



北京大学

本科生毕业论文

题目：基于固体神经网络波函数的原
子间受力计算

姓 名：钱昱冰
学 号：1900011338
院 系：物理学院
专 业：物理学
研究方向：物理学
导 师：陈基

二〇二三年五月

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。

摘要

精确的第一性原理计算在物理和材料科学等领域占有重要地位。近些年来，随着神经网络的应用，第一性原理计算得到了快速的发展，尤其是基于变分蒙特卡洛（VMC）的神经网络波函数方法。继分子体系之后，神经网络波函数方法也被用于模拟固体体系，但它们大多数都是关注体系能量和电子密度的计算。在这项研究中，我们在之前的工作上更进一步，在固体体系中实现并测试了若干个在 VMC 中常用的力的估计量，并探究了固体神经网络方法中的原子间受力计算方案。我们的结果展示了该方案的正确性和精度，表明了固体神经网络方法强大能力。我们还讨论了各个估计量的精度和计算成本，从而为未来的应用提供指导。此外，我们也讨论了常用网络输入特征在计算原子间受力时的局限性。

关键词：神经网络；变分蒙特卡洛；周期性体系

Interatomic force calculation with neural network wavefunction for solids

Yubing Qian (Physics)

Directed by Prof. Ji Chen

ABSTRACT

Accurate first principles calculations are of fundamental importance in fields such as physics and materials science. Recently, first-principles calculations have been developed rapidly with the application of neural networks, especially the variational Monte Carlo (VMC) based neural network wavefunction methods. Following molecular systems, neural network wave function methods have also been applied to model solid systems, but most of the works focus on the energies and electron densities. In this study, we take a step further by implementing and testing several force estimators commonly used in traditional VMC in the solid system and exploring the interatomic forces in the the DeepSolid neural network. Our results show that our calculations are correct and accurate, which implies the power of solid neural network. We also discuss the accuracy and computational cost of different estimators, thus providing guidance for future applications. In addition, we also discuss the limitations of the commonly used network input feature in the calculation of interatomic forces.

KEYWORDS: Neural network; Variational Monte Carlo; Periodic systems

目录

第一章 引言	1
第二章 方法和计算细节	3
2.1 DeepSolid 方法	3
2.2 原子间受力的估计量	5
2.2.1 最简单估计量	5
2.2.2 相反采样估计量	5
2.2.3 带修正的最简单和相反采样估计量	6
2.2.4 Assaraf-Caffarel 估计量	7
2.2.5 空间坐标扭曲估计量	7
2.3 实现	8
第三章 结果和讨论	10
3.1 收敛网络下的能量和空间坐标扭曲估计量	10
3.2 各估计量的比较	11
3.2.1 垂直方向	11
3.2.2 平行方向	13
3.3 不同估计量的单步相对开销	15
3.4 不同的网络距离特征	17
第四章 结论	19
参考文献	20

第一章 引言

精确的第一性原理计算一直是材料模拟的重要基础。尽管我们对基础的电磁相互作用和量子力学原理有了很清楚的认识，但是要从基础理论出发，得到宏观材料的一些性质并不是一件简单的事情，因为这中间涉及到的多体薛定谔方程，我们是做不到严格求解的。不过这也不妨碍人们继续开发各种第一性原理计算方法，因为我们可以采取一些近似或者经验性的方法来简化计算的复杂度。近年来随着硬件水平的提高和神经网络算法的进步，基于神经网络的第一性原理计算方法得到了快速的发展，这些方法能够得到很精确的结果^[1-5]。它们中大多数是将神经网络方法和传统的变分蒙特卡罗方法（VMC）相结合，算法基本沿用 VMC 的框架，但是把试探波函数换成表达能力极强的神经网络。这些方法主要关注的是一些原子和分子体系的能量，利用神经网络方法往往可以得到很精确的能量结果。

尽管也有工作利用神经网络 VMC 方法计算电子气体^[6]。但是对实际的周期性固体体系，直到李向等人开发了深度固体神经网络（DeepSolid）^[7-8]才能够用神经网络波函数方法计算。DeepSolid 是基于费米神经网络（FermiNet^[3,9-10]）的改进方案。FermiNet 能够在一些原子和分子体系上得到很精确的基态能量，但是网络本身不能满足周期性边界条件，也就不能直接处理固体体系。李向等人对费米神经网络做了一系列改进，包括使用推广的 Bloch 形式的波函数，使用满足晶格周期性的距离和位矢作为神经网络输入特征，使得神经网络满足周期性边界条件，并且能够较好地收敛。

之前的基于固体神经网络 VMC 工作主要还是关注能量以及电子密度的计算。而对于其他的可观测量，比如原子间受力，直接计算有极大的统计误差，还会有一些数值上的不稳定性，需要一些额外的工作和计算才能得到准确的结果。但在材料研究中却离不开这些可观测量，比如结构弛豫和声子色散的计算都需要力。过去几十年来，在传统 VMC 模拟中，人们提出了不同的方案，但这些方法还没有在基于固体神经网络的 VMC 中应用和测试过。例如，在 DeepSolid（包括 FermiNet）中，尖端条件^[11]没有被精确地写出来；在 DeepSolid 中使用的周期性输入特征也不一定能够直接运用于原子间受力的计算，所以我们不能保证神经网络能得到正确的原子间受力。我们在之前的研究中^[12]，基于 FermiNet 对分子体系中的原子间受力做了很多测试，展示了神经网络方法在计算原子间受力上的优越性，这有望促进神经网络方法在未来材料建模中的应用。但对于材料研究真正关心的固体体系中的原子间受力计算，还没有得到实现和测试。

在本次工作中，我们在 DeepSolid 的基础上实现了适用于复值波函数的原子间受

力计算方案，并在一维和三维的固体体系中测试了它们的效果，展示了原子间受力计算的正确性和精度。我们还对不同的估计量进行了精度和单步相对开销上的比较，希望能够为受力计算提供了指导。此外我们还分析了不同的网络输入特征，指出了默认的输入特征在计算原子间受力时的局限性。

第二章 方法和计算细节

2.1 DeepSolid 方法

研究一个固体体系的电子结构就要求解关于体系内所有电子的定态薛定谔方程:

$$\hat{H}\psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_{\text{elec}}}) = E\psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_{\text{elec}}}) \quad (2.1)$$

其中 $\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)$ ，是第 i 个电子的空间和自旋坐标。同时，电子服从费米-狄拉克统计，所以波函数 ψ 需要满足在 $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n_{\text{elec}}})$ 的置换下是反对称的。

即使我们已经使用了玻恩-奥本海默近似认为原子核是固定不动的，但宏观大的电子数目依然使得求解方程异常困难。所以人们常常使用超胞近似，即在周期性边界条件下，只考虑有限个电子和原子核的情况。这个时候哈密顿量就可以写成

$$\hat{H}_S = \sum_i -\frac{1}{2}\nabla_i^2 + \frac{1}{2} \sum_{\substack{\mathbf{L}_S, i, j \\ i \neq j}} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{L}_S|} - \sum_{\mathbf{L}_S, i, I} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I + \mathbf{L}_S|} + \frac{1}{2} \sum_{\substack{\mathbf{L}_S, I, J \\ I \neq J}} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J + \mathbf{L}_S|} \quad (2.2)$$

其中 \mathbf{r}_i 和 \mathbf{R}_I 分别是第 i 个电子和第 I 个原子核在超胞内的空间坐标， Z_I 是第 I 个原子核的电荷。与开放边界条件不同，在周期性边界条件下，我们需要包含对超胞格矢 $\{\mathbf{L}_S\}$ 的求和，也就是需要考虑电子和原子核与它们的像之间的相互作用，从而更好地近似真实的固体体系。

另一点使得固体体系和分子体系不同的是，体系和哈密顿量的对称性对波函数的对称性有了额外的要求^[13]。具体说来，当所有电子一起移动一个原胞格矢，或者某个电子移动一个超胞格矢时，波函数只能分别改变一个相位。也就是说

$$\psi(\mathbf{r}_1 + \mathbf{L}_p, \mathbf{r}_2 + \mathbf{L}_p, \dots, \mathbf{r}_{n_{\text{elec}}} + \mathbf{L}_p) = e^{i\mathbf{k}_p \cdot \mathbf{L}_p} \psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n_{\text{elec}}}) \quad (2.3)$$

$$\psi(\mathbf{r}_1 + \mathbf{L}_S, \mathbf{r}_2, \dots, \mathbf{r}_{n_{\text{elec}}}) = e^{i\mathbf{k}_S \cdot \mathbf{L}_S} \psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n_{\text{elec}}}) \quad (2.4)$$

为了简洁起见，上面的式子中略去了自旋轨道。正是因为这些类似平面波的平移对称性要求，使得一般来说固体体系的波函数需要是一个复值函数。这又给固体体系的计算和模拟增加了难度。

一般来说，如果波函数 ψ_θ 依赖于一些参数 θ ，并且是一个反对称的函数，那么就可以用基态的能量期望值 E_θ 作为损失函数来优化波函数到基态。

$$E_{v; \theta} = \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \langle E_L \rangle_{|\psi_\theta|^2} \quad (2.5)$$

其中 $\langle \cdot \rangle_{|\psi_\theta|^2}$ 表示在 $|\psi_\theta|^2$ 的分布下计算期望值，而 E_L 是局部能量，定义为

$$E_{L;\theta} = \frac{\hat{H}\psi_\theta}{\psi_\theta} \quad (2.6)$$

上述积分可以用蒙特卡洛方法与重要性采样计算，这就是通常说的 VMC 方法^[14]。

DeepSolid 是一个构造满足上述平移对称性条件的反对称神经网络波函数的方案，从而代替传统的波函数拟设进行 VMC 优化计算。DeepSolid 借鉴了传统的 Bloch 波函数组成行列式的形式，但与 FermiNet 相同，每一个 Bloch 轨道都不再只是一个电子坐标的函数，而是所有电子坐标的函数。

$$\psi_{\mathbf{k}_S, \mathbf{k}_p}(\mathbf{r}) = \begin{vmatrix} e^{i\mathbf{k}_1 \cdot \mathbf{r}_1} u_{\mathbf{k}_1}(\mathbf{r}_1; \{\mathbf{r}_{/1}\}) & \cdots & e^{i\mathbf{k}_n \cdot \mathbf{r}_1} u_{\mathbf{k}_n}(\mathbf{r}_1; \{\mathbf{r}_{/1}\}) \\ \vdots & \ddots & \vdots \\ e^{i\mathbf{k}_1 \cdot \mathbf{r}_n} u_{\mathbf{k}_1}(\mathbf{r}_n; \{\mathbf{r}_{/n}\}) & \cdots & e^{i\mathbf{k}_n \cdot \mathbf{r}_n} u_{\mathbf{k}_n}(\mathbf{r}_n; \{\mathbf{r}_{/n}\}) \end{vmatrix} \quad (2.7)$$

其中 $\{\mathbf{r}_{/i}\}$ 表示除电子 i 以外其他所有原子的坐标，并且我们还要求函数 $u_{\mathbf{k}}$ 不依赖于 $\{\mathbf{r}_{/i}\}$ 中电子坐标的具体顺序，也即函数 $u_{\mathbf{k}}$ 对于 i 以外任意电子的置换都保持不变。这些轨道 $u_{\mathbf{k}}$ 都是由深度神经网络计算得到的，这也使得 DeepSolid 具有和 FermiNet 类似的强大表达能力。为了使得训练更加容易，并且满足函数的连续性，这些电子坐标并不是直接作为神经网络的输入，而是需要先构造输入距离特征 $d(\mathbf{r})$

$$d(\mathbf{r}) = \frac{\sqrt{\text{tr}(\mathbf{A}\mathbf{M}\mathbf{A}^\top)}}{2\pi}, \quad \mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) \quad (2.8)$$

$$\mathbf{M}_{ij} = f^2(\omega_i)\delta_{ij} + g(\omega_i)g(\omega_j)(1 - \delta_{ij}), \quad \omega_i = \mathbf{r} \cdot \mathbf{b}_i$$

其中 \mathbf{a}_i 和 \mathbf{b}_i 分别为格矢和倒格矢，而 $f(\omega)$ 和 $g(\omega)$ 可以有很多种取法，其中一种称为 NU 型^[7,15]

$$\begin{aligned} f_{\text{NU}}(\omega) &= |\omega| \left(1 - \frac{|\omega/\pi|^3}{4} \right) \\ g_{\text{NU}}(\omega) &= \omega \left(1 - \frac{3}{2}|\omega/\pi| + \frac{1}{2}|\omega/\pi|^2 \right) \end{aligned} \quad (2.9)$$

另一种称为 SIN 型^[6]

$$\begin{aligned} f_{\text{SIN}}^2(\omega) &= 2 - 2\cos(\omega) \\ g_{\text{SIN}}(\omega) &= \sin(\omega) \end{aligned} \quad (2.10)$$

这样构造的距离特征就能够满足周期性条件，保证边界上的连续性，以及渐进的尖端条件。如果没有特殊说明，所有的结果都采用 SIN 型距离特征。

尽管 DeepSolid 具有强大的表达能力，理论上只要神经网络的规模足够，优化工作做得足够好，就可以得到高精度的能量结果，但现实中神经网络的计算和训练开销

很大，我们没有能力计算很大的超胞，使得这里超胞近似的效果不好，最终的结果受尺寸效应影响较大，并不一定能反映真实体系的结果。

2.2 原子间受力的估计量

2.2.1 最简单估计量

在固体体系的能量计算中，我们需要使用 Ewald 求和^[16]的技巧计算库伦势 E_{Coulomb} ，而不是直接对所有粒子的库伦势求和。这是因为库仑力是长程力，对不同粒子和不同格矢的求和是条件收敛的，即结果与求和次序有关^[16]。Ewald 求和的技巧简单来说，就是把原本的点电荷密度转化为两部分求和，一部分是原本的点电荷密度，加上以各个点电荷为中心、电荷大小相等符号相反的高斯分布的电荷密度，另一部分就是与新加的高斯分布电荷符号相反的分布。由于屏蔽效应，第一部分的求和在实空间容易收敛，而平滑的密度分布使得第二部分在倒空间容易收敛。

作为最简单的估计（Bare），我们可以直接计算

$$\mathbf{F}_{I;\text{Bare}} = -\partial_{\mathbf{R}_I} E_{\text{Coulomb}} \quad (2.11)$$

得到原子核 I 上的受力。

值得说明的是，尽管在 Ewald 求和中有屏蔽的效应，但这并没有消除点电荷的 δ 形分布。所以在电子靠近原子核时，力的大小依然表现为 r^{-2} ，这就使得 \mathbf{F}_I 的方差理论上可以是无穷大，也就是在受力计算中所谓的无限方差问题^[17-18]。当然在实际计算中，电子不会距离原子核无限近，方差也必然是一个有限值，只是它的值很大且不会随着波函数的优化而变小。无限方差问题的存在，使得我们需要寻找其他具有更好的统计性质的估计量来计算原子间受力。在过去几十年中，基于传统的 VMC，人们提出了很多不同的方案，我们将在下面具体介绍。

2.2.2 相反采样估计量

相反采样（antithetic sampling）是一种数值上减小统计误差的手段，其基本思想是为每个采样点匹配一个受力近似大小相等方向相反的点，从而抵消一部分统计波动^[19]。具体做法就是选择一个内层电子截断半径 r_{core} ，对每个采样的电子坐标 \mathbf{r}_i ，寻找与之最近的原子核 \mathbf{R}_I ，如果 $|\mathbf{r}_i - \mathbf{R}_I| < r_{\text{core}}$ ，则将 \mathbf{r}_i 关于 \mathbf{R}_I 做中心对称变换得到 \mathbf{r}'_i ，否则 \mathbf{r}'_i 依然等于 \mathbf{r}_i 。然后再计算所有电子做完变换后的原子核受力 $\mathbf{F}'_{I;\text{Bare}}$ 。为了保持采样的正确性，我们需要给 $\mathbf{F}'_{I;\text{Bare}}$ 加一个额外的权重

$$w = \left| \frac{\psi(\mathbf{r}'_1, \mathbf{r}'_2, \dots, \mathbf{r}'_{n_{\text{elec}}})}{\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n_{\text{elec}}})} \right|^2 \quad (2.12)$$

最后相反采样估计量（Antithetic）的表达式就是

$$\mathbf{F}_{I; \text{Antithetic}} = \frac{\mathbf{F}_{I; \text{Bare}} + w\mathbf{F}'_{I; \text{Bare}}}{2} \quad (2.13)$$

在本工作中，我们选择内层电子截断半径 $r_{\text{core}} = 0.5 \text{ Bohr}$ 。另外值得一提的是，在固体中，体系往往具有较高的中心对称性，即使原子核相对平衡位置有一个小的位移，也可以有 $w \rightarrow 1$ 从而抵消波动。而一般的分子只有当 $|\mathbf{r}_i - \mathbf{R}_I| \ll r_{\text{core}}$ 的时候才会有 $w \rightarrow 1$ 。

2.2.3 带修正的最简单和相反采样估计量

Assaraf 和 Caffarel 等人提出^[17]，我们可以在原本的系统哈密顿量 \hat{H}_0 的基础上定义了一个新的与参数 λ 有关的哈密顿量

$$\hat{H}(\lambda) = \hat{H}_0 + \lambda\hat{O} \quad (2.14)$$

其中为了表达的简单和结果的一般性，这里先假设我们关心的是个一般的可观测量 \hat{O} 的期望值。那么根据 Hellmann-Feynman 定理可得

$$\frac{\langle \psi_0 | \hat{O} | \psi_0 \rangle}{\langle \psi_0 | \psi_0 \rangle} = \left. \frac{dE_0}{d\lambda} \right|_{\lambda=0} = \left. \frac{dE_v}{d\lambda} \right|_{\lambda=0} + \varepsilon \quad (2.15)$$

但我们无法取得准确的基态波函数 ψ_0 和基态 E_0 ，只能根据试探波函数 ψ_T 得到能量的期望值 E_v 。这就会使计算结果与实际结果相差 ε ，但这个偏差会随着 ψ_T 接近 ψ_0 而降低。简单计算可以发现

$$\frac{dE_v}{d\lambda} = \partial_\lambda \frac{\langle \psi_T | \hat{H} | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \left\langle \hat{O} + 2\Re(E_L \partial_\lambda \log \psi_T^*) - 2E_v \Re(\partial_\lambda \log \psi_T^*) \right\rangle_{|\psi_T|^2} \quad (2.16)$$

其中 \Re 表示取实部。第一个导数可以变为偏导是因为优化目标函数 E_v 在收敛时对其他参数 θ 的导数应该接近零。

由 2.16 式可知，我们直接如 2.11 式计算其实是有系统误差的，应该包含 2.16 式中后面两项修正项：

$$\mathbf{F}_{I; \text{Bare-ZB}} = \mathbf{F}_{I; \text{Bare}} - 2\Re(E_L \partial_{\mathbf{R}_I} \log \psi_T^*) + 2E_v \Re(\partial_{\mathbf{R}_I} \log \psi_T^*) \quad (2.17)$$

这就是带修正的最简单估计量（Bare-ZB），其中 ZB 是 zero bias 的缩写。同理，我们也可以给相反采样估计量也加上同样的修正项，就得到了带修正的相反采样估计量（Antithetic-ZB）。新加的修正项在 ψ_T 严格等于 ψ_0 时严格为零，但在 ψ_T 不等于 ψ_0 时可以修正基于 ψ_T 采样而非 ψ_0 带来的误差。

2.2.4 Assaraf-Caffarel 估计量

在 2.16 式的基础上, Assaraf 和 Caffarel 等人提出了一种新的估计量^[17], 这里称为 Assaraf-Caffarel 估计量, 或 AC 估计量。这个估计量一开始是基于实值波函数提出的, 推广到复值波函数的形式为:

$$\mathbf{F}_{I; \text{AC-ZVZB}} = \mathbf{F}_{I; \text{Bare}} - \frac{(\hat{H} - E_L^*)\tilde{\psi}_I}{\psi_T} - 2\Re\left(E_L \frac{\tilde{\psi}_I}{\psi_T}\right) + 2E_v \Re\left(\frac{\tilde{\psi}_I}{\psi_T}\right) \quad (2.18)$$

其中 $\tilde{\psi}_I$ 是 $\partial_{\mathbf{R}_I}\psi_T$ 的近似函数。2.18 式中后面两项和 2.17 式中的相同, 都是为了减小系统误差。第二项的出现主要是为了降低统计误差, 从而解决无限方差的问题, 这也是下标中 ZV (zero variance) 的由来。

尽管 Assaraf-Caffarel 估计量的复值波函数形式并不复杂, 但它并不能很好地应用于固体体系。原因是一般 Assaraf-Caffarel 估计量会使用基于类氢原子波函数的导数近似 $\tilde{\psi}_I$, 但在固体中的波函数与孤立氢原子相去甚远。所以这里我们不在固体中实现和讨论 Assaraf-Caffarel 估计量。

2.2.5 空间坐标扭曲估计量

事实上我们还可以把 2.16 式分子上的 \hat{H} 换成局部能量 E_L :

$$\frac{dE_v}{d\lambda} = \partial_\lambda \frac{\langle \psi_T | E_L | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \langle \partial_\lambda E_L + 2(E_L - E_v) \Re(\partial_\lambda \log \psi_T) \rangle_{|\psi_T|^2} \quad (2.19)$$

这么做的动机是因为如果 ψ_T 接近于 ψ_0 , 那么 E_L 应该接近于一个常数, 会具有更好的统计性质。换言之, 在电子靠近原子核处, 波函数的尖端条件^[11]会保证势能项的和动能项的发散相互抵消, 得到一个合理的能量值, 从而可以避免出现无限方差的问题。这就得到了未扭曲的估计量 (No-SWCT):

$$\mathbf{F}_{I; \text{No-SWCT}} = -\partial_{\mathbf{R}_I} E_L + 2(E_v - E_L) \Re(\partial_{\mathbf{R}_I} \log \psi_T) \quad (2.20)$$

尽管未扭曲的估计量和空间坐标扭曲 (SWCT) 一点关系都没有, 但是为了和 Sorella 等人^[20]保持一致, 我们依然采用这个名称。

虽然尖端条件可以保证能量不会在原子核附近发散, 但能量的导数可能依然存在问题。空间坐标扭曲估计就是基于未扭曲估计的改进。这种方法最初是由 Umrigar 在 1989 年提出的^[21]。Filippi 和 Umrigar 在 2000 年扩展了这个想法^[22], 然后 Sorella 和 Capriotti 在 2010 年利用自动微分重新实现了这个方法^[20]。

空间坐标扭曲的核心思想是这样的。假设 QMC 没有统计误差或者有办法消除, 我们可以使用有限差分的方法计算受力, 即分别计算两个原子核位置略有不同的构型的

能量并做差。这样电子的分布会有变化。如果电子和原子核靠得很近，那么 \mathbf{r}_i 处的电子几乎就会跟着原子核移动到 $\bar{\mathbf{r}}_i$ 。如果我们依然用之前的波函数做重要性采样计算蒙特卡洛积分，积分的计算效率会很低，而且可能以高于实际的概率采样到能量导数较大的点，从而影响到受力的计算效率和精度。所以我们在计算时应该考虑到电子坐标的变化趋势。用公式写出来就是

$$\bar{\mathbf{r}}_i = \mathbf{r}_i + \Delta \mathbf{R}_I \omega_I(\mathbf{r}_i) \quad (2.21)$$

$$\omega_I(\mathbf{r}) = \frac{\sum_{\mathbf{L}_S} f(|\mathbf{r} - \mathbf{R}_I - \mathbf{L}_S|)}{\sum_{\mathbf{L}_S} \sum_{J=1}^{n_{\text{atom}}} f(|\mathbf{r} - \mathbf{R}_J - \mathbf{L}_S|)} \quad (2.22)$$

其中 f 是一个快速衰减的函数，所以在实际计算对 \mathbf{L}_S 求和只取最近邻计算。这里我们与 Filippi 和 Umrigar 一样^[22]，选择 $f(r) = r^{-4}$ 。那么空间坐标扭曲估计量就是

$$\begin{aligned} \mathbf{F}_{I;\text{SWCT}} = & -\partial_{\mathbf{R}_I} E_L - \sum_{i=1}^{n_{\text{elec}}} \omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} E_L \\ & + 2(E_v - E_L) \Re \left[\partial_{\mathbf{R}_I} \log \psi_T + \sum_{i=1}^{n_{\text{elec}}} \left(\omega_I(\mathbf{r}_i) \partial_{\mathbf{r}_i} \log \psi_T + \frac{1}{2} \partial_{\mathbf{r}_i} \omega_I(\mathbf{r}_i) \right) \right] \end{aligned} \quad (2.23)$$

2.3 实现

我们基于 DeepSolid 软件库^[8]实现了上述的几个估计量。本工作中的算法是之前利用 FermiNet 计算原子间受力的工作^[12]的扩展。为了进行充分的蒙特卡洛采样，我们需要进行多步循环，在循环中我们会计算并存储局部能量以及估计量每一项的值，并进行 50 步马尔科夫链蒙特卡洛 (MCMC)，使得在可承受范围内，两步结果之间的关联尽量小。在循环结束后我们会对所有的局部能量做平均得到 E_v ，再得出最终的受力计算结果。我们还会主循环之前先运行 5000 步 MCMC，也即 MCMC 预热，来优化电子构型。

与 FermiNet 上的工作^[12]相同，在 DeepSolid 上也需要对局部能量和势能 Ewald 求和的部分做一些修改，使得我们能够利用 Python JAX^[23] 的自动微分 (AD) 方法得到良好定义的能量关于原子坐标的梯度。此外，这里也同样采用了四分位距 (IQR) 的稳健统计方法。所谓 IQR 就是第一四分位数 (Q_1) 和第三四分位数 (Q_3) 的差值。所有大于 $Q_3 + 3\text{IQR}$ 和小于 $Q_1 - 3\text{IQR}$ 的数据都被剪切到这个边界上，从而去除一些不合理的数据点。同时，所有标准误差都上对剪切之后的数据利用考虑了马尔可夫链自相关的分块方法^[24-25]给出。

表 2.1 计算设置

来源	名称	值
DeepSolid ^[7]	硬件	32GB V100 GPU
	批量大小	4096
	单电子层隐藏神经元	256
	双电子层隐藏神经元	32
	网络层数	3
	行列式个数	1
	训练迭代次数	Li 原子 2×10^5 轮; $5 \times 1 \times 1$ 氢链和 $1 \times 1 \times 1$ 原胞氢化锂 3×10^5 轮
	优化器	KFAC-JAX ^[26]
	其他设置	默认
受力计算 ^[12]	MCMC 预热步数	100
	主循环中 MCMC 步数	50
	额外能量推理步数	0

在训练和计算中的设置如表 2.1 所示。本工作主要关注固体中的原子间受力估计的实现和收敛后的表现，所以神经网络的设置基本固定不变。

第三章 结果和讨论

3.1 收敛网络下的能量和空间坐标扭曲估计量

和分子体系不同，固体中的原子间受力不能使用莫斯势拟合，也缺乏相关工作拟合实验数据，还很少有人发表一些高精度方法的计算结果。就比如理论研究者最常研究的一维氢链，现有的研究^[27-28]主要是关心氢链的状态方程，也就是所有氢原子均匀排布，仅改变原子的间距，计算这些不同构型下的能量。但是对于原子间受力计算而言，如果关心的是由两个氢原子组成的晶胞内的一个氢原子的受力，则应该保持晶格大小不变，改变晶格内的两个氢原子的间距。而目前没有这方面的发表的工作，所以我们这里无法与其他人的结果比较。但是空间坐标扭曲估计量在之前非周期性体系的测试中是最精确的，统计误差也最小^[12,20]。如果在周期性体系中空间坐标扭曲估计量的结果和能量曲线相比没有明显的问题，我们就认为代码实现和计算收敛性上没有明显问题，于是选择相信空间坐标扭曲估计量能够给出好的结果。在 3.2.1 节中，我们将初步比较不同的估计量，进一步说明选择空间坐标扭曲估计量的合理性。所以在本节中，我们将只展示空间坐标扭曲估计量与势能曲线的结果。

我们计算了 $5 \times 1 \times 1$ 的氢链体系，每个晶胞内包含两个氢原子。我们保持晶格常数 $R = 4.0 \text{ Bohr}$ 不变，改变晶胞内的两个氢原子的间距 r ，使用 DeepSolid 计算，得到了一条能量曲线，即图 3.1a 中的蓝色方块。图中清楚得展示了 $r = 2.0 \text{ Bohr}$ 并不是系统的一个稳定平衡点，这与 Peierls 不稳定性^[29]的结论一致，即均匀排布的一维氢链不稳定，会倾向于二聚化。我们又在各点计算了空间坐标扭曲估计得到的平行于氢链方向的受力，也就是图 3.1a 中绿色短线的斜率。可以发现空间坐标扭曲估计的结果和能量曲线吻合得很好。

我们也计算了 $1 \times 1 \times 1$ 的原胞氢化锂体系，保持晶格常数 $a = 4.0 \text{ \AA} \approx 7.6 \text{ Bohr}$ 不变，沿其中一条体对角线移动体心的氢原子。如 3.1b 所示，我们把这个氢原子距离偏离中心的距离记为 x 。和氢链一样，我们画出了在不同 x 处的体系能量和空间坐标扭曲估计的结果，分别为图 3.1b 中的蓝色方块和绿色短线的斜率。需要指出的是，我们发现由于 $1 \times 1 \times 1$ 的体系太小，尺寸效应明显，计算出来的能量曲线甚至在 $x = 0$ 处不是稳定平衡点，所以这里只选取了 x 从 1.0 Bohr 到 3.0 Bohr 的一段展示。不过即使能量曲线与实际情况有较大差异，但空间坐标扭曲估计的结果和能量曲线吻合得很好，说明它能够很好地给出 $1 \times 1 \times 1$ 原胞氢化锂体系的受力计算结果。

我们也尝试了使用莫斯势（图 3.1a 中黄色区域）和二次函数（图 3.1b 中黄色区

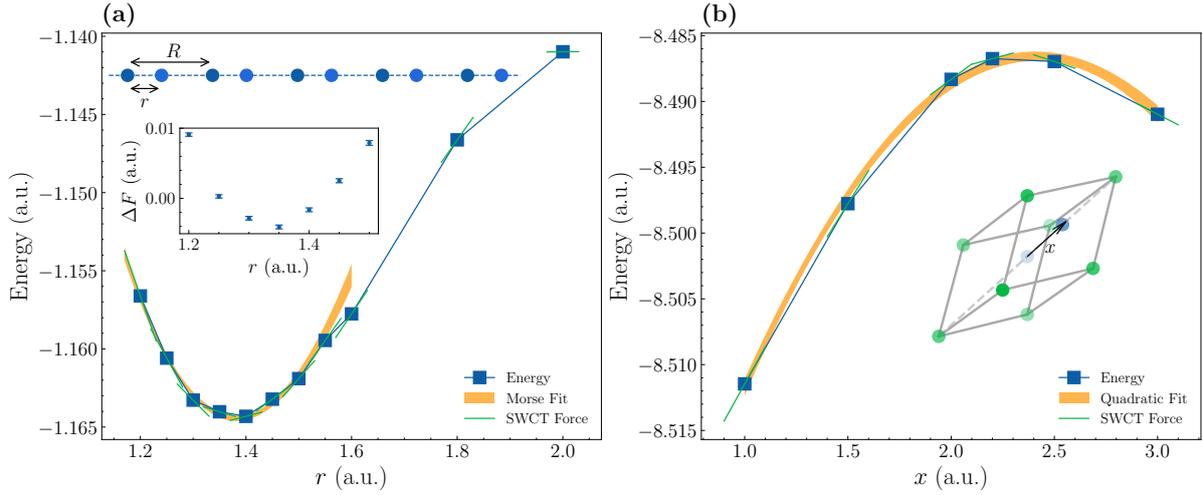


图 3.1 (a) $5 \times 1 \times 1$ 氢链保持 $R = 4.0$ Bohr, 改变胞内氢原子间距 r 的能量曲线 (蓝色方块) 和计算 2000 步的空间坐标扭曲估计量的受力计算结果 (绿色短线的斜率), 以及在 $r \leq 1.50$ Bohr 使用莫斯势拟合能量的结果和误差 (黄色区域)。第一幅插图为氢链体系的示意图。第二幅插图为空间坐标扭曲估计的结果和拟合的莫斯势导数的差。(b) $1 \times 1 \times 1$ 的原胞氢化锂保持晶格常数 $a = 4.0 \text{ \AA}$ 不变, 沿其中一条体对角线移动中心氢原子的能量曲线 (蓝色方块) 和计算 2000 步的空间坐标扭曲估计量的受力计算结果 (绿色短线的斜率), 以及使用二次函数拟合能量的结果和误差 (黄色区域)。插图为原胞氢化锂体系的示意图, 其中绿色表示锂原子核, 蓝色表示氢原子核, 电子没有画出。

域) 拟合能量结果, 希望能够利用拟合结果的导数和受力计算的结果做比较, 但拟合效果较差。对于氢链, 我们选择了 $r \leq 1.50$ Bohr 的区域进行莫斯势拟合, 拟合的结果与能量结果有一定的差距。我们把空间坐标扭曲估计的结果和莫斯势的导数的差画在图 3.1a 的插图中, 很容易发现空间坐标扭曲估计和莫斯势导数的差有一个明显的抛物线形状, 这表明莫斯势拟合确实存在一个系统性的误差, 并不适合用于进行受力结果的比较。尽管不适合比较, 但也至少说明空间坐标扭曲估计的误差会在 10^{-2} a.u. 以内。对于氢化锂体系, 使用二次函数拟合的结果具有较大误差, 甚至不能得到正确的平衡位置, 所以我们不使用拟合的结果和受力计算进行比较。

总之, 空间坐标扭曲估计的计算结果上合理的, 它在氢链和氢化锂上和能量曲线吻合得很好, 以它作为受力结果的比较基准具有一定的合理性。

3.2 各估计量的比较

3.2.1 垂直方向

由于对称性, 原子核在垂直于移动方向上的受力应该是零, 所以我们只关心平行方向的结果。但是由于蒙特卡洛方法随机性的存在, 垂直方向的受力计算结果并不是零, 而且垂直方向上的误差和平行方向有着相同的表现^[12]。虽然在 3.1 节中我们已经看到了空间坐标扭曲估计量和能量曲线的良好吻合, 但依然缺乏证据表明在我们考虑

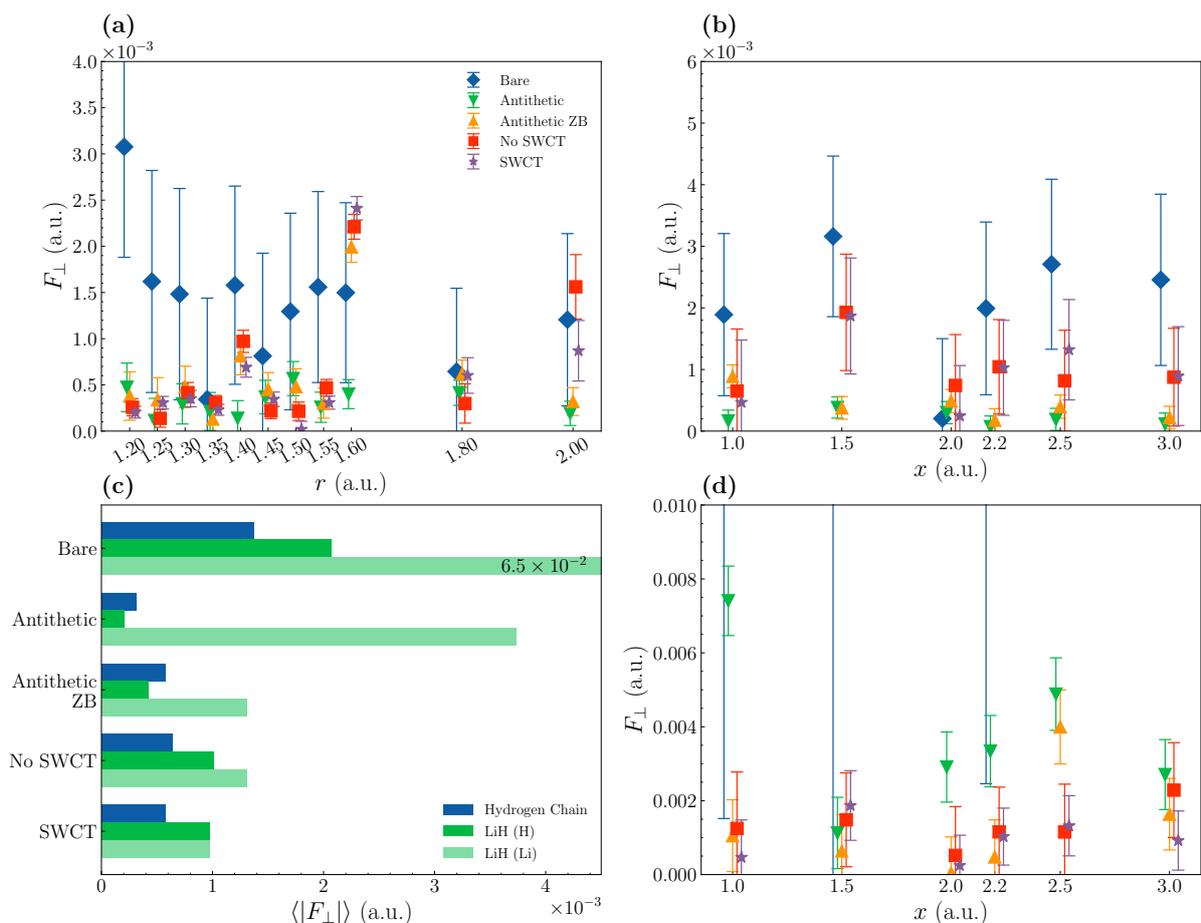


图 3.2 (a) $5 \times 1 \times 1$ 氢链不同 r 处的最简单估计 (蓝色菱形)、相反采样估计 (绿色倒三角)、带修正的相反采样估计 (黄色上三角) 和未扭曲估计 (红色方块) 受力计算结果与空间坐标扭曲估计 (紫色五角星) 在垂直方向受力计算结果的绝对值。除空间坐标扭曲估计和未扭曲估计只计算了 2000 步以外, 每个估计量都计算了 10000 步。(b) 与 (a) 类似, 但计算的是 $1 \times 1 \times 1$ 的原胞氢化锂中的氢原子在不同 x 处的结果, 且空间坐标扭曲估计和未扭曲估计计算了 5000 步。(c) 不同估计量在不同 r 处 (氢链体系, 蓝色) 和不同 x 处 (氢化锂体系, 深绿为氢原子上的受力, 浅绿为锂原子上的受力) 的垂直方向受力计算结果的绝对值 $|F_{\perp}|$ 做平均。(d) 与 (b) 类似, 但展示的是锂原子上的结果。为了更清楚地展示各个估计量的结果, (a) (b) (d) 图绘制时使用了一些错位。

的各个估计量中, 空间坐标扭曲估计量是最可信的, 能够用来作为参考值, 正如我们在 3.2.2 节中将要做的那样。由于垂直方向的受力的理论值应该严格为零, 通过比较不同估计量在垂直方向的受力计算结果, 恰好可以给我们一个参考, 找出表现最好的估计量, 以便我们进一步研究平行方向的受力。

我们使用了最简单估计 (Bare)、相反采样估计 (Antithetic)、带修正的相反采样估计 (Antithetic-ZB)、未扭曲估计量 (No-SWCT) 和空间坐标扭曲估计量 (SWCT) 计算了氢链不同 r 处的垂直方向受力 (图 3.2a) 以及氢化锂中的氢原子和锂原子在不同 x 处的垂直方向受力 (图 3.2b、d)。我们很容易注意到最简单估计的结果误差较大, 统计误差也很大, 不能作为参考值。而且虽然最简单估计在氢原子上的结果与其他几个估计量相比还算过得去, 但是在锂原子上的结果就远差于其他这些估计量的结

果，其误差以及统计误差要大一个数量级，在图 3.2c、d 中甚至画不下。这也表明固体中的受力计算确实也受到了无限方差问题的影响，对于锂原子这样有着两个内壳层电子的原子来讲，相比于氢原子，电子有更大的概率处于原子核附近，也就是蒙特卡洛方法有更大的纪律采样到更大的库伦力，受到无限方差问题影响也就更严重。正是由于计算锂原子上的受力更具挑战性，也更能代表一般的具有内壳层电子的原子的情况，在接下来的讨论中，我们会关注氢化锂中锂原子上的受力而不是氢原子上的。

我们没有考虑带修正的最简单估计量，因为修正很可能被最简单估计量的统计误差淹没。而在使用相反采样后，无论是氢原子还是锂原子上的受力误差都有了明显的下降。但在锂原子上，相反采样估计量尽管有效地降低了受力计算的统计误差，但和其他的估计量相比，依然具有较大的系统误差。这一系统误差能够被带修正的相反采样估计很好得消除，达到与未扭曲估计量和空间坐标扭曲估计量相近的水平。

总体而言，在氢原子和锂原子垂直方向的受力上，带修正的相反采样估计、未扭曲估计量和空间坐标扭曲估计量是精度最高的三个估计量，它们的精度大致相等。在氢原子上，它们的表现很相近，而且受到蒙特卡洛的统计误差影响较大，会出现带修正的估计反而比不带修正的估计误差更大的情况。在锂原子上，空间坐标扭曲估计量更胜一筹，有着更小的误差。所以在 3.2.2 中的讨论中，我们将使用空间坐标扭曲估计量的结果作为参考值，来进一步比较各个估计量。

3.2.2 平行方向

与 3.2.1 节相同，我们使用了各个估计量计算了氢链和氢化锂在各个构型下的受力，并与空间坐标扭曲估计 (SWCT) 做比较，如图 3.3a、b 所示。同时为了直观起见，我们将这三个估计量在不同构型处和空间坐标扭曲估计量结果的差的绝对值的平均值画在了图 3.3c 中。总体来看平行方向的结果和垂直方向很相近，没有很大的出入，这也说明使用空间坐标扭曲估计作为参考值是具有合理性的。具体来说，在氢链上，这几个估计量彼此之间吻合得很好，与空间坐标扭曲估计量的差距都只在 10^{-4} 到 10^{-3} a.u. 的数量级，而结构优化和声子计算所需要的原子间受力的精度大约在 10^{-3} a.u. 的数量级，所以这里展示的结果在这些任务的精度要求之内。在氢化锂的锂原子上的受力也和氢链上的结果相似，除了最简单估计量和其他估计量相比具有很大的误差和统计误差以外，不过这也和垂直方向上的现象一致。在氢原子上，最简单估计量和空间坐标扭曲估计量相比，误差也能和其他的估计量保持在一个数量级。

至于相反采样估计，考虑到计算步数上的差异，其统计误差与空间坐标扭曲估计量相近，说明相反采样的技巧确实能够有效地降低统计误差。在往相反采样估计里加入修正项后，总体而言受力计算的误差比起没有加入修正的时候有所降低，在氢链的

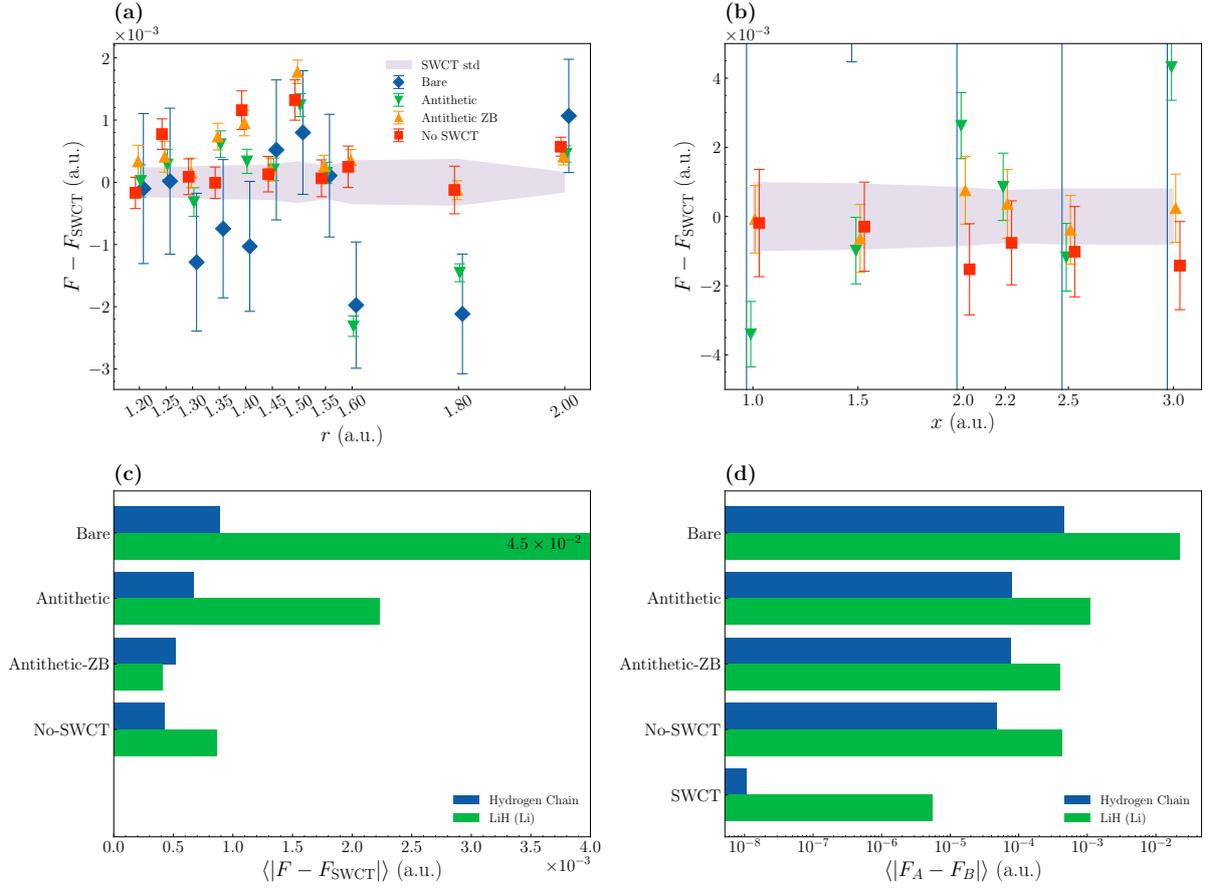


图 3.3 (a) $5 \times 1 \times 1$ 氢链不同 r 处的最简单估计 (蓝色菱形)、相反采样估计 (绿色倒三角)、带修正的相反采样估计 (黄色上三角) 和未扭曲估计 (红色方块) 在平行方向上的受力计算结果与空间坐标扭曲估计的差 $F - F_{\text{SWCT}}$ 。紫色阴影区域代表了空间坐标扭曲估计的统计误差。除空间坐标扭曲估计和未扭曲估计只计算了 2000 步以外, 每个估计量都计算了 10000 步。(b) 与 (a) 类似, 但计算的是 $1 \times 1 \times 1$ 的原胞氢化锂中的锂原子在不同 x 处的结果, 且空间坐标扭曲估计和未扭曲估计计算了 5000 步。(c) 与 (a) (b) 类似, 但将不同构型的受力差的绝对值 $|F - F_{\text{SWCT}}|$ 做了平均。(d) 不同估计量计算得到的晶胞内两个原子的受力差值。

氢原子上降低的幅度比较有限, 而且并不是在所有情况下加入修正项都能降低误差, 只有在 $r = 1.6 \text{ Bohr}$ 和 $r = 1.8 \text{ Bohr}$ 处相反采样估计误差较大时, 加入修正项才能较明显地降低误差; 而在氢化锂的锂原子上, 加入修正项后各个构型处的受力计算误差有着较为明显的降低。这一方面可能是由于蒙特卡洛的随机性, 马尔可夫链上前后的自相关性^[24]会低估统计误差, 使得在相反采样估计误差较小时, 修正项被统计噪声淹没。另一方面也可能是由于空间坐标扭曲估计也并不足够准确, 与之相比不能反映实际的误差。

未扭曲的估计量的在氢链上与空间坐标扭曲估计量的结果是最接近的, 但在氢化锂的锂原子上却不如带修正的相反采样估计。但正如 3.2.1 节中所展示的, 这三个估计量总体来说误差和统计误差都具有相近的表现, 这些结论会较大地受到统计误差的影响。

同时，我们也注意到了尽管理论上同一晶胞内的两个原子的受力应该大小相等方向相反，但不同估计量的受力计算结果却并不是这样，如图 3.3d 所示。空间坐标扭曲估计量对两个原子的受力给出了一致的结果，平均差距在氢链上只有 10^{-8} a.u.，在氢化锂的锂原子上只有 10^{-5} a.u.，远低于受力计算的误差和标准误差。但对于其他四个估计量则并非如此。两个原子的受力差值达到了 10^{-4} 到 10^{-3} a.u. 的数量级，氢化锂中锂原子上的最简单估计甚至达到了 10^{-2} a.u.，与这几个估计量的误差相仿。这也从另一个方面验证了空间坐标扭曲估计量相比于其他几个估计量更加可信。这一现象从理论上也比较好理解：由于体系平移对称性的存在，如果所有原子核和电子都移动一个 Δx 的距离，则能量的结果应该不变。也就是说通过对局部能量求导并使用空间坐标扭曲变换，两个原子上的受力理应是相差无几的。但对于未扭曲的估计量，电子并没有跟着原子核移动，就没有这种性质。对于其他几个基于 Ewald 库伦势求导的估计量，本质上还是在进行电子坐标采样计算，对两个原子的库仑力计算并没有如空间坐标扭曲估计量那样的关联，同一个电子坐标采样出来的点，可以对一个原子有较大的库仑力，但对另一个原子只有很小的库仑力。这也就解释了这些估计量两个原子的受力差值与受力结果的误差相仿的原因。相反采样能够在一定程度上减少这种情况的出现，所以两个原子的受力差值相比于最简单估计量有了一些下降。而相反采样的修正项则并没有这样的作用，带不带修正的相反采样估计在原子的受力差值上几乎没有区别。

3.3 不同估计量的单步相对开销

在 3.2 节中，我们比较了在氢链上不同估计量的精度。本节中我们将比较不同估计量的单步耗时情况，以及随着体系大小的变化。为了公平地比较不同的估计量在不同体系大小下的表现，我们将从单步耗时 $T_{\text{estimator}}$ 中扣除了空循环的用时 T_0 ，即扣除了 MCMC 采样的用时，同时又除以了单步的训练时间 T_{train} 作为基准，得到了相对开销 RC：

$$\text{RC} = \frac{T_{\text{estimator}} - T_0}{T_{\text{train}}} \quad (3.1)$$

我们在不同大小的氢链体系下测量了不同估计量的单步计算耗时，相对开销与体系大小的关系如图 3.4 所示。图中清楚地展示了最简单估计量速度最快，相反采样估计量次之，带修正的相反采样估计量更慢一些，最慢的是空间坐标扭曲估计量。同时，最简单估计量和相反采样估计量的相对开销随着体系的增大而减小，说明它们具有比训练神经网络更低的标度。带修正的相反采样估计量的相对开销则随着体系的增大相对平缓，只是很有限地增大。而空间坐标扭曲估计量就很明显随着体系的增大它的相

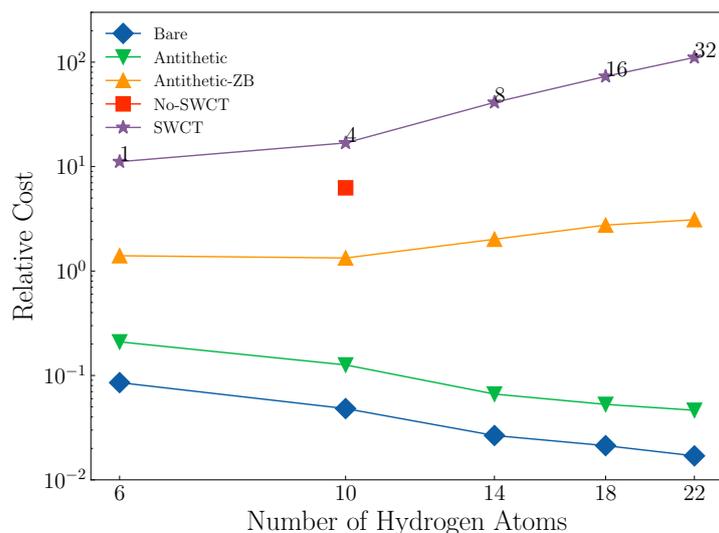


图 3.4 最简单估计量（蓝色方块）、相反采样估计量（绿色倒三角）、带修正的相反采样估计量（黄色上三角）、未扭曲估计（红色方块），以及空间坐标扭曲估计量（紫色五角星）在 $3 \times 1 \times 1$ 、 $5 \times 1 \times 1$ 、 $7 \times 1 \times 1$ 、 $9 \times 1 \times 1$ 、 $11 \times 1 \times 1$ 氢链上的单步相对开销，即每步耗时 $T_{\text{estimator}}$ 减去空循环（只进行 MCMC 采样而不计算力）的时间 T_0 ，再除以训练的单步耗时 T_{train} 。其中空间坐标扭曲估计量上标注的数字表示了分的批数 n_c 。所有时间都是在 4 张 V100 GPU 上计算得到的。

对开销也一直在增大。对于未扭曲的估计量，从理论还是实际上计算量都是空间坐标扭曲估计量的一半，故没有必要绘制整条曲线。需要说明的是空间坐标扭曲估计量由于需要计算高阶导数，对显存的消耗很大，我们的显存并不允许我们一次性计算完一整个批量的电子构型，需要分成 n_c 批计算，也就是图 3.4 中在紫色五角星边上标注的黑色数字。分多批计算会带来额外的开销，使得空间坐标扭曲估计量的相对开销增大。另外，虽然图 3.4 中展示的批数是按指数增加的，但这是因为 n_c 必须是批量大小的因子，并不意味着显存的消耗必然是按照指数增长的。事实上，在较小的体系中，空间坐标扭曲估计量的显存消耗大约是 $O(n^4)$ 的标度^[12]。

从理论上讲，在自动微分中，一个函数 $f(\mathbf{x})$ 和它的梯度 $\nabla f(\mathbf{x})$ 具有相同的渐进复杂度，但是如果计算拉普拉斯算符 $\nabla^2 f(\mathbf{x})$ ，则会使得渐进复杂度增加一个输入维度的因子。在氢链上， $n_{\text{atom}} = n_{\text{elec}} = n$ 。如果我们认为计算 DeepSolid 波函数的复杂度是 $O(n^3)$ ^[3]，那么计算局部能量的复杂度就是 $O(n^4)$ 。所以最简单估计量和相反采样估计量的复杂度就会是 $O(n^3)$ ，相对开销应当随着体系的增大而减小，这与图 3.4 中展示的一致；而带修正的相反采样估计量和空间坐标扭曲估计量的复杂度是 $O(n^4)$ ，相对开销会随着体系的增大基本不变，这与图 3.4 中和上文分析的一致。虽然实际上在较小的体系上波函数 $O(n^3)$ 的计算并未占主导，可能实际的波函数计算体现出来只有 $O(n^2)$ ，但相对开销的结论却不会改变。

总的来说，虽然空间坐标扭曲估计量的计算结果很精确，但是由于显存的原因，相对开销却随体系的增大而减小，所以对于较大的体系，空间坐标扭曲估计量的开销会

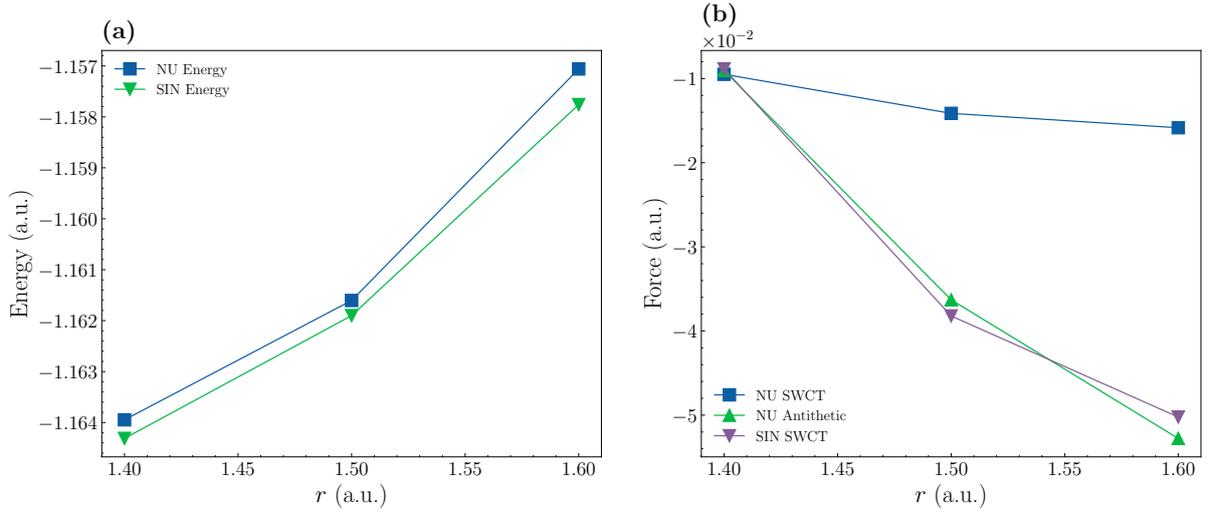


图 3.5 (a) $5 \times 1 \times 1$ 氢链不同 r 处的使用 SIN 型（红色倒三角）和 NU 型（蓝色方块）网络距离特征的能量结果。(b) 不同 r 处的 SIN 型距离特征下的空间坐标扭曲估计量（红色倒三角）、NU 型距离特征下的空间坐标扭曲估计量（蓝色方块）和相反采样估计量的受力计算结果。标准误差没有在图内画出，因为它太小了。

变得无法忍受。此时可以选择使用带修正的相反采样估计或者开销更低的估计量。

3.4 不同的网络距离特征

DeepSolid 的默认是采用 2.9 式中的 NU 型距离输入特征，但我们到目前为止一直是展示的 SIN 型特征的结果。这是因为 NU 型特征存在一些问题。我们比较了相同超参数（表 2.1）在 $5 \times 1 \times 1$ 氢链不同 r 处使用 SIN 型和 NU 型距离特征的能量和受力计算结果，如图 3.5 所示。从图 3.5a 中可以很明显地发现 NU 型特征的能量结果都高于 SIN 型特征，根据能量的变分性可知 NU 型特征训练得到的能量和波函数都比 SIN 型特征的差。这说明 NU 型特征相比于 SIN 型特征具有优化与收敛上的问题。这可以从 $g(\omega)$ 的展开上理解。对于 $g_{\text{NU}}(\omega)$ 而言，除了必要的 ω 的一阶项来保证 $\omega \rightarrow 0$ 时回到原本的欧式距离以外，最低的就是关于 ω 的二阶项；但对于 g_{SIN} ，除去一阶项以外就是 ω 的三阶项。这使得如果使用 SIN 型距离特征，在氢链这个体系上，与氢链垂直方向的像的影响得以迅速衰减，从而更加容易优化和收敛到更好的基态。

如果仅仅是能量和波函数结果差一些，也不能说 NU 型特征具有很大的问题。但是图 3.5 显示，尽管使用 NU 型距离特征得到的能量曲线差不多就是把 SIN 型的曲线向上进行了一个平移，使用 NU 型输入特征的空间坐标扭曲估计量的受力计算结果却存在严重的问题，它和 SIN 型距离特征下的空间坐标扭曲估计量的结果存在巨大偏离。这并不是由于 NU 型输入特征得到的波函数质量很差，因为使用 NU 型输入特征的相反采样估计量的受力计算结果和 SIN 型的空间坐标扭曲估计量的结果吻合得很不

错。主要问题还是出在空间坐标扭曲估计量所需的高阶导数上。我们计算局部能量的时候，需要计算波函数对电子坐标的二阶导数，而计算空间坐标扭曲估计量时，需要计算局部能量对电子坐标和原子坐标的导数，也就是需要计算波函数的三阶导数。可是 NU 型输入特征的三阶导数在 $\pm\pi$ 处是不连续的，这就使得使用 NU 型输入特征计算空间坐标扭曲估计量会存在严重的问题。但是相反采样估计量不需要计算波函数的高阶导数，所以不会受到这个不连续性的影响。

第四章 结论

本工作在周期性体系中扩展了原有的原子间受力计算方案，并实现了一些估计量。我们在一维氢链和三维氢化锂的固体体系中进行了测试，展示了原子间受力计算的正确性和精度，体现了使用神经网络波函数方法计算固体体系的强大能力。我们还对不同的估计量进行了精度和单步相对开销上的比较，为受力计算提供了指导：如果追求精度，应该使用空间坐标扭曲估计量，但要承受随着体系增大而不断增大的相对开销；如果想要可控的相对开销，可以使用带修正的相反采样估计量；如果只是追求速度，至少也应该使用相反采样估计量。此外我们还分析了不同的网络输入特征，指出默认的 NU 型输入特征并不适合用于计算受力，应该使用 SIN 型特征。

参考文献

- [1] Han J, Zhang L, E W. Solving Many-Electron Schrödinger Equation Using Deep Neural Networks[J]. Journal of Computational Physics, 2019, 399: 108929.
- [2] Choo K, Mezzacapo A, Carleo G. Fermionic Neural-Network States for Ab-Initio Electronic Structure[J]. Nature Communications, 2020, 11(1): 2368.
- [3] Pfau D, Spencer J S, Matthews A G D G, et al. Ab Initio Solution of the Many-Electron Schrödinger Equation with Deep Neural Networks[J]. Physical Review Research, 2020, 2(3): 033429.
- [4] Scherbela M, Reisenhofer R, Gerard L, et al. Solving the Electronic Schrödinger Equation for Multiple Nuclear Geometries with Weight-Sharing Deep Neural Networks[J]. Nature Computational Science, 2022, 2(5): 331-341.
- [5] Ren W, Fu W, Wu X, et al. Towards the Ground State of Molecules via Diffusion Monte Carlo on Neural Networks[J]. Nature Communications, 2023, 14(1): 1860.
- [6] Cassella G, Sutterud H, Azadi S, et al. Discovering Quantum Phase Transitions with Fermionic Neural Networks[J]. Physical Review Letters, 2023, 130(3): 036401.
- [7] Li X, Li Z, Chen J. Ab Initio Calculation of Real Solids via Neural Network Ansatz[J]. Nature Communications, 2022, 13(1): 7895.
- [8] DeepSolid[CP/OL]. (2023-04-11) [2023-04-19]. <https://github.com/bytedance/DeepSolid>.
- [9] Spencer J S, Pfau D, Botev A, et al. Better, Faster Fermionic Neural Networks[EB/OL]. (2020-11-13) [2023-05-02]. <http://arxiv.org/abs/2011.07125>. arXiv: 2011.07125 [physics]. preprint.
- [10] FermiNet: Fermionic Neural Networks[CP/OL]. (2022-06-17) [2022-06-26]. <https://github.com/deepmind/ferminet>.
- [11] Kato T. On the Eigenfunctions of Many-Particle Systems in Quantum Mechanics[J]. Communications on Pure and Applied Mathematics, 1957, 10(2): 151-177.
- [12] Qian Y, Fu W, Ren W, et al. Interatomic Force from Neural Network Based Variational Quantum Monte Carlo[J]. The Journal of Chemical Physics, 2022.
- [13] Rajagopal G, Needs R J, James A, et al. Variational and Diffusion Quantum Monte Carlo Calculations at Nonzero Wave Vectors: Theory and Application to Diamond-Structure Germanium[J]. Physical Review B, 1995, 51(16): 10591-10600.
- [14] Kent P R C. Techniques and Applications of Quantum Monte Carlo[D/OL]. University of Cambridge. 1999 [2023-02-24]. <https://web.ornl.gov/~kentpr/thesis/>.
- [15] Whitehead T M, Michael M H, Conduit G J. Jastrow Correlation Factor for Periodic Systems [J]. Physical Review B, 2016, 94(3): 035157.
- [16] Toukmaji A Y, Board J A. Ewald Summation Techniques in Perspective: A Survey[J]. Computer Physics Communications, 1996, 95(2-3): 73-92.

-
- [17] Assaraf R, Caffarel M. Zero-Variance Zero-Bias Principle for Observables in Quantum Monte Carlo: Application to Forces[J]. *The Journal of Chemical Physics*, 2003, 119(20):10536-10552.
- [18] Hammond B L, Lester W A, Reynolds P J. Monte Carlo Methods in *Ab Initio* Quantum Chemistry: vol. 1[M]. WORLD SCIENTIFIC, 1994.
- [19] Chiesa S, Ceperley D M, Zhang S. Accurate, Efficient, and Simple Forces Computed with Quantum Monte Carlo Methods[J]. *Physical Review Letters*, 2005, 94(3):036404.
- [20] Sorella S, Capriotti L. Algorithmic Differentiation and the Calculation of Forces by Quantum Monte Carlo[J]. *The Journal of Chemical Physics*, 2010, 133(23):234111.
- [21] Umrigar C J. Two Aspects of Quantum Monte Carlo: Determination of Accurate Wavefunctions and Determination of Potential Energy Surfaces of Molecules[J]. *International Journal of Quantum Chemistry*, 1989, 36(S23):217-230.
- [22] Filippi C, Umrigar C J. Correlated Sampling in Quantum Monte Carlo: A Route to Forces [J]. *Physical Review B*, 2000, 61(24):R16291-R16294.
- [23] JAX: Autograd and XLA[CP/OL]. (2023-04-19) [2023-04-19]. <https://github.com/google/jax>.
- [24] Wolff U. Monte Carlo Errors with Less Errors[J]. *Computer Physics Communications*, 2004, 156(2):143-153.
- [25] Hesse D. Analysis of Monte Carlo Data[CP/OL]. (2022-08-26) [2023-04-28]. <https://github.com/dhesse/py-uwerr>.
- [26] KFAC-JAX - Second Order Optimization with Approximate Curvature in JAX[CP/OL]. (2023-04-02) [2023-04-19]. <https://github.com/deepmind/kfac-jax>.
- [27] Stella L, Attaccalite C, Sorella S, et al. Strong Electronic Correlation in the Hydrogen Chain: A Variational Monte Carlo Study[J]. *Physical Review B*, 2011, 84(24):245117.
- [28] Simons Collaboration on the Many-Electron Problem, Motta M, Ceperley D M, et al. Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods[J]. *Physical Review X*, 2017, 7(3):031059.
- [29] Peierls R E, Peierls R E. *Quantum Theory of Solids*[M]. Oxford, New York: Oxford University Press, 1996. 238 pp.

致谢

感谢字节跳动字节跳动研究部的李向博士对本工作的支持，我在与他的多次讨论中学习到了很多，得到了有益的启发。感谢字节跳动字节跳动研究部对本工作提供的GPU算力支持，也感谢字节跳动 AI Lab 量子方向的所有成员，在对本工作的讨论中提出了很多宝贵的建议。

感谢导师对本工作在选题上的启发，以及对工作以及文章写作的悉心指导。感谢组内师兄师姐对本工作的支持和鼓励。

感谢本科室友为本文的写作提供了安静和谐的环境。

感谢北京大学以及物理学院提供了良好的平台与教学科研资源，使我们能够较为顺利地完成本项工作。

北京大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名：钱昱冰 日期：2023年5月22日

学位论文使用授权说明

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

- 按照学校要求提交学位论文的印刷本和电子版本；
- 学校有权保留学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
- 学校可以采用影印、缩印、数字化或其它复制手段保存论文；
- 因某种特殊原因需要延迟发布学位论文电子版，授权学校一年/两年/三年以后，在校园网上全文发布。

（保密论文在解密后遵守此规定）

论文作者签名：钱昱冰 导师签名：陈昱
日期：2023年5月22日