# Enhanced quantum state reconstruction with artificial neural networks

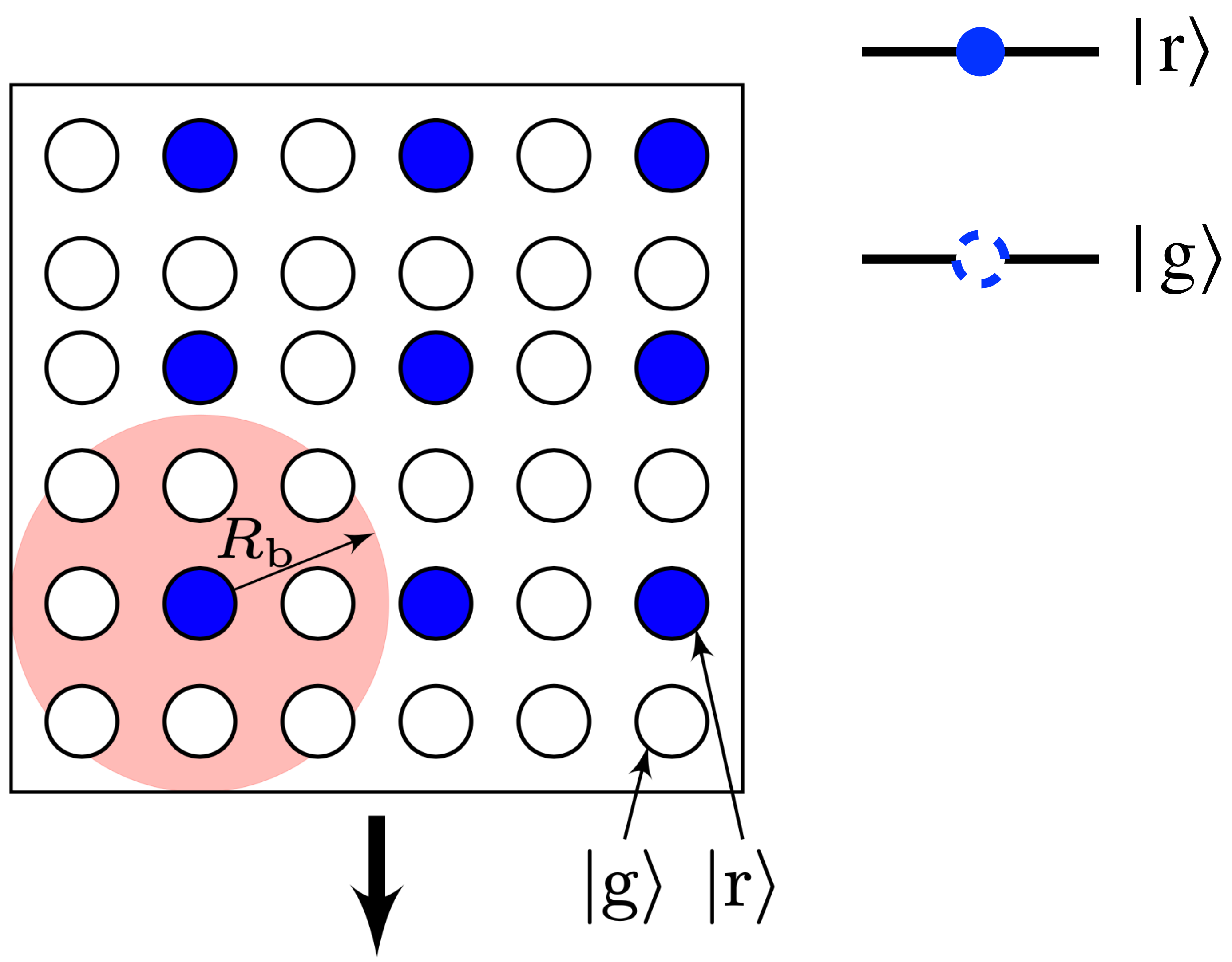
**Stef Czischek**

June 20, 2023

uOttawa

APRIQuOt

# Rydberg atom arrays



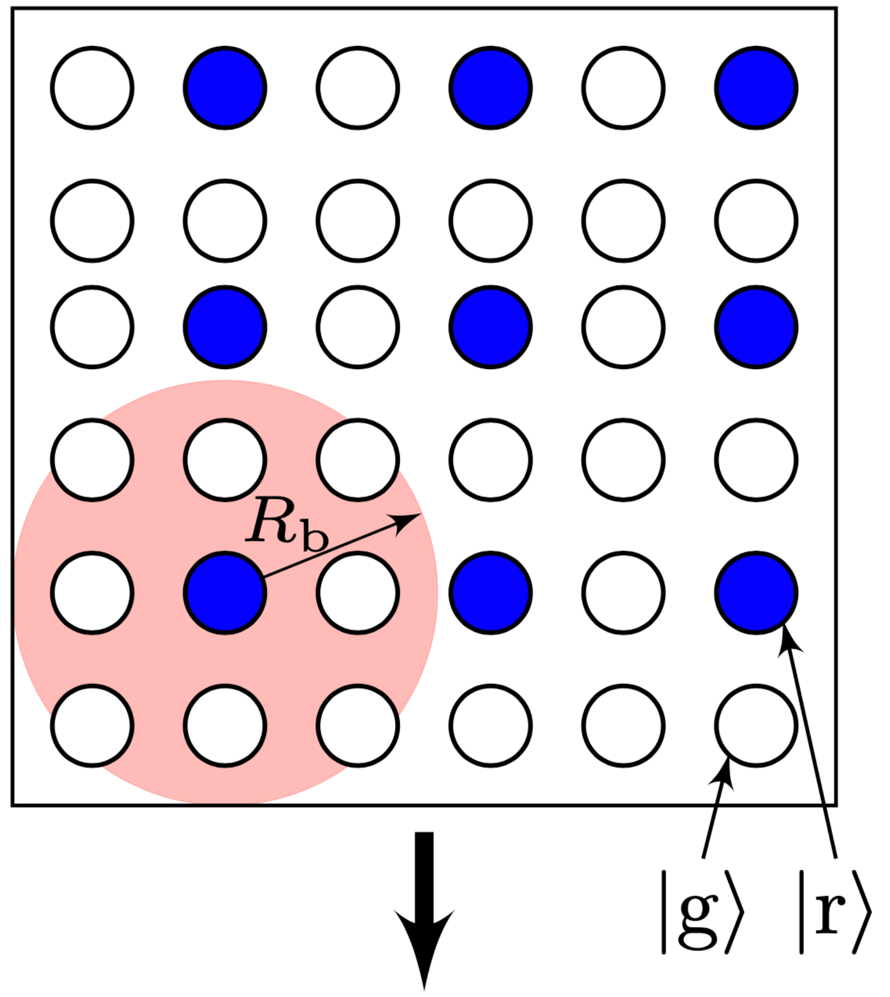

Projective measurement

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

$N = L \times L$

atoms on
square lattice

# Rydberg atom arrays

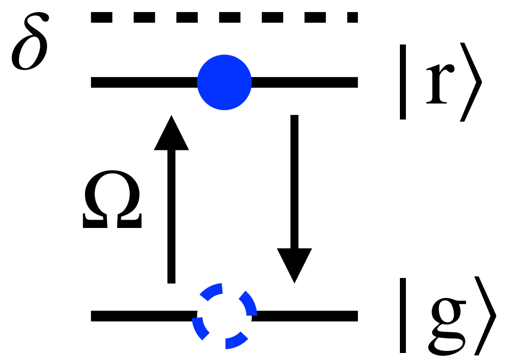



Projective measurement

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

$N = L \times L$

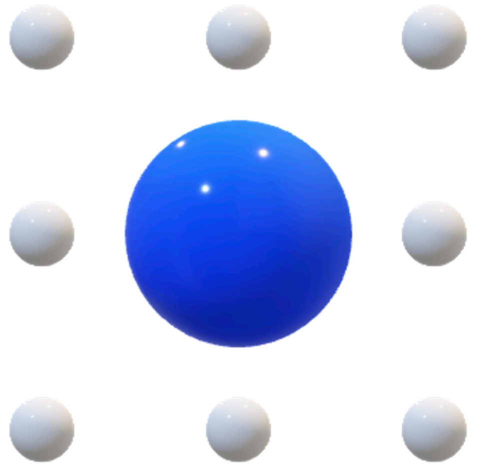
atoms on square lattice




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

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

van der Waals interaction:
penalize two excitations within $R_b$
(Rydberg blockade)



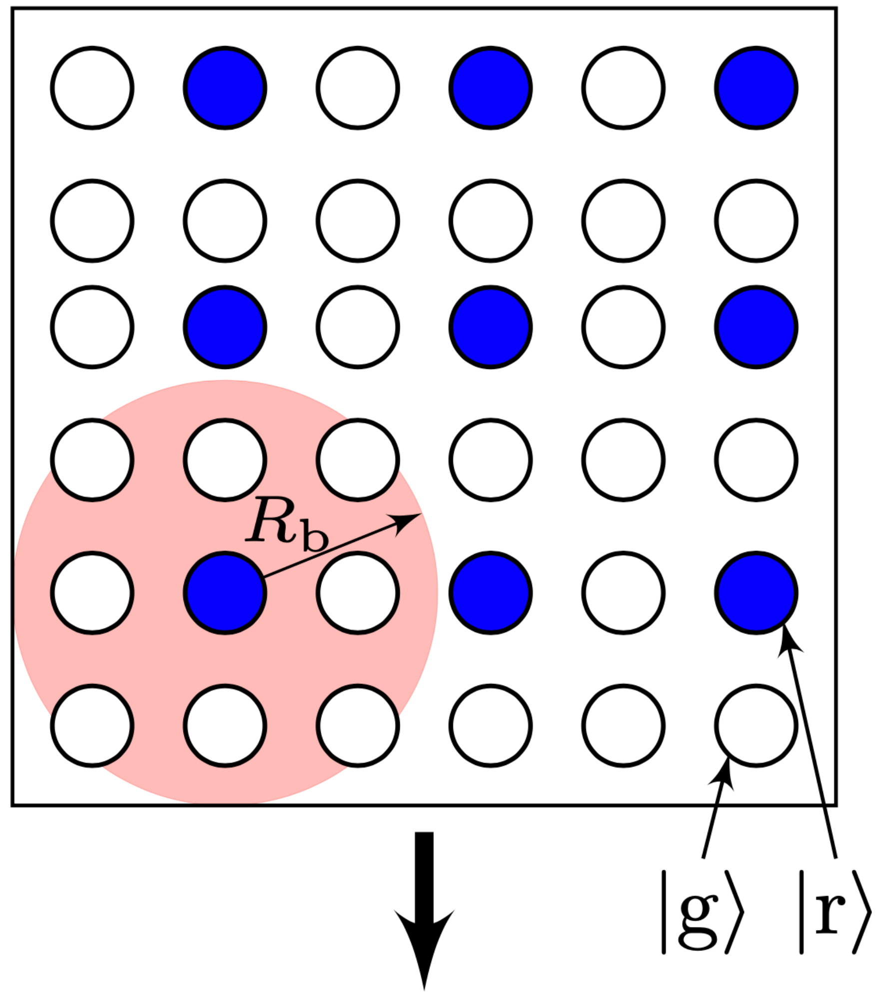
$$V_{ij} = \frac{\Omega R_b^6}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^6}$$

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

Laser driving:
detuning $\delta$, Rabi frequency $\Omega$

Stefanie.Czischek@uOttawa.ca

# Rydberg atom arrays

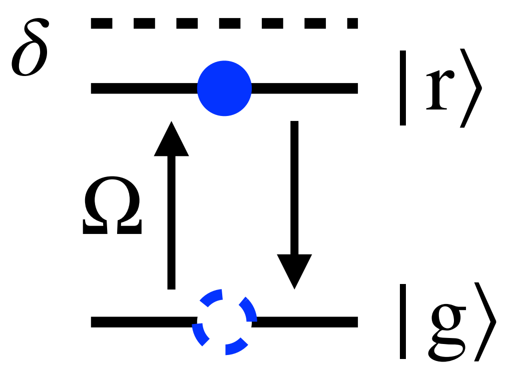



Projective measurement

$|\boldsymbol{\sigma}\rangle = |g\ r\ g \dots g\ g\rangle$

$N = L \times L$

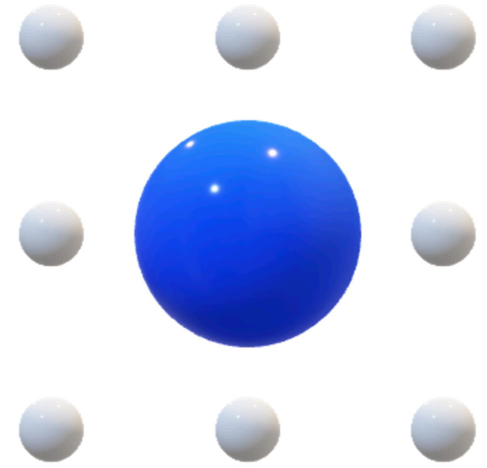
atoms on square lattice




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

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

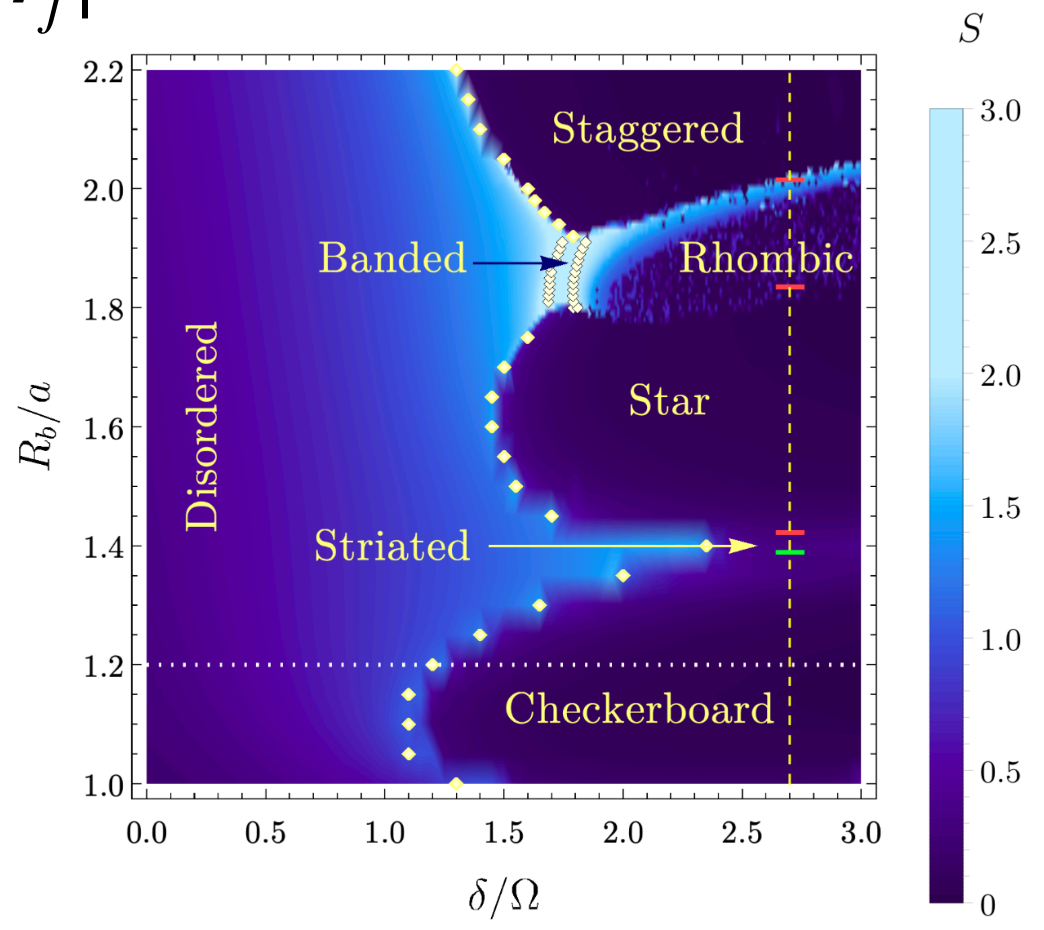
van der Waals interaction: penalize two excitations within $R_b$ (Rydberg blockade)



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

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

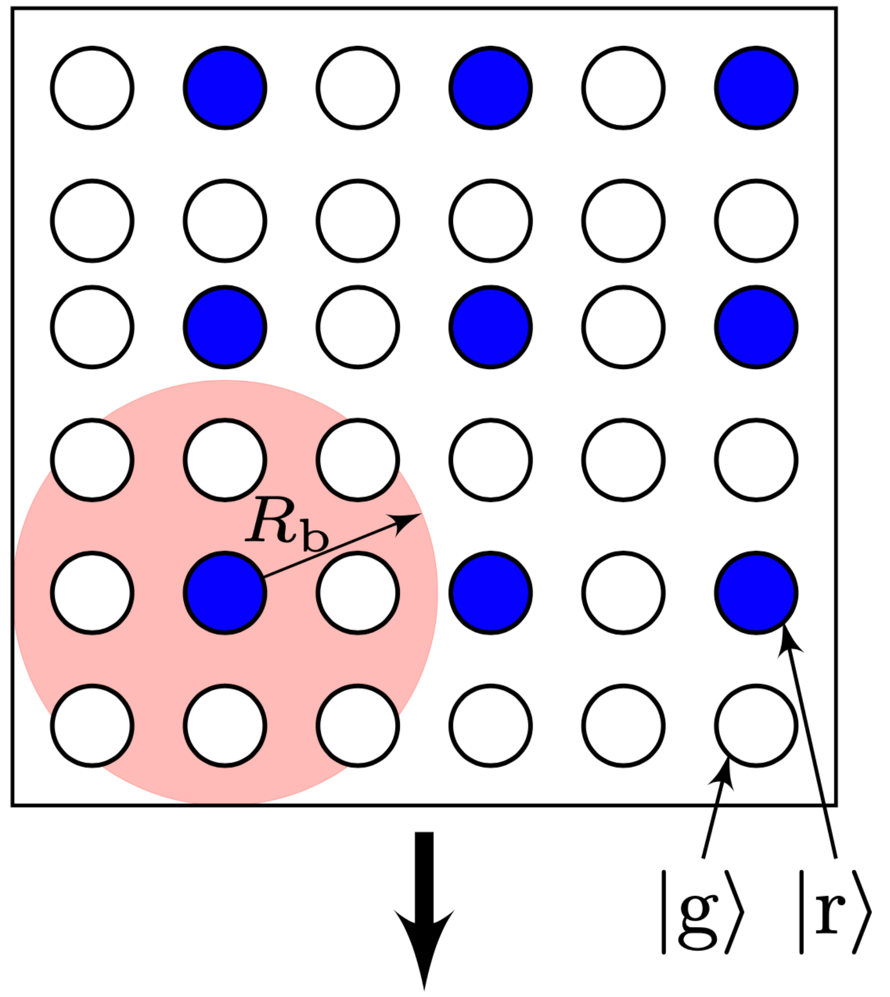
Laser driving: detuning $\delta$, Rabi frequency $\Omega$




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

Stefanie.Czischek@uOttawa.ca

# Rydberg atom arrays

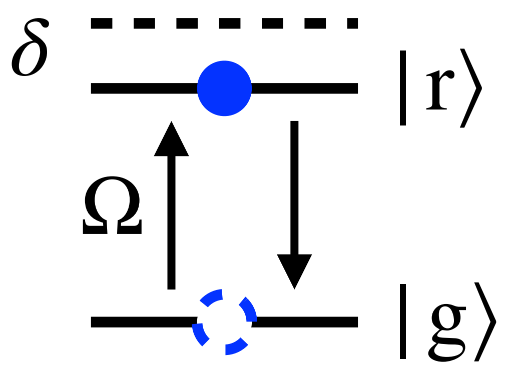



Projective measurement

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

$N = L \times L$

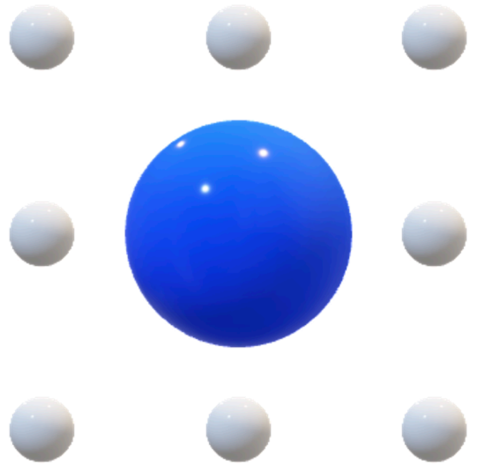
atoms on square lattice




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

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

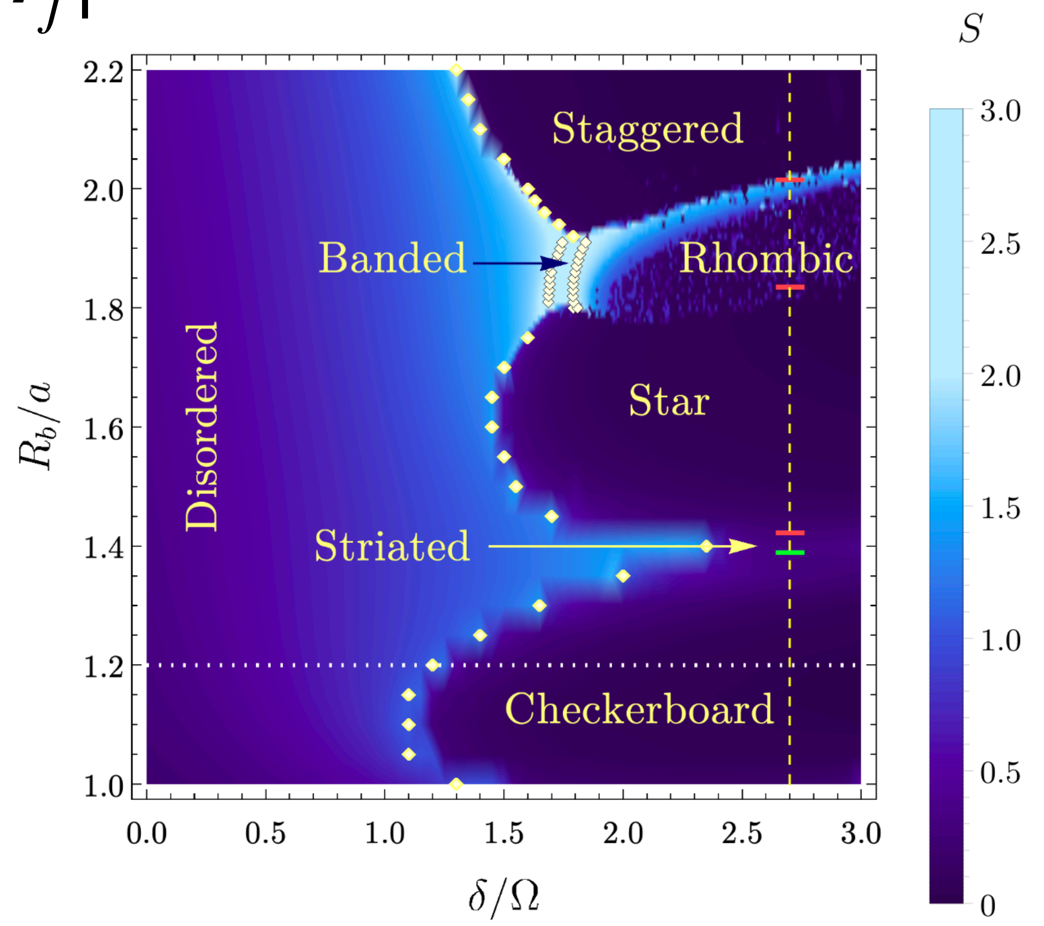
van der Waals interaction:
penalize two excitations within $R_b$
(Rydberg blockade)



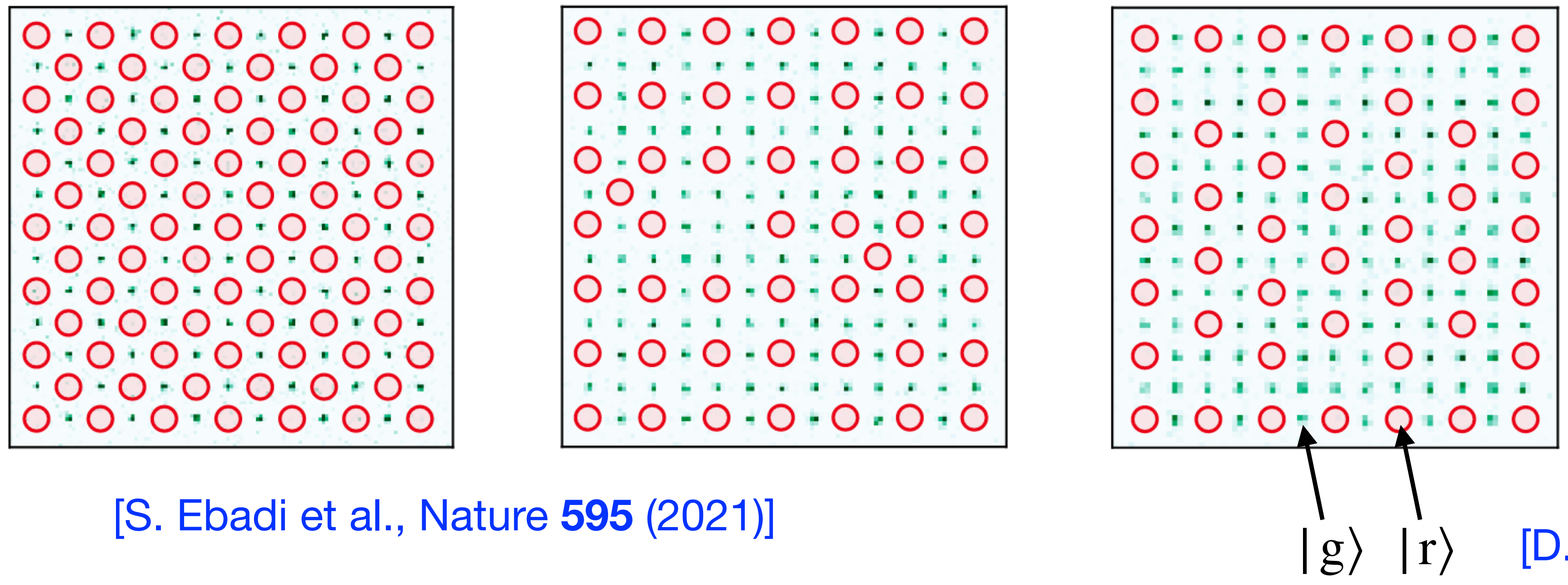
$$V_{ij} = \frac{\Omega R_b^6}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^6}$$

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

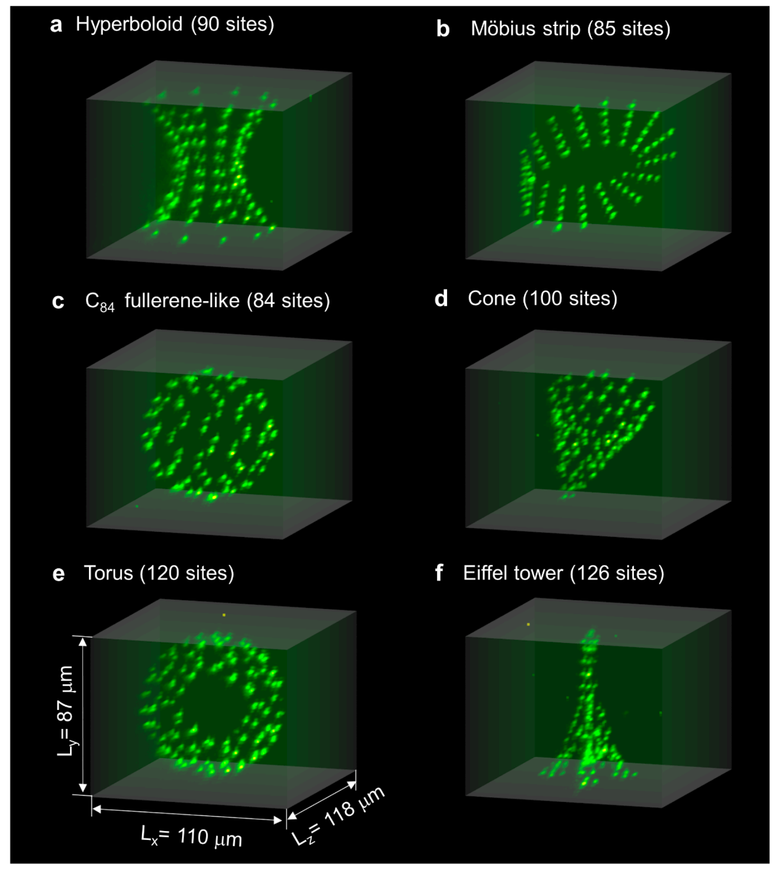
Laser driving:
detuning $\delta$, Rabi frequency $\Omega$




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




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

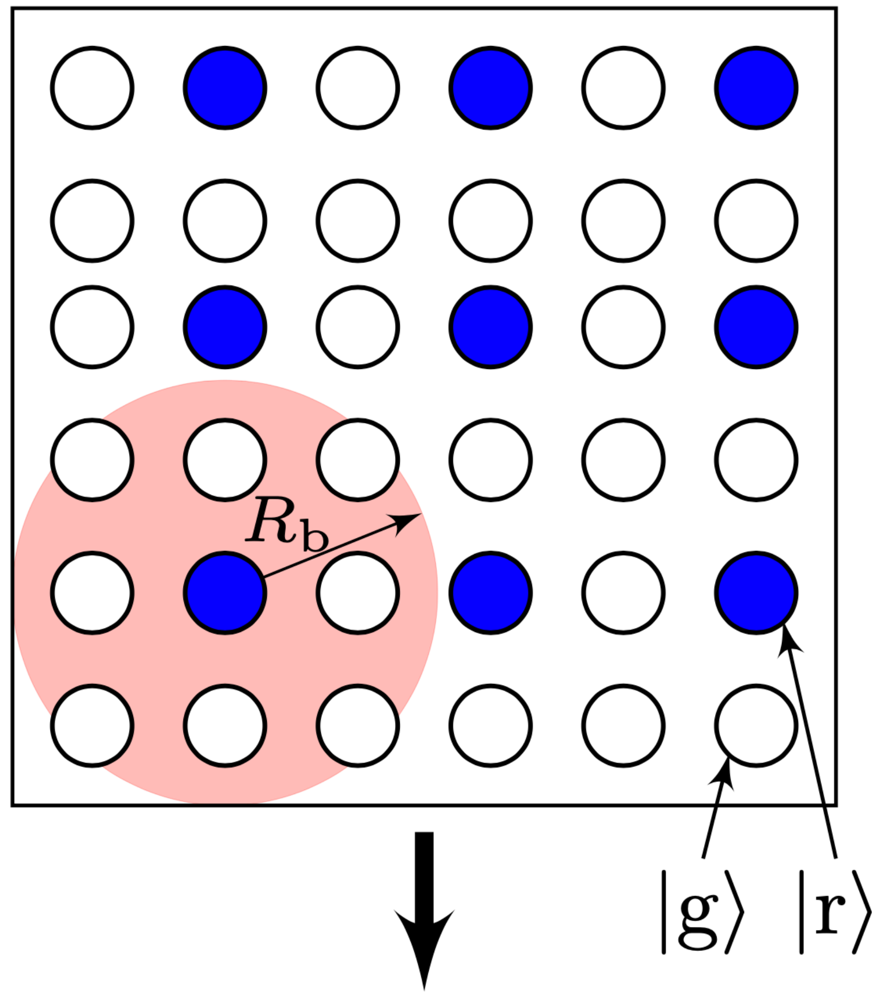



[D. Barredo et al., Nature **561** (2018)]



uOttawa

# Rydberg atom arrays

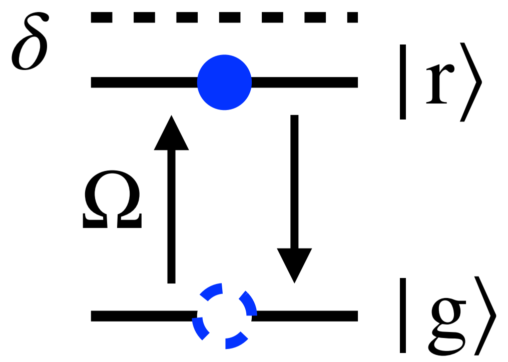



Projective measurement

$|\boldsymbol{\sigma}\rangle = |\text{g r g} \dots \text{g g}\rangle$

$N = L \times L$

atoms on square lattice



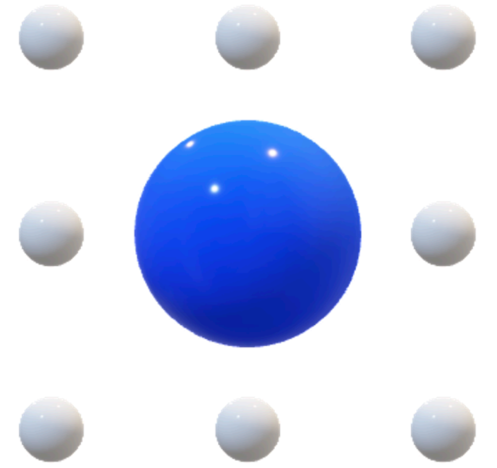

$\hat{\sigma}_i^x = |\text{g}\rangle_i\langle \text{r}|_i$

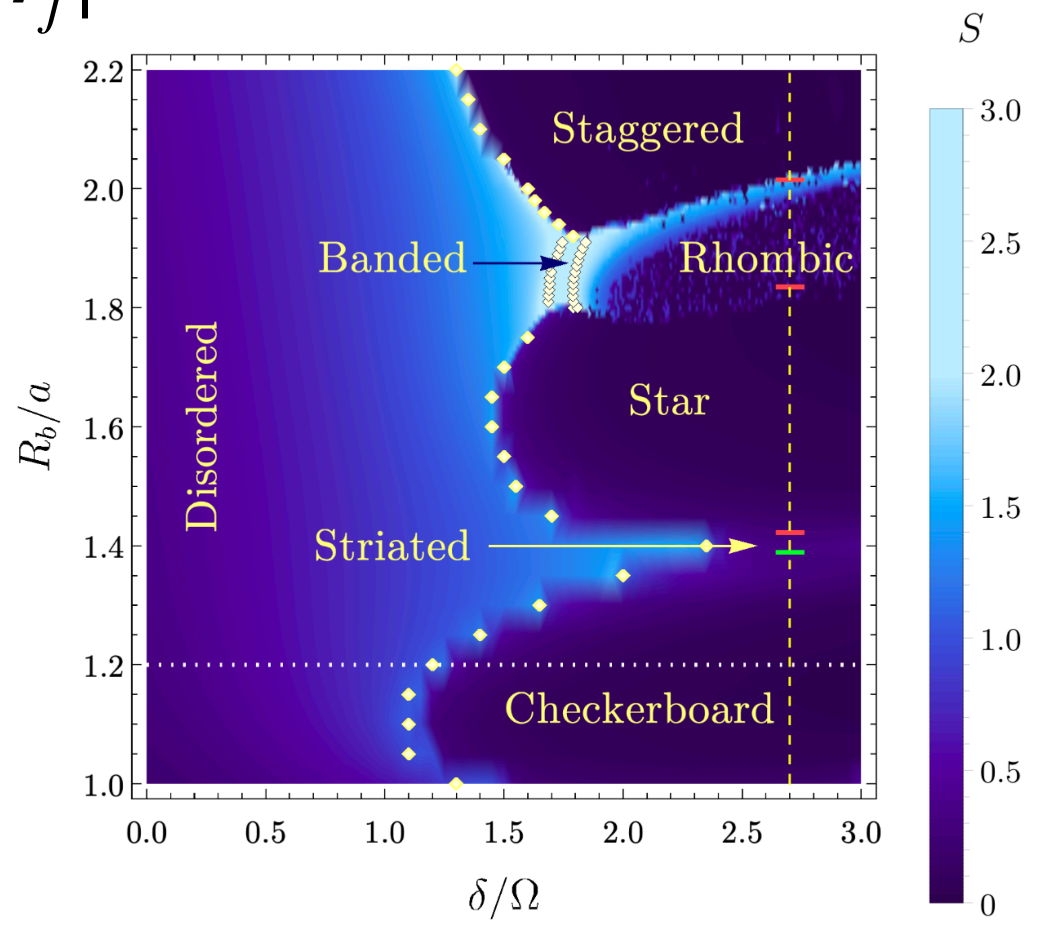
$\hat{n}_i = |\text{r}\rangle_i\langle \text{r}|_i$

Stoquastic:
Positive, real-valued ground states
All information covered in $\left|\Psi(\boldsymbol{\sigma})\right|^2$

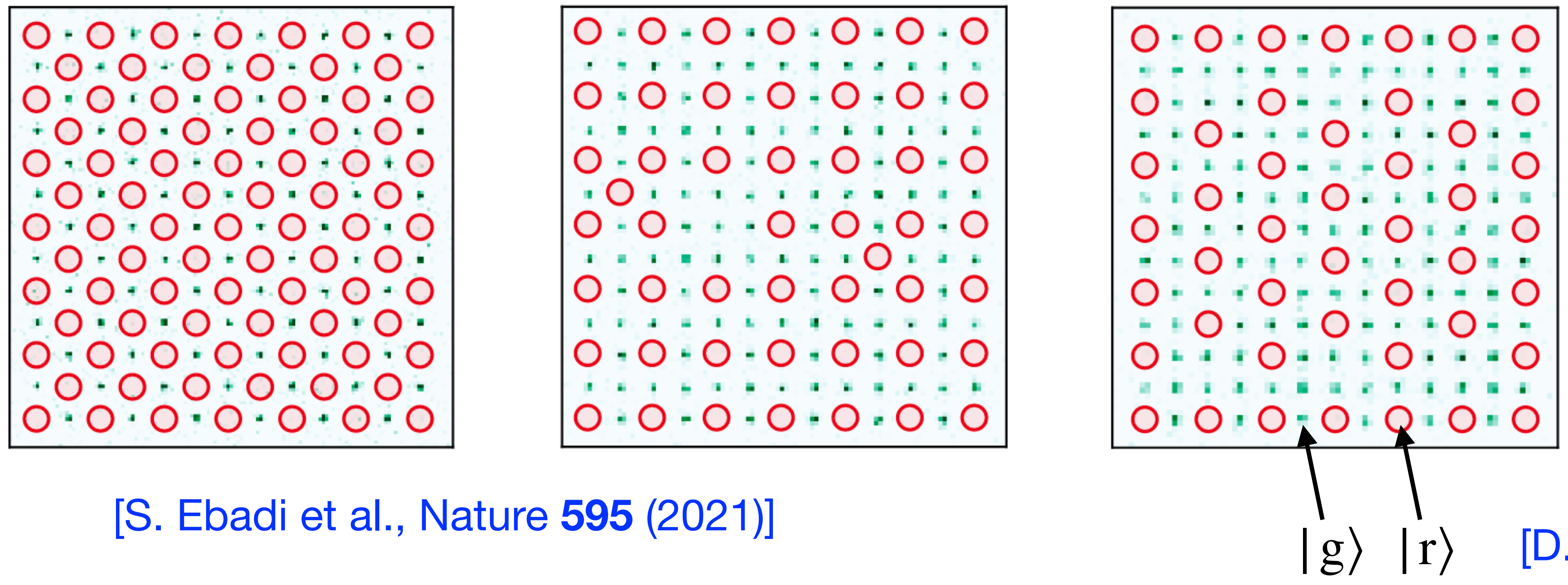
$$\hat{H} = -\frac{\Omega}{2}\sum_{i=1}^{N}\hat{\sigma}_i^x - \delta\sum_{i=1}^{N}\hat{n}_i + \sum_{i,j} V_{ij}\hat{n}_i\hat{n}_j$$

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

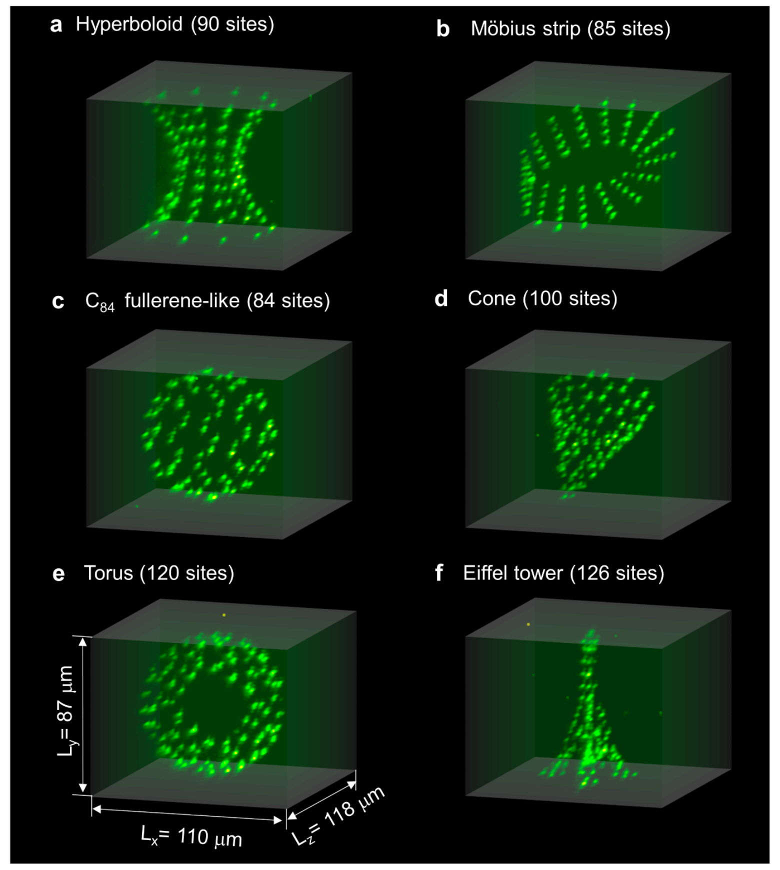





[R. Samajdar et al., PRL **124** (2020)]




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




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

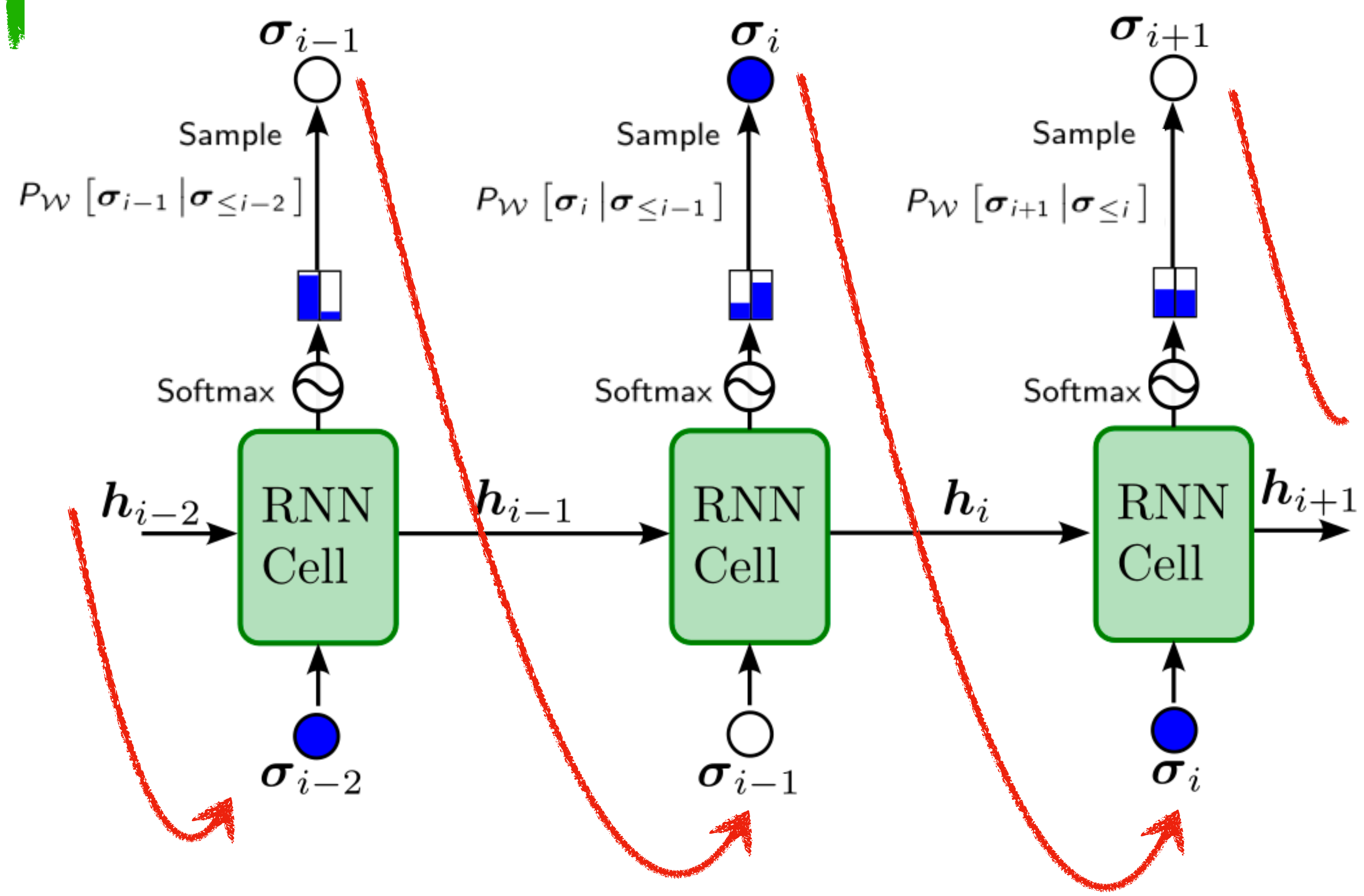
Stefanie.Czischek@uOttawa.ca

uOttawa

# Recurrent neural network quantum states

[O. Sharir et al., PRL **124** (2020)]

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



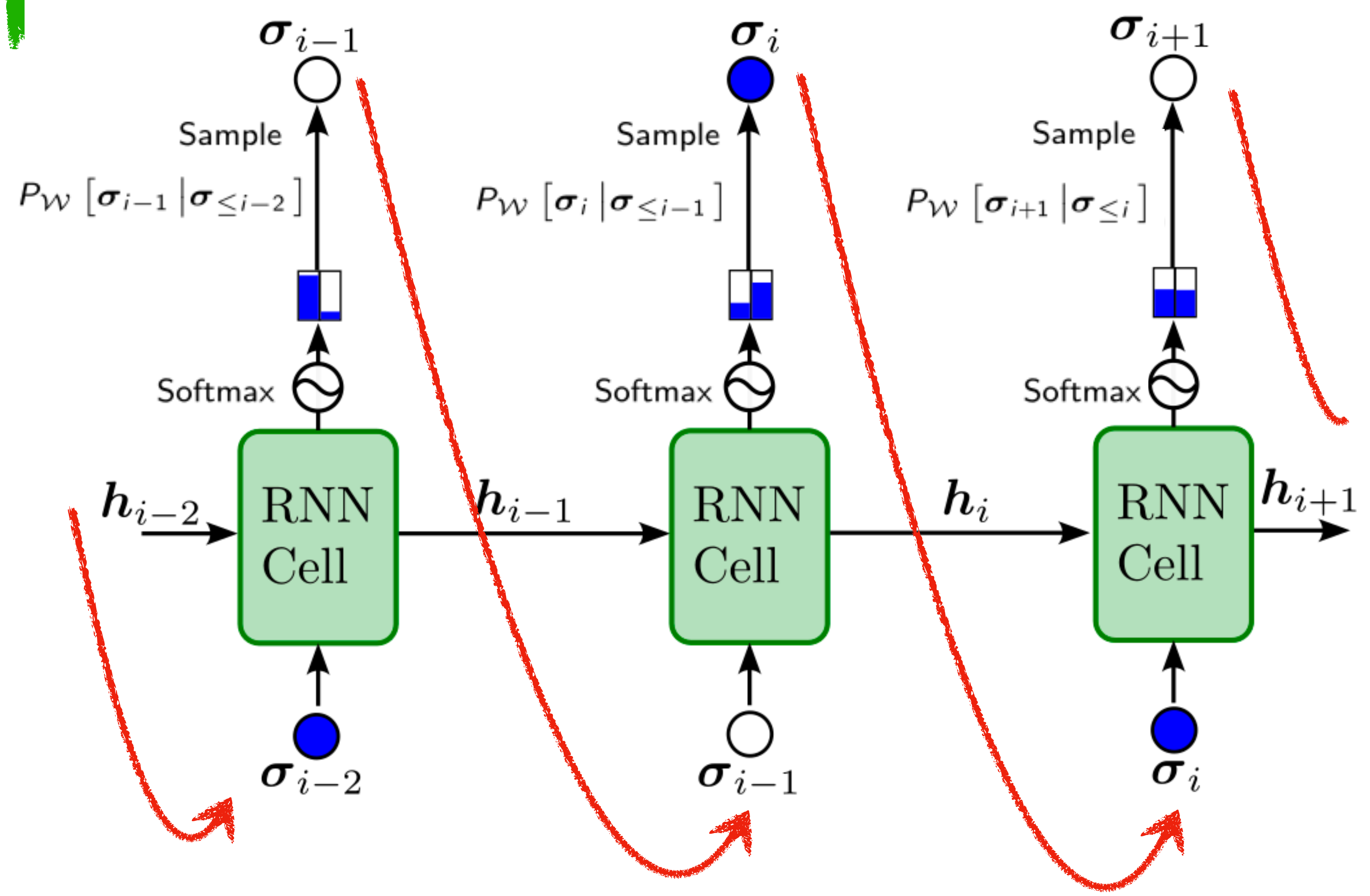

- RNN cell: $N_{\mathrm{h}}$ hidden units, variational parameters $\mathscr{W}$

Stefanie.Czischek@uOttawa.ca

# Recurrent neural network quantum states

[O. Sharir et al., PRL **124** (2020)]

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



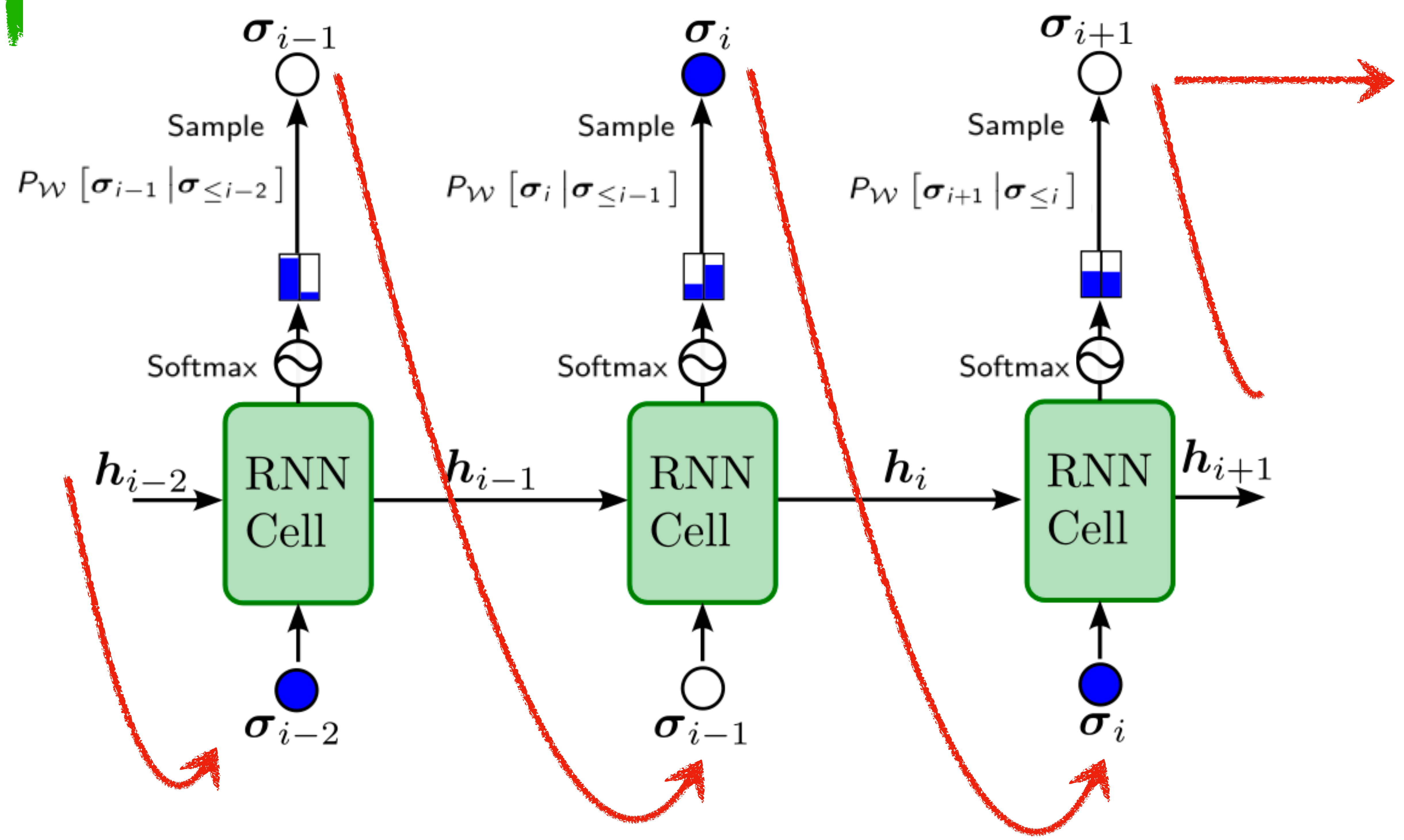

- RNN cell: $N_\mathrm{h}$ hidden units, variational parameters $\mathscr{W}$
- RNN encodes squared wavefunction amplitudes

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

Stefanie.Czischek@uOttawa.ca

# Recurrent neural network quantum states

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




$|\Psi_{\mathscr{W}}\rangle = |$○●○○●●○○●$\rangle$

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

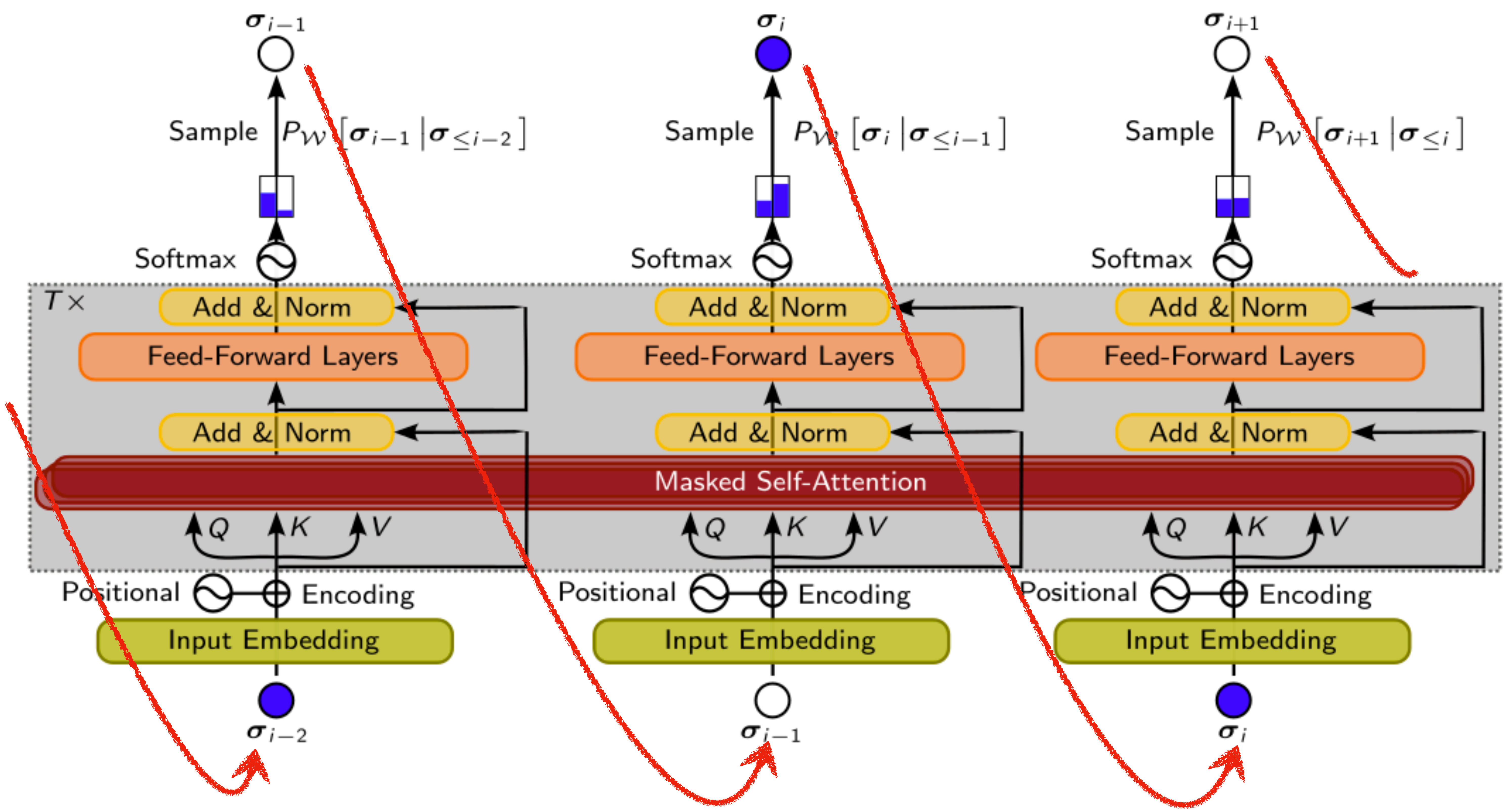
$$\left|\Psi(\boldsymbol{\sigma})\right|^2 \approx \left|\Psi_{\mathscr{W}}(\boldsymbol{\sigma})\right|^2 = p_{\mathrm{RNN}}(\boldsymbol{\sigma};\mathscr{W}) = \prod_i p_{\mathrm{RNN}}(\sigma_i\,|\,\sigma_{i-1},\ldots,\sigma_1;\mathscr{W})$$

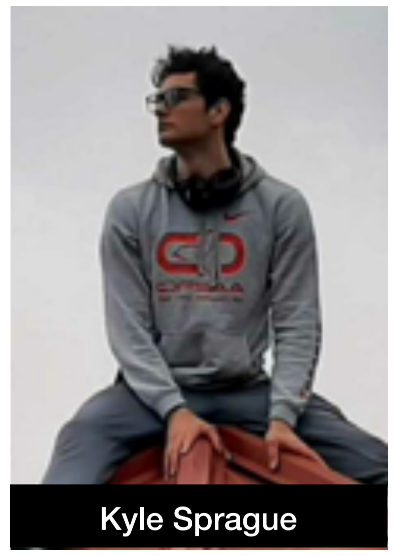
$$\langle E \rangle = \sum_{\boldsymbol{\sigma}} \left|\Psi_{\mathscr{W}}(\boldsymbol{\sigma})\right|^2 H_{\mathrm{loc}}(\boldsymbol{\sigma}) \approx \frac{1}{N_\mathrm{s}} \sum_{\boldsymbol{\sigma}\sim p_{\mathrm{RNN}}(\boldsymbol{\sigma};\mathscr{W})} H_{\mathrm{loc}}(\boldsymbol{\sigma}) \qquad H_{\mathrm{loc}}(\boldsymbol{\sigma})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

Stefanie.Czischek@uOttawa.ca

uOttawa

# Transformer quantum states



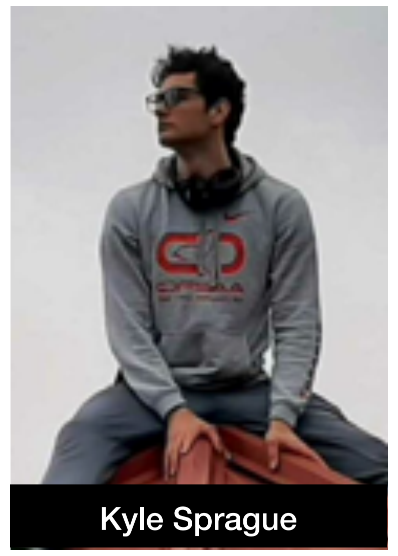



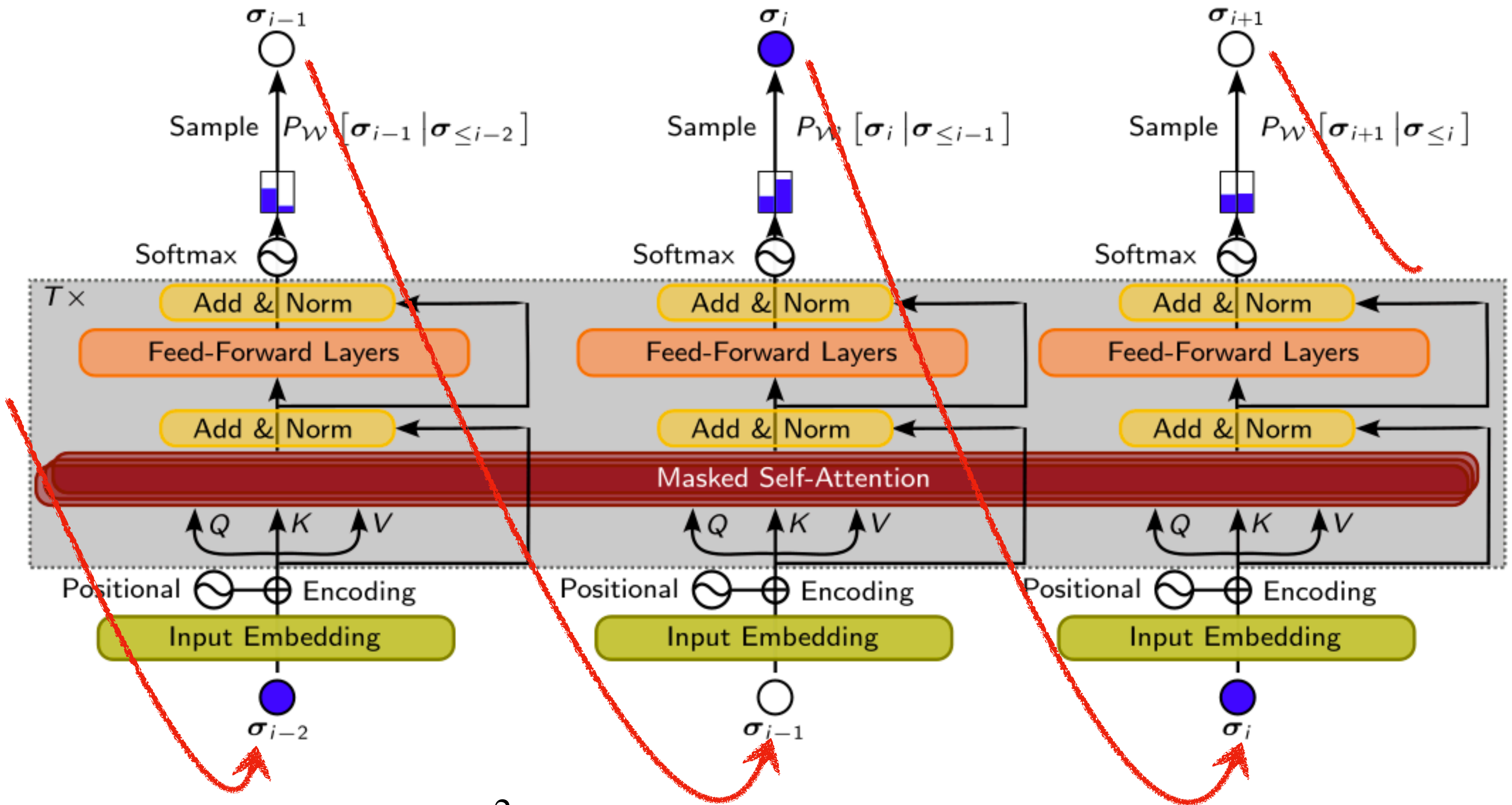

- Attention: trained connections to all elements
- Wave function encoded similar to RNN

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

Stefanie.Czischek@uOttawa.ca

# Transformer quantum states



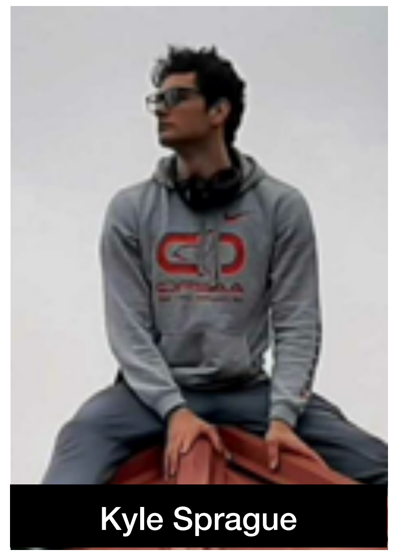



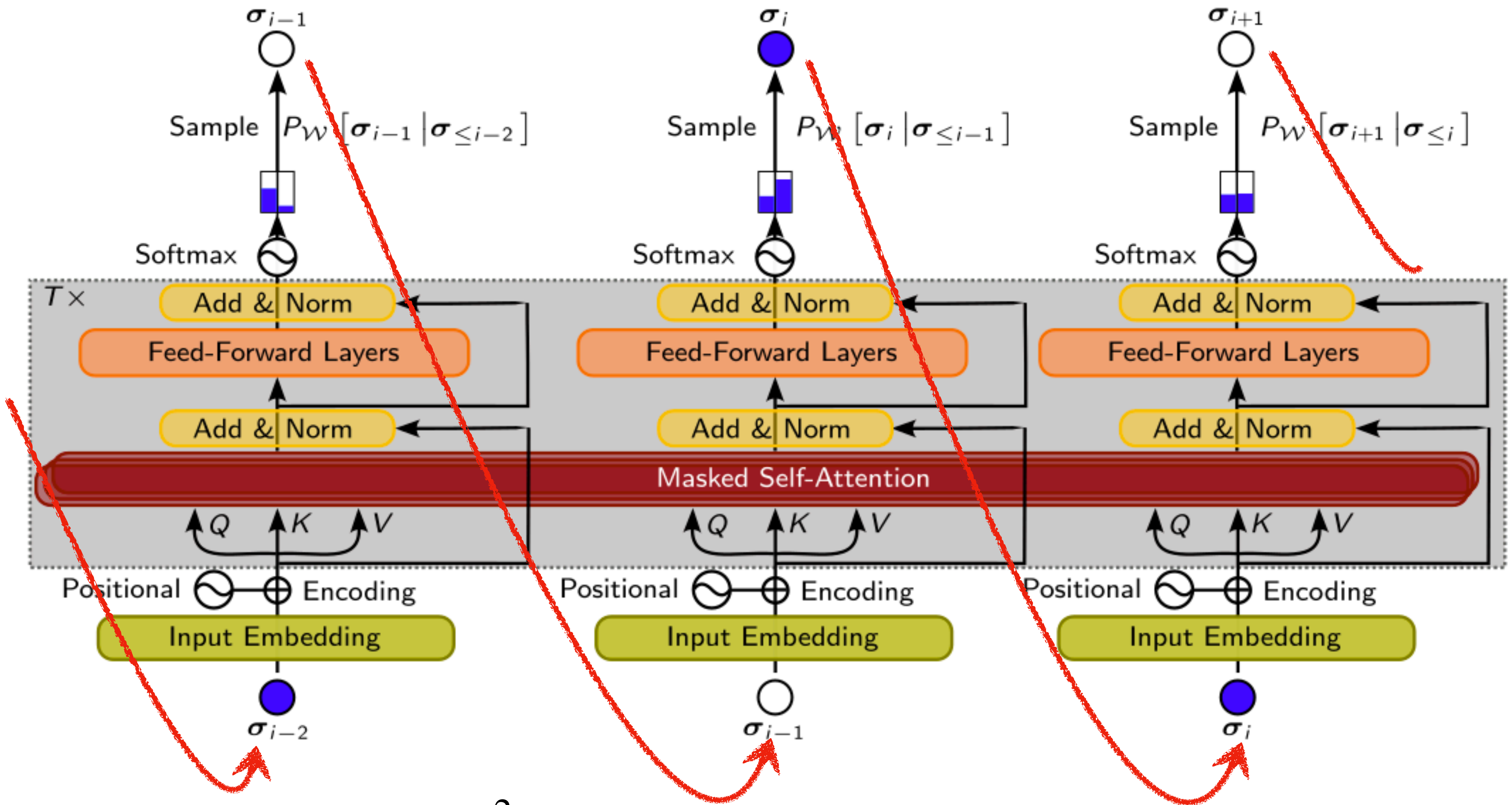

$$\left|\Psi(\boldsymbol{\sigma})\right|^2 \approx p_{\mathrm{TF}}(\boldsymbol{\sigma};\mathscr{W}) = \prod_i P_{\mathscr{W}}[\sigma_i \,|\, \sigma_{\leq i-1}]$$

- Attention: trained connections to all elements
- Wave function encoded similar to RNN

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

# Transformer quantum states



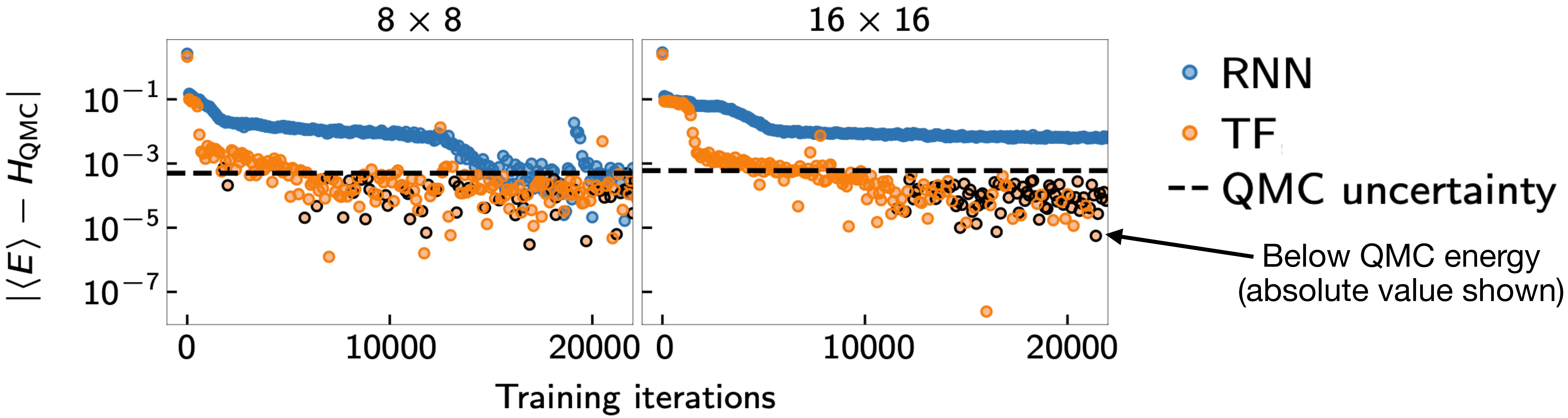

$$|\Psi(\boldsymbol{\sigma})|^2 \approx p_{\mathrm{TF}}(\boldsymbol{\sigma}; \mathscr{W}) = \prod_i P_{\mathscr{W}}[\sigma_i | \sigma_{\leq i-1}]$$

- Attention: trained connections to all elements
- Wave function encoded similar to RNN

Do transformers perform better than RNNs?

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

# Performance comparison



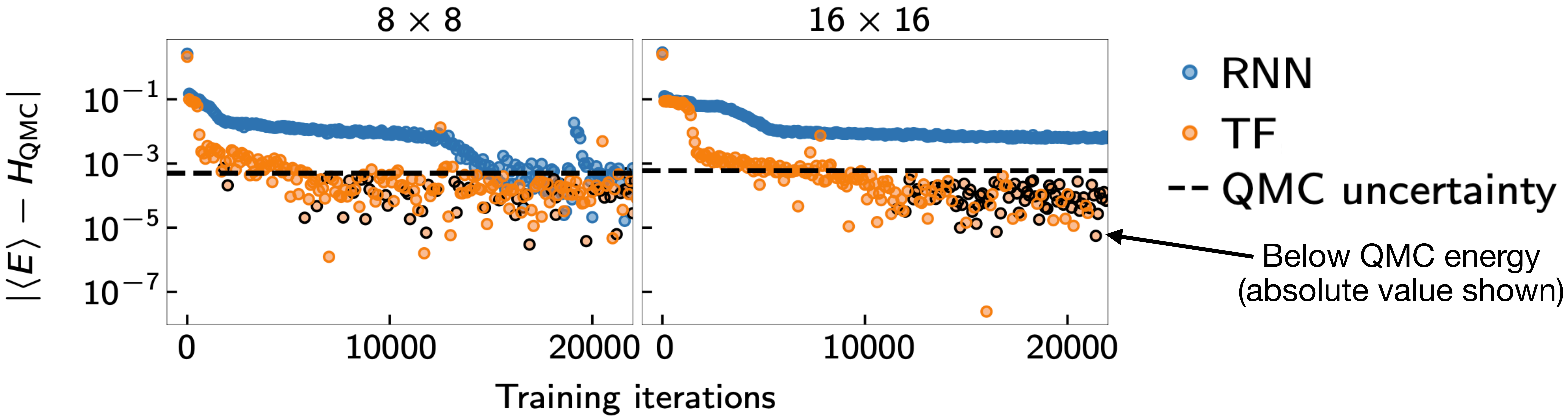

- $\langle E \rangle$: Energy evaluated on $512$ neural network samples
- $H_{\mathrm{QMC}}$: Energy evaluated on $7 \times 10^4$ quantum Monte Carlo samples

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

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

Stefanie.Czischek@uOttawa.ca

# Performance comparison




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

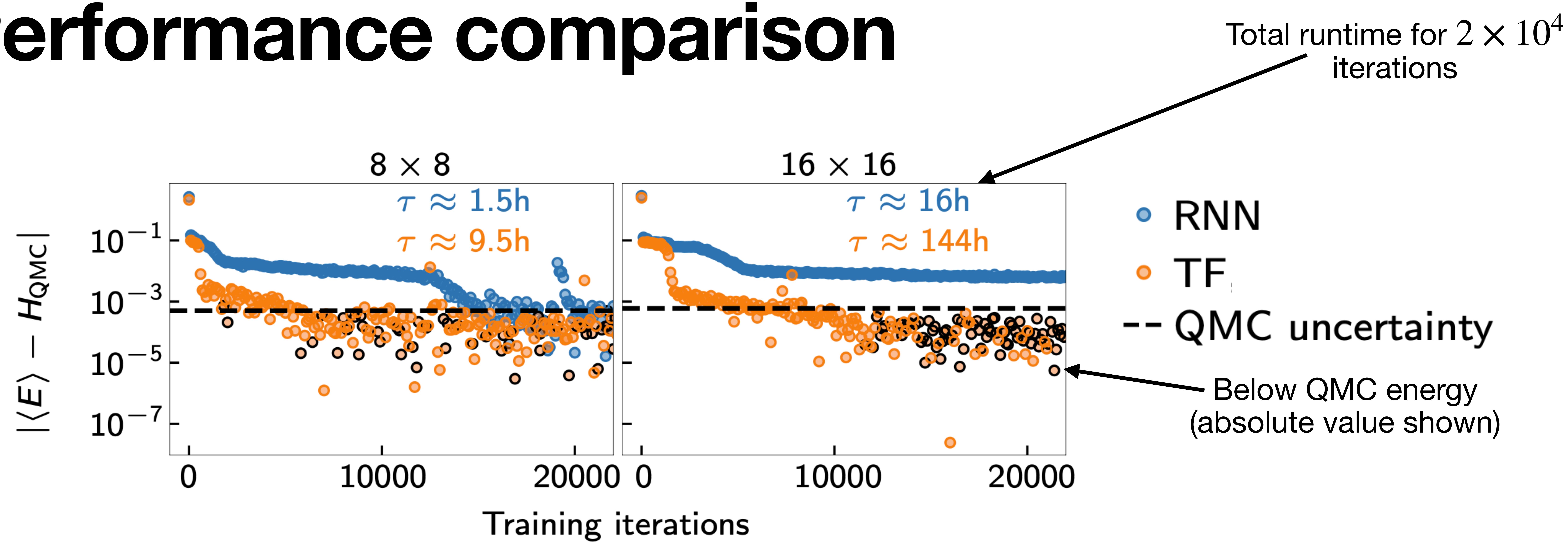
- Transformers outperform RNNs

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

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

Stefanie.Czischek@uOttawa.ca

# Performance comparison




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

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

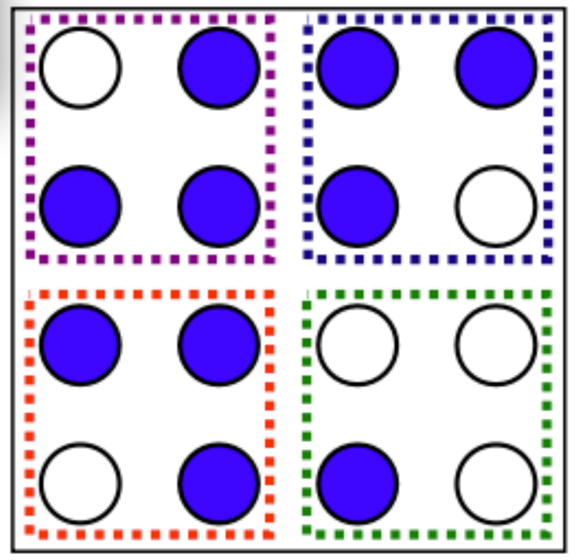
We cannot scale to larger system sizes!

# Patched neural network approach

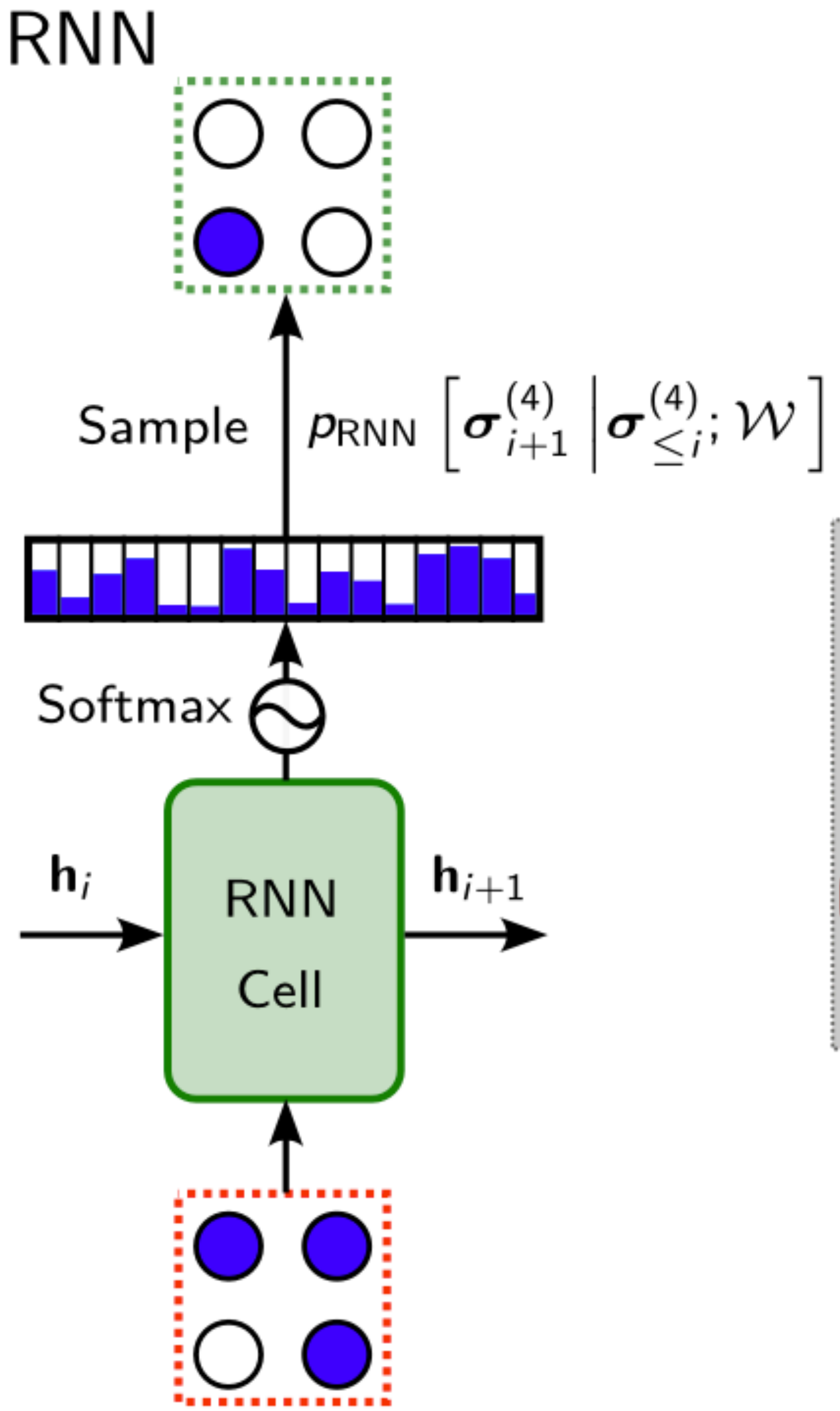
**Transformer variational wave functions for frustrated quantum spin systems**
Luciano Loris Viteritti, Riccardo Rende, Federico Becca

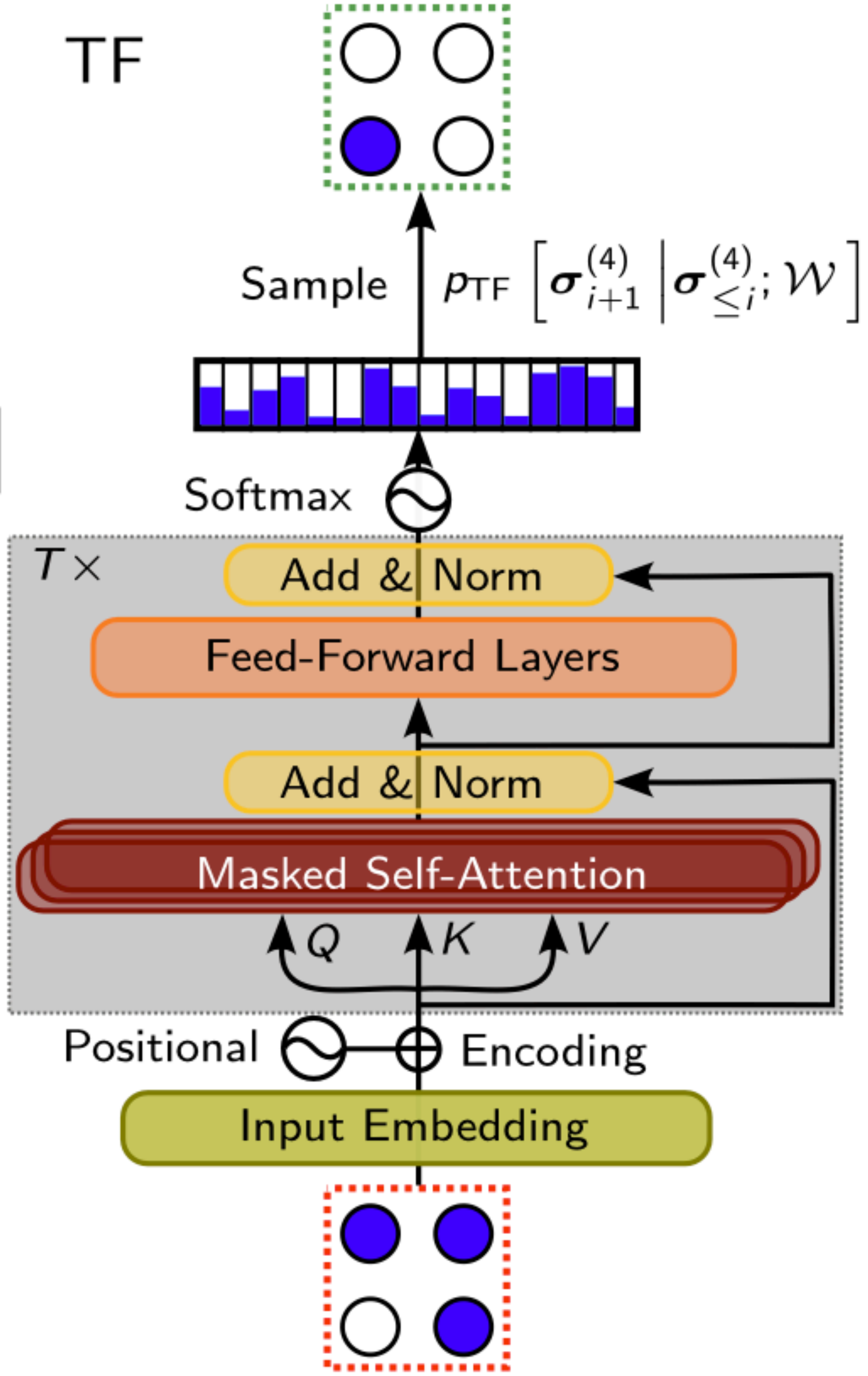
**Towards Neural Variational Monte Carlo That Scales Linearly with System Size**
Or Sharir, Garnet Kin-Lic Chan, Anima Anandkumar

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



- Input patch of $2 \times 2$ atoms
- Sample from output probability over $2^{(2\times2)}$ states




Stefanie.Czischek@uOttawa.ca

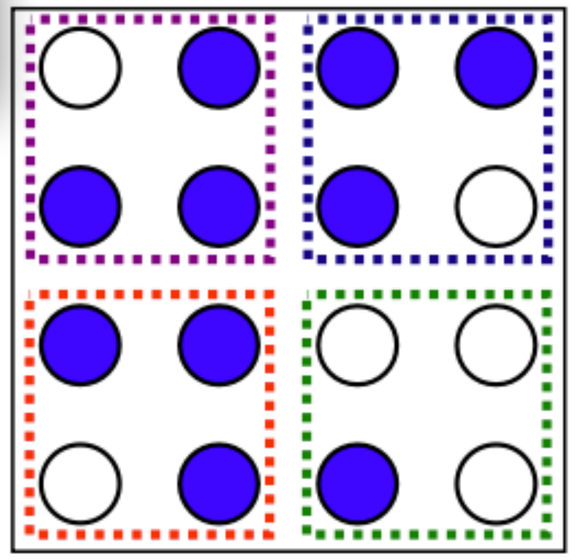
uOttawa

# Patched neural network approach

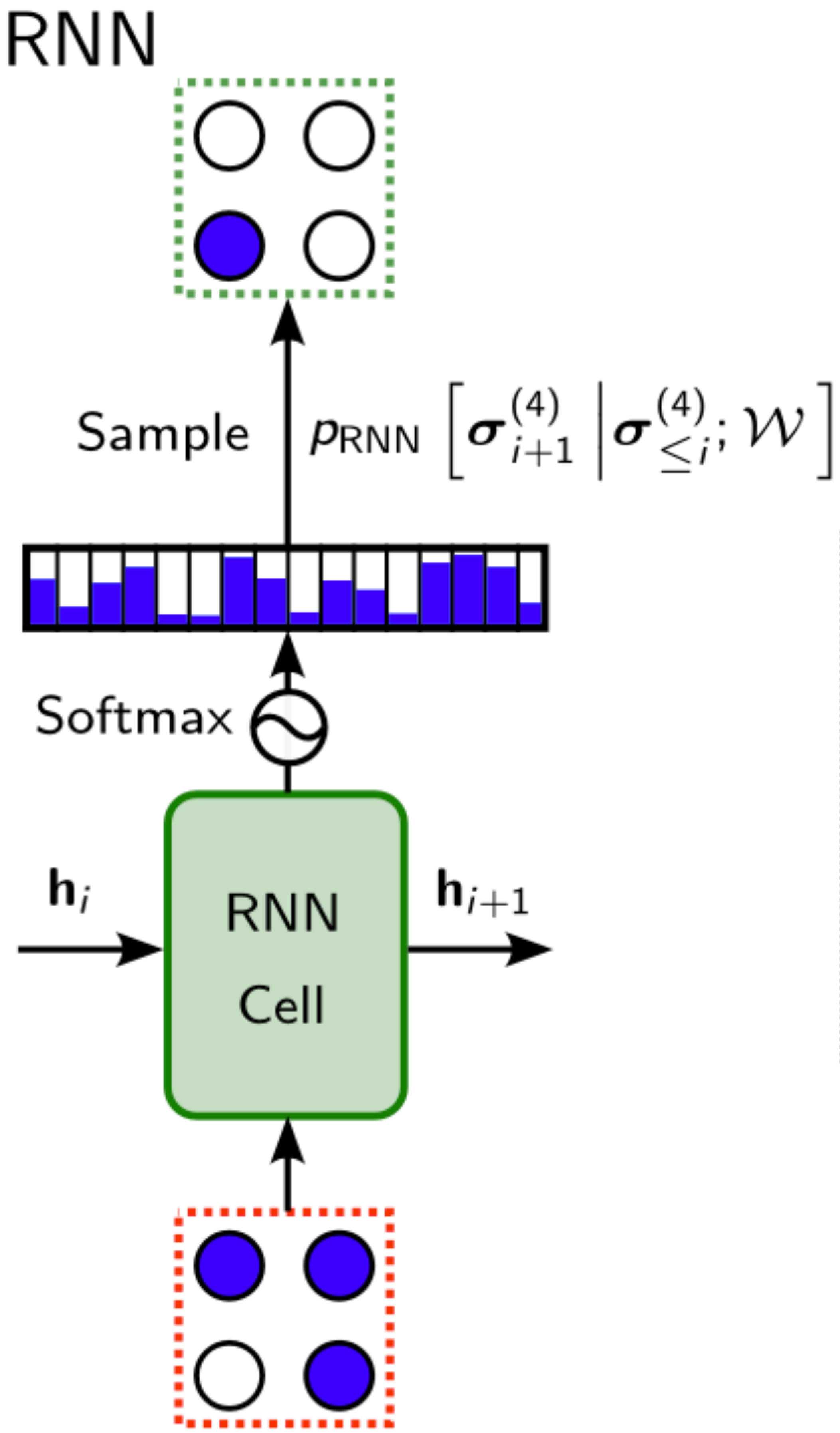
**Transformer variational wave functions for frustrated quantum spin systems**
Luciano Loris Viteritti, Riccardo Rende, Federico Becca

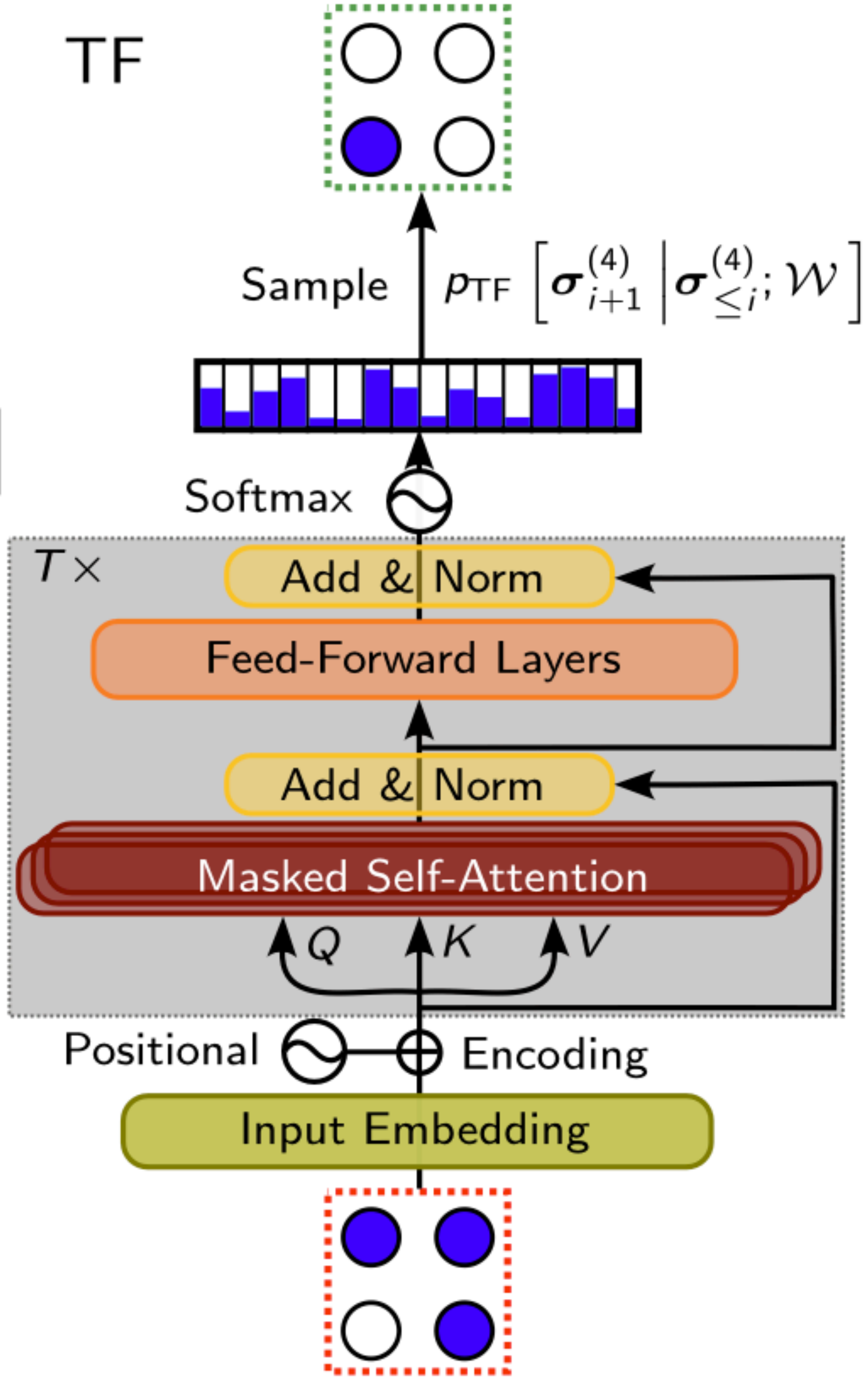
**Towards Neural Variational Monte Carlo That Scales Linearly with System Size**
Or Sharir, Garnet Kin-Lic Chan, Anima Anandkumar

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



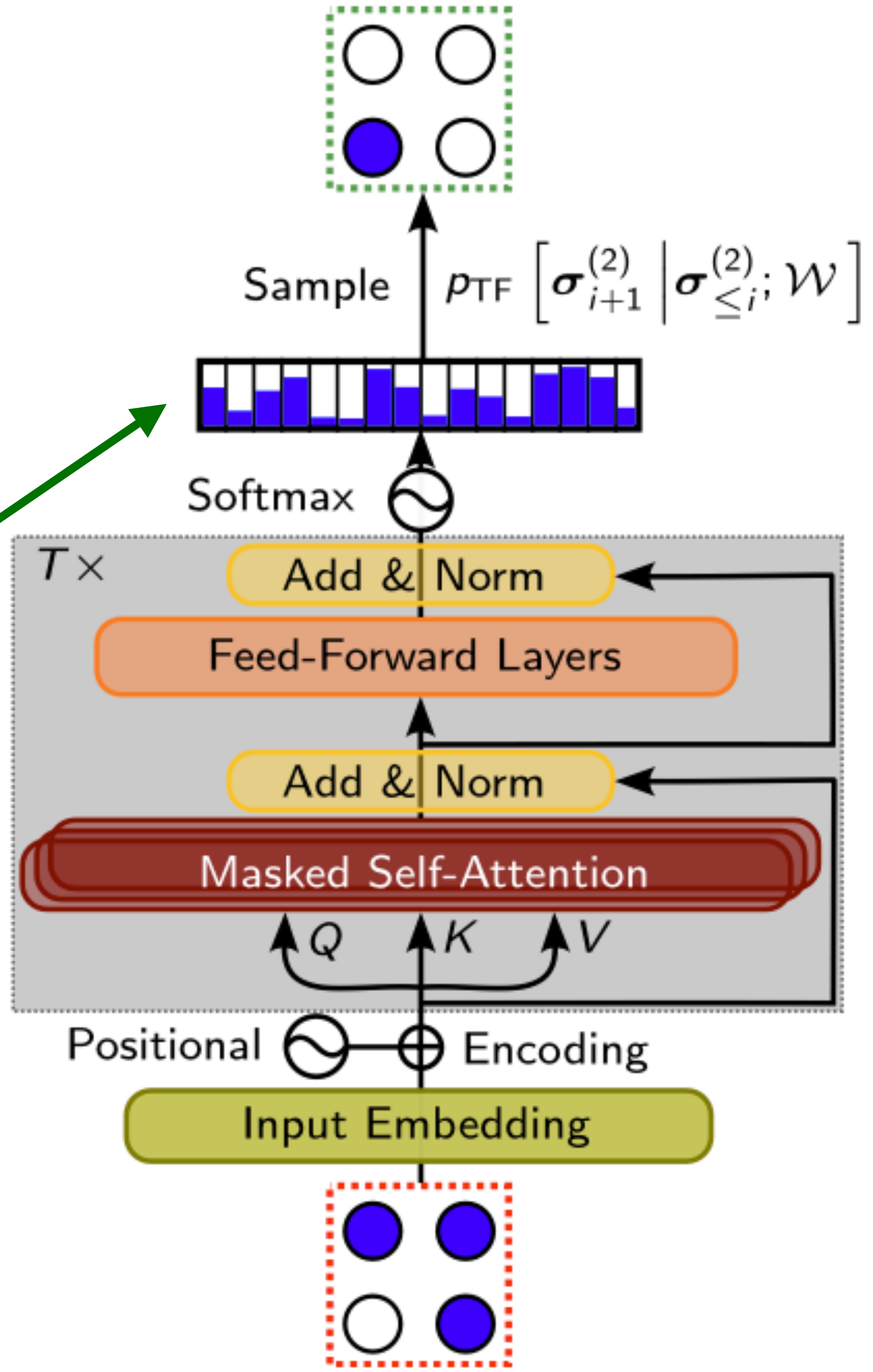
- Input patch of $2 \times 2$ atoms
- Sample from output probability over $2^{(2\times 2)}$ states
- Sequence length divided by four
- Local correlations directly encoded



# Large, patched transformers

- Larger input patches
  - Shorter runtimes
  - Comparable accuracies

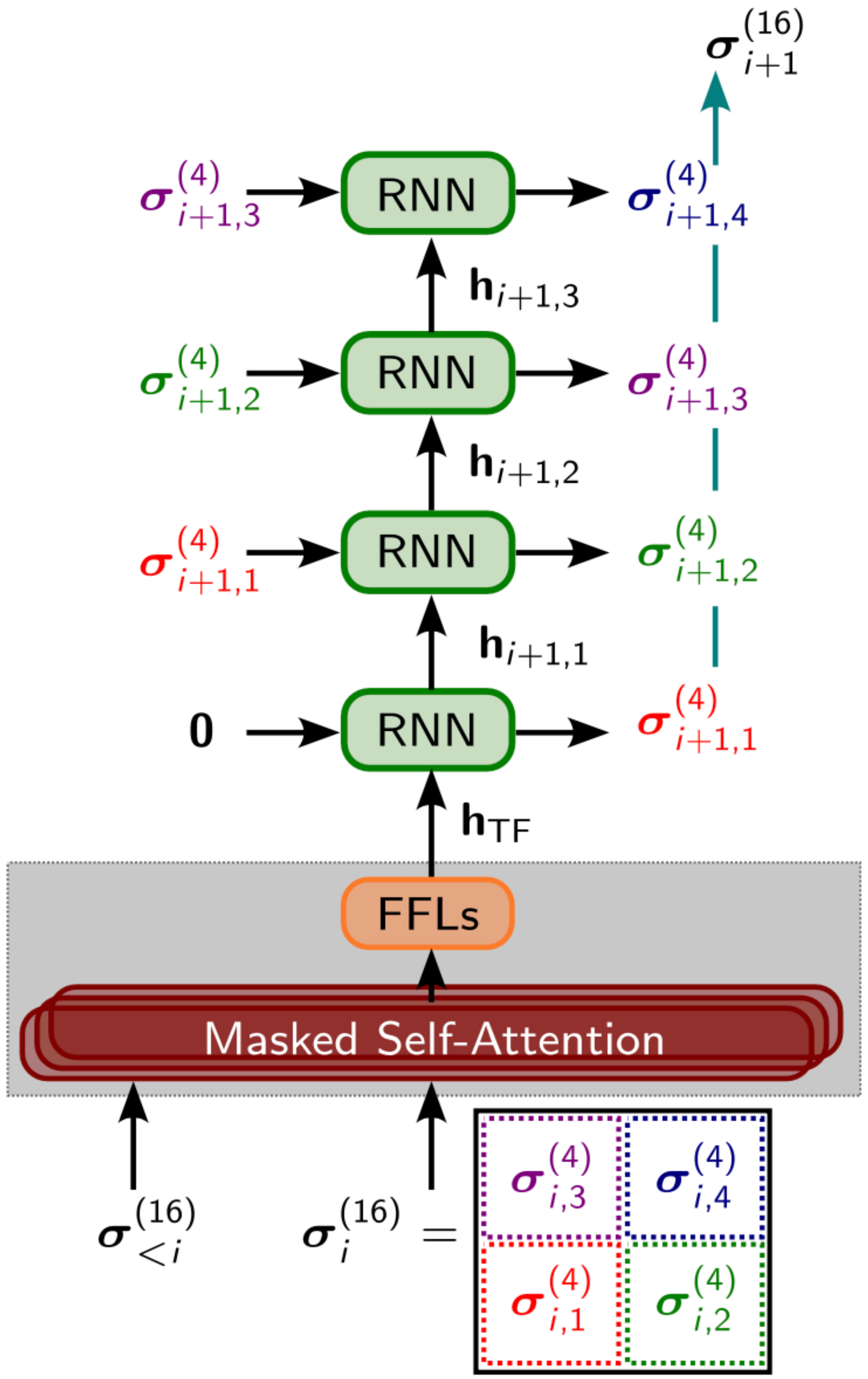


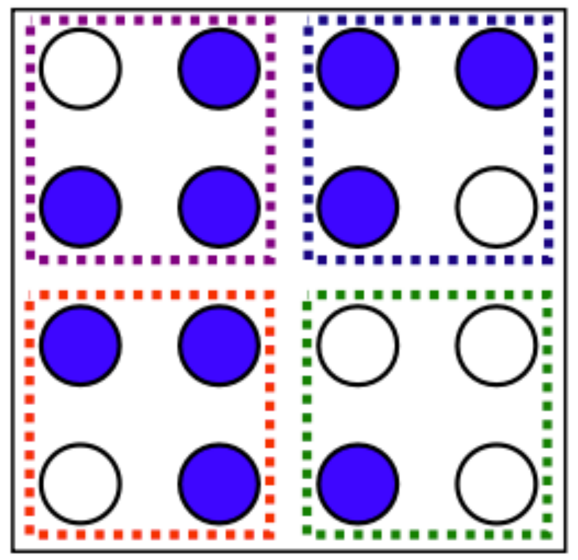

Output dimension scales exponentially
with the patch size!

Stefanie.Czischek@uOttawa.ca

# Large, patched transformers

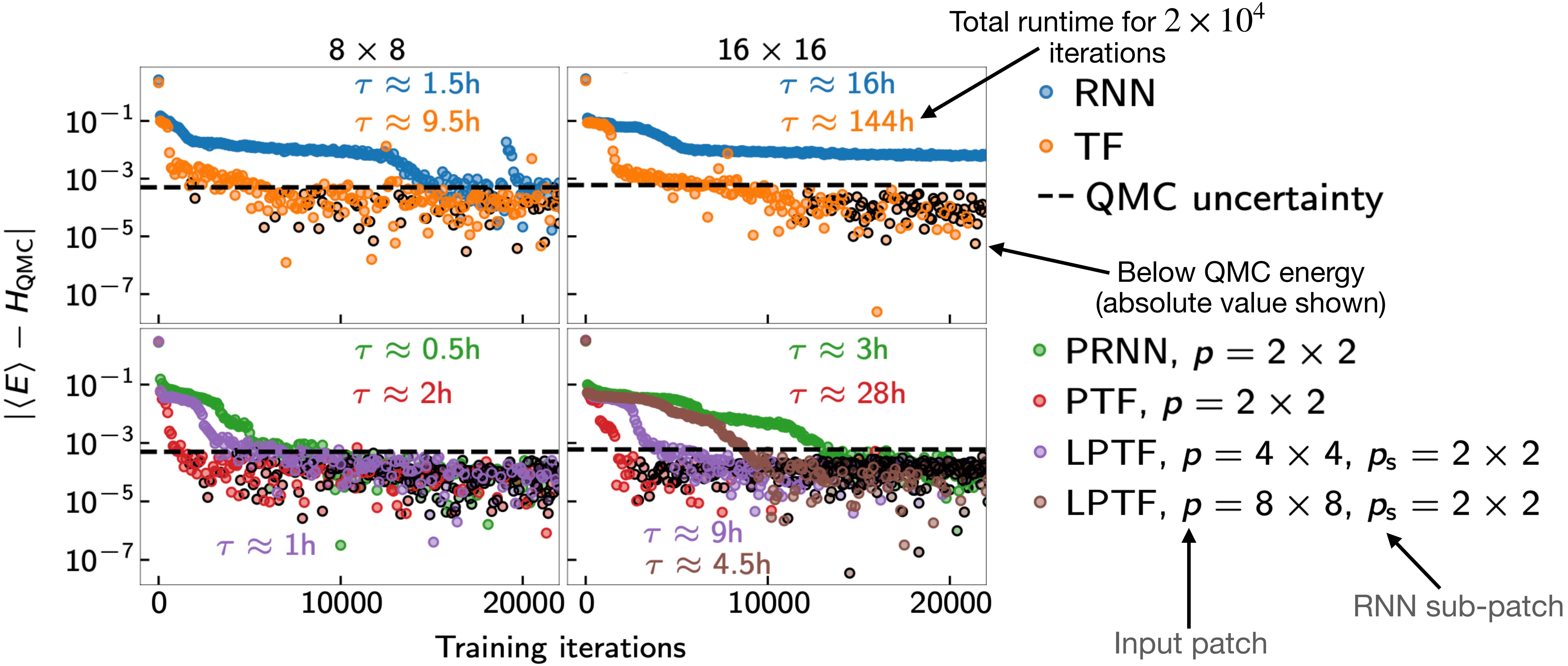
- Larger input patches
    - Shorter runtimes
    - Comparable accuracies
- Use an additional RNN to break down patch size
    - Gain power of transformer on large patches
    - Efficient RNN reduces output size



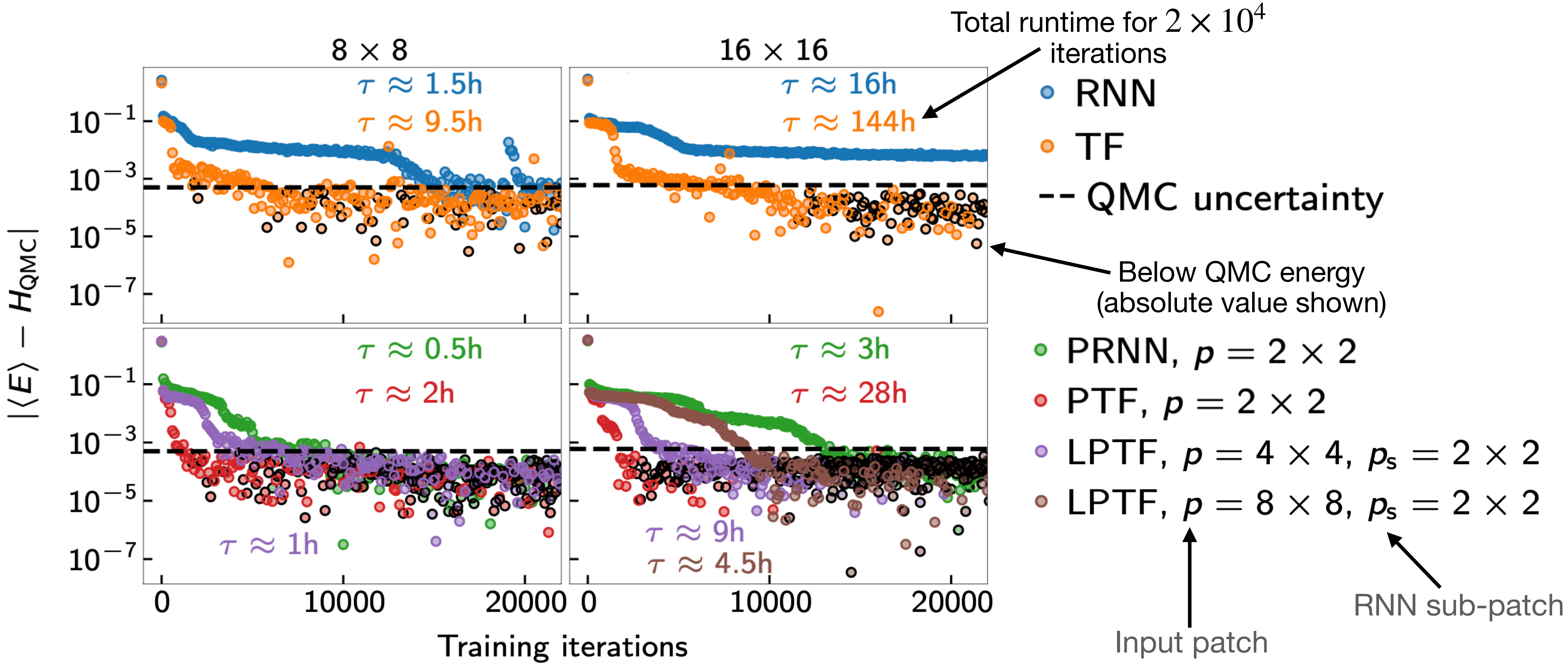



Stefanie.Czischek@uOttawa.ca

# Large, patched transformer: performance

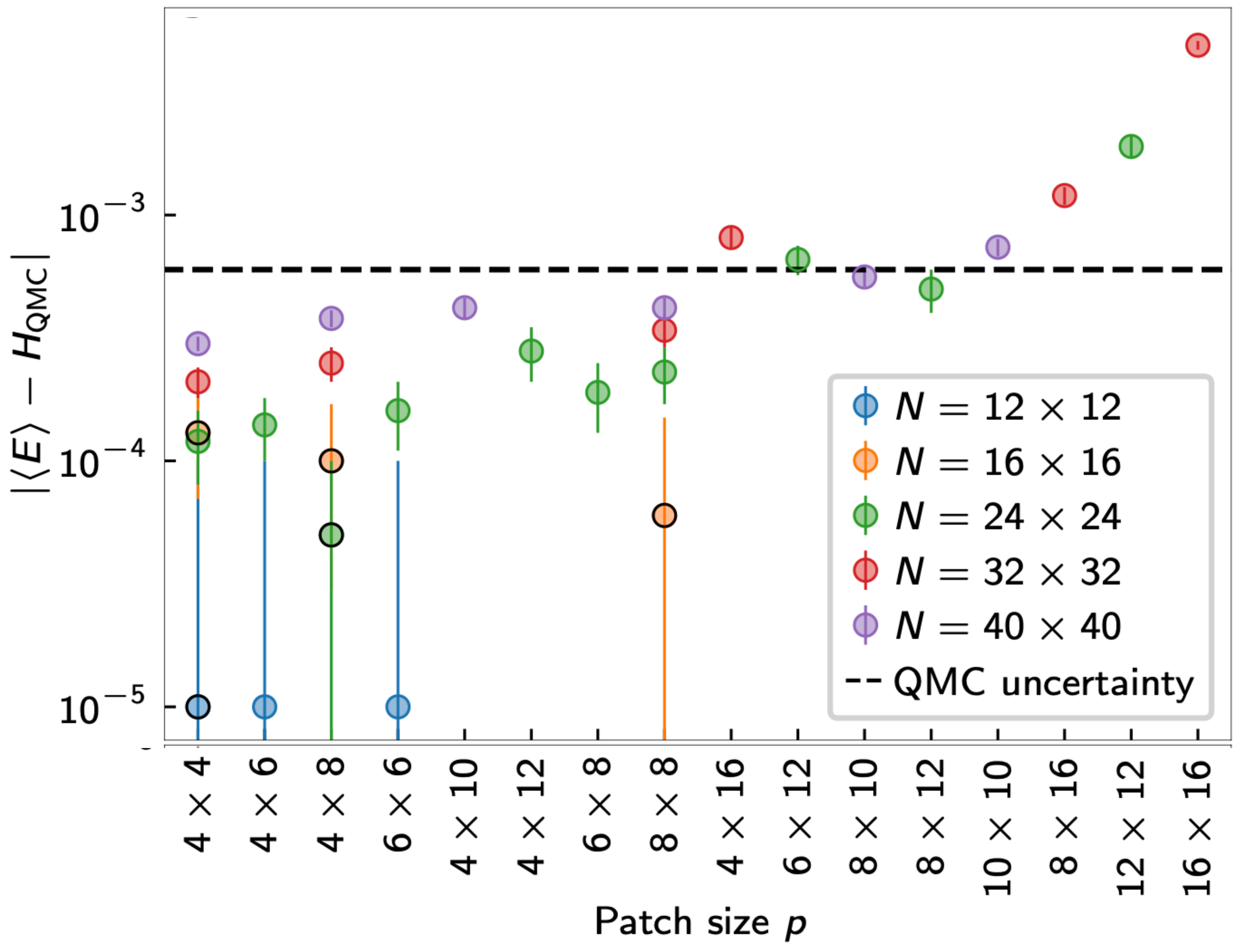


# Large, patched transformer: performance



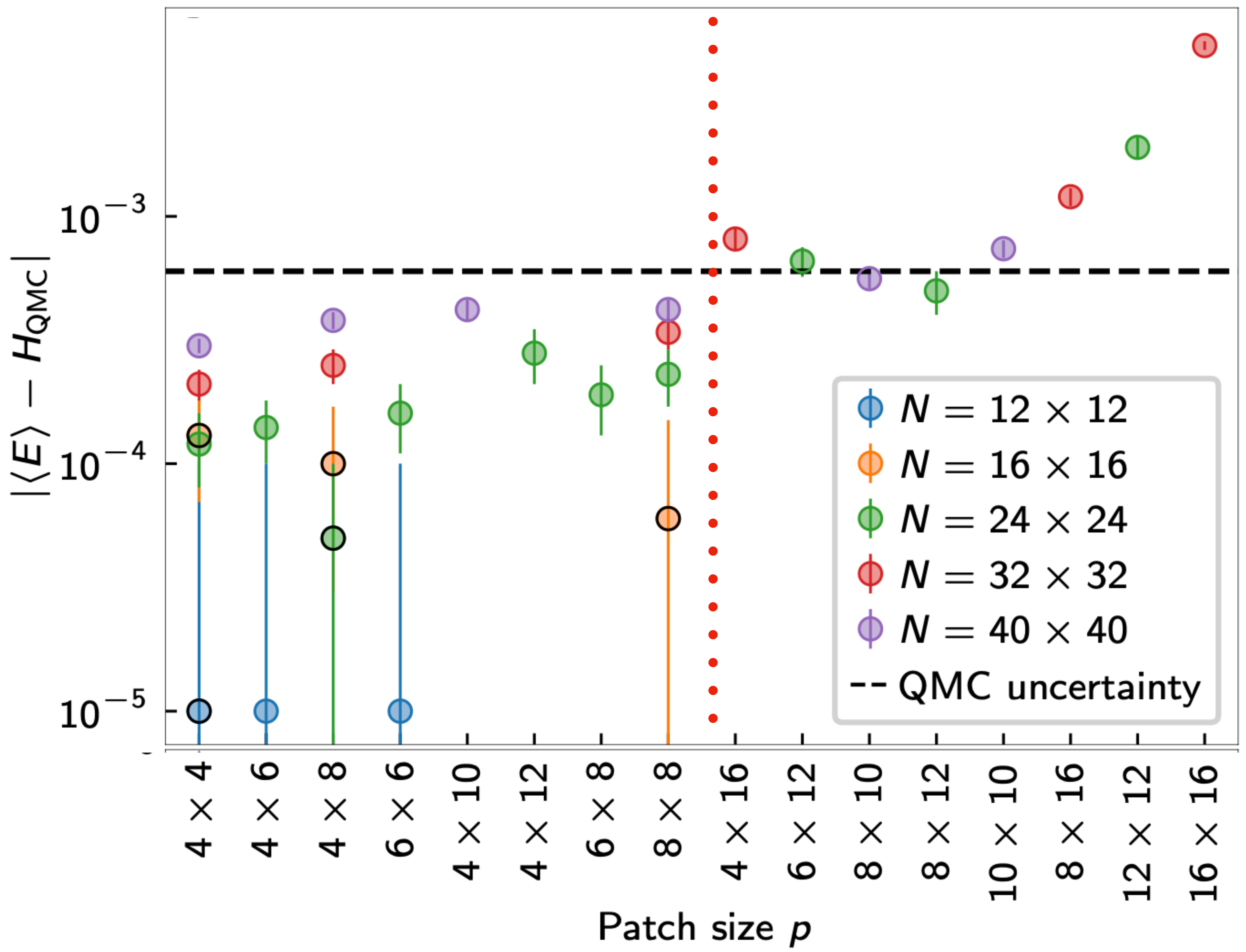

Reasonable runtimes and high accuracies!

# Going bigger




Stefanie.Czischek@uOttawa.ca

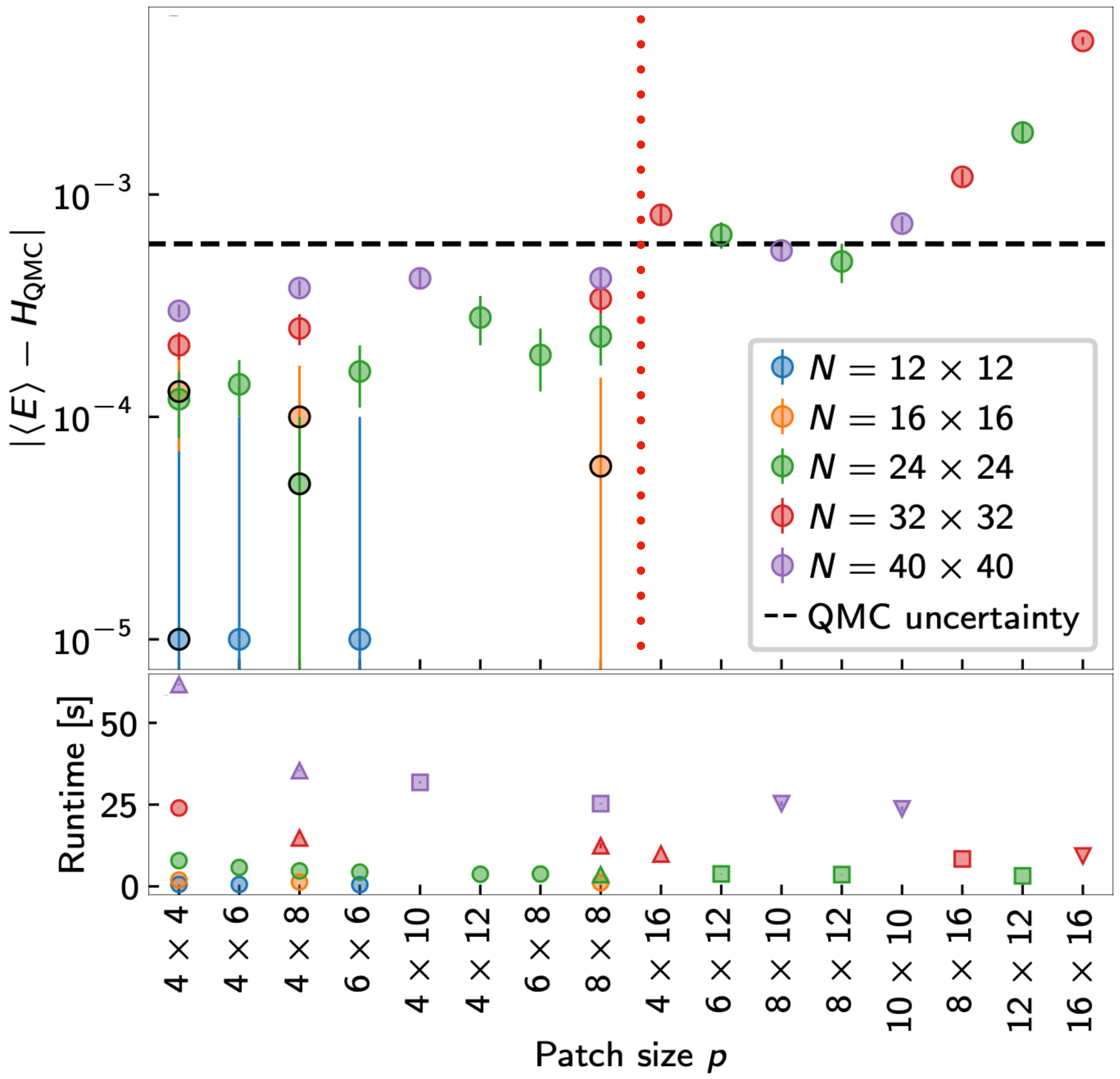
# Going bigger




- Patch too large: RNN expressivity limited and amount of information increased
- Accuracies below QMC uncertainty below $p = 8 \times 8$

Stefanie.Czischek@uOttawa.ca

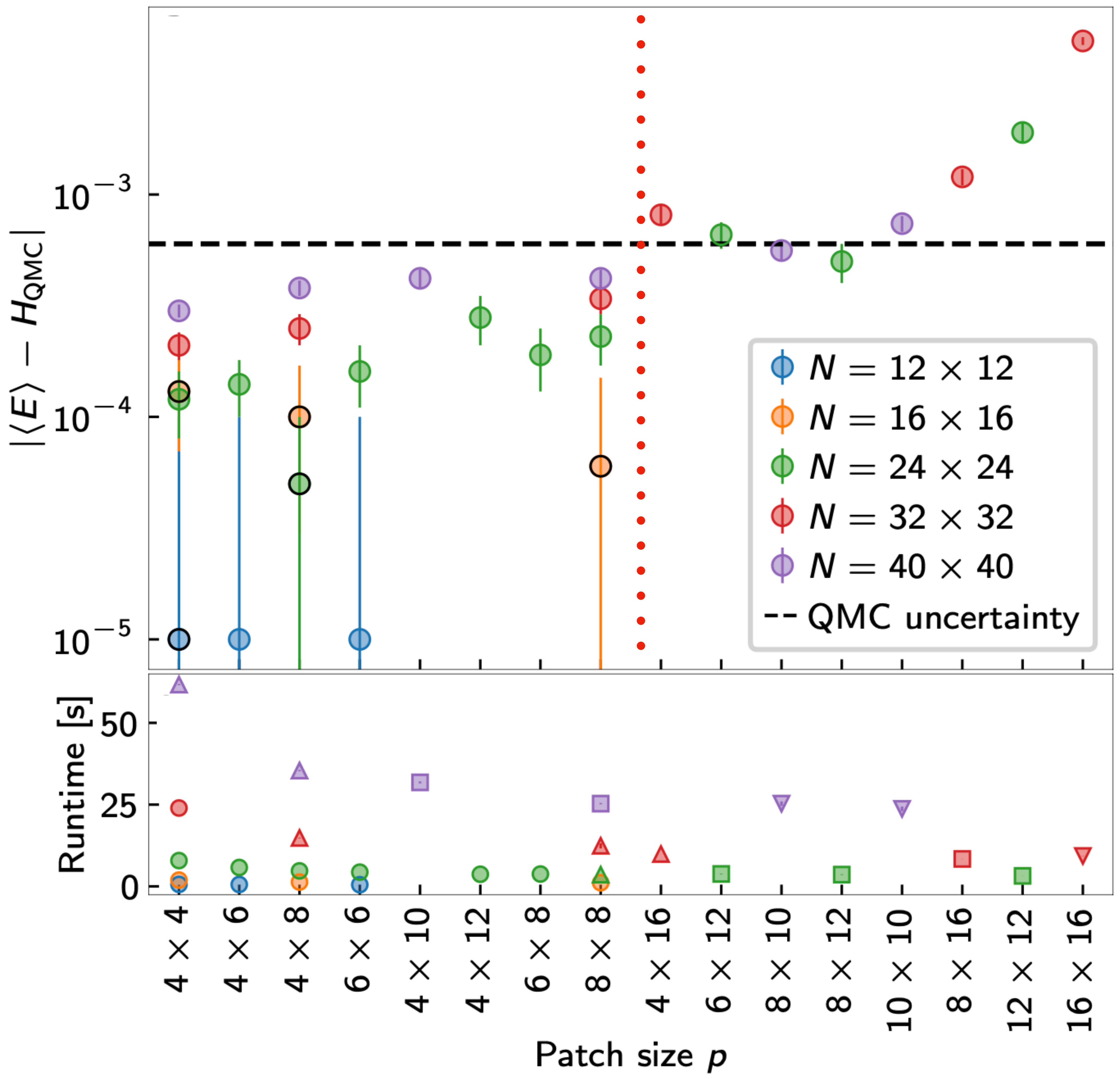
# Going bigger




- 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)

Stefanie.Czischek@uOttawa.ca

# Going bigger



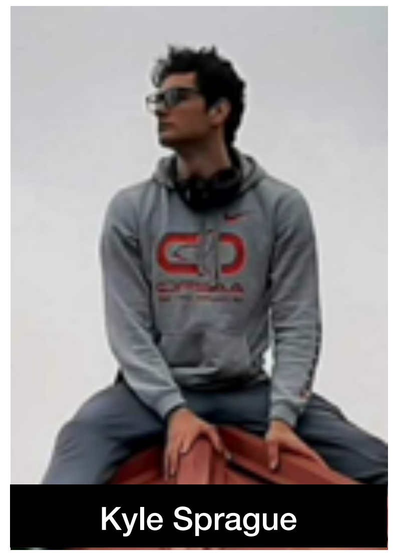

- Patch too large: RNN expressivity limited and amount of information increased
- Accuracies below QMC uncertainty below $p = 8 \times 8$
- Run times saturate for large patches (implementation detail)

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

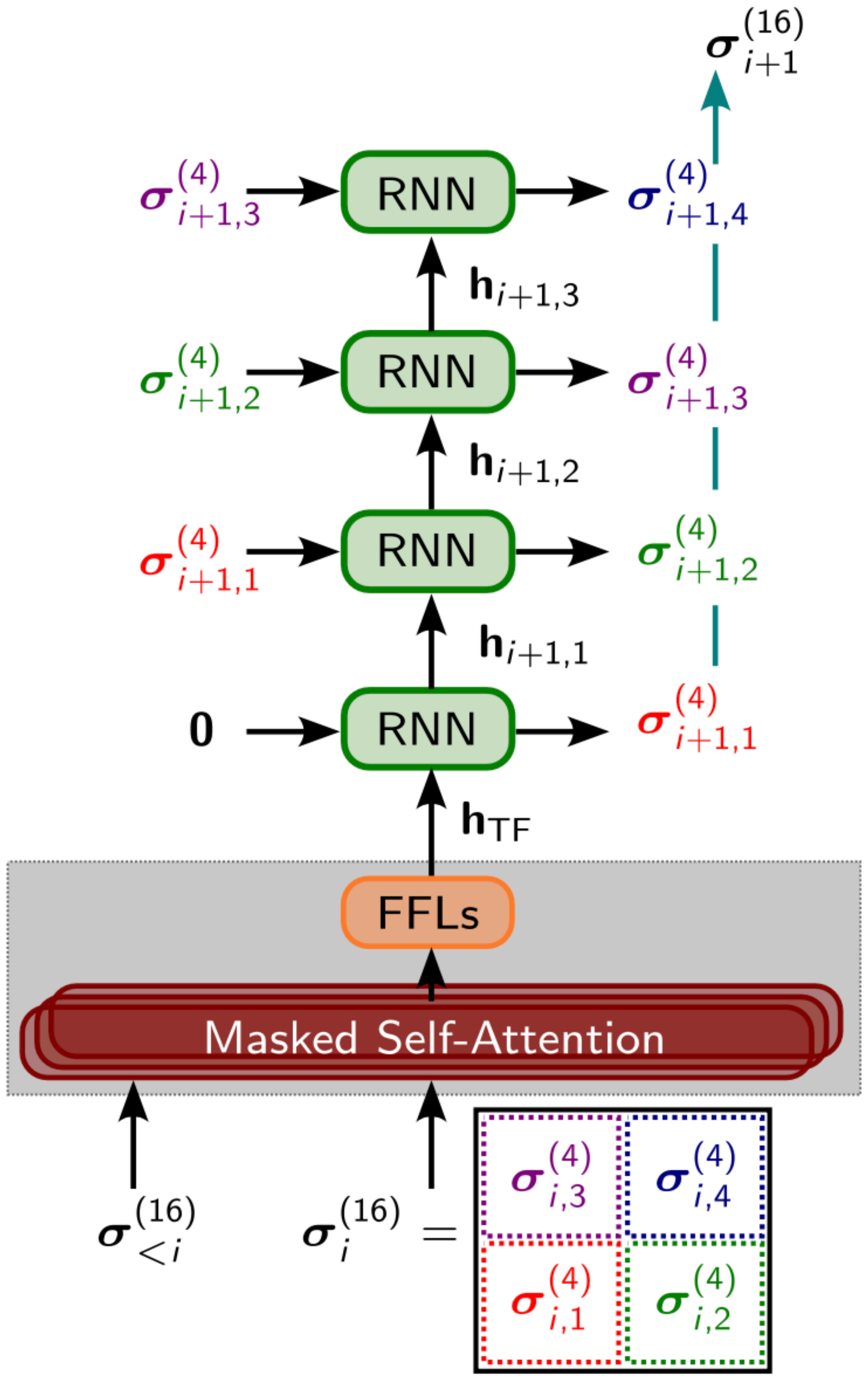
Stefanie.Czischek@uOttawa.ca

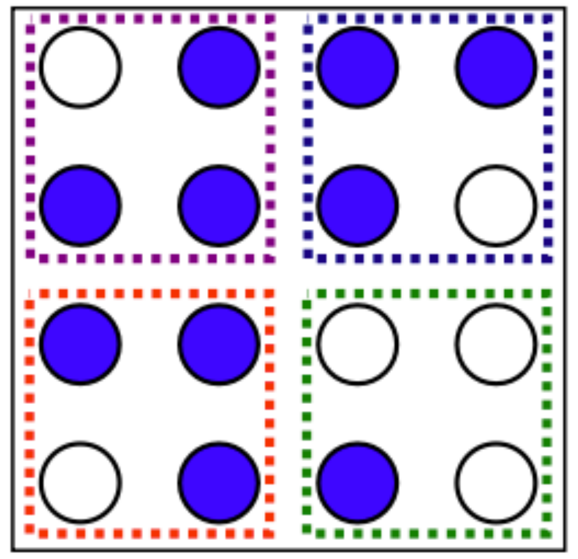
# Summary

[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