Towards the Optimizer that NQS Deserves

M. Drissi, J.W.T Keeble, J. Rozalén Sarmiento & A. Rios, Phil. Trans. R. Soc. A. 382 (2024)



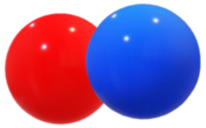

J. Rozalén Sarmiento, M. Drissi, J.W.T. Keeble, A. Rios

CPAN 2025, FNUC. November 19th, Valencia, 2025

ICCUB Institut de Ciències del Cosmos, Universitat de Barcelona

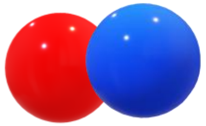
Excelencia María de Maeztu 04/2025–03/2031
Ministerio de Ciencia, Innovación y Universidades
Agencia Estatal de Investigación
FNUC Red Temática de Física Nuclear

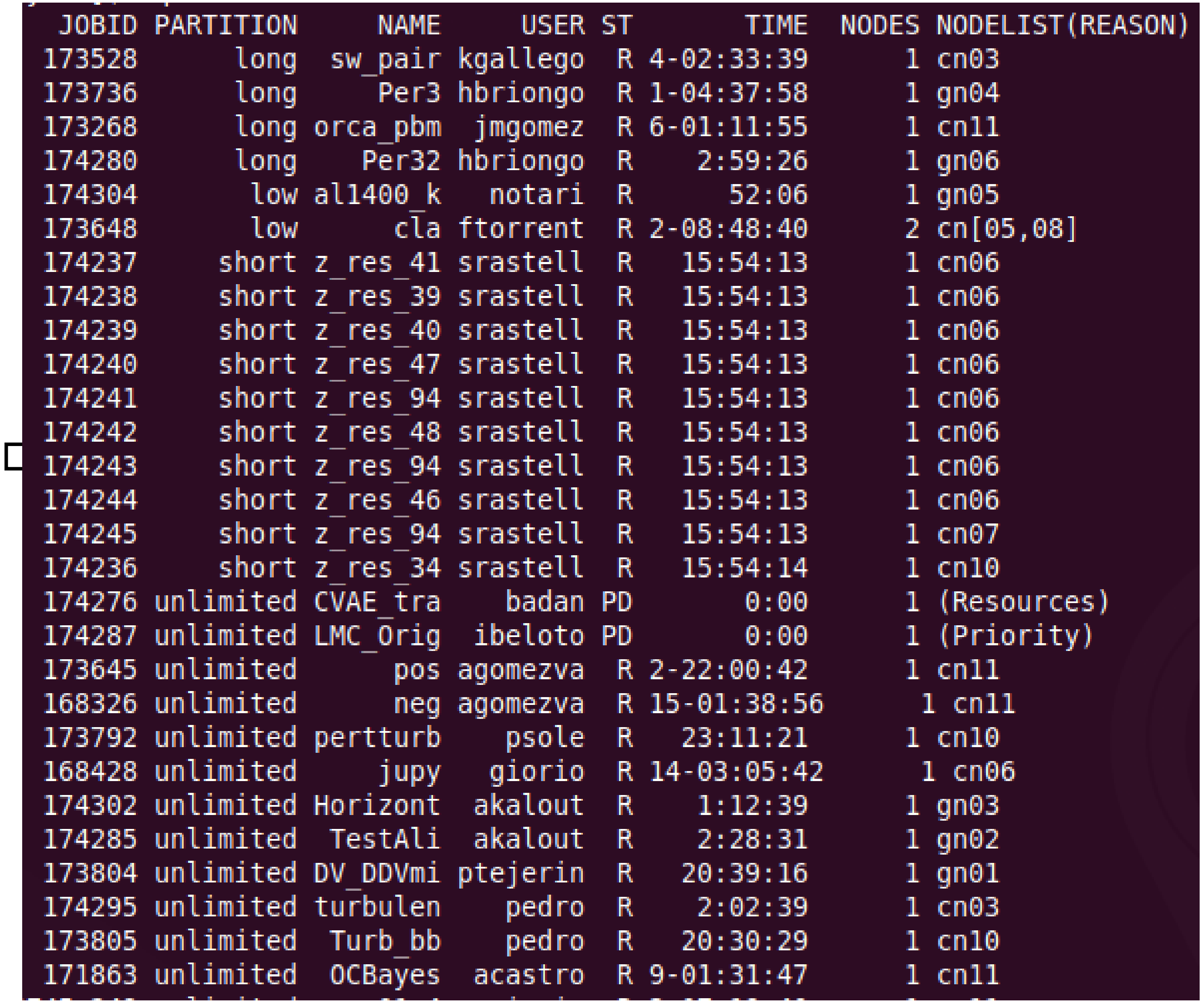
The computation problem



$$\psi_\theta(X, \sigma, \tau)$$

The computation problem



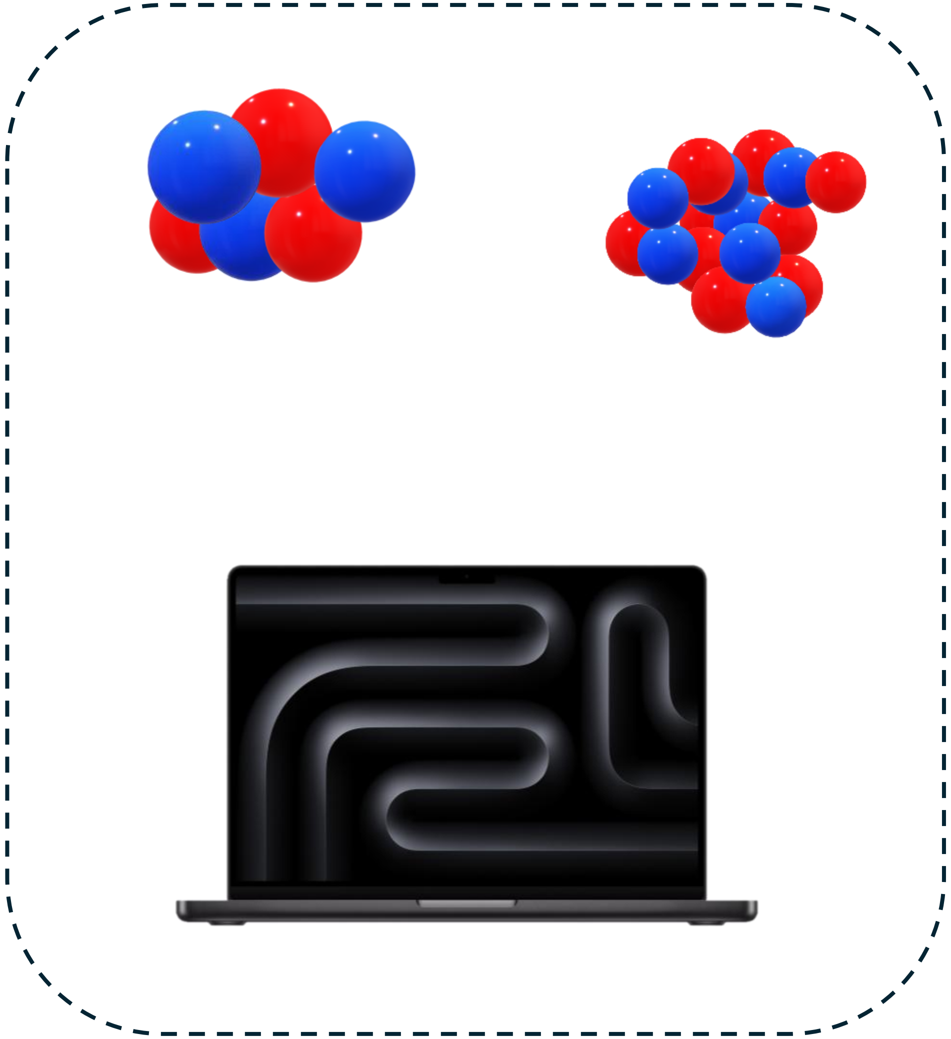


```
 JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
173528      long  sw_pair kgallego  R 4-02:33:39      1 cn03
173736      long     Per3 hbriongo  R 1-04:37:58      1 gn04
173268      long orca_pbm  jmgomez  R 6-01:11:55      1 cn11
174280      long    Per32 hbriongo  R    2:59:26      1 gn06
174304       low al1400_k   notari  R      52:06      1 gn05
173648       low      cla ftorrent  R 2-08:48:40      2 cn[05,08]
174237     short z_res_41 srastell  R   15:54:13      1 cn06
174238     short z_res_39 srastell  R   15:54:13      1 cn06
174239     short z_res_40 srastell  R   15:54:13      1 cn06
174240     short z_res_47 srastell  R   15:54:13      1 cn06
174241     short z_res_94 srastell  R   15:54:13      1 cn06
174242     short z_res_48 srastell  R   15:54:13      1 cn06
174243     short z_res_94 srastell  R   15:54:13      1 cn06
174244     short z_res_46 srastell  R   15:54:13      1 cn06
174245     short z_res_94 srastell  R   15:54:13      1 cn07
174236     short z_res_34 srastell  R   15:54:14      1 cn10
174276 unlimited CVAE_tra    badan PD       0:00      1 (Resources)
174287 unlimited LMC_Orig  ibeloto PD       0:00      1 (Priority)
173645 unlimited      pos agomezva  R 2-22:00:42      1 cn11
168326 unlimited      neg agomezva  R 15-01:38:56      1 cn11
173792 unlimited pertturb    psole  R   23:11:21      1 cn10
168428 unlimited     jupy   giorio  R 14-03:05:42      1 cn06
174302 unlimited Horizont  akalout  R    1:12:39      1 gn03
174285 unlimited  TestAli  akalout  R    2:28:31      1 gn02
173804 unlimited DV_DDVmi ptejerin  R   20:39:16      1 gn01
174295 unlimited turbulen    pedro  R    2:02:39      1 cn03
173805 unlimited  Turb_bb    pedro  R   20:30:29      1 cn10
171863 unlimited  OCBayes  acastro  R 9-01:31:47      1 cn11
```

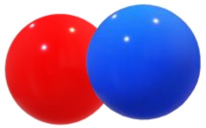
$$\psi_\theta(X, \sigma, \tau)$$

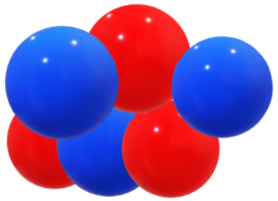
Exponential Scaling Problem

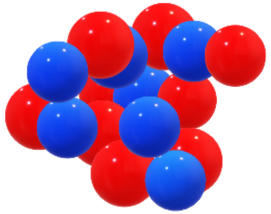
Exponential Scaling Problem

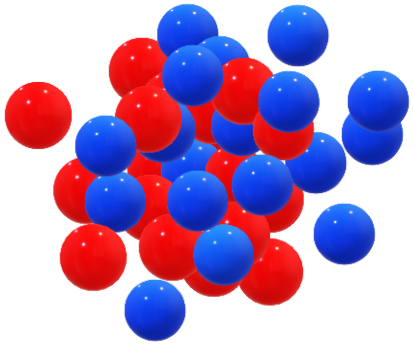


Exponential Scaling Problem



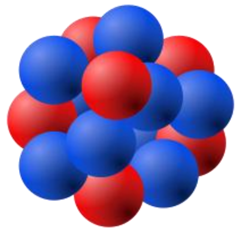






Neural Quantum States

- **Goal:** Solve for the wave function of a nucleus, ψ

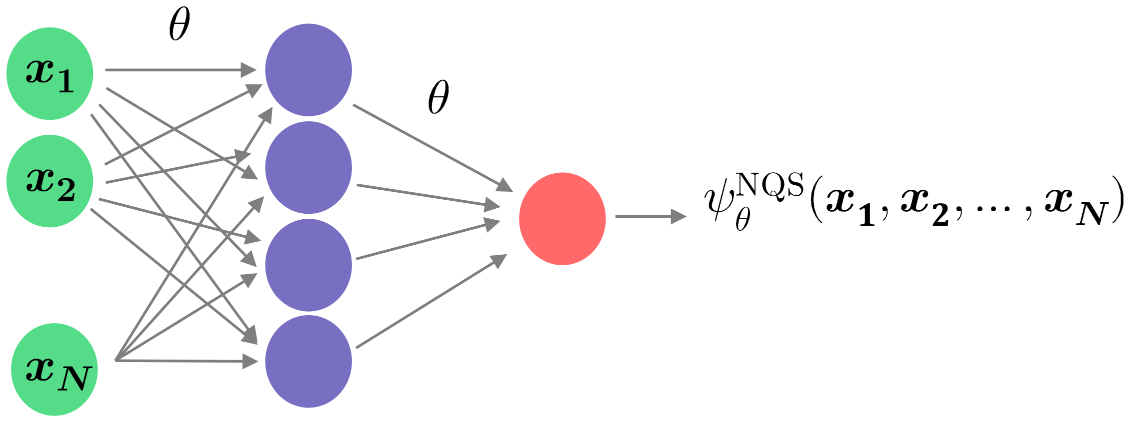


$\longrightarrow |\psi\rangle$

- **How?** Rayleigh-Ritz variational principle

$$\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \geq E_{\mathrm{GS}}$$

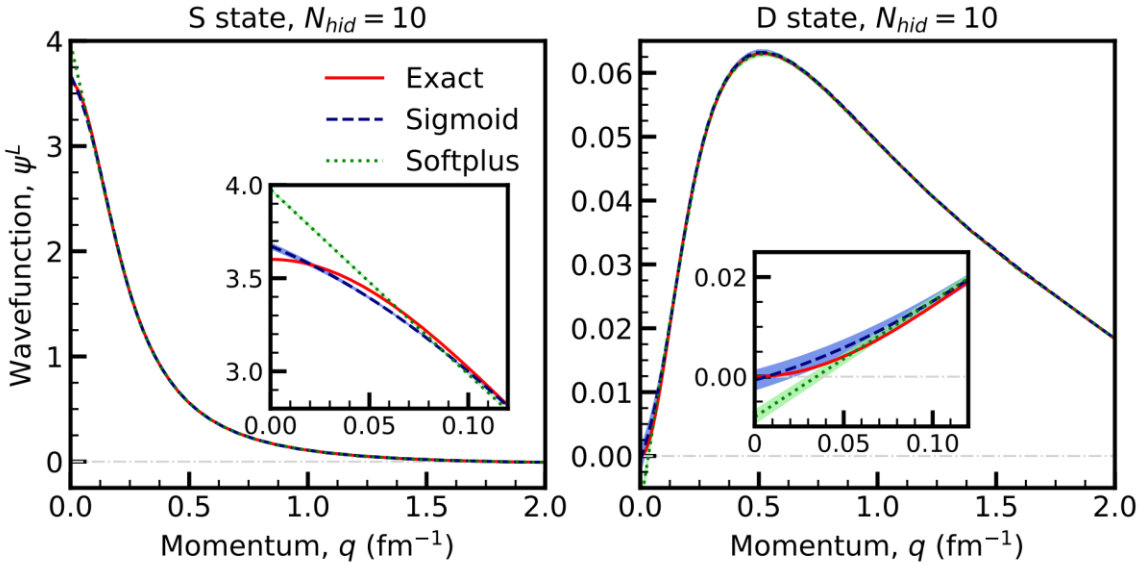
$$|\psi_\theta\rangle = \int |x\rangle \; \psi_\theta(\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_N}) \mathrm{d}^{3\mathrm{N}} x$$

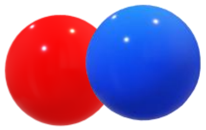


- **NQS Ansatz:** $\psi_{\mathrm{NQS}}(\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_N}; \boldsymbol{\theta})$

NQS for Nuclei

2020

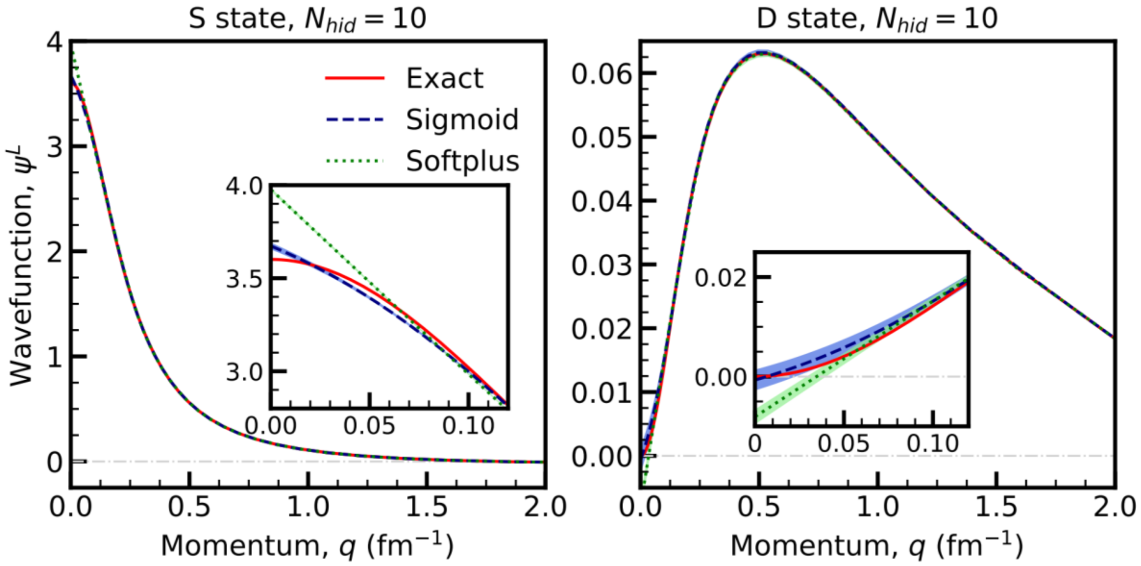


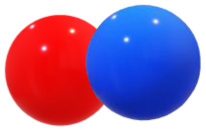


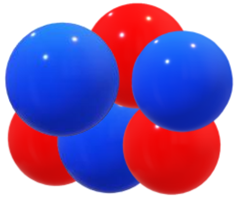
J. Keeble & A. Rios, Phys. Lett. B **809** (2020)

NQS for Nuclei

2020 → …





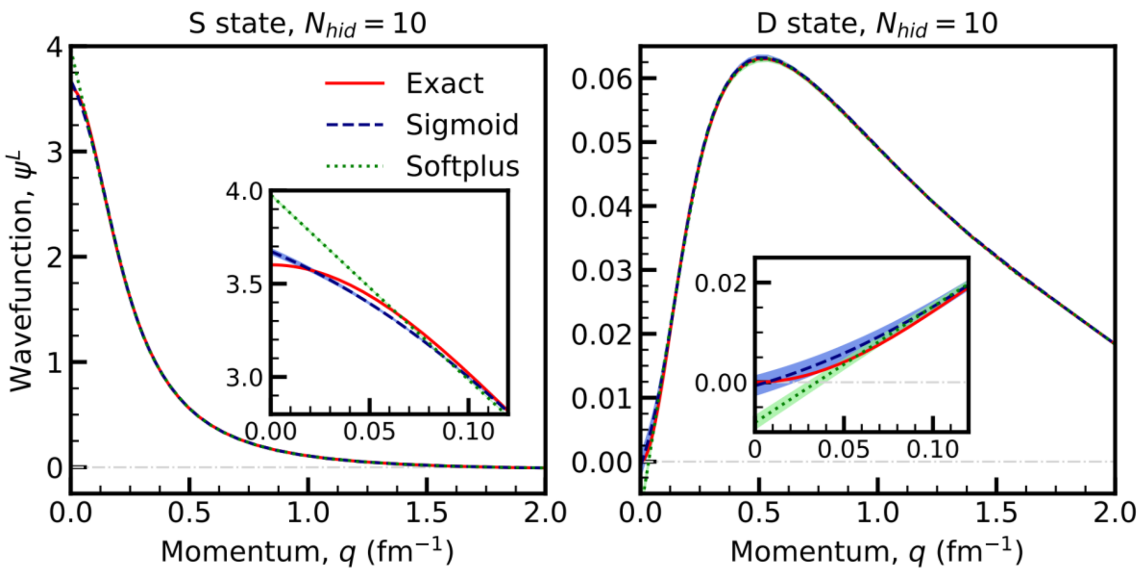


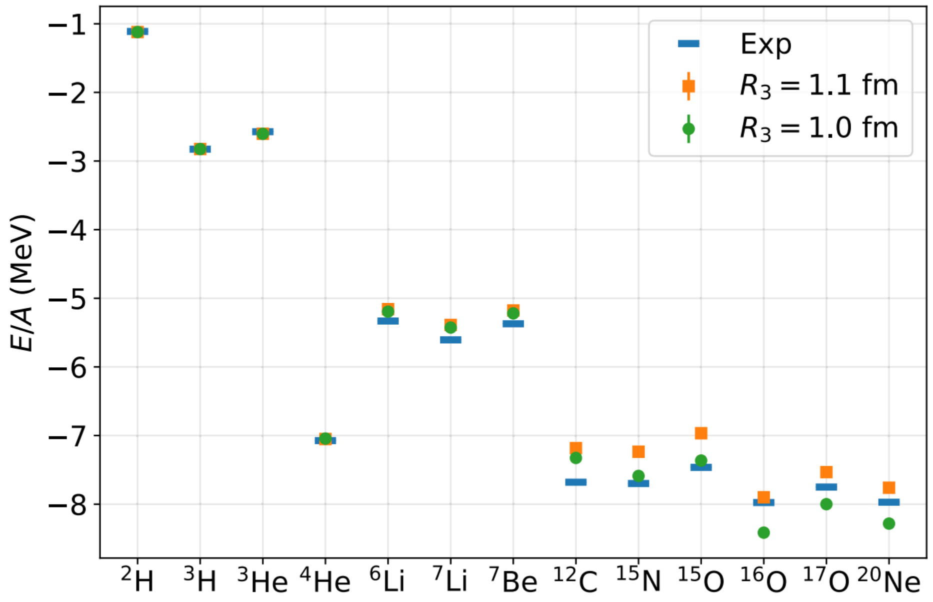
J. Keeble & A. Rios, Phys. Lett. B **809** (2020)

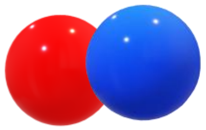
A. Gnech et al, Few-Body Syst. **63** (2022)

NQS for Nuclei

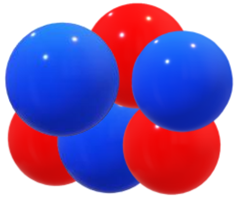
2020 → … → **2024**





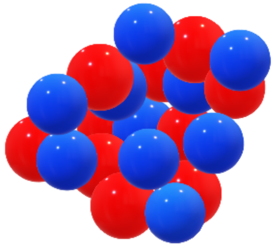


J. Keeble & A. Rios, Phys. Lett. B **809** (2020)



A. Gnech et al, Few-Body Syst. **63** (2022)

…



A. Gnech et al, PRL **133** (2024)

Optimisation:
A Big Challenge in NQS

The Optimisation Step

$$\theta_{n+1} = \theta_n - \alpha \nabla E(\theta_n)$$

- **Learning rate:** How big the steps are
- **Gradient:** In which direction we move

The Optimisation Step

$$\theta_{n+1} = \theta_n - \alpha G^{-1} \nabla E(\theta_n)$$

- **Learning rate:** How big should the step be?
- **Gradient:** In which direction do we move?
- **Metric tensor:** What is the geometry of the energy landscape?

The Optimisation Step

$$\theta_{n+1} = \theta_n - \alpha G^{-1} \nabla E(\theta_n)$$

- **Learning rate:** How big should the step be?
- **Gradient:** In which direction do we move?
- **Metric tensor:** What is the geometry of the energy landscape?

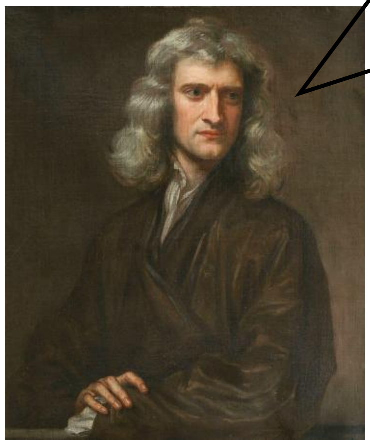
$\alpha?$ $G?$

The Metric Tensor

The Metric Tensor

$$\theta_{n+1} = \theta_n - \alpha G^{-1} \nabla E(\theta_n)$$

$$G(\theta_n) = H(\theta_n)$$



$$G(\theta_n) = \mathrm{QGT}(\theta_n)$$

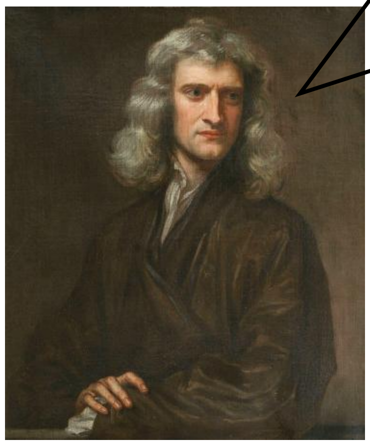
Modern Literature

The Metric Tensor

$$\theta_{n+1} = \theta_n - \alpha G^{-1} \nabla E(\theta_n)$$

$$G(\theta_n) = H(\theta_n)$$

$$G(\theta_n) = \mathrm{QGT}(\theta_n)$$



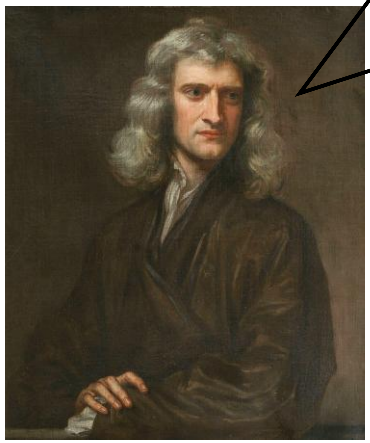
Modern
Literature

- ✓ Very precise
- ✓ $\alpha = 1$
- ✗ Costly to compute

The Metric Tensor

$$\theta_{n+1} = \theta_n - \alpha G^{-1} \nabla E(\theta_n)$$

$$G(\theta_n) = H(\theta_n)$$



$$G(\theta_n) = \mathrm{QGT}(\theta_n)$$

Modern Literature

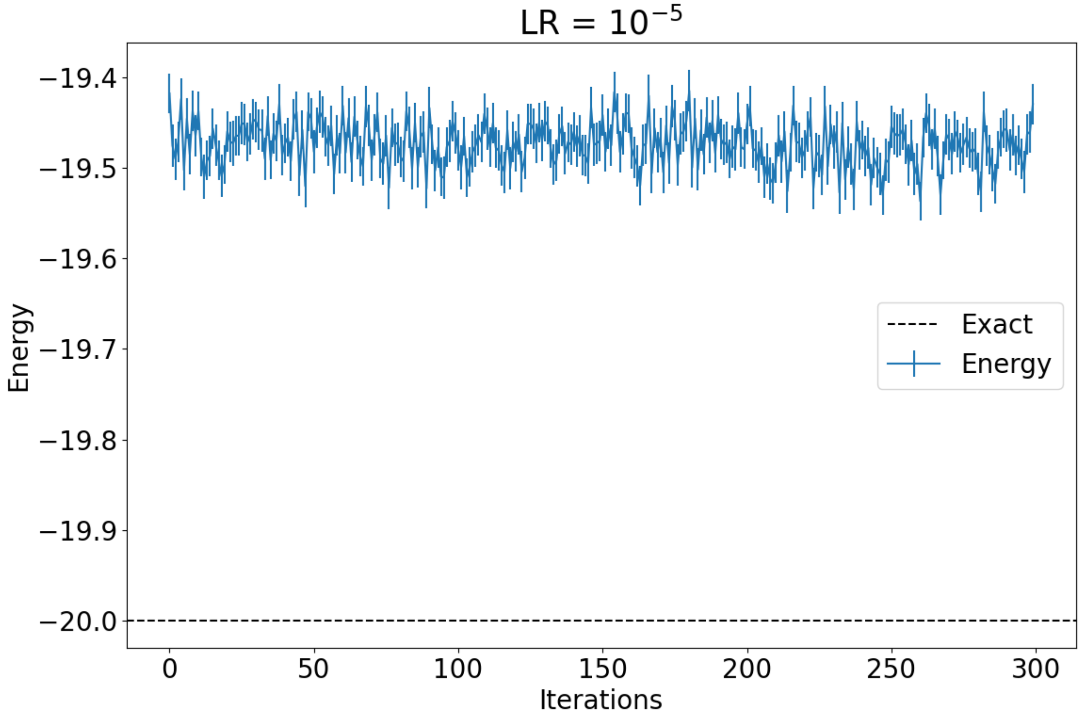
- ✓ Very precise
- ✓ $\alpha = 1$
- ✗ Costly to compute

- ✓ Quite good
- ✗ $\alpha \ll 1$
- ✓ Cheap to compute

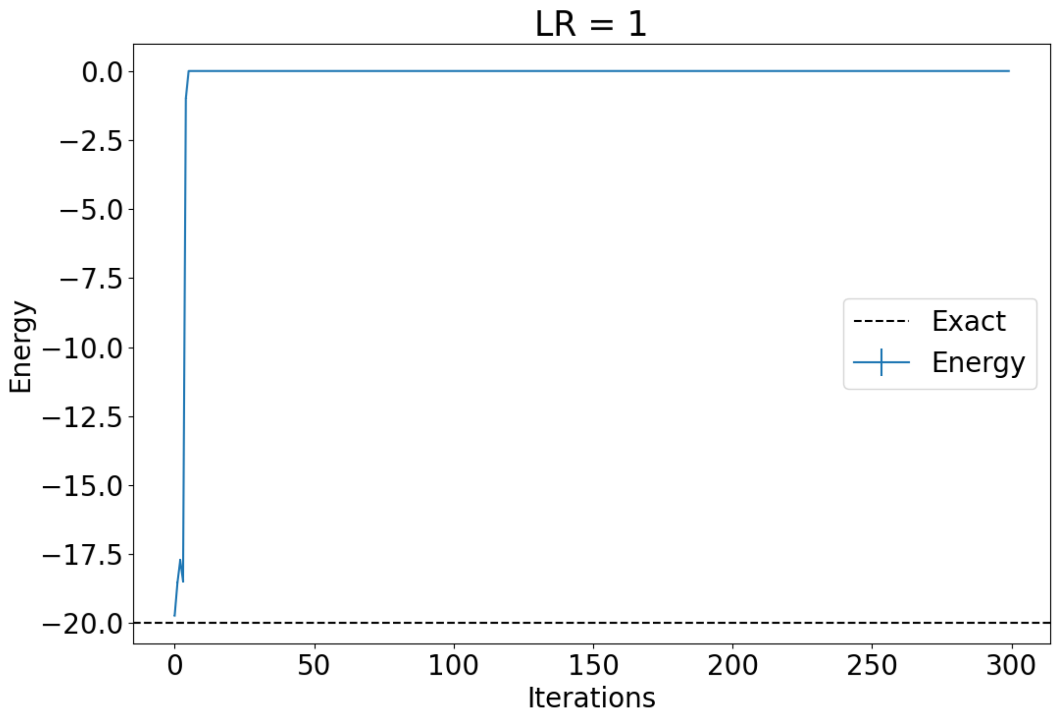
The Learning Rate

A Common Problem...

α is too small



α is too large



$$H = -\sum_{i=1}^{N} \sigma_i^{(z)} \sigma_{i+1}^{(z)} + V \sum_{j=1}^{N} \sigma_j^{(x)}$$

- N=20 particles
- QGT

Can we do better?

Decisional Gradient Descent

A new proposal

- **QGT:** based on the Hilbert distance between states, $< \phi_i | \phi_j >$

$$\mathrm{QGT}(\theta)_{\theta_i\theta_j} = \mathbb{E}_{X\sim p_\theta}\left[\left(\partial_{\theta_\mathrm{i}} \ln \psi_\theta\left(\mathrm{X}\right) - \mathbb{E}_{X\sim p_\theta}[\partial_{\theta_\mathrm{i}} \ln \psi_\theta\left(\mathrm{X}\right)]\right)^* \left(\partial_{\theta_\mathrm{j}} \ln \psi_\theta\left(\mathrm{X}\right) - \mathbb{E}_{X\sim p_\theta}\left[\partial_{\theta_\mathrm{j}} \ln \psi_\theta\left(\mathrm{X}\right)\right]\right)\right]$$

- **DGD:** based on the Hamiltonian distance between states, $< \phi_i | H | \phi_j >$

$$G_{\mathrm{VMC}}(\theta)_{\theta_i\theta_j} \equiv \frac{1}{4}\sum_{k=1}^{A} \mathbb{E}_{X\sim p_\theta}\left[\partial_{\theta_i}\partial_{x_k} \ln p_\theta(X)\ \partial_{\theta_j}\partial_{x_k} \ln p_\theta(X)\right]$$

A new proposal

- **QGT:** based on the Hilbert distance between states, $< \phi_i | \phi_j >$

$$\mathrm{QGT}(\theta)_{\theta_i\theta_j} = \mathbb{E}_{X\sim p_\theta}\left[\left(\partial_{\theta_\mathrm{i}} \ln\psi_\theta\left(\mathrm{X}\right) - \mathbb{E}_{X\sim p_\theta}\left[\partial_{\theta_\mathrm{i}} \ln\psi_\theta\left(\mathrm{X}\right)\right]\right)^*\left(\partial_{\theta_\mathrm{j}} \ln\psi_\theta\left(\mathrm{X}\right) - \mathbb{E}_{X\sim p_\theta}\left[\partial_{\theta_\mathrm{j}} \ln\psi_\theta\left(\mathrm{X}\right)\right]\right)\right]$$

- **DGD:** based on the Hamiltonian distance between states, $< \phi_i | H | \phi_j >$

$$G_{\mathrm{VMC}}(\theta)_{\theta_i\theta_j} \equiv \frac{1}{4}\sum_{k=1}^{A} \mathbb{E}_{X\sim p_\theta}\left[\partial_{\theta_i}\partial_{x_k} \ln p_\theta(X)\ \partial_{\theta_j}\partial_{x_k} \ln p_\theta(X)\right]$$

$\hookrightarrow$ Similar to the Hessian $\longrightarrow$ $\alpha \simeq 1$

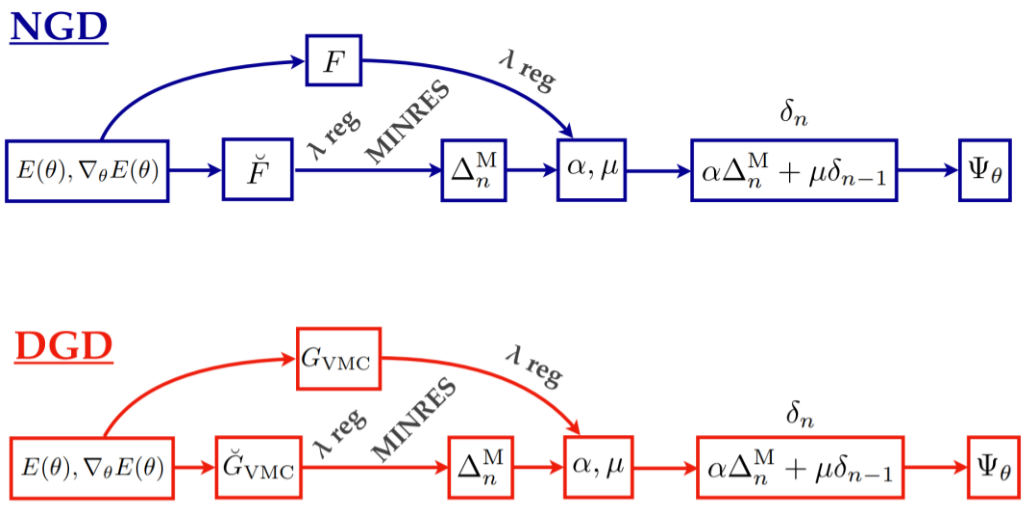
Results: Fermions on the continuum

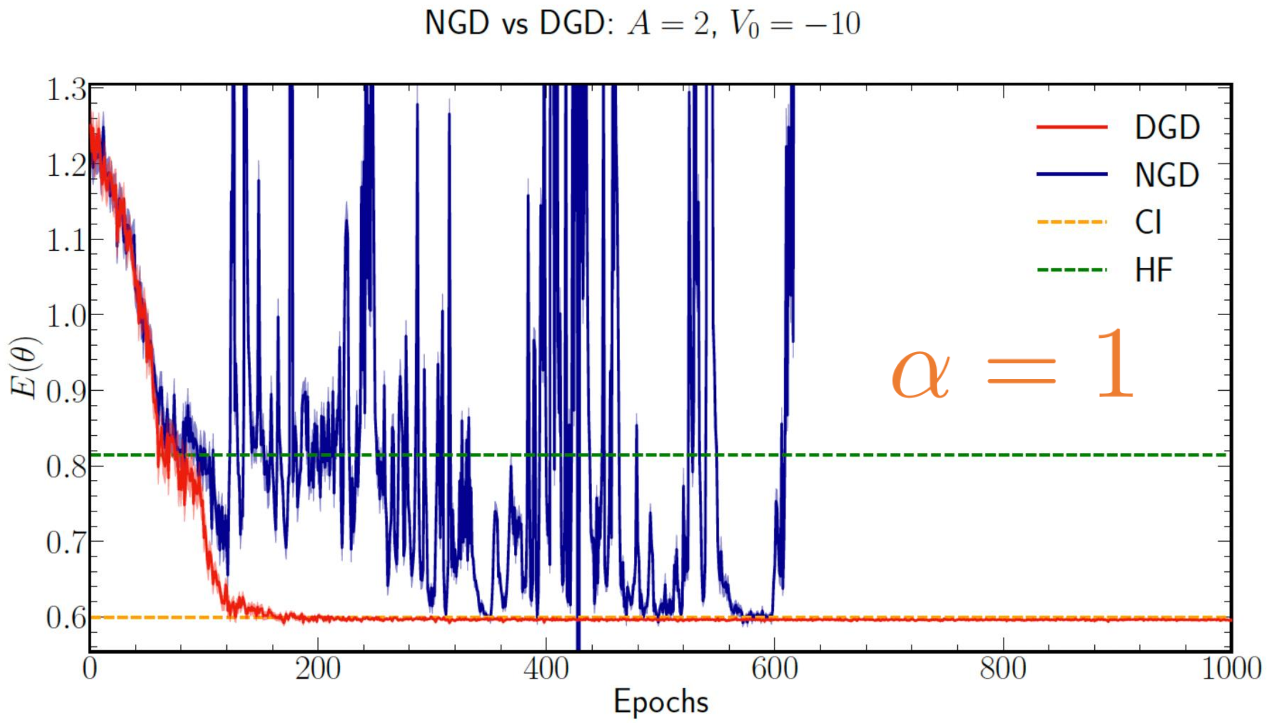
DGD vs. QGT

- **System:** chain of **1D spinless fermions**

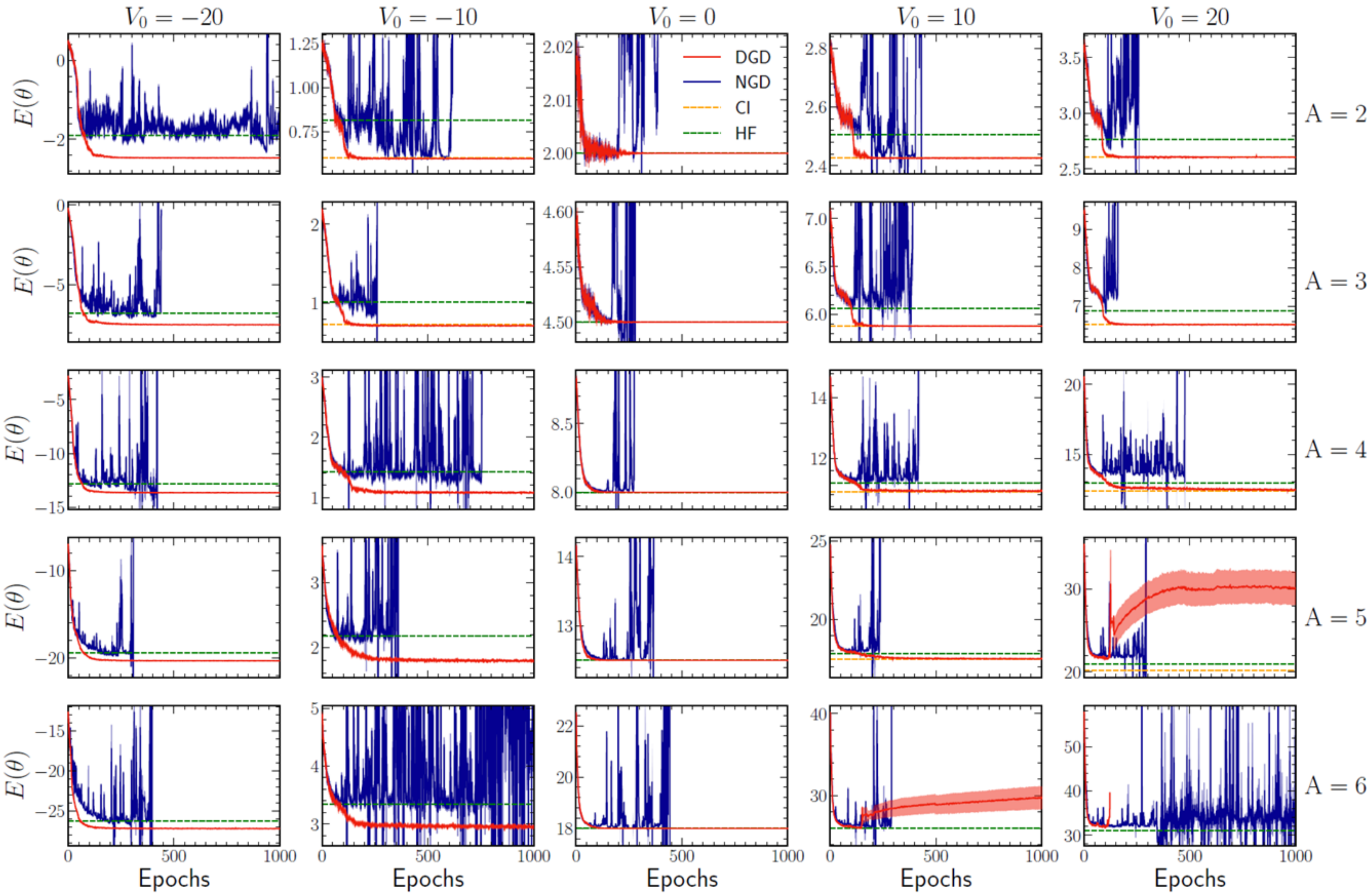
$$H = -\frac{1}{2}\sum_{i=1}^{A}\nabla_i^2 + \frac{1}{2}\sum_{i=1}^{A}x_i^2 + \frac{V_0}{\sqrt{2\pi}\sigma_0}\sum_{i<j} e^{-\frac{(x_i-x_j)^2}{2\sigma_0^2}}$$

- **Comparison:** **Natural Gradient Descent** vs. **Decisional Gradient Descent**





DGD vs. QGT



M. Drissi, J.W.T Keeble, J. Rozalén Sarmiento & A. Rios, Phil. Trans. R. Soc. A. **382** (2024)

Results: Fermions on a lattice

DGD for lattice systems

- **Transverse-Field Ising Model:**

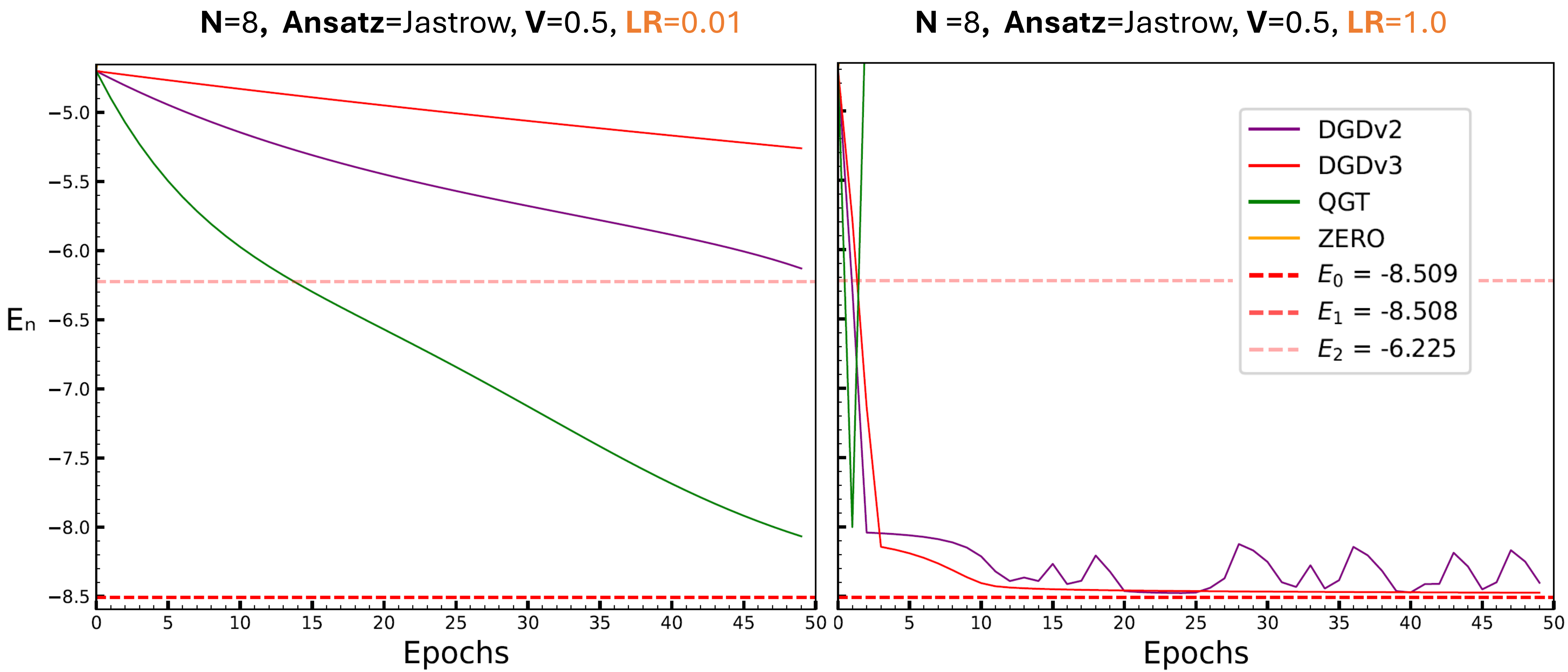
$$G(\theta_n)_{kl} = 4\mathrm{e}^{\beta(c^{\mathrm{MFA}}-\mu)} \prod_{i=1}^{N} \cosh \frac{\beta \mathrm{d}_{\mathrm{i}}^{\mathrm{MFA}}}{2} \times \mathrm{QGT}(\theta_n)_{kl}$$
$$+ \sum_{j=1}^{N} \tanh \frac{-\beta d_j^{MFA}}{2} \times E_{\sigma\sim|\psi_\theta|^2} \left\{ \mathrm{sgn}(\sigma_j) \left| \frac{T_j[\psi_\theta](\sigma)}{\psi_\theta(\sigma)} \right| \left(\partial_k \ln T_j\,[\psi_\theta](\sigma) \right)^* \left(\partial_{\mathrm{l}} \ln T_j\,[\psi_\theta](\sigma) \right) \right\}$$

- **J_1-J_2 Heisenberg:**

$$G(\theta_n)_{kl} = \left(\frac{1}{\beta} - \mu + \mathrm{J}_1 \mathrm{N}_{<\mathrm{i,j}>} + \mathrm{J}_2 \mathrm{N}_{<<\mathrm{i,j}>>} \right) \times 2\mathrm{Re}(\mathrm{QGT}(\theta_n)_{kl}) \; + 2Re \left\{ \mathrm{J}_1 \sum_{<i,j>} \underset{\sigma\sim|\psi_\theta|^2}{E} \left[(\mathrm{sgn}(\sigma_i)\mathrm{sgn}(\sigma_j) \right] \right\}$$
$$\times \left(\partial_k \ln \psi_\theta\,(\sigma) - \frac{\psi_\theta\left(f_{ij}(\sigma)\right)}{\psi_\theta(\sigma)} \partial_k \ln \psi \left(f_{ij}(\sigma) \right) - \left(1 - \frac{\psi_\theta\left(f_{ij}(\sigma)\right)}{\psi_\theta(\sigma)} \right) \times E_{\sigma\sim|\psi_\theta|^2}[\partial_k \ln \psi_\theta\,(\sigma)] \right)$$
$$\times \left(\partial_{\mathrm{l}} \ln \psi_\theta\,(\sigma) - \frac{\psi_\theta\left(f_{ij}(\sigma)\right)}{\psi_\theta(\sigma)} \partial_l \ln \psi \left(f_{ij}(\sigma) \right) - \left(1 - \frac{\psi_\theta\left(f_{ij}(\sigma)\right)}{\psi_\theta(\sigma)} \right) \times E_{\sigma\sim|\psi_\theta|^2}[\partial_l \ln \psi_\theta\,(\sigma)] \right)$$

DGD vs. QGT: Ising

$$H = -\sum_{i=1}^{N} \sigma_i^{(z)} \sigma_{i+1}^{(z)} + V \sum_{j=1}^{N} \sigma_j^{(x)}$$

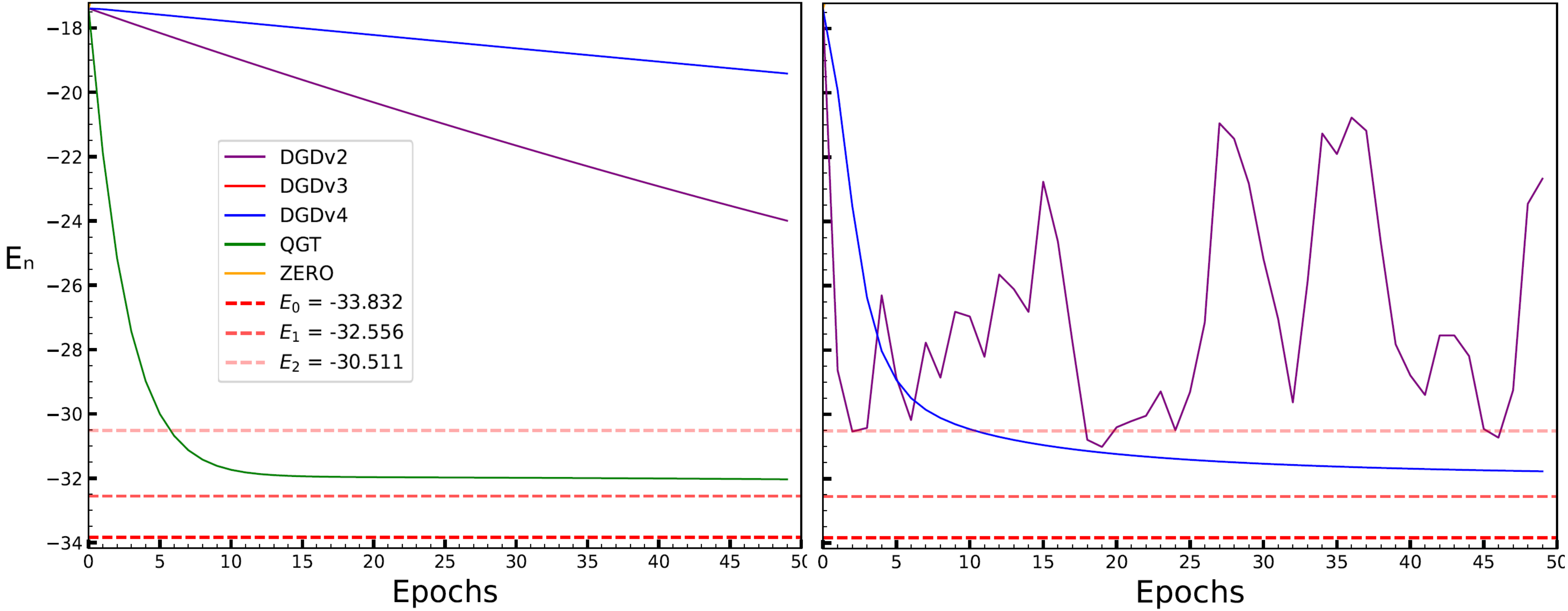


DGD vs. QGT: J_1-J_2

$$H = J_1 \sum_{<i,j>} \vec{S_i} \cdot \vec{S_j} + J_2 \sum_{\ll i,j \gg} \vec{S_i} \cdot \vec{S_j}$$

N=16 (4x4), **Ansatz**=Jastrow, $\mathbf{J_2}$=0.5, **LR**=0.01

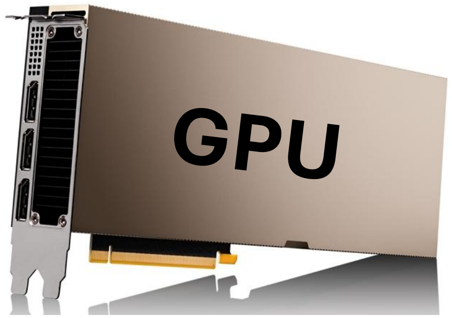
N=16 (4x4), **Ansatz**=Jastrow, $\mathbf{J_2}$=0.5, **LR**=1.0

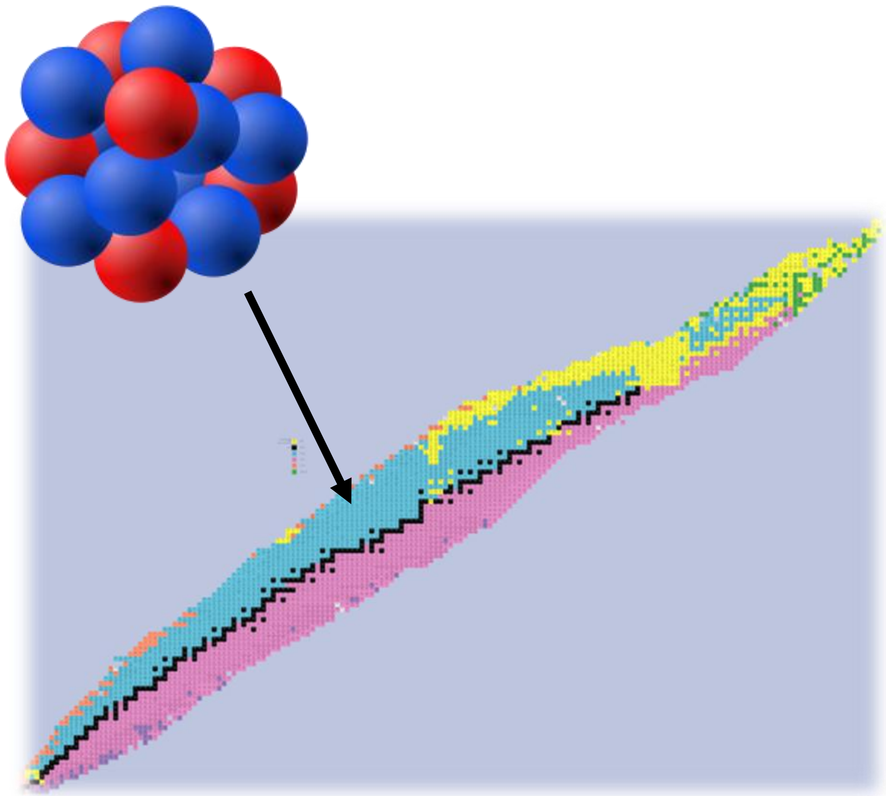


Final Thoughts

Takeaway

$$G(\theta_n) +$$





Thank you

J. Rozalén Sarmiento, M. Drissi, J.W.T. Keeble, A. Rios

https://doi.org/10.1098/rsta.2024.0057 arXiv:2401.17550

M. Drissi, J.W.T Keeble, J. Rozalén Sarmiento & A. Rios, Phil. Trans. R. Soc. A. **382** (2024)

javi.rozalen@icc.ub.edu

ICCUB Institut de Ciències del Cosmos UNIVERSITAT DE BARCELONA

EXCELENCIA MARÍA DE MAEZTU 04/2025–03/2031

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

AGENCIA ESTATAL DE INVESTIGACIÓN

FNUC RED TEMÁTICA DE FÍSICA NUCLEAR