Combining data-driven and Hamiltonian-driven training for learning Quantum Ground States

Schuyler Moss

September 12th, 2022

UNIVERSITY OF WATERLOO

PI QUANTUM INTELLIGENCE LAB

International Conference on Recent Progress in Many-Body Theories XXI

Combining data-driven and Hamiltonian-driven training for learning Quantum Ground States

Schuyler Moss

September 12th, 2022

Stefanie Czischek, **M. Schuyler Moss**, et al. Phys. Rev. B **105**, 205108 – Published 9 May 2022

https://arxiv.org/abs/2203.04988

Ejaaz Merali (Not Pictured)

UNIVERSITY OF WATERLOO

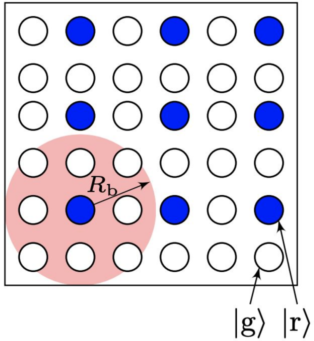
PI QUANTUM INTELLIGENCE LAB

International Conference on Recent Progress in Many-Body Theories XXI

Learning Quantum Ground states

Of what?

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



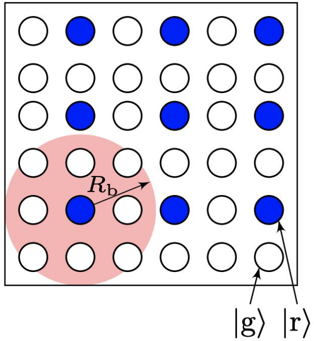
Learning Quantum Ground states

Of what?

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

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

“spin flip”
(off diagonal)



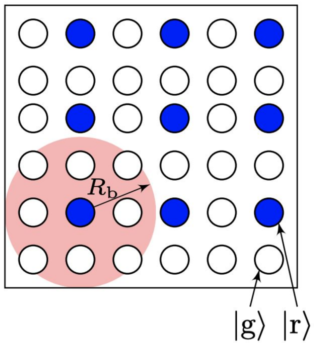
Learning Quantum Ground states

Of what?

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

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

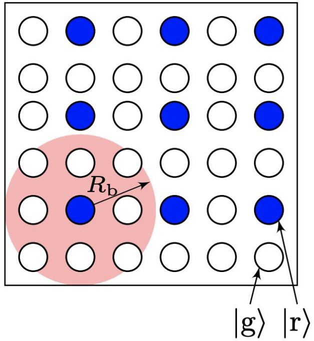
occupation operator
(diagonal)



Learning Quantum Ground states

Of what?

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



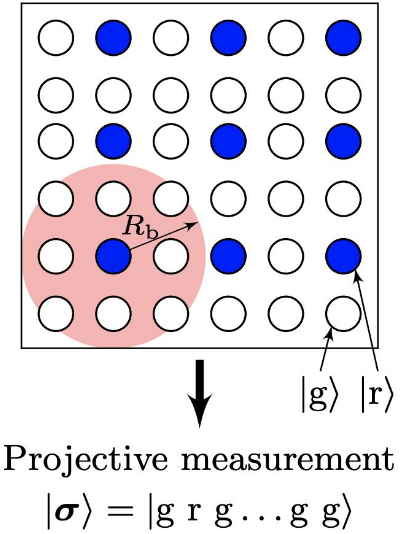
$$V_{ij} = \frac{\Omega R_b^6}{|\mathbf{r}_i - \mathbf{r}_j|^6}$$

Van der Waals potential

Learning Quantum Ground states

Of what?

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



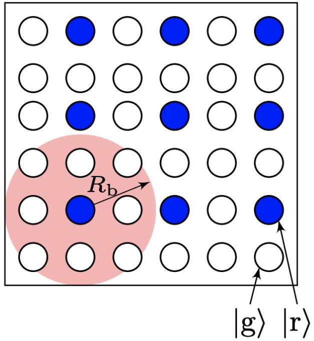
UNIVERSITY OF WATERLOO

QUANTUM INTELLIGENCE LAB

Learning Quantum Ground states

Of what?

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



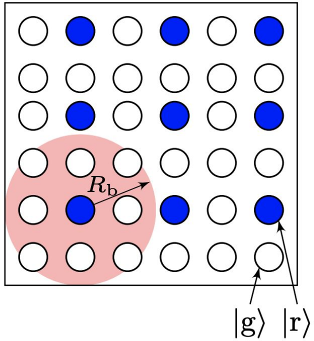
Rydberg Basis:

- Hamiltonian is Stoquastic
- Measurements in this basis are informationally complete

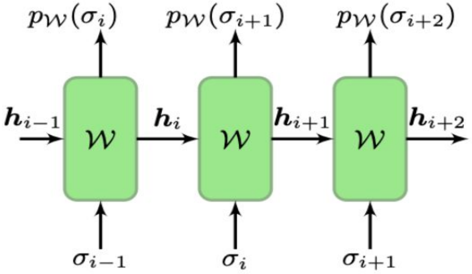
Learning Quantum Ground states

Of what?

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



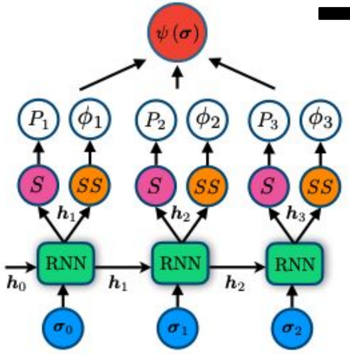
With what?



where $\quad \Psi_{\mathcal{W}}(\boldsymbol{\sigma}) = \sqrt{\prod_i p_{\mathcal{W}}(\sigma_i)}$

Hibat-Allah et. al. Recurrent neural network wave functions. *Physical Review Research*, *2*(2), p.023358.

Aside #1 on RNN Wavefunctions



$$|\Psi\rangle = \sum_{\sigma} \exp(i\phi(\sigma))\sqrt{P(\sigma)}|\sigma\rangle$$

$$\phi(\sigma) = \sum_{n=1}^{N} \phi_n$$

$$P(\sigma) \equiv \prod_{n=1}^{N} P_n$$

Hibat-Allah et. al. Recurrent neural network wave functions. *Physical Review Research*, *2*(2), p.023358.

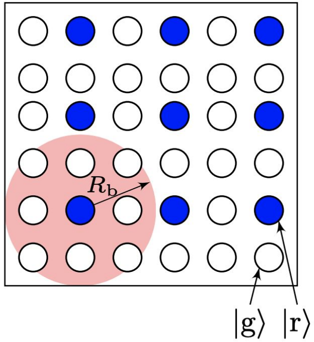
Not limited to Stoquastic Hamiltonians!!!!

Aside #2 on RNN Wavefunctions

The many nice properties:

- Autoregressive Neural Network
  - Chain rule of probabilities: $$P(x_1, x_2, ..., x_N) = \prod_i^N p(x_i | x_{j<i})$$
  - Efficient sampling
  - Encodes a ***normalized*** probability distribution

- Natural interpretation for lattice systems



UNIVERSITY OF WATERLOO

PI QUANTUM INTELLIGENCE LAB

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

$$\downarrow$$

$$H_{\mathrm{RNN}} = \frac{\langle\Psi_{\mathcal{W}}|\hat{H}|\Psi_{\mathcal{W}}\rangle}{\langle\Psi_{\mathcal{W}}|\Psi_{\mathcal{W}}\rangle}$$

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

$$H_{\text{RNN}} = \frac{\sum_{\sigma} |\langle \Psi_{\mathcal{W}} | \sigma \rangle|^2 \frac{\langle \sigma | \hat{H} | \Psi_{\mathcal{W}} \rangle}{\langle \sigma | \Psi_{\mathcal{W}} \rangle}}{\sum_{\sigma'} |\langle \Psi_{\mathcal{W}} | \sigma' \rangle|^2}$$

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

$$H_{\mathrm{RNN}} = \frac{\sum_\sigma \boxed{|\langle \Psi_{\mathcal{W}}|\sigma\rangle|^2} \frac{\langle \sigma|\hat{H}|\Psi_{\mathcal{W}}\rangle}{\langle \sigma|\Psi_{\mathcal{W}}\rangle}}{\boxed{\sum_{\sigma'} |\langle \Psi_{\mathcal{W}}|\sigma'\rangle|^2}}$$

$P(\sigma)$

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

$$H_{\text{RNN}} = \frac{\sum_\sigma |\langle\Psi_{\mathcal{W}}|\sigma\rangle|^2 \boxed{\frac{\langle\sigma|\hat{H}|\Psi_{\mathcal{W}}\rangle}{\langle\sigma|\Psi_{\mathcal{W}}\rangle}}}{\sum_{\sigma'}|\langle\Psi_{\mathcal{W}}|\sigma'\rangle|^2}$$

$$H_{\text{loc}}(\sigma)$$

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

$$H_{\mathrm{RNN}} = \sum_{\sigma} P(\sigma)H_{\mathrm{loc}}(\sigma)$$

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

$$H_{\mathrm{RNN}} = \sum_{\sigma} P(\sigma) H_{\mathrm{loc}}(\sigma)$$

$$H_{RNN} \approx \frac{1}{N_s}\sum_{\boldsymbol{\sigma}\sim p_{RNN}(\boldsymbol{\sigma};\mathcal{W})} H_{loc}(\boldsymbol{\sigma})$$

Hamiltonian-Driven

Carleo, Giuseppe, and Matthias Troyer. "Solving the quantum many-body problem with artificial neural networks." *Science* 355.6325 (2017): 602-606.

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

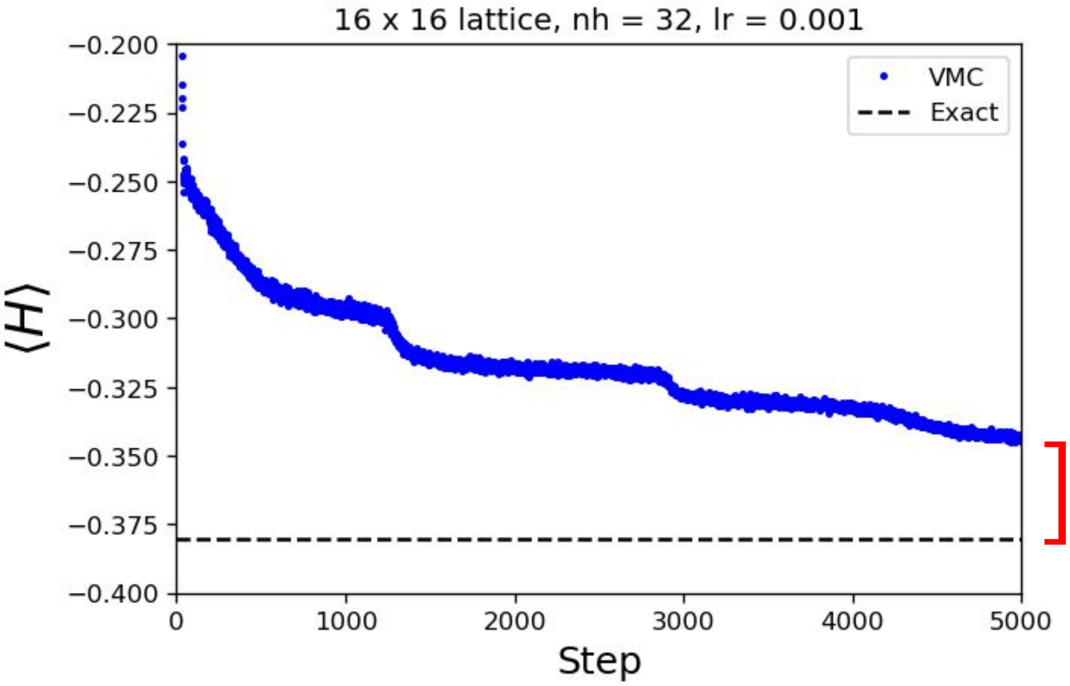
$$H_{\mathrm{RNN}} = \sum_{\boldsymbol{\sigma}} P(\sigma) H_{\mathrm{loc}}(\sigma)$$

$$H_{RNN} \approx \boxed{\frac{1}{N_s}\sum_{\boldsymbol{\sigma}\sim p_{RNN}(\boldsymbol{\sigma};\mathcal{W})} H_{loc}(\boldsymbol{\sigma}) = \mathcal{L}_H(\mathcal{W})}$$

UNIVERSITY OF WATERLOO

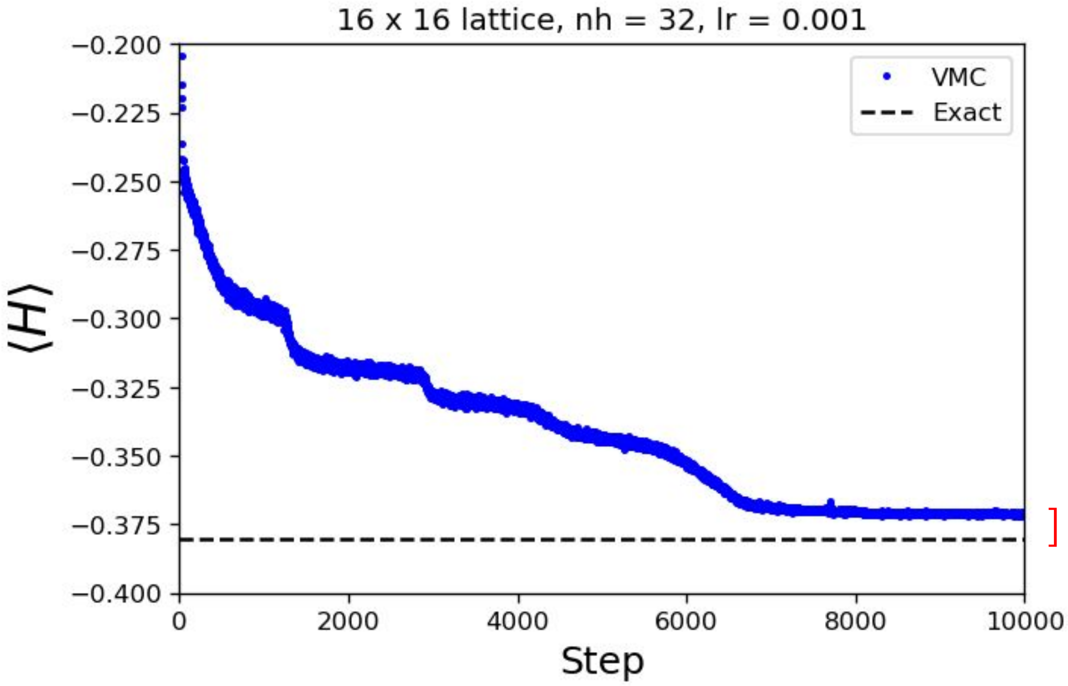
QUANTUM INTELLIGENCE LAB

Optimization Challenges (variational training)



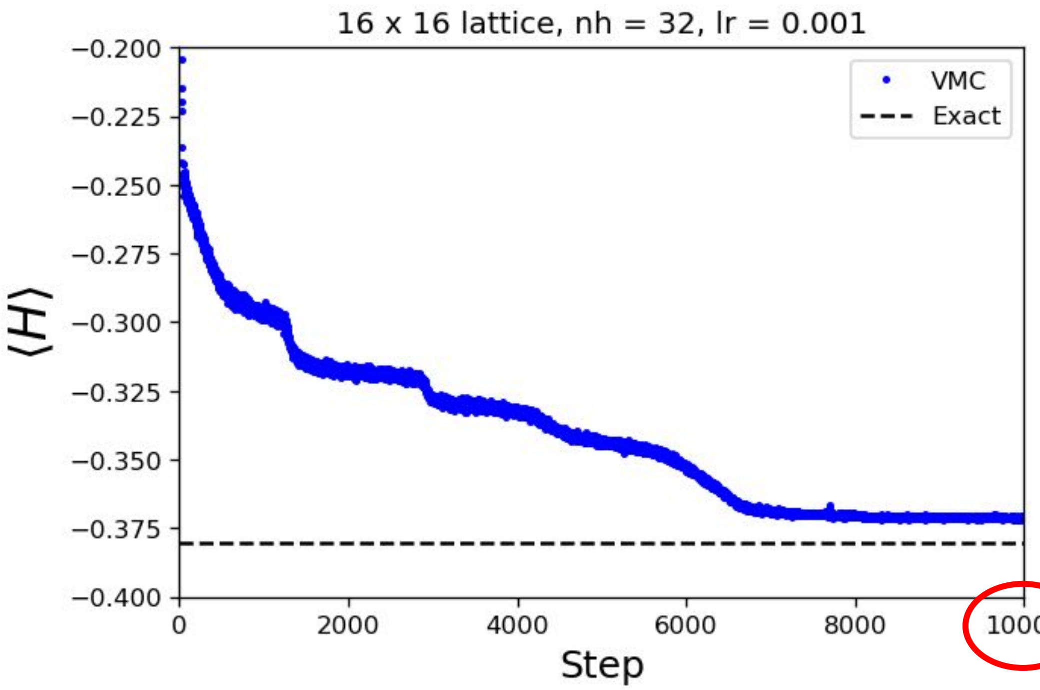
not great…

Optimization Challenges (variational training)



] better… but…

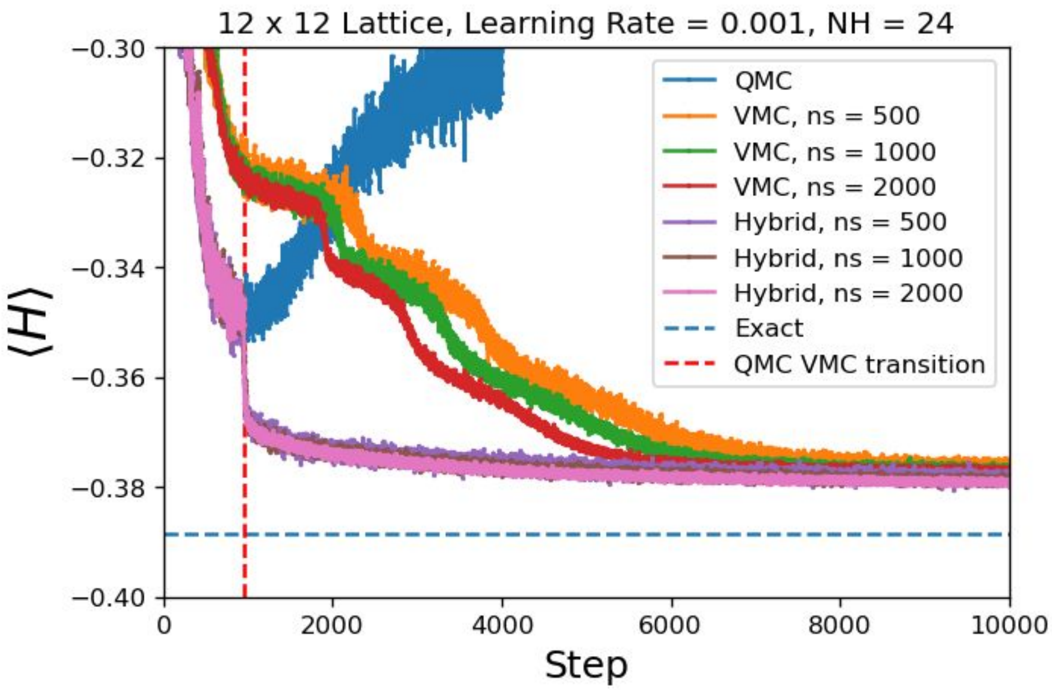
Optimization Challenges (variational training)



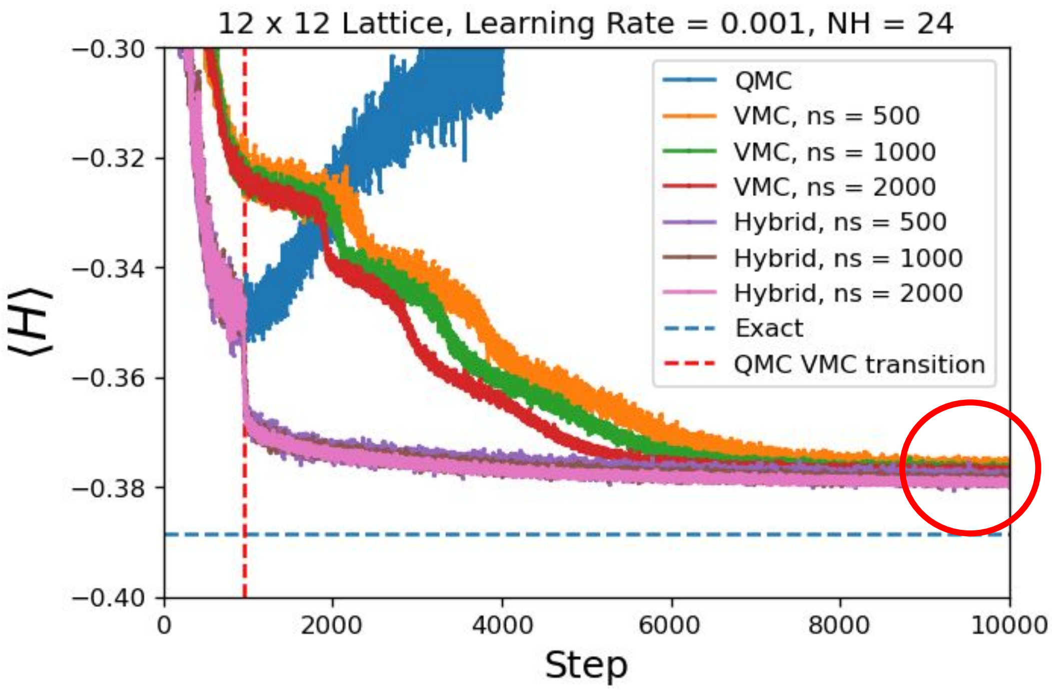
] better… but…

Very long runtime

Aside on Energy Estimation



Aside on Energy Estimation

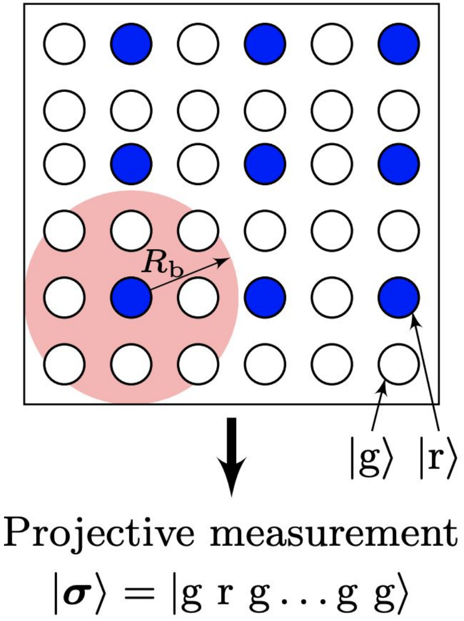


reasonable increases in ns don't result in significant improvements

Data-Driven Training

Torlai *et al.* Neural-network quantum state tomography. *Nature Phys* 14, 447–450 (2018).

Merali, Ejaaz *et al.* "Stochastic series expansion quantum monte carlo for rydberg arrays." *arXiv preprint arXiv:2107.00766* (2021).



Data-Driven Training

Torlai *et al.* Neural-network quantum state tomography. *Nature Phys* 14, 447–450 (2018).

$$\mathcal{L}_{KL}(\mathcal{W}) = \sum_{\{\boldsymbol{\sigma}\}} p_{\mathcal{D}}(\boldsymbol{\sigma}) \log \frac{p_{\mathcal{D}}(\boldsymbol{\sigma})}{p_{RNN}(\boldsymbol{\sigma}; \mathcal{W})}$$

Data-Driven Training

Torlai *et al.* Neural-network quantum state tomography. *Nature Phys* 14, 447–450 (2018).

$$\mathcal{L}_{KL}(\mathcal{W}) = \sum_{\{\boldsymbol{\sigma}\}} p_{\mathcal{D}}(\boldsymbol{\sigma}) \log \frac{p_{\mathcal{D}}(\boldsymbol{\sigma})}{p_{RNN}(\boldsymbol{\sigma};\mathcal{W})}$$

$$\approx -S_{\mathcal{D}} - \sum_{\boldsymbol{\sigma} \in \mathcal{D}} p_{\mathcal{D}}(\boldsymbol{\sigma}) \log p_{\mathrm{RNN}}(\boldsymbol{\sigma})$$

Data-Driven Training

Torlai *et al.* Neural-network quantum state tomography. *Nature Phys* 14, 447–450 (2018).

$$\mathcal{L}_{KL}(\mathcal{W}) = \sum_{\{\boldsymbol{\sigma}\}} p_{\mathcal{D}}(\boldsymbol{\sigma}) \log \frac{p_{\mathcal{D}}(\boldsymbol{\sigma})}{p_{RNN}(\boldsymbol{\sigma};\mathcal{W})}$$

$$\approx -S_{\mathcal{D}} - \sum_{\boldsymbol{\sigma}\in\mathcal{D}} p_{\mathcal{D}}(\boldsymbol{\sigma}) \log p_{\mathrm{RNN}}(\boldsymbol{\sigma})$$

$$\approx -S_{\mathcal{D}} - \frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{\sigma}\in\mathcal{D}} \log p_{RNN}(\boldsymbol{\sigma};\mathcal{W})$$

Data-Driven Training

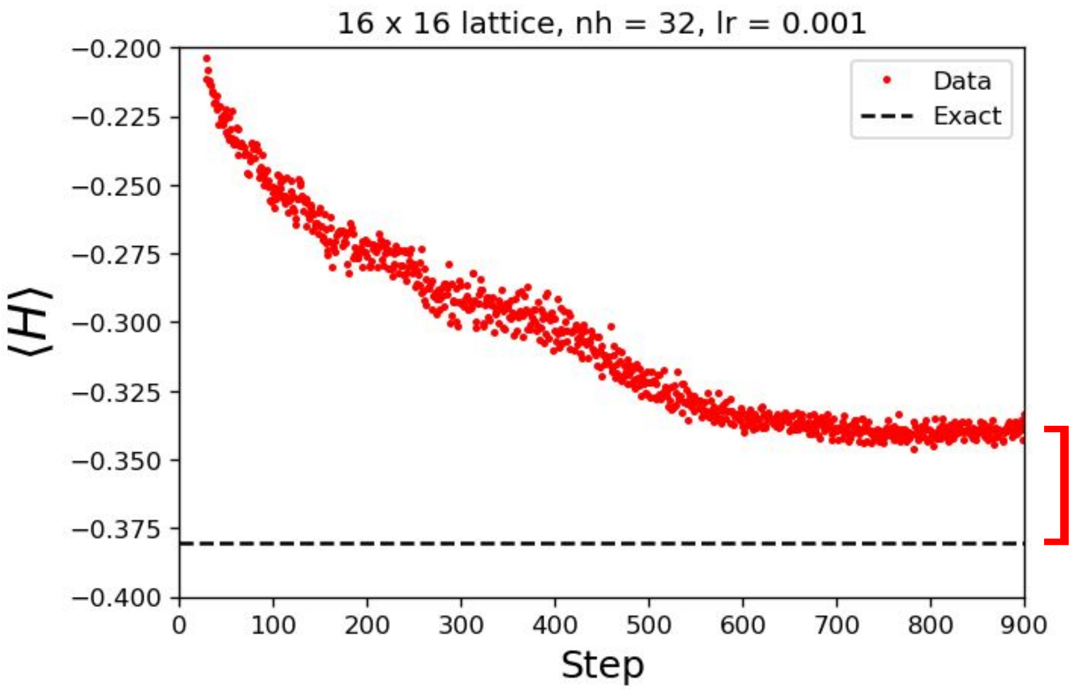
Torlai *et al.* Neural-network quantum state tomography. *Nature Phys* 14, 447–450 (2018).

$$\boxed{\mathcal{L}_{KL}(\mathcal{W})} = \sum_{\{\boldsymbol{\sigma}\}} p_{\mathcal{D}}(\boldsymbol{\sigma}) \log \frac{p_{\mathcal{D}}(\boldsymbol{\sigma})}{p_{RNN}(\boldsymbol{\sigma};\mathcal{W})}$$

$$\approx -S_{\mathcal{D}} - \sum_{\boldsymbol{\sigma}\in\mathcal{D}} p_{\mathcal{D}}(\boldsymbol{\sigma}) \log p_{\mathrm{RNN}}(\boldsymbol{\sigma})$$

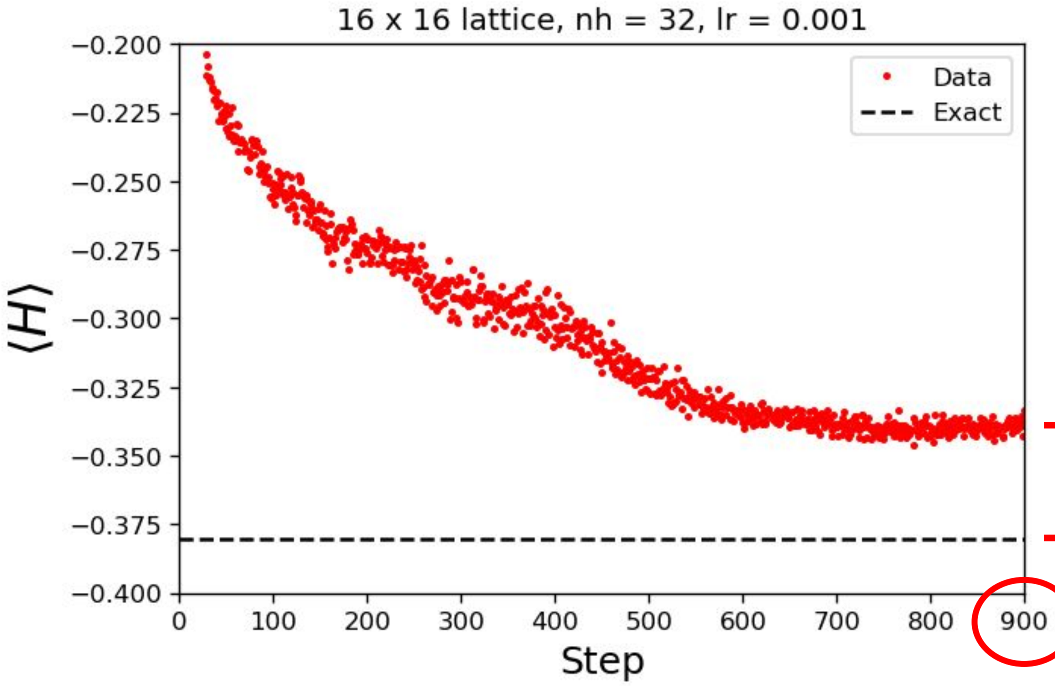
$$\approx -S_{\mathcal{D}} \boxed{- \frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{\sigma}\in\mathcal{D}} \log p_{RNN}(\boldsymbol{\sigma};\mathcal{W})}$$

Optimization Challenges (training with data)



Again,
not great…

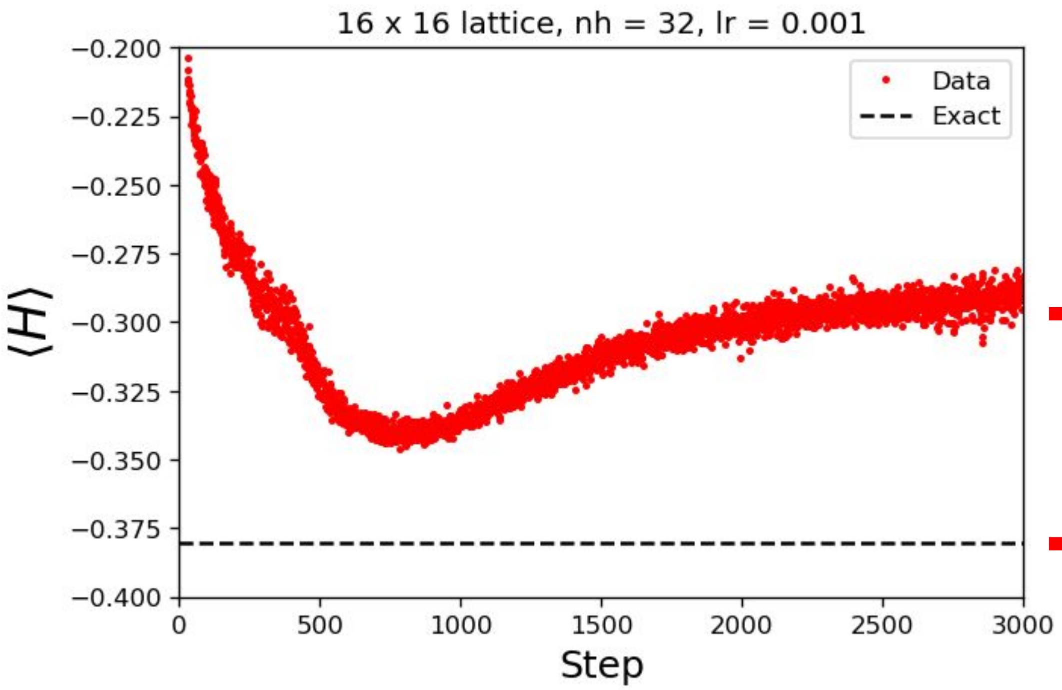
Optimization Challenges (training with data)



Again, not great…

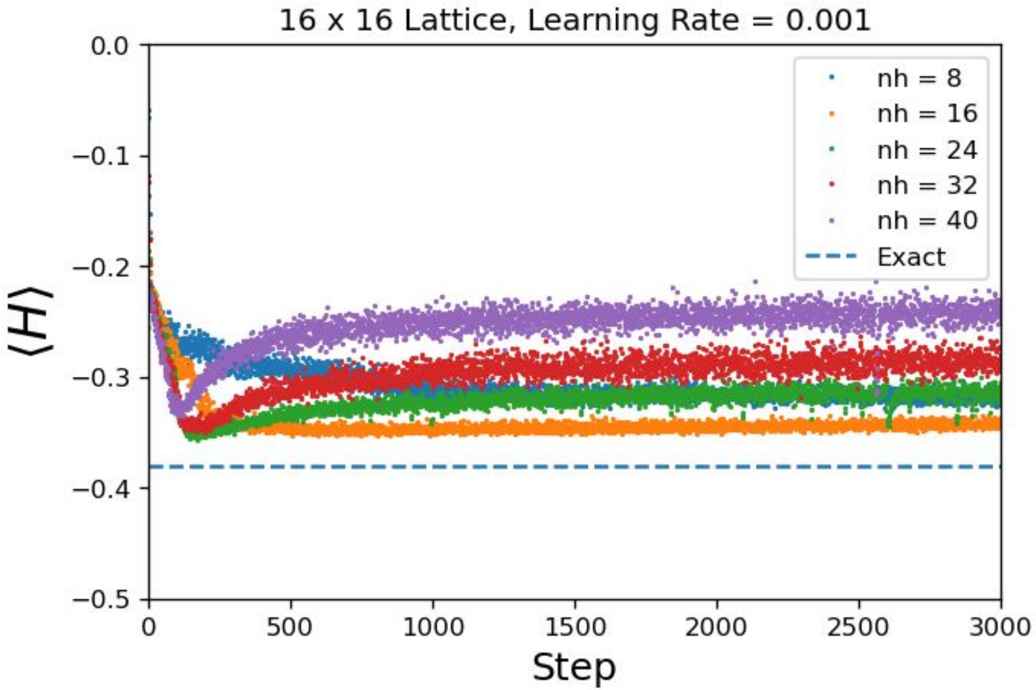
but… much faster.

Optimization Challenges (training with data)

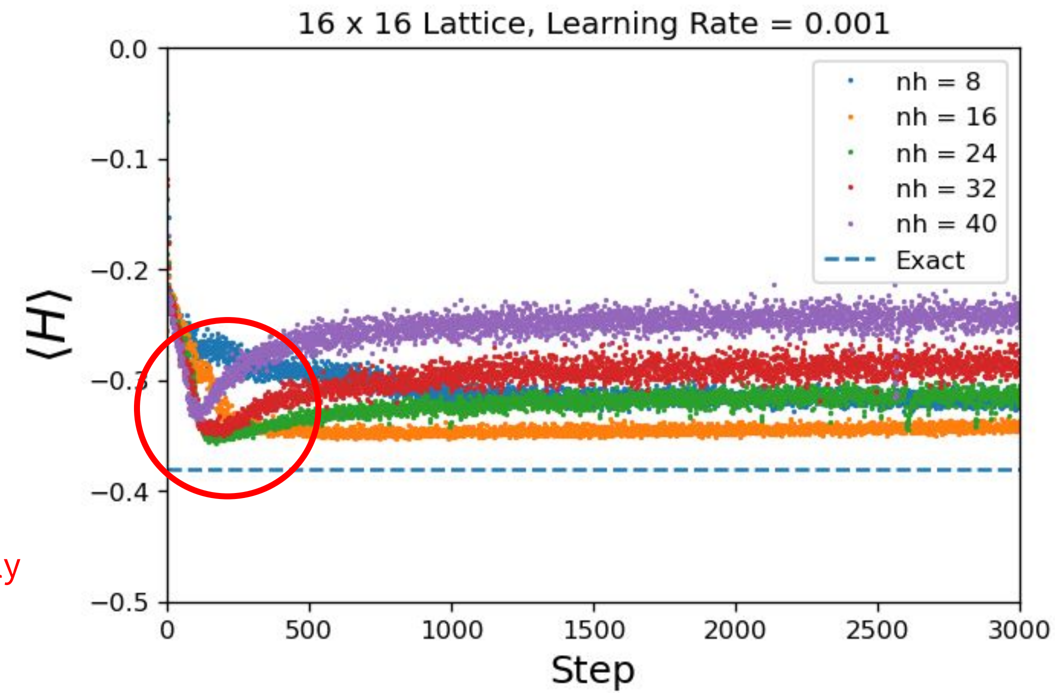


Really bad.

Aside on Overfitting

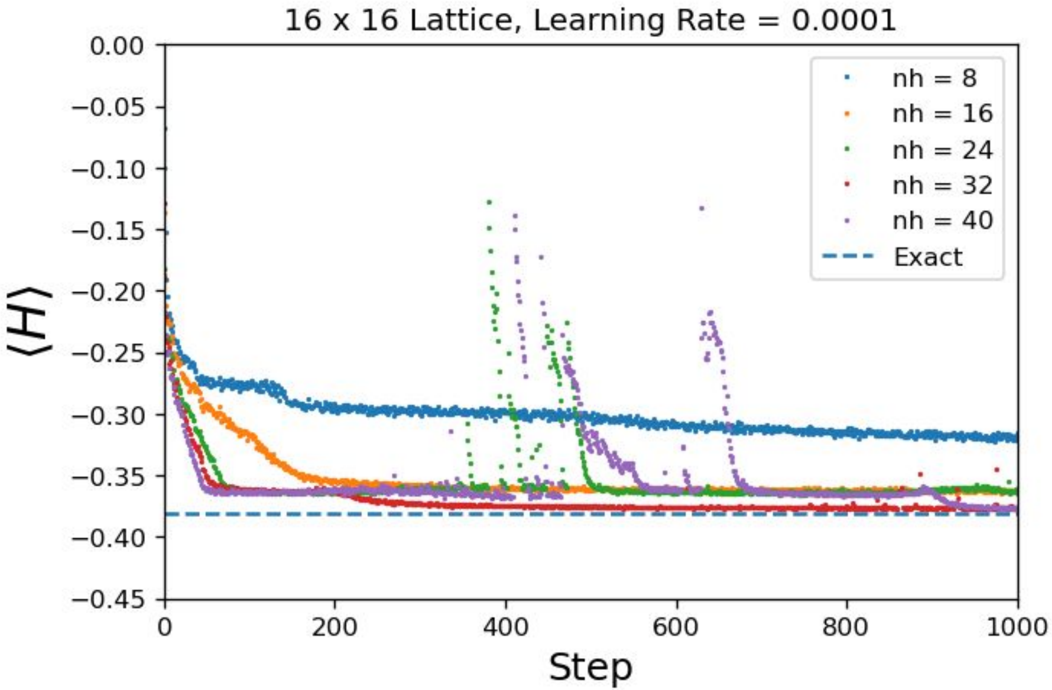


Aside on Overfitting

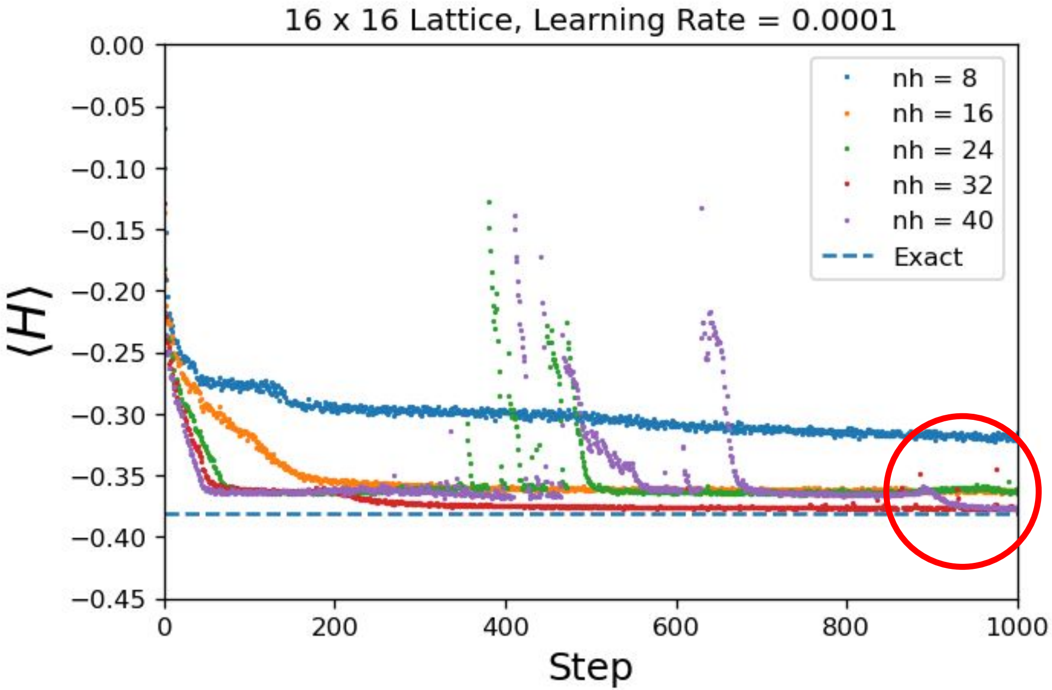


Overfitting occurs consistently

Aside on Expressiveness

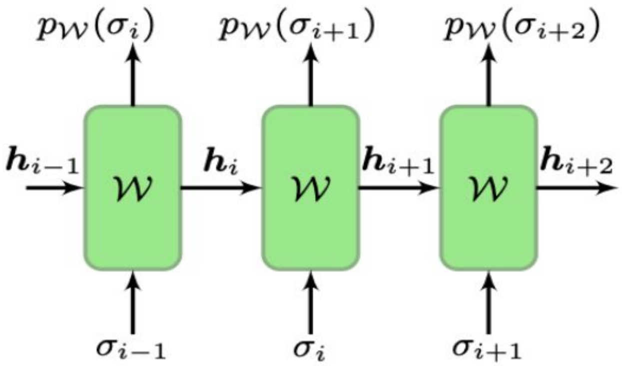


Aside on Expressiveness

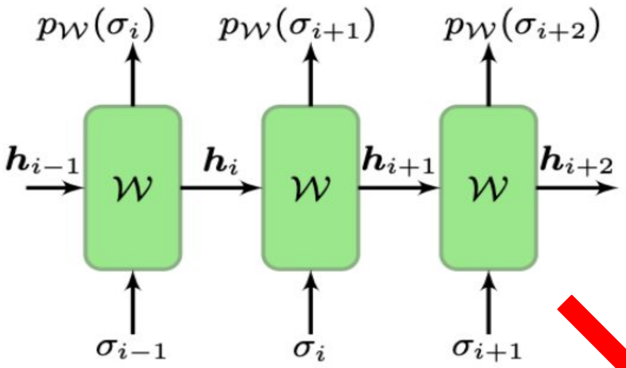


Given the full data set, some system sizes are expressive enough

Proposed Solution



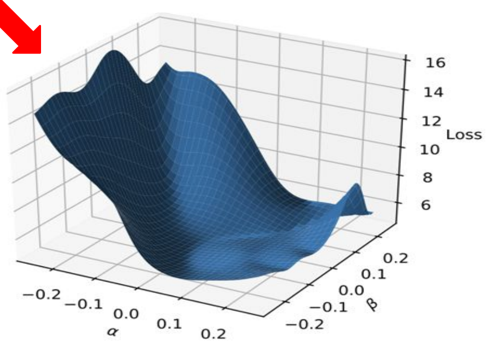
Proposed Solution



train variationally

Hibat-Allah et. al. Recurrent neural network wave functions. *Physical Review Research*, *2*(2), p.023358.

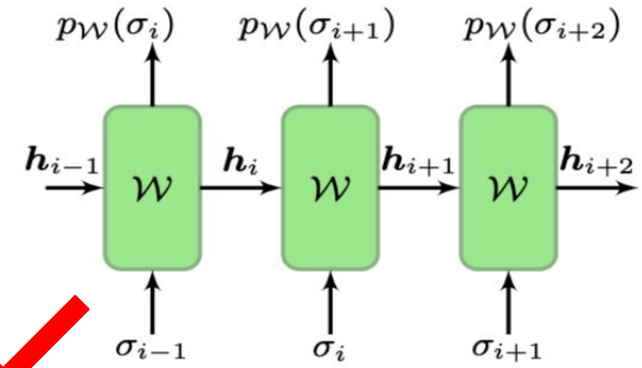
$$\langle H \rangle$$

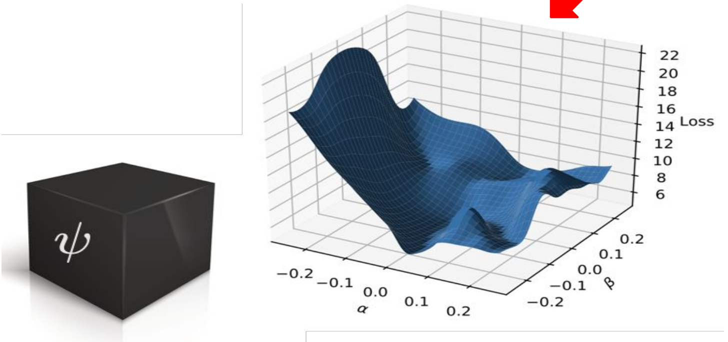


Proposed Solution

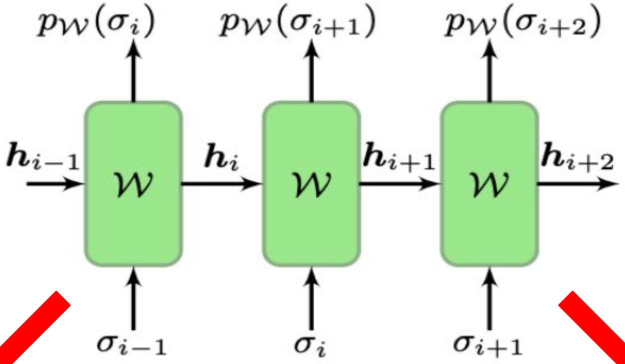
train with data

Torlai *et al.* Neural-network quantum state tomography. *Nature Phys* 14, 447–450 (2018).



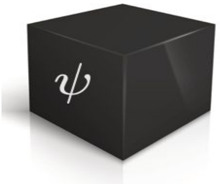


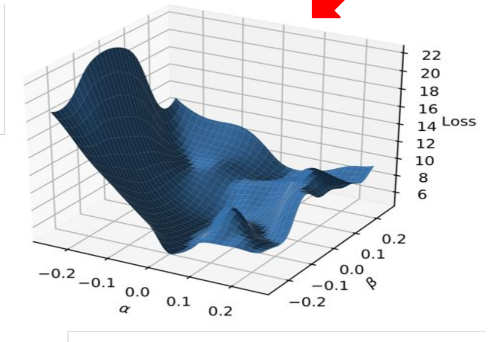
Proposed Solution



train with data

Torlai *et al.* Neural-network quantum state tomography. *Nature Phys* 14, 447–450 (2018).





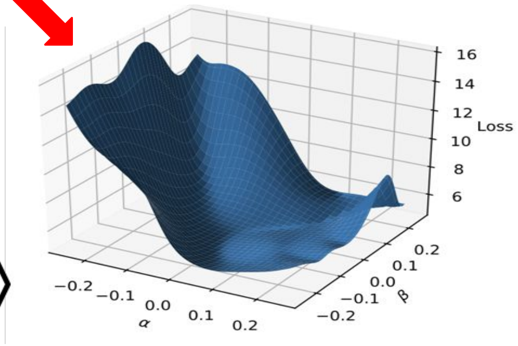
&

??

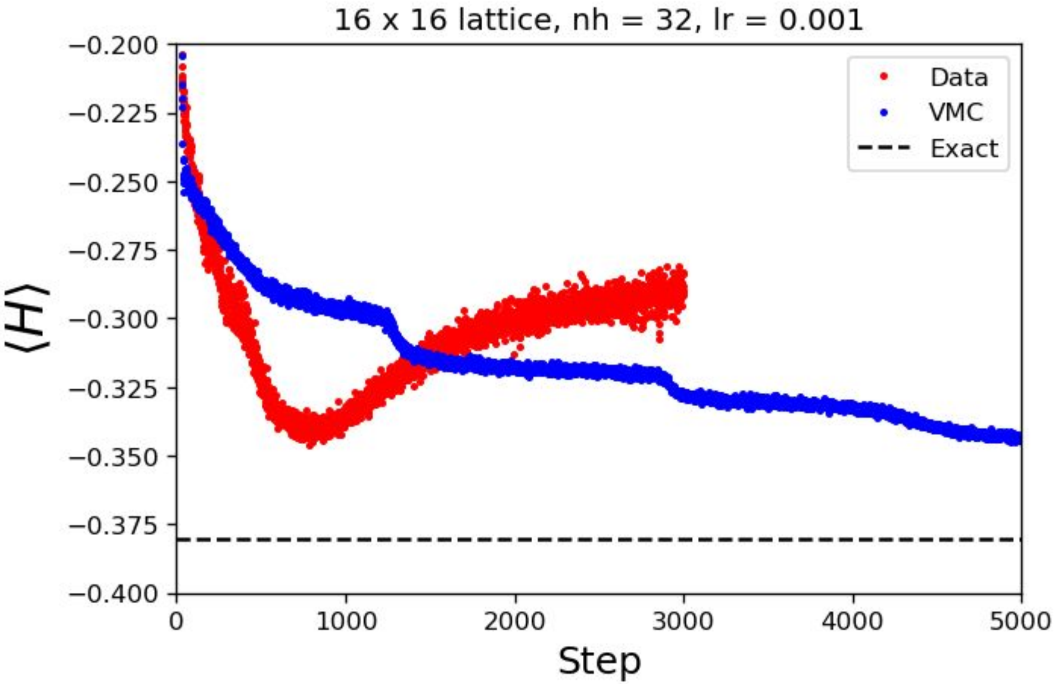
$\langle H \rangle$

train variationally

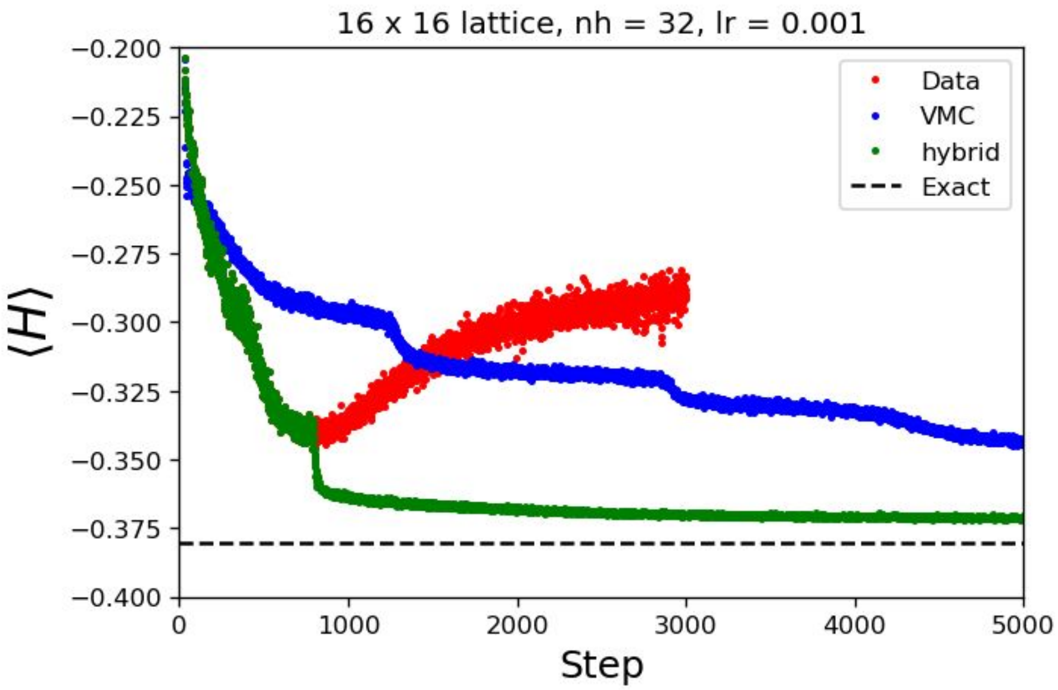
Hibat-Allah et. al. Recurrent neural network wave functions. *Physical Review Research*, *2*(2), p.023358.



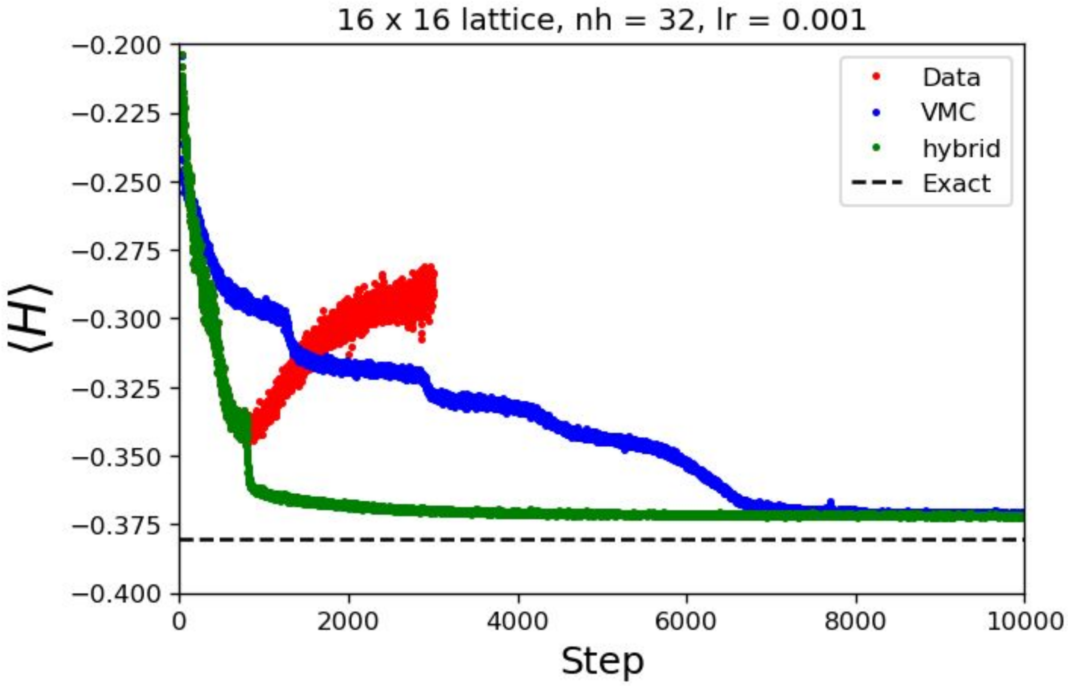
Combining Training Methods



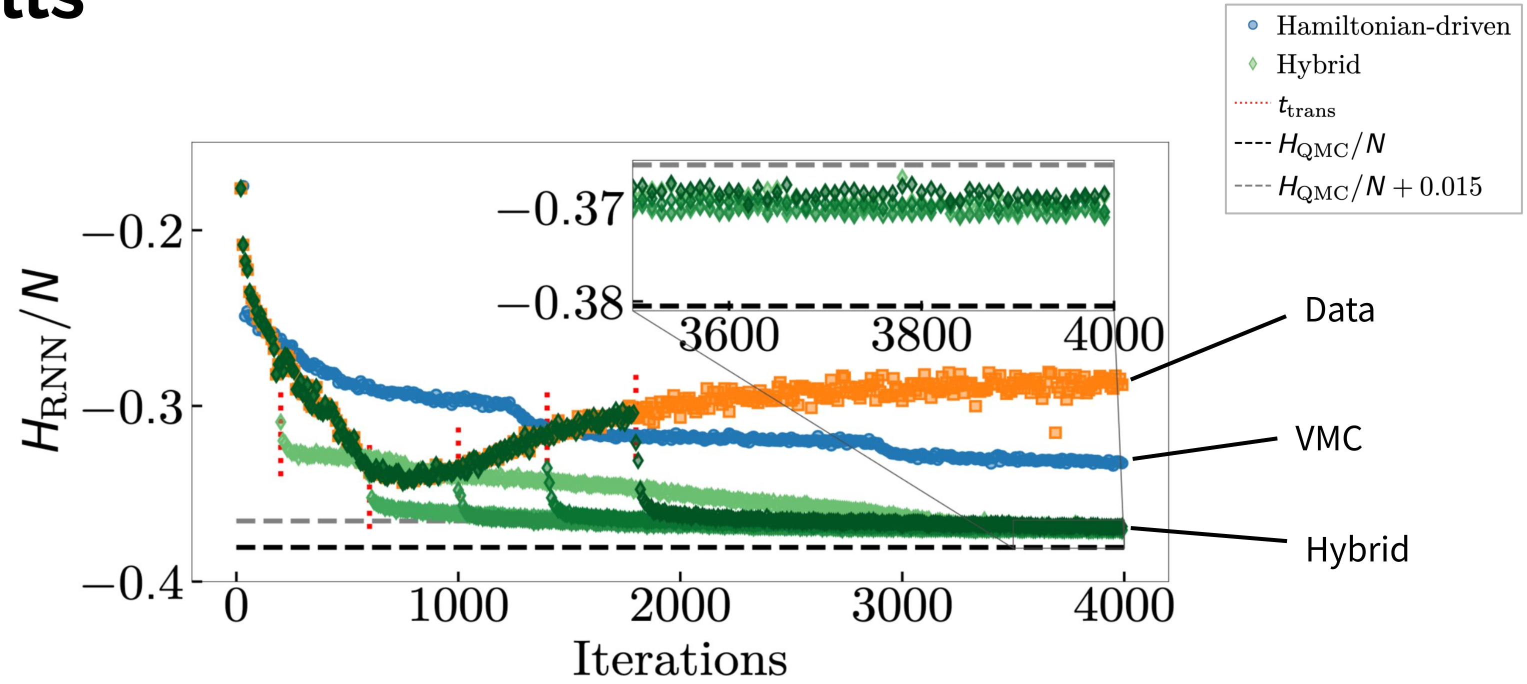
Combining Training Methods



Combining Training Methods

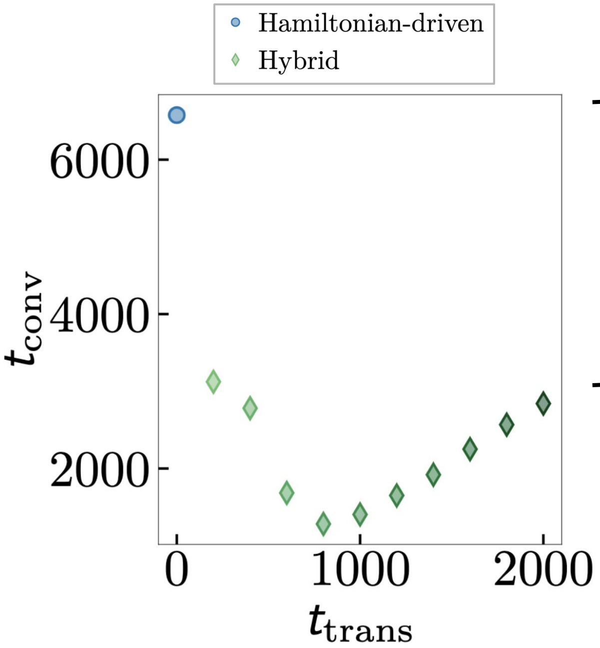


Results



Stefanie Czischek, **M. Schuyler Moss**, et al. Phys. Rev. B **105**, 205108 – Published 9 May 2022

Results



Huge improvement (even for non-optimal transition point)

Stefanie Czischek, **M. Schuyler Moss**, et al. Phys. Rev. B **105**, 205108 – Published 9 May 2022

Conclusions & Open Questions

Conclusions:

1. Can train the same network with data AND variationally with a simple change in the loss function.
2. Results in a speedup in the time to convergence.

Conclusions & Open Questions

Conclusions:

1. Can train the same network with data AND variationally with a simple change in the loss function.
2. Results in a speedup in the time to convergence.

AND can be viewed as error mitigation!

Bennewitz, Elizabeth R., et al. "Neural error mitigation of near-term quantum simulations." *Nature Machine Intelligence* 4.7 (2022): 618-624.

Conclusions & Open Questions

Conclusions:

1. Can train the same network with data AND variationally with a simple change in the loss function.
2. Results in a speedup in the time to convergence.

AND can be viewed as error mitigation!

Bennewitz, Elizabeth R., et al. "Neural error mitigation of near-term quantum simulations." *Nature Machine Intelligence* 4.7 (2022): 618-624.

Open Questions:

1. Is there a "loss schedule" that would result in accuracy improvements in addition to the improved convergence time?

Conclusions & Open Questions

Conclusions:

1. Can train the same network with data AND variationally with a simple change in the loss function.
2. Results in a speedup in the time to convergence.

AND can be viewed as error mitigation!

Bennewitz, Elizabeth R., et al. "Neural error mitigation of near-term quantum simulations." *Nature Machine Intelligence* 4.7 (2022): 618-624.

Open Questions:

1. Is there a "loss schedule" that would result in accuracy improvements in addition to the improved convergence time? (hybrid accuracy better than VMC only)

Conclusions & Open Questions

Conclusions:

1. Can train the same network with data AND variationally with a simple change in the loss function.
2. Results in a speedup in the time to convergence.

AND can be viewed as error mitigation!

Bennewitz, Elizabeth R., et al. "Neural error mitigation of near-term quantum simulations." *Nature Machine Intelligence* 4.7 (2022): 618-624.

Open Questions:

1. Is there a "loss schedule" that would result in accuracy improvements in addition to the improved convergence time? (hybrid accuracy better than VMC only)
2. What does it mean that using both loss functions results in comparable final accuracies?