arXiv:2611.33150v1 [cond-mat.quant-gas] 24 Nov 2026

跨越玻色子相变的参数化神经网络波函数

Daniil S. Antonenko, [1, *] Liang Fu, [2] 和 Leonid I. Glazman[1]

[1] *Department of Physics, Yale University, New Haven, Connecticut 06520, USA* [2] *Department of Physics, MIT, Cambridge, USA*（日期：2026年11月25日）

我们提出了 *Boseformer*，这是一种通用的基于注意力的神经网络架构，用于表示具有玻色量子统计的粒子的波函数，包括超流体相和晶体相。该模型经过训练，可生成作为哈密顿量参数函数的波函数，从而描述这两种相之间的量子相变。通过变分蒙特卡罗优化完成训练后，该模型可以作为基础模型描述这一相变，而无需针对各个参数值重新训练。我们在具有 $1/r^3$ 排斥相互作用的二维玻色子系统上展示了该方法。

代码可用性： https://github.com/andanya/boseformer

一、引言

近年来，神经变分波函数，或称神经量子态（NQS），已成为强大的拟设¨，能够近似多种多样的量子态[1]。它们通常与变分蒙特卡罗（VMC）优化方案配合使用，已成功应用于多种电子与自旋晶格模型，以及量子化学[2,3] 和凝聚态物理[4–6]中的连续系统。

利用 NQS 研究玻色子系统受到的关注较少（但请参阅文献 7–9）。费米子波函数的反对称性通常通过广义轨道的行列式来保证[2,3]。将这一方法简单推广至玻色子系统时，会以永久式取代行列式，但这在数值上不可行：计算永久式的难度要大得多，需要 $O(N2^N)$ 次运算，而计算行列式只需 $O(N^3)$ 次运算[10]。另一种方法是采用置换不变网络来生成对称输出，DeepSets 架构便实现了这种方法[7,11]。

在本研究中，我们提出*Boseformer*，一种不同的基于Transformer的架构[12]，用于表示玻色子的完整多体波函数，并展示了其优势。我们使用广义轨道的乘积来生成依赖于所有粒子坐标的多体玻色子波函数；见图2。如下文所示，我们的架构能够自然地表示超流态，也能成功描述玻色子的晶体相。

我们的下一项重大进展是开发一种参数化（基础）模型架构[13–15]，它利用单个网络生成以哈密顿量参数为条件的波函数，并将这些参数作为网络的输入。在这种设置下，网络经过优化，可同时在选定的一组参数点上求得基态。训练完成后，网络无需重新训练，即可预测哈密顿量参数连续范围内的多体波函数。值得注意的是，我们证明了该方法甚至适用于跨越量子相变的波函数。

我们将该方法应用于具有偶极-偶极排斥相互作用的二维玻色子系统，以展示其效果。我们的模型成功预测了该系统中跨越超流体到晶体相变的波函数；参见图 1。所得变分波函数的能量与标准扩散蒙特卡罗（DMC）方法的结果相符，该方法通常用于研究此类系统。与此同时，与 DMC 不同，我们的方法能够显式生成波函数，因此可以计算任意基态可观测量。

II. 物理系统

我们研究一种连续模型，其中包含 N 二维玻色子粒子，具有抛物线色散和成对排斥相互作用。通常，随着玻色子间相互作用增强，超流体凝聚态会逐渐耗尽，量子关联的重要性也随之增加；相关研究曾结合氦开展，见参考文献16和17。对于长程相互作用，晶体相在能量上可能与超流体相竞争[18–20]。

在本研究中，我们聚焦于 $1/r^3$ 排斥相互作用，例如，TMD 双层中面外激子之间的偶极–偶极相互作用[21,22]就与之相关。以 $\hbar = m = 1$ 为单位，哈密顿量写作：

$$\mathcal{H} = -\frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\nabla}_i^2 + \frac{\lambda}{\sqrt{\pi n}} \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|^3}. \qquad (1)$$

我们在此将相互作用强度写成 $\lambda/\sqrt{\pi n}$，理由如下。简单的量纲分析表明，在固定物理偶极相互作用强度的情况下，每粒子动能随粒子密度按 n 缩放，而相互作用能则按 $n^{3/2}$ 缩放。随着密度增加，相互作用能不断增大，并在某个与无关的 n取值处，与动能变得相当。此前对偶极

采用 DMC 方法研究玻色子[18,20] 发现，在 $\lambda \approx 30$处发生了从超流体到三角晶体相的一级相变。请注意，相互作用占主导地位

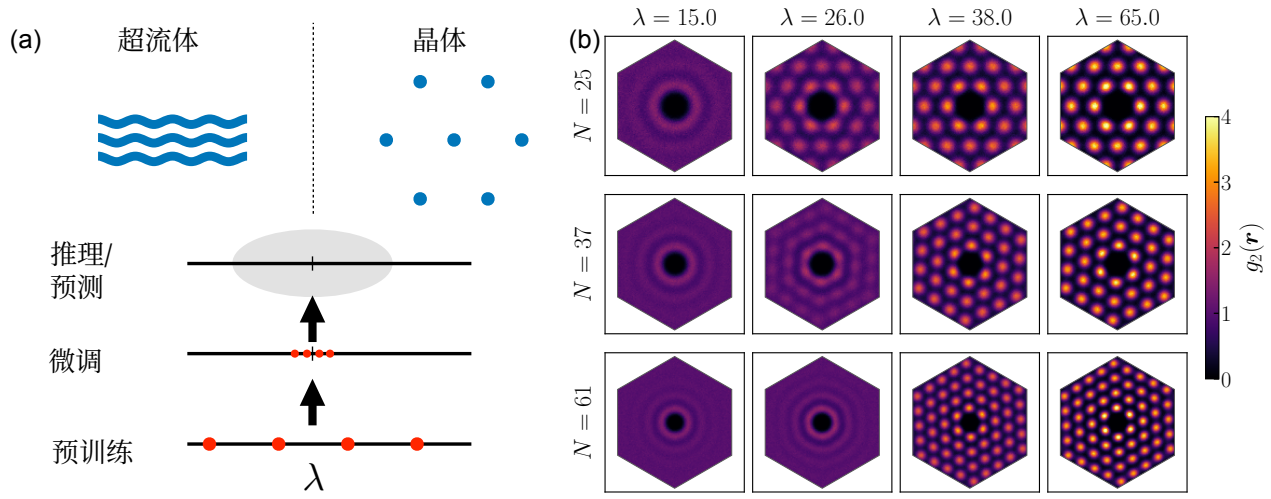


图 1。（a） *Boseformer*的工作流程：首先，在所关注区域内选定的一组相互作用参数 λ 下，通过最小化变分能量对网络进行训练；随后，将其微调至相变附近或其他感兴趣的区域；最后，可用其预测某一区域内参数对应的波函数，而无需针对每个 λ 分别进行变分优化。（b） N 二维玻色子在具有 $1/r^3$ 排斥相互作用时跨越超流体到晶体的相变过程中的对关联函数 $g_2(\boldsymbol{r})$。每一行均通过使用单个预训练模型进行推理生成。

并导致高密度下形成晶体相。这使我们的系统有别于存在库仑相互作用的系统，后者在相反的低密度极限下形成维格纳晶体。

下面，我们将考虑具有周期性边界条件的环面几何。为恢复平移不变性，应将相互作用项提升为对所有周期性映像求和，以确保平移不变性：

$$U = \frac{\lambda}{2\sqrt{\pi n}} \sum_{\mathbf{R}} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{R}|^3}, \tag{2}$$

其中 $\mathbf{R}$ 在由模拟晶胞定义的晶格上求和。具体而言，该求和通过 Ewald求和[23–25]完成。

为了尽可能减小有限尺寸效应，我们让模拟晶胞的几何形状与强耦合下形成的三角晶体相相匹配。具体而言，我们研究由原始晶格矢量 $\boldsymbol{a}_1 = L(1,\ 0)$ 和 $\boldsymbol{a}_2 = L(1/2,\ \sqrt{3}/2)$ 定义的菱形。我们发现，按模拟粒子数 N 缩放其尺寸较为方便，这样粒子密度便能保持在 $n = 1/\pi$。由此可确保超流体相和晶体相具有一致的热力学极限。

由于模型哈密顿量中不存在周期势（1），模拟晶胞与晶体晶胞之间的整比性完全由算术整比判据决定。具体而言，当粒子数为 L¨奥斯基数[26,27], 时，菱形晶胞与三角晶格整比。

$$N = m^2 + mn + n^2, \qquad m, n \in \mathbb{Z}, \tag{3}$$

我们将在计算中使用它。

当某些 i、j 对应的 $\mathbf{r}_i$ 接近 $\mathbf{r}j$ 时，式 (1) 中的粒子间相互作用会变得奇异，这可能给数值方法带来困难。标准做法是在变分波函数中加入 Jastrow因子[28]，以抵消相互作用中最奇异的部分，并避免浪费神经网络容量来刻画粒子位置接近时的关联空穴。为了得到合适的 Jastrow前因子，我们采用二维中粒子在 λ/r^3 势能作用下的零能二体散射问题的解，$f_2(r) = K_0(2\sqrt{\lambda/r})$[18]。从二体波函数中提出 $f_2(r)$，可得到 Jastrow 的动能贡献 $-\nabla^2 \log f_2 - (\nabla \log f_2)^2$,，它会抵消当 $r \to 0$ 时的主要 λ/r^3 发散。我们使用的确切 Jastrow因子形式如下：

$$J_\lambda(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = \prod_{i<j} K_0\left(2\sqrt{\frac{\lambda}{\rho_{ij}}}\right), \tag{4}$$

其中 ρ_{ij} 是第 III 节中引入的光滑周期距离。在粒子间距离较小时，它接近$|\boldsymbol{r}_i - \boldsymbol{r}_j|$，同时在整个周期性晶胞内保持光滑，从而确保拉普拉斯算子的数值计算能够平滑进行。

III. 神经网络架构

本节将详细介绍我们工作中使用的神经网络架构。它以一组粒子坐标 $R = (\boldsymbol{r}_1,\ \ldots\ldots,\ \boldsymbol{r}_N)$ 为输入，

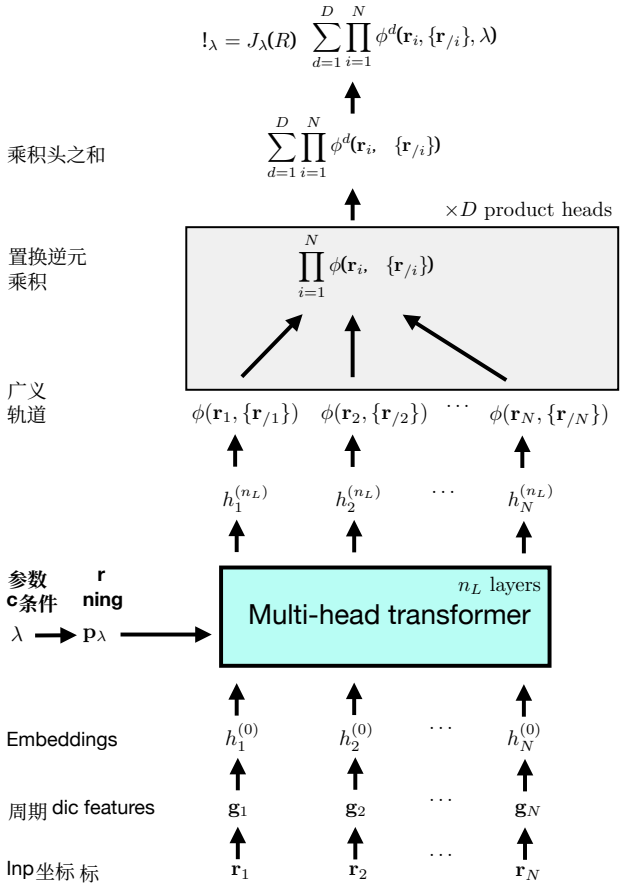


图 2. 表示 N 个玻色子粒子的多体波函数的神经网络架构。每个粒子的坐标 $\mathbf{r}_i$，以及一些哈密顿量参数 λ，构成模型的输入。与每个粒子对应的数据流以等变方式交换信息，见第 III 节。因此，网络输出一个置换不变的波函数，体现了玻色统计。

以及相互作用强度 λ。输出为波函数振幅的对数 $\log\psi$。由于玻色统计，$\psi_\lambda(R)$ 应在任意粒子坐标置换下保持不变，这可通过使用等变神经网络来实现。

下面，为了具体说明，我们给出二维粒子的表示，因为本文重点研究二维情形；推广到其他维度也很直接。

我们首先为每个粒子（索引为 i）生成周期性输入特征，以体现环面几何：

$$g_i^{1,a} = \sin \boldsymbol{G}_a \cdot \boldsymbol{r}_i, \tag{5a}$$
$$g_i^{2,a} = \cos \boldsymbol{G}_a \cdot \boldsymbol{r}_i, \tag{5b}$$

其中 $a = 1,2$ 枚举了模拟晶胞的原始倒格矢 $\boldsymbol{G}_a$。此外，我们引入粒子[4]之间的周期性距离

用于生成周期性版本的 Jastrow 因子；详情请参见附录 A。

作为神经网络的输入，我们将 $g^{m,a}_i$（索引值 $m = 1, 2$，$a = 1, 2$）组合成一个四维向量$\boldsymbol{g}_i$。然后，通过一个维度为 $d \times 4$ 的线性变换 W_{emb}，将其映射到嵌入空间（维度为 d）：

$$\boldsymbol{h}_i^{(0)} = W_{\text{emb}}\boldsymbol{g}_i + \boldsymbol{b}. \tag{6}$$

每个粒子 i 的嵌入向量 $\boldsymbol{h}_i^{(0)}$ 进入 Transformer 的残差流。

神经网络架构的骨干网络是一种 Psiformer 类型的[3] Transformer 网络，由 n_L层组成，每层包含多头缩放点积注意力（MHSA）和多层感知机（MLP），详情见附录 A。值得注意的是，我们使用置换等变 Transformer，这意味着每个粒子 i 的流会根据其他粒子集合中的信息进行更新 $\{\mathbf{r}_{/i}\}$，且更新方式在对 $\{\mathbf{r}_{/i}\}$进行任意置换时保持不变。这对于确保波函数具有玻色统计至关重要。为了使模型输出依赖于相互作用强度 λ，我们在每个 Transformer 层中引入与可学习的相互作用标记 $\boldsymbol{p}_\lambda$的耦合。后者通过一个小型 MLP 在每个网络中生成一次：

$$\boldsymbol{p}_\lambda = W_2 \tanh\left(W_1\lambda + b_1\right) + b_2. \tag{7}$$

Transformer 的输出用于生成 D 广义轨道 ϕ_d，这些轨道具有置换等变性。这意味着每个轨道都是 N 个坐标向量的函数：它可以任意依赖第一个自变量，而对其余 $N-1$ 个自变量进行任意置换 σ 都不会改变函数值。网络以第 i 个粒子的坐标作为第一个自变量，生成每个 D 个轨道的值 $\phi_d\left(\mathbf{r}_i, \{\mathbf{r}_{/i}\}\right)$。置换不变性表示为：

$$\phi_d(\mathbf{r}_i, \{\mathbf{r}_{/i}\}) = \phi_d(\mathbf{r}_i, \sigma\{\mathbf{r}_{/i}\}). \tag{8}$$

λ条件下的轨道值通过对最终残差流值 $\boldsymbol{h}^{(n_L)}{}_i$ 应用的 D 个独立线性变换 W^d_{orb} 获得，适用于每个粒子：

$$\log \phi^d(\boldsymbol{r}_i, \{\boldsymbol{r}_{/i}\}, \lambda) = W^d_{\text{orb}}\boldsymbol{h}_i^{(n_L)}. \tag{9}$$

最后，系统的多体波函数如下获得：

$$\psi_\lambda(R) = J_\lambda(R)\sum_{d=1}^{D}\prod_{i=1}^{N}\phi^d(\boldsymbol{r}_i, \{\boldsymbol{r}_{/i}\}, \lambda), \tag{10}$$

其中 $J(R)$ 是粒子间的 Jastrow前因子（4）。轨道值的对数在一个注意力头内求和，然后使用标准的对数和指数技巧合并各个注意力头。由于 ϕ_i^d 的等变性，输出（10）满足玻色统计，即在粒子坐标的任意置换下保持不变。重要的是，如果轨道 ϕ_i^d 仅依赖于 $\boldsymbol{r}_i$

而不依赖其他粒子的坐标（注意力关闭时便会如此），那么它们的乘积就对应于处于状态 ϕ^d 的理想玻色凝聚体。使用完整的广义轨道$\phi^d(\boldsymbol{r}_i, \{\boldsymbol{r}_{/i}\}, \lambda)$ 可以超越这一点，同时保留有效表征超流相的能力。

至此，我们完成了对本研究所使用神经网络架构的介绍。其实际的 JAX 实现可在代码仓库中找到[29]，，该实现基于 Psiformer 代码构建[3,30]。网络参数列于附录I的表A中。

IV. 训练过程

本节将介绍上述神经网络的无监督训练过程。对于每个粒子数，我们分别训练一个 λ参数化模型$N = 7$: 19、25、37 和 61。为优化网络参数，我们通常采用变分蒙特卡罗方法。由于我们的目标是训练一个将相互作用强度 λ 作为输入之一的参数化模型，因此训练会针对一组相互作用强度 Λ_{tr}同时进行。损失函数为：

$$\mathcal{L} = \frac{1}{|\Lambda_{\mathrm{tr}}|} \sum_{\lambda \in \Lambda_{\mathrm{tr}}} E_\lambda^{\mathrm{var}} \to \min_{\{\theta\}}, \tag{11}$$

其中 $\{\theta\}$是所有网络参数，而每个 λ 对应的变分能量为

$$E_\lambda^{\mathrm{var}} = \frac{\langle \psi_\lambda(\theta) | \mathcal{H}(\lambda) | \psi_\lambda(\theta) \rangle}{\langle \psi_\lambda(\theta) | \psi_\lambda(\theta) \rangle}. \tag{12}$$

如果神经网络能够精确表示每个 λ处的基态，那么损失函数的最小值将对应于所有 λ 在 Λ_{tr}中的基态的理想表示。在我们的变分方法中，我们期望网络能够学习一系列 λ的波函数近似表示，其中可能包含相变。

我们选择一组 λ （见表 I），使其包含相变两侧的点。所得网络已经能够生成超流体和晶体波函数。然而，为了提高准确性，我们发现，在相变附近更窄的区域内对所得网络进行微调很有用， Λ_{ft}。在推理时，我们根据基础模型或微调模型中变分能量较低者进行选择。此外，我们还训练了一个单独的λ模型，以检查条件模型的性能，并将其与针对每个 λ分别进行优化的设置进行比较。

变分蒙特卡罗优化的常规细节见参考文献3。使用Metropolis马尔可夫链蒙特卡罗（马尔可夫链蒙特卡罗）对蒙特卡罗行走者进行采样[31]，采样所依据的密度由 $|\psi_\lambda(R)|^2$给出，其中 ψ_λ 取自神经网络输出。由于每个 λ 都对应其自身的波函数，因此我们将所有行走者划分为训练集中 λ 的每个取值对应一组。为控制离群值，我们采用了标准的截断和重新居中流程，但按组分别处理，并使用固定的 λ，因为具有不同 λ 的构型能量可能存在显著差异。

二阶 KFAC 优化器[32] 用于优化模型参数 $\{\theta\}$。我们采用了递减的学习率 $\eta(t) = 0.02(1 + t/t_0)^{-\gamma}$，其中 $\gamma = 1.5$。有关参数值的更多细节见表 I。优化结果表明，损失函数在优化过程中持续下降，随后趋于平稳，见附录中的图 6 。我们将在下一节研究所得模型，并展示它能够在较小的 λ 下成功预测超流相，在较大的 λ[33] 下成功预测晶体。

V. 相变过程中的可观测量

本节将介绍如何计算由训练完成的网络得出的物理量，以识别并刻画其在每个相互作用强度下所预测的量子序。由于我们的方法能够显式地产生基态波函数，计算可观测量非常直接，只需进行一次蒙特卡罗推理即可。蒙特卡罗过程与训练期间相同，但不更新网络参数。对于作为网络输入给定的相互作用值 λ，我们仅运行数千步的马尔可夫链蒙特卡罗采样，从 $|\psi_\lambda(\boldsymbol{r})|^2$ 中抽取构型，并在这些构型上计算可观测量。通过这种方式，我们从多个可观测量中获得超流体—晶体相变的证据，而所有这些可观测量都由一个横跨两个相的网络计算得到。

A. 能量及其在动能和势能之间的划分

变分蒙特卡罗能量被用作网络训练期间的损失函数。在附录 A（图 6）中，我们展示了其对训练迭代次数的依赖关系，结果表明该能量单调下降并趋于饱和。

在图 4 （绿色曲线）中，我们绘出了所得变分能量随相互作用强度的变化关系， λ，对应于一个含有 $N = 19$ 个玻色子的系统。该能量通过对训练好的模型运行推理马尔可夫链蒙特卡罗链（不更新网络参数），并在平衡后对最后几千次迭代取平均而得。在图中，我们将能量除以 $Nn^{3/2}\lambda$ 这一归一化因子的选取依据是晶体相的经典能量表达式，$E_{\mathrm{cl}} = 4.45Nn^{3/2}\lambda$，该表达式以点划线表示。我们观察到，缩放后的能量 $E/(Nn^{3/2}\lambda)$ 随 λ 减小，但仍高于经典晶体能量。

我们将本方法得到的能量与扩散蒙特卡罗（DMC）方法进行基准比较。我们的 DMC 计算结果复现了文献 18中的结果，具体如下

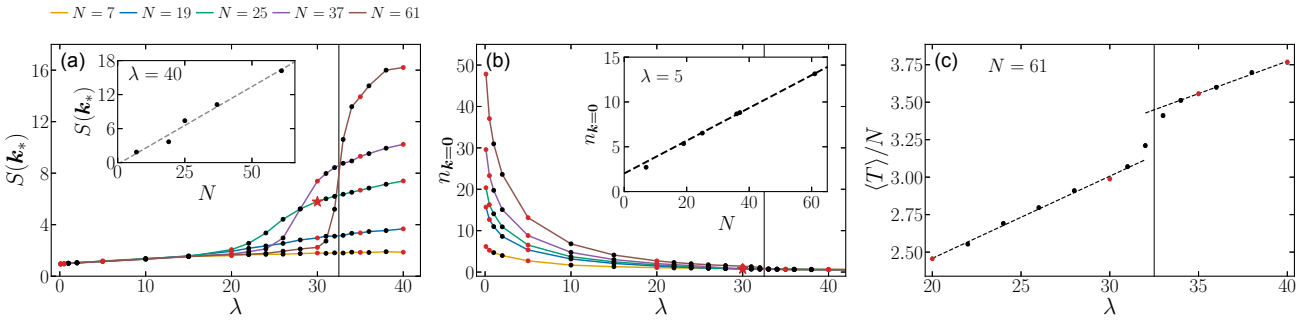


图 3。（a）对于不同粒子数 N，布拉格峰高度 $S(\boldsymbol{k}_*)$ 随相互作用强度的变化。红色星号表示一次独立的单次-λ 优化结果，展示了其与参数化模型的一致性。插图：在 $\lambda = 40$ 处，布拉格峰高度随 N 的标度关系，显示出线性依赖关系 $S(\boldsymbol{k}_*) \propto N$，符合晶体相中的预期。（b）对于不同粒子数 N，跨越相变时凝聚态占据数 $n_{\boldsymbol{k}=0}$ 随相互作用强度 λ 的变化。插图：在 $\lambda = 5$ 处，$n_{\boldsymbol{k}=0}$ 随粒子数呈线性标度关系，表明处于超流相。（c）在 $N = 61$ 处，系统动能随相互作用强度 λ 的变化。该依赖关系在超流体到晶体的相变附近发生跳变。虚线表示远离相变、各相内部的线性拟合。在各面板中，红点表示网络训练期间使用的 λ 值（Λ_{tr} 和 Λ_{ft}），黑点则表示未重新优化时的模型预测。在 $\lambda = 32.5$ 处的细竖线表示根据 *Boseformer* 得到的约 $N = 61$ 相变位置，该位置在三个面板中保持一致。

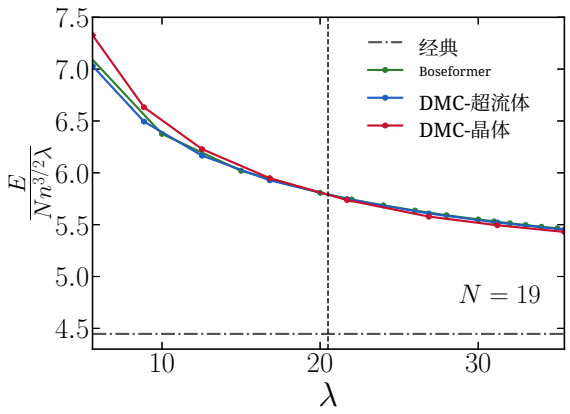


图 4. 系统的基态能量，其中包含 $N = 19$ 个玻色子，随相互作用强度 λ 的变化。绿色显示变分蒙特卡罗对 *Boseformer* 的估计，并与扩散蒙特卡罗（DMC）方法的预测进行比较，数据引自参考文献 18；蓝线（红线）分别表示超流相（晶体相）中的 DMC 能量。两种方法的结果在 1% 的精度范围内一致。黑色点划线表示晶体的经典能量，竖直虚线表示超流与晶体 DMC 分支的交点。

图 4 所示系统的结果以蓝色（超流体分支）和红色（晶体分支）表示。对于 $N = 19$ 个粒子，超流体到晶体的一级相变发生在$\lambda = 20.5$，这不同于 $N \to \infty$ 的外推值 $\lambda = 30$。与 DMC 不同，我们的方法不会分别生成超流体分支和晶体分支，而是经过训练直接生成能量最低的在每个 λ下占优的相。结果表明，总体而言，我们的方法所得能量与 DMC 计算的结果相符，精度约为 1%。在图 5(a) 中，我们放大相变区域，并以纯 Jastrow 波函数 $\psi(R) = J_\lambda(R$ 的基准能量归一化；该波函数是超流相的一种最简考虑相互作用的试探函数，只有一个经过优化的变分参数 R_0。可以看到，在超流相中，DMC 与 *Boseformer* 的结果更为一致；而在晶体相中，*Boseformer* 的能量略高，并更接近 DMC 的超流分支。不过，正如结构因子所示，*Boseformer* 成功描述了有序结构向晶体的重构。尽管 *Boseformer* 的能量预测较为准确，它还能够直接生成波函数，用于计算任意可观测量。

超流体到晶体的相变并未在总能量关于λ的依赖关系中直观地显现出来，该结果由*Boseformer*获得。然而，通过分别绘制总能量中的动能和势能部分，可以观察到相变。在图3(c)中，我们展示了每粒子动能 $\langle T \rangle / N$ 随 λ 变化的曲线，系统中含有 $N = 61$ 个粒子。从图中可以看出，在 $\lambda \approx 32$处，动能发生跃变，表明发生了相变。

B. 成对关联函数

接下来，我们研究玻色子的成对关联函数 $g_2(\boldsymbol{r})$。它非常适合用于观察向晶体相的转变，应当用它替代MCMC平均的普通粒子密度 $\rho(\boldsymbol{r})$。事实上，由于平移不变性，后者在晶体相中可能仍保持均匀

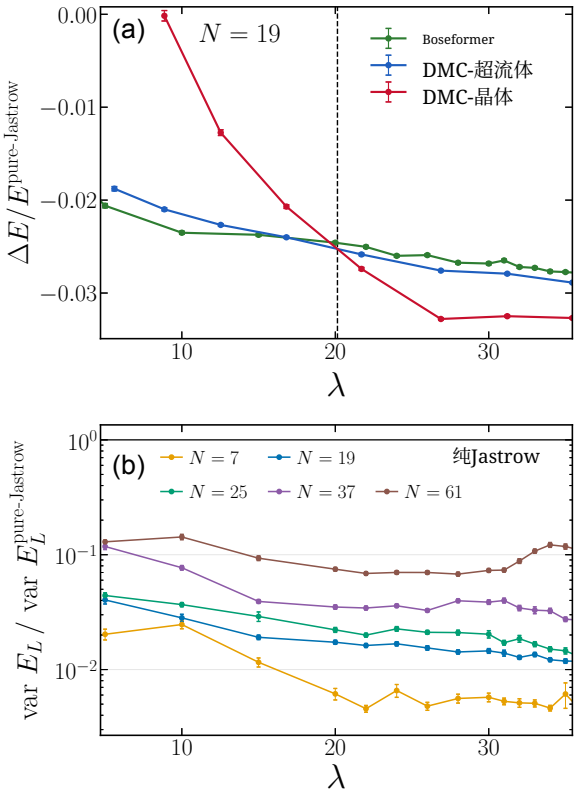


图 5。（a）对变分能量差进行放大显示 ΔE，即我们的方法与单参数纯Jastrow试探波函数（4）之间的差异，以绿色显示。 *Boseformer* 与通过扩散蒙特卡罗（DMC）方法获得的超流体（蓝色）和晶体（红色）分支的能量进行比较。（b）局域能量方差（见第 VE节），以纯Jastrow波函数基线的局域能量方差归一化。

于周期性模拟胞中。与之相反，即使对涌现晶体的随机平移取平均， $g_2(\mathbf{r})$ 仍能展现晶体结构。我们将其定义为

$$g_2(\mathbf{r}) = \frac{|A|}{N(N-1)} \left\langle \sum_{i \neq j} \delta(\boldsymbol{r} - \boldsymbol{r}_{ij}^{\min}) \right\rangle, \qquad (13)$$

其中， A 是模拟胞的面积， $\delta(\boldsymbol{r})$是模拟环面上的周期性狄拉克δ函数， $\boldsymbol{r}_{ij}^{\min}$是粒子 i 与j之间的最小镜像距离。对于理想气体中不相关的粒子，归一化对应于 $g_2(\boldsymbol{r}) = 1$ 。实际上， $g_2(\mathbf{r})$是在MCMC推断运行的最后一次迭代中，将 $\boldsymbol{r}_{ij}^{\min}$计算为直方图得到的。

不同玻色子数下的结果 N 以及不同相互作用强度下的结果，以表格形式展示于图1(b) 中，其中 λ 和 N取不同值。显而易见，图中展示了随着 λ增大而发生的超流体到晶体的相变；随着 N 增大，相变变得更加显著。在超流体中，唯一的特征是 $g_2(\boldsymbol{r})$ 在 $\boldsymbol{r} = 0$ 处出现一个凹陷这是由玻色子间排斥所致。随着 λ 增大，首先可以观察到各向同性的同心密度调制，随后它们呈现六边形畸变和颗粒状特征，最终在 λ较大时发展成清晰可见的三角晶格图样。在下一小节中，我们将通过研究结构因子对这一过程进行定量描述。

C. 结构因子与布拉格峰

作为下一个观测量，我们计算静态结构因子 $S(\boldsymbol{k})$，即傅里叶空间中的密度-密度关联函数。它可以通过对蒙特卡罗游走者进行如下平均来估计：

$$S(\boldsymbol{k}) = \frac{1}{N} \left\langle \left| \sum_{j=1}^{N} e^{i\boldsymbol{k}\boldsymbol{r}_j} \right|^2 \right\rangle. \qquad (14)$$

准动量 $\boldsymbol{k}$ 取自模拟晶胞的倒格矢，即 $\boldsymbol{G}_{ij} = i\boldsymbol{G}_1 + j\boldsymbol{G}_2$。在存在晶体序时，结构因子会形成由晶体几何结构和晶格常数决定的布拉格峰。我们在该设置中得到的晶体具有三角晶格，并且与模拟晶胞相容。因此，存在六个主要的布拉格峰，它们与 $\boldsymbol{k} = 0$点等距，并通过 a C_6 旋转对称性相互关联。为估计主要布拉格峰 $S(\boldsymbol{k}_*)$ 的高度，我们在第一个相容的三角倒空间壳层上的 $S(\boldsymbol{k})$ 中寻找最高极大值，并对位于 $\boldsymbol{k}$点、由绕原点 [的 C_6旋转得到的六个值取平均，因此形式上有 $S(\boldsymbol{k}_*)^{\text{estim.}} = (1/N_{\mathcal{R}}) \sum_{\mathcal{R}} S(\mathcal{R}\boldsymbol{k}_*)$]。

不同粒子数 N 下的 $S(\boldsymbol{k}_*)$结果见图 3(a)，显示出明显的相变行为——对于足够大的 N， $S(\boldsymbol{k}_*)$在λ 处、数值为 20–40 的范围内急剧增大。值得注意的是，转变的陡峭程度以及大 λ时 $S(\boldsymbol{k}_*)$的值均随 N增大。后者表明系统处于晶体相，此时 $S(\boldsymbol{k}_*)$的值是广延量，应随 N线性增长。为检验这一标度关系，我们在图 3(a)插图中绘出了$S(\boldsymbol{k}_*)$对 N 的依赖关系，其中 $\lambda = 40$ ，结果确实显示出线性关系。

D. 凝聚态占据数 $n_{k=0}$

接下来，我们估算零动量凝聚态占据数 $n_{\boldsymbol{k}=0}$，它表示超流体的非对角长程有序。该量可通过单体密度矩阵[16,17]的零动量傅里叶谐波表示

$$\rho_1(\boldsymbol{r}, \boldsymbol{r}') = \frac{N}{\langle \psi | \psi \rangle} \int d\boldsymbol{r}_2 \ldots d\boldsymbol{r}_N \qquad (15)$$
$$\psi^*(\boldsymbol{r}, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)\, \psi(\boldsymbol{r}', \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N).$$

代入 $\boldsymbol{r}' = \boldsymbol{r} + \boldsymbol{s}$ 并提出 $|\psi(R)|^2$，得到

$$n_{\boldsymbol{k}=0} = \frac{N}{A}\int d\boldsymbol{s}\left\langle \frac{\psi(R_{\boldsymbol{s}})}{\psi(R)}\right\rangle, \tag{16}$$

其中 R 是粒子坐标数组，$R_{\boldsymbol{s}} = (\boldsymbol{r}_1 + \boldsymbol{s}\ \boldsymbol{r}_2\ \ldots\ \boldsymbol{r}_N$ $\langle\ldots\rangle$ R ～、 、 、) 表示对 ( ) 取平均。我们在蒙特卡罗游走者上对其进行评估，并在晶胞内均匀采样位移向量 $\boldsymbol{s}$ 。$|\psi\ R\ |^2$

我们在图 $n_{\boldsymbol{k}=0}$ 3(b) 中展示了结果。我们发现，在较小的 λ下，凝聚态占据数非零，并随粒子数 N线性增长，这表明系统处于超流相。在 $\lambda \approx 25$–35 时，凝聚态占据数骤降，并且在更大的 λ 下仍严重耗尽，这与晶体相一致。所得结果没有显示出超流性与结晶化共存的超固态相的证据。

E. 局域能量的方差

这里我们研究局域能量 $E_L = \mathcal{H}\psi/\psi$的涨落，以刻画变分方法的收敛性。对变分波函数$|\psi(R)|^2$所定义的分布上的 E_L 取平均，得到该态的哈密顿量平均值 $\langle E_L\rangle = E^{\mathrm{var}}$；而 E_L 的方差则提供了变分波函数优劣的信息。事实上，对于真实基态（或能量为 E的任意哈密顿量本征态），$E_L = E$ 恒等于零，且方差为零。因此，变分近似的质量通常用 V评分[34, 35], $V_0 = N\mathrm{var}(E_L)$ $/\langle E_L\rangle^2$来表征；该评分将方差按变分能量归一化，并额外乘以因子 N，以适用于非广延热力学极限。

然而，在我们的情形中，超流体的基态能量在 $\lambda \to 0$ 处趋于零，因此在此极限下按能量归一化并无参考意义。我们改为按纯Jastrow波函数的方差 $\psi(R) = J_\lambda(R)$进行归一化；此前我们已将其用作变分能量的基准。我们在图5(b)中给出了方差 E_L/var $E_L^{\text{pure-Jastrow}}$ var 的结果。我们发现，与基准相比，我们的方法使局域能量的方差降低了1至3个数量级，这表明其近似效果显著更好。

VI. 讨论

在本研究中，我们提出了一种适用于相互作用玻色子的通用神经量子波函数，并展示了其在一种具有 $1/r^3$ 排斥相互作用[20]的典型玻色子系统中的应用。我们的结果一方面验证了该方法及神经网络架构的有效性，另一方面独立证实了文献18.所预测的从超流体到晶体的量子相变。

在实验上，该物理模型可应用于TMD材料中具有面外偶极矩的层间激子[21,22]。

我们将模型的性能与扩散蒙特卡罗（DMC）进行了比较。DMC是一种强大的方法，能够使包含多达 $N \sim 10^3$ 个粒子的系统收敛到基态，但也有其局限性：它无法显式表示多体波函数；对于可观测量，只有能量以及与哈密顿量对易的可观测量，其自然估计才是无偏的。要用DMC研究超流体到晶体的相变，必须使用专门制备的试探波函数，这会使该方法偏向两个相中的一个。相比之下，*Boseformer*经过训练，可在每个 λ直接预测最优波函数。它成功捕捉了相变，并提供了一种适用于相变全过程的通用试探形式。在一级相变处，波函数的跳变被单个预训练网络所产生的陡峭交叉平滑化。随着粒子数增加，交叉会变得更加陡峭。尽管网络容量可能会限制相变的陡峭程度，即使在热力学极限下也是如此，但目前的表现已足以观察和研究相变。

玻色子系统中一个引人入胜的理论可能性，是超固态的存在；它兼具非零超流刚度和晶体有序。先前的研究[36] 得出结论：若无额外扰动，哈密顿量（1）不会出现这种相。我们的研究未发现任何超固态，证实了这一结论。然而，要在我们的方法框架内可靠地否定超固态假说，还需要将研究扩展到规模更大的系统 N 。更广泛地说，未来研究方向可包括将计算模拟扩展到粒子数更多的系统，并将该方法应用于各种玻色子哈密顿量。

AI 使用声明

Boseformer 架构的初始实现由一位作者（D. A.）完成，未借助 AI 编程工具。随后，作者使用了 Claude 和 ChatGPT 的多个模型，进一步开发、运行并分析*Boseformer*的结果。作者对科学内容和写作承担责任。

致谢

D. A. 感谢 Khachatur Nazaryan 和 Timothy Zaklama 提供的宝贵讨论和意见。本研究在耶鲁大学开展，得到了 NSF资助项目 DMR-241018 的支持。

附录A：神经网络架构与优化过程的详细信息

这里我们将详细介绍本研究所用神经网络的架构。本研究使用的具体参数值见表 I。

1. 周期性粒子间距

我们在工作中使用的变分波函数中的 Jastrow前因子（4）通过周期性距离定义，

$$\tilde{r}_{ij} = \frac{1}{2\pi}\sqrt{\boldsymbol{\alpha}^T\Gamma\boldsymbol{\alpha} + \boldsymbol{\beta}^T\Gamma\boldsymbol{\beta}}, \tag{A1}$$

其中，度量 $\Gamma = A^T A$ 由模拟晶胞的实空间晶格矢量 $A = [\boldsymbol{a}_1|\boldsymbol{a}_2]$表示，而相位分量则通过倒易矢量 $\boldsymbol{G}_a$定义：

$$\alpha_a = 1 - \cos \boldsymbol{G}_a \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j), \tag{A2}$$
$$\beta_a = \sin \boldsymbol{G}_a \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j). \tag{A3}$$

当周期性距离 $\tilde{r}_{ij}$ 相较于周期性晶胞的大小足够小时，它可以近似粒子间距离 $|\boldsymbol{r}_i - \boldsymbol{r}_j|$；对于任意参数值，它在该晶胞中都平滑且周期性。此外，我们还引入了一种截断周期性距离

$$\rho_{ij} = R_0 \tanh \frac{\tilde{r}_{ij}}{R_0}, \tag{A4}$$

其中 $R_0 = \zeta L$ 是调节 ρ_{ij} 在尖点区域之外饱和程度的参数。“匹配分数” ζ 在*Boseformer*中固定，见表I。如上所述， ρ_{ij} 的值用于 Jastrow 前因子，见公式（4），但不会输入神经网络。

2. Transformer架构细节

我们神经网络架构的核心模块是一个等变的 Psiformer 类型[3] Transformer 网络。它有 H 个注意力头，满足关系式 $d = Hd_k$，其中d 是残差流维度， d_k 是每个注意力头的维度。每个粒子都保留自己的残差流 $\boldsymbol{h}_i$，并且仅通过注意力收集其他粒子的信息。除选定粒子的残差流外，所有残差流值构成的无序集合记为 $\{\boldsymbol{h}_{/i}\}$。注意力和 MLP 模块的权重在各粒子间共享，但每一层都有各自的注意力权重和 MLP 权重。这些特性共同确保了等变性：对于任意粒子，对其余粒子进行任意置换都不会改变该粒子的残差流。因此，注意力表示粒子之间的通信，而 MLP 则在每个流内部进行计算，具体流程如下

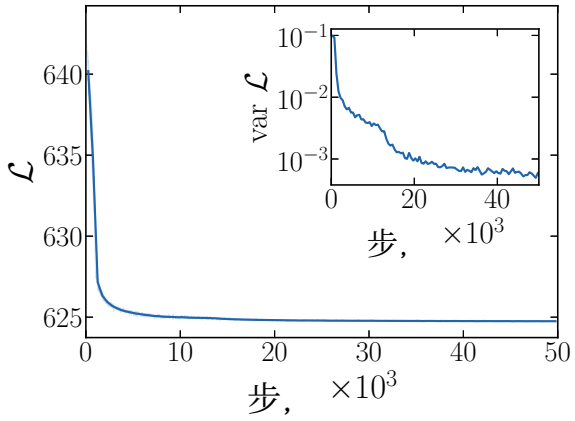


图 6. $N = 25$ 个粒子训练期间的损失函数（行走者上的平均局域能量）。结果显示损失单调下降并趋于饱和。插图：批次平均损失在迭代次数中的运行方差，权重随各次迭代的时间间隔呈指数衰减。优化过程中，该方差降低了两个数量级。

每个流中的 tation。流程如下：

低位：

$$\boldsymbol{h}_i \to \mathrm{LN}\left(\boldsymbol{h}_i + \mathrm{MHSA}(\boldsymbol{h}_i, \{\boldsymbol{h}_{/i}\})\right), \tag{A5}$$
$$\boldsymbol{h}_i \to \mathrm{LN}\left(\boldsymbol{h}_i + \mathrm{MLP}(\boldsymbol{h}_i)\right), \tag{A6}$$
$$\boldsymbol{h}_i \to \boldsymbol{h}_i + \tanh W_l \boldsymbol{p}_\lambda, \tag{A7}$$

其中，LN 是后残差层归一化[37]，用于归一化每个粒子的残差流。多头自注意力计算每个注意力头的输出

$$a_i^k = \sum_j \mathrm{softmax}_j \left[\frac{(W_Q^k \boldsymbol{h}_i)\cdot(W_K^k \boldsymbol{h}_j)}{\sqrt{d_k}}\right] W_V^k \boldsymbol{h}_j, \tag{A8}$$

使用维度为 $Hd_k \times d$ 的查询、键和值矩阵（W_Q、 W_K 和 W_V），其中包含分别对应每个注意力头的块（ W_Q^k、 W_K^k 和 W_V^k）。所得的注意力头输出经过拼接后，使用矩阵 W_O 投影回残差流维度

$$\mathrm{MHSA}(\boldsymbol{h}_i, \{\boldsymbol{h}_{/i}\}) = W_O[a_i^1, \ldots, a_i^H], \tag{A9}$$

，然后将其加到残差流中。MLP 模块由一个两层的 tanh激活感知器表示，该感知器限制在每个粒子的独立残差流中。Psiformer 的一项重要扩展是加入操作（A7），该操作使模型以哈密顿量参数：相互作用强度为条件λ。如正文所述，相互作用标记 $\boldsymbol{p}_\lambda$通过可训练线性变换获得，并在整个网络中共享。在实际应用方面，注意力部分使用 JAX 的融合函数jax.nn 点积注意力 以提高效率。

神经网络和优化器的参数见表I。优化过程中损失函数的典型行为表现为单调收敛；见图6。

表 I。本文所用 *Boseformer* 神经网络和优化器的参数。

组	参数	值
骨干网络	嵌入维度 d	256
	层数 n_L	4
	注意力头数 H	4
	注意力头维度 d_k	64
	MLP 隐藏维度	(256,)
	轨道数 D	16
λ-条件	λ-嵌入隐藏维度	64
	λ 令牌的维度（$\boldsymbol{p}_\lambda$）	256
Jastrow	匹配分数 ζ	1.0
	网络参数总数。	$1.9 \cdot 10^6$
Ewald 求和。实空间壳层数 N		4
	倒易空间壳层数 N	12
多-λ	训练集 Λ_{tr}	[0.1, 0.5, 5, 20, 50, 80]
	微调集 Λ_{ft}	[20, 25, 30, 35, 40]
MCMC	采样器	随机游走 Metropolis
	Metropolis 步数/迭代	30
	初始步长	0.02
	自适应频率	100
	预热步数	300
优化器	算法	KFAC
	迭代次数	$5 \cdot 10^4$ (基础)， $2 \cdot 10^4$ (微调)
	批次大小	2048–4096
	学习率	0.02
	学习率延迟 t_0	10^4 （基础）， $3 \cdot 10^4$ （微调）
	学习率衰减指数 γ	1.5
	局部能量裁剪	3.0
	KFAC阻尼	0.01
	KFAC范数约束	0.003

附录B：与扩散蒙特卡罗（DMC）进行基准比较

如第V节所示，我们将我们的方法与扩散蒙特卡罗（DMC）进行比较。在不存在磁场时，对于玻色子而言，DMC 不受符号问题影响。DMC 是该系统的标准选择，其主要优势在于：当蒙特卡罗游走者的数量趋于无穷且时间步长趋于无穷小时，DMC 会收敛至精确的基态能量 $\Delta\tau$。然而，DMC 不会显式提供基态波函数。一般而言，对于能量以外的可观测量，或与哈密顿量对易的算符以外的可观测量，其标准混合估计量都存在偏差。

我们遵循文献18，并使用 $N_W = 2048$ 重要性采样在虚时间中推进蒙特卡罗游走者[38]。类似地，对于每个 λ，我们使用独立的引导波函数和初始构型，以跟踪液体分支和晶体分支。对于超流体，我们使用基于（4）的 Jastrow引导波函数；对于晶体，则额外使用一个将粒子钉扎在三角晶格格点上的高斯因子。为匹配*Boseformer* 设置，我们采用相同的菱形模拟晶胞、密度、粒子数 N和 Ewald 求和过程[20]。我们采用时间步长 $\Delta\tau \sim 10^{-5}\lambda^2$；我们检查了有限 $\Delta\tau$ 会引入关于 $\Delta\tau$呈线性的能量偏差。可以使用 $\Delta\tau \to 0$ 外推来提高 DMC 方法的精度。

值得注意的是，超流体分支与晶体分支的交点强烈依赖于粒子数N。对于 *Boseformer*相关的 $N = 19$–61，拟合得到的分支交点位于 $\lambda = 18$–21，而$N \to \infty$外推（通过拟合$E_N/N = E_\infty + C/\sqrt{N}$依赖关系完成， N最大至 $\sim 10^3$）则再现了 $\lambda_c \approx 30$，与参考文献 18一致。

正如我们在正文中所示（见图 4），对于粒子数为 $N = 7$–61 的系统，我们的方法收敛得到的变分能量通常与 DMC 的结果相差不到 1%。与此同时，只需一个模型，它就能在一系列哈密顿量参数值下提供完整的多体波函数。

* daniil.antonenko@yale.edu

[1] G. Carleo 和 M. Troyer，《利用人工神经网络求解量子多体问题》，《科学》355，602（2017）。[2] D. Pfau、J. S. Spencer、A. G. de G. Matthews 和 W. M. C. Foulkes，《利用深度神经网络从头算求解多电子薛定谔方程》，《物理评论研究》2，033429（2020）。[3] I. von Glehn、J. S. Spencer 和 D. Pfau，《用于从头算量子化学的自注意力试探波函数》，载于 *11th ICLR*（2023）。[4] G. Cassella、H. Sutterud、S. Azadi、N. D. Drummond、D. Pfau、J. S. Spencer 和 W. M. C. Foulkes，《利用费米子神经网络发现量子相变》，《物理评论快报》130，036401（2023）。[5] Y. Teng、D. D. Dai 和 L. Fu，《利用自注意力神经网络求解分数量子霍尔问题》，《物理评论B》111，205117（2025）。[6] M. Geier、K. Nazaryan、T. Zaklama 和 L. Fu，《用于求解固体中关联电子问题的自注意力神经网络》，《物理评论B》112，045119（2025）。[7] G. Pescia、J. Han、A. Lovato、J. Lu 和 G. Carleo，《连续空间周期性系统的神经网络量子态》，《物理评论研究》4，023138（2022）。[8] W. Freitas 和 S. A. Vitiello，《深度神经网络与变分蒙特卡罗方法在小型 ^4He_N 团簇中的协同作用》，Quantum 7，1209（2023）。[9] D. Linteau、G. Pescia、J. Nys、G. Carleo 和 M. Holzmann，《二维氦-4的相图与晶体熔化》，《物理评论快报》134，246001（2025）。[10] H. J. Ryser，*Combinatorial Mathematics*，Carus 数学专著，第14卷（美国数学协会，1963）。[11] M. Zaheer、S. Kottur、S. Ravanbakhsh、B. P´oczos、R. Salakhutdinov 和 A. J. Smola，《深度集合》，载于 *Adv. Neural Inf. Process. Syst.*，第30（Curran Associates, Inc.，2017）卷，第3391页。[12] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、 ̈L. Kaiser 和 I. Polosukhin，《注意力机制就是你所需要的一切》，载于 *Adv. Neural Inf. Process. Syst.*，第30卷（Curran Associates, Inc.，2017），第5998页。[13] Y.-H. Zhang 和 M. Di Ventra，《Transformer量子态：用于量子多体问题的多用途模型》，《物理评论B》107，075147（2023）。[14] T. Zaklama、M. Geier 和 L. Fu，《大电子模型：通用基态预测器》，（2026），arXiv:2603.02346 [cond-mat.str-el]。[15] K. Nazaryan 和 L. Fu，《QERNEL：可扩展的大电子模型》，（2026），arXiv:2604.26018 [cond-mat.str-el]。[16] O. Penrose 和 L. Onsager，《玻色-爱因斯坦凝聚与液氦》，《物理评论》104，576（1956）。[17] S. Giorgini、J. Boronat 和 J. Casulleras，《二维液氦的扩散蒙特卡罗研究》，《物理评论B》54，6099（1996）。[18] G. E. Astrakharchik、J. Boronat、I. L. Kurbakov 和 Yu. E. Lozovik，《二维偶极子系统中的量子相变》，《物理评论快报》98，060405（2007）。[19] H. P. B ̈uchler、E. Demler、M. Lukin、A. Micheli、N. Prokof'ev、G. Pupillo 和 P. Zoller，《冷极性分子构成的强关联二维量子相：Con-

.....控制相互作用势的形状”，《物理评论快报》98, 060404（2007）。[20] C. Mora、O. Parcollet 和 X. Waintal，“偶极玻色子晶体的量子熔化”，Phys. Rev. B 76, 064511（2007）。[21] O. L. Berman 和 R. Ya. Kezerashvili，“过渡金属二硫族化物双层中双组分玻色气体的高温超流性”，Phys. Rev. B 93, 245410（2016）。[22] W. Li、X. Lu、S. Dubey、L. Devenica 和 A. Srivastava，“范德华异质结构中局域化层间激子之间的偶极相互作用”，Nat. Mater. 19, 624（2020）。[23] P. P. Ewald，“光学和静电晶格势的计算”，Ann. Phys. (Berlin) 369, 253（1921）。[24] B. R. A. Nijboer 和 F. W. de Wette，“关于晶格和的计算”，Physica 23, 309（1957）。[25] M. Mazars，“计算机模拟中的长程相互作用及准二维系统”，Phys. Rep. 500, 43（2011）。[26] J. H. Conway 和 N. J. A. Sloane，*Sphere Packings Lattices and Groups*，第3版，数学科学基础丛书，第290卷（Springer，纽约，1999）。[27] R. Connelly 和 W. Dickinson，“周期性平面圆盘堆积”，Philos. Trans. R. Soc. A 372, 20120039（2014）。[28] R. J astrow，“强作用力下的多体问题”，Phys. Rev. 98, 1479（1955）。[29] D. S. Antonenko，“*Boseformer*：用于相互作用玻色子的Transformer 量子态”，https://github.com/andanya/boseformer（2026）。[30] J. S. Spencer、D. Pfau 和 FermiNet Contributors，“FermiNet”，http://github.com/deepmind/ferminet（2020）。[31] N. Metropolis、A. W. Rosenbluth、M. N. Rosenbluth、A. H. Teller 和 E. Teller，“利用高速计算机计算状态方程”，《化学物理杂志》21, 1087（1953）。[32] J. Martens 和 R. Grosse，“利用克罗内克因子近似曲率优化神经网络”，载于 *ICML 2015*，37，《机器学习研究论文集》，第卷（PMLR，2015），第2408页。[33] 虽然对于大多数 N，我们观察到模型在较大的 λ下收敛到相称晶序，但也观察到对于 $N = 36$，系统有时会收敛到带有一个空位的 $N = 37$ 态。随后在更窄的参数区域内进行微调即可解决这一问题。[34] D. Wu、R. Rossi、F. Vicentini、N. Astrakhantsev、F. Becca、X. Cao、J. Carrasquilla、F. Ferrari、A. Georges、M. Hibat-Allah、M. Imada、A. M. L ̈auchli、G. Mazzola、A. Mezzacapo、A. Millis、J. Robledo Moreno、T. Neupert、Y. Nomura、J. Nys、O. Parcollet、R. Pohle、I. Romero、M. Schmid、J. M. Silvester、S. Sorella、L. F. Tocchio、L. Wang、S. R. White、A. Wietek、Q. Yang、Y. Yang、S. Zhang 和 G. Carleo，“量子多体问题的变分基准”，《科学》386, 296（2024）。[35] R. Rende、A. Sinibaldi、L. L. Viteritti、R. Wiersema、A. Georges 和 G. Carleo，“神经网络量子态的标度律”，arXiv:2606.02794（2026），10.48550/arXiv.2606.02794。[36] I. L. Kurbakov、Yu. E. Lozovik、G. E. Astrakharchik 和 J. Boronat，“二维偶极晶体的准平衡超固相”，Phys. Rev. B 82, 014508

（2010年）。[37] J. L. Ba、J. R. Kiros 和 G. E. Hinton，“层归一化”，（2016年），arXiv:1607.06450 [stat.ML]。[38] P. J. Reynolds、D. M. Ceperley、B. J. Alder 和 W. A. Lester, Jr.，《用于分子的固定节点量子蒙特卡罗》，《化学物理杂志》**77**，5593（1982年）。