Framework for efficient *ab initio* electronic structure with Gaussian Process States

Yannic Rath [*] and George H. Booth [†]
Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom

(Received 2 February 2023; revised 18 April 2023; accepted 27 April 2023; published 10 May 2023)

We present a general framework for the efficient simulation of realistic fermionic systems with modern machine-learning-inspired representations of quantum many-body states, towards a universal tool for *ab initio* electronic structure. These machine-learning-inspired *Ansätze* have recently come to the fore in both a (first-quantized) continuum and discrete Fock space representations, where, however, the inherent scaling of the latter approach for realistic interactions has so far limited practical applications. With application to the "Gaussian Process State," a recently introduced *Ansatz* inspired by systematically improvable kernel models in machine learning, we discuss different choices to define the representation of the computational Fock space. We show how local representations are particularly suited for stochastic sampling of expectation values, while also indicating a route to overcome the discrepancy in the scaling compared with continuum-formulated models. We are able to show competitive accuracy for systems with up to 64 electrons, including a simplified (yet fully *ab initio*) model of the Mott transition in three-dimensional hydrogen, indicating a significant improvement over similar approaches, even for moderate numbers of configurational samples.

DOI: 10.1103/PhysRevB.107.205119

I. INTRODUCTION

Interactions of many electrons with each other and their environmental nuclear potentials give rise to almost all the complexity in chemical and materials science. Accurate simulations of these quantum particles with their known interactions can describe emergent properties of a system and are therefore a key challenge linking the electronic scale and trustworthy predictions of relevant physical observables from fundamental physical principles. However, these quantum problems are formulated in a Hilbert space which inherently scales exponentially with number of particles, and hence numerically tractable many-body wave function approaches typically require approximations, effectively compressing the information in this space.

Many of these approximate wave function approaches are based around explicitly imposing an appropriate functional form of the many-electron state. These well-established representations are generally directly informed by exploiting some physical characteristics or intuition of the state which is exploited in order to describe them compactly, such as Laughlin [1], BCS [2], or Gutzwiller [3] states. However, since the structure of the target state depends specifically on the underlying physics of the studied system, most introduced state approximations for many-electron wave functions are not universally suitable for all systems of interest. This makes these representations successful for specific classes of systems. This approach can also encompass more flexible forms which are still nevertheless restricted in their applicability, such as coupled-cluster wave functions [4] (which require low-rank correlations) or tensor networks [5] (which require low entanglement).

The framework of variational Monte Carlo (VMC) [6] makes it possible to use any functional form as a model for the many-body wave function, as long as it can be efficiently sampled in a chosen computational basis. Optimization of its parameters and extraction of many-body expectation values of interest are then enabled through efficient stochastic sampling of the configuration space. In recent years, this has enabled wave function models inspired by classical machine learning (ML) to come to the fore due to their ability to describe complicated functions of many variables in a black-box and efficient fashion [7–14]. Importantly, such ML models, e.g., neural network architectures or kernel models, are typically not limited by rigid imposed functional forms, and in principle can be improved systematically to arbitrary accuracy to model the many-body correlations in the quantum state. Although the speed of this convergence in desired expectation values with the complexity of the model is not guaranteed, the systematic and unbiased ability to describe many-body effects without restriction in rank or range represents probably the most important advantage over other established models typically used in VMC such as (Slater-)Jastrow *Ansätze* [6,15].

With increasingly many successful applications of ML-inspired models for quantum many-body wave functions often challenging the state of the art [16,17], this route is considered a promising candidate for a truly universal quantum many-body method. In this paper, we build on the Gaussian Process State (GPS) [18–20], an ML-inspired wave function

[*] yannic.rath@kcl.ac.uk
[†] george.booth@kcl.ac.uk

model motivated from Bayesian kernel models. This *Ansatz* takes a particularly simple form which has been shown to practically reach accuracies comparable to similarly systematically improvable neural-network-inspired states. We apply the state to electronic problems interacting with the physical long-ranged Coulomb interaction as a step towards realistic electronic structure and discuss the challenges which arise in this context for VMC methods in a Fock space picture. We ameliorate many of these difficulties when sampling dense Hamiltonians in this representation by combining the GPS with approaches previously considered in the orbital-space VMC literature [21], in particular constructing the Fock space from localized basis functions. This also paves the way to achieve a reduced (asymptotic) scaling of the method by exploiting the emerging natural sparsity of the Hamiltonian in this representation [22–25], which would bring the scaling into agreement with continuum real-space VMC [23,26], as has recently also been the focus of ML-inspired quantum states [27–37]. This opens the possibility to practically treat larger systems.

We demonstrate the applicability of the methodology in combination with our state parametrization, with application to different test systems, reaching competitive energies compared with established methodologies in the investigation of electronic systems with the real Coulomb interaction. Our results include the description of a correlation-driven metal-insulator transition in a minimal 64-atom model of a correlated hydrogen material, representing a system size beyond what has been discussed with related approaches and towards realistic materials science applications. Furthermore, we expect the discussed developments to be able to be combined seamlessly with other ML-inspired models, including recent parametrizations specifically tailored for electronic systems [38–40]. The following section introduces the GPS *Ansatz* to model the electronic wave function which we utilize in the VMC framework for *ab initio* quantum chemical calculations, which we outline in Sec. III. Finally, we present benchmarking results for one- and three-dimensional hydrogen materials in Sec. IV.

II. GAUSSIAN PROCESS STATES

The general procedure for approximating the ground state of a system with VMC is conceptually simple: Having defined a functional form for the wave function in the computational basis, defining a mapping from basis states $|x\rangle$ to the configurational wave function amplitudes $\langle\Psi|x\rangle$, expectation values are evaluated by stochastic sampling from the computational basis. Through a numerical minimization, the chosen parametrization of the state can be optimized to find a suitable approximation of the (generally unknown) many-body target state, here considered to be the electronic ground state of the *ab initio* chemical system. Key to the success is the choice of the trial wave function *Ansatz* and its ability to faithfully represent the physics of the target state as well as exhibiting as compact a form as possible to facilitate optimization. These properties have recently been well served by the application of traditional machine learning models which, if carefully designed, do not require a low scaling in entanglement of the target state for efficient representation [9,41].

Various neural network architectures have recently been applied as a model to represent *ab initio* wave functions in both a first-quantized [27–37] and second-quantized perspective [42–46]. In this paper, we follow the latter approach, in which we construct a computational basis from Fock states identifying the electronic occupancies of a finite number of molecular orbitals. We use a similar ML-inspired model for the wave function that was recently introduced, dubbed the Gaussian Process State (GPS) [18–20]. The GPS representation can be derived from the application of a general kernel model, as in Gaussian process regression or kernel ridge regression. Kernel models in machine learning recast the problem into a very high dimensional feature space, at which point the data can be described via a linear model. By allowing the effective dimension of this feature space to be systematically enlarged (and, in this paper, variationally optimized), the expressiveness of the GPS can be improved systematically. In this way, the GPS represents a universal approximator of a target state, not restricted to a rigid functional form or specific correlation characteristics.

Here, we utilize the recent formulation of the GPS, which can be viewed as a model supported by a set of M unentangled product states as data points explicitly driving the representation [20]. The number of product states, M, in the following referred to as the "support dimension" of the model, serves as the single hyperparameter of the model controlling its complexity (and hence both its expressibility and number of parameters). Therefore, in keeping with the approach of other ML-inspired *Ansätze*, it can in principle span any state in the Hilbert space as M increases.

The considered GPS model associates many-body configurations with their wave function amplitudes according to a simple form, given by

$$\Psi(x) = \langle x|\Psi\rangle = \exp\left(\sum_{\alpha=1}^{M}\prod_{i=1}^{L}\epsilon_{\alpha,i,x_i}\right), \tag{1}$$

which is specified by $M\times L\times 4$ continuous variational parameters in the tensor ϵ. Within this parametrization, each local occupancy of the L spatial orbitals, denoted by x_i, is used as an index into the tensor of variational parameters. The local occupancy x_i can therefore take one of four values depending on whether the orbital is unoccupied, singly occupied with a spin-up or spin-down electron, or doubly occupied with electrons of both spin types. The model generalizes a previous incarnation of the GPS, which considered a more rigid "squared-exponential" form of the kernel based on the Hamming distance metric between "classical" configurations, thereby supporting the model with fixed integer occupancies for each degree of freedom [18]. The form of Eq. (1) can be considered a completely flexible parametrization of such a kernel function, allowing for a fully variational identification and weighting of the dominant correlation features, and can be efficiently evaluated for arbitrary configurations of the state.

The model can also be viewed from the perspective of a tensor network state, being analogous to an exponentiated matrix product state (MPS) [5] for which the matrices are constrained to be diagonal. Equivalently, the state is a CANDECOMP/PARAFAC (CP) factorization [47,48] of the log of the wave function amplitude tensor. The

"diagonal" nature of the matrices makes the expressible amplitudes independent of the orbital ordering, and the act of exponentiation yields a combination of all possible products of terms contributing to the linear combination of product states in the exponential. This makes it possible to capture nontrivial entanglement and results in a product over the correlation features reminiscent of correlator product states [49,50], however, without any restriction on the ranks and ranges of the correlations which are described. We show in Ref. [20] how this state can also be represented by a neural network, with a specific architecture, exposing a duality between kernel and neural network approaches which has been previously explored in the ML community [51–53].

III. EFFICIENT FOCK SPACE VMC FOR *AB INITIO* FERMIONS

A. Electronic VMC in second quantization

In this paper, we aim to utilize the representative power of the GPS to describe the electronic ground state of molecular systems. We first review second-quantized VMC for *ab initio* fermions, particularly focusing on practical approaches for our specific context. The *ab initio* electronic structure Hamiltonian in the Born-Oppenheimer approximation can be expressed in a discretized basis of molecular spin-orbitals according to [54]

$$\hat{H} = \sum_{ij}^{2L} h_{ij}^{(1)}\, \hat{c}_i^\dagger \hat{c}_j + \frac{1}{2}\sum_{ijkl}^{2L} h_{ijkl}^{(2)}\, \hat{c}_i^\dagger \hat{c}_k^\dagger \hat{c}_l \hat{c}_j. \tag{2}$$

This Hamiltonian describes the interactions of the electrons occupying the various $2L$ degrees of freedom via the creation and annihilation operators $\hat{c}^\dagger$ and $\hat{c}$ satisfying fermionic commutation relations. The Hamiltonian matrix elements are defined via the one-electron integrals $h_{ij}^{(1)}$ capturing the single-particle contributions from their kinetic energy and interaction with the fixed external potential. The two-electron integrals describe instantaneous electron-electron interactions via the Coulomb interaction, defined as

$$h_{ijkl}^{(2)} = \int d\mathbf{r} \int d\mathbf{r}'\, \frac{\phi_i^*(\mathbf{r})\phi_j(\mathbf{r})\phi_k^*(\mathbf{r}')\phi_l(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{3}$$

which are evaluated with respect to the molecular orbital functions $\phi_i(\mathbf{r})$ defined across the real space. The spin-orbital labels i, j, k, l can be understood as compound indices indexing the two-dimensional spin degree of freedom, together with the spatial degree of freedom. Here, we work in a restricted basis of L spatial orbitals, which are the same irrespective of the spin component. In this paper, the molecular orbitals are obtained as contracted functions of the underlying linear combination of atomic orbitals as defined by various tabulated quantum chemical basis sets [55]. With the molecular orbitals defining the computational basis of the problem, we can use the GPS, or other general *Ansätze*, as the model mapping the orbital occupancies of an instantaneous electronic configuration to wave function amplitudes.

We evaluate the expectation via stochastic sampling. In particular, the variational energy of the state is computed as

$$E = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \left\langle \frac{[\hat{H}\Psi](x)}{\Psi(x)} \right\rangle_{p(x)\sim|\Psi(x)|^2}, \tag{4}$$

where the expectation value is approximated by drawing a finite number of samples according to the un-normalized probability distribution $|\langle x|\Psi\rangle|^2$. We generate samples with standard Markov chains utilizing the Metropolis-Hastings algorithm, in which moves are proposed based on arbitrarily ranged single electron hops of the configuration, also ensuring that configurations are taken from the correct particle number and spin magnetization sector. In second quantization, the local energy terms of Eq. (4) are formally evaluated as

$$\frac{[\hat{H}\Psi](x)}{\Psi(x)} = \frac{\sum_{x'} \hat{H}_{x,x'}\Psi(x')}{\Psi(x)}, \tag{5}$$

where $\hat{H}_{x,x'} = \langle x|\hat{H}|x'\rangle$ denotes a Hamiltonian matrix element. Due to the k-local nature of the Hamiltonian in Eq. (2), which contains at most quartic dependence on the fermionic operators, each local energy evaluation thus involves $O[L^4]$ terms. Excluding vanishing terms due to the particle-conservation symmetry for a fixed number of N electrons, the local energy evaluation then requires the evaluation of $O[N^2\times(2L-N)^2]$ wave function amplitudes [21].

The second-quantization approach allows for the incorporation of the antisymmetry directly into the many-body basis (and therefore operators expressed in the basis), as achieved via the commutation relations of the constituent operators in Eq. (2). The scaling above, however, contrasts with first-quantized representations of quantum states (where an explicitly antisymmetrized *Ansatz* for the state must be imposed). In these, the configurations directly represent an arrangement of the N electrons in real space, which avoids the variational approximation in second-quantized representations associated with the restriction to the fixed subspace spanned by the basis set. Due to the fact that the real-space first-quantized Hamiltonian acting on a single configuration only considers an analytically tractable semilocal term for the one-body operators, and a quadratically scaling electron-electron part depending on all electron pairs, the number of terms to consider in the evaluation of Eq. (5) scales only quadratically with the number of electrons, rather than $O[N^2\times(2L-N)^2]$ [26].

B. Basis choice

In the second-quantized representations, as utilized in this paper, there is a further freedom concerning the choice of representation of the molecular orbitals. The choice of the molecular orbitals is not unique as any unitary single-body rotation applied to the set of orbitals $\{\phi_i\}$ yields another valid representation. Such a change of the basis does not alter the physical observable characteristics of the (typically inaccessible) target state, whose expectation values are independent of the change. However, it will change the amplitudes of the wave function for each configuration in the chosen computational basis. Consequently, it likely also affects the accuracy of the *Ansatz* (away from the exact $M\to\infty$ limit) and the rate of convergence to this exact limit, as well as the efficiency by which the space can be sampled via a stochastic Markov

process. Furthermore, symmetries of the model are often easier to exploit in symmetry-preserving representations (where the single-particle basis transforms onto itself appropriately under action of the symmetry operations), constraining the choices available. With the ability to incorporate structure into the basis of the second- quantized formulation, making an appropriate choice is thus also a key contributing factor to the overall success of the method. While this choice can also be automatically tuned based on variational principles [56], here we discuss different conceptual paradigms to construct the molecular orbital basis.

A canonical choice for the molecular orbitals can easily be obtained by applying a self-consistent mean-field method such as Hartree-Fock (HF) [54], which finds an orthogonal set of molecular orbitals based on the defined atomic orbitals. By setting up the orbitals according to a mean-field calculation, the resulting basis representation automatically incorporates a large amount of physical information making this the standard basis choice for a large number of electronic structure methods. In particular, the HF wave function is then easily obtained as a single basis configuration in which the N energetically lowest orbitals are occupied.

As a consequence, target wave functions for relatively weakly correlated systems typically exhibit a very peaked probability structure, with dominant weights only for configurations which differ by few excitations from the HF configuration. While the sparsity of the wave function is a central cornerstone of the success of post-HF methods, such as coupled-cluster approaches, it can be expected to cause additional difficulties for a reliable model optimization in a VMC context. Indeed, it was noted in Ref. [42] that a key bottleneck in the application of neural network quantum states (NQSs) in second-quantized bases consists of difficulties with the VMC optimization of the state for approximations of a peaked target distribution. This resulted in significant numbers of configurational samples being required to achieve suitable exploration of the Hilbert space to achieve the full potential of the chosen neural network *Ansatz*. Though the general method overall only scales linearly in the number of samples, and it is easily parallelizable over the samples, the ability to optimize the *Ansatz* with as few samples as possible is crucial in order to scale the method up to larger systems.

While the canonical orbitals respect a natural ordering according to single-particle energies, the obtained orbital functions will typically be delocalized over the physical space. As an alternative to the canonical construction of the orbitals, in this paper, we consider orbital functions constructed to fulfill locality requirements, for which we expect some practical advantages outlined below. Different approaches have been proposed to construct an orthogonal set of localized orbital functions, $\{\phi_i\}$, commonly based on either a direct orthogonalization of the underlying atomic orbitals [57–59] or a numerical optimization of a locality measure [60–62]. Here, we consider the Boys localization scheme, an approach of the latter category. This constructs the orbitals via minimization of the orbital width, which can equivalently be formulated as a maximization of the quantity [63]

$$\mathcal{L} = \sum_i \left| \int d\mathbf{r}\, (\phi_i^*(\mathbf{r})\, \mathbf{r}\, \phi_i(\mathbf{r})) \right|^2 \tag{6}$$

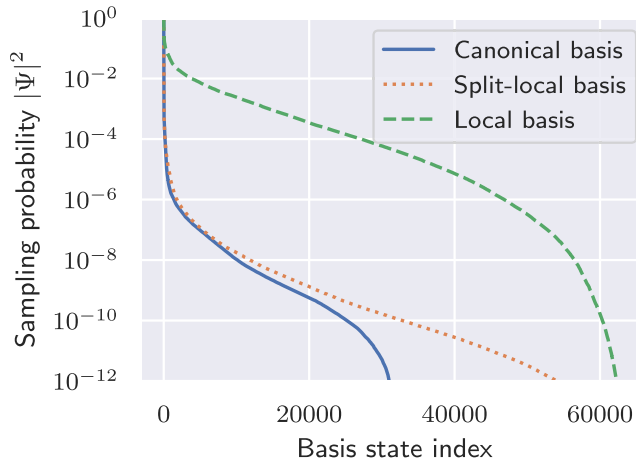

FIG. 1. Ground-state sample distribution of $|\Psi(\mathbf{x})|^2$, for a linear chain of ten hydrogen atoms with interatomic separation of $R = 1.8\ a_0$, described in a minimal basis set (STO-6G). The amplitudes associated with computational basis states are plotted in descending order from left to right with respect to different choices of the molecular orbital basis: the canonical basis from HF orbitals (blue solid curve), a split-localized basis (orange dotted curve), and a localized basis (green dashed curve), according to the Boys localization criteria. Sampling probabilities are rescaled to give a value of 1 for the most dominant configuration, and amplitudes smaller than 10^{-12} are not displayed.

while ensuring orthogonality of the orbital functions by only allowing optimization via unitary rotations of orthogonalized basis functions.

In a local basis, we generally expect that the approximated wave function amplitudes follow a broader distribution across the computational basis, in turn improving the ability to faithfully sample expectation values required for the optimization. This is exemplified for the ground-state wave function of a small one-dimensional system of ten hydrogen atoms represented in Fig. 1. The figure visualizes the sampling probability distribution across the computational basis for three different basis choices. In the canonical basis constructed from HF wave functions, a strongly peaked distribution becomes apparent. For this, only 176 basis configurations reach an occupational probability of more than 0.01% relative to that of the most strongly weighted configuration. This strongly contrasts with the structure emerging for a localized orbital choice, giving a probability distribution spreading across a significantly broader section of the Hilbert space. In the considered example, the localized basis gives a sampling distribution of the target state in which 27 164 basis configurations are sampled that are at least 0.01% as likely as the most probably sampled configuration. Lastly, we also present the distribution for a split-localized representation, in which the occupied and virtual orbitals are localized separately. This construction is a common choice within density matrix renormalization group (DMRG) calculations due to a reduction of the orbital entanglement within regimes in which both the single-particle effects and the local many-body interactions contribute significantly to this entanglement [64]. Furthermore, it still allows the mean-field state to be represented as a single configuration in the Hilbert space. For the

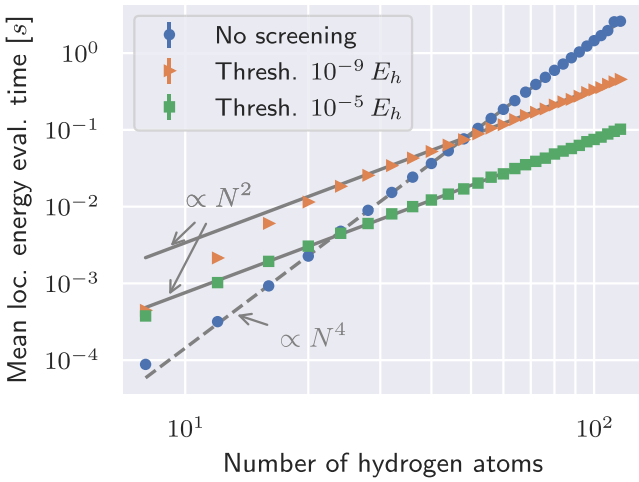

FIG. 2. Mean local energy evaluation time (mean loc. energy eval. time) as a function of the number of atoms in a one-dimensional hydrogen chain (fixed internuclear spacing of 1.8 a_0), described in a minimal atomic basis set (STO-6G). The local energy evaluations were performed with a uniform wave function model. The figure displays the mean evaluation time with an efficient full contraction over all allowed terms (blue circles), giving an asymptotic scaling of $O[N^4]$, as well as implementations with a prescreened selection of nonvanishing contributions with pruning thresholds 10^{-5} E_h (green squares) and 10^{-9} E_h (orange triangles), resulting in an asymptotic scaling of $O[N^2]$ in the chosen basis with localized orbitals. The calculations were performed on a single Intel Xeon Gold 6142 CPU core.

considered example, it still results in a sampling distribution of the target state which is concentrated around few configurations, and it is therefore not expected to help alleviate the sampling difficulties within VMC approaches.

In addition to changing the target distribution, the utilization of localized orbitals moreover paves the way for a reduction of the computational complexity of the local energy evaluation—typically the main computational bottleneck within practical implementations. If the molecular orbitals are sufficiently localized, many of the two-electron integrals $h^{(2)}_{i,j,k,l}$ as defined in Eq. (3) vanish for pairs of orbitals with large separation. By efficiently pruning the vanishing terms from the local energy evaluation [22–25], we can obtain an asymptotic reduction to $O[N^2]$ terms, thus in line with real-space formulations of the problem.

To highlight the scaling reduction which can be achieved in a local basis, we report the mean evaluation time of the local energy evaluation for linear chains of hydrogen atoms of variable lengths in Fig. 2. To efficiently prune vanishing terms in the local energy evaluation, we utilize a sparse data structure listing the elements of the electron integral tensors above a chosen threshold for all possible orbital indices (one-electron terms) or index pairs (two-electron terms). For each (single or double) electron annihilation in the application of the terms in Eq. (2), the summation can then be performed by only contracting over the nonvanishing elements. In the figure, we compare the mean evaluation time for a full contraction over all terms with the implementation exploiting the sparsity through the preselection of nonvanishing one- and two-electron integrals. The timings were obtained with a uniform state *Ansatz* and do therefore not take into account a scaling dependency from the evaluation complexity of chosen wave function *Ansätze* (potentially utilizing low-rank updates for the efficient evaluation), though this will not materially affect this leading-order scaling step.

While the utilization of the sparse data structure comes with an additional (constant) computational overhead associated with checking the validity of an electronic move, the plot clearly confirms the asymptotic scaling of $O[N^2]$ through the preselection of terms. This demonstrates a clear computational advantage when the pruning is applied in a local basis as the systems increase beyond a certain size. With an aggressive pruning threshold of 10^{-5} E_h, we observe an advantage as the chains increase beyond ≈25 hydrogen atoms, and for a smaller threshold of 10^{-9} E_h, the crossover point is obtained at ≈50 atoms. Although the specific timing details and crossover points will be system and implementation specific, given the nuclear separation of 1.8 a_0 chosen in our tests, this suggests that a computational advantage can generally be expected for the *ab initio* Coulomb interaction if any linear dimension of the system is of the order of ≈50 a_0 or greater. While the pruning of vanishing terms was implemented, in the following we consider system sizes which are not in this asymptotic limit, and therefore we still allow for a full contraction of the local energy contributions with moderate computational resources. However, though this thresholding is therefore not used, this reduction of the asymptotic scaling can nevertheless become a helpful tool for pushing the approach to larger systems which would otherwise be inaccessible.

C. Practical considerations of the GPS

Having set up the computational basis of the problem, we describe the quantum state as a GPS associating amplitudes with basis states according to Eq. (1). We optimize the parametrized *Ansatz* by minimization of the stochastically approximated energy expectation value with the standard stochastic reconfiguration method [6,65] until convergence of the variational energy is observed. For all the numerical tests discussed in the following, we used a moderate number of ≈10 000 samples for the approximation of expectation values. Our implementation is based on the NETKET software package [66,67], with additional functionality, including the GPS model definition and the implementation of the *ab initio* Hamiltonian, publicly available via the GPSKET plug-in library (with scripts to generate results in this paper included in the repository). To set up the molecular orbitals and obtain the one- and two-electron integrals, we utilized the PYSCF package [68,69].

A central element of the state approximation is the evaluation of the local energy for each sampled configuration, as defined in Eq. (5). This requires the evaluation of the Hamiltonian matrix elements, $\hat{H}_{x,x'}$, as well as the amplitudes, $\Psi(x')$, for all configurations connected via a nonzero Hamiltonian matrix element to $\Psi(x)$. As is presented in more detail in Appendix A, the evaluation of the contributing terms can be performed efficiently for the GPS by considering local updates to precomputed quantities. This allows us to evaluate the amplitudes of a GPS *Ansatz* for each connected

configuration, $\Psi(x')$, via local updates, with a complexity of $O[M]$, therefore independent of the system size. The evaluation of the Hamiltonian matrix elements involves a computation of a parity prefactor ensuring the fermionic commutation relations depending on a chosen normal ordering of the orbitals [70]. The computation of parity prefactors is equivalent to the evaluation of Jordan-Wigner type mappings from fermions to spin or qubit degrees of freedom [71], which we can also evaluate in constant time for each term in the Hamiltonian with appropriate setup. Overall, each local energy evaluation therefore scales as $O[M\times N^2\times(L-N)^2]$, not taking into account any pruning of (approximately) vanishing terms.

The canonical basis directly builds upon mean-field simulations, making it trivial to recover the HF level of accuracy with a wave function for which all but one amplitude vanishes—a distribution easily represented as a GPS. In the local basis, however, it is not immediately obvious how mean-field properties can be recovered with this *Ansatz*, and we often found it difficult to reliably reach the uncorrelated approximation in our simulations with a GPS. Furthermore, there are questions raised regarding the ordering of the fermionic degrees of freedom.

A specific ordering of the orbitals is required to define a normal order for the evaluation of parity prefactors in the Hamiltonian [70]. While we can define a natural choice to order the orbitals in a canonical basis via the single-particle energy level, ordering the orbitals becomes ambiguous and ill defined for all but one-dimensional systems represented in a local basis. However, we show in Appendix B that the effect of *all* possible orbital reorderings on the sign structure can be efficiently captured in the span of the GPS model by increasing the support dimension of the model polynomially in system size (quadratically). Put another way, any GPS model with support dimension M is able to have all possible observables reproduced under any fermionic orbital reordering, by a model with support dimension $M + O[L^2]$. This is a manifestation of the nonlocal correlation in the amplitudes (in this case their sign) that the GPS model can describe, allowing this changing sign structure from fermionic orbital reordering to be expressed in a polynomially compact fashion. However, the explicit construction suggests that, in general, we require a support dimension scaling quadratically with the number of orbitals to be able to span a state which is invariant to the choice of the ordering. This is still a relatively high scaling, which would reduce the efficiency of the method.

As a practical alternative, we can instead augment the GPS with an explicitly antisymmetric reference state, such as a single Slater determinant (SD), allowing the effect of orbital reorderings to be entirely subsumed within this reference, and avoiding the ambiguities of orbital ordering without requiring an explicit scaling of the support dimension with system size. This can replicate the success of such constructions for fermionic lattice models [11,18,72] and furthermore allows us to incorporate the properties of the uncorrelated physics without an increase in the support dimension, or significant impact on the overall computational cost [21]. This also ensures that we can rigorously describe the mean-field character of the state, using the GPS in a similar spirit to the Jastrow factor in standard Slater-Jastrow *Ansätze* [6]. However, by simultaneously optimizing the reference state and GPS, this construction does not limit the ability to (theoretically) approach exactness of the description by increasing the GPS support dimension [40].

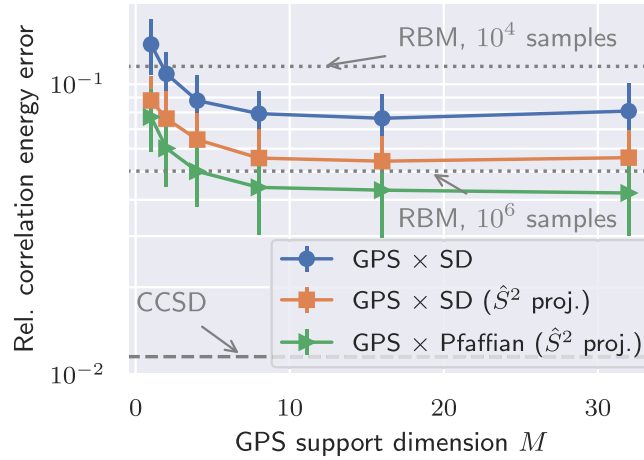

FIG. 3. Relative (Rel.) error in the correlation energy (with respect to the exact ground-state energy) obtained for a water molecule (6-31G basis, geometry as specified in Ref. [42]), in relation to the support dimension of the GPS which is augmented by different reference states. Co-optimized reference *Ansätze* include an unrestricted SD (blue circles), a spin-projected unrestricted SD (orange squares), and a spin-projected antiparallel Pfaffian *Ansatz* (green triangles). The error bars were computed from the standard deviation of the variational energy over the last 50 optimization steps, rescaled by the exact correlation energy. The figure includes reference values achieved with an RBM, optimized with 10^4 and 10^6 samples, from Ref. [42], as well as a CCSD value.

As a simple example, we replicate the description of a water molecule in a 6-31G basis set, as discussed in Ref. [42]. That work discusses the restricted Boltzmann machine (RBM) neural network architecture, comprising a single hidden layer of neurons, as the *Ansatz* for the state. The number of hidden nodes in the network serves as the model's main hyperparameter equivalent to the support dimension in the GPS, controlling its flexibility and computational cost, which is comparable between the two models. It was shown that the achieved accuracy strongly depends on the number of Monte Carlo samples, which was attributed to a particularly peaked sampling distribution of the target, indicating the use of a canonical basis. While it is one of the larger systems discussed in the *ab initio* study with RBM, the system still allows for a treatment with full configuration interaction techniques providing an exact baseline reference. Figure 3 shows results achieved with a variationally optimized GPS, augmented by a reference state optimized alongside the GPS, as a function of the support dimension. We present results for three different reference states: a single unrestricted SD, an $\hat{S}^2$-projected unrestricted SD, and a $\hat{S}^2$-projected antiparallel Pfaffian state [73,74]. While we were not able to achieve similar accuracies with a sole GPS, with the augmentation, our results mostly improve upon the accuracy reported for the RBM with only $\approx 10^4$ samples.

A single SD without explicit spin projection, together with the most simple GPS with a support dimension $M = 1$, gives a

corresponding relative correlation energy error of ≈14%. The error can be reduced further either by increasing the support dimension, thus adding further variational flexibility in the GPS part, or by allowing more flexibility in the reference state. At $M = 1$, the spin-projected SD decreases the relative energy error by approximately 5%, which is then further decreased by roughly another percent through utilization of the spin-projected Pfaffian. Although we see an improvement of the description for small support dimensions, the energetic values saturate for support dimensions larger than $M = 8$. With a full $\hat{S}^2$-projected Pfaffian reference state, the relative correlation energy error converges to a value of ≈4.2%, marginally improving upon the overall best reported RBM result of Ref. [42], which was obtained with the maximum considered number of 10^6 samples, with an RBM hidden unit density of one hidden neuron per spin-orbital. Nonetheless, the results do not match the accuracy obtained from coupled-cluster calculations based on single and double excitations (CCSD). The lack of further accuracy improvements with the GPS of larger support dimensions suggests that additional optimization difficulties limit the manifestation of the systematic improvability suggested by the *Ansatz* construction. We have confirmed that this limitation is, in fact, not caused by shortcomings in the stochastic approximation of expectation values, but also persists similarly if expectation values are evaluated from a contraction over the full computational basis without stochastic noise. The lack of systematic improvability can therefore be attributed either to fundamental limitations of the model in the considered limit or to difficulties with faithfully finding the optimal model parameters, often identified as a notoriously hard challenge for machine-learning-inspired *Ansätze*, which can suffer from restrictive parameter landscapes or generalization difficulties [75–77]. Although more than 90% of the correlation energy is captured in our results, further practical improvements to the algorithm are required to reach arbitrary accuracies to match the high level of accuracy obtained from coupled-cluster calculations for this system. However, whereas coupled-cluster approaches are particularly successful for systems exhibiting relatively weak degrees of electronic correlation, the GPS model does not particularly target this limit, and we expect a more general applicability of the model.

IV. TOWARDS EXTENDED HYDROGEN MATERIALS

To test the assertion of applicability in more strongly correlated *ab initio* systems, we benchmark the methodology for simple arrays of hydrogen atoms, described in a minimal basis set representation, already giving rise to rich quantum phenomena of condensed matter systems driven by strong electronic correlation. Conceptually, such hydrogen materials share a high degree of similarity to Fermi-Hubbard models, as well as the ability to change physical correlation regimes via changes in bond lengths. However, these hydrogen models are extended to general quartic electron interactions, as well as single-particle Hamiltonian terms, which range across the full system. They have therefore become a common testing ground for electronic structure methods [25,78–81], for which we benchmark the ability of the GPS *Ansatz* to capture the strong electronic correlation emerging in these systems as the interatomic spacing is increased, while ensuring that the interactions remain more faithful to the true Coulombic one.

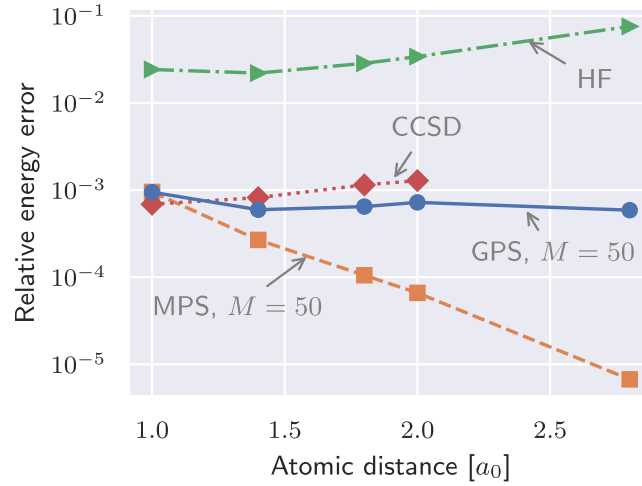

FIG. 4. Relative energy errors for linear hydrogen chains of 50 atoms at different atomic separations as obtained by different approaches. Data points include results from HF (green triangles), the GPS description as outlined in the main text (blue circles), CCSD (red diamonds), and DMRG calculations with fixed bond dimension of $M = 50$ (orange squares). Errors are evaluated by comparison to DMRG results converged with the MPS bond dimension. DMRG, HF, and CCSD results are taken from Ref. [25].

In Fig. 4 we describe obtained accuracies for a one-dimensional hydrogen chain comprising 50 atoms at different interatomic separations. The quasi-one-dimensionality of the system limits its entanglement and makes it possible for DMRG to provide essentially exact descriptions for these systems in a local basis representation, and we compare our results to DMRG results with a converged MPS bond dimension from Ref. [25]. The quasi-one-dimensional nature also avoids complicated nodal structures in the described wave function solely emerging due to fermionic ordering ambiguities, and we were able to achieve competitive accuracies solely with a GPS of practically manageable support dimension, which we chose as $M = L = 50$, not requiring the inclusion of a reference state to capture mean-field characteristics and avoid orbital ordering ambiguities. In addition to the GPS results, the figure also includes results obtained with HF, results obtained with CCSD (where calculations could be converged), and DMRG results with fixed MPS bond dimension of $M = 50$, all taken from Ref. [25].

The electronic correlation contributes significantly to the physical characteristics of the system. This manifests in an inability to reach reasonable accuracy with HF methods, giving relative energy errors greater than 1% for all considered separations, which goes up to ≈7.6% for an atomic separation of 2.8 a_0 as the correlations become more significant. The optimization of the GPS model within the VMC framework, on the other hand, consistently reaches an error of slightly less than 0.1% for all geometries. This level of accuracy is mostly in agreement with that achieved from converged CCSD calculations at equilibrium and weaker correlation regimes. Importantly, however, the GPS also reaches this accuracy for the largest considered separation where the CCSD calculation could not be converged, indicating a good consistency of the GPS across different physical regimes.

Interestingly, the relative energy error of the GPS matches that of an MPS with bond dimension equal to the GPS support dimension at 1 a_0, despite the significantly smaller number of variational parameters in the GPS. Nevertheless, the relative energy error from the MPS decreases as the geometry becomes more stretched, indicating a decay of entanglement rank between the orbitals. While the GPS does not follow this accuracy improvement, being able to reach a consistent level of accuracy across different correlation regimes is a good indication of the model's general ability to compress the target state efficiently.

While the MPS description explicitly builds on a one-dimensional structure to define an exactly contractible representation, the stochastically optimized GPS is not explicitly tailored for one-dimensional systems, also enabling efficient descriptions for higher-dimensional systems [18,20]. By augmenting the pure GPS with an explicitly antisymmetric reference state for fermionic systems, the description becomes entirely independent of the imposed fermionic orbital ordering. Furthermore, it allows us to capture mean-field effects efficiently, while allowing for systematic improvements of the variational flexibility. To highlight the ability of the augmented model to simulate systems for which no numerically exact methods are available, we study a cubic array comprising 4×4×4 hydrogen atoms and simultaneously stretch all bonds symmetrically, breaking all bonds simultaneously. This constitutes a larger system size than was considered in comparable NQS studies, and a careful implementation of the methodology allows us to present state-of-the-art benchmarks for this challenging system with long-range interactions and tuneable correlation strength. While it is still somewhat contrived (especially due to the limited basis size), it is an important first step towards realistic periodic and extended systems with this methodology. We report results achieved with a GPS of fixed support dimension, $M = 96$, optimized together with a SD of fixed magnetization, in Fig. 5. Based on our experience, we expect that such a support dimension of the order of the number of orbitals allows us to reach a high accuracy level beyond which diminishing returns are found with additional increase in the support dimension.

In the main panel of the figure, we show the energy per atom as the atomic separation is varied. The optimized GPS *Ansatz* predicts a similar equilibrium geometry as obtained from the HF baseline, also indicated in the figure, giving an energy minimum for an atomic spacing of about 1.5 Å. Crucially, however, we are able to achieve results significantly improving upon the HF level of accuracy, with an increasingly large discrepancy between the augmented GPS and the HF energies as the atomic separations get larger, resulting in a significant reduction of the harmonic frequency of the symmetric vibrations about equilibrium due to the correlation. With the ability to model local correlation properties with the GPS, we observe the expected asymptotic convergence of the energy as the cluster is dissociated, which cannot be captured based on mean-field considerations lacking the required charge fluctuations.

To compare the accuracy of the obtained energy values, the figure also includes the energy values obtained from a variational two-body reduced density matrix approach with approximate N-representability enforced via the "DQG" conditions, as discussed in Ref. [80]. While the VMC framework

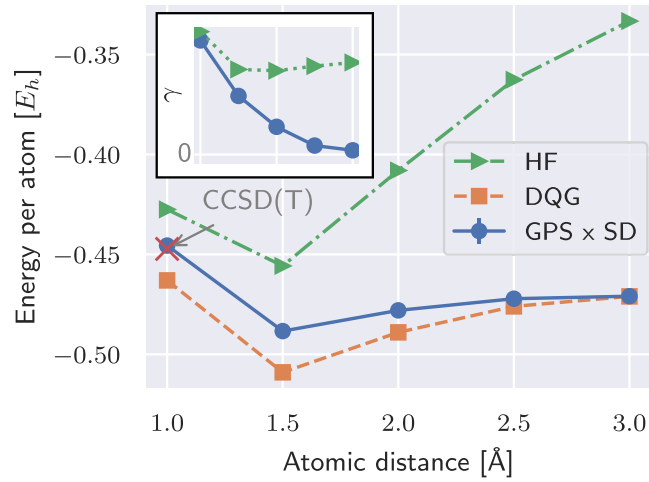

FIG. 5. Results for a cubic system of 4×4×4 hydrogen atoms (STO-6G atomic orbital basis) at different interatomic distances. The main panel shows energies per atom obtained with GPS ($M = 96$) multiplied by an unrestricted SD (blue circles), as well as HF results (green triangles), a single CCSD(T) value (red cross), and DQG results taken from Ref. [80] (orange squares). The inset shows the electronic mobility coefficient γ as defined in Eq. (7) from the VMC optimized state and the HF wave function.

always produces an upper bound to the exact ground-state energy, the DQG energies represent a lower bound to this value. Both methods give good agreement in the limit of large atomic separation, predicting an energy per atom of approximately $-0.471\ E_h$ at 3 Å spacing between the atoms, confirming this as an accurate approximation of the energy. For less stretched geometries, however, we obtain an increasing discrepancy between the two approaches. While it was not possible to converge coupled-cluster calculations for larger separations, at a distance of 1 Å, we obtained an energetic comparison value from CCSD with perturbative inclusion of triple excitations [CCSD(T)] of approximately $-0.447\ E_h$ per atom. Although this is marginally smaller than the one suggested by our VMC calculation, its agreement with our VMC value is significantly better than its agreement with the DQG result, which increases our confidence that we obtained a highly accurate approximation with the GPS.

Going beyond the evaluation of the energy expectation values, we also confirm that the system undergoes a phase transition from a metal to a Mott insulator as we increase the separation between hydrogen atoms. In keeping with the analysis of Ref. [80], we characterize this transition by quantifying the instantaneous electronic mobility from the coherences of the one-body reduced density matrix. More specifically, we evaluate the root mean square of its off-diagonal in a local (atomic) orbital basis, given by

$$\gamma = \sqrt{\frac{\sum_{a\neq b=1}^{2L}\sum_{i,j=1}^{2L} C_{a,i}\,C_{b,j}\,|\langle\Psi|\hat{c}_i^\dagger\hat{c}_j|\Psi\rangle|^2}{2L\times(2L-1)}}, \tag{7}$$

where the coefficients $C_{a,i}$ represent the change from the molecular basis with orbitals labeled by i to the atomic basis labeled by index a, and the expectation values $\langle\Psi|\hat{c}_i^\dagger\hat{c}_j|\Psi\rangle$ are again evaluated via stochastic sampling. We report the

electronic mobility coefficient for the dissociation of the cubic hydrogen material from our simulations in the inset of Fig. 5. We observe a decay of the electronic mobility to zero, not captured on the mean-field level of accuracy. Being able to predict this breakdown of instantaneous electron transfers induced through quantum many-body interactions underlines the applicability of the method to understand and describe quantum phenomena with realistic interactions driving technologically relevant material properties in which few, if any, reliable alternative approaches exist.

V. CONCLUSIONS AND OUTLOOK

With the unparalleled progress of artificial intelligence (AI) technology in recent years, the exploitation of dualities and synergies with quantum many-body problems will likely provide an increasingly important research area in the near future. This work provides further exemplification of how powerful ML frameworks can provide tools to extract quantum physical properties from fundamental principles. We have highlighted the general applicability of modern ML-inspired functional forms in a second-quantized VMC framework for chemical predictions from *ab initio* principles. In particular, we considered the choice of a local basis for alleviation of practical difficulties in the VMC optimization driven by stochastic sampling from the computational basis. While mean-field characteristics are implicitly incorporated in a canonical molecular orbital basis, if required, we incorporate these by inclusion of appropriate reference states within a local basis representation. We have exemplified the methodology utilizing the recently developed GPS *Ansatz*, corresponding to a particularly simple functional form inspired by kernel models and Bayesian regression, controlled by a single hyperparameter to describe the expressiveness in the constructed chosen computational basis. Providing a benchmark level of accuracy, and with a larger system size than was used in comparable approaches, we describe a metal-to-insulator transition driven by quantum correlations in a three-dimensional hydrogen material comprising 64 atoms. The description in a local basis, for which a natural sparsity of the Hamiltonian emerges, also provides natural extensions of the approach to enable larger-scale simulations.

Based on a variety of recent benchmarks for prototypical lattice models, we expect the presented results to be largely independent of the chosen ML-inspired functional form to define the *Ansatz*. The discussed framework is, amongst others, equivalently applicable to similarly motivated *Ansätze* constructed from neural network representations, such as the RBM, which we anticipate to reach comparable results. While the definite confirmation of this assumption requires further benchmarks, there is increasing evidence that results achievable with highly expressive variational functional forms in a VMC context are often limited by shortcomings of the optimization procedures rather than the model's expressivity [76,82]. Indeed, we were practically not able to observe a general improvability of the model to arbitrary target accuracies for fermionic systems. While the theoretical expressivity of the GPS can be improved by increasing its support dimension, even for a relatively simple testing system we observed a saturation to an accuracy limit beyond which no further improvements materialized. We therefore particularly consider further developments of techniques to overcome such limitations to be of major importance in order to extend the abilities of ML-inspired formalisms to study chemical properties from *ab initio* simulations in a second-quantized framework. This might be achieved through modifications to the optimization strategies [17,83–85] or by considering different paradigms to construct the electronic state, e.g., by following the construction of fully flexible, explicitly antisymmetrized representations [39,40] as these are commonly applied with great success in the real-space picture [27–37].

The code used in this paper is provided [86], with the input and scripts to generate all results available [87].

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the Air Force Office of Scientific Research under Award No. FA8655-22-1-7011, as well as the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 759063. We are grateful to the U.K. Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (Grants No. EP/P020194/1 and No. EP/T022213/1). Furthermore, we acknowledge the use of the high-performance computing environment CREATE at King's College London [88].

APPENDIX A: FAST EVALUATION OF LOCAL ENERGY TERMS FOR GPS

In order to evaluate the local energy for a configuration $|x\rangle$ from the Hamiltonian in Eq. (2), amplitudes arising from the model due to one-electron (two-electron) operators of the form

$$E_{i,j,(k,l)} = \frac{\langle x|\hat{c}_i^\dagger(\hat{c}_k^\dagger\hat{c}_l)\hat{c}_j|\Psi\rangle}{\langle x|\Psi\rangle} \tag{A1}$$

need to be evaluated. The indices i, j, k, l label different spin-orbitals, and the additional operators $\hat{c}_k^\dagger\hat{c}_l$ are only present for two-electron terms.

For a model associating amplitudes to basis states from the Fock basis, the evaluation of the terms generally involves two steps. Firstly, the connected Fock state, $|x'\rangle = P_{x,x'}\,\hat{c}_j^\dagger(\hat{c}_l^\dagger\hat{c}_k)\hat{c}_i\,|x\rangle$, needs to be identified. This involves the evaluation of a parity prefactor $P_{x,x'}$, emerging through the fermionic commutation relations. The value of $P_{x,x'}$ is zero if the final electronic occupancy in every degree of freedom does not agree with that of $|x\rangle$. Otherwise, it is ± 1, depending on the number of electrons which are passed by the (double) electron move according to the chosen ordering of orbitals [70]. The identification of the connected configuration and the evaluation of the parity prefactor are equivalent to the evaluation of Pauli operator strings in a Jordan-Wigner transformation from fermionic degrees of freedom to spin or qubit systems [71]. With the identification of a connected Fock state, the local energy contribution is given as

$$E_{i,j,(k,l)} = P_{x,x'}\frac{\langle x'|\Psi\rangle}{\langle x|\Psi\rangle} \tag{A2}$$

and therefore also requires the evaluation of the amplitude ratio $\frac{\langle x'|\Psi\rangle}{\langle x|\Psi\rangle}$ defined from the *Ansatz*.

Due to the large number of terms which contribute to the local energy, it is often helpful to consider more efficient local updates to evaluate each term. This commonly involves appropriate setup for the sample configuration $|x\rangle$ to enable a more efficient calculation for the connected configurations. By storing the cumulative electron counts in the different orbitals in the setup, the parity prefactor, directly obtained through counting the number of passed electrons, can be evaluated in constant time for each connected configuration.

Similar to the fast update of Slater-Jastrow models [21], we can also utilize update equations for the GPS *Ansätze* considered in this paper. These exploit the fact that each connected configuration $|x'\rangle$ gives an electronic occupancy which differs by at most the occupancy of four orbitals compared with the sampled configuration $|x\rangle$. To evaluate the corresponding amplitude update for the GPS as specified in Eq. (1) of the main text, its amplitude, evaluated for configuration $|x\rangle$, can be represented as

$$\Psi(x) = \exp\left(\sum_{\alpha=1}^{M} \varphi_\alpha(x)\right). \tag{A3}$$

Here, $\varphi_\alpha(x)$ denote un-normalized product state amplitudes defined via the variational parameters of the GPS, ϵ, as a product over the spatial orbitals according to $\varphi_\alpha(x) = \prod_{i=1}^{L} \epsilon_{\alpha,i,x_i}$. Using precomputed product state amplitudes, $\varphi_\alpha(x)$, the amplitude for the connected configuration evaluates to

$$\Psi(x') = \exp\left(\sum_{\alpha=1}^{M} \varphi_\alpha(x) \times \prod_{\bar{i}\in\{i,j,(k,l)\}} \left(\frac{\epsilon_{\alpha,\bar{i},x'_{\bar{i}}}}{\epsilon_{\alpha,\bar{i},x_{\bar{i}}}}\right)\right), \tag{A4}$$

where the set $\{i, j, (k, l)\}$ contains (at most) four indices, labeling the orbitals with changed occupancy. Using the precomputed amplitude of the central configuration, $\Psi(x)$, as well as the M product state amplitudes, $\varphi_\alpha(x)$, each term in the local energy can therefore be evaluated in $O[M]$ time. This scaling, independent of the number of orbitals, is generally unaffected by the inclusion of a mean-field-type reference state, for which similar low-rank updates can be performed [21].

APPENDIX B: ORBITAL REORDERING INVARIANCE OF GPS

While the span of representable amplitudes (including sign) is independent of the imposed ordering of the orbitals for a GPS, the quality of results (in the absence of a reference state) will still depend on this chosen ordering for fermionic systems. This is due to the fact that the sign structure of the modeled target amplitudes in the Fock basis changes under a change of the ordering, due to the different parity prefactors $P_{x,x'}$ in Eq. (A2). We show in this Appendix that we can construct a GPS with support dimension scaling at most quadratically with the system size which is able to express *all* changes induced in the sign structure of a fermionic state due to a different normal ordering.

For two different choices to define the normal ordering, basis states from the original computational basis can be related to basis states from a basis with changed orbital ordering via a sign transformation. Assuming a particular choice of molecular orbitals and letting $|x\rangle$ be the states from the associated computational basis, we express the basis states as

$$|x\rangle = c^\dagger_{r(0)} c^\dagger_{r(1)} \cdots c^\dagger_{r(N)} |0\rangle. \tag{B1}$$

Here, $r(i)$ labels the spin-orbital which is occupied by the electron with index i, and we choose the electron labels such that the indices satisfy $r(0) < r(1) < \cdots < r(N)$ (we make the choice of sorting the electrons such that all the spin-up electrons are followed by all the spin-down electrons). While there might be a natural choice of this ordering for some systems and molecular orbital choices (e.g., by energy level in the canonical basis or by position in space for one-dimensional systems in a local basis), ambiguities emerge in other cases, e.g., for larger basis sets, or local degrees of freedom in more than one dimension.

To compare the sign structure emerging through a reordering of the orbitals, we consider a relabeling of the orbitals to give a different set of Fock basis states,

$$|\tilde{x}\rangle = c^\dagger_{\tilde{r}(0)} c^\dagger_{\tilde{r}(1)} \cdots c^\dagger_{\tilde{r}(N)} |0\rangle. \tag{B2}$$

The electron labels are again chosen to fulfill a normal ordering, $\tilde{r}(0) < \tilde{r}(1) < \cdots < \tilde{r}(N)$, however, now defined with respect to a different linear order. This is specified by a permutation $\mathcal{P}$ defining a one-to-one mapping $\mathcal{P}(r) = \tilde{r}$ from an index r to a new index $\tilde{r}$ labeling the same physical orbital. Each new basis state can be associated with one from the original basis multiplied by a configuration-dependent sign according to

$$|x\rangle = (-1)^{\mathcal{N}_{x,\tilde{x}}} |\tilde{x}\rangle. \tag{B3}$$

The sign relating the two basis states is simply the parity of the permutation exchanging the creation operators in the definition of the basis states from one order into the other. It can be expressed as $(-1)^{\mathcal{N}_{x,\tilde{x}}}$, where $\mathcal{N}_{x,\tilde{x}}$ is the number of pairwise electron exchanges required to reassign the electron labels so that the list of orbital indices $(\tilde{r}(0), \tilde{r}(1), \ldots, \tilde{r}(N))$ satisfies the order $\mathcal{P}^{(-1)}(\tilde{r}(0)) < \mathcal{P}^{(-1)}(\tilde{r}(1)) < \cdots < \mathcal{P}^{(-1)}(\tilde{r}(N))$.

The additional sign structure, $(-1)^{\mathcal{N}_{x,\tilde{x}}}$, in the wave function amplitudes, solely emerging through a reordering of the orbitals, can be represented as a GPS with support dimension M scaling at most quadratically in the number of molecular orbitals. To show this, we decompose the number of pairwise exchanges, $\mathcal{N}_{x,\tilde{x}}$, as a sum over all pairwise next-element exchanges in the permutation $\mathcal{P}$, according to

$$\mathcal{N}_{x,\tilde{x}} = \sum_{\sigma\in\{\uparrow,\downarrow\}} \sum_{\{(a,b)\}} n_{a,\sigma}(x) n_{b,\sigma}(x). \tag{B4}$$

Here, the (a, b) sum runs over all index pairs which need to be exchanged so that the original list of orbital indices $(1, \ldots, L)$ is iteratively brought into the order $(\mathcal{P}(1), \ldots, \mathcal{P}(L))$ by only exchanging indices which are directly adjacent. The occupation number $n_{i,\sigma}(x)$ gives the number of electrons occupying spin channel σ of spatial orbital i in the many-electron configuration $|x\rangle$, thus either evaluating to 1 or zero. If (and only if) the configuration $|x\rangle$ has an electron with the same spin in both of the orbitals labeled by a and b, then two creation operators in the construction of Eq. (B1) are exchanged, resulting in

an additional (-1) prefactor from the commutation relations. Iteratively applying the pairwise exchanges of adjacent creation operators according to the permutation of orbitals $\mathcal{P}$ then yields the desired representation together with the induced sign transformation.

The representation of the sign transformation as a GPS follows directly from the representation of $\mathcal{N}_{x,\tilde{x}}$ according to Eq. (B4). Specifically, we can represent the sign structure $(-1)^{\mathcal{N}_{x,\tilde{x}}}$ as a GPS by associating each support point index, α, with a term from Eq. (B4), i.e., a unique pair of orbital indices from the set $\{(a,b)\}$ and associated spin value $\sigma \in \{\uparrow,\downarrow\}$. Based on the definition of the GPS amplitudes according to Eq. (1), the representation is, e.g., obtained with the parameter choice

$$\begin{aligned} \epsilon_{\alpha,a,\sigma} = \epsilon_{\alpha,a,\uparrow\downarrow} &= i\pi, \\ \epsilon_{\alpha,b,\sigma} = \epsilon_{\alpha,b,\uparrow\downarrow} &= 1, \\ \epsilon_{\alpha,a,\bar{\sigma}} = \epsilon_{\alpha,b,\bar{\sigma}} = \epsilon_{\alpha,a,\cdot} = \epsilon_{\alpha,b,\cdot} &= 0, \\ \epsilon_{\alpha,i\notin\{(a,b)\},l} &= 1, \end{aligned}$$

where $\bar{\sigma}$ denotes the inversion of spin σ and index l runs over all possible local occupancies, $l \in \{\uparrow,\downarrow,\uparrow\downarrow,\cdot\}$.

With this construction, the GPS representation of the sign structure relating two different orbital orderings therefore requires a support dimension of $M = 2|\{(a,b)\}|$, where $|\{(a,b)\}|$ corresponds to the number of pairwise index exchanges in the permutation of indices specified by $\mathcal{P}$. Any such permutation comprises at most $O(L^2)$ pairwise exchanges, consequently limiting the required support dimension (and therefore the total number of variational parameters) of the constructed GPS to scale at most quadratically with the number of orbitals. This also means that we can always define a GPS with a support dimension M increased by at most $O(L^2)$ so that it exactly spans any other GPS with support dimension M with a changed ordering of the molecular orbitals in the definition of the basis states. Though the span of a GPS with given support dimension is therefore not fully invariant under changes to the orbital ordering, it always contains a subset of states which can be represented independent of the orbital ordering if the support dimension is scaled quadratically with the system size. Nonetheless, an exact invariance of the state, independent of the support dimension, is also obtained through the inclusion of an explicitly antisymmetrized reference state acting in a first-quantized picture, such as a single Slater determinant.

[1] R. B. Laughlin, Anomalous Quantum Hall Effect: An Incompressible Quantum Fluid with Fractionally Charged Excitations, Phys. Rev. Lett. **50**, 1395 (1983).

[2] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Theory of superconductivity, Phys. Rev. **108**, 1175 (1957).

[3] M. C. Gutzwiller, Effect of Correlation on the Ferromagnetism of Transition Metals, Phys. Rev. Lett. **10**, 159 (1963).

[4] R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, Rev. Mod. Phys. **79**, 291 (2007).

[5] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, Ann. Phys. (Amsterdam) **349**, 117 (2014).

[6] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).

[7] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[8] G. Carleo, Y. Nomura, and M. Imada, Constructing exact representations of quantum many-body systems with deep neural networks, Nat. Commun. **9**, 5322 (2018).

[9] D.-L. Deng, X. Li, and S. Das Sarma, Quantum Entanglement in Neural Network States, Phys. Rev. X **7**, 021021 (2017).

[10] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 662 (2017).

[11] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted Boltzmann machine learning for solving strongly correlated quantum systems, Phys. Rev. B **96**, 205152 (2017).

[12] Y. Huang and J. E. Moore, Neural Network Representation of Tensor Network and Chiral States, Phys. Rev. Lett. **127**, 170601 (2021).

[13] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Neural-Network Quantum States, String-Bond States, and Chiral Topological States, Phys. Rev. X **8**, 011006 (2018).

[14] Z.-A. Jia, B. Yi, R. Zhai, Y.-C. Wu, G.-C. Guo, and G.-P. Guo, Quantum neural network states: A brief review of methods and applications, Adv. Quantum Technol. **2**, 1800077 (2019).

[15] R. Jastrow, Many-body problem with strong forces, Phys. Rev. **98**, 1479 (1955).

[16] Y. Nomura and M. Imada, Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy, Phys. Rev. X **11**, 031034 (2021).

[17] C. Roth, A. Szabó, and A. MacDonald, High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks, arXiv:2211.07749.

[18] A. Glielmo, Y. Rath, G. Csányi, A. De Vita, and G. H. Booth, Gaussian Process States: A Data-Driven Representation of Quantum Many-Body Physics, Phys. Rev. X **10**, 041026 (2020).

[19] Y. Rath, A. Glielmo, and G. H. Booth, A Bayesian inference framework for compression and prediction of quantum states, J. Chem. Phys. **153**, 124108 (2020).

[20] Y. Rath and G. H. Booth, The quantum Gaussian process state: A kernel-inspired state with quantum support data, Phys. Rev. Res. **4**, 023126 (2022).

[21] E. Neuscamman, The Jastrow antisymmetric geminal power in Hilbert space: Theory, benchmarking, and application to a novel transition state, J. Chem. Phys. **139**, 194105 (2013).

[22] A. Mahajan and S. Sharma, Efficient local energy evaluation for multi-Slater wave functions in orbital space quantum Monte Carlo, J. Chem. Phys. **153**, 194108 (2020).

[23] H. Wei and E. Neuscamman, Reduced scaling Hilbert space variational Monte Carlo, J. Chem. Phys. **149**, 184106 (2018).

[24] I. Sabzevari and S. Sharma, Improved speed and scaling in orbital space variational Monte Carlo, J. Chem. Theory Comput. **14**, 6276 (2018).
[25] J. Hachmann, W. Cardoen, and G. K.-L. Chan, Multireference correlation in long molecules with the quadratic scaling density matrix renormalization group, J. Chem. Phys. **125**, 144101 (2006).
[26] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).
[27] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).
[28] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, Nat. Chem. **12**, 891 (2020).
[29] J. S. Spencer, D. Pfau, A. Botev, and W. M. C. Foulkes, Better, faster fermionic neural networks, arXiv:2011.07125.
[30] L. Gerard, M. Scherbela, P. Marquet, and P. Grohs, Gold-standard solutions to the Schrödinger equation using deep learning: How much physics do we need? in *Advances in Neural Information Processing Systems*, edited by A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho (2022).
[31] J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, G. Carleo, and F. Noé, Ab-initio quantum chemistry with neural-network wavefunctions, arXiv:2208.12590.
[32] X. Li, Z. Li, and J. Chen, Ab initio calculation of real solids via neural network ansatz, Nat. Commun. **13**, 7895 (2022).
[33] I. von Glehn, J. S. Spencer, and D. Pfau, A self-attention ansatz for ab-initio quantum chemistry, in *The Eleventh International Conference on Learning Representations* (2023).
[34] J. Han, L. Zhang, and W. E, Solving many-electron Schrödinger equation using deep neural networks, J. Comput. Phys. **399**, 108929 (2019).
[35] N. Gao and S. Günnemann, Generalizing neural wave functions, arXiv:2302.04168.
[36] N. Gao and S. Günnemann, *Ab-initio* potential energy surfaces by pairing GNNs with neural wave functions, in *International Conference on Learning Representations* (2022).
[37] N. Gao and S. Günnemann, Sampling-free inference for *ab-initio* potential energy surface networks, in *The Eleventh International Conference on Learning Representations* (2023).
[38] P. Lopez Rios, Backflow and pairing wave function for quantum Monte Carlo methods, Ph.D. thesis, University of Cambridge, 2016.
[39] D. Luo and B. K. Clark, Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions, Phys. Rev. Lett. **122**, 226401 (2019).
[40] J. R. Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, Proc. Natl. Acad. Sci. USA **119**, e2205789119 (2022).
[41] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum Entanglement in Deep Learning Architectures, Phys. Rev. Lett. **122**, 065301 (2019).
[42] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for *ab-initio* electronic structure, Nat. Commun. **11**, 2368 (2020).
[43] T. D. Barrett, A. Malyshev, and A. I. Lvovsky, Autoregressive neural-network wavefunctions for *ab initio* quantum chemistry, Nat. Mach. Intell. **4**, 351 (2022).
[44] T. Zhao, J. Stokes, and S. Veerapaneni, Scalable neural quantum states architecture for quantum chemistry, arXiv:2208.05637.
[45] Y. Wu, X. Xu, D. Poletti, Y. Fan, C. Guo, and H. Shang, A real neural network state for quantum chemistry, arXiv:2301.03755.
[46] P.-J. Yang, M. Sugiyama, K. Tsuda, and T. Yanai, Artificial neural networks applied as molecular wave function solvers, J. Chem. Theory Comput. **16**, 3513 (2020).
[47] H. A. L. Kiers, Towards a standardized notation and terminology in multiway analysis, J. Chemom. **14**, 105 (2000).
[48] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, SIAM Rev. **51**, 455 (2009).
[49] H. J. Changlani, J. M. Kinder, C. J. Umrigar, and G. K.-L. Chan, Approximating strongly correlated wave functions with correlator product states, Phys. Rev. B **80**, 245116 (2009).
[50] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, Ground-state properties of quantum many-body systems: Entangled-plaquette states and variational Monte Carlo, New J. Phys. **11**, 083026 (2009).
[51] R. M. Neal, Bayesian learning for neural networks, Ph.D. thesis, University of Toronto, 1995.
[52] J. Lee, J. Sohl-Dickstein, J. Pennington, R. Novak, S. Schoenholz, and Y. Bahri, Deep neural networks as Gaussian processes, in *International Conference on Learning Representations 2018* (ICLR, Appleton, WI, 2018).
[53] A. G. de G. Matthews, J. Hron, M. Rowland, R. E. Turner, and Z. Ghahramani, Gaussian process behaviour in wide deep neural networks, in *International Conference on Learning Representations 2018* (ICLR, Appleton, WI, 2018).
[54] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry Introduction to Advanced Electronic Structure Theory* (Dover, Mineola, NY, 2012).
[55] F. Jensen, Atomic orbital basis sets, WIREs Comput. Mol. Sci. **3**, 273 (2013).
[56] J. R. Moreno, J. Cohn, D. Sels, and M. Motta, Enhancing the expressivity of variational neural, and hardware-efficient quantum states through orbital rotations, arXiv:2302.11588.
[57] P.-O. Löwdin, On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals, J. Chem. Phys. **18**, 365 (1950).
[58] A. E. Reed, R. B. Weinstock, and F. Weinhold, Natural population analysis, J. Chem. Phys. **83**, 735 (1985).
[59] F. Aquilante, T. B. Pedersen, A. S. de Merás, and H. Koch, Fast noniterative orbital localization for large molecules, J. Chem. Phys. **125**, 174101 (2006).
[60] J. M. Foster and S. F. Boys, Canonical configurational interaction procedure, Rev. Mod. Phys. **32**, 300 (1960).
[61] C. Edmiston and K. Ruedenberg, Localized atomic and molecular orbitals, Rev. Mod. Phys. **35**, 457 (1963).
[62] J. Pipek and P. G. Mezey, A fast intrinsic localization procedure applicable for *ab initio* and semiempirical linear combination of atomic orbital wave functions, J. Chem. Phys. **90**, 4916 (1989).
[63] D. A. Kleier, T. A. Halgren, J. H. Hall, and W. N. Lipscomb, Localized molecular orbitals for polyatomic molecules. I. A comparison of the Edmiston-Ruedenberg and Boys localization methods, J. Chem. Phys. **61**, 3905 (1974).

[64] R. Olivares-Amaya, W. Hu, N. Nakatani, S. Sharma, J. Yang, and G. K.-L. Chan, The *ab-initio* density matrix renormalization group in practice, J. Chem. Phys. **142**, 034102 (2015).

[65] S. Sorella, Generalized Lanczos algorithm for variational quantum Monte Carlo, Phys. Rev. B **64**, 024512 (2001).

[66] G. Carleo, K. Choo, D. Hofmann, J. E. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, NetKet: A machine learning toolkit for many-body quantum systems, SoftwareX **10**, 100311 (2019).

[67] 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 **7** (2022).

[68] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, PySCF: The Python-based simulations of chemistry framework, WIREs Comput. Mol. Sci. **8**, e1340 (2018).

[69] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu *et al.*, Recent developments in the PySCF program package, J. Chem. Phys. **153**, 024109 (2020).

[70] A. Altland and B. D. Simons, *Condensed Matter Field Theory* (Cambridge University Press, Cambridge, 2006).

[71] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, A comparison of the Bravyi–Kitaev and Jordan–Wigner transformations for the quantum simulation of quantum chemistry, J. Chem. Theory Comput. **14**, 5617 (2018).

[72] S. Humeniuk, Y. Wan, and L. Wang, Autoregressive neural Slater-Jastrow ansatz for variational Monte Carlo simulation, arXiv:2210.05871.

[73] M. Bajdich, L. Mitas, G. Drobný, L. K. Wagner, and K. E. Schmidt, Pfaffian Pairing Wave Functions in Electronic-Structure Quantum Monte Carlo Simulations, Phys. Rev. Lett. **96**, 130201 (2006).

[74] T. Misawa, S. Morita, K. Yoshimi, M. Kawamura, Y. Motoyama, K. Ido, T. Ohgoe, M. Imada, and T. Kato, mVMC—Open-source software for many-variable variational Monte Carlo method, Comput. Phys. Commun. **235**, 447 (2019).

[75] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, Phys. Rev. Res. **2**, 033075 (2020).

[76] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, SciPost Phys. **10**, 147 (2021).

[77] J. T. Frank and M. J. Kastoryano, Learning neural network quantum states with the linear method, arXiv:2104.11011.

[78] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, and S. Zhang (Simons Collaboration on the Many-Electron Problem), Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods, Phys. Rev. X **7**, 031059 (2017).

[79] L. Stella, C. Attaccalite, S. Sorella, and A. Rubio, Strong electronic correlation in the hydrogen chain: A variational Monte Carlo study, Phys. Rev. B **84**, 245117 (2011).

[80] A. V. Sinitskiy, L. Greenman, and D. A. Mazziotti, Strong correlation in hydrogen chains and lattices using the variational two-electron reduced density matrix method, J. Chem. Phys. **133**, 014104 (2010).

[81] T. Tsuchimochi and G. E. Scuseria, Strong correlations via constrained-pairing mean-field theory, J. Chem. Phys. **131**, 121102 (2009).

[82] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Generalization properties of neural network approximations to frustrated magnet ground states, Nat. Commun. **11**, 1593 (2020).

[83] D. Kochkov and B. K. Clark, Variational optimization in the AI era: Computational graph states and supervised wave-function optimization, arXiv:1811.12423.

[84] A. Lovato, C. Adams, G. Carleo, and N. Rocco, Hidden-nucleons neural-network quantum states for the nuclear many-body problem, Phys. Rev. Res. **4**, 043178 (2022).

[85] A. Chen and M. Heyl, Efficient optimization of deep neural quantum states toward machine precision, arXiv:2302.01941.

[86] https://github.com/BoothGroup/GPSKet.

[87] https://github.com/BoothGroup/GPSKet/tree/master/scripts/GPS_for_ab_initio.

[88] King's College London, King's Computational Research, Engineering and Technology Environment (CREATE), Retrieved January 13, 2023, from https://doi.org/10.18742/rnvf-m076.