COMPUTATIONAL CHEMISTRY

Accelerating quantum molecular simulations

Variational Monte Carlo is one of the most accurate methods to solve the many-electron Schrödinger equation, but suffers from high computational cost. A recent study uses a weight-sharing technique to accelerate the neural network-based variational Monte Carlo method, allowing accurate and effective simulations of molecules.

Huan Tran

The wave function encodes an essential amount of information for calculating the properties of a quantum system[1], and thus, accurately solving the Schrödinger equation for the wave function is a critical problem in multiple disciplines, for example, physics, quantum chemistry, and materials science. Except for simple systems such as the hydrogen atom, this equation cannot be solved analytically. When the complexity in the system increases, the numerical solution also becomes out of reach, mainly due to quantum many-body interactions, for instance nuclei–nuclei, nuclei–electron, and electron–electron interactions. The Born–Oppenheimer approximation[1] separates the nuclei and electron's degrees of freedom, thus reducing the complexity in nuclei–electron interactions, but the many-electron problem remains cumbersome to be solved effectively. Density functional theory, which maps the many-electron system onto an effective non-interacting system of 'electrons', offers a solution that better balances the accuracy and computational cost, hence being widely used to study many physical, chemical, and biological systems. When higher accuracy is needed, other computationally more expensive methods, as schematically shown in Fig. 1a, must be used. Among these methods, a popular option is variational Monte Carlo (VMC)[2]. A major advantage of VMC over other methods, for example, Hatree–Fock, is that it can be used with arbitrary forms of the wave function. During the last several years, artificial intelligence has inspired many efforts to improve VMC[3–8], and the work by Scherbela et al.[9] published in this issue of *Nature Computational Science*, is an example of such a recent achievement.

Given a wave function ansatz, VMC optimizes it to minimize the energy of its described quantum system, and drives it towards the ground state. A good ansatz must be as close as possible to the unknown exact solution while satisfying the anti-symmetry requirement. Choosing a good ansatz is therefore a challenge. The traditional choice is the Slater–Jastrow ansatz[2], obtained by multiplying the Slater determinant with a (Jastrow) factor. The Slater–Jastrow ansatz, which sometimes undergoes a backflow transformation, determines the accuracy and the computational cost of VMC with respect to other methods, as shown in Fig. 1a. Recently, deep neural networks (DNN) have been used as an ansatz for VMC[3–8]. The main idea is to represent the wave function by a DNN that accepts the atomic configurations as inputs, and outputs the wave function amplitude and phase factor[3]. Then, the DNN is trained along with the VMC optimizations. In a DNN, weights are the main parameters that specify the strength of the connections among the artificial neurons, visualized as arrows in Fig. 1b. Therefore, the weights of a DNN determine the wave function that is represented throughout the DNN-based VMC approach. Methods of this family yield superior accuracy and computational cost scaling that can be comparable with or better than the highly accurate coupled-cluster singles, doubles and perturbative triples, known as CCSD(T)[10]. However, as depicted in the inset of Fig. 1a, because the constant prefactor in the scaling is large due to hundreds of thousands of parameters of the DNN, they remain slower than CCSD(T) for small molecular systems. Even the

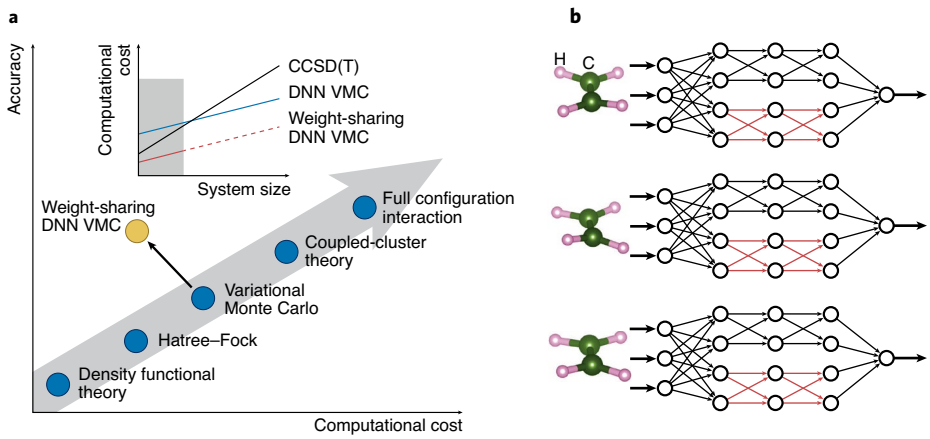

Fig. 1 | Weight-sharing technique to accelerate DNN-based VMC. a, A schematic view of the available methods for solving the many-electron Schrödinger equation. In the inset of **a**, the computational cost scalings (given in the log scale) of CCSD(T), DNN-based VMC, and weight-sharing DNN-based VMC methods, are given. DNN-based VMC (blue line) scales better (or has smaller slope) than CCSD(T) (black line) but because of its prefactor (the intercept with the *y* axis), it is slower than CCSD(T) in the small system size region (shaded region). The weight-sharing DNN-based VMC technique (red line) brings the blue line down, accelerating the ab initio simulations of molecules; the red dashed line represents large molecular sizes beyond this study. **b**, A depiction of three DNN wave functions for three geometries in weight-sharing DNN-based VMC. In each DNN, circles represent artificial neurons, arranged in layers. Following the arrow, input information, which comes from the left, is processed in the neurons, and outputs are passed to the neurons in the next layers. Each arrow is associated with a weight factor, specifying how important this line of transferred information is among the inputs of the neuron accepting it. Red arrows schematically indicate the weights that are kept the same across the geometries. In **b**, three geometries of the ethene molecule — each having six atoms (2 carbon atoms and 4 hydrogen atoms) — are used to visualize the inputs of these DNNs.

computation of medium-size molecules needs days or weeks of calculations on highly optimized modern hardware, thus the DNN-based VMC approach is generally computationally too expensive for sizes of practical interest.

The work by Scherbela et al.[9] proposes a solution for this problem that can substantially reduce the prefactor and accelerate the DNN-based VMC. It was motivated by realistic scenarios when the many-electron Schrödinger equation should be solved for a large number of nuclear geometries. Examples of these situations include the calculations of chemical reaction pathways, which involves multiple continuously deformed nuclear geometries, and the generation of big datasets for artificial intelligence in materials informatics[11]. The main idea of Scherbela et al.[9] is that by constructing and optimizing the wave function ansatzes for these geometries concurrently using a technique called weight sharing, multiple advantages can be obtained, as depicted in Fig. 1a.

Each DNN that is used as a wave function in VMC[3–8] could have hundreds of thousands of parameters, that is, the weights, that have to be optimized during the training process. Within the weight-sharing DNN-based VMC, which resembles deep transfer learning, 75% to 95% of these parameters are shared — that is — they are kept as the same among different nuclear geometries in each optimization step. When the parameters of one geometry are optimized, the shared set of parameters will be updated across other geometries. The first advantage of the weight-sharing DNN-based VMC is that the number of independent parameters is much smaller than constructing and optimizing the wave function ansatzes independently. In the study, the authors demonstrated a substantial acceleration of about one order of magnitude compared to independent optimization without weight sharing. Second, weight sharing plays a role of regularization, forcing the current wave functions in each iteration to be comparably good in approximating the real wave functions of the geometries considered. In other words, this regularization makes VMC faster and more stable than independently using regular DNN-based methods for multiple geometries.

Data size and quality are enormously important for applying artificial intelligence methods in disciplines like quantum chemistry and materials informatics. The technique demonstrated by Scherbela et al.[9] can greatly accelerate the DNN-based VMC method for the simulation of small molecules (up to ten atoms as demonstrated in the study), and thus, can further enable the fast generation of high-qualify molecular datasets. Within this context, DeepErwin, the open-source package that implements weight-sharing DNN-based VMC and is available in GitHub, could be a useful tool for the community. Given the momentum gained recently in this field[3–9], we can anticipate more developments in the future. A promising future avenue, which remains very challenging, is to extend the weight-sharing technique to effectively calculate the wave function and other properties of large molecules, which have sizes corresponding to the region of the dashed red line in the inset of Fig. 1a. ❐

Huan Tran ✉

School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA.
✉*e-mail: huan.tran@mse.gatech.edu*

Published online: 19 May 2022
https://doi.org/10.1038/s43588-022-00237-w

References

1. Griffiths, D. J. & Schroeter, D. F. *Introduction to Quantum Mechanics* (Cambridge Univ. Press, 2018).
2. Foulkes, W., Mitas, L., Needs, R. & Rajagopal, G. *Rev. Mod. Phys.* **73**, 33 (2001).
3. Carleo, G. & Troyer, M. *Science* **355**, 602–606 (2017).
4. Kessler, J., Calcavecchia, F. & Kühne, T. D. *Adv. Theory Simul.* **4**, 2000269 (2021).
5. Han, J., Zhang, L. & Weinan, E. *J. Comput. Phys.* **399**, 108929 (2019).
6. Pfau, D., Spencer, J. S., Matthews, A. G. & Foulkes, W. M. C. *Phys. Rev. Res.* **2**, 033429 (2020).
7. Gao, N. & Günnemann, S. Preprint at https://arxiv.org/abs/2110.05064 (2021).
8. Hermann, J., Schätzle, Z. & Noé, F. *Nat. Chem.* **12**, 891–897 (2020).
9. Scherbela, M., Reisenhofer, R., Gerard, L., Marquetand, P. & Grohs, P. G. *Nat. Comput. Sci.* https://doi.org/10.1038/s43588-022-00228-x (2022).
10. Bartlett, R. J. & Musiał, M. *Rev. Mod. Phys.* **79**, 291 (2007).
11. Huang, B. & von Lilienfeld, O. A. *Chem. Rev.* **121**, 10001–10036 (2021).

Competing interests

The author declares no competing interests.