Adaptive-basis sample-based neural diagonalization for quantum many-body systems

Simone Cantori[1,2⋆], Luca Brodoloni[1,2], Edoardo Recchi[1], Emanuele Costa[3,4], Bruno Juliá-Díaz[3,4] and Sebastiano Pilati[1,2]

1 School of Science and Technology, Physics Division, University of Camerino, I-62032 Camerino (MC), Italy
2 INFN-Sezione di Perugia, 06123 Perugia, Italy
3 Departament de Física Quàntica i Astrofísica, Facultat de Física, Universitat de Barcelona, 08028 Barcelona, Spain
4 Institut de Ciències del Cosmos de la Universitat de Barcelona, ICCUB, 08028 Barcelona, Spain

⋆ simone.cantori@unicam.it ,

Abstract

Estimating ground-state energies of quantum many-body systems is challenging due to the exponential growth of Hilbert space. Sample-based diagonalization (SBD) addresses this by projecting the Hamiltonian onto a subspace of selected basis configurations but works only for concentrated ground-state wave functions. We propose two neural network-enhanced SBD methods: sample-based neural diagonalization (SND) and adaptive-basis SND (AB-SND). Both leverage autoregressive neural networks for efficient sampling; AB-SND also optimizes a basis transformation to concentrate the wave function. We explore classically tractable single- and two-spin rotations, and more expressive unitaries implementable on quantum computers. On quantum Ising models, SND performs well for concentrated states, while AB-SND consistently outperforms SND and standard SBD in less concentrated regimes.

Received Date
Accepted Date
Published Date

Contents

---

1 Introduction

The accurate calculation of ground-state properties of quantum many-body systems is one of the central challenges in quantum chemistry and condensed matter physics. The exponential growth of the Hilbert space with the system size makes exact solutions intractable for large systems, necessitating the development of approximate computational methods. Deep learning methods have emerged as promising tools to address this challenge [1–3]. For example, supervised learning approaches have been used to predict ground-state energies based on labeled training data [4–7]. On the other hand, the introduction of neural quantum states (NQS), which represent wave functions using neural network (NN) architectures, has opened new possibilities for variational Monte Carlo simulations, circumventing the need of labeled data [8–10].

In quantum chemistry, a standard approach to tackle the problem of the Hilbert space size is represented by selected configuration interaction methods [11–14]. These employ predefined wave-function ansatzes, Monte Carlo sampling, or other empirical criteria to select a set of relevant basis configurations $|x^{(l)}\rangle$, labeled by the index l. The corresponding Hamiltonian matrix elements $\langle x^{(l)}|H|x^{(m)}\rangle$ are evaluated to define a subspace Hamiltonian matrix. The ground-state energy is then approximated by computing the lowest eigenvalue of this matrix. Recently, these approaches have also been adopted in the context of quantum computing, under the name of Sample-based Diagonalization (SBD) [15–17]. The key idea is to employ quantum circuits to sample relevant configurations, leading to what is dubbed Sample-based Quantum Diagonalization (SQD). In principle, quantum circuits might allow for the sampling of classically intractable distributions [18]. Machine learning algorithms have also been used to select relevant configurations [19–22]. Yet, the problem of how to efficiently truncate the Hilbert space, while minimizing the introduced approximation, is still open. In fact, SBD approaches are known to perform well only when the ground-state wave function is concentrated on the chosen computational basis [15], which means that its amplitudes are not negligible only on a small subset of basis elements. This strongly limits the regime of applicability of SBD methods.

In this article, we introduce two NN enhanced SBD approaches: sample-based neural diagonalization (SND) and its extension based on an adaptive basis, which we refer to as adaptive-basis SND (AB-SND). Both methods employ autoregressive NNs to efficiently sample basis configurations relevant for the estimation of the ground-state energy, as illustrated in Fig. 1. While SND operates in a fixed computational basis, AB-SND incorporates a basis transformation, allowing for improved performance when the ground state is not concentrated in

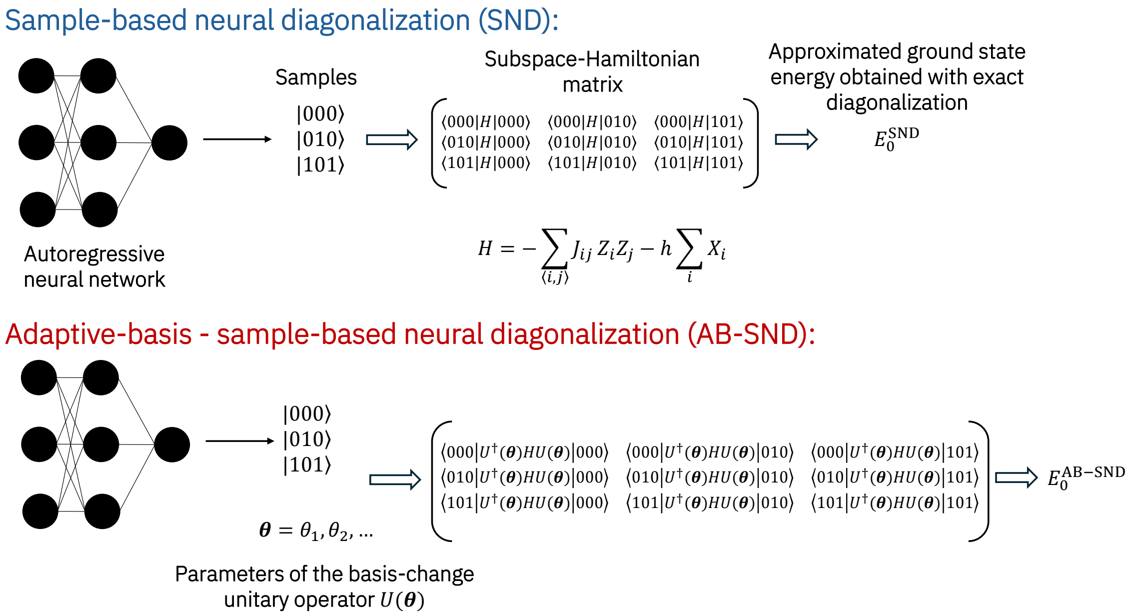

Figure 1: Scheme of the SND and AB-SND methods. In SND (top), an autoregressive neural network generates the bitstrings that define the subspace for diagonalization. In AB-SND (bottom), we also optimize some basis-transformation parameters $\boldsymbol{\theta}$, enabling us to perform sampling in a rotated basis where the ground state is more concentrated.

the original computational basis, a regime in which standard SBD techniques are doomed to fail. In our implementation, the basis change is performed using parameterized single-spin and two-spin rotations, which are efficiently computable on classical hardware. Additionally, we explore more expressive basis transformations, describing how quantum computers allow their implementation for large systems. The testbeds we consider are one-dimensional (1D) and two-dimensional (2D) ferromagnetic quantum Ising models and a 2D quantum spin-glass model. These models are chosen because the concentration of their ground state can be tuned, allowing us to test SBD approaches in different regimes of computational hardness, and because the benchmark ground-state energy can be exactly computed. Notably, we find that the AB-SND approach allows us to considerably extend the regime of applicability of SBD methods, in particular, when performed with the more general basis transformations.

The rest of this article is organized as follows: Sec. 2 provides the details of the methodology of both SND and AB-SND. Sec. 3 presents our numerical results for different quantum Ising models, comparing the performance of the proposed approaches to standard SBD techniques based on sampling from NQS variational ansatzes. In Sec. 4, we discuss the possibility of implementing AB-SND using quantum hardware, highlighting its potential for hybrid quantum-classical schemes. In Sec. 5, we summarize our findings and discuss future directions. Finally, the Appendices provide additional details on: the definition of the loss functions and their gradient calculations for SND (Appendix A) and AB-SND (App. B); the failure of standard SBD when the ground state is not concentrated (App. C); the regime of maximum error in AB-SND (App. D); the optimization of angle parameters via stochastic sampling (App. E); the effective temperature scaling for efficient sampling of unique configurations (App. F); the autoregressive NN architectures (App. G).

2 Methods

As in standard SBD approaches, the SND methods we introduce hereafter aim to approximate the ground-state energy of a quantum system by projecting the Hamiltonian onto a subspace spanned by a selected set of basis configurations. In our implementation, these configurations are selected from the standard computational basis, which consists of tensor products of the single-qubit basis states $|0\rangle$ and $|1\rangle$, namely, the eigenstates of the Pauli-Z operator. For a system of N spins, the computational basis states $|x\rangle = |x_1 x_2 \dots x_N\rangle$ correspond to bitstrings x, where each component $x_i \in \{0,1\}$. Given a set of S unique configurations, $\{|x^{(l)}\rangle\}_{l=1,\dots S}$, one constructs a subspace Hamiltonian by evaluating the matrix elements $\langle x^{(l)}|H|x^{(m)}\rangle$. The lowest eigenvalue E of this subspace matrix provides a variational upper bound for the ground-state energy, which converges to the exact value for $S \to 2^N$. Clearly, this limit is computationally impractical for system sizes $N \gg 10$. Yet, suitable criteria to select the subset of basis configurations lead to accurate approximations for feasible value of S.

In our framework, the selected configurations are sampled using autoregressive NNs, which can be trained to minimize E. More precisely, as in Ref. [15], we define the loss function to be minimized during training as

$$L = \sum_k P(S^{(k)}) E^{(k)}, \tag{1}$$

where $S^{(k)}$ represents a batch of bitstrings and $E^{(k)}$ is the lowest eigenvalue of the subspace Hamiltonian built on the k-th batch of bitstrings. In our case, the probability $P(S^{(k)})$ of sampling $S^{(k)}$ is given by the autoregressive NN. The minimization is performed using stochastic gradient-based methods, and the derivation of the gradient of L with respect to the weights of the NN is shown in the Appendix A.

The testbeds considered in this article are quantum Ising models described by the following Hamiltonian:

$$H = -\sum_{\langle i,j\rangle} J_{ij} Z_i Z_j - h \sum_i X_i \,, \tag{2}$$

where Z_i and X_i are Pauli operators acting on spin i, J_{ij} represents the interaction strength between the nearest-neighbor spins i and j, and h is the transverse field strength. Specifically, we consider three variants of this model:

1. 1D ferromagnetic transverse field Ising model (1D-TFIM), with $J_{ii+1} = 1$ for $i = 1, \dots, N$, and periodic boundary conditions, i.e., the spin $N+1$ is identified with the spin 1;
2. 2D ferromagnetic TFIM (2D-TFIM), with $J_{ij} = 1$ for i and j nearest-neighbor spins on a square lattice, and open boundary conditions;
3. 2D Edward-Anderson model (2D-EAM) on a square lattice, with J_{ij} randomly sampled from a normal distribution $\mathcal{N}(0,1)$ with zero mean and unit variance, and periodic boundary conditions.

It is worth pointing out that quantum Monte Carlo simulations of quantum Ising models are not affected by the negative sign problem. Thus, by adopting these models as testbeds for the SBD approaches, we have the opportunity to make comparisons against unbiased estimates of the ground-state energy, even beyond the 1D case where the Jordan-Wigner theory provides the exact solution. In particular, the 2D-EAM represents a challenging testbed due to the presence of disordered frustrated interactions. In addition, tuning h allows us to control the concentration of the ground state [23], thus testing the SBD approaches in different regimes. Indeed, for $h \to \infty$, the ground state of these models tends to an equally weighted superposition of the computational basis elements, i.e. $|++\cdots+\rangle$ (with $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$). In the opposite limit $h \to 0$, quantum fluctuations vanish, leading to a more concentrated ground-state wave function in the chosen computational basis. It is worth emphasizing that generic SBD methods, including our SND method without adaptive basis changes, are expected to perform well only for relatively small h. Numerical results confirming this expectation are shown in the Appendix A.

To extend the regime of applicability of SBD approaches, we introduce the AB-SND method. AB-SND improves the SND strategy by incorporating a basis transformation defined by a parameterized unitary operator $U(\boldsymbol{\theta})$, which maps the original computational basis to a rotated basis in which the ground state is more concentrated. As in SND, an autoregressive NN generates bitstring samples, but these are rotated by $U(\boldsymbol{\theta})$. The subspace Hamiltonian is then constructed from transformed matrix elements $\langle x^{(l)}|U(\boldsymbol{\theta})^{\dagger}HU(\boldsymbol{\theta})|x^{(m)}\rangle$, as shown in Fig. 1. In most of our experiments, we use a combination of single-spin rotations $U(\boldsymbol{\theta}) = \bigotimes_{i=1,\dots N} U_i(\theta_i)$, where

$$U_i(\theta_i) = \begin{bmatrix} \cos\frac{\theta_i}{2} & -\sin\frac{\theta_i}{2} \\ \sin\frac{\theta_i}{2} & \cos\frac{\theta_i}{2} \end{bmatrix}, \tag{3}$$

and each angle θ_i is an independent parameter for spin i. Because the rotations act independently on each spin and the Hamiltonian is composed of local Pauli operators, we can efficiently compute the transformed Hamiltonian $U^{\dagger}HU$ using classical hardware. In addition to single-spin rotations, we also implement non-overlapping two-spin rotations, which increase the expressive power of the basis transformation while remaining classically tractable. The unitary operator is defined as a composition of R_Y, R_{ZZ}, and R_X gates that act on each spin in the pair. The matrix representations of these gates are as follows:

$$\begin{gathered} R_Y(\alpha) = \begin{bmatrix} \cos\frac{\alpha}{2} & -\sin\frac{\alpha}{2} \\ \sin\frac{\alpha}{2} & \cos\frac{\alpha}{2} \end{bmatrix}, \\ R_X(\beta) = \begin{bmatrix} \cos\frac{\beta}{2} & -i\sin\frac{\beta}{2} \\ -i\sin\frac{\beta}{2} & \cos\frac{\beta}{2} \end{bmatrix}, \\ R_{ZZ}(\gamma) = \begin{bmatrix} e^{-i\frac{\gamma}{2}} & 0 & 0 & 0 \\ 0 & e^{i\frac{\gamma}{2}} & 0 & 0 \\ 0 & 0 & e^{i\frac{\gamma}{2}} & 0 \\ 0 & 0 & 0 & e^{-i\frac{\gamma}{2}} \end{bmatrix}. \end{gathered} \tag{4}$$

As mentioned above, the gates $R_{ZZ}(\gamma)$ act on non-overlapping spin pairs $\{0,1\}$, $\{2,3\}$, etc. However, the AB-SND framework is compatible with more general, potentially strongly entangling basis transformations, which could be implemented on quantum hardware. The use of quantum circuits for evaluating subspace matrix elements is discussed in Sec. 4. For AB-SND, the optimization of the angle parameter $\boldsymbol{\theta}$ can be approached in different ways, as discussed in the Appendices B and E.

3 Results

Hereafter, we analyze the performances of SND and AB-SND powered by local basis transformations on the three quantum spin models described in Sec. 2. As a benchmark, we consider a more standard SBD approach in which the configurations are sampled from the (squared modulus) exact ground-state wave function or from a very accurate approximation obtained via a variational Monte Carlo simulation. It is worth emphasizing that this procedure assumes

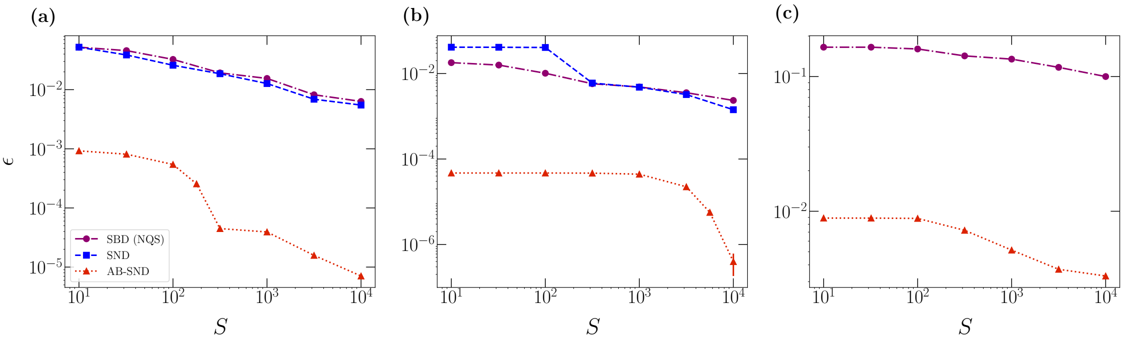

Figure 2: Relative error ϵ as a function of the number of unique configurations S used to build the subspace Hamiltonian. Panels (a), (b), and (c) display the relative errors for the 1D-TFIM, the 2D-TFIM, and the 2D-EAM with $N = 50$, $N = 100$, and $N = 64$ spins, respectively. The transverse field is $h = 0.5$ for the 1D-TFIM and the 2D-TFIM, while it is $h = 1$ for the 2D-EAM. In panel (b), we take into account the statistical uncertainty of the quantum Monte Carlo simulations used to determine the unbiased estimate of the ground-state energy.

that an accurate representation of the ground state can be obtained through a complementary computational technique. This allows us to execute standard SBD under very favorable conditions, thus representing a stringent benchmark for novel SND approaches. We sample configurations from an NQS in the form of a restricted Boltzmann machine. The latter is optimized using the NetKet library [24, 25]. Our numerical tests show that, as a sampling engine for SBD approaches, the NQS ansatz performs essentially as well as the exact ground state, at least for the system sizes for which the latter can be computed. Thus, in the following, we mostly adopt NQS sampling, unless otherwise specified.

To quantify the performance of the various SBD approaches, we compute the relative error $\epsilon = |\frac{y-E_0}{E_0}|$, where y represents the energy estimate from a given method, and E_0 is the exact ground-state energy. The latter is computed via the Jordan-Wigner transformations for the 1D model [26, 27], while for the 2D models we employ continuous-time projection quantum Monte Carlo simulations [28, 29]. The latter provide unbiased estimates, affected only by very small statistical uncertainties.

In Fig. 2, we analyze the relative errors in the three testbed models, considering relatively weak transverse fields h for which the ground state is concentrated in the chosen computational basis. As expected, all three SBD techniques perform well, showing a systematic accuracy improvement with the number of unique configurations S used to build the subspace. However, the SND performance deteriorates for the 2D-EAM (not shown). This effect may be attributed to the rugged energy landscapes occurring in spin-glass phases [30]. In this testbed, the standard SBD method based on NQS-sampled configurations performs better, but still reaches errors as large as 10% for computationally practical values of S. Notably, thanks to the additional variational flexibility introduced by the learnable basis transformation, the AB-SND method displays a systematic performance improvement with S also in the 2D-EAM. In fact, it consistently outperforms the other SBD approaches we consider, in all three testbeds.

In Fig. 3, the relative energy error is plotted as a function of the transverse field h, for different numbers of unique configurations S. As expected, the performance of SND and standard NQS-based SBD methods rapidly deteriorates as h increases, denoting the limited regime of applicability of these approaches. Instead, the AB-SND method, here implemented with single-spin rotations, is accurate also at significantly larger transverse fields. In fact, it reaches small relative errors also in the large h limit. This indicates that adaptive single-spin rotations enable a continuous interpolation between the original computational basis, in which

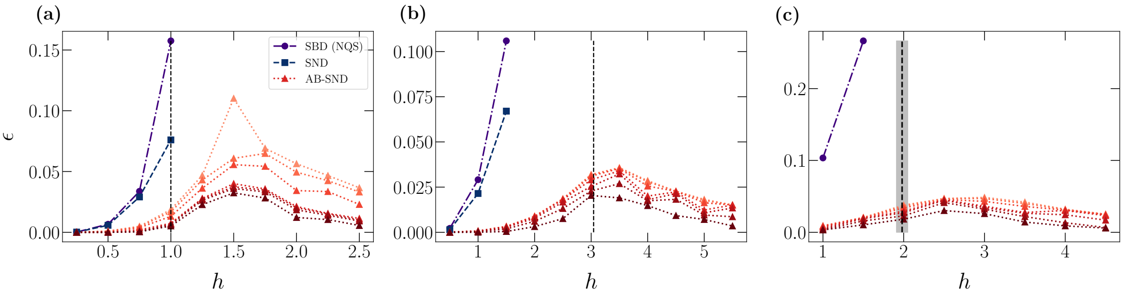

Figure 3: Relative error ϵ as a function of the transverse field h, for different numbers of unique configurations S. For AB-SND, darker colors correspond to larger S, with $S \in \{10, 10^{1.5}, 10^2, \ldots, 10^4\}$, while for SBD from the NQS and SND we consider only $S = 10^4$. Panel (a) represents the errors for the 1D-TFIM with $N = 50$ spins, panel (b) corresponds to the 2D-TFIM with $N = 100$ spins, and panel (c) to the 2D-EAM with $N = 64$ spins. The vertical dashed lines represent the corresponding quantum critical points, namely, the ferromagnetic transitions in the 1D-TFIM and the 2D-TFIM, and the spin-glass transition in the 2D-EAM. For the latter, the uncertainty is represented by the gray bar [29].

the ground state is concentrated in the $h \to 0$ limit, and the basis built using eigenstates of the X Pauli operator $|+\rangle, |-\rangle$, in which concentration occurs in the opposite limit. Sizable inaccuracies occur in the intermediate regime, approximately in the region around the quantum phase transitions occurring in the three testbed models, namely, the paramagnetic to ferromagnetic transition in the 1D-TFIM and the 2D-TFIM, and the quantum spin-glass transition in the 2D-EAM [29]. In the Appendix D, we provide numerical evidence that in the large S limit, the peak of the energy error approaches the critical point $h_c = 1$ of the 1D-TFIM.

In Fig. 4, we analyze how the computational cost scales with the system size N. Specifically, we determine the number of unique configurations S required to reach a relative error of 1%, considering the 1D-TFIM. A standard SBD approach based on NQS sampling displays a problematic scaling already for $h \gtrsim 0.5$, making it impractical to reach system sizes $N \simeq 100$ keeping the target accuracy. In this regime, the number of configurations required by the AB-SND approach powered by single-spin rotations is still essentially independent of N, denoting the important role of the basis change. However, the scaling approaches an exponential behavior slightly beyond the critical point $h_c = 1$, while it improves again for transverse fields $h \gg 1$. Better accuracies in the critical regime $h \simeq 1$ can be obtained within the AB-SND approach with more general basis transformations, as discussed in the following.

In Fig. 5, we compare the accuracies of the AB-SND approaches powered by single-spin rotations and by non-overlapping two-spin unitary operators. The latter approach allows introducing some entanglement and provides a more expressive basis transformation, while remaining classically tractable. In fact, we find that, at and slightly beyond the critical regime $h \gtrsim 1$, two-spin rotations lead to a sizable accuracy improvement compared to the single-spin case and, of course, compared to standard SBD based on the NQS-approximated ground-state sampling. Even better accuracies can be obtained by implementing classically intractable basis transformations using quantum hardware, as discussed in Section 4.

An additional important challenge for all stochastic SBD methods is the decreasing efficiency of sampling unique configurations as the number of samples increases. This phenomenon, also noted in Ref. [31], can be addressed in our framework through effective temperature scaling during inference. This procedure is discussed in Appendix F).

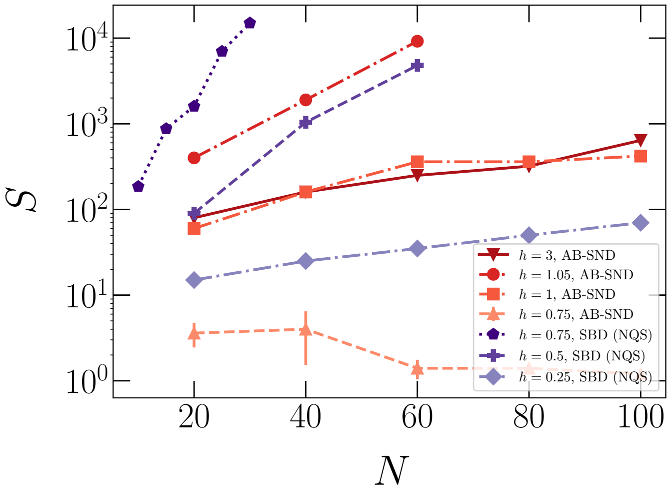

Figure 4: Number of unique configurations S required to reach relative errors below $\epsilon = 0.01$ as a function of the number of spins N for the 1D-TFIM at different transverse fields h. We compare the performance of a standard SBD approach powered by NQS sampling (shades of blue) with the one of the AB-SND with single-spin rotations (shades of red). The error-bars for the AB-SND method for $h = 0.5$ and $h = 0.75$ represent the estimated standard deviation of the average over five training processes.

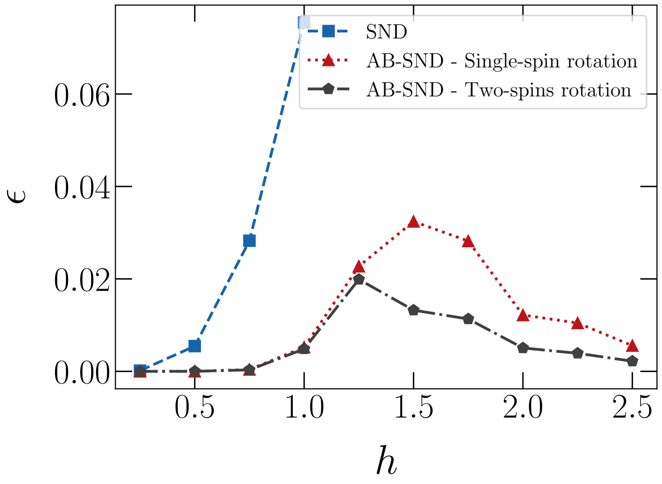

Figure 5: Relative error ϵ as a function of transverse field h, for $S = 10^4$ unique configurations in the truncated basis. The testbed is the 1D-TFIM with $N = 50$ spins. We compare the performances of the SND approach, with the ones of AB-SND approaches featuring single-spin and non-overlapping two-spin rotations.

4 Basis change using a quantum computer

For the quantum Ising models we consider, which feature up to two-spin couplings, basis changes based on combinations of single-spin and non-overlapping two-spin transformations can be efficiently performed on classical computers. More expressive basis changes could be efficiently performed using quantum circuits. Such transformations could further improve the performance of AB-SND approaches, especially near critical points. Indeed, a variational quantum eigensolver (VQE) algorithm has been shown to be able to accurately solve the 1D-TFIM also at criticality by optimizing a unitary transformation $U(\boldsymbol{\theta})$. This is achieved by minimizing the expectation value $\langle 0|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|0\rangle$ [32], which is estimated using (typically large) shot numbers. This is equivalent to a specific AB-SND strategy with only a single sampled configuration, namely, the state $|0\rangle = |00\ldots0\rangle$. The general AB-SND approach extends VQE by including the sampling of more basis elements, which leads to a finite matrix whose lowest eigenvalue is to be determined. Hereafter, we discuss how to implement a generic basis-change procedure using quantum computers. First, we set up a quantum circuit representing a parametrized basis-change unitary operator $U(\boldsymbol{\theta})$. Then, we can calculate the subspace-Hamiltonian elements $\langle x^{(l)}|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|x^{(m)}\rangle$ using the approach introduced in Ref. [33]. The diagonal terms $H_l = \langle x^{(l)}|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|x^{(l)}\rangle$ can be computed as standard expectation values. The off-diagonal terms can be calculated noticing that (using $i=\sqrt{-1}$)

$$\Re\langle x^{(l)}|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|x^{(m)}\rangle = H_{l+m} - \frac{H_l}{2} - \frac{H_m}{2}\,, \tag{5}$$

and

$$\Im\langle x^{(l)}|U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})|x^{(m)}\rangle = -H_{l+im} + \frac{H_l}{2} + \frac{H_m}{2}\,, \tag{6}$$

where

$$H_{l+m} = \frac{1}{\sqrt{2}}(\langle x^{(l)}| + \langle x^{(m)}|)U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})\frac{1}{\sqrt{2}}(|x^{(l)}\rangle + |x^{(m)}\rangle) \tag{7}$$

and

$$H_{l+im} = \frac{1}{\sqrt{2}}(\langle x^{(l)}| - i\langle x^{(m)}|)U^\dagger(\boldsymbol{\theta})HU(\boldsymbol{\theta})\frac{1}{\sqrt{2}}(|x^{(l)}\rangle + i|x^{(m)}\rangle)\,. \tag{8}$$

Finally, in our framework, both the parameters $\boldsymbol{\theta}$ and the weights of the autoregressive NN from which the bitstrings $x^{(l)}$ are sampled can be optimized as explained for the AB-SND procedure with single-spin rotations. To test this approach, we implement a small numerical experiment using a classical simulation of quantum circuits with $N=6$ qubits for the 1D-TFIM. The circuit ansatz we choose is rather shallow. It features one layer of R_Y gates acting on each qubit, and two blocks including R_{ZZ} operators acting on all nearest-neighbor pairs and R_X rotations acting on each qubit. As shown in Fig. 6, the improvement of the AB-SND predictions based on the circuit-based basis change compared with the case of single-spin transformations is significant. With the more general transformation, appreciable inaccuracies occur only very close to the critical point $h_c = 1$. We expect even better performances to be obtained by implementing basis transformations using deeper quantum circuits.

5 Conclusions

In this article, we introduced SND and its basis-adaptive extension, dubbed AB-SND. These are two NN-assisted sample-based diagonalization techniques for estimating ground-state energies of quantum many-body systems. Our tests show that AB-SND offers significant improvements over conventional SBD methods, particularly in regimes where the ground state is delocalized in the computational basis.

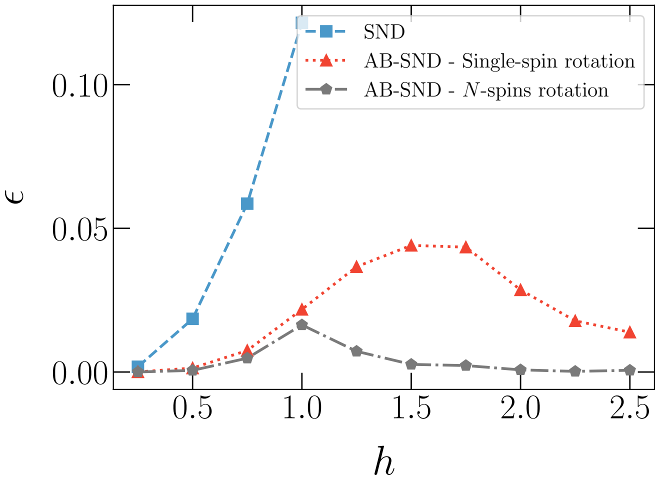

Figure 6: Relative error ϵ as a function of the transverse field h, for $S = 16$ unique configurations. The testbed model is the 1D-TFIM with $N = 6$ spins. We compare the accuracies of the SND approach with the ones of the AB-SND methods with single-spin rotations and with a general basis change performed by the (simulated) quantum circuit described in the text.

To extend the AB-SND approach beyond basis changes based on single-spin and two-spin operators, we explored the integration of AB-SND with quantum computing. As we explained, this allows the implementation of more general basis transformations. To test this procedure, we implemented a small-scale proof-of-concept experiment using classically simulated quantum circuits. In this setting, we used parametrized entangling gates to define a more expressive unitary $U(\boldsymbol{\theta})$, which led to improved performance compared to local spin rotations. Although these results are currently limited to small systems simulated classically, they suggest that AB-SND can be extended to hybrid quantum-classical workflows and could benefit from access to real quantum hardware in the future. In fact, the AB-SND approach driven by quantum circuits represents an extension of the VQE algorithm beyond the case of a single initial state.

Further developments could include the use of more expressive basis-change circuits, improved optimization strategies in complex energy landscapes, and systematic explorations of the performance of AB-SND methods in larger or more strongly correlated systems. By combining neural sampling with learnable basis transformations, AB-SND provides a flexible and scalable framework for studying quantum many-body problems across a wide range of regimes.

The essential scripts used in this study are available on GitHub [34].

Acknowledgements

Funding information Support from the following sources is acknowledged: PNRR MUR project PE0000023-NQSTI; PRIN 2022 MUR project “Hybrid algorithms for quantum simulators” – 2022H77XB7; PRIN-PNRR 2022 MUR project “UEFA” – P2022NMBAJ; National Centre for HPC, Big Data and Quantum Computing (ICSC), CN00000013 Spoke 7 – Materials & Molecular Sciences, under project INNOVATOR; CINECA awards ISCRA IsCc2_REASON and INF25_lincoln, for the availability of high-performance computing resources and support; EuroHPC Joint Undertaking for awarding access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland), through EuroHPC Development and Regular Access calls. E.R. acknowledges support from the Catalonia Quantum Academy Student Research Fellowship 2024 for

his research internship at the University of Barcelona.

A Loss function of SND and its gradient

As in Eq. (1), one can write the probability of a set of configurations $S^{(k)}$ as $P(S^{(k)}) = \prod_{x^{(l)} \in S^{(k)}} P(x^{(l)})$, where $P(x^{(l)})$ is the probability of sampling the bitstring $x^{(l)}$ defined by an autoregressive NN. The loss function L and its derivative with respect to a parameter of the network ω can be calculated as follows:

$$\begin{aligned} L &= \sum_k \Big(\prod_{x^{(l)} \in S^{(k)}} P(x^{(l)}) \Big) E^{(k)} \Longrightarrow \\ \frac{\partial L}{\partial \omega} &= \sum_k \frac{\partial}{\partial \omega} \Big(\prod_{x^{(l)} \in S^{(k)}} P(x^{(l)}) \Big) E^{(k)} \\ &= \sum_k \Big[\sum_{x^{(l)} \in S^{(k)}} \Big(\frac{\partial P(x^{(l)})}{\partial \omega} \prod_{x^{(m)} \neq x^{(l)}} P(x^{(m)}) \Big) \Big] E^{(k)}, \end{aligned} \tag{A.1}$$

where $x^{(l)}$ and $x^{(m)}$ are bitstrings in the batch $S^{(k)}$. With a straightforward rearrangement, the derivative can be rewritten as

$$\begin{aligned} \frac{\partial L}{\partial \omega} &= \sum_k \Big[\sum_{x^{(l)} \in S^{(k)}} \Big(\frac{\partial P(x^{(l)})}{\partial \omega} \frac{P(x^{(l)})}{P(x^{(l)})} \prod_{x^{(m)} \neq x^{(l)}} P(x^{(m)}) \Big) \Big] E^{(k)} \\ &= \sum_k P(S^{(k)}) \Big(\sum_{x^{(l)} \in S^{(k)}} \frac{\partial \log(P(x^{(l)}))}{\partial \omega} \Big) E^{(k)}, \end{aligned} \tag{A.2}$$

and the stochastic estimator is given by

$$\frac{\partial L}{\partial \omega} \simeq \frac{1}{K} \sum_{k=1}^{K} \Big(\sum_{x^{(l)} \in S^{(k)}} \frac{\partial \log(P(x^{(l)}))}{\partial \omega} \Big) E^{(k)}, \tag{A.3}$$

where the batches of bitstrings $S^{(k)}$ are sampled according to $P(S^{(k)})$. A baseline term is useful to stabilize the training process [9,10]. In our framework, we set it equal to the average energy over the K batches $\overline{E^{(k)}}$. Therefore, the loss function for SND reads:

$$L = \frac{1}{K} \sum_{k=1}^{K} \Big(\sum_{x^{(l)} \in S^{(k)}} \log(P(x^{(l)})) \Big) (E^{(k)} - \overline{E^{(k)}}). \tag{A.4}$$

B Loss function of AB-SND and its gradient

For the AB-SND method, the derivative with respect to the parameters of the NN used to sample basis configurations $\frac{\partial L}{\partial \omega}$ is calculated as discussed in the previous section. However, here we also want to optimize the basis-change parameters $\boldsymbol{\theta}$ in order to minimize the estimated ground-state energy. Notably, the rotation angles $\boldsymbol{\theta}$ are also used as a condition for the NN that generates the spin configurations. For this, they are provided as inputs preceding the spin values. Similarly to the SND method, one obtains:

$$\frac{\partial L}{\partial \theta_i} = \sum_k \frac{\partial}{\partial \theta_i} \Big(\prod_{x^{(l)} \in S^{(k)}} P(x^{(l)} | \boldsymbol{\theta}) \Big) E^{(k)} + \Big(\prod_{x^{(l)} \in S^{(k)}} P(x^{(l)} | \boldsymbol{\theta}) \Big) \frac{\partial}{\partial \theta_i} E^{(k)}. \tag{B.1}$$

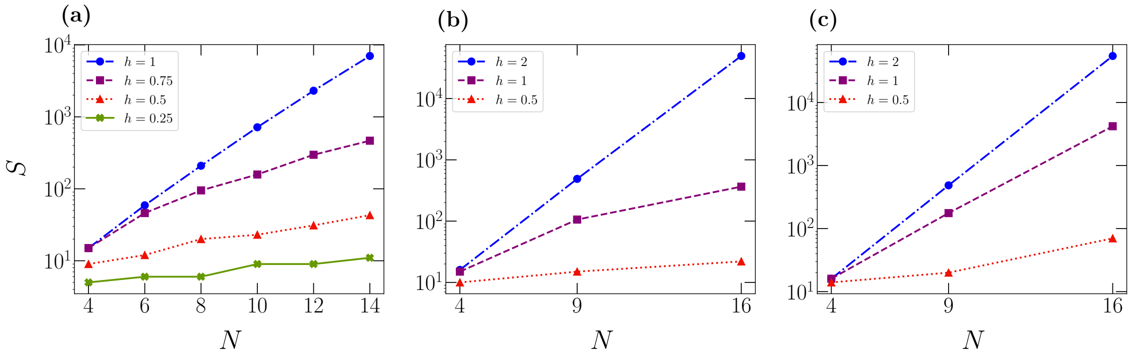

Figure 7: Number of unique configurations S required to reach the relative error $\epsilon = 0.01$ as a function of number of spins N using a standard SBD approach with configurations sampled from the exact ground state. Panels (a), (b), and (c) display results for the 1D-TFIM, the 2D-TFIM, and the 2D-EAM, respectively. Different datasets in each panel correspond to different transverse fields h.

From the Hellmann–Feynman theorem, one can write

$$\frac{\partial E^{(k)}(\theta_i)}{\partial \theta_i} = \left\langle \psi_0^{(k)} \middle| \frac{\partial \hat{H}}{\partial \theta_i} \middle| \psi_0^{(k)} \right\rangle, \tag{B.2}$$

where $\psi_0^{(k)}$ is the estimated ground-state wave function. In AB-SND, we use a basis-change unitary operator $U(\boldsymbol{\theta})$, so

$$\frac{\partial E^{(k)}(\theta_i)}{\partial \theta_i} = \left\langle \psi_0^{(k)} \middle| \frac{\partial (U^\dagger(\boldsymbol{\theta})\hat{H}U(\boldsymbol{\theta}))}{\partial \theta_i} \middle| \psi_0^{(k)} \right\rangle. \tag{B.3}$$

This quantity can be calculated using the parameter-shift rule [35, 36]. Alternatively, if $U(\boldsymbol{\theta})$ is implemented using classical algorithms, the gradient can be calculated using automatic differentiation, e.g., via the Pytorch library [37]. Finally, the derivative with respect to a generic rotation angle reads:

$$\frac{\partial L}{\partial \theta_i} \simeq \frac{1}{K}\sum_{k=1}^{K}\left[\left(\sum_{x^{(l)}\in S^{(k)}} \frac{\partial \log(P(x^{(l)}|\boldsymbol{\theta}))}{\partial \theta_i}\right)E^{(k)} + \left\langle \psi_0^{(k)} \middle| \frac{\partial (U^\dagger(\boldsymbol{\theta})\hat{H}U(\boldsymbol{\theta}))}{\partial \theta_i} \middle| \psi_0^{(k)} \right\rangle\right]. \tag{B.4}$$

We also implement an alternative approach to optimize the rotation angles, which avoids the multiple diagonalization steps used in the parameter-shift rule. This approach involves sampling the angles $\boldsymbol{\theta}$ from an additional autoregressive NN. It is detailed in Appendix E.

C Failure of standard SBD approaches at large transverse field

The accuracy of SBD approaches noticeably depends on how the computational basis elements used to build the truncated Hamiltonian matrix are sampled. In Fig. 7, we show that even when the exact ground-state wave function is used for sampling, without adaptive basis rotations the SBD method fails when the ground-state wave function is not strongly concentrated in the chosen computational basis. In fact, beyond the small transverse field regime $h \ll 1$, the truncated basis size S required to reach the target accuracy of 1% approaches an exponential scaling with the system size N.

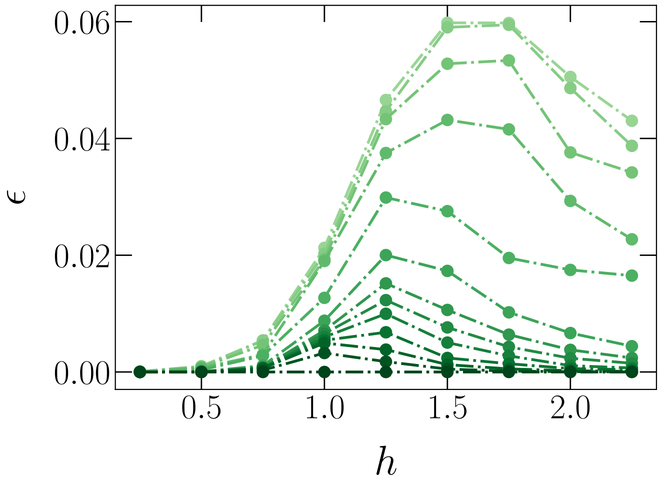

Figure 8: Relative error ϵ as a function of the transverse field h, for different numbers of unique configuration S. Darker colors mean higher S, with $S \in \{4, 8, 16, 32, 64, 128, 256, 384, 512, 640, 768, 896, 1024\}$). The configurations are sampled from the exact ground state. Adaptive single-spin rotations are then applied, as explained in the main text. These results are for the 1D-TFIM with $N = 10$ spins.

D Most challenging regime for AB-SND

In Fig. 8 we report numerical evidence showing that, in the large S limit, the peak of the energy error obtained via the AB-SND method drifts towards the critical point of the ferromagnetic quantum phase transition. The chosen testbed is the 1D-TFIM with $N = 10$ spins. The configurations are sampled from the exact ground state.

E Loss function of AB-SND with sampled rotational parameters

Instead of optimizing the basis change using the Hellmann–Feynman theorem, one can employ an additional autoregressive NN to sample the parameters $\boldsymbol{\theta}$ that define the basis-change unitary operator. In this alternative procedure, the probability of a set of spin configurations is written as: $P(S^{(k)}) = P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}) P_{\boldsymbol{\omega}}(S^{(k)}|\boldsymbol{\theta}^{(k)}) = P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}) \prod_{x^{(l)} \in S^{(k)}} P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)})$, where $\boldsymbol{\nu}$ denotes the weights of the neural network responsible for sampling the parameters $\boldsymbol{\theta}^{(k)}$, while $\boldsymbol{\omega}$ denotes the weights of the neural network responsible for sampling the bitstrings in the batch $S^{(k)}$. The former, is an autoregressive NN that provides the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ of a Von Mises distribution [38] from which the basis-change parameters $\boldsymbol{\theta}^{(k)}$ are sampled. Due to the autoregressive architecture, μ_i and κ_i depend on the angles θ_j, with $j < i$. The rotation angles $\boldsymbol{\theta}^{(k)}$ are also used as conditions for the other autoregressive NN that defines the probabilities $P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)})$ of each bitstrings $x^{(l)}$. The loss function to be minimized is defined as:

$$L = \sum_k \int d\boldsymbol{\theta}^{(k)} \Big[P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}) \prod_{x^{(l)} \in S^{(k)}} P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)}) \Big] E^{(k)} \tag{E.1}$$

and its derivatives are

$$\begin{aligned}\frac{\partial L}{\partial \omega} &= \sum_k \int d\boldsymbol{\theta}^{(k)} \Big[P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}) \frac{\partial}{\partial \omega} \prod_{x^{(l)} \in S^{(k)}} P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)}) \Big] E^{(k)} \\ &= \sum_k \int d\boldsymbol{\theta}^{(k)} \Big[P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}) P_{\boldsymbol{\omega}}(S^{(k)}|\boldsymbol{\theta}^{(k)}) \Big(\sum_{x^{(l)} \in S^{(k)}} \frac{\partial \log(P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)}))}{\partial \omega} \Big) \Big] E^{(k)} \\ &\simeq \frac{1}{K} \sum_{k=1}^{K} \Big[\sum_{x^{(l)} \in S^{(k)}} \frac{\partial \log(P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)}))}{\partial \omega} \Big] E^{(k)} ,\end{aligned} \tag{E.2}$$

and

$$\begin{aligned}\frac{\partial L}{\partial \nu} &= \sum_k \int d\boldsymbol{\theta}^{(k)} \Big[\frac{\partial P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)})}{\partial \nu} \prod_{x^{(l)} \in S^{(k)}} P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)}) \Big] E^{(k)} \\ &= \sum_k \int d\boldsymbol{\theta}^{(k)} \Big[\frac{\partial P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)})}{\partial \nu} \frac{P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)})}{P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)})} \prod_{x^{(l)} \in S^{(k)}} P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)}) \Big] E^{(k)} \\ &= \sum_k \int d\boldsymbol{\theta}^{(k)} P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}) P_{\boldsymbol{\omega}}(S^{(k)}|\boldsymbol{\theta}^{(k)}) \Big[\frac{\partial \log(P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}))}{\partial \nu} \Big] E^{(k)} \\ &\simeq \frac{1}{K} \sum_{k=1}^{K} \Big[\frac{\partial \log(P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)}))}{\partial \nu} \Big] E^{(k)} ,\end{aligned} \tag{E.3}$$

where $\omega \in \boldsymbol{\omega}$ and $\nu \in \boldsymbol{\nu}$ denote single weights of the corresponding neural networks.

Therefore, including the baseline term, the loss is evaluated as:

$$L = \frac{1}{K} \sum_{k=1}^{K} \Big[\log(P_{\boldsymbol{\nu}}(\boldsymbol{\theta}^{(k)})) + \Big(\sum_{x^{(l)} \in S^{(k)}} \log(P_{\boldsymbol{\omega}}(x^{(l)}|\boldsymbol{\theta}^{(k)})) \Big) \Big] (E^{(k)} - \overline{E^{(k)}}) . \tag{E.4}$$

Sampling $\boldsymbol{\theta}^{(k)}$ from a conditional autoregressive NN is conceptually and practically appealing. Yet, the test results visualized in Fig. 9 indicate that this approach does not perform better than the gradient-based optimization described in Appendix B. The accuracy shows a similar improvement rate as a function of S, with an approximately constant upward shift, denoting a marginally worse performance.

F Problem of sampling unique configurations

As the size S of the configuration set increases, the probability of sampling already included configurations rapidly rises. This leads to a problematic computational cost for sampling unique configurations. To overcome this problem, we introduce the effective temperature parameter T [39]. This controls the shape of the output distribution by tuning the width of the softmax activation function in the final NN layer. The NN produces two outputs for each spin, and the softmax turns these outputs into probabilities of sampling 0 or 1. Specifically, if Y_0 and Y_1 are the two outputs, then

$$\mathrm{Softmax}(Y_q) = \frac{\exp(Y_q/T)}{\sum_{r \in \{0,1\}} \exp(Y_r/T)} . \tag{F.1}$$

Complementary strategies were introduced in Ref. [31]. During training, we set $T = 1$, but in the inference phase we increase T to have a broader distribution, thus favoring the sampling of different outputs. This effect is demonstrated in Fig. 10. Indeed, while with $T = 1$ the

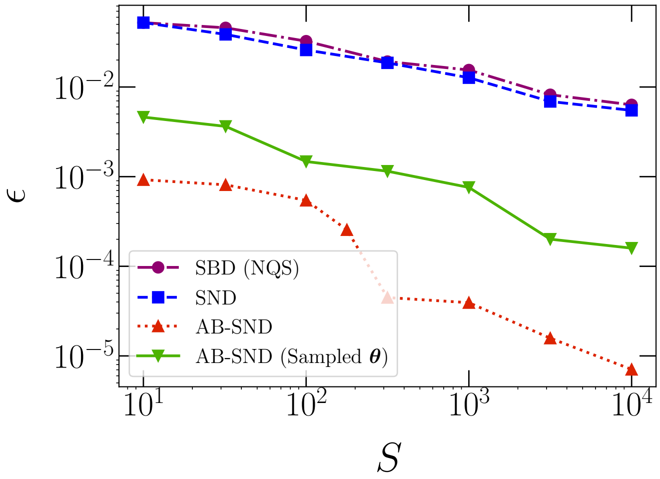

Figure 9: Relative error ϵ as a function of the number of unique configurations S used to build the subspace Hamiltonian. These results are for the 1D-TFIM with $h = 0.5$ and $N = 50$. We compare the accuracies of a standard SBD approach powered by NQS sampling, the SND approach, and two AB-SND approaches. The (green) upside-down triangles refer to the AB-SND approach with rotation parameters $\boldsymbol{\theta}^{(k)}$ sampled from an autoregressive NN, as explained in the text.

ratio between the number of unique configurations S and of total configurations N_s decreases almost exponentially fast, slightly larger effective temperatures suffice to significantly increase the number of unique configurations, thus drastically decreasing the computational cost of sampling.

Importantly, increasing T to values appropriate for efficient sampling does not reduce the performance of SND approaches. This is demonstrated in Fig. 11, where one observes that, for $S \gtrsim 10^2$, values of $T \in [1, 1.6]$ provide comparable accuracies for the 1D-TFIM.

G Autoregressive neural network and hyperparameters

We use a transformer encoder [40] with causal mask, two layers, four attention heads, and embedding size equal to 64. During training, we use $K = 16$ batches for SND and $K = 4$ batches for AB-SND, and number of sampled bitstrings equal to $BS = 128$. It is worth emphasizing that the latter does not coincide with the number of unique configurations S used during inference. Also, the rare repeated configurations are simply discarded. A comparison of the performances obtained with different values of K and BS is shown in Fig. 12. Notably, the accuracy of the SND method is not significantly affected by different choices for these parameters.

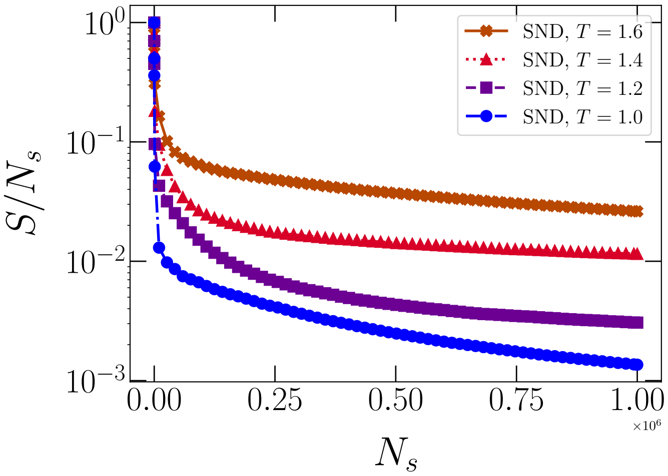

Figure 10: Ratio between number of unique configurations S and total number of samples N_s as a function of N_s for different values of the effective temperature T. The samples are obtained via SND for the 1D-TFIM with $N = 50$ and $h = 0.5$.

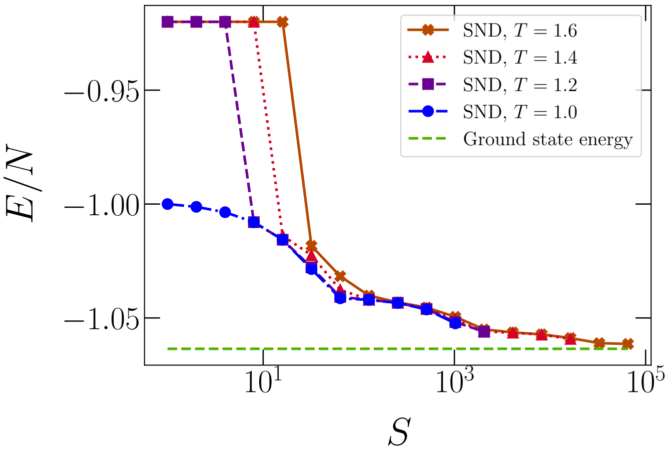

Figure 11: Energy per spin E/N obtained via SND as a function of the number of unique configurations S used to build the subspace Hamiltonian. The different datasets correspond to different effective temperatures T. The testbed model is the 1D-TFIM with $N = 50$ spin and transverse field $h = 0.5$.

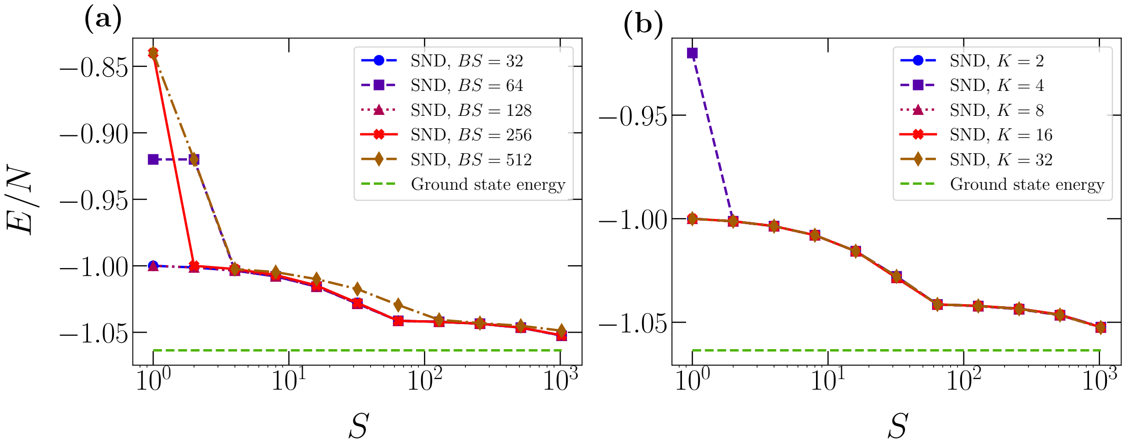

Figure 12: Energy per spin E/N obtained with SND as a function of the number of unique configurations S used to build the subspace Hamiltonian for the 1D-TFIM featuring $N = 50$ spins and $h = 0.5$. Panels (a) and (b) display results for different values of BS and K, respectively.

References

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborová, *Machine learning and the physical sciences*, Rev. Mod. Phys. **91**, 045002 (2019), doi:10.1103/RevModPhys.91.045002.

[2] J. Carrasquilla, *Machine learning for quantum matter*, Advances in Physics: X **5**(1), 1797528 (2020), doi:10.1080/23746149.2020.1797528.

[3] H. J. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. A. L. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo *et al.*, *Roadmap on machine learning in electronic structure*, Electronic Structure **4**(2), 023004 (2022), doi:10.1088/2516-1075/ac572f.

[4] V. Dunjko and H. J. Briegel, *Machine learning & artificial intelligence in the quantum domain: a review of recent progress*, Reports on Progress in Physics **81**(7), 074001 (2018), doi:10.1088/1361-6633/aab406.

[5] N. Saraceni, S. Cantori and S. Pilati, *Scalable neural networks for the efficient learning of disordered quantum systems*, Phys. Rev. E **102**, 033301 (2020), doi:10.1103/PhysRevE.102.033301.

[6] K. Mills, K. Ryczko, I. Luchak, A. Domurad, C. Beeler and I. Tamblyn, *Extensive deep neural networks for transferring small scale learning to large scale systems*, Chem. Sci. **10**, 4129 (2019), doi:10.1039/C8SC04578J.

[7] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, *Prediction errors of molecular machine learning models lower than hybrid DFT error*, Journal of Chemical Theory and Computation **13**(11), 5255 (2017), doi:10.1021/acs.jctc.7b00577, PMID: 28926232.

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

[9] O. Sharir, Y. Levine, N. Wies, G. Carleo and A. Shashua, *Deep autoregressive models for the efficient variational simulation of many-body quantum systems*, Phys. Rev. Lett. **124**, 020503 (2020), doi:10.1103/PhysRevLett.124.020503.

[10] T. D. Barrett, A. Malyshev and A. I. Lvovsky, *Autoregressive neural-network wavefunctions for ab initio quantum chemistry*, Nature Machine Intelligence **4**(4), 351 (2022), doi:10.1038/s42256-022-00461-z.

[11] V. Abraham and N. J. Mayhall, *Selected configuration interaction in a basis of cluster state tensor products*, Journal of Chemical Theory and Computation **16**(10), 6098–6113 (2020), doi:10.1021/acs.jctc.0c00141.

[12] L. Craciunescu, A. W. Prentice and M. J. Paterson, *Selected configuration interaction for high accuracy and compact wave functions: Propane as a case study*, The Journal of Chemical Physics **162**(3), 034102 (2025), doi:10.1063/5.0233542.

[13] E. Giner, A. Scemama and M. Caffarel, *Using perturbatively selected configuration interaction in quantum Monte Carlo calculations*, Canadian Journal of Chemistry **91**(9), 879 (2013), doi:10.1139/cjc-2013-0017.

[14] T. Yanagisawa, *Quantum Monte Carlo diagonalization for many-fermion systems*, Phys. Rev. B **75**, 224503 (2007), doi:10.1103/PhysRevB.75.224503.

[15] J. Robledo-Moreno, M. Motta, H. Haas, A. Javadi-Abhari, P. Jurcevic, W. Kirby, S. Martiel, K. Sharma, S. Sharma, T. Shirakawa, I. Sitdikov, R.-Y. Sun *et al.*, *Chemistry beyond the scale of exact diagonalization on a quantum-centric supercomputer*, Science Advances **11**(25) (2025), doi:10.1126/sciadv.adu9991.

[16] N. Yoshioka, M. Amico, W. Kirby, P. Jurcevic, A. Dutt, B. Fuller, S. Garion, H. Haas, I. Hamamura, A. Ivrii, R. Majumdar, Z. Minev *et al.*, *Krylov diagonalization of large many-body hamiltonians on a quantum processor*, Nature Communications **16**(1) (2025), doi:10.1038/s41467-025-59716-z.

[17] K. Kanno, M. Kohda, R. Imai, S. Koh, K. Mitarai, W. Mizukami and Y. O. Nakagawa, *Quantum-selected configuration interaction: Classical diagonalization of Hamiltonians in subspaces selected by quantum computers*, doi:https://arxiv.org/abs/2302.11320.

[18] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell *et al.*, *Quantum supremacy using a programmable superconducting processor*, Nature **574**(7779), 505 (2019), doi:https://doi.org/10.1126/sciadv.adu9991.

[19] J. P. Coe, *Machine learning configuration interaction*, Journal of Chemical Theory and Computation **14**(11), 5739–5749 (2018), doi:10.1021/acs.jctc.8b00849.

[20] P. Bilous, L. Thirion, H. Menke, M. W. Haverkort, A. Pálffy and P. Hansmann, *Neural-network-supported basis optimizer for the configuration interaction problem in quantum many-body clusters: Feasibility study and numerical proof*, Phys. Rev. B **111**, 035124 (2025), doi:10.1103/PhysRevB.111.035124.

[21] P. Bilous, A. Pálffy and F. Marquardt, *Deep-learning approach for the atomic configuration interaction problem on large basis sets*, Phys. Rev. Lett. **131**, 133002 (2023), doi:10.1103/PhysRevLett.131.133002.

[22] M. Rano and D. Ghosh, *Efficient machine learning configuration interaction for bond breaking problems*, The Journal of Physical Chemistry A **127**(16), 3705 (2023), doi:10.1021/acs.jpca.2c09103.

[23] J. Yu, J. R. Moreno, J. T. Iosue, L. Bertels, D. Claudino, B. Fuller, P. Groszkowski, T. S. Humble, P. Jurcevic, W. Kirby, T. A. Maier, M. Motta *et al.*, *Quantum-centric algorithm for sample-based Krylov diagonalization*, doi:https://arxiv.org/abs/2501.09702.

[24] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola *et al.*, *Netket: A machine learning toolkit for many-body quantum systems*, SoftwareX p. 100311 (2019), doi:10.1016/j.softx.2019.100311.

[25] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev and G. Carleo, *NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems*, SciPost Phys. Codebases p. 7 (2022), doi:10.21468/SciPostPhysCodeb.7.

[26] P. Pfeuty, *The one-dimensional Ising model with a transverse field*, Annals of Physics **57**(1), 79 (1970), doi:https://doi.org/10.1016/0003-4916(70)90270-8.

[27] A. P. Young and H. Rieger, *Numerical study of the random transverse-field Ising spin chain*, Phys. Rev. B **53**, 8486 (1996), doi:10.1103/PhysRevB.53.8486.

[28] F. Becca and S. Sorella, *Quantum Monte Carlo approaches for correlated systems*, Cambridge University Press (2017).

[29] L. Brodoloni and S. Pilati, *Zero-temperature Monte Carlo simulations of two-dimensional quantum spin glasses guided by neural network states*, Phys. Rev. E **110**, 065305 (2024), doi:10.1103/PhysRevE.110.065305.

[30] C.-Y. Park and M. J. Kastoryano, *Geometry of learning neural quantum states*, Phys. Rev. Res. **2**, 023232 (2020), doi:10.1103/PhysRevResearch.2.023232.

[31] P. Reinholdt, K. M. Ziems, E. R. Kjellgren, S. Coriani, S. P. A. Sauer and J. Kongsted, *Fundamental limitations in sample-based quantum diagonalization methods*, doi:https://arxiv.org/abs/2501.07231.

[32] W. W. Ho and T. H. Hsieh, *Efficient variational simulation of non-trivial quantum states*, SciPost Physics **6**(3) (2019), doi:10.21468/scipostphys.6.3.029.

[33] A. Asthana, A. Kumar, V. Abraham, H. Grimsley, Y. Zhang, L. Cincio, S. Tretiak, P. A. Dub, S. E. Economou, E. Barnes and N. J. Mayhall, *Quantum self-consistent equation-of-motion method for computing molecular excitation energies, ionization potentials, and electron affinities on a quantum computer*, Chem. Sci. **14**, 2405 (2023), doi:10.1039/D2SC05371C.

[34] S. Cantori, *Adaptive-basis sample-based neural diagonalization for quantum many-body systems*, GitHub repository, https://github.com/simonecantori/Sample-based-Neural-Diagonalization.

[35] J. Li, X. Yang, X. Peng and C.-P. Sun, *Hybrid quantum-classical approach to quantum optimal control*, Phys. Rev. Lett. **118**, 150503 (2017), doi:10.1103/PhysRevLett.118.150503.

[36] K. Mitarai, M. Negoro, M. Kitagawa and K. Fujii, *Quantum circuit learning*, Phys. Rev. A **98**, 032309 (2018), doi:10.1103/PhysRevA.98.032309.

[37] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga and A. Lerer, *Automatic differentiation in pytorch*, In *NIPS-W* (2017).

[38] K. Mardia and P. Jupp, *Directional Statistics*, Wiley Series in Probability and Statistics. Wiley, ISBN 9780470317815 (2009), https://books.google.it/books?id=PTNiCm4Q-M0C.

[39] H. Xuan, B. Yang and X. Li, *Exploring the impact of temperature scaling in softmax for classification and adversarial robustness*, doi:https://arxiv.org/abs/2502.20604.

[40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, *Attention is all you need*, doi:https://arxiv.org/abs/1706.03762.