Title: Talk - tbc

Speakers: Stefanie Czischek

Collection: Machine Learning for Quantum Many-Body Systems

Date: June 15, 2023 - 11:15 AM

URL: https://pirsa.org/23060046

Variational Monte Carlo with Large, Patched Transformers

Stef Czischek

June 15, 2023

Transformers: Natural language processing

English (detected) Swedish

Text translation

Översättning av text

Alternatives:

Översättning av texter

Textöversättning

DeepL Translator

[https://www.deepl.com/translator#en/sv/Text translation]

If you have any questions about my rate or if you find it necessary to increase or decrease the scope for this project, please let me know.

Clarity · Rephrase sentence

If you have any questions about my rate or ~~if you~~ find it necessary to increase or decrease ~~the scope~~ ~~for~~ this ~~project~~ project's scope, please let me know.

Rephrase Dismiss

grammarly

[https://www.grammarly.com/]

ChatGPT

Great move! Here's the current state of the Tic-Tac-Toe board:

```markdown
X |   | O
---------
  | O |
---------
X |   | X
```

It's my turn now, I will place an O in the bottom-left square. The board now looks like this:

```markdown
X |   | O
---------
  | O |
---------
O |   | X
```

It's your turn again. Which square would you like to place your X on?

[https://openai.com/product/chatgpt]

Stefanie.Czischek@uOttawa.ca

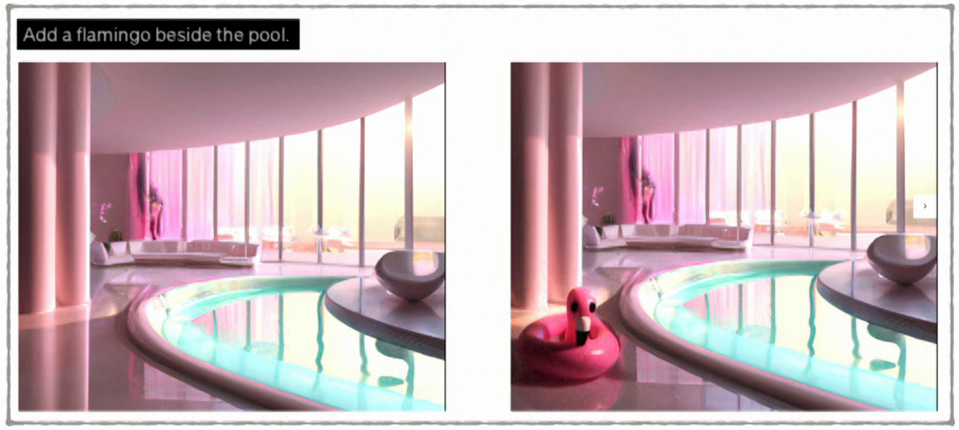
uOttawa

Transformers: Image generation/processing

DALL·E 2

DALL·E 2 is an AI system that can create realistic images and art from a description in natural language.

[https://openai.com/product/dall-e-2]



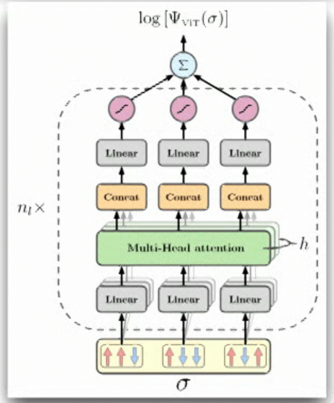
Stefanie.Czischek@uOttawa.ca

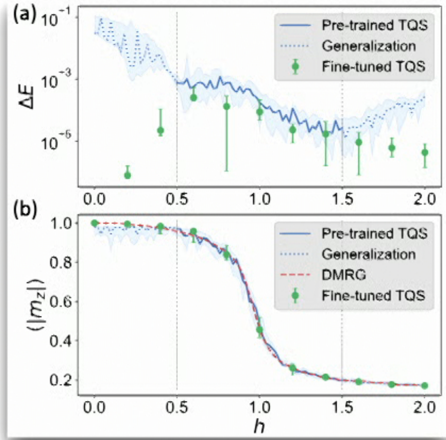
uOttawa

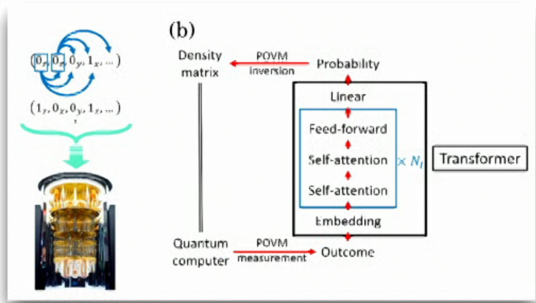
Transformers in quantum many-body physics

Towards Neural Variational Monte Carlo That Scales Linearly with System Size
Or Sharir, Garnet Kin-Lic Chan, Anima Anandkumar

Transformer variational wave functions for frustrated quantum spin systems
Luciano Loris Viteritti, Riccardo Rende, Federico Becca







Attention-based quantum tomography
Peter Cha[6,1], Paul Ginsparg[2], Felix Wu[2], Juan Carrasquilla[3,4], Peter L McMahon[5] and Eun-Ah Kim[1]
Published 23 November 2021 • © 2021 The Author(s). Published by IOP Publishing Ltd
Machine Learning: Science and Technology, Volume 3, Number 1
Citation Peter Cha *et al* 2022 *Mach. Learn.: Sci. Technol.* 3 01LT01
DOI 10.1088/2632-2153/ac362b

Attention-Based Transformer Networks for Quantum State Tomography
Hailan Ma, Zhenhong Sun, Daoyi Dong, Chunlin Chen, Herschel Rabitz

Transformer quantum state: A multipurpose model for quantum many-body problems
Yuan-Hang Zhang and Massimiliano Di Ventra
Phys. Rev. B 107, 075147 – Published 22 February 2023

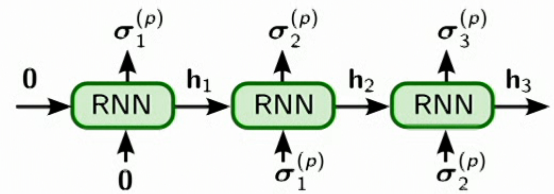
Unified Quantum State Tomography and Hamiltonian Learning Using Transformer Models: A Language-Translation-Like Approach for Quantum Systems
Zheng An, Jiahui Wu, Muchun Yang, D. L. Zhou, Bei Zeng

Stefanie.Czischek@uOttawa.ca

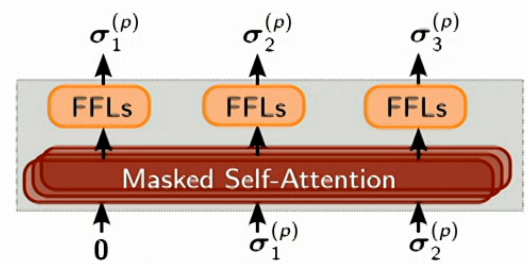
uOttawa

Transformers for variational Monte Carlo

- Variational Monte Carlo: ground state search for a given Hamiltonian
- Recurrent neural networks as a powerful wavefunction ansatz
  [Hibat-Allah et al., Phys Rev Res **2** (2020)]
  - Limitations for long-range correlations



- Can transformers do better?
  - Attention mechanism can describe correlations
    [Vaswani et al., arXiv:1706.03762 (2017)]



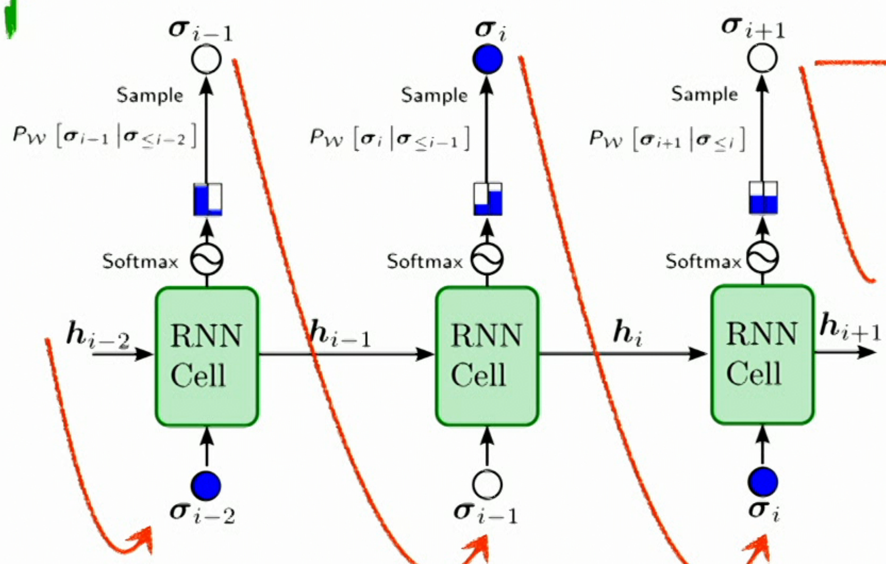
[SC, K. Sprague, arXiv:2306.03921 (2023)]

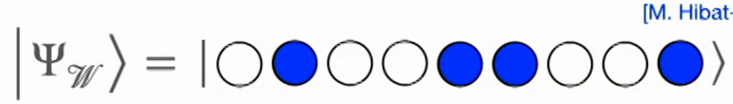
Stefanie.Czischek@uOttawa.ca

uOttawa

Recurrent neural network quantum states

[O. Sharir et al., PRL **124** (2020)]
[M. Hibat-Allah et al., PRR **2** (2020)]





- RNN cell: N_{h} hidden units, variational parameters $\mathscr{W}$
- RNN encodes squared wavefunction amplitudes
- Qubit samples: projective measurements
- Training: minimize energy expectation value $\langle E \rangle$

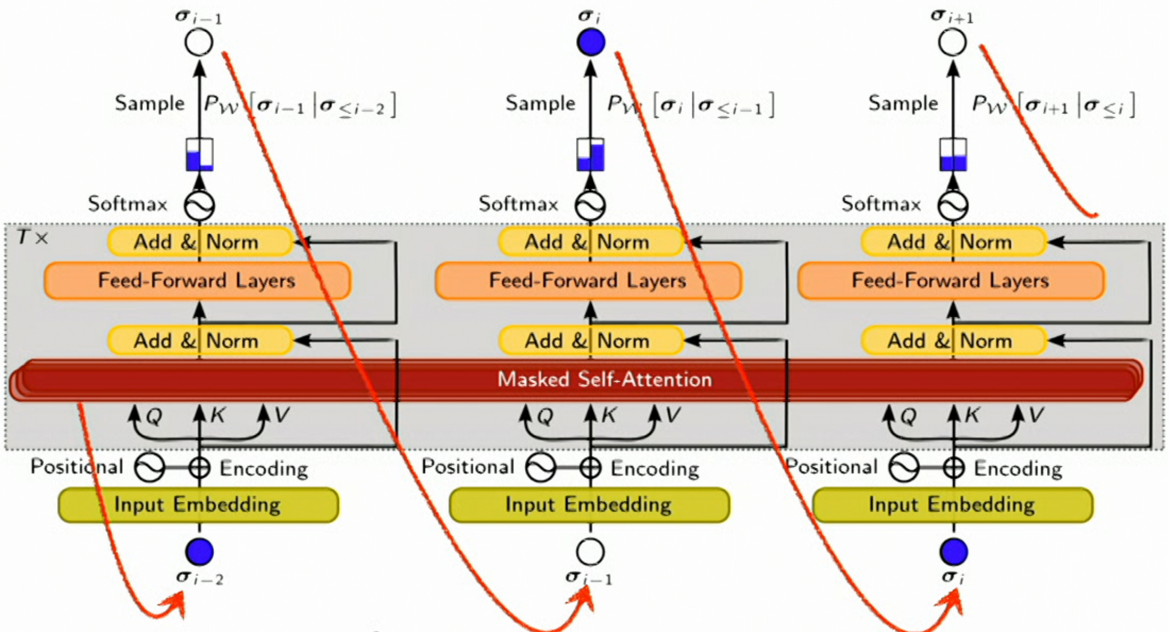
$$\left|\Psi(\boldsymbol{\sigma})\right|^2 \approx \left|\Psi_{\mathscr{W}}(\boldsymbol{\sigma})\right|^2 = p_{\mathrm{RNN}}(\boldsymbol{\sigma};\mathscr{W}) = \prod_i p_{\mathrm{RNN}}(\sigma_i\,|\,\sigma_{i-1},\ldots,\sigma_1;\mathscr{W})$$

$$\langle E \rangle = \sum_{\boldsymbol{\sigma}} \left|\Psi_{\mathscr{W}}(\boldsymbol{\sigma})\right|^2 H_{\mathrm{loc}}(\boldsymbol{\sigma}) \approx \frac{1}{N_{\mathrm{s}}} \sum_{\boldsymbol{\sigma}\sim p_{\mathrm{RNN}}(\boldsymbol{\sigma};\mathscr{W})} H_{\mathrm{loc}}(\boldsymbol{\sigma}) \qquad H_{\mathrm{loc}}(\boldsymbol{\sigma}) = \frac{\langle \boldsymbol{\sigma}\,|\,\hat{H}\,|\,\Psi_{\mathscr{W}}\rangle}{\langle \boldsymbol{\sigma}\,|\,\Psi_{\mathscr{W}}\rangle}$$

Stefanie.Czischek@uOttawa.ca

uOttawa

Transformer quantum states

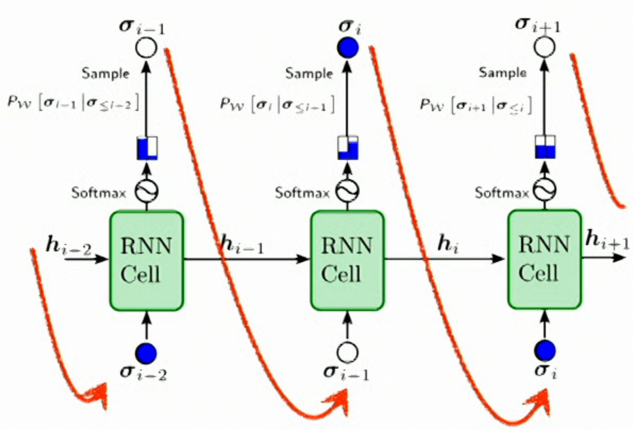


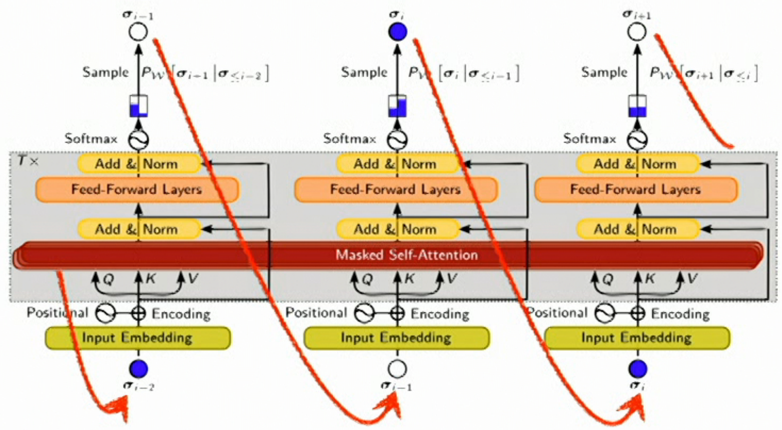
$$\left|\Psi(\boldsymbol{\sigma})\right|^2 \approx p_{\mathrm{TF}}(\boldsymbol{\sigma};\mathcal{W}) = \prod_i P_{\mathcal{W}}[\sigma_i|\sigma_{\leq i-1}]$$

- Attention: trained connections to all elements
- Mask: only previous elements get influence
- Positional encoding: encode sequence order
- Wave function encoded similar to RNN

Stefanie.Czischek@uOttawa.ca

RNN vs. Transformers





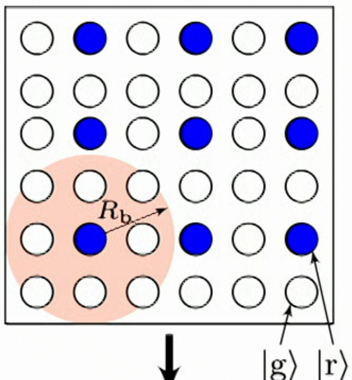
- Recurrent: information encoded in hidden state
- Attention: direct access to whole sequence

How does this feature affect quantum state representations?

Stefanie.Czischek@uOttawa.ca

uOttawa

Rydberg atom arrays

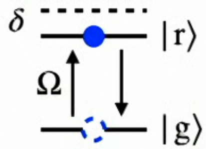


Projective measurement

$|\boldsymbol{\sigma}\rangle = |\mathrm{g\ r\ g} \ldots \mathrm{g\ g}\rangle$

$N = L \times L$

atoms on square lattice



$\hat{\sigma}_i^x = |g\rangle_i\langle r|_i$

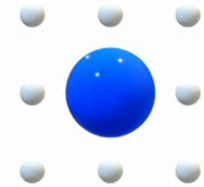
$\hat{n}_i = |r\rangle_i\langle r|_i$

Stoquastic:

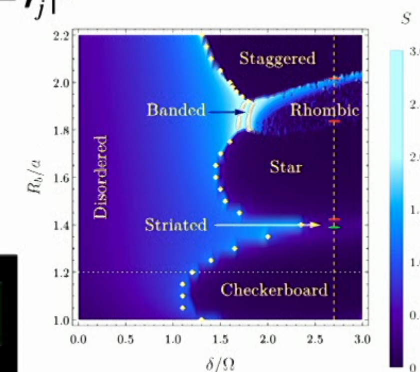
Positive, real-valued ground states

All information covered in $\left|\Psi(\boldsymbol{\sigma})\right|^2$

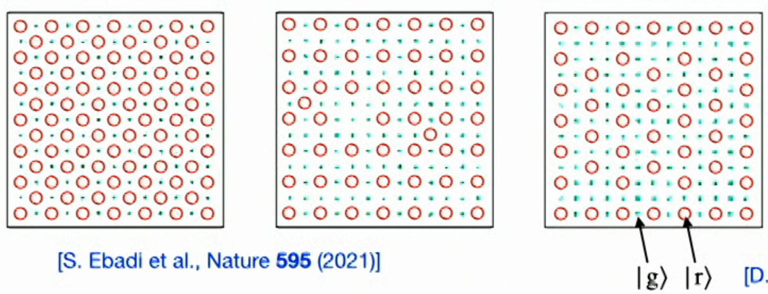
$$\hat{H} = -\frac{\Omega}{2}\sum_{i=1}^{N}\hat{\sigma}_i^x - \delta\sum_{i=1}^{N}\hat{n}_i + \sum_{i,j} V_{ij}\hat{n}_i\hat{n}_j$$



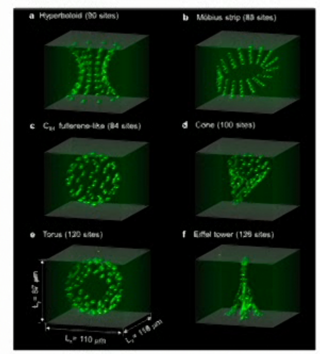
$$V_{ij} = \frac{\Omega R_b^6}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^6}$$



[R. Samaidar et al., PRL **124** (2020)]



[S. Ebadi et al., Nature **595** (2021)]

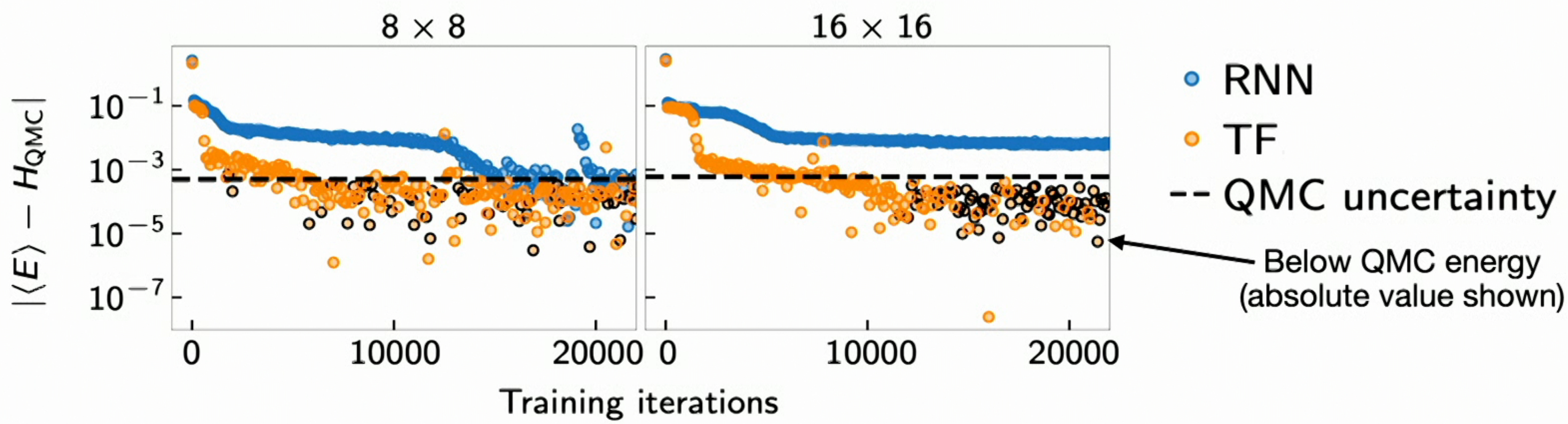


[D. Barredo et al., Nature **561** (2018)]

Stefanie.Czischek@uOttawa.ca

uOttawa

Performance comparison



- $\langle E\rangle$: Energy evaluated on 512 neural network samples
- H_{QMC}: Energy evaluated on 7×10^4 quantum Monte Carlo samples

- Transformers outperform RNNs

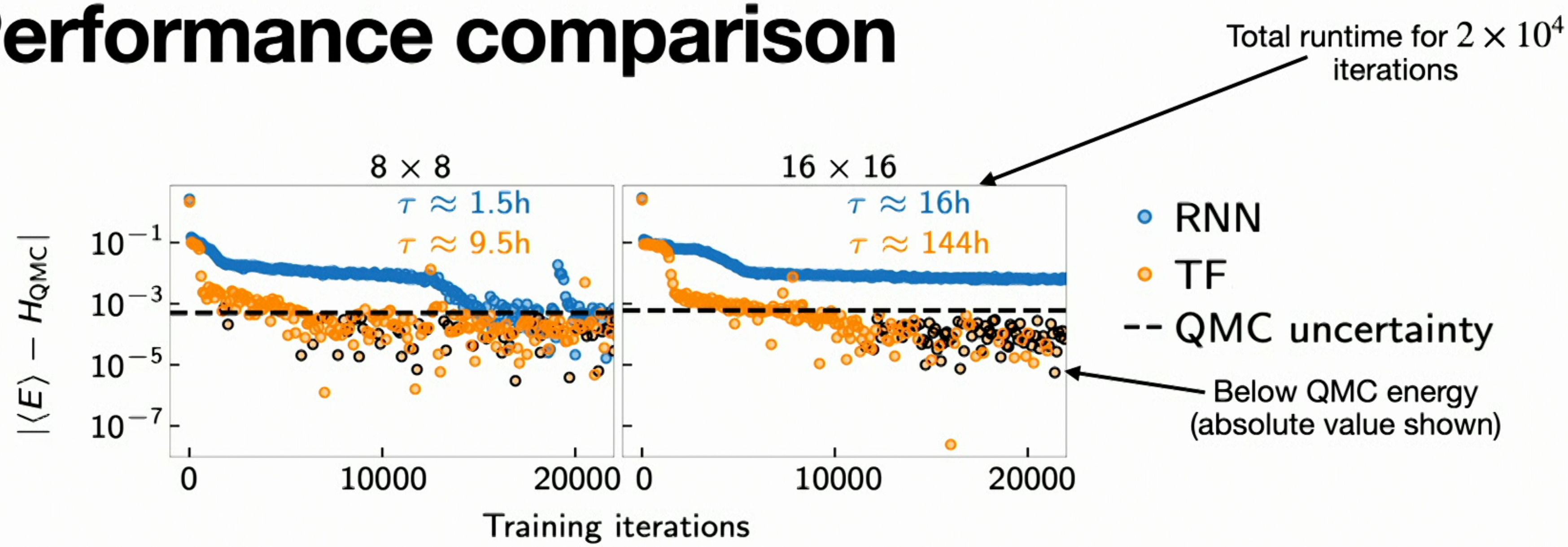
$$\hat{H} = -\frac{\Omega}{2}\sum_{i=1}^{N}\hat{\sigma}_i^x - \delta\sum_{i=1}^{N}\hat{n}_i + \sum_{i,j} V_{ij}\hat{n}_i\hat{n}_j$$

$$\Omega = \delta = 1 \qquad V_{ij} = \frac{7}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^6}$$

Stefanie.Czischek@uOttawa.ca

uOttawa

Performance comparison



- $\langle E \rangle$: Energy evaluated on 512 neural network samples
- H_{QMC}: Energy evaluated on 7×10^4 quantum Monte Carlo samples

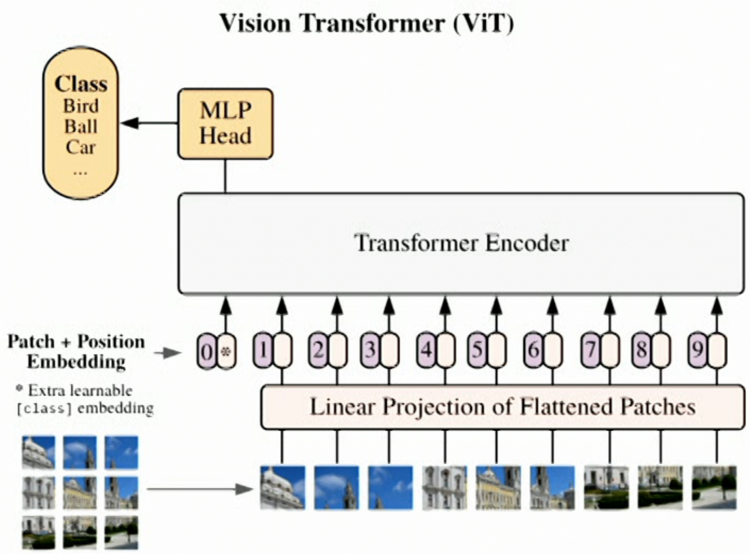
- Transformers outperform RNNs
- But at a high computational cost…

We cannot scale to larger system sizes!

Stefanie.Czischek@uOttawa.ca

uOttawa

Vision Transformer

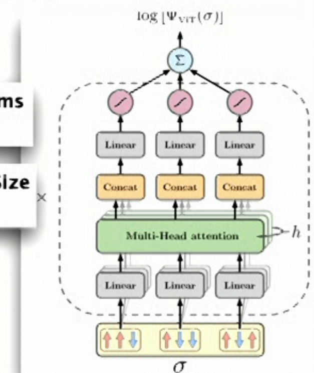


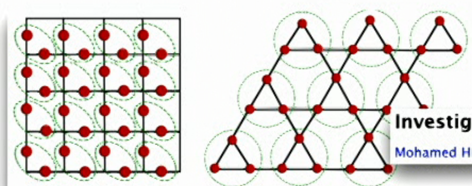
[Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at scale, arXiv:2010.11929 (2020)]

- Consider image patches instead of single pixels
  - Shorter sequence length and computation time
- The same approach can be used for qubit systems

Transformer variational wave functions for frustrated quantum spin systems
Luciano Loris Viteritti, Riccardo Rende, Federico Becca

Towards Neural Variational Monte Carlo That Scales Linearly with System Size
Or Sharir, Garnet Kin-Lic Chan, Anima Anandkumar



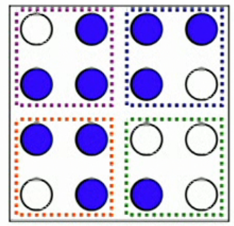


Investigating Topological Order using Recurrent Neural Networks
Mohamed Hibat-Allah, Roger G. Melko, Juan Carrasquilla

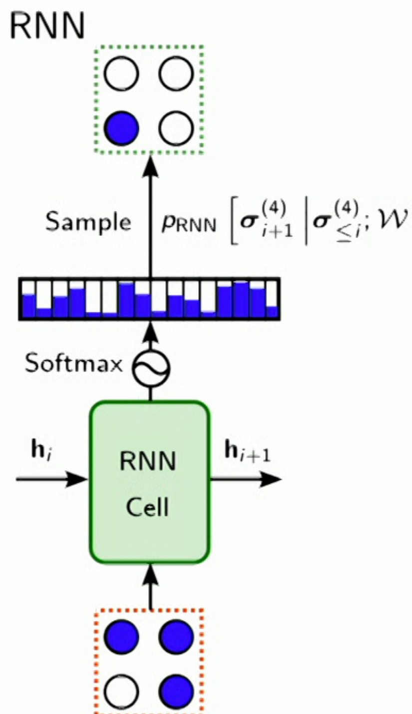
Stefanie.Czischek@uOttawa.ca

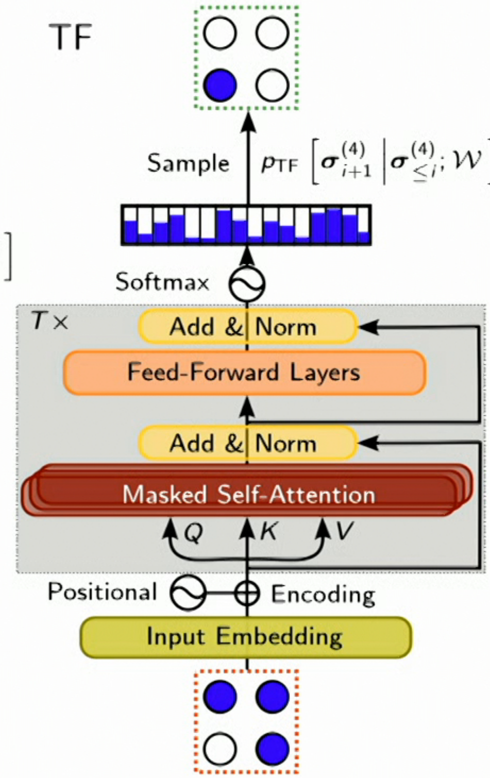
uOttawa

Patched neural network approach



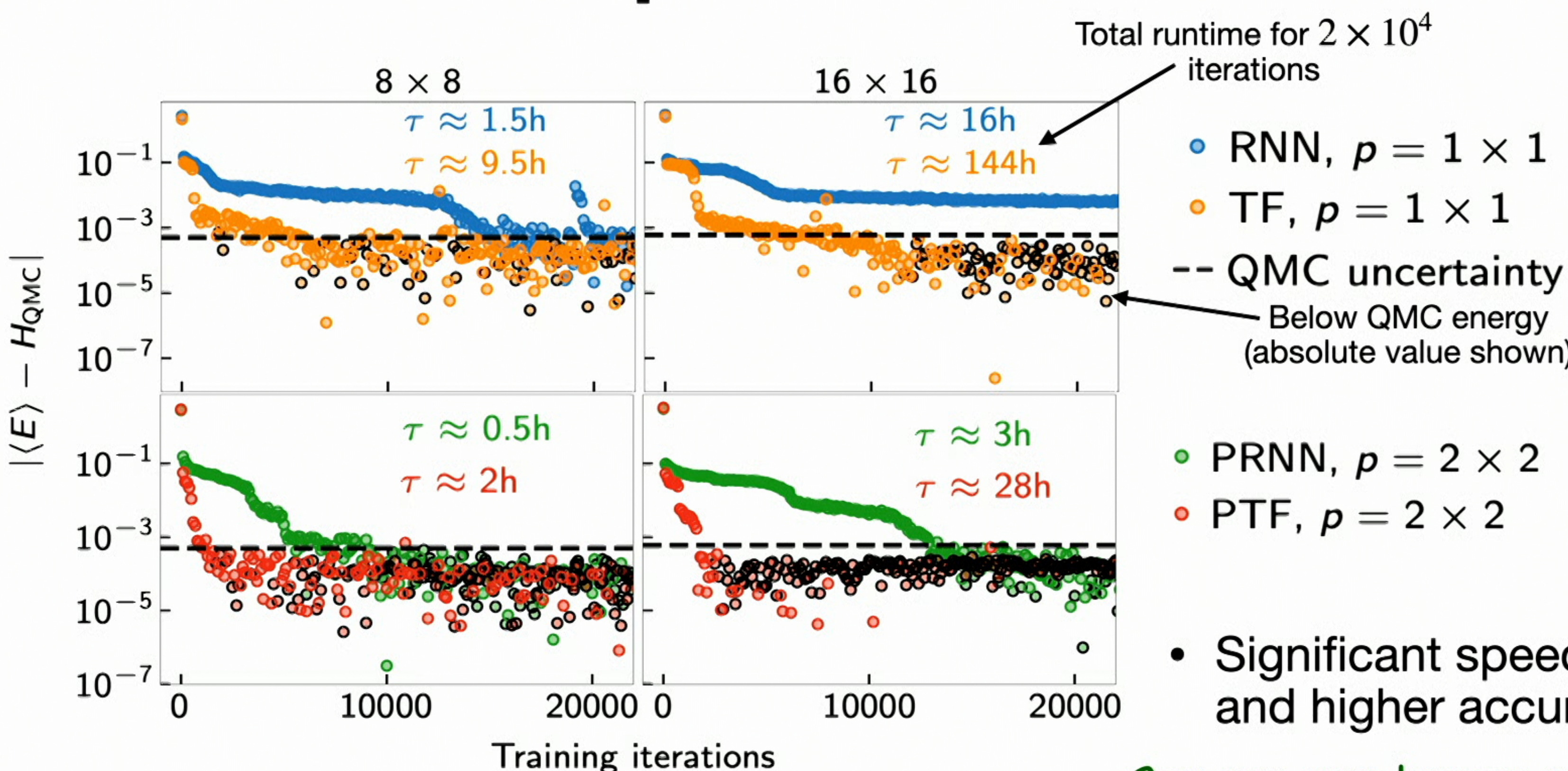
- Input patch of 2×2 atoms
- Sample from output probability over $2^{(2\times 2)}$ states
- Sequence length divided by four
- Local correlations directly encoded





Stefanie.Czischek@uOttawa.ca

Performance of patched networks



- Significant speedup and higher accuracies!

Can we use larger patches?

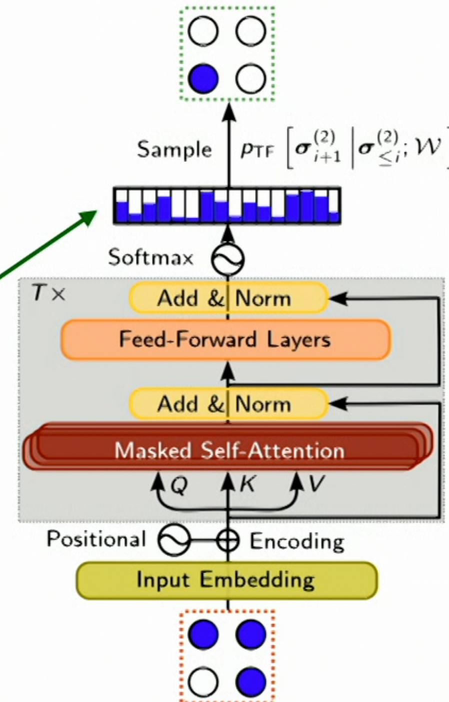
Stefanie.Czischek@uOttawa.ca

uOttawa

Large, patched transformers

- Larger input patches
  - Shorter runtimes
  - Comparable accuracies

Output dimension scales exponentially with the patch size!

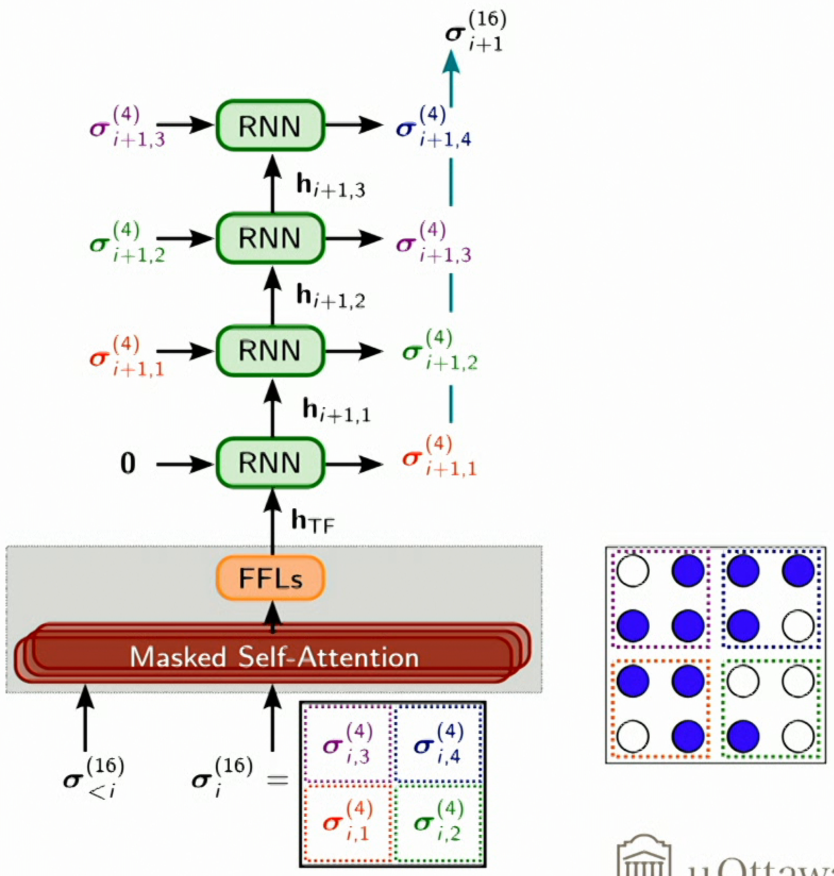


Stefanie.Czischek@uOttawa.ca

uOttawa

Large, patched transformers

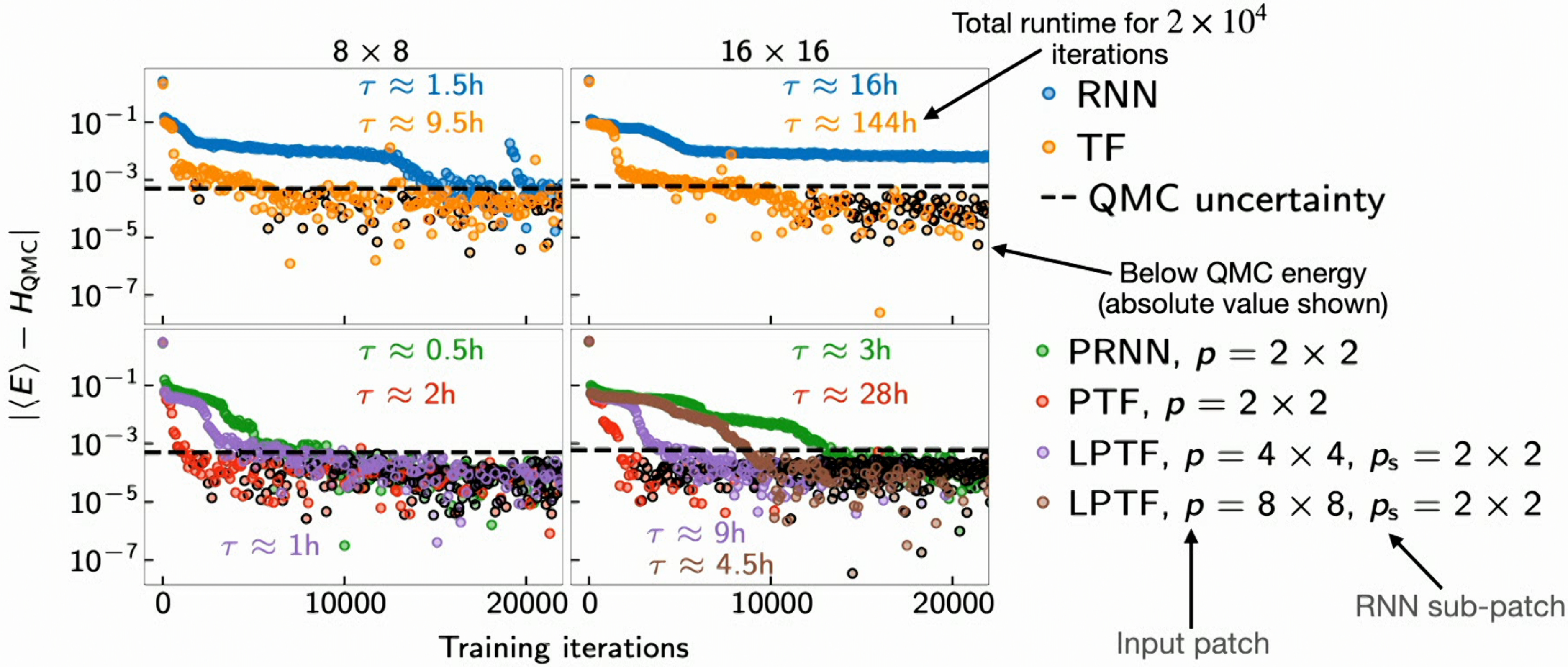
- Larger input patches
  - Shorter runtimes
  - Comparable accuracies
- Use an additional RNN to break down patch size
  - Gain power of transformer on large patches
  - Efficient RNN reduces output size



Stefanie.Czischek@uOttawa.ca

uOttawa

Large, patched transformer: performance

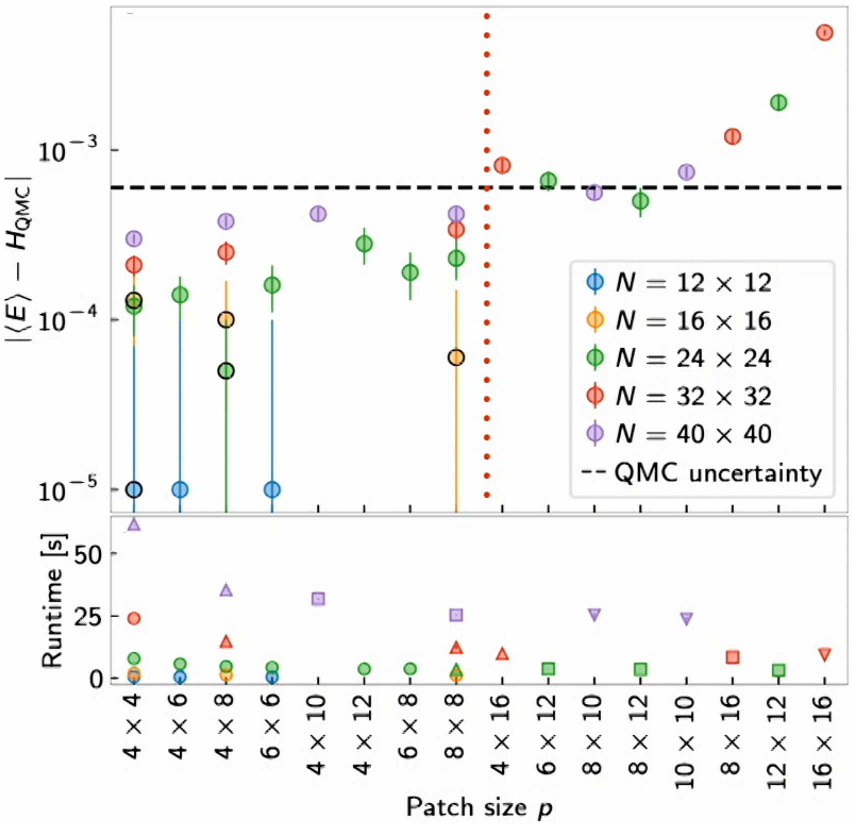


Reasonable runtimes and high accuracies!

Stefanie.Czischek@uOttawa.ca

uOttawa

Going bigger



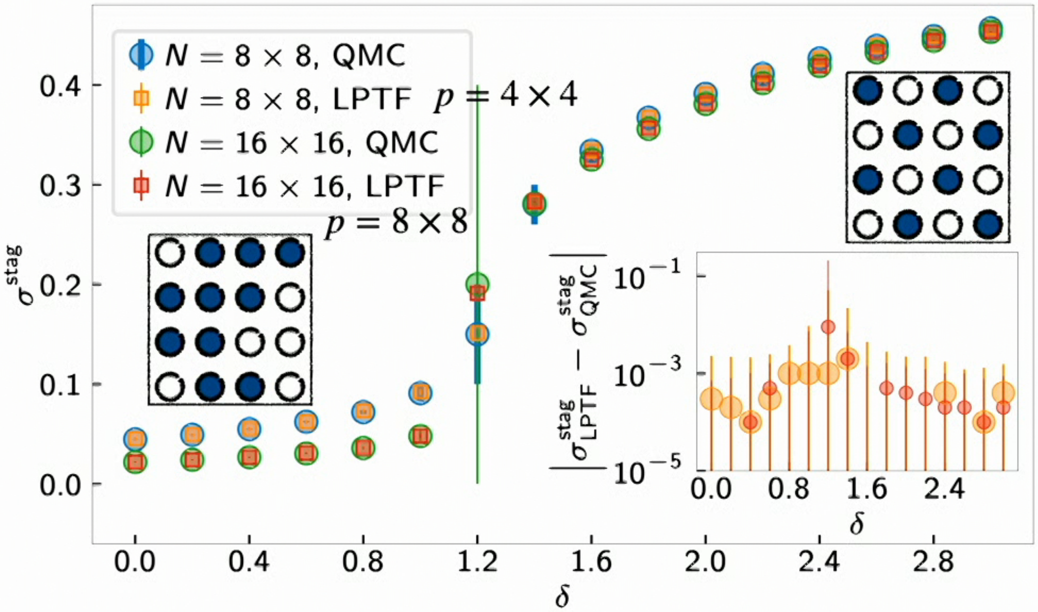
- Patch too large: RNN expressivity limited and amount of information increased
- Accuracies below QMC uncertainty below $p = 8 \times 8$
- Run times saturate for large patches (implementation detail)

Choosing patches around $p = 8 \times 8$, we can model enormous system sizes at high accuracies and low costs!

Stefanie.Czischek@uOttawa.ca

uOttawa

Different phases of matter



$$\hat{H} = -\frac{\Omega}{2}\sum_{i=1}^{N}\hat{\sigma}_i^x - \delta\sum_{i=1}^{N}\hat{n}_i + \sum_{i,j}V_{ij}\hat{n}_i\hat{n}_j$$

$$\Omega = 1 \qquad V_{ij} = \frac{3}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^6}$$

- Phase transition between disordered and checkerboard phase
- Order parameter: staggered magnetization

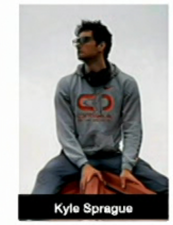
$$\sigma^{\mathrm{stag}} = \left\langle \left| \sum_{i=1}^{N} (-1)^i \frac{n_i - 1/2}{N} \right| \right\rangle$$

Stefanie.Czischek@uOttawa.ca

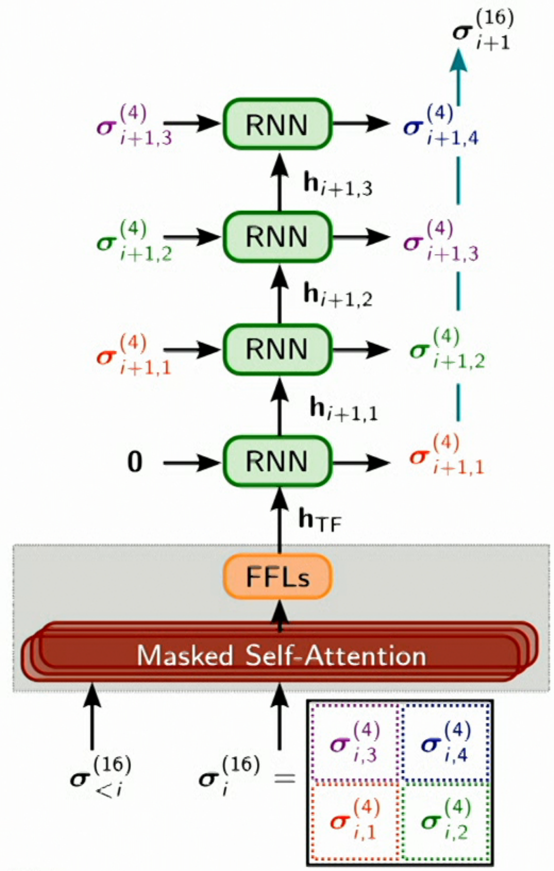
uOttawa

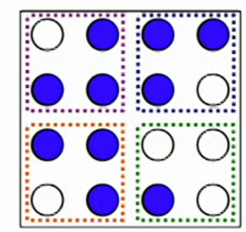
Summary

[SC, K. Sprague, arXiv:2306.03921 (2023)]



- Considering patches of atoms leads to higher accuracies and shorter runtimes
- The large, patched transformer shows remarkable results beyond state-of-the-art simulations
  - Combines transformer and RNN
- The approach can be used for arbitrary qubit systems
- The chosen transformer models are still small…





Stefanie.Czischek@uOttawa.ca

APRIQuOt

uOttawa