct optimization, but it is much

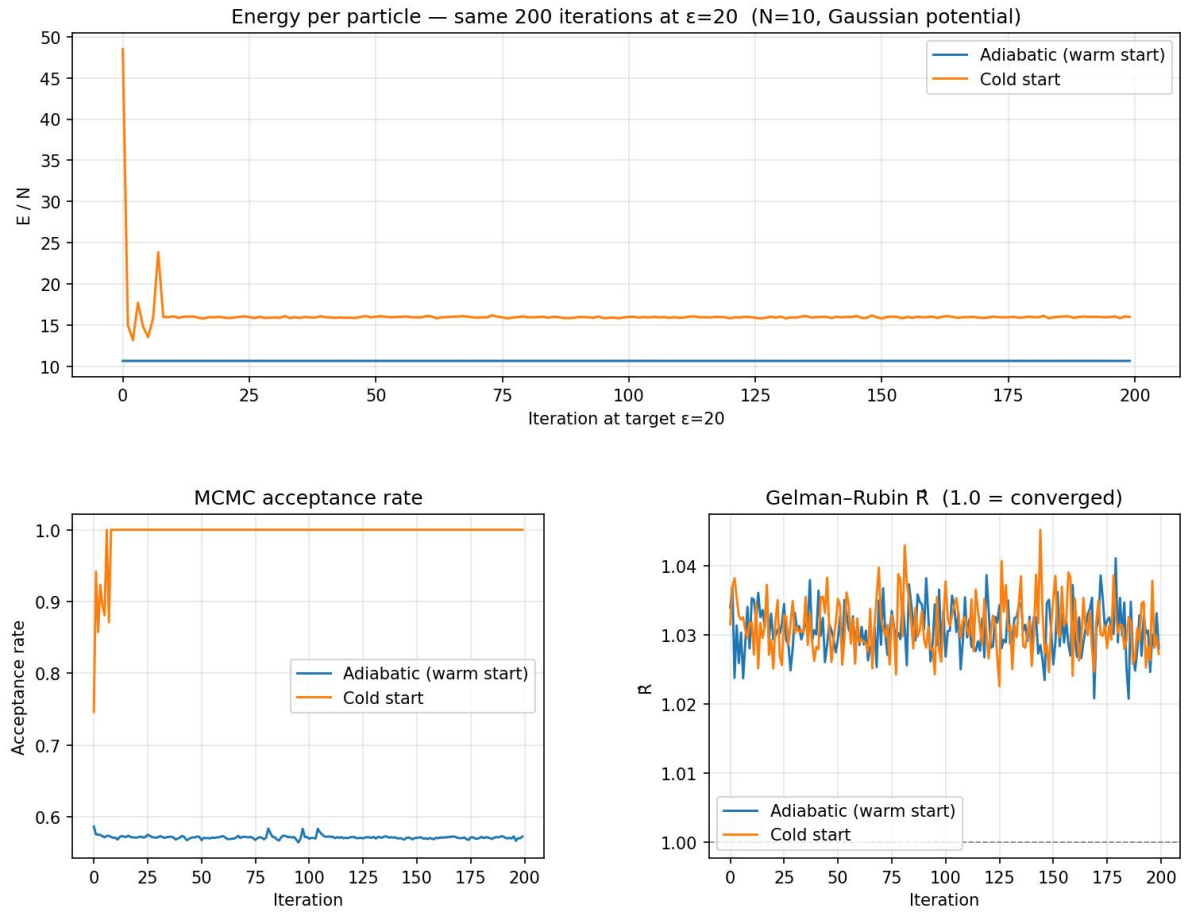


Figure 4.7: Comparison between direct training and adiabatic continuation for the Gaussian benchmark at large interaction strength. Initializing each run from the optimized parameters at a nearby value of ϵ leads to more stable convergence than training from a random initialization directly at $\epsilon = 20$.

more stable. The benefit is that each new stage begins from parameters that already encode the qualitative structure of the wave function, so the optimizer only needs to adapt to an incremental deformation of the Hamiltonian rather than discover the full solution from scratch.

Conceptually, the Hamiltonian parameter plays the role of an effective adiabatic time. When the steps in ϵ are sufficiently small, the variational optimum evolves smoothly, and the optimization can track this path through model space. In our Gaussian tests, this continuation procedure succeeds in regimes where random initialization fails, showing that adiabatic evolution is a practical training heuristic for difficult continuous-space NQS problems.

4.12. Main term and series

Because KAN layers represent the learned map as a sum of elementary contributions, it is natural to ask how much of this capacity is actually needed after training. To answer this question, we perform a post-training truncation analysis in which the learned weight matrices are projected onto a lower-dimensional subspace. This allows us to separate the dominant contribution of the model–the "main term" carried by the leading modes–from the higher-order corrections contained in the remaining series of smaller components.

Operationally, we take a trained Deep Sets–KAN model, compute a low-rank approximation of the relevant learned matrices, and then reevaluate observables without retraining. If a truncated model reproduces the same physical quantities as the full model, this indicates that the underlying wave function is effectively low dimensional. If the observables degrade quickly under truncation, then a larger number of learned modes is essential. In schematic form, this analysis amounts to replacing a learned matrix W by a truncated approximation

$$W^{(r)} = U_r \Sigma_r V_r^\top, \tag{4.54}$$

where only the first r dominant singular modes are retained.

The workflow is therefore straightforward: we first train the full model, then truncate it to a chosen effective rank, and finally recompute the correlation functions and energies. The resulting observables reveal how much of the learned many-body structure is already captured by the leading modes and how this dependence changes with system size.

Figure 4.8 shows the correlation functions obtained after truncation for several particle numbers. For smaller systems, the correlation profiles remain close to those of the full model even after a substantial reduction of the effective rank. This indicates that a relatively small number of dominant modes already captures the main many-body structure. For larger systems, however, the curves become more sensitive to truncation and develop visibly stronger distortions, which implies that the learned representation must retain more modes to describe the state accurately.

The energy comparison in Figure 4.9 leads to the same conclusion. The normalized energy is fairly insensitive to truncation for smaller particle numbers, but it deteriorates more rapidly for larger systems as the effective rank is reduced. Taken together, these results suggest that the model contains a genuine low-rank core, yet

the number of physically relevant modes grows with system size. This observation is consistent with the idea that small bosonic systems can be described by a compact set of collective features, whereas larger systems require a richer hierarchy of correlations.

4.13. Activations

One advantage of KAN-based models over standard MLPs is that the learned one-dimensional edge functions can be inspected directly. This makes it possible to analyze not only the final variational energy, but also the internal functional structure that the model uses to represent correlations. In the present setting, such an analysis is especially useful because the network operates on pair-based inputs: by examining the learned edge maps, we can see which regions of pair distance are used most strongly and how this dependence changes with the particle number.

For the one-dimensional Gaussian benchmark, the most informative object is the family of learned activations in the pair-feature map. Since the relevant input is a periodic pair distance, it is sufficient to display the functions only on the fundamental interval $[0, 1]$. This representation is natural because distances outside this interval are redundant under the minimum-image convention used in the periodic geometry.

Figure 4.10 complements this picture by plotting differences between learned activations. This provides a more sensitive diagnostic of convergence than the raw curves themselves. Small differences indicate that the internal representation has stabilized, while larger deviations highlight channels in which the model is still reorganizing its description of the many-body state. In practice, this analysis shows that part of the learned basis is shared across system sizes, whereas another part adapts progressively as the particle number increases.

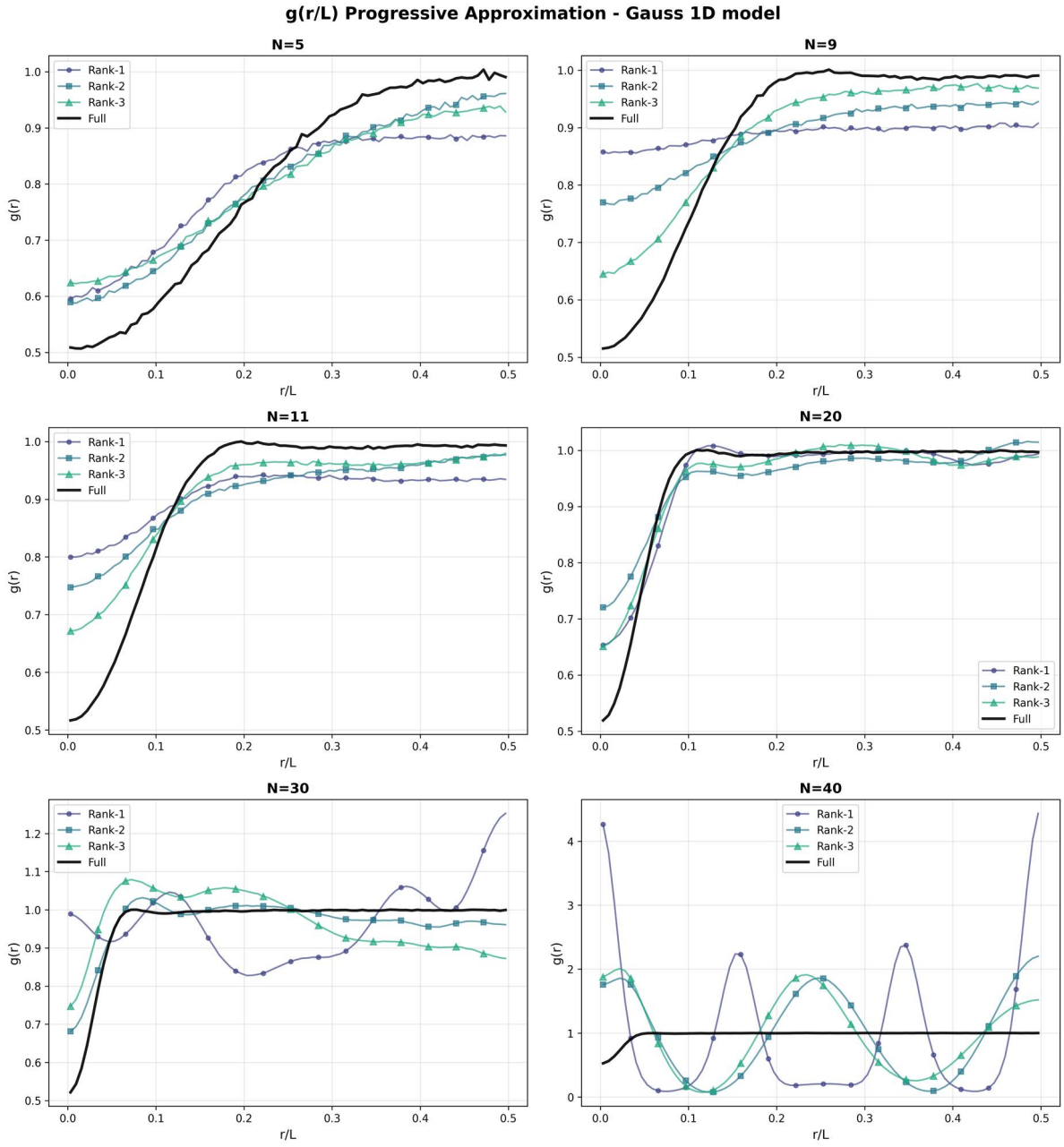


Figure 4.8: Correlation functions obtained after post-training truncation of the learned KAN representation for several particle numbers. Small systems remain close to the full model under moderate truncation, whereas larger systems are more sensitive to the removal of subleading modes.

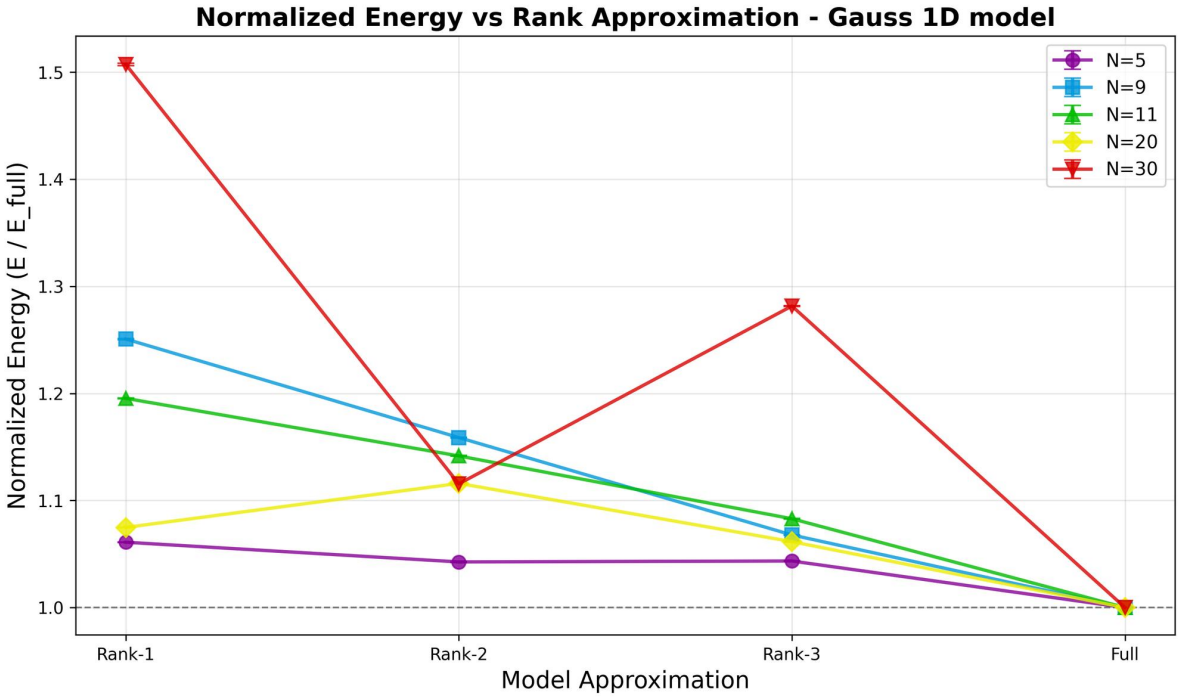


Figure 4.9: Normalized energy as a function of the retained effective rank after truncation. The energy stays close to the full-model value for smaller systems over a wider range of ranks, while larger systems require more retained modes to maintain the same accuracy.

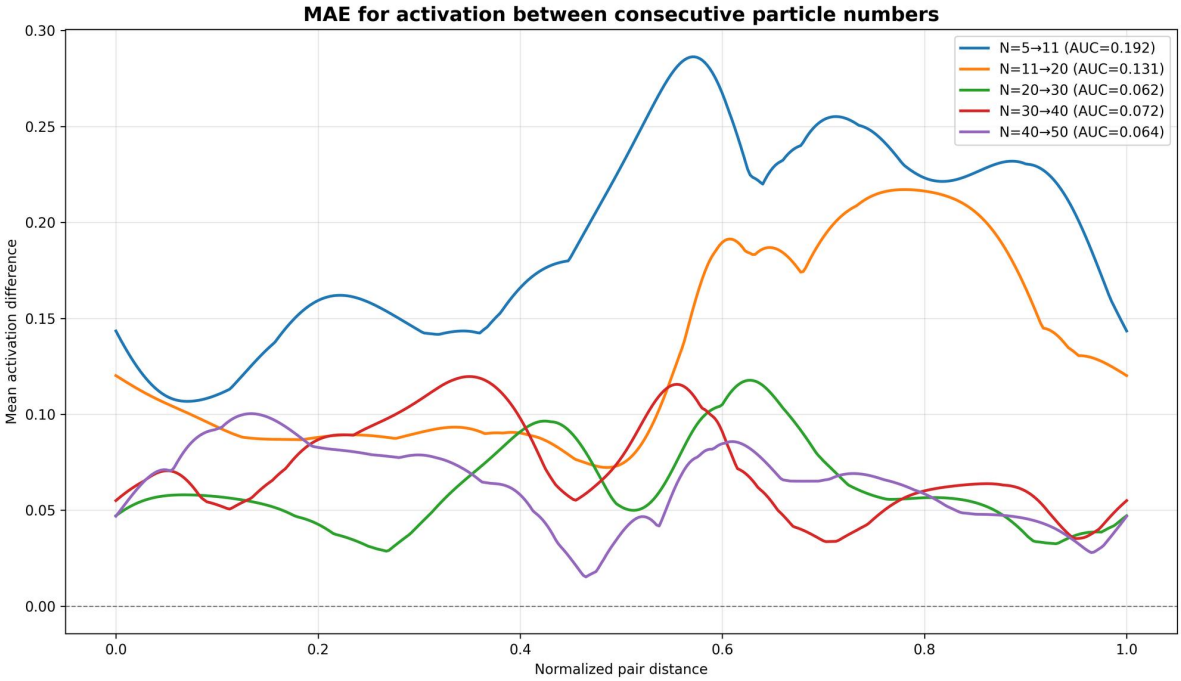


Figure 4.10: Differences between corresponding learned activation functions for the one-dimensional Gaussian benchmark. This diagnostic highlights which channels remain stable across system sizes and which continue to adapt as the many-body correlations become more complex.

5. Results

The results below combine three complementary types of evidence: small-system validation against finite-basis exact diagonalization whenever ED is tractable, variational calculations for larger particle numbers beyond the ED regime, and comparisons with Jastrow baselines in regimes where more expressive correlations become important. To reach the larger systems efficiently, we use an iterative particle-growth strategy: the model is first optimized for smaller N, and the learned parameters are then transferred to the next particle number. This substantially accelerates convergence and, in difficult regimes, can be essential for stable training.

5.1. Particle-Addition Pretraining

We use particle-addition pretraining to improve convergence as the system size grows. The idea is to first optimize the ansatz for a smaller number of particles and then use the resulting parameters to initialize the calculation for the next particle number. Because the density is kept fixed throughout the study, the optimal wave function for $N + M$ where M<N particles is typically close to that for N particles, so the previously learned representation provides a physically meaningful starting point.

In practice, this strategy reduces the risk of poor local minima and makes the optimization substantially more robust. For the repulsive systems considered here, the smaller-N problems are easier to optimize and already encode the main short-range avoidance and collective structure. Transferring these parameters to the next particle number therefore allows the model to refine an already reasonable state rather than rediscover the full many-body structure from a random initialization.

5.2. Gaussian Interaction

The Gaussian interaction in the form (3.5) serves as our primary baseline benchmark because it is smooth, purely pairwise, and free of cusp singularities. This makes it the cleanest setting in which to assess the variational quality of the ansatz without the additional complication of singular short-distance physics. For small systems, the Gaussian benchmark is anchored to the finite-basis ED comparison reported in Table 1 and Fig. 5.1, which provide a representative validation example in the regime where ED is still feasible. Table 1 compares finite-basis ED energies with the corresponding KAN–NQS results for several system sizes, while the accompanying

Table 1: Finite-basis exact-diagonalization benchmark for the Gaussian model compared with the corresponding KAN–NQS results for small system sizes.

N	E_{ED}	E_{NQS}	Variance
2	0.73876371	0.73884842	1.02×10^{-4}
3	1.07404973	1.07412500	4.83×10^{-4}
5	1.19917370	1.19917960	3.25×10^{-4}

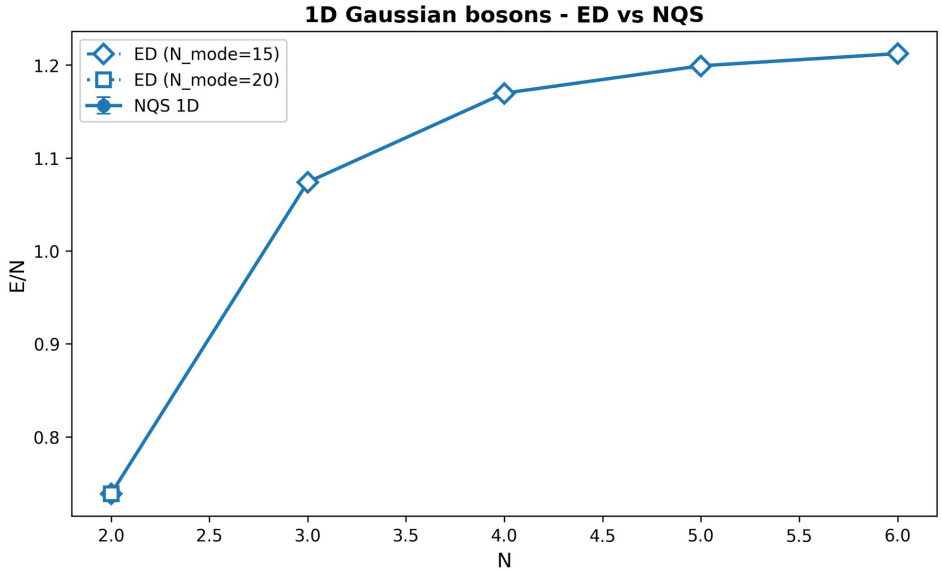


Figure 5.1: Finite-basis ED and KAN–NQS energies for the small-system one-dimensional Gaussian benchmark. The plot visualizes the close agreement summarized in Table 1.

NQS variances provide an additional measure of optimization stability. The discrepancy between ED and NQS decreases rapidly with particle number, reaching near agreement already by $N = 5$, while the variances remain small throughout. Presenting the table together with the Gaussian benchmark results keeps this small-system validation next to the setting where it is used most directly. For larger systems, where ED rapidly becomes impractical, we report the NQS and Jastrow baseline energies. In this way, the Gaussian results play a dual role: they first validate the ansatz against a quasi-exact benchmark and then demonstrate its scalability to larger particle numbers in one, two, and three spatial dimensions.

The correlation function serves as a fundamental diagnostic of how particles or degrees of freedom influence each other beyond mean-field descriptions. Physically, it measures the statistical dependence between observables at different points in space or time, thereby capturing the extent to which the presence or motion of one particle constrains another. Incorporating a correlation function into a variational

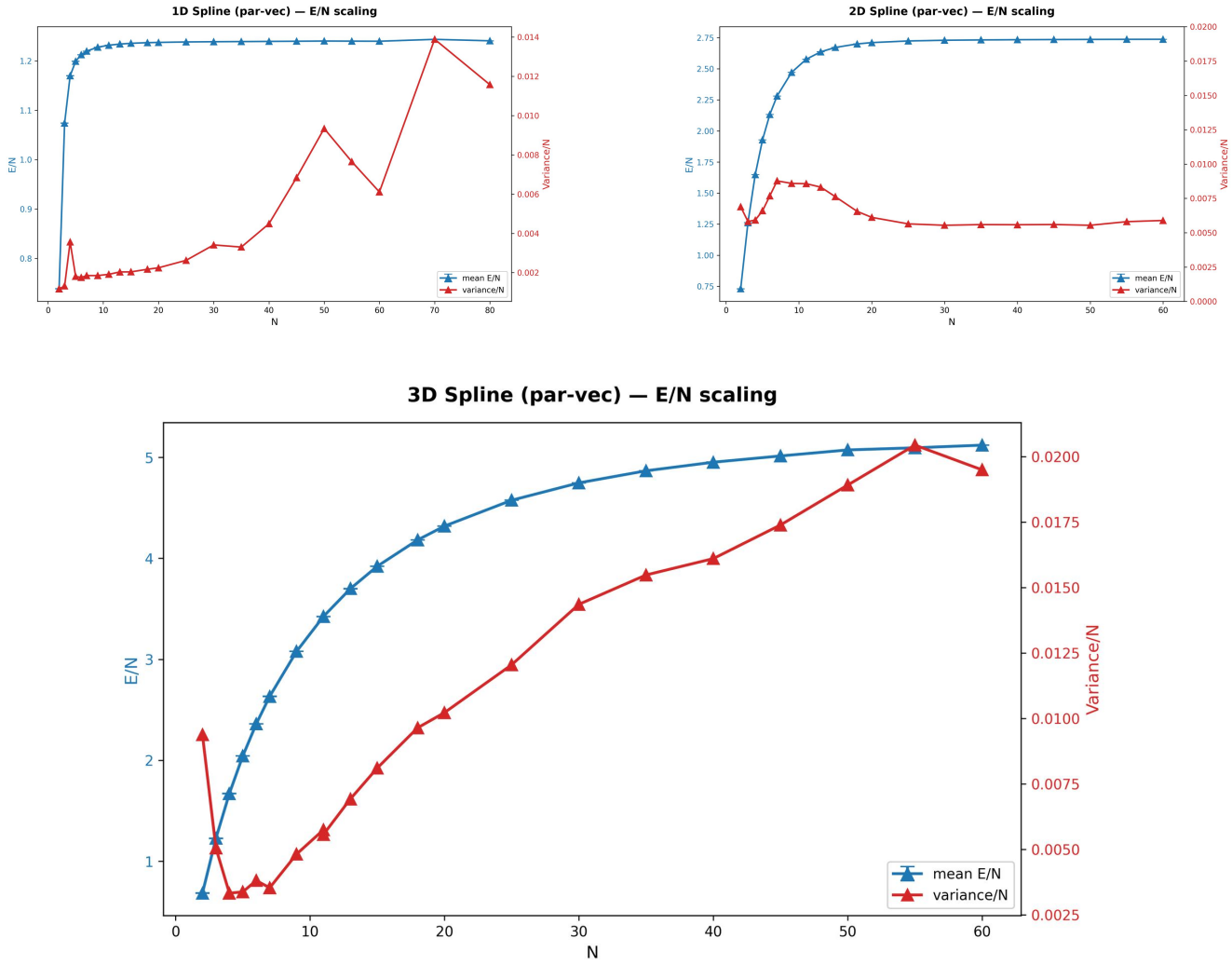


Figure 5.2: Gaussian-interaction energies per particle in one, two, and three spatial dimensions as functions of the particle number.

ansatz is essential because it builds in the correct physical structure of interactions from the outset, ensuring that the wave function respects cusp conditions and faithfully represents emergent collective behavior. In this sense, the correlation function bridges microscopic interactions with macroscopic observables, making it a central object for both analytical theory and numerical modeling. In the present work, the reported pair-correlation functions are evaluated as the radial distribution function $g(r)$ from histograms of pair distances,

$$g(r_k) = \frac{L^d \,\mathrm{hist}(r_k)}{M\, N^2\, J(r_k)\, \delta r}, \tag{5.1}$$

where $r_k \in [0, L/2]$ is the center of the k-th histogram bin at which $g(r)$ is evaluated. In the plots, the quantity on the horizontal axis is written as c/L, where c denotes the bin center ($c = r_k$), so that $c/L \in [0, 1/2]$ is simply the bin-center position

normalized by the box length. Furthermore, $\mathrm{hist}(r_k)$ is the total number of pair distances accumulated over all Monte Carlo samples that fall in the interval $[r_k - \delta r/2,\, r_k + \delta r/2]$, M is the number of Monte Carlo samples, N is the number of particles, L is the box side length, d is the spatial dimension, and the bin width is chosen as $\delta r = L/(2\,\mathrm{bins})$. The surface-element (Jacobian) factor is

$$J(r) = \begin{cases} 1, & d = 1, \\ 2\pi r, & d = 2, \\ 4\pi r^2, & d = 3. \end{cases} \tag{5.2}$$

This normalization ensures that $g(r)$ measures the pair density relative to the uniform-density reference appropriate to the simulation cell.

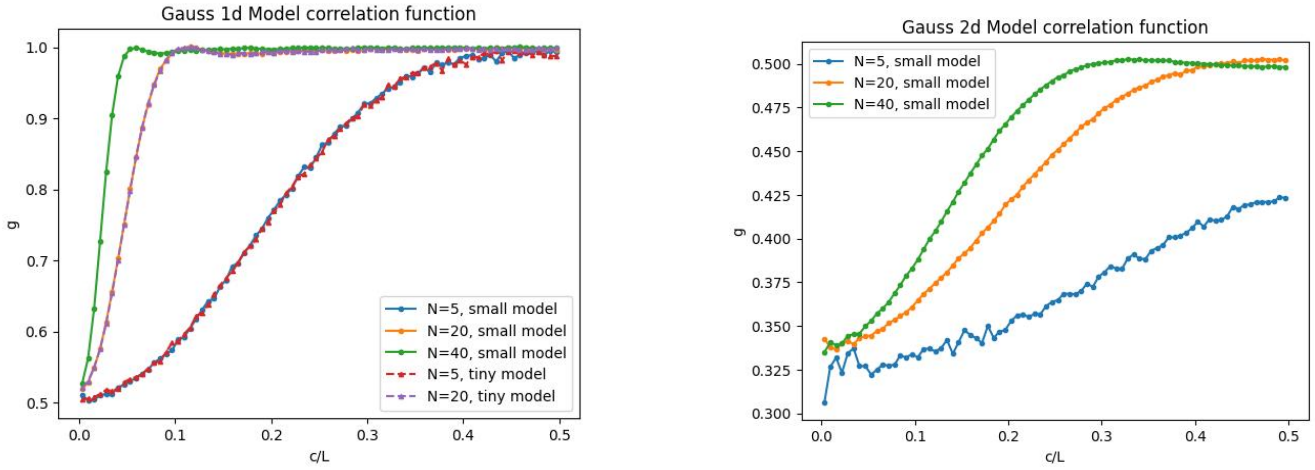


Figure 5.3: Pair-correlation functions for the Gaussian benchmark in one and two spatial dimensions.

5.3. Calogero–Sutherland Model

The Calogero–Sutherland model is a substantially harder benchmark than the Gaussian case because the inverse-square interaction is singular at short distance. As two particles approach each other, the local energy becomes extremely sensitive to the detailed short-range behavior of the wave function. A smooth generic neural ansatz therefore has difficulty representing the correct near-coincidence structure directly, which leads to large optimization noise and unstable gradients.

To control this problem, we incorporate a Kato-type cusp factor into the variational ansatz. In practice, this means that the singular short-distance part of the

wave function is treated explicitly, while the neural network is responsible only for learning the remaining regular many-body contribution. This separation is important both physically and numerically: the cusp encodes the correct local behavior imposed by the interaction, and at the same time it suppresses the pathological variance that would otherwise appear in the Monte Carlo estimate of the local energy.

The resulting optimization dynamics are qualitatively different from the smooth-potential case. Once the cusp is included, the energy reaches the correct scale very early in training and then changes only modestly, while the variance keeps decreasing over a longer time window. This suggests that the cusp captures the main short-distance physics from the start, while the network later improves the longer-range structure of the state. The fact that the energy settles before the variance becomes small also fits a non-convex loss landscape: several parameter choices already give the right singular behavior, but more optimization is needed to obtain a stable and accurate wave function.

This behavior is visible in Fig. 5.4, where the learning curves remain stable across the system sizes shown. The main effect of the neural optimization is therefore not only to recover the singular physics already encoded by the cusp, but also to improve upon that initial physically motivated baseline. Because the Calogero–Sutherland model is integrable, an exact benchmark is available. For the representative case shown here, the optimized variational ansatz yields an energy of -156.3160 with variance 0.08215, in excellent agreement with the exact value -156.317.

In addition to the energy, we computed the pair-correlation function for the Calogero–Sutherland model. This observable is particularly informative here because it directly reflects the repulsive correlation hole generated by the inverse-square interaction and therefore probes whether the learned wave function captures the correct spatial organization of particles beyond the energy alone.

Figure 5.5 shows that the ansatz reproduces the strong suppression of close particle encounters expected for this model. At small separations, the correlation function is reduced because configurations with nearby particles are energetically penalized by the singular interaction. At larger separations, the structure of the curve reflects the nontrivial ordering induced by the long-range repulsion on the ring. The systematic evolution with particle number further indicates that the model is not merely fitting a few-body artifact, but is learning a many-body state whose correlation pattern remains consistent as the system size increases.

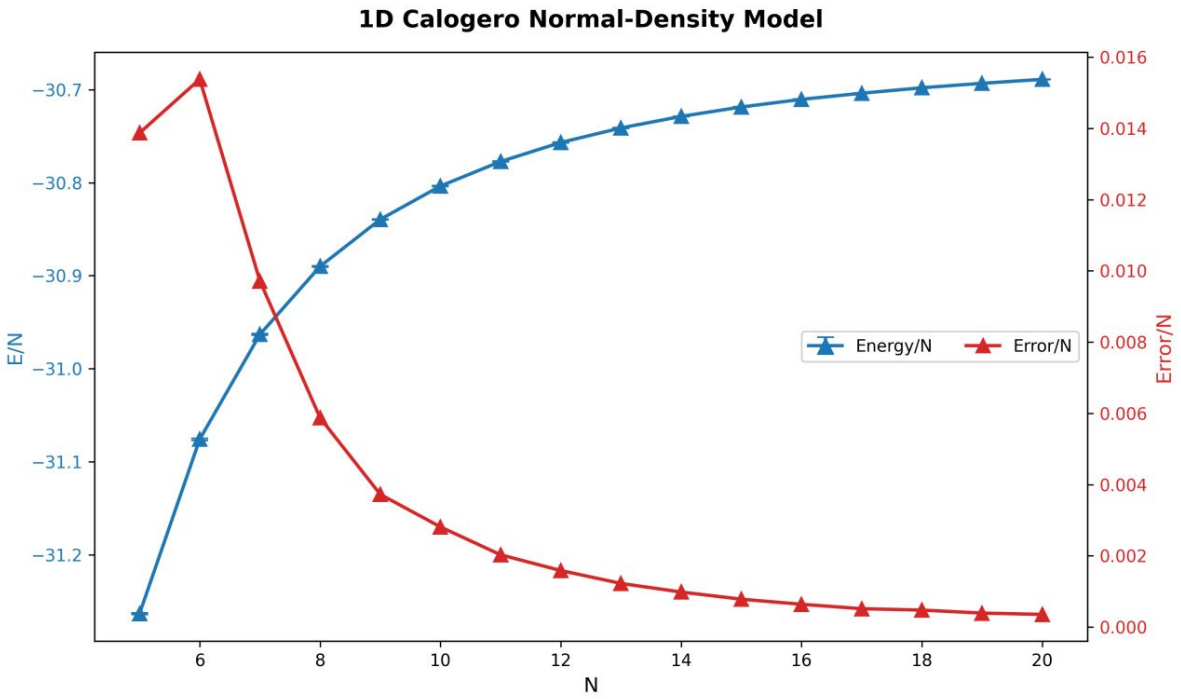


Figure 5.4: Energy during optimization for the one-dimensional Calogero–Sutherland model at coupling constant $g = 5$, shown for systems of up to 20 particles.

5.4. Weight analysis

Beyond the variational benchmarks, it is useful to ask how much of the nominal capacity of the spline-based KAN layers is actually used after training. To address this question, we analyze the learned weight matrices through their singular values. The singular-value spectrum provides a compact measure of effective dimensionality: if only a few modes dominate, then the trained representation behaves as low rank even when the raw parameter count is large.

Figure 5.6 compares the singular-value spectra for several model sizes. In all cases, the spectra decay markedly rather than remaining flat, which indicates that only a subset of directions contributes strongly to the final representation. The effective rank of the learned layers can therefore be substantially smaller than the nominal parameter count. This, in turn, supports the idea that spline-based KAN layers are often overparameterized relative to the complexity actually needed for the target wave function and motivates pruning or low-rank compression as natural follow-up strategies.

5.5. Comparison with the Jastrow Baseline

In the previous paragraphs we showed that the DeepSet–KAN ansatz is a flexible variational representation for bosonic wave functions. In this section, we highlight

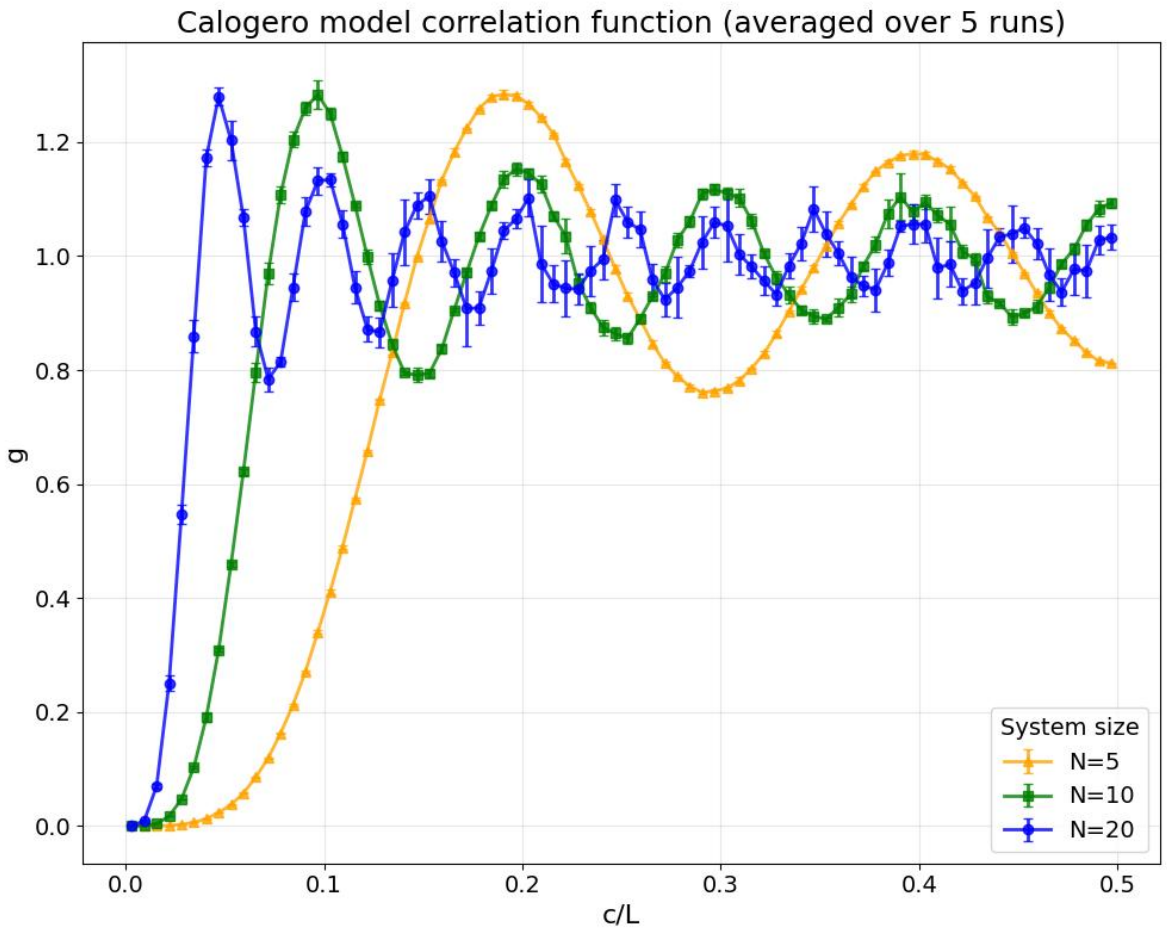


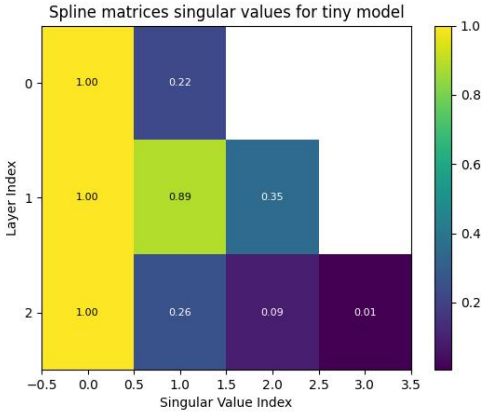
Figure 5.5: Correlation function for 1d Calogero-Sutherland for 5,10,20 particles.

regimes in which this additional flexibility is not only convenient but also physically relevant. As a reference baseline, we compare against Jastrow-type ansatzes, which are very effective when the ground state is dominated by pair correlations but become restrictive once higher-order or more collective structures become important.

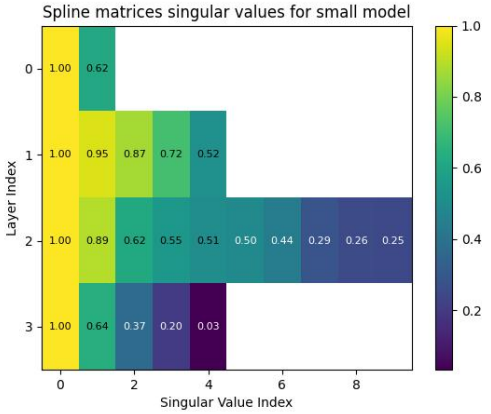
The three examples below were chosen to probe different mechanisms by which the Jastrow form can become insufficient: a Gaussian interaction with large interaction scale in two dimensions, an explicitly three-body interaction, and a soft-core Lennard-Jones-type potential. In each case, the central question is whether the extra expressive power of the KAN-enhanced architecture produces a measurable variational improvement over the pair-product baseline.

5.5.1 Gaussian Interaction

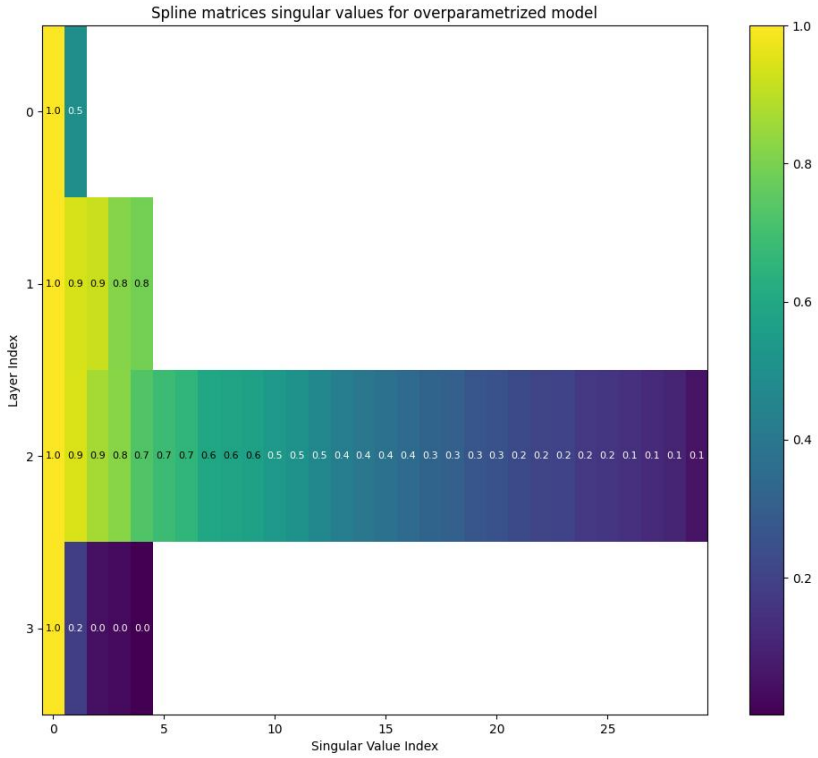
For the Gaussian model on a two-dimensional plane at large $\varepsilon = 30$, the Jastrow ansatz remains a strong baseline because the interaction is smooth and purely pairwise. Nevertheless, the comparison plot shows that the KAN-enhanced wave func-



(a) Singular-value spectra of the three spline layers in the tiny model.



(b) Singular-value spectra of the three spline layers in the small model.



(c) Singular-value spectra of the four spline layers in the overparameterized model.

Figure 5.6: Singular-value analysis of learned spline-layer weight matrices for tiny, small, and overparameterized models for Gauss 1D model. The decay of the spectra indicates that the effective rank can be lower than the nominal parameter count, which motivates pruning and low-rank compression.

tion can further reduce the variational energy beyond the level achieved by the Jastrow reference. This indicates that even for a nominally two-body potential, the effective structure of the optimized many-body state is not exhausted by a simple

Table 2: Quantitative comparison for the two-dimensional Gaussian benchmark at large ε, averaged over the last 50 optimization iterations. All energies, variances, and statistical uncertainties are reported per particle.

N	E_{J}/N	E_{NQS}/N	$\mathrm{Var}_{\mathrm{J}}/N$	$\mathrm{Var}_{\mathrm{NQS}}/N$	σ_{J}/N	σ_{NQS}/N
4	25.239830	25.230297	1.084405	1.256764	0.002843	0.003047
8	37.690089	37.686412	2.084764	2.137031	0.002778	0.002854
12	42.437423	42.436876	0.958809	0.941979	0.001552	0.001534

product of pair factors.

An important feature of the plot is that the difference is not only visible in the final energy, but also in the optimization trajectory and the accompanying uncertainty band. This makes the figure useful not merely as a final benchmark, but as evidence that the richer ansatz explores a lower-energy part of parameter space more systematically. In other words, the advantage here should be interpreted as a genuine improvement of the variational family rather than an isolated optimization artifact.

The averaged results from the last 50 optimization iterations are summarized in Table 2. For all three system sizes, the NQS ansatz reaches a slightly lower energy per particle than the Jastrow baseline, with $\Delta E/N = -9.533 \times 10^{-3}$ at $N = 4$, -3.677×10^{-3} at $N = 8$, and -5.47×10^{-4} at $N = 12$. The quantitative advantage is therefore modest and decreases with system size. The variance does not show the same systematic trend: it is slightly higher for the NQS ansatz at $N = 4$ and $N = 8$, and slightly lower at $N = 12$. This makes the Gaussian benchmark qualitatively different from the three-body case: here the main signal is a small but consistent energy improvement rather than a dramatic reduction in variance.

5.5.2 Three-Body Interaction

The three-body benchmark provides the clearest conceptual separation between the two ansatzes. A standard Jastrow construction is built from one-body and two-body factors and therefore has no natural mechanism for representing irreducible three-particle correlations. By contrast, the Deep Sets–KAN ansatz aggregates learned features over the full configuration and can therefore encode correlations that are not reducible to independent pair contributions.

The corresponding plot should therefore be interpreted as a structural test of the ansatz rather than merely a numerical comparison. If the KAN-based model

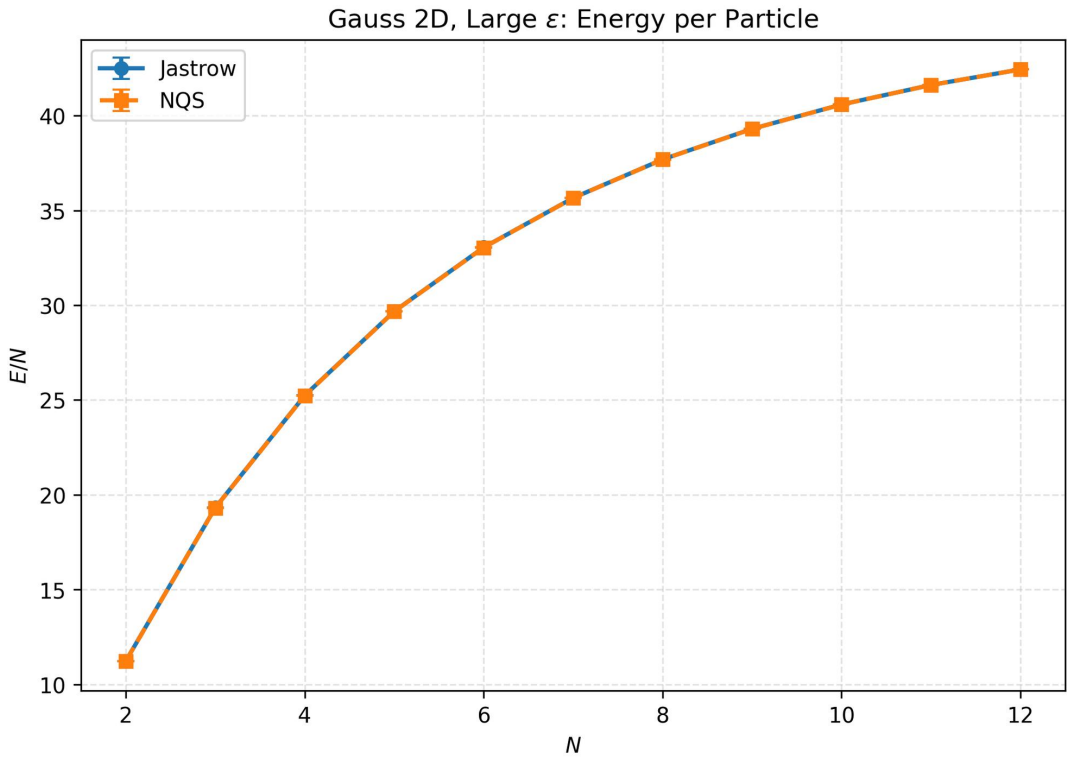


Figure 5.7: Energy per particle for the Jastrow baseline and the KAN-enhanced NQS ansatz for the two-dimensional Gaussian model at large $\varepsilon = 30$ and density $n = 1.2$. The plot illustrates both the variational energies and their difference, highlighting the regime in which the more expressive ansatz improves upon the pair-product description.

reaches lower variational energies, this supports the conclusion that the network is capturing genuine many-body information that the Jastrow baseline cannot represent explicitly. This example is particularly important because it shows that the advantage of the architecture is tied to the structure of the interaction itself, not simply to the presence of more trainable parameters.

This interpretation is reinforced by the quantitative comparison in Table 3, where all values are averaged over the last 50 optimization iterations and reported per particle. The NQS ansatz yields a lower energy than the Jastrow reference for all three system sizes: the energy difference decreases from 1.6363×10^{-2} at $N = 4$ to 6.698×10^{-3} at $N = 12$, but it remains systematic throughout. An equally important signal is the variance reduction. At $N = 4$, the NQS variance is lower by more than a factor of three, and even at $N = 12$ it remains lower by about 30%. This persistent suppression of the variance indicates that the KAN-based wave function is not only energetically favorable, but also provides a more stable and physically faithful representation of the three-body correlations.

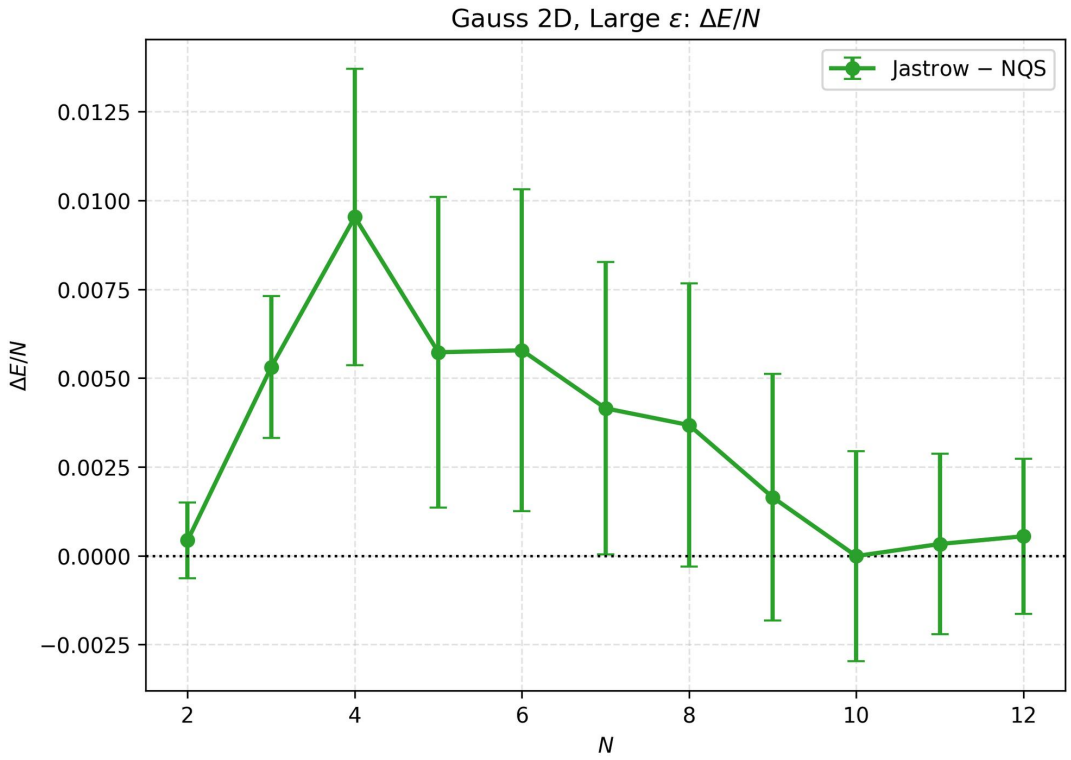


Figure 5.8: Difference in energy per particle between the Jastrow baseline and the KAN-enhanced NQS ansatz for the two-dimensional Gaussian model at large $\varepsilon = 30$ and density $n = 1.2$. The plot illustrates both the variational energies and their difference, highlighting the regime in which the more expressive ansatz improves upon the pair-product description.

5.5.3 Lennard-Jones-like Soft-Core Interaction

The Lennard-Jones-like soft-core problem is intermediate between the previous two cases. Unlike the three-body benchmark, the Hamiltonian is still purely pairwise, but the potential now combines a short-range repulsive core with an attractive tail. This is the characteristic structure of Lennard-Jones interactions, which are commonly used as an effective model for neutral atoms and molecules. Even though the microscopic interaction remains two-body, the competition between repulsion at short distances and attraction at intermediate ranges can produce nontrivial collective organization as the particle number grows.

The comparison plot is therefore useful for showing how a nominally pairwise Lennard-Jones-type model can still require a more expressive variational description. A Jastrow ansatz is naturally well suited to the dominant pair structure of the problem and can capture the basic repulsive core together with part of the attractive shell structure. The KAN-enhanced model, however, has additional freedom to reorganize the wave function globally and to encode more complicated collective

Table 3: Quantitative comparison for the three-body benchmark, averaged over the last 50 optimization iterations. All energies, variances, and statistical uncertainties are reported per particle.

N	E_{J}/N	E_{NQS}/N	$\mathrm{Var}_{\mathrm{J}}/N$	$\mathrm{Var}_{\mathrm{NQS}}/N$	σ_{J}/N	σ_{NQS}/N	$\Delta E/N$
4	0.187755	0.171392	0.905324	0.256112	0.002637	0.001384	−0.016363
8	0.262859	0.251582	1.153117	0.675665	0.002080	0.001584	−0.011277
12	0.281198	0.274500	1.198279	0.915352	0.001728	0.001531	−0.006698

Table 4: Quantitative comparison for the Lennard–Jones-type soft-core benchmark, averaged over the last 50 optimization iterations.

N	E_{J}/N	E_{Ψ}/N	$\mathrm{Var}_{\mathrm{J}}/N$	Var_{Ψ}/N	$\Delta E/N$	Significance
11	−0.375436	−0.378041	0.01371	0.01178	−0.002605	-4.9σ
20	−0.718835	−0.720362	0.01402	0.01719	−0.001528	-3.5σ
30	−1.089383	−1.093170	0.02205	0.01992	−0.003787	-9.3σ

responses to this competing attraction–repulsion balance. When the learned ansatz attains a lower variational energy, the natural interpretation is that these additional collective adjustments matter quantitatively even for this pair-potential benchmark.

This interpretation is supported by the averaged results from the last 50 optimization iterations, summarized in Table 4. For all three system sizes, the Ψ ansatz yields a lower energy per particle than the Jastrow baseline. The energy gain is modest in absolute magnitude, but statistically significant in every case: $\Delta E/N = -2.605 \times 10^{-3}$ for $N = 11$, -1.528×10^{-3} for $N = 20$, and -3.787×10^{-3} for $N = 30$, corresponding to improvements of 4.9σ, 3.5σ, and 9.3σ, respectively. The variance pattern is more mixed: the Ψ ansatz has lower variance at $N = 11$ and $N = 30$, while at $N = 20$ its variance is slightly higher. Since the variational energy is the primary measure of wave-function quality, the consistent energy improvement across all sizes is the main conclusion.

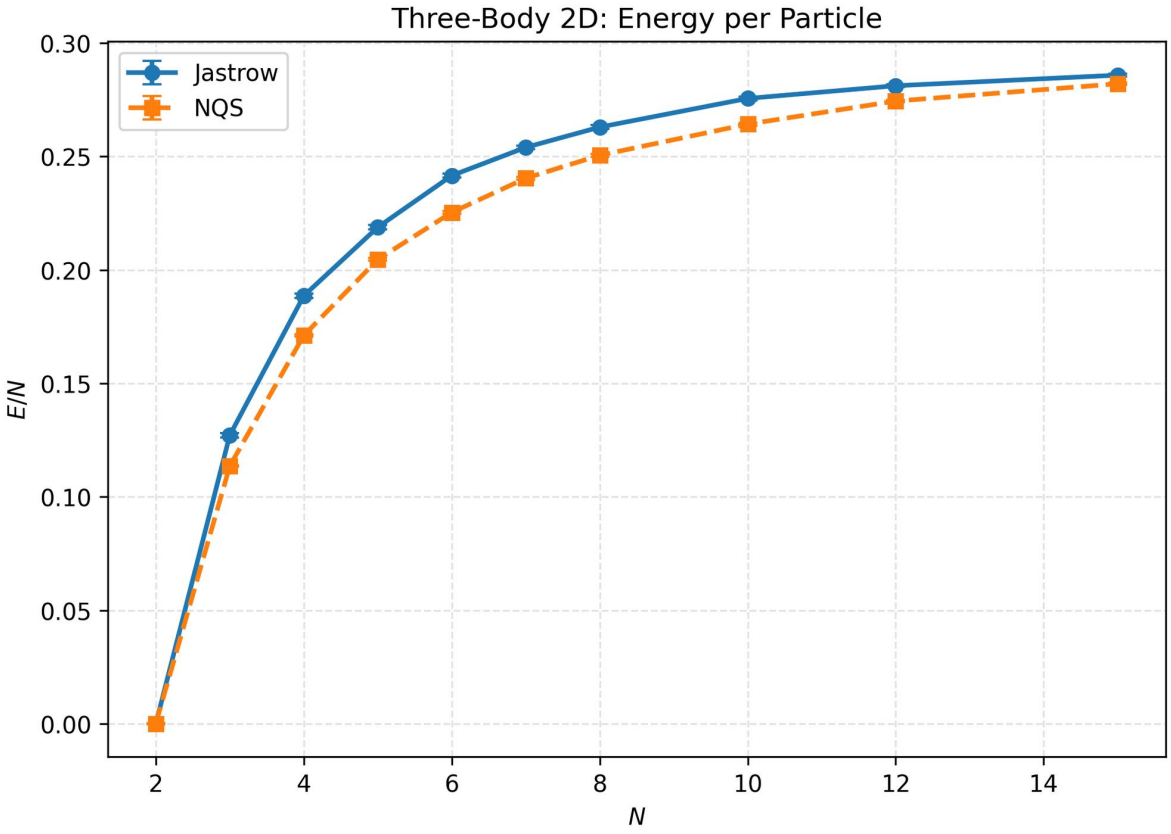


Figure 5.9: Energy per particle for the three-body interaction. Since the Jastrow ansatz is intrinsically pair-based, the comparison emphasizes the ability of the KAN-enhanced wave function to capture irreducible many-body correlations.

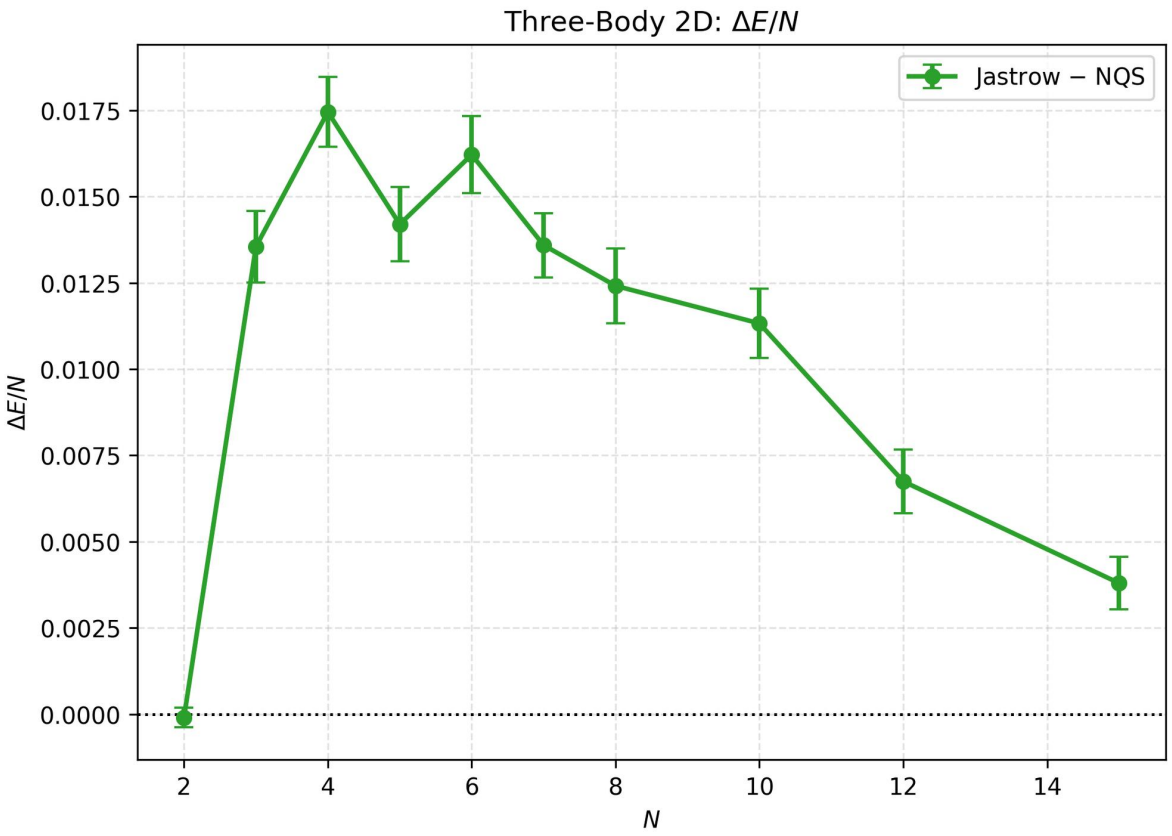


Figure 5.10: Difference in energy per particle between the Jastrow ansatz and the KAN-enhanced NQS for the three-body interaction potential.

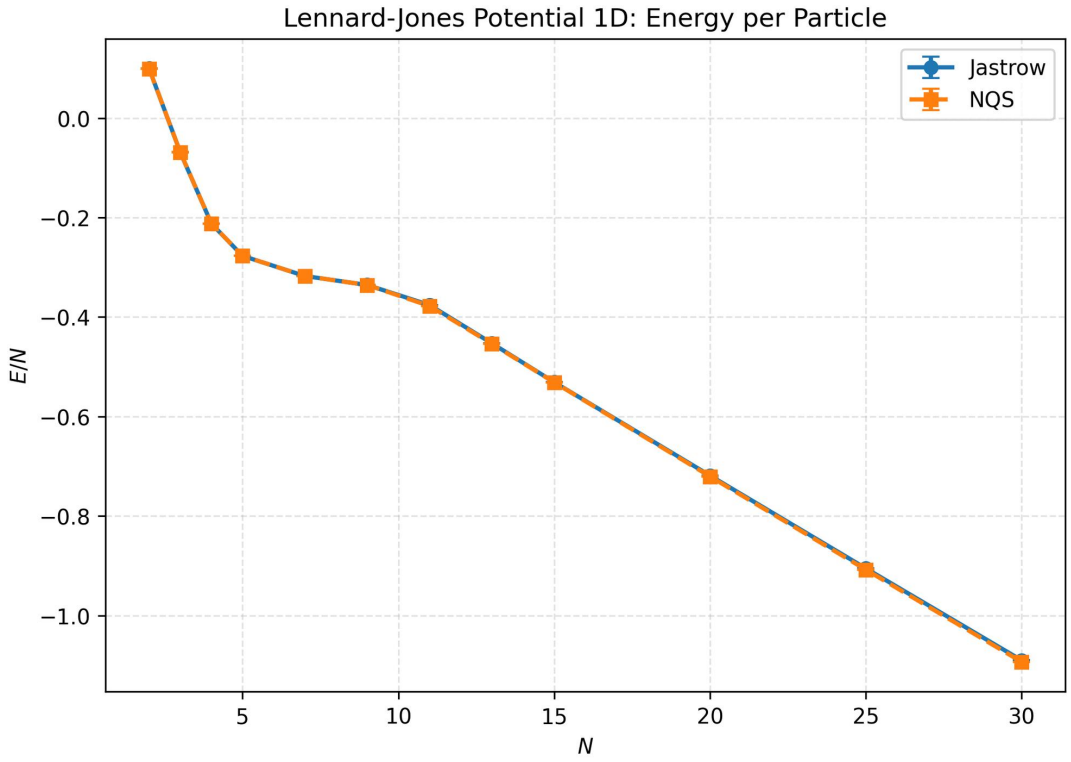


Figure 5.11: Energy per particle for the Jastrow baseline and the KAN-enhanced ansatz for a Lennard–Jones-type soft-core interaction. The figure illustrates how long-range pair forces can generate effective collective structure that benefits from a more expressive wave-function representation.

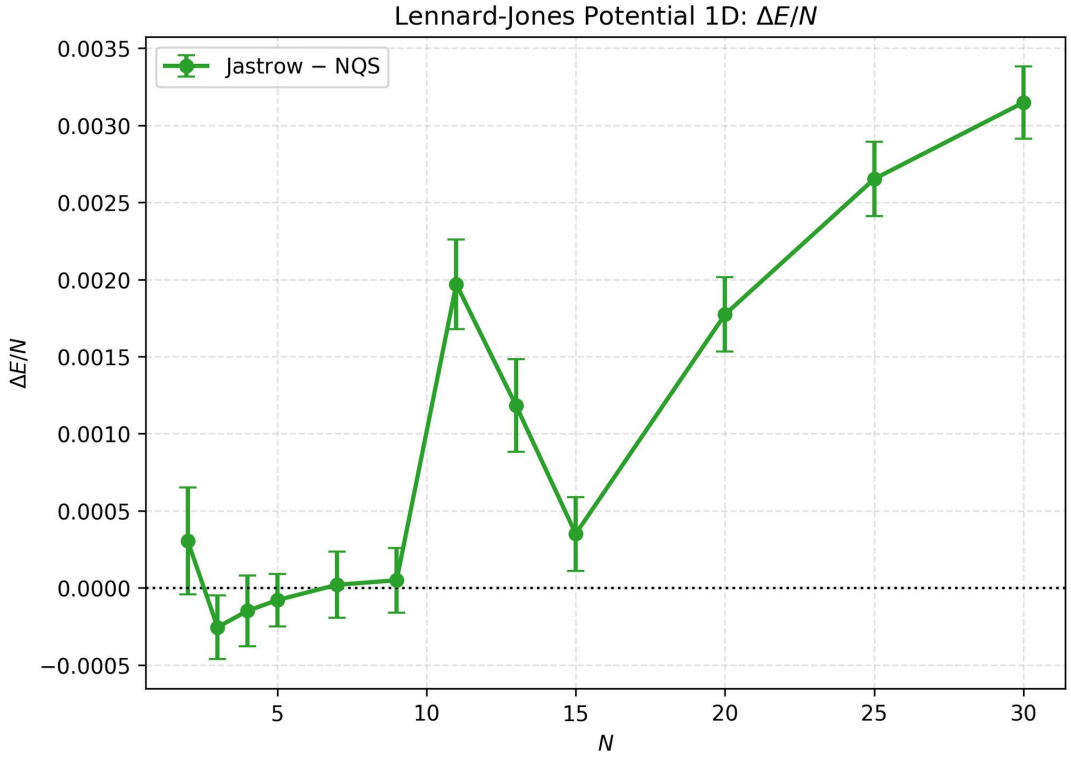


Figure 5.12: Difference in energy per particle between the Jastrow baseline and the KAN-enhanced ansatz for a Lennard–Jones-type soft-core interaction. The figure illustrates how long-range pair forces can generate effective collective structure that benefits from a more expressive wave-function representation.

6. Conclusions

In this work, we introduced a KAN-enhanced neural quantum state for continuous bosonic systems with periodic geometry. The ansatz combines the permutation-invariant Deep Sets construction with KAN layers and explicit correlation-aware inputs, so that the main physical symmetries are built into the architecture from the outset. The central result of the manuscript is that this framework is not only viable in practice but also already competitive across a broad set of benchmark problems relevant to continuous-space many-body physics.

On the architectural side, we benchmarked several of the most commonly used KAN layer types within the same variational pipeline. Among the tested implementations, our PRKAN variant gives the best overall trade-off between expressivity, optimization stability, and parameter efficiency. This is an important practical outcome of the study: the work does not only propose a KAN-based ansatz in abstract form, but also identifies a concrete layer choice that appears to be the most effective for the class of bosonic problems considered here.

On the physical side, we tested the approach on several models with clear physical relevance, including Gaussian interactions in one, two, and three dimensions, the Calogero–Sutherland model, and additional many-body benchmarks used to probe the limitations of simpler variational forms. Across these examples, the Deep Sets–KAN construction consistently learns accurate ground-state properties and, in direct comparisons, improves on the widely used Jastrow baseline. This is particularly significant in regimes where pair-product structures become restrictive: by maintaining a macroscopically homogeneous unit density across our benchmarks, the systematic reduction in ground-state energy confirms that the KAN-enhanced Deep Sets ansatz successfully captures non-trivial, higher-order collective correlations. Furthermore, our analysis of the pair-correlation functions $g(r)$ demonstrates that the network faithfully reconstructs the emergent spatial structure of the bosonic fluid—such as the characteristic short-range repulsion hole—and scales stably toward the thermodynamic limit. For the singular Calogero–Sutherland problem, the explicit cusp term remains essential in practice, as it handles the non-analytic short-distance structure and suppresses the local-energy variance during optimization.

From an experimental perspective, the ability of our framework to accurately reproduce these pair-correlation profiles carries immediate relevance for the physics of ultracold atomic gases in optical lattices and laser fields. In modern labora-

tory setups, these exact spatial microstructures and density-density fluctuations are primary observables accessible via time-of-flight imaging techniques following free expansion, as well as noise-correlation spectroscopy. Consequently, the KAN–NQS framework establishes itself as a predictive tool capable of benchmarking analog quantum simulators of continuum quantum matter.

Overall, the results support the conclusion that KAN-based Deep Sets are a promising replacement for more rigid handcrafted baselines in continuous bosonic systems. At the same time, the present work opens several clear directions for continuation. The most immediate next steps are to extend the framework to a broader range of physically relevant Hamiltonians, to study larger and more challenging systems, and to develop a more efficient parameterization strategy that preserves the expressive advantages of the current model while reducing computational overhead. In that sense, the present manuscript establishes both a working benchmark and a strong foundation for a more general parameter-efficient framework for quantum many-body problems in continuous space.

References

1. John M. Martyn, Khadijeh Najafi, and Di Luo. Variational neural-network ansatz for continuum quantum field theory. *Physical Review Letters*, 131(8), August 2023. ISSN 1079-7114. doi: 10.1103/physrevlett.131.081601. URL http://dx.doi.org/10.1103/PhysRevLett.131.081601.
2. Tosio Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Communications on Pure and Applied Mathematics*, 10(2):151–177, 1957. doi: https://doi.org/10.1002/cpa.3160100201. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.3160100201.
3. Gabriel Pescia, Jiequn Han, Alessandro Lovato, Jianfeng Lu, and Giuseppe Carleo. Neural-network quantum states for periodic systems in continuous space. *Physical Review Research*, 4(2):023138, 2022. doi: 10.1103/PhysRevResearch.4.023138. URL http://dx.doi.org/10.1103/PhysRevResearch.4.023138.
4. Hannah Lange, Anka Van de Walle, Atiye Abedinnia, and Annabelle Bohrdt. From architectures to applications: A review of neural quantum states, 2024. URL https://arxiv.org/abs/2402.09402.
5. 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. URL http://dx.doi.org/10.1126/science.aag2302.
6. Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Quantum entanglement in neural network states. *Phys. Rev. X*, 7:021021, May 2017. doi: 10.1103/PhysRevX.7.021021. URL https://link.aps.org/doi/10.1103/PhysRevX.7.021021.
7. Xun Gao and Lu-Ming Duan. Efficient representation of quantum many-body states with deep neural networks. *Nature Communications*, 8(1), 2017. doi: 10.1038/s41467-017-00705-2. URL http://dx.doi.org/10.1038/s41467-017-00705-2.
8. Or Sharir, Yoav Levine, Noam Wies, Giuseppe Carleo, and Amnon Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Physical Review Letters*, 124(2), January 2020. doi: 10.1103/physrevlett.124.020503. URL http://dx.doi.org/10.1103/PhysRevLett.124.020503.
9. Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla. Recurrent neural network wave functions. *Physical Review*

Research, 2(2), 2020. doi: 10.1103/physrevresearch.2.023358. URL http://dx.doi.org/10.1103/PhysRevResearch.2.023358.

10. Yuan-Hang Zhang and Massimiliano Di Ventra. Transformer quantum state: A multipurpose model for quantum many-body problems. *Physical Review B*, 107 (7), February 2023. doi: 10.1103/physrevb.107.075147. URL http://dx.doi.org/10.1103/PhysRevB.107.075147.
11. Kenny Choo, Giuseppe Carleo, Nicolas Regnault, and Titus Neupert. Symmetries and many-body excitations with neural-network quantum states. *Physical Review Letters*, 121(16), October 2018. doi: 10.1103/physrevlett.121.167204. URL http://dx.doi.org/10.1103/PhysRevLett.121.167204.
12. Di Luo, Zhuo Chen, Kaiwen Hu, Zhizhen Zhao, Vera Mikyoung Hur, and Bryan K. Clark. Gauge-invariant and anyonic-symmetric autoregressive neural network for quantum lattice models. *Physical Review Research*, 5(1), March 2023. doi: 10.1103/physrevresearch.5.013216. URL http://dx.doi.org/10.1103/PhysRevResearch.5.013216.
13. James Stokes, Saibal De, Shravan Veerapaneni, and Giuseppe Carleo. Continuous-variable neural-network quantum states and the quantum rotor model, 2021. URL https://arxiv.org/abs/2107.07105.
14. Gabriel Pescia, Jiequn Han, Alessandro Lovato, Jianfeng Lu, and Giuseppe Carleo. Neural-network quantum states for periodic systems in continuous space. *Physical Review Research*, 4(2), May 2022. doi: 10.1103/physrevresearch.4.023138. URL http://dx.doi.org/10.1103/PhysRevResearch.4.023138.
15. Jane Kim, Gabriel Pescia, Bryce Fore, Jannes Nys, Giuseppe Carleo, Stefano Gandolfi, Morten Hjorth-Jensen, and Alessandro Lovato. Neural-network quantum states for ultra-cold Fermi gases, 2023. URL https://arxiv.org/abs/2305.08831.
16. Kenny Choo, Titus Neupert, and Giuseppe Carleo. Two-dimensional frustrated J_1–J_2 model studied with neural network quantum states. *Physical Review B*, 100(12):125124, 2019. doi: 10.1103/PhysRevB.100.125124. URL http://dx.doi.org/10.1103/PhysRevB.100.125124.
17. Michele Ruggeri, Saverio Moroni, and Markus Holzmann. Nonlinear network description for many-body quantum systems in continuous space. *Physical Review Letters*, 120(20):205302, 2018. doi: 10.1103/PhysRevLett.120.205302. URL http://dx.doi.org/10.1103/PhysRevLett.120.205302.

18. Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic Schrödinger equation. *Nature Chemistry*, 12:891–897, 2020. doi: 10.1038/s41557-020-0544-y. URL http://dx.doi.org/10.1038/s41557-020-0544-y.

19. Gleb Arutyunov. *Elements of Classical and Quantum Integrable Systems*. Springer, Cham, Switzerland, 2019. ISBN 978-3-030-24197-1. doi: 10.1007/978-3-030-24198-8. URL http://dx.doi.org/10.1007/978-3-030-24198-8.

20. Phillip Weinberg and Marin Bukov. Quspin: a python package for dynamics and exact diagonalisation of quantum many body systems part i: spin chains. *SciPost Physics*, 2(1), February 2017. ISSN 2542-4653. doi: 10.21468/scipostphys.2.1.003. URL http://dx.doi.org/10.21468/SciPostPhys.2.1.003.

21. Jutho Haegeman, J. Ignacio Cirac, Tobias J. Osborne, and Frank Verstraete. Calculus of continuous matrix product states. *Physical Review B*, 88(8), August 2013. ISSN 1550-235X. doi: 10.1103/physrevb.88.085118. URL http://dx.doi.org/10.1103/PhysRevB.88.085118.

22. D. M. Ceperley. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.*, 67:279–355, Apr 1995. doi: 10.1103/RevModPhys.67.279. URL https://link.aps.org/doi/10.1103/RevModPhys.67.279.

23. Robert Jastrow. Many-body problem with strong forces. *Phys. Rev.*, 98:1479–1484, Jun 1955. doi: 10.1103/PhysRev.98.1479. URL https://link.aps.org/doi/10.1103/PhysRev.98.1479.

24. W. L. McMillan. Ground state of liquid He^4. *Phys. Rev.*, 138:A442–A451, Apr 1965. doi: 10.1103/PhysRev.138.A442. URL https://link.aps.org/doi/10.1103/PhysRev.138.A442.

25. Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov-Arnold Networks, 2025. URL https://arxiv.org/abs/2404.19756.

26. Runpeng Yu, Weihao Yu, and Xinchao Wang. Kan or mlp: A fairer comparison, 2024. URL https://arxiv.org/abs/2407.16674.

27. Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander Smola. Deep sets, 2018. URL https://arxiv.org/abs/1703.06114.

28. Liang Fu. Fermi sets: Universal and interpretable neural architectures for fermions, 2026. URL https://arxiv.org/abs/2601.02508.

29. T. M. Whitehead, M. H. Michael, and G. J. Conduit. Jastrow correlation factor for periodic systems. *Physical Review B*, 94(3), 2016. ISSN 2469-9969. doi: 10.1103/physrevb.94.035157. URL http://dx.doi.org/10.1103/PhysRevB.94.035157.

30. Ao Chen and Markus Heyl. Empowering deep neural quantum states through efficient optimization. *Nature Physics*, 20(9):1476–1481, 2024. ISSN 1745-2481. doi: 10.1038/s41567-024-02566-1. URL http://dx.doi.org/10.1038/s41567-024-02566-1.

31. Chae-Yeun Park and Michael J. Kastoryano. Geometry of learning neural quantum states. *Phys. Rev. Res.*, 2:023232, May 2020. doi: 10.1103/PhysRevResearch.2.023232. URL https://link.aps.org/doi/10.1103/PhysRevResearch.2.023232.

32. Hoang-Thang Ta, Duy-Quy Thai, Anh Tran, Grigori Sidorov, and Alexander Gelbukh. PRKAN: Parameter-Reduced Kolmogorov-Arnold Networks, 2025. URL https://arxiv.org/abs/2501.07032.

33. Benjamin C. Koenig, Suyong Kim, and Sili Deng. LeanKAN: a parameter-lean Kolmogorov-Arnold network layer with improved memory efficiency and convergence behavior. *Neural Networks*, 192:107883, December 2025. doi: 10.1016/j.neunet.2025.107883. URL http://dx.doi.org/10.1016/j.neunet.2025.107883.

34. Jamison Moody and James Usevitch. Automatic grid updates for Kolmogorov-Arnold networks using layer histograms, 2025. URL https://arxiv.org/abs/2511.08570.

35. Vinicius Hernandes, Thomas Spriggs, Saqar Khaleefah, and Eliska Greplova. Adiabatic fine-tuning of neural quantum states enables detection of phase transitions in weight space, 2025. URL https://arxiv.org/abs/2503.17140.