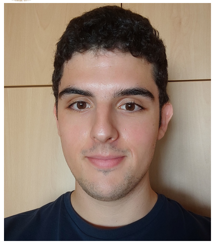
Neural Quantum States for Quantum Many-Body Physics

JWT Keeble

M. Drissi



J. Rozalén Sarmiento

A. Rojo-Francàs

B. Julià-Díaz

Dr Arnau Rios Huguet

Institute of Cosmos Sciences

Universitat de Barcelona

ECT* Workshop

18 April 2024

Departament de Física Quàntica i Astrofísica

@HadNucAtUB

Outlook

- **Nuclear physics**

- Neural Quantum States

- Deuteron

  JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
  J Rozalen-Sarmiento, JWT Keeble & A. Rios, Eur. Phys. J. Plus **139**, 189 (2024)

- 1D spinless (polarised) systems (A>2)

  J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

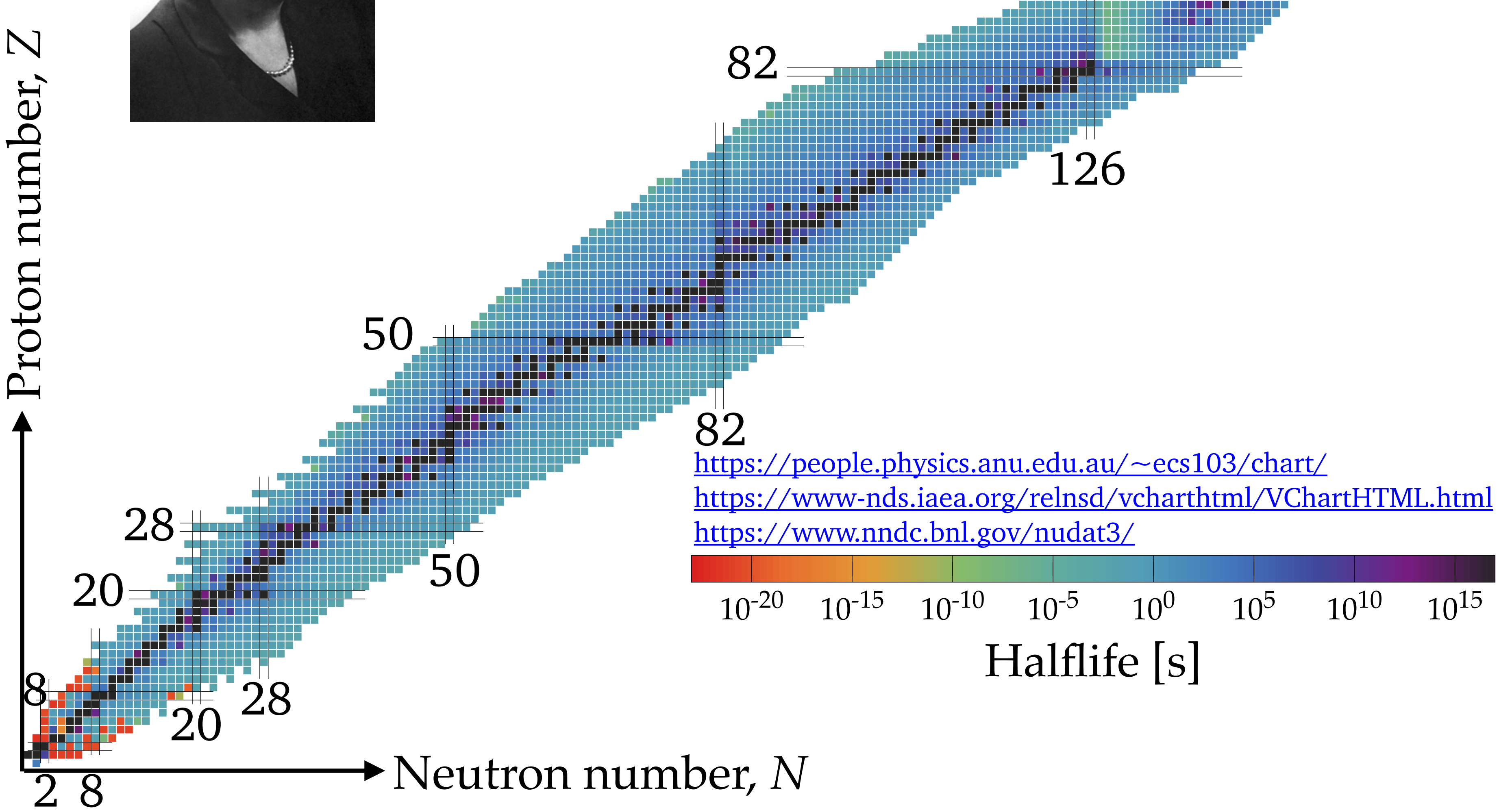
- Optimisation strategy

  M. Drissi, J.W.T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Isotope chart

Maria Goeppert Mayer

Nobel Prize (1963)



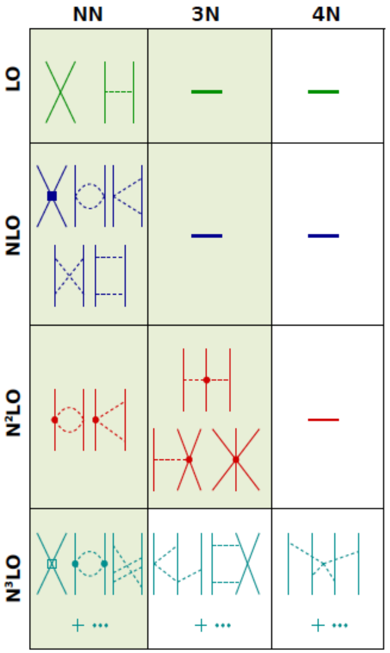
https://people.physics.anu.edu.au/~ecs103/chart/

https://www-nds.iaea.org/relnsd/vcharthtml/VChartHTML.html

https://www.nndc.bnl.gov/nudat3/

Ab initio for isotopes

χ-EFT interactions

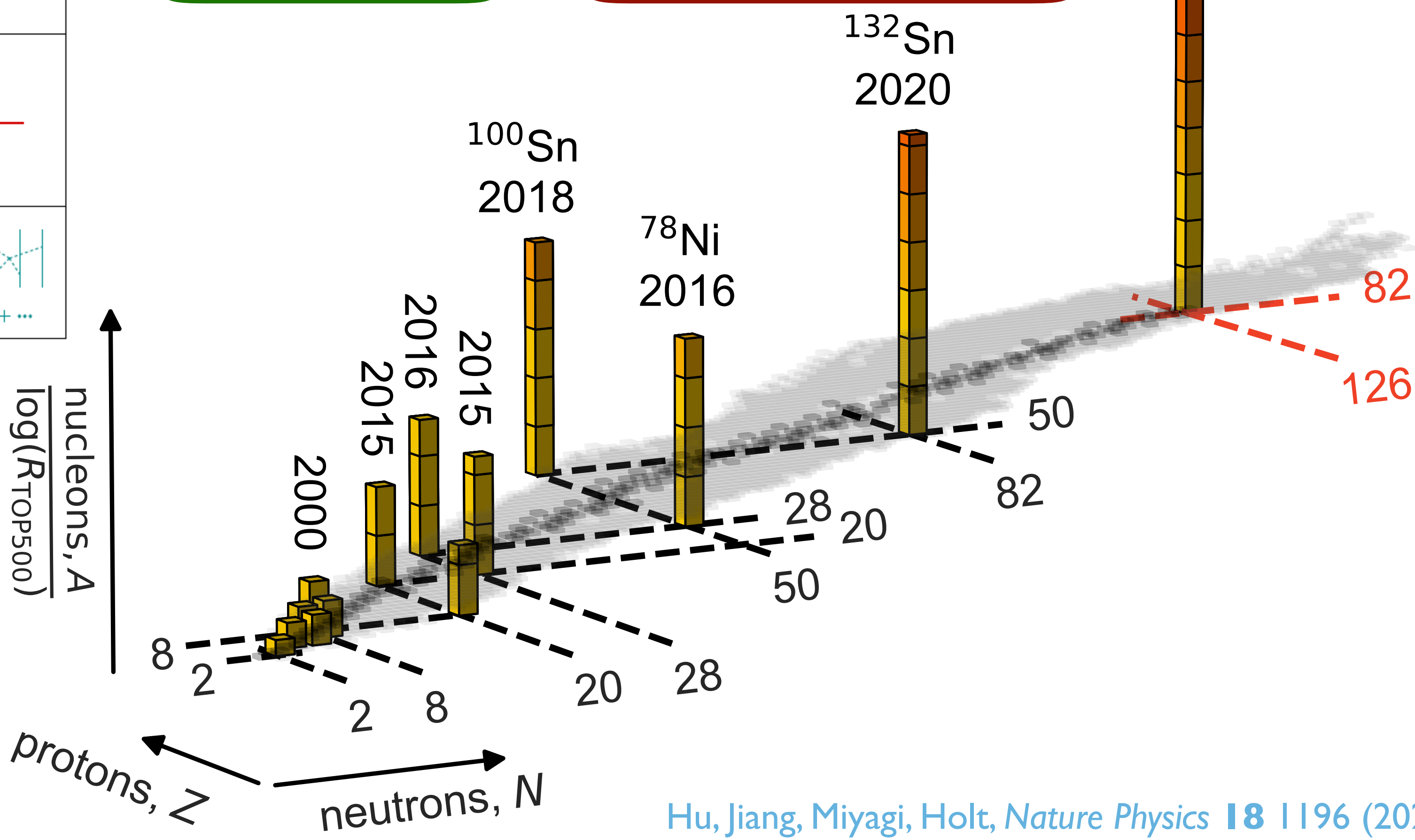


Renormalisation
SRG

$$\frac{dH_s}{ds} = [\eta_s, H_s]$$

Many-body methods

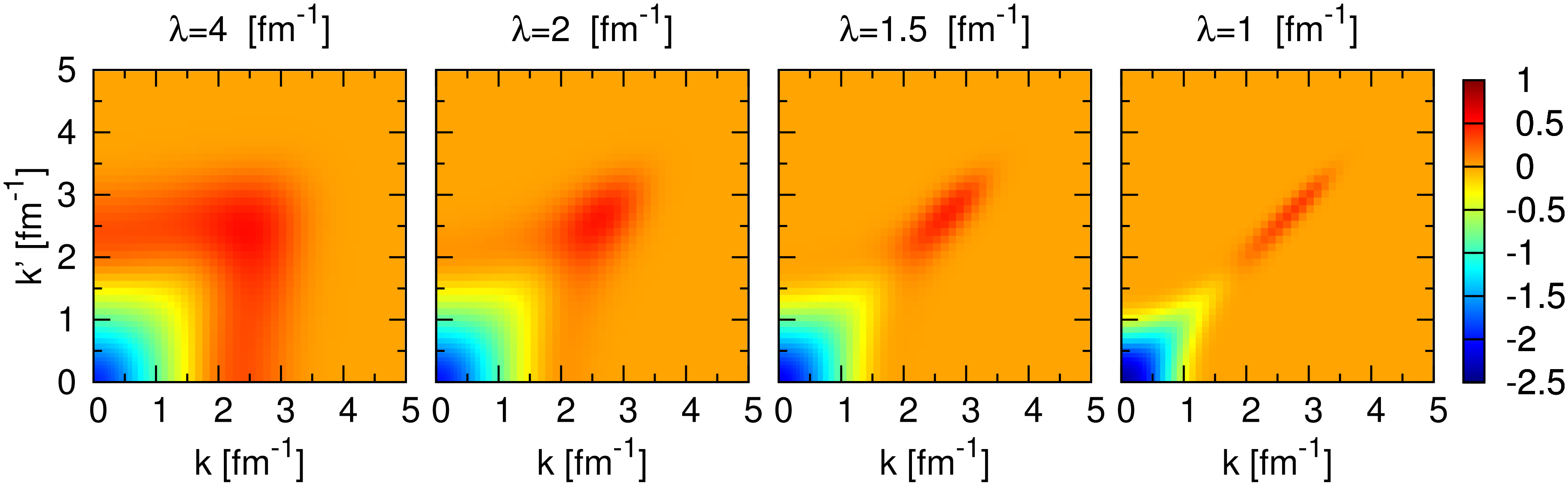
- No Core Shell Model
- Coupled Cluster
- Green's functions
- Neural quantum states?



Hu, Jiang, Miyagi, Holt, *Nature Physics* **18** 1196 (2022)
Comment: Rios, *A historic match for nuclei and neutron stars, Nature Physics* **18** 1137 (2022)

Similarity Renormalization Group

$^1S_0$ NN matrix elements from N3LO

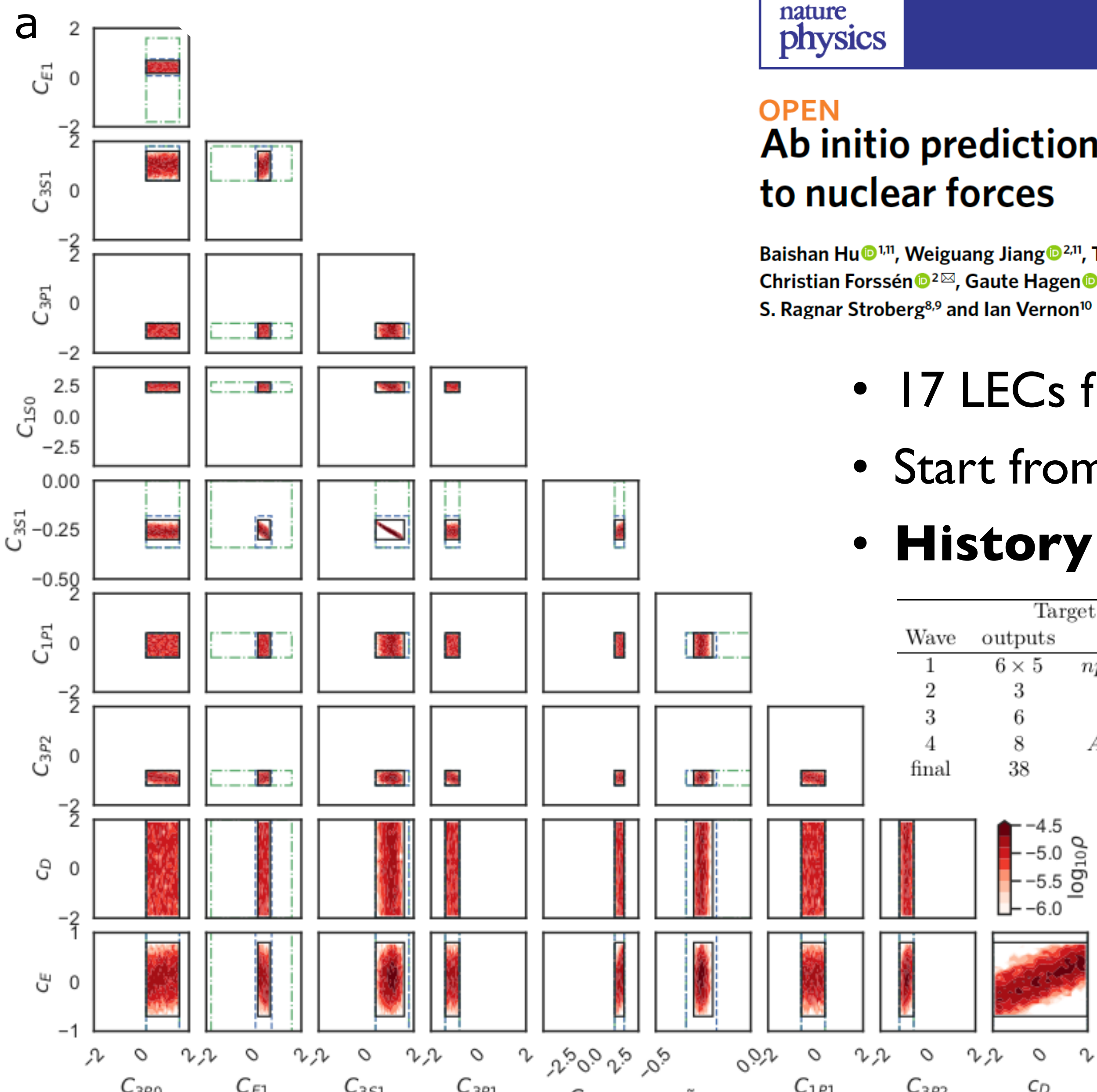


$$H_s = U(s) H U^\dagger(s) \qquad \frac{dH_s}{ds} = [\eta_s, H_s] = [[T_{\rm rel}, H_s], H_s] \qquad \lambda = s^{-1/4}$$

- Series of unitary transformation
- Observables unaltered at 2-body level, but…
- Induces 3-, 4- and up to A-body forces
- If these can be treated perturbatively, calculation is easier
- **Fixed point?**

Bogner, Furnstahl, Schwenk, *Prog. Part. Nucl. Phys.* **65** 94 (2010)

History matching



nature physics
ARTICLES
https://doi.org/10.1038/s41567-022-01715-8
Check for updates
OPEN

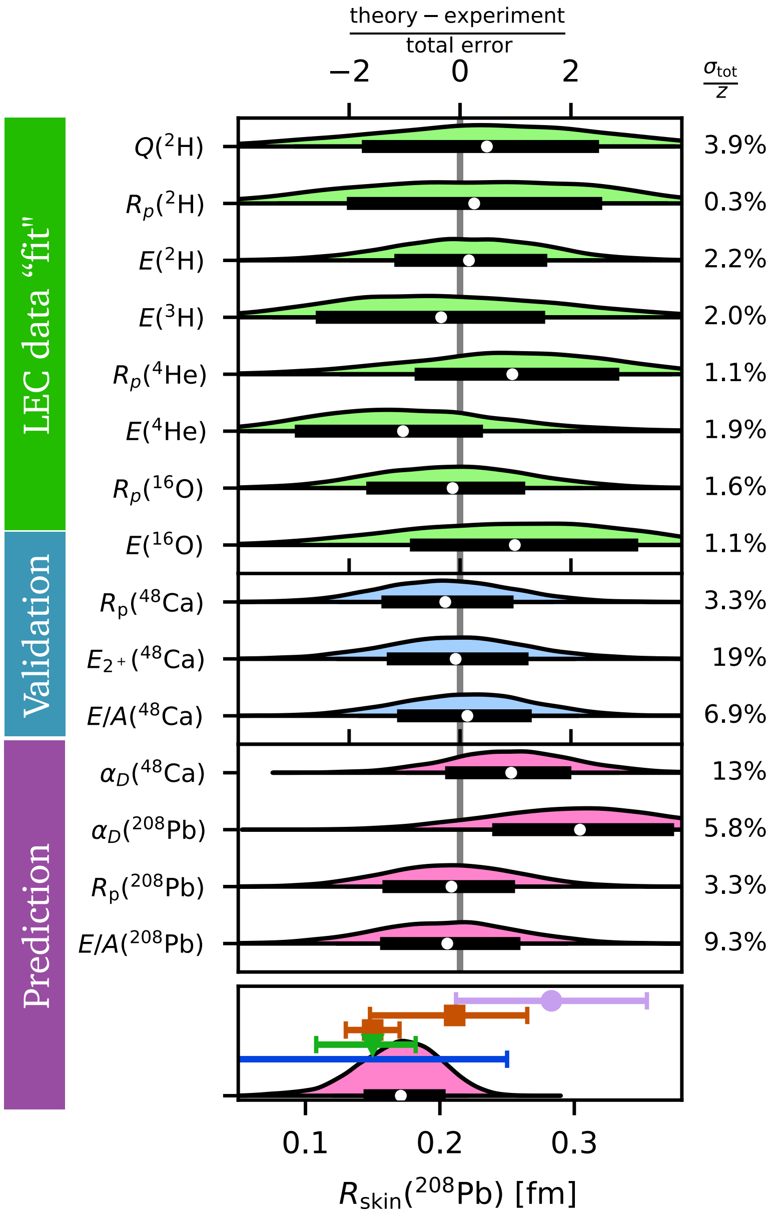
Ab initio predictions link the neutron skin of ^{208}Pb to nuclear forces

Baishan Hu[1,11], Weiguang Jiang[2,11], Takayuki Miyagi[1,3,4,11], Zhonghao Sun[5,6,11], Andreas Ekström[2], Christian Forssén[2], Gaute Hagen[1,5,6], Jason D. Holt[1,7], Thomas Papenbrock[5,6], S. Ragnar Stroberg[8,9] and Ian Vernon[10]

- 17 LECs from χEFT
- Start from **10^9** realisations
- **History matching** reduces to **34**

Wave	Target set $\mathcal{Z}$ outputs	Target set $\mathcal{Z}$ systems	Active inputs	Input samples	NI fraction	Proportion space NI
1	6×5	np scattering	5–7	10^5–10^7	10^{-1}–10^{-4}	$3.2 \cdot 10^{-6}$
2	3	$A = 2$	7	10^8	$2.3 \cdot 10^{-4}$	$3.0 \cdot 10^{-7}$
3	6	$A = 2$–4	13	10^8	$3.4 \cdot 10^{-5}$	$9.2 \cdot 10^{-8}$
4	8	$A = 2$–4, 16	17	$5 \cdot 10^8$	$8.6 \cdot 10^{-6}$	same
final	38	all above	17	4,337	$7.8 \cdot 10^{-3}$	same

History matching



nature physics

ARTICLES

https://doi.org/10.1038/s41567-022-01715-8

Check for updates

OPEN

Ab initio predictions link the neutron skin of ^{208}Pb to nuclear forces

Baishan Hu[1,11], Weiguang Jiang[2,11], Takayuki Miyagi[1,3,4,11], Zhonghao Sun[5,6,11], Andreas Ekström[2], Christian Forssén[2]✉, Gaute Hagen[1,5,6], Jason D. Holt[1,7], Thomas Papenbrock[5,6], S. Ragnar Stroberg[8,9] and Ian Vernon[10]

- 17 LECs from χEFT
- Start from **10^9** realisations
- **History matching** reduces to **34**

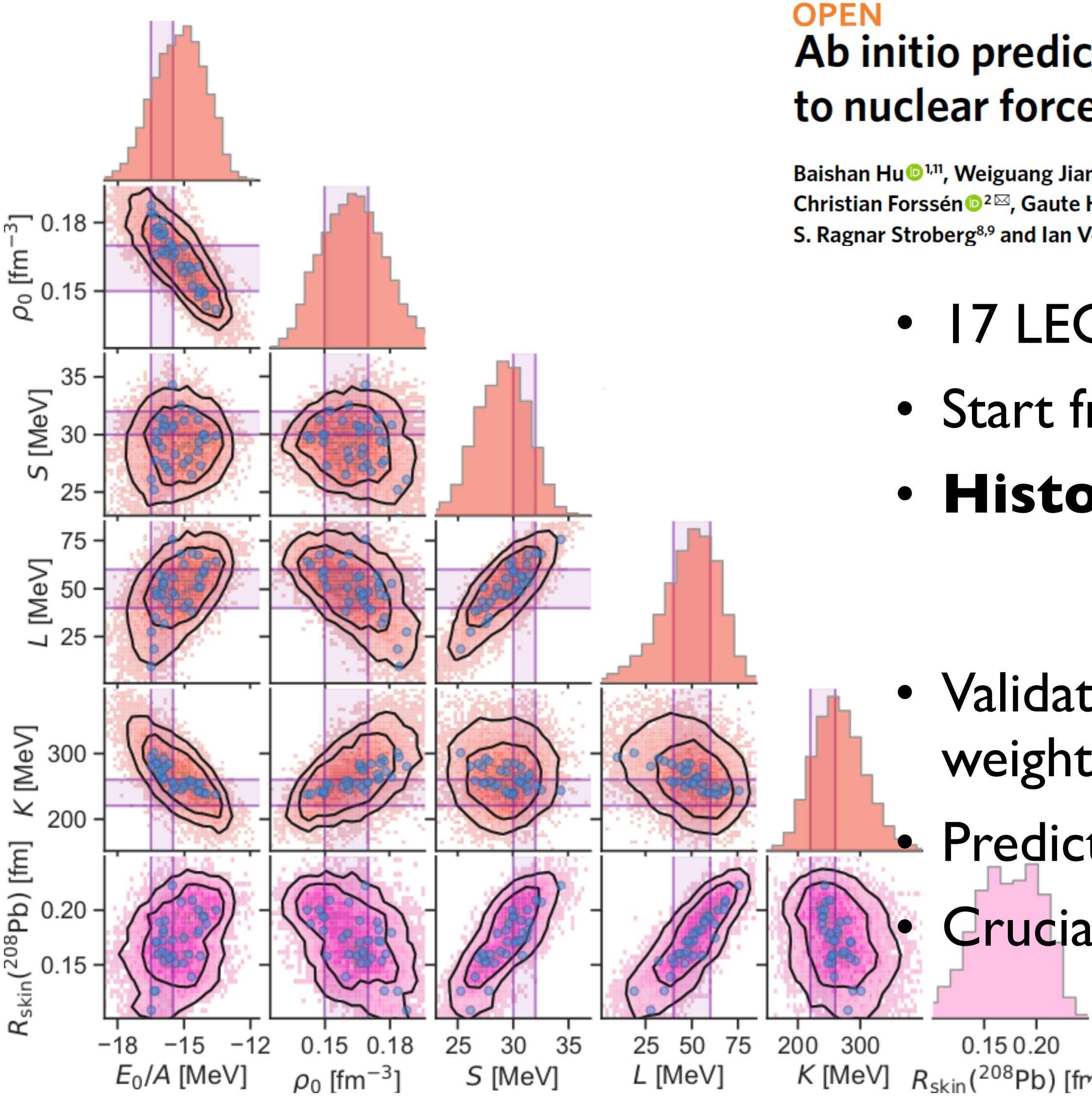
- Validate data in ^{48}Ca (importance weights)
- Predict ^{208}Pb & nuclear matter
- Crucial role of **emulators** (GPs)

History matching

nature physics
ARTICLES
https://doi.org/10.1038/s41567-022-01715-8
Check for updates
OPEN

Ab initio predictions link the neutron skin of ^{208}Pb to nuclear forces

Baishan Hu[1,11], Weiguang Jiang[2,11], Takayuki Miyagi[1,3,4,11], Zhonghao Sun[5,6,11], Andreas Ekström[2], Christian Forssén[2]✉, Gaute Hagen[1,5,6], Jason D. Holt[1,7], Thomas Papenbrock[5,6], S. Ragnar Stroberg[8,9] and Ian Vernon[10]



- 17 LECs from χEFT
- Start from **10^9** realisations
- **History matching** reduces to **34**

- Validate data in ^{48}Ca (importance weights)
- Predict ^{208}Pb & nuclear matter
- Crucial role of **emulators** (GPs)

Outlook

- Nuclear physics

- **Neural Quantum States**

- Deuteron

  JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
  J Rozalen-Sarmiento, JWT Keeble & A. Rios, Eur. Phys. J. Plus **139**, 189 (2024)

- 1D spinless (polarised) systems (A>2)

  J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

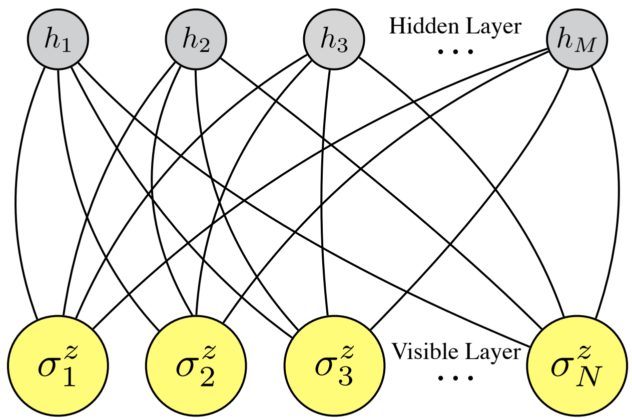
- Optimisation strategy

  M. Drissi, J.W.T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Neural quantum states

$$E_\theta = \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle}$$

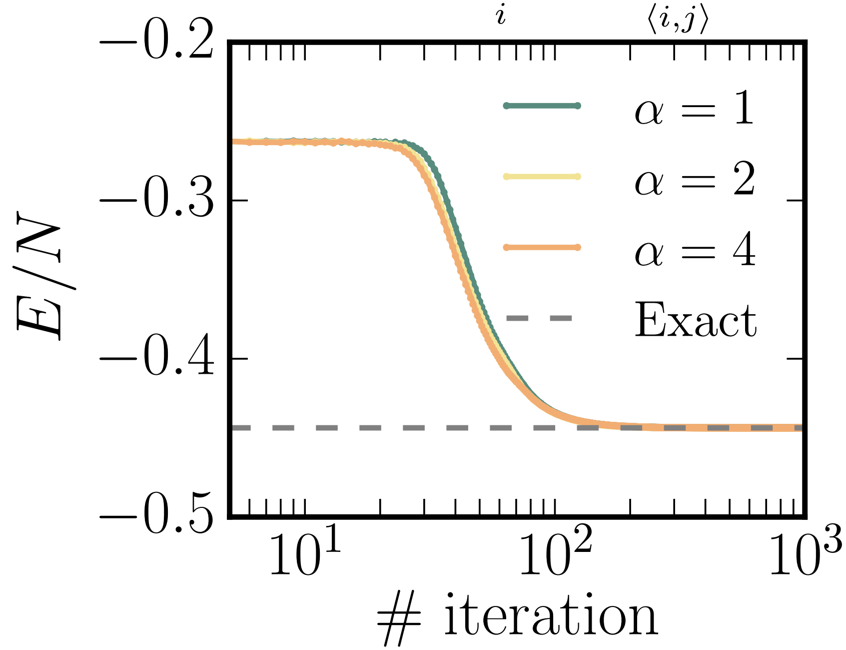
$$|\psi_\theta\rangle =$$



G. Carleo & M. Troyer, Science ***355****, 602 (2017)*
H. Saito, J. Phys. Soc. Japan ***86****, 093001 (2017)*
H. Saito, J. Phys. Soc. Japan ***87****, 074002 (2018)*

1D Heisenberg model

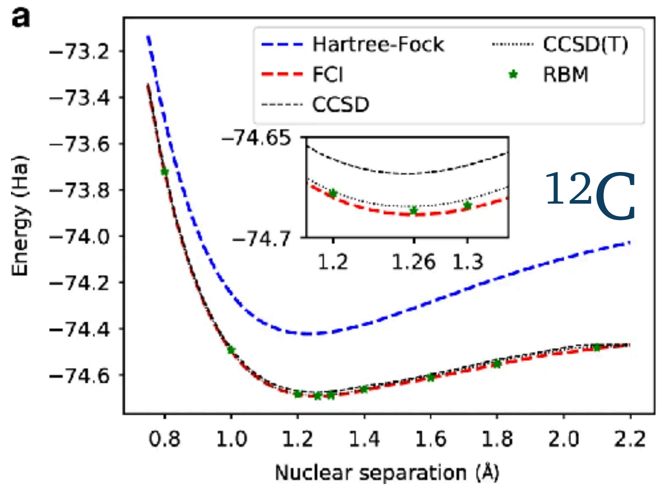
$$\mathcal{H}_{\mathrm{TFI}} = -h \sum_i \sigma_i^x - \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z$$



- **Machine learning** techniques to solve the many-body problem
- **Variational Artificial Neural Networks** (now NQS)
  - ANNs **efficiently represent** high-dimensional data (coarse graining?)
  - ANNs are setup for **variational** problems
  - ANNs can be used in **dynamics**

NQSs in quantum chemistry

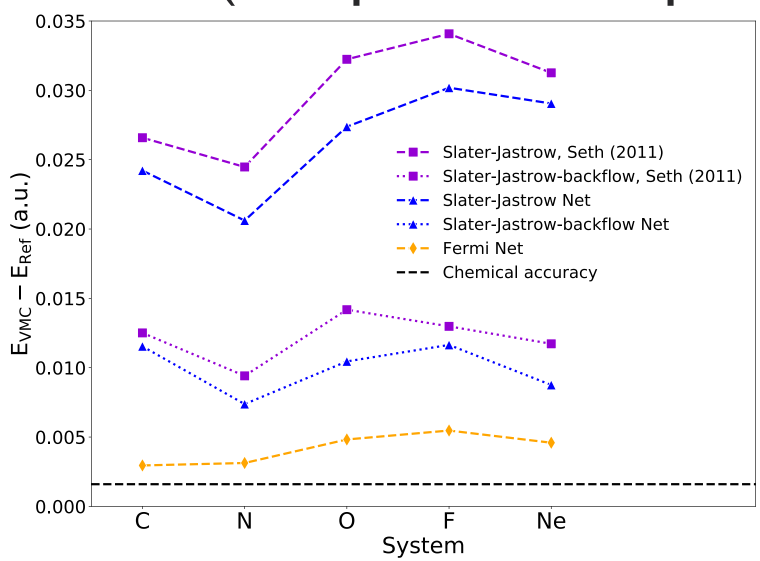
NetKet framework (Zurich & Flatiron)



Choo, Carleo et al *Nat Comm* **11** 2368 (2019)

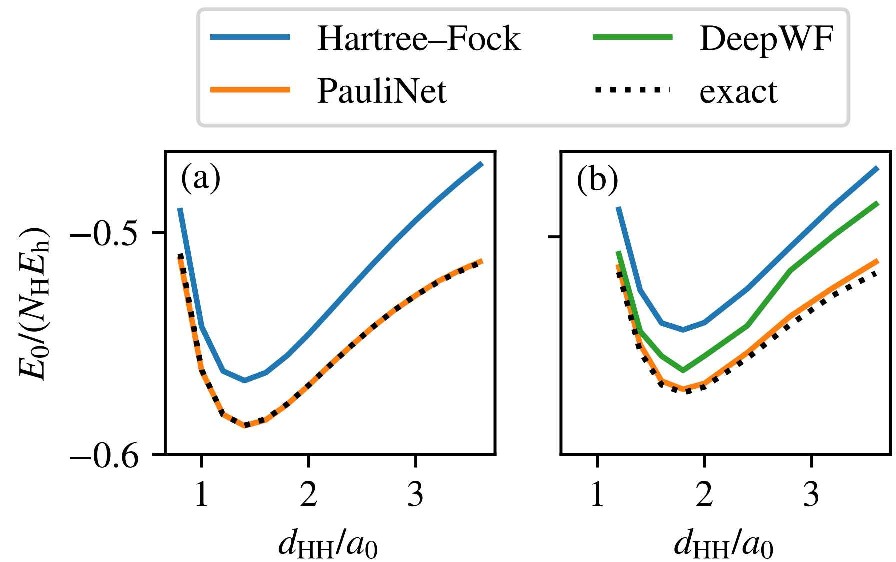
https://www.netket.org/

FermiNet (DeepMind & Imperial)



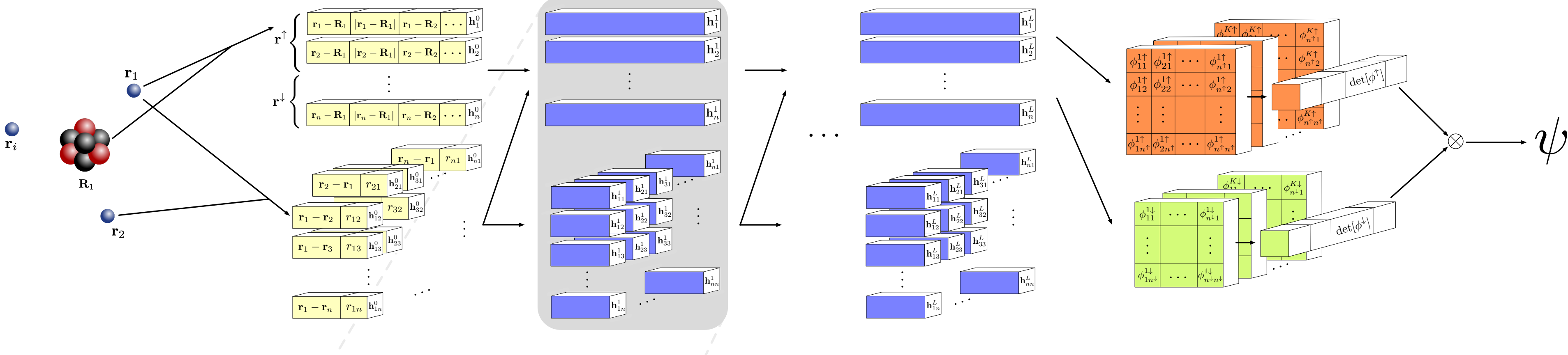
Pfau et al *Phys Rev Res* **2** 033429 (2020)

PauliNet (FU & TU Berlin)



Hermann et al *Nat Chem* **12** 891 (2020)

FermiNet



Input: N-particle positions: $\mathbf{r}_i - \mathbf{R}$ & $\mathbf{r}_i - \mathbf{r}_j$

Network:

1) Single-particle orbitals (with **backflow**)

$$\phi_i(\mathbf{r}_j) \rightarrow \phi_i(\mathbf{r}_j; \mathbf{r}_1, \ldots, \mathbf{r}_{j-1}, \mathbf{r}_{j+1}, \ldots, \mathbf{r}_n) = \phi_i(\mathbf{r}_j; \{\mathbf{r}_{/j}\})$$

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

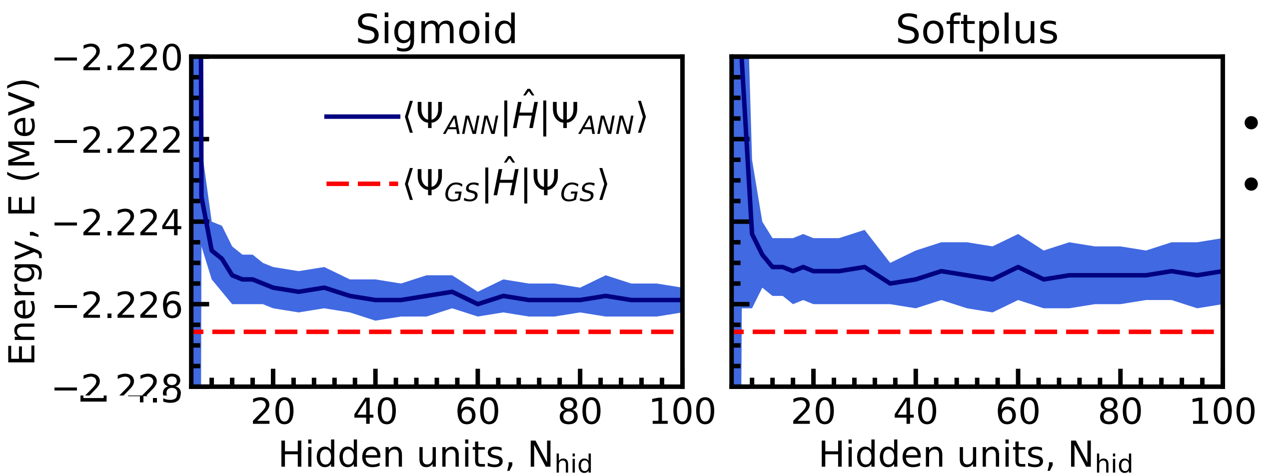
2) N-body wavefunction

$$\psi(\mathbf{r}_1^{\uparrow}, \ldots, \mathbf{r}_{n\downarrow}^{\downarrow}) = \sum_k \omega_k \left(\det\left[\phi_i^{k\uparrow}(\mathbf{r}_j^{\uparrow}; \{\mathbf{r}_{/j}^{\uparrow}\}; \{\mathbf{r}^{\downarrow}\})\right] \; \det\left[\phi_i^{k\downarrow}(\mathbf{r}_j^{\downarrow}; \{\mathbf{r}_{/j}^{\downarrow}\}); \{\mathbf{r}^{\uparrow}\};\right]\right)$$

Pfau et al *Phys Rev Res* **2** 033429 (2020)

Nuclear Physics

JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020) arxiv:1911.13092
J Rozalén, JWT Keeble & A Rios, EPJPlus arXiv:2205.12795

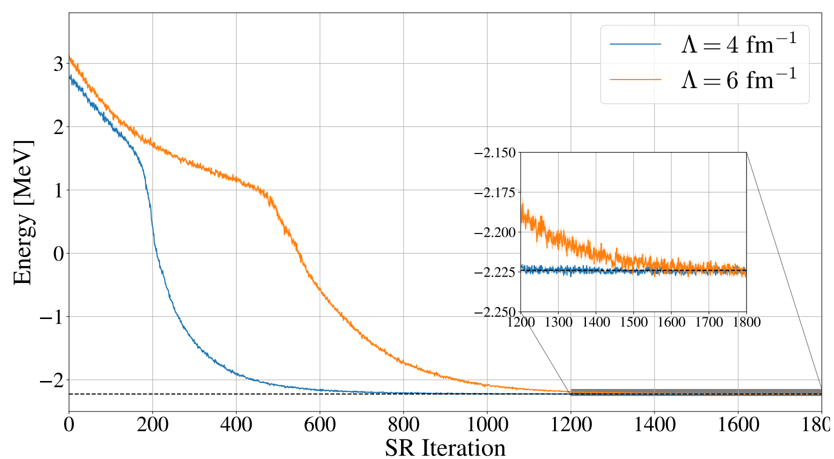


- Momentum basis
- EM N3LO

- **A<16, Neutron matter**

Adams, Carleo, **Lovato** & Rocco, Phys. Rev. Lett. **127** 022502 (2021) arXiv:2007.14282
Gnech, Adams, **Lovato** et al, Few-Body Systems **63** 7 (2022) arXiv:2108.06836
Lovato, Adams, Carleo & Rocco, Phys. Rev. Research **4** 043178 (2022) arXiv:2206.10021

Argonne NATIONAL LABORATORY



- π-less EFT
- Gaussian int
- Spin?

Other systems

FONDAZIONE BRUNO KESSLER

ECT*
EUROPEAN CENTRE
FOR THEORETICAL STUDIES
IN NUCLEAR PHYSICS AND RELATED AREAS

Many-body quantum physics with machine learning

4–8 Sept 2023
ECT*
Europe/Rome timezone

Enter your search term

https://indico.ectstar.eu/event/177/

Outlook

- Nuclear physics

- Neural Quantum States

- **Deuteron**

  JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
  J Rozalen-Sarmiento, JWT Keeble & A. Rios, Eur. Phys. J. Plus **139**, 189 (2024)

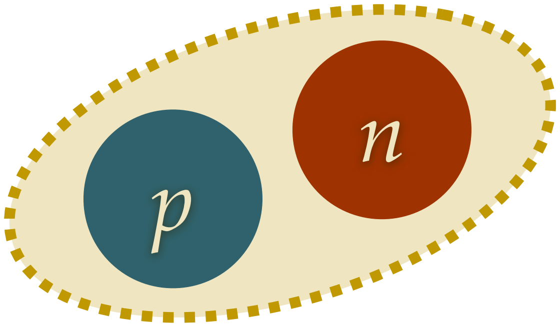
- 1D spinless (polarised) systems (A>2)

  J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

- Optimisation strategy

  M. Drissi, J.W.T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Why the deuteron?



- **Simplest** nuclear (np) bound state
- Experimental info **well-known**
- **Computationally** inexpensive benchmark
- Already requires 2 **states coupling** $^3SD_1$
- Solved in **momentum** space

$$\hat{H} = \hat{H}_{CM} + \hat{H}_r$$

$$\hat{H}_r \psi(q) = E_d \psi(q)$$

$$\hat{H}_r = \frac{q^2}{2\mu} + V(q, q')$$

Deuteron

E_d=2.224575(9) MeV

r_d=1.97535(85) fm

Q=0.2859(3) fm2

JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
arxiv:1911.13092

PYTORCH

- **Scalability** of Neural Networks
- Direct calculation of **gradients** via **Autograd**
- Portability to multiple GPUs

Why the deuteron?

- **Simplest** nuclear (np) bound state
- Experimental info **well-known**
- **Computationally** inexpensive benchmark
- Already requires 2 **states coupling** $^3SD_1$
- Solved in **momentum** space

$$\hat{H} = \hat{H}_{CM} + \hat{H}_r$$

$$\hat{H}_r \psi(q) = E_d \psi(q)$$

$$\hat{H}_r = \frac{q^2}{2\mu} + V(q, q')$$

Deuteron

E_d=2.224575(9) MeV

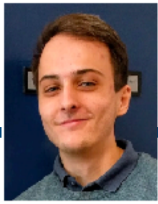
r_d=1.97535(85) fm

Q=0.2859(3) fm2

JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
arxiv:1911.13092

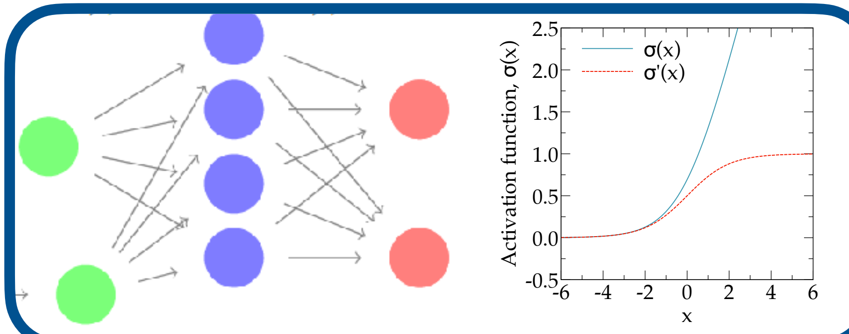
PYTORCH

- **Scalability** of Neural Networks
- Direct calculation of **gradients** via **Autograd**
- Portability to multiple GPUs

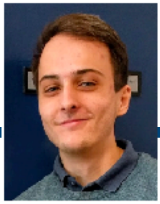


NQSs three-step process

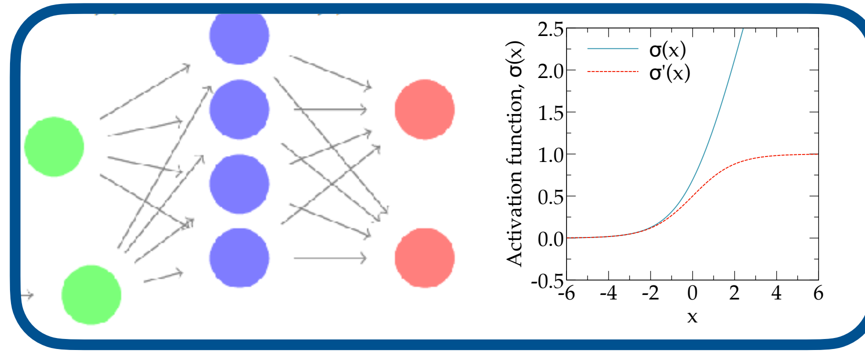
1. Network architecture & initialisation



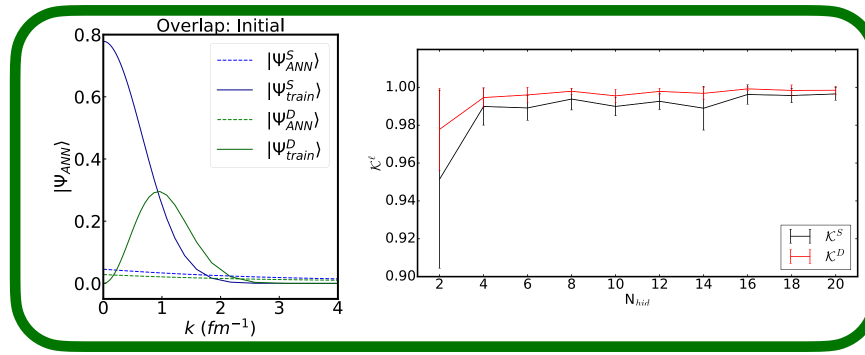
NQSs three-step process

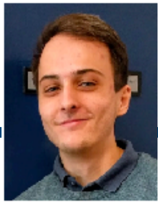


1. Network architecture & initialisation



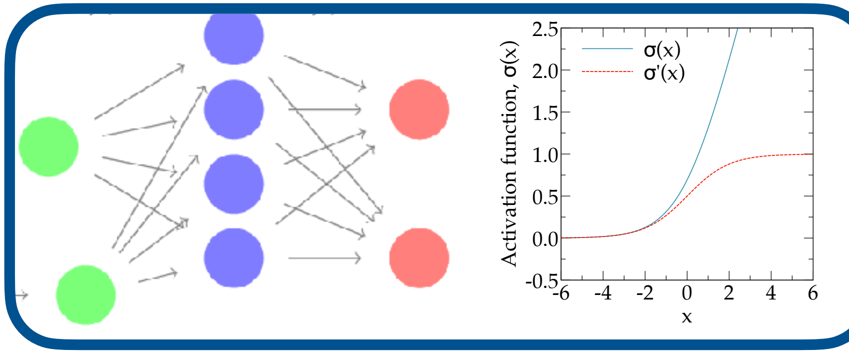
2. Network pre-training



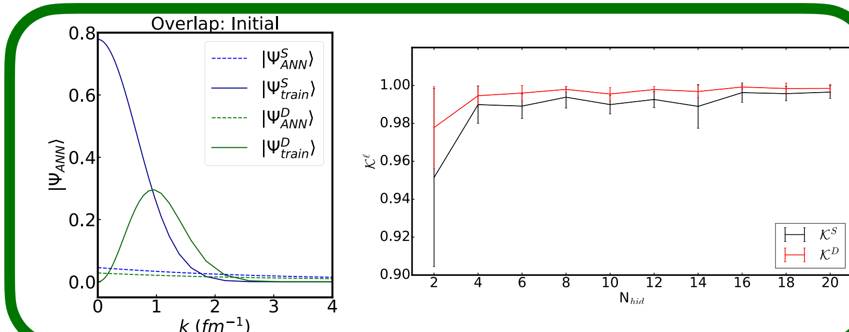


NQSs three-step process

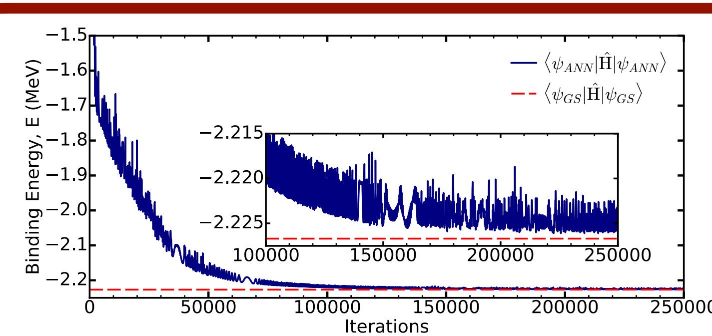
1. Network architecture & initialisation



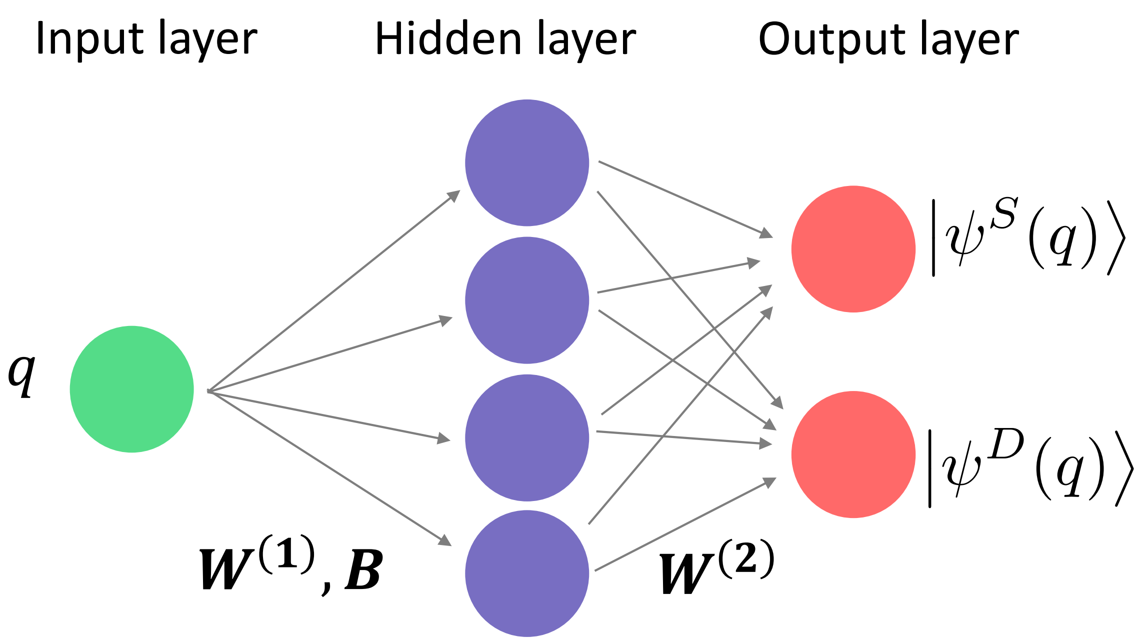
2. Network pre-training



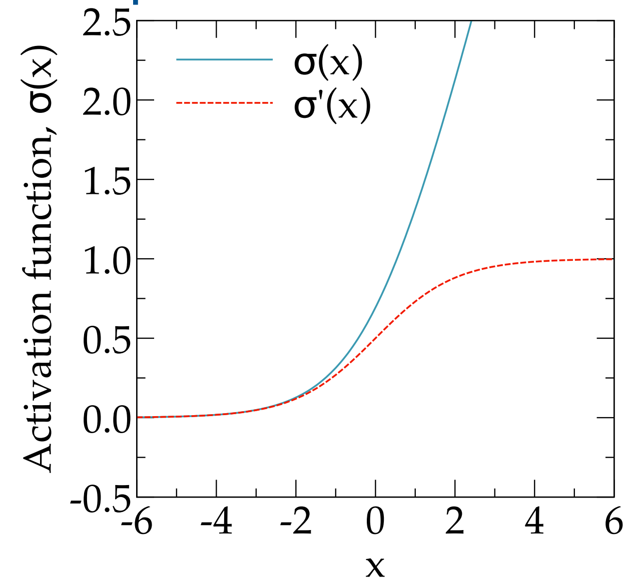
3. Energy minimisation



Network architecture & initialisation



Softplus $\sigma(x) = \ln(1+e^x)$



Wavefunction ansatz

$$\psi^L_{\mathrm{ANN}}(q) = \sum_{i=1}^{N_{hid}} \mathcal{W}^{(2)}_{i,L}\, \sigma\left(\mathcal{W}^{(1)}_i q + b_i\right)$$

# of parameters

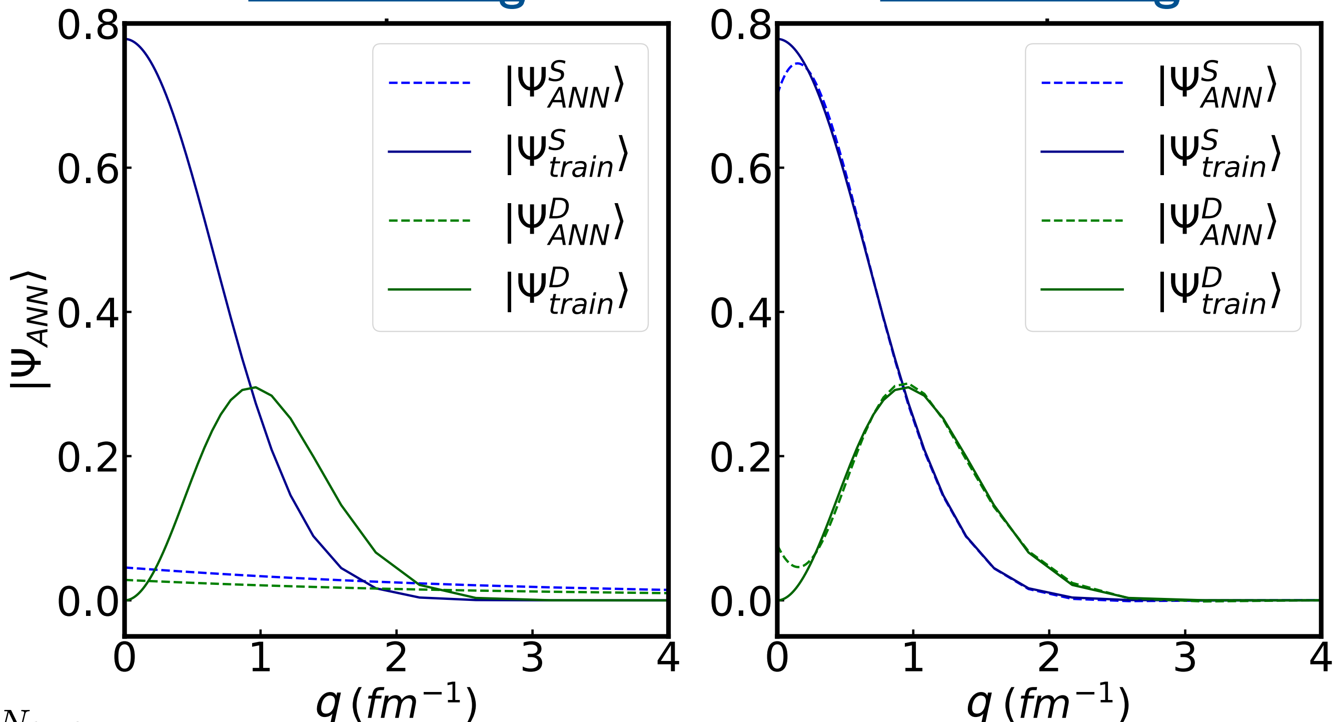
$$N_{\mathcal{W}} = 4 \times N_{\mathrm{hid}}$$

- Hidden layer provides **representability** (UAT)
- Architecture is **minimal**

JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
arxiv:1911.13092

Step 2: Network training

Pre-training | Post-training



$$\psi^L_{\mathrm{ANN}}(q) = \sum_{i=1}^{N_{hid}} \mathcal{W}^{(2)}_{i,L}\, \sigma\left(\mathcal{W}^{(1)}_i q + b_i\right)$$

Initialisation

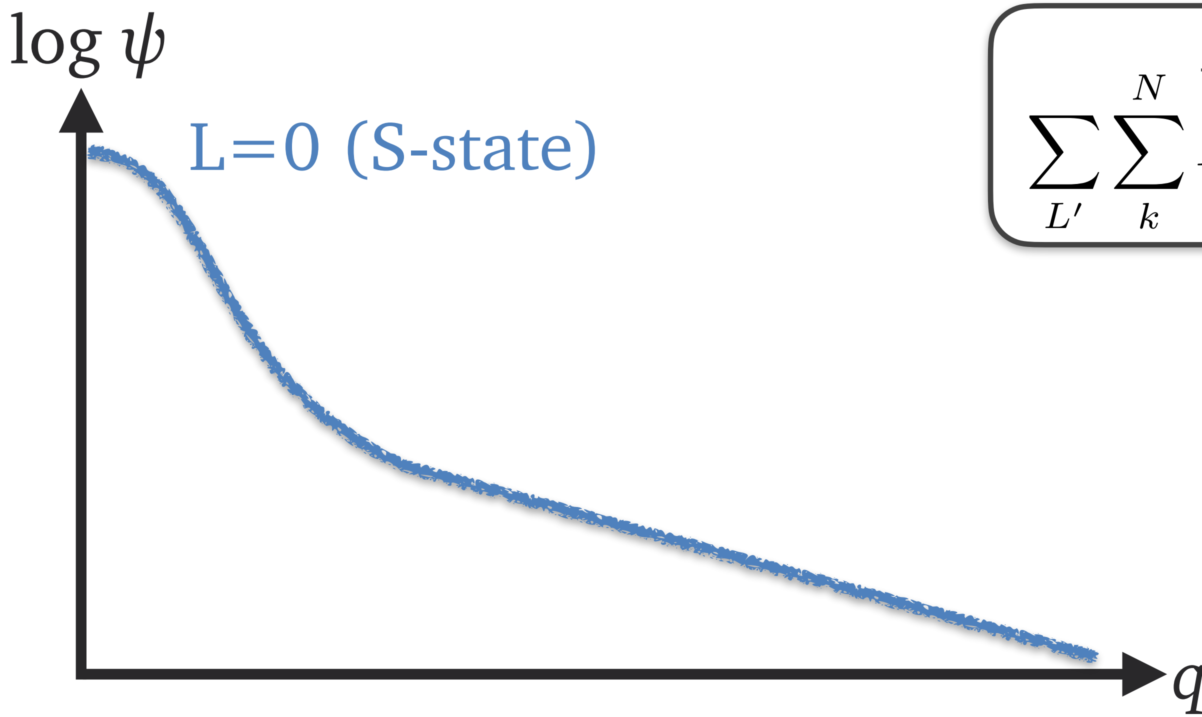
$$\mathcal{W}^{(1)} \in [-1,0), \quad \mathcal{W}^{(2)} \in [0,1)$$
$$b^{(1)} \in [-1,1)$$

Training

Target wavefunction

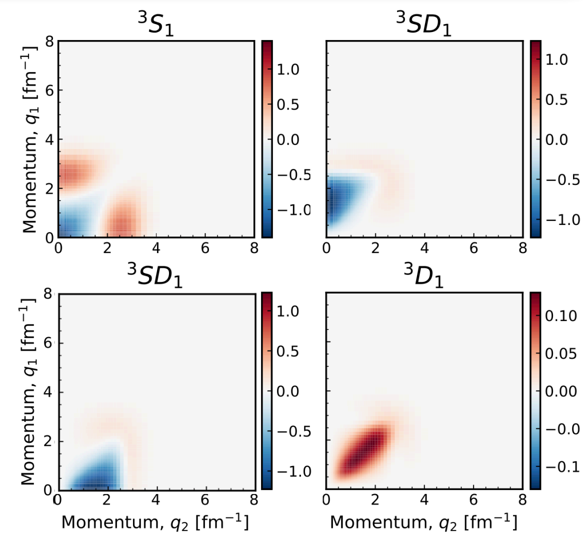
$$\psi^L_{\mathrm{targ}}(q) = \mathcal{N}_L\, q^L e^{-\frac{b^2 q^2}{2}}$$

Momentum discretisation



Effective 1-body eq

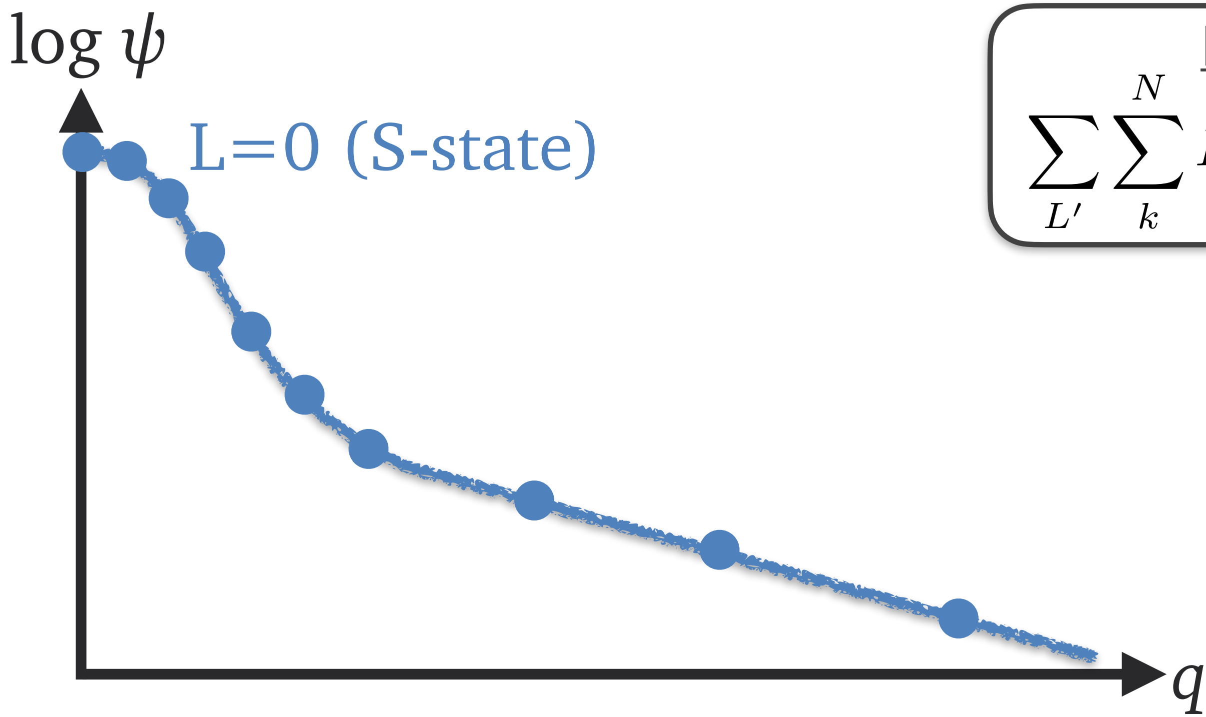
$$\sum_{L'} \sum_{k}^{N} H_{ik}^{LL'} \left|\psi^{L}(q_k)\right\rangle = E \left|\psi^{L}(q_i)\right\rangle$$



$$\langle\psi|K|\psi\rangle = \sum_{L} \int_0^\infty dq q^2 \frac{q^2}{2\mu} \left|\psi^L(q)\right|^2$$

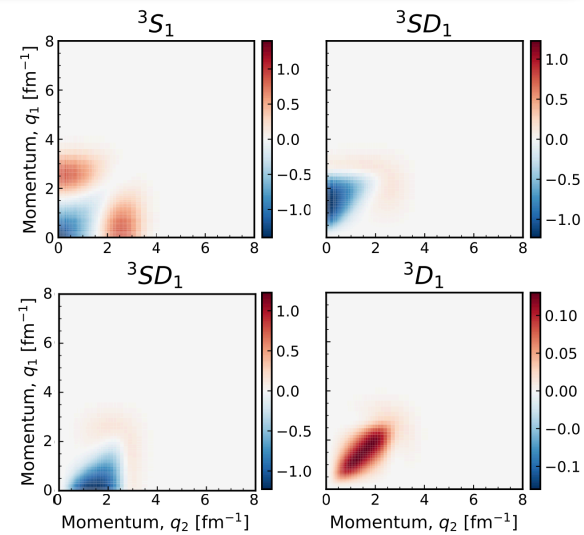
- **No** need for **derivatives** in momentum space
- Tangential transform on *N*=*64* Gauss-Legendre mesh
- Training & validation **set**: *N*=*64* points

Momentum discretisation



Effective 1-body eq

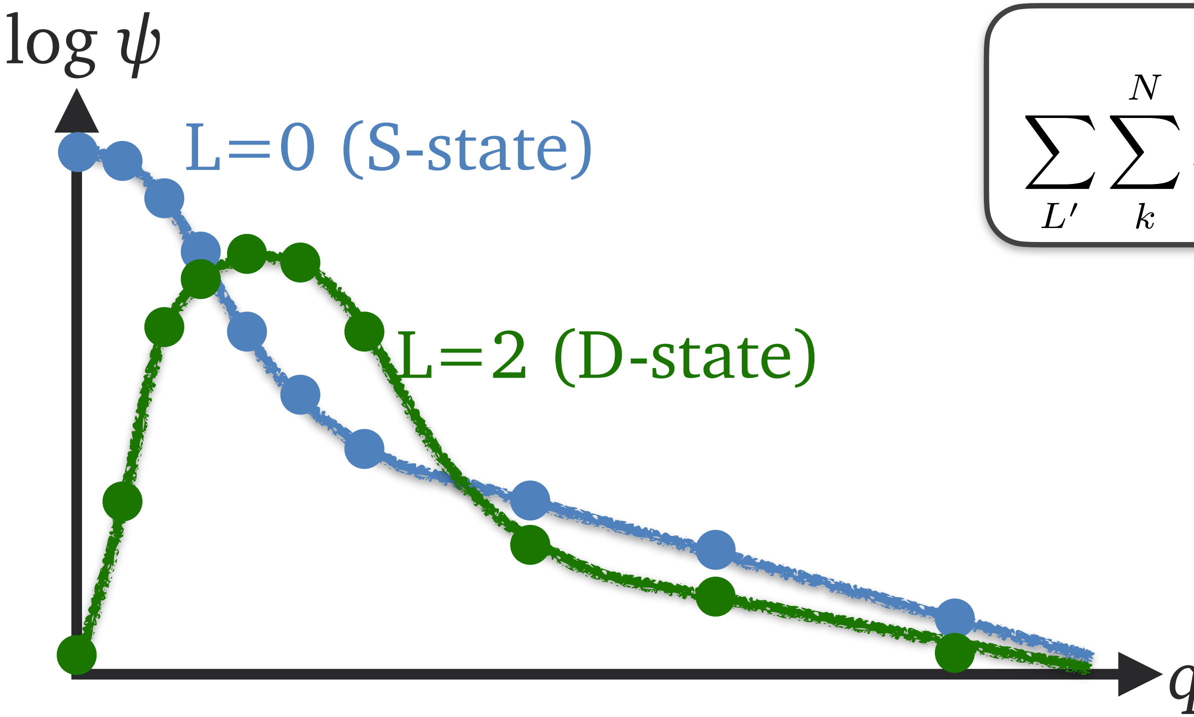
$$\sum_{L'} \sum_{k}^{N} H_{ik}^{LL'} \left|\psi^L(q_k)\right\rangle = E \left|\psi^L(q_i)\right\rangle$$



$$\langle\psi|K|\psi\rangle = \sum_L \int_0^\infty dq q^2 \frac{q^2}{2\mu} \left|\psi^L(q)\right|^2 \approx \sum_L \sum_i^N w_i q_i^2 \frac{q_i^2}{2\mu} \left|\psi^L(q_i)\right|^2$$

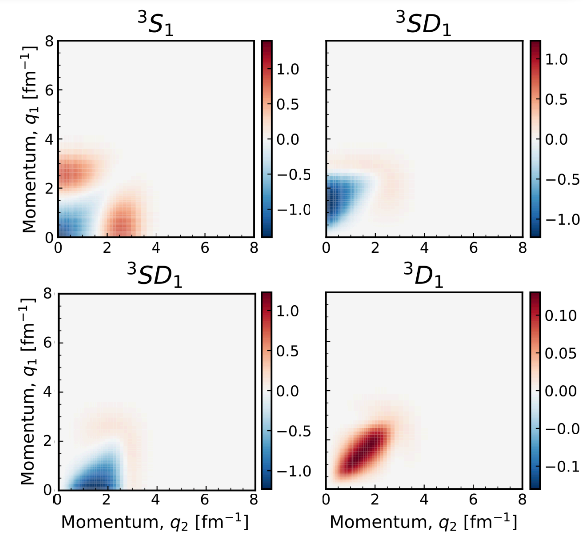
- **No** need for **derivatives** in momentum space
- Tangential transform on *N*=64 Gauss-Legendre mesh
- Training & validation **set**: *N*=64 points

Momentum discretisation



Effective 1-body eq

$$\sum_{L'} \sum_{k}^{N} H_{ik}^{LL'} \left|\psi^{L}(q_k)\right\rangle = E \left|\psi^{L}(q_i)\right\rangle$$



$$\langle\psi|K|\psi\rangle = \sum_L \int_0^\infty dq q^2 \frac{q^2}{2\mu} \left|\psi^L(q)\right|^2 \approx \sum_L \sum_i^N w_i q_i^2 \frac{q_i^2}{2\mu} \left|\psi^L(q_i)\right|^2$$

- **No** need for **derivatives** in momentum space
- Tangential transform on *N*=64 Gauss-Legendre mesh
- Training & validation **set**: *N*=64 points

Energy minimisation - movie

$$\frac{\partial E}{\partial \mathcal{W}} = \frac{2}{\langle\psi|\psi\rangle}\left\{\langle\frac{\partial \psi}{\partial \mathcal{W}}|\hat{H}|\psi\rangle - E\langle\frac{\partial \psi}{\partial \mathcal{W}}|\psi\rangle\right\}$$

RMS prop

$$\mathcal{W}_{t+1} = \mathcal{W}_t - \frac{\alpha}{\sqrt{\mathcal{S}_t + \epsilon}}\frac{\partial \mathcal{C}}{\partial \mathcal{W}_t}$$

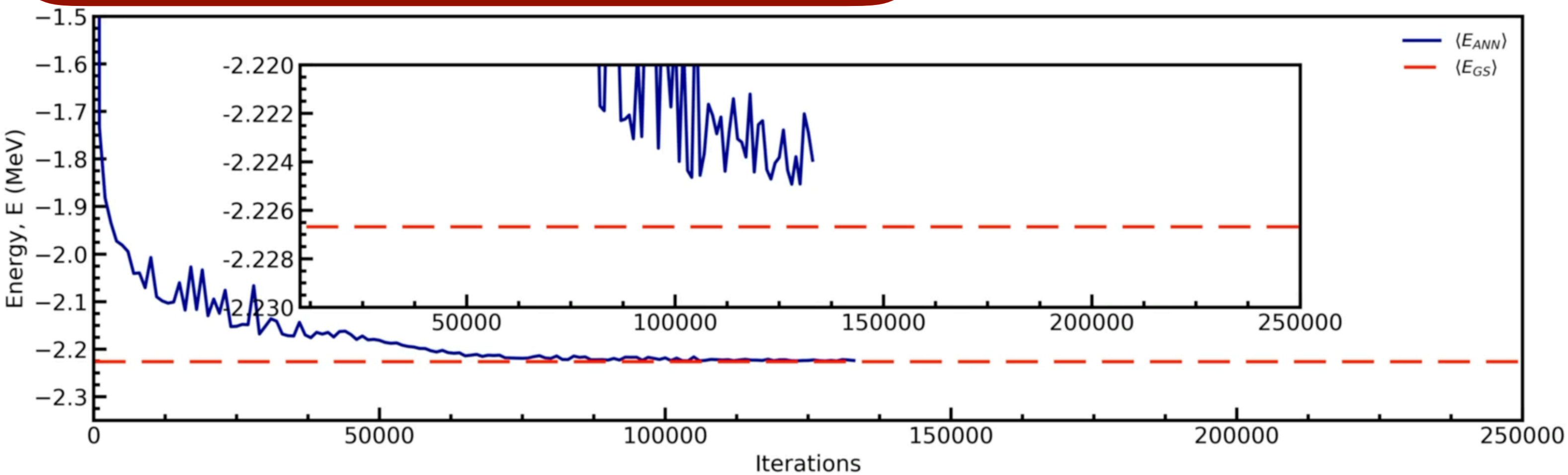
$$\mathcal{S}_t = \gamma \mathcal{S}_{t-1} + (1-\gamma)\left(\frac{\partial \mathcal{C}}{\partial \mathcal{W}_t}\right)^2$$

Hyperparameters

$$\alpha = 10^{-2}$$

$$\epsilon = 10^{-8}$$

$$\gamma = 0.9$$



JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
arxiv:1911.13092

Energy minimisation - movie

$$\frac{\partial E}{\partial \mathcal{W}} = \frac{2}{\langle \psi | \psi \rangle} \left\{ \langle \frac{\partial \psi}{\partial \mathcal{W}} | \hat{H} | \psi \rangle - E \langle \frac{\partial \psi}{\partial \mathcal{W}} | \psi \rangle \right\}$$

RMS prop

$$\mathcal{W}_{t+1} = \mathcal{W}_t - \frac{\alpha}{\sqrt{\mathcal{S}_t + \epsilon}} \frac{\partial \mathcal{C}}{\partial \mathcal{W}_t}$$

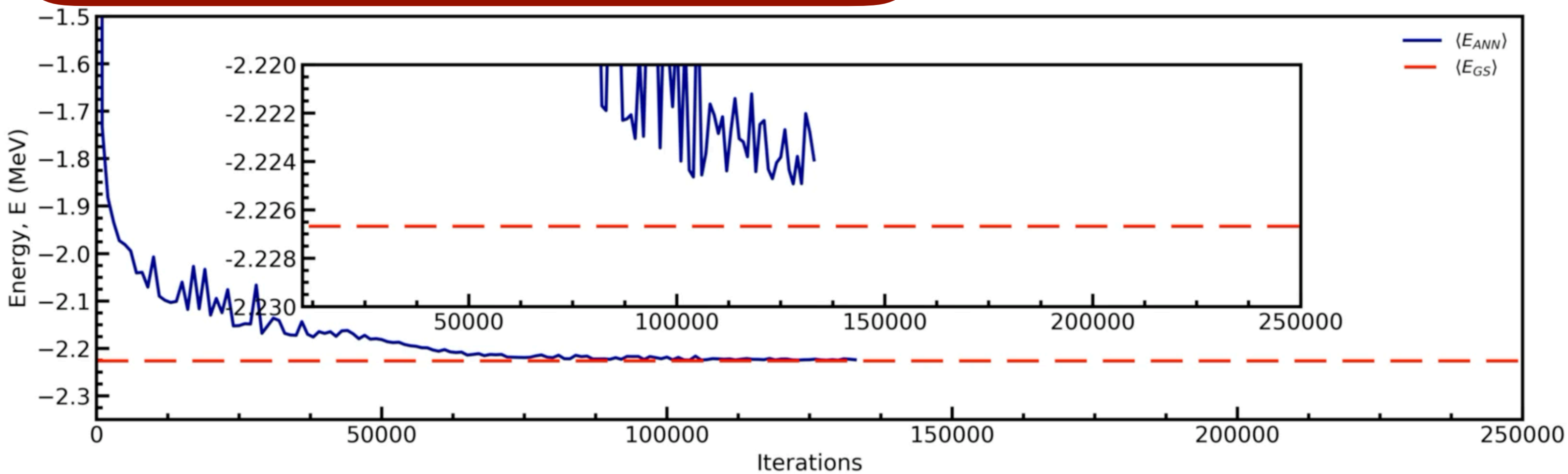
$$\mathcal{S}_t = \gamma \mathcal{S}_{t-1} + (1-\gamma) \left(\frac{\partial \mathcal{C}}{\partial \mathcal{W}_t} \right)^2$$

Hyperparameters

$$\alpha = 10^{-2}$$

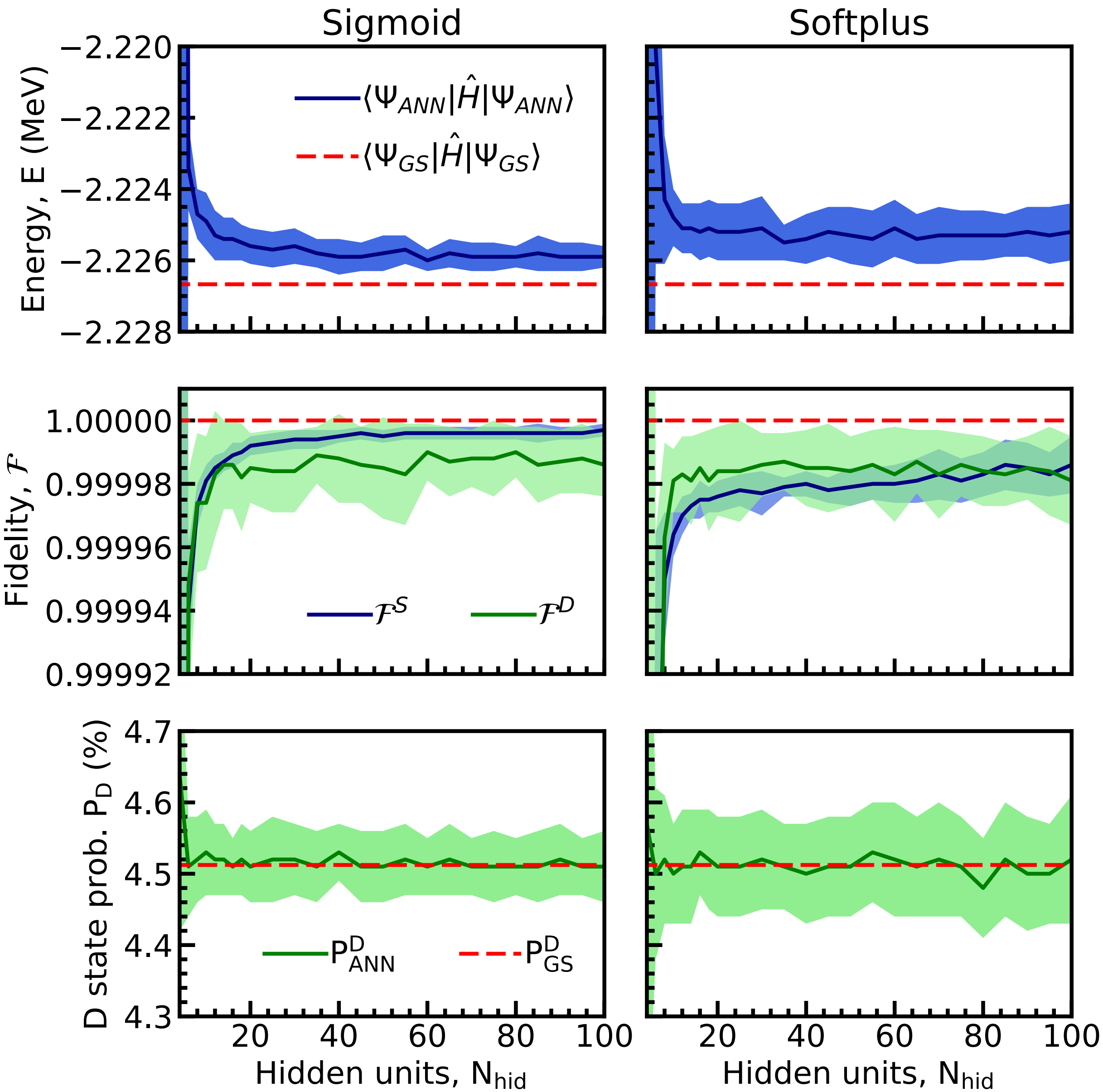
$$\epsilon = 10^{-8}$$

$$\gamma = 0.9$$



JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
arxiv:1911.13092

Results: dependence on N_{hid}



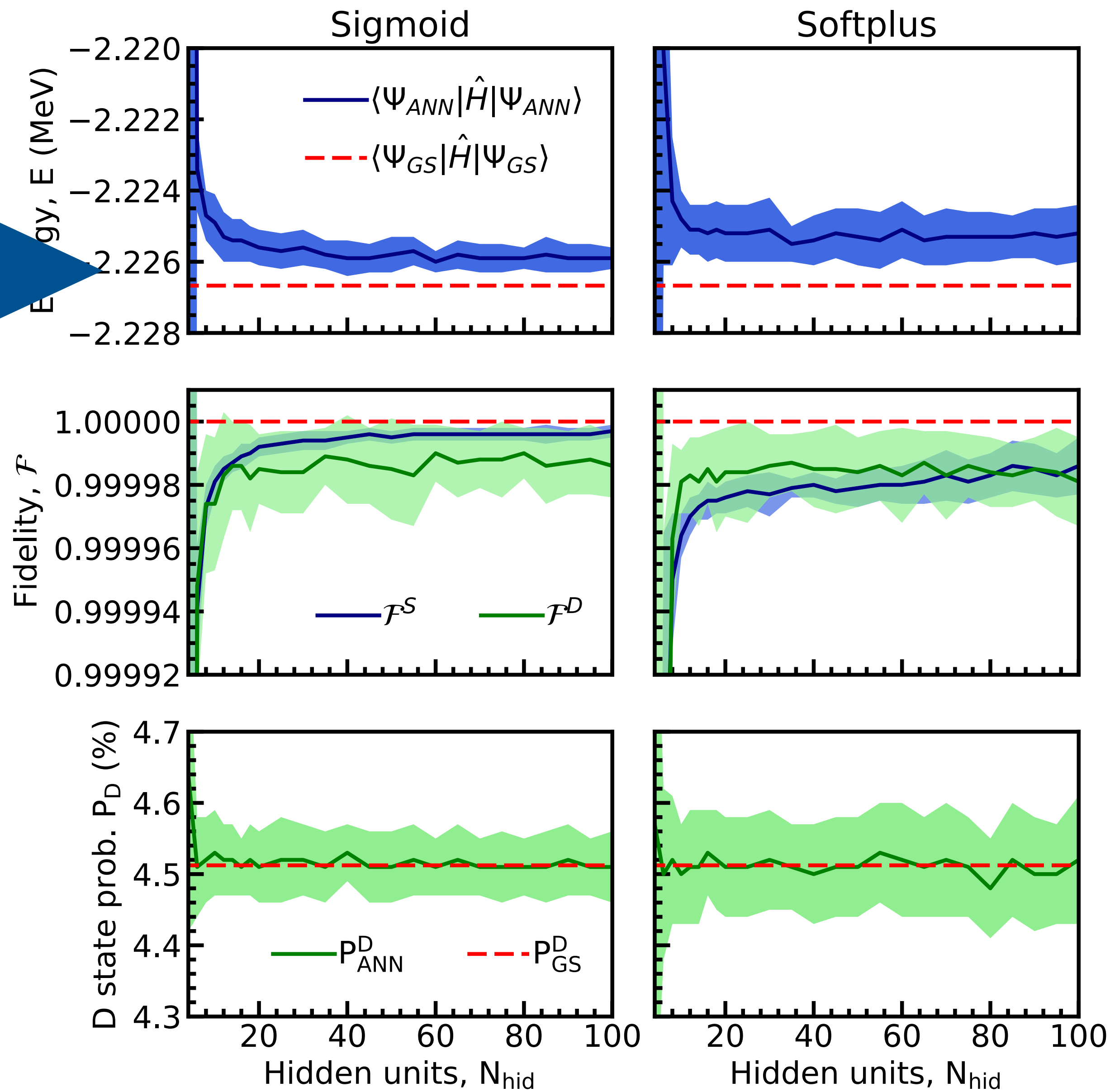
Fidelity

$$\mathcal{F}^L = \frac{\langle \psi^L_{\rm GS} | \psi^L_{\rm ANN} \rangle^2}{\langle \psi^L_{\rm GS} | \psi^L_{\rm GS} \rangle \langle \psi^L_{\rm ANN} | \psi^L_{\rm ANN} \rangle}$$

JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020); arxiv:1911.13092

Results: dependence on N_{hid}

keV accuracy
No bias-variance

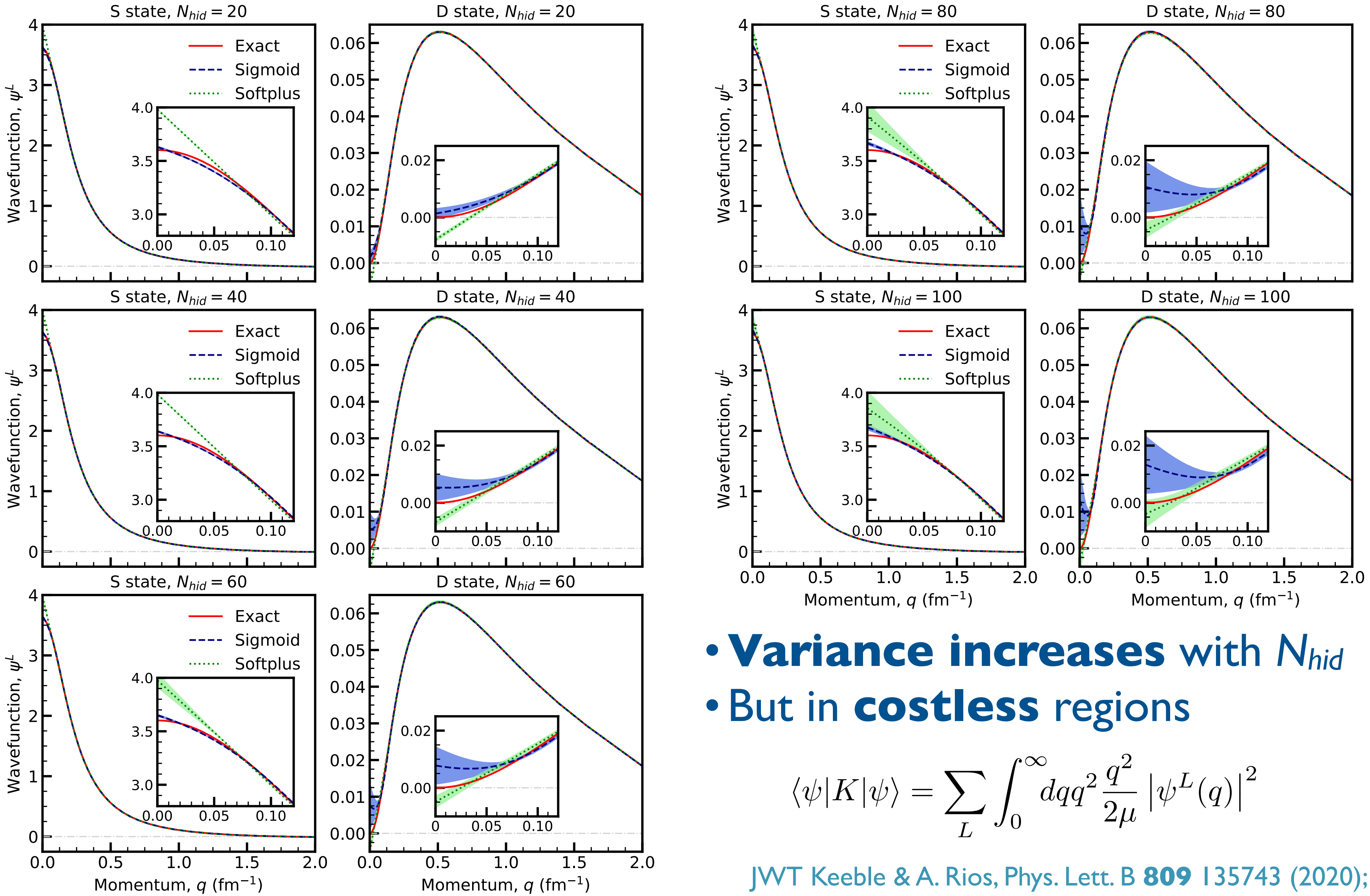


Fidelity

$$\mathcal{F}^L = \frac{\langle\psi^L_{\rm GS}|\psi^L_{\rm ANN}\rangle^2}{\langle\psi^L_{\rm GS}|\psi^L_{\rm GS}\rangle\langle\psi^L_{\rm ANN}|\psi^L_{\rm ANN}\rangle}$$

JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020); arxiv:1911.13092

More on wavefunctions

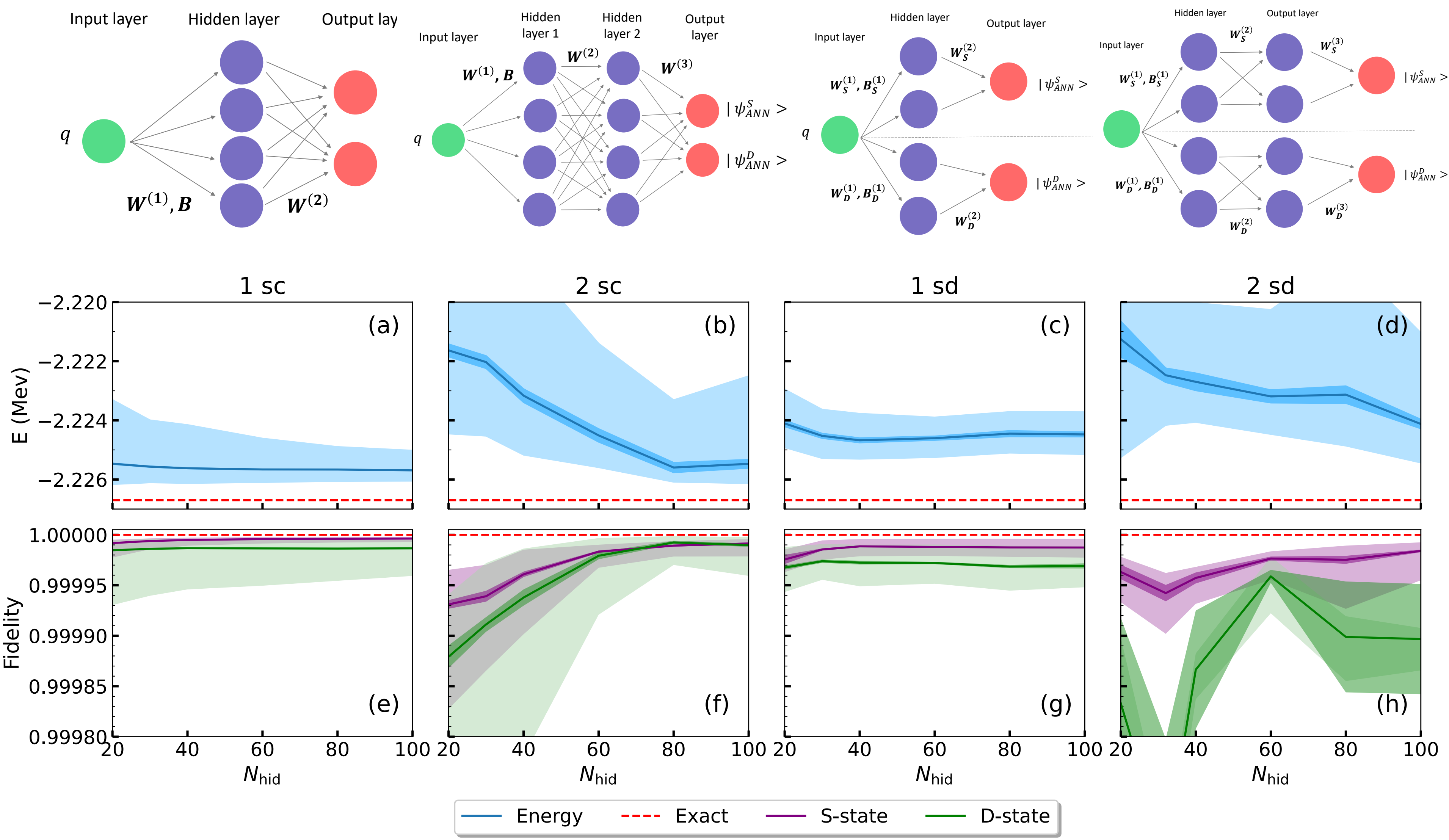


- **Variance increases** with N_{hid}
- But in **costless** regions

$$\langle \psi | K | \psi \rangle = \sum_L \int_0^\infty dq q^2 \frac{q^2}{2\mu} \left| \psi^L(q) \right|^2$$

JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020); arxiv:1911.13092

Different architectures



J Rozalen-Sarmiento, JWT Keeble & A. Rios, Eur. Phys. J. Plus **139**, 189 (2024)
arxiv:2205.12795

Outlook

- Nuclear physics

- Neural Quantum States

- Deuteron

  JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
  J Rozalen-Sarmiento, JWT Keeble & A. Rios, Eur. Phys. J. Plus **139**, 189 (2024)

- **1D spinless (polarised) systems (A>2)**

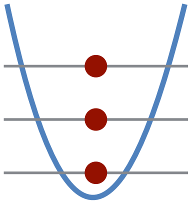
  J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

- Optimisation strategy

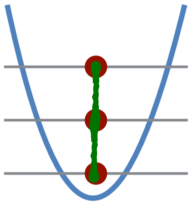
  M. Drissi, J.W.T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Towards the many-body regime

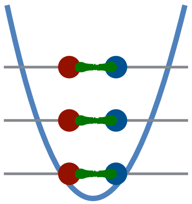
Deep-VMC approach



1. **Free** spinless fermions in 1D ✔

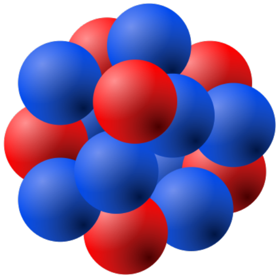


2. **Interacting** spinless fermions in 1D ✔



3. **Interacting spin-full** fermions in 1D ✘

4. 3D nuclei ✘

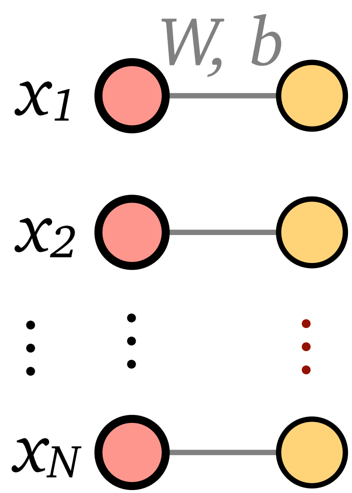


PhD Thesis J Keeble

Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \ldots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \ldots, x_N)}$$

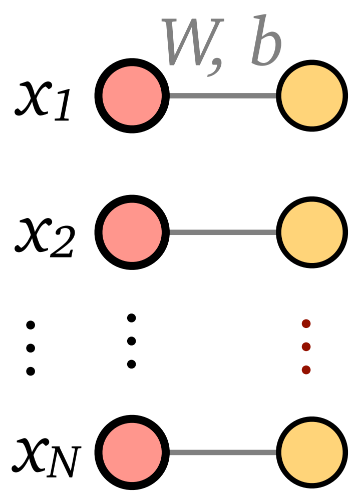
$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & 0 & \cdots & 0 \\ 0 & W & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} Wx_1 + b \\ Wx_2 + b \\ \vdots \\ Wx_N + b \end{pmatrix}$$



Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \ldots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \ldots, x_N)}$$

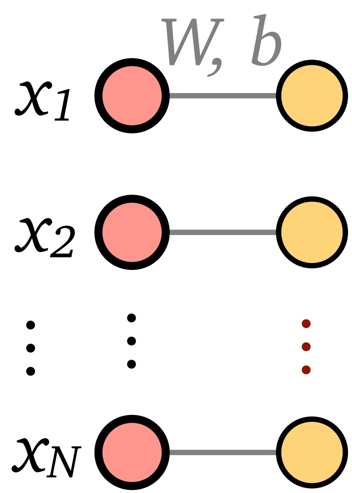
$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & 0 & \cdots & 0 \\ 0 & W & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} Wx_1 + b \\ Wx_2 + b \\ \vdots \\ Wx_N + b \end{pmatrix}$$



Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \ldots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \ldots, x_N)}$$

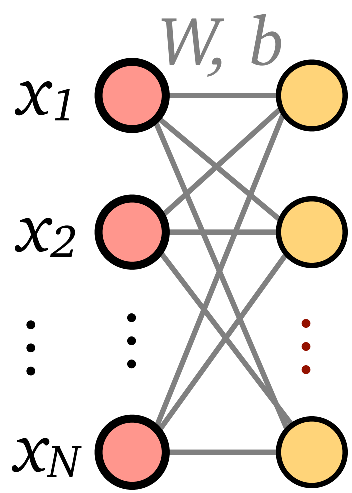
$$\begin{pmatrix} f_1 \\ f_N \\ \vdots \\ f_2 \end{pmatrix} = \begin{pmatrix} W & 0 & \cdots & 0 \\ 0 & W & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_N \\ \vdots \\ x_2 \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} Wx_1 + b \\ Wx_N + b \\ \vdots \\ Wx_2 + b \end{pmatrix}$$



Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \ldots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \ldots, x_N)}$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & V & \cdots & V \\ V & W & \cdots & V \\ \vdots & \vdots & & \vdots \\ V & V & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} W x_1 + V \sum_{i \neq 1} x_i + b \\ W x_2 + V \sum_{i \neq 2} x_i + b \\ \vdots \\ W x_N + V \sum_{i \neq N} x_i + b \end{pmatrix}$$

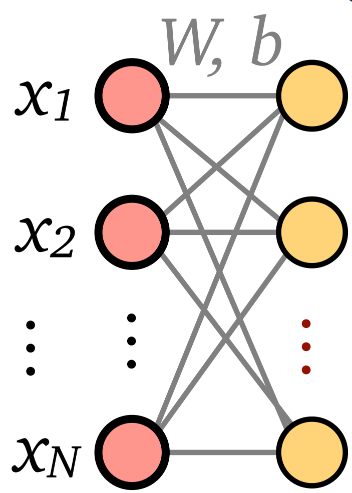


Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \ldots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \ldots, x_N)}$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & V & \cdots & V \\ V & W & \cdots & V \\ \vdots & \vdots & & \vdots \\ V & V & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} Wx_1 + V\sum_{i\neq 1} x_i + b \\ Wx_2 + V\sum_{i\neq 2} x_i + b \\ \vdots \\ Wx_N + V\sum_{i\neq N} x_i + b \end{pmatrix}$$

$$\boxed{Wx_i + V\sum_{j\neq i} x_j + b = (W-V)x_i + V\sum_j x_j + b}$$



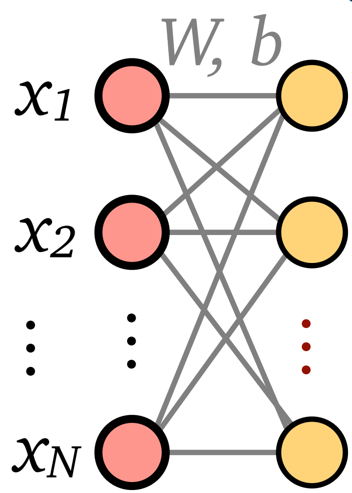
Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)},\ldots,x_{\pi(N)}) = f_{\pi(i)}(x_1,\ldots,x_N)}$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & V & \cdots & V \\ V & W & \cdots & V \\ \vdots & \vdots & & \vdots \\ V & V & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} W x_1 + V\sum_{i\neq 1} x_i + b \\ W x_2 + V\sum_{i\neq 2} x_i + b \\ \vdots \\ W x_N + V\sum_{i\neq N} x_i + b \end{pmatrix}$$

$$\boxed{W x_i + V\sum_{j\neq i} x_j + b = (\overset{W'}{\cancel{W-V}})x_i + V\sum_{j} x_j + b}$$

π-invariant



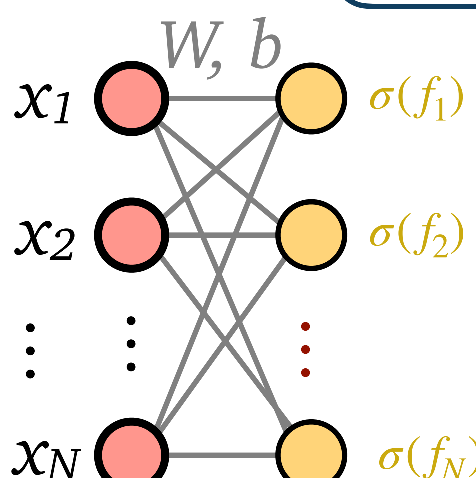
Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \ldots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \ldots, x_N)}$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & V & \cdots & V \\ V & W & \cdots & V \\ \vdots & \vdots & & \vdots \\ V & V & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} Wx_1 + V\sum_{i\neq 1} x_i + b \\ Wx_2 + V\sum_{i\neq 2} x_i + b \\ \vdots \\ Wx_N + V\sum_{i\neq N} x_i + b \end{pmatrix}$$

$$\boxed{Wx_i + V\sum_{j\neq i} x_j + b = (\overset{W'}{\cancel{W-V}})x_i + V\sum_j x_j + b}$$

π-invariant



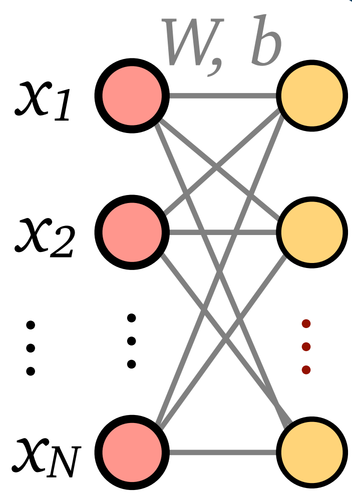
Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \ldots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \ldots, x_N)}$$

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & V & \cdots & V \\ V & W & \cdots & V \\ \vdots & \vdots & & \vdots \\ V & V & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} W x_1 + V \sum_{i \neq 1} x_i + b \\ W x_2 + V \sum_{i \neq 2} x_i + b \\ \vdots \\ W x_N + V \sum_{i \neq N} x_i + b \end{pmatrix}$$

$$\boxed{W x_i + V \sum_{j \neq i} x_j + b = (\overset{W'}{\cancel{W - V}}) x_i + V \sum_j x_j + b}$$

π-invariant



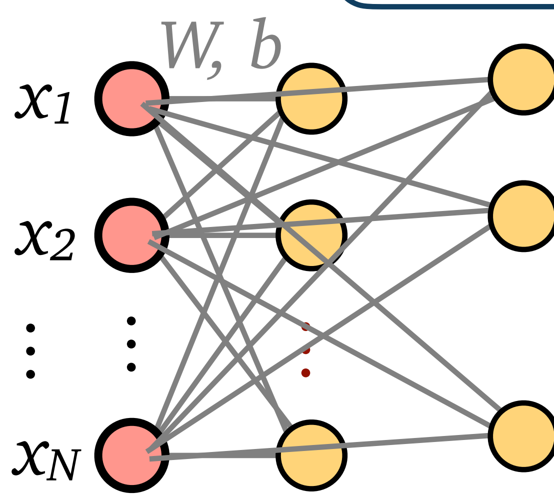
Antisymmetry? Equivariance!

$$\boxed{f_i(x_{\pi(1)}, \dots, x_{\pi(N)}) = f_{\pi(i)}(x_1, \dots, x_N)}$$

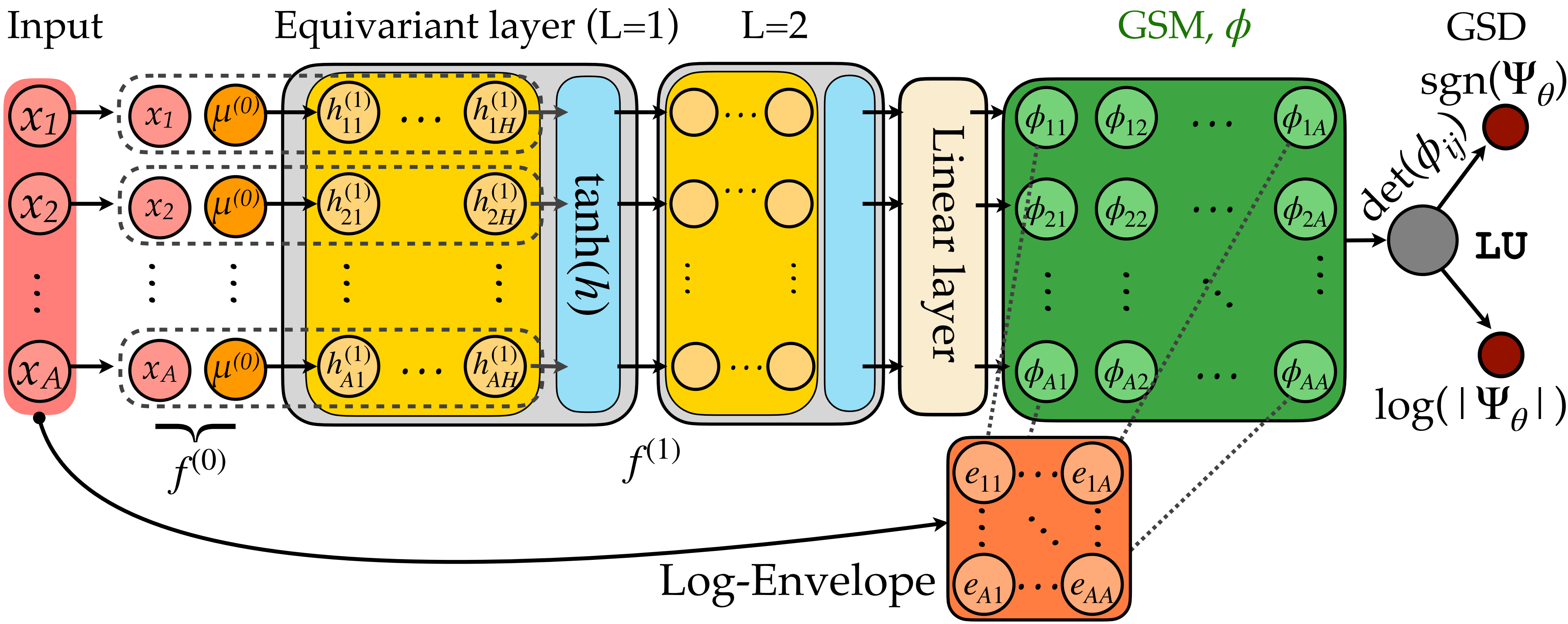
$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} W & V & \cdots & V \\ V & W & \cdots & V \\ \vdots & \vdots & & \vdots \\ V & V & & W \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} b \\ b \\ \vdots \\ b \end{pmatrix} = \begin{pmatrix} W x_1 + V \sum_{i \neq 1} x_i + b \\ W x_2 + V \sum_{i \neq 2} x_i + b \\ \vdots \\ W x_N + V \sum_{i \neq N} x_i + b \end{pmatrix}$$

$$\boxed{W x_i + V \sum_{j \neq i} x_j + b = (\overset{W'}{\cancel{W - V}}) x_i + V \sum_j x_j + b}$$

π-invariant



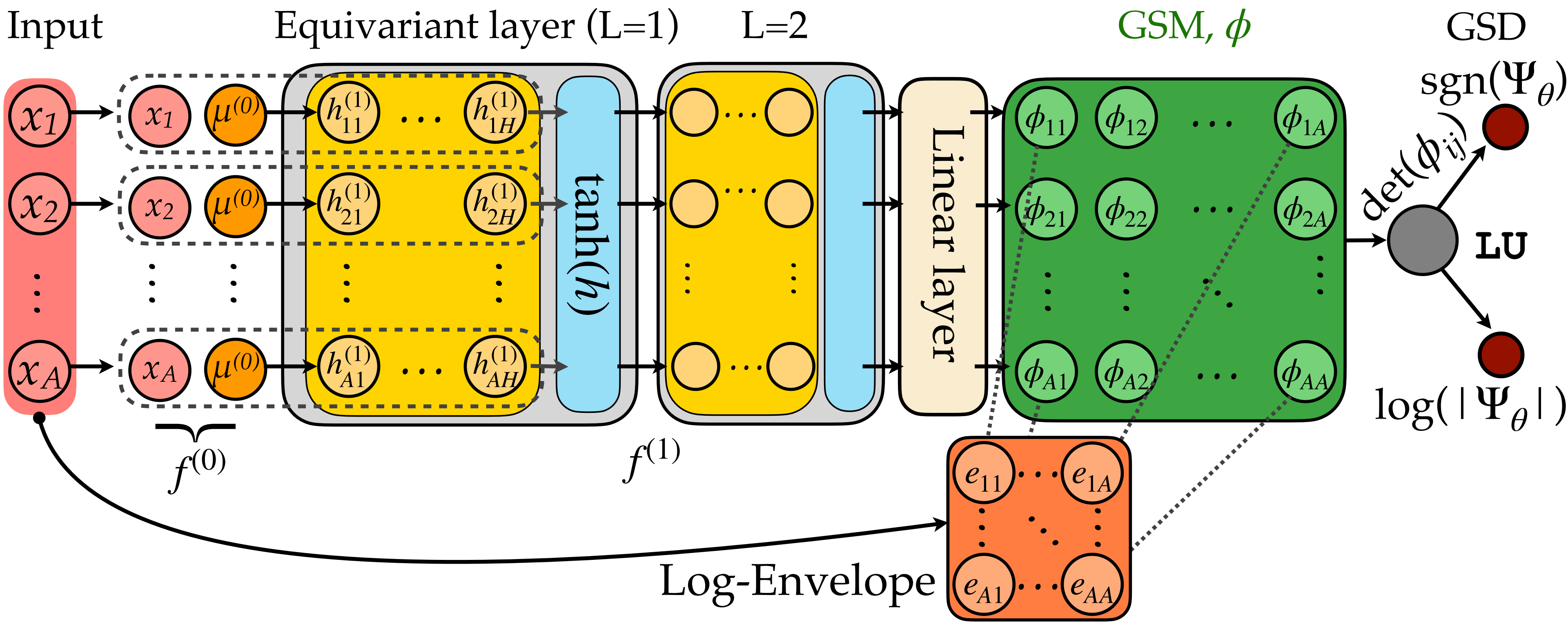
NQS architecture



- Permutation **equivariance**
- D=1 determinants
- Fixed **latent width** *H=64*
- *L*=2 layers
- Envelope layer (boundaries)
- **Linear increase** of # parameters increase in *A*

J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

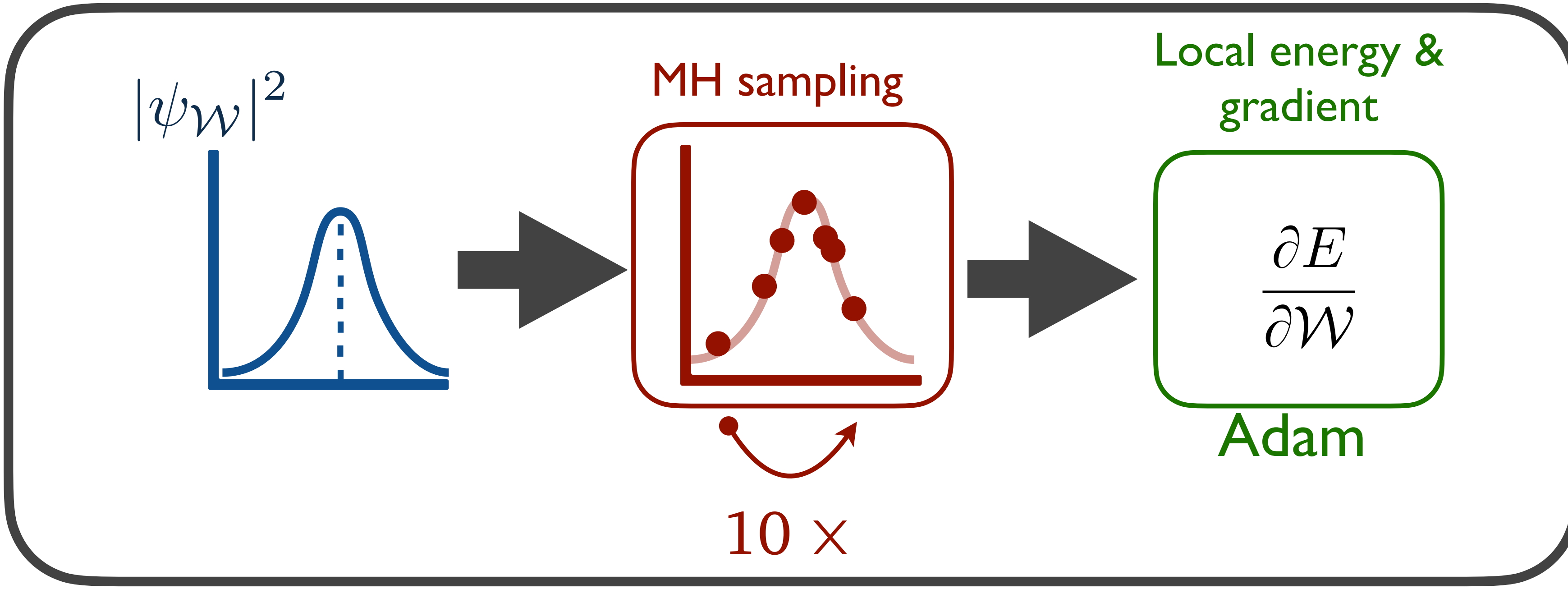
NQS architecture



- Permutation **equivariance**
- D=1 determinants
- Fixed **latent width** *H=64*
- *L*=2 layers
- Envelope layer (boundaries)
- **Linear increase** of # parameters increase in *A*

J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

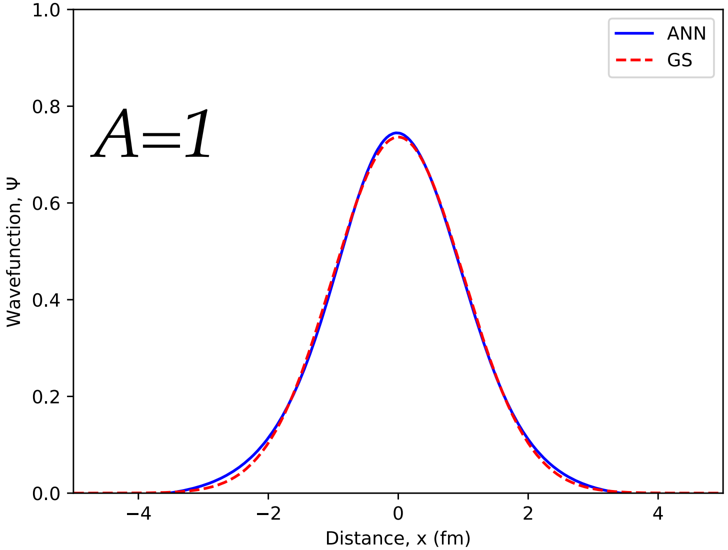
MCMC

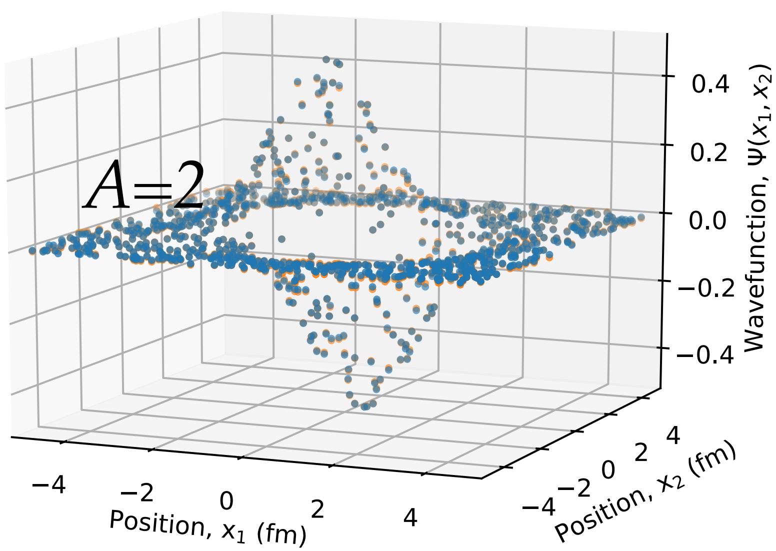


- *4096* walkers in **real** space
- **Local energy** & **gradient** every 10 MH epochs
- Total of **100,000** epochs
- Energy computed from *10^4* batches × *4096* walkers
- Width of Metropolis-Hastings adapted on-the-fly
- For KE, laplacian wrt input needed

$$-\sum_i \frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_i^2} \Psi(x_1, \cdots, x_A)$$

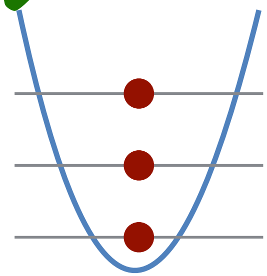
Trapped spinless fermions in 1D HO

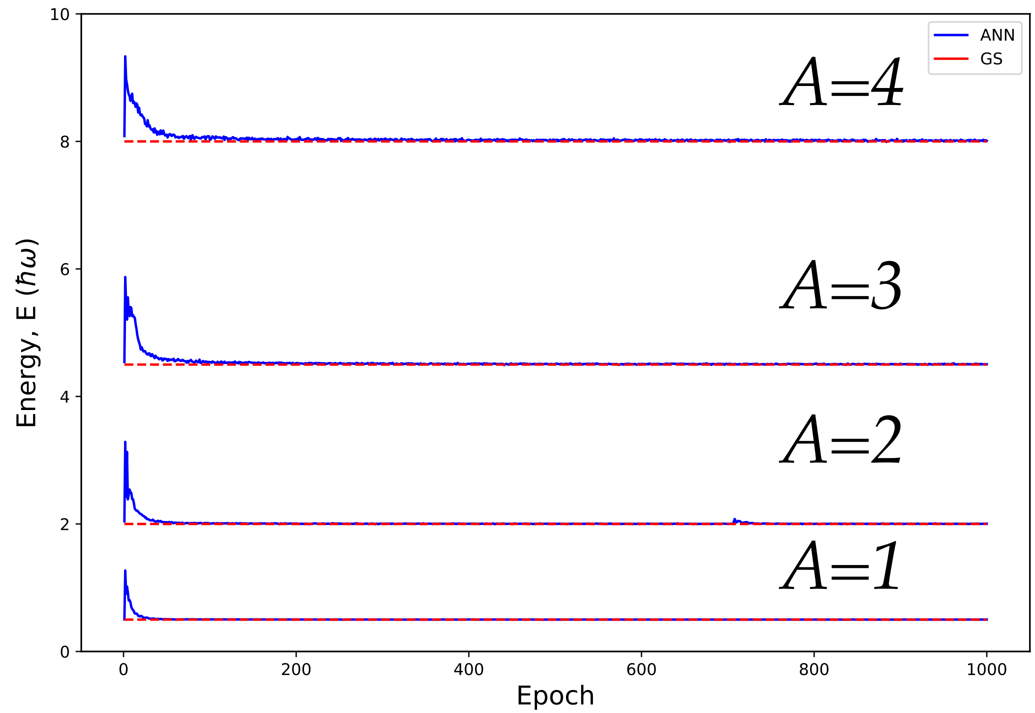




- **Antisymmetrization** ✔

$$E = \frac{A^2}{2}$$

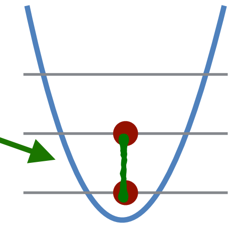


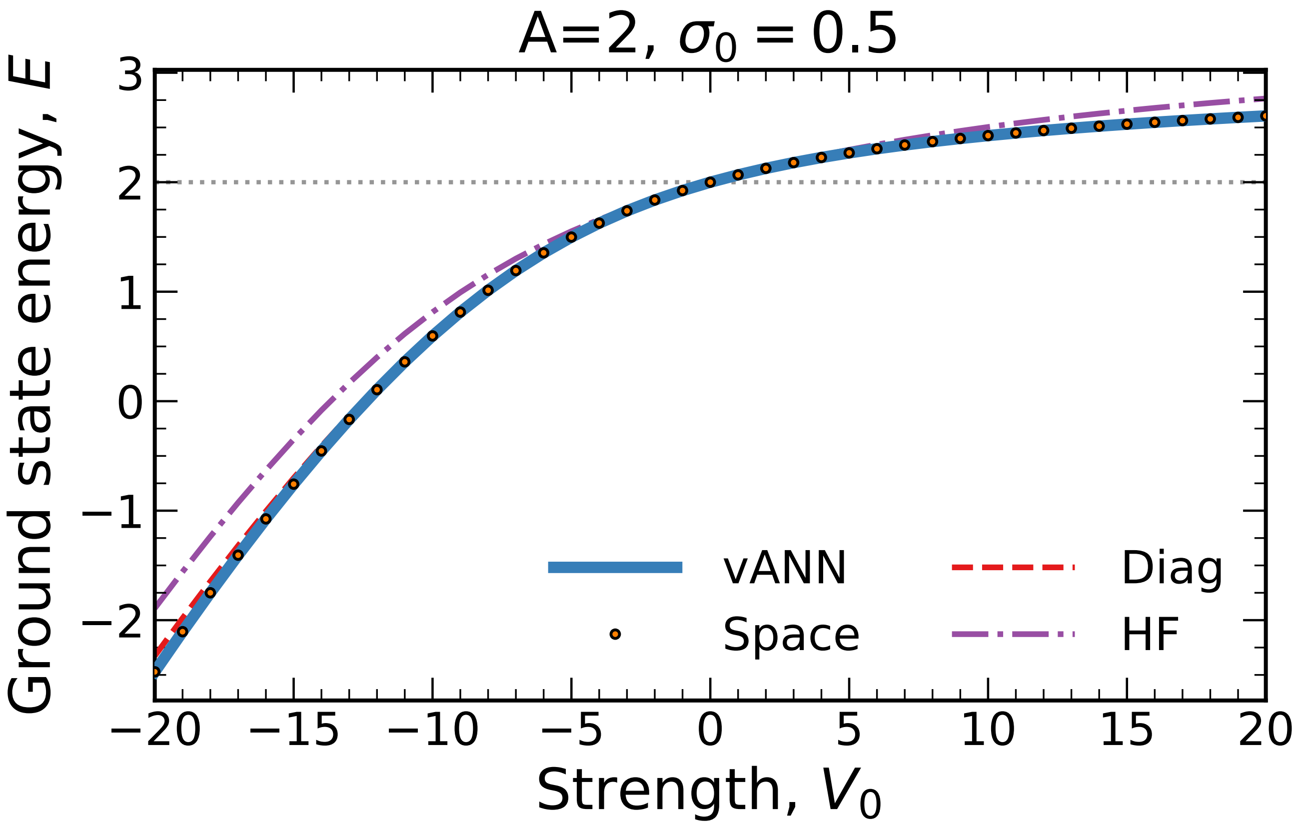


J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

A=2, finite range

$$\hat{H} = -\sum_i \frac{\partial^2_{x_i}}{2} + \sum_i \frac{x_i^2}{2} + \boxed{\sum_{i<j} \frac{V_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(x_i - x_j)^2}{2\sigma_0^2}\right)}$$



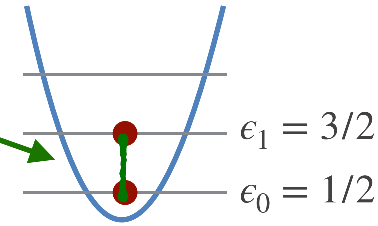


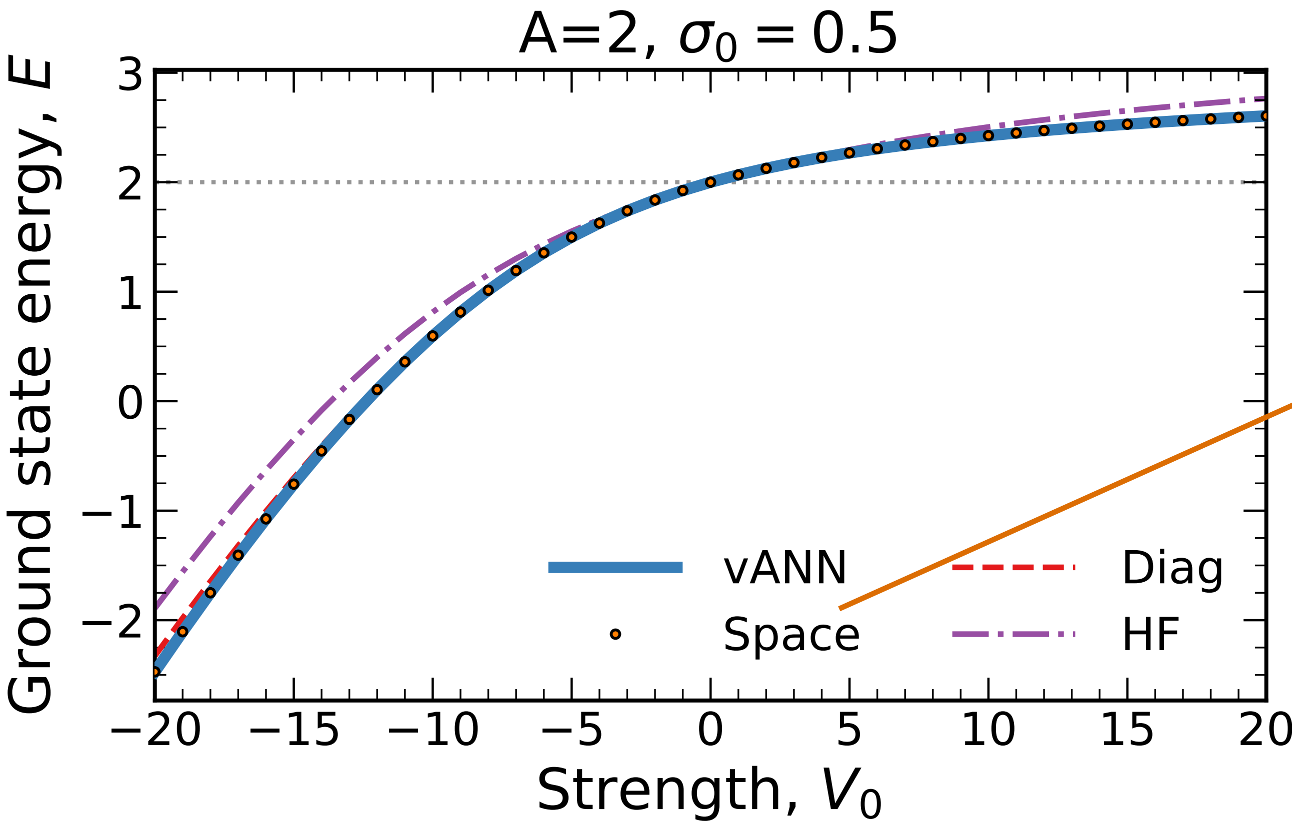
$$\Psi(x_1, x_2) = \begin{vmatrix} \phi_1(x_1, x_2) & \phi_1(x_2, x_1) \\ \phi_2(x_1, x_2) & \phi_2(x_2, x_1) \end{vmatrix}$$

J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

A=2, finite range

$$\hat{H} = -\sum_i \frac{\partial^2_{x_i}}{2} + \sum_i \frac{x_i^2}{2} + \boxed{\sum_{i<j} \frac{V_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(x_i - x_j)^2}{2\sigma_0^2}\right)}$$





$\Psi(x_1, x_2) = \psi(X)\phi(r)$

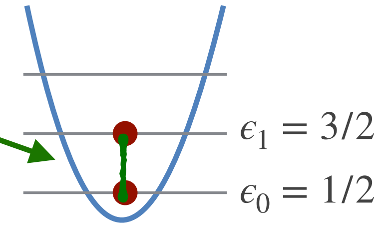
1. ψ is HO
2. Solve ϕ on 1D space grid

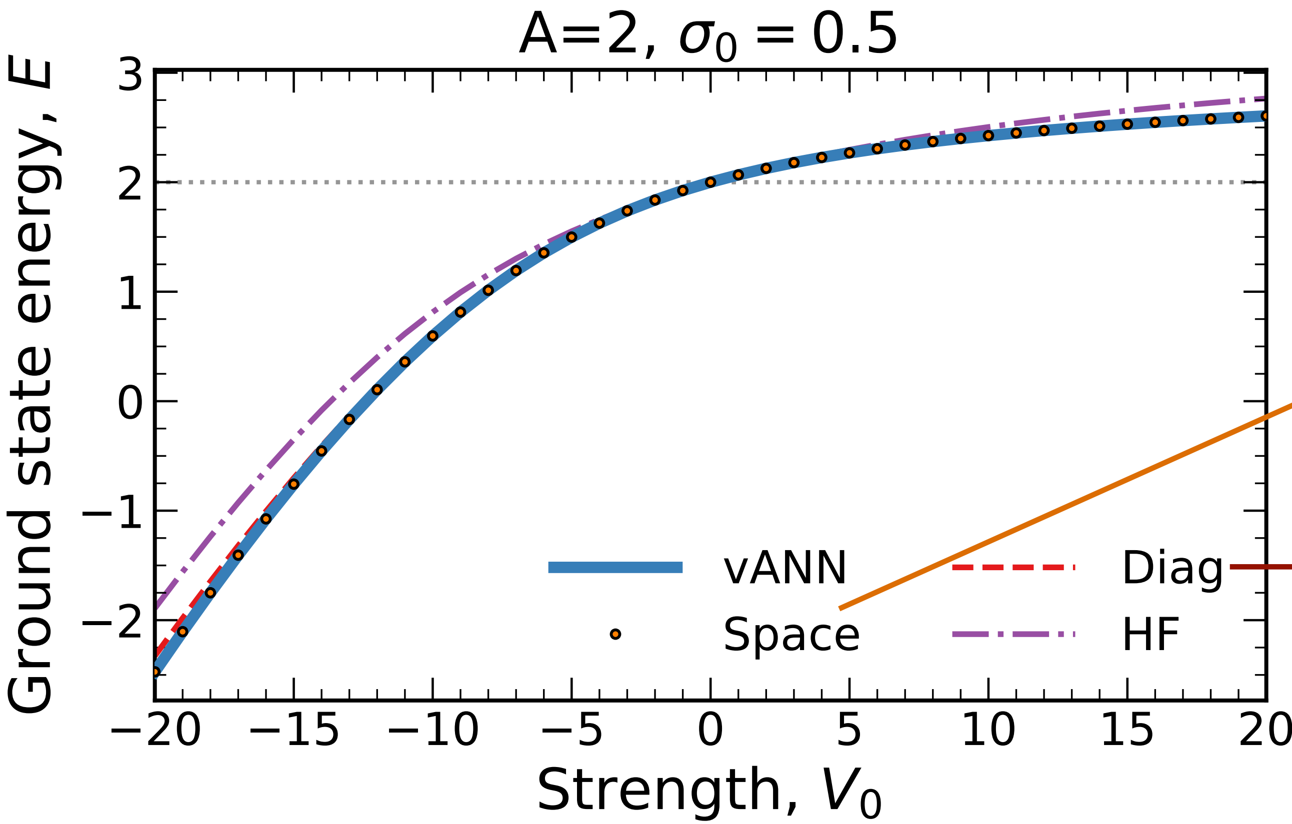
$$\Psi(x_1, x_2) = \begin{vmatrix} \phi_1(x_1, x_2) & \phi_1(x_2, x_1) \\ \phi_2(x_1, x_2) & \phi_2(x_2, x_1) \end{vmatrix}$$

J. W. T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

A=2, finite range

$$\hat{H} = -\sum_i \frac{\partial^2_{x_i}}{2} + \sum_i \frac{x_i^2}{2} + \boxed{\sum_{i<j} \frac{V_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(x_i - x_j)^2}{2\sigma_0^2} \right)}$$





$\Psi(x_1, x_2) = \psi(X)\phi(r)$

1. ψ is HO
2. Solve ϕ on 1D space grid

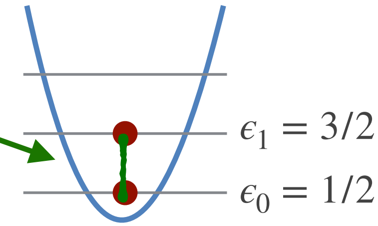
Full CI on HO basis

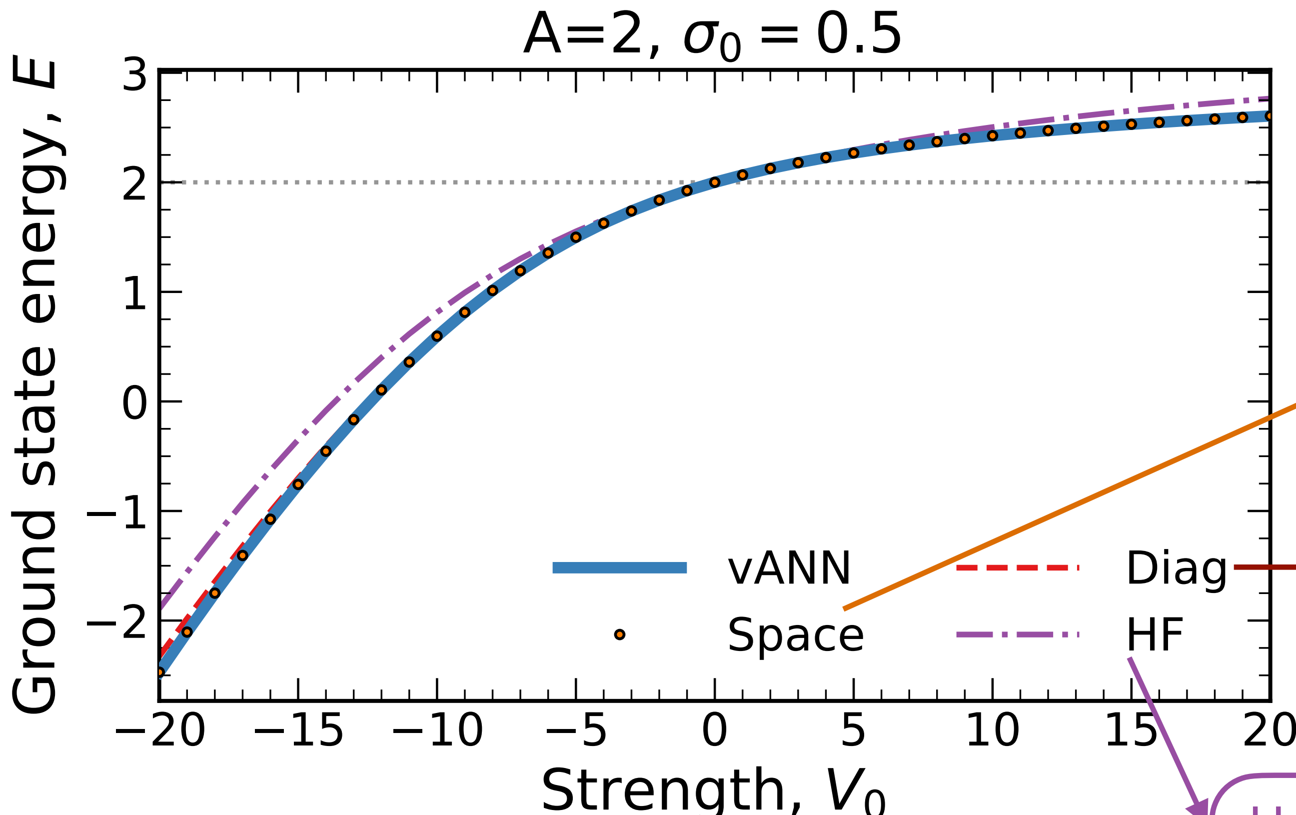
$$\Psi(x_1, x_2) = \begin{vmatrix} \phi_1(x_1, x_2) & \phi_1(x_2, x_1) \\ \phi_2(x_1, x_2) & \phi_2(x_2, x_1) \end{vmatrix}$$

J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

A=2, finite range

$$\hat{H} = -\sum_i \frac{\partial^2_{x_i}}{2} + \sum_i \frac{x_i^2}{2} + \boxed{\sum_{i<j} \frac{V_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(x_i - x_j)^2}{2\sigma_0^2}\right)}$$





$\Psi(x_1, x_2) = \psi(X)\phi(r)$

1. ψ is HO
2. Solve ϕ on 1D space grid

Full CI on HO basis

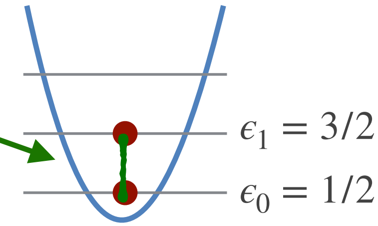
Hartree-Fock:
1 Slater determinant
No backflow

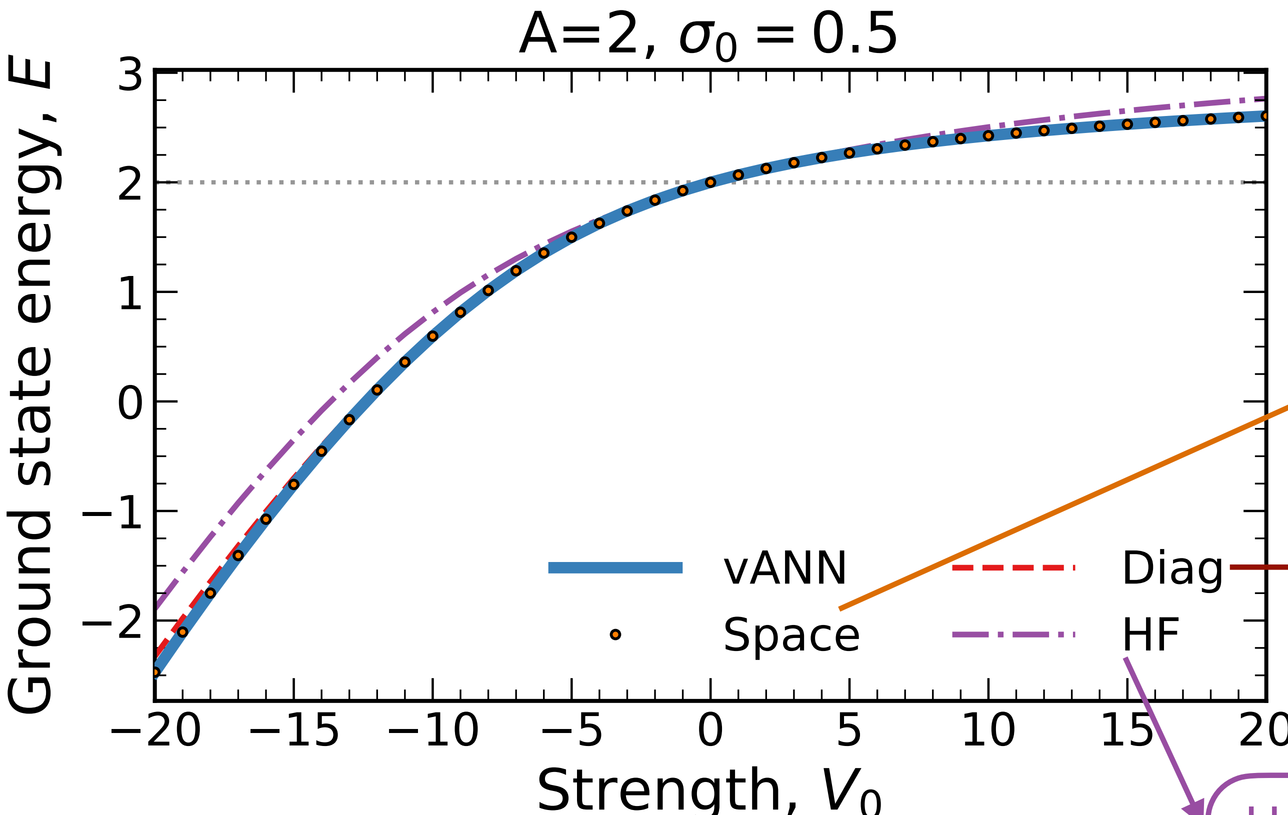
$$\Psi(x_1, x_2) = \begin{vmatrix} \phi_1(x_1, x_2) & \phi_1(x_2, x_1) \\ \phi_2(x_1, x_2) & \phi_2(x_2, x_1) \end{vmatrix}$$

J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

A=2, finite range

$$\hat{H} = -\sum_i \frac{\partial^2_{x_i}}{2} + \sum_i \frac{x_i^2}{2} + \boxed{\sum_{i<j} \frac{V_0}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(x_i - x_j)^2}{2\sigma_0^2}\right)}$$





$\Psi(x_1, x_2) = \psi(X)\phi(r)$

1. ψ is HO
2. Solve ϕ on 1D space grid

Full CI on HO basis

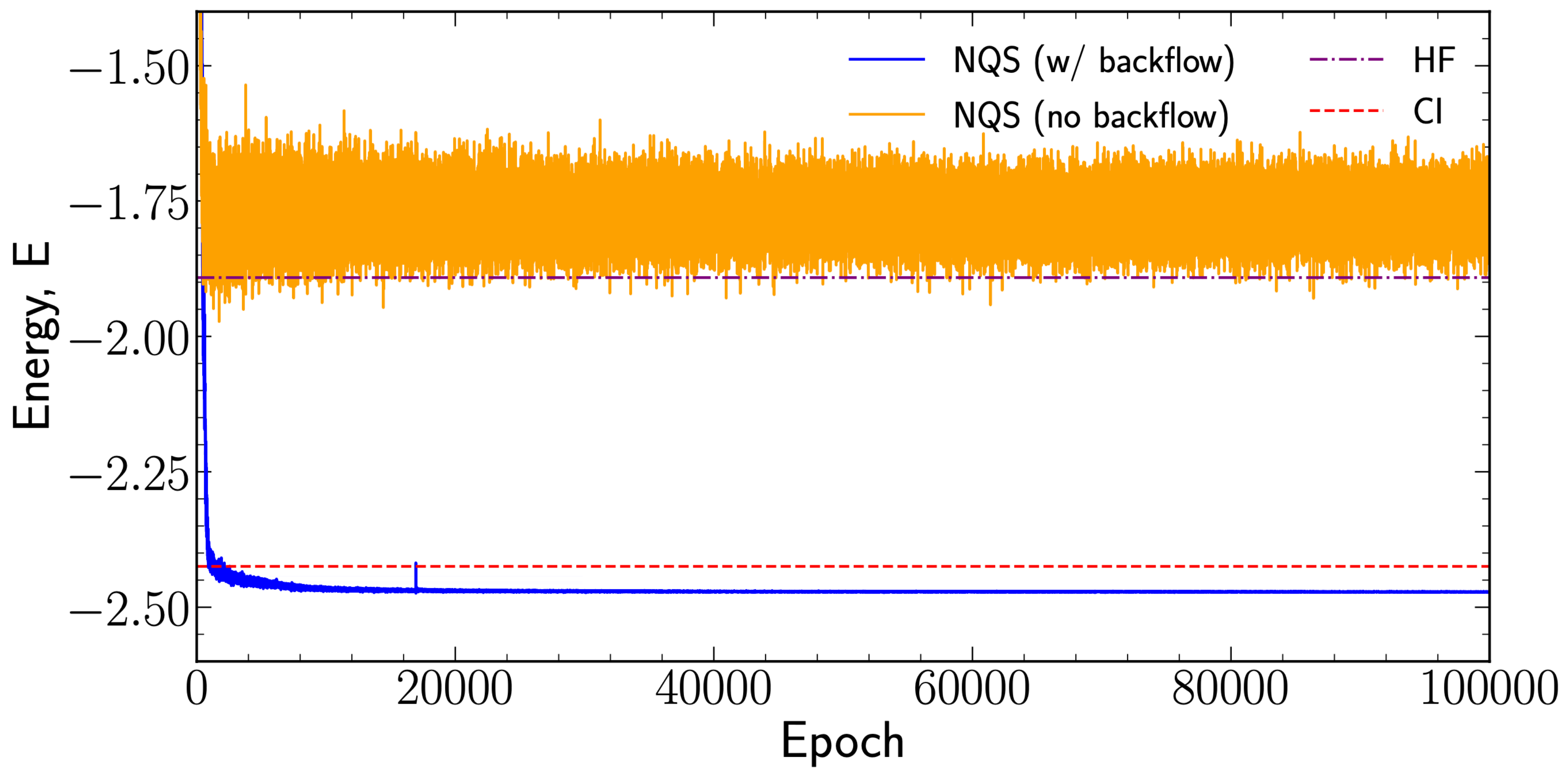
Hartree-Fock:
1 Slater determinant
No backflow

$$\Psi(x_1, x_2) = \begin{vmatrix} \phi_1(x_1, x_2) & \phi_1(x_2, x_1) \\ \phi_2(x_1, x_2) & \phi_2(x_2, x_1) \end{vmatrix}$$

J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

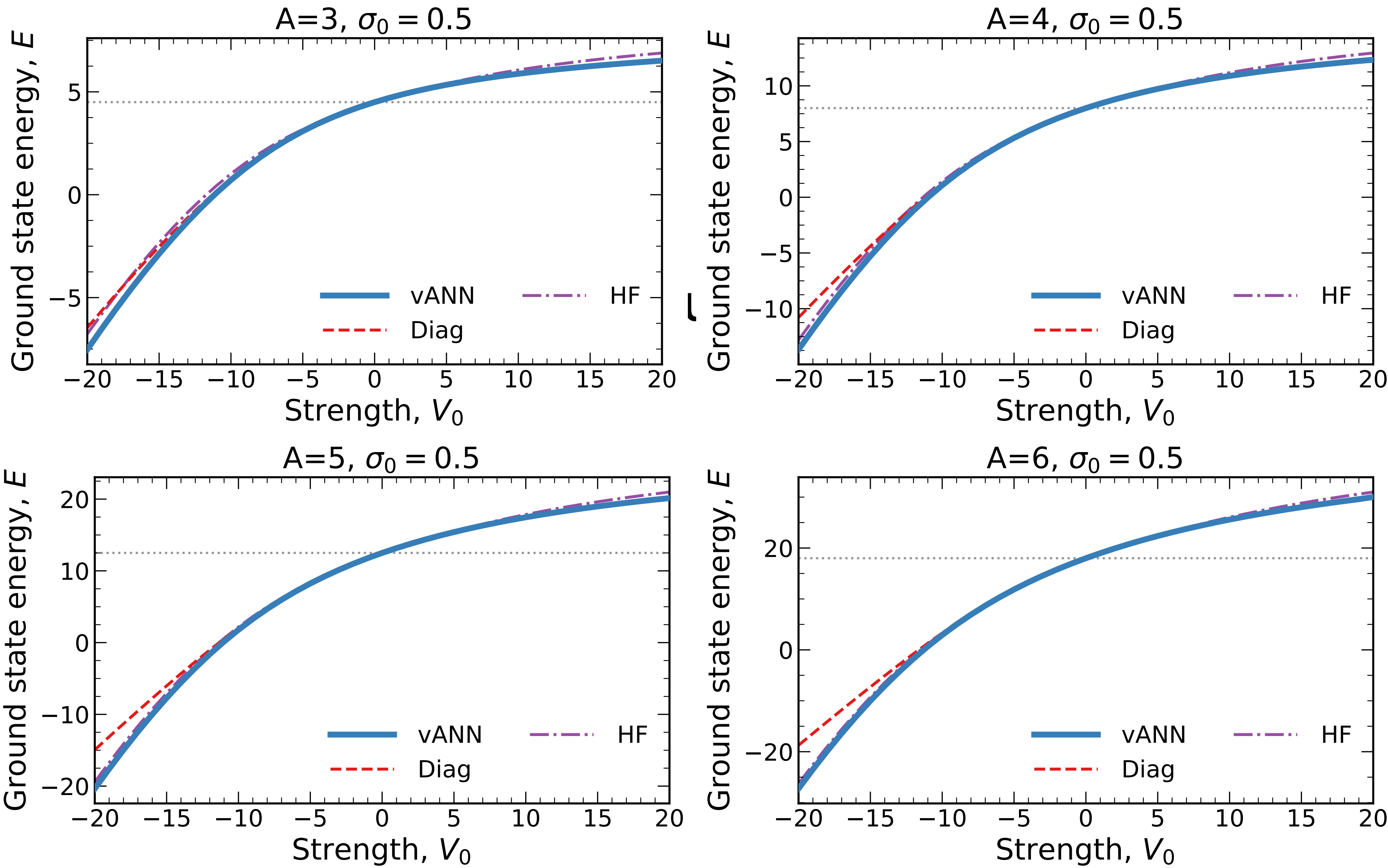
Backflow correlations

A=2, σ_0=0.5, V_0=-20



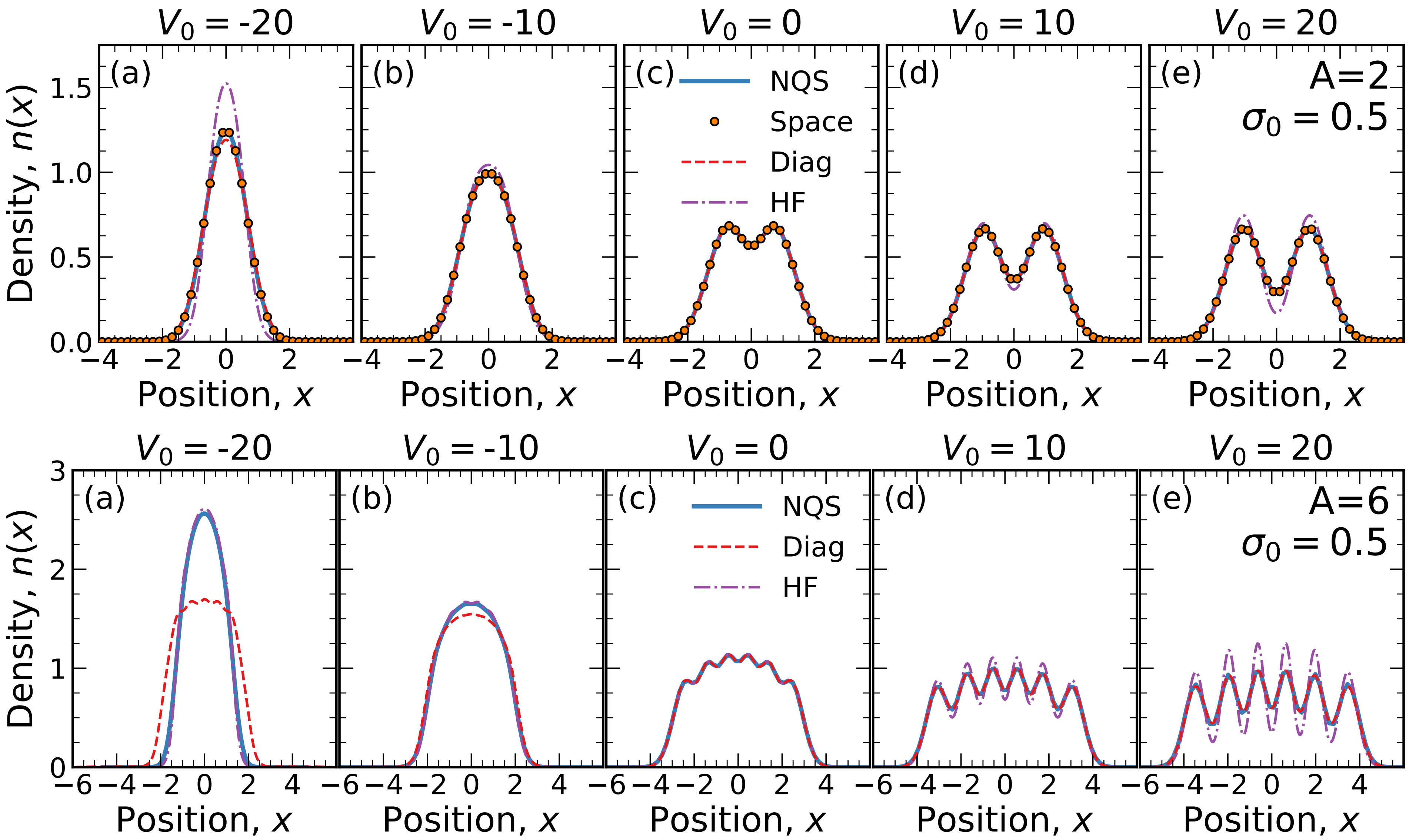
J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

Energy A>2

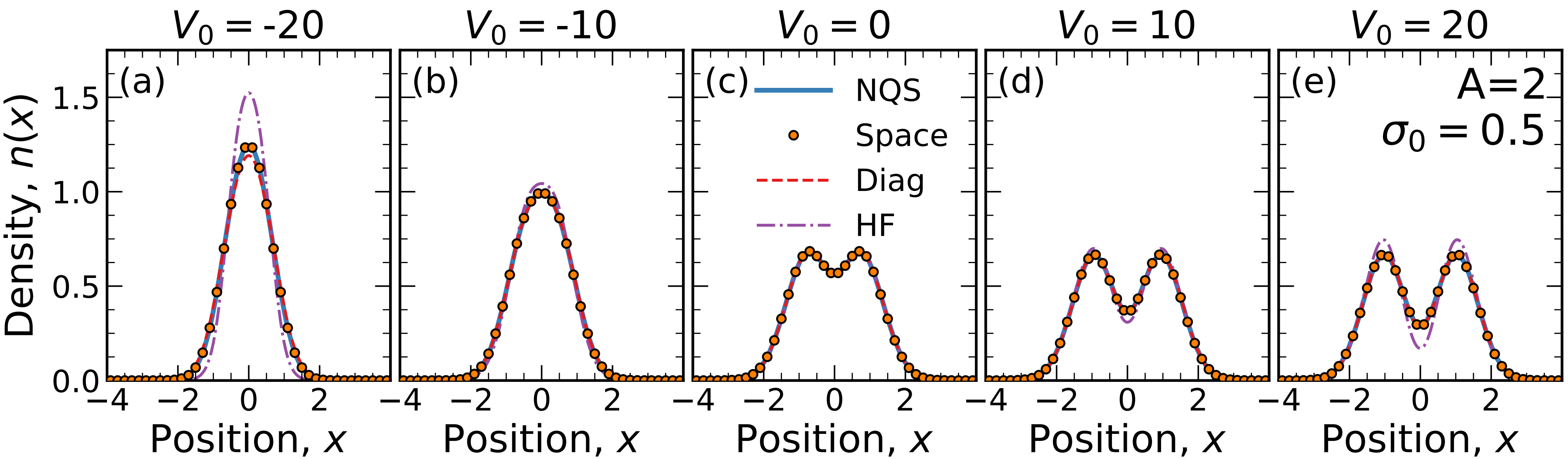


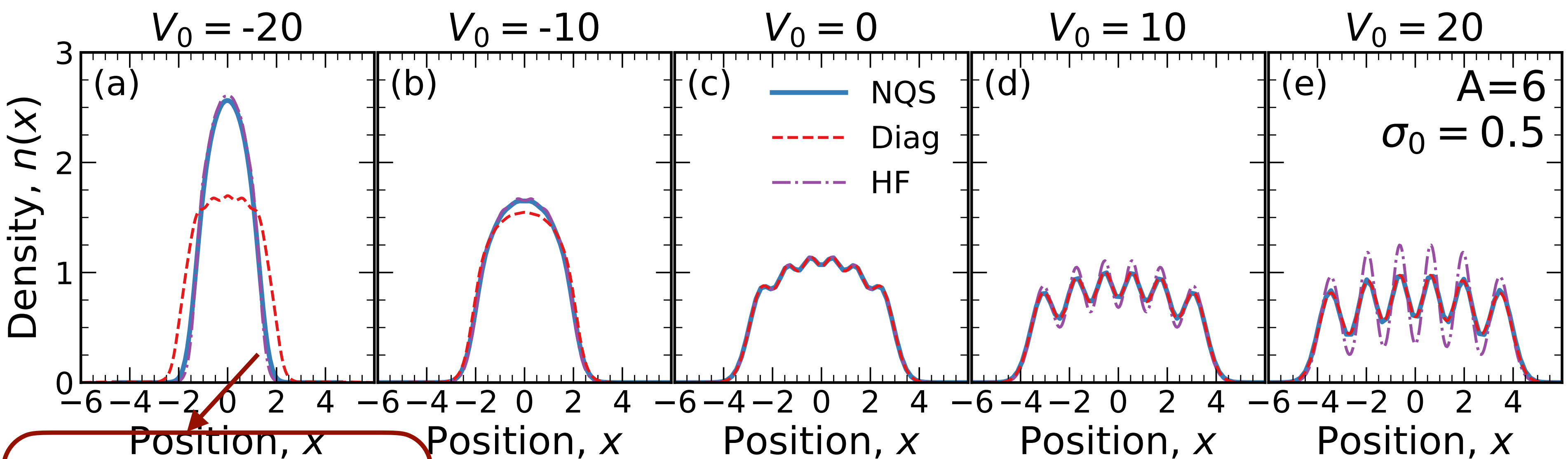
J. W. T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

Local density



Local density

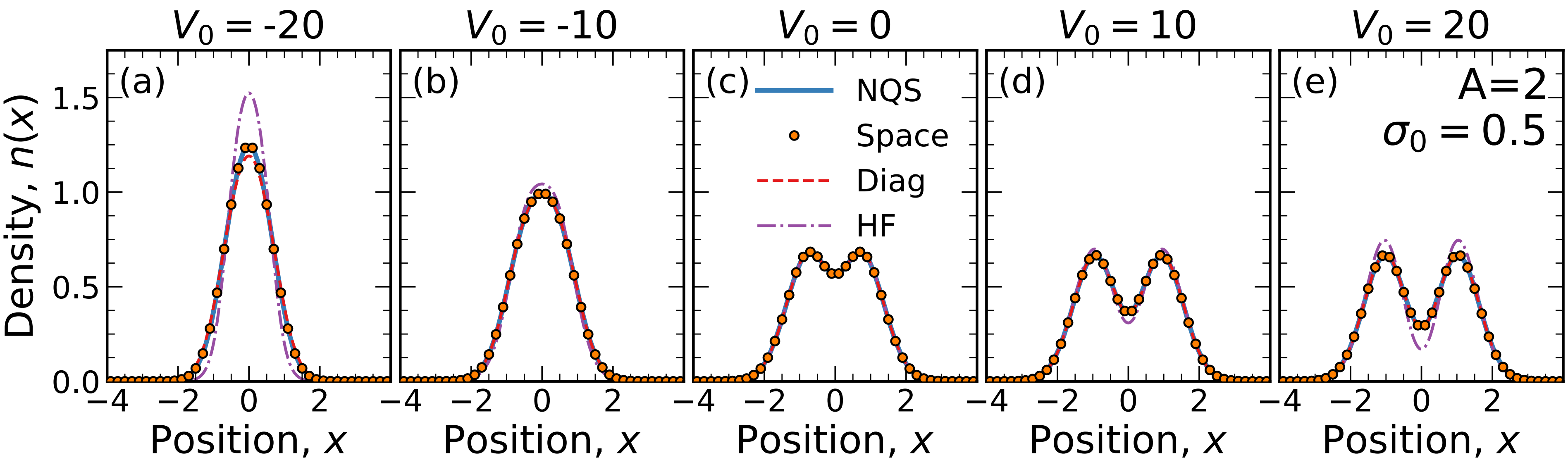


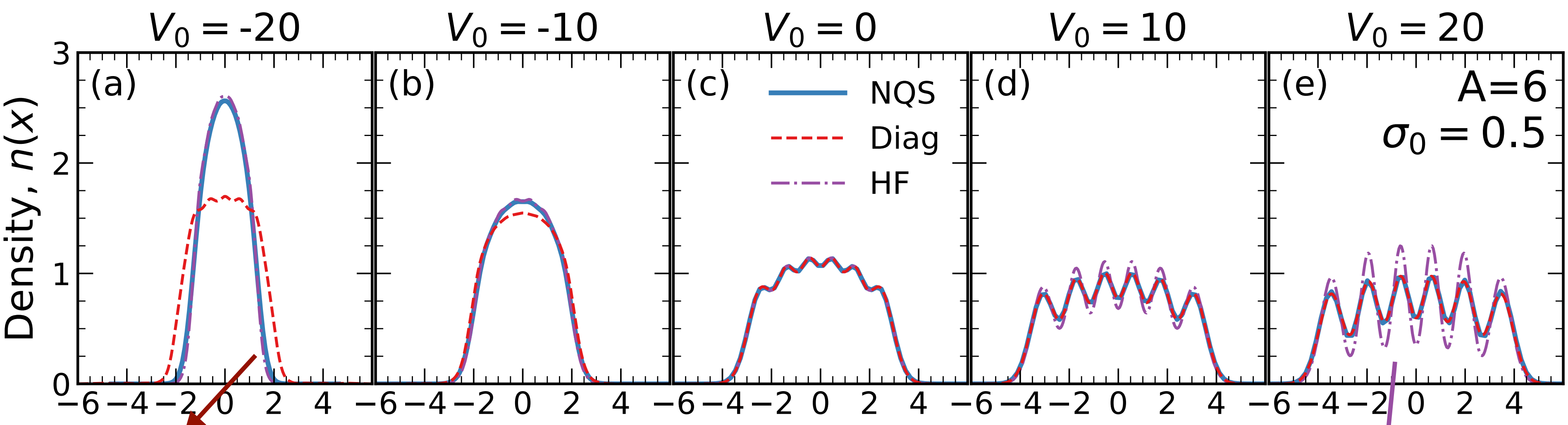


Boson-Fermion duality

Girardeau & Olshanii, arXiv:cond-mat/0309396
Valiente, PRA 103, L021302 (2021)

Local density



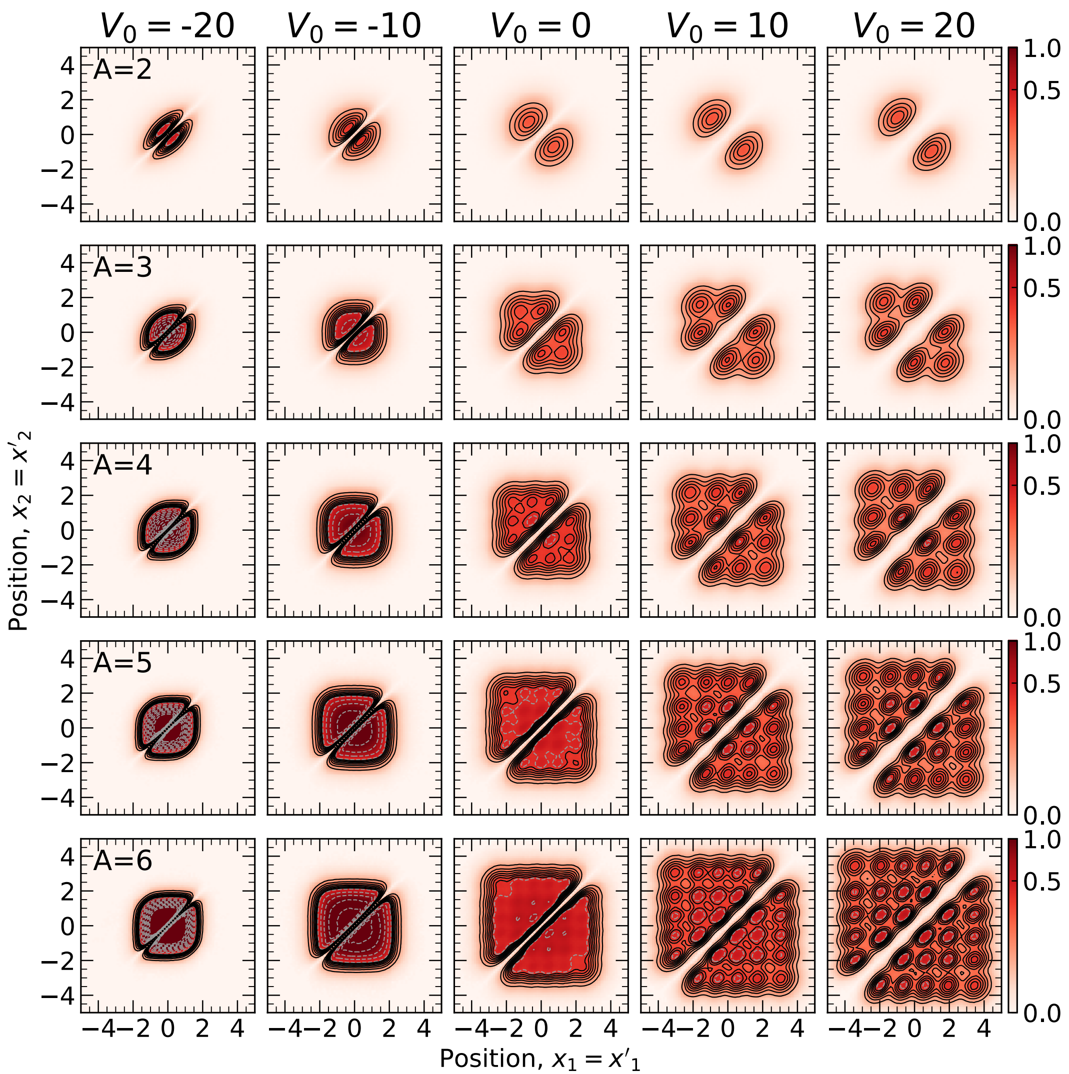


Boson-Fermion duality

Wigner crystallisation

Girardeau & Olshanii, arXiv:cond-mat/0309396
Valiente, PRA 103, L021302 (2021)

Pair distribution function

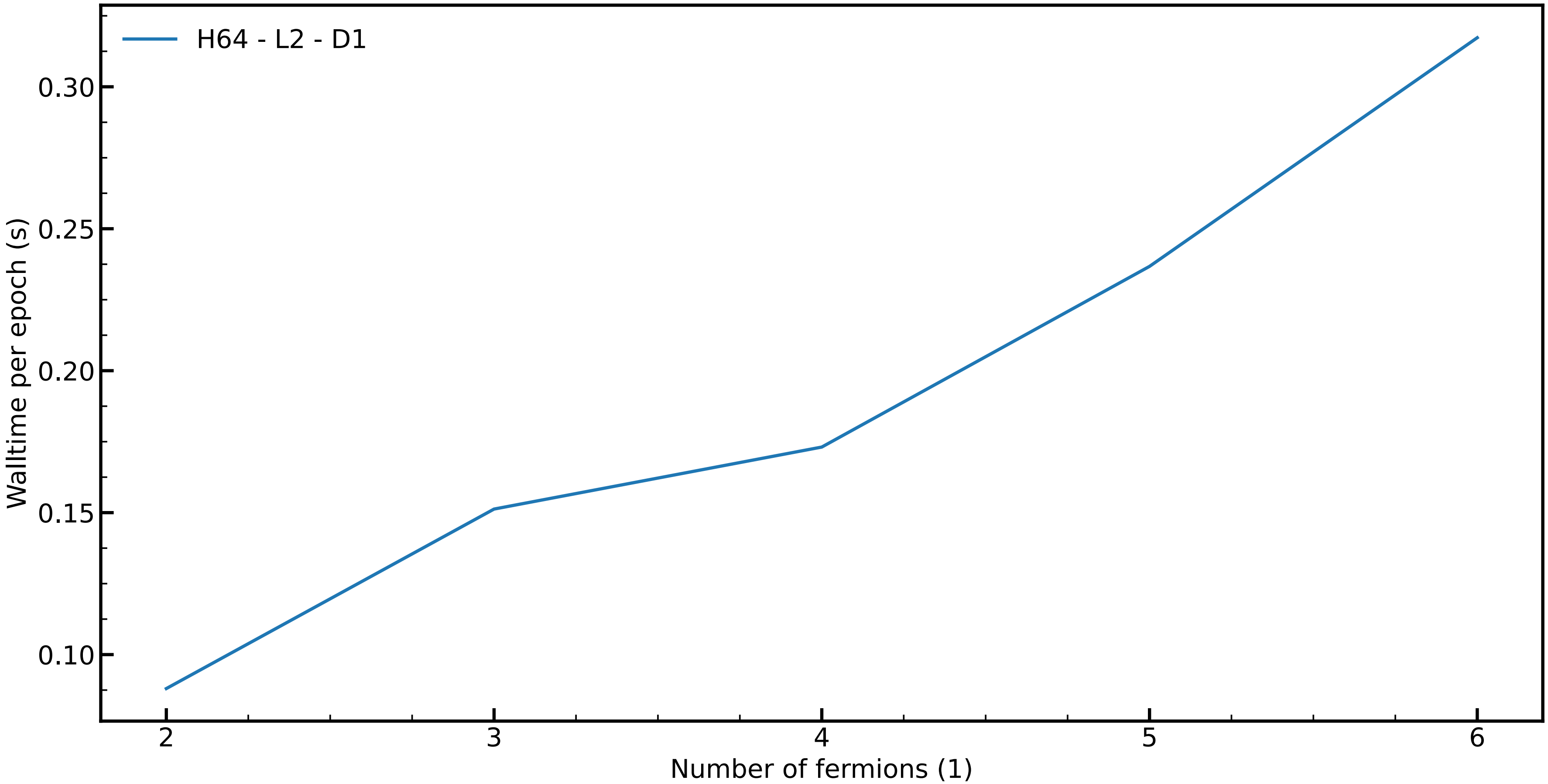


Free case

$$g(x_1, x_2) = \frac{e^{-x_1^2 - x_2^2}}{\pi}(x_1 - x_2)^2 \mathcal{G}^{(A)}(x_1, x_2)$$

J. W. T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

Numerical scaling



- Not so steep in practice!

Outlook

- ◉ Nuclear physics

- ◉ Neural Quantum States

- ◉ Deuteron

  JWT Keeble & A. Rios, Phys. Lett. B **809** 135743 (2020)
  J Rozalen-Sarmiento, JWT Keeble & A. Rios, Eur. Phys. J. Plus **139**, 189 (2024)

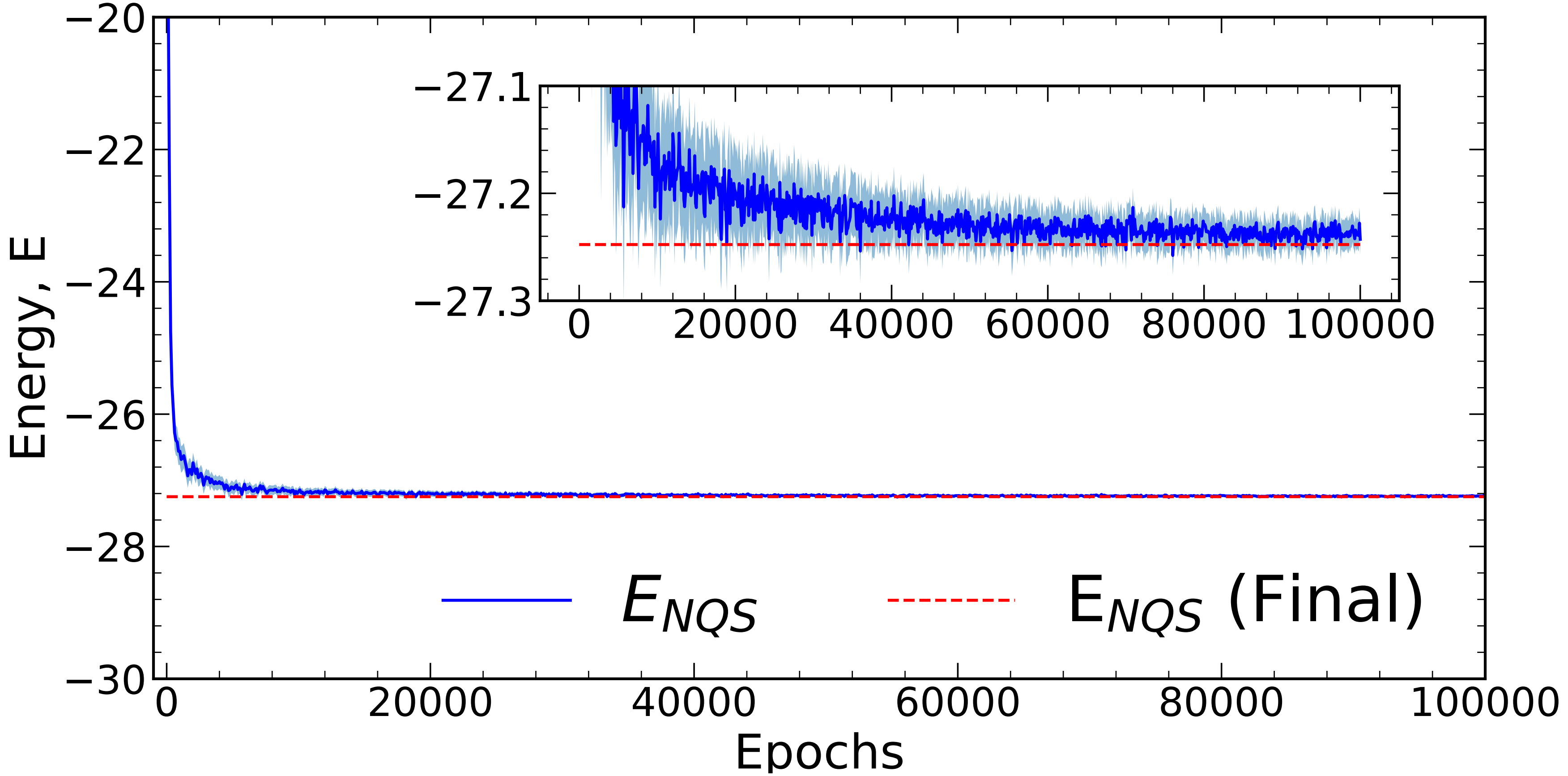
- ◉ 1D spinless (polarised) systems (A>2)

  J.W.T. Keeble, M. Drissi, A. Rojo-Francàs, B. Juliá-Díaz & A. Rios, Phys. Rev. A **108**, 063320 (2023), arxiv:2304.04725

- ◉ **Optimisation strategy**

  M. Drissi, J.W.T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Motivation



−20
−22
−24
−26
−28
−30
Energy, E
0
20000
40000
60000
80000
100000
Epochs
−27.1
−27.2
−27.3
E_{NQS}
E_{NQS} (Final)

Global optimization problem

- **Cost function:** $E(\theta)$
- **Goal**

$$E^* = \min_{\theta \in \mathbb{R}^D} E(\theta)$$

$$\theta^* = \operatorname{argmin}_{\theta \in \mathbb{R}^D} E(\theta)$$

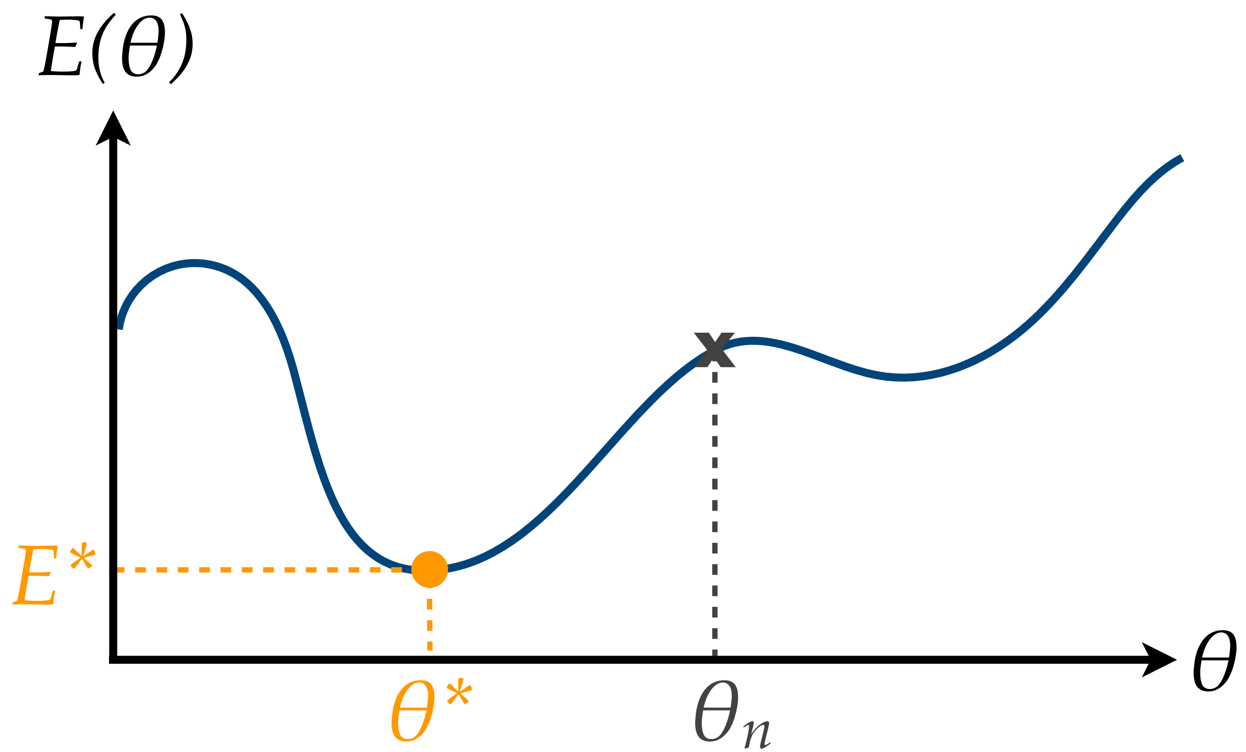
- **Problems**
  - Dimensionality of the problem $D \sim 10^5$
  - $E(\theta)$ highly nonlinear
- **Solution**
  - Solve a sequence of linear/quadratic optimizations

Iterative argument

- Epoch n

$$E_{n+1} = E(\theta_{n+1})$$
$$\theta_{n+1} = \theta_n + \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$$

- with **local model** $M_n(\delta)$ & **trust region** T_n

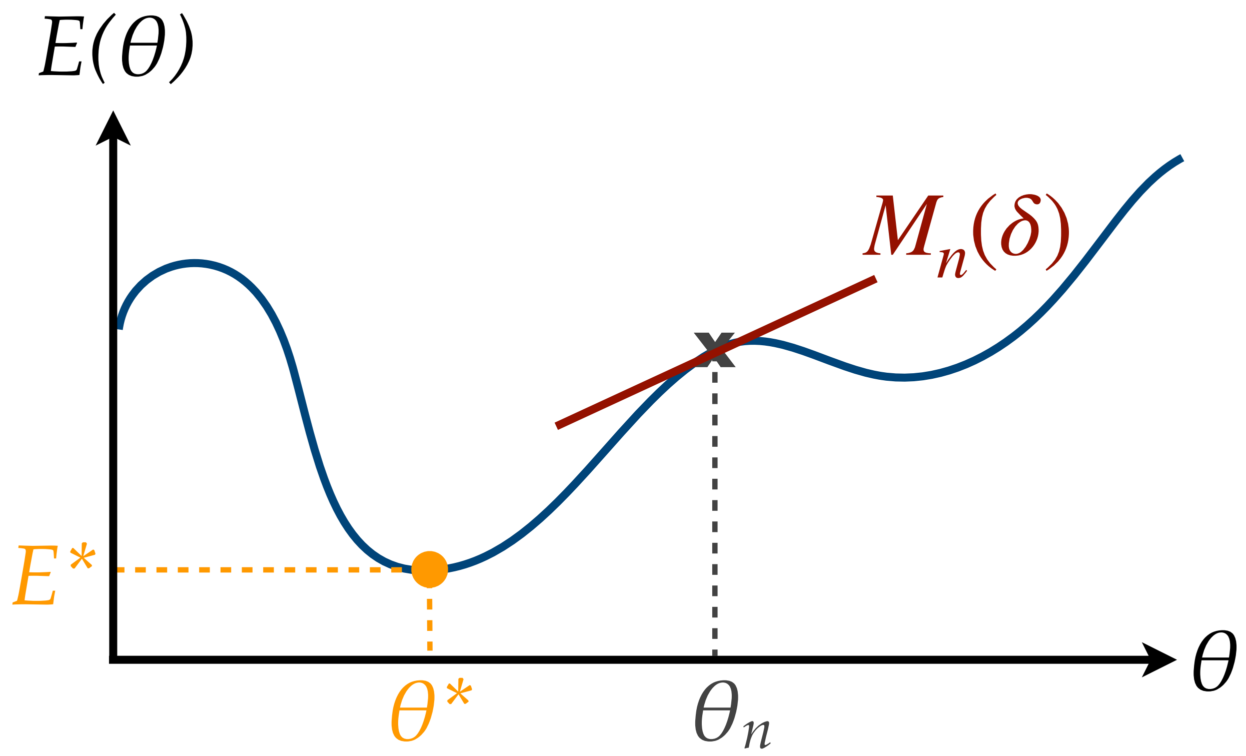


Iterative argument

- Epoch n

$$E_{n+1} = E(\theta_{n+1})$$
$$\theta_{n+1} = \theta_n + \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$$

- with **local model** $M_n(\delta)$ & **trust region** T_n

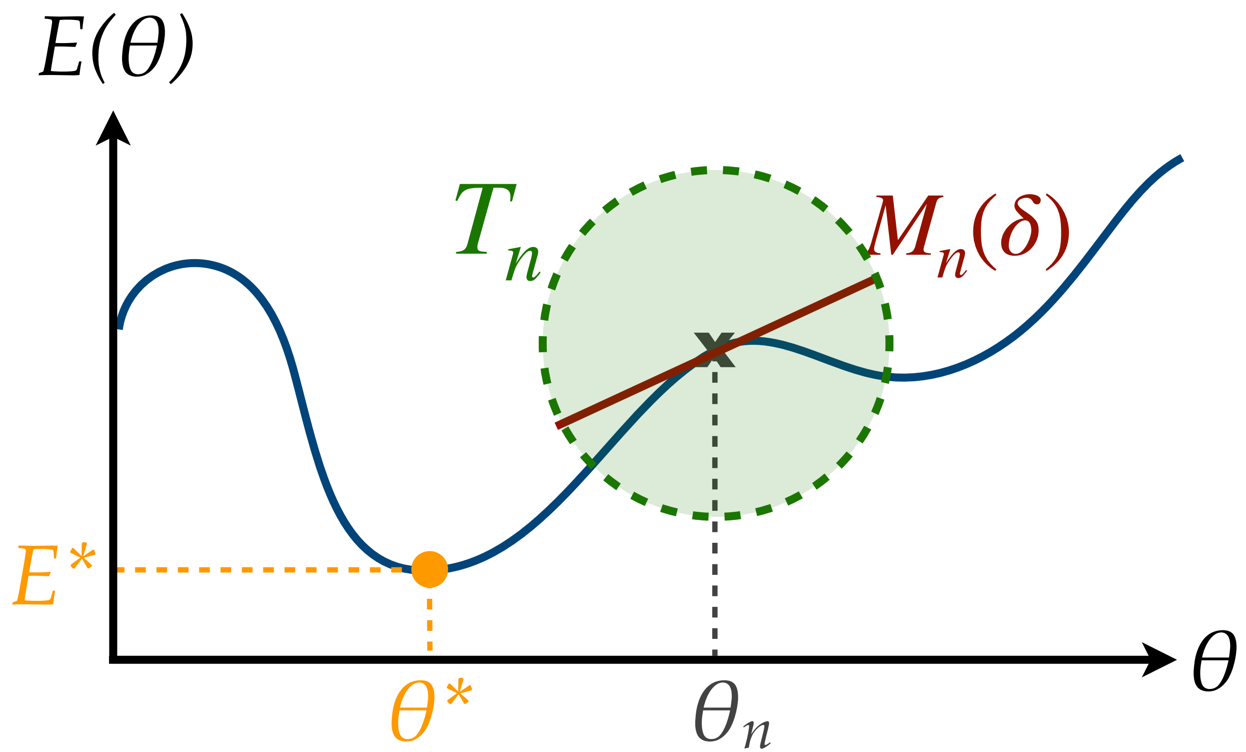


Iterative argument

- Epoch n

$$E_{n+1} = E(\theta_{n+1})$$
$$\theta_{n+1} = \theta_n + \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$$

- with **local model** $M_n(\delta)$ & **trust region** T_n

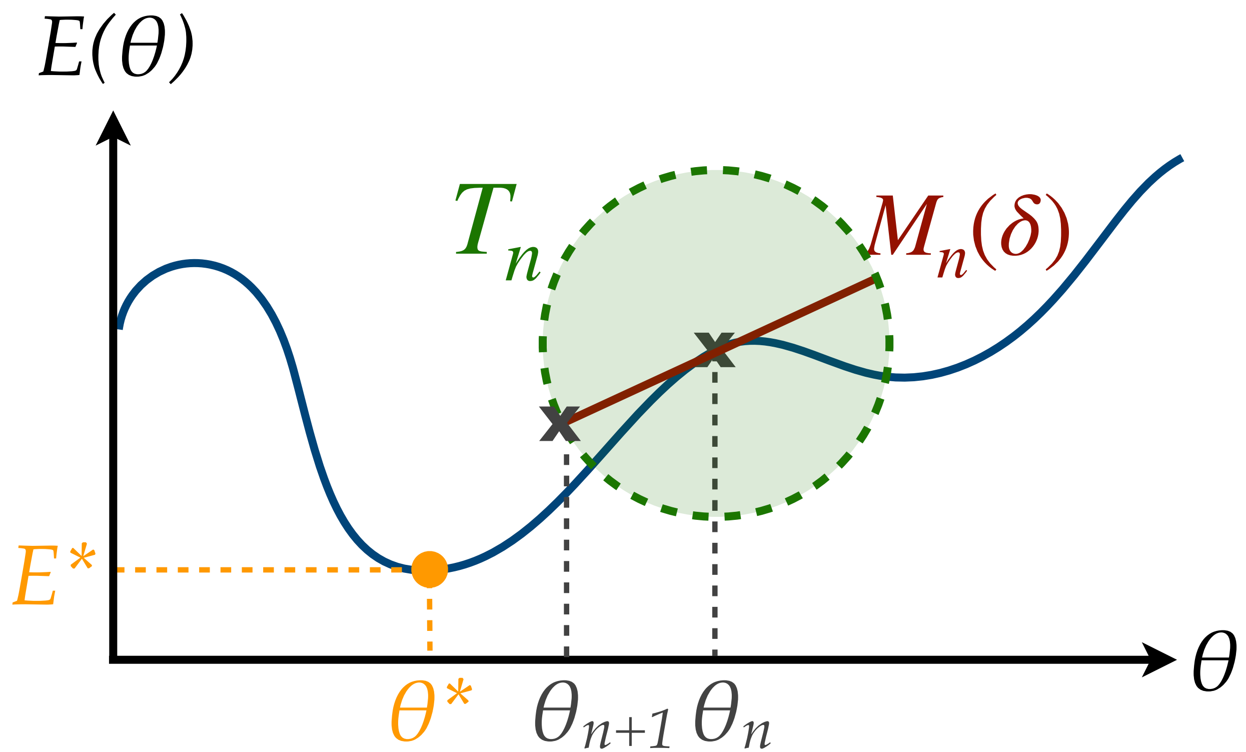


Iterative argument

- Epoch n

$$E_{n+1} = E(\theta_{n+1})$$
$$\theta_{n+1} = \theta_n + \text{argmin}_{\delta \in T_n} M_n(\delta)$$

- with **local model** $M_n(\delta)$ & **trust region** T_n

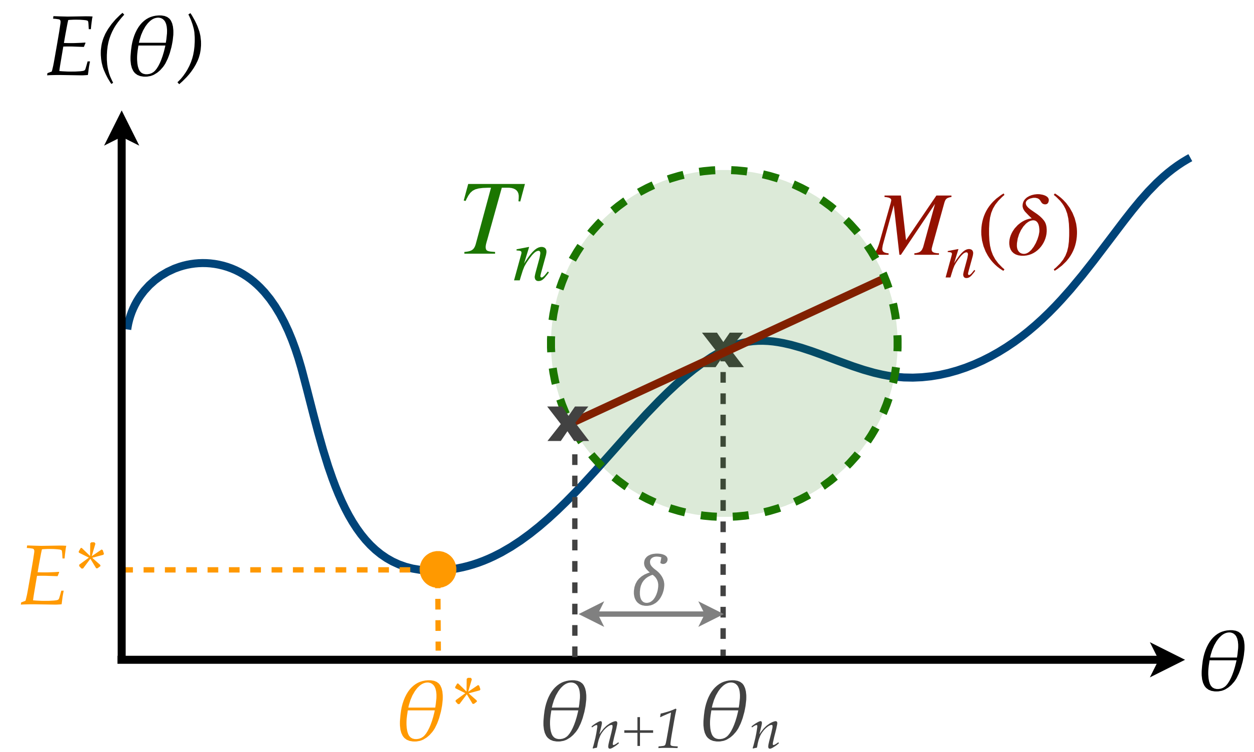


Iterative argument

- Epoch n

$$E_{n+1} = E(\theta_{n+1})$$
$$\theta_{n+1} = \theta_n + \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$$

- with **local model** $M_n(\delta)$ & **trust region** T_n



Regularized quadratic model

- Model
- Regularization

$$M_n(\delta) = \frac{1}{2}\delta^T Q\delta + L^T\delta + C$$
$$M_n(\delta) \leftarrow M_n(\delta) + \frac{1}{2}\delta^T R_n \delta$$

Gradient descent

$$M_n(\delta) = \nabla E(\theta_n)^T\delta + E(\theta_n)$$

$$T_n \equiv \left\{\delta : ||\delta||_2 \leq \alpha ||\nabla E(\theta_n)||_2\right\}$$

$\alpha \equiv$ Learning rate

$$\boxed{\delta_n = -\alpha\nabla E(\theta_n)}$$

Natural gradient descent

Hessian matrix

$$H_{ij} = \partial_{\theta_i}\partial_{\theta_j}E(\theta)$$

$$M_n(\delta) = \frac{1}{2}\delta^T H\delta + \nabla E(\theta)^T\delta + C$$

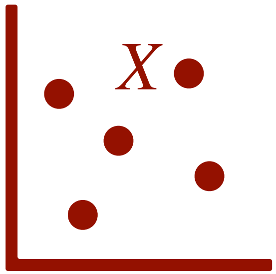
$$T_n \equiv \left\{\delta : \delta^T H(\theta_n)\delta \leq r^2\right\}$$

$$\boxed{\delta_n = -H^{-1}(\theta_n)\nabla E(\theta_n)}$$

Information geometry

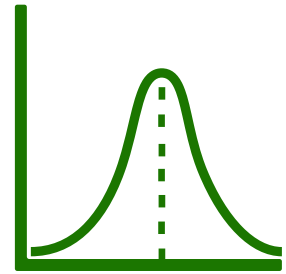
Data

q=target distribution



Model

p_θ =prob. model



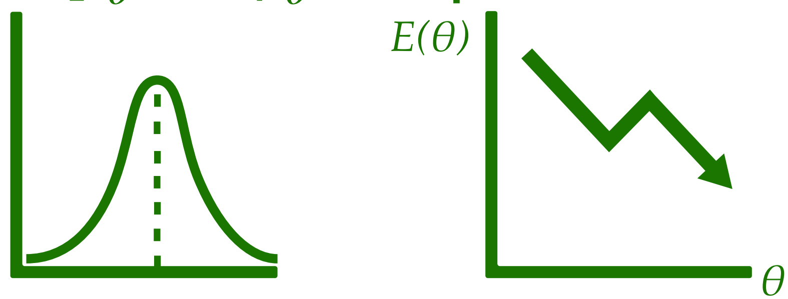
- **Cost function:** cross entropy $L(\theta) = -\mathbb{E}_{X\sim q}\left[\ln p_\theta(X)\right]$
- Kullback-Leibler divergence $D_{KL}(p_{\theta_1}, p_{\theta_2}) = -\mathbb{E}_{X\sim p_{\theta_1}}\left[\ln p_{\theta_1}(X) - \ln p_{\theta_2}(X)\right]$
- KL near minimum $D_{KL}(p_\theta, p_{\theta+\delta\theta}) \approx \frac{1}{2}\delta^T F(\theta)\delta + O(\delta^3)$
- **Fisher** information matrix/“metric” $F_{ij} = \mathbb{E}\left[\partial_{\theta_i} \ln p_\theta \, \partial_{\theta_j} \ln p_\theta\right]$
- Update rule: *natural gradient descent* $\delta_n = -F^{-1}(\theta_n)\nabla E(\theta_n)$

Amari (1997), Amari *Information Geometry & its Applications*

NQS version

Stochastic reconfiguration

$$p_\theta = |\psi_\theta|^2 = \text{prob. model}$$



- **Cost function:** min distance in imaginary time step

$$|\phi\rangle = e^{-\epsilon\hat{H}}|\psi_\theta\rangle \approx (1-\epsilon\hat{H})|\psi_\theta\rangle \Rightarrow L(\theta) = \min_\theta D_{FS}(\phi, \psi_\theta)$$

- Fubini-Study distance $$D_{FS}(\psi_{\theta_1}, \psi_{\theta_2}) = \arccos\sqrt{\frac{\langle\psi_{\theta_1}|\psi_{\theta_2}\rangle\langle\psi_{\theta_2}|\psi_{\theta_1}\rangle}{\langle\psi_{\theta_1}|\psi_{\theta_1}\rangle\langle\psi_{\theta_2}|\psi_{\theta_2}\rangle}}$$
- FS near minimum $$D_{FS}(\psi_\theta, \psi_{\theta+\delta\theta}) \approx \frac{1}{2}\delta^T G(\theta)\delta + O(\delta^3)$$
- Quantum geometric tensor $$G_{ij} = \left\langle \partial_{\theta_i}\psi_\theta \middle| \partial_{\theta_i}\psi_\theta \right\rangle - \left\langle \partial_{\theta_i}\psi_\theta \middle| \psi \right\rangle \left\langle \psi_\theta \middle| \partial_{\theta_j}\psi \right\rangle$$
- Update rule $$\delta_n = -G^{-1}(\theta_n)\nabla E(\theta_n)$$

S. Sorella, Phys. Rev. B 71, 241103 (2005)

Directional updates & KFAC

- Update rule for δ sometimes **too crude**. Call it Δ_n
- Add **scaling factor** α + **momentum** μ & minimise:

$$M_n(\alpha\Delta_n + \mu\delta_{n-1})$$

- Optimal values:

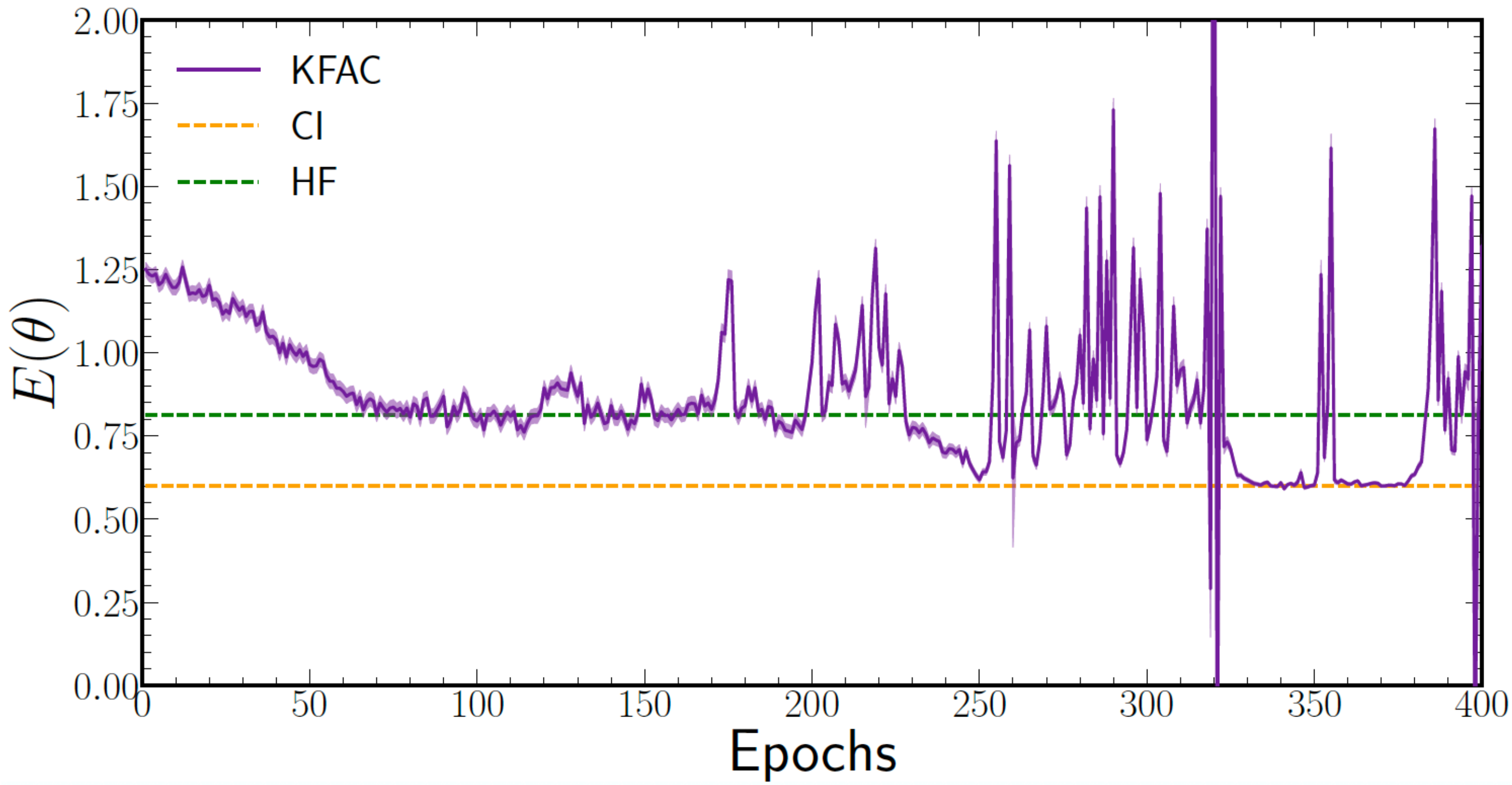
$$\begin{pmatrix}\alpha^* \\ \mu^*\end{pmatrix} = -\begin{pmatrix}\Delta_n^T F(\theta_n)\Delta_n & \Delta_n^T F(\theta_n)\delta_{n-1} \\ \delta_{n-1}^T F(\theta_n)\Delta_n & \delta_{n-1}^T F(\theta_n)\delta_{n-1}\end{pmatrix}^{-1} \begin{pmatrix}\nabla E(\theta_n)^T \Delta_n \\ \nabla E(\theta_n)^T \delta_{n-1}\end{pmatrix}$$

- Here, we can use **exact** Fisher (no need for inversion)
- Approximate inverse Fisher by KFAC version

KFAC: Martens, Grosse, arXiv.1503.05671

Results: KFAC

KFAC: $A = 2,\ V_0 = -10$



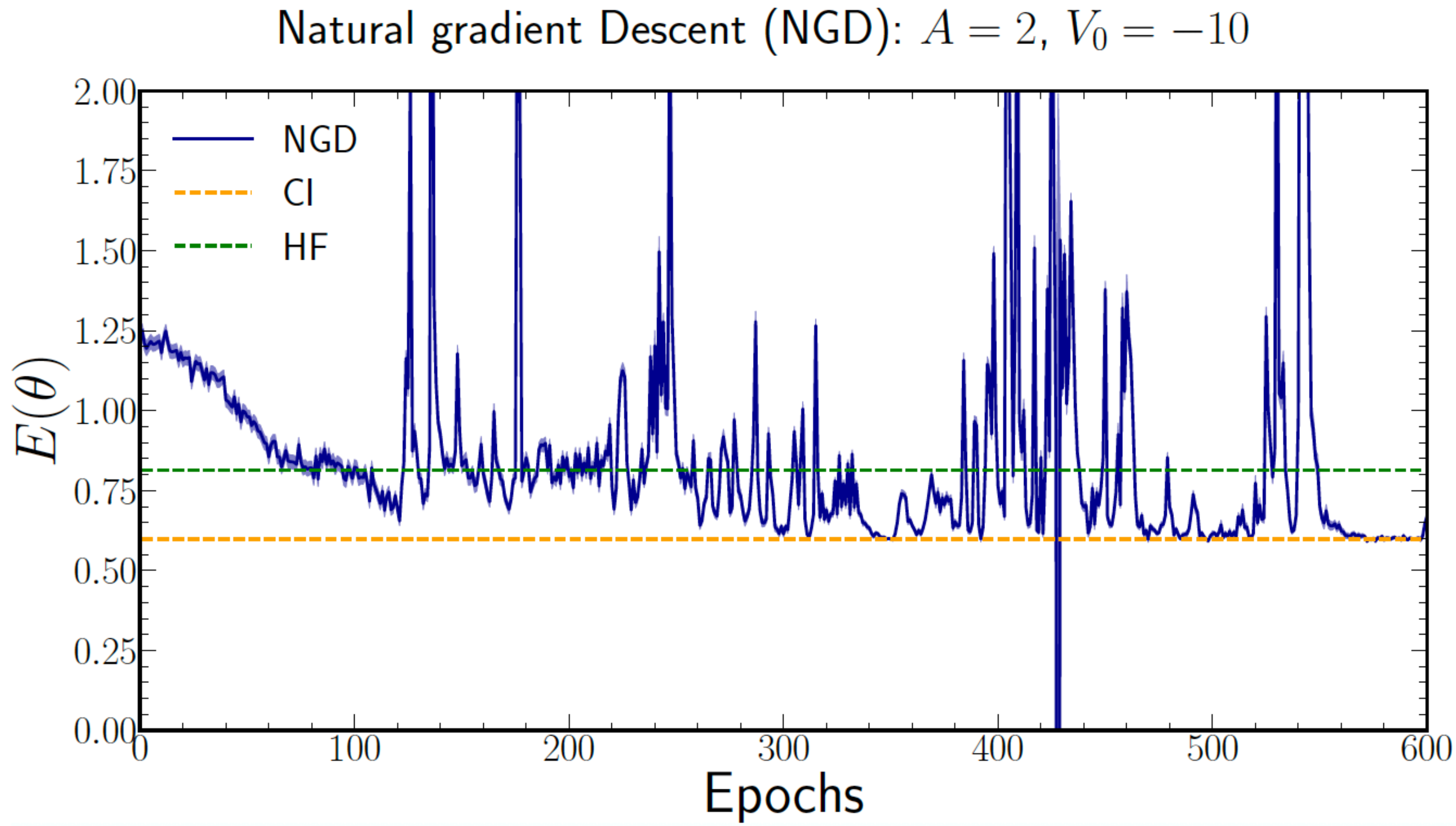
KFAC

Supervised learning?

$$\begin{tikzcd}
 & F \arrow[rrd, bend left, "\lambda\text{ reg}"] & & & \delta_n & \\
E(\theta), \nabla_\theta E(\theta) \arrow[r] \arrow[ru, bend left] & \breve{F}_{\mathrm{KFAC}} \arrow[r, "\gamma\text{ reg}"] & \Delta_n^{\mathrm{KFAC}} \arrow[r] & \alpha, \mu \arrow[r] & \alpha\Delta_n^{\mathrm{KFAC}} + \mu\delta_{n-1} \arrow[r] & \Psi_\theta
\end{tikzcd}$$

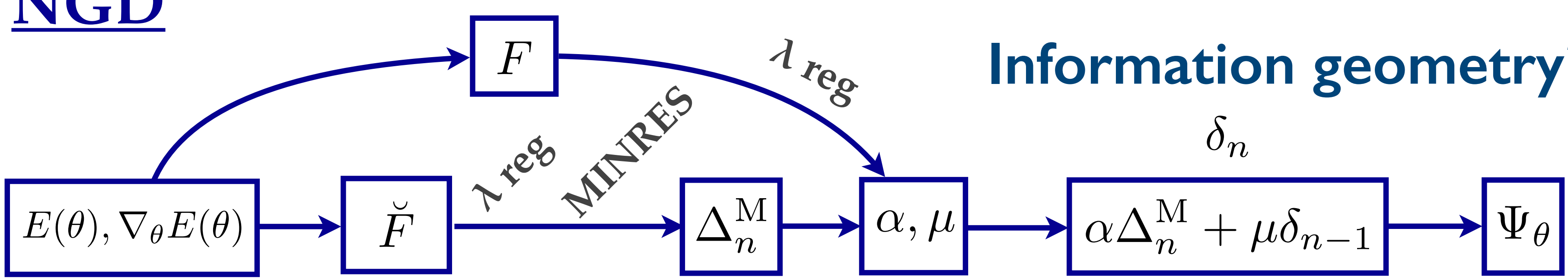
M. Drissi, J. W. T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Results: NGD



NGD

Information geometry?



M. Drissi, J. W. T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Decision geometry 101

- **Cost function** from score $\mathcal{L}(\theta) \equiv -\mathbb{E}_{X\sim p_\theta}\left[\mathcal{S}(X, p_\theta)\right]$
- **Scoring rule:** $\mathcal{S}(X, p_\theta)$ is any integrable function that takes as inputs a probability distribution, p, and a sample, X, and returns a real valued quantity.
- **Expected score:** $S(p, q) \equiv \mathbb{E}_{X\sim p}\left[\mathcal{S}(X, q)\right]$
- **Proper** scoring rules: $S(p, p) \leq S(p, q)$

Dawid, *Ann. Inst. Stat. Math.* 59, 77–93 (2007);
Gneiting, Raftery, *J. Am. Stat. Assoc.* 102, 359–378 (2007)

- Generalized entropy & cross-entropy: $S(p, p)$ $S(p, q)$
- Divergence: $D_S(p, q) = S(p, q) - S(p, p)$
- Divergence near minimum: $D_S(p_\theta,\ p_{\theta+\delta}) = \frac{1}{2}\delta^T G_S(\theta)\delta + O\left(\delta^3\right)$
- G is **semipositive definite**.
- Update rule: decision gradient descent $\delta_n = -G^{-1}(\theta_n)\nabla\mathscr{L}(\theta_n)$

Decision geometry for VMC

- **Information geometry:** $\mathcal{S}(X, p_\theta) = -\ln p_\theta(X)$ ✘
- **VMC scoring rule:** $\mathcal{S}_{\text{VMC}}(X, p_\theta) = -E_{L,\theta}(X)$ ✔
- **Cost function** from score $\mathcal{L}(\theta) = -\mathbb{E}_{X\sim p_\theta}\left[\mathcal{S}(X, p_\theta)\right] = E(\theta)$
- Local energy: $E_{L,\theta}(X) \equiv \Psi_\theta(X)^{-1} H \Psi_\theta(X)$

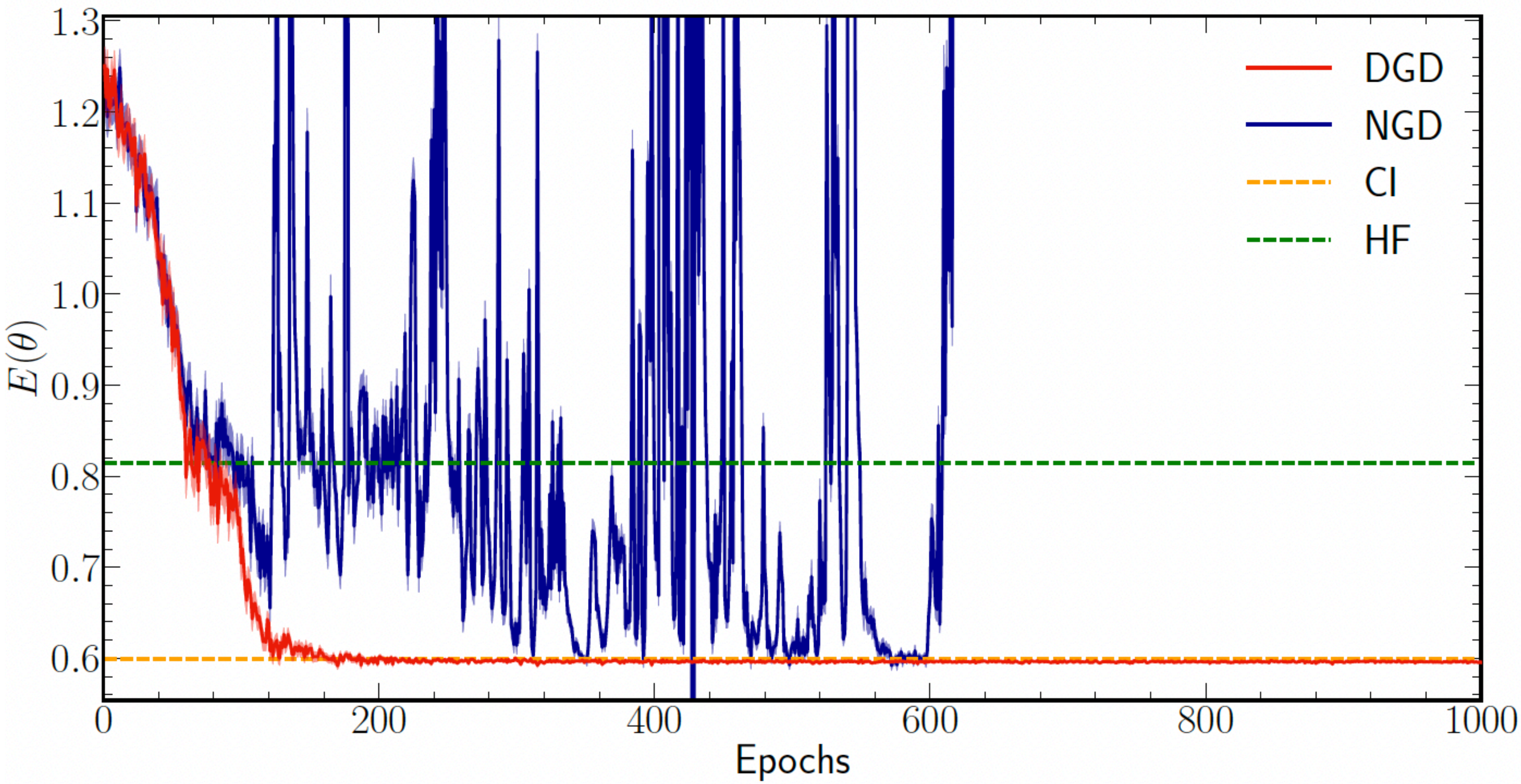
$$= -\frac{1}{2}\boxed{\sum_{i=1}^{A}\left[\partial^2_{x_i} \ln|\Psi_\theta(X)| + \left(\partial_{x_i} \ln|\Psi_\theta(X)|\right)^2\right]} + V(X)$$

Hyvärinen score!

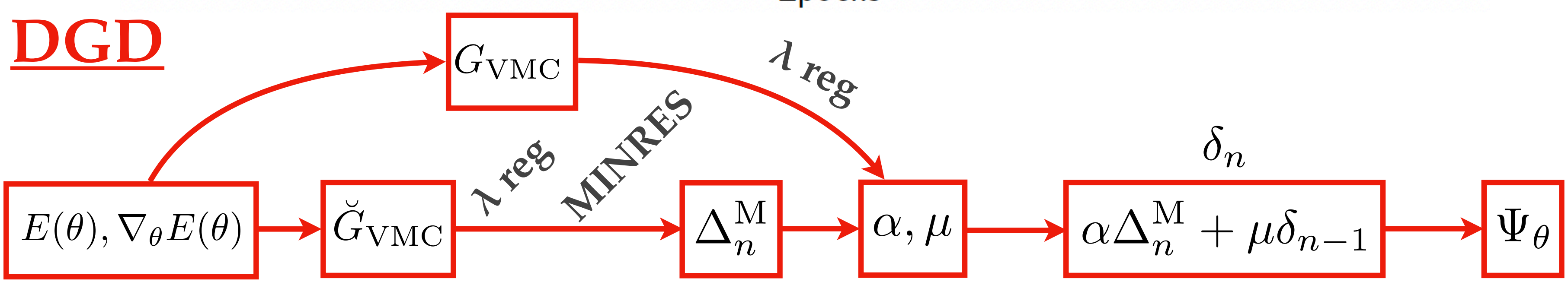
- Metric: $G_{\text{VMC}}(\theta)_{\theta_i\theta_j} \equiv \frac{1}{4}\sum_{k=1}^{A}\mathbb{E}_{X\sim p_\theta}\left[\partial_{\theta_i}\partial_{x_k} \ln p_\theta(X)\ \partial_{\theta_j}\partial_{x_k} \ln p_\theta(X)\right]$
- Update rule: $\delta_n = -\, G^{-1}_{VMC}(\theta_n)\, \nabla E(\theta_n)$

Results

NGD vs DGD: $A = 2$, $V_0 = -10$

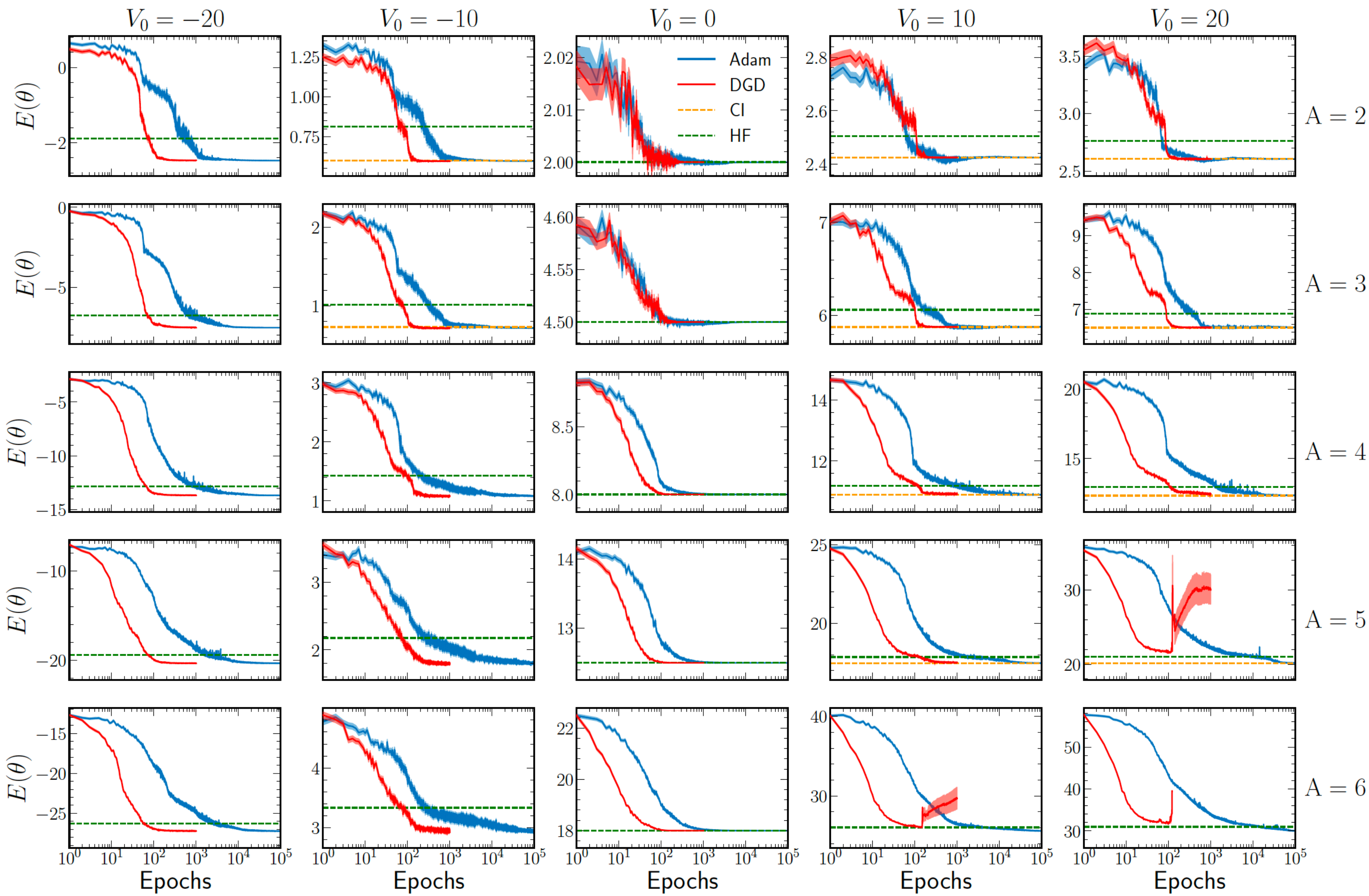


DGD



M. Drissi, J. W. T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Results

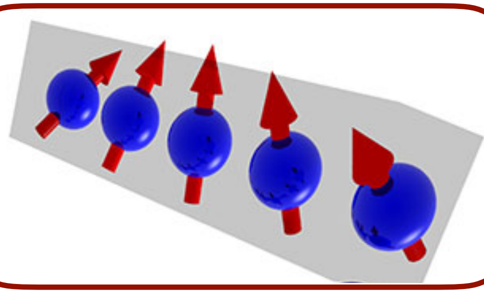


M. Drissi, J. W. T. Keeble, J. Rozalen-Sarmiento & A. Rios, arxiv:2401.17550

Conclusions

- **Machine Learning** techniques can solve **nuclear** & **condensed matter** many-body problems
- **Potential**: same as VMC, possibly less expensive
- **Challenges**
  - **Antisymmetry** (dealt with)
  - **Symmetries in general** (not dealt just yet)
  - **Numerical scaling** in **different** systems
  - **Optimization strategies**

$$P\Psi \rightarrow \Psi$$



Thank you!

UNIVERSITY OF SURREY

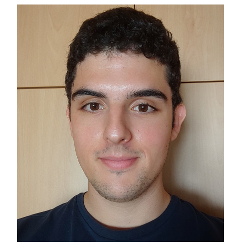
TRIUMF

UNIVERSITAT DE BARCELONA

JWT Keeble

M. Drissi

J. Rozalén
J. Gómez Micó
A. Rojo-Francás
B. Juliá-Díaz



arnau.rios@ub.edu twitter: @riosarnau
https://sites.google.com/view/arnaurios/

Funding from

Science & Technology Facilities Council

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Investigación Programa Ramón y Cajal

AGENCIA ESTATAL DE INVESTIGACIÓN

UNIVERSITY OF SURREY UNIVERSITAT DE BARCELONA Departament de Física Quàntica i Astrofísica