Learning quantum states from noisy data

Giacomo Torlai
CCQ, Flatiron Institute
gtorlai@uwaterloo.ca

Brian Timar
Caltech
timarbrian@gmail.com

Evert van Nieuwenburg
Caltech
evert.v.nieuwenburg@gmail.com

Manuel Endres
Caltech
mendres@caltech.edu

Roger G. Melko
University of Waterloo
rgmelko@uwaterloo.ca

Abstract

We extract restricted Boltzmann machine (RBM) wavefunctions from data produced by a Rydberg atom quantum simulator, and apply a novel regularization technique to mitigate the effects of measurement errors in the training data. Reconstructions of modest complexity are able to capture observables not accessible to experimentalists.

1 Introduction

Quantum state tomography [1] is an important tool for reconstructing generic quantum states, but traditional techniques require a number of measurements scaling exponentially in the system size [2]. In certain cases, methods that exploit particular entanglement properties [3, 4, 5, 6] allow for more efficient tomography of states prepared in experiment. However, such approaches still involve explicit reconstruction of local density operators [7, 3], incurring a significant computational overhead – especially in the presence of measurement errors. In order to facilitate the characterization of simulators that are currently being realized experimentally [8], a reconstruction method which can efficiently extract physical quantities of interest directly from noisy experimental datasets is highly desirable. As the quantum state may be regarded as a structured distribution over measurement outcomes, unsupervised learning methods are naturally suited to this task. We demonstrate how a simple modification of the restricted Boltzmann machine can be trained in a noise-resilient fashion to produce nontrivial observables, using data from a real quantum simulator.

1.1 Related work

Several efficient tomographic methods have been developed based on targeting states with particular entanglement properties [3, 4, 5, 6] or symmetries [9]. Unsupervised learning methods for reconstruction have recently been developed: restricted Boltzmann Machines were first applied to problems in classical statistical physics [10], and later to the generic many-body state tomography problem [11, 12]. Generalizations of this wavefunction approach to mixed states have been developed [13], as well as alternative methods based on generative modeling of POVM distributions [14].

Experimentally, reconstructions of seven- and fourteen-qubit systems [5, 6] have been demonstrated, using compressed sensing and matrix product state-based methods. For low-dimensional systems, supervised methods have been applied to denoising in a tomographic context [15].

The problem of denoising has been studied extensively in the machine learning [16, 17] and biology [18, 19] communities. Modifications to the RBM similar to the one we use here have been applied to occluded image restoration [20].

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

1.2 Contributions

We demonstrate unsupervised learning of states produced by a strongly interacting quantum simulator. We implement an efficient procedure for denoising the full probability distribution from bit-flip-type measurement errors, by incorporating a dedicated "noise layer" in the network architecture. We test the validity of our approach by comparing predictions of the trained RBMs with numerical results for observables that are inaccessible in experiment.

2 Methods

2.1 Experimental system

Our experiment [21, 22] consists of a Rydberg atom quantum simulator, a neutral-atom system for realizing Ising-type quantum spin models. In the experiments utilized in this study, a one-dimensional array of $N = 8$ trapped Rubidium atoms is prepared. Each atom can occupy a ground state $|g\rangle$ or an excited (Rydberg) state $|r\rangle$. When subjected to a uniform laser drive, the effective Hamiltonian of the many-body system may be written as [21]

$$\hat{H}(\Omega, \Delta) = -\Delta \sum_{i=1}^{N} \hat{n}_i - \frac{\Omega}{2} \sum_{i=1}^{N} \hat{\sigma}_i^x + \sum_{i<j} \frac{V_{nn}}{|i-j|^6} \hat{n}_i \hat{n}_j, \tag{1}$$

where V_{nn} is the interaction strength between Rydberg atoms at adjacent sites, $\hat{\sigma}_i^\alpha$, with $\alpha = x, y, z$, are the Pauli pseudo-spin operators at site i (defined as $\hat{\sigma}_i^z = |r_i\rangle\langle r_i| - |g_i\rangle\langle g_i|$, $\hat{\sigma}_i^x = |r_i\rangle\langle g_i| + h.c.$, etc), and $\hat{n}_i = \frac{1}{2}(1 + \hat{\sigma}_i^z)$ projects onto the Rydberg state at site i. The parameters Ω, Δ denote the effective Rabi frequency and detuning, respectively, which characterize the laser drive, and can be varied in time as $\Omega(t), \Delta(t)$. The atoms are initially pumped into the state $|g\,g\,g\,g\,g\ldots\rangle$, coinciding with the ground state of Hamiltonian (1) at $t = 0$, and then evolve adiabatically under a "sweep" of the laser parameters $\Omega(t), \Delta(t)$ for a time T_{ev}. Our goal is to demonstrate efficient, accurate reconstruction of all intermediate states in the sweep.

Measurement data consists of a collection of N-bit strings $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_N)$, with $\tau_j = 0, 1$ indicating that atom j was recorded as being in the ground $|g\rangle$ or Rydberg state $|r\rangle$ respectively. There are small measurement error probabilities $p(1|0) \sim 1\%$, $p(0|1) \sim 4\%$ [23] for an atom in the ground state to be recorded as excited and vice-versa, resulting in imperfect experimental data $\boldsymbol{\tau}$. The distribution sampled in experiment is therefore related to the underlying state as $P_{\text{exp}}(\boldsymbol{\tau}) = \sum_{\boldsymbol{\sigma}} p(\boldsymbol{\tau}|\boldsymbol{\sigma})\langle\boldsymbol{\sigma}|\hat{\rho}|\boldsymbol{\sigma}\rangle$, where $p(\boldsymbol{\tau}|\boldsymbol{\sigma}) = \prod_{j=1}^{N} p(\tau_j|\sigma_j)$ is the probability of recording bitstring $\boldsymbol{\tau}$ when the atoms are prepared in the state $|\boldsymbol{\sigma}\rangle$.

2.2 Wavefunction model

As Hamiltonian (1) is stoquastic [25], its ground state has real, positive amplitudes in the occupation number basis $|\boldsymbol{\sigma}\rangle = |\sigma_1, \ldots, \sigma_N\rangle$, defined as the simultaneous eigenstates of $\hat{n}_1, \ldots, \hat{n}_N$. Therefore, if the state of the simulator evolves adiabatically and with negligible loss of purity, it is uniquely characterized by the probability distribution $p(\boldsymbol{\sigma})$ it assigns to projective measurements in the $|\boldsymbol{\sigma}\rangle$ basis, and may be written as the pure state $|\psi\rangle = \sum_{\boldsymbol{\sigma}} \sqrt{p(\boldsymbol{\sigma})}|\boldsymbol{\sigma}\rangle$.

We parametrize $p(\boldsymbol{\sigma})$ with an RBM [26, 27], a generative model composed of a stochastic neural network with two layers of binary units, a visible layer $\boldsymbol{\sigma}$ and a hidden layer $\boldsymbol{h}$. The marginal distribution over visible variables is

$$p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}) = \frac{1}{Z_{\boldsymbol{\lambda}}} \sum_{\boldsymbol{h}} e^{\,\boldsymbol{h}^\top \boldsymbol{W}\boldsymbol{\sigma} + \boldsymbol{b}\cdot\boldsymbol{\sigma} + \boldsymbol{c}\cdot\boldsymbol{h}} \tag{2}$$

where $Z_{\boldsymbol{\lambda}}$ is a normalization constant, and the real-valued network parameters are $\boldsymbol{\lambda} = \{\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{c}\}$, with $\boldsymbol{W}$ being the weights connecting the two layers and $\boldsymbol{b}$ ($\boldsymbol{c}$) the visible (hidden) bias vectors. We use the visible layer of the RBM to define the projective measurement distribution $p(\boldsymbol{\sigma})$ of the pure state: $\psi_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}) = \langle\boldsymbol{\sigma}|\psi_{\boldsymbol{\lambda}}\rangle = \sqrt{p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})}$.

Three-layer extension The experimental measurement process is incorporated into our model via a third binary layer, the so-called *noise layer*, which represents the observed POVM outcomes

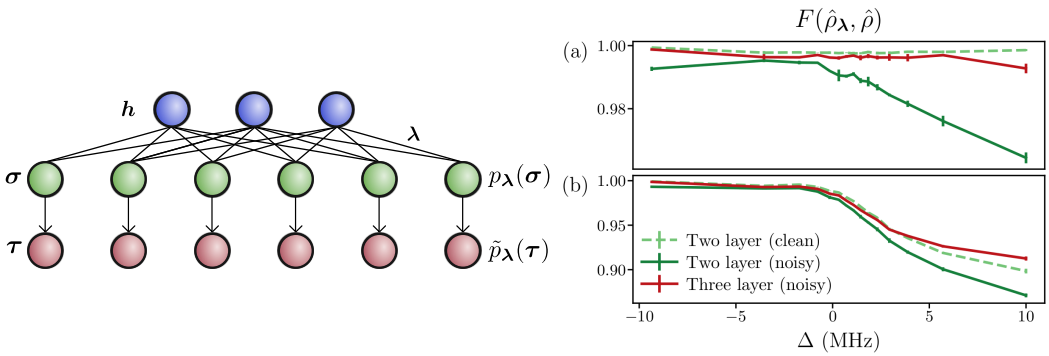

Figure 1: Model summary. Left: graphical model used for training on noise-corrupted data. For given measurement error rates, the pre-noise RBM distribution $p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})$ induces a distribution $\tilde{p}_{\boldsymbol{\lambda}}(\boldsymbol{\tau})$ over observations $\boldsymbol{\tau}$, which can be used to fit the RBM parameters via standard log-likelihood methods. Right: a summary of training quality on simulated data, both with (b) and without (a) a realistic decoherence model. The quantum state fidelity F between the reconstruction $\hat{\rho}_{\boldsymbol{\lambda}}$ and the ground-truth state $\hat{\rho}$ – a measure $0 \leq F \leq 1$ of reconstruction quality [24] – is plotted as a function of detuning for a simulation of the experimental sweep. The solid red and green lines are the fidelity after training on noise-corrupted data with and without noise-layer regularization, respectively; as a benchmark, the dashed green line shows the fidelity of a two-layer RBM trained on noise-free data.

$\boldsymbol{\tau}$. The measurement error rates $p(\boldsymbol{\tau}|\boldsymbol{\sigma})$ are included as connections between the visible and noise layers (Fig. 1), by assigning a probability $\tilde{p}_{\boldsymbol{\lambda}}(\boldsymbol{\tau}) = \sum_{\boldsymbol{\sigma}} p(\boldsymbol{\tau}|\boldsymbol{\sigma}) p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})$ to the measurement result $\boldsymbol{\tau}$. For a dataset D of observations $\boldsymbol{\tau}$, the RBM is trained to minimize the negative log-likelihood $\mathcal{L}_{\boldsymbol{\lambda}} = -\frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{\tau} \in \mathcal{D}} \log \tilde{p}_{\boldsymbol{\lambda}}(\boldsymbol{\tau})$. The corresponding gradient is

$$\nabla_{\boldsymbol{\lambda}} \mathcal{L}_{\boldsymbol{\lambda}} = \mathbb{E}_{\boldsymbol{\sigma} \sim p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})} \left[\nabla_{\boldsymbol{\lambda}} \mathcal{E}_{\text{eff}}(\boldsymbol{\sigma}) \right] - \frac{1}{|\mathcal{D}|} \sum_{\boldsymbol{\tau} \in D} \mathbb{E}_{\boldsymbol{\sigma} \sim \tilde{p}_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}|\boldsymbol{\tau})} \left[\nabla_{\boldsymbol{\lambda}} \mathcal{E}_{\text{eff}}(\boldsymbol{\sigma}) \right] \tag{3}$$

where $\mathcal{E}_{\text{eff}}(\boldsymbol{\sigma}) = \boldsymbol{b} \cdot \boldsymbol{\sigma} + \sum_j \log\left(1 + e^{\sum_i W_{ji}\sigma_i + c_j}\right)$, and $\tilde{p}_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}|\boldsymbol{\tau}) = \frac{p(\boldsymbol{\tau}|\boldsymbol{\sigma}) p_{\boldsymbol{\lambda}}(\boldsymbol{\sigma})}{\tilde{p}_{\boldsymbol{\lambda}}(\boldsymbol{\tau})}$. This posterior probability can be computed efficiently using Gibbs sampling, since the conditional

$$\tilde{p}_{\boldsymbol{\lambda}}(\boldsymbol{\sigma}|\boldsymbol{\tau}, \boldsymbol{h}) = \prod_i \frac{p(\boldsymbol{\tau}|\sigma_i) p_{\boldsymbol{\lambda}}(\sigma_i|\boldsymbol{h})}{\sum_{\sigma_i'=0,1} p(\boldsymbol{\tau}|\sigma_i') p_{\boldsymbol{\lambda}}(\sigma_i'|\boldsymbol{h})} = \prod_i \tilde{p}_{\boldsymbol{\lambda}}(\sigma_i|\boldsymbol{\tau}, \boldsymbol{h}) \tag{4}$$

is tractable.

During training, the noise layer acts as a buffer between the noisy data and the RBM, preventing the parameters from $\boldsymbol{\lambda}$ from fitting to spurious features in the data produced by measurement errors. This *noise layer regularization* significantly improves the fidelity between $|\psi_{\boldsymbol{\lambda}}\rangle$ and the state $\hat{\rho}$ underlying the data; tests (Fig. 1a,b) based on numerical simulation of our experiment result in fidelities greater than 90% for the full many-body state at the end of the sweep.

2.3 Dataset construction and training

At fifteen subsequent time-steps t, the sweep was halted and a dataset of around 3,000 measurements $\boldsymbol{\tau}$ was sampled from the state $\hat{\rho}(t)$. These were used to train three-layer models with $2N = 16$ hidden units; a scaling analysis in the hidden layer size was performed to ensure that the networks had sufficient capacity. Training was performed using stochastic gradient descent with a decayed learning rate, the gradients being estimated via contrastive divergence [28] with $k = 30$ sampling steps. We found it beneficial to set the model's error rates to zero for the first epoch of training.

After training the networks, standard sampling methods can be applied to compute expectation values of observables, with a computational cost scaling polynomially in the network size; in this work the Hilbert space dimension was small enough that observables could be computed exactly.

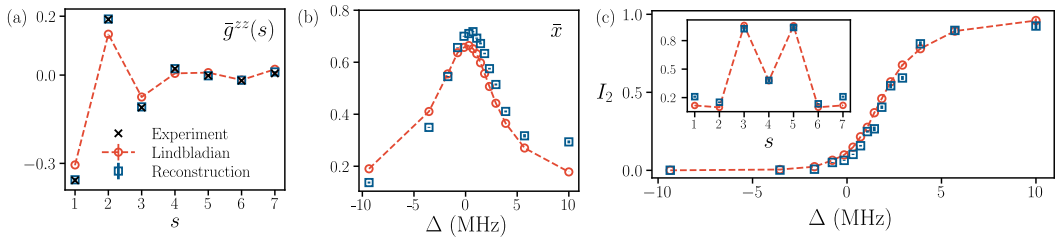

Figure 2: Reconstructed observables. Comparison of the RBM reconstruction (squares) with the experiment results (crosses) and the predictions from our Lindbladian master equation (circles).The values reported in (a) for the RBM and Lindbladian observables are computed including the known measurement error rates $p(0|1) = 0.04$, $p(1|0) = 0.01$. (a) Nearest-neighbor correlations $\bar{g}^{zz}(s)$ in the z basis, defined as $\bar{g}^{zz}(s) = \frac{1}{N-s}\sum_{i=1}^{N-s}\langle\hat{\sigma}_i^z\hat{\sigma}_{i+s}^z\rangle_c$, at $\Delta = 10$MHz. Here $\langle\hat{\sigma}_i^z\hat{\sigma}_j^z\rangle_c = \langle\hat{\sigma}_i^z\hat{\sigma}_j^z\rangle - \langle\hat{\sigma}_i^z\rangle\langle\hat{\sigma}_j^z\rangle$ denotes the two-point correlation function. (b) Spatial average $\bar{x}$ of the transverse field $\langle\hat{\sigma}_i^x\rangle$. (c) Rényi mutual information. The quantum (Rényi) mutual information I_2 is defined as $I_2(s) = S_2(\hat{\rho}_s^A) + S_2(\hat{\rho}_s^B) - S_2(\hat{\rho})$, where $S_2(\hat{\rho}) = -\log \mathrm{Tr}\hat{\rho}^2$ is the second-order Rényi entropy, $\hat{\rho}$ is the (mixed) state of the whole system, and $\hat{\rho}_s^A, \hat{\rho}_s^B$ are the reduced density matrices for the subsystems $A_s = \{1, ..., s\}$, $B_s = \{s+1, ...N\}$ defined by a partitioning of the system at bond $(s, s+1)$. The mutual information is plotted for a partition at bond (3,4), as a function of detuning. Inset: The mutual information $I_2(s)$ as a function of the cut bond s for $\Delta = 10$ MHz.

3 Results

Fig. 2 presents some observables reconstructed from experimental data. The networks learn the strong two-body correlations $\langle\hat{\sigma}_i^z\hat{\sigma}_j^z\rangle_c$ present in the experimental data, as seen in Fig 2a; such statistics, which are diagonal in the measurement basis, provide a simple check on the quality of RBM training. We also compare the results of the reconstruction process to the exact solutions of a Lindblad master equation for the full many-body evolution – our simulation predicts significantly weaker correlations, suggesting our model for the sweep dynamics is partially incomplete. In experimentally inaccessible quantities, such as the average transverse field (Fig. 2b), the reconstructions are in excellent agreement with simulation, albeit predicting somewhat larger values in the ordered phase.

From RBM wavefunctions, the quantum Rényi entropy – which requires specialized or hardware-specific protocols to access directly in experiment [29, 30] – may also be extracted in a scalable fashion by applying a state-replication and swap procedure virtually [31, 11]. In fact, for pure experimental states, positive-pure ansatzes such as the RBM wavefunction provide a lower bound on the mutual information defined by the Rényi entropy ([32], [33]), regardless of the sign structure of the true state. We demonstrate a reconstruction of the mutual information defined by the second-order Rényi entropy in Fig. 2c, finding that the RBM values are in remarkable agreement with the results of numerical simulation.

4 Conclusions

We have demonstrated unsupervised learning of quantum states from data produced by a Rydberg-atom quantum simulator. RBMs were trained in a noise-regularized fashion in a single measurement basis, then queried to produce a variety of observables not directly accessible in experiment. This approach can be integrated without alteration into experimental platforms where a positive wavefunction ansatz is a valid approximation, such as Bose-Hubbard experiments and some non-frustrated quantum spin simulators [34, 35, 36, 37]. If measurements in multiple bases are available, the RBM reconstruction protocol can also be easily adapted to reconstruct non-positive and complex wavefunctions [11], although noise-layer regularization cannot be directly applied.

Developing more powerful denoising methods for quantum systems would be quite valuable to experimentalists. It is interesting to note that the noise regularization used in this work does not require passing gradients through binary units, instead obtaining all learning signals through suitably conditioned Gibbs sampling. It may be worth developing more sophisticated models which can support a binary prior relevant to quantum experiments.

References

[1] K. Banaszek, M. Cramer, and D. Gross, New Journal of Physics **15**, 125020 (2013).

[2] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür, and R. Blatt, Nature **438**, 643 (2005).

[3] M. Cramer, M. B. Plenio, S. T. Flammia, R. Somma, D. Gross, S. D. Bartlett, O. Landon-Cardinal, D. Poulin, and Y.-K. Liu, Nature Communications **1**, 149 (2010).

[4] J. Y. Lee and O. Landon-Cardinal, Phys. Rev. A **91**, 062128 (2015).

[5] C. A. Riofrío, D. Gross, S. T. Flammia, T. Monz, D. Nigg, R. Blatt, and J. Eisert, Nature Communications **8**, 15305 (2017).

[6] B. P. Lanyon, C. Maier, M. Holzäpfel, T. Baumgratz, C. Hempel, P. Jurcevic, I. Dhand, A. S. Buyskikh, A. J. Daley, M. Cramer, M. B. Plenio, R. Blatt, and C. F. Roos, Nature Physics **13**, 1158 (2017).

[7] M. Ježek, J. Fiurášek, and Z. Hradil, Physical Review A **68**, 012305 (2003).

[8] J. Preskill, Quantum **2**, 79 (2018).

[9] G. Tóth, W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter, Phys. Rev. Lett. **105**, 250403 (2010).

[10] G. Torlai and R. G. Melko, Physical Review B **94**, 165134 (2016).

[11] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nature Physics **14**, 447 (2018).

[12] G. Torlai and R. G. Melko, arXiv e-prints , arXiv:1905.04312 (2019), arXiv:1905.04312 [quant-ph] .

[13] G. Torlai and R. G. Melko, Physical Review Letters **120**, 240503 (2018).

[14] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Nature Machine Intelligence **1**, 155 (2019).

[15] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. Biamonte, and S. Kulik, arXiv:1904.05902 [cond-mat, physics:quant-ph] (2019), arXiv: 1904.05902.

[16] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, in *Proceedings of the 25th international conference on Machine learning - ICML '08* (ACM Press, New York, New York, USA, 2008) pp. 1096–1103.

[17] D. Ulyanov, A. Vedaldi, and V. Lempitsky, (2017), arXiv:1711.10925 .

[18] J. Batson and L. Royer, (2019), arXiv:1901.11365 .

[19] A. Krull, T.-O. Buchholz, and F. Jug, (2018), arXiv:1811.10980 .

[20] Yichuan Tang, R. Salakhutdinov, and G. Hinton, in *2012 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2012) pp. 2264–2271.

[21] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Nature **551**, 579 (2017).

[22] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletic, M. Greiner, and M. D. Lukin, Science **354**, 1024 (2016).

[23] H. Levine, A. Keesling, A. Omran, H. Bernien, S. Schwartz, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Physical Review Letters **121**, 123603 (2018).

[24] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge, 2010).

[25] S. Bravyi, D. P. Divincenzo, R. Oliveira, and B. M. Terhal, Quantum Info. Comput. **8**, 361 (2008).

[26] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cognitive Science **9**, 147 (1985).

[27] P. Smolensky, in *Parallel Distributed Processing*, edited by D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group (MIT Press, Cambridge, MA, USA, 1986) Chap. Information Processing in Dynamical Systems: Foundations of Harmony Theory, pp. 194–281.

[28] G. E. Hinton, Neural Comput. **14**, 1771 (2002).

[29] R. Islam, R. Ma, P. M. Preiss, M. Eric Tai, A. Lukin, M. Rispoli, and M. Greiner, Nature **528**, 77 (2015).

[30] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, (2018), arXiv:1806.05747 .

[31] M. B. Hastings, I. González, A. B. Kallin, and R. G. Melko, Physical Review Letters **104**, 157201 (2010).

[32] Y. Zhang, T. Grover, and A. Vishwanath, Physical Review Letters **107**, 067202 (2011).

[33] T. Grover and M. P. A. Fisher, Physical Review A **92**, 042308 (2015).

[34] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, Nature **462**, 74 (2009).

[35] C. Weitenberg, M. Endres, J. F. Sherson, M. Cheneau, P. Schauß, T. Fukuhara, I. Bloch, and S. Kuhr, Nature **471**, 319 (2011).

[36] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Science **353**, 794 (2016).

[37] H. Labuhn, D. Barredo, S. Ravets, S. de Léséleuc, T. Macrì, T. Lahaye, and A. Browaeys, Nature **534**, 667 (2016).