

Learning the Ground State with Neural Quantum States

J. Rozalén Sarmiento, A. Rios

HadNucMat Workshop, 2025, Barcelona



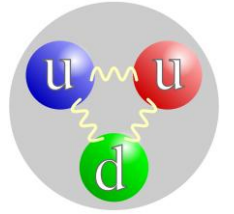
Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA



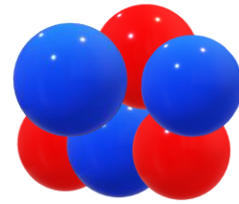
UNIVERSITAT_{DE}
BARCELONA

“Ab Initio” Nuclear Structure

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_i (i\gamma^\mu (D_\mu)_{ij} - m\delta_{ij}) \psi_j - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}$$

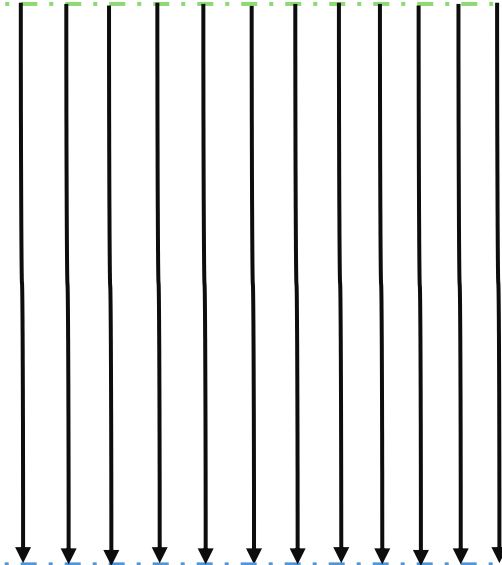
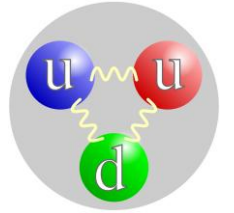


$$\hat{H}_{\text{nuc}} = \hat{T} + \hat{V}_{\text{nuc}}$$



“Ab Initio” Nuclear Structure

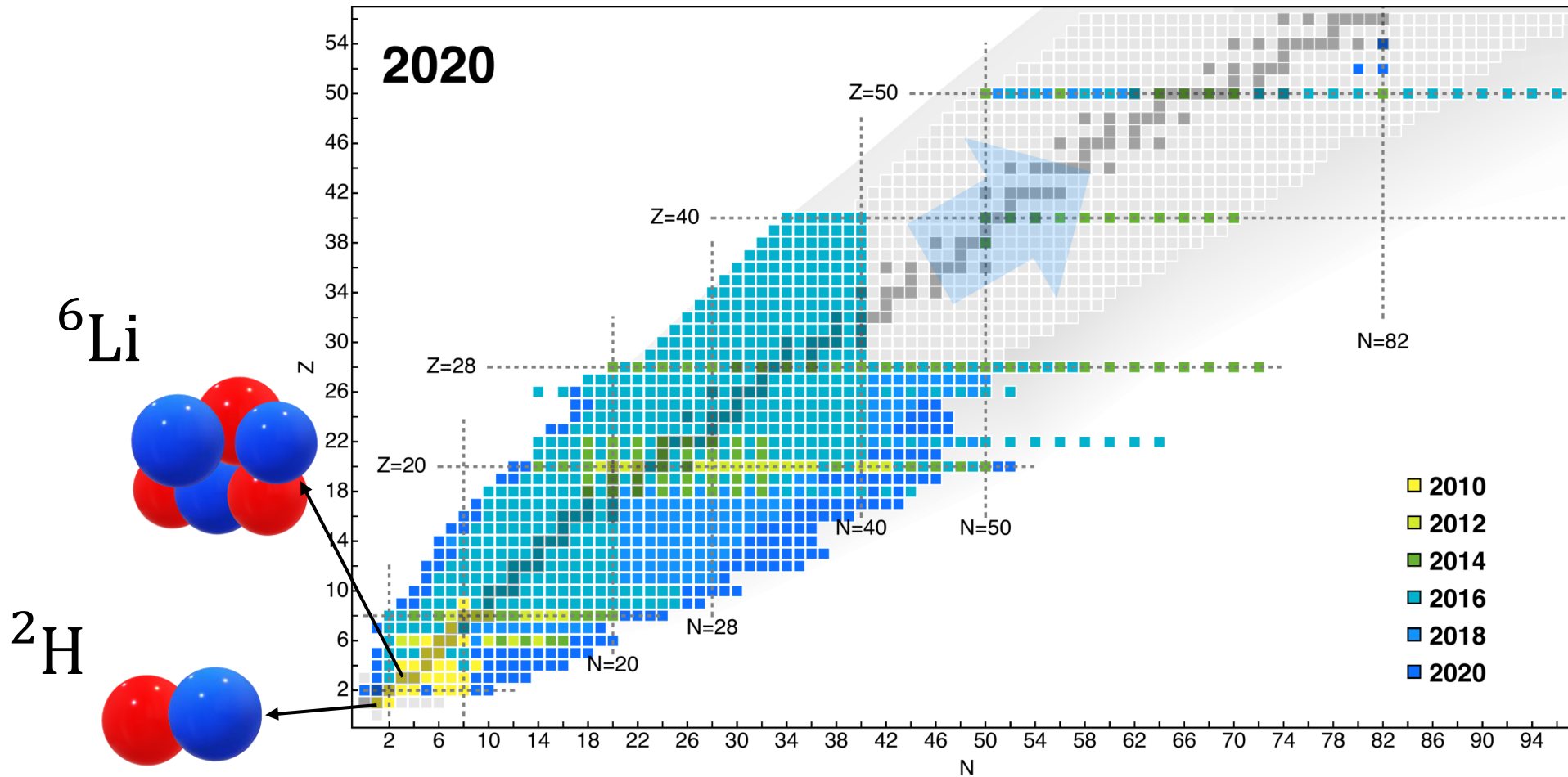
$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_i (i\gamma^\mu (D_\mu)_{ij} - m\delta_{ij}) \psi_j - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}$$



$$\hat{H}_{\text{nuc}}^{(i)} = \hat{T} + \hat{V}_{\text{nuc}}$$

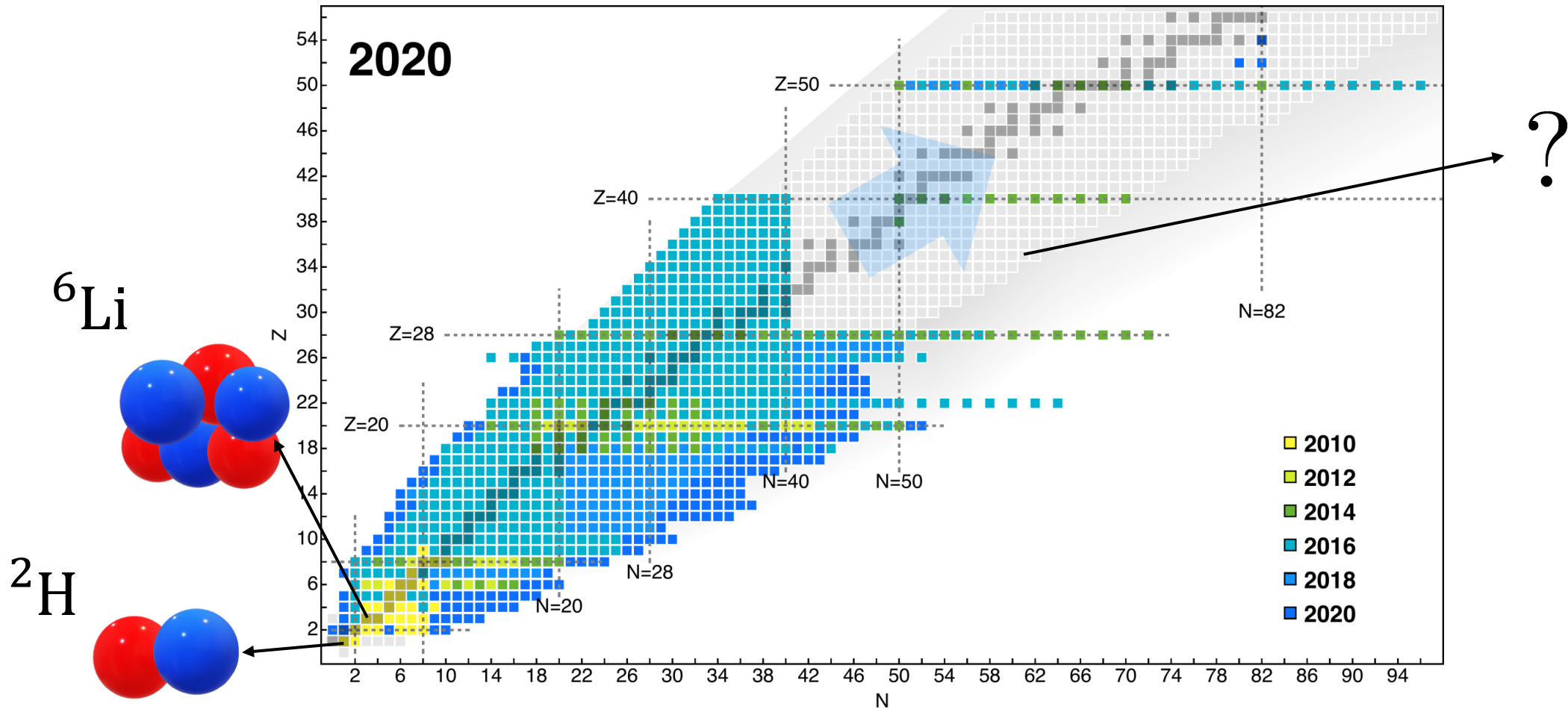


Ab Initio Nuclear Structure



H. Hergert, Front. Phys. **07** (2020)

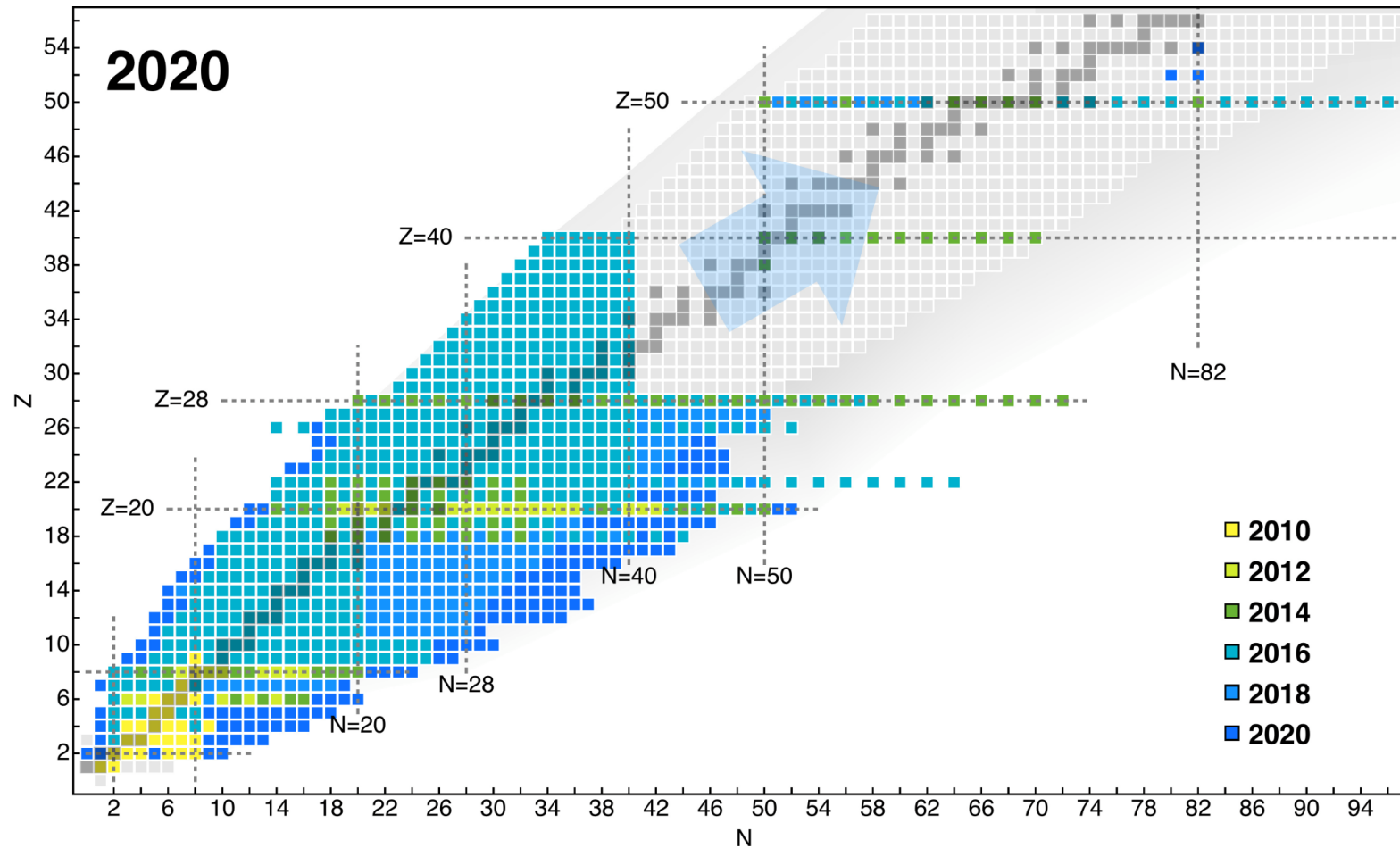
Ab Initio Nuclear Structure



H. Hergert, Front. Phys. **07** (2020)

Ab Initio Nuclear Structure

- IT-NCSM
- MR-IMSRG
- VS-IMSRG
- ADC

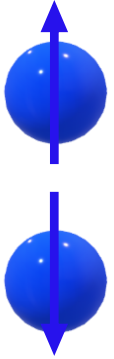


- Lattice EFT
- CCSD
- Λ -CCSD

H. Hergert, Front. Phys. **07** (2020)

Exponential Scaling Problem

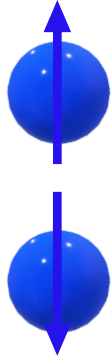
$$N=1$$



$$2^N=2$$

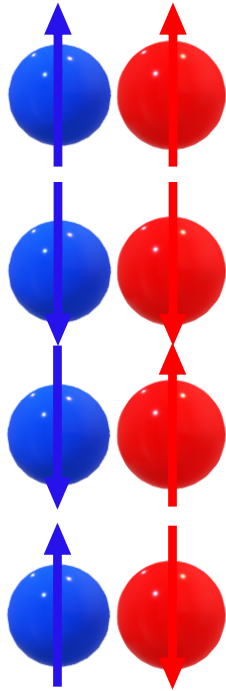
Exponential Scaling Problem

$N=1$



$$2^N=2$$

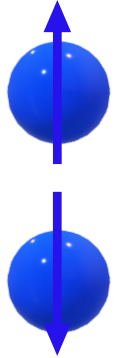
$N=2$



$$2^N=4$$

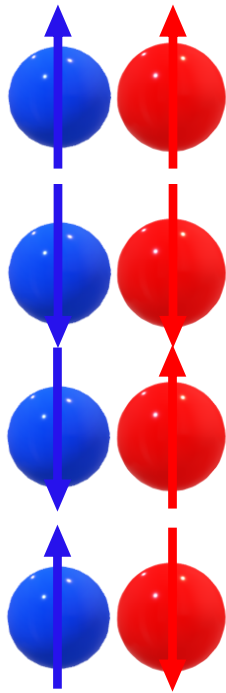
Exponential Scaling Problem

$N=1$



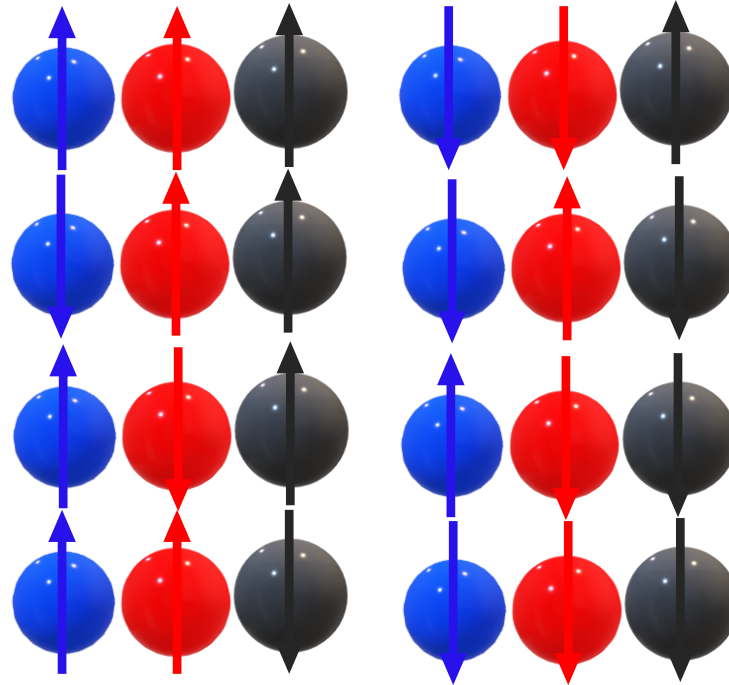
$$2^N=2$$

$N=2$



$$2^N=4$$

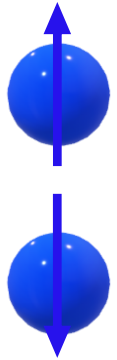
$N=3$



$$2^N=8$$

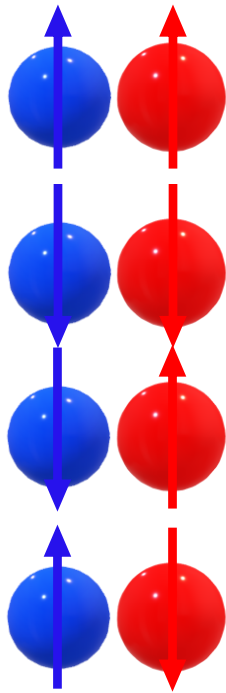
Exponential Scaling Problem

$N=1$



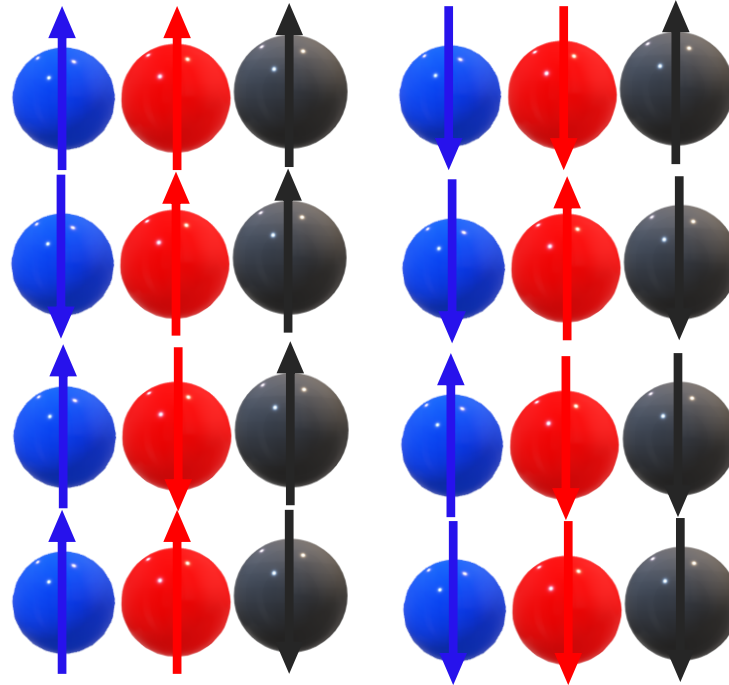
$$2^N=2$$

$N=2$



$$2^N=4$$

$N=3$



$$2^N=8$$

$N=80$ (Zr)

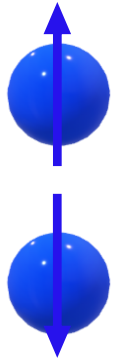


$$2^N=1,208,925,819,614,629,174,706,176$$

$$(1,2 \times 10^{24})$$

Exponential Scaling Problem

$N=1$

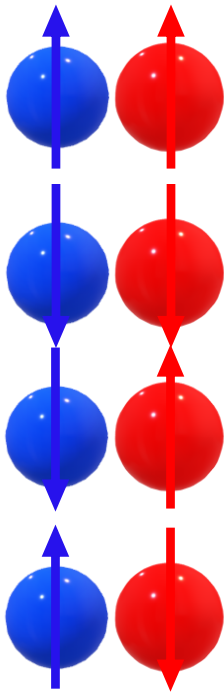


$$2^N=2$$



=16B

$N=2$

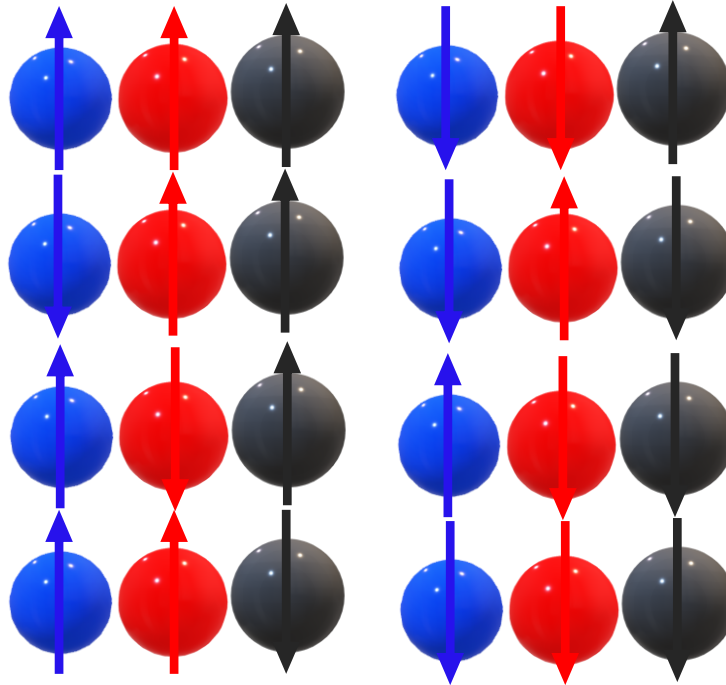


$$2^N=4$$



=32B

$N=3$



$$2^N=8$$



=64B

$N=80$ (Zr)



$$2^N=1,208,925,819,614,629,174,706,176$$



$\sim 10^{25}$ B

The Solution: AI



How can AI help us?

Neural Quantum States

- **Variational principle** $\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \geq E_{\text{GS}}$ $\longrightarrow$ Find θ that minimises $\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle}$

Neural Quantum States

- **Variational principle** $\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \geq E_{\text{GS}}$ $\longrightarrow$ Find θ that minimises $\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle}$

- **Variational Monte Carlo**

$$|\psi_\theta\rangle = \int \psi_\theta(X) dX = \int \psi_\theta(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_N$$

$$\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \mathbb{E}_{X \sim p_\theta} \left[\frac{\langle X | \hat{H} | \psi_\theta \rangle}{\langle X | \psi_\theta \rangle} \right] \simeq \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{\langle X_i | \hat{H} | \psi_\theta \rangle}{\langle X_i | \psi_\theta \rangle} \quad \text{“Local energy”}$$

Neural Quantum States

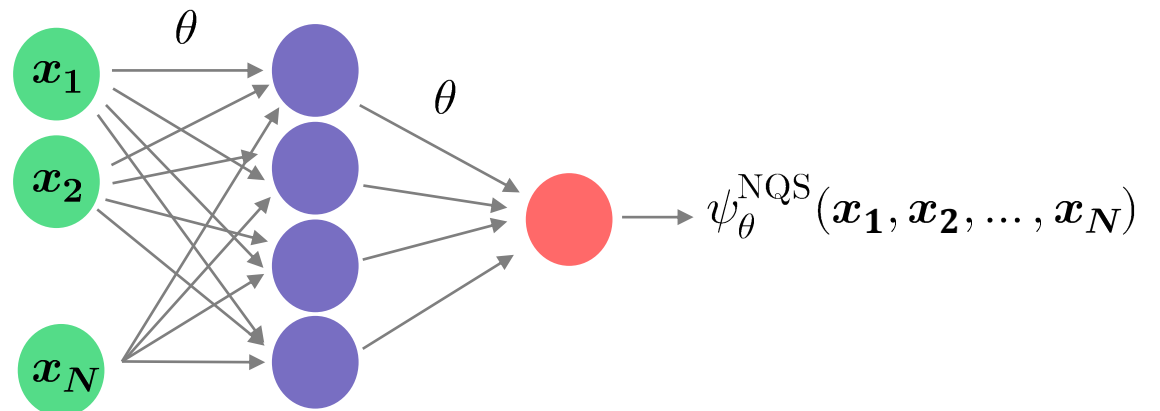
- **Variational principle** $\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \geq E_{\text{GS}} \longrightarrow$ Find θ that minimises $\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle}$

- **Variational Monte Carlo**

$$|\psi_\theta\rangle = \int \psi_\theta(X) dX = \int \psi_\theta(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_N$$

$$\frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \mathbb{E}_{X \sim p_\theta} \left[\frac{\langle X | \hat{H} | \psi_\theta \rangle}{\langle X | \psi_\theta \rangle} \right] \simeq \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{\langle X_i | \hat{H} | \psi_\theta \rangle}{\langle X_i | \psi_\theta \rangle} \quad \text{“Local energy”}$$

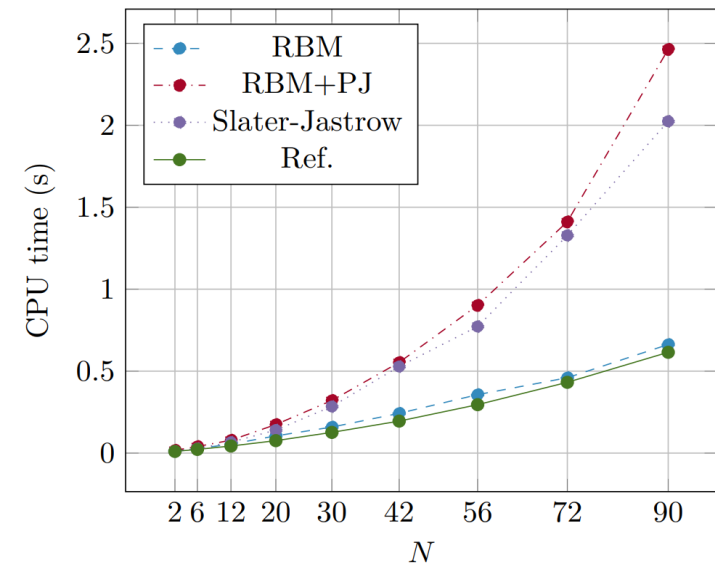
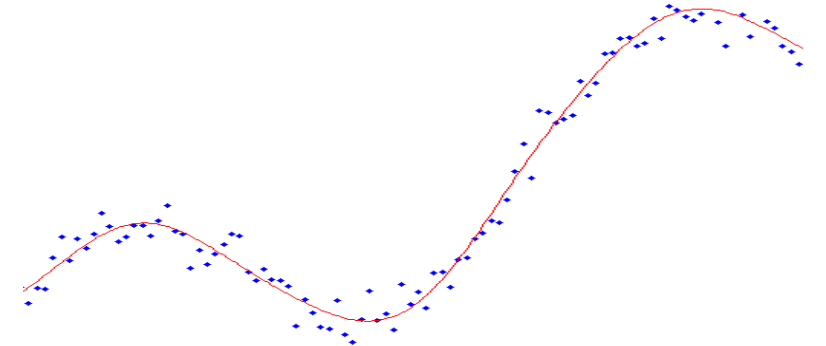
- **NQS Ansatz:** $\psi_{\text{NQS}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; \boldsymbol{\theta})$



Why Neural Networks?

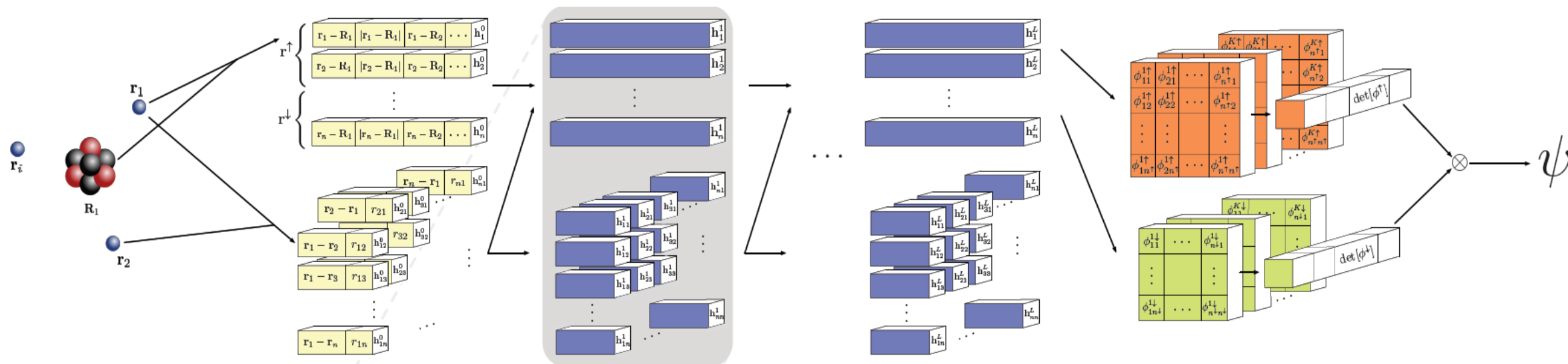
- **NNs have “ ∞ power”**: a neural network can approximate any continuous function.
- **Space complexity**: polynomial scaling of memory resources... possibly!

$$\text{Memory}(A) \sim \mathcal{O}(A^n), \quad A = Z + N$$

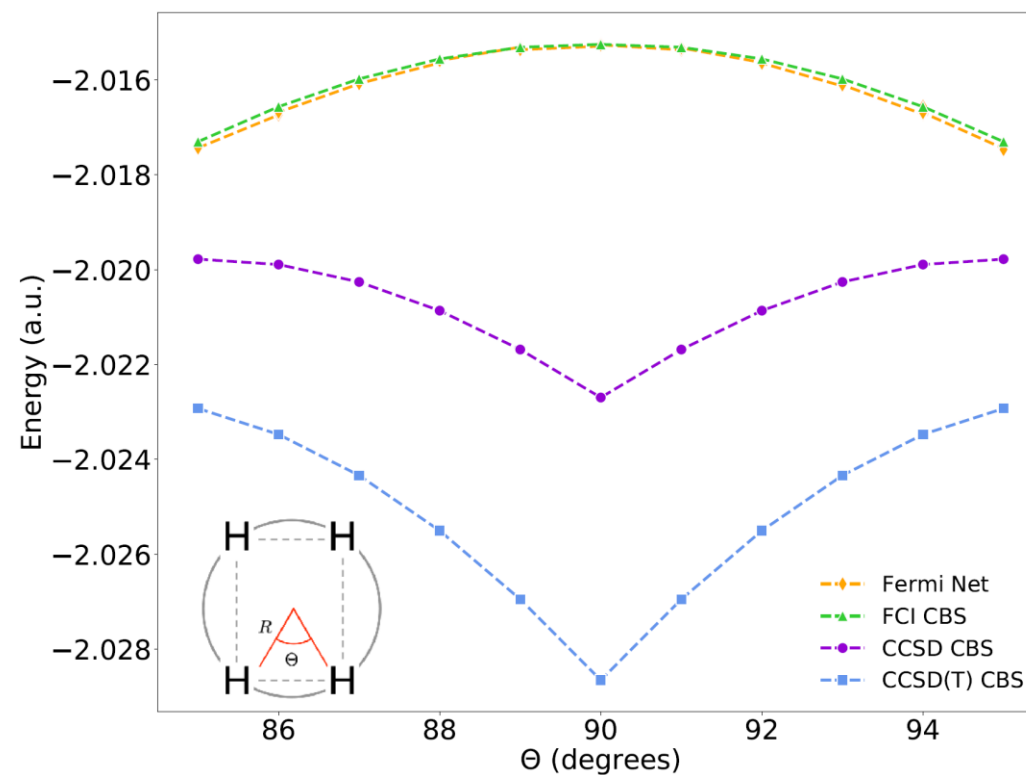
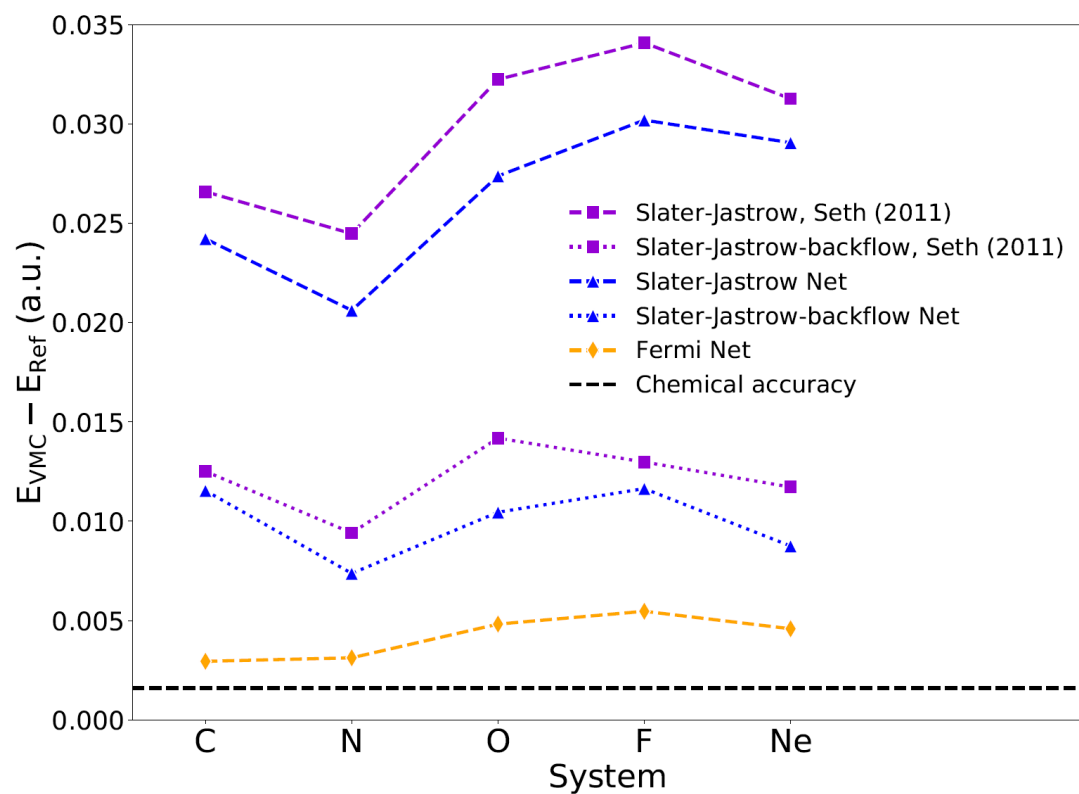


NQS showcase

FermiNet (Quantum Chemistry)

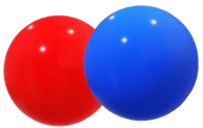
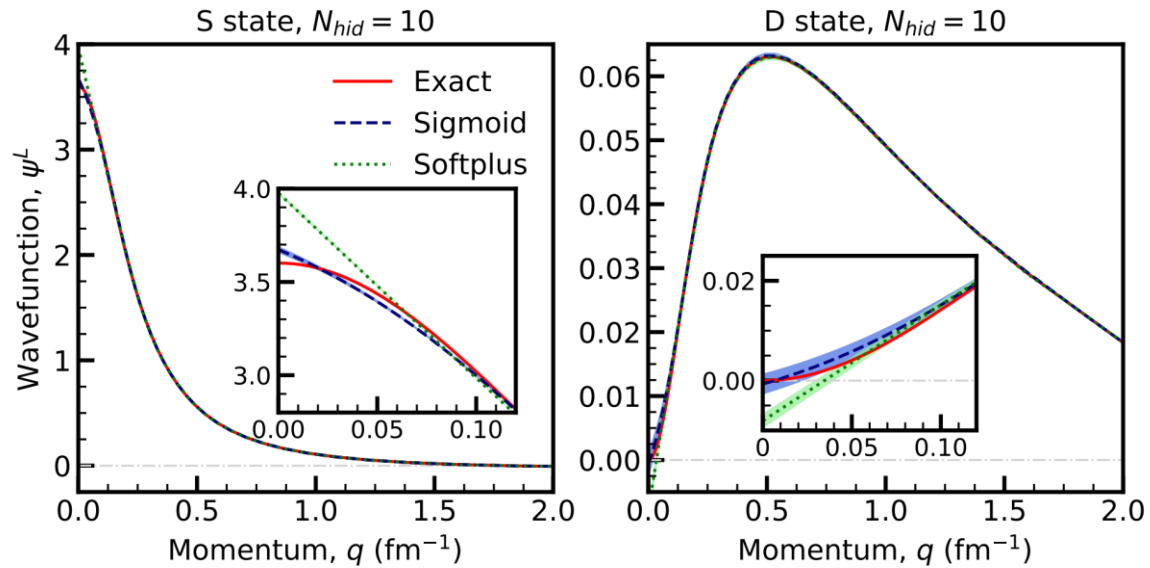


FermiNet (Quantum Chemistry)



NQS for Nuclei

2020

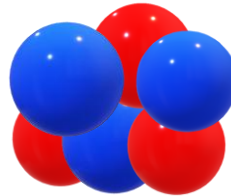
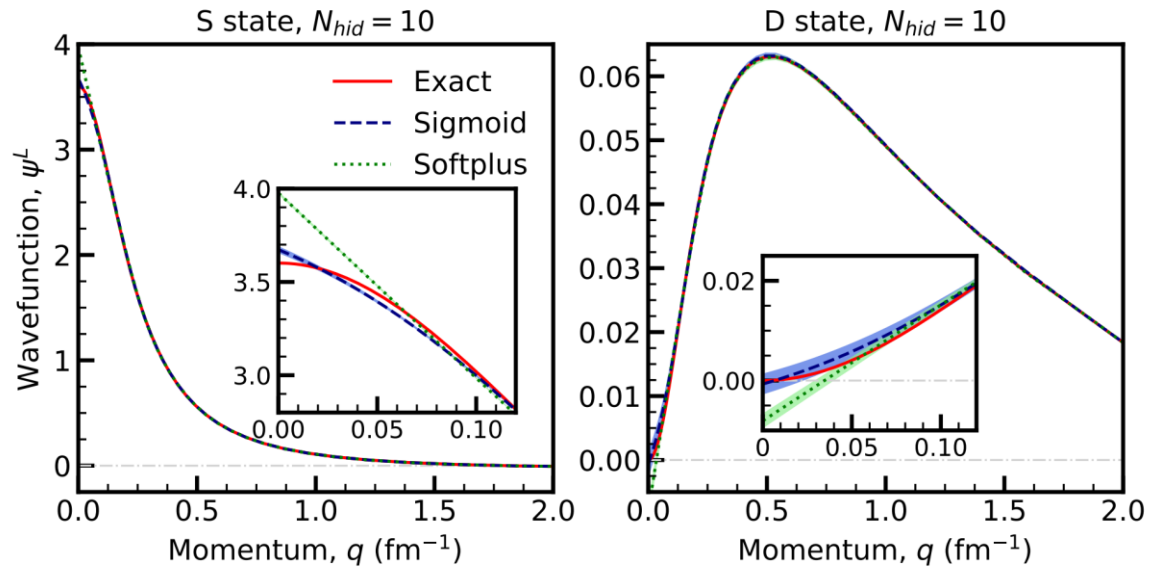


NQS for Nuclei

2020



...



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

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

NQS for Nuclei

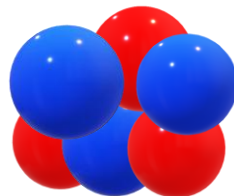
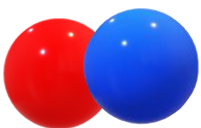
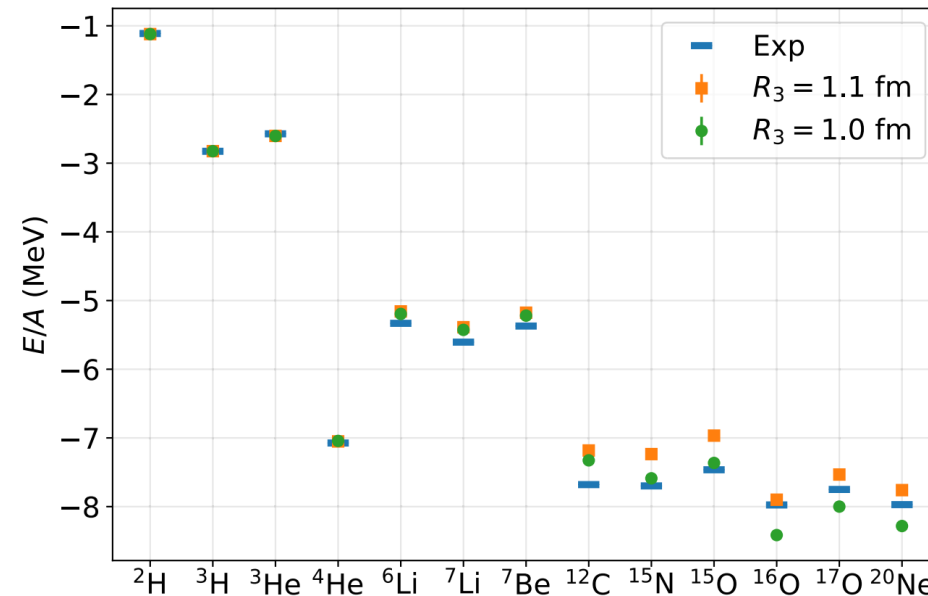
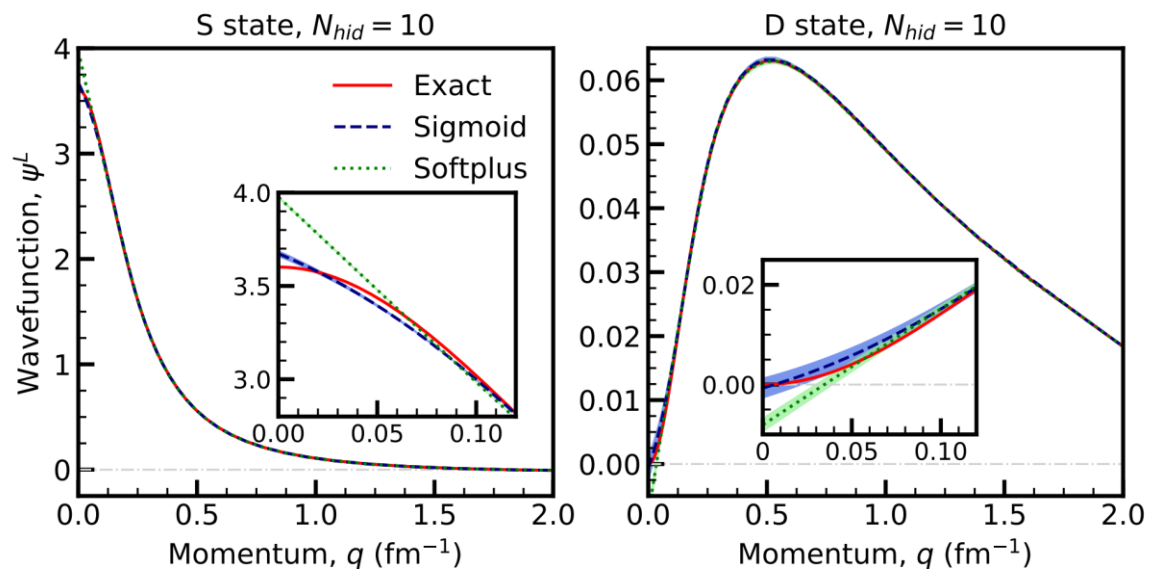
2020



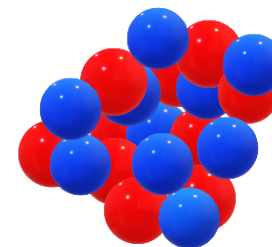
...



2024



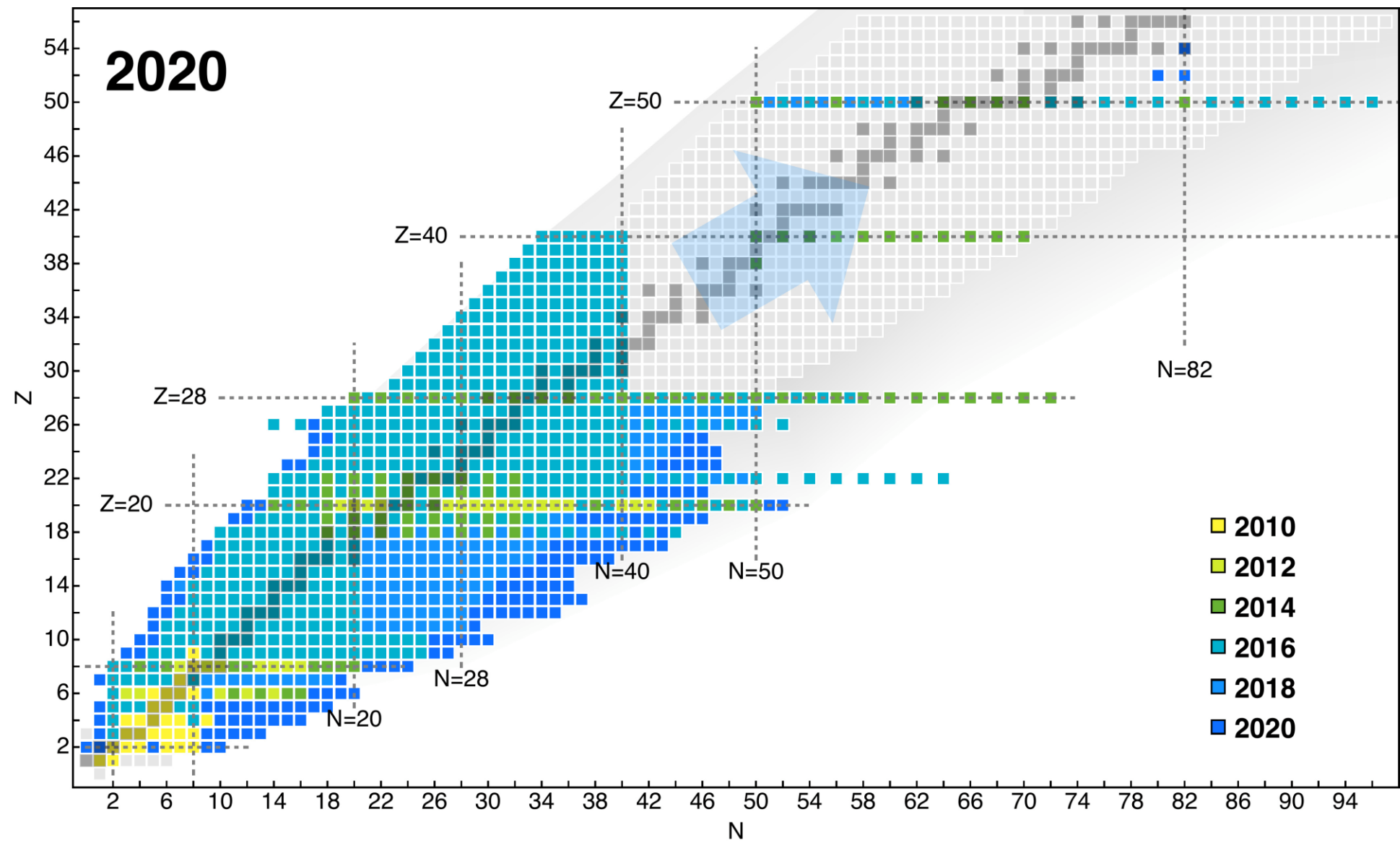
...

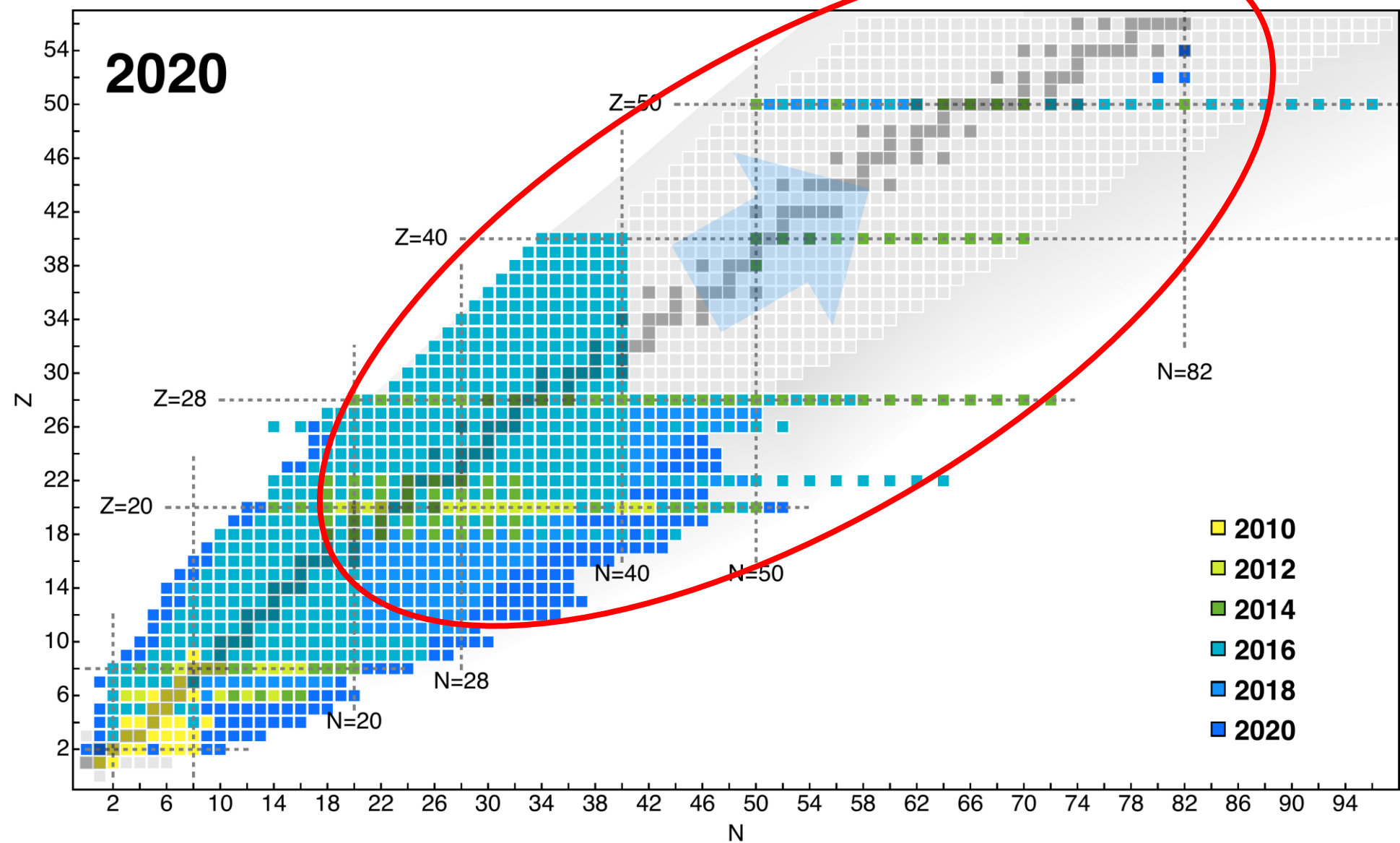


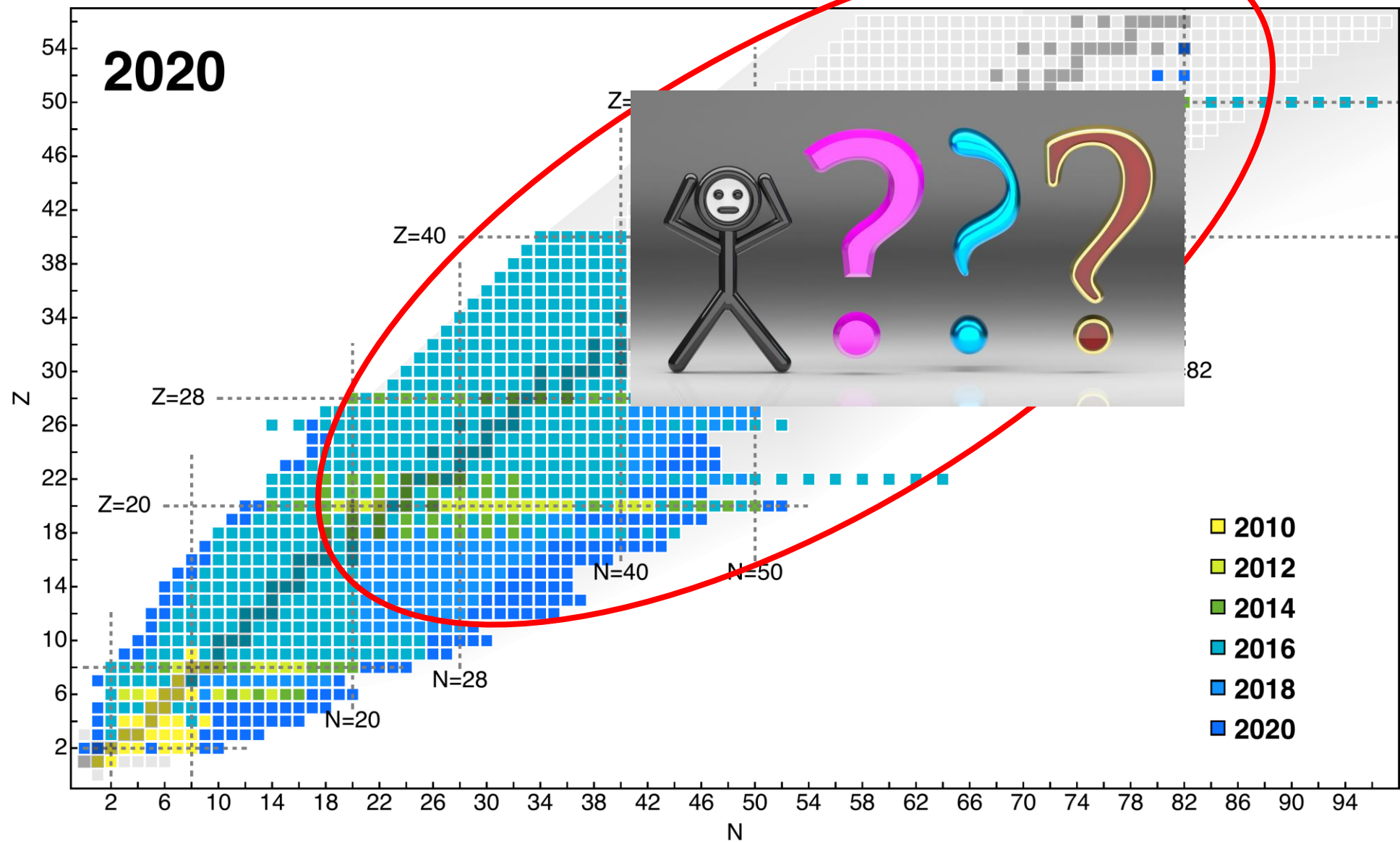
J. Keeble & A. Rios, Phys. Lett. B **809**
(2020)

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

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

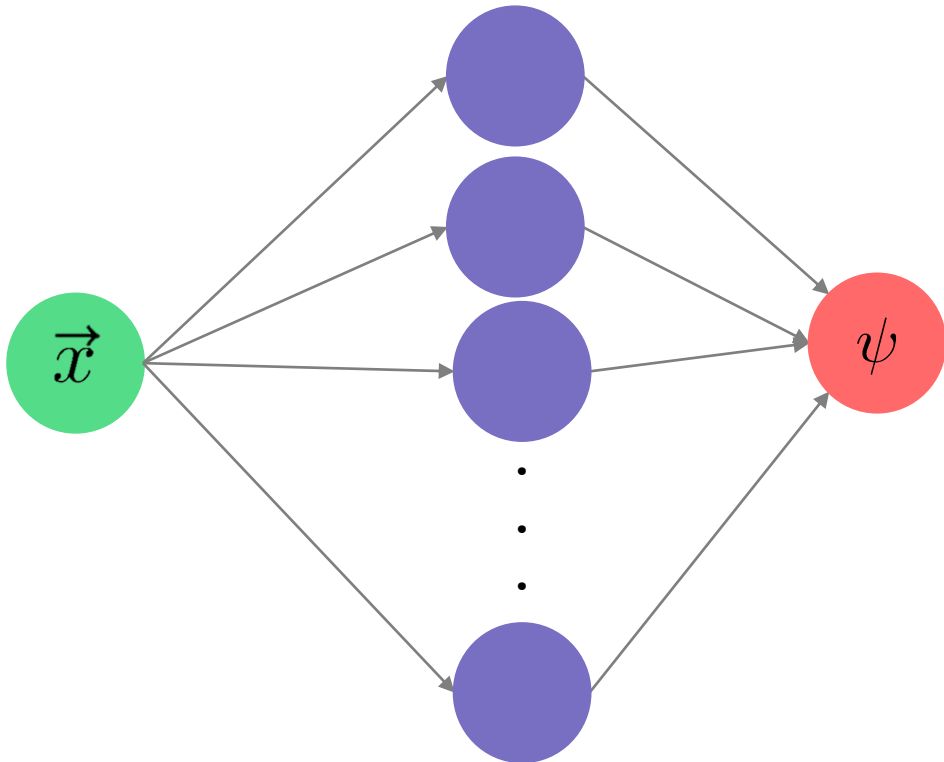






1. Neural Network Architecture

Physical properties in NNs



- **Particle exchange symmetry**

$$\psi(x_1, x_2, \dots, x_N) = \pm \psi(x_2, x_1, \dots, x_N)$$

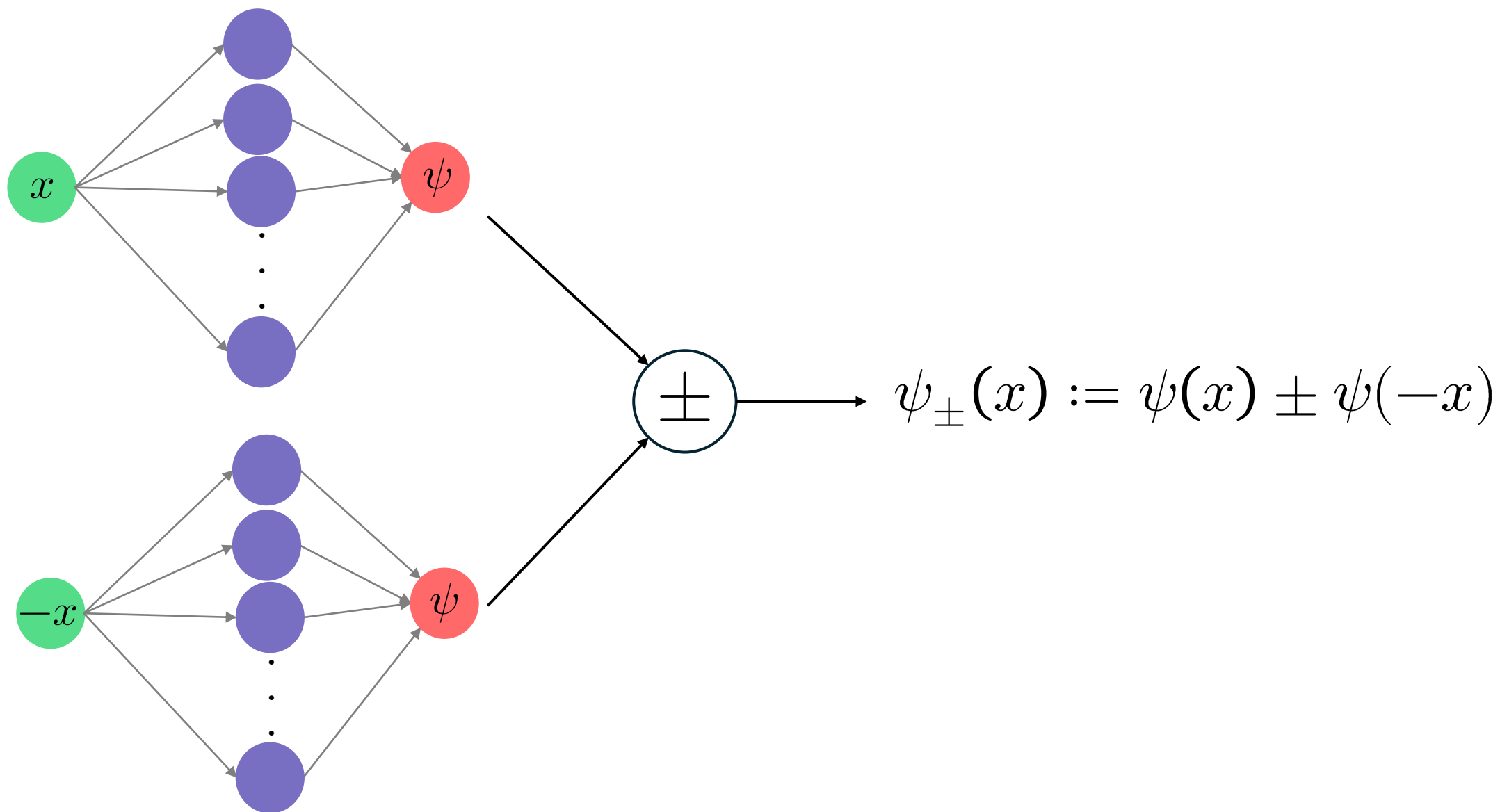
- **Spherical symmetry**

$$\psi(\vec{x}) = \psi(R\vec{x})$$

- **Time-reversal symmetry**

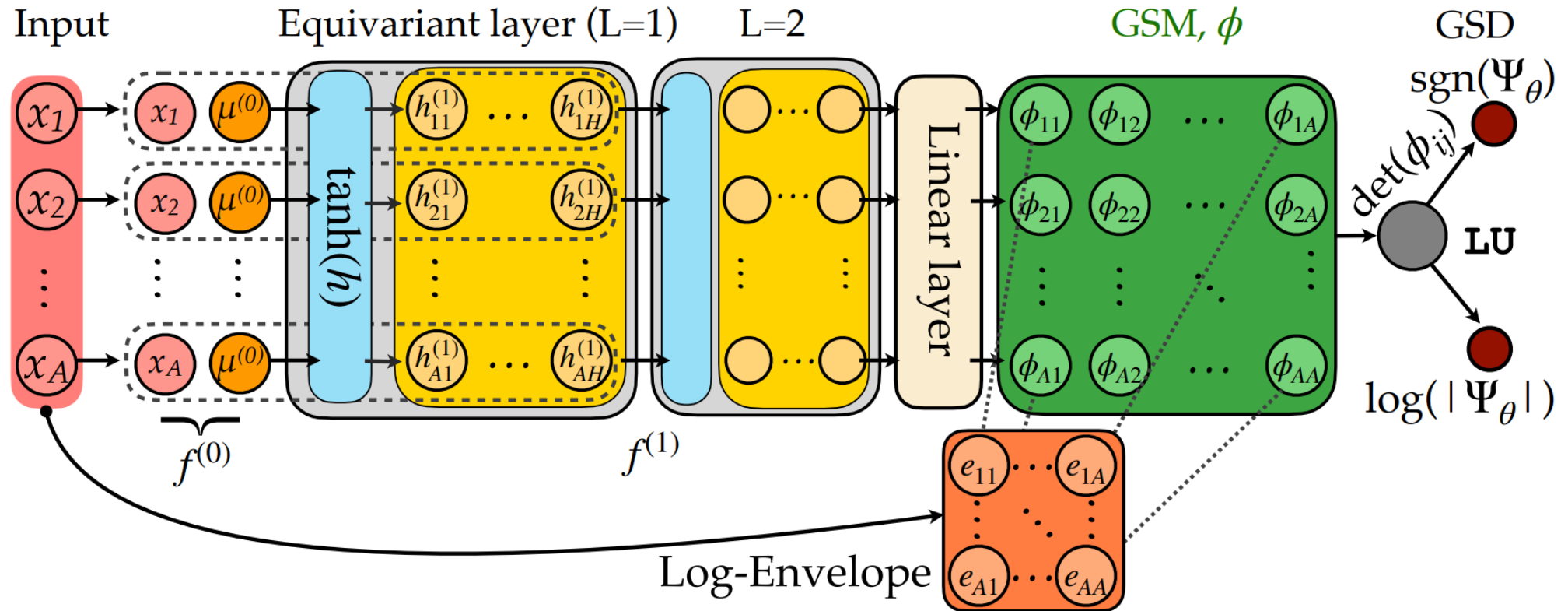
$$\psi(\vec{x}, \sigma) = \psi(T(\vec{x}, \sigma))$$

Reflection Symmetry



Particle Exchange Symmetry

$$\psi_{\text{NQS}}(x_1, x_2, \dots, x_N) = \varphi_{\text{EQUIV}} \circ \det \phi_{\text{GSM}}(x_1, x_2, \dots, x_N)$$

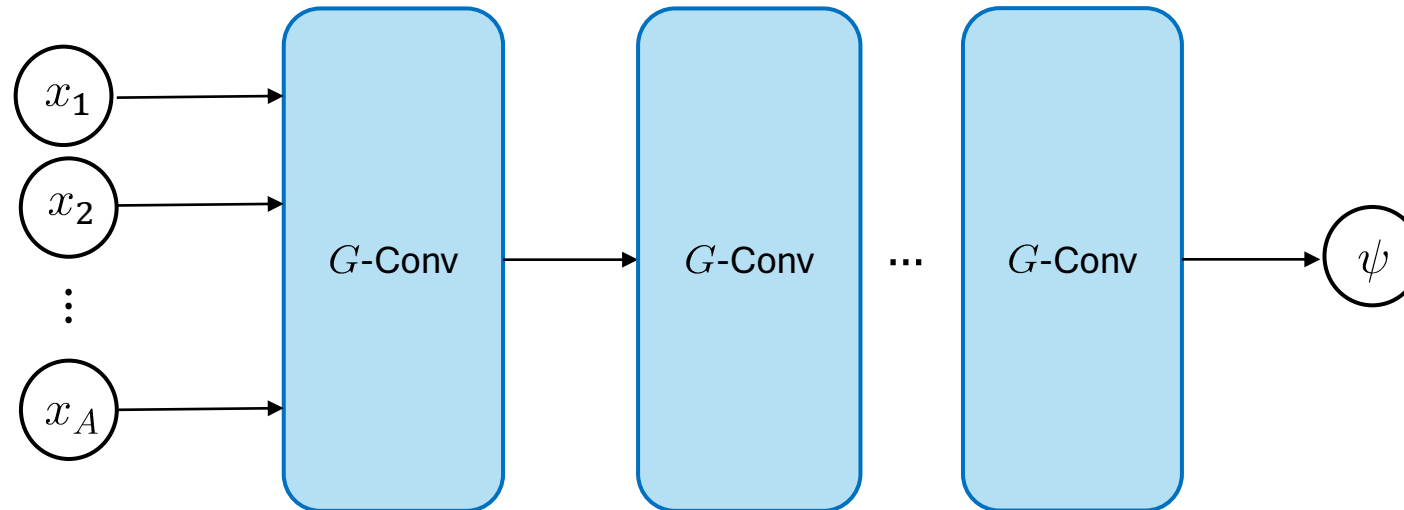


Towards a universal recipe...

- **Continuous groups:** what about Lie groups? $SU(2)$ $SO(3)$
- **Mixed groups:** what about product groups? $SU(2) \times S_N$

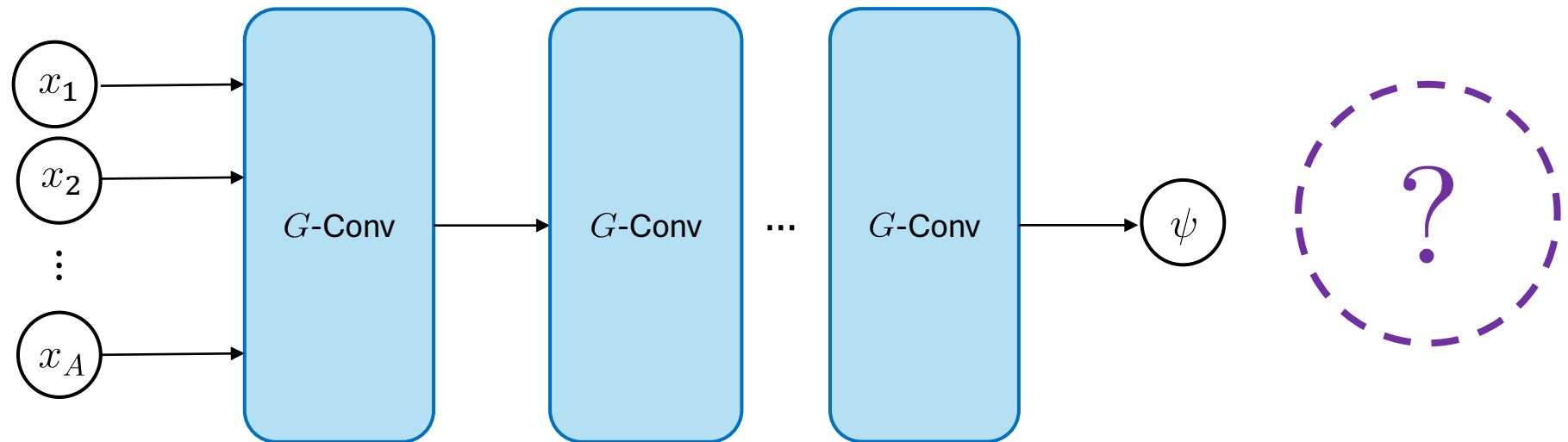
Towards a universal recipe...

- **Continuous groups:** what about Lie groups? $SU(2)$ $SO(3)$
- **Mixed groups:** what about product groups? $SU(2) \times S_N$



Towards a universal recipe...

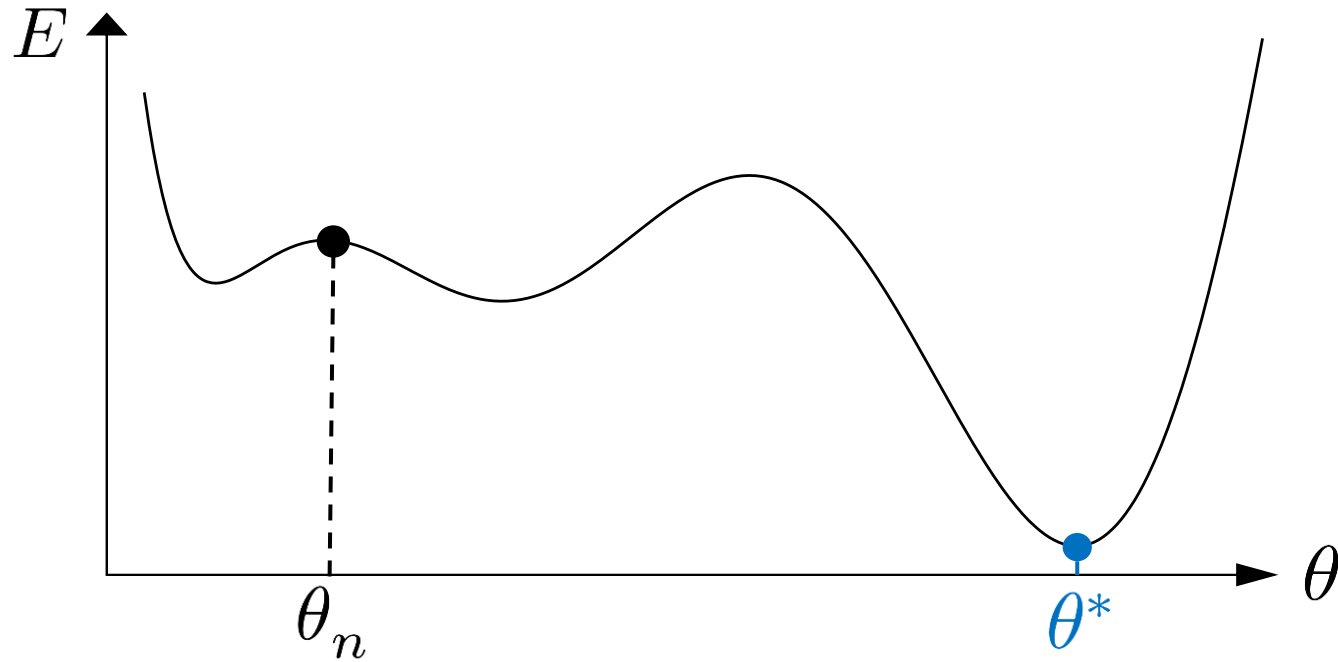
- **Continuous groups:** what about Lie groups? $SU(2)$ $SO(3)$
- **Mixed groups:** what about product groups? $SU(2) \times S_N$



2. Optimisation Strategy

Gradient Descent

- **Local Linear Model:** $M_n(\delta) = \textcolor{red}{L}^T \delta + C$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$

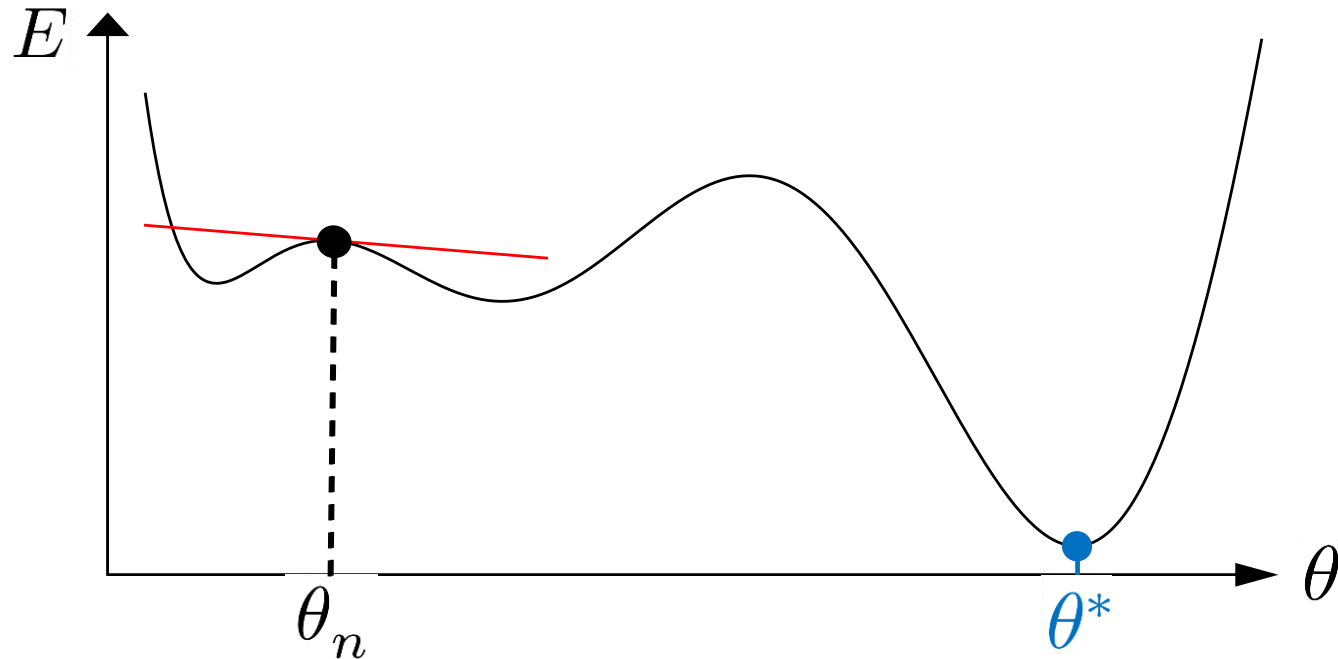


Gradient Descent

- **Local Linear Model:** $M_n(\delta) = \textcolor{red}{L}^T \delta + C$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$

$$\delta_n = -\alpha \nabla E(\theta_n)$$

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



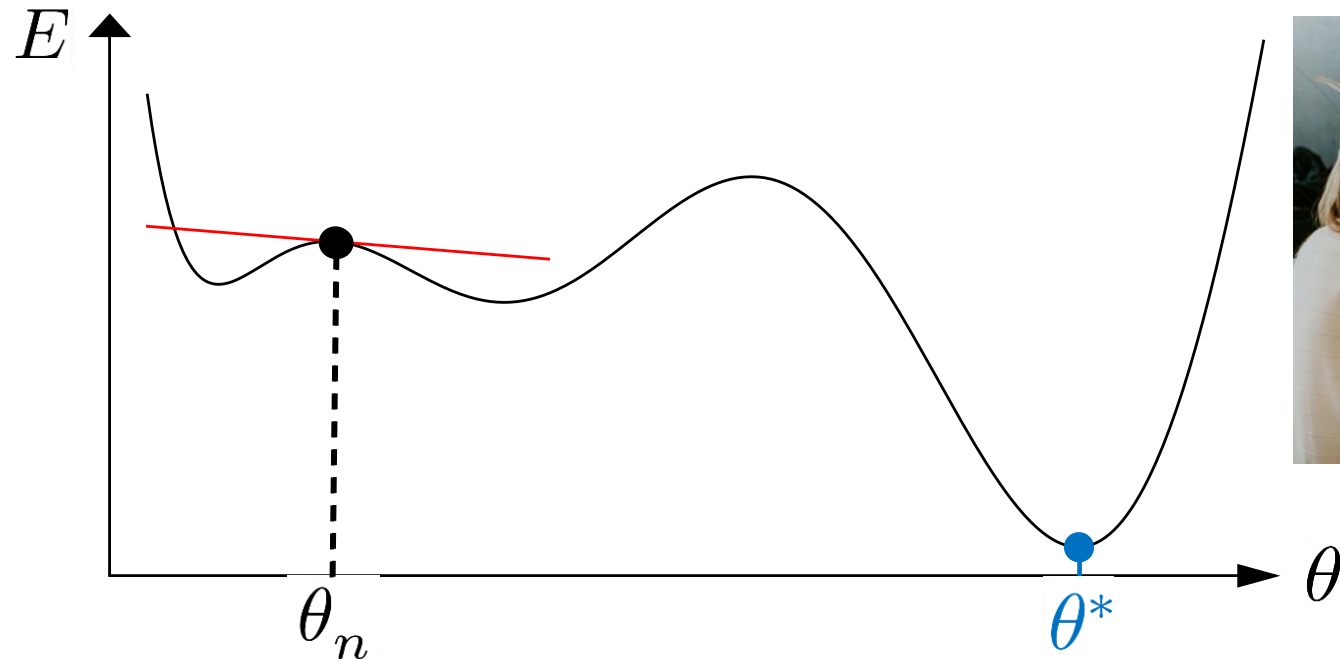
Gradient Descent



- **Local Linear Model:** $M_n(\delta) = \textcolor{red}{L}^T \delta + C$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$

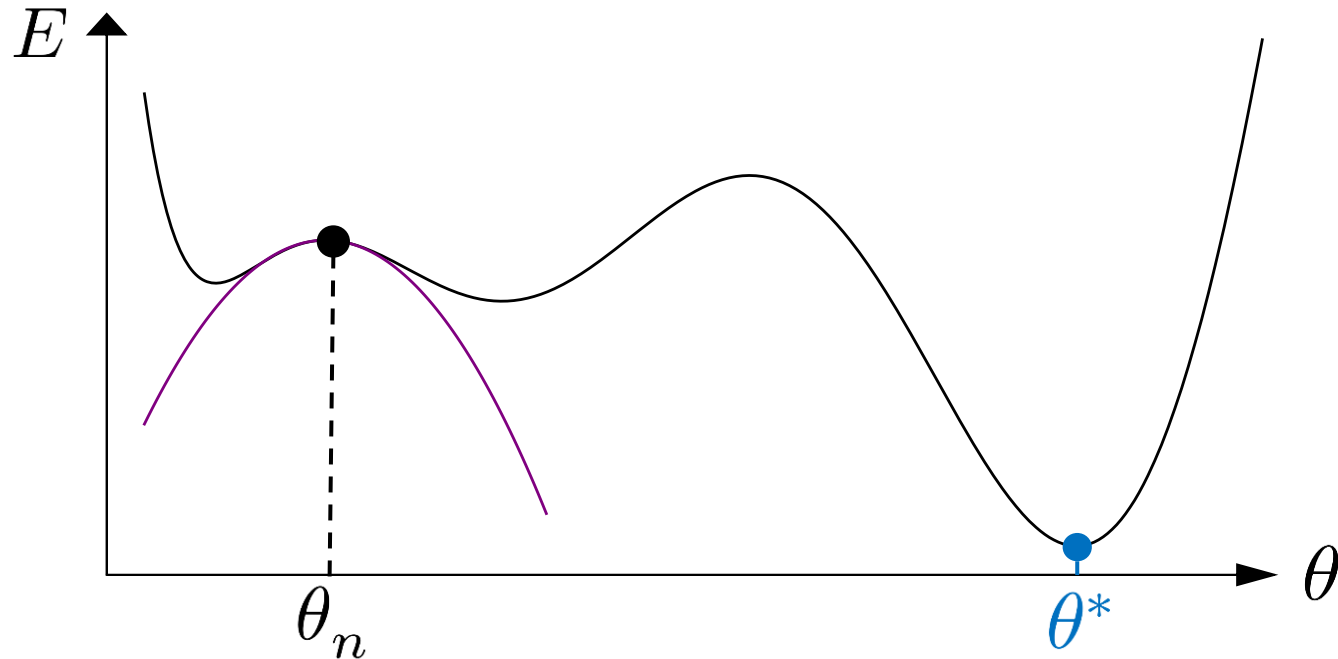
$$\delta_n = -\alpha \nabla E(\theta_n)$$

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



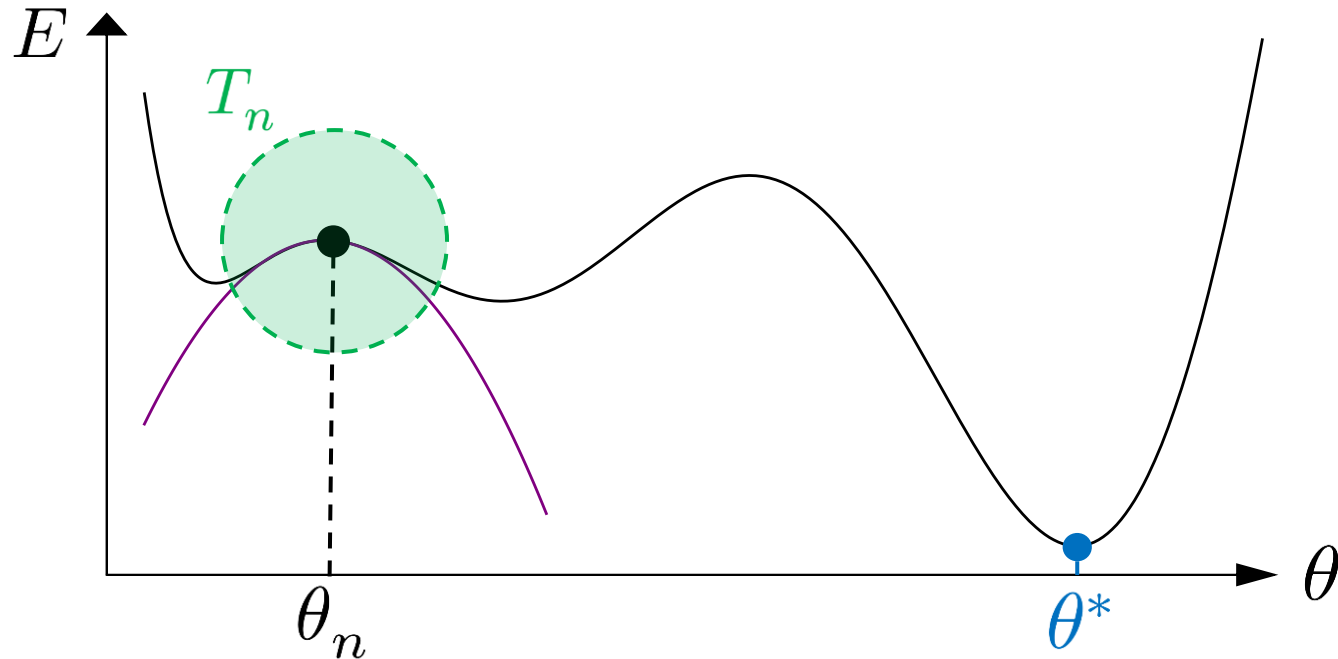
Second-order optimisation

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + \textcolor{red}{L}^T \delta + C$
- **Trust region:** $\textcolor{teal}{T}_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$



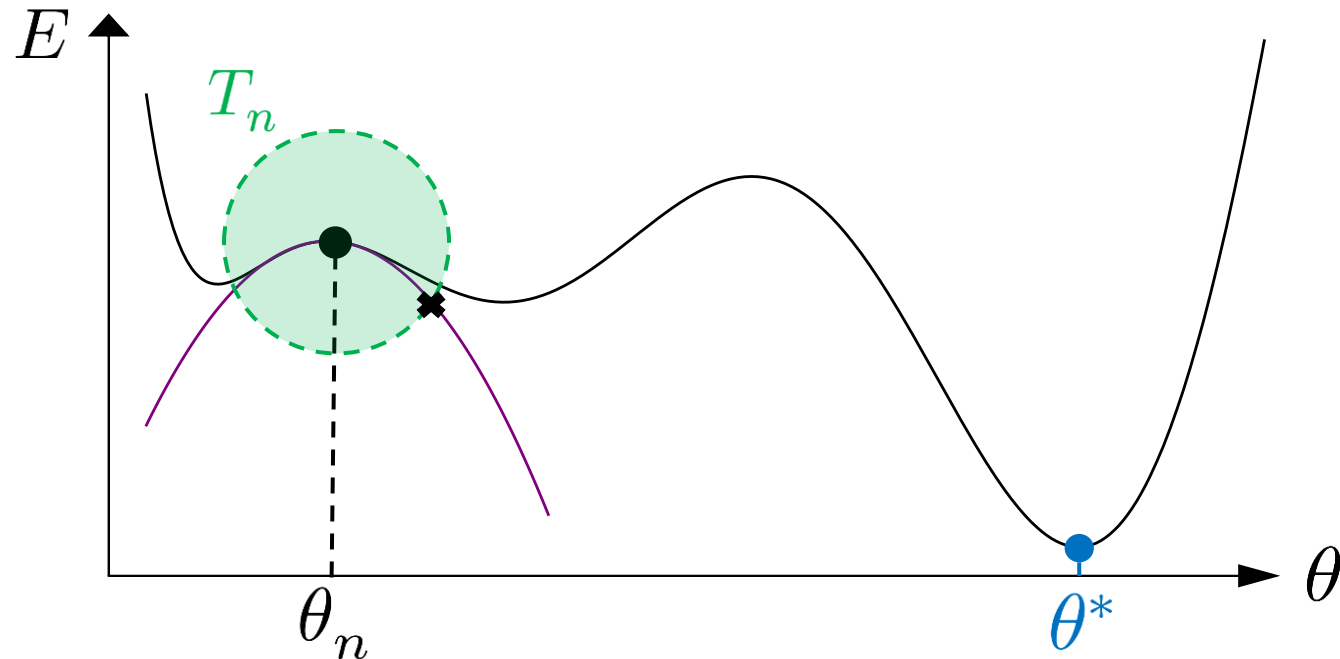
Second-order optimisation

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q\delta + \textcolor{red}{L}^T \delta + C$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$



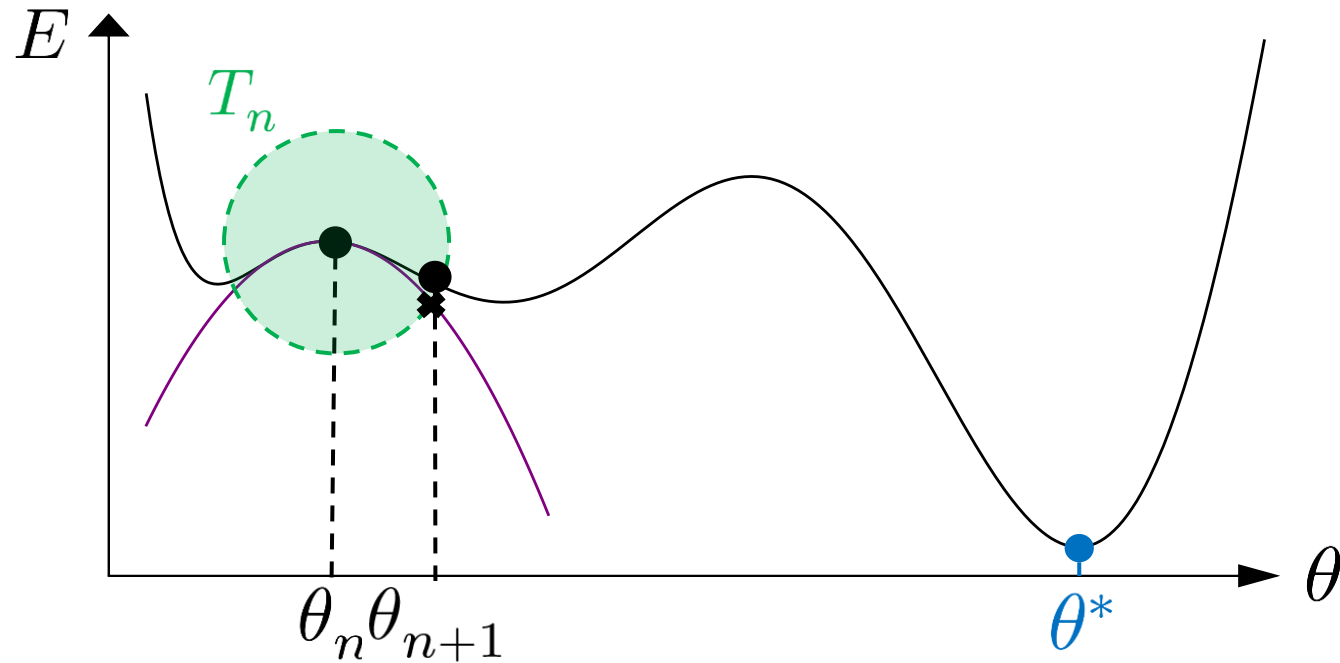
Second-order optimisation

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q\delta + \textcolor{red}{L}^T \delta + C$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$



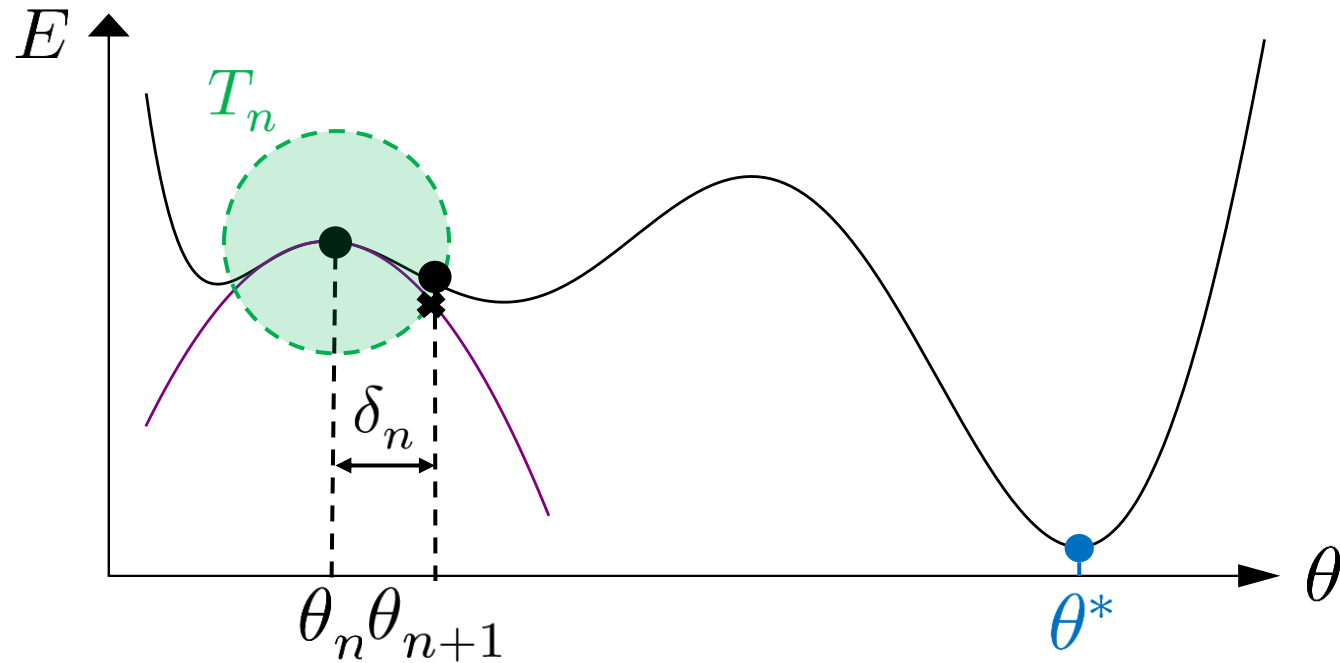
Second-order optimisation

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q\delta + \textcolor{red}{L}^T \delta + C$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$



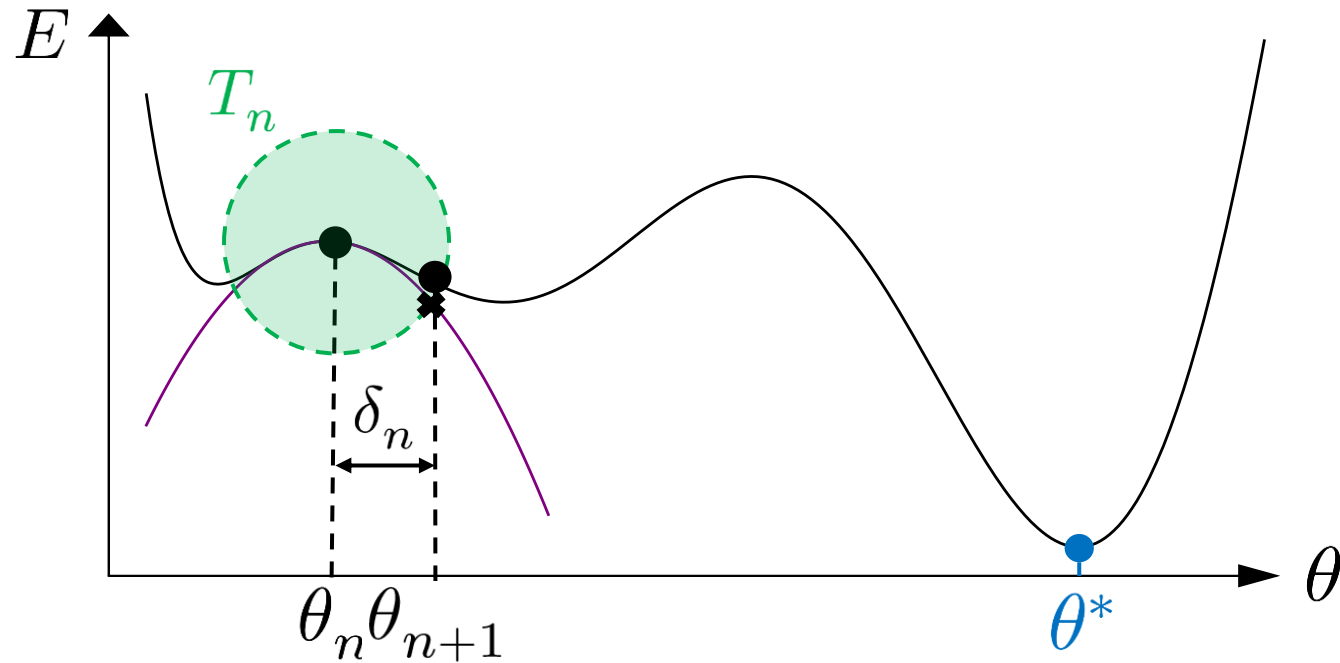
Second-order optimisation

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + \textcolor{red}{L}^T \delta + C$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \operatorname{argmin}_{\delta \in T_n} M_n(\delta)$



Second-order optimisation

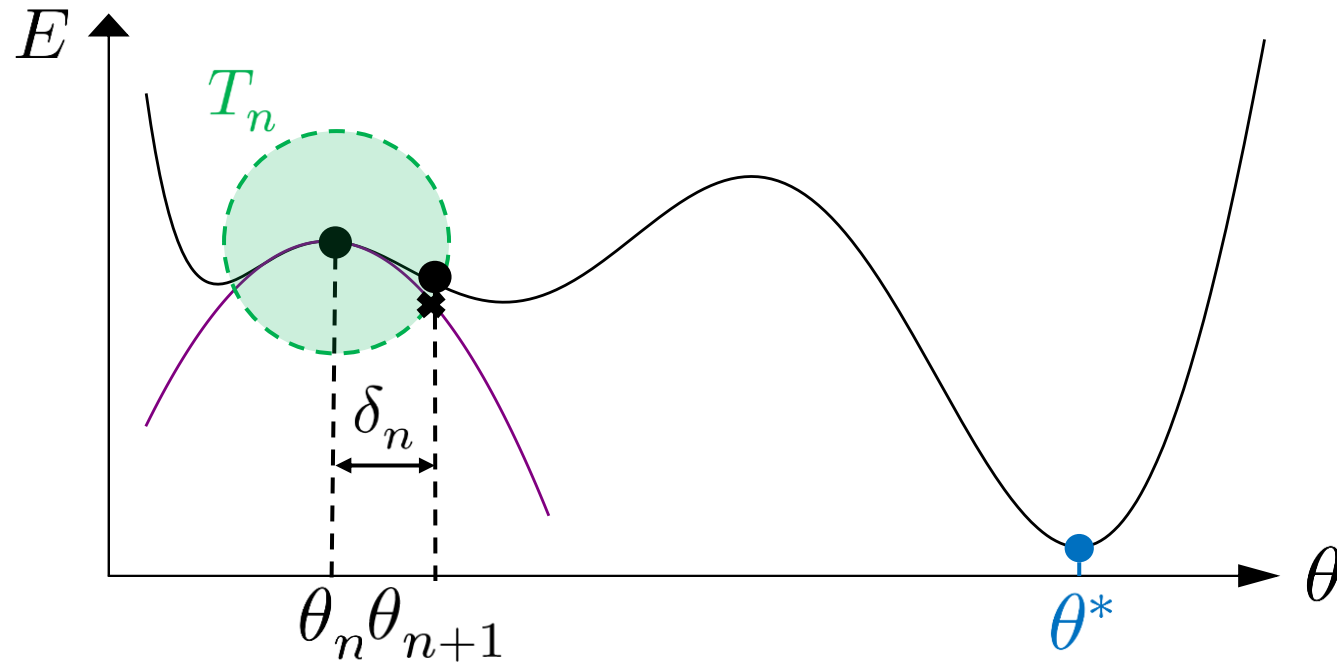
- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + \textcolor{red}{L}^T \delta + C$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n = - (Q_n + \lambda_n R_n)^{-1} \nabla E(\theta_n)$



Second-order optimisation

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n = -(Q_n + \lambda_n R_n)^{-1} \nabla E(\theta_n)$

$$L = \nabla_{\theta} E$$

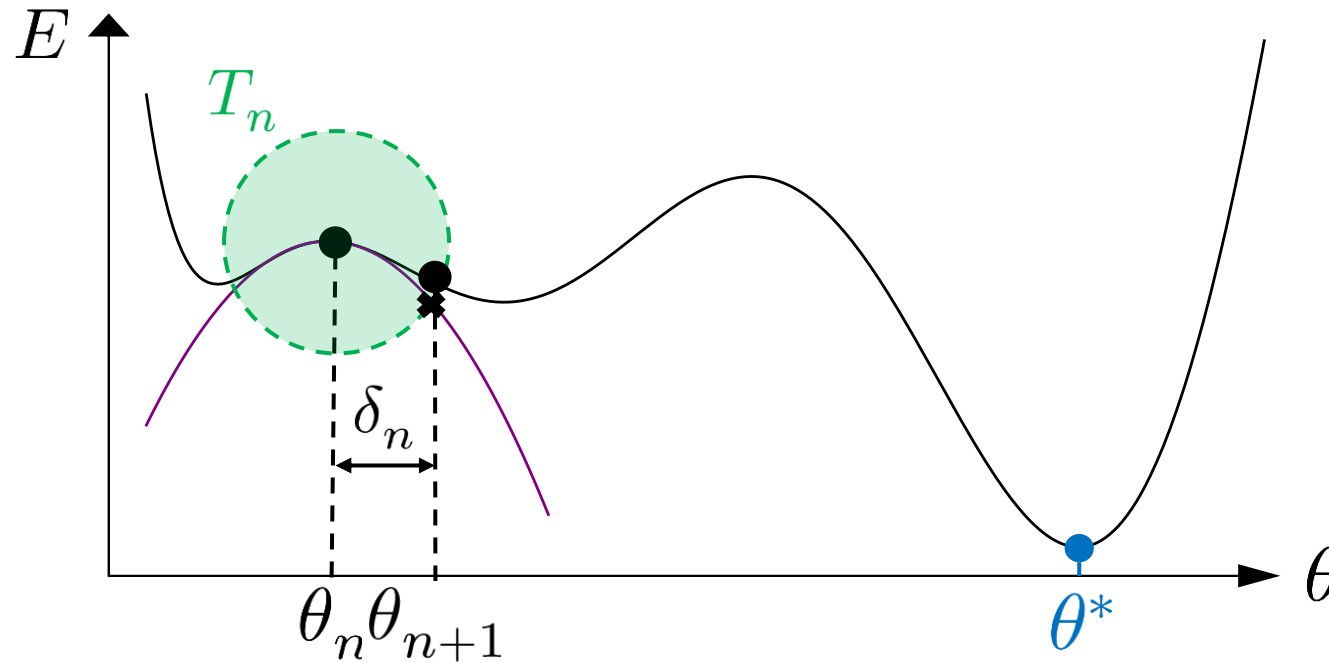


Second-order optimisation

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n = -(Q_n + \lambda_n R_n)^{-1} \nabla E(\theta_n)$

$$L = \nabla_{\theta} E$$

$$Q?$$



Stochastic Reconfiguration

“Quantum Geometric Tensor”

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$, $Q_{ij}(\theta) = \left\langle \frac{\partial \psi_\theta}{\partial \theta_i}, \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle - \left\langle \frac{\partial \psi_\theta}{\partial \theta_i}, \psi_\theta \right\rangle \left\langle \psi_\theta, \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle$
- **Trust region:** $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$, $R_n = 1$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n = -(Q_n + \lambda_n R_n)^{-1} \nabla E(\theta_n)$

Stochastic Reconfiguration

“Quantum Geometric Tensor”

- Local Quadratic Model: $M_n(\delta) = \frac{1}{2} \delta^T Q \delta + L^T \delta + C$, $Q_{ij}(\theta) = \left\langle \frac{\partial \psi_\theta}{\partial \theta_i}, \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle - \left\langle \frac{\partial \psi_\theta}{\partial \theta_i}, \psi_\theta \right\rangle \left\langle \psi_\theta, \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle$
- Trust region: $T_n = \{\delta: \delta^T R_n \delta \leq r^2\}$, $R_n = 1$
- Update: $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n = -(Q_n + \lambda_n R_n)^{-1} L_n$

Quantum Natural Gradient

James Stokes¹, Josh Izaac², Nathan Killoran², and Giuseppe Carleo³

¹Center for Computational Quantum Physics and Center for Computational Mathematics, Flatiron Institute, 1900 University Avenue, Boulder, CO 80506, USA

²Xanadu, 777 Bay Street, Toronto, Canada

³Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010 USA

A quantum generalization of Natural Gradient Descent is presented as part of a general-purpose optimization framework for variational quantum circuits. The optimization dynamics is interpreted as moving in the steepest descent direction with respect to the Quantum Information Geometry, corresponding to the real part of the Quantum Geometric Tensor (QGT), also known as the Fubini-Study metric tensor. An efficient algorithm is presented for computing a block-diagonal approximation to the Fubini-Study metric tensor for parametrized quantum circuits, which may be of independent interest.

1 Introduction

Variational optimization of parametrized quantum circuits is an integral component for many hybrid quantum-classical algorithms, which are arguably the most promising applications of quantum computing today. The success of these algorithms hinges on the quality of the initial state and the choice of the parameter space.

parameters, including derivative-free methods such as Nelder-Mead [12] or SPSSA [33]. Recent work exploiting direct access to the gradient has been demonstrated in deep learning, where circuits have been designed to compute gradients with minimal overhead in terms of function evaluations.

One motivation for this work is theoretical: in the limit of large system size, the expected error in the objective function of the best known zeroth-order optimization algorithms scales polynomially with the dimension d of the parameter space, while the error of the Quantum Natural Gradient Descent (QNGD) scales as $1/d$. Another motivation is practical: the success of stochastic gradient descent in deep neural networks, where the parameter space is non-convex and the loss landscape is highly non-convex, suggests that the application of QNGD to quantum circuits may be of independent interest.

The application of QNGD to quantum circuits may be of independent interest.



15715v2 [cond-mat.str-el] 2 Aug 2024

A simple Large-Scale

Riccardo Rende,^{1,*} Luciano Loris V

¹International School for Advanced Studies
²Dipartimento di Fisica,

Neural-network architecture functions. These networks require optimization using traditional methods, which are effective with a limited number of parameters. Here, we leverage the power of deep learning to optimize the quantum Fisher information matrix in the J_1 - J_2 Heisenberg benchmark in highly-frustrated spin systems, demonstrating the scalability and efficiency of the proposed method to investigate unknown

I. INTRODUCTION

Deep learning has become crucial in fields, with neural networks being the key to many breakthroughs. Well-known examples include convolutional Neural Networks for image recognition and Deep Transformers for language-related tasks. The success of deep networks comes from their ability to learn complex patterns in data, often in the billions, which allow for training these networks on large datasets. However, to successfully train these large models, one needs to navigate the complex non-convex landscape associated with the optimization of the parameters.

The most used methods rely on stochastic gradient descent (SGD), where the gradient of the loss function is estimated from a randomly selected subset of the data. Over the years, variations of SGD, such as Adam [6] or AdamW [7], have

Geometric
Chaos
Institute for

Combining insights from the natural method with neural network ground state estimation of quantum states, in practice, little is known about the hidden details of the algorithm of the quantum Fisher information matrix initial dynamics, but the contrast to the spectral properties at convergence do not reveal a new measure of coherence that, generically, the value, suggesting that correlated stable representations of the

I. INTRODUCTION

Recently the fields of machine learning and information science have seen a lot of progress. On the one hand, a number of promising results have been obtained suggesting the potential of quantum or classical machine learning to

information science have seen a lot of progress. On the one hand, a number of promising results have been obtained suggesting the potential of quantum or classical machine learning to

v2 [cond-mat.dis-nn] 21 Feb 2023

Efficient optimization of deep neural quantum states toward machine precision

Ao Chen and Markus Heyl
Center for Electronic Correlations and Magnetism,
University of Augsburg, 86135 Augsburg, Germany

Neural quantum states (NQSs) have emerged as a novel promising numerical method to solve the quantum many-body problem. However, it has remained a central challenge to train modern large-scale deep network architectures to desired quantum state accuracy, which would be vital in utilizing the full power of NQSs and making them competitive or superior to conventional numerical approaches. Here, we propose a minimum-step stochastic reconfiguration (MinSR) method that reduces the optimization cost by orders of magnitude while keeping similar accuracy as compared to conventional stochastic reconfiguration. MinSR allows for accurate training on unprecedentedly deep NQS with up to 64 layers and more than 10^5 parameters in the spin-1/2 Heisenberg J_1 - J_2 models on the square lattice. We find that this approach yields better variational energies as compared to existing numerical results and we further observe that the accuracy of our ground state calculations approaches different levels of machine precision on modern GPU and TPU hardware. The MinSR method opens up the potential to make NQS superior as compared to conventional computational methods with the capability to address yet inaccessible regimes for two-dimensional quantum matter in the future.

As a fundamental concept in quantum physics, the ground state wave function plays a central role in understanding the behavior of many-body quantum systems. The accurate numerical solution of ground states, however, becomes an extraordinary challenge for existing numerical methods, especially in complex and large two-dimensional systems. The respective challenges depend on the individual utilized method, such as the “curse of dimensionality” in exact diagonalization (ED) [1], the notorious sign problem [2] in quantum Monte Carlo (QMC) [3], or the entanglement growth and matrix contraction complexity in tensor network (TN) methods [4].

Recently, the neural quantum state (NQS) has been introduced as a promising alternative for the calculation of ground states of quantum matter by means of artificial

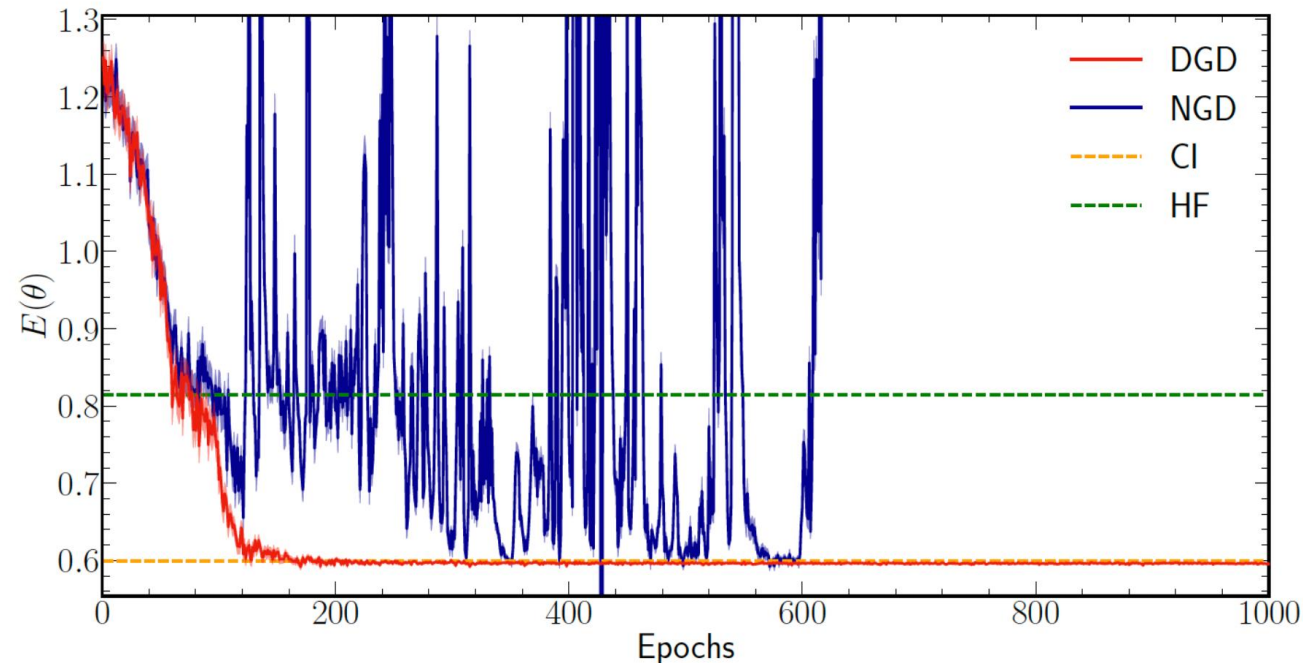
nature of entanglement in the spin chain. We are far from such a detailed understanding of machine learning inspired methods.

Thus it is natural that some studies have related complex RBM states to tensor network states [17, 18]. But these studies are mostly based on constructing abstract mappings between RBM wave functions and tensor net-

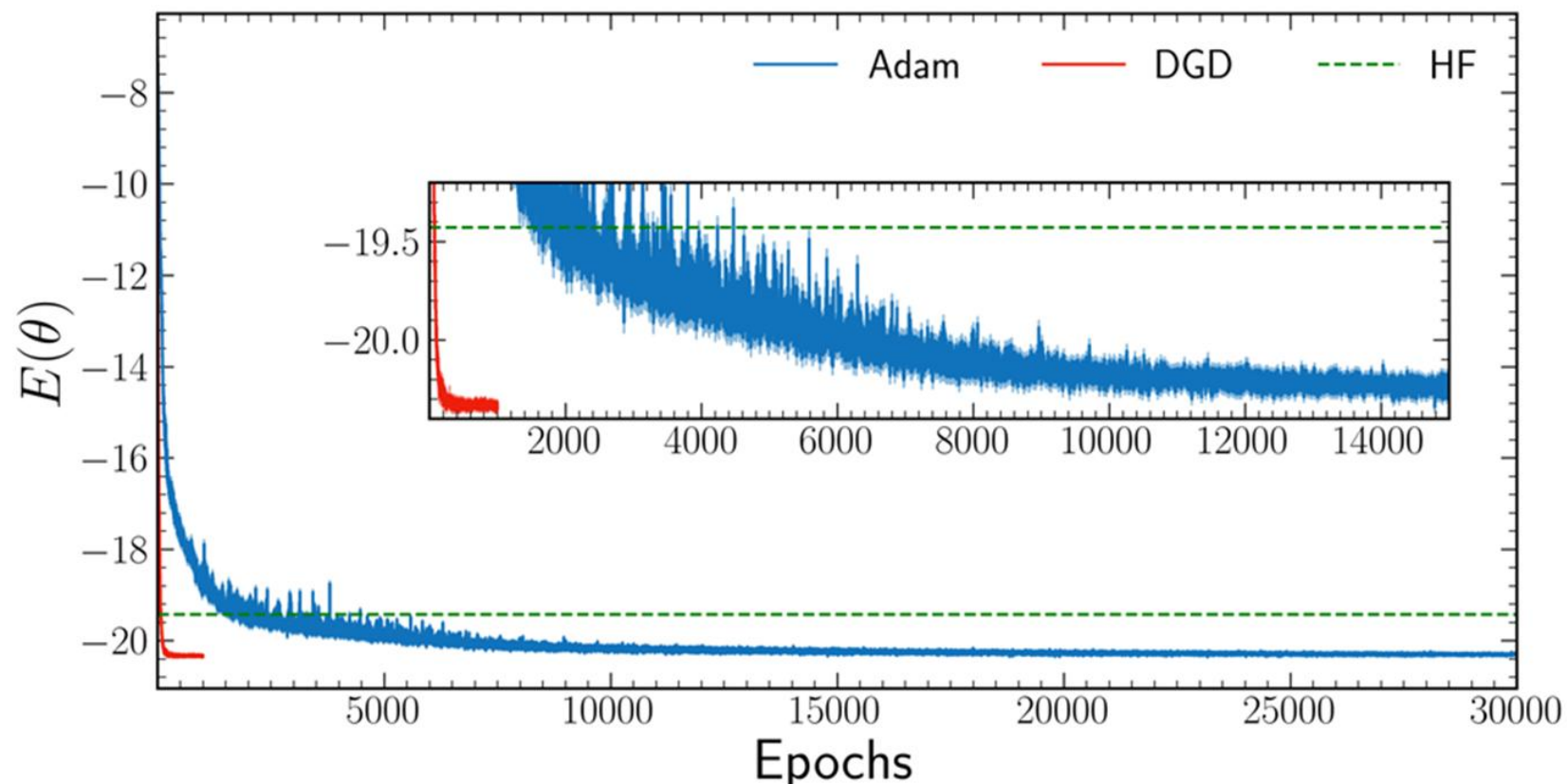
works, we find significantly lower variational energies outperforming conventional numerical approaches up to lattice sizes of 16×16 spins. Most importantly, we observe that our ground state results reach different levels of machine precision. Thus, with MinSR we are able to reach the frontier where the sole limitation of applying the NQS approach to complex two-dimensional quantum matter is not anymore the expressive power of the neural network but rather the inherent numerical precision of the computing device. This is of key importance to exploit the full power of NQS for the calculation of ground states in the future opening up the potential to address yet inaccessible regimes of quantum many-body systems also in higher spatial dimensions.

Decisional Gradient Descent

- **System:** chain of **1D spin-polarized fermions** $H = -\frac{1}{2} \sum_{i=1}^A \nabla_i^2 + \frac{1}{2} \sum_{i=1}^A x_i^2 + \frac{V_0}{\sqrt{2\pi}\sigma_0} \sum_{i<j} e^{-\frac{(x_i-x_j)^2}{2\sigma_0^2}}$
- **Quadratic term:** $Q_{\text{VMC}}(\theta)_{\theta_i \theta_j} \equiv \frac{1}{4} \sum_{k=1}^A E_{X \sim p_\theta} [\partial_{\theta_i} \partial_{x_k} \ln p_\theta(X) \partial_{\theta_j} \partial_{x_k} \ln p_\theta(X)]$

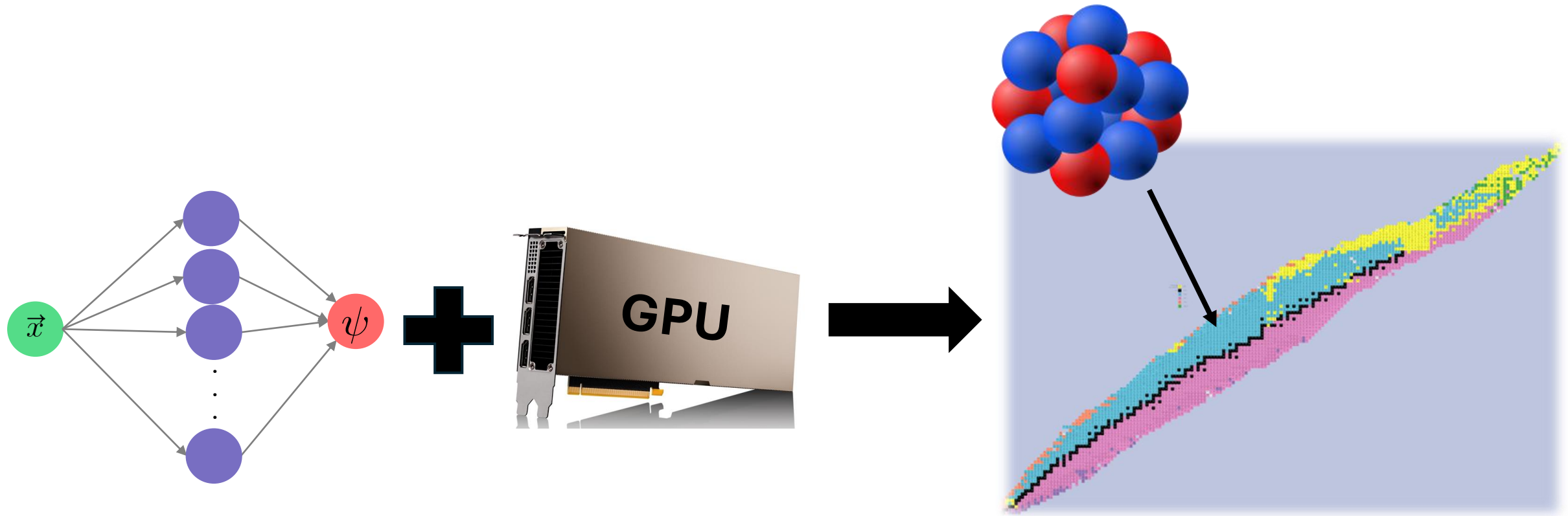


DGD vs. ADAM



Conclusions

Takeaway



Thank you



J. Rozalén Sarmiento, A. Rios



javi.rozalen@icc.ub.edu



Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA



EXCELENCIA
MARÍA
DE MAEZTU
2020-2024



**UNIVERSITAT_{DE}
BARCELONA**