Learning Density Functionals from Noisy Quantum Data

Emiel Koridon,[1,2,*] Felix Frohnert,[1,*] Eric Prehn,[1] Evert van Nieuwenburg,[1] Jordi Tura,[1] and Stefano Polla[1,†]

[1] $\langle aQa^L \rangle$ *Applied Quantum Algorithms — Lorentz Insitute for Theoretical physics & Leiden Institute of Advanced Computer Science, Universiteit Leiden, The Netherlands*

[2] *Theoretical Chemistry, Vrije Universiteit Amsterdam, The Netherlands*

“Could NISQ devices be useful training data generators for classical ML models?”

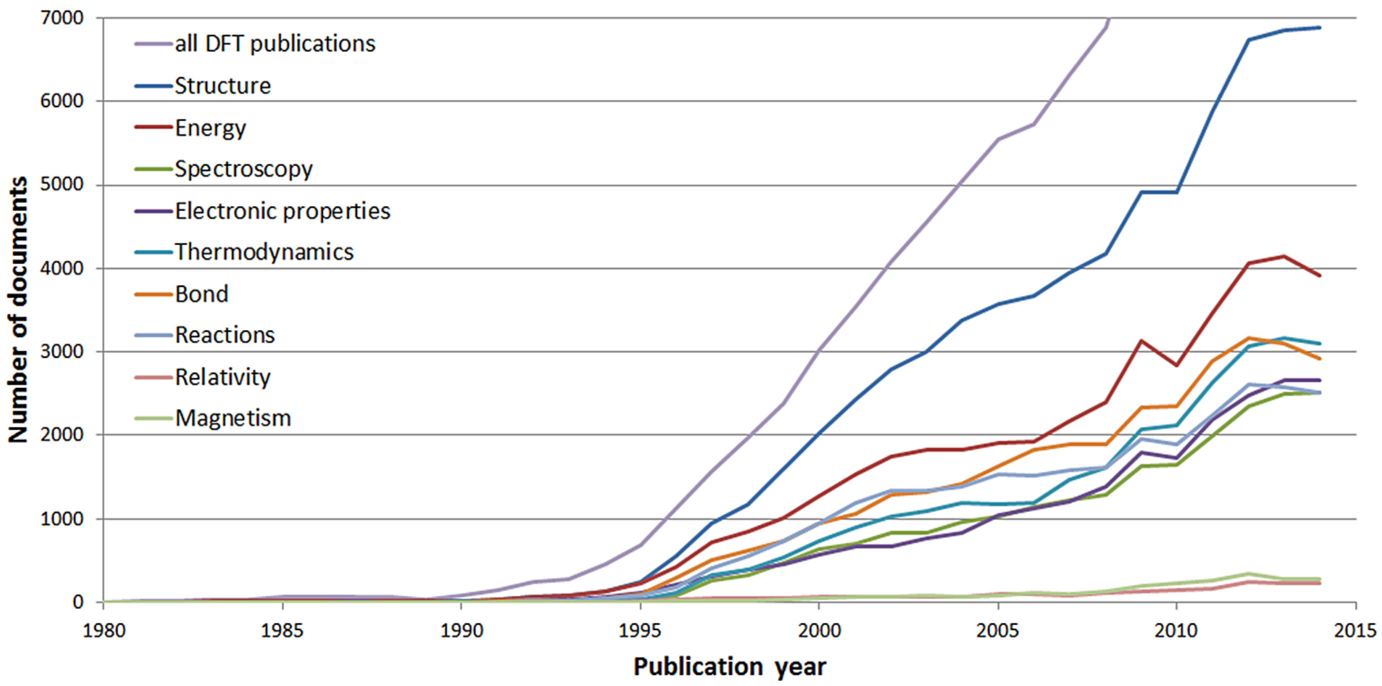
2025 Mach. Learn.: Sci. Technol. 6 025020

Outlook

- Handwavy intro to DFT
- Fermi-Hubbard model
- ML workflow
- Quantum data
- Results: model errors and density optimization
- Conclusion + outlook
- Sneak-peek: auto-encoders

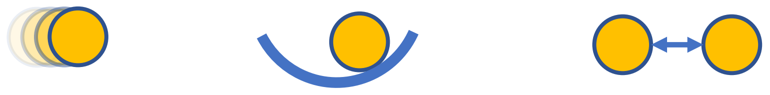
Density functional theory

- Workhorse of computational chemistry
- A good balance of accuracy and computational efficiency
- Can fail for strongly correlated systems



Density functional theory

$$\hat{H} = \overbrace{\hat{T}}^{\text{kinetic}} + \overbrace{\hat{V}}^{\text{potential}} + \overbrace{\hat{U}}^{\text{interaction}}$$



One-body operator

Two-body operator

Diagonal in plane-wave basis (Fourier space)

Diagonal in position basis (real space)

n := the number of orbitals / sites / basis functions used to discretize space

DFT: Hohenberg-Kohn I

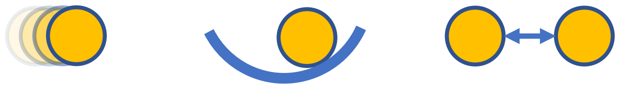
Density

n dimensional

$$\rho_{\mathrm{GS}}$$

Potential

$$\boldsymbol{\mu} \text{ or } v(\mathbf{r})$$



2^n dimensional

$$|\psi_{\mathrm{GS}}\rangle\,,\ E_{\mathrm{GS}} \xleftarrow{\text{diagonalization}} \hat{H}_{\boldsymbol{\mu}} = \hat{T} + \hat{V}_{\boldsymbol{\mu}} + \hat{U}$$

State vector

Full Hamiltonian

n := the number of orbitals / sites / basis functions used to discretize space

DFT: Hohenberg-Kohn I

The ground-state electron density $\rho(\boldsymbol{r})$ uniquely determines the external potential $v(\boldsymbol{r})$, and therefore all properties of the ground state.

	Density		Potential
n dimensional	ρ_{GS}	$\longrightarrow$	$\boldsymbol{\mu}$ or $v(\mathbf{r})$
	$\uparrow$		$\downarrow$
2^n dimensional	$\vert\psi_{\mathrm{GS}}\rangle\, ,\; E_{\mathrm{GS}}$	$\xleftarrow{\text{diagonalization}}$	$\hat{H}_{\boldsymbol{\mu}} = \hat{T} + \hat{V}_{\boldsymbol{\mu}} + \hat{U}$
	State vector		Full Hamiltonian

n := the number of orbitals / sites / basis functions used to discretize space

DFT: Hohenberg-Kohn I

The ground-state electron density $\rho(\boldsymbol{r})$ uniquely determines the external potential $v(\boldsymbol{r})$, and therefore all properties of the ground state.

n dimensional

Density

ρ_{GS}

Potential

$\boldsymbol{\mu}$ or $v(\mathbf{r})$

DFT

2^n dimensional

$|\psi_{\mathrm{GS}}\rangle\ ,\ E_{\mathrm{GS}}$

diagonalization

$$\hat{H}_{\boldsymbol{\mu}} = \hat{T} + \hat{V}_{\boldsymbol{\mu}} + \hat{U}$$

State vector

Full Hamiltonian

n := the number of orbitals / sites / basis functions used to discretize space

DFT: Hohenberg-Kohn I & II

$$\hat{H}_{\boldsymbol{\mu}} = \overbrace{\hat{T}}^{\text{kinetic}} + \overbrace{\hat{V}_{\boldsymbol{\mu}}}^{\text{potential}} + \overbrace{\hat{U}}^{\text{interaction}}$$

Wavefunction theory **DFT**

$$|\psi_{\mathrm{GS}}\rangle \in \mathbb{C}^{2^n} \longrightarrow \rho_{\mathrm{GS}} \in \mathbb{R}^n$$

$$\begin{aligned} E_{\mathrm{GS}} &= \langle \psi_{\mathrm{GS}} | \, H \, | \psi_{\mathrm{GS}} \rangle \\ &= \min_{|\psi\rangle} \langle \psi | H | \psi \rangle \end{aligned}$$

n := the number of orbitals / sites / basis functions used to discretize space

DFT: Hohenberg-Kohn I & II

$$\hat{H}_{\boldsymbol{\mu}} = \overbrace{\hat{T}}^{\text{kinetic}} + \overbrace{\hat{V}_{\boldsymbol{\mu}}}^{\text{potential}} + \overbrace{\hat{U}}^{\text{interaction}}$$

Wavefunction theory **DFT**

$$|\psi_{\mathrm{GS}}\rangle \in \mathbb{C}^{2^n} \longrightarrow \rho_{\mathrm{GS}} \in \mathbb{R}^n$$

$$\mathcal{F}\left[\boldsymbol{\rho}_{\mathrm{GS}}\right] := \langle\psi_{\mathrm{GS}}|\, T + U \,|\psi_{\mathrm{GS}}\rangle$$

$$\begin{aligned} E_{\mathrm{GS}} &= \langle\psi_{\mathrm{GS}}|\, H \,|\psi_{\mathrm{GS}}\rangle \\ &= \min_{|\psi\rangle} \langle\psi|H|\psi\rangle \end{aligned}$$

n := the number of orbitals / sites / basis functions used to discretize space

DFT: Hohenberg-Kohn I & II

$$\hat{H}_{\boldsymbol{\mu}} = \overbrace{\hat{T}}^{\text{kinetic}} + \overbrace{\hat{V}_{\boldsymbol{\mu}}}^{\text{potential}} + \overbrace{\hat{U}}^{\text{interaction}}$$

Wavefunction theory **DFT**

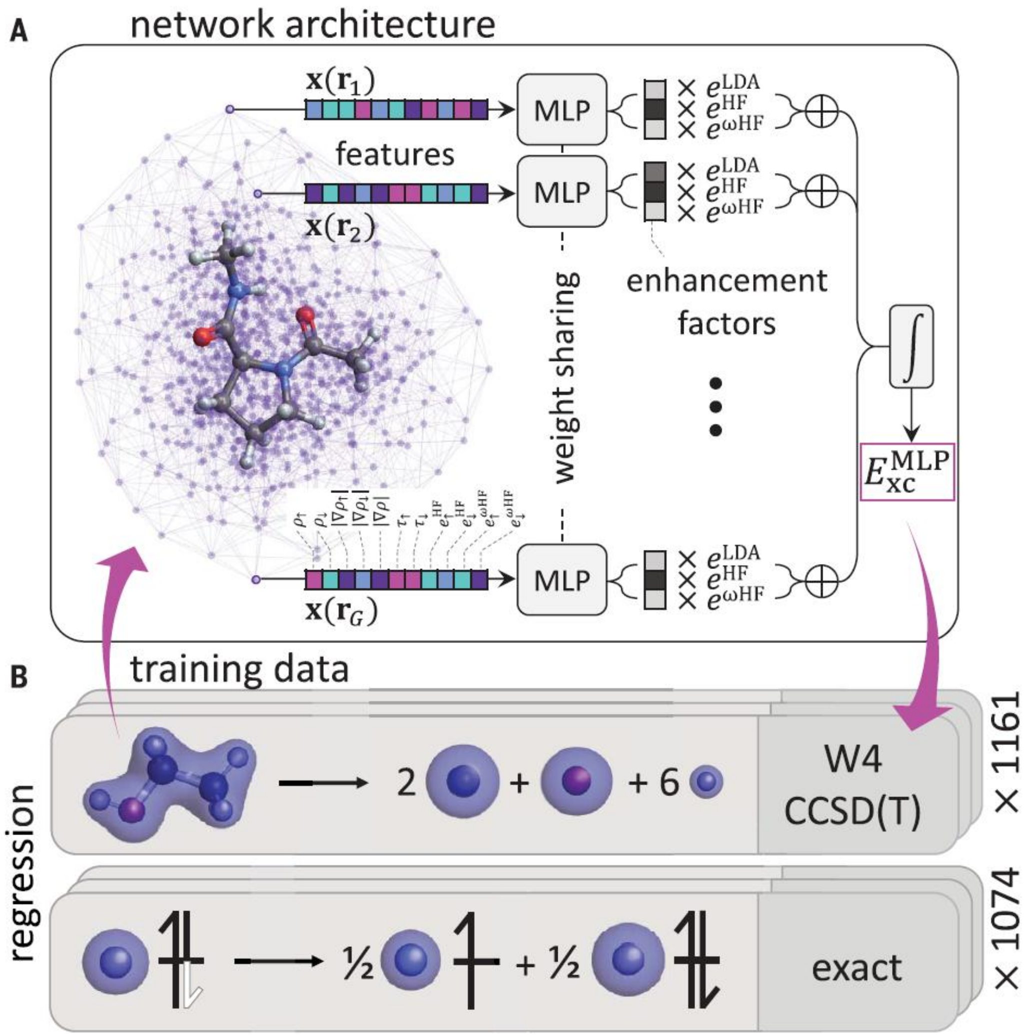
$$|\psi_{\text{GS}}\rangle \in \mathbb{C}^{2^n} \longrightarrow \rho_{\text{GS}} \in \mathbb{R}^n$$

$$\mathcal{F}\left[\boldsymbol{\rho}_{\text{GS}}\right] := \langle\psi_{\text{GS}}|\, T + U\, |\psi_{\text{GS}}\rangle$$

$$\begin{aligned} E_{\text{GS}} &= \langle\psi_{\text{GS}}|\, H\, |\psi_{\text{GS}}\rangle \\ &= \min_{|\psi\rangle} \langle\psi|H|\psi\rangle \end{aligned}$$

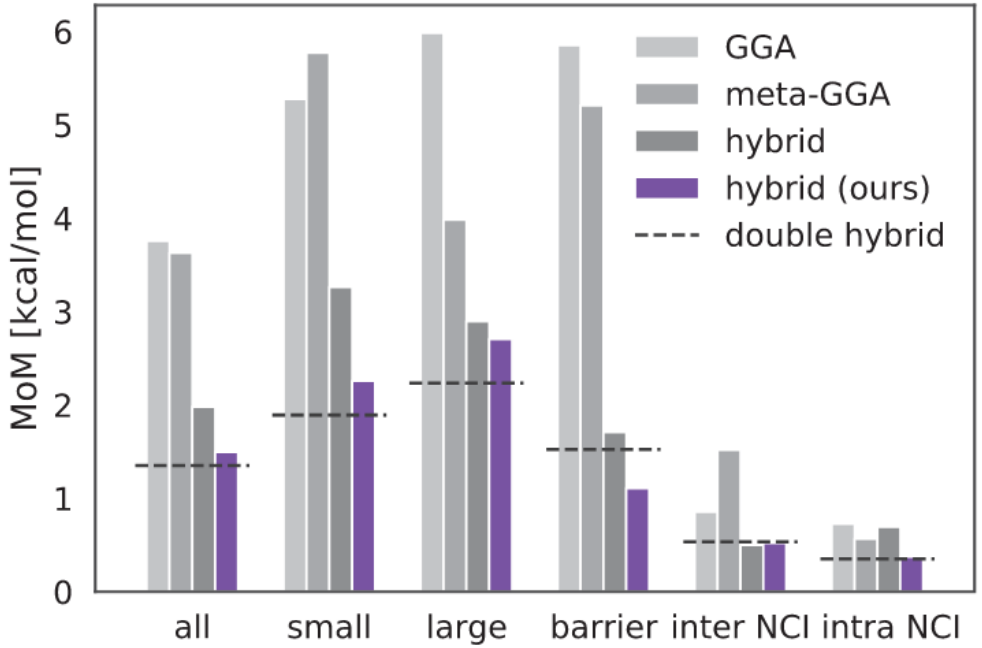
$$\begin{aligned} E_{\text{GS}} &= \mathcal{F}\left[\boldsymbol{\rho}_{\text{GS}}\right] + \mathrm{Tr}[\rho_{\text{GS}}\hat{V}] \\ &= \min_{\rho} \mathcal{F}[\rho] + \mathrm{Tr}[\rho\hat{V}] \end{aligned}$$

Machine learning a functional



DeepMind model

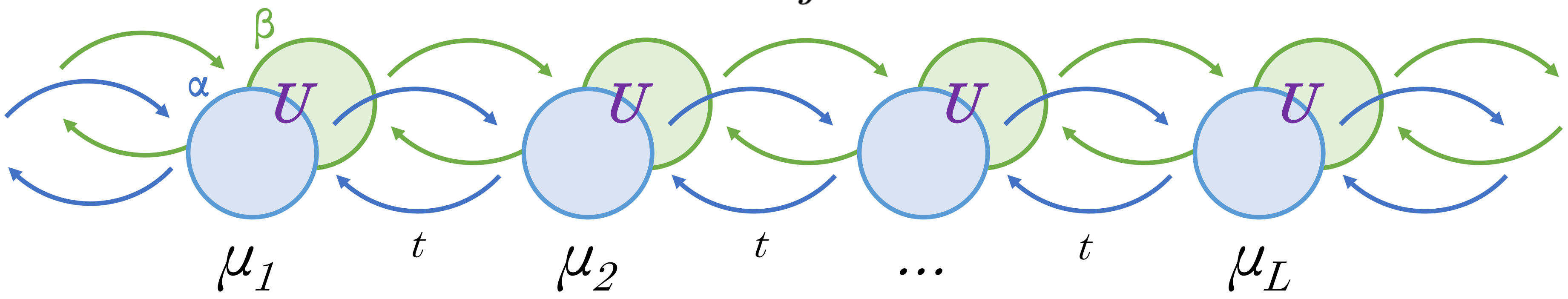
[J. Kirkpatrick *et al.*, *Science* **374**, dec 2021]



$$\hat{n}_j = \hat{n}_{j\uparrow} + \hat{n}_{j\downarrow} = \langle c^\dagger_{j\uparrow} c_{j\uparrow} \rangle + \langle c^\dagger_{j\downarrow} c_{j\downarrow} \rangle$$

Fermi-Hubbard model

$$H_{\boldsymbol{\mu}} = \hat{T} + \hat{U} + \sum_j \mu_j \cdot \hat{n}_j$$



$$\hat{T} = -t \sum_{i\sigma} \left(\hat{c}^\dagger_{i+1,\sigma} \hat{c}_{i\sigma} + \hat{c}^\dagger_{i\sigma} \hat{c}_{i+1,\sigma} \right)$$

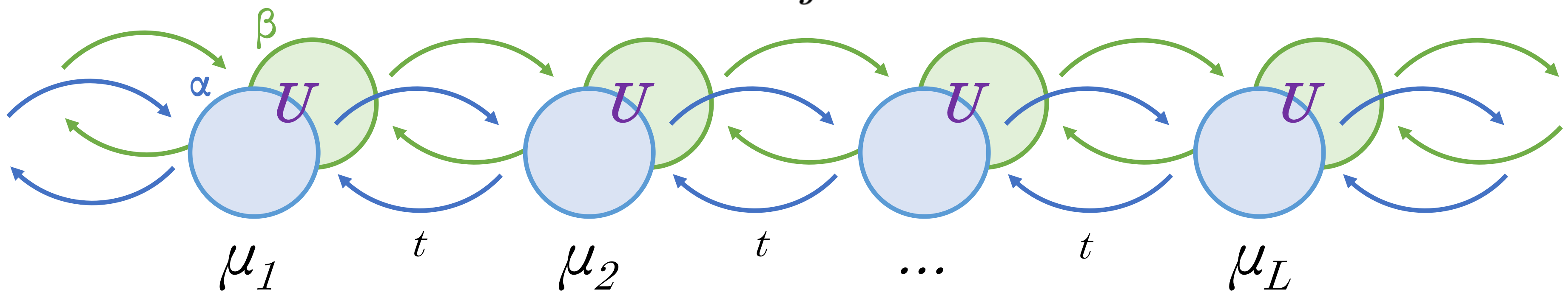
$$\hat{U} = U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}$$

$$\hat{n}_j = \hat{n}_{j\uparrow} + \hat{n}_{j\downarrow} = \langle c^\dagger_{j\uparrow} c_{j\uparrow} \rangle + \langle c^\dagger_{j\downarrow} c_{j\downarrow} \rangle$$

Fermi-Hubbard model

$$\rho_j = \langle \hat{n}_j \rangle$$

$$H_{\boldsymbol{\mu}} = \hat{T} + \hat{U} + \sum_j \mu_j \cdot \hat{n}_j$$



$$\hat{T} = -t \sum_{i\sigma} \left(\hat{c}^\dagger_{i+1,\sigma} \hat{c}_{i\sigma} + \hat{c}^\dagger_{i\sigma} \hat{c}_{i+1,\sigma} \right)$$

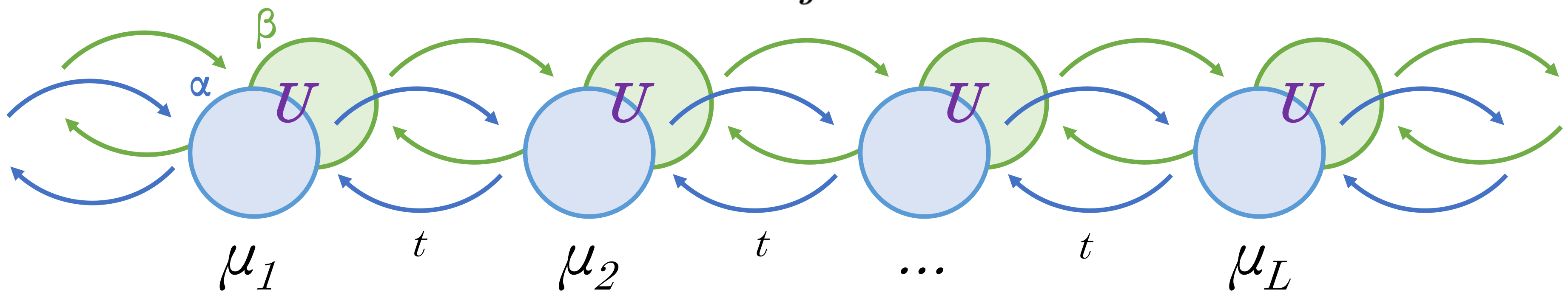
$$\hat{U} = U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}$$

$$\hat{n}_j = \hat{n}_{j\uparrow} + \hat{n}_{j\downarrow} = \langle c^\dagger_{j\uparrow} c_{j\uparrow} \rangle + \langle c^\dagger_{j\downarrow} c_{j\downarrow} \rangle$$

Fermi-Hubbard model

$$H_{\boldsymbol{\mu}} = \hat{T} + \hat{U} + \sum_j \mu_j \cdot \hat{n}_j$$

$$\rho_j = \langle \hat{n}_j \rangle$$



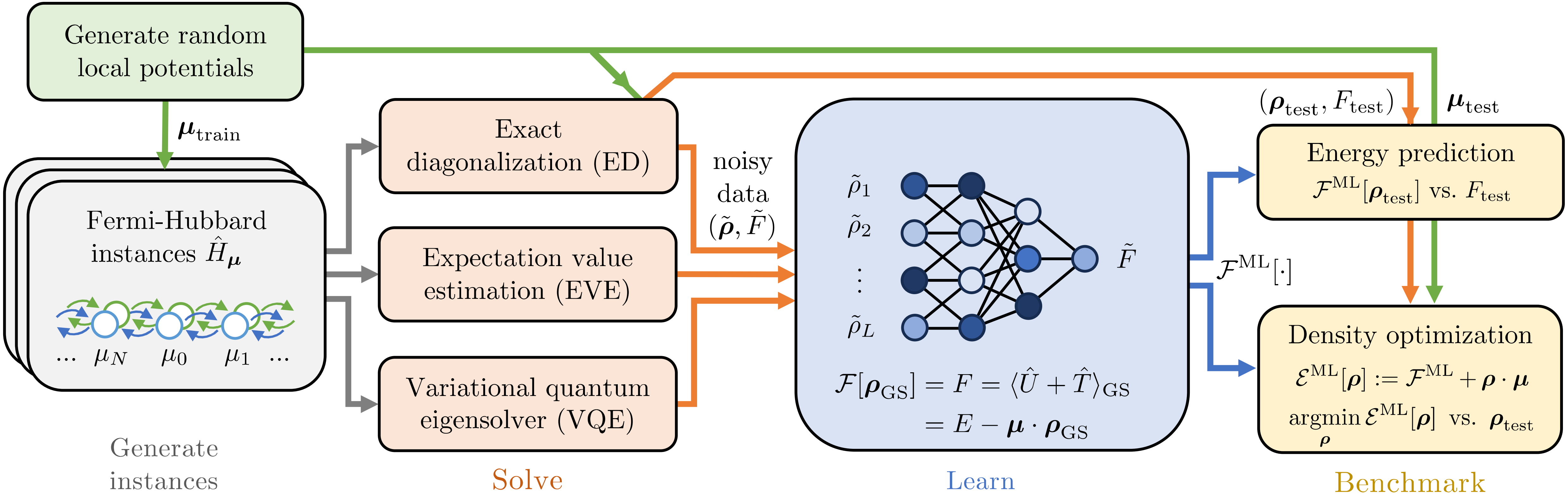
$$\hat{T} = -t \sum_{i\sigma} \left(\hat{c}^\dagger_{i+1,\sigma} \hat{c}_{i\sigma} + \hat{c}^\dagger_{i\sigma} \hat{c}_{i+1,\sigma} \right)$$

$$\hat{U} = U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}$$

$$E_{\mathrm{GS}} = \mathcal{F}\left[\boldsymbol{\rho}_{\mathrm{GS}}\right] + \boldsymbol{\rho} \cdot \boldsymbol{\mu},$$

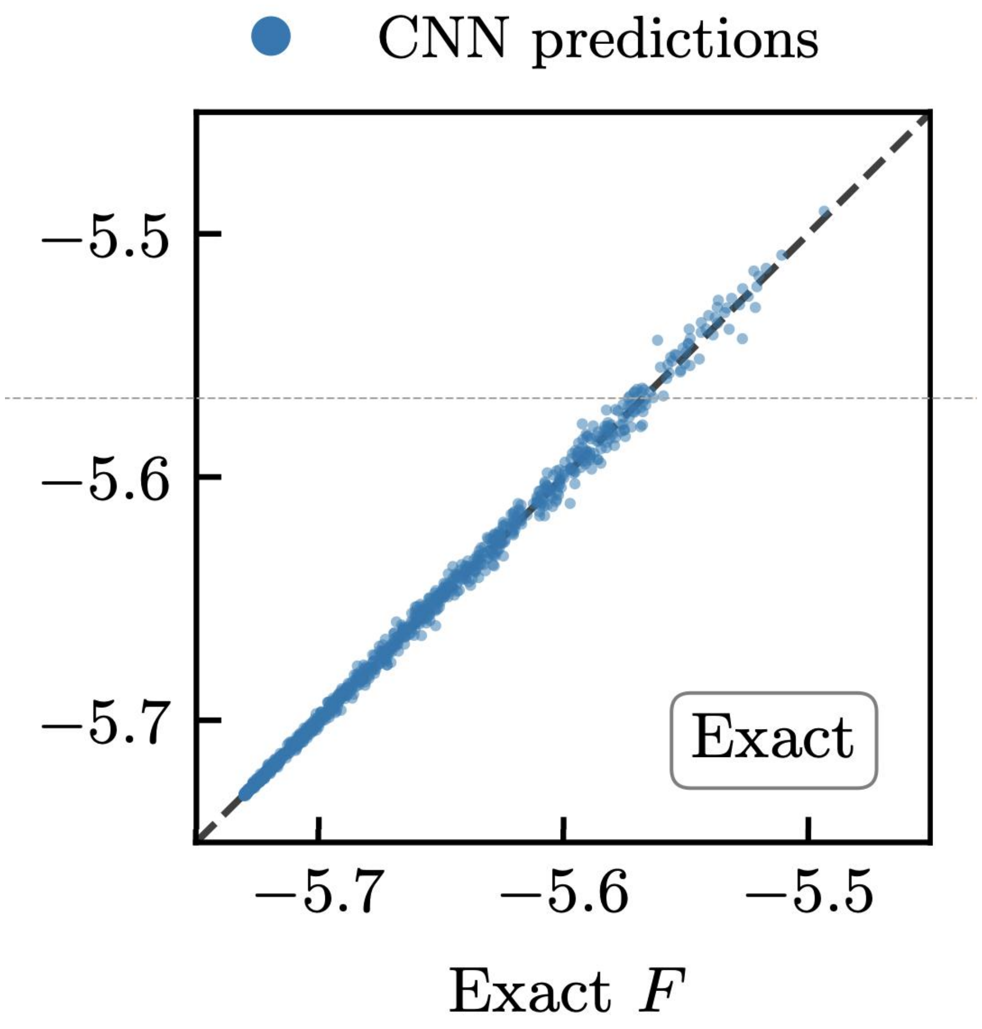
$$\mathcal{F} : \boldsymbol{\rho}_{\mathrm{GS}} \to \overbrace{\langle \psi_{\mathrm{GS}} | T + U | \psi_{\mathrm{GS}} \rangle}^{:=F}$$

The workflow



$L = 8$

Benchmark: exact diagonalization

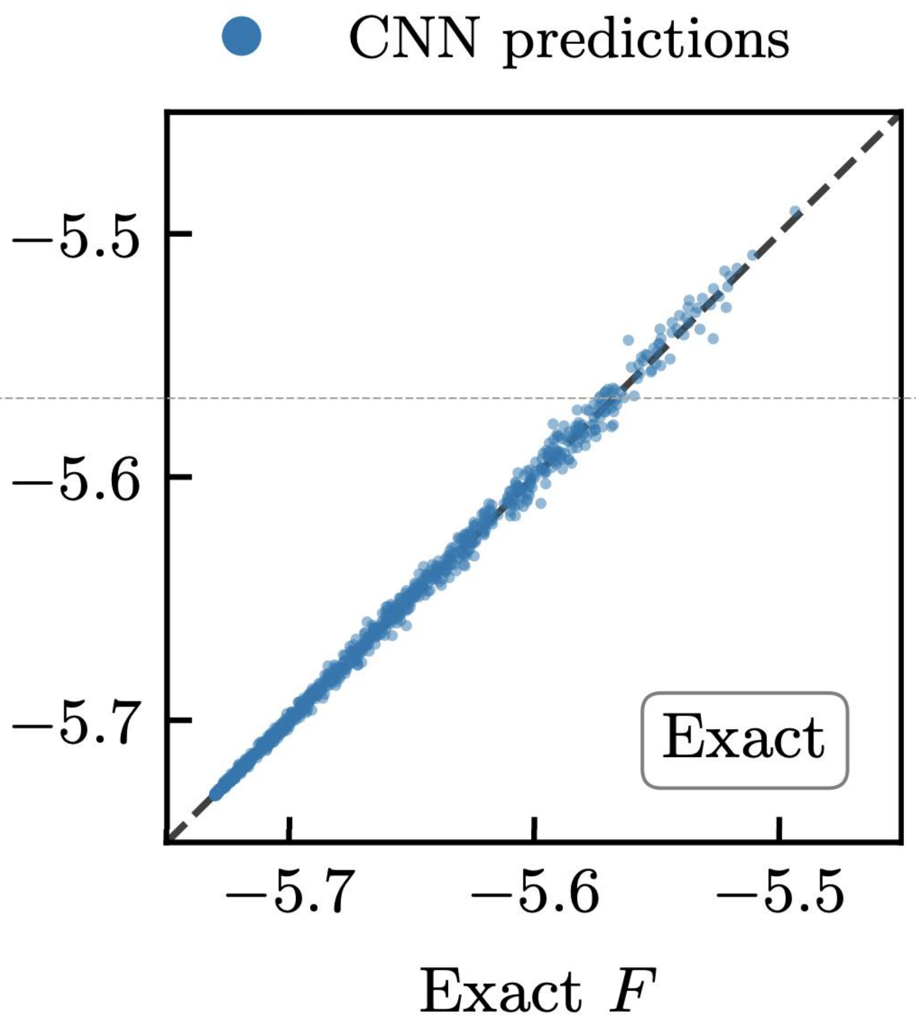


Features	Label
ρ_{GS} →	$F = \langle \hat{T} + \hat{U} \rangle_{\mathrm{GS}}$

- ML model: CNN *[Nelson et al. PRB 99 (2019)]*

$L = 8$

Benchmark: exact diagonalization

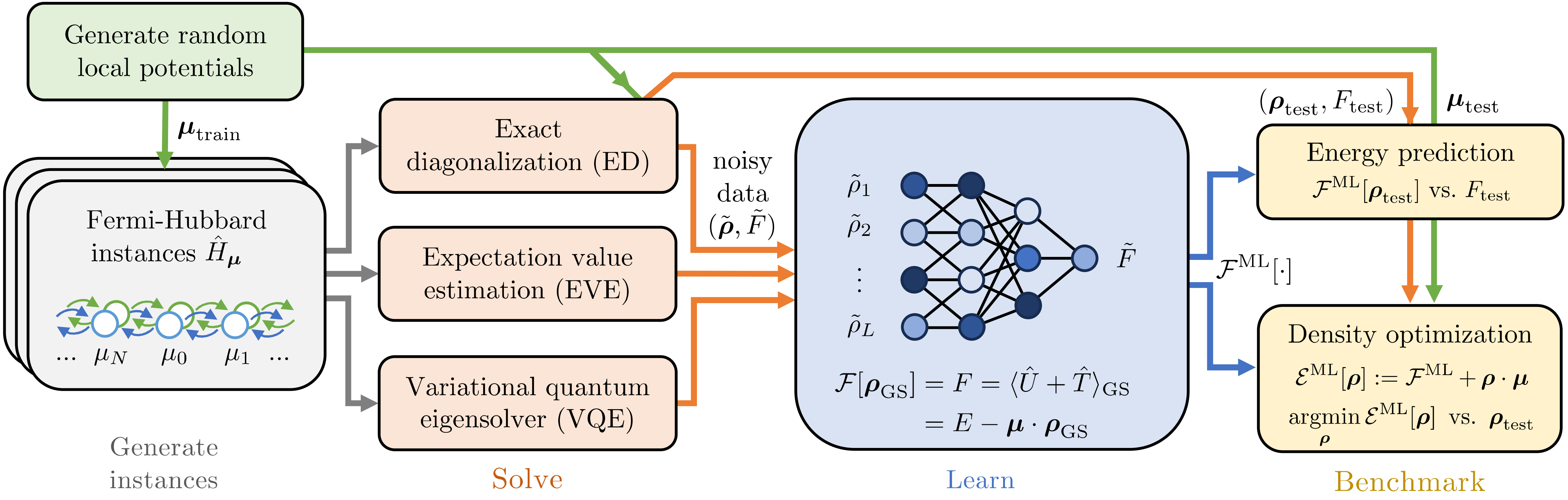


Features | Label

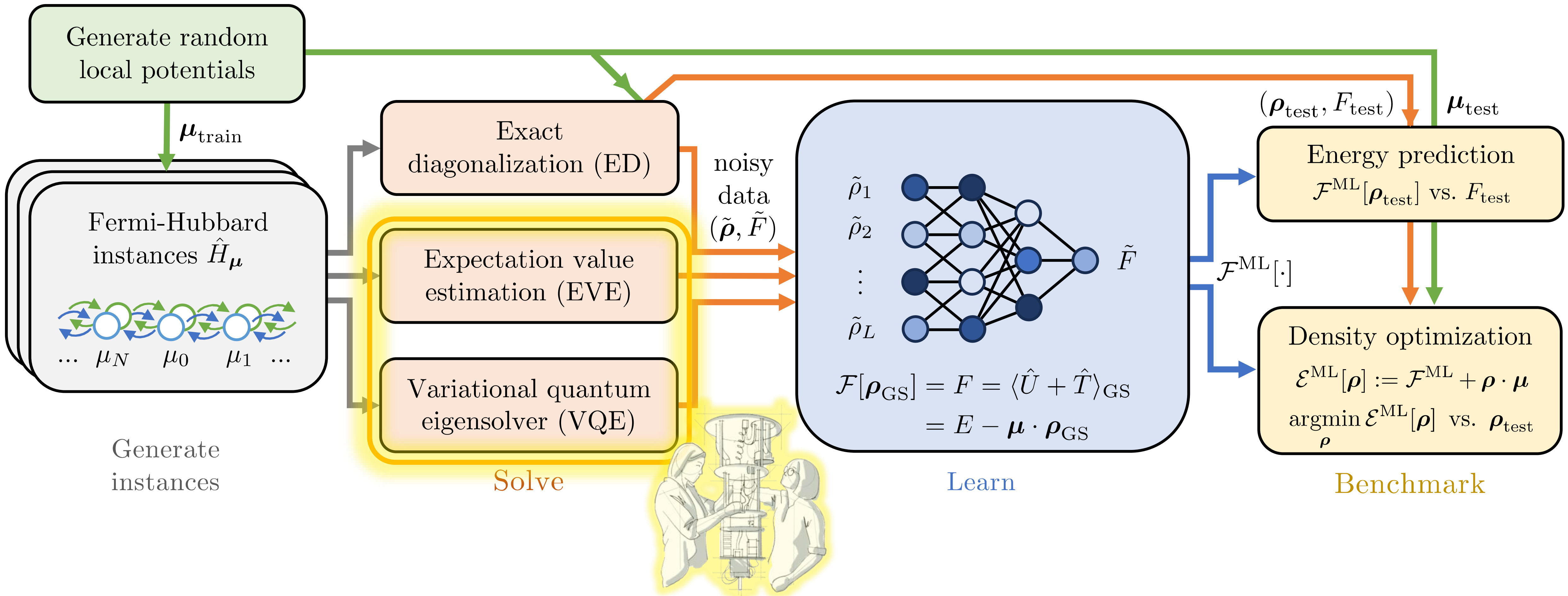
$\rho_{\mathrm{GS}} \longrightarrow F = \langle \hat{T} + \hat{U} \rangle_{\mathrm{GS}}$

- ML model: CNN *[Nelson et al. PRB 99 (2019)]*
- 5-fold x-val, batch normalization, early stopping
- 1000 training points
- Data augmentation using translational + mirror symmetry (1'000 → 16'000 datapoints)

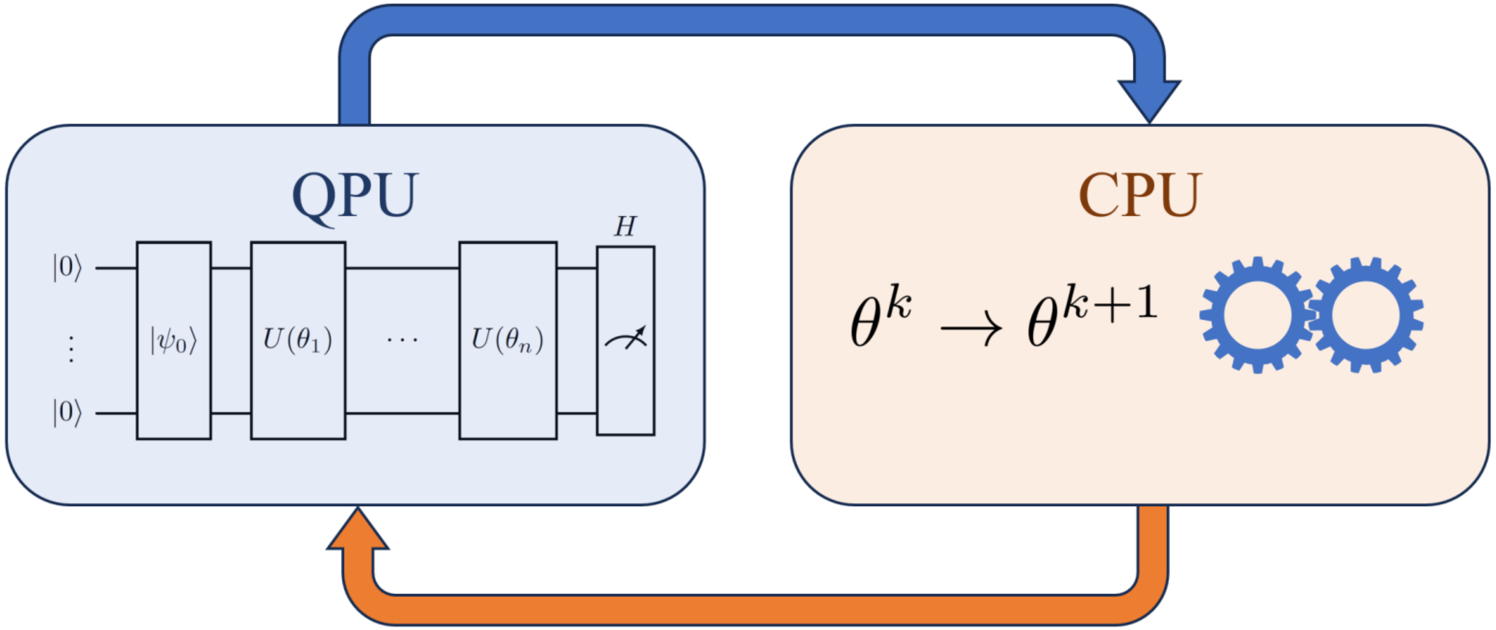
The workflow



The workflow



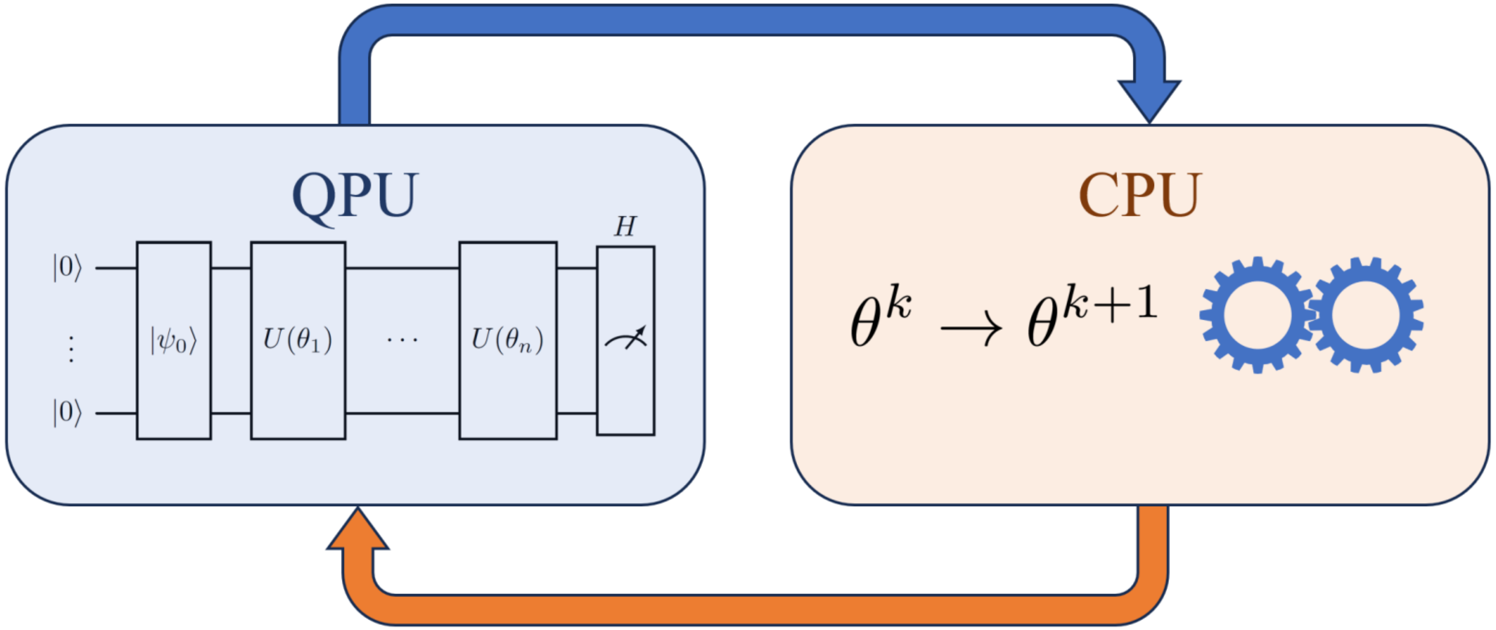
Near-term variational algorithms



Near-term variational algorithms

Error from:

- Circuit error

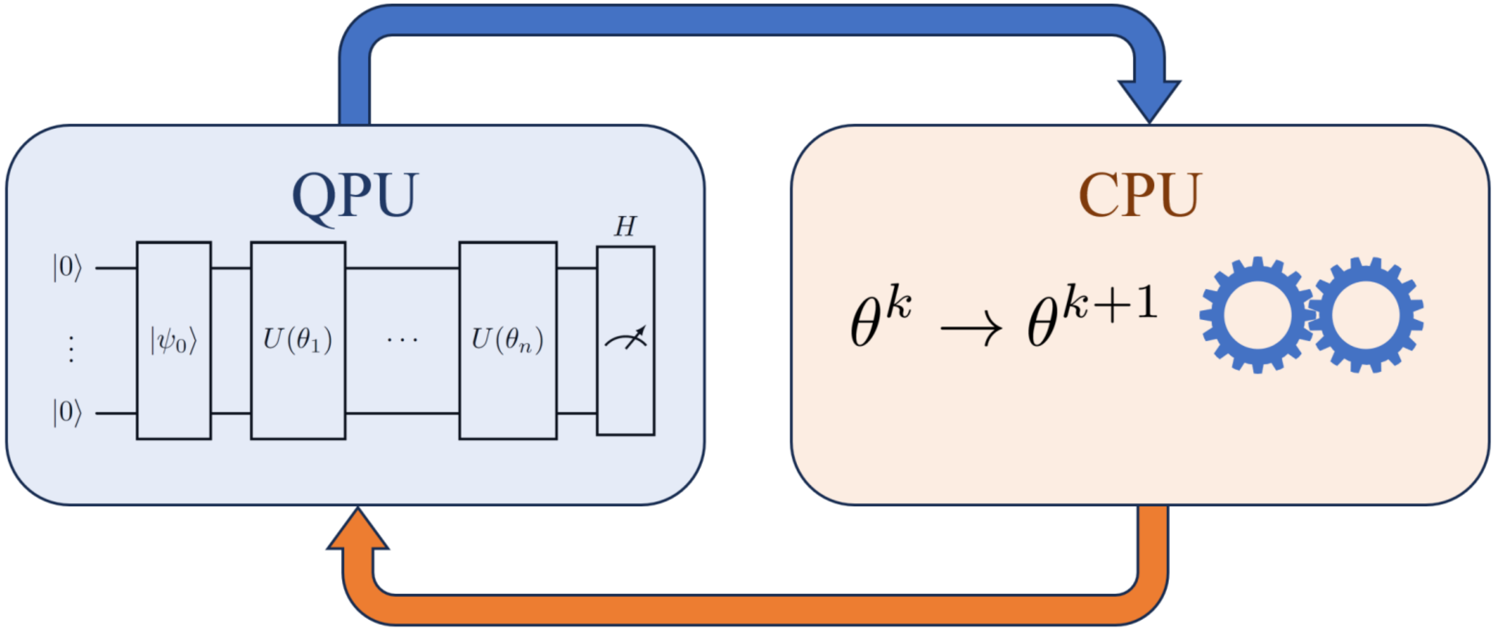


Near-term variational algorithms

Error from:

- Circuit error

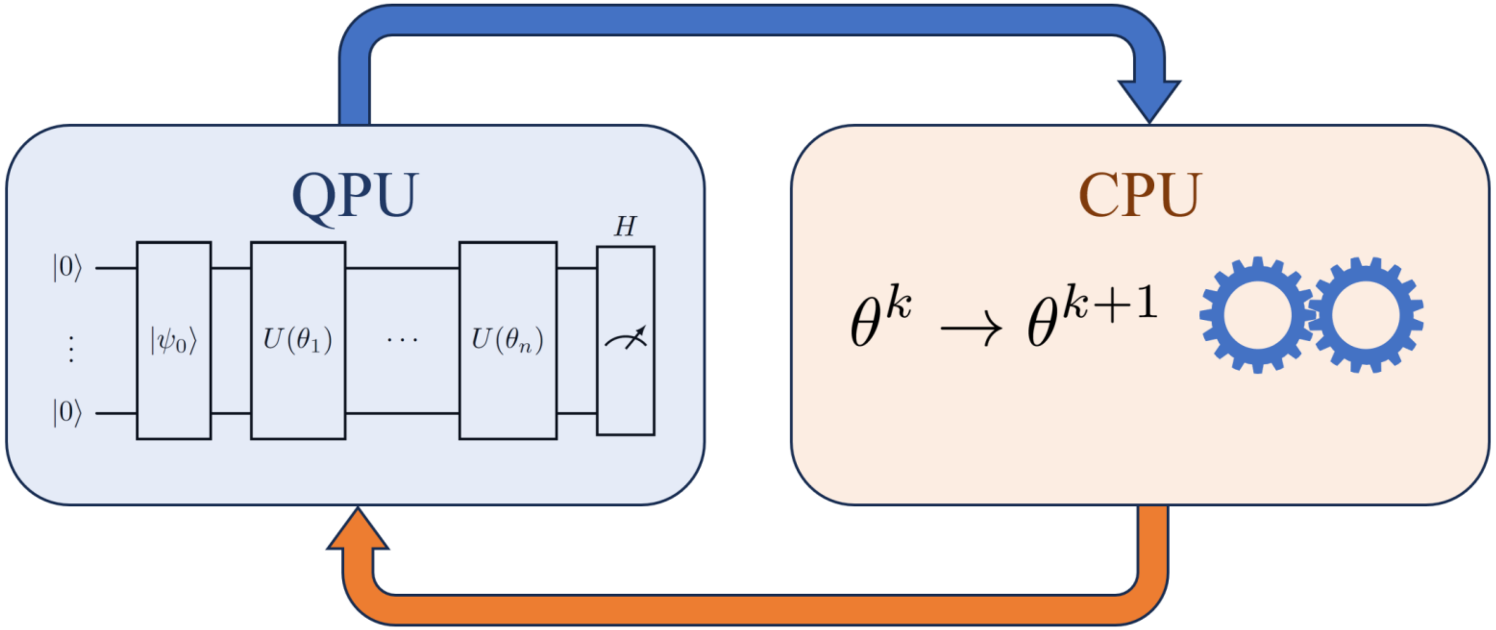
- Sampling error



Near-term variational algorithms

Error from:

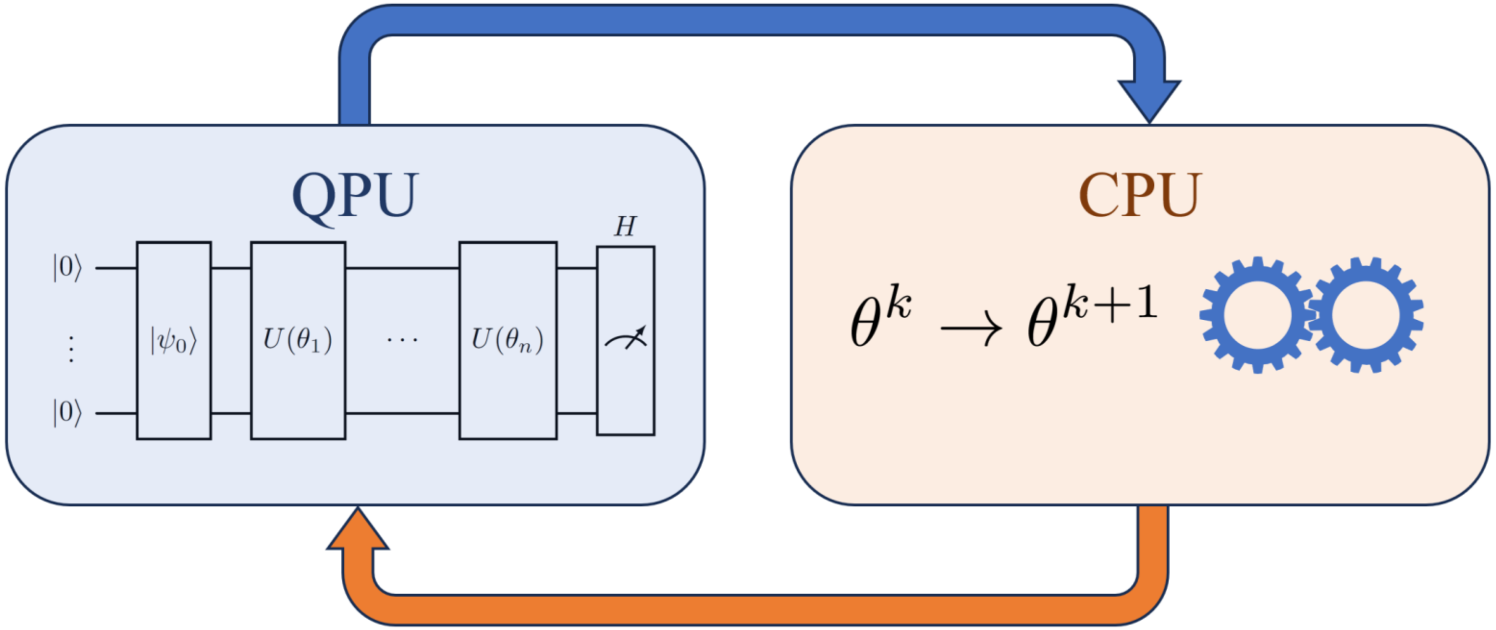
- Circuit error
- Sampling error
- Algorithmic errors
  - Ansatz expressibility
  - Optimization error (local minima)



Near-term variational algorithms

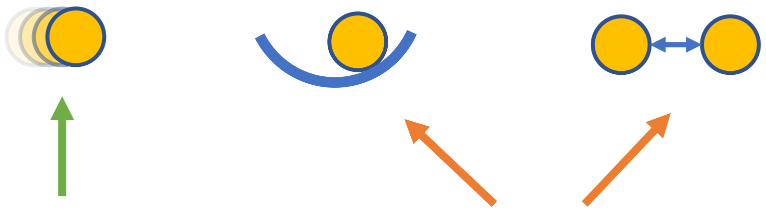
Error from:

- Circuit error
- Sampling error
- Algorithmic errors
  - Ansatz expressibility
  - Optimization error (local minima)



Sampling noise: Expectation Value Estimation (EVE)

$$\hat{H} = \overbrace{\hat{T}}^{\text{kinetic}} + \overbrace{\hat{V}}^{\text{potential}} + \overbrace{\hat{U}}^{\text{interaction}}$$

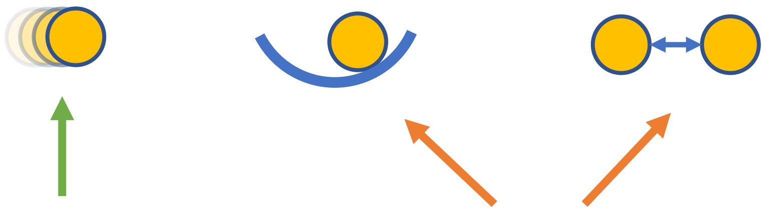


Diagonal in
plane-wave basis
(Fourier space)

Diagonal in
position basis
(real space)

Sampling noise: Expectation Value Estimation (EVE)

$$\hat{H} = \overbrace{\hat{T}}^{\text{kinetic}} + \overbrace{\hat{V}}^{\text{potential}} + \overbrace{\hat{U}}^{\text{interaction}}$$



Diagonal in plane-wave basis (Fourier space)

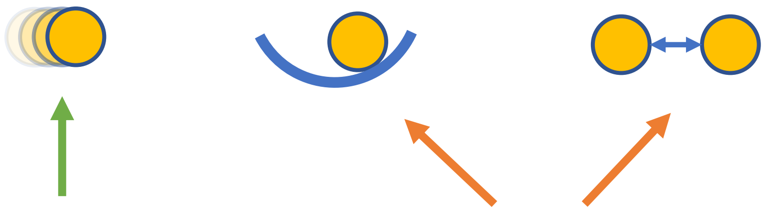
Diagonal in position basis (real space)

$$\langle \hat{U} \rangle = U \sum_i \langle \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \rangle$$

$$\langle \hat{T} \rangle = -2t \sum_k \sum_{\sigma=\uparrow,\downarrow} \cos\left(\frac{2\pi k}{L}\right) \langle \hat{n}_{k,\sigma} \rangle$$

Sampling noise: Expectation Value Estimation (EVE)

$$\hat{H} = \overbrace{\hat{T}}^{\text{kinetic}} + \overbrace{\hat{V}}^{\text{potential}} + \overbrace{\hat{U}}^{\text{interaction}}$$



Diagonal in
plane-wave basis
(Fourier space)

Diagonal in
position basis
(real space)

$M :=$ number of samples
in real space
and Fourier space

$$\langle \hat{U} \rangle = U \sum_i \langle \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \rangle$$

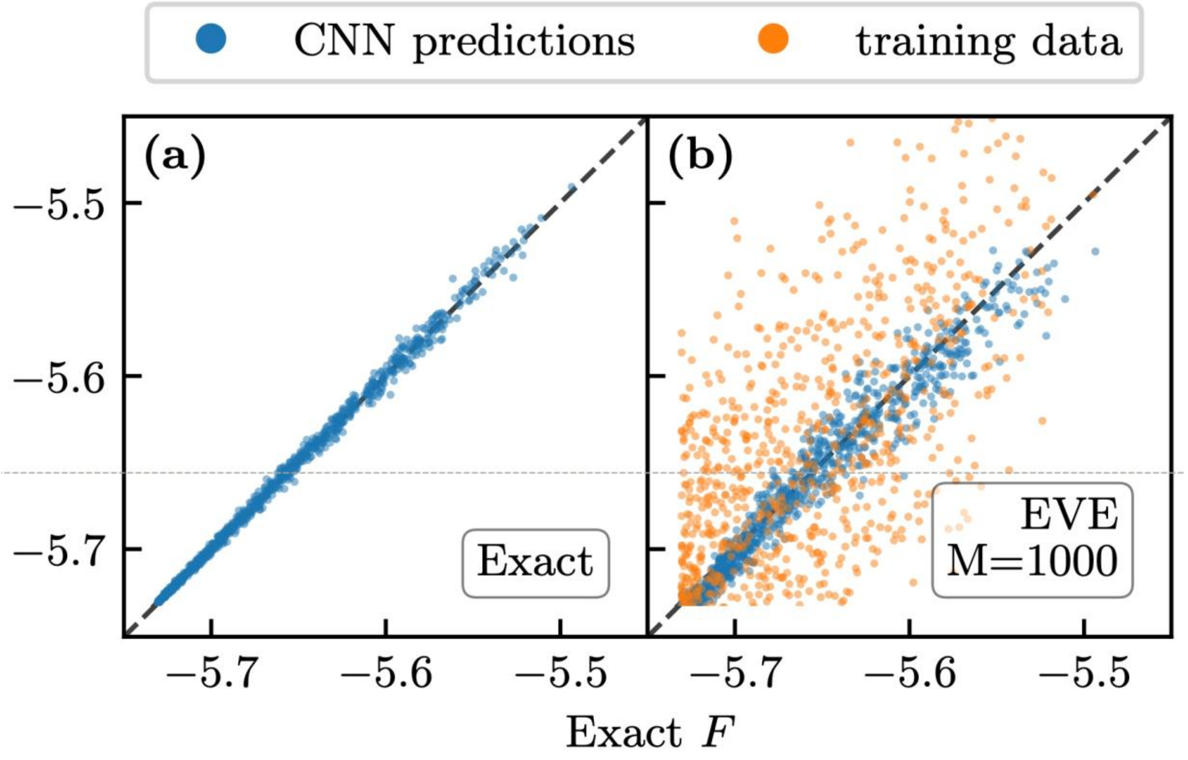
$$\langle \hat{T} \rangle = -2t \sum_k \sum_{\sigma=\uparrow,\downarrow} \cos\left(\frac{2\pi k}{L}\right) \langle \hat{n}_{k,\sigma} \rangle$$

$$\tilde{\rho}_j = \widetilde{\langle \hat{n}_j \rangle} \qquad \tilde{E} = \widetilde{\langle \hat{T} \rangle} + \widetilde{\langle \hat{U} \rangle}$$

$$\epsilon = \mathcal{O}\left(\frac{1}{\sqrt{M}}\right)$$

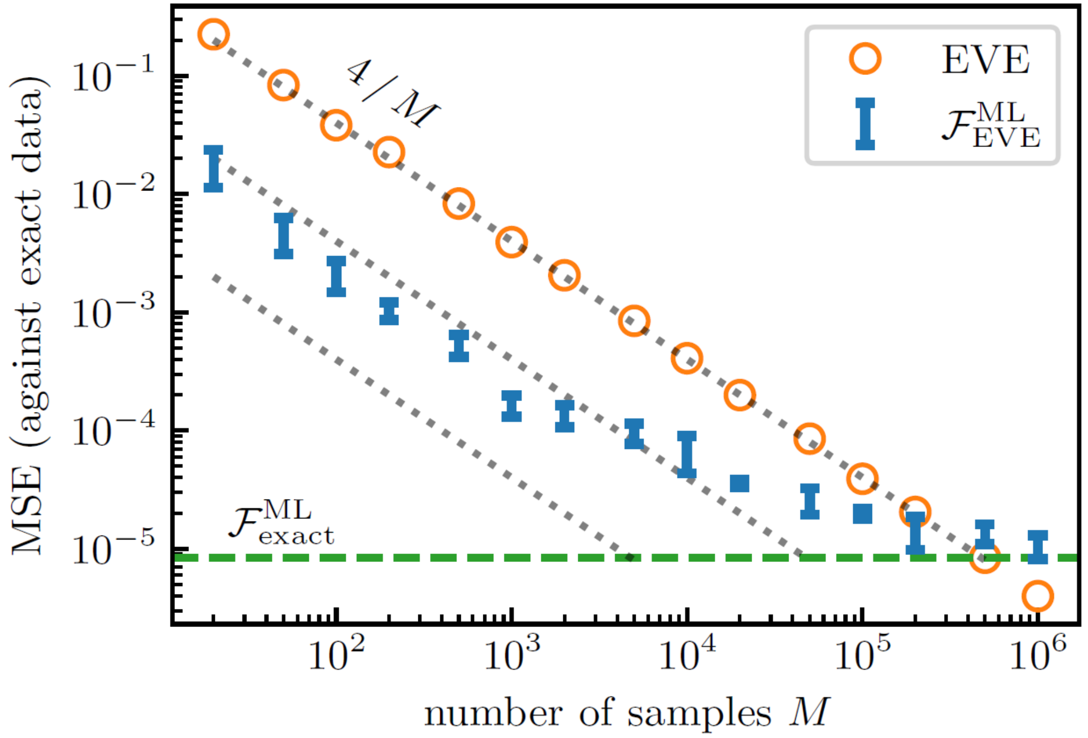
$L = 8$

Training data + results: EVE

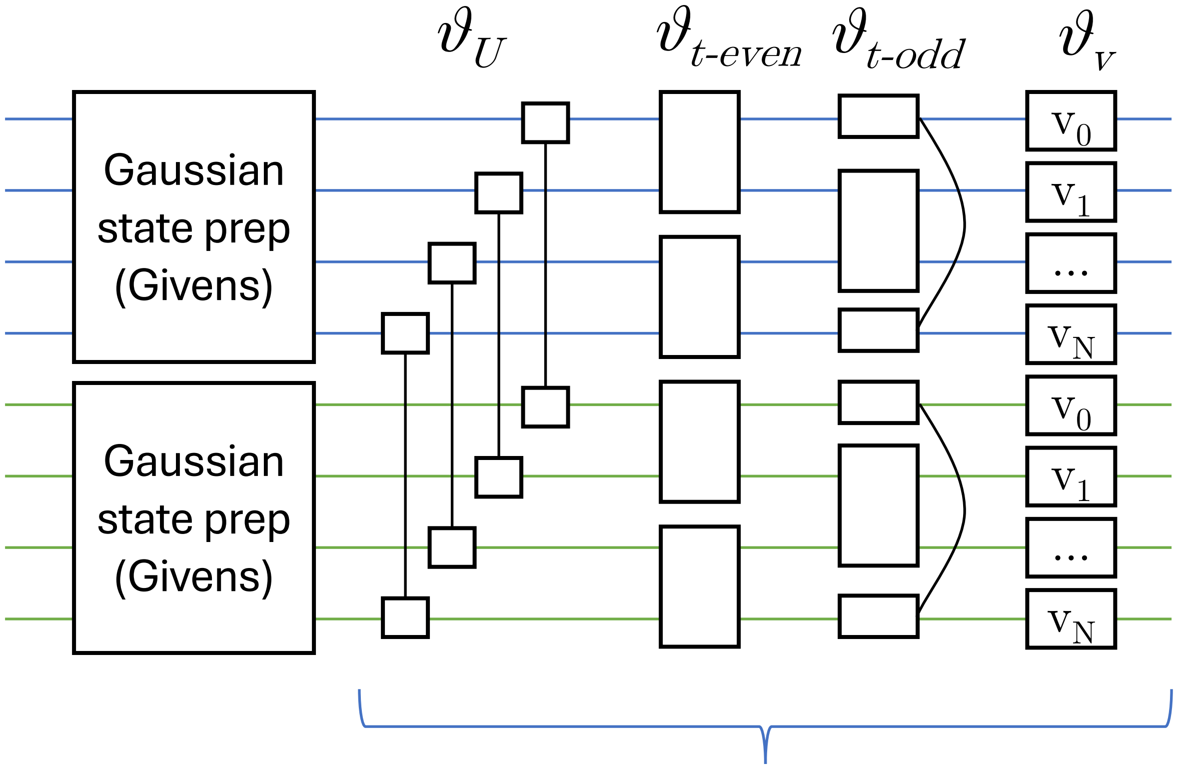


$M :=$ number of samples
in real space
and Fourier space

Expectation value estimation
(Unbiased sampling noise)



VQE: Variational Hamiltonian Ansatz



Repeat *depth* times

nature communications

Article https://doi.org/10.1038/s41467-022-33335-4

Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer

Received: 14 March 2022

Accepted: 14 September 2022

Published online: 11 October 2022

Stasja Stanisic[1], Jan Lukas Bosse[1,2], Filippo Maria Gambetta[1], Raul A. Santos[3], Wojciech Mruczkiewicz[4], Thomas E. O'Brien[4], Eric Ostby[4] & Ashley Montanaro[1,2]

VQE: Quantum-number preserving fabrics

New J. Phys. **23** (2021) 113010 https://doi.org/10.1088/1367-2630/ac2cb3

New Journal of Physics

The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG

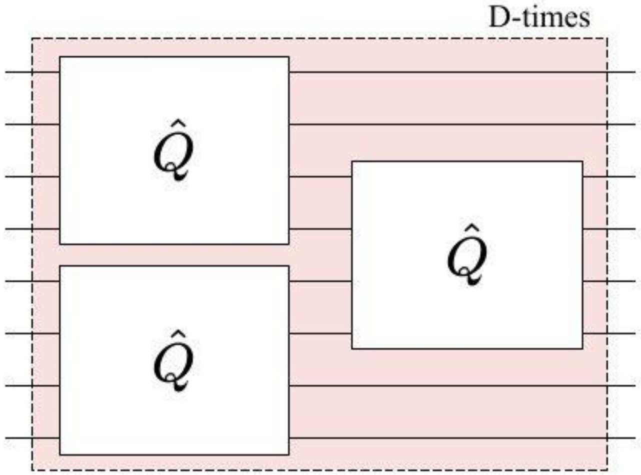
IOP Institute of Physics

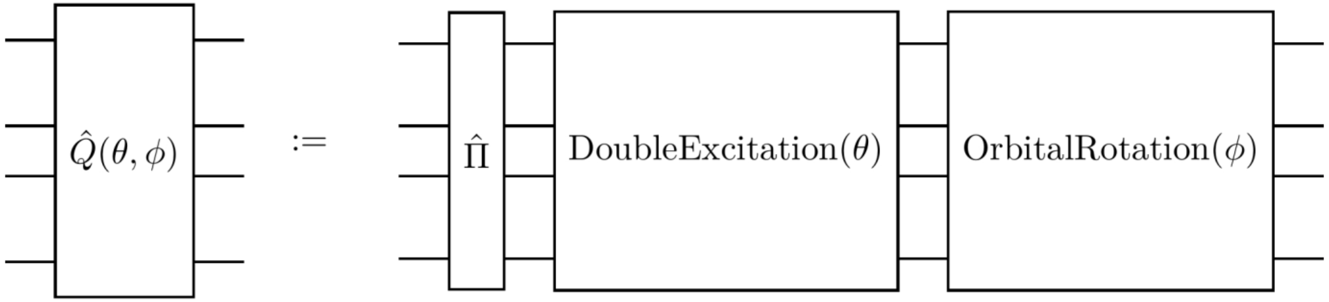
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

PAPER

Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems

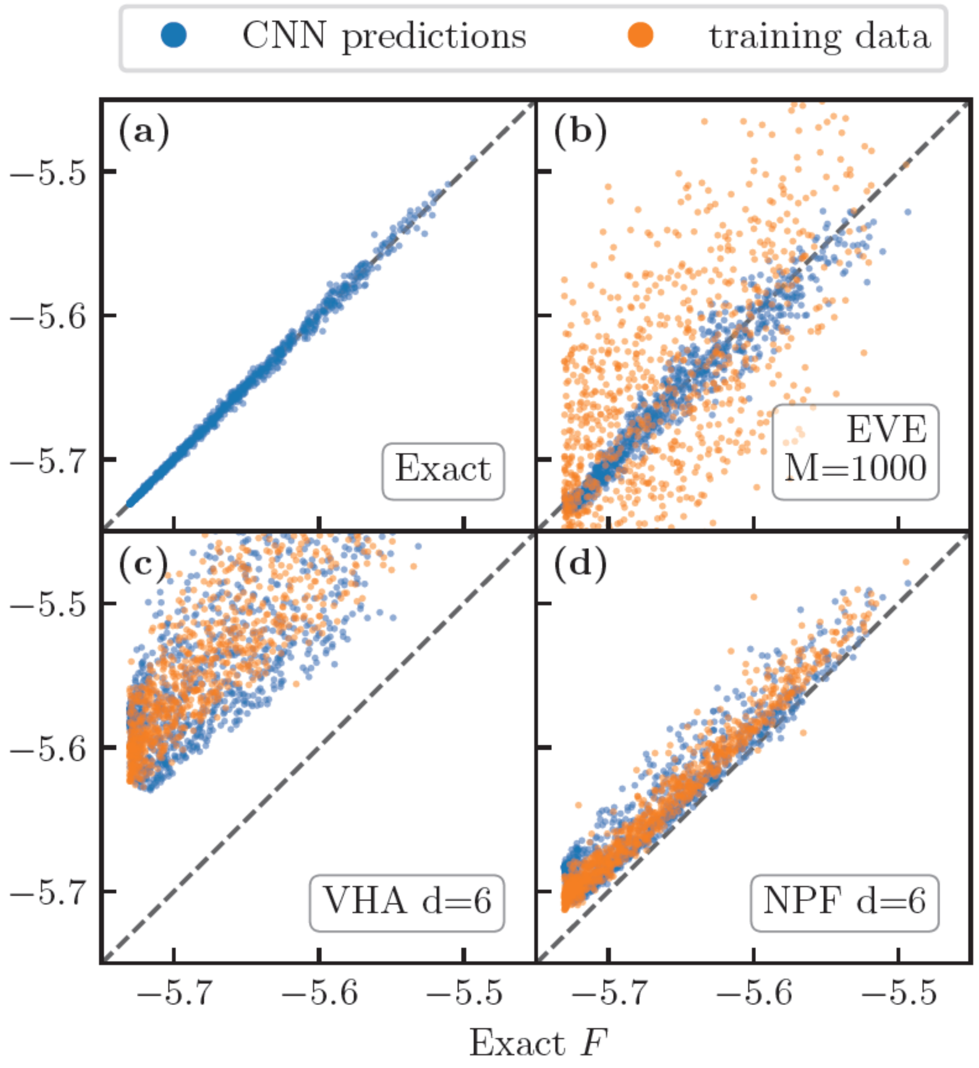
Gian-Luca R Anselmetti[1], David Wierichs[2], Christian Gogolin[1] and Robert M Parrish[3,*]





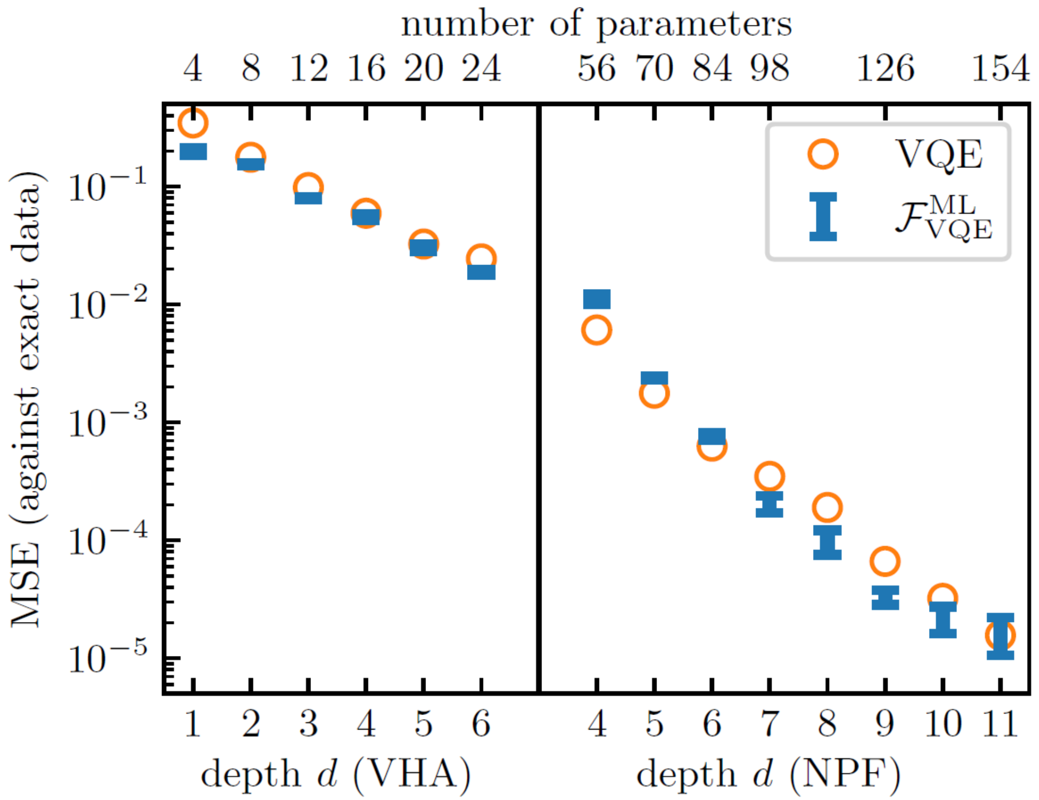
$L = 8$

Training data & results: VQE



VQE on state-vector simulator

Representability + optimization

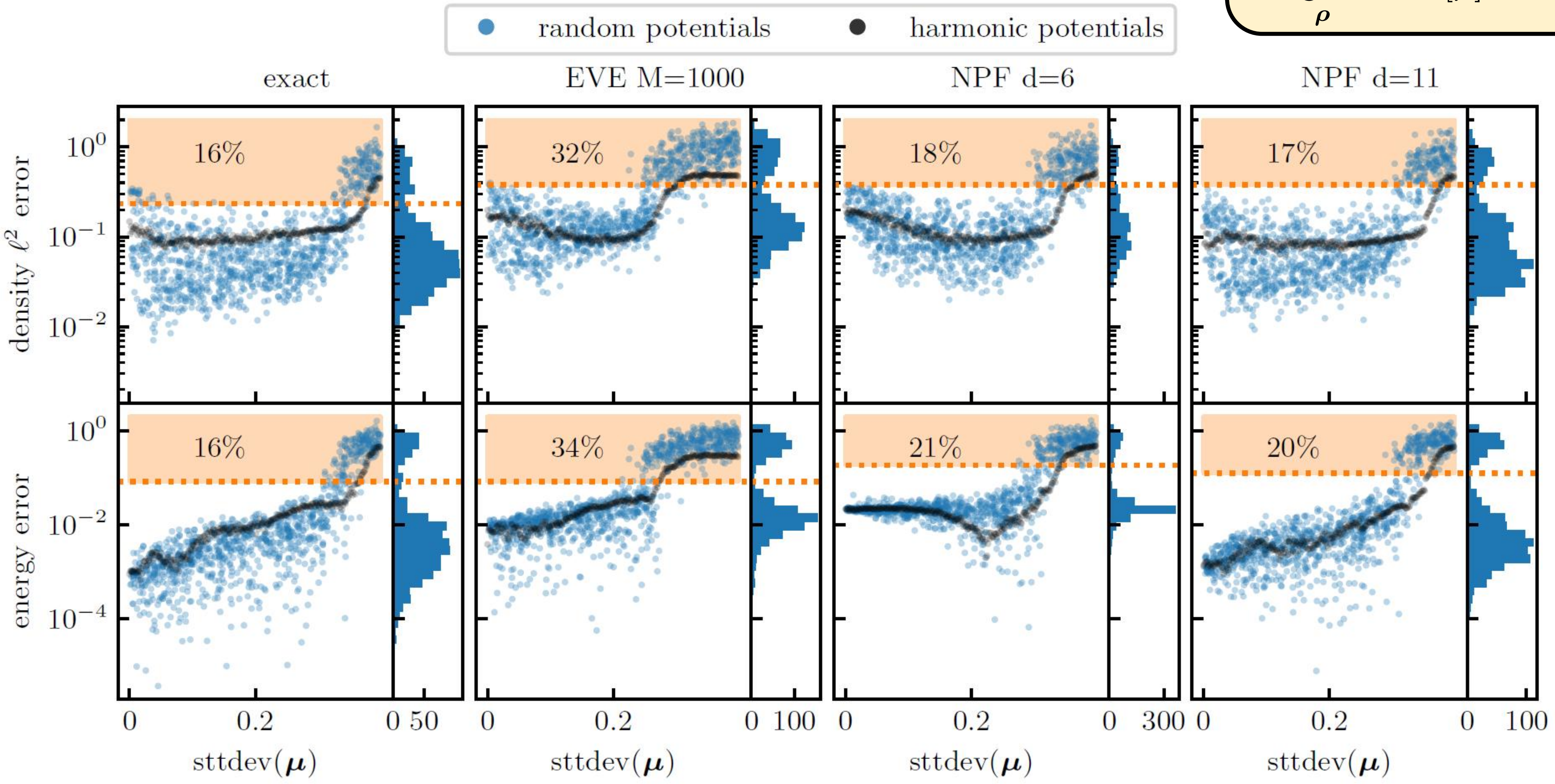


Density optimization

Density optimization

$\mathcal{E}^{\mathrm{ML}}[\boldsymbol{\rho}] := \mathcal{F}^{\mathrm{ML}} + \boldsymbol{\rho} \cdot \boldsymbol{\mu}$

$\operatorname{argmin}_{\boldsymbol{\rho}} \mathcal{E}^{\mathrm{ML}}[\boldsymbol{\rho}]$ vs. $\boldsymbol{\rho}_{\mathrm{test}}$



Conclusion & outlook

- Meaningful functionals can be learned from a small amount of noisy training data
- Significant improvement on unbiased noise
- Can be used in density optimization on new instances

https://github.com/StefanoPolla/dftqml

Conclusion & outlook

https://github.com/StefanoPolla/dftqml

- Meaningful functionals can be learned from a small amount of noisy training data
- Significant improvement on unbiased noise
- Can be used in density optimization on new instances

- Larger models
- Advanced ML de-noising
- DFT offers other learning targets, which is best?
- Transfer learning, learning from mixed datasets, learn from lower bounds