Aalto-yliopisto
Aalto-universitetet
Aalto University

Institute Q

Finnish Quantum Flagship
FQF

PHYS-EV0007 - Machine Learning from and for Quantum Science

Anouar Moustaj

Institute Q

Aalto University School of Science

InnoQ

Innovation and ecosystem development

ResQ

fundamental and applied research and infrastructure development

EduQ

education and talent development

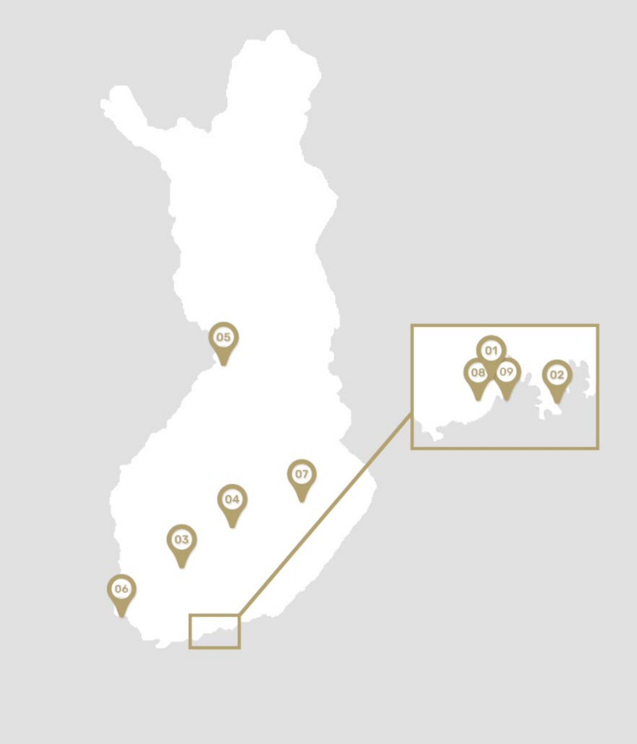
Finnish Quantum Flagship

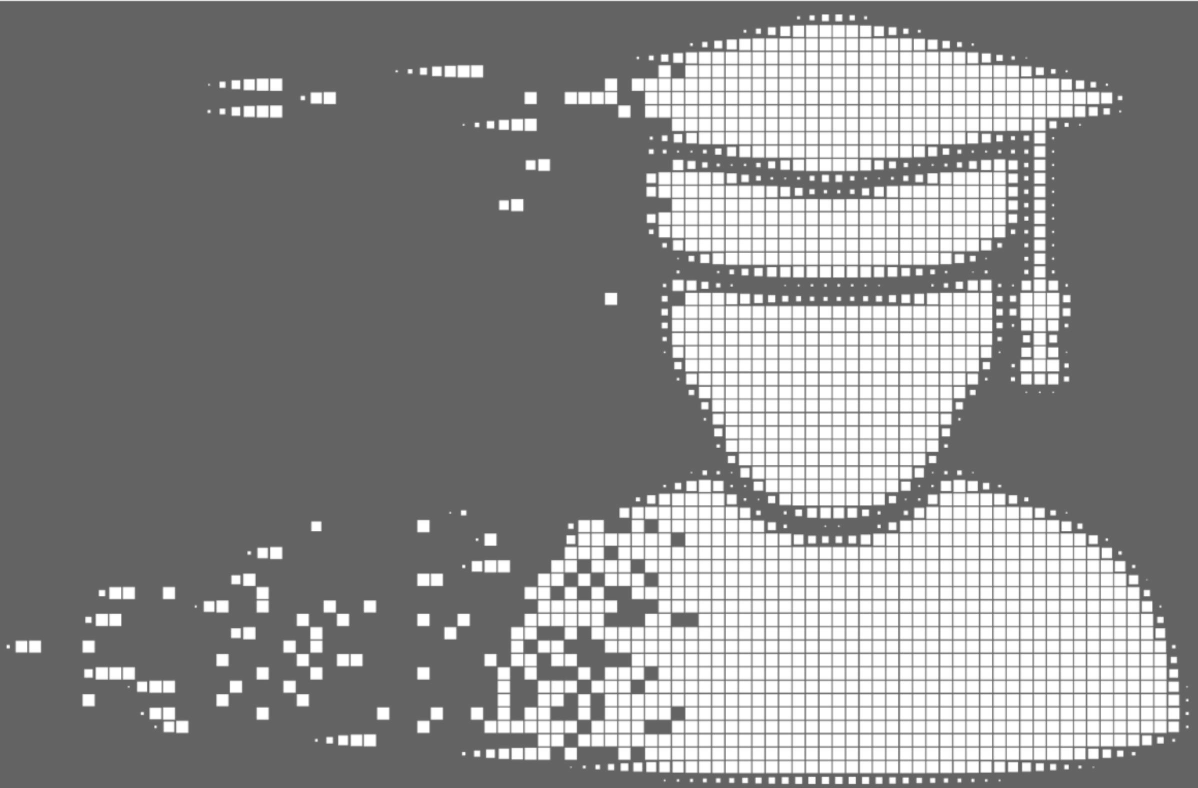
Aalto University
School of Science

Finnish Quantum Flagship

Hosted within InstituteQ, the Finnish Quantum Flagship is an 8-year project jointly funded by its various host organisations and the Research Council of Finland.

The host organisations are Aalto University (coordinator), VTT Technical Research Centre of Finland, the University of Helsinki, the University of Jyväskylä, Tampere University, the University of Oulu, and CSC-IT Centre for Science.





Aalto University
School of Science

About the Quantum Doctoral Programme

This doctoral programme (QDOC) is training the next line of quantum technology experts as part of the **Finnish Quantum Flagship's** eight-year master plan.

The Finnish Ministry of Education announced on February 7, 2024 its funding plans for Finland's new PhD pilot programme, which aims to recruit 1,000 new doctoral students of various disciplines.

90

quantum doctoral positions

€23M

in funding (2024-2027)

Aalto-yliopisto
Aalto-universitetet
Aalto University

Institute

Finnish Quantum Flagship

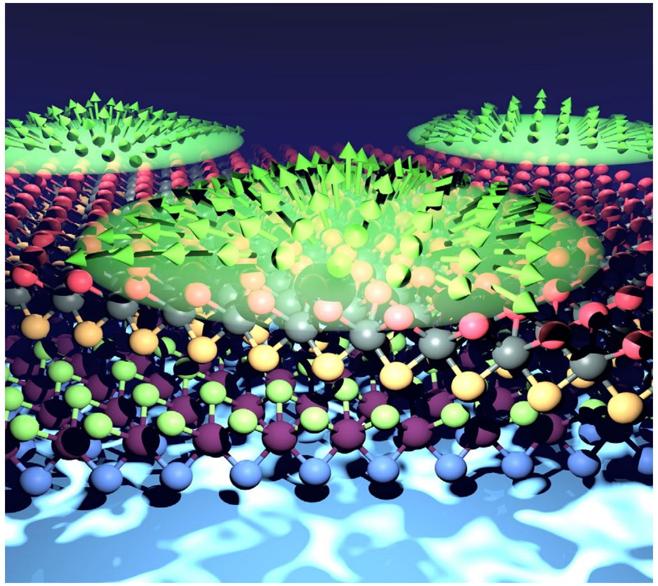
PHYS-EV0007 - Machine Learning from and for Quantum Science

Introduction

CQM group

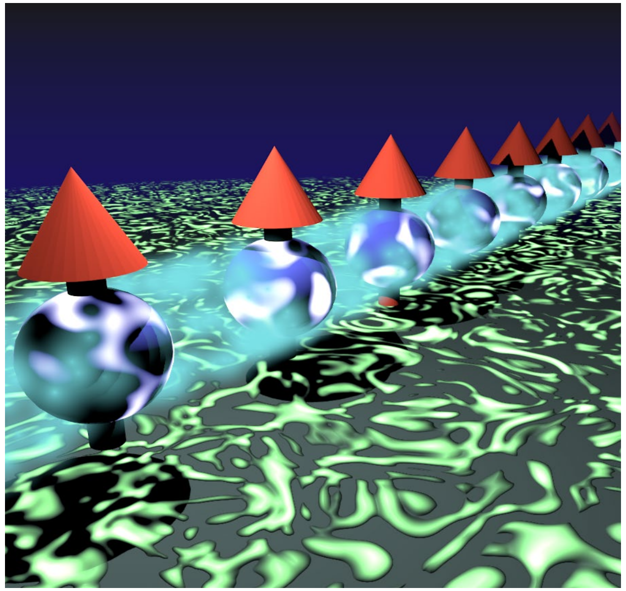
Aalto University
School of Science

Theory of van der Waals quantum materials



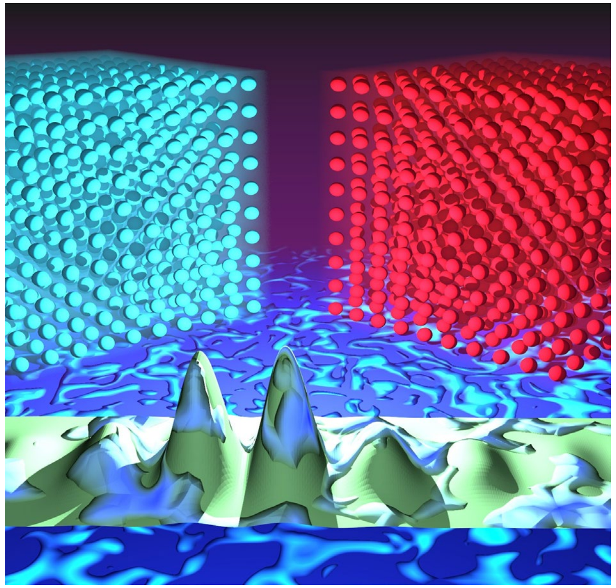
Theoretical material science

Emergence in quantum many-body physics



Theoretical condensed matter physics

Machine learning quantum materials



Machine learning for quantum technologies

Image courtesy: Jose Lado

Learning outcomes

Aalto University
School of Science

- Various ways machine learning is used in quantum science
- How quantum science shapes classical learning algorithms
- Research and Presentation
- Hands-on computational experience
- Independent project management

Course structure

- Project-based (groups of 2)
- Presentation and report: 60% of grade
- Computational task: 40% of grade

- Week 1: 2 lectures
- Week 2-4: Projects
- Week 5-6: Presentations
- Deadline project submission: 25.05.2026

Aalto University
School of Science

Course timeline

Aalto University
School of Science

L01	Tomorrow, 12:15 » 14:00	LECTURE
L01	Wednesday, 22 April, 12:15 » 14:00	LECTURE
CANCELLED L01	Monday, 27 April, 12:15 » 14:00	
L01	Monday, 4 May, 12:15 » 14:00	Q&A
L01	Monday, 11 May, 12:15 » 14:00	Q&A
L01	Monday, 18 May, 12:15 » 14:00	PRESENTATIONS
L01	Monday, 25 May, 12:15 » 14:00	PRESENTATIONS

- Friday: Projects online at 13:00
- Choice: select 3 in order of preference by sending an email (first come first serve)
- Start of project: Monday

Evaluation: Presentation

Aalto University
School of Science

- Independent literature research
- Understanding of the topic
- Clarity of explanations
- Use of sources
- Presentation quality (slides, figures, etc)
- Respect of time (20+5 min)
- Asking questions (for audience)

Evaluation: Computational part

Aalto University
School of Science

- Provide a Git page on GitLab by 25/05/2025
- Notebook with
  - Markdown cells with explanations of what you are doing
  - Code cells with minimal examples and figures
- File for larger simulations (if done)
- Figures for larger simulations (if done)
- Clear mention of contributions

Course content

Aalto University
School of Science

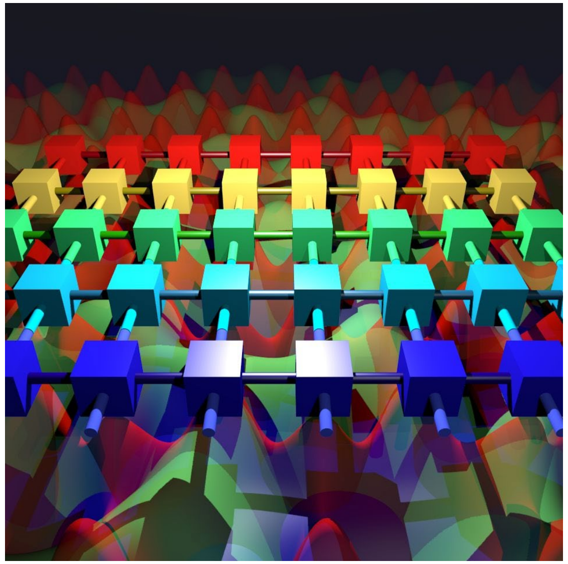
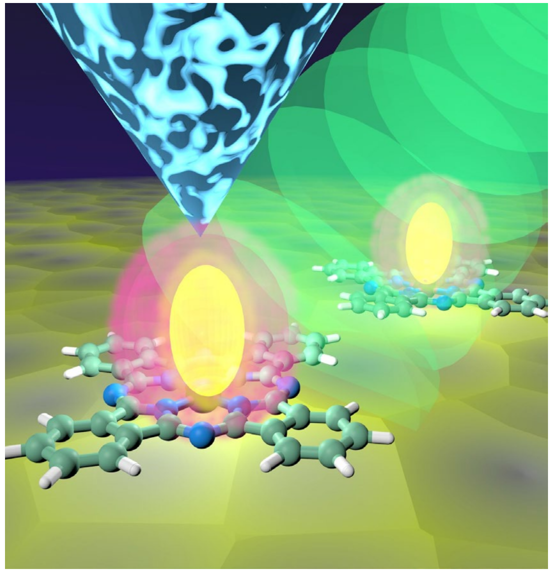
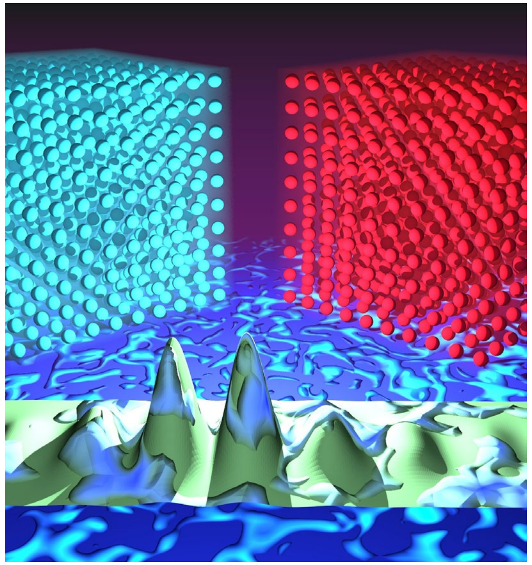
Learning Tensor Networks	Hamiltonian Learning	Neural Network Quantum States
		

Image courtesy: Jose Lado

Aalto-yliopisto
Aalto-universitetet
Aalto University

PHYS-EV0007 - Machine Learning from and for Quantum Science

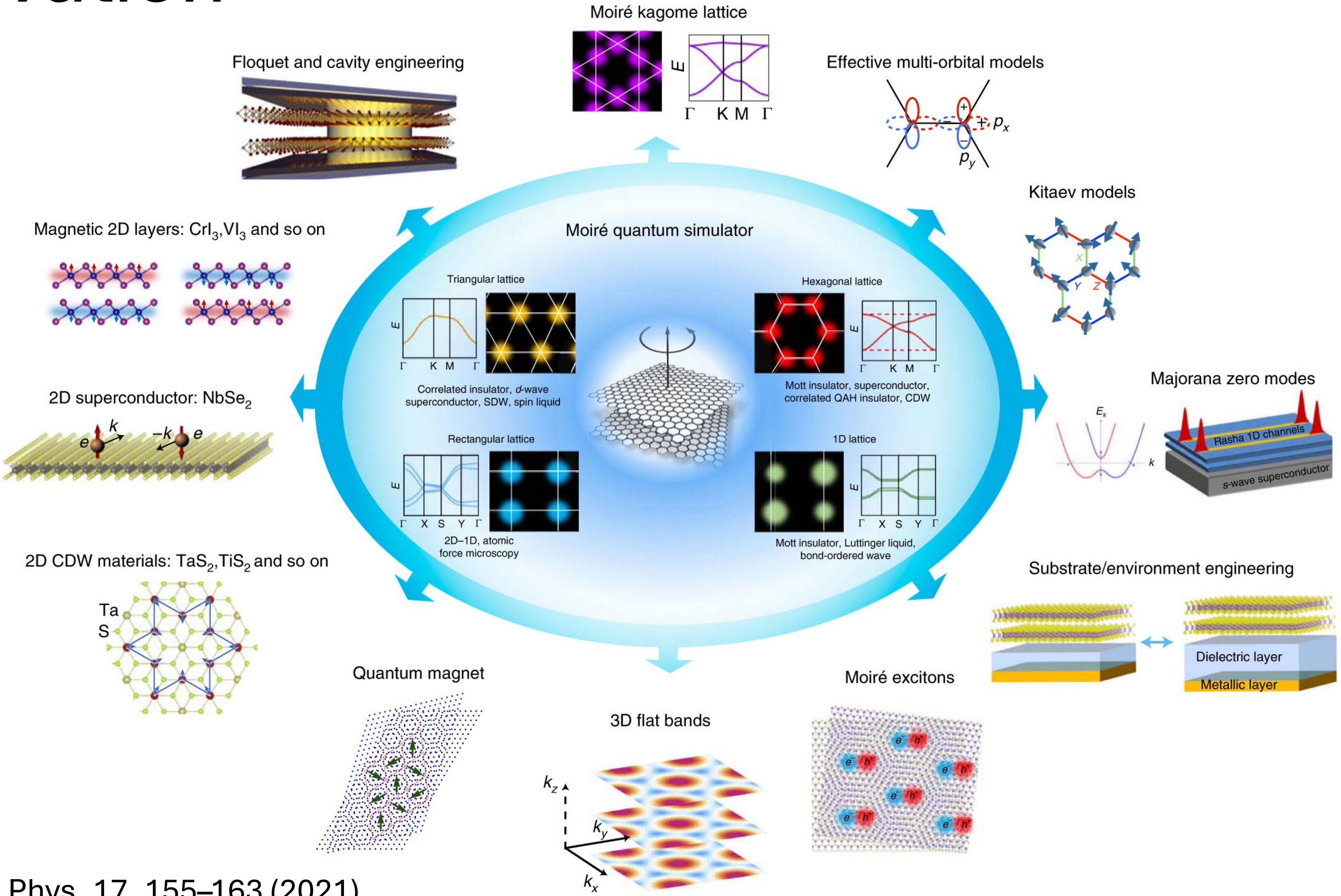
Topic 1: Active Tensor Networks Learning for Tight-Binding Models

Content

Aalto University
School of Science

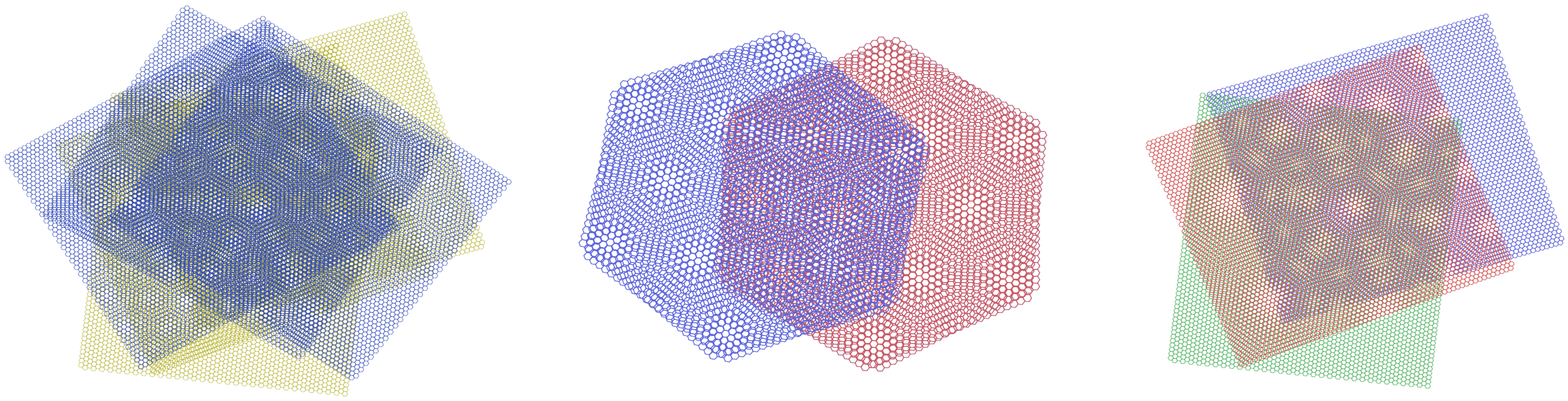
Motivation

Aalto University
School of Science



Kennes et al. – Nat. Phys. 17, 155–163 (2021)

Motivation – Limitations



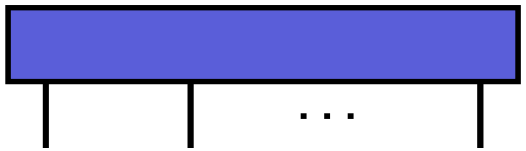
Sizes $\gtrsim 10^9$ → Storage issues: not enough RAM

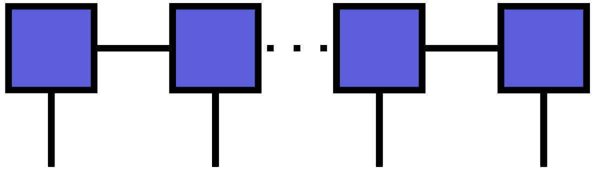
→ Computation times: days to years

Quantum many-body physics

Aalto University
School of Science

- Decades of dealing with exponentially large objects
- Tensor-network schemes are now very powerful



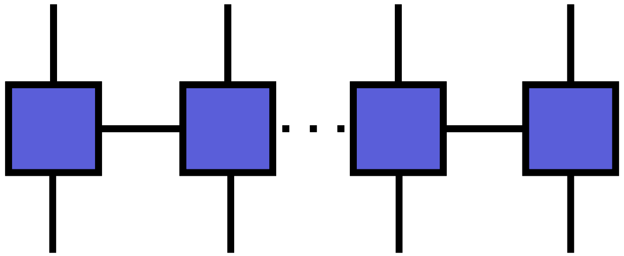


Aalto University
School of Science

Tight-Binding Hamiltonians as MPOs

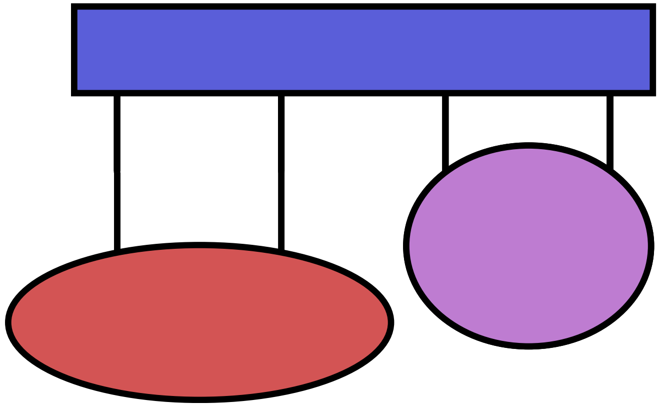
$$\hat{H} = \sum_{ij,\sigma} t_{ij} c^\dagger_{i\sigma} c_{j\sigma} + \sum_{i,\sigma} V_i c^\dagger_{i\sigma} c_{i\sigma} + \sum_i U_i c^\dagger_{i\uparrow} c_{i\uparrow} c^\dagger_{i\downarrow} c_{i\downarrow}$$

$$\hat{\mathcal{H}} =$$



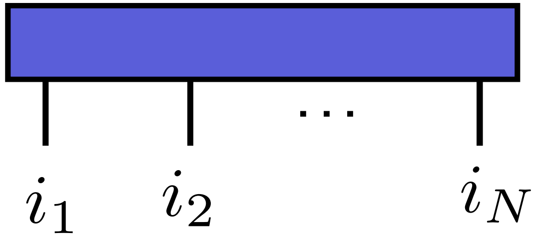
Tensor Networks Crash Course

Aalto University
School of Science

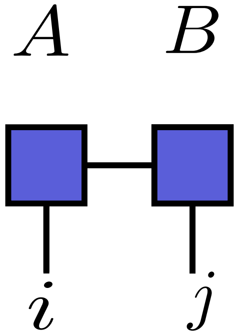


Tensor Network

$$C_{i_1 i_2 \cdots i_N}$$



$$C_{ij} = \sum_k A_{ik} B_{kj} =$$



Tensor Networks Crash Course

N Qubit states in general

$$2^N$$

$$|\psi\rangle = \sum_{i_1 i_2 \cdots i_N} C_{i_1 i_2 \cdots i_N} |i_1 i_2 \cdots i_N\rangle$$

$N > 40$ → TB memory issues

Tensor Networks Crash Course

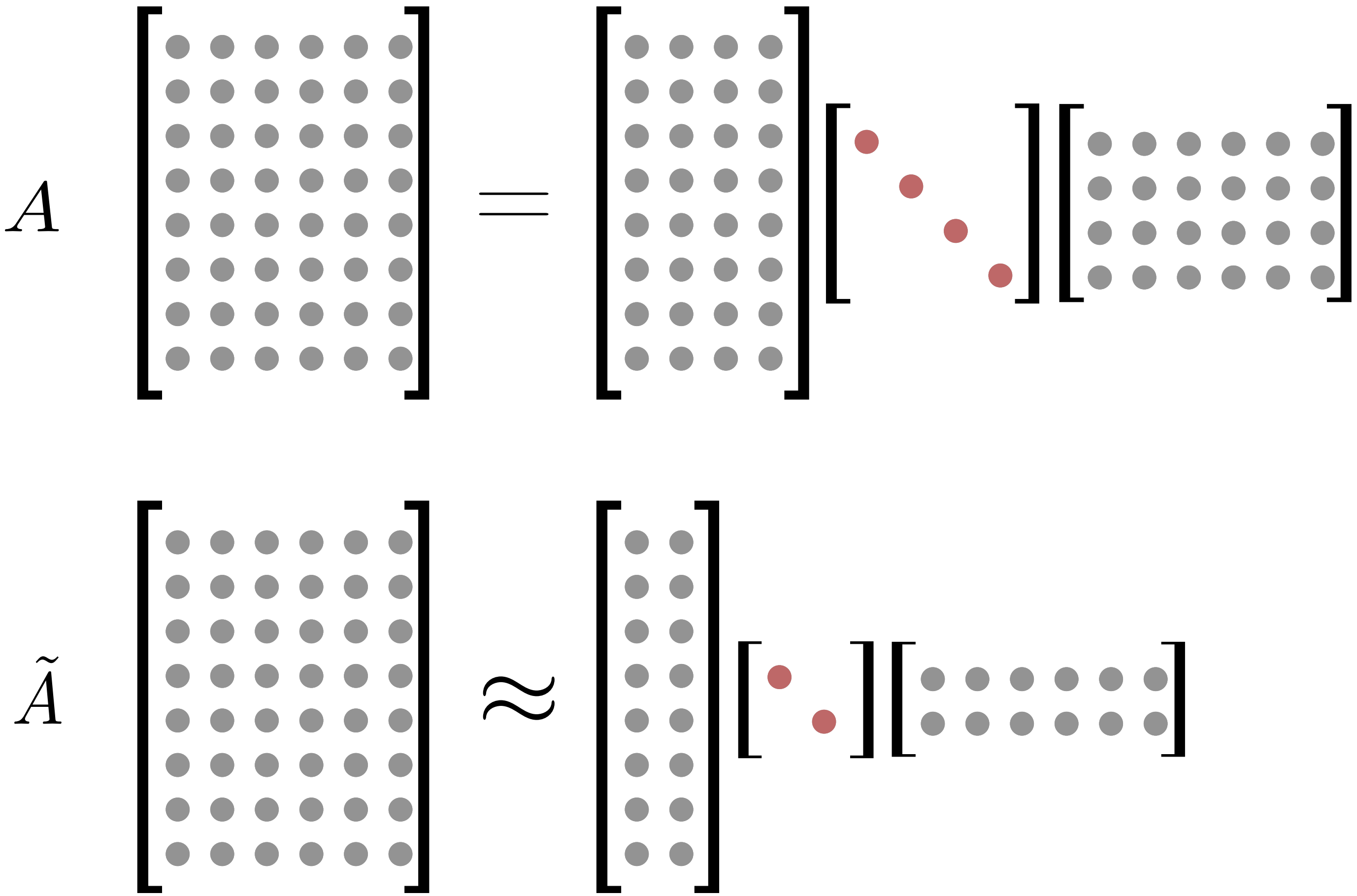
- Simplest product states

$$2N$$

$$|\psi\rangle = \sum_{i_1 i_2 \cdots i_N} C^{(1)}_{i_1} C^{(2)}_{i_2} \cdots C^{(N)}_{i_N} |i_1 i_2 \cdots i_N\rangle$$

- Commonly: somewhere in between

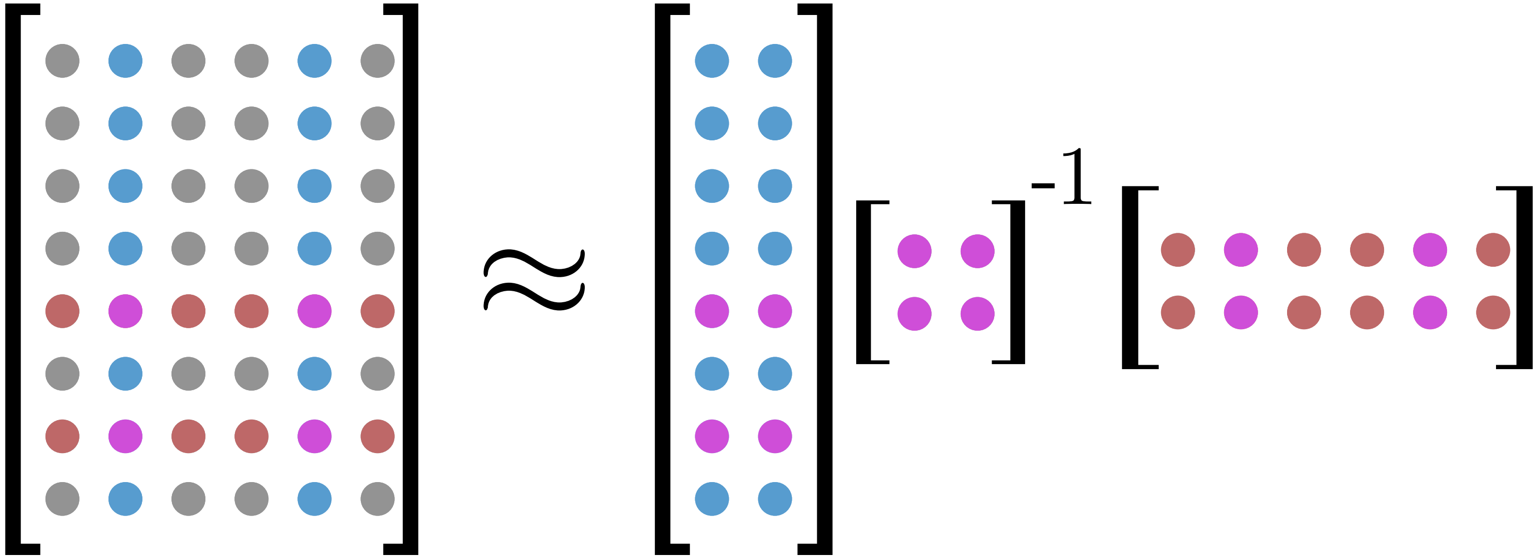
Tensor Networks – Compression: SVD



Error $\epsilon = ||A - \tilde{A}||_F^2 = \sum_{r+1}^{Q} \lambda_k^2$

Aalto University
School of Science

Tensor Networks – Cross Interpolation



$$\epsilon_{\mathrm{CI}} = (1+\delta)\epsilon_{\mathrm{SVD}}$$

BUT less complex

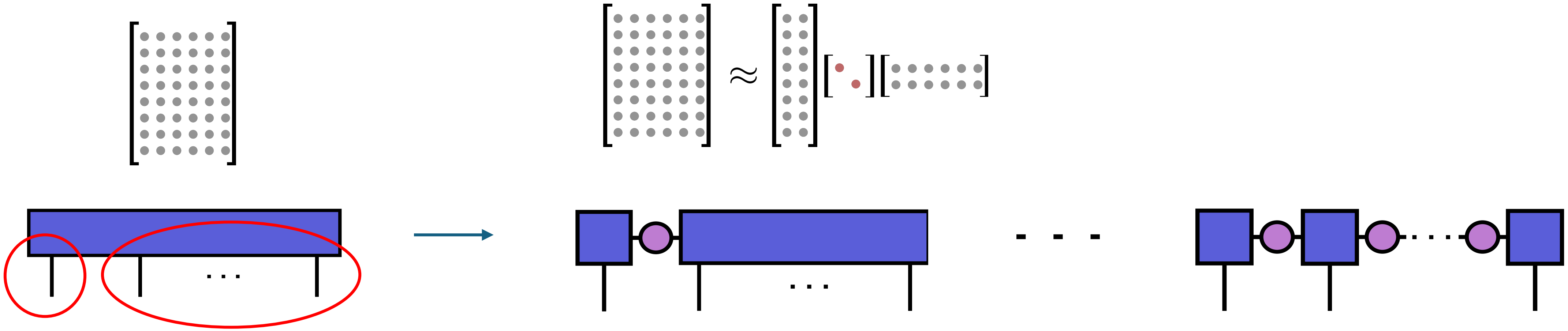
AND less storage

$$SVD \rightarrow \mathcal{O}(mnk)$$

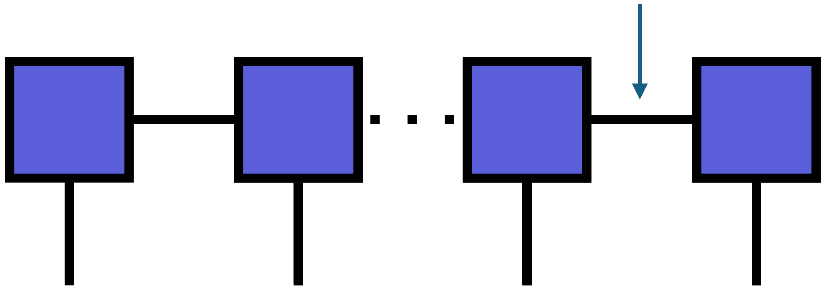
$$MCI \rightarrow \mathcal{O}((m+n)k^2)$$

$$MCI \rightarrow \mathcal{O}((m+n)k)$$

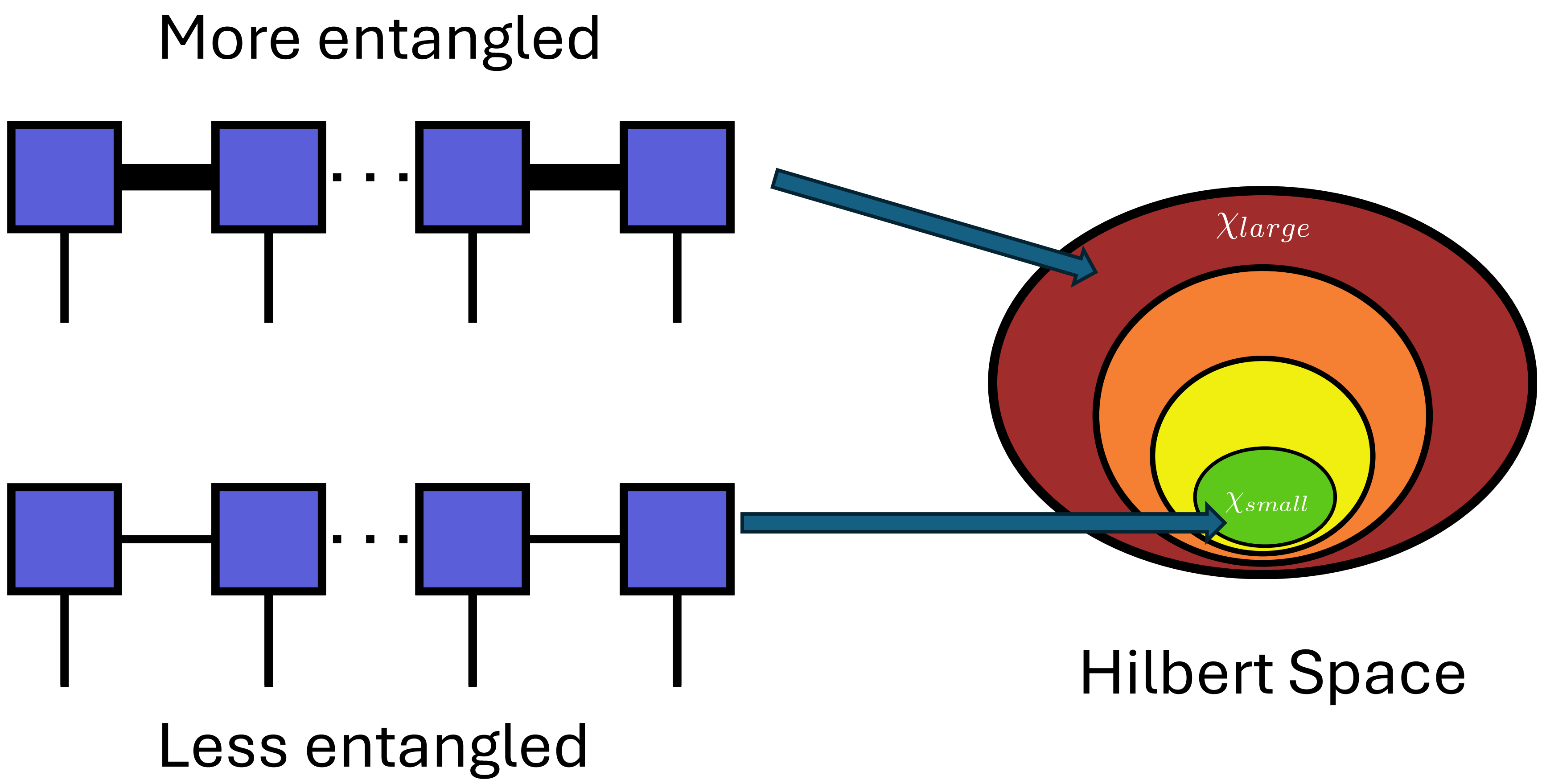
Tensor Networks – Decomposition



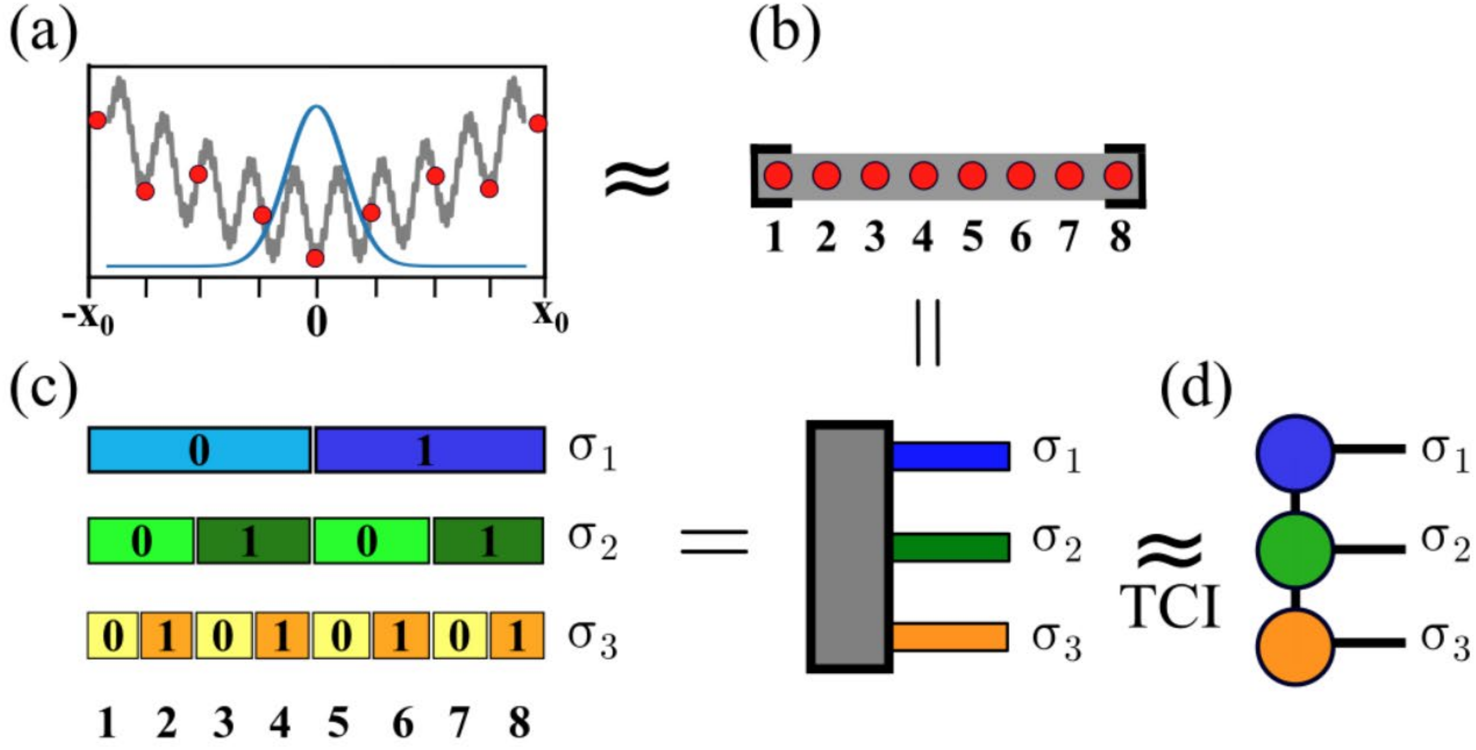
TT/MPS $\chi \sim e^S$



Tensor Networks – Decomposition



Quantics Representation of functions



$$f_m = f\left(m = \sum_{n=1}^{L} \sigma_n 2^{L-n}\right) = f_{\sigma_1 \sigma_2 \cdots \sigma_R}$$

M. Niedermeier – PhD thesis (2025)

Quantics Representation of functions

Aalto University
School of Science

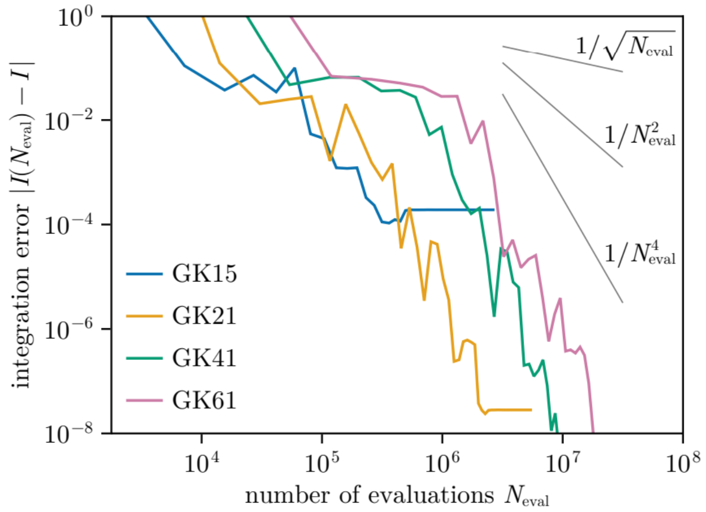


Figure 2: Convergence of the 10-dimensional integral I of Eq. (5). $I(N_{eval})$ is computed using TCI with 15, 21, 41 and 61-point Gauss–Kronrod quadrature in each dimension, and N_{eval} is the number of evaluations of the integrand. With 41- and 61-point quadrature, the value converges to $I = -5.4960415218049$. Convergence of the lower-order quadrature rules is limited by the number of discretization points.

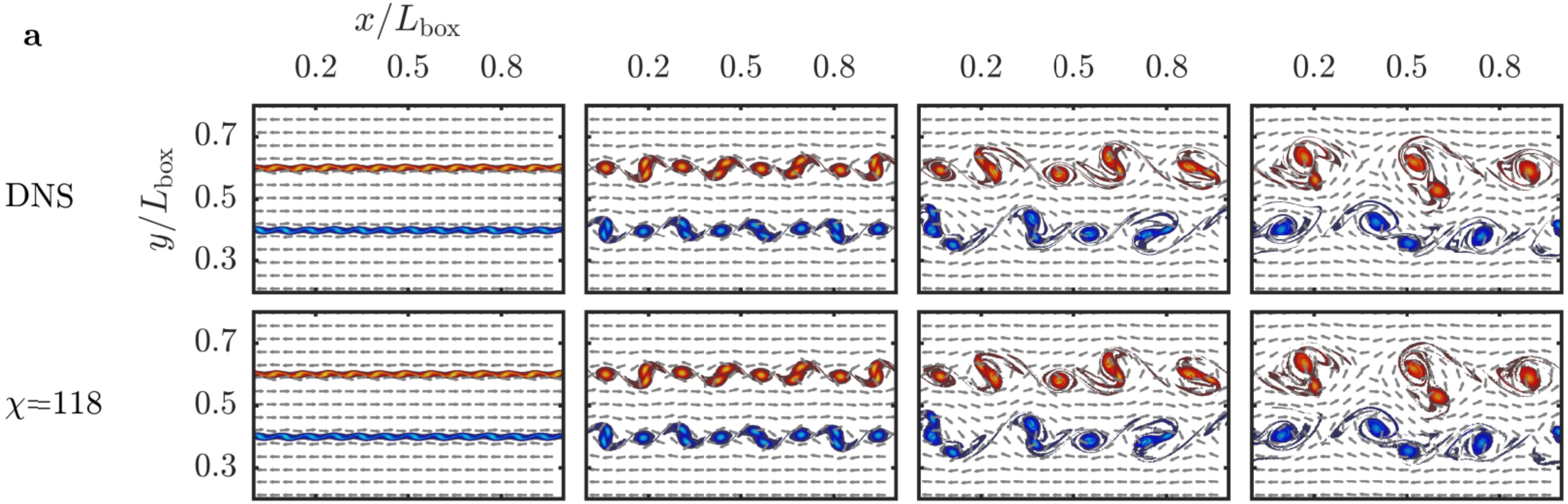
$$I = 10^3 \int_{[-1,+1]^{10}} \mathrm{d}^{10}\mathbf{x}\ \cos\Big(10 \sum_{\ell=1}^{10} x_\ell^2\Big)\ \exp\Big[-10^{-3}\Big(\sum_{\ell=1}^{10} x_\ell\Big)^4\Big]$$

Y.N. Fernandez et al. SciPost Phys. 18 104 (2025)

Quantics Representation of functions

A Quantum Inspired Approach to Exploit Turbulence Structures

Nikita Gourianov[1],* Michael Lubasch[2], Sergey Dolgov[3], Quincy Y. van den Berg[1], Hessam Babaee[4], Peyman Givi[4], Martin Kiffner[5,1], and Dieter Jaksch[1,5,6]†



Aalto University
School of Science

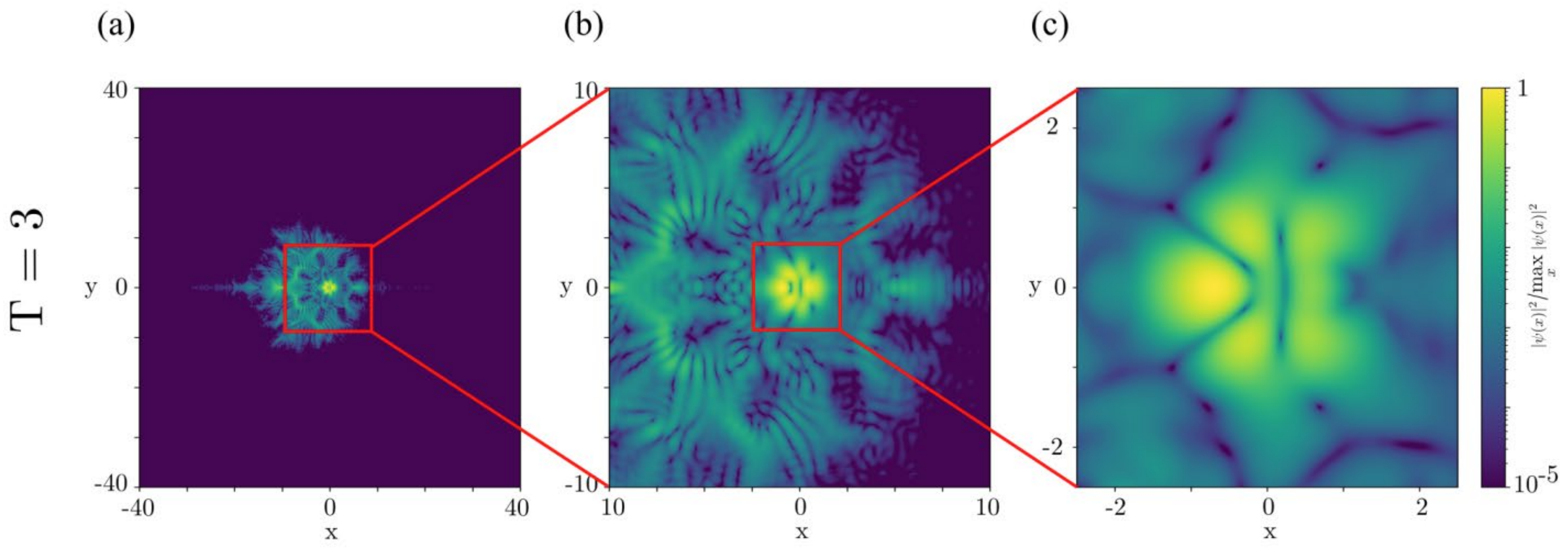
Quantics Representation of functions

Solving the Gross-Pitaevskii equation on multiple different scales using the quantics tensor train representation

Marcel Niedermeier,[1,2] Adrien Moulinas,[2] Thibaud Louvet,[2] Jose L. Lado,[1] and Xavier Waintal[2]

[1] *Department of Applied Physics, Aalto University, 02150 Espoo, Finland*
[2] *Univ. Grenoble Alpes, CEA, Grenoble INP, IRIG, Pheliqs, 38000 Grenoble, France*

(Dated: July 8, 2025)



Quantics Representation of functions

Aalto University
School of Science

Learning Feynman Diagrams with Tensor Trains

Yuriel Núñez-Fernández,[1,*] Matthieu Jeannin,[1] Philipp T. Dumitrescu,[2] Thomas Kloss,[1,3] Jason Kaye,[2,4] Olivier Parcollet,[2,5] and Xavier Waintal[1,†]

Tensorized orbitals for computational chemistry

Nicolas Jolly,[1] Yuriel Núñez Fernández,[2] and Xavier Waintal[1,*]
[1]*Université Grenoble Alpes, CEA, Grenoble INP, IRIG, PHELIQS, 38000 Grenoble, France*
[2]*Université Grenoble Alpes, CNRS, Institut Néel, 38000 Grenoble, France*

Solving the Gross-Pitaevskii equation on multiple different scales using the quantics tensor train representation

Marcel Niedermeier,[1,2] Adrien Moulinas,[2] Thibaud Louvet,[2] Jose L. Lado,[1] and Xavier Waintal[2]
[1]*Department of Applied Physics, Aalto University, 02150 Espoo, Finland*
[2]*Univ. Grenoble Alpes, CEA, Grenoble INP, IRIG, Pheliqs, 38000 Grenoble, France*

Permutation of Tensor-Train Cores for Computing Moments on Stochastic Differential Equations

Kayo Kinjo[1*], Rihito Sakurai[2], Tatsuya Kishimoto[1], and Jun Ohkubo[1]

A Quantum Inspired Approach to Exploit Turbulence Structures

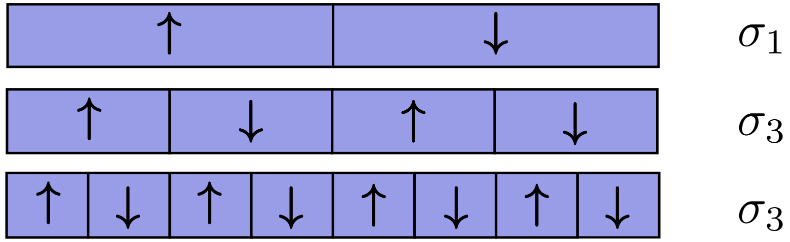
Nikita Gourianov[1,*] Michael Lubasch[2], Sergey Dolgov[3], Quincy Y. van den Berg[1], Hessam Babaee[4], Peyman Givi[4], Martin Kiffner[5,1], and Dieter Jaksch[1,5,6†]

TENSOR TRAIN REPRESENTATIONS OF GREEKS FOR FOURIER-BASED OPTION PRICING OF MULTI-ASSET OPTIONS *

RIHITO SAKURAI[†], KOICHI MIYAMOTO[‡], AND TSUYOSHI OKUBO[†§]

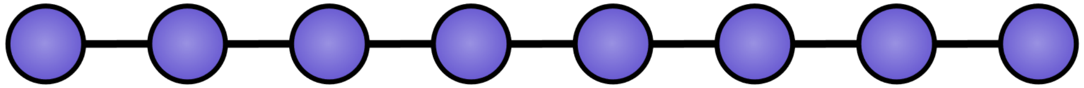
Aalto University
School of Science

Tight-Binding Hamiltonians as MPOs

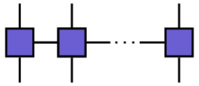


$L = 3$

$c_1^\dagger |0\rangle$ $c_2^\dagger |0\rangle$ $c_3^\dagger |0\rangle$ $c_4^\dagger |0\rangle$ $c_5^\dagger |0\rangle$ $c_6^\dagger |0\rangle$ $c_7^\dagger |0\rangle$ $c_8^\dagger |0\rangle$



$|\uparrow\uparrow\uparrow\rangle$ $|\uparrow\uparrow\downarrow\rangle$ $|\uparrow\downarrow\uparrow\rangle$ $|\uparrow\downarrow\downarrow\rangle$ $|\downarrow\uparrow\uparrow\rangle$ $|\downarrow\uparrow\downarrow\rangle$ $|\downarrow\downarrow\uparrow\rangle$ $|\downarrow\downarrow\downarrow\rangle$

H_{ij} —— Tensor Network Representation ——→ $\mathcal{H}_{\alpha\alpha'} \equiv$ 

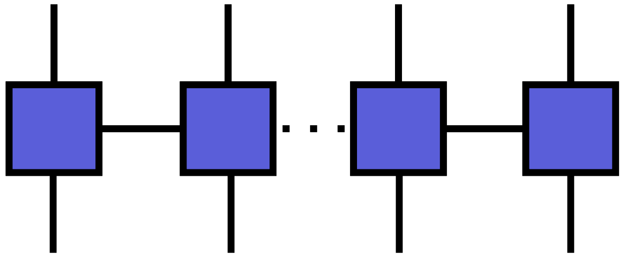
$2^L \times 2^L$ $\qquad$ L sites MPO

Aalto University
School of Science

Tight-Binding Hamiltonians as MPOs

$$\hat{H} = \sum_{ij,\sigma} t_{ij} c^{\dagger}_{i\sigma} c_{j\sigma} + \sum_{i,\sigma} V_i c^{\dagger}_{i\sigma} c_{i\sigma} + \sum_{i} U_i c^{\dagger}_{i\uparrow} c_{i\uparrow} c^{\dagger}_{i\downarrow} c_{i\downarrow}$$

$$\hat{\mathcal{H}} =$$



TN KPM: Spectral Function

Aalto University
School of Science

$$f(\omega) = \delta(\omega - \hat{\mathcal{H}})$$

$$\mu_n = T_n(\mathcal{H})$$

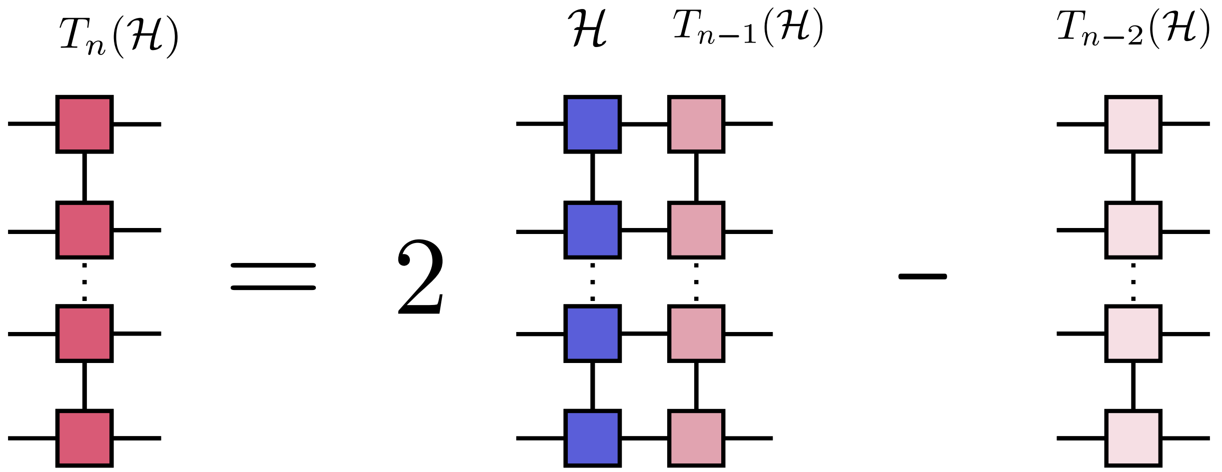
Repeated MPO contractions and applications of TCI to accelerate calculations

$$\delta(\omega - \hat{\mathcal{H}}) \approx \frac{1}{\pi\sqrt{1-\omega^2}} \left[\hat{I} + 2\sum_{n=1}^{N} T_n(\hat{\mathcal{H}}) T_n(\omega) \right]$$

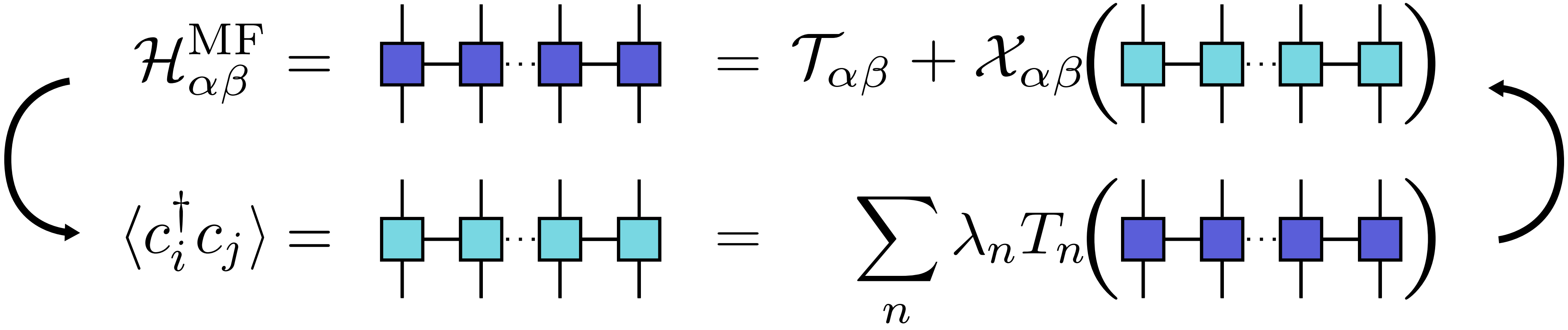
TN KPM: Spectral Function

$$f(\omega) = \delta(\omega - \hat{\mathcal{H}}) \qquad \mu_n = T_n(\mathcal{H})$$

Repeated MPO contractions and applications of TCI to accelerate calculations



TN KPM: Mean-field calculations

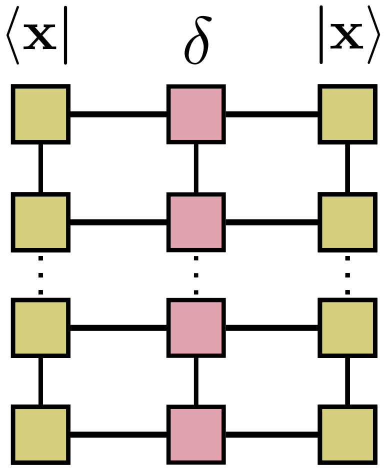


Y. Sun et al. arXiv 2503.04373 (2025)

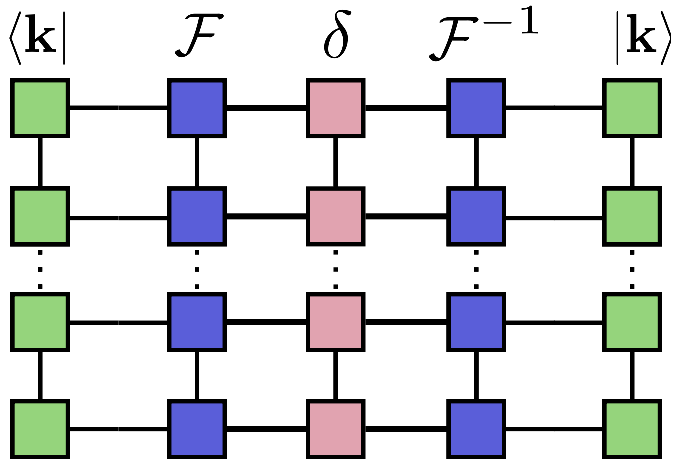
TN KPM: Spectral functions

- Contract with MPS $|\mathbf{x}\rangle$ → $\rho(\mathbf{x},\omega) = \langle \mathbf{x}|\, \delta(\omega - \hat{\mathcal{H}})\, |\mathbf{x}\rangle$
- Contract with MPS $|\mathbf{k}\rangle$ → $\mathcal{A}(\mathbf{k},\omega) = \langle \mathbf{k}|\, \hat{\mathcal{F}}\delta(\omega - \hat{\mathcal{H}})\hat{\mathcal{F}}^{-1}\, |\mathbf{k}\rangle$

$\rho(\mathbf{x}\ \ \omega) =$



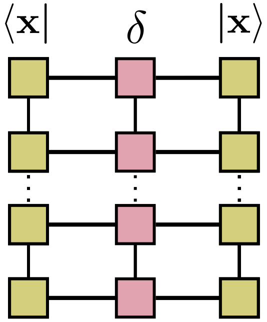
$\mathcal{A}(\mathbf{k},\omega) =$



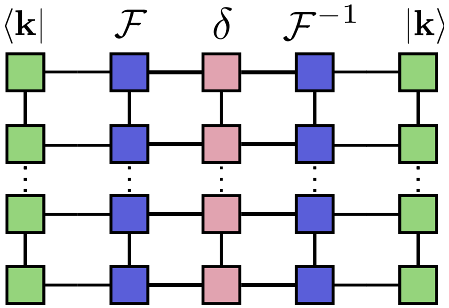
Aalto University
School of Science

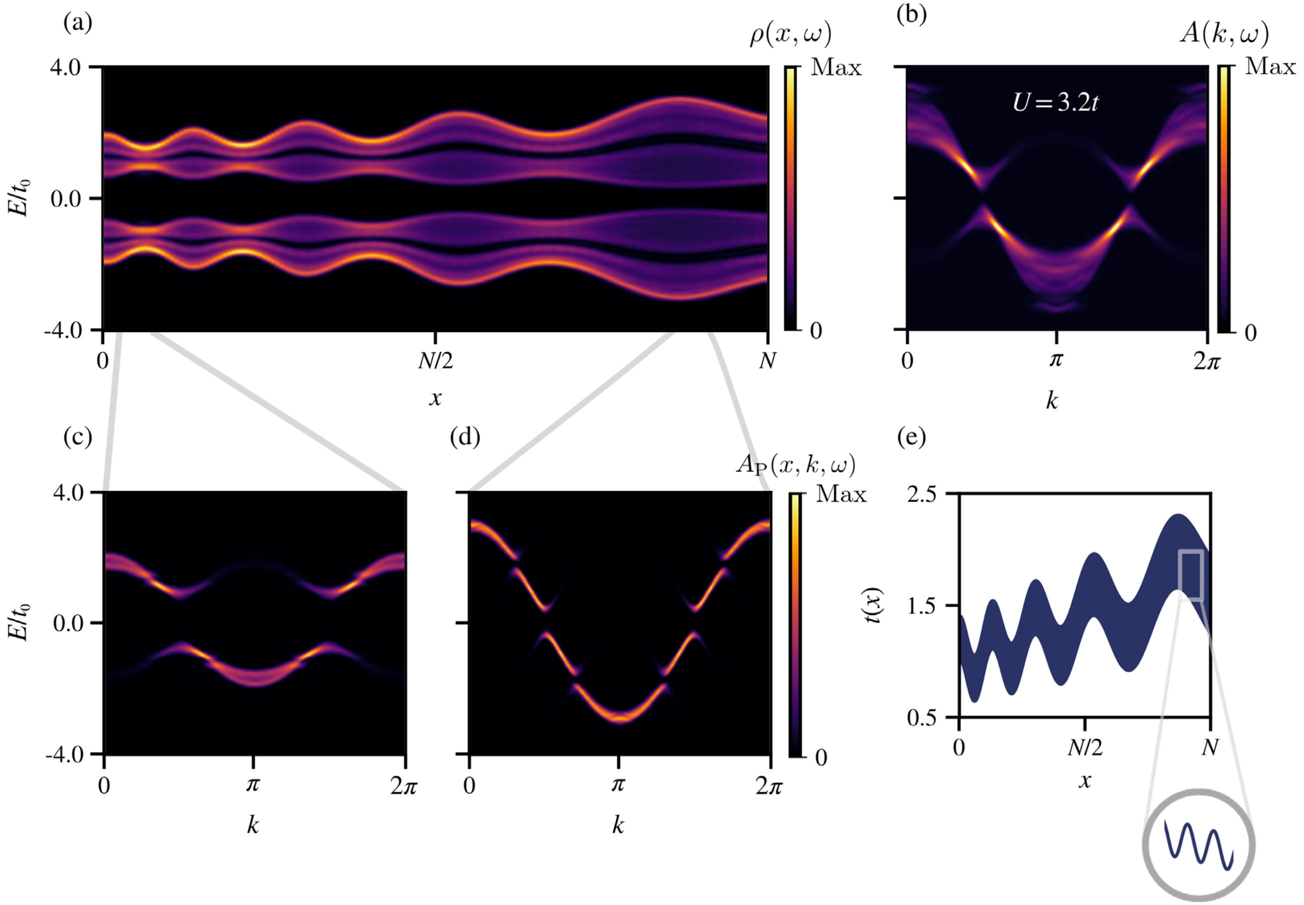
Examples

$$\rho(\mathbf{x}\ \ \omega) =$$



$$\mathcal{A}(\mathbf{k}, \omega) =$$



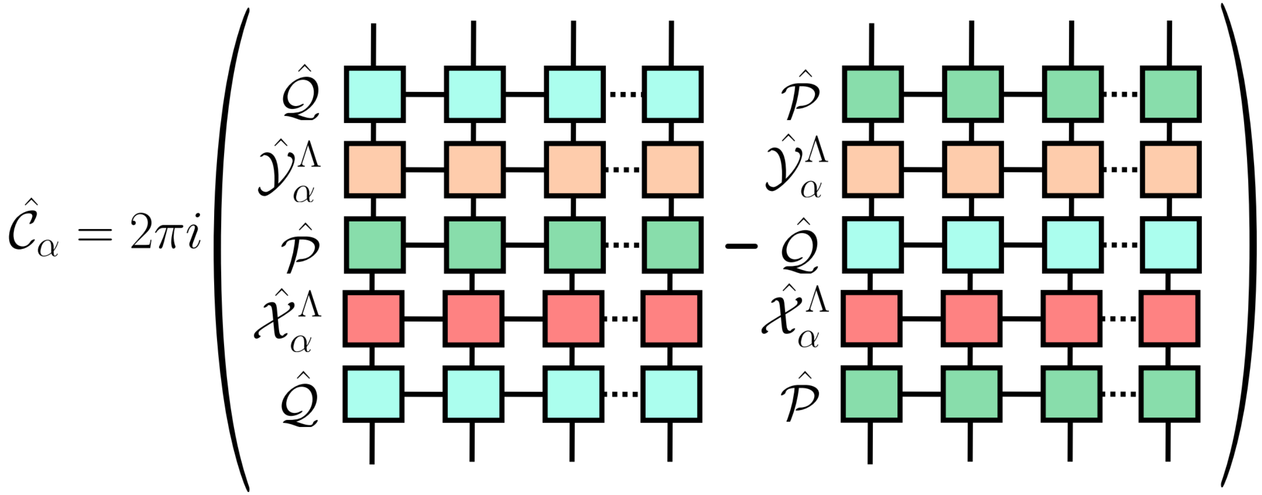


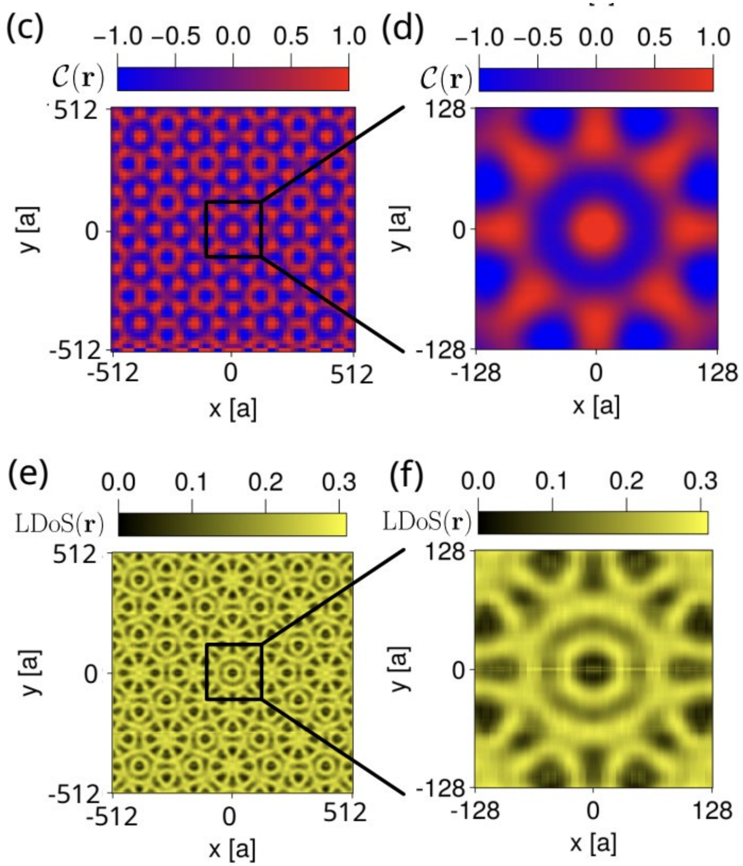
A. Moustaj et al. arXiv 2512.18397 (2025)

Aalto University
School of Science

TN KPM: Chern Marker

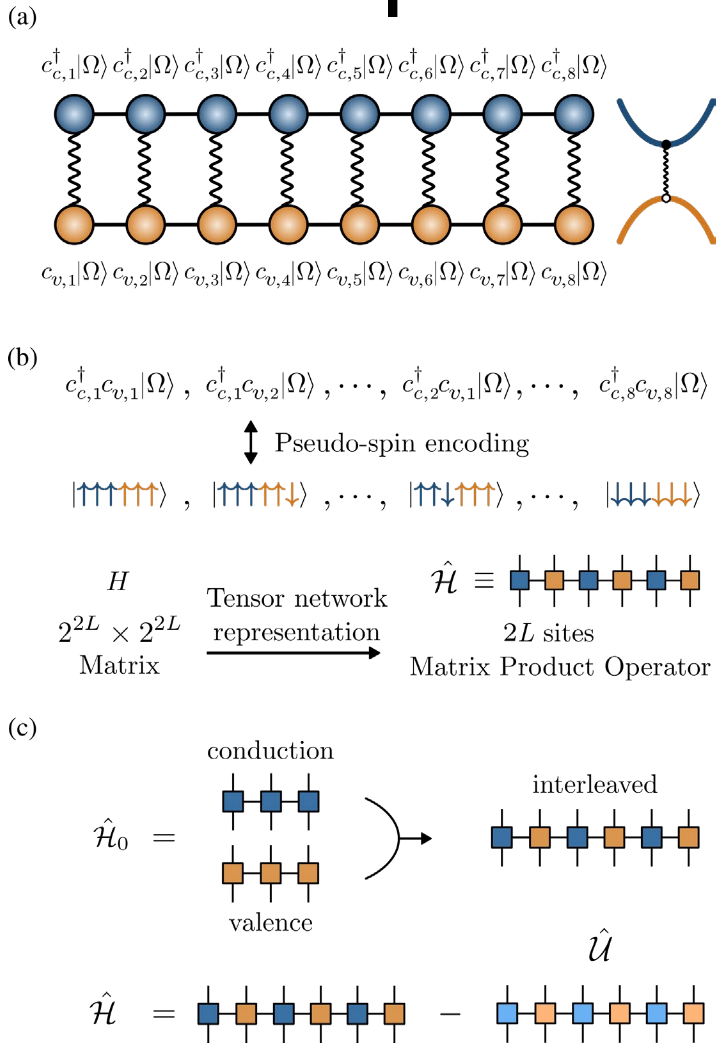
$$\hat{\mathcal{C}}_\alpha = 2\pi \mathrm{Im} \hat{\mathcal{P}} \left[\hat{\mathcal{Q}} \hat{\mathcal{X}}_\alpha^\Lambda , \hat{\mathcal{P}} \hat{\mathcal{Y}}_\alpha^\Lambda \right] \hat{\mathcal{P}}$$

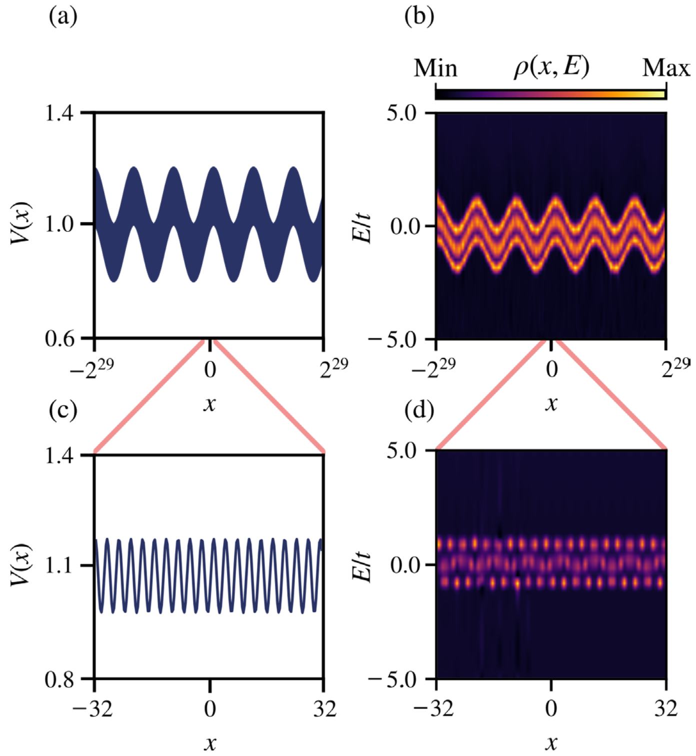


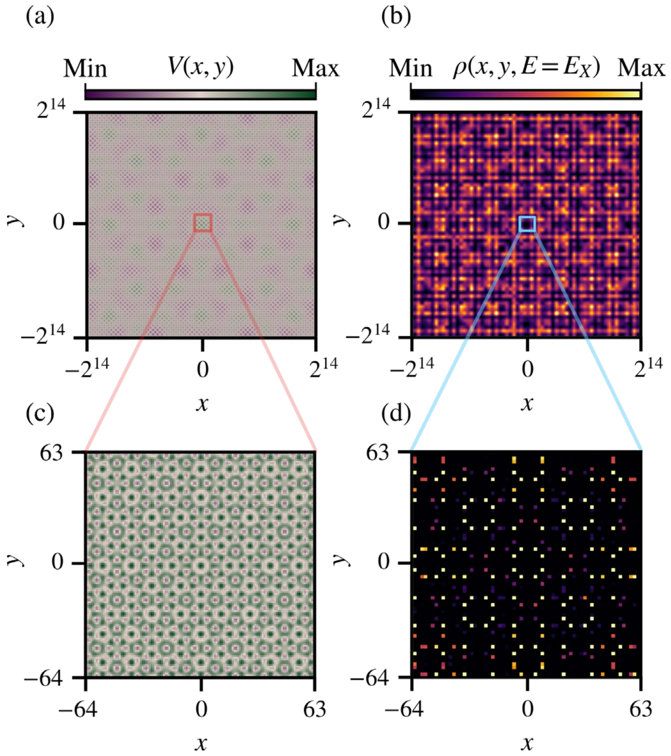


T. Antao et al. arXiv 2506.05230 (2025)

Examples: excitons (beyond MF)







A. Moustaj, Y. Sun et al. arXiv 2603.02011(2026)

Aalto-yliopisto
Aalto-universitetet
Aalto University

PHYS-EV0007 - Machine Learning from and for Quantum Science

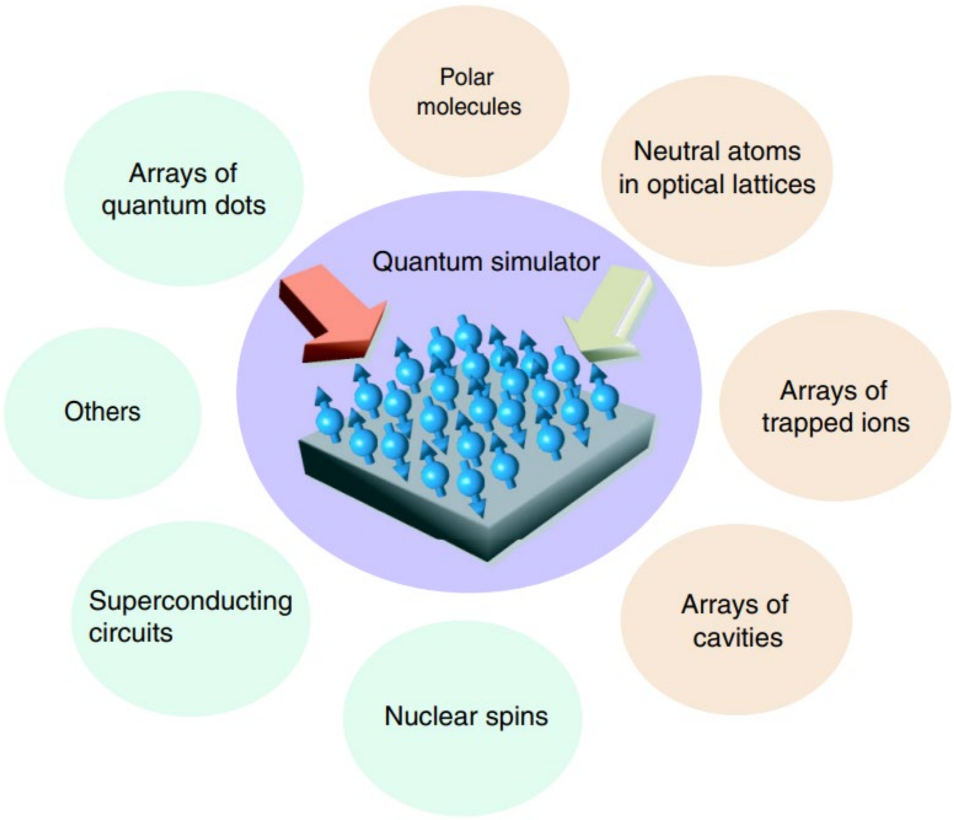
Topic 2: Hamiltonian Learning

Content

1. Motivation
2. General Concept
3. Examples
4. Supervised Learning
5. ML good practice

Aalto University
School of Science

Motivation: Quantum Simulators



$$\hat{H}(\lambda_i) = \sum_i \lambda_i \hat{h}_i$$

I.M. Georgescu et al. – Rev. Mod. Phys. 86, 153 (2014)

General Concept: Inverse problem

Aalto University
School of Science

Theory

Hamiltonian

$$\hat{H}(\lambda_i) = \sum_i \lambda_i \hat{h}_i$$

Predict →

Observable

$$O(\lambda_i) = \langle \Psi(\lambda_i) | \hat{O} | \Psi(\lambda_i) \rangle$$

Experiment

Samples

$$\{O_i\}_i^N$$

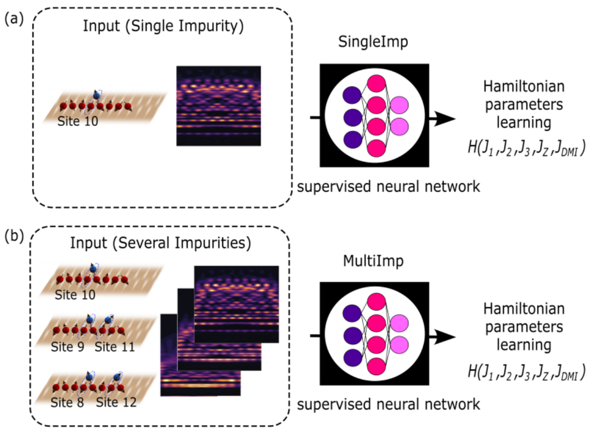
Determine? →

Hamiltonian

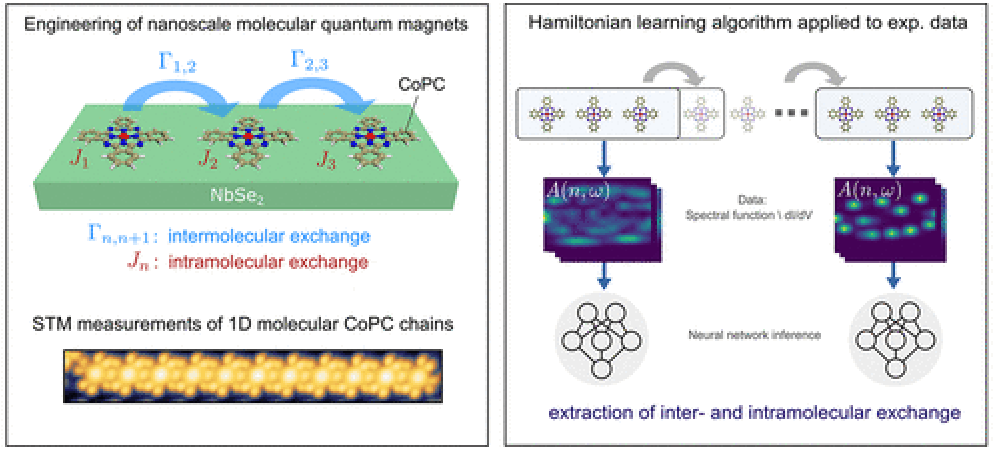
$$\hat{H}(\lambda_i)$$

Aalto University
School of Science

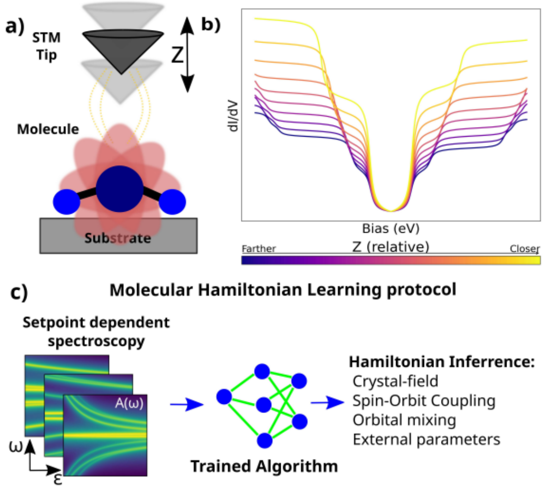
Different Applications



N. Karjalainen et al. – arXiv 2510.18613 (2025)



R. Koch et al. – Nano Lett. 25, 36, 13435-13440 (2025)



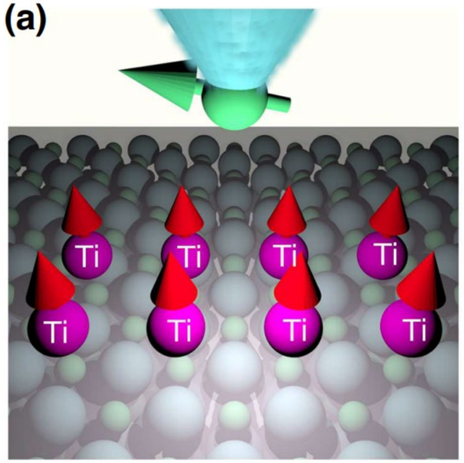
G. Lupi et al. – arXiv 2601.19371 (2026)

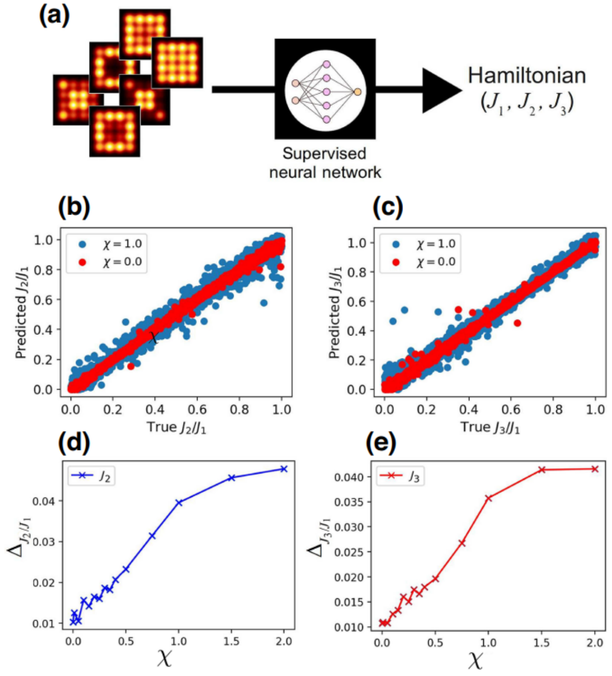
Example: Quantum Magnets

$$\hat{H} = \sum_{\langle i,j \rangle} J_{ij} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j$$

$$A(j,\omega) = \langle \Omega | \hat{S}_j \delta(\omega - \hat{H} - E_\Omega) \hat{S}_j | \Omega \rangle$$

$$\frac{\mathrm{d}^2 I}{dV^2} \propto A(j,\omega)$$

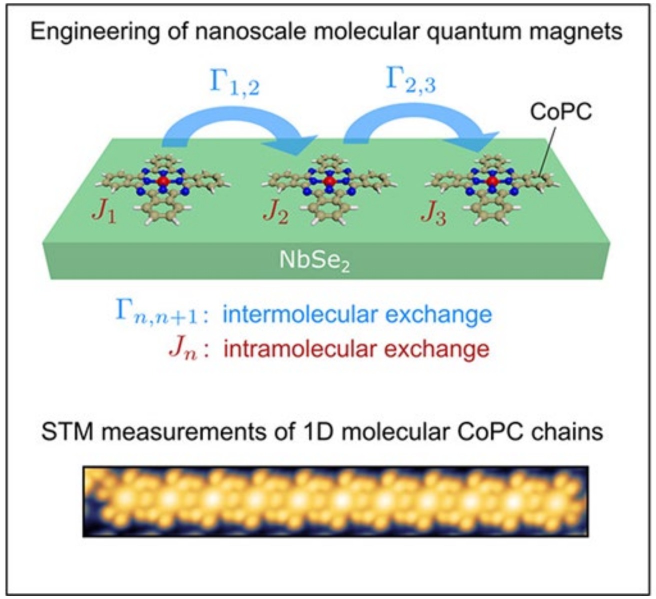


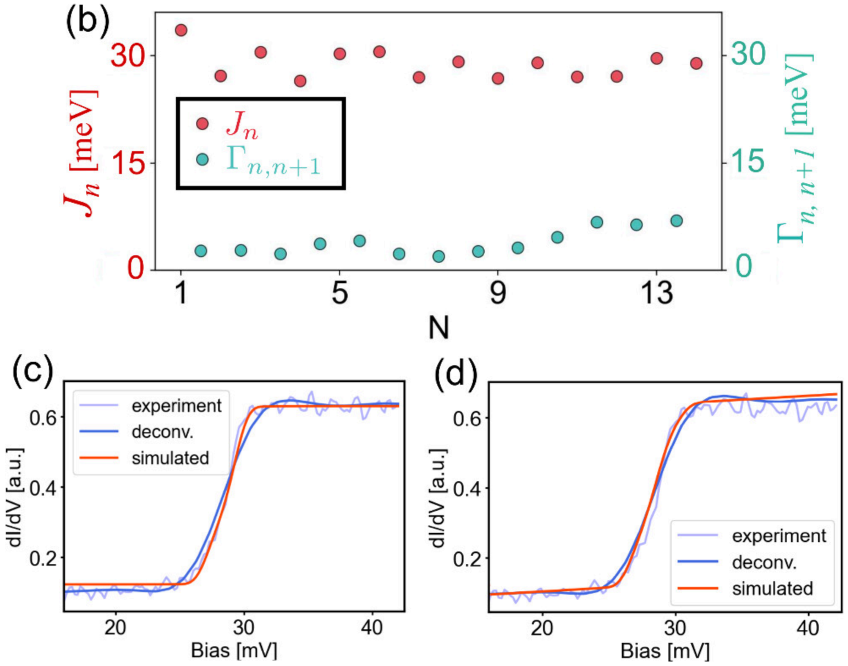


N. Karjalainen et al. – Phys. Rev. App. 20, 024054 (2023)

Example: Triplon Excitations

$$\hat{H} = \sum_n J_n \hat{\mathbf{S}}_n \cdot \hat{\mathbf{K}}_n + \sum_{\langle n,m \rangle} \Gamma_{nm} \hat{\mathbf{K}}_n \cdot \hat{\mathbf{K}}_m$$



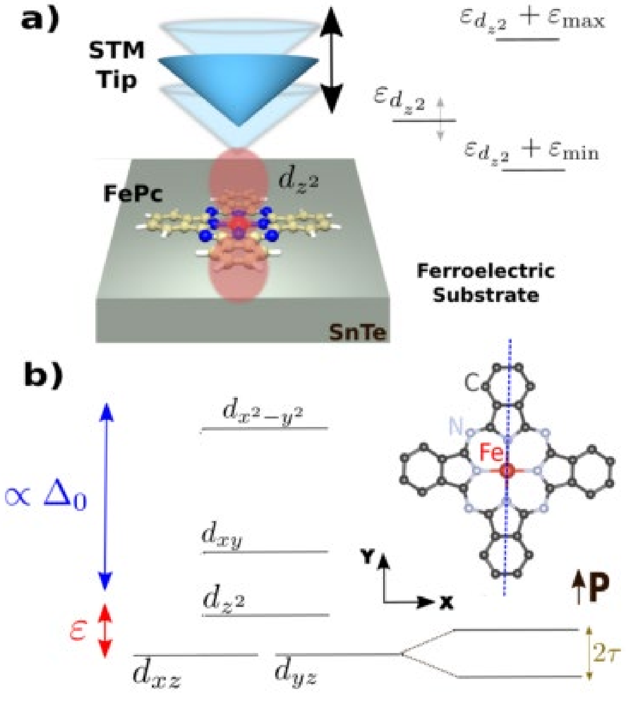


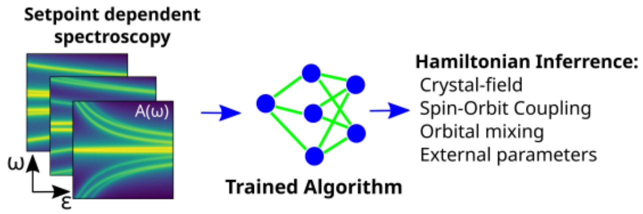
R. Koch et al. – Nano Lett. 25, 36, 13435-13440 (2025)

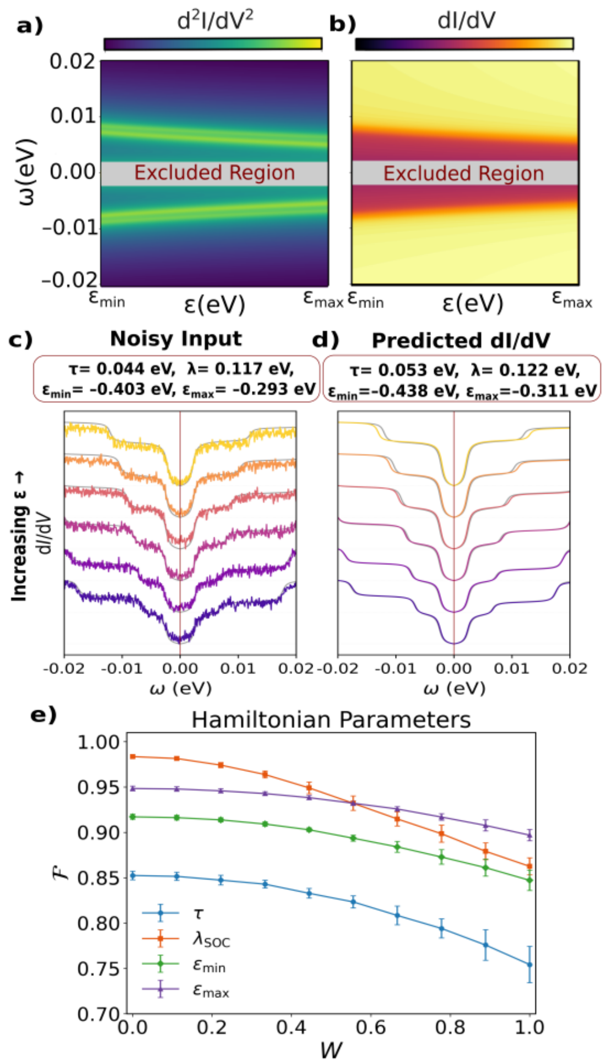
Example: Molecular H Learning

Aalto University
School of Science

$$\hat{H} = \hat{H}_{CF}(z) + \hat{H}_{\text{Coulomb}} + \hat{H}_{\text{SOC}}$$







G. Lupi et al. – arXiv 2601.19371 (2026)

Aalto University
School of Science

Supervised Learning Protocol

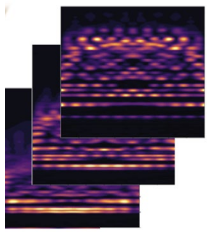
Input data

(x_1, y_1)

(x_2, y_2)

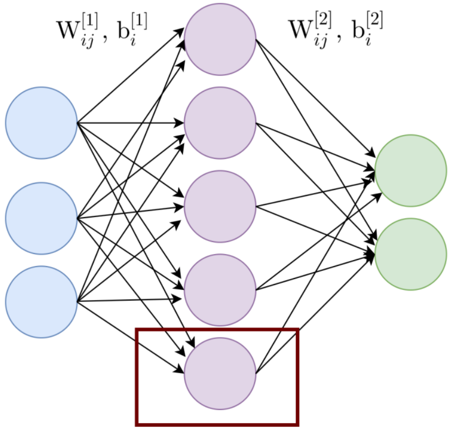
$\cdots$

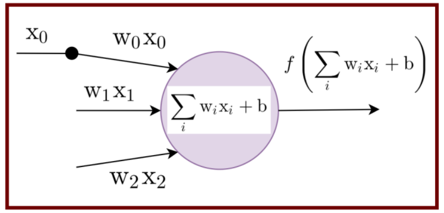
(x_n, y_n)



(J_1, J_2, J_3)

Neural Network





Output + optimization

$$f(x_n) \approx y_n$$

Minimize

$$R(X, \theta) \propto \sum_i \mathcal{L}\{y_i, f(x_i, \theta)\}$$

Data preprocessing important!

- Add noise to input
- Feature scaling -> all inputs have similar amplitudes
- Mean normalization -> all inputs have the same range
- Dimensionality reduction: PCA
- Etc …

- All help optimize learning

Hamiltonian learning: regression

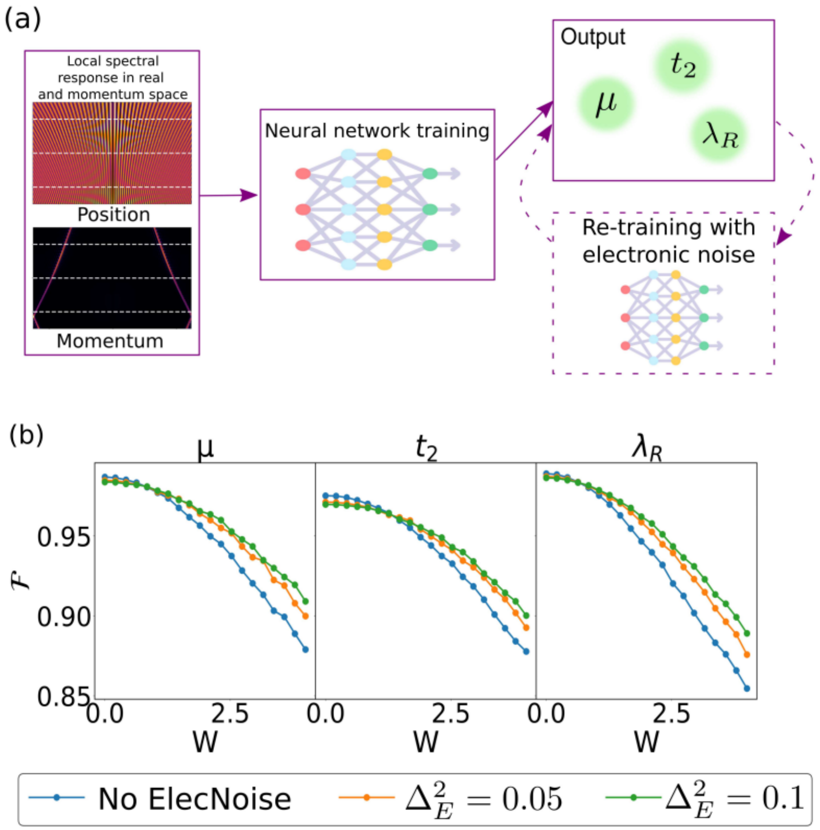
- Regression task: minimize MSE

$$\mathcal{L}\{y_i, f(x_i,\theta)\} = [y_i - f(x_i,\theta)]^2$$

- Add noise to input

- Fidelity

$$\mathcal{F}_{\Lambda_n} = \frac{\langle \Lambda_n^{\text{pred}} \Lambda_n^{\text{true}} \rangle - \langle \Lambda_n^{\text{pred}} \rangle \langle \Lambda_n^{\text{true}} \rangle}{\sqrt{\text{var}(\Lambda_n^{\text{pred}})}\sqrt{\text{var}(\Lambda_n^{\text{true}})}},$$



G. Lupi, J. Lado – Phys. Rev. App. 23, 054077 (2025)

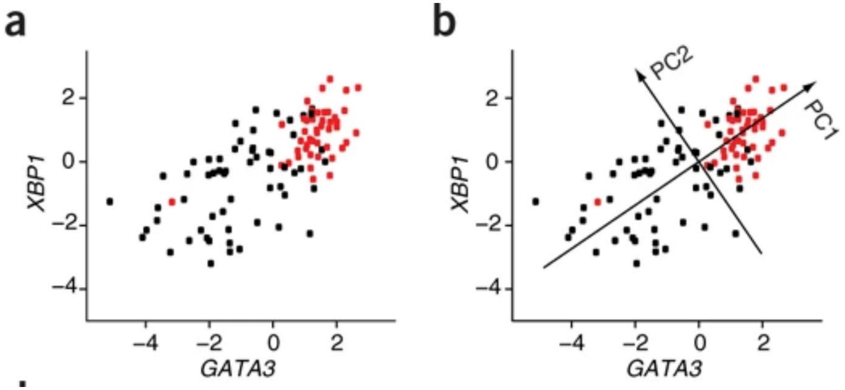
PCA

- Capture dimensions of maximum variance

$$C = \frac{1}{N} X^T X \quad \Rightarrow \quad Cv_i = \lambda_i v_i$$

v_i: $\mathrm{i^{th}}$ PC vector

λ_i: variance of v_i



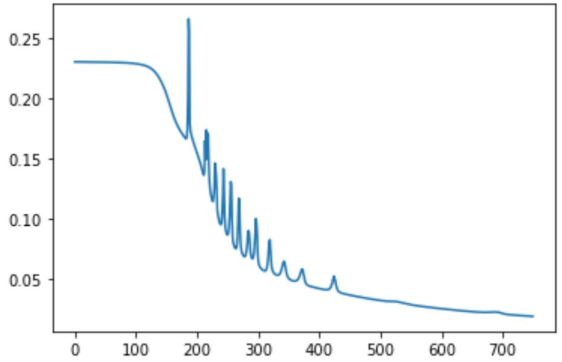
- Goal: drop PCV of the lowest k variance while minimizing error

M. Ringnér – Nat. Biotech. 26, 303-304 (2008)

Gradient descent – learning rate

Aalto University
School of Science

- Gradient descent: $\theta_i \to \theta_i - \alpha \nabla J$
- Repeat until convergence
- α is the learning rate: choose small enough but not too small
- Choose optimization algorithm
  - ADAM: updates α depending on ∇ size



ML good practice: model evaluation

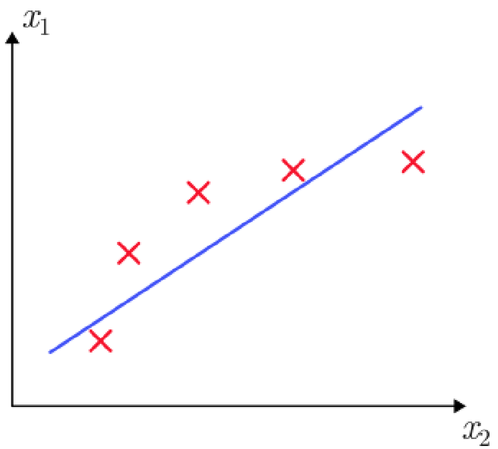
Aalto University School of Science

- Split data in
  - Training set: S_{tr}
  - Test set: $S_{\mathrm{te}} \text{ with } |S_{\mathrm{te}}| < |S_{\mathrm{tr}}|$
  - X-validation: $S_{\mathrm{te}} \to \tilde{S}_{\mathrm{te}} \cup S_{\mathrm{cv}}$
- Calculate cost functions $J_{\mathrm{tr}},\ J_{\mathrm{te}},\ J_{\mathrm{cv}}$
- Check that $J_{\mathrm{te}}/\ J_{\mathrm{tr}} < 1$: fraction of “misclassification”
- Check bias versus variance

ML good practice: bias versus variance

Aalto University School of Science

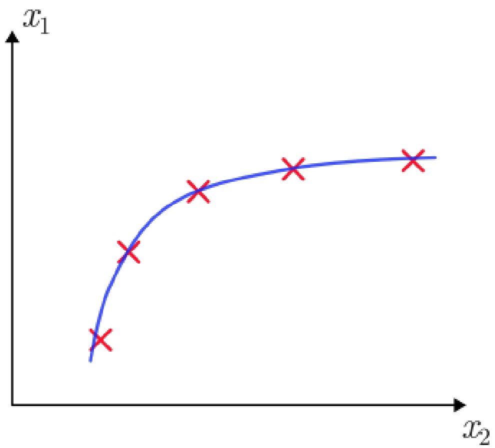
High bias (underfit)



J_{tr} high

J_{cv} high

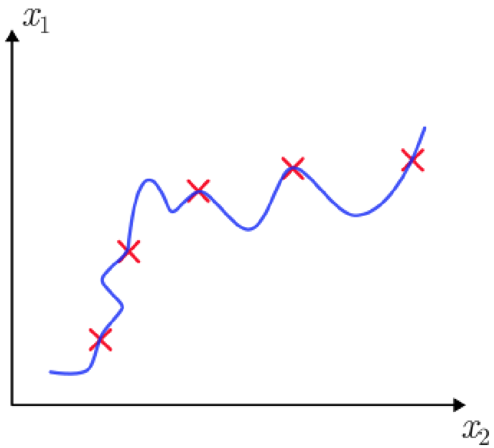
Just right



J_{tr} low

J_{cv} low

High variance (overfit)



J_{tr} low

J_{cv} high

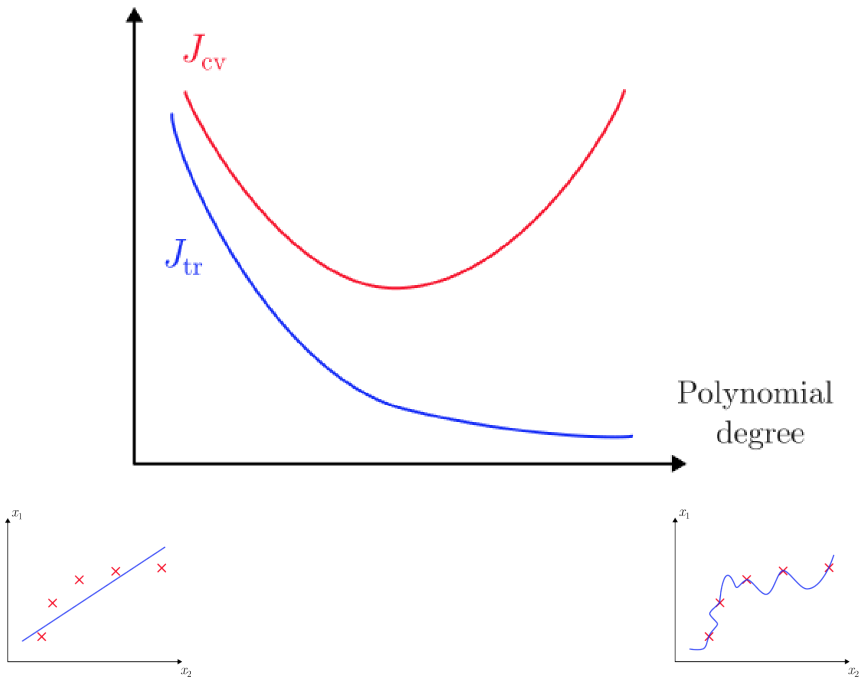
Gradient descent – regularization

- Overfitting: reduce weights in polynomial regression

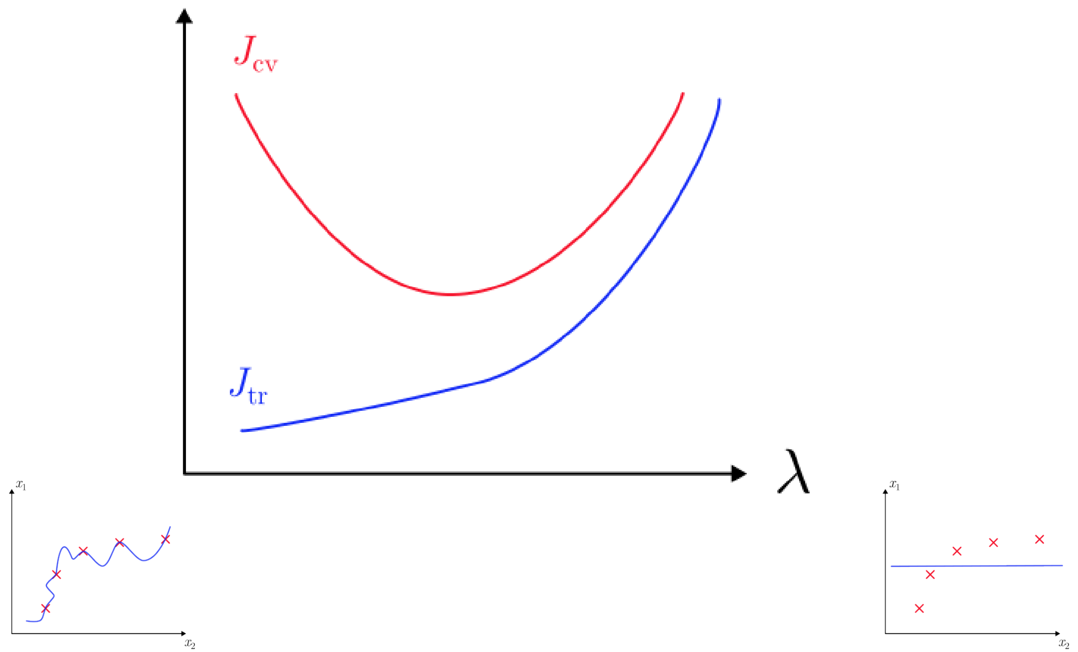
$$J \to J + \lambda \sum_i w_i^2$$

- Gradient descent: $w_i \to w_i(1 - 2\alpha\lambda) - \alpha\partial_w J$

ML good practice: bias versus variance

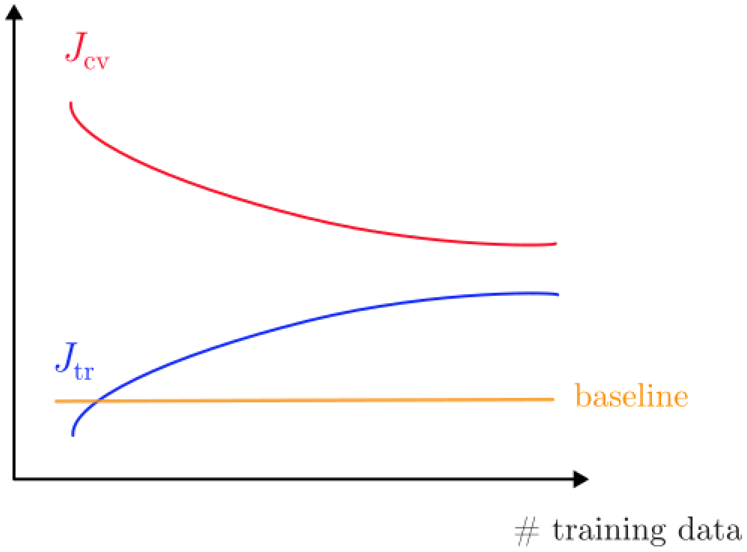


Regularization



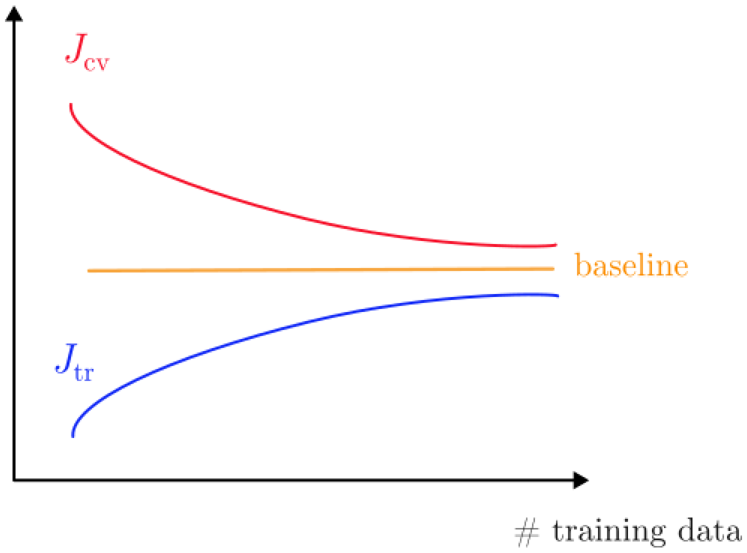
ML good practice: learning curves

High bias
(underfit)



More data can't help

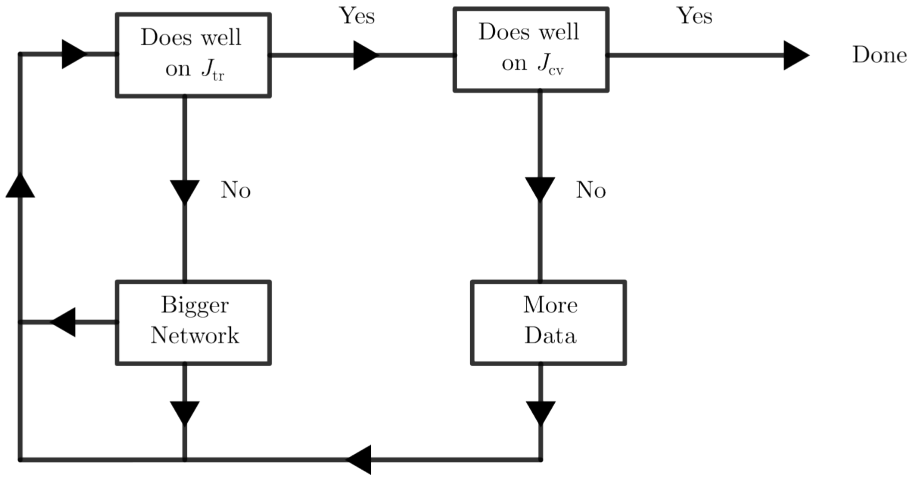
High variance
(overfit)



More data can help

ML good practice: Size of NN

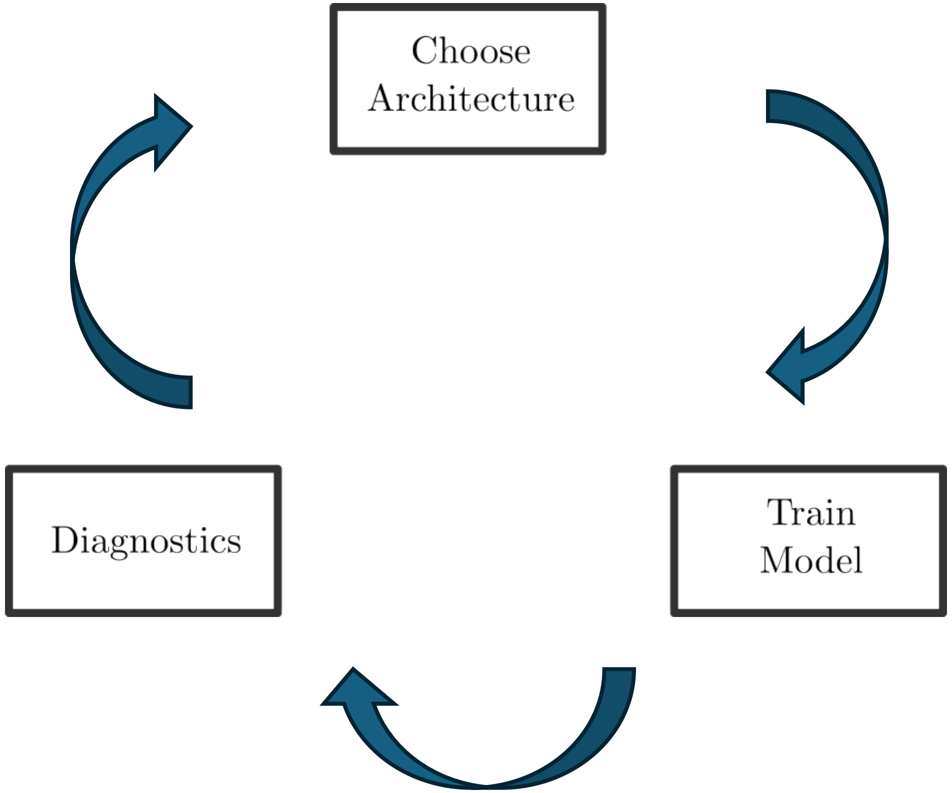
- Large NN are low bias machines



- Large NN do better with good regularization

ML good practice: development loop

- Make picture of development loop



Simulation tools: Physics

- DMRGpy: high-level library for many-body problems
  - Only need to know second quantization formalism
- Pyqula: high-level library for single-particle problems
- Can also use any library of your choice
- Can also build Hamiltonians from scratch (not recommended)

Simulation tools: ML

Aalto University
School of Science

- Scikit-learn
- Tensorflow - Keras
- Pytorch
- Jax

Aalto-yliopisto
Aalto-universitetet
Aalto University

PHYS-EV0007 - Machine Learning from and for Quantum Science

Topic 3: Neural Network Quantum States

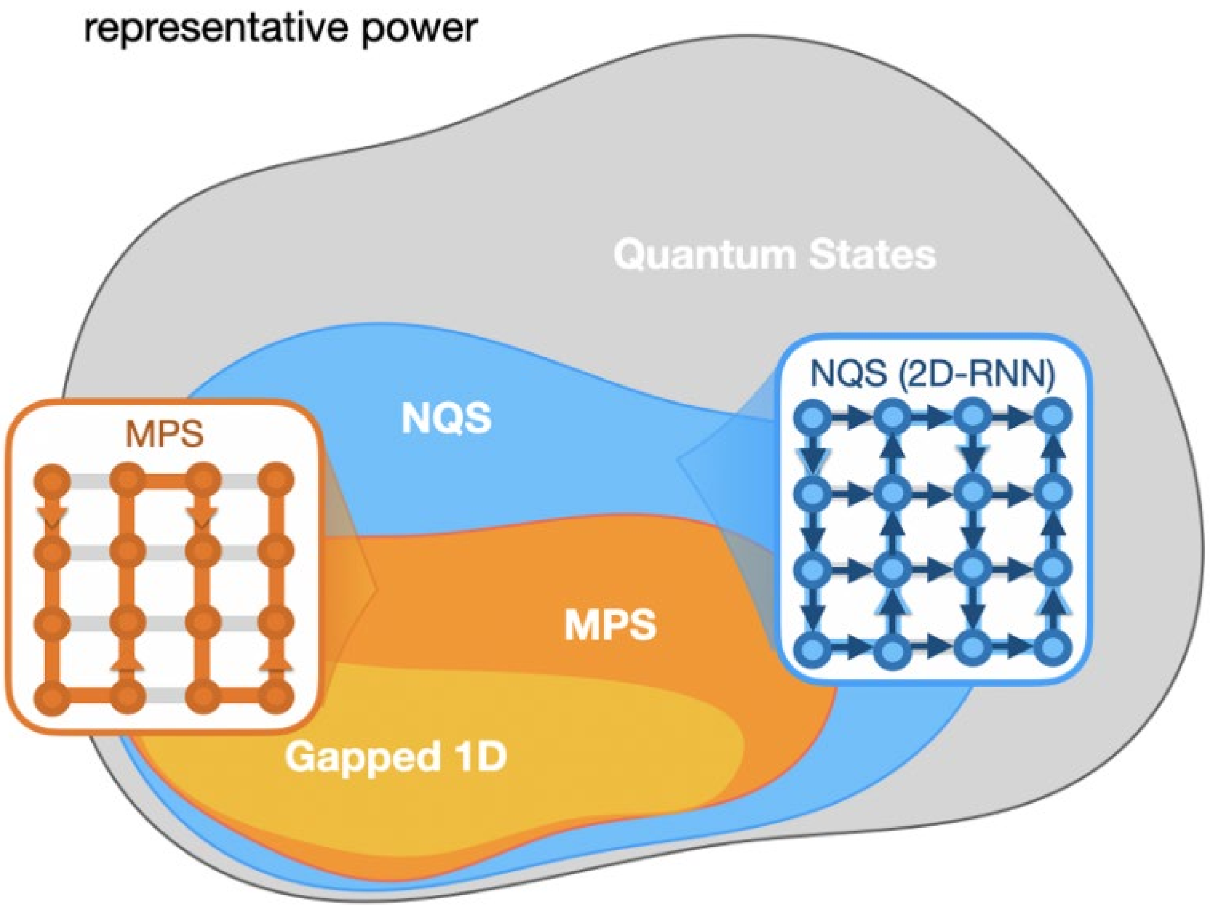
Content

1. Motivation
2. General Picture
3. Examples
4. Framework
5. Architectures

Motivation

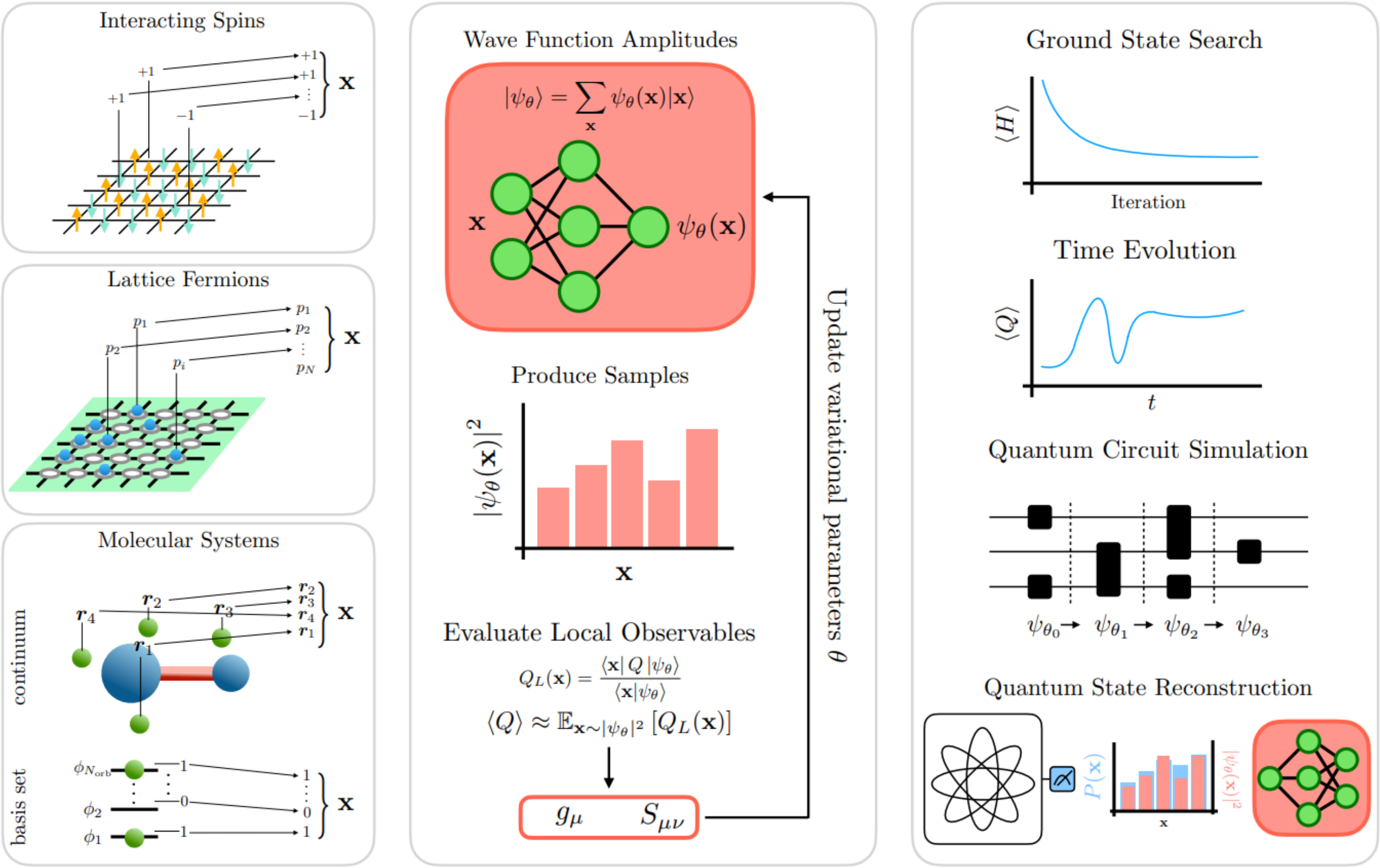
$$M^N$$

$$|\psi\rangle = \sum_{i_1 i_2 \cdots i_N} C_{i_1 i_2 \cdots i_N} |i_1 i_2 \cdots i_N\rangle$$



H. Lange et al. – Quantum Sci. Technol. 9, 040501 (2024)

General Picture

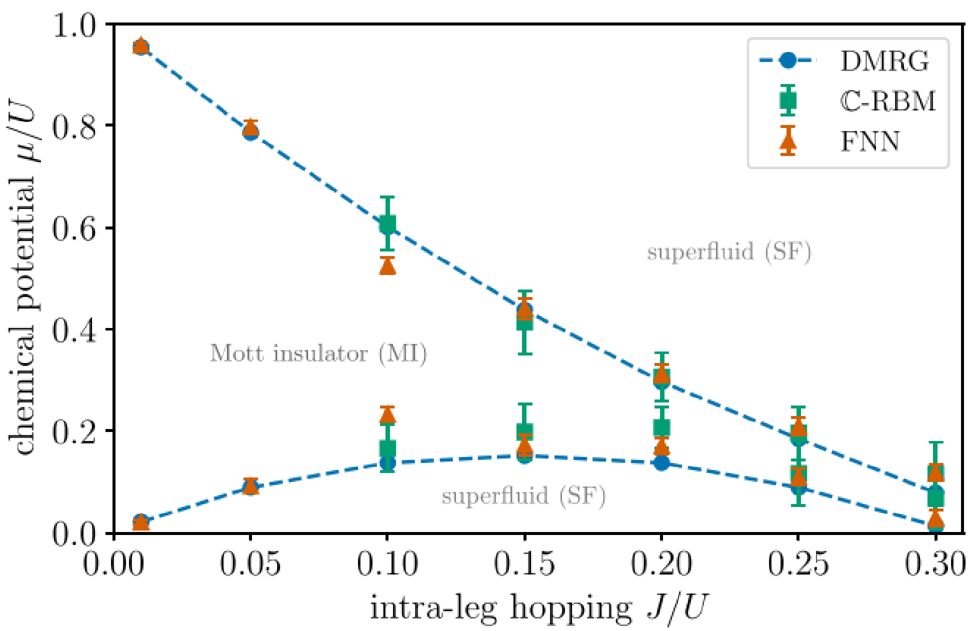


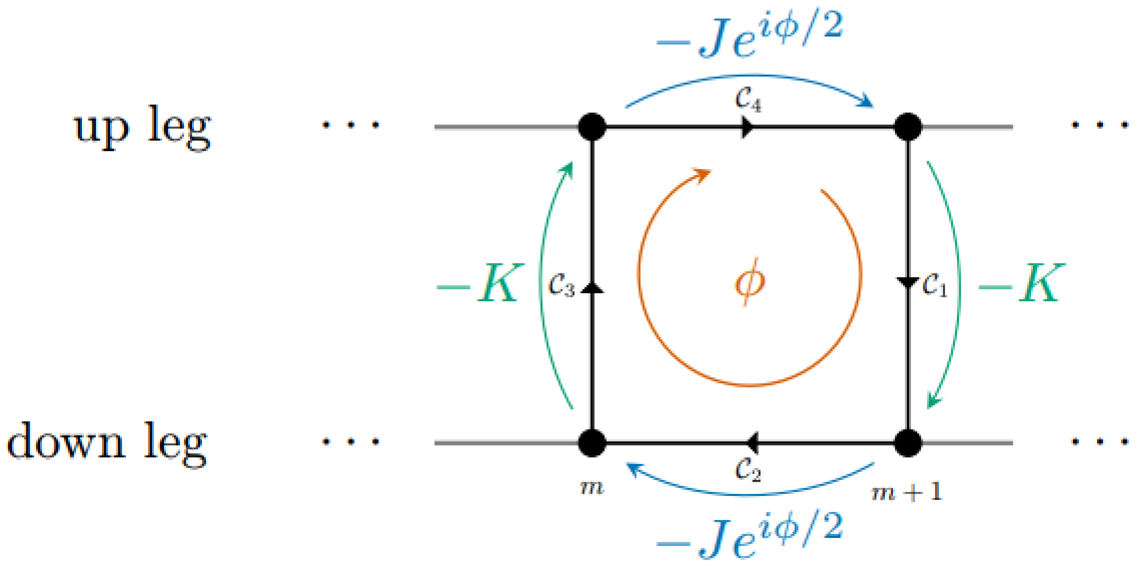
M. Medvidović et al. – Eur. Phys. J. Plus 139, 631 (2024)

Examples: Bose-Hubbard model

Aalto University
School of Science

$$H = -J \sum_{m=1}^{L} \sum_{\ell \in \{u,d\}} \left(e^{i\sigma_\ell \phi/2}\, a^\dagger_{\ell,m+1} a_{\ell,m} + \mathrm{H.c.} \right) - K \sum_{m=1}^{L} \left(a^\dagger_{u,m} a_{d,m} + \mathrm{H.c.} \right) + \sum_{m=1}^{L} \sum_{\ell \in \{u,d\}} \left[\frac{U}{2} n_{\ell,m}(n_{\ell,m} - 1) - \mu n_{\ell,m} \right]$$

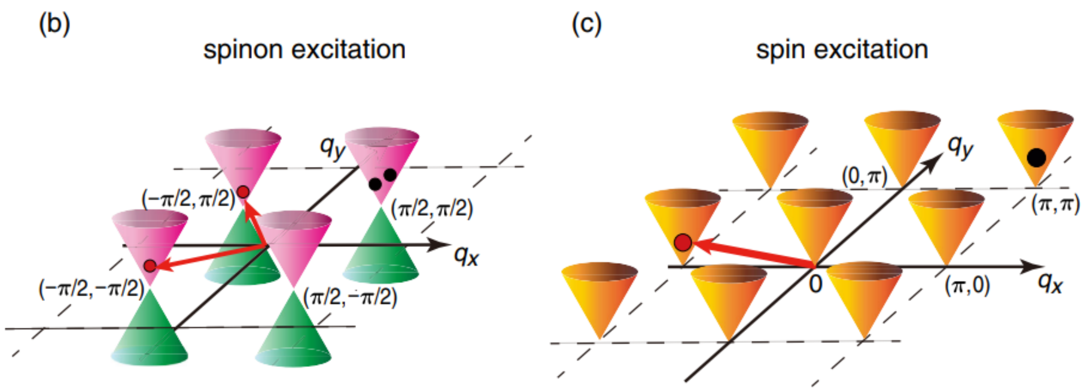


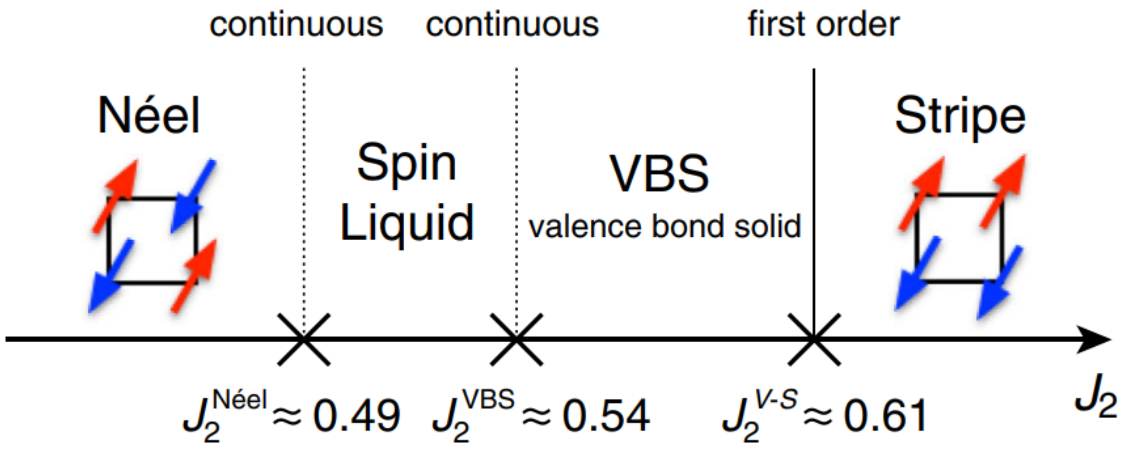


K. Çeven et al. – Phys. Rev. A 106, 063320 (2022)

Examples: Heisenberg Models

$$\hat{H} = J_1 \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j + J_2 \sum_{\langle\langle i,j \rangle\rangle} \mathbf{S}_i \cdot \mathbf{S}_j$$

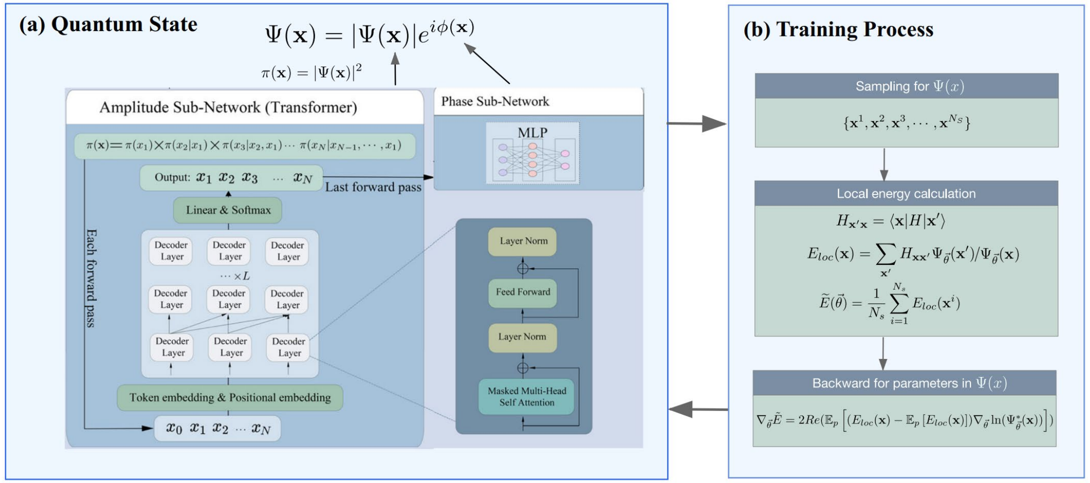


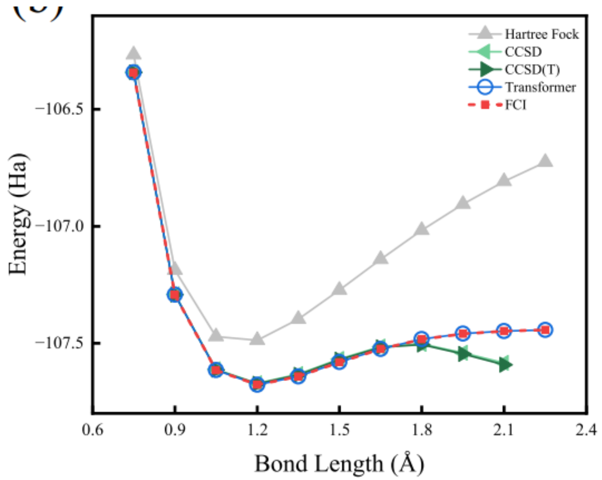


Y. Nomura et al. – Phys. Rev. X 11, 031034 (2021)

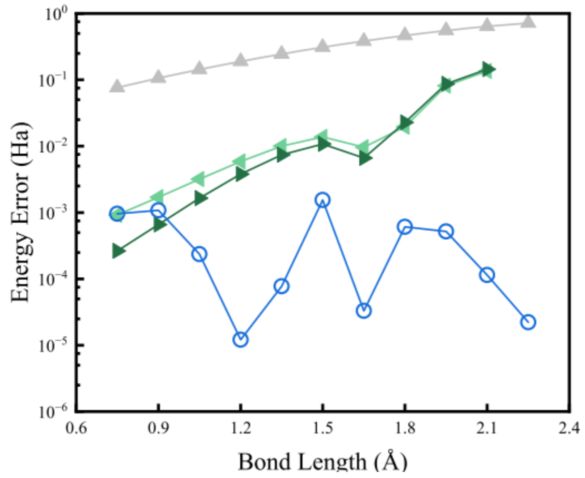
Examples: quantum chemistry

$$\hat{H}^e = \sum_{p,q} h_q^p \, \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{p,q,r,s} g_{r,s}^{p,q} \, \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s$$





N2 molecule

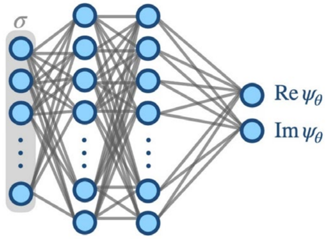


H. Shang et al. – Nat. Comm. 16, 8464 (2025)

Framework

- Aim: Find most optimal many body state psi on a variational manifold

$$|\psi_\theta\rangle = \sum_{\mathbf{x}} \psi_\theta(\mathbf{x}) \, |\mathbf{x}\rangle$$



- Optimization depends on problem: Typically finding GS of Hamiltonian

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

- Let's work with that to show how it works

Local Observable Trick

Aalto University
School of Science

$$E_\theta = \frac{\sum_{\mathbf{x}} \langle \psi_\theta | \mathbf{x} \rangle \langle \mathbf{x} | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \sum_{\mathbf{x}} p_\theta(\mathbf{x}) H_\theta(\mathbf{x})$$

$$p_\theta(\mathbf{x}) = \frac{|\psi_\theta(\mathbf{x})|^2}{\sum_{\mathbf{x}} |\psi_\theta(\mathbf{x})|^2} \qquad H_\theta(\mathbf{x}) = \frac{\langle \mathbf{x} | \hat{H} | \psi_\theta \rangle}{\langle \mathbf{x} | \psi_\theta \rangle}$$

Sampling

Aalto University
School of Science

- Impossible to calculate the full EV. Sample:

$$E_\theta \approx \frac{1}{N_s} \sum_{\mathbf{x}_i} H_\theta(\mathbf{x}_i) \equiv \langle\langle \hat{H} \rangle\rangle_\theta$$

Sampling: Monte Carlo

- Impossible to calculate the full EV. Sample:

$$E_\theta \approx \frac{1}{N_s} \sum_{\mathbf{x}_i} H_\theta(\mathbf{x}_i) \equiv \langle\langle \hat{H} \rangle\rangle_\theta$$

- E.G. Metropolis

$$p_{\text{accept}} = \min\left(1,\, p(\mathbf{x}')/p(\mathbf{x})\right)$$

Sampling: Autoregressive

- Impossible to calculate the full EV. Sample:

$$E_\theta \approx \frac{1}{N_s} \sum_{\mathbf{x}_i} H_\theta(\mathbf{x}_i) \equiv \langle\langle \hat{H} \rangle\rangle_\theta$$

$$p(\mathbf{x}) = p(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \prod_i p(\mathbf{x}_i | \mathbf{x}_{<i}) =$$
$$= p(\mathbf{x}_1)\, p(\mathbf{x}_2 | \mathbf{x}_1) \,\cdots\, p(\mathbf{x}_N | \mathbf{x}_{N-1}, \ldots, \mathbf{x}_2, \mathbf{x}_1)$$

1. Sample $\mathbf{x}_i \sim p(\mathbf{x}_i | \mathbf{x}_{<i})$ and concatenate to $\mathbf{x}_{<i}$.
2. Use samples to define $p(\mathbf{x}_{i+1} | \mathbf{x}_{<i+1})$.

VMC: Stochastic Gradient Descent

Aalto University
School of Science

- Optimize using variational principle:

$$\frac{\langle\psi_\theta|\,\hat{H}\,|\psi_\theta\rangle}{\langle\psi_\theta|\psi_\theta\rangle} = E_\theta \geq E_0$$

$$\nabla_\theta E_\theta \approx \nabla_\theta \langle\langle \hat{H} \rangle\rangle_\theta \quad \Longrightarrow \quad |\psi_{\theta+\delta\theta}\rangle \propto e^{-\delta\tau \hat{H}}\,|\psi_\theta\rangle$$

$$|\mathrm{GS}\rangle \propto \lim_{\tau\to\infty} e^{-\tau \hat{H}}\,|\psi_\theta\rangle$$

M. Medvidović et al. – Eur. Phys. J. Plus 139, 631 (2024)

Stochastic Quantum Geometry Picture

$$|\psi_{\theta+\delta\theta}\rangle \propto e^{-\delta\tau \hat{H}} |\psi_\theta\rangle \quad \Longrightarrow \quad S_{\mu\nu}\frac{\mathrm{d}\theta^\nu}{\mathrm{d}\tau} = g_\mu$$

$$S_{\mu\nu} = 2\mathrm{Re}\left[\langle\langle \hat{O}_\mu \hat{O}_\nu \rangle\rangle_\theta - \langle\langle \hat{O}_\mu \rangle\rangle_\theta \langle\langle \hat{O}_\nu \rangle\rangle_\theta\right]$$

Quantum Metric Tensor

$$g_\mu = 2\mathrm{Re}\left[\langle\langle \hat{O}_\mu \hat{H} \rangle\rangle_\theta - \langle\langle \hat{O}_\mu \rangle\rangle_\theta \langle\langle \hat{H} \rangle\rangle_\theta\right]$$

Stochastic gradient $\nabla_\theta E_\theta$

$$\partial_\mu |\psi_\theta\rangle = \hat{O}_\mu |\psi_\theta\rangle \quad \Longrightarrow \quad \hat{O}_\mu = \sum_{\mathbf{x}} \partial_\mu \psi(\mathbf{x}) |\mathbf{x}\rangle \langle \mathbf{x}|$$

M. Medvidović et al. – Eur. Phys. J. Plus 139, 631 (2024)

Aalto University
School of Science

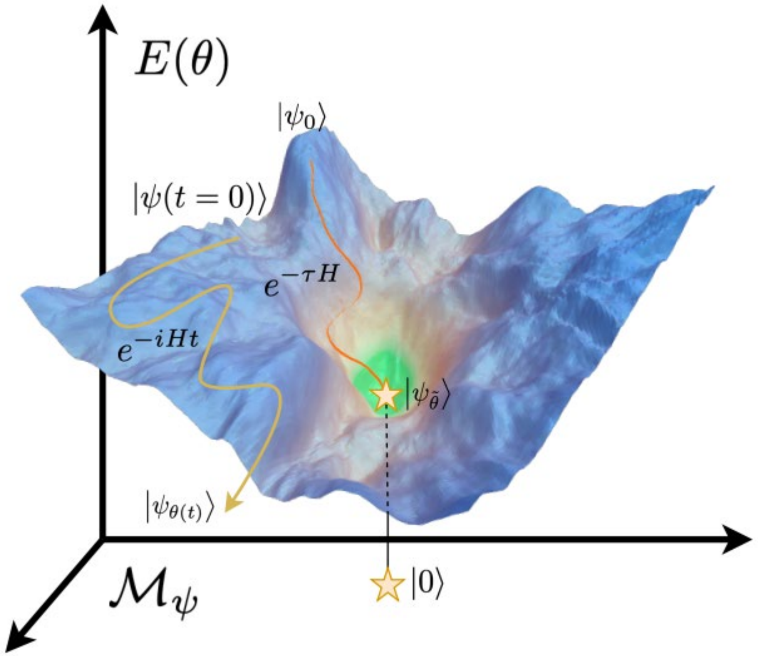
Stochastic Quantum Geometry Picture

Aalto University
School of Science

- Quantum Metric Tensor measures infinitesimal overlap on $\mathcal{M}_\psi$
- Aka Quantum Fisher Information:

$$F(|\psi_\theta\rangle, |\psi_{\theta+\delta\theta}\rangle) = \frac{|\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle|^2}{\langle\psi_\theta|\psi_\theta\rangle\langle\psi_{\theta+\delta\theta}|\psi_{\theta+\delta\theta}\rangle}$$

$$= 1 - \frac{1}{2}S_{\mu\nu}\delta\theta^\mu\delta\theta^\nu + \mathcal{O}\left(\delta\theta^3\right)$$



M. Medvidović et al. – Eur. Phys. J. Plus 139, 631 (2024)

Different Approximations to $S\frac{\mathrm{d}\theta}{\mathrm{d}\tau} = g$

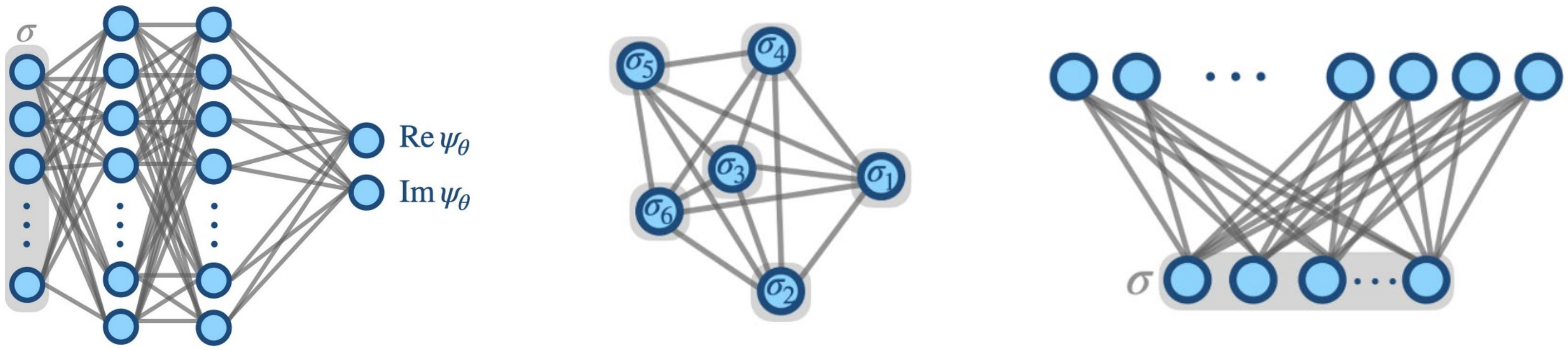
- Quantum Natural Gradient (QNG): $\theta_{i+1} = \theta_i - \delta\tau S^{-1} g$
- Kronecker-factored approximate curvature (KFAC)
- Stochastic Gradient Descent (SGD) : $S \approx \hat{1} \Longrightarrow \theta_{i+1} = \theta_i - \alpha \nabla_\theta E_\theta$
- $S = \hat{1} \Longrightarrow$ regular GD with ADAM or other optimizers

More info on geometric picture: L Hackl et al. – Scipost Phys. 9, 048 (2020)

Neural Network Quantum States

- W.F. is the output of a LNN with L parameters:

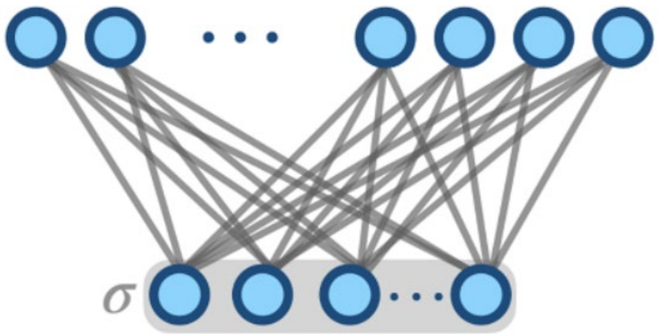
$$\psi_\theta(\mathbf{x}) \equiv F(\mathbf{x}; \theta_1, \cdots, \theta_L)$$



H. Lange et al. – Quantum Sci. Technol. 9, 040501 (2024)

Aalto University
School of Science

Architectures: RBM



- First widely used architecture for spins

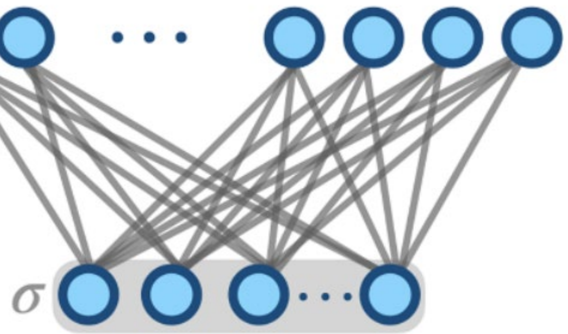
$$F(\boldsymbol{\sigma};\boldsymbol{\theta}) = \sum_{\{\mathbf{h}\}} \exp\left(\sum_{ij} W_{ij}\sigma_i^z h_j + \sum_i \sigma_i^z a_i + \sum_i h_i b_i\right)$$

Hidden spins

Parameters

- Restriction: Connections are only between physical spins and hidden units
- Partition function of hidden spins

Architectures: RBM

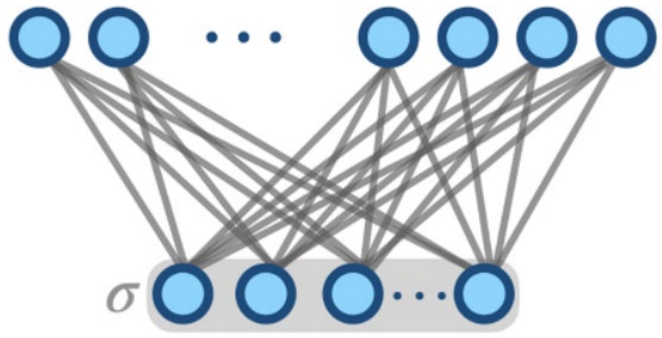


- Advantage: explicit form

$$F(\boldsymbol{\sigma};\boldsymbol{\theta}) = \exp\left\{\sum_i a_i \sigma_i^z\right\} \prod_j 2\cosh\left(\sum_i W_{ij}\sigma_i^z + b_j\right)$$

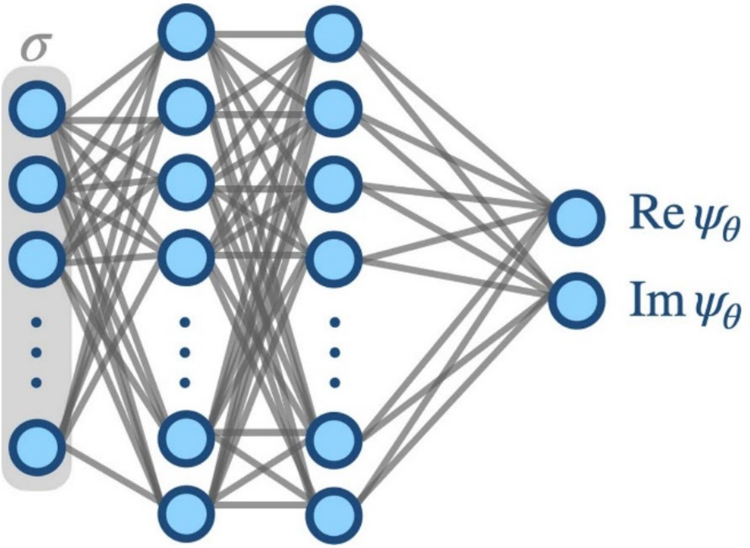
- Variational energy efficiently computed in polynomial time

Architectures: RBM



- Spin systems
- Topological States
- Bosons
- Fermions
- Molecules

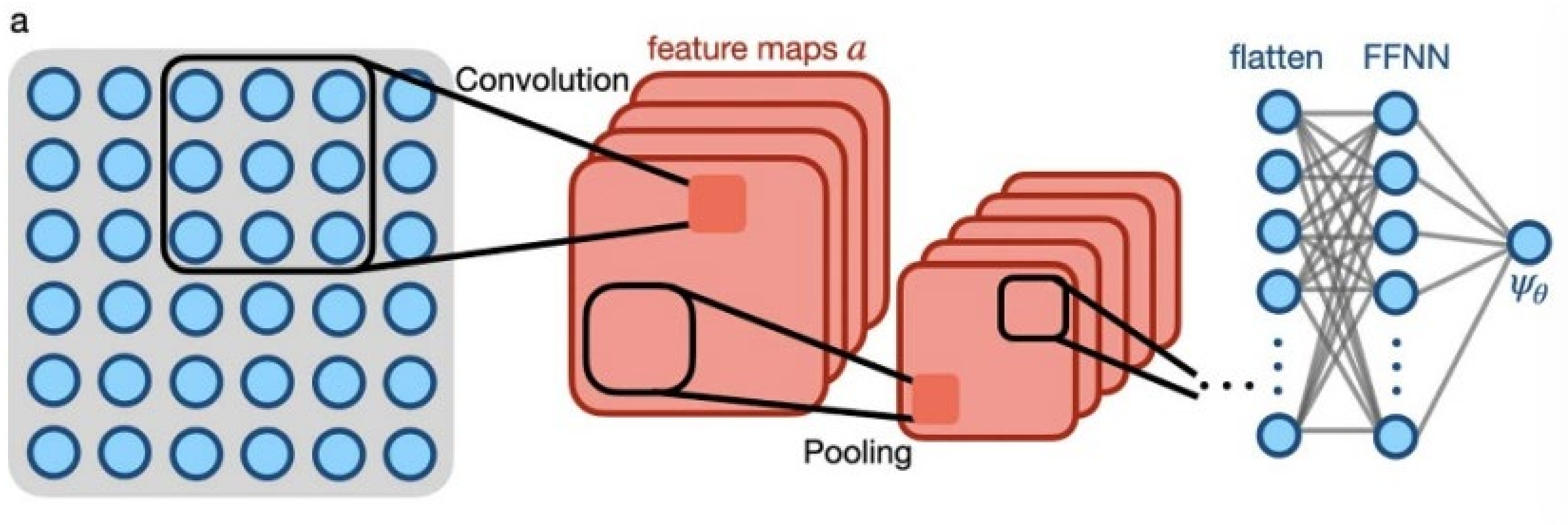
Architectures: other



Feed Forward
Neural Network

- Frustrated Spin systems
- Bosonic systems

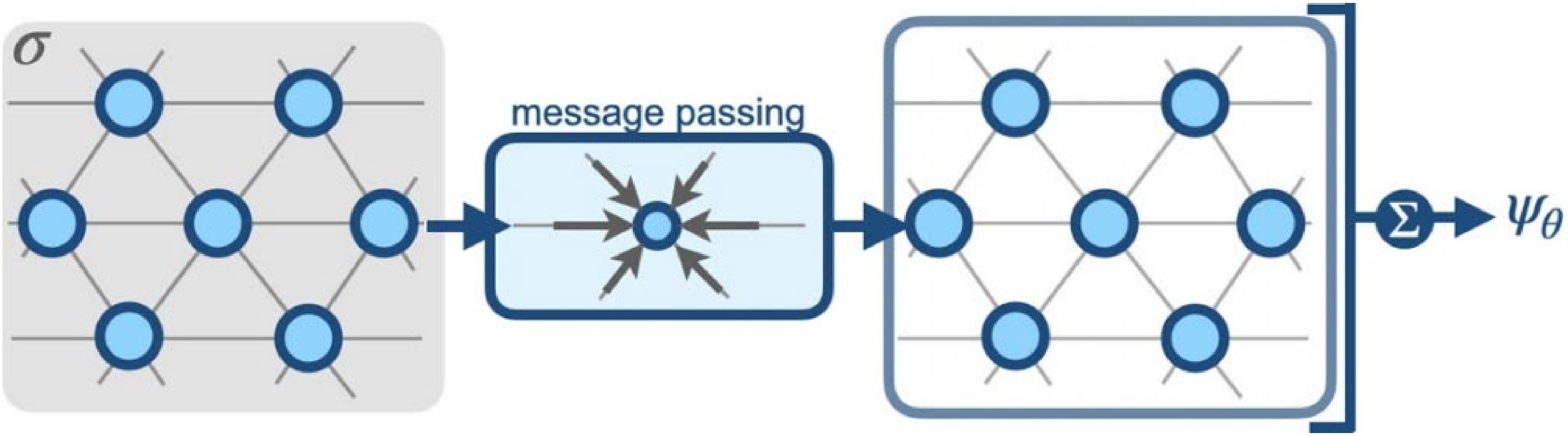
Architectures: other



Convolutional Neural Network

- Frustrated Spin systems
- Lattice symmetries incorporated

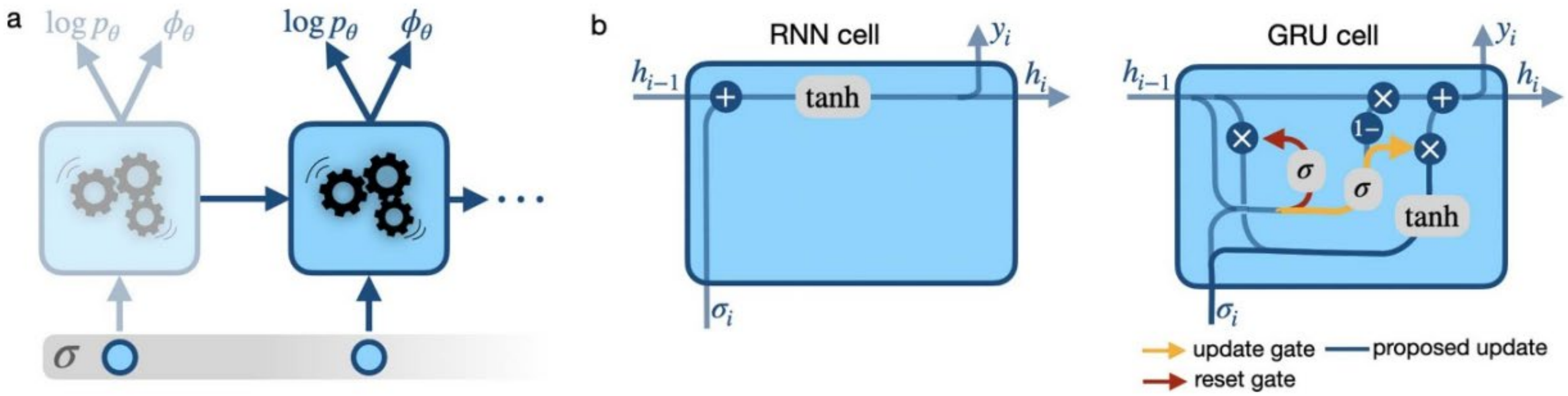
Architectures: other



Graph Neural Network

- Spin systems
- Bosonic systems
- Any lattice geometry

Architectures: other

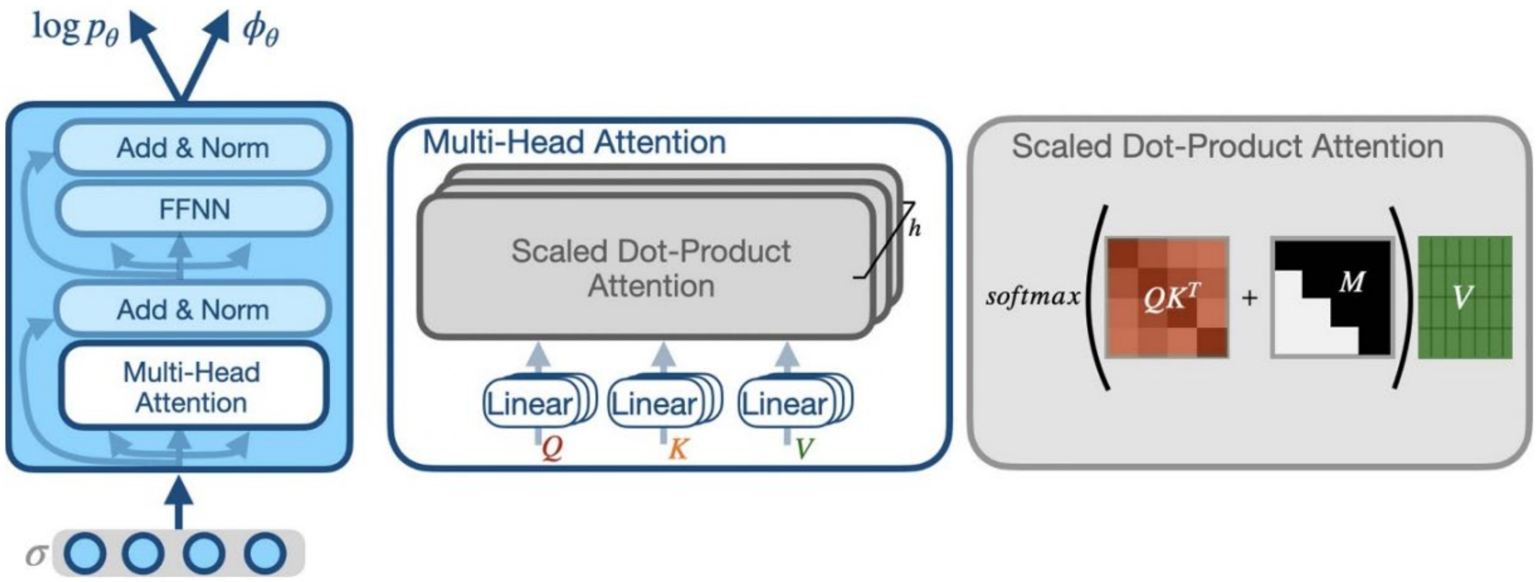


Recurrent
Neural Network

- Fermions
- Bosons
- Spin glass

- Spin systems
- Topological Systems
- Rydberg states

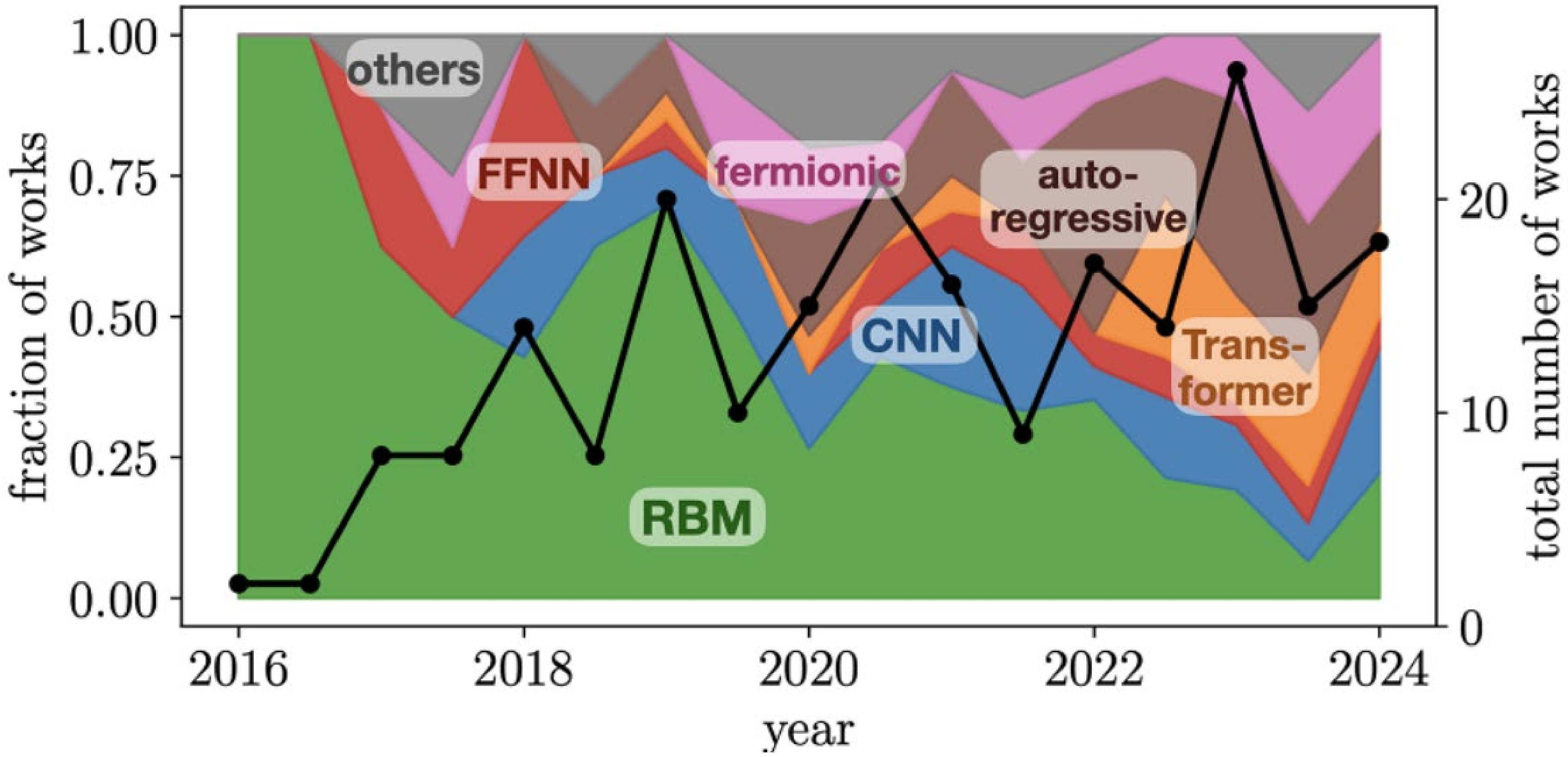
Architectures: other



Transformers

- Spin systems
- Bosonic systems

Architecture popularity



H. Lange et al. – Quantum Sci. Technol. 9, 040501 (2024)

Open Source Toolbox

- NetKet (GS search, dynamics, tomography)
- jVMC (efficient VMC, GS search, dynamics)
- FermiNet (GS search for atoms and molecules)
- QuCumber (RBM-based tomography)