CHIN. PHYS. LETT. Vol. 37, No. 1 (2020) 018401

Predicting Quantum Many-Body Dynamics with Transferable Neural Networks *

Ze-Wang Zhang(张泽旺)[1], Shuo Yang(杨硕)[2], Yi-Hang Wu(吴亦航)[1], Chen-Xi Liu(刘晨曦)[1], Yi-Min Han(韩翊民)[1], Ching-Hua Lee[3,4], Zheng Sun(孙政)[1], Guang-Jie Li(李光杰)[1], Xiao Zhang(张笑)[1]**

[1]*School of Physics, Sun Yat-sen University, Guangzhou 510275*
[2]*State Key Laboratory of Low-Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing 100084*
[3]*Department of Physics, National University of Singapore, 117542, Singapore*
[4]*Institute of High Performance Computing, 138632, Singapore*

(Received 24 October 2019)

Advanced machine learning (ML) approaches such as transfer learning have seldom been applied to approximate quantum many-body systems. Here we demonstrate that a simple recurrent unit (SRU) based efficient and transferable sequence learning framework is capable of learning and accurately predicting the time evolution of the one-dimensional (1D) Ising model with simultaneous transverse and parallel magnetic fields, as quantitatively corroborated by relative entropy measurements between the predicted and exact state distributions. At a cost of constant computational complexity, a larger many-body state evolution is predicted in an autoregressive way from just one initial state, without any guidance or knowledge of any Hamiltonian. Our work paves the way for future applications of advanced ML methods in quantum many-body dynamics with knowledge only from a smaller system.

PACS: 84.35.+i, 05.50.+q, 02.70.−c DOI: 10.1088/0256-307X/37/1/018401

Machine learning (ML) approaches, particularly neural networks (NNs), have achieved great success in solving real-world industrial and social problems.[1−8] Inspired by its widespread applicability, ML was soon adopted by condensed matter physicists in the modeling of quantum many-body behavior and phase transition discovery.[9−19] Compared to many advanced applications, it is natural to ask if recent progress in these more sophisticated ML architectures can benefit or even revolutionize the modeling of quantum systems. For instance, can quantum many-body dynamics be "learned" through transferable learning?[20,21]

Thus, the main objective of this work is to demonstrate the novel application of NNs in the transferable learning and prediction of the evolution of a many-body wavefunction, an otherwise computationally intensive task that has not been solved by generative models. Focusing on static problems, it is proven that deep NNs like the restricted Boltzmann machine (RBM) can represent most physical states,[22] and a recent work based on deep CNNs shows the ability to circumvent the need for Markov Chain sampling on the two-dimensional interacting spin model of larger systems.[23] Lately, physical properties of spin Hamiltonians are reproduced by the deep Boltzmann machine (DBM), as an alternative to the standard path integral.[24] Our approach is fundamentally in contrast with conventional approaches in computing many-body dynamics: instead of evolving the wavefunction explicitly with the Hamiltonian, we directly predict the dynamical wavefunction from the initial state by propagating it with an efficient and transferable framework based on unified spin encoding, chain encoding and SRU[25] module. With the same level of parallelism as feed-forward CNNs and scalable context-dependent capacity of recurrent connections, our proposed framework is naturally suited for learning many-body systems with unified parameters, although they have never been harnessed for an exact quantum state evolution; in our scenario, a 1D Ising model with both parallel and transverse magnetic field.

Inspired by end-to-end training[26] and domain adaptation,[27,28] we specialize in the many-body dynamics of a 1D Ising chain with transverse and parallel magnetic fields. Comparison with exact conventionally computed results with up to seven spins reveals high predictive accuracy, as quantified by the relative entropy as well as magnetization. Indeed, our SRU-propagated wavefunction shows a strong grasp of the periodicity in the time evolution, despite being unaware of the Hamiltonian that sets the energy (inverse periodicity) scale. Encouraged by circumventing the problem of exponential computational complexity through unified encoding mechanisms and parallel recurrent connections, we hope that such encouraging results from our pioneering transferable learning approach will inspire further applications of transferable learning methods to build a shared model suited for quantum systems with vast spin variables.

We consider a 1D Ising spin chain composed of N spin variables with local transverse (g) and parallel (h) magnetic fields, described by the Hamiltonian

$$H = -\sum_{i}^{N-1} \sigma_i^z \otimes \sigma_{i+1}^z - h\sum_{i}^{N} \sigma_i^z - g\sum_{i}^{N} \sigma_i^x, \quad (1)$$

*Supported by the National Natural Science Foundation of China under Grant Nos 11874431 and 11804181, the National Key R&D Program of China under Grant No 2018YFA0306800, and the Guangdong Science and Technology Innovation Youth Talent Program under Grant Nos 2016TQ03X688 and 2018YFA0306504, and the Research Fund Program of the State Key Laboratory of Low-Dimensional Quantum Physics under Grant No ZZ201803.

**Corresponding author. Email: zhangxiao@mail.sysu.edu.cn

where σ^x and σ^z denote the Pauli matrices, and i denotes the spin variable. When the magnetic field is parallel ($g = 0$) or transverse ($h = 0$), the Hamiltonian is exactly solvable. However, when $g \neq 0$ and $h \neq 0$, the dynamics of N spins must be numerically computed in the 2^N-dimensional many-body Hilbert space spanned by direct product states Ψ of single-spin wavefunctions ψ_i:

$$\Psi = \prod_i^N \otimes \psi_i = \prod_i^N \otimes \begin{pmatrix} \phi_i^{\uparrow} \\ \phi_i^{\downarrow} \end{pmatrix}, \quad \dim \Psi = 2^N. \quad (2)$$

Wavefunction dynamics can be exactly computed through unitary time evolution of the Hamiltonian

$$\begin{aligned} |\Psi(t)\rangle &= \exp\left(-i\frac{H}{\hbar}t\right)|\Psi(0)\rangle \\ &= V\exp\left(-i\frac{E}{\hbar}t\right)V^{-1}|\Psi(0)\rangle, \end{aligned} \quad (3)$$

where $E = V^{-1}HV$ is the diagonal eigenenergy matrix.

The 2^N-dimensional N-body wave function $|\Psi(t)\rangle$ quickly becomes expensive to compute as N increases. Inspired by the sequence generation model,[29] our framework (as shown in Fig. 1(b)) is composed of a spin encoding layer, chain encoding layer, SRU layers, and spin decoding layer, which instead attempts to predict its time evolution based on prior knowledge of the time evolution behavior of known training states. This training (learning) only has to be performed once for the relatively inexpensive prediction of any number of initial states. Importantly, the training and prediction process captures solely the intrinsic evolution patterns of the wavefunctions, and does not involve any explicit knowledge about the Hamiltonian. Moreover, as we shall explain, our SRU-based framework is transferable. We next outline the broad principles behind our NN approach of predicting quantum state evolution, for details see the Supplementary Material. The vanilla SRU NN with peephole connections (Fig. 1(a)) substitutes inherent matrix multiplication with parallelizable element-wise multiplication operations ($\odot$ in Fig. 1(a)) associated with c_{t-1}, hence the calculation of f_t does not have to wait until the whole c_{t-1} is updated. With the help of spin encoding and decoding layers, the amount of trained parameters is fixed, and thus the complexity has an upper bound instead of increase exponentially.

Our procedure occurs in two main stages: the training stage and the inference stage. We first "train" or optimize the weight parameters of our SRU-based framework in a teacher forcing mode[30] by feeding it with a large number of training sequences, which are the time-evolved wavefunction data of 10^4 randomly chosen initial 2-spin to 7-spin state sequences sampled over 500 time steps, obtained via conventional exact diagonalization (ED). The SRU-based framework is fully optimized by the Adam optimization algorithm[31] to minimize the mean squared error between the ED-evolved and SRU-evolved states at all time steps in a mini-batch (see the Supplementary Material).

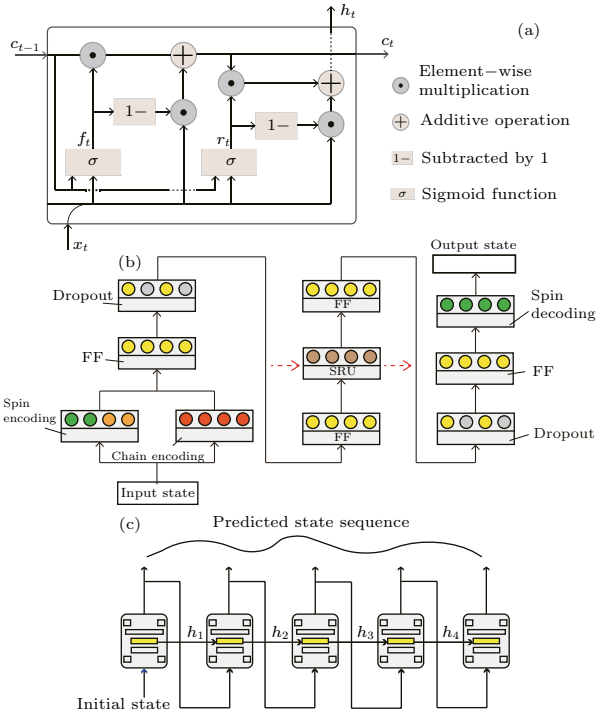

Fig. 1. (a) Details of a vanilla SRU cell with forget gate f_t, skip gate r_t, memory cell c_t and weighted connections from previous memory cell c_{t-1} to both forget gate f_t and skip gate r_t. X_t is the input and h_t is the output hidden state of SRU cell. The forget gate decides what to forget from previous memory cell c_{t-1} and what redundant information to drop when adapting to other systems. The skip gate, along with current memory cell c_t, decides what to skip from input and what to output as h_t. Here $\sigma(x) = \frac{e^x}{e^x+1}$ is the activation function. (b) Proposed architecture we take in this study as a block composed of a spin encoding layer, a chain encoding layer, SRU layers and a spin decoding layer at each time step. Spin encoding, chain encoding and spin decoding are all feed-forward layers (labeled as FF). (c) Autoregressive procedure of generating new quantum states, given an initial state at the beginning. Each time-evolved quantum state is predicted from our proposed block by SRU's cell memory information (shown as arrows in the middle) and previous predicted state.

Once trained well, the SRU-based framework is ready to predict the evolution of arbitrarily given initial states. As sketched in Fig. 1(c), the initial state $|\Psi(t=0)\rangle$ enters the leftmost block at $t = 0$, and is then processed by a spin encoding layer, a chain encoding layer, and two fully connected layers, and its output is propagated as an input state to the next block with hidden layers h_t. The output of each block denotes a new quantum state at a certain time step. The combination of memory cell c_t and hidden output h_t acts as effective context-dependent behaviors. As illustrated in Fig. 1(a) and further elaborated in the Supplementary Material, context-dependent information kept in memory cell c_t is modified by its previous value c_{t-1}, new input x_t interacted with forgotten gate and skipped gate at that time step, as well as "hidden" information on h_{t-1} from the previous SRU cell. Based on the already optimized SRU-

based framework, the final predicted quantum state as a function of time would be generated from one fully connected layer and the spin decoding layer as shown in Fig. 1(b).

We report very encouraging agreements between wavefunctions evolved by $e^{-iHt/\hbar}$ as computed by ED, and wavefunction evolutions as predicted by our SRU-based framework. As for the 1D Ising model, we set the local transverse magnetic field g to be -1.05, parallel magnetic field h to be 0.5 and Δt, the time interval to be 0.002, and keep this setting constant for all computation. We find that the maximum energy eigenvalue is about $0.1 \gg 0.002$, proving that the chosen time interval is small enough. The number of spin variables studied (2 to 7) decides the cost of exactly computing the 10^4 different time evolutions over 0.2 s (100 time steps) prior to training the network, since the time complexity of the ED method is $\mathcal{O}(2^n)$.

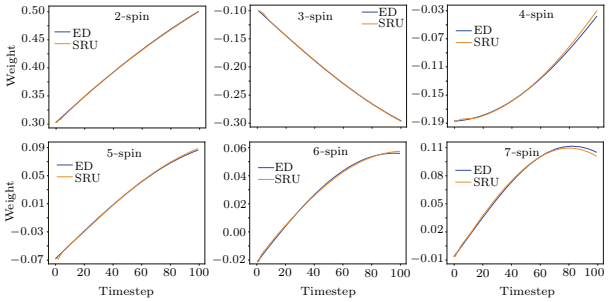

Fig. 2. Output (SRU-based) and target (ED-based) wavefunction magnitude (y-axis) for 2-spin to 7-spin systems, with the initial states given in the Supplementary Material. We plot the curves of the total 100 time steps.

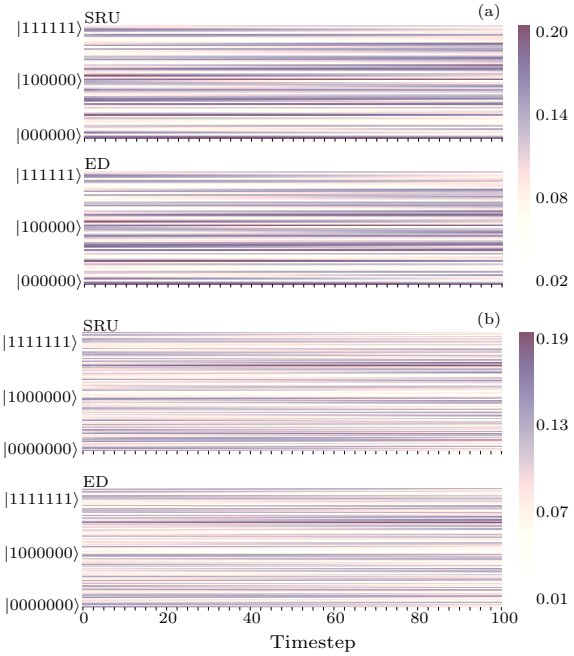

Fig. 3. Comparisons of all components of coefficients for SRU-based prediction and ED-based simulation in each transverse lattice with different colors for both (a) six-spin and (b) seven-spin systems.

Concretely, we visually illustrate the comparison for the evolution of a typical state from 2-spin to 7-spin in Fig. 2. These states are evolved from arbitrarily chosen initial states from the test set. Saliently, the evolution predicted by the SRU-based model accurately reproduces that from exact computations at the beginning 100 time steps. To confirm that this agreement is not just due to a fortuitous choice of component, we look at the evolution across all components of the same states in Fig. 3.

To further quantify the agreement of SRU and ED wavefunction evolutions, we compute the relative entropy (Kullback–Leibler divergence)[32] of their distributions over 1000 test wavefunctions sequences. For discrete probability distributions P and Q, the relative entropy is defined as

$$D_{KL}(P||Q) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right). \tag{4}$$

Given ED-computed wavefunction coefficient vectors M^{ED} and SRU-predicted coefficient vectors M^{SRU}, at time t and basis vector x, the P and Q variables are taken as

$$P_{n,t,x} = \frac{|M^{\mathrm{ED}}_{n,t,x}|}{\sum_{x=1}^{2^N} |M^{\mathrm{ED}}_{n,t,x}|}, \tag{5}$$

$$Q_{n,t,x} = \frac{|M^{\mathrm{SRU}}_{n,t,x}|}{\sum_{x=1}^{2^N} |M^{\mathrm{SRU}}_{n,t,x}|}, \tag{6}$$

where n labels the test sequence. Hence the mean relative entropy (MRE) at each time step t is

$$D_{KL}(P||Q)(t) = \frac{1}{1000} \sum_{n=1}^{1000} \sum_{x=1}^{2^N} P_{n,t,x} \log \frac{P_{n,t,x}}{Q_{n,t,x}}, \tag{7}$$

and measures the amount of information lost when the distribution Q from SRU predictions is used to represent the distribution P from ED results. The smaller the value of $D_{KL}(P||Q)(t)$, the more accurate their agreement is.

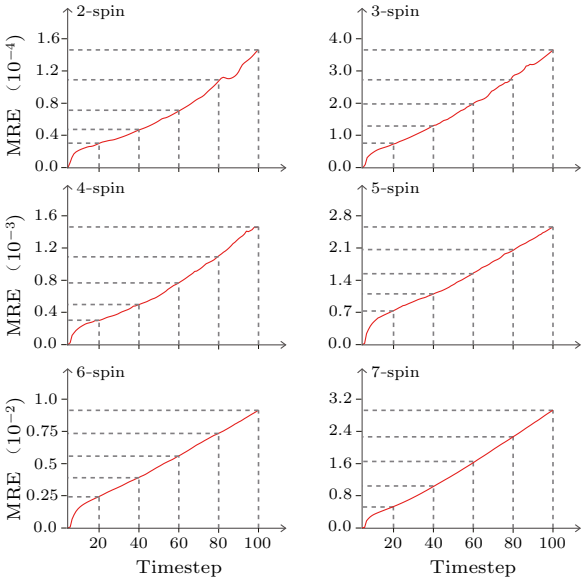

Fig. 4. MRE of generating long sequences in different systems. The MRE increases linearly with time steps.

In Fig. 4, we show the MRE varies significantly across time steps during prediction of test set. In

all six systems, the order of relative entropy is always within 0.03. Evidently, with the increase of time steps, the relative entropy generally shows an upward trend and increases linearly (see the Supplementary Material), which is caused by the accumulation of errors during the conditional generation without any external guidance, though already suppressed dropout layers.

Owing to the unified encoding and parallelism, our SRU-based NN is becoming increasingly more superior over the ED method in terms of efficiency, as the number of spins and batch size increase. Table 1 summarizes the results. When the number of spins gets larger, e.g. 6 and 7, the advantage of our SRU-based framework on inference speed becomes more and more obvious. This is attributed to its constant computational complexity. In addition, when we enlarge the batch size to 256 for 7 spins, our model demonstrates a speed of 130 times faster than the ED-based method.

Table 1. Comparison of time consumption (seconds) between the ED and SRU-based methods. Three independent runs are performed to generate sequences of different batch sizes, and the average time consumption is reported. BS denotes the batch size, and the bold black font indicates that our SRU-based method is superior to ED with increasing spin number and batch size.

	BS = 1		BS = 64		BS = 128		BS = 256	
System	ED	Ours	ED	Ours	ED	Ours	ED	Ours
2-spin	0.015	0.425	1.1	**0.74**	2.3	**0.83**	4.6	**0.69**
3-spin	0.035	0.425	2.2	**0.74**	4.4	**0.69**	8.7	**0.71**
4-spin	0.059	0.425	3.8	**0.77**	7.6	**0.66**	15.1	**0.72**
5-spin	0.271	0.425	17.5	**0.82**	34.9	**0.76**	69.7	**0.79**
6-spin	0.556	**0.425**	35.2	**0.71**	70.5	**0.66**	141.4	**0.98**
7-spin	1.15	**0.425**	73.1	**0.80**	146.3	**0.91**	292.5	**2.18**

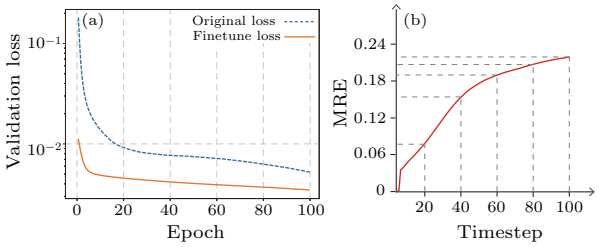

Fig. 5. (a) The validation loss for predicting the dynamics of the 8-spin system by finetuning and training from scratch, respectively. (b) The MRE of the 8-spin system.

After obtaining the base model trained with data sets of 2 to 7 spins by 300 epochs, we may continue to finetune it with the data set of the 8-spin system. We also train it from scratch for comparison. The results in Fig. 5(a) shows that the validation loss by finetuning the base model is much lower than training from scratch, demonstrating that our NN has already learned transferable features from smaller systems. The MRE of the 8-spin system is shown in Fig. 5(b).

In summary, we have successfully applied a transferable NN approach based on SRU networks to approximate the state evolution of dynamic quantum many-body systems with high accuracy and superior scalability. Our work encourages future applications of advanced ML methods in quantum many-body dynamics in a Hamiltonian-agnostic manner. One possibility is to predict the behavior of large and inhomogeneous systems lacking training data by just learning from a smaller-sized system.[33,34] Applications of these advancements in ML to quantum many-body problems are left to future work.

Xiao Zhang thanks Yingfei Gu, Meng Cheng, Yi Zhang for discussions.

References

[1] Bengio Y, Courville A and Vincent P 2013 *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 1798
[2] Krizhevsky A, Sutskever I and Hinton G E 2017 *Commun. ACM* **60** 84
[3] Gatys L A, Ecker A S and Bethge M 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* p 2414
[4] Van Den Oord A, Dieleman S and Zen H 2016 *CoRR abs/1609.03499*
[5] Sun Z, Liu J and Zhang Z 2018 *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* p 1864
[6] Wu Y, Schuster M and Chen Z 2016 arXiv:1609.08144 [cs.CL]
[7] Voulodimos A, Doulamis N, Doulamis A and Protopapadakis E 2018 *Comput. Intell. Neurosci.* **2018**
[8] Deng L, Hinton G and Kingsbury B 2013 *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* p 8599
[9] Van Nieuwenburg E P, Liu Y H and Huber S D 2017 *Nat. Phys.* **13** 435
[10] Cai Z and Liu J 2018 *Phys. Rev. B* **97** 035116
[11] Torlai G and Melko R G 2016 *Phys. Rev. B* **94** 165134
[12] Torlai G and Melko R G 2017 *Phys. Rev. Lett.* **119** 030501
[13] Deng D L, Li X and Sarma S D 2017 *Phys. Rev. X* **7** 021021
[14] Gray J, Banchi L, Bayat A and Bose S 2018 *Phys. Rev. Lett.* **121** 150503
[15] Ch'ng K, Carrasquilla J, Melko R G and Khatami E 2017 *Phys. Rev. X* **7** 031038
[16] Broecker P, Carrasquilla J, Melko R G and Trebst S 2017 *Sci. Rep.* **7** 8823
[17] Amin M H, Andriyash E, Rolfe J, Kulchytskyy B and Melko R 2018 *Phys. Rev. X* **8** 021050
[18] Zhang Y and Kim E A 2017 *Phys. Rev. Lett.* **118** 216401
[19] Carleo G and Troyer M 2017 *Science* **355** 602
[20] Pan S J and Yang Q 2009 *IEEE Trans. Knowl. Data Eng.* **22** 1345
[21] Torrey L and Shavlik J 2010 *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* (IGI Global) p 242
[22] Gao X and Duan L M 2017 *Nat. Commun.* **8** 662
[23] Sharir O, Levine Y, Wies N, Carleo G and Shashua A 2019 arXiv:1902.04057 [cond-mat.dis-nn]
[24] Carleo G, Nomura Y and Imada M 2018 *Nat. Commun.* **9** 5322
[25] Lei T, Zhang Y, Wang S I, Dai H and Artzi Y 2018 *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* p 4470
[26] Graves A and Jaitly N 2014 *International Conference on Machine Learning* p 1764
[27] Pan S J, Tsang I W, Kwok J T and Yang Q 2010 *IEEE Trans. Neural Netw.* **22** 199
[28] Glorot X, Bordes A and Bengio Y 2011 *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* p 513
[29] Sutskever I, Vinyals O and Le Q V 2014 *Advances in Neural Information Processing Systems* p 3104
[30] Williams R J and Zipser D 1989 *Neural Comput.* **1** 270
[31] Kingma D P and Ba J 2014 arXiv:1412.6980 [cs.LG]
[32] Vedral V 2002 *Rev. Mod. Phys.* **74** 197
[33] Hinton G, Vinyals O and Dean J 2015 arXiv:1503.02531 [stat.ML]
[34] Kingma D P, Salimans T and Jozefowicz R 2016 *Advances in Neural Information Processing Systems* p 4743