

Title: Machine Learning Lecture - 230330

Speakers: Lauren Hayward

Collection: Machine Learning for Many-Body Physics (2022/2023)

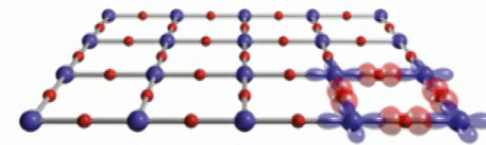
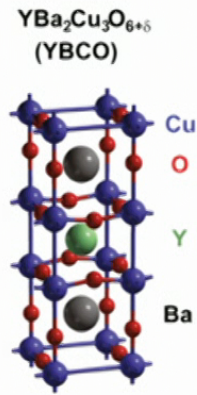
Date: March 30, 2023 - 9:00 AM

URL: <https://pirsa.org/23030037>

Outline

- ML and (condensed matter) physics
- Research by Perimeter students
 - Supervised & unsupervised learning in physics
 - The role of industry connections in research
- The frontier of quantum + machine learning

The Explainability crisis in Condensed matter physics



$$H = -t \sum_{\langle ij \rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + c_{i\sigma} c_{j\sigma}^\dagger) + U \sum_i n_{i\uparrow} n_{i\downarrow}$$

Condensed matter physics and (machine) learning



Philip W. Anderson

Nobel Prize 1977
(Semiconductors, superconductivity, magnetism)

Harvard, Bell Labs, Cambridge, Princeton

The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe

The constructionist hypothesis breaks down when confronted with the twin difficulties of scale and complexity.

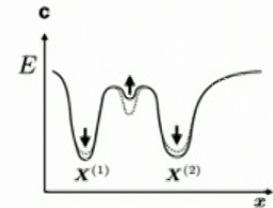
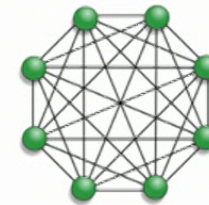
Condensed matter physics and (machine) learning



John Hopfield

Buckley Prize, 1968

"For their joint work combining theory and experiment which has advanced the understanding of the interaction of light with solids"



G. Torlai, PhD Thesis
<https://uwspace.uwaterloo.ca/handle/10012/14196>

"Over time, his research meandered from hard physics to neuroscience, where he applied his knowhow from the former to construct an artificial neural network capable of modeling certain functions of the human brain.

Decades later, these fundamental concepts have helped to unleash the tide of "deep learning" technologies that allow machines to observe, remember, and learn on their own."

"The Hopfield neural net was not only an intriguing demonstration of possibility and a proof by demonstration of the information-handling capabilities of neural nets, but also was an enormous stimulant in reviving the perceptron and other machine-learning programs."

Outline

- ML and (condensed matter) physics
- Research by Perimeter students
 - Supervised & unsupervised learning in physics
 - The role of industry connections in research
- The frontier of quantum + machine learning

Quantum Physics

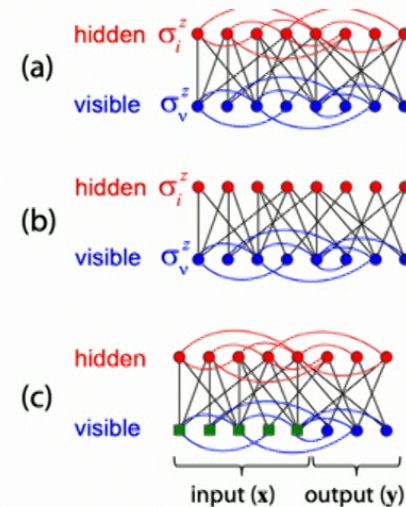
[Submitted on 8 Jan 2016]

Quantum Boltzmann Machine

Mohammad H. Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchitsky, Roger Melko

Inspired by the success of Boltzmann Machines based on classical Boltzmann distribution, we propose a new machine learning approach based on quantum Boltzmann distribution of a transverse-field Ising Hamiltonian. Due to the non-commutative nature of quantum mechanics, the training process of the Quantum Boltzmann Machine (QBM) can become nontrivial. We circumvent the problem by introducing bounds on the quantum probabilities. This allows us to train the QBM efficiently by sampling. We show examples of QBM training with and without the bound, using exact diagonalization, and compare the results with classical Boltzmann training. We also discuss the possibility of using quantum annealing processors like D-Wave for QBM training and application.

$$H = - \sum_a \Gamma_a \sigma_a^x - \sum_a b_a \sigma_a^z - \sum_{a,b} w_{ab} \sigma_a^z \sigma_b^z$$



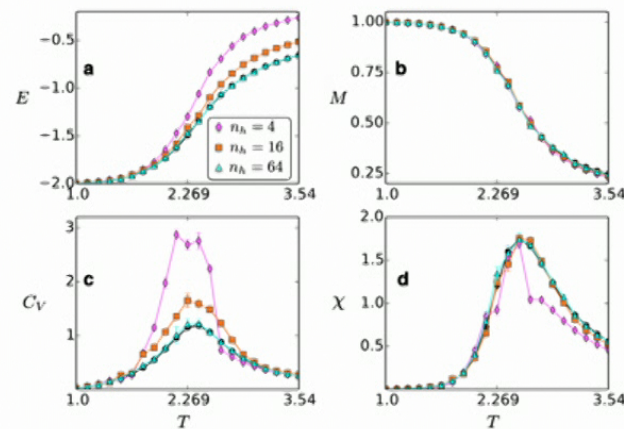
[Submitted on 8 Jun 2016]

Learning Thermodynamics with Boltzmann Machines

Giacomo Torlai, Roger G. Melko

A Boltzmann machine is a stochastic neural network that has been extensively used in the layers of deep architectures for modern machine learning applications. In this paper, we develop a Boltzmann machine that is capable of modelling thermodynamic observables for physical systems in thermal equilibrium. Through unsupervised learning, we train the Boltzmann machine on data sets constructed with spin configurations importance-sampled from the partition function of an Ising Hamiltonian at different temperatures using Monte Carlo (MC) methods. The trained Boltzmann machine is then used to generate spin states, for which we compare thermodynamic observables to those computed by direct MC sampling. We demonstrate that the Boltzmann machine can faithfully reproduce the observables of the physical system. Further, we observe that the number of neurons required to obtain accurate results increases as the system is brought close to criticality.

Comments: 8 pages, 5 figures

Subjects: **Statistical Mechanics (cond-mat.stat-mech)**; Disordered Systems and Neural Networks (cond-mat.dis-nn); Machine Learning (cs.LG)

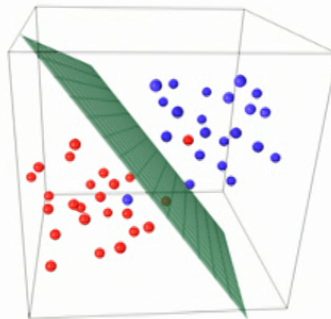
Condensed Matter > Statistical Mechanics

[Submitted on 19 Apr 2017]

Kernel methods for interpretable machine learning of order parameters

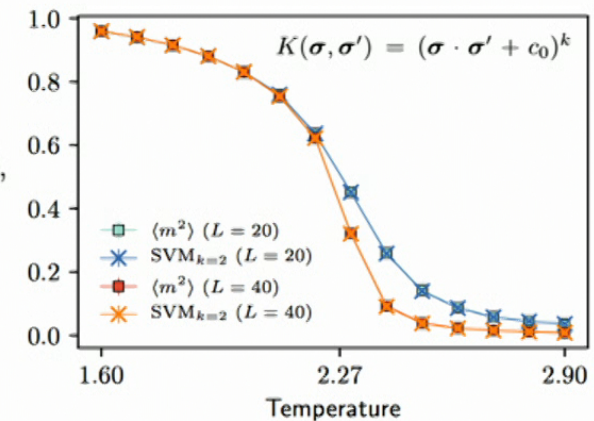
Pedro Ponte, Roger G. Melko

Machine learning is capable of discriminating phases of matter, and finding associated phase transitions, directly from large data sets of raw state configurations. In the context of condensed matter physics, most progress in the field of supervised learning has come from employing neural networks as classifiers. Although very powerful, such algorithms suffer from a lack of interpretability, which is usually desired in scientific applications in order to associate learned features with physical phenomena. In this paper, we explore support vector machines (SVMs) which are a class of supervised kernel methods that provide interpretable decision functions. We find that SVMs can learn the mathematical form of physical discriminators, such as order parameters and Hamiltonian constraints, for a set of two-dimensional spin models: the ferromagnetic Ising model, a conserved-order-parameter Ising model, and the Ising gauge theory. The ability of SVMs to provide interpretable classification highlights their potential for automating feature detection in both synthetic and experimental data sets for condensed matter and other many-body systems.



- Optimal hyperplane for linearly separable patterns
- Extend to patterns that are not linearly separable by transformations of original data to map into new space – the **Kernel** function

$$d(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i K(\mathbf{x}^{(i)}, \mathbf{x}) + b,$$



Condensed Matter > Strongly Correlated Electrons

[Submitted on 1 Feb 2021 (v1), last revised 12 Sep 2021 (this version, v2)]

Machine-Learned Phase Diagrams of Generalized Kitaev Honeycomb Magnets

Nihal Rao, Ke Liu, Marc Machaczek, Lode Pollet

We use a recently developed interpretable and unsupervised machine-learning method, the tensorial kernel support vector machine (TK-SVM), to investigate the low-temperature classical phase diagram of a generalized Heisenberg-Kitaev- Γ (J - K - Γ) model on a honeycomb lattice. Aside from reproducing phases reported by previous quantum and classical studies, our machine finds a hitherto missed nested zigzag-stripy order and establishes the robustness of a recently identified modulated $S_3 \times Z_3$ phase, which emerges through the competition between the Kitaev and Γ spin liquids, against Heisenberg interactions. The results imply that, in the restricted parameter space spanned by the three primary exchange interactions -- J , K , and Γ , the representative Kitaev material α - RuCl_3 lies close to the boundaries of several phases, including a simple ferromagnet, the unconventional $S_3 \times Z_3$ and nested zigzag-stripy magnets. A zigzag order is stabilized by a finite Γ' and/or J_3 term, whereas the four magnetic orders may compete in particular if Γ' is anti-ferromagnetic.

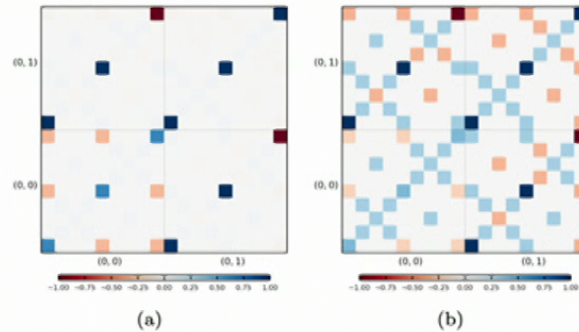


FIG. 14. Representative blocks of the $C_{\mu\nu}$ matrices of the ZZ/D_{2h} phase learned by a rank-2 TK-SVM with the eight-spin D_{2h} magnetic cell, away from (a) and at (b) the $O(3)$ point. Blocks are labeled by the spin indices (i, j) . Nonvanishing entries in a block correspond to correlations between quadratic components $S_i^\alpha S_j^\beta$ and $S_i^{\alpha'} S_j^{\beta'}$. Negative elements in the $(0, 0)$ block reflect the spin normalization $|\vec{S}| = 1$. Non-trivial entries in (a) are the diagonal ones in each 9×9 sub-block.

[Submitted on 1 Jun 2016 (v1), last revised 6 Jun 2016 (this version, v2)]

Discovering Phase Transitions with Unsupervised Learning

Lei Wang

Unsupervised learning is a discipline of machine learning which aims at discovering patterns in big data sets or classifying the data into several categories without being trained explicitly. We show that unsupervised learning techniques can be readily used to identify phases and phases transitions of many body systems. Starting with raw spin configurations of a prototypical Ising model, we use principal component analysis to extract relevant low dimensional representations the original data and use clustering analysis to identify distinct phases in the feature space. This approach successfully finds out physical concepts such as order parameter and structure factor to be indicators of the phase transition. We discuss future prospects of discovering more complex phases and phase transitions using unsupervised learning techniques.

Comments: corrected typos, fixed links in references

Subjects: **Statistical Mechanics (cond-mat.stat-mech)**; Machine Learning (stat.ML)

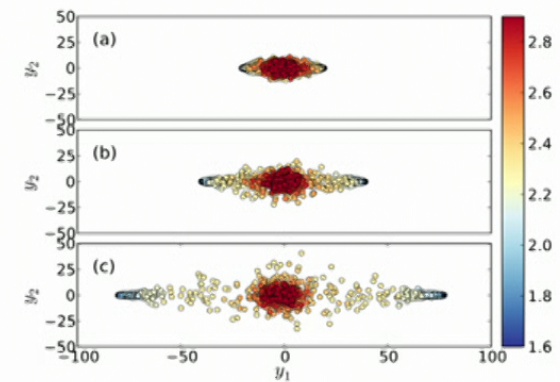


Figure 2: Projection of the samples onto the plane of the leading two principal components. The color bar on the right indicates the temperature T/J of the samples. The panels (a-c) are for $N = 20^2, 40^2$ and 80^2 sites respectively.



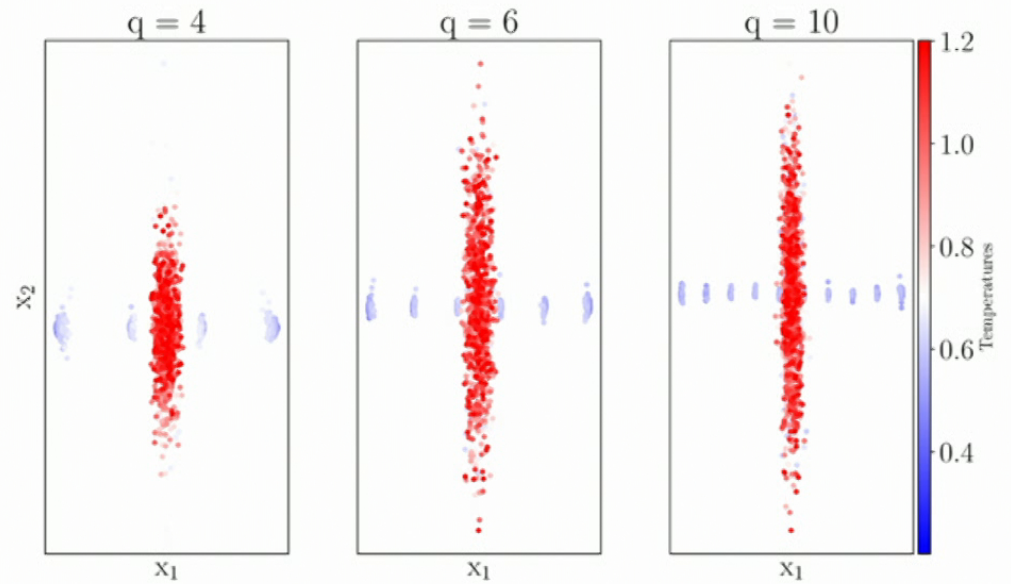
Distinguishing phases and detecting local and non-local order using t-SNE and Monte Carlo methods

Matthew Duschenes

An essay submitted
for partial fulfilment of
Perimeter Scholars International

June, 2018

PCA of Pott's model



$$H_{\text{Potts}} = -J \sum_{\langle ij \rangle} \delta_{s_i s_j} \quad s_i \in \{1, 2, \dots, q\}$$

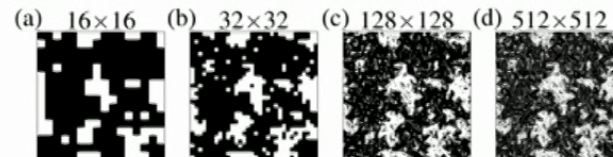
Condensed Matter > Statistical Mechanics

[Submitted on 4 Oct 2018 (v1), last revised 30 Jan 2019 (this version, v2)]

Super-resolving the Ising model with convolutional neural networks

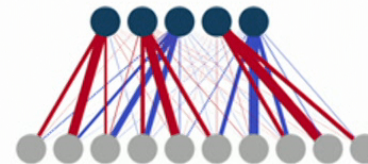
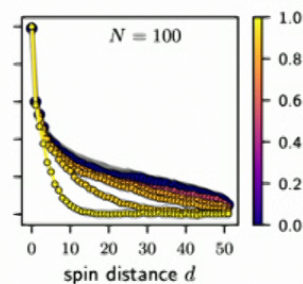
Stavros Efthymiou, Matthew J. S. Beach, Roger G. Melko

Machine learning is becoming widely used in condensed matter physics. Inspired by the concept of image super-resolution, we propose a method to increase the size of lattice spin configurations using deep convolutional neural networks. Through supervised learning on Monte Carlo (MC) generated spin configurations, we train networks that invert real-space renormalization decimations. We demonstrate that super-resolution can reproduce thermodynamic observables that agree with MC calculations for the one and two-dimensional Ising model at various temperatures. We find that it is possible to predict thermodynamic quantities for lattice sizes larger than those used in training by extrapolating the parameters of the network. We use this method to extrapolate the exponents of the 2D Ising critical point towards the thermodynamic limit, which results in good agreement with theory.



$$H = - \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

FIG. 2. Critical configurations obtained using the weight extrapolation idea presented in Section II C. We show the original Monte Carlo configuration in (a) and the results after (b) one, (c) three and (d) five consecutive super-resolutions.



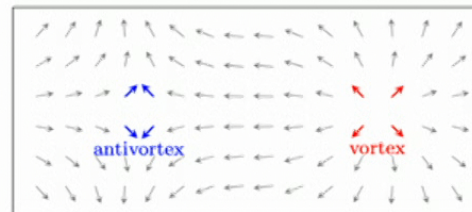
A. Golubeva

[Submitted on 26 Oct 2017]

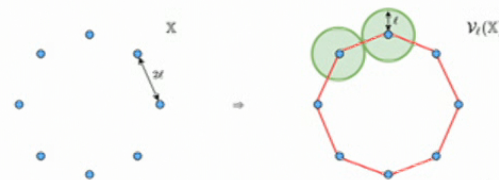
Machine learning vortices at the Kosterlitz–Thouless transition

Matthew J. S. Beach, Anna Golubeva, Roger G. Melko

Efficient and automated classification of phases from minimally processed data is one goal of machine learning in condensed matter and statistical physics. Supervised algorithms trained on raw samples of microstates can successfully detect conventional phase transitions via learning a bulk feature such as an order parameter. In this paper, we investigate whether neural networks can learn to classify phases based on topological defects. We address this question on the two-dimensional classical XY model which exhibits a Kosterlitz–Thouless transition. We find significant feature engineering of the raw spin states is required to convincingly claim that features of the vortex configurations are responsible for learning the transition temperature. We further show a single-layer network does not correctly classify the phases of the XY model, while a convolutional network easily performs classification by learning the global magnetization. Finally, we design a deep network capable of learning vortices without feature engineering. We demonstrate the detection of vortices does not necessarily result in the best classification accuracy, especially for lattices of less than approximately 1000 spins. For larger systems, it remains a difficult task to learn vortices.



$$\mathcal{H}_{XY} = -J \sum_{\langle ij \rangle} \cos(\theta_i - \theta_j)$$



Dan Sehayek

Quantum Physics

[Submitted on 20 Aug 2019 (v1), last revised 26 Aug 2019 (this version, v2)]

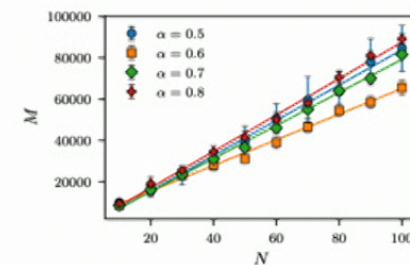
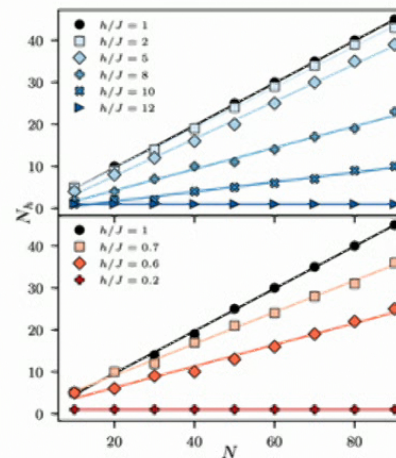
The learnability scaling of quantum states: restricted Boltzmann machines

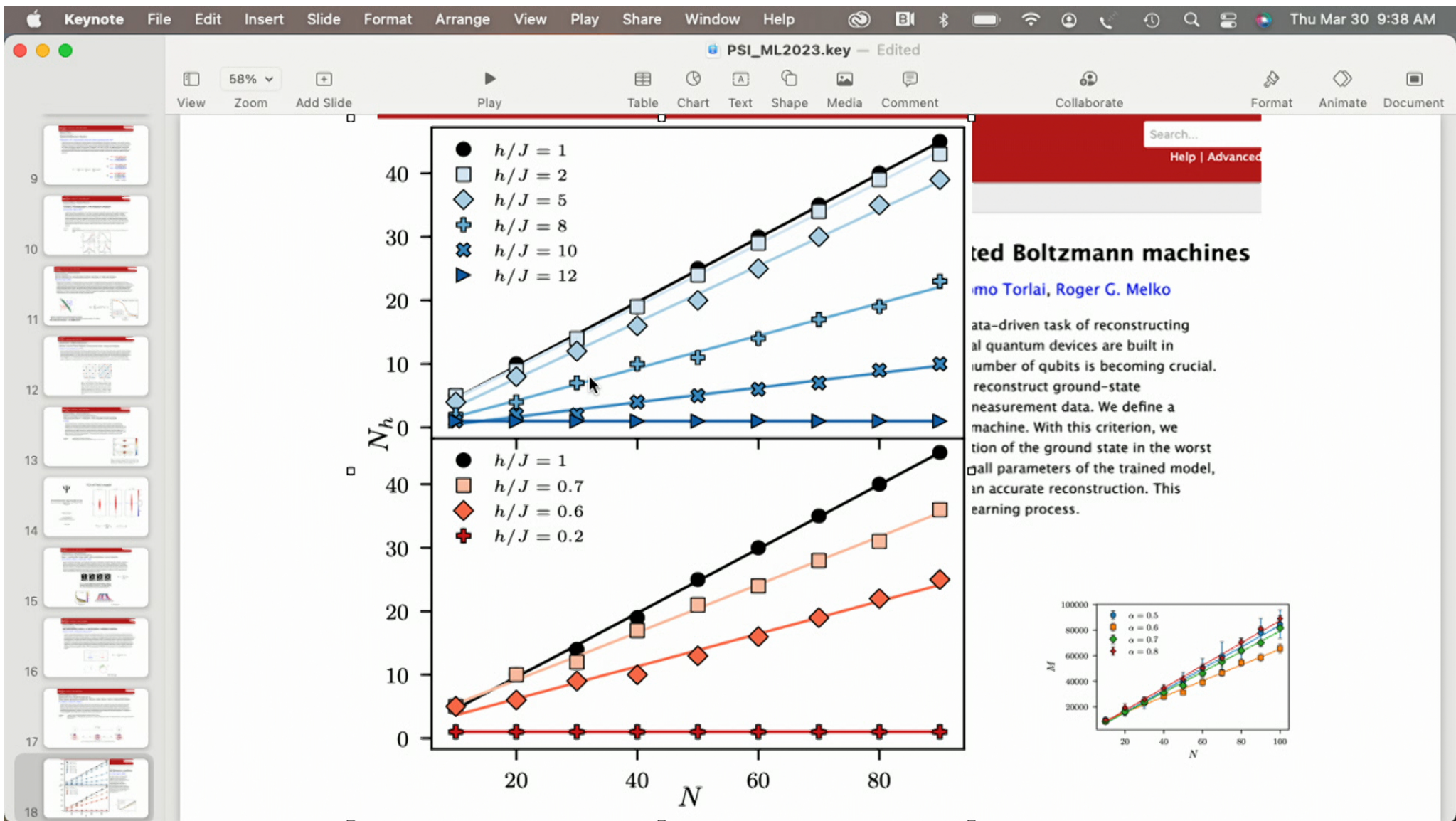
Dan Sehayek, Anna Golubeva, Michael S. Albergo, Bohdan Kulchytskyy, Giacomo Torlai, Roger G. Melko

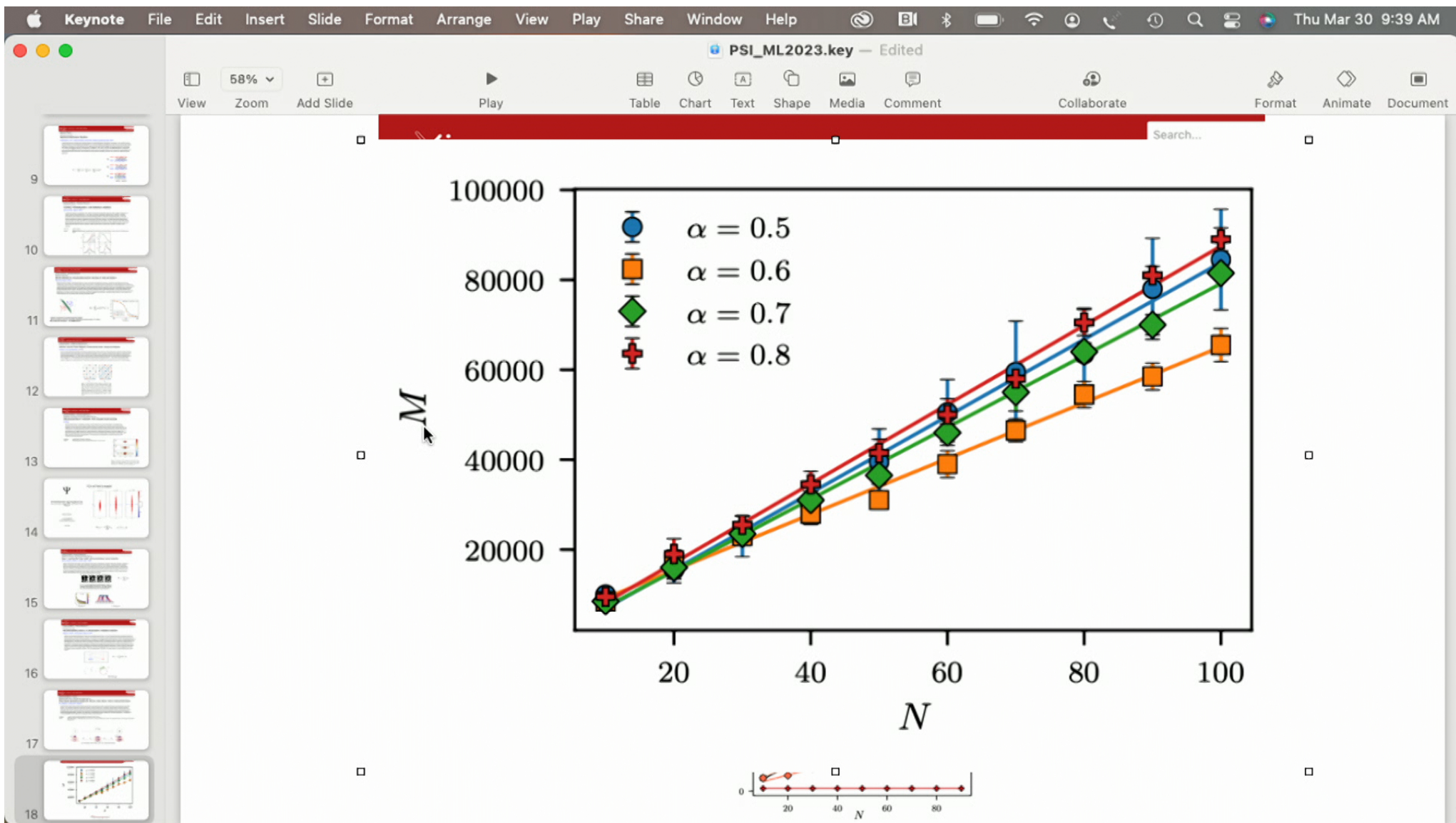
Generative modeling with machine learning has provided a new perspective on the data-driven task of reconstructing quantum states from a set of qubit measurements. As increasingly large experimental quantum devices are built in laboratories, the question of how these machine learning techniques scale with the number of qubits is becoming crucial. We empirically study the scaling of restricted Boltzmann machines (RBMs) applied to reconstruct ground-state wavefunctions of the one-dimensional transverse-field Ising model from projective measurement data. We define a learning criterion via a threshold on the relative error in the energy estimator of the machine. With this criterion, we observe that the number of RBM weight parameters required for accurate representation of the ground state in the worst case – near criticality – scales quadratically with the number of qubits. By pruning small parameters of the trained model, we find that the number of weights can be significantly reduced while still retaining an accurate reconstruction. This provides evidence that over-parametrization of the RBM is required to facilitate the learning process.

Comments: 8 pages, 5 figures

Subjects: Quantum Physics (quant-ph)







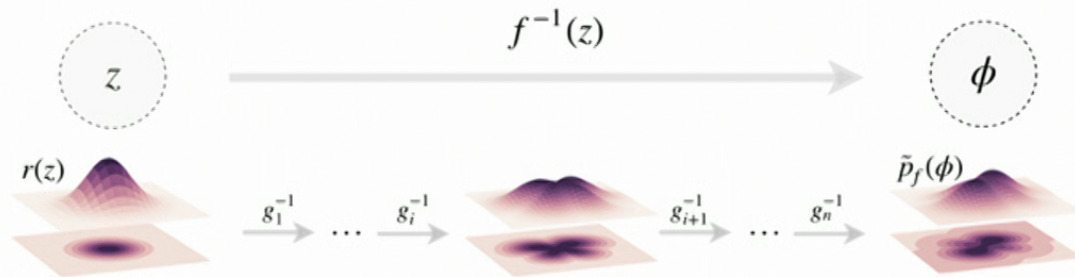
High Energy Physics - Lattice

*[Submitted on 26 Apr 2019 (v1), last revised 9 Sep 2019 (this version, v3)]***Flow-based generative models for Markov chain Monte Carlo in lattice field theory**

M. S. Albergo, G. Kanwar, P. E. Shanahan

A Markov chain update scheme using a machine-learned flow-based generative model is proposed for Monte Carlo sampling in lattice field theories. The generative model may be optimized (trained) to produce samples from a distribution approximating the desired Boltzmann distribution determined by the lattice action of the theory being studied. Training the model systematically improves autocorrelation times in the Markov chain, even in regions of parameter space where standard Markov chain Monte Carlo algorithms exhibit critical slowing down in producing decorrelated updates. Moreover, the model may be trained without existing samples from the desired distribution. The algorithm is compared with HMC and local Metropolis sampling for ϕ^4 theory in two dimensions.

Comments: 13 pages, 7 figures; corrected normalization conventions in eqns. 20 and 23

Subjects: **High Energy Physics - Lattice (hep-lat)**; Disordered Systems and Neural Networks (cond-mat.dis-nn); Statistical Mechanics (cond-mat.stat-mech); Machine Learning (cs.LG)

(a) Normalizing flow between prior and output distributions

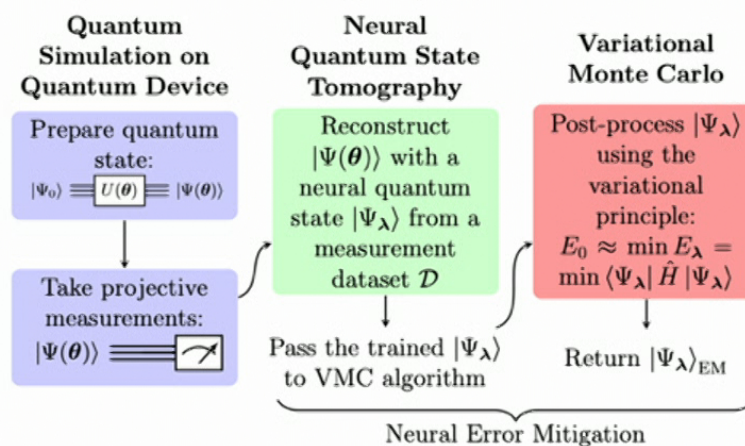
Quantum Physics

[Submitted on 17 May 2021]

Neural Error Mitigation of Near-Term Quantum Simulations

Elizabeth R. Bennewitz, Florian Hopfmueller, Bohdan Kulchytskyy, Juan Carrasquilla, Pooya Ronagh

One of the promising applications of early quantum computers is the simulation of quantum systems. Variational methods for near-term quantum computers, such as the variational quantum eigensolver (VQE), are a promising approach to finding ground states of quantum systems relevant in physics, chemistry, and materials science. These approaches, however, are constrained by the effects of noise as well as the limited quantum resources of near-term quantum hardware, motivating the need for quantum error mitigation techniques to reduce the effects of noise. Here we introduce *neural error mitigation*, a novel method that uses neural networks to improve estimates of ground states and ground-state observables obtained using VQE on near-term quantum computers. To demonstrate our method's versatility, we apply neural error mitigation to finding the ground states of H_2 and LiH molecular Hamiltonians, as well as the lattice Schwinger model. Our results show that neural error mitigation improves the numerical and experimental VQE computation to yield low-energy errors, low infidelities, and accurate estimations of more-complex observables like order parameters and entanglement entropy, without requiring additional quantum resources. Additionally, neural error mitigation is agnostic to both the quantum hardware and the particular noise channel, making it a versatile tool for quantum simulation. Applying quantum many-body machine learning techniques to error mitigation, our method is a promising strategy for extending the reach of near-term quantum computers to solve complex quantum simulation problems.



Quantum Physics

[Submitted on 20 Aug 2019 (v1), last revised 26 Aug 2019 (this version, v2)]

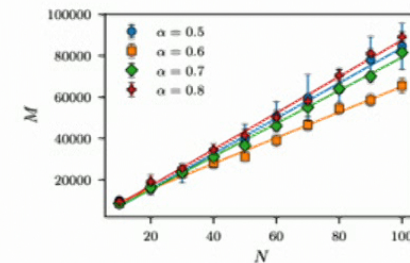
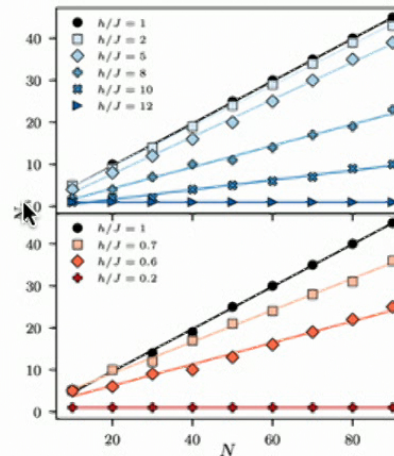
The learnability scaling of quantum states: restricted Boltzmann machines

Dan Sehayek, Anna Golubeva, Michael S. Albergo, Bohdan Kulchytskyy, Giacomo Torlai, Roger G. Melko

Generative modeling with machine learning has provided a new perspective on the data-driven task of reconstructing quantum states from a set of qubit measurements. As increasingly large experimental quantum devices are built in laboratories, the question of how these machine learning techniques scale with the number of qubits is becoming crucial. We empirically study the scaling of restricted Boltzmann machines (RBMs) applied to reconstruct ground-state wavefunctions of the one-dimensional transverse-field Ising model from projective measurement data. We define a learning criterion via a threshold on the relative error in the energy estimator of the machine. With this criterion, we observe that the number of RBM weight parameters required for accurate representation of the ground state in the worst case – near criticality – scales quadratically with the number of qubits. By pruning small parameters of the trained model, we find that the number of weights can be significantly reduced while still retaining an accurate reconstruction. This provides evidence that over-parametrization of the RBM is required to facilitate the learning process.

Comments: 8 pages, 5 figures

Subjects: Quantum Physics (quant-ph)



Quantum Physics

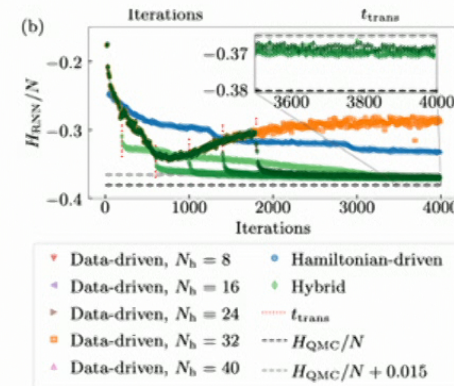
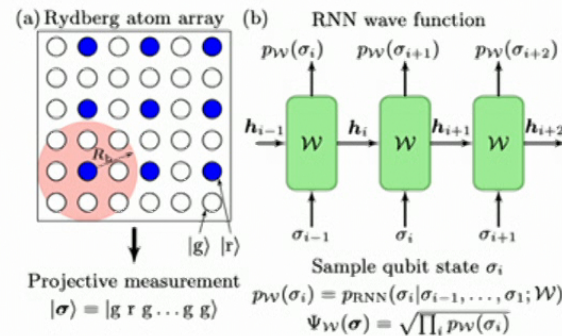
[Submitted on 9 Mar 2022 (v1), last revised 9 May 2022 (this version, v2)]

Data-Enhanced Variational Monte Carlo Simulations for Rydberg Atom Arrays

Stefanie Czischek, M. Schuyler Moss, Matthew Radzihovsky, Ejaaz Merali, Roger G. Melko

Rydberg atom arrays are programmable quantum simulators capable of preparing interacting qubit systems in a variety of quantum states. Due to long experimental preparation times, obtaining projective measurement data can be relatively slow for large arrays, which poses a challenge for state reconstruction methods such as tomography. Today, novel groundstate wavefunction ansätze like recurrent neural networks (RNNs) can be efficiently trained not only from projective measurement data, but also through Hamiltonian-guided variational Monte Carlo (VMC). In this paper, we demonstrate how pretraining modern RNNs on even small amounts of data significantly reduces the convergence time for a subsequent variational optimization of the wavefunction. This suggests that essentially any amount of measurements obtained from a state prepared in an experimental quantum simulator could provide significant value for neural-network-based VMC strategies.

Comments: 7 pages, 3 figures

Subjects: **Quantum Physics (quant-ph)**; Disordered Systems and Neural Networks (cond-mat.dis-nn); Computational Physics (physics.comp-ph)

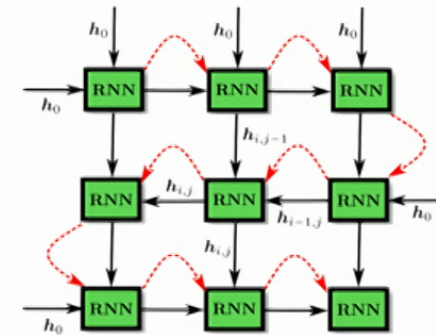
Condensed Matter > Disordered Systems and Neural Networks

*[Submitted on 7 Feb 2020 (v1), last revised 20 Jun 2020 (this version, v4)]***Recurrent Neural Network Wave Functions**

Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, Juan Carrasquilla

A core technology that has emerged from the artificial intelligence revolution is the recurrent neural network (RNN). Its unique sequence-based architecture provides a tractable likelihood estimate with stable training paradigms, a combination that has precipitated many spectacular advances in natural language processing and neural machine translation. This architecture also makes a good candidate for a variational wave function, where the RNN parameters are tuned to learn the approximate ground state of a quantum Hamiltonian. In this paper, we demonstrate the ability of RNNs to represent several many-body wave functions, optimizing the variational parameters using a stochastic approach. Among other attractive features of these variational wave functions, their autoregressive nature allows for the efficient calculation of physical estimators by providing independent samples. We demonstrate the effectiveness of RNN wave functions by calculating ground state energies, correlation functions, and entanglement entropies for several quantum spin models of interest to condensed matter physicists in one and two spatial dimensions.

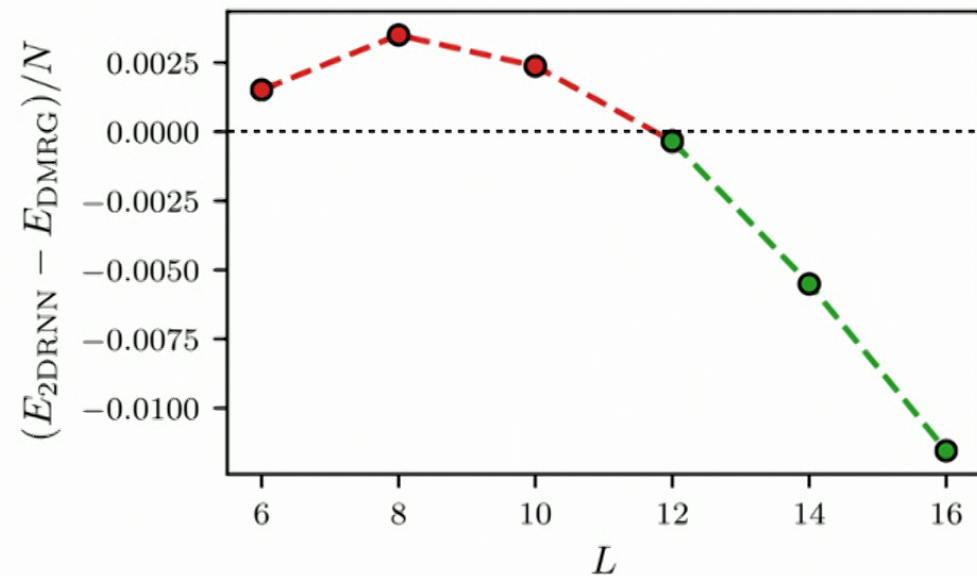
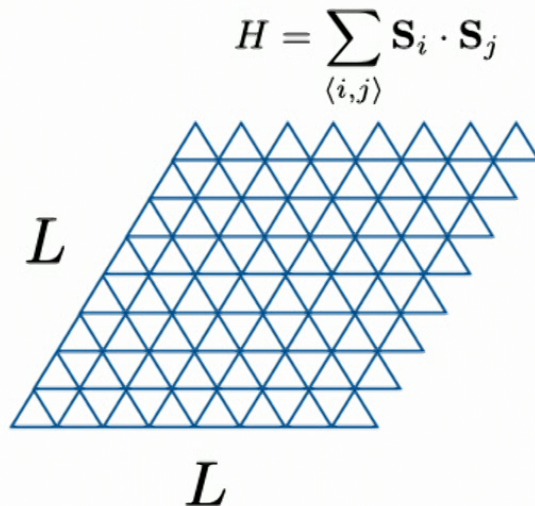
Comments: The GitHub link to the open-source code is fixed

Subjects: **Disordered Systems and Neural Networks (cond-mat.dis-nn)**; Strongly Correlated Electrons (cond-mat.str-el); Computational Physics (physics.comp-ph); Quantum Physics (quant-ph)

Application: variational wavefunction optimization

arXiv:2207.14314

$$\mathcal{L} = \langle \psi_\lambda | H | \psi_\lambda \rangle$$



- Spin-1/2 Heisenberg on the triangular lattice
- 2D RNN wavefunction energies begin to beat DMRG around size 12×12
- Used 1.6M parameters (0.1% of the parameters of the DMRG)

GPT-4 visual input example, Solving École Polytechnique's Physics Exam:

User Answer question I.1.a. Think step-by-step.

I. Principe de la détection de rayonnement avec un bolomètre

Comme illustré sur la figure 1 un bolomètre est constitué d'un absorbeur qui reçoit le rayonnement que l'on désire détecter. Sa température T , supposée uniforme, est mesurée à l'aide d'un thermomètre incorporé, constitué d'un matériau conducteur dont la résistance $R(T)$ varie avec la température T ; cette variation est caractérisée par le coefficient $\alpha = \frac{1}{R} \frac{dR}{dT}$. L'ensemble possède la capacité thermique C_{th} .

Un barreau, conducteur thermique, homogène, de longueur L , de section S et de conductivité thermique λ et sans échanges thermiques latéraux, relie le bolomètre à un thermostat de température T_b fixe.

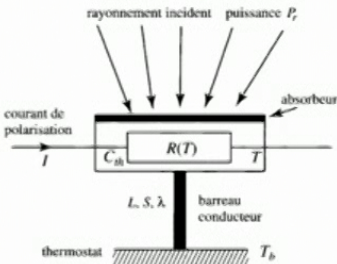


Figure 1 - Schéma d'un bolomètre

I.1. Dans cette première partie, on étudie le comportement du bolomètre en régime stationnaire, à la température de fonctionnement $T = T_0$.

I.1.a) Exprimer la température $T(x)$ en tout point du barreau conducteur repéré par son abscisse x ; on prendra $T(0) = T_0$ et $T(L) = T_b$.

GPT-4 To answer question I.1.a, we need to express the temperature $T(x)$ in every point of the conductive bar, which is characterized by the abscissa x . We are given that $T(0) = T_0$ and $T(L) = T_b$.

For a stationary regime, the heat equation for the one-dimensional heat conduction in the bar is:
$$d^2T/dx^2 = 0$$

Integrating once, we get:
$$dT/dx = C_1$$

Integrating a second time:
$$T(x) = C_1x + C_2$$

To determine the constants C_1 and C_2 , we use the boundary conditions:
 $T(0) = T_0$, which gives $C_2 = T_0$
 $T(L) = T_b$, which gives $C_1 = (T_b - T_0)/L$

Finally, the expression for $T(x)$ is:
$$T(x) = (T_b - T_0) * (x/L) + T_0$$

This represents a linear temperature profile along the conductive bar, with a slope of $(T_b - T_0)/L$.

Table 15. Example prompt demonstrating GPT-4's visual input capability. The prompt consists of a question which requires understanding a physics problem with a diagram written in French and solving it using chain-of-thought prompting.

| Exam | GPT-4 | GPT-4 (no vision) | GPT-3.5 |
|------------------------------------------------|-------------------------|-------------------------|------------------------|
| Uniform Bar Exam (MBE+MEE+MPT) | 298 / 400 (~90th) | 298 / 400 (~90th) | 213 / 400 (~10th) |
| LSAT | 163 (~88th) | 161 (~83rd) | 149 (~40th) |
| SAT Evidence-Based Reading & Writing | 710 / 800 (~93rd) | 710 / 800 (~93rd) | 670 / 800 (~87th) |
| SAT Math | 700 / 800 (~89th) | 690 / 800 (~89th) | 590 / 800 (~70th) |
| Graduate Record Examination (GRE) Quantitative | 163 / 170 (~80th) | 157 / 170 (~62nd) | 147 / 170 (~25th) |
| Graduate Record Examination (GRE) Verbal | 169 / 170 (~99th) | 165 / 170 (~96th) | 154 / 170 (~63rd) |
| Graduate Record Examination (GRE) Writing | 4 / 6 (~54th) | 4 / 6 (~54th) | 4 / 6 (~54th) |
| USABO Semifinal Exam 2020 | 87 / 150 (99th - 100th) | 87 / 150 (99th - 100th) | 43 / 150 (31st - 33rd) |
| USNCO Local Section Exam 2022 | 36 / 60 | 38 / 60 | 24 / 60 |
| Medical Knowledge Self-Assessment Program | 75 % | 75 % | 53 % |
| Codeforces Rating | 392 (below 5th) | 392 (below 5th) | 260 (below 5th) |
| AP Art History | 5 (86th - 100th) | 5 (86th - 100th) | 5 (86th - 100th) |
| AP Biology | 5 (85th - 100th) | 5 (85th - 100th) | 4 (62nd - 85th) |
| AP Calculus BC | 4 (43rd - 59th) | 4 (43rd - 59th) | 1 (0th - 7th) |
| AP Chemistry | 4 (71st - 88th) | 4 (71st - 88th) | 2 (22nd - 46th) |
| AP English Language and Composition | 2 (14th - 44th) | 2 (14th - 44th) | 2 (14th - 44th) |
| AP English Literature and Composition | 2 (8th - 22nd) | 2 (8th - 22nd) | 2 (8th - 22nd) |
| AP Environmental Science | 5 (91st - 100th) | 5 (91st - 100th) | 5 (91st - 100th) |
| AP Macroeconomics | 5 (84th - 100th) | 5 (84th - 100th) | 2 (33rd - 48th) |
| AP Microeconomics | 5 (82nd - 100th) | 4 (60th - 82nd) | 4 (60th - 82nd) |
| AP Physics 2 | 4 (66th - 84th) | 4 (66th - 84th) | 3 (30th - 66th) |
| AP Psychology | 5 (83rd - 100th) | 5 (83rd - 100th) | 5 (83rd - 100th) |
| AP Statistics | 5 (85th - 100th) | 5 (85th - 100th) | 3 (40th - 63rd) |
| AP US Government | 5 (88th - 100th) | 5 (88th - 100th) | 4 (77th - 88th) |
| AP US History | 5 (89th - 100th) | 4 (74th - 89th) | 4 (74th - 89th) |
| AP World History | 4 (65th - 87th) | 4 (65th - 87th) | 4 (65th - 87th) |
| AMC 10 | 30 / 150 (6th - 12th) | 36 / 150 (10th - 19th) | 36 / 150 (10th - 19th) |
| AMC 12 | 60 / 150 (45th - 66th) | 48 / 150 (19th - 40th) | 30 / 150 (4th - 8th) |
| Introductory Sommelier (theory knowledge) | 92 % | 92 % | 80 % |
| Certified Sommelier (theory knowledge) | 86 % | 86 % | 58 % |
| Advanced Sommelier (theory knowledge) | 77 % | 77 % | 46 % |
| Leetcode (easy) | 31 / 41 | 31 / 41 | 12 / 41 |
| Leetcode (medium) | 21 / 80 | 21 / 80 | 8 / 80 |
| Leetcode (hard) | 3 / 45 | 3 / 45 | 0 / 45 |

Table 1. GPT performance on academic and professional exams. In each case, we simulate the conditions and scoring of the real exam. We report GPT-4's final score graded according to exam-specific rubrics, as well as the percentile of test-takers achieving GPT-4's score.

Language models are demonstrating scalability

Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877–1901.

| Model Name | n_{params} | n_{layers} | d_{model} | n_{heads} | d_{head} | Batch Size | Learning Rate |
|-----------------------|---------------------|---------------------|--------------------|--------------------|-------------------|------------|----------------------|
| GPT-3 Small | 125M | 12 | 768 | 12 | 64 | 0.5M | 6.0×10^{-4} |
| GPT-3 Medium | 350M | 24 | 1024 | 16 | 64 | 0.5M | 3.0×10^{-4} |
| GPT-3 Large | 760M | 24 | 1536 | 16 | 96 | 0.5M | 2.5×10^{-4} |
| GPT-3 XL | 1.3B | 24 | 2048 | 24 | 128 | 1M | 2.0×10^{-4} |
| GPT-3 2.7B | 2.7B | 32 | 2560 | 32 | 80 | 1M | 1.6×10^{-4} |
| GPT-3 6.7B | 6.7B | 32 | 4096 | 32 | 128 | 2M | 1.2×10^{-4} |
| GPT-3 13B | 13.0B | 40 | 5140 | 40 | 128 | 2M | 1.0×10^{-4} |
| GPT-3 175B or "GPT-3" | 175.0B | 96 | 12288 | 96 | 128 | 3.2M | 0.6×10^{-4} |

Table 2.1: Sizes, architectures, and learning hyper-parameters (batch size in tokens and learning rate) of the models which we trained. All models were trained for a total of 300 billion tokens.

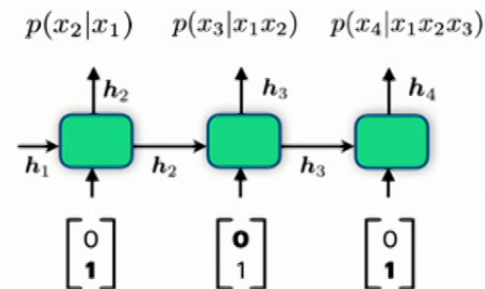
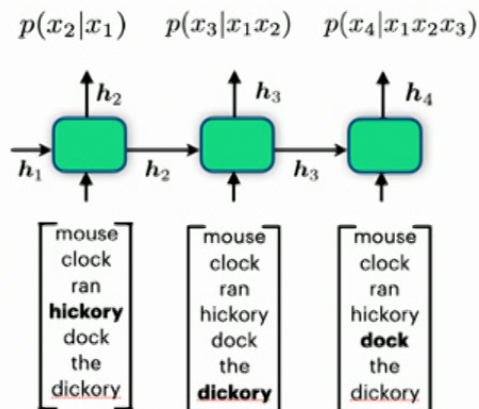
GPT-3 was trained with data from [CommonCrawl](#), [WebText](#), Wikipedia, and a corpus of books.

- Generative pre-trained transformer... overheard at aspen: "GPT4 cost \$100M to train"
- GPT are parameterized probabilistic models, whose most basic function is to predict the next word in a sequence given the words that precede it

Sequences: generative autoregressive models

Conditional probabilities of the next word (or qubit). In a sequence they obey the *autoregressive* property, i.e. chain rule

$$p(x_1, \dots, x_N) = \prod_{j=1}^N p(x_j | x_{<j}) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \dots$$

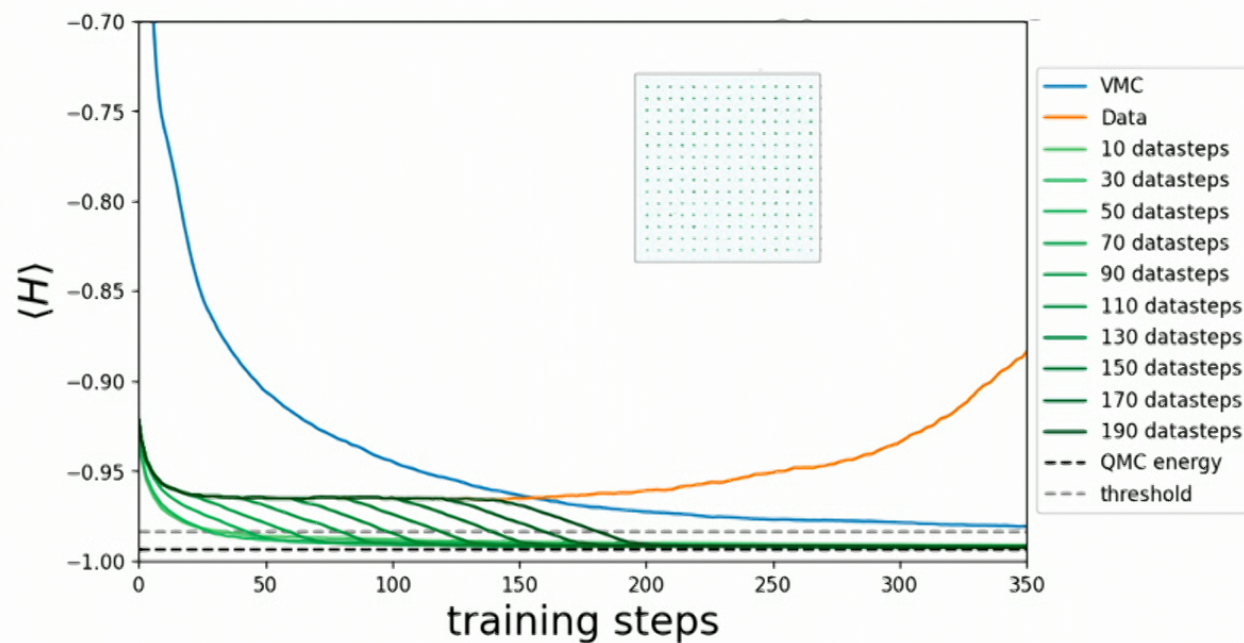


256 qubits 1000 shots

- 16x16 lattice, near the disordered-checkerboard transition
- state is adiabatically prepared by slowly varying detuning
- 1000 projective measurements per detuning parameter
- Used to train a state-of-the-art 2D RNN wavefunction



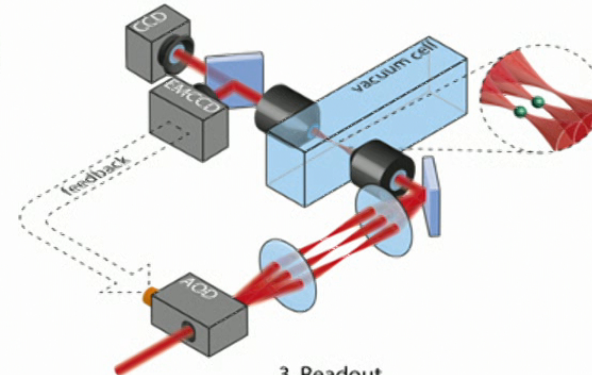
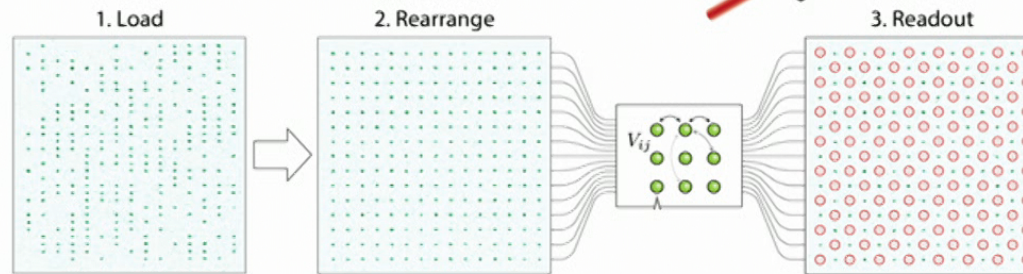
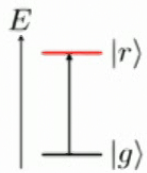
M. Schuyler Moss



Experimental Rydberg Arrays

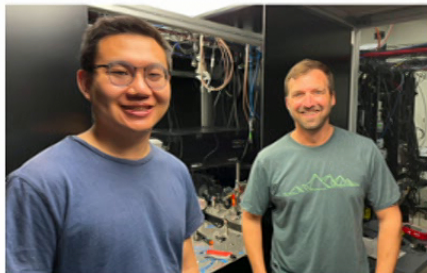
Ebadi et. al. arXiv:2012.12281
Nature 595, 227 (2021)

^{87}Rb



HARVARD UNIVERSITY

IQEra



Emergence resurgence

arXiv > cs > arXiv:2206.07682

Search...

Help | Advanced S

Computer Science > Computation and Language

[Submitted on 15 Jun 2022 (v1), last revised 26 Oct 2022 (this version, v2)]

Emergent Abilities of Large Language Models

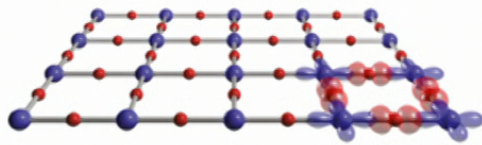
Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, William Fedus

Scaling up language models has been shown to predictably improve performance and sample efficiency on a wide range of downstream tasks. This paper instead discusses an unpredictable phenomenon that we refer to as emergent abilities of large language models. We consider an ability to be emergent if it is not present in smaller models but is present in larger models. Thus, emergent abilities cannot be predicted simply by extrapolating the performance of smaller models. The existence of such emergence implies that additional scaling could further expand the range of capabilities of language models.

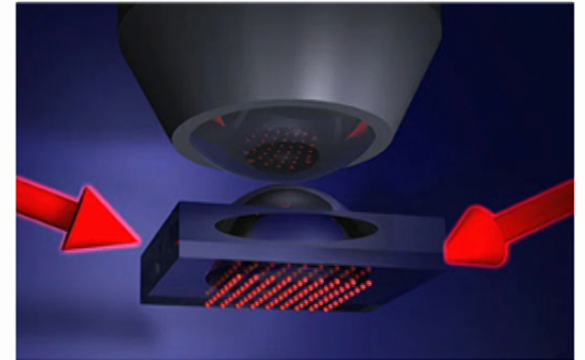
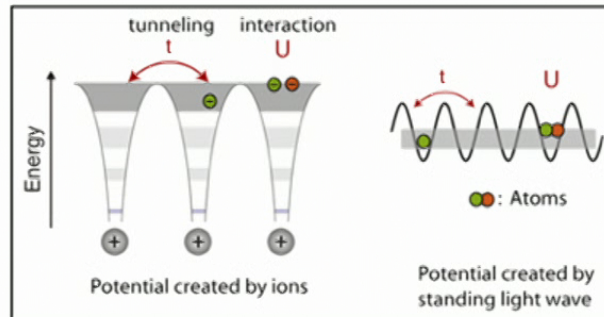
Comments: Transactions on Machine Learning Research (TMLR), 2022

Subjects: **Computation and Language (cs.CL)**

Quantum “simulation” (or emulation)



$$H = -t \sum_{\langle ij \rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + c_{i\sigma} c_{j\sigma}^\dagger) + U \sum_i n_{i\uparrow} n_{i\downarrow}$$



Can the emergence of unexplained macroscopic phenomena be understood by building quantum computers?

Data obtained on challenging models could be used to drastically improve variational calculations...

PIQuIL / RydbergGPT
Private
Edit Pins
Up

Code
Issues
Pull requests
Discussions
Actions
Projects
Security
Insights
Settings

main
1 branch
0 tags
Go to file
Add file
Code

| | |
|-------------------------------------------------------|----------------------------------------------------|
| yhteoh Update README.md 4abe2e2 5 days ago 23 commits | David Fitzek davidfitzek · Outside Collaborator |
| LLM dev team @PIQuIL/llm-dev-team · 3 members | |

examples Added skeleton for rydberggpt package last week
resources Updated architecture diagram last week
rydberggpt Updated skeleton for rydberggpt package last week
.gitignore Updated readme last week
README.md Update README.md 5 days ago

README.md

RydbergGPT

A large language model (LLM) for Rydberg atom array physics.

Architecture

Vanilla transformer architecture taken from [Attention is All You Need](#).

Many of you may be aware of an open letter circulating from the Future of Life Institute, asking for a pause on giant AI experiments released yesterday.



Thanks!

