

# Electronic Supplementary Information to: Force and stress calculation with neural network wavefunction for solids

Yubing Qian,<sup>a,b</sup> Xiang Li,<sup>b</sup> and Ji Chen<sup>\*a,c</sup>

## Contents

<b>1 More details for force estimators</b>	<b>1</b>
1.1 The infinite variance problem . . . . .	1
1.2 Assaraf–Caffarel estimator . . . . .	1
1.2.1 Minimal form . . . . .	1
1.2.2 More complicated forms . . . . .	2
1.3 Space warp coordinate transformation estimator . . . . .	2
1.3.1 Fast warp estimator . . . . .	2
1.4 Antithetic variates technique . . . . .	2
<b>2 Implementation</b>	<b>3</b>
2.1 Neural network hyperparameters . . . . .	3
2.2 Settings for evaluation time . . . . .	3

## 1 More details for force estimators

### 1.1 The infinite variance problem

If  $\psi_T$  is assumed to be in the vicinity of the exact ground state,  $\mathbf{F}_I$  can be calculated with Hellmann–Feynman theorem:

$$\mathbf{F}_I = -\frac{\partial}{\partial \mathbf{R}_I} \frac{\langle \psi_T | \hat{H}_S | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \frac{\langle \psi_T | -\partial_{\mathbf{R}_I} \hat{H}_S | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle}. \quad (1)$$

Considering that only the Coulomb potential part of local energy  $E_L^P$  is relevant, the bare estimator  $\mathbf{F}_I^{\text{Bare}}$  takes the following form in the  $\mathbf{X}$ -representation:

$$\mathbf{F}_I^{\text{Bare}} = \langle -\partial_{\mathbf{R}_I} E_L^P(\mathbf{X}) \rangle, \quad (2)$$

In molecular systems, the  $r^{-2}$  Coulomb force results in the infinite variance problem during the Monte Carlo sampling process<sup>1,2</sup>. In solid systems, the summation of Coulomb potential is carried out with the Ewald summation technique for a better convergence<sup>3,4</sup>. The summation of the Coulomb potential is split into a short-range part and a long-range part. The short-range part includes the contribution of original point charge surrounded by a Gaussian charge density distribution of equal charge and opposite sign, and the corresponding potential is  $\text{erfc}(ar)/r$  instead of  $1/r$ , where  $a$  is the width of the Gaussian distribution. However, when  $r$  is small,  $\text{erfc}(ar)/r$  is still proportional to  $1/r$ , and the infinite variance problem is still relevant.

<sup>a</sup> School of Physics, Peking University, Beijing 100871, People’s Republic of China.

<sup>b</sup> ByteDance Research, Zhonghang Plaza, No. 43, North 3rd Ring West Road, Haidian District, Beijing, People’s Republic of China.

<sup>c</sup> Interdisciplinary Institute of Light-Element Quantum Materials, Frontiers Science Center for Nano-Optoelectronics, Peking University, Beijing 100871, People’s Republic of China, E-mail: ji.chen@pku.edu.cn

### 1.2 Assaraf–Caffarel estimator

Assaraf–Caffarel estimator is introduced in the main text, and it reads

$$\mathbf{F}_I^{\text{AC}} = \mathbf{F}_I^{\text{Bare}} + \left\langle -\frac{(\hat{H}_S - E_L^*) \bar{\psi}_I}{\psi_T} - 2\text{Re}(E_L^* \partial_{\mathbf{R}_I} \log \psi_T) + 2E_{\text{VMC}} \text{Re}(\partial_{\mathbf{R}_I} \log \psi_T) \right\rangle, \quad (3)$$

where  $\bar{\psi}_I$  is an approximation to  $\partial_{\mathbf{R}_I} \psi_0$ .

#### 1.2.1 Minimal form

A simple approximation of  $\bar{\psi}_I$  was proposed by Assaraf and Caffarel for molecular systems:

$$\bar{\psi}_I^{\text{mol}} = \mathbf{Q}_I^{\text{mol}} \psi_T, \quad \mathbf{Q}_I^{\text{mol}} = Z_I \sum_i \frac{\mathbf{r}_i - \mathbf{R}_I}{|\mathbf{r}_i - \mathbf{R}_I|}. \quad (4)$$

By substituting  $\bar{\psi}_I^{\text{mol}}$  into the second term on the right-hand side of Eq. (3), we found that

$$-\frac{(\hat{H}_S - E_L^*) \bar{\psi}_I}{\psi_T} = \frac{1}{2} \sum_i \nabla_i^2 \mathbf{Q}_I^{\text{mol}} + \sum_i \left( \nabla_i \mathbf{Q}_I^{\text{mol}} \right) \left( \nabla_i \psi_T \right) \quad (5)$$

$$= -Z_I \sum_i \frac{\mathbf{r}_i - \mathbf{R}_I}{|\mathbf{r}_i - \mathbf{R}_I|^3} + \sum_i \left( \nabla_i \mathbf{Q}_I^{\text{mol}} \right) \left( \nabla_i \psi_T \right), \quad (6)$$

where the  $r^{-2}$  singularity from  $\mathbf{F}_I^{\text{Bare}}$  is exactly cancelled, and thus removing the infinite variance. Here we dub the estimator “min-AC-mol”, which stands for the minimal form of AC estimator for molecular systems.

However, the choice in Eq. (4) is not applicable in periodic systems, because the summation does not converge. A new form  $\bar{\psi}_I$  suitable for solids is needed to cancel the  $\text{erfc}(ar)/r^2$  divergence from the short-range part of Ewald summation. A new “min-AC-solid” estimator is introduced in the main text with  $\bar{\psi}_I^{\text{solid}} = \mathbf{Q}_I^{\text{solid}} \psi_T$ , and gradient of the  $x$  component of  $\mathbf{Q}_I$  satisfies

$$\nabla_{Q_{Ix}}^{\text{solid}} = -\sum_i \sum_{\mathbf{L}} Z_I \left( \frac{\text{erfc}(ar_{i\mathbf{L}}) x_{i\mathbf{L}}^2}{r_{i\mathbf{L}}^3} - \frac{\text{erfc}(ar_{i\mathbf{L}})}{r_{i\mathbf{L}}} - \frac{a}{\sqrt{\pi}} \text{Ei}(-a^2 r_{i\mathbf{L}}^2), \right. \\ \left. \frac{\text{erfc}(ar_{i\mathbf{L}}) x_{i\mathbf{L}} y_{i\mathbf{L}}}{r_{i\mathbf{L}}^3}, \frac{\text{erfc}(ar_{i\mathbf{L}}) x_{i\mathbf{L}} z_{i\mathbf{L}}}{r_{i\mathbf{L}}^3} \right), \quad (7)$$

where  $r_{i\mathbf{L}} = |\mathbf{r}_i - \mathbf{R}_I - \mathbf{L}|$ , and  $x_{i\mathbf{L}}, y_{i\mathbf{L}}, z_{i\mathbf{L}}$  are its components. The parameter  $a$  and range of  $\mathbf{L}$  are the same as those in the short-range Ewald summation. And  $\text{Ei}(-a^2 r^2)$  is the exponential integral function<sup>6</sup>, which can be calculated efficiently in an approximate way<sup>7</sup>:

$$-\text{Ei}(-x) = E_1(x) = \left( A^{-7.7} + B \right)^{-0.13} \quad (8)$$

where

$$A = \ln \left[ \left( \frac{0.56146}{x} + 0.65 \right) (1+x) \right] \quad (9)$$

$$B = x^4 e^{7.7x} (2+x)^{3.7} \quad (10)$$

With the form of  $\nabla \mathbf{Q}_I^{\text{solid}}$  given, it is straightforward to verify that  $\frac{1}{2} \nabla^2 \mathbf{Q}_I^{\text{solid}}$  cancels the  $\text{erfc}(ar)/r^2$  divergence and  $\nabla \times \nabla \mathbf{Q}_I^{\text{solid}}$  is zero, making sure that Eq. (7) is a valid gradient. Although it must exist, the explicit form of  $\mathbf{Q}_I^{\text{solid}}$  is not needed, as shown in Eq. (5).

### 1.2.2 More complicated forms

Another natural choice is  $\tilde{\psi}_I = \partial_{\mathbf{R}_I} \psi_T$ . Since we only focus on the real part, it is essentially the same as:

$$\mathbf{F}_I = - \frac{\partial}{\partial \mathbf{R}_I} \frac{\langle \psi_T | E_L | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \frac{\langle \psi_T | \mathbf{F}_I^{\text{no-SWCT}} | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle}, \quad (11)$$

$$\mathbf{F}_I^{\text{no-SWCT}} = \langle -\partial_{\mathbf{R}_I} E_L + 2(E_{\text{VMC}} - E_L) \text{Re}(\partial_{\mathbf{R}_I} \log \psi_T) \rangle, \quad (12)$$

### 1.3 Space warp coordinate transformation estimator

Space warp coordinate transformation (SWCT) is a variance reduction technique to further improve the no-SWCT estimator<sup>8-12</sup>. To demonstrate the SWCT method more clearly, let us first consider a finite displacement case, where the  $I$ -th nucleus is displaced by  $\Delta \mathbf{R}_I$ . By using correlated sampling, the energy of the original geometry ( $E_0$ ) and the displaced geometry ( $E_1$ ) are evaluated using the same Monte Carlo samples from the original geometry, and the variance of  $E_1 - E_0$  goes to zero linearly as  $\Delta \mathbf{R}_I \rightarrow 0^9$ . This is the picture behind the no-SWCT estimator. However, after the displacement, the ground state wavefunction and electron probability distribution will change accordingly, and the core electrons will closely follow the movement of the nucleus. To avoid bad samples and deterioration of accuracy, a coordinate transformation is applied instead of directly using original Monte Carlo samples:

$$\tilde{\mathbf{r}}_i = \mathbf{r}_i + \Delta \mathbf{R}_I \omega_I(\mathbf{r}_i), \quad \omega_I(\mathbf{r}) = \frac{\sum_{\mathbf{L}_S} f(|\mathbf{r} - \mathbf{R}_I - \mathbf{L}_S|)}{\sum_{\mathbf{L}_S} \sum_I f(|\mathbf{r} - \mathbf{R}_I - \mathbf{L}_S|)}, \quad (13)$$

With  $\Delta \mathbf{R}_I \rightarrow 0$ , the Jacobian of the transformation still has a non-zero contribution to the force, and the SWCT estimator can be formulated as<sup>10-12</sup>

$$\mathbf{F}_I^{\text{SWCT}} = \left\langle -\partial_{\mathbf{R}_I} E_L - \sum_i \omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} E_L + 2(E_{\text{VMC}} - E_L) \text{Re} P_{\text{SWCT}} \right\rangle, \quad (14)$$

where

$$P_{\text{SWCT}} = \partial_{\mathbf{R}_I} \log \psi_T + \sum_i \left( \omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} \log \psi_T + \frac{1}{2} \partial_{\mathbf{r}_i} \omega_I(\mathbf{r}_i) \right). \quad (15)$$

The effectiveness of SWCT can be easily seen in an isolated atom case where  $\omega_I = 1$ , and the terms cancel exactly due to translational symmetry, leading to zero variance. Besides, the relation  $1 - \omega_I \ll 1$  is guaranteed near the nuclei, making the SWCT estimator better at variance reduction.

### 1.3.1 Fast warp estimator

The formula of fast-warp estimator is introduced in the main text, and here we elaborate the derivation and explanation. The most expensive and difficult part of the SWCT estimator is the derivatives of the kinetic energy. The aim of the fast-warp estimator is to remove this part, enabling faster evaluation and better variance reduction.

We start from the first term in Eq. (14):

$$\langle \partial_{\mathbf{R}_I} E_L \rangle = \langle \partial_{\mathbf{R}_I} \frac{\hat{H}_S \psi_T}{\psi_T} \rangle = \langle \partial_{\mathbf{R}_I} E_L^{\text{P}} \rangle + \left\langle \frac{(\hat{H}_S - E_L) \partial_{\mathbf{R}_I} \psi_T}{\psi_T} \right\rangle \quad (16a)$$

$$= \langle \partial_{\mathbf{R}_I} E_L^{\text{P}} \rangle + \left\langle \frac{(E_L^* - E_L) \partial_{\mathbf{R}_I} \psi_T}{\psi_T} \right\rangle. \quad (16b)$$

Eq. (16b) holds because the Hamiltonian operator is Hermitian and can act on the bra state  $\langle \psi_T |$ . In the no-SWCT and SWCT estimator, Eq. (16b) is not used because Eq. (16a) can reduce the variance in  $\partial_{\mathbf{R}_I} \frac{\hat{H}_S \psi_T}{\psi_T}$ . However, this is not needed in the fast-warp estimator and Eq. (16b) is used instead.

We then shift our focus to the second term of Eq. (14):

$$\left\langle \sum_i \omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} E_L \right\rangle = \left\langle \sum_i \omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} E_L^{\text{P}} \right\rangle + \left\langle \omega_I(\mathbf{r}_i) \frac{(\hat{H} - E_L) \partial_{\mathbf{r}_i} \psi_T}{\psi_T} \right\rangle. \quad (17)$$

It is impossible to do the same trick directly as Eq. (16b). However, we can move  $\omega$  to the right-hand side of the Hamiltonian operator:

$$\omega_I(\mathbf{r}_i) \frac{(\hat{H} - E_L) \partial_{\mathbf{r}_i} \psi_T}{\psi_T} = \frac{(E_L^* - E_L) [\omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} \psi_T]}{\psi_T} + \frac{1}{2} \frac{[\nabla_i^2 \omega_I(\mathbf{r}_i)] \partial_{\mathbf{r}_i} \psi_T}{\psi_T} + \frac{(\nabla_i \omega_I(\mathbf{r}_i)) (\nabla_i \partial_{\mathbf{r}_i} \psi_T)}{\psi_T}, \quad (18)$$

where the last two terms of Eq. (18) are simplified based on the fact that  $\nabla_j \omega_I(\mathbf{r}_i) = \delta_{ij} \nabla_i \omega_I(\mathbf{r}_i)$ .

After some organization, we obtain the expression of the fast-warp estimator:

$$\mathbf{F}_I^{\text{fast-warp}} = \left\langle -\partial_{\mathbf{R}_I} E_L^{\text{P}} - \sum_i \omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} E_L^{\text{P}} - \frac{1}{2} \frac{[\nabla_i^2 \omega_I(\mathbf{r}_i)] \partial_{\mathbf{r}_i} \psi_T}{\psi_T} - \frac{(\nabla_i \omega_I(\mathbf{r}_i)) (\nabla_i \partial_{\mathbf{r}_i} \psi_T)}{\psi_T} + 2E_{\text{VMC}} \text{Re} P_{\text{SWCT}} - 2\text{Re}(E_L^* P_{\text{SWCT}}) \right\rangle. \quad (19)$$

### 1.4 Antithetic variates technique

Antithetic variates<sup>13,14</sup> can be used in combination with other estimators. The basic idea is to reduce the variance of the expectation by using pairs of negatively correlated samples. In the context of force calculation in VMC, given a Monte Carlo sample with a set of electron coordinates  $\mathbf{X} = \{\mathbf{r}_i\}$ , our aim is to find  $\mathbf{X}' = \{\mathbf{r}'_i\}$  which contributes approximately the same magnitude of atomic force but on the opposite direction. And then, the force contribution is summed over with an additional weight

$w = |\psi(\mathbf{X}')|^2/|\psi(\mathbf{X})|^2$  to keep the importance sampling correct:

$$\mathbf{F}_I^{\text{antithetic}} = \frac{1}{2}(\mathbf{F}_I(\mathbf{X}) + w\mathbf{F}_I(\mathbf{X}')). \quad (20)$$

If the antithetic variables  $\mathbf{X}'$  further satisfies  $w \approx 1$ , then the statistical fluctuation can be greatly reduced.

The process for finding  $\mathbf{X}'$  is given as follows<sup>15</sup>. For each electron coordinate  $\mathbf{r}_i$ , we first find its closest nucleus  $\mathbf{R}_I$ . If the distance  $|\mathbf{r}_i - \mathbf{R}_I|$  is less than the core electron cutoff radius  $r_{\text{core}}$ , we choose  $\mathbf{r}'_i = 2\mathbf{R}_I - \mathbf{r}_i$  which is the mirror of  $\mathbf{r}_i$  with respect to  $\mathbf{R}_I$ . Otherwise, we simply let  $\mathbf{r}'_i = \mathbf{r}_i$ . The purpose of such differential treatment is to maximize the negative correlation between  $\mathbf{X}$  and  $\mathbf{X}'$  since core electrons contributes the most, while avoiding bad samples and maintaining  $w \approx 1$ .

## 2 Implementation

The estimators for solid systems are implemented within the DeepSolid software<sup>16</sup>. The algorithm used in this work is summarized as follows.

Markov chain Monte Carlo (MCMC) is used for sampling, and the value of the estimators and local energies are evaluated and stored along the Markov chain. To minimize the correlation with an affordable cost, 5000 steps of MCMC warm-up is performed before the first evaluation, and 50 steps of MCMC is performed between evaluation steps. At the end, we average all the local energies to get  $E_{\text{VMC}}$ , and obtain a list of force results from each evaluation step.

The estimators are implemented with Python JAX<sup>17</sup> code, and the automatic differentiation feature from JAX is used for calculating derivatives with respect to all nuclei coordinates in all directions in one go. The Python functions for local energy and Ewald summation require some modifications to obtain well-defined gradients of local energies. To get robust statistical results, interquartile range (IQR) method is used to remove the outliers. IQR is the difference between the first quartile ( $Q_1$ ) and the third quartile ( $Q_3$ ). All data less than  $Q_1 - 3\text{IQR}$  or greater than  $Q_3 + 3\text{IQR}$  are clipped to the boundaries. Also, all standard errors given are based on the clipped data with Monte Carlo autocorrelation considered<sup>18,19</sup>.

### 2.1 Neural network hyperparameters

The default settings for the network training process are listed in Table 1.

Table 1 Default training settings

Name	Value
Type of distance feature	tri
Pretrain	No
Numerical precision	Float32
Batch size	4096
Hidden units per one-electron layer	256
Hidden units per two-electron layer	32
Number of layers	3
Number of determinants	1
Training iterations for lithium hydride	$10^5$
Training iterations for graphene	$2 \times 10^5$
Other settings	Default <sup>16</sup>

### 2.2 Settings for evaluation time

In the main text, we compared the force variance and evaluation time of different estimators. The settings for the evaluation are listed in Table 2, and the hyperparameters for the neural network are the same as those listed in Table 1.

Table 2 Settings for evaluation time

Name	Value
Hardware	4 V100 GPU cards
System	equilibrium graphene
Evaluation steps	200
Split chunks per batch	SWCT: 8, Others: 1

## Notes and references

- 1 B. L. Hammond, W. A. Lester and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*, WORLD SCIENTIFIC, 1994, vol. 1.
- 2 A. N. Badinski, *PhD thesis*, University of Cambridge, 2008.
- 3 P. P. Ewald, *Annalen der Physik*, 1921, **369**, 253–287.
- 4 A. Y. Toukmaji and J. A. Board, *Computer Physics Communications*, 1996, **95**, 73–92.
- 5 R. Assaraf and M. Caffarel, *The Journal of Chemical Physics*, 2003, **119**, 10536–10552.
- 6 *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, ed. M. Abramowitz and I. A. Stegun, Dover Publications, New York, NY, 0009th edn, 1965.
- 7 P. H. Giao, *Groundwater*, 2003, **41**, 387–390.
- 8 C. J. Umrigar, *International Journal of Quantum Chemistry*, 1989, **36**, 217–230.
- 9 C. Filippi and C. J. Umrigar, *Physical Review B*, 2000, **61**, R16291–R16294.
- 10 S. Sorella and L. Capriotti, *The Journal of Chemical Physics*, 2010, **133**, 234111.
- 11 K. Nakano, C. Attaccalite, M. Barborini, L. Capriotti, M. Casula, E. Coccia, M. Dagrada, C. Genovese, Y. Luo, G. Mazzola, A. Zen and S. Sorella, *The Journal of Chemical Physics*, 2020, **152**, 204121.
- 12 K. Nakano, A. Raghav and S. Sorella, *The Journal of Chemical Physics*, 2022, **156**, 034101.
- 13 J. M. Hammersley and K. W. Morton, *Mathematical Proceedings of the Cambridge Philosophical Society*, 1956, **52**, 449–475.
- 14 B. J. T. Morgan, *Wiley StatsRef: Statistics Reference Online*, John Wiley & Sons, Ltd, 2014.
- 15 S. Chiesa, D. M. Ceperley and S. Zhang, *Physical Review Letters*, 2005, **94**, 036404.
- 16 X. Li, *DeepSolid: A library combining solid quantum Monte Carlo and neural network*, Bytedance Inc., 2023, <https://github.com/bytedance/deepsolid>.
- 17 J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne and Q. Zhang, *JAX: composable transformations of Python+NumPy programs*, 2018, <http://github.com/google/jax>.

18 U. Wolff, *Computer Physics Communications*, 2004, **156**, 143–153.

19 D. Hesse, *py-uwerr: Analysis of Monte Carlo Data*, 2022, <https://github.com/dhesse/py-uwerr>.