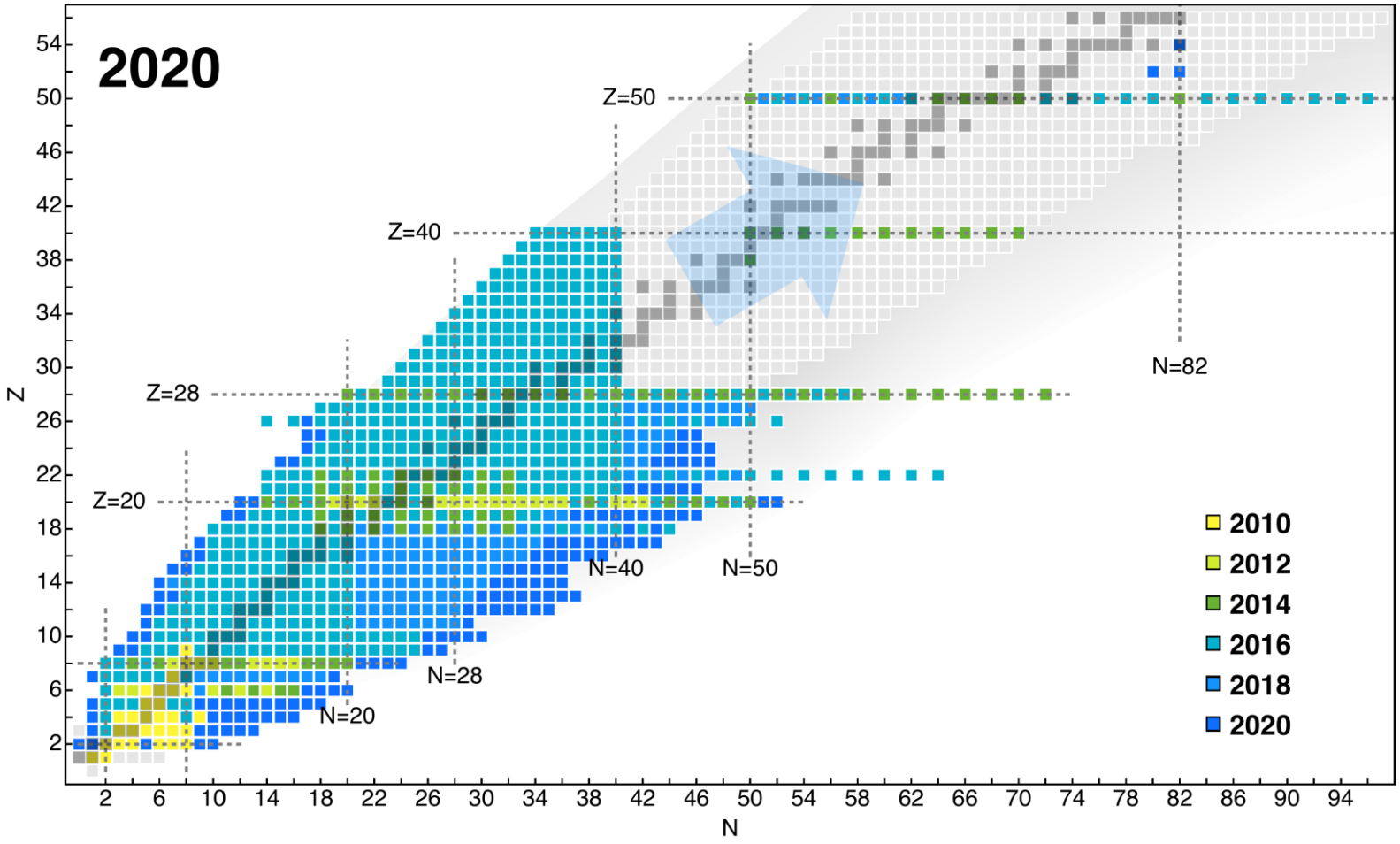
Decisional Gradient Descent: A new Optimizer for Variational Monte Carlo

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

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

ICCUB | Institut de Ciències del Cosmos UNIVERSITAT DE BARCELONA
EXCELENCIA MARÍA DE MAEZTU 2020-2024
CPAN Ingenio 2010

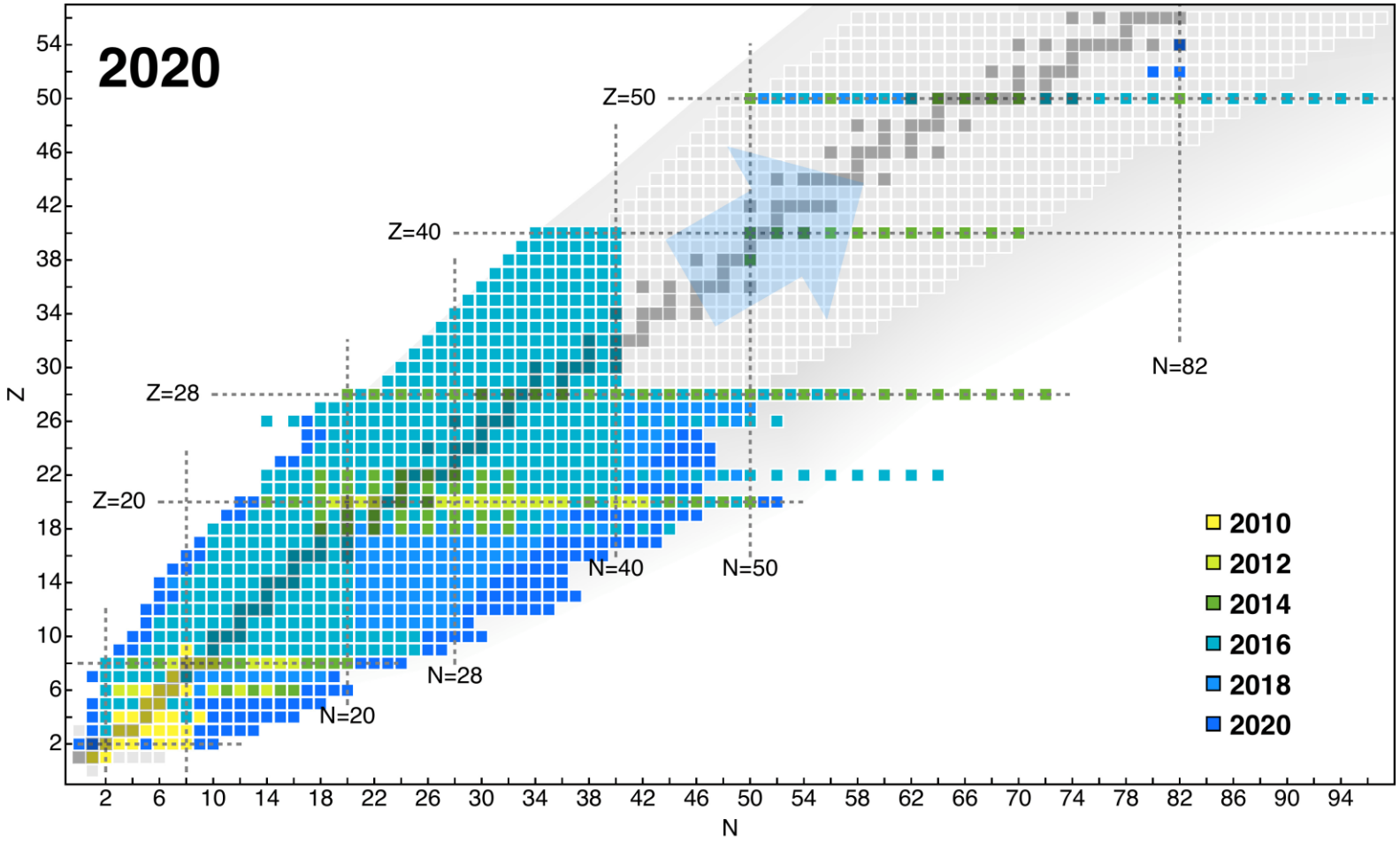
Ab Initio Nuclear Structure



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

Ab Initio Nuclear Structure

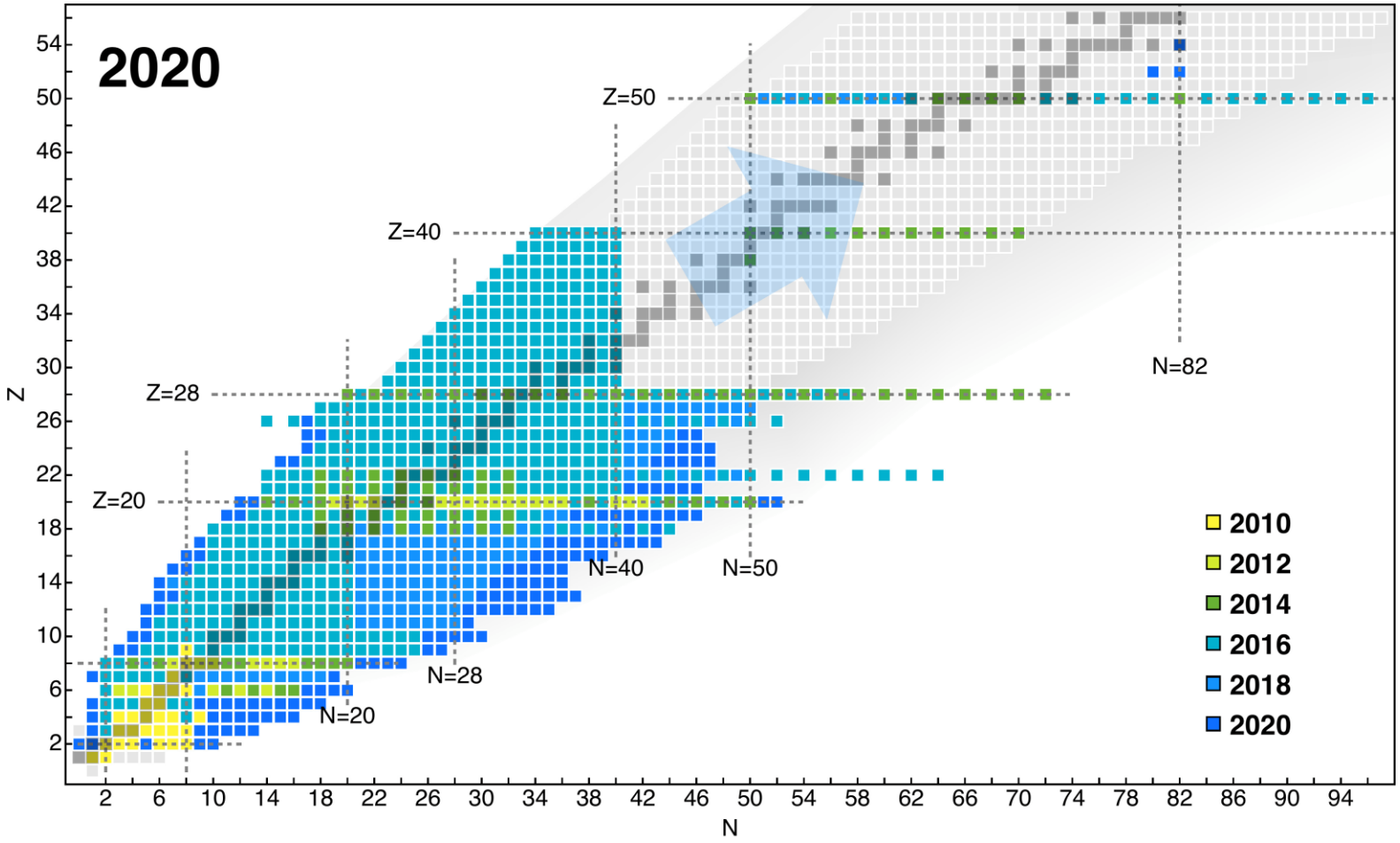
- IT-NCSM



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

Ab Initio Nuclear Structure

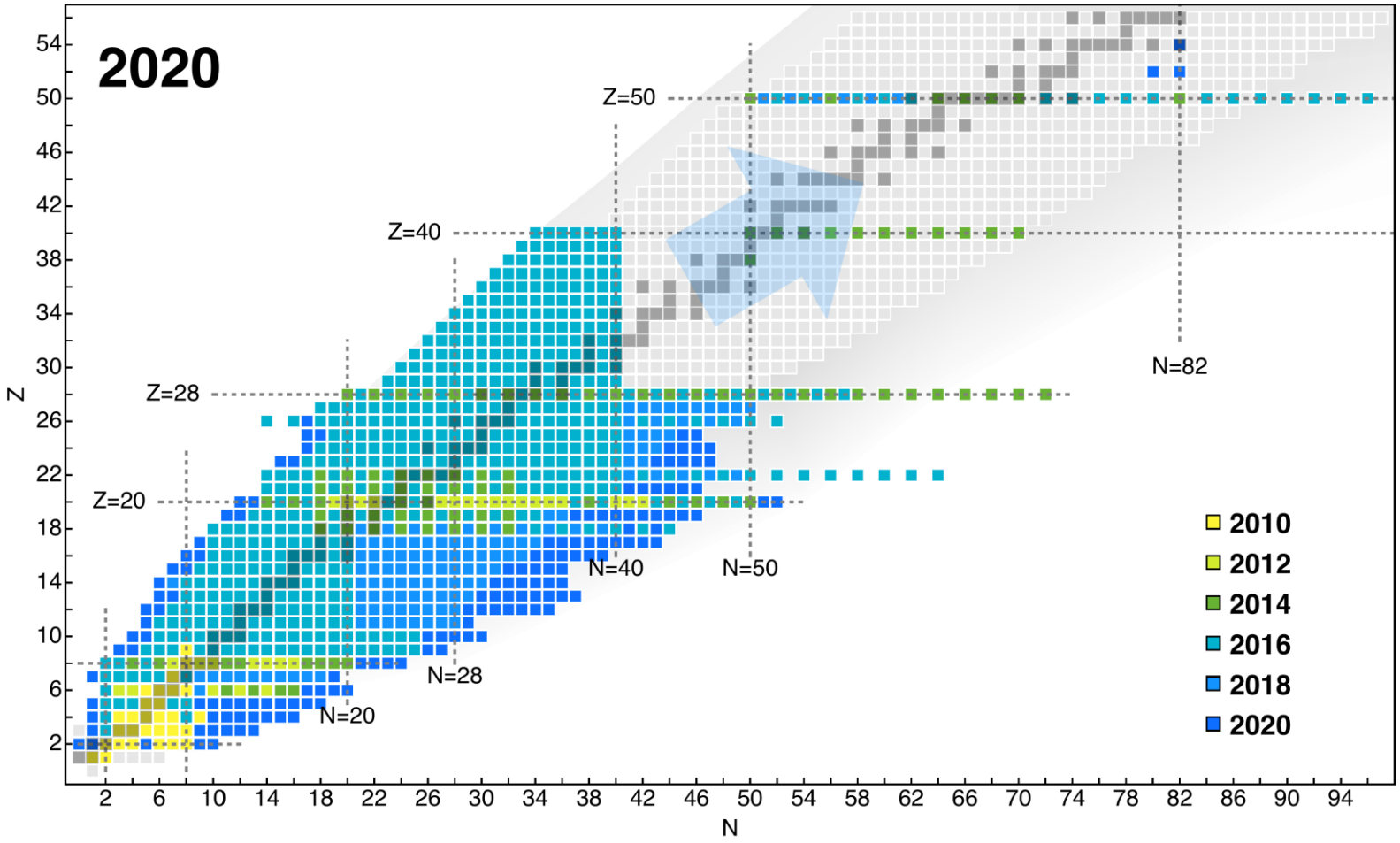
- IT-NCSM
- MR-IMSRG



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

Ab Initio Nuclear Structure

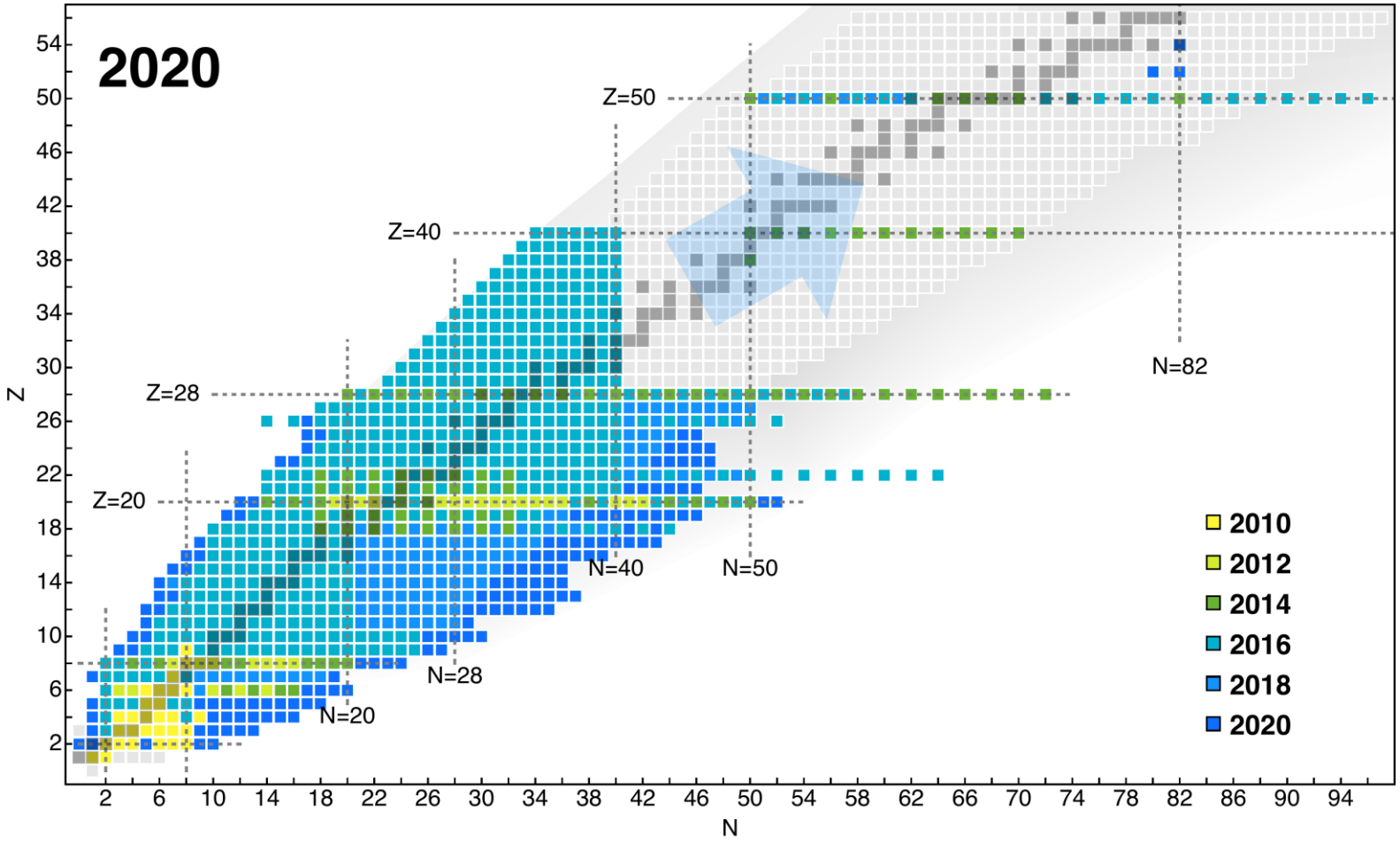
- IT-NCSM
- MR-IMSRG
- VS-IMSRG



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

Ab Initio Nuclear Structure

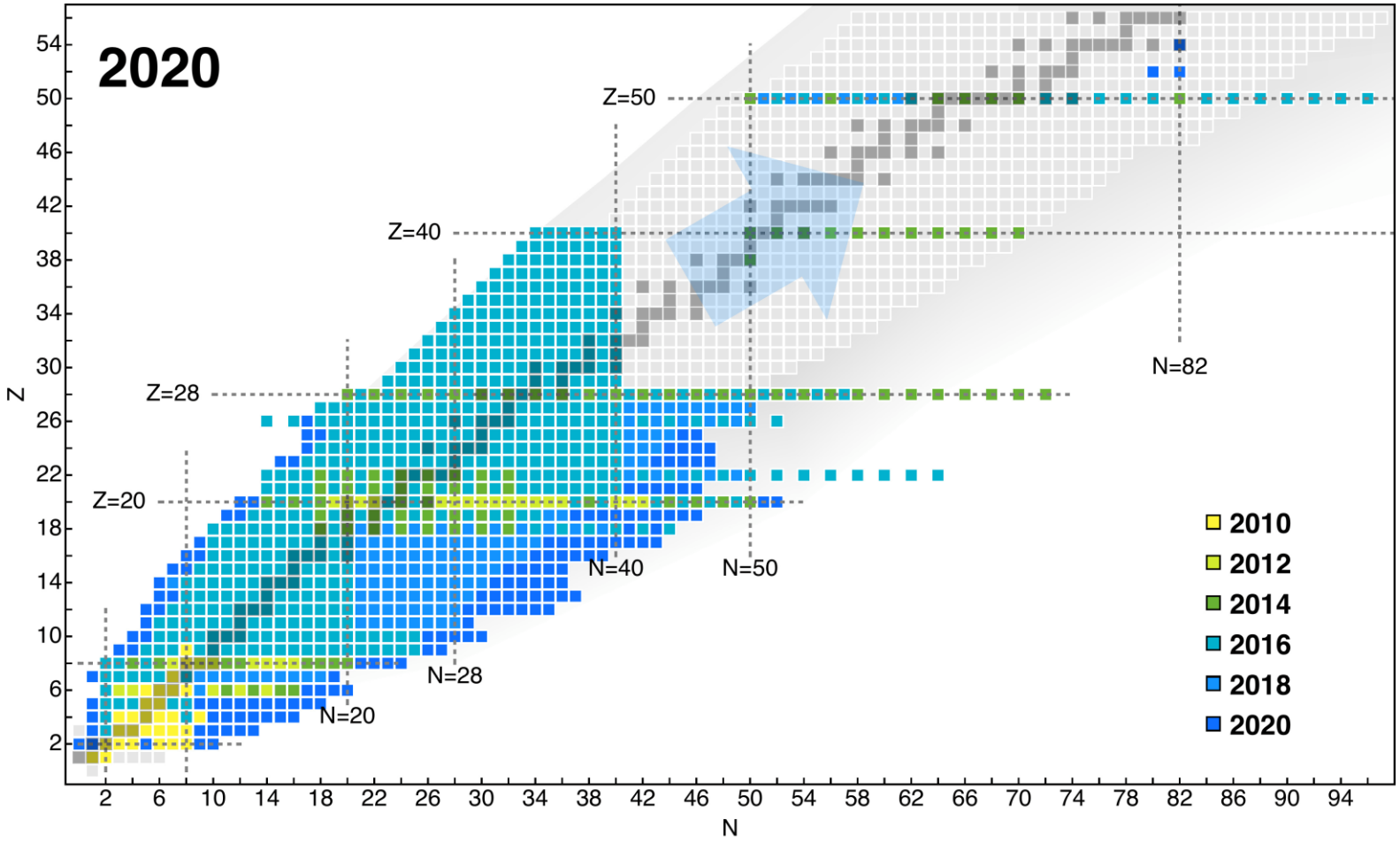
- IT-NCSM
- MR-IMSRG
- VS-IMSRG
- ADC



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

Ab Initio Nuclear Structure

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

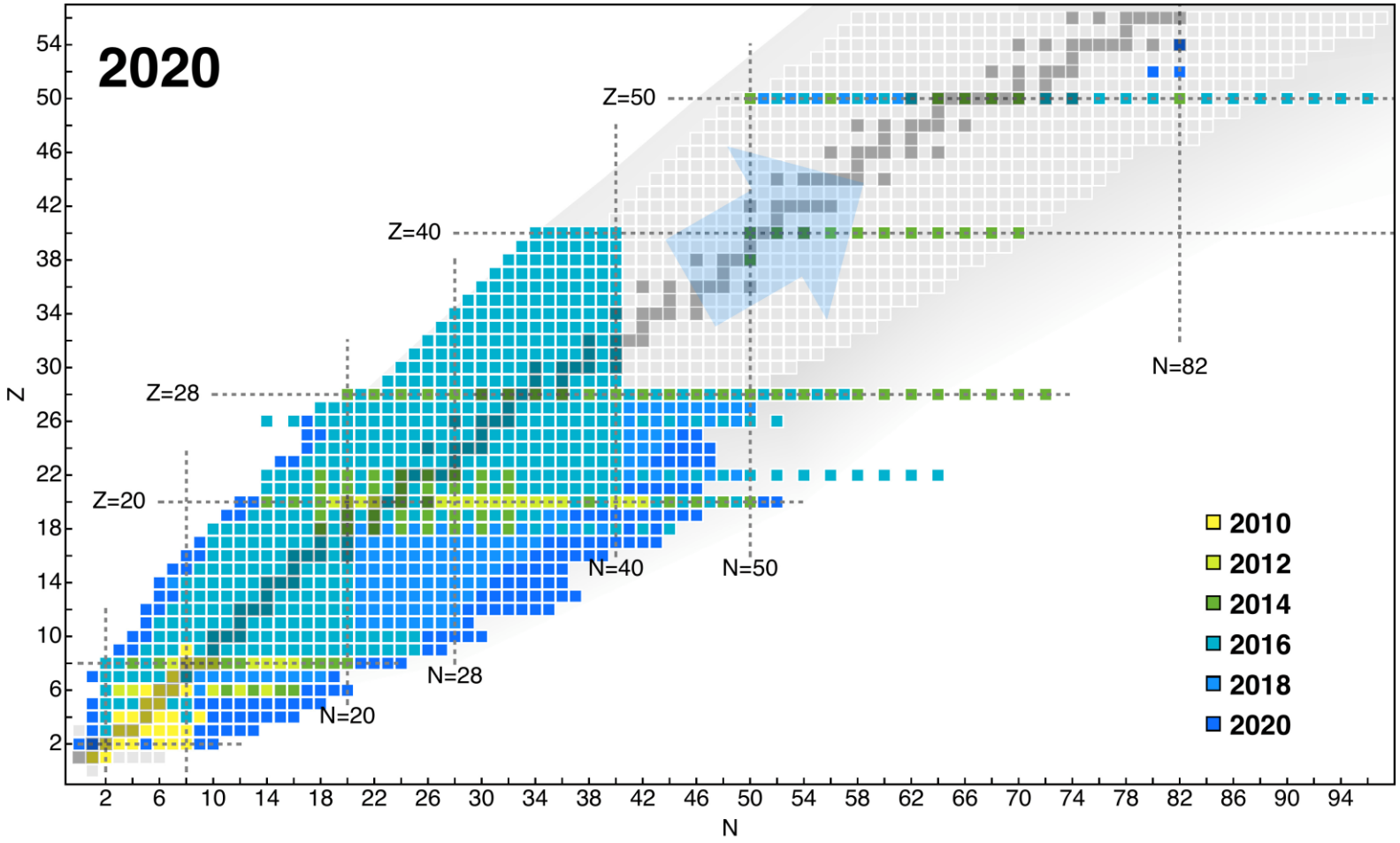


- Lattice EFT

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

Ab Initio Nuclear Structure

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

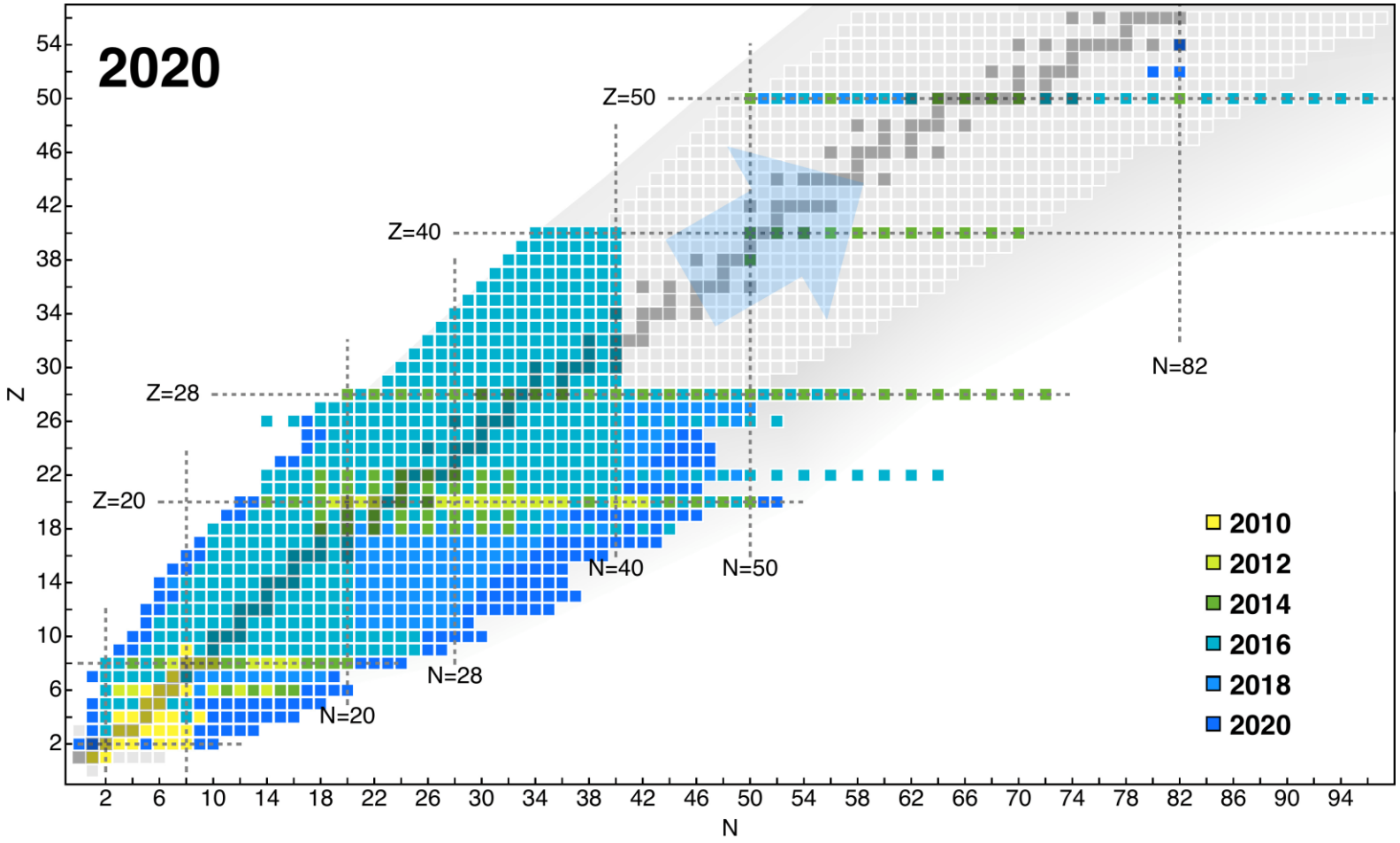


- Lattice EFT
- CCSD

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

Ab Initio Nuclear Structure

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

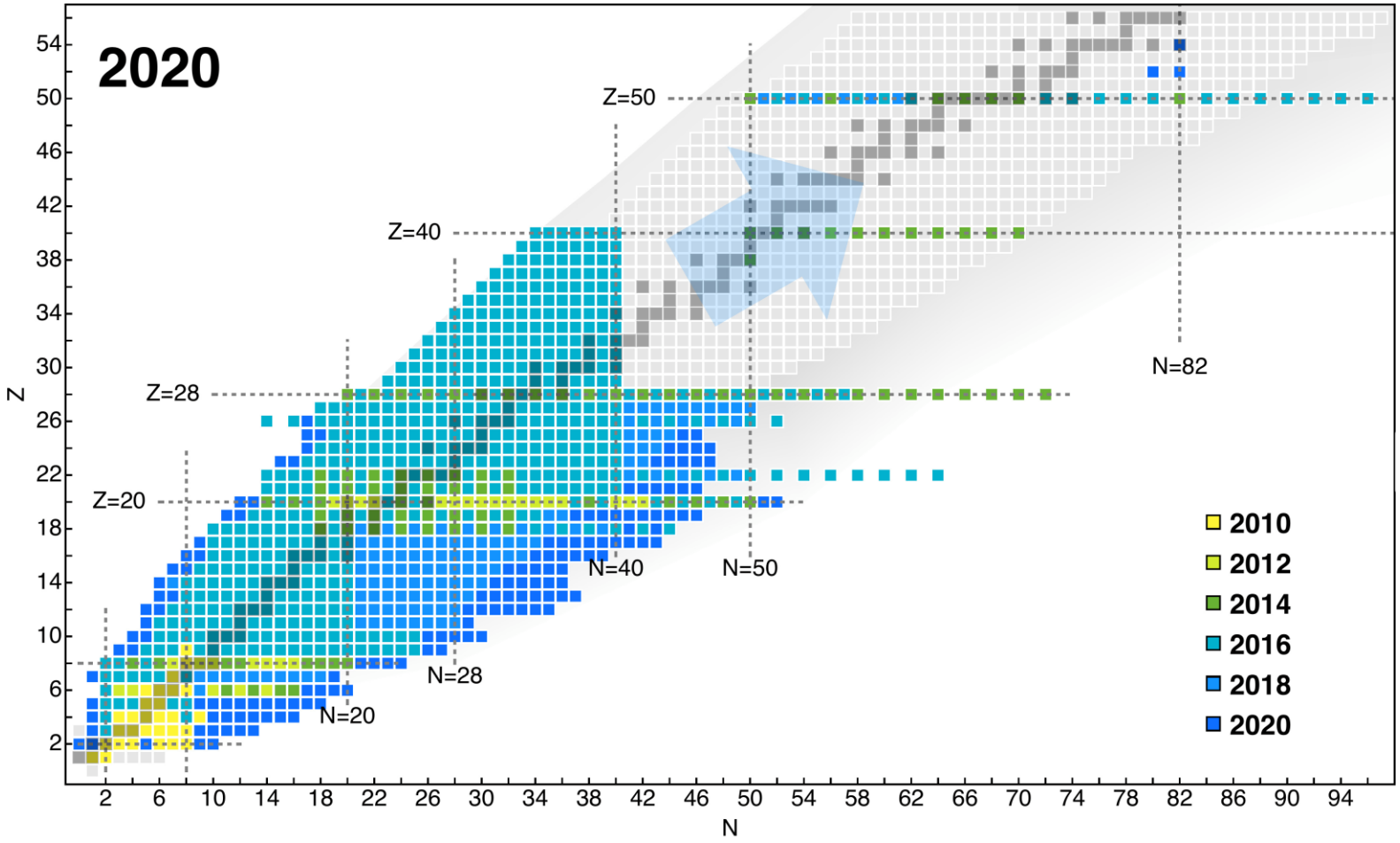


- Lattice EFT
- CCSD
- Λ-CCSD

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

Ab Initio Nuclear Structure

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

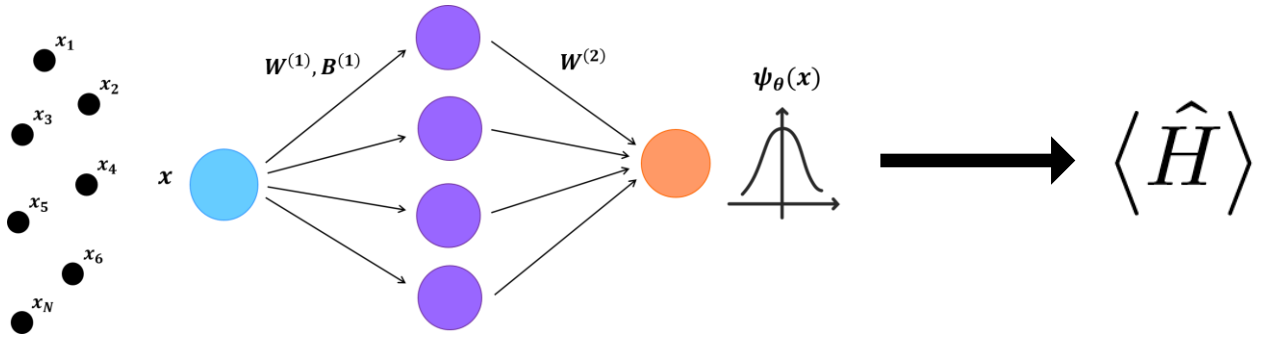


- Lattice EFT
- CCSD
- Λ-CCSD
- **NQS**

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

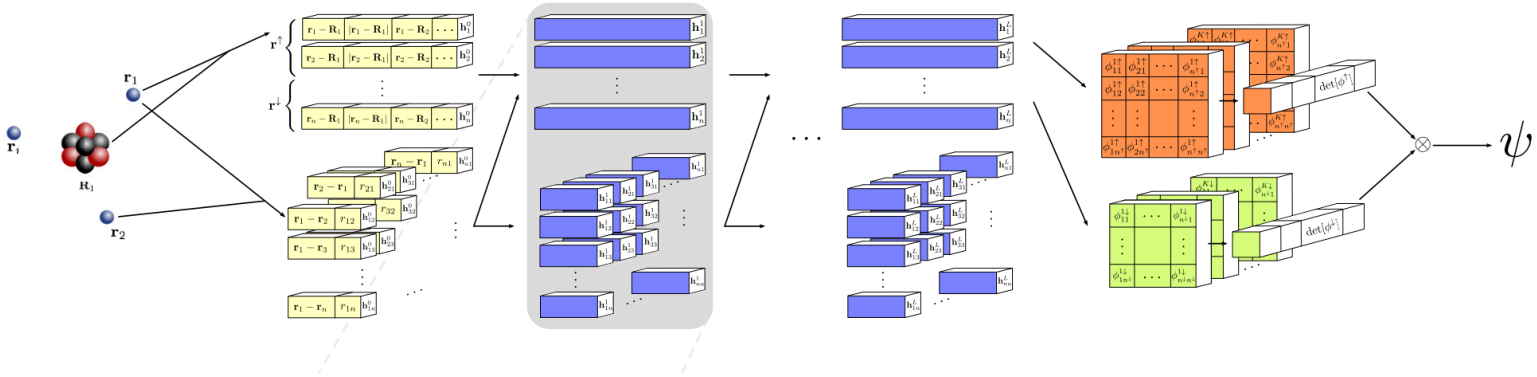
Neural Quantum States

- **NQS = Variational Monte Carlo + NNs**



G. Carleo & M. Troyer, Science **355** (2017)

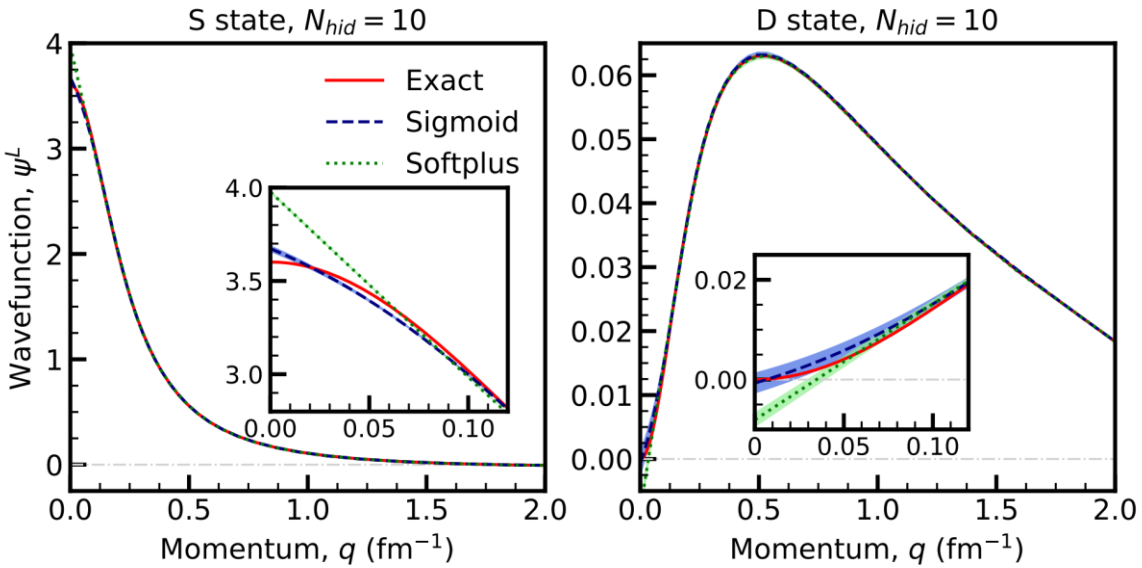
- **FermiNet:**

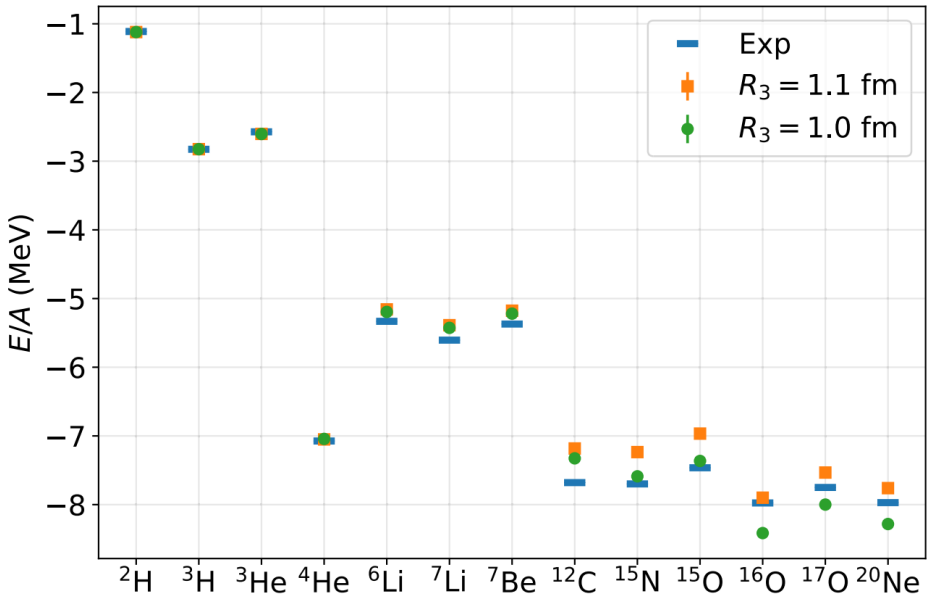


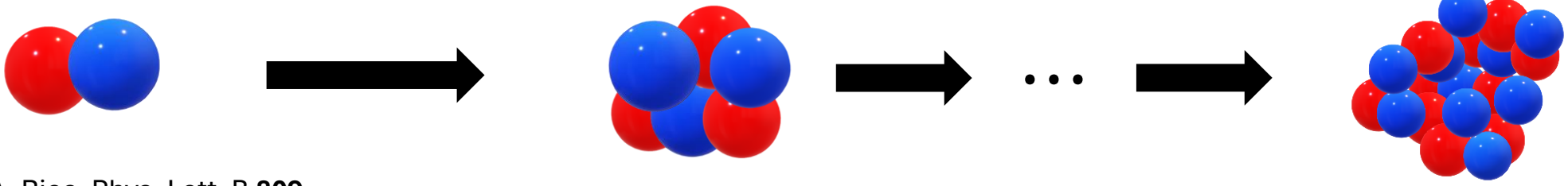
D. Pfau et al, Phys. Rev. Research **2** (2020)

NQS for Nuclei

2020 → … → **2024**







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

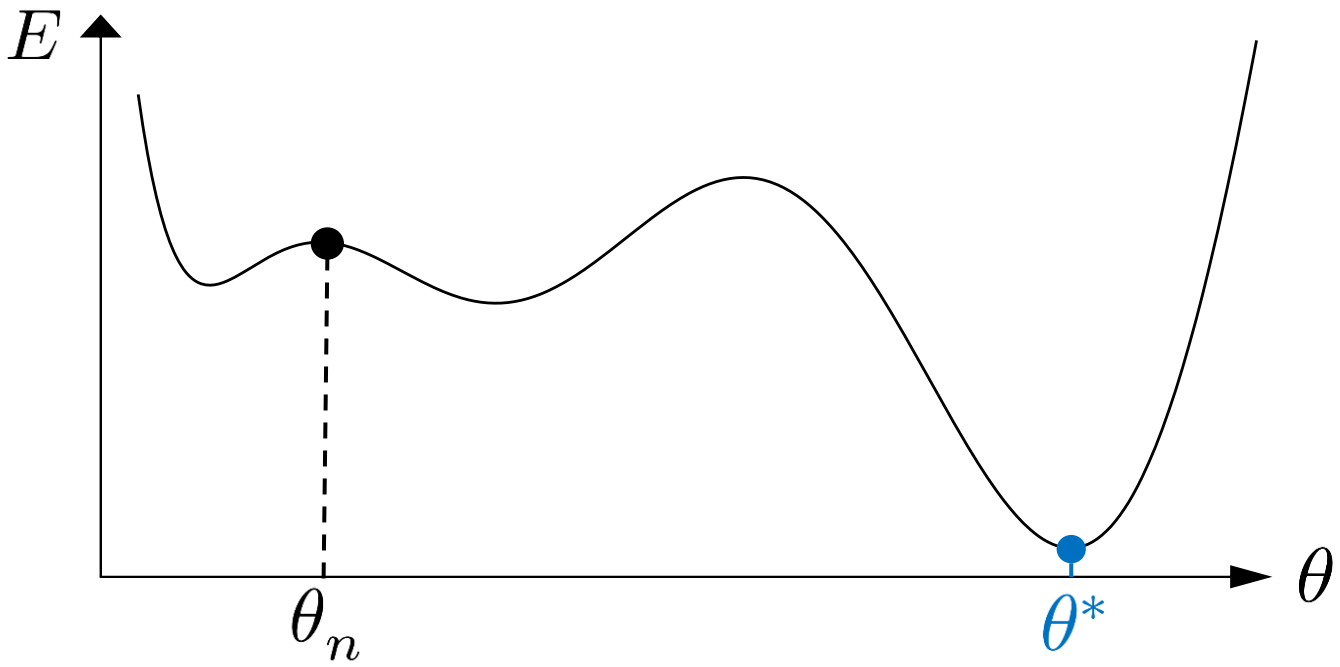
A. Gnech et al, Few-Body Syst. **63** (2022)

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

Part I: Optimisation

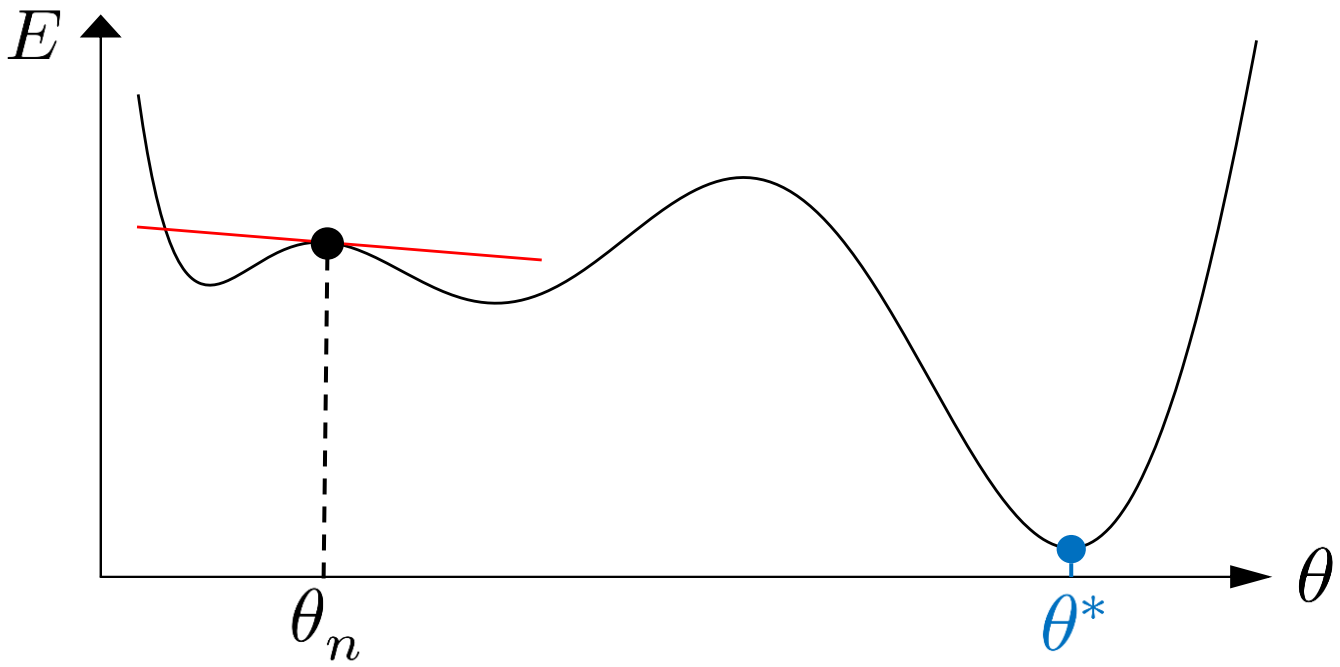
Learning the ground state

- **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 $\delta_n \equiv \text{argmin}_{\delta \in T_n} M_n(\delta)$



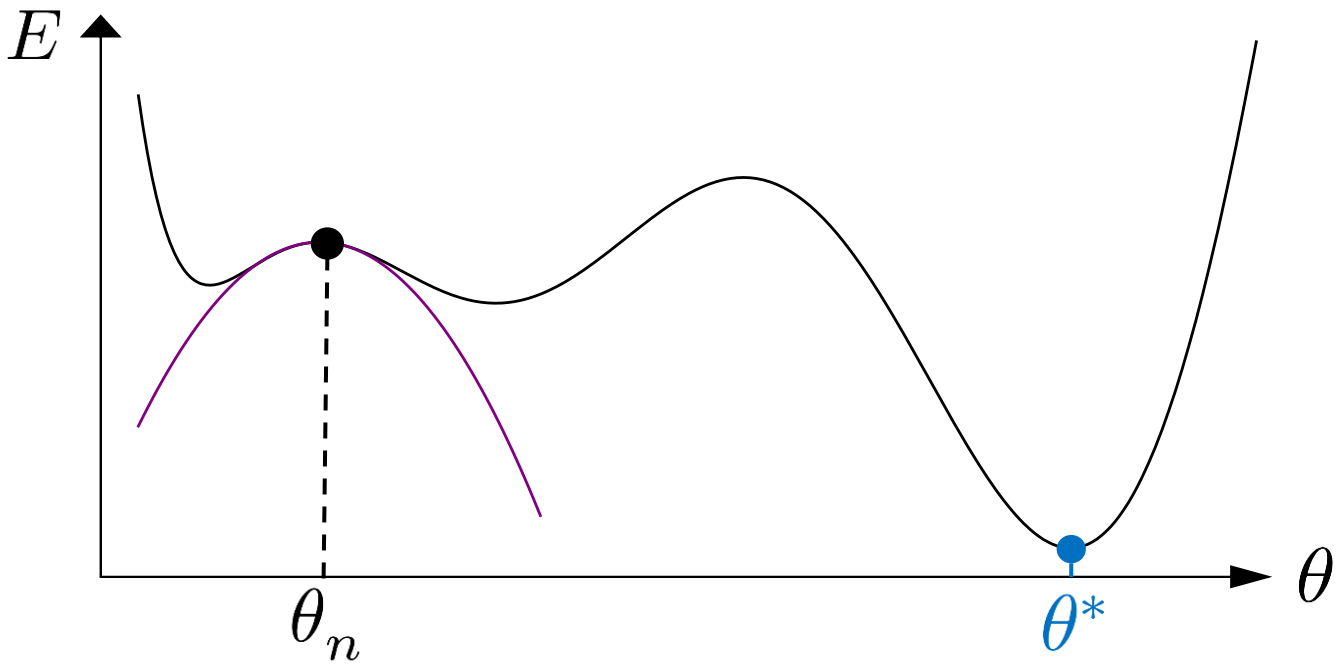
Learning the ground state

- **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 $\delta_n \equiv \text{argmin}_{\delta \in T_n} M_n(\delta)$



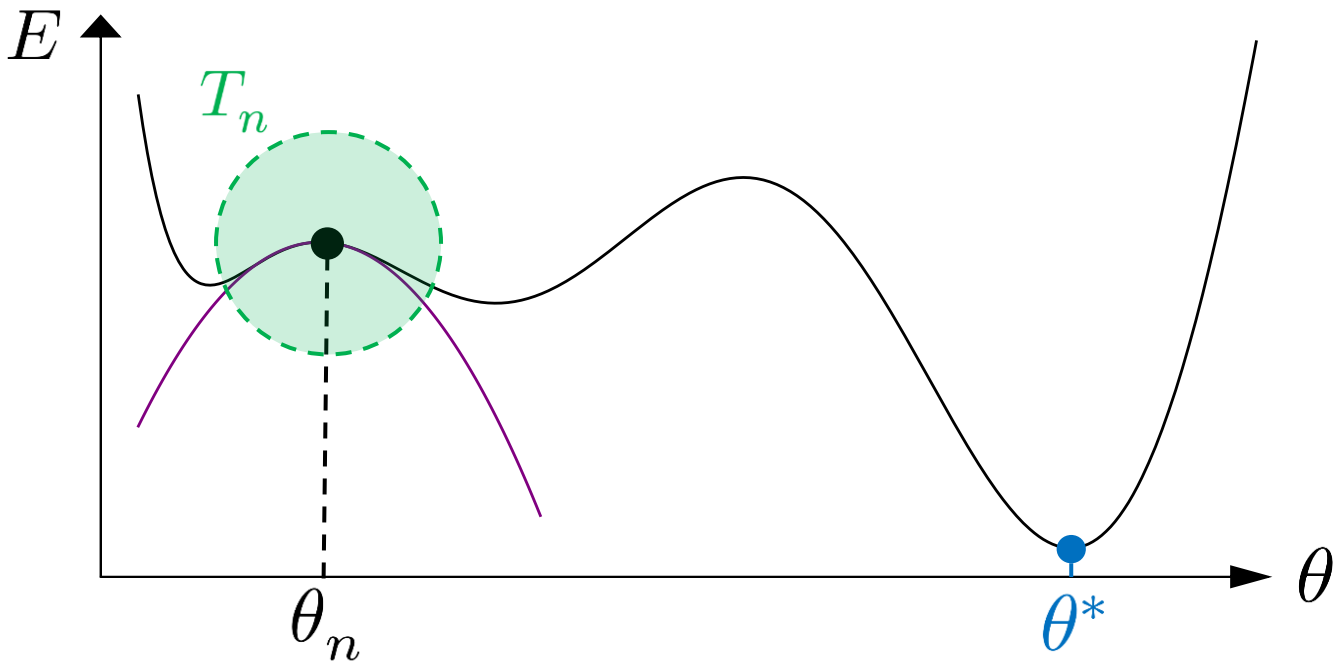
Learning the ground state

- **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 $\delta_n \equiv \text{argmin}_{\delta \in T_n} M_n(\delta)$



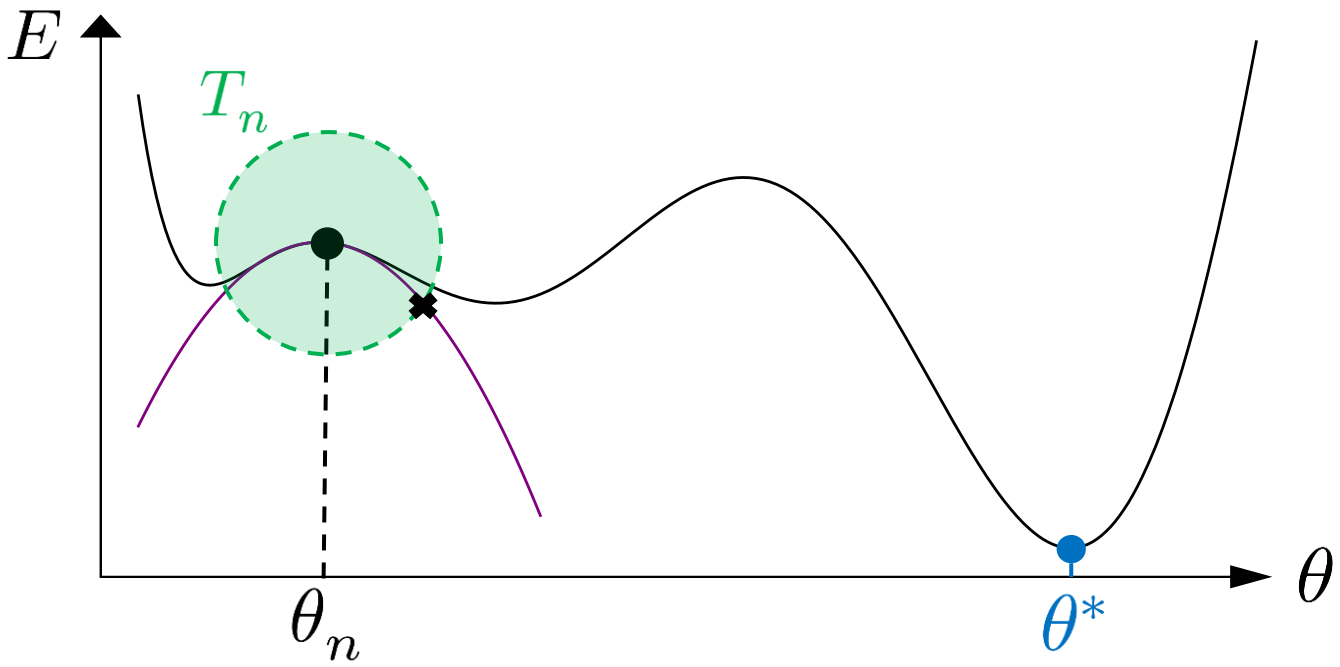
Learning the ground state

- **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 $\delta_n \equiv \text{argmin}_{\delta \in T_n} M_n(\delta)$



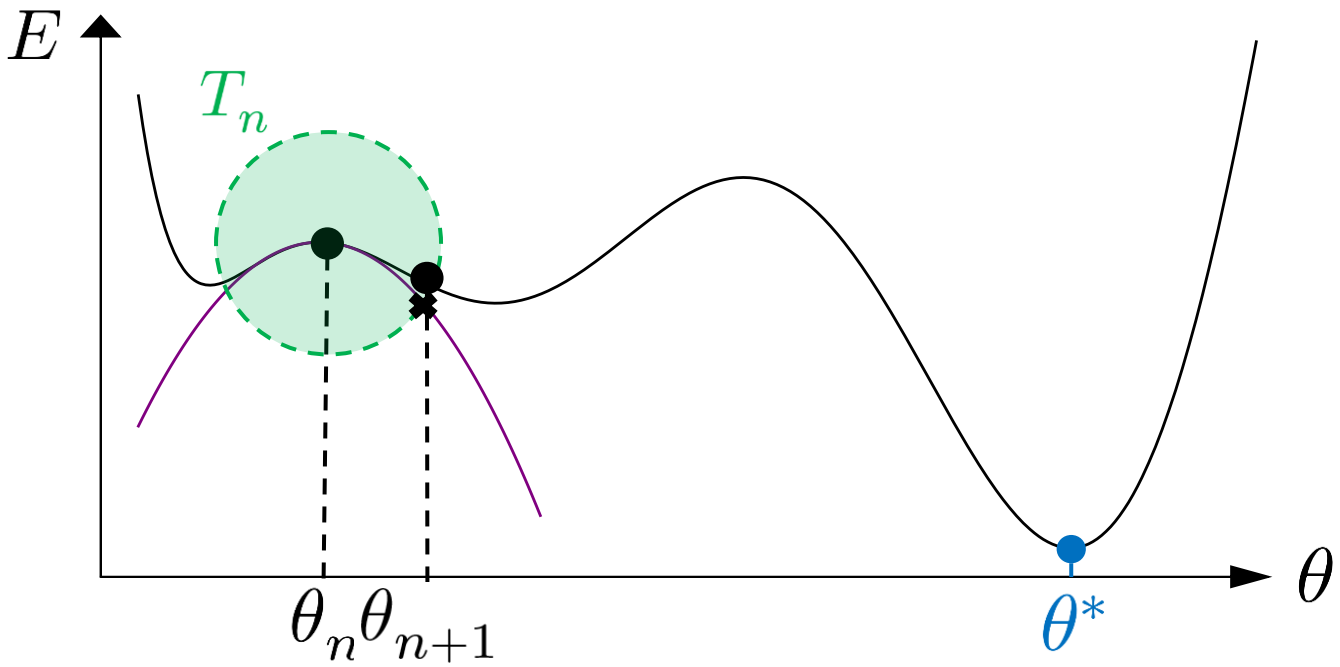
Learning the ground state

- **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 $\delta_n \equiv \text{argmin}_{\delta \in T_n} M_n(\delta)$



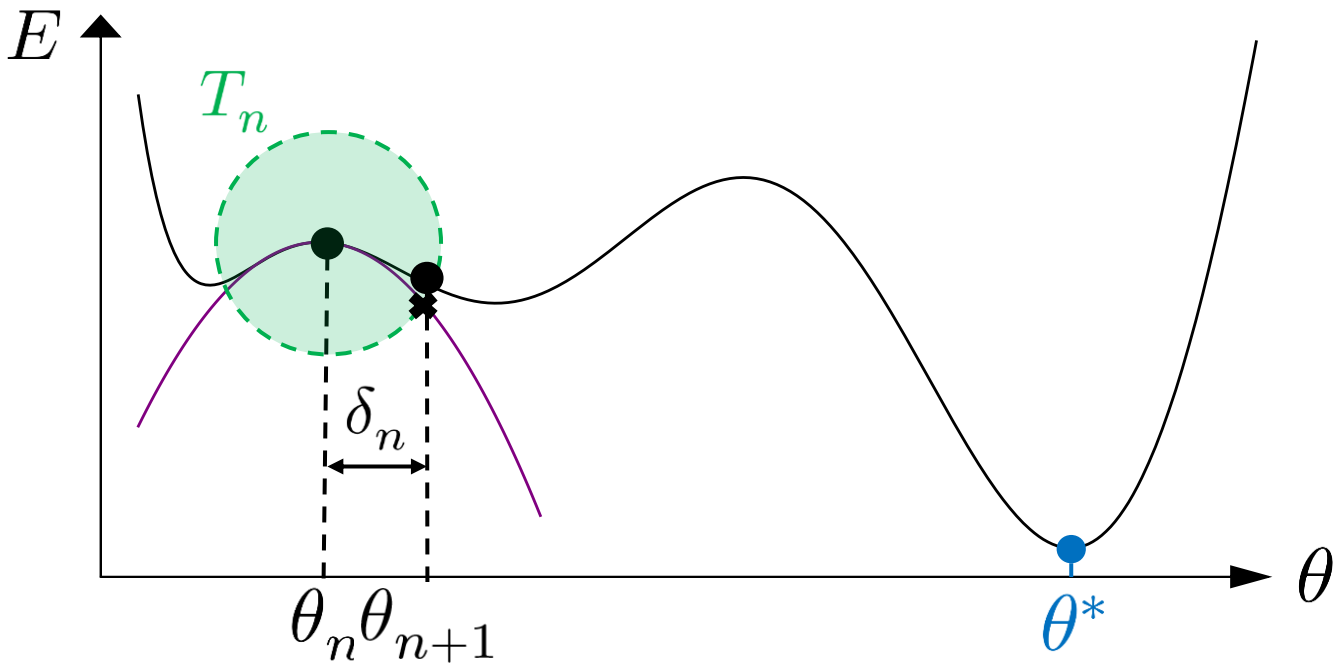
Learning the ground state

- **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 $\delta_n \equiv \text{argmin}_{\delta \in T_n} M_n(\delta)$



Learning the ground state

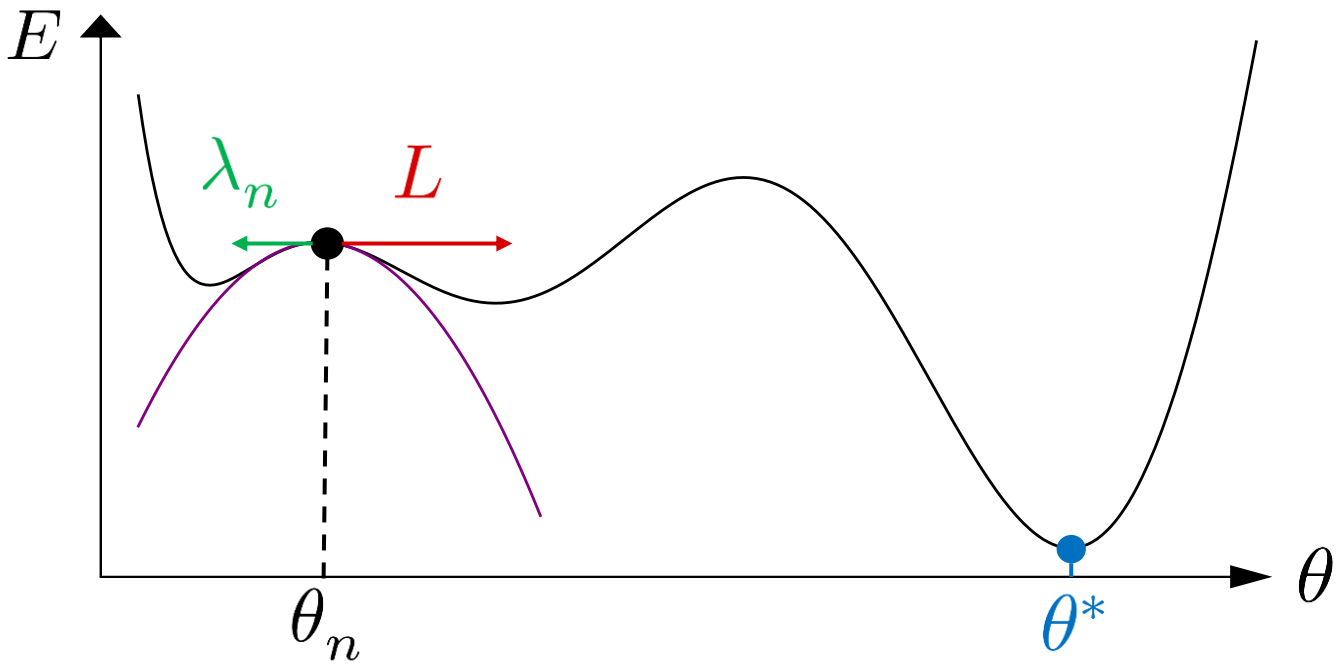
- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$
- **Trust region:** $T_n = \{\delta : \delta^T R_n \delta \leq r^2\}$
- **Update:** $\theta_{n+1} = \theta_n + \delta_n$ where $\delta_n \equiv \text{argmin}_{\delta \in T_n} M_n(\delta)$



Learning the ground state

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$
- **Trust region:** $T_n = \{\delta : \delta^T R_n \delta \leq r^2\}$
- **Tikhonov regularisation:** $\delta_n \equiv \text{argmin}_{\delta \in T_n} \; M_n(\delta) = \text{argmin}_\delta \left(M_n(\delta) + \frac{1}{2}\lambda_n \delta^\mathsf{T} R_n \delta \right)$

dampening

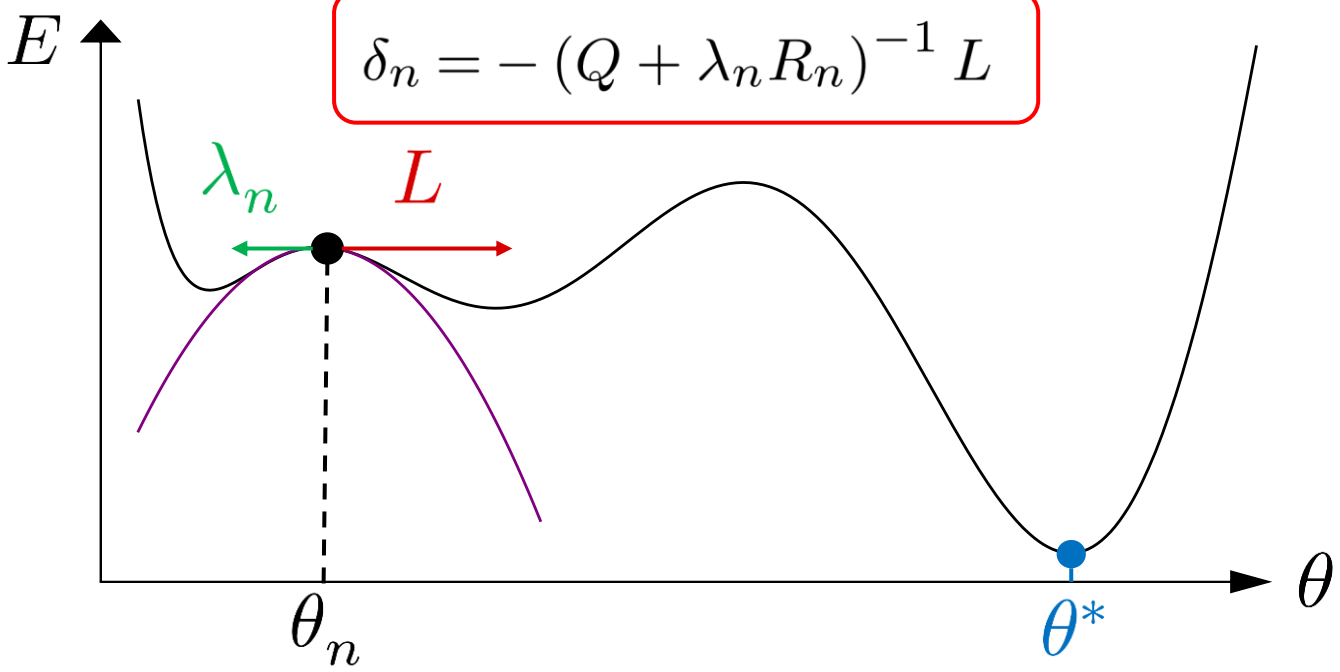


Learning the ground state

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$
- **Trust region:** $T_n = \{\delta : \delta^T R_n \delta \leq r^2\}$
- **Tikhonov regularisation:** $\delta_n \equiv \text{argmin}_{\delta \in T_n} \; M_n(\delta) = \text{argmin}_\delta \left(M_n(\delta) + \frac{1}{2}\lambda_n \delta^\mathsf{T} R_n \delta \right)$

dampening

$$\delta_n = -\left(Q + \lambda_n R_n\right)^{-1} L$$

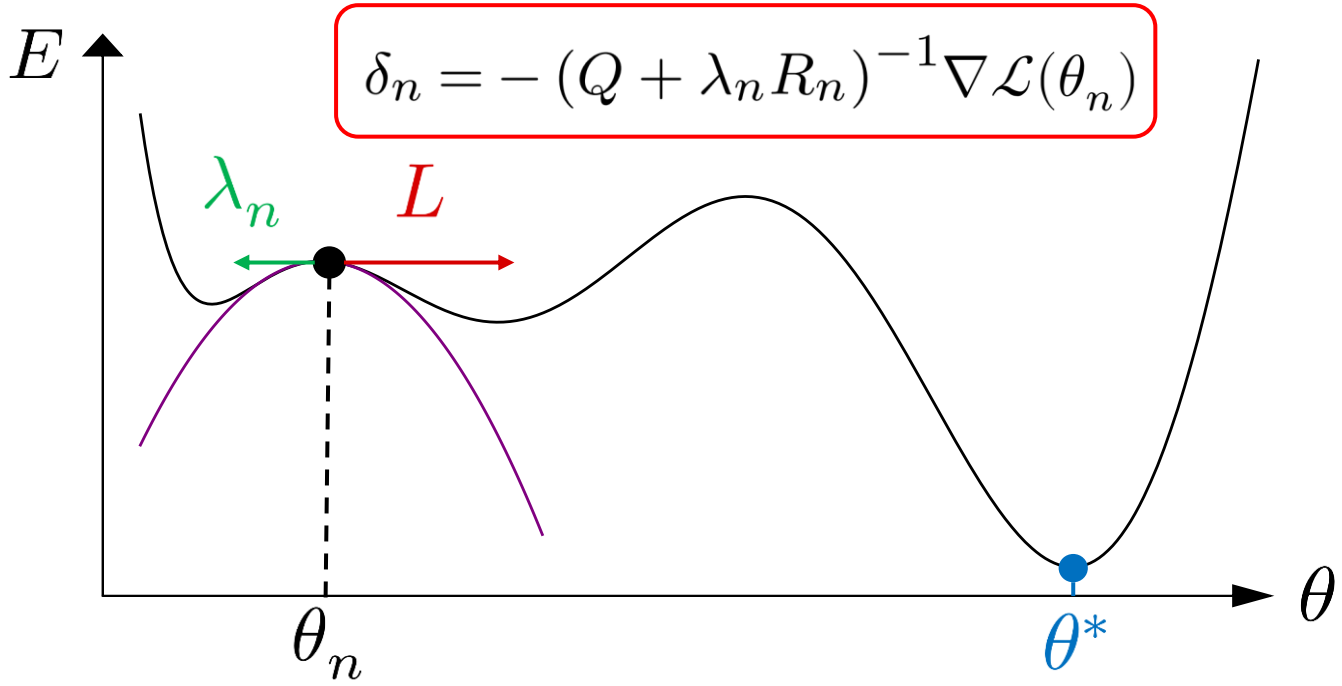


Learning the ground state

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$
- **Trust region:** $T_n = \{\delta : \delta^T R_n \delta \leq r^2\}$
- **Tikhonov regularisation:** $\delta_n \equiv \text{argmin}_{\delta \in T_n} \; M_n(\delta) = \text{argmin}_\delta \left(M_n(\delta) + \frac{1}{2}\lambda_n \delta^\mathsf{T} R_n \delta \right)$

dampening

$$\delta_n = -\left(Q + \lambda_n R_n\right)^{-1} \nabla \mathcal{L}(\theta_n)$$

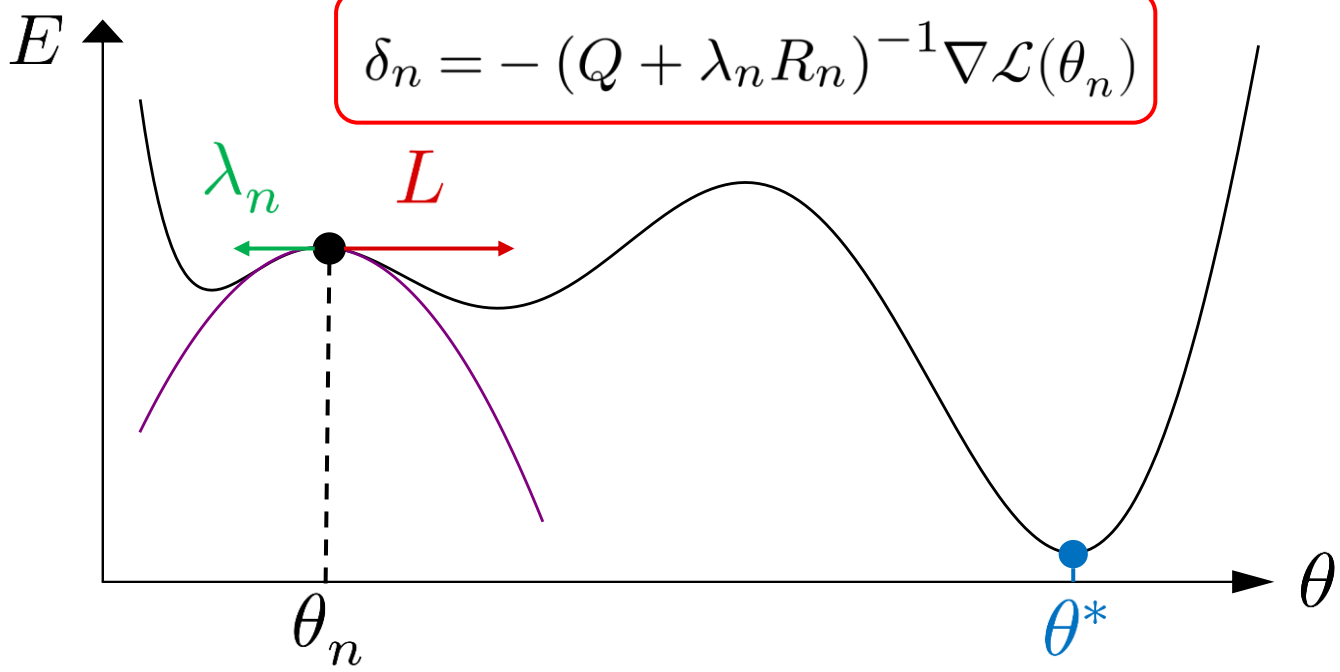


Learning the ground state

- **Local Quadratic Model:** $M_n(\delta) = \frac{1}{2}\delta^T Q \delta + L^T \delta + C$
- **Trust region:** $T_n = \{\delta : \delta^T R_n \delta \leq r^2\}$
- **Tikhonov regularisation:** $\delta_n \equiv \text{argmin}_{\delta \in T_n} \; M_n(\delta) = \text{argmin}_\delta \left(M_n(\delta) + \frac{1}{2}\lambda_n \delta^\mathsf{T} R_n \delta \right)$

dampening

$$\delta_n = -\left(Q + \lambda_n R_n\right)^{-1} \nabla \mathcal{L}(\theta_n)$$



$Q?$

Part II: Decision Geometry

Decision Geometry 101

- **Loss function** from a scoring rule $\mathcal{L}(\theta) \equiv -\mathbb{E}_{X\sim p_\theta}\left[\mathcal{S}(X, p_\theta)\right]$
- **Expected score** $S(p, q) \equiv \mathbb{E}_{X\sim p}\left[\mathcal{S}(X, q)\right]$
- **Divergence** $D_S(p, q) \equiv S(p, q) - S(p, p) \geq 0$
- **Taylor-expansion** of divergence $D_S(p_\theta, p_{\theta+\delta}) = \frac{1}{2}\delta^T G_S(\theta)\delta + \mathcal{O}(\delta^3)$
- G_S is **positive semi-definite**

probability

local scoring rule

A. P. Dawid, Ann. Inst. Stat. Math. **59** (2007);
T. Gneiting & A. Raftery, J. Am. Stat. Assoc. **102** (2007)

Decision Geometry 101

- **Loss function** from a scoring rule $\mathcal{L}(\theta) \equiv -\mathbb{E}_{X\sim p_\theta}\left[\mathcal{S}(X, p_\theta)\right]$
- **Expected score** $S(p,q) \equiv \mathbb{E}_{X\sim p}\left[\mathcal{S}(X,q)\right]$
- **Divergence** $D_S(p,q) \equiv S(p,q) - S(p,p) \geq 0$
- **Taylor-expansion** of divergence $D_S(p_\theta, p_{\theta+\delta}) = \frac{1}{2}\delta^T G_S(\theta)\delta + \mathcal{O}(\delta^3)$
- G_S is **positive semi-definite** → Q

probability

local scoring rule

metric

$$\delta_n = -\alpha G_S^{-1}(\theta_n)\nabla\mathcal{L}(\theta_n)$$

A. P. Dawid, Ann. Inst. Stat. Math. **59** (2007);
T. Gneiting & A. Raftery, J. Am. Stat. Assoc. **102** (2007)

Natural Gradient Descent

- **Loss function** $\mathcal{L}^{\sup}(\theta) = \mathbb{E}_{X\sim q}\left[-\ln p_\theta(X)\right]$

$$p_\theta(X) = |\psi_\theta(X)|^2$$

- **Kullback-Leibler divergence** $D_{\mathrm{KL}}(p_\theta,\ p_{\theta'}) \equiv \mathbb{E}_{X\sim p_\theta}\left[-\ln p_{\theta'}(X) - (-\ln p_\theta(X))\right]$

- **Fisher Information Matrix (FIM)** $F(\theta)_{\theta_i\theta_j} \equiv \mathbb{E}_{X\sim p_\theta}\left[\partial_{\theta_i}\ln p_\theta(X)\partial_{\theta_j}\ln p_\theta(X)\right]$

$$\delta_n^{\text{NGD-reg}} = -\alpha F^{-1}(\theta_n)\nabla\mathcal{L}(\theta_n)$$

S. Amari, Electron. Commun. Jpn. **66** (1983)

A new proposal

$$p_\theta(X) = |\psi_\theta(X)|^2$$

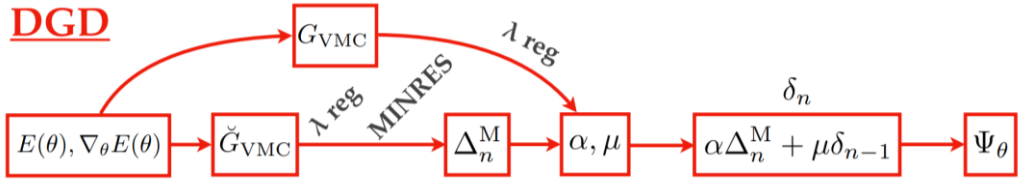
- **Local score** informed by a local estimator of the loss: $\mathcal{S}_{\mathrm{VMC}}(X, p_\theta) \equiv -E_{L,\theta}(X)$

- **Divergence** $D_{\mathrm{VMC}}(p_\theta, p_{\theta'}) = \mathbb{E}_{X\sim p_\theta}\left[E_{L,\theta}(X) - E_{L,\theta'}(X)\right]$

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

- **Quadratic model** $M_n^{\mathrm{DGD}}(\delta) = \frac{1}{2}\delta^{\mathsf{T}} G_{\mathrm{VMC}}(\theta_n)\delta + \nabla E(\theta_n)^{\mathsf{T}}\delta + E(\theta_n)$

- **Decisional Gradient Descent (DGD)**



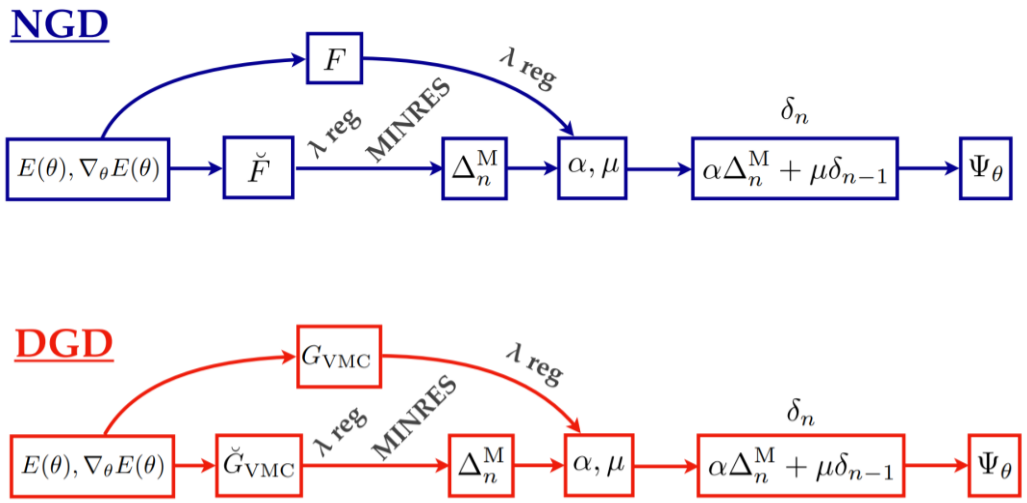
Part III: Results

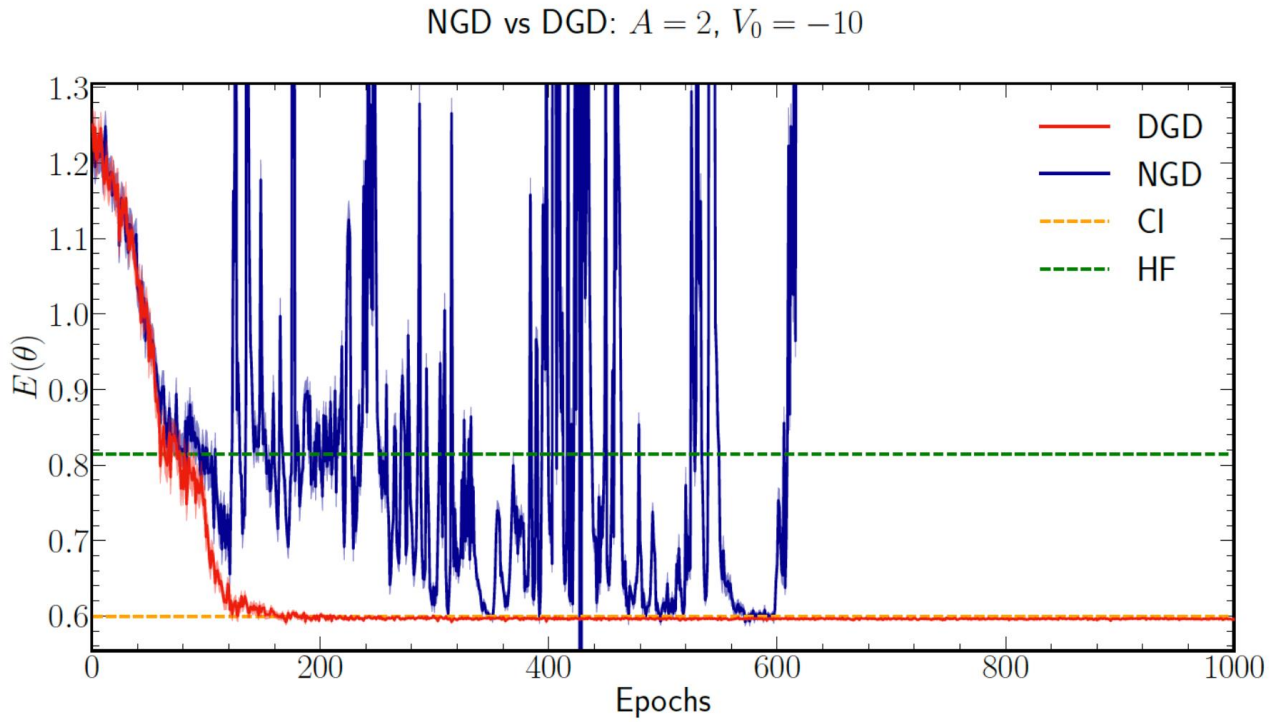
Results

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

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

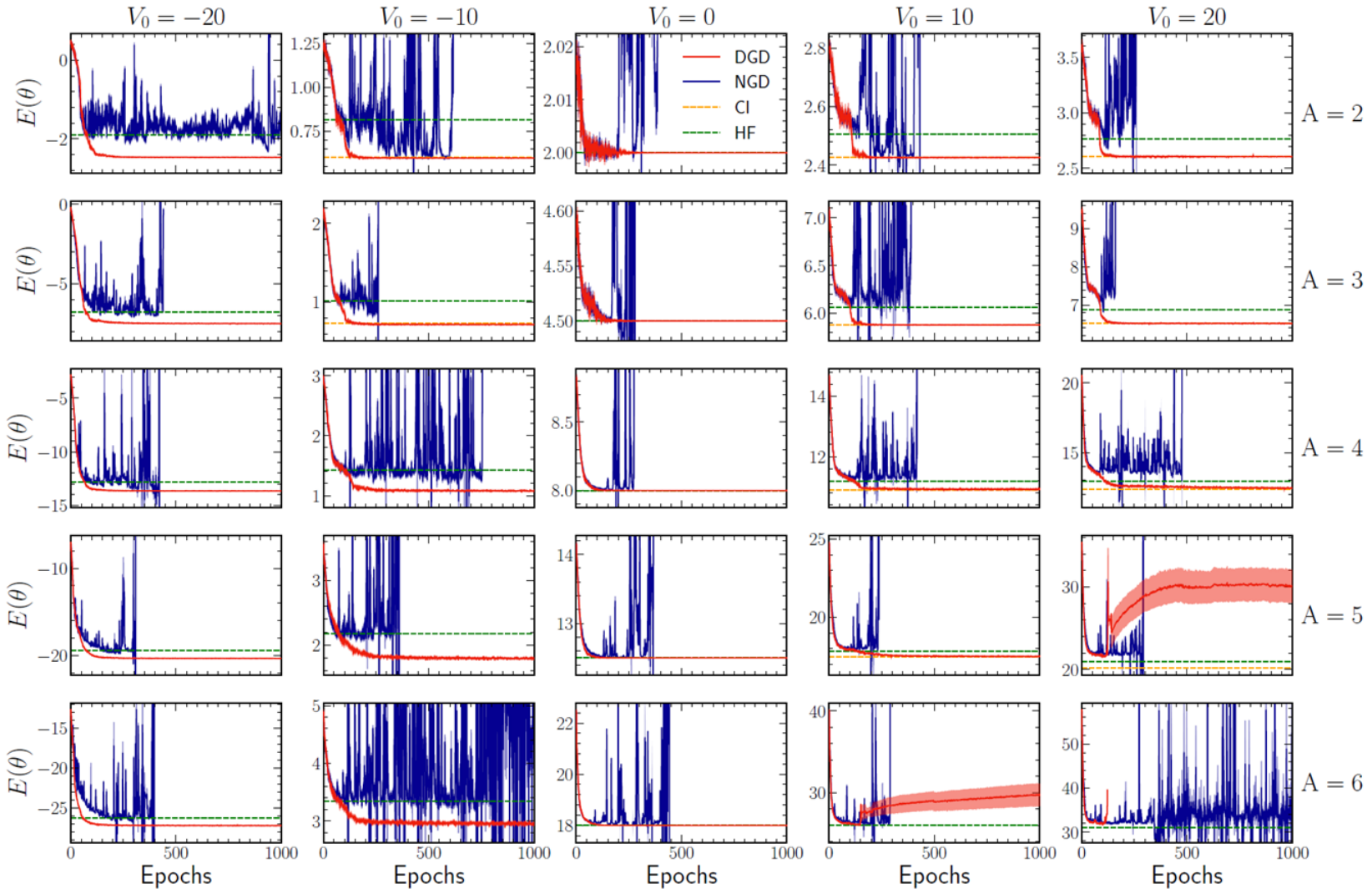
- **Comparison:** **Natural Gradient Descent** vs. **Decisional Gradient Descent**





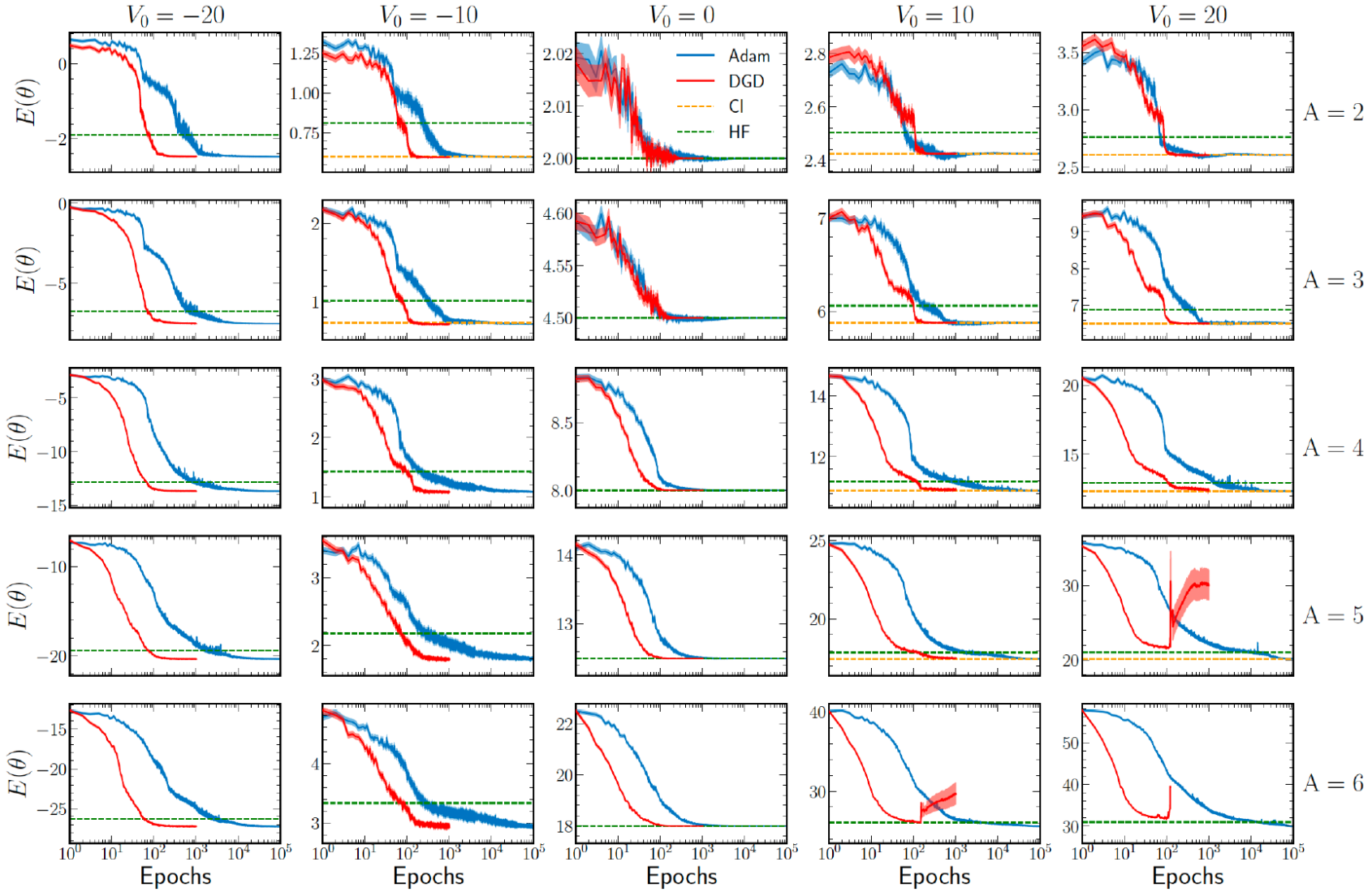
M. Drissi, J.W.T Keeble, J. Rozalén Sarmiento & A. Rios, Phil. Trans. R. Soc. A. **382** (2024)

DGD vs. NGD



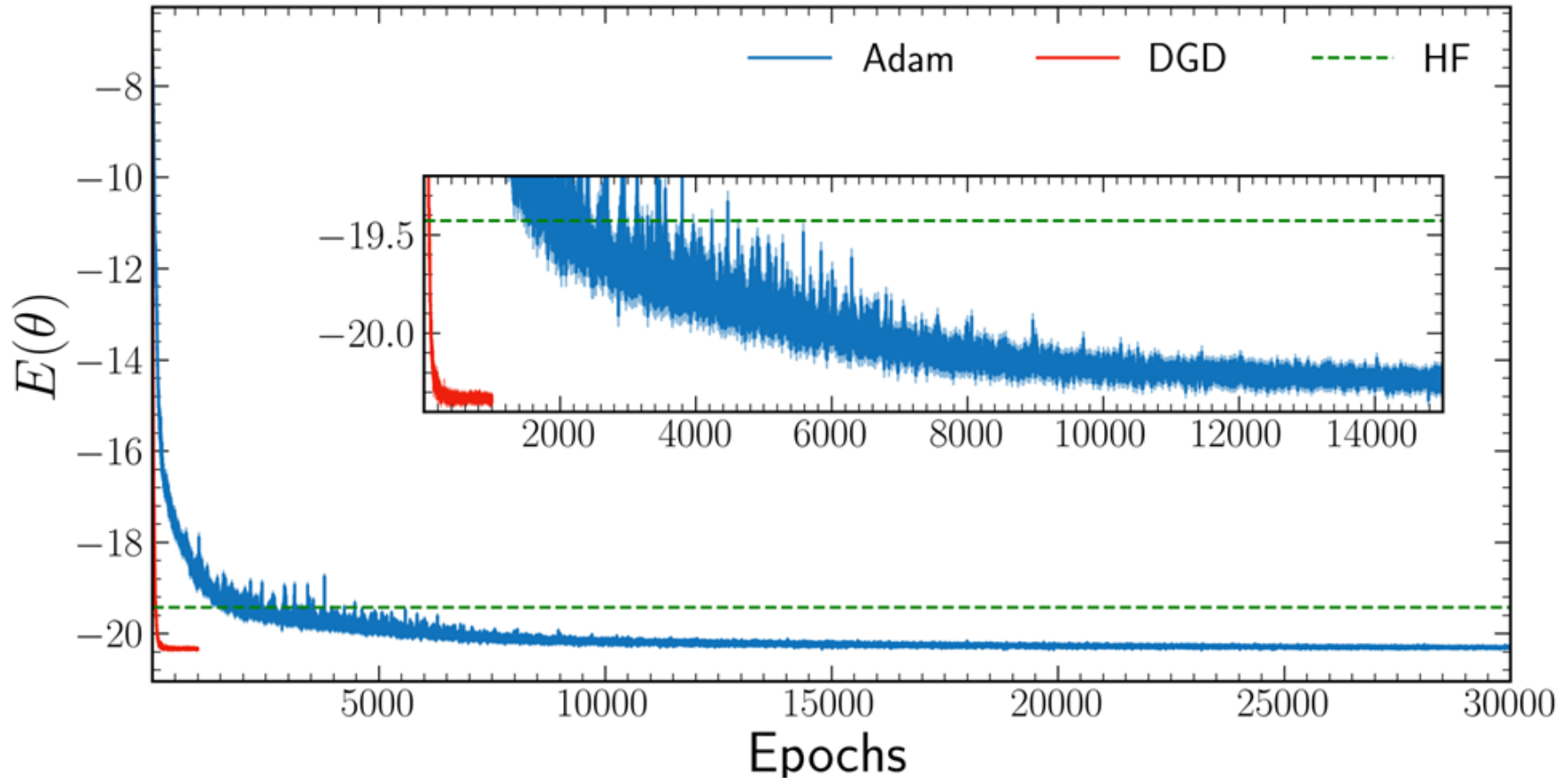
M. Drissi, J.W.T Keeble, J. Rozalén Sarmiento & A. Rios, Phil. Trans. R. Soc. A. **382** (2024)

DGD vs. ADAM



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

DGD vs. ADAM



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

Future Outlook

- Nuclear wave functions are **complex objects** $\psi_\theta(X) \in \mathbb{C}$
- In systems with complex wave functions, we need to include the **phase** in the metric
- Systems with **discrete degrees of freedom** (e.g. condensed matter) → work in progress

Thank you

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

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

javi.rozalen@icc.ub.edu

ICCUB | Institut de Ciències del Cosmos UNIVERSITAT DE BARCELONA | EXCELENCIA MARÍA DE MAEZTU 2020–2024

CPAN Ingenio 2010