



COLLÈGE
DE FRANCE
— 1530 —

Chaire de Physique
de la Matière Condensée
Antoine Georges

Réseaux de Neurons, Apprentissage et Physique Quantique

ρ_1
⋮
 ρ_N

Cours 5:

Représentations d'états quantiques fermioniques
utilisant les réseaux de neurones (II)

Ψ_1
 Ψ_2
 Ψ_3
 Ψ_4
 Ψ_5
 Ψ_6
 Ψ_7
 Ψ_8
⋮
 Ψ_{2N}

Cycle 2022-2023
30 mai 2023



COLLÈGE
DE FRANCE
— 1530 —

Chaire de Physique
de la Matière Condensée
Antoine Georges

Machine Learning and Neural Networks for Quantum Physics

ρ_1
⋮
 ρ_N

Lecture 4

Fermionic Quantum States with Neural Networks (II)

Ψ_1
 Ψ_2
 Ψ_3
 Ψ_4
 Ψ_5
 Ψ_6
 Ψ_7
 Ψ_8
⋮
 Ψ_{2N}

Cycle 2022-2023
May 23, 2023

Seminar (11h30): Giulio Biroli Renormalisation group and machine learning: the wavelet-conditional RG



Renormalisation group and machine learning: the Wavelet-Conditional RG

Giulio Biroli
ENS, Paris

Reconstructing, or generating, high dimensional distributions starting from data is a central problem in machine learning and data sciences. I will present a method –The Wavelet Conditional Renormalization Group –that combines ideas from physics (renormalization group theory) and computer science (wavelets, stable representations of operators). The Wavelet Conditional Renormalization Group allows to reconstruct in a very efficient way classes of high dimensional probability distributions hierarchically from large to small spatial scales, and to perform RG directly from data. It allows to bridge the gap between approaches based on physical intuition and modern machine learning algorithms. I will present the method and then show its applications to data from statistical physics and cosmology. I shall also discuss the interesting insights that our method offers on the interplay between structures of data and architectures of deep neural networks.

Mailing List

(Weekly announcement of lecture and seminar, etc.)

Send email to: listes-diffusion.cdf@college-de-france.fr

Subject line: subscribe chaire-pmc.ipcdf

...or: unsubscribe chaire-pmc.ipcdf

You can also just send me an email to be placed on the list

Website:

<https://www.college-de-france.fr/site/antoine-georges/index.htm>

Lectures are video recorded
and available on the website

In the previous lecture:

- The many fermion problem: introduction, notations
- Slater Determinants
- Hartree-Fock
- Some classic variational wave functions: Slater determinant, BCS/pairs, Gutzwiller, Jastrow-Slater, Backflow, ...
- Optimizing wave functions with Monte Carlo

Today's menu:

- Variational wave functions (cont'd): Jastrow, Backflow
- Fermionic Neural Wave Functions (determinantal or not)
- Some examples – with a particular focus on Hidden Fermion Determinantal States (HFDS)
- Some applications

Jastrow-Slater Wave Function

Continuous space: Slater determinant with (possibly long-range) symmetric weight:

$$e^{\frac{1}{2} \sum_{ij} u(r_i - r_j)} \psi_{\text{SD}}(r_1, \dots, r_N)$$

General form also applicable to lattice:

$$|\Psi_{\text{JS}}\rangle = e^{\frac{1}{2} \sum_{pq} g_{pq} \hat{n}_p \hat{n}_q} |\Psi_{\text{SD}}\rangle$$
$$(1 \leq p, q \leq M)$$

Variational Description of Mott Insulators

Manuela Capello,^{1,2} Federico Becca,^{1,2} Michele Fabrizio,^{1,2,3} Sandro Sorella,^{1,2} and Erio Tosatti^{1,2,3}

¹International School for Advanced Studies (SISSA), Via Beirut 2-4, I-34014 Trieste, Italy

²INFM-Democritos National Simulation Centre, Trieste, Italy

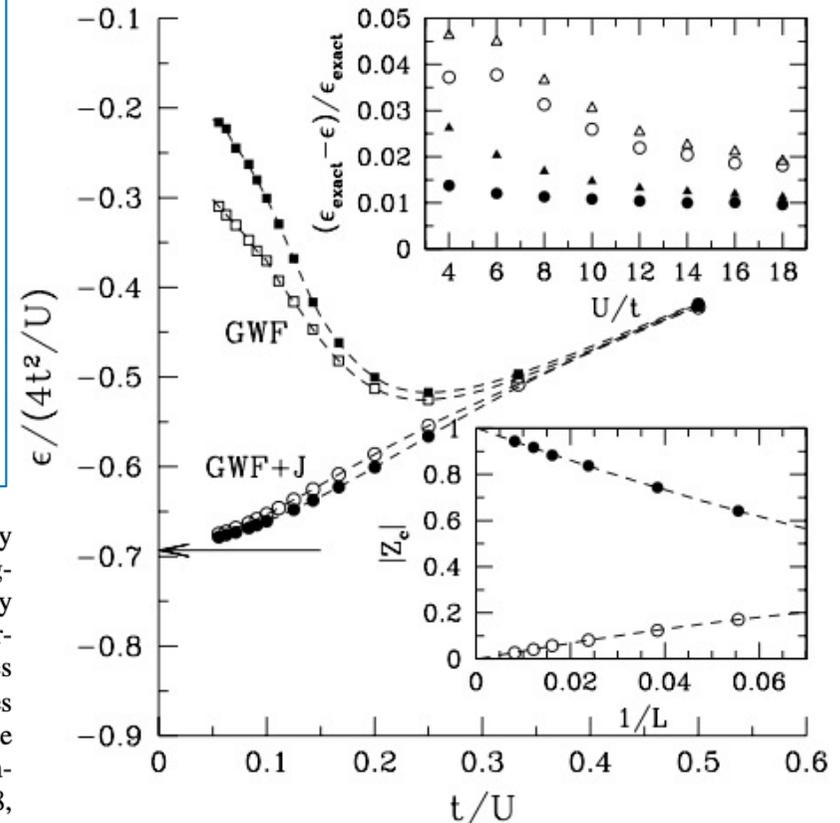
³International Centre for Theoretical Physics (ICTP), P.O. Box 586, I-34014 Trieste, Italy

(Received 17 March 2004; published 20 January 2005)

The Gutzwiller wave function for a strongly correlated model can, if supplemented with a long-range Jastrow factor, provide a proper variational description of Mott insulators, so far unavailable. We demonstrate this concept in the prototypical one-dimensional $t - t'$ Hubbard model, where at half-filling we reproduce all known phases, namely, the ordinary Mott undimerized insulator with power-law spin correlations at small t'/t , the spin-gapped metal above a critical t'/t and small U , and the dimerized Mott insulator at large repulsion.

- JSWF applied to Hubbard model in $d=1$
- In contrast to GWF, able to reproduce Mott insulator at $\frac{1}{2}$ -filling (with no symmetry breaking)
- Long-range Jastrow crucial
- $d=2$: JSWF is not enough see: Capello et al PRB 73, 245116 (2006)

FIG. 2. Energy per site ϵ for the simple GWF ($L = 18$, empty squares, and $L = 82$, full squares) and for the WF with long-range density-density Jastrow (GWF + J) ($L = 18$, empty circles, and $L = 82$, full circles), for $t' = 0$. The Slater determinant is a simple Fermi sea, with $\Delta_q = 0$. The arrow indicates the exact energy per site of the Heisenberg model, and the lines are guides to the eye. Top inset: accuracy of the WF with all the density-density Jastrow independently minimized (same symbols as before) and with the analytic parametrization ($L = 18$, empty triangles, and $L = 82$, full triangles). Bottom inset: Z_c for the GWF + J (full circles) and for the GWF (open circles); lines are three parameter fits.



'Backflow'

- General idea: allow the 1-particle orbitals entering the Slater determinant
- to depend on the coordinates on all other particles.
- First proposed in the continuum (Helium 4) by Feynman and Cohen
Phys Rev 102, 1189 (1956)
- Applied to the lattice only in 2008 and onwards:
Tocchio et al. PRB 78, 041101 (2008); PRB 83, 185138 (2011);
PRB 94, 195126 (2016)

$$\bar{\mathbf{r}}_1 = b(\mathbf{r}_1; \mathbf{r}_2, \dots, \mathbf{r}_N) , \text{ etc.}$$

$$\psi_{\text{SD}}(r_1, \dots, r_N) = \det [\phi_{\alpha_i}(\bar{r}_j)]$$

For example: $\phi_{\alpha}(\bar{r}_i) = \phi_{\alpha}(r_i) + \sum_j c_{ij}[x] \phi_{\alpha}(r_j)$

More generally:

$$\det \begin{bmatrix} \Phi_{x_1,1}[n(x)] & \dots & \Phi_{x_1,N}[n(x)] \\ \vdots & \ddots & \vdots \\ \Phi_{x_N,1}[n(x)] & \dots & \Phi_{x_N,N}[n(x)] \end{bmatrix}$$

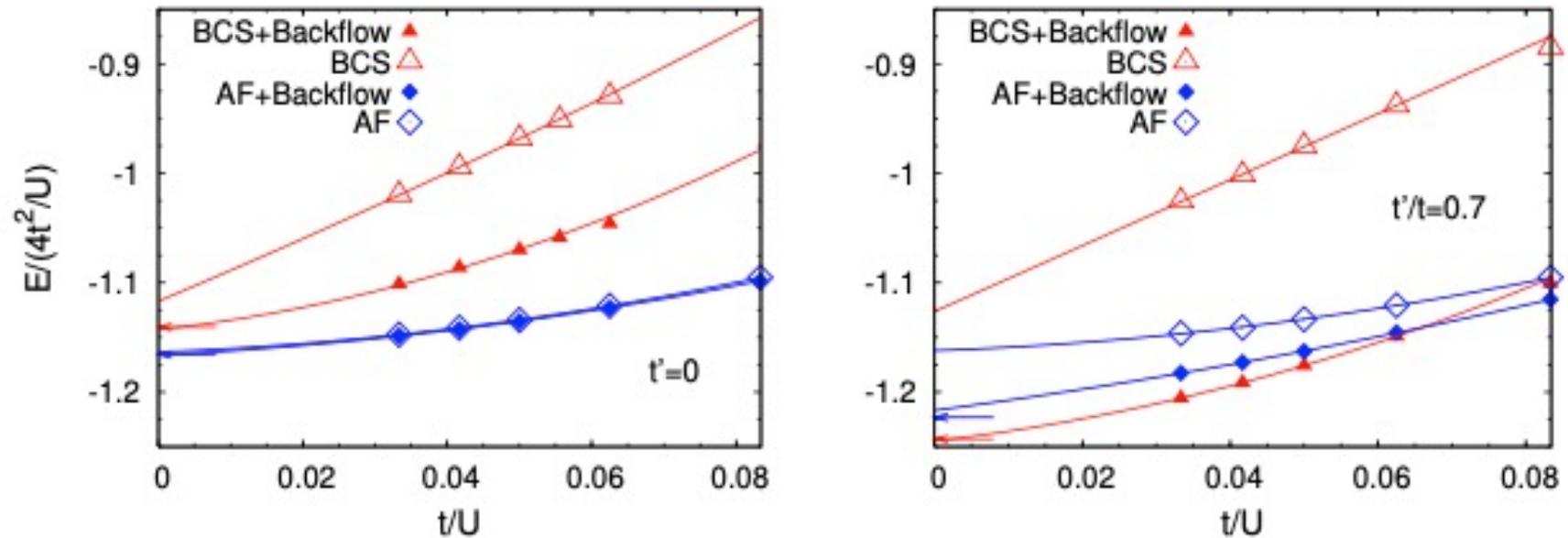


Fig. 2: Energies per site (in units of $J = 4t^2/U$) for the two-dimensional Hubbard model at half filling, for both the unfrustrated ($t' = 0$) and frustrated ($t'/t = 0.7$) case. The cases with and without backflow correlations are reported (for the BCS state). The results for the wave function with antiferromagnetic order and no BCS pairing are also shown. Arrows indicate the energies per site for the corresponding fully-projected states in the Heisenberg model.

Hidden Mott transition and large- U superconductivity in the two-dimensional Hubbard model

Luca F. Tocchio, Federico Becca, and Sandro Sorella

*CNR-IOM-Democritos National Simulation Centre and International School for Advanced Studies (SISSA),
Via Bonomea 265, I-34136, Trieste, Italy*

(Received 29 July 2016; revised manuscript received 26 September 2016; published 14 November 2016)

We consider the one-band Hubbard model on the square lattice by using variational and Green's function Monte Carlo methods, where the variational states contain Jastrow and backflow correlations on top of an uncorrelated wave function that includes BCS pairing and magnetic order. At half-filling, where the ground state is antiferromagnetically ordered for any value of the on-site interaction U , we can identify a hidden critical point U_{Mott} , above which a finite BCS pairing is stabilized in the wave function. The existence of this point is reminiscent of the Mott transition in the paramagnetic sector and determines a separation between a Slater insulator (at small values of U), where magnetism induces a potential energy gain, and a Mott insulator (at large values of U), where magnetic correlations drive a kinetic energy gain. Most importantly, the existence of U_{Mott} has crucial consequences when doping the system: We observe a tendency for phase separation into hole-rich and hole-poor regions only when doping the Slater insulator, while the system is uniform by doping the Mott insulator. Superconducting correlations are clearly observed above U_{Mott} , leading to the characteristic dome structure in doping. Furthermore, we show that the energy gain due to the presence of a finite BCS pairing above U_{Mott} shifts from the potential to the kinetic sector by increasing the value of the Coulomb repulsion.

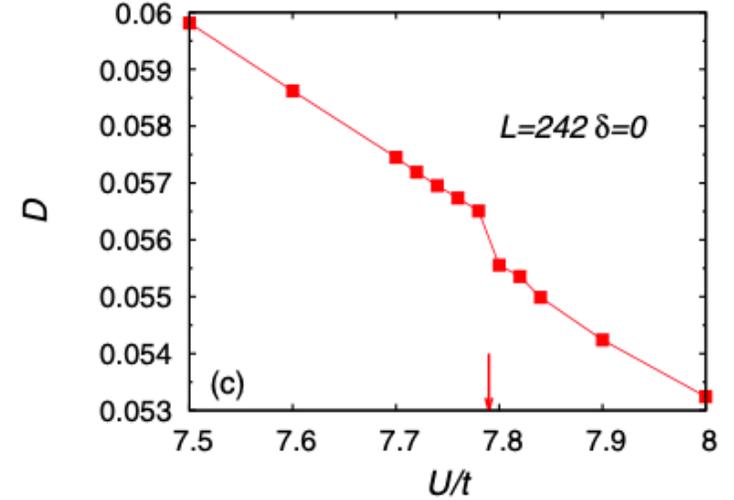
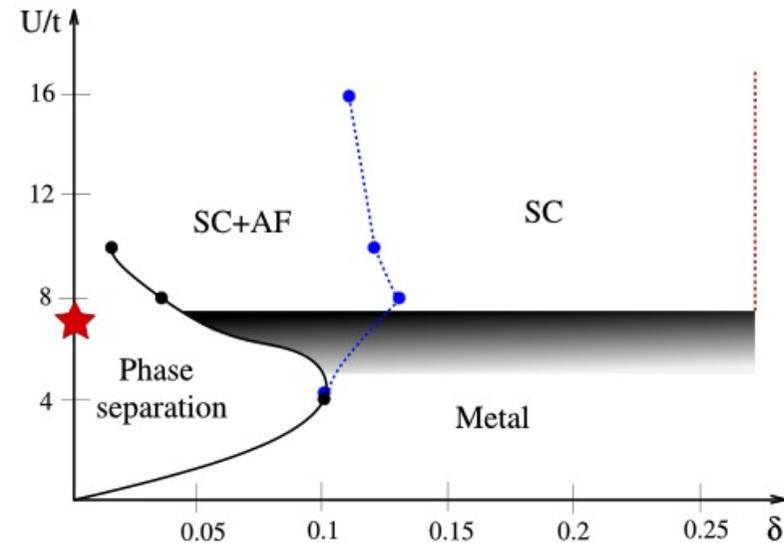


FIG. 7. Schematic phase diagram as obtained by using a combined VMC and GFMC (with FN approximation) approach. The red star labels the location of the hidden Mott transition U_{Mott}/t at half-filling. The black line with black dots denotes the boundary of the phase-separation region, that shrinks for $U/t \gtrsim U_{\text{Mott}}/t$. The curve is left open for $U/t > 10$, since we cannot exclude the presence of phase separation very close to half-filling. The dashed blue line with blue dots marks the disappearance of Δ_{AF} in the optimal variational state. The dashed red line indicates the boundary of the region where sizable pairing correlations are detected. Finally, in the shaded gray region finite-size effects are strong and precise results cannot be obtained in the thermodynamic limit.



Backflow with Neural Networks

PHYSICAL REVIEW LETTERS **120**, 205302 (2018)

Nonlinear Network Description for Many-Body Quantum Systems in Continuous Space

Michele Ruggeri,¹ Saverio Moroni,² and Markus Holzmann^{3,4}

¹Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany

²DEMOCRITOS National Simulation Center, Istituto Officina dei Materiali del CNR and SISSA,
Via Bonomea 265, I-34136 Trieste, Italy

³Univ. Grenoble Alpes, CNRS, LPMMC, 3800 Grenoble, France

⁴Institut Laue Langevin, BP 156, F-38042 Grenoble Cedex 9, France

 (Received 8 November 2017; published 17 May 2018)

We show that the recently introduced iterative backflow wave function can be interpreted as a general neural network in continuum space with nonlinear functions in the hidden units. Using this wave function in variational Monte Carlo simulations of liquid ^4He in two and three dimensions, we typically find a tenfold increase in accuracy over currently used wave functions. Furthermore, subsequent stages of the iteration procedure define a set of increasingly good wave functions, each with its own variational energy and variance of the local energy: extrapolation to zero variance gives energies in close agreement with the exact values. For two dimensional ^4He , we also show that the iterative backflow wave function can describe both the liquid and the solid phase with the same functional form—a feature shared with the shadow wave function, but now joined by much higher accuracy. We also achieve significant progress for liquid ^3He in three dimensions, improving previous variational and fixed-node energies.

DOI: [10.1103/PhysRevLett.120.205302](https://doi.org/10.1103/PhysRevLett.120.205302)

PHYSICAL REVIEW LETTERS **122**, 226401 (2019)

Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions

Di Luo and Bryan K. Clark

*Institute for Condensed Matter Theory and Department of Physics,
University of Illinois at Urbana-Champaign, Illinois 61801, USA*

 (Received 25 August 2018; revised manuscript received 12 January 2019; published 4 June 2019)

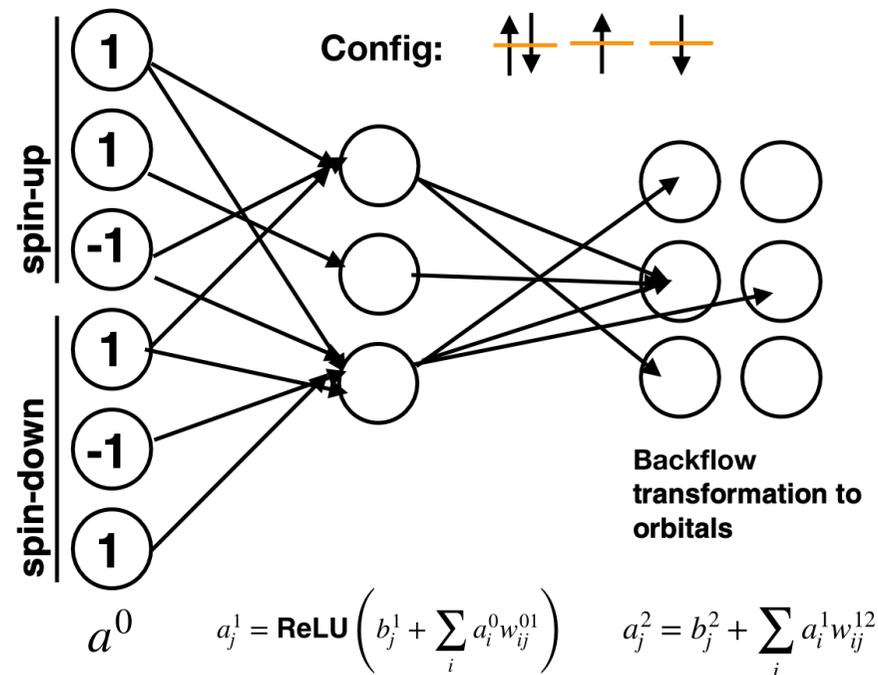
Obtaining an accurate ground state wave function is one of the great challenges in the quantum many-body problem. In this Letter, we propose a new class of wave functions, neural network backflow (NNB). The backflow approach, pioneered originally by Feynman and Cohen [*Phys. Rev.* **102**, 1189 (1956)], adds correlation to a mean-field ground state by transforming the single-particle orbitals in a configuration-dependent way. NNB uses a feed-forward neural network to learn the optimal transformation via variational Monte Carlo calculations. NNB directly dresses a mean-field state, can be systematically improved, and directly alters the sign structure of the wave function. It generalizes the standard backflow [L. F. Tocchio *et al.*, *Phys. Rev. B* **78**, 041101(R) (2008)], which we show how to explicitly represent as a NNB. We benchmark the NNB on Hubbard models at intermediate doping, finding that it significantly decreases the relative error, restores the symmetry of both observables and single-particle orbitals, and decreases the double-occupancy density. Finally, we illustrate interesting patterns in the weights and bias of the optimized neural network.

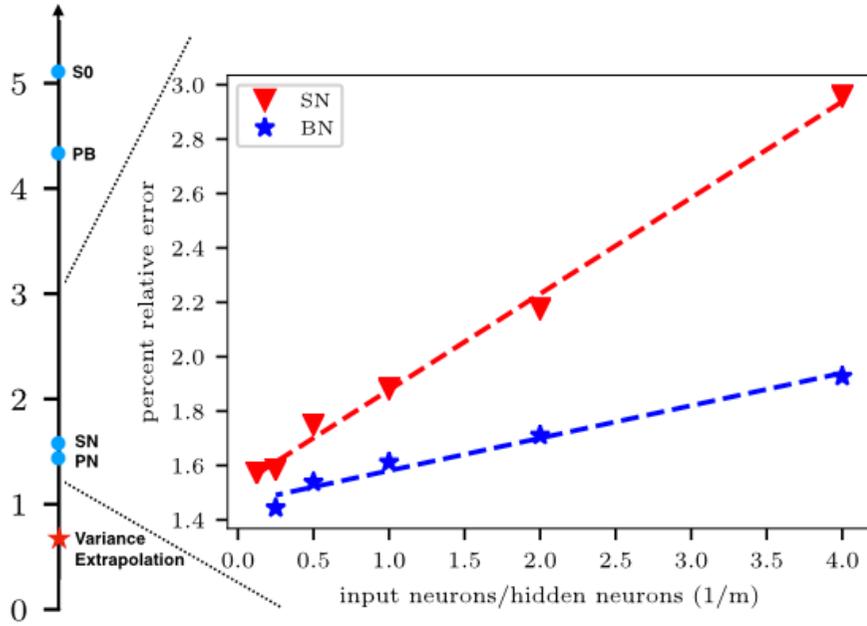
DOI: [10.1103/PhysRevLett.122.226401](https://doi.org/10.1103/PhysRevLett.122.226401)

Di Luo and Clark, PRL 122, 226401 (2019)

NN parametrized form of the backflow term for each orbital:

$$\phi_{k\sigma}^b(r_{i\sigma}; \mathbf{r}_{/i}) = \phi_{k\sigma}(r_{i\sigma}) + a_{ki,\sigma}^{\text{NN}}(\{r_{j\sigma}, j \neq i\})$$





Much better accuracy is obtained with a flexible parametrization of the backflow as provided by a NN, in comparison to a priori parametrizations.

FIG. 2. Top left: Percentage relative error from the exact ground-state energy of Eq. (7) ($E = -11.868$ [64]) on 4×4 Hubbard model at $U/t = 8, n = 0.875$, for various classes of wave functions. The star is the variance extrapolation result of Ψ_{PN} (see Supplemental Material [57] Sec. I). Top right: Percentage relative energy error, $(E_{\text{exact}} - E_{\text{NNB}})/E_{\text{exact}} \times 100\%$, as a function of $1/m$ for NNB. Statistical error bars are shown, but are smaller than the marker size. Bottom: Variance extrapolated energy per site for Hubbard model with $U/t = 8, n = 0.875$ with system size $L \times 4$ for $L = 4, 8, 12, 16$. The dash line is the DMRG energy per site ($-0.7659 \pm 4 \times 10^{-5}$) for system size $\infty \times 4$ (PBC, open) [65].

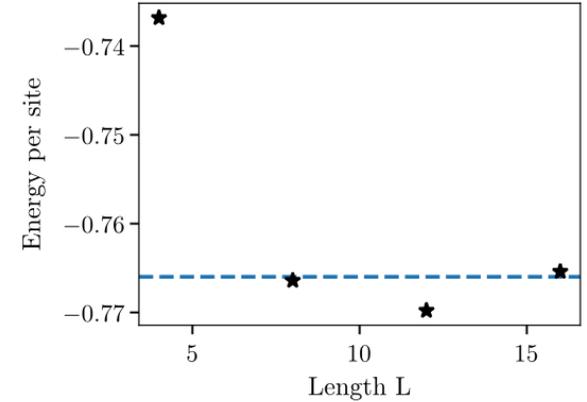


TABLE I. Wave function *Ansätze*

Method	Backflow transformation	Mean field	Variational functions
Ψ_{S0}	not applicable	Eq. (1)	$\phi_{k\uparrow}(r_{i,\uparrow}), \phi_{k\downarrow}(r_{i,\downarrow}), v_{ij}$
Ψ_{SN}	$\phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) = \phi_{k\sigma}(r_{i,\sigma}) + a_{ki,\sigma}^{\text{NN}}(\mathbf{r})$	Eq. (1)	$a_{ki,\sigma}^{\text{NN}}(\mathbf{r}), v_{ij}$
Ψ_{PB}	$\phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) = \phi_{k\sigma}(r_{i,\sigma}) + \theta_{1\sigma} \sum_j t D_i H_j \phi_{k\sigma}(r_{j,\sigma}) + \theta_{2\sigma} \sum_j t D_i H_j \phi_{k\sigma}(r_{j,\sigma})$	Eq. (3)	$\theta_{1\sigma}, \theta_{2\sigma}, S_{kl}, v_{ij}$
Ψ_{PN}	$\phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) = \phi_{k\sigma}(r_{i,\sigma}) + a_{ki,\sigma}^{\text{NN}}(\mathbf{r}); S_{kl}(\mathbf{r}) = S_{kl} + d_{kl}^{\text{NN}}(\mathbf{r})$	Eq. (3)	$a_{ki,\sigma}^{\text{NN}}(\mathbf{r}), d_{kl}^{\text{NN}}(\mathbf{r}), v_{ij}$

Neural Quantum States and ML: The Dawn of a New Era for Fermionic Variational Wave Functions

Two broad classes of methods:

1. Antisymmetry guaranteed by structure of the wave-function – usually Slater determinant or superpositions of SDs (“1st quantization”)

$$\Psi(x_1, \dots, x_N) = \varepsilon_\sigma \Psi(x_{\sigma(1)}, \dots, x_{\sigma(N)})$$

2. Wave-function represented on the basis of configurations in Fock space – antisymmetry of amplitudes is insured by taking care of the sign between $|x\rangle$ and $|n(x)\rangle$ (“2nd quantization”)

$$|\Psi\rangle = \frac{1}{N!} \sum_x \psi(x) |x\rangle = \sum_n \psi(n) |n\rangle$$

$$\psi(x) = \varepsilon_{\sigma(x)} \psi[n(x)]$$

The second method can only be used by choosing a one-particle basis (always the case for lattice models) while the first can be used directly in the continuum. Beware of computational overhead re: interaction tensor U_{ijkl} if a basis is used!

The non-determinantal ('2nd quantized') approaches are a quite direct extension of what we saw earlier for spin models

Two early papers:

ARTICLE

DOI: [10.1038/s41467-018-06598-z](https://doi.org/10.1038/s41467-018-06598-z)

OPEN

Nature Communications, 2018

Quantum machine learning for electronic structure calculations

Rongxin Xia¹ & Sabre Kais^{1,2,3}

[Note: 'Quantum' ML in the title is not an essential aspect]

ARTICLE

<https://doi.org/10.1038/s41467-020-15724-9>

OPEN

Nature Communications, 2020



Fermionic neural-network states for ab-initio electronic structure

Kenny Choo¹✉, Antonio Mezzacapo²✉ & Giuseppe Carleo³✉

Indeed, fermionic creation and annihilation operators can be represented as spin operators:

$$c_p \sim \sigma_p^- , \quad c_p^+ \sim \sigma_p^+ , \quad c_p^+ c_p \sim \frac{1}{2} (\sigma_p^z + 1)$$

BUT, importantly, one must take care of the anticommutation:

$$\{c_p, c_q^+\} = \delta_{pq}$$

$$|x\rangle \equiv |\{p_1, \dots, p_N\}\rangle = \varepsilon_{\sigma(x)} |(p_1, \dots, p_N)_{can.ord.}\rangle \leftrightarrow \varepsilon_{\sigma(x)} |n(x)\rangle$$

This can be formalized as a Jordan-Wigner ‘string’ (or other encoding such as Bravyi-Kitaev):

$$c_p \rightarrow \left[\prod_{q=1}^{p-1} \sigma_q^z \right] \sigma_p^- , \quad c_p^+ \rightarrow \left[\prod_{q=1}^{p-1} \sigma_q^z \right] \sigma_p^+$$

This is basically bookkeeping for keeping track of the sign of the permutation bringing (x) to its canonically ordered form – can indeed be coded as bookkeeping when accepting a move $x \rightarrow x'$ and calculating matrix elements by sampling

For more info on fermion to qubit mapping see e.g. Nys and Carleo arXiv:2205.00733

The wave function amplitude on configuration x can be encoded as a NN, e.g. an RBM, as in previous lectures and VMC performed, etc.

Computational cost: the bottleneck is the computation of the 'local energy' which costs $N_{\text{par}} \cdot N_{\text{ord}} \cdot N_{\text{conf}}$ with:

- N_{par} the number of parameters in the NN
- N_{ord} the number of modes we encounter when reordering the configuration
- N_{conf} the number of distinct configurations encountered when calculating the matrix elements of the hamiltonian for a given sample

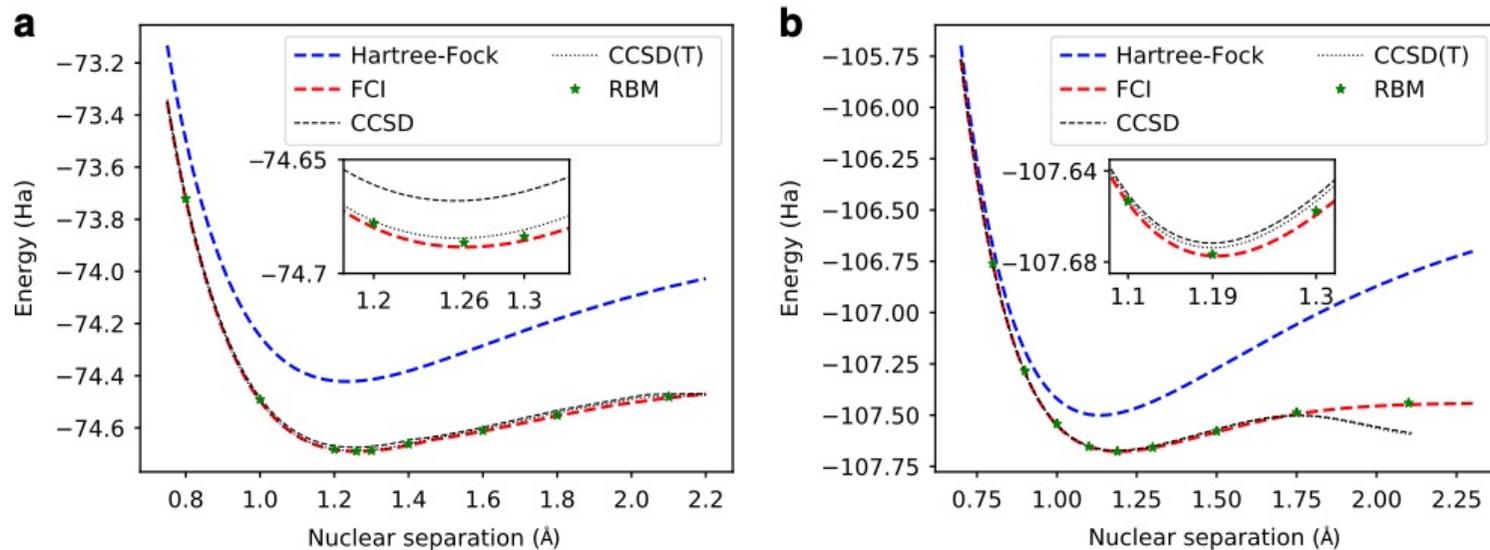
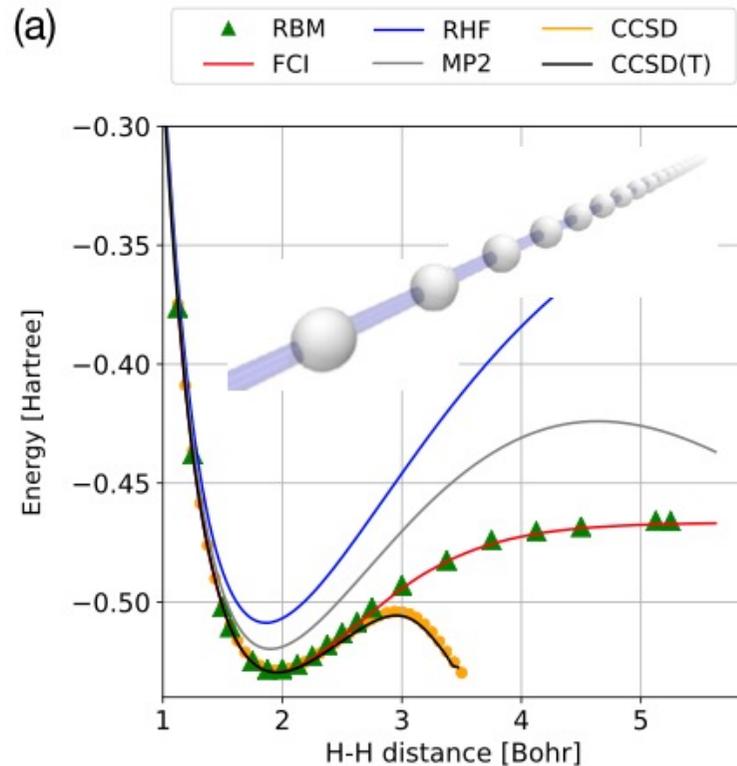


Fig. 1 Dissociation profiles. The accuracy of fermionic neural-network quantum states compared with other quantum chemistry approaches. Shown here are dissociation curves for **a** C₂ and **b** N₂, in the STO-3G basis with 20 spin-orbitals. The RBM used has 40 hidden units, and it is compared to both coupled-cluster approaches (CCSD, CCSD(T)) and FCI energies.

Extension to the solid-state (periodic systems)

N.Yoshioka, W.Mizokami and F.Nori Comm. Phys. 2021



$$H = \sum_{pq} \sum_{\mathbf{k}} t_{pq}^{\mathbf{k}} c_{p\mathbf{k}}^{\dagger} c_{q\mathbf{k}} + \frac{1}{2} \sum_{pqrs} \sum_{\mathbf{k}_p \mathbf{k}_q \mathbf{k}_r \mathbf{k}_s} v_{pqrs}^{\mathbf{k}_p \mathbf{k}_q \mathbf{k}_r \mathbf{k}_s} c_{p\mathbf{k}_p}^{\dagger} c_{q\mathbf{k}_q} c_{r\mathbf{k}_r}^{\dagger} c_{s\mathbf{k}_s}$$

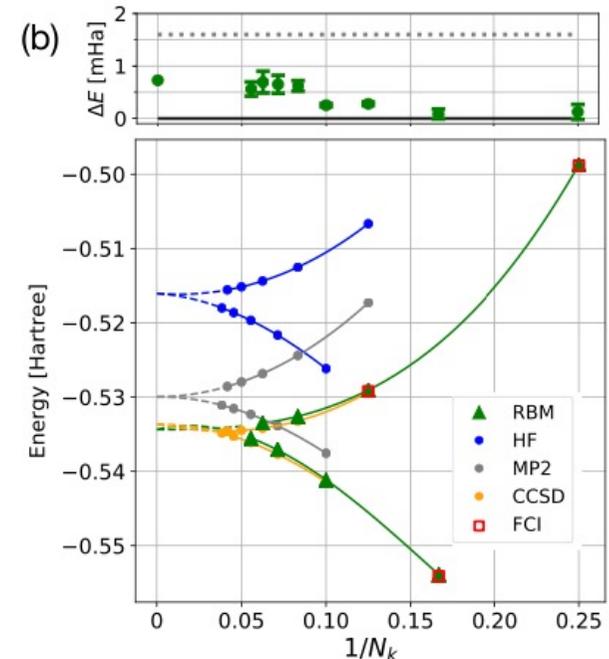
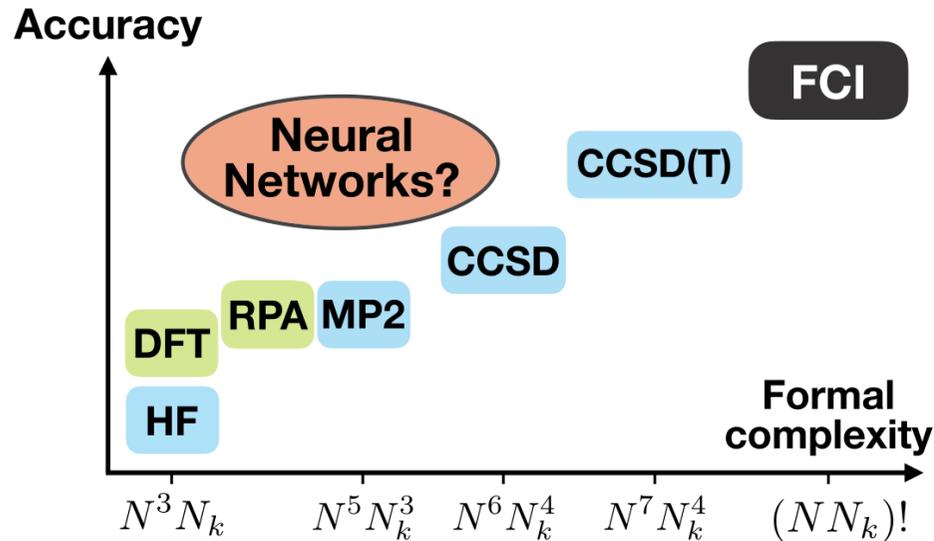


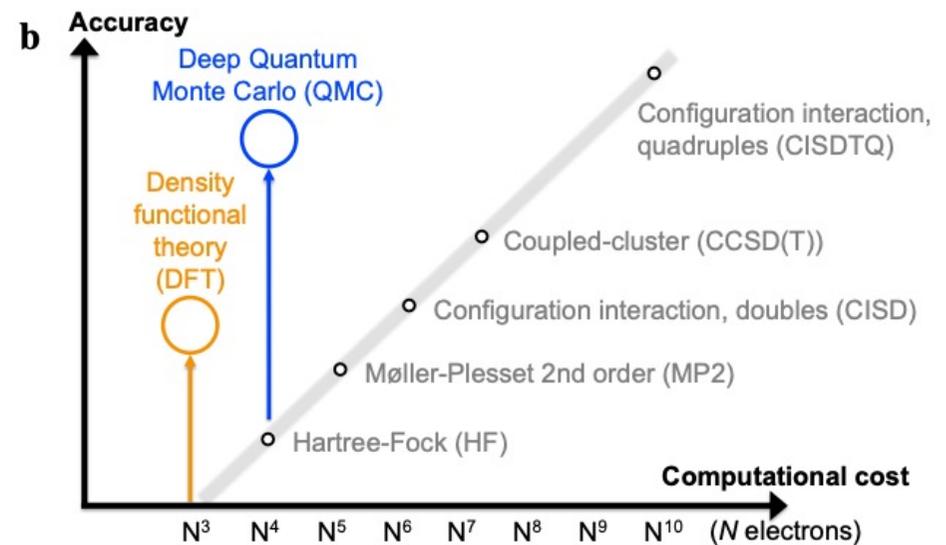
Fig. 2 Solving the ground state of the linear hydrogen chain using the minimal STO-3G basis set. **a** The potential energy curve calculated by the restricted Boltzmann machine (RBM) agrees with the full configuration interaction (FCI) method within chemical accuracy (1.6 mHa) for any atom separation d_H . This indicates that the RBM states are capable of describing both the weakly and strongly interacting regimes, where gold-standard techniques, such as coupled-cluster singles and doubles (CCSD) shown by the yellow line and CCSD with perturbative triple excitations (CCSD(T)) in black line, break down. The results by restricted Hartree-Fock (RHF) and second-order Møller-Plesset perturbation theory are indicated by blue and gray lines, respectively. A unit cell consists of four hydrogen atoms placed at even intervals, and two k -points are sampled from a uniform grid. **b** Finite-size scaling of the ground-state energy up to $N_k = 18$ and its deviation from the FCI ($N_k \leq 8N_k \leq 8$) or CCSD(T) ($N_k > 8N_k > 8$), ΔE , at near-equilibrium $d_H = 2$. The results show excellent agreement with conventional methods even in the thermodynamic limit $N_k \rightarrow \infty$. Here, the unit cell consists of a single hydrogen atom, and hence the maximum number of spin orbitals considered here is 36. The error bars denote the standard deviation of the estimation by the Monte Carlo sampling.

The goal: cutting computational cost with NNs



Yoshioka et al. Comm Phys 2021

Hermann et al. arXiv:2208.12590
(A recent review article)



Determinant based/ 1^{st} quantized approaches: some examples

- **FermiNet:** D.Pfau et al. (Google DeepMind + Imperial College) Phys Rev Research 2, 033429 (2020) and subsequent papers (e.g. arXiv:2011.07125)
- **PauliNet:** J.Hermann, Z.Schätzle and F.Noé (FU+TU, Berlin) Nature Chemistry 12, 891 (2020) and subsequent papers
- **HFDS (Hidden Fermions Determinantal State)** J.Robledo-Moreno, G.Carleo, A.G. and J.Stokes PNAS, 119 Vol 32 (2022) and subsequent papers e.g. A.Lovato et al. PRR 4, 043178 (2022)
- **Also:**
- **Early paper:** Nomura et al. PRB 96, 205152 (2017) [RBM parametrized Jastrow etc.]
- J.Stokes et al. PRB 102, 205122 (2020)
- NN backflows, see above: Ruggeri et al. PRL 120, Di Luo and Clark PRL 122
- Several relevant works not quoted here, for a recent short review see J.Hermann et al. arXiv:2208.12590

FermiNet

D.Pfau et al. Phys Rev Research 2, 033429 (2020) and subsequent papers

Directly in continuum space
(no basis set)

$$\hat{H}\psi(\mathbf{x}_1, \dots, \mathbf{x}_n) = E\psi(\mathbf{x}_1, \dots, \mathbf{x}_n),$$

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{iI} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{I>J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|},$$

$$\psi(\mathbf{r}_1^\uparrow, \dots, \mathbf{r}_{n^\downarrow}^\downarrow) = \sum_k \omega_k \left(\det [\phi_i^{k\uparrow}(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}^\uparrow\}); \{\mathbf{r}^\downarrow\}] \right. \\ \left. \times \det [\phi_i^{k\downarrow}(\mathbf{r}_j^\downarrow; \{\mathbf{r}_{/j}^\downarrow\}); \{\mathbf{r}^\uparrow\};] \right) \quad \text{Sum of determinants}$$

$$\phi_i^{k\alpha}(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) = (\mathbf{w}_i^{k\alpha} \cdot \mathbf{h}_j^{L\alpha} + g_i^{k\alpha}) \\ \times \sum_m \pi_{im}^{k\alpha} \exp(-|\boldsymbol{\Sigma}_{im}^{k\alpha}(\mathbf{r}_j^\alpha - \mathbf{R}_m)|),$$

Orbitals parametrized as deep NN
with backflow etc.

Algorithm 1: FermiNet evaluation.

Require: walker configuration $\{\mathbf{r}_1^\uparrow, \dots, \mathbf{r}_{n^\uparrow}^\uparrow, \mathbf{r}_1^\downarrow, \dots, \mathbf{r}_{n^\downarrow}^\downarrow\}$

Require: nuclear positions $\{\mathbf{R}_I\}$

1: **for each** electron i, α **do**

2: $\mathbf{h}_i^{\ell\alpha} \leftarrow \text{concatenate}(\mathbf{r}_i^\alpha - \mathbf{R}_I, |\mathbf{r}_i^\alpha - \mathbf{R}_I| \forall I)$

3: $\mathbf{h}_{ij}^{\ell\alpha\beta} \leftarrow \text{concatenate}(\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta, |\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta| \forall j, \beta)$

4: **end for**

5: **for each layer** $\ell \in \{0, L-1\}$ **do**

6: $\mathbf{g}^{\ell\uparrow} \leftarrow \frac{1}{n^\uparrow} \sum_i^{n^\uparrow} \mathbf{h}_i^{\ell\uparrow}$

7: $\mathbf{g}^{\ell\downarrow} \leftarrow \frac{1}{n^\downarrow} \sum_i^{n^\downarrow} \mathbf{h}_i^{\ell\downarrow}$

8: **for each electron** i, α **do**

9: $\mathbf{g}_i^{\ell\alpha\uparrow} \leftarrow \frac{1}{n^\uparrow} \sum_j^{n^\uparrow} \mathbf{h}_{ij}^{\ell\alpha\uparrow}$

10: $\mathbf{g}_i^{\ell\alpha\downarrow} \leftarrow \frac{1}{n^\downarrow} \sum_j^{n^\downarrow} \mathbf{h}_{ij}^{\ell\alpha\downarrow}$

11: $\mathbf{f}_i^{\ell\alpha} \leftarrow \text{concatenate}(\mathbf{h}_i^{\ell\alpha}, \mathbf{g}^{\ell\uparrow}, \mathbf{g}^{\ell\downarrow}, \mathbf{g}_i^{\ell\alpha\uparrow}, \mathbf{g}_i^{\ell\alpha\downarrow})$

12: $\mathbf{h}_i^{\ell+1\alpha} \leftarrow \tanh(\text{matmul}(\mathbf{V}^l, \mathbf{f}_i^{\ell\alpha}) + \mathbf{b}^l) + \mathbf{h}_i^{\ell\alpha}$

13: $\mathbf{h}_{ij}^{\ell+1\alpha\beta} \leftarrow \tanh(\text{matmul}(\mathbf{W}^l, \mathbf{h}_{ij}^{\ell\alpha\beta}) + \mathbf{c}^l) + \mathbf{h}_{ij}^{\ell\alpha\beta}$

14: **end for**

15: **end for**

16: **for each determinant** k **do**

17: **for each orbital** i **do**

18: **for each electron** j, α **do**

19: $e \leftarrow \text{envelope}(\mathbf{r}_j^\alpha, \{\mathbf{r}_i^\alpha - \mathbf{R}_I\})$

20: $\phi_i(\mathbf{r}_j^\alpha; \{\mathbf{r}_j^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) = (\text{dot}(\mathbf{w}_i^{k\alpha}, \mathbf{h}_i^{L\alpha}) + g_i^{k\alpha})e$

21: **end for**

22: **end for**

23: $D^{k\uparrow} \leftarrow \det[\phi_i^{k\uparrow}(\mathbf{r}_j^\uparrow; \{\mathbf{r}_j^\uparrow\}; \{\mathbf{r}^\downarrow\})]$

24: $D^{k\downarrow} \leftarrow \det[\phi_i^{k\downarrow}(\mathbf{r}_j^\downarrow; \{\mathbf{r}_j^\downarrow\}; \{\mathbf{r}^\uparrow\})]$

25: **end for**

26: $\psi \leftarrow \sum_k \omega_k D^{k\uparrow} D^{k\downarrow}$

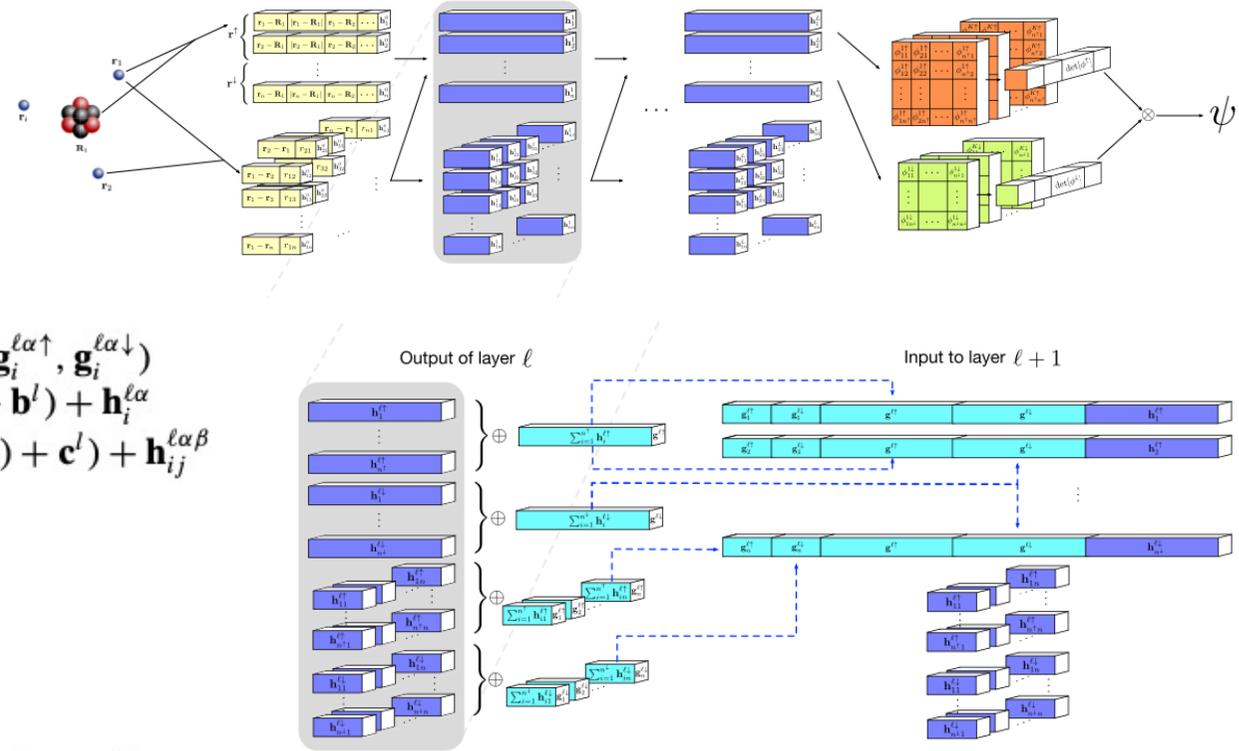
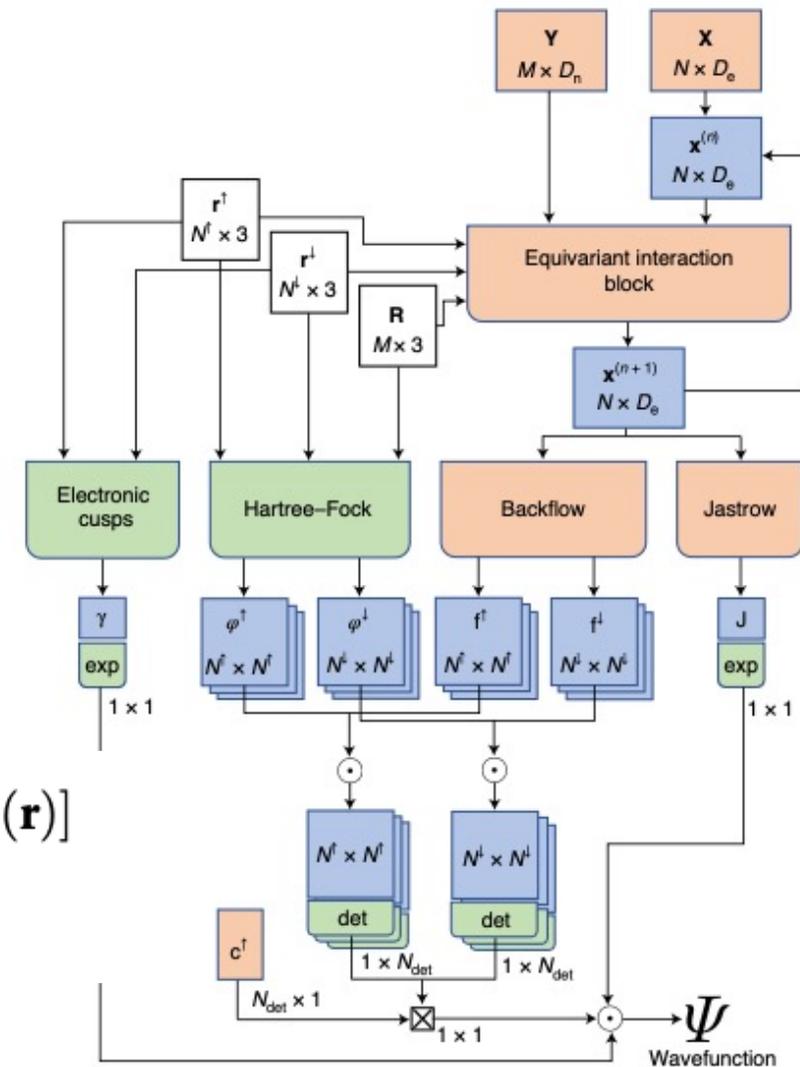


FIG. 1. The Fermionic neural network (FermiNet). Top: Global architecture. Features of one or two electron positions are inputs to different streams of the network. These features are transformed through several layers, a determinant is applied, and the wave function at that position is given as output. Bottom: Detail of a single layer. The network averages features of electrons with the same spin together, then concatenates these features to construct an equivariant function of electron position at each layer.

PauliNet

J.Hermann, Z.Schätzle and F.Noé
 Nature Chemistry 12, 891 (2020)

Fig. 2 | Architecture of the newly developed PauliNet wavefunction ansatz. The information flows from the input electron and nuclear coordinates, \mathbf{r} and \mathbf{R} , to the output wavefunction value, Ψ . Modelling the wavefunction via Jastrow and backflow functions is common in QMC, but here these functions are learned with DNNs. N_{det} number of determinants.



$$\Psi_{\theta}(\mathbf{r}) = e^{\gamma(\mathbf{r}) + J_{\theta}(\mathbf{r})} \sum_p c_p \det[\tilde{\varphi}_{\theta, \mu_p}^{\uparrow}(\mathbf{r})] \det[\tilde{\varphi}_{\theta, \mu_p}^{\downarrow}(\mathbf{r})]$$

$$\tilde{\varphi}_{\mu_i}(\mathbf{r}) = \varphi_{\mu}(\mathbf{r}_i) f_{\theta, \mu_i}(\mathbf{r})$$

Slater-Jastrow + Backflow
 More physical input than
 more agnostic FermiNet

N Number of electrons	Neural network (trainable function)	Trainable array
N^{\uparrow} Number of spin-up electrons	Fixed function	Input array
N^{\downarrow} Number of spin-down electrons		Hidden array
M Number of nuclei		
D_e Embedding dimension, electrons		
D_n Embedding dimension, nuclei		

TABLE II. Ground-state energy at equilibrium geometry for diatomics and small molecules. The percentage of correlation energy captured by the FermiNet relative to the exact energy (where available) or CCSD(T)/CBS is given in the rightmost column. If no citation is provided, then the number was from our own calculation. Geometries for larger molecules are given in Appendix G.

Molecule	Bond length (a_0)	FermiNet (E_h)	CCSD(T) (E_h)			HF (E_h)		Exact (E_h)	% corr
			aug-cc-pCVQZ	aug-cc-pCV5Z	CBS	CBS			
LiH	3.015	-8.07050(1)	-8.0687	-8.0697	-8.070696	-7.98737	-8.070548 [46]	99.94(1)	
Li ₂	5.051	-14.99475(1)	-14.9921	-14.9936	-14.99507	-14.87155	-14.9954 [47]	99.47(1)	
NH ₃	—	-56.56295(8)	-56.5535	-56.5591	-56.5644	-56.2247	—	99.57(2)	
CH ₄	—	-40.51400(7)	-40.5067	-40.5110	-40.5150	-40.2171	—	99.66(3)	
CO	2.173	-113.3218(1)	-113.3047	-113.3154	-113.3255	-112.7871	—	99.32(3)	
N ₂	2.068	-109.5388(1)	-109.5224	-109.5327	-109.5425	-108.9940	-109.5423 [47]	99.36(2)	
Ethene	—	-78.5844(1)	-78.5733	-78.5812	-78.5888	-78.0705	—	99.16(2)	
Methylamine	—	-95.8554(2)	-95.8437	—	-95.8653	-95.2628	—	98.36(3)	
Ozone	—	-225.4145(3)	-225.3907	-225.4119	-225.4338	-224.3526	—	98.42(3)	
Ethanol	—	-155.0308(3)	-155.0205	—	-155.0545	-154.1573	—	97.36(4)	
Bicyclobutane	—	-155.9263(6)	-155.9216	—	-155.9575	-154.9372	—	96.94(5)	

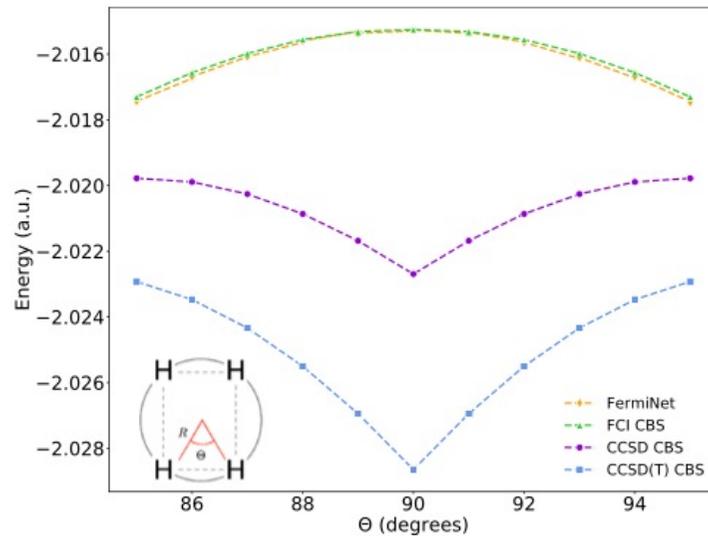


FIG. 4. The H_4 rectangle, $R = 3.2843a_0$. Coupled cluster methods incorrectly predict a cusp and energy minimum at $\Theta = 90^\circ$, while the FermiNet approach agrees with exact FCI results.

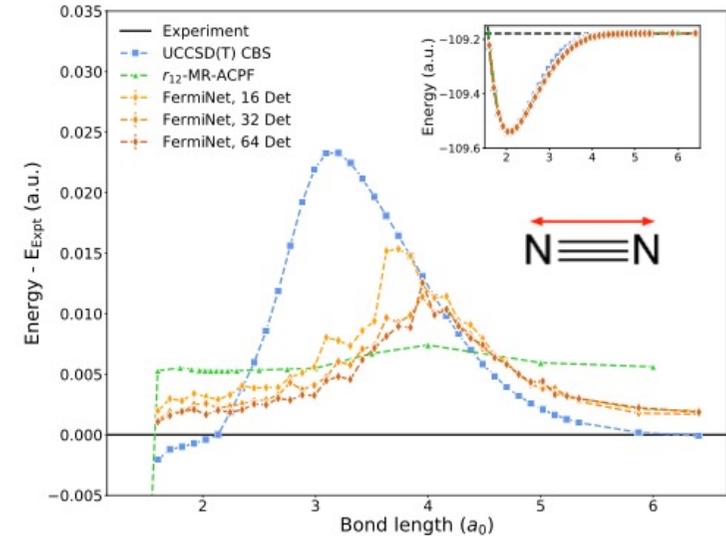


FIG. 5. The dissociation curve for the nitrogen triple-bond. The difference from experimental data [53] is given in the main panel. In the region of largest UCCSD(T) error, the FermiNet prediction is comparable to highly accurate r_{12} -MR-ACPF results [54].

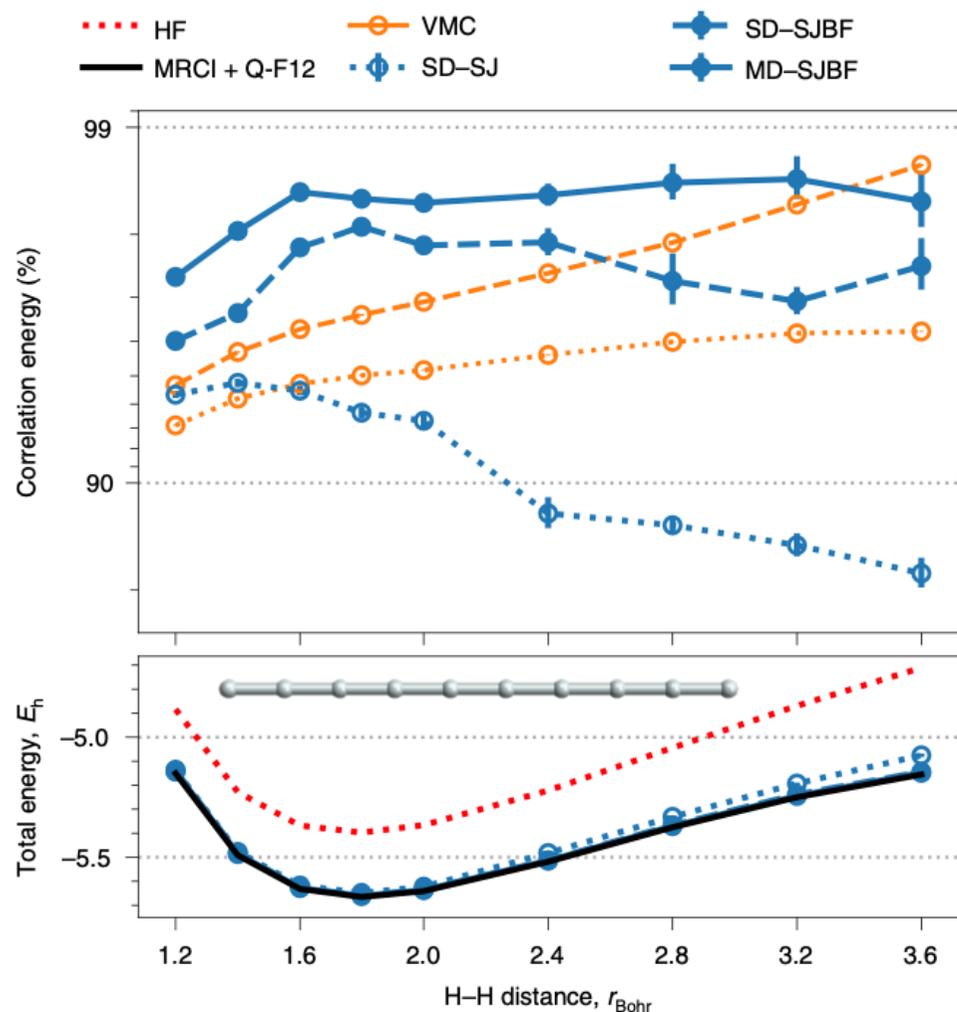


Fig. 6 | PauliNet captures strong correlation in H_{10} along the dissociation curve. PauliNet results (blue) with single determinant or 16 determinants (MD) and with or without backflow are shown. The backflow plays a much larger role than multiple determinants. PauliNet outperforms highly specialized VMC ansatzes (orange) of the single-determinant (dotted) and multideterminant geminal (dashed) form by Motta et al.³⁵. The correlation energy is calculated with respect to multireference configuration-interaction (MRCI) results, also by Motta et al. E_h , Hartree energy; r_{Bohr} , Bohr radius.

Comparison FermiNet/PauliNet

Spencer et al. arXiv:2011.07125

Hermann et al. arXiv:2208.12590

Sketch from Monino et al.
arXiv:2204.05098

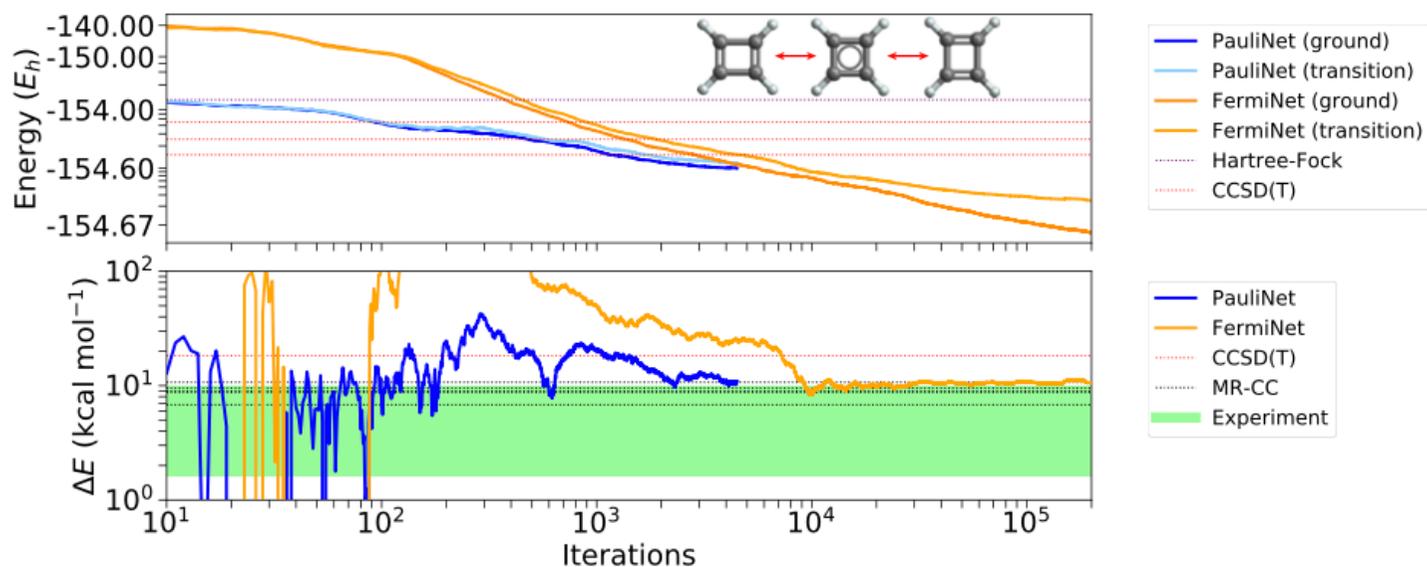
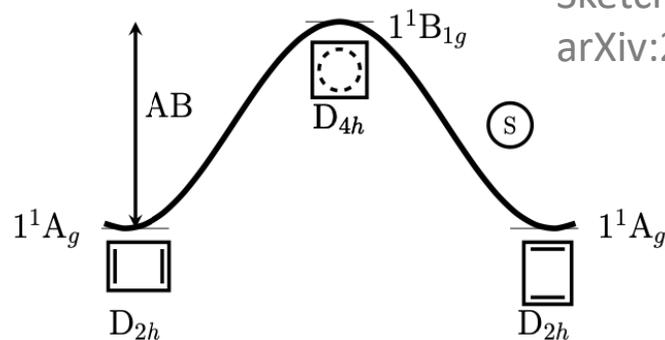


Figure 2: Automerization of cyclobutadiene and comparison with the PauliNet. The geometries of the ground state (left and right) and transition state (center) are shown in the top figure. Top: total energy of the FermiNet and PauliNet. Energy is plotted on a log scale zeroed at $-154.68 E_h$. While the FermiNet is initialized well above the PauliNet, the converged energy is $\sim 70 mE_h$ lower – ~ 44 kcal/mol. Bottom: energy difference between the ground and transition state for the PauliNet and FermiNet. Both the PauliNet and FermiNet are at the very upper end of the experimentally-measured values. The FermiNet agrees well with the highest multireference coupled cluster (MR-CC) result. All results other than the FermiNet are from [7].

Hidden Fermion Determinantal States (HFDS)

PNAS

RESEARCH ARTICLE

PHYSICS

 OPEN ACCESS



Fermionic wave functions from neural-network constrained hidden states

Javier Robledo Moreno^{a,b,1} , Giuseppe Carleo^{c,d}, Antoine Georges^{a,e,f,g} , and James Stokes^{a,h}

PNAS 2022 Vol. 119 No. 32 e2122059119

- For a recent extension to continuum space and application to nuclear matter, see: A.Lovato et al. Phys Rev Research 4, 043178 (2022)

Inspired by auxiliary 'slave' particle methods

Key idea: Enlarge Hilbert space and project back onto the physical Hilbert space by imposing a constraint.

Example: 'Slave boson' for $U=\infty$

Local Hilbert space:

$$|0\rangle \rightarrow b^+ |0\rangle, \quad |\sigma\rangle = c_\sigma^+ |0\rangle \rightarrow f_\sigma^+ |0\rangle \quad (\text{'Holon' / 'Spinon'})$$

Impose constraint on each site: $\forall i, \quad b_i^+ b_i + \sum_{\sigma} f_{i\sigma}^+ f_{i\sigma} = 1$

This constraint prevents double occupancies, as well as (unphysical) states with more than one boson per site. It must be imposed on all sites.

Configuration in enlarged Hilbert space is uniquely related to the physical configuration in the physical (projected) Hilbert space:

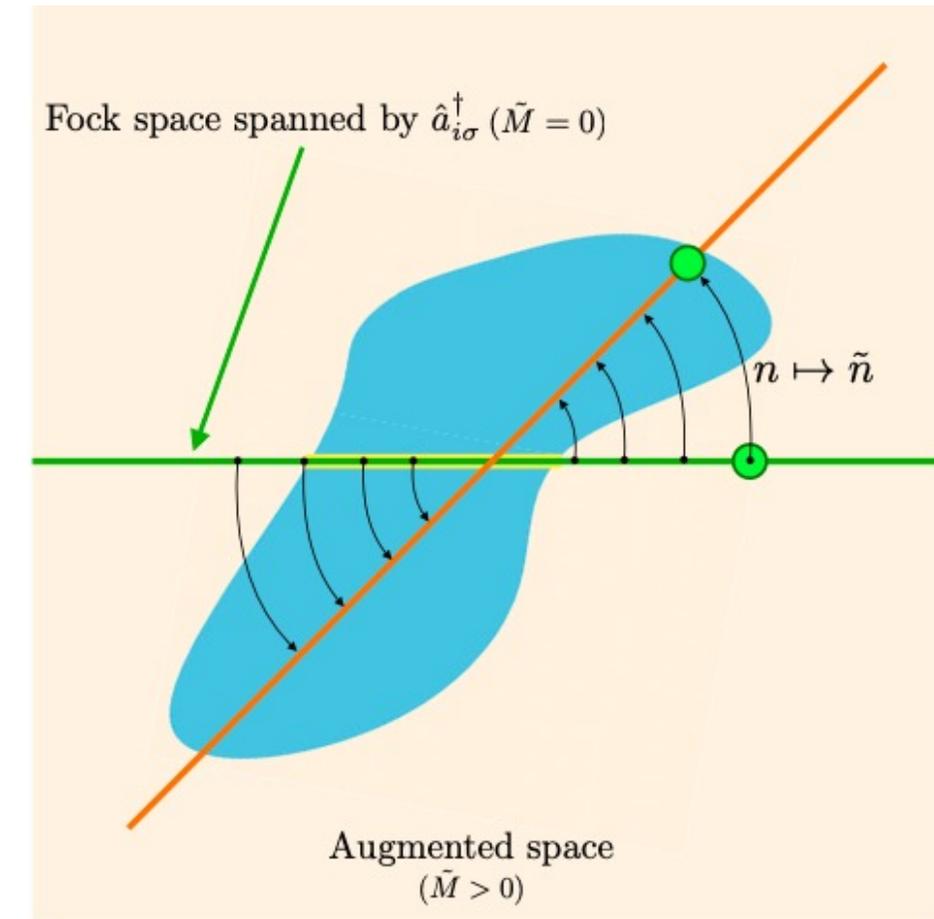
$$|n_c = 0\rangle \rightarrow |n_b = 1, n_{f\sigma} = 0\rangle, \quad |n_{c\sigma} = 1\rangle \rightarrow |n_b = 0, n_{f\sigma} = 1\rangle$$

HFDS: The Basic Concept

HFDS enlarges the Hilbert space with *fermionic* 'hidden' modes.

The wave function is a Slater determinant in the enlarged space, subject to a constraint for projecting it back to the physical space.

This results in a correlated/entangled wave-function



- Target correlated state
- SDs in augmented space
- SDs in physical space

Fig. 1. Depiction of the geometrical interpretation of the hidden fermion formalism. The Fock space spanned by the visible-fermionic modes $\hat{a}_{i\sigma}^\dagger$ is represented by the green horizontal line. The augmented Fock space is represented by the light orange plane (plane of the paper). The orange diagonal line represents the subspace in the augmented Fock space that is isomorphic to the physical Hilbert space after applying the constraint function (black arrows). The collection of SDs in the augmented space is represented by the blue shape, and the intersection with the subspace of just visible DOFs is marked in yellow. This intersection corresponds to the physical Hartree-Fock states. The constraint function changes the collection of states that represent the physical Hilbert space bringing the target correlated state close to a Slater determinant in the enlarged space.

HFDS formalism (done on the board)

- Example: 2-site Hubbard
- General formalism
- Special cases; Jastrow, Gutzwiller etc
- NN parametrization
- See lecture notes on website

VIII. HIDDEN FERMIONS DETERMINANTAL STATE

A. Simple example: 2-site Hubbard model

Recall that the ground state in the $N_\uparrow = N_\downarrow = 1$ sector is:

$$|\Psi_0\rangle = \cos\theta \frac{1}{\sqrt{2}}[c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger] + \sin\theta \frac{1}{\sqrt{2}}[c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{1\downarrow}^\dagger] \quad (8.1)$$

with:

$$\frac{1}{\tan\theta} = -u \left[1 + \sqrt{1 + \frac{1}{u^2}} \right], \quad u = \frac{U}{4t} \quad (8.2)$$

This wave-function cannot be written as a product of two separate Slater determinants for spin-up and spin-down electrons (however, it can if the one-particle orbital are allowed to mix the two spins and one projects back onto the sector with $N_\uparrow = N_\downarrow$).

we are going to show that Ψ_0 can be written as a projected SD by introducing a single hidden fermion which can occupy two modes, denoted (for this example) S and D (S and D stand for ‘single’ and ‘double’, respectively). We consider the matrix defining the new orbitals:

$$\Phi_{p\alpha}, \quad p = (1\uparrow, 2\uparrow, 1\downarrow, 2\downarrow, S, D), \quad \alpha = (1, 2, 3) = (\uparrow, \downarrow, h) \quad (8.3)$$

$$\Phi = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & c \\ 0 & 0 & ce^{-g} \end{pmatrix} \quad (8.4)$$

c is simply a normalization: $c^2 = 1/[1 + e^{-2g}]$ i.e. $c = e^{g/2}/\sqrt{2 \cosh g}$ so that $\Phi^\dagger \Phi = \mathbb{I}$.

This corresponds to the operators creating electrons in the three 1-particle wave-functions defined by Φ :

$$\varphi_\uparrow^+ = \frac{1}{\sqrt{2}}[c_{1\uparrow}^\dagger + c_{2\uparrow}^\dagger] \quad (8.5)$$

$$\varphi_\downarrow^+ = \frac{1}{\sqrt{2}}[c_{1\downarrow}^\dagger + c_{2\downarrow}^\dagger] \quad (8.6)$$

$$h^+ = c[h_S^+ + e^{-g} h_D^+] \quad (8.7)$$

We form the Slater determinant with $N_\uparrow = N_\downarrow = N_h = 1$:

$$|\Psi_{SD}\rangle = \varphi_\uparrow^+ \varphi_\downarrow^+ h^+ |0\rangle \quad (8.8)$$

and project it subject to the constraint that *the hidden fermion occupies state $|S\rangle$ if the physical configuration has no double occupancy and state $|D\rangle$ if the physical configuration has one double occupancy* - note that with two electrons at most one double occupancy can occur. This can be written:

$$h_s^+ h_S = \hat{P}_{\hat{D}=0} \equiv [1 - \hat{n}_{1\uparrow} \hat{n}_{1\downarrow}] [1 - \hat{n}_{2\uparrow} \hat{n}_{2\downarrow}], \quad h_D^+ h_D = \hat{P}_{\hat{D}=1} \equiv \hat{n}_{1\uparrow} \hat{n}_{1\downarrow} + \hat{n}_{2\uparrow} \hat{n}_{2\downarrow} \quad (8.9)$$

The hidden fermion acts as a ‘tag’: singly occupied configurations are given the tag ‘S’, and double the tag ‘D’. Note that the expressions of the projector above are valid only in the $N = 2$ sector considered here. The full projector imposing these two constraints reads (note that for a given fermionic mode a , $h_a^\dagger h_a$ is a projector):

$$\hat{P} = \hat{P}_0 h_S^\dagger h_S + \hat{P}_1 h_D^\dagger h_D \quad (8.10)$$

The HFDS state is the projected SD:

$$|\Psi_{\text{HFDS}}\rangle = |\Psi_{\text{SD}}\rangle = \hat{P} \varphi_\uparrow^\dagger \varphi_\downarrow^\dagger h^+ |0\rangle \quad (8.11)$$

Note that in contrast to most ‘slave particle’ theories, the constraint is not quadratic in the physical electron operators - the theory is not geared at performing analytical mean-field approximations but rather at optimizing the constraint. The HFDS state reads (dropping the subscript):

$$|\Psi\rangle = \frac{c}{2} e^{-g} [c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger] |0\rangle \otimes |D\rangle + \frac{c}{2} [c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{1\downarrow}^\dagger] |0\rangle \otimes |S\rangle \quad (8.12)$$

The occupation numbers of the hidden fermions are determined in a unique way by that of the physical electrons. Explicitly, in second quantized notations with $(1 \uparrow, 2 \uparrow, 1 \downarrow, 2 \downarrow; S, D)$ ordering:

$$\left[\begin{array}{l} (1 \uparrow, 1 \downarrow) \equiv (1, 0, 1, 0) \rightarrow (1, 0, 1, 0; 0, 1) \Rightarrow ce^{-g}/2 \\ (2 \uparrow, 2 \downarrow) \equiv (0, 1, 0, 1) \rightarrow (0, 1, 0, 1; 0, 1) \Rightarrow ce^{-g}/2 \\ (1 \uparrow, 2 \downarrow) \equiv (1, 0, 0, 1) \rightarrow (1, 0, 0, 1; 1, 0) \Rightarrow c/2 \\ (2 \uparrow, 1 \downarrow) \equiv (0, 1, 1, 0) \rightarrow (0, 1, 1, 0; 1, 0) \Rightarrow c/2 \end{array} \right] \quad (8.13)$$

Hence, the basis states in the enlarged HS can be labelled $|n, f(n)\rangle$ and the components of the HFDS state read:

$$\psi(n) \equiv \langle n, f(n) | \Psi \rangle = \det[\mathbf{n} \star \Phi_{\text{phys}}] ce^{-gD(n)} \quad (8.14)$$

in which Φ_{phys} is the matrix restricted to the physical sector:

$$\Phi_{\text{phys}} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (8.15)$$

and \star denote the component by component product:

$$[n \star \Phi]_{p\alpha} \equiv n_p \Phi_{p\alpha} \quad (8.16)$$

If alternatively we use configurations in first quantized form $x = (i_1 \sigma_1, i_2 \sigma_2)$, the wave-function amplitude reads:

$$\psi(x) = \varepsilon_{\sigma(x)} \psi[n(x)] \quad (8.17)$$

with $\sigma(x)$ the permutation bringing x to its canonically ordered form.

Note that for all of the 4 physical states above, $\det[\mathbf{n} \star \Phi_{\text{phys}}] = 1/2$. We thus recognize that the HFDS provides an exact description of the ground-state with:

$$e^g = \tan \theta \quad (8.18)$$

and the expected limits $g \rightarrow 0, \theta \rightarrow \pi/4$ (uncorrelated limit, 2 electrons in bonding state) and $g \rightarrow 1, \theta \rightarrow \pi/2$ (Heitler-London limit, singlet state with no double occupancy).

B. Gutzwiller wave-function

This is easily generalized to the GWF for an arbitrary number of sites. We still need only a single hidden fermion, but it can now occupy $N/2 + 1$ corresponding to the possible values of the double-occupancy $D = 0, \dots, N/2$. The above expressions generalize to:

$$\Phi = \begin{pmatrix} \boxed{\Phi_{i\sigma,\alpha}^{\text{phys}}} & \boxed{0} \\ 0 & c \\ 0 & ce^{-g} \\ \dots & \dots \\ 0 & ce^{-gN/2} \end{pmatrix} \quad (8.19)$$

So that ($\alpha = 1, \dots, N$):

$$\varphi_\alpha^+ = \sum_{i\sigma} c_{i\sigma}^+ \Phi_{i\sigma,\alpha}^{\text{phys}} \quad (8.20)$$

$$h^+ = c \sum_{d=0}^{N/2} e^{-gd} h_d^+ \quad (8.21)$$

and we impose the constraint, for each configuration, that the hidden fermion is in mode d iff the double occupancy $D(n) = d$. This corresponds to the projector:

$$\hat{P} = \sum_{d=0}^{N/2} \hat{P}_d h_d^+ h_d \quad (8.22)$$

or, using first-quantized numbering $f(n) \in \{0, 1, \dots, N/2\}$ for the location of the hidden fermion:

$$f(n) = \sum_i n_{i\uparrow} n_{i\downarrow} \quad (8.23)$$

The HFDS is now:

$$|\Psi\rangle = \hat{P} \varphi_1^+ \dots \varphi_N^+ h^+ |0\rangle \quad (8.24)$$

and it is easy to check that the HDS amplitude coincides with the Gutzwiller *ansatz*:

$$\psi(n) = \det[\mathbf{n} \star \Phi_{\text{phys}}] ce^{-gD(n)} = \psi_G(n) \quad (8.25)$$

C. General HFDS

$$\Phi = \begin{pmatrix} \Phi_{p\alpha} & \Phi_{p\tilde{\alpha}} \\ \Phi_{\tilde{p}\alpha} & \Phi_{\tilde{p}\tilde{\alpha}} \end{pmatrix} = \begin{pmatrix} \phi_v & \chi_v \\ \phi_h & \chi_h \end{pmatrix}, \quad \dim = \begin{pmatrix} M \times N & M \times \tilde{N} \\ \tilde{M} \times N & \tilde{M} \times \tilde{N} \end{pmatrix} \quad (8.26)$$

In this expression, $\alpha = 1, \dots, N$ runs over physical (or ‘visible’) electrons, $\tilde{\alpha} = 1, \dots, \tilde{N}$ runs over hidden fermions, $p = 1, \dots, M$ runs over physical modes and $\tilde{p} = 1, \dots, \tilde{M}$ runs over hidden modes. We have N physical fermions and \tilde{N} hidden fermions:

$$\varphi_\alpha^+ = \sum_{p=1}^M c_p^+ \Phi_{p\alpha} + \sum_{\tilde{p}=1}^{\tilde{M}} h_{\tilde{p}}^+ \Phi_{\tilde{p}\alpha} \quad (8.27)$$

$$\chi_{\tilde{\alpha}}^+ = \sum_{p=1}^M c_p^+ \Phi_{p\tilde{\alpha}} + \sum_{\tilde{p}=1}^{\tilde{M}} h_{\tilde{p}}^+ \Phi_{\tilde{p}\tilde{\alpha}} \quad (8.28)$$

The hidden fermion configuration is uniquely determined by that of the visible/physical ones:

$$\tilde{n} = F(n) \text{ , } \tilde{x} = f(x) \quad (8.29)$$

The first expression refers to the occupation number in second-quantized notation: $n = \{n_p\}$, $\tilde{n} = \{n_{\tilde{p}}\}$ (recall: $p = i\sigma$), while the second one refers to a configuration in first quantized form (without imposing canonical ordering): $x = (x_1, \dots, x_N)$, $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_{\tilde{N}})$ with e.g. $x_j = (i_j, \sigma_j)$. The function f is thus a fully symmetric function of its arguments.

Denoting by \hat{P}_f the projector associated with this constraint the HFDS wave function reads:

$$|\Psi\rangle = \hat{P}_f \varphi_1^+ \cdots \varphi_N^+ \chi_1^+ \cdots \chi_{\tilde{N}}^+ |0\rangle \quad (8.30)$$

and its amplitude over a given configuration reads:

$$\psi(x) = \det \begin{pmatrix} \phi_v(x) & \chi_v(x) \\ \phi_h[f(x)] & \chi_h[f(x)] \end{pmatrix} = \varepsilon_{\sigma(x)} \det [(n(x), F(n(x)) \star \Phi)] \quad (8.31)$$

The last line expresses that the matrix Φ is ‘sliced’ according to the configuration x . It also makes explicit that the wave-function amplitude is antisymmetric, thanks to the fact that the constraint function $f(x)$ is symmetric (hence exchanging e.g. x_1 and x_2 amounts to exchange the first to lines in the determinant). The permutation $\sigma(x)$ is the one that brings x to its canonically ordered form.

D. Relation to other variational states

We have already seen that the Gutzwiller wave-function is a particular case of HFDS. This can also be shown for:

- Jastrow. Simply choose $\chi_v = \phi_h = 0$, i.e. no off diagonal elements in Φ between the visible and hidden sectors. Then:

$$\psi(x) = \det[\chi_h(f(x))] \det[\phi_v(x)] \quad (8.32)$$

The first term is a fully symmetric Jastrow factor.

- Configuration interaction - see JRM thesis
- Backflow - see JRM thesis

E. NN parametrization

The amplitudes $\phi_v(x), \chi_v(x); \phi_h[f(x)], \chi_h[f(x)]$ are all variational parameters. The hidden-sector components of the determinant are parametrized using a NN. Because we only need the amplitudes $\phi_h[f(x)], \chi_h[f(x)]$ we can directly parametrize those, rather than parametrize explicitly the constraint *which also allows us to not specify explicitly the numbers of modes \tilde{M}* . The NN parametrization, in a sense, deals with a continuous set of modes! We don’t have access to either these modes, or to the constraint function $f(x)$.

For each configuration of the physical/visible fermions x , the entry layer of the network is the bit-string $n(x)$, hence insuring explicit symmetry of the output under a permutation of x . Each row i of the hidden submatrix $[\phi_h(x), \chi_h(x)]$ is parametrized by a distinct NN, which can be chosen as a multilayer perceptron with non-linear activation functions. Here is an example with two layers and a tanh activation:

$$[\phi_1^h[f_i(x)], \dots, \chi_{\tilde{N}}^h[f_i(x)]] = \tanh \left(\tanh \left(\mathbf{n} \cdot \mathbf{w}_i^{(1)} + \mathbf{b}_i^{(1)} \right) \cdot \mathbf{w}_i^{(2)} + \mathbf{b}_i^{(2)} \right) \quad (8.33)$$

(the tanh is taken component by component).

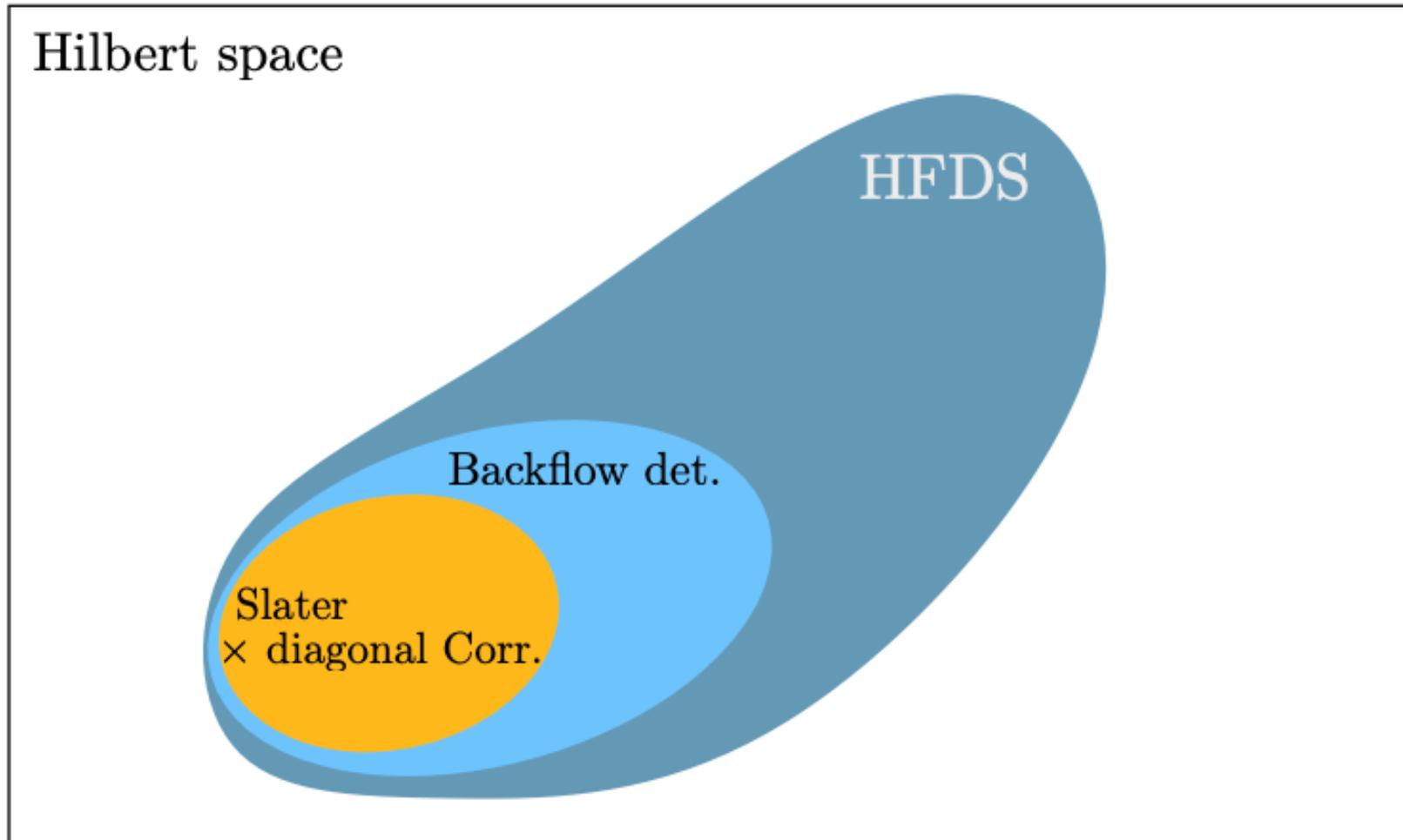
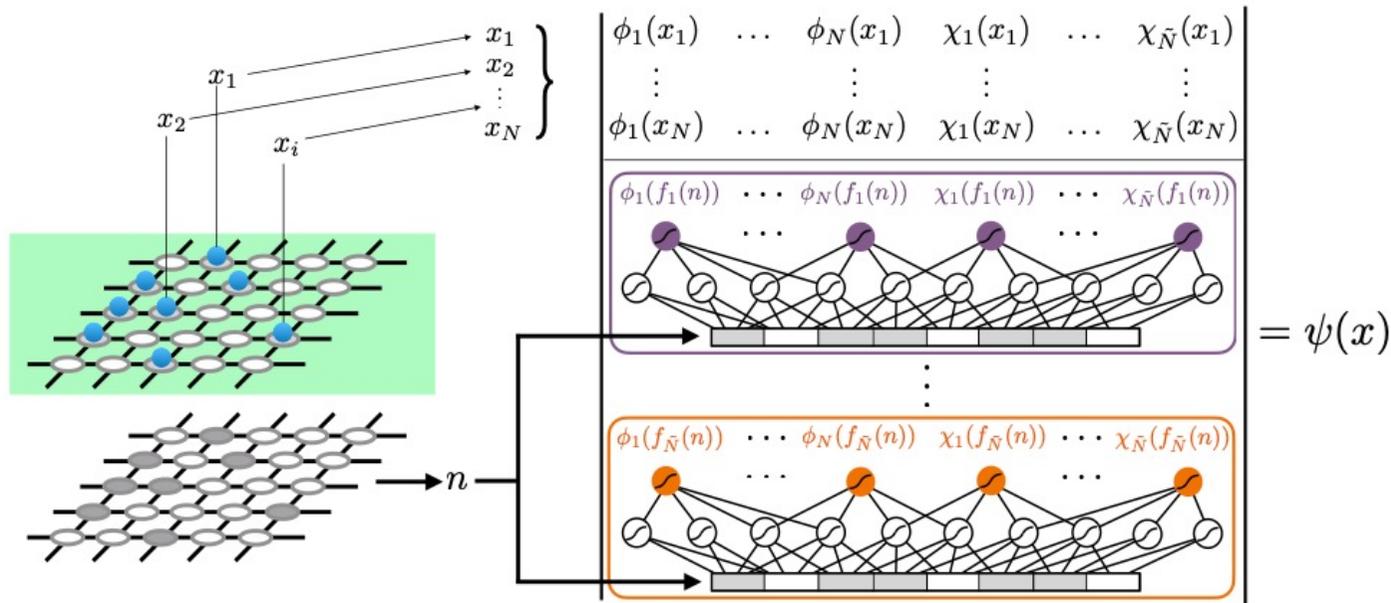


Figure 6.2: Venn diagram of the expressive power of classically tractable (amplitudes can be evaluated in polynomial time) determinant-based trial states, in the Hilbert space spanned by a finite number of basis elements (discrete degrees of freedom). HFDS stands for hidden fermion determinant state, in this case with $\tilde{N} = N$. Backflow det. stands for a single $N \times N$ backflow determinant. Slater \times Corr. is the $N \times N$ Slater determinant multiplied by a diagonal correlation factor.

Figure courtesy Javier Robledo-Moreno

Parametrization of a Hidden Fermion determinantal State by neural networks



Each row of the hidden sector is parametrized by a different perceptron.
For example, with 2 layers:

$$[\phi_1^h[f_i(x)], \dots, \chi_{\tilde{N}}^h[f_i(x)]] = \tanh \left(\tanh \left(\mathbf{n} \cdot \mathbf{w}_i^{(1)} + \mathbf{b}_i^{(1)} \right) \cdot \mathbf{w}_i^{(2)} + \mathbf{b}_i^{(2)} \right)$$

HFDS with physically motivated constraints: benchmark on 4×4 Hubbard model at $\frac{1}{2}$ -filling

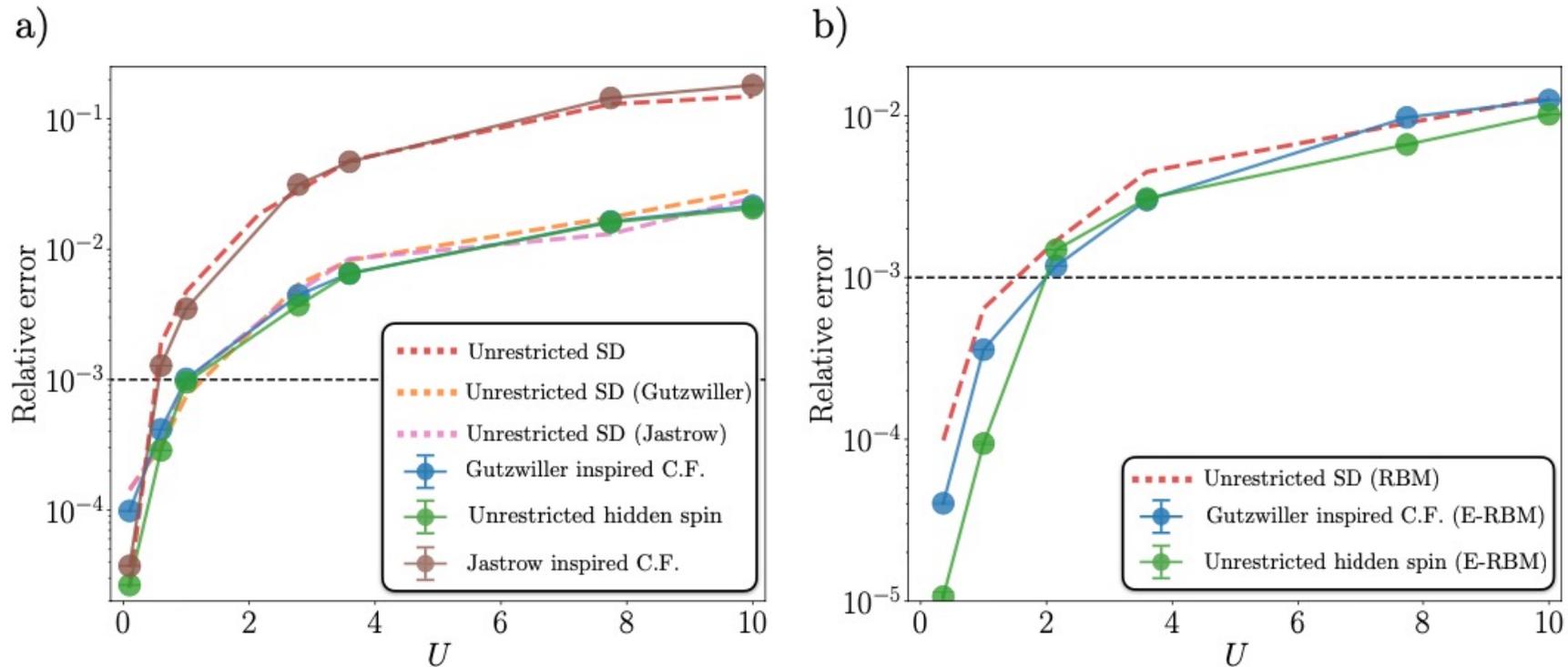


Fig. 1. Benchmarks of physically motivated constraint functions with ED energies in the 4×4 Hubbard model at $n = 1/2$ average physical site occupation. Results from standard wave function *ansätze* are shown as dashed lines for comparison purposes. (a) Relative error in the ground-state energy as a function of the coupling constant U . The different constraint function *ansätze* are a single Slater determinant in the augmented Fock space with no projections. (b) Same as panel (a) including a complex RBM projection factor both in the control unrestricted HF *ansatz* and a E-RBM factor in the the hidden fermion *ansätze*.

Numerical results with NN parametrization: 4*4 Hubbard model at 1/2-filling

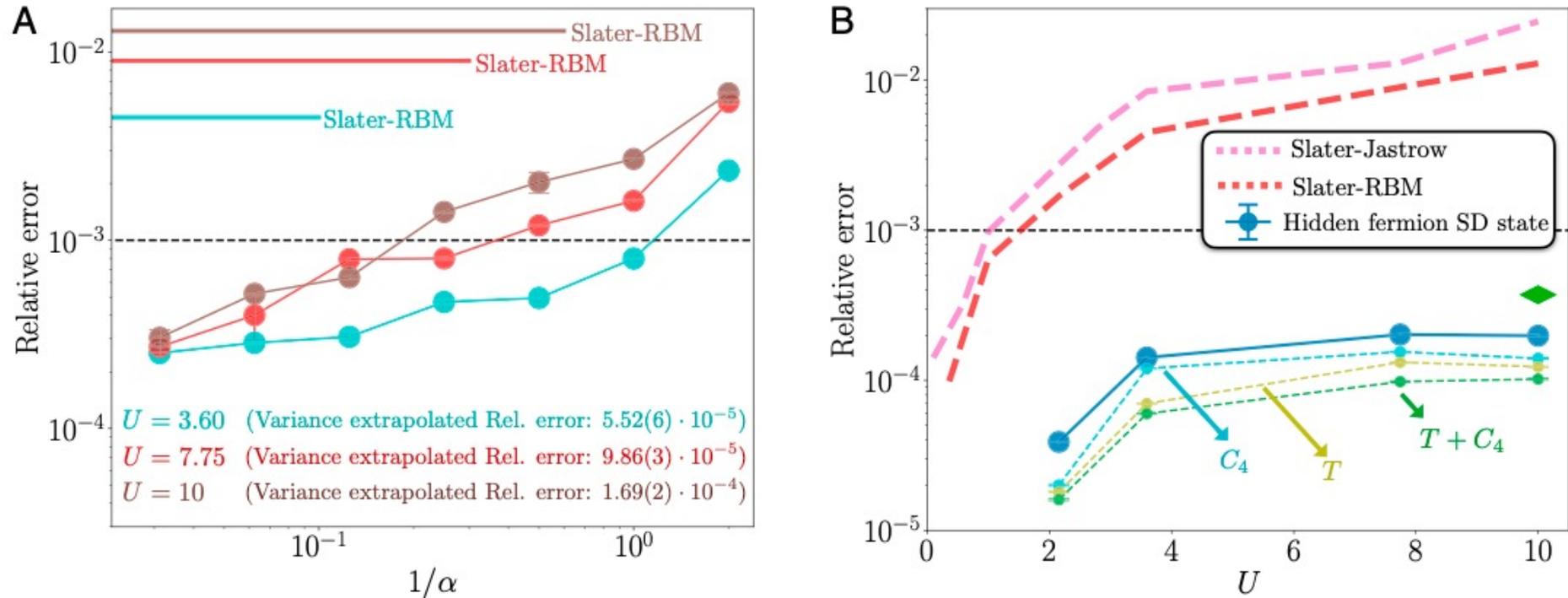


Fig. 3. Exact diagonalization benchmarks of the ground-state energy in the 4×4 lattice with periodic boundary conditions. (A) Relative error in the ground-state energy as a function of the inverse of the width density α of the single-hidden-layer neural networks parameterizing the rows of the hidden submatrix. Average physical site occupancy is $n = 1/2$ and $\tilde{N} = 8$. Different values of U are considered, as indicated by each color. The error for a Slater-RBM ansatz (main text) with hidden neuron density $\alpha = 32$, at the same values of U , is included for comparison. Indicated is also the relative error from the variance-extrapolated energy for each value of U (see *SI Appendix* for details). (B) Relative error in the ground-state energy as a function of the coupling constant U , at $n = 5/8$ average site occupancy (first closed shell) and $\tilde{N} = 10$. The rows of the hidden submatrix are given by single-hidden-layer neural networks with $\alpha = 64$. The errors from Slater-Jastrow and Slater-RBM ansätze are included for comparison. The green diamond is the relative error found with the state-of-the-art, tensor-network-based ansatz from ref. 46. Shown is also the relative error according to the projection of the converged hidden-fermion determinant state to the subspace of invariant wave functions under the action of $\pi/2$ rotations (C_4) and the group of all possible translations T with $K = 0$ momentum, separately and together.

Relative error as a function of hidden fermion number and depth of NN (4*4 Hubbard model at 1/2 filling)

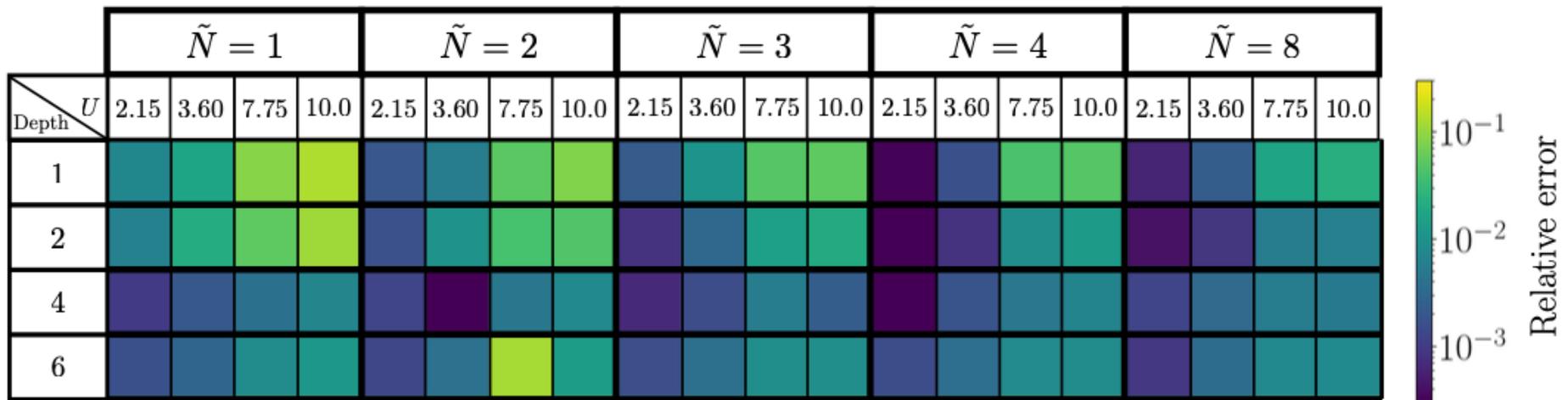


Fig. 2. Effect of the number of hidden fermions and depth of the fully connected neural network that parametrizes the hidden sub-matrix in the expressive power of the hidden fermion determinant *ansatz*. The scale shows the relative error in the ground-state energy for different values of U given N_h and the neural network depth. The results correspond to the 4×4 Hubbard model at $n = 1/2$ filling.

NN-HFDS: Stripe order in the Hubbard model at 1/8 doping on $4 \times L$ cylinders (up to $L=16$)

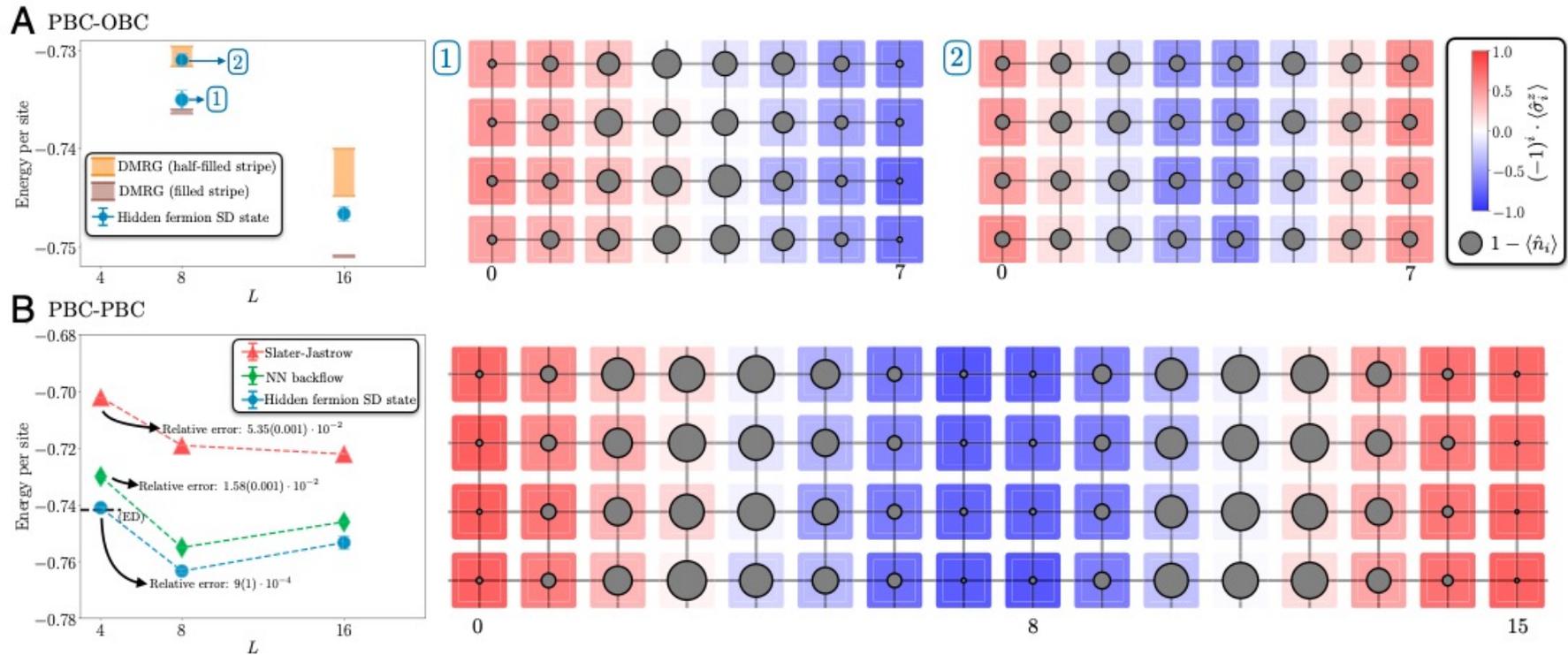


Fig. 4. Energy per site and competing charge and spin orders in the $4 \times L$ rectangular lattice at 1/8 hole doping ($n = 0.875$) and $U = 8$. (A) Periodic boundary conditions on the short side of the cylinder and open on the long side (PBC-OBC). *Left panel* compares the hidden-fermion determinant-state energies with DMRG energies. The width of the DMRG symbols shows the range of converged variational energies for different bond dimensions used in ref. 48. For $L = 8$, blue points labeled as 1 and 2 correspond to filled and half-filled stripes. *Right panel* shows the hole and staggered spin distribution for both metastable configurations. The diameter of the gray circles is proportional to the hole density. (B) Periodic boundary conditions along both sides of the rectangles (PBC-PBC). *Left panel* compares the hidden-fermion determinant-state energies with the Slater-Jastrow and neural-network backflow ansätze (from ref. 9). The dashed horizontal line marks the ED (4×4 with PBCs from ref. 51) energy. In the 4×4 lattice the relative error in the ground-state energy is displayed for each ansatz. *Right panel* shows the hole and staggered spin distributions in the 4×16 lattice.