arXiv:2304.01798v1 [hep-lat] 4 Apr 2023

用于格点场论模拟的局域性约束自回归条件归一化流

Dinesh P. R.[*]

摘要

基于归一化流的采样方法已成功应对传统上与格点量子场论模拟相关的计算挑战。后续研究将作用量积分的规范不变性和平移不变性纳入底层神经网络，从而使这些模型能够高效训练和推理。本文将作用量积分的局域性纳入考虑，由此简化了条件归一化流的输入域；这类流在自回归过程中对恒定时间子晶格进行采样，我们将其称为局域自回归条件归一化流（l-ACNF）。我们发现，在二维晶格上对 ϕ^4 理论进行采样时，l-ACNF 模型的自相关时间比在完整晶格上采样的同类归一化流模型短数个数量级。

1 引言

求解具有较大耦合的量子场论中的路径积分，需要将底层时空离散化为晶格，并使用马尔可夫链蒙特卡罗（MCMC）算法对场进行数值采样——这称为格点量子场论[9]。对于较大的晶格尺寸，以及会导致晶格间距较小、相关长度较大的作用量参数取值，MCMC 方法往往会出现较长的相关时间，导致计算成本呈指数级增长；这一现象称为临界减速（CSD）[17]。尽管已有少数针对特定模型开发的非局域更新算法来应对 CSD [13, 16], , 但它们无法应用于包括量子色动力学（QCD）在内的许多关键理论。

近年来，人们开始探索基于机器学习的方法 [19, 18] ，用于构建晶格上统计场论的生成模型。特别是，一种先训练归一化流、再进行独立 Metropolis-Hastings 抽样的技术 [2] ，已被证明在对二维 ϕ^4 理论建模时，几乎可以消除临界慢化（CSD）。后续研究 [7, 3]提出了在使用归一化流对规范场论建模时，将作用量积分的规范不变性纳入模型的方法。当存在局域 $SU（N）$ 对称性时，这种方法消除了流模型所表示的变分分布中的冗余自由度，并降低了模型训练和推断的计算成本。这解决了机器学习中一个众所周知的问题，通常称为维数灾难[8]： 随着输入／输出数据规模（维数）的增长，训练时间和模型复杂度预计会呈指数级增长。尽管研究表明，深度神经网络在许多情况下能够打破这一灾难[11], , 但这只在目标函数依赖于输入的低维投影这一近似假设下成立[4]。当相关长度与晶格尺寸成正比时，这种近似对于晶格场论的生成模型而言相当糟糕，因此消除冗余自由度至关重要。另一项研究 [6] 探索了在底层分布呈多峰特征（例如对称性破缺相）时，改善归一化流收敛不佳问题的多种方法。该研究还引入了掩码卷积网络，作为耦合层的条件映射，以强制实现（破缺的）平移不变性，并在分布中体现一种局域性概念，使局域依赖更强。

在本研究中，我们提出一种条件归一化流的自回归过程，更严格地强制作用量积分满足局域性。这样一来，我们将条件归一化流的有效输入规模从上界 L^d 降至 $O(L^{d-1})$，从而进一步按照应对

[*]项目助理，印度科学教育与研究学院 浦那分校。电子邮箱： dinesh.pr@students.iiserpune.ac.in

维数灾难的思路消除不必要的自由度。我们将其称为局部自回归条件归一化流 (l-ACNF)。在第 2.1 节简要介绍研究系统和符号后，我们在第 2.3 节详细分析自回归条件分布（第 2.2 节介绍）的依赖关系。第 2.4 节讨论 l-ACNF 模型的数学结构及其配套的 Metropolis-Hamilton 过程。在第 3 节，我们描述了使用 l-ACNF 在二维晶格上对 ϕ^4 理论进行采样的数值实现。

2 数学背景

2.1 标量晶格场论

研究系统由一个边长为 L、位于 d 维空间中的超立方晶格组成，其中每个位置都用一个 d 维向量 $\boldsymbol{r} \in [1, L]^d$ 标记。对于开放边界条件，晶格位置 $\boldsymbol{r} \notin [1, L]^d$ 不存在。对于周期性边界条件，我们将 $\boldsymbol{r} \in \mathbb{Z}^d$ 的每个分量映射为 $r_i \to (r_i - 1) \bmod L + 1$, 。系统的状态/构型由标量场值 $\phi(\boldsymbol{r})$ 描述。构型遵循玻尔兹曼分布：

$$p\left(\phi(\boldsymbol{r})\right) = e^{-S[\phi]}/Z \tag{1}$$

其中，作用量 $S[\phi]$ 是场 ϕ 的泛函，而 Z 是称为配分函数的归一化常数。对于具有 ϕ^4 相互作用的标量场理论，作用量为：

$$S[\phi] = \sum_{\boldsymbol{r} \in [1,L]^d} \left[\phi(\boldsymbol{r}) \sum_{\bar{\boldsymbol{r}}} \square\left(\boldsymbol{r}, \bar{\boldsymbol{r}}\right) \phi(\bar{\boldsymbol{r}}) + m^2 \phi(\boldsymbol{r})^2 + \lambda \phi(\boldsymbol{r})^4 \right] \tag{2}$$

其中 m, λ 是裸质量和耦合。晶格近似中的达朗贝尔算子为：

$$\phi(\boldsymbol{r}) \sum_{\bar{\boldsymbol{r}}} \square\left(\boldsymbol{r}, \bar{\boldsymbol{r}}\right) \phi(\bar{\boldsymbol{r}}) = \sum_{\mu=1}^{d} 2\phi(\boldsymbol{r})^2 - \phi(\boldsymbol{r})\phi(\boldsymbol{r} - \hat{\mu}) - \phi(\boldsymbol{r})\phi(\boldsymbol{r} + \hat{\mu})$$

其中 $\phi(\boldsymbol{r})$可以取任意实数值。

2 中的作用量 $S[\phi]$具有许多有趣的数学性质。除 $\phi(\boldsymbol{r})$ 的幂次项外，它只包含最近邻乘积/相互作用项 $\phi(\boldsymbol{r})\phi(\boldsymbol{r} - \hat{\mu})$ 和 $\phi(\boldsymbol{r})\phi(\boldsymbol{r} + \hat{\mu})$。因此，作用量具有局域性，这也是许多基本物理理论的常见特征。作用量还对变换$\phi(\boldsymbol{r}) \to -\phi(\boldsymbol{r})$以及晶格的各种离散旋转和反射（超立方对称群）保持对称/不变。

2.2 自回归过程

1 中的玻尔兹曼分布取决于大量随机变量 $L^d = N$，通常难以从中采样或直接计算概率密度函数（PDF）。一种规避方法是将其建模为各个随机变量的条件分布之积。首先，将晶格位置的标签从向量 $\boldsymbol{r} \in [1, L]^d$ 改为整数排序排序 $k \in [1, N]$。我们将采用一种特定的排序方式，将向量标签映射为：

$$k(\boldsymbol{r}) = \left(\sum_{i=1}^{d} (r_i - 1) L^{i-1} \right) + 1 \tag{3}$$

利用通过反复应用贝叶斯定理得到的链式法则，我们可以将概率分解为按上述排序排列的 N 条件概率的乘积：

$$\begin{aligned} p(\{\phi_k | k \in [1, N]\}) &= p(\phi_1, \phi_2 \ldots \phi_N) = p(\phi_1) p(\phi_2|\phi_1) \ldots p(\phi_N|\phi_{N-1} \ldots, \phi_1) \\ \implies \log p(\{\phi_k | k \in [1, N]\}) &= \log p(\phi_1) + \log p(\phi_2|\phi_1) + \ldots \log p(\phi_N|\phi_{N-1} \ldots, \phi_1) \quad (4) \\ &= \sum_{k \in [1,N]} \log p(\phi_k|\phi_{<k}) \quad (5) \end{aligned}$$

这被称为自回归过程。如果我们能够按顺序计算并采样条件概率，就可以得到整个玻尔兹曼分布的一个样本及其对数概率。从 1 开始， $p(\phi_k|\phi_{<k})$ 的精确形式在解析上仍然难以处理。因此，采用这一原理的方法通常会优化基于机器学习的变分分布与未归一化玻尔兹曼分布[19]之间的 KL 散度。

2.3 $d-1$ 维局域相互作用的依赖集

伊辛模型与开放边界条件 考察 5 中第 k 个条件概率 $p(\phi_k|\phi_{<k})$，其分布通常取决于 $\phi_{<k}=\{\phi_{k-1},\dots\phi_1\}$ 中的 $k-1$ 个值。这意味着，如果晶格点 N 的数量很大，这些分布的复杂度可能会急剧增加，而这通常正是我们关注的情形。不过，对于具有最近邻相互作用的系统，依赖集 会小得多 [12]。不失一般性，可以通过最近邻伊辛模型更容易地说明这一点，其作用量为：

$$S[\phi]=-\beta J\sum_{\mu=1}^{d}\sum_{\boldsymbol{r}}\phi(\boldsymbol{r}-\hat{\mu})\phi(\boldsymbol{r}) \tag{6}$$

其中 $\phi(\boldsymbol{x})$ 取值为 ±1。目前我们假设采用开放边界条件，之后再放宽这一假设。将伊辛模型的玻尔兹曼分布重述为自回归过程[2]：

$$\prod_{k=1}^{N}p(\phi_k|\phi_{<k})=p(\phi)=\exp\left(-\beta J\sum_{\mu=1}^{d}\sum_{k}\phi_k\phi_{k-\hat{\mu}}\right)/Z \tag{7}$$

根据贝叶斯定理，我们可以将该条件概率与前 k 个和 $k-1$ 个自旋的无条件联合概率联系起来，而后者又可写成玻尔兹曼分布的积分形式：

$$p(\phi_k|\phi_{<k})=\frac{p(\phi_1,\dots\phi_k)}{p(\phi_1,\dots\phi_{k-1})}=\frac{\sum_{\phi_{k+1}\dots\phi_N}p(\phi)}{\sum_{\phi_k\dots\phi_N}p(\phi)}$$

展开伊辛模型的 $p(\phi)$：

$$p(\phi_k|\phi_{<k})=\frac{\sum_{\phi_N,\dots\phi_{k+1}}\exp\left(-\beta J\sum_{l=k}^{N}\left(\phi_l\sum_{\mu}\phi_{l-\hat{\mu}}\right)+\delta(\phi_{<k})\right)}{\sum_{\phi_N,\dots\phi_k}\exp\left(-\beta J\sum_{l=k}^{N}\left(\phi_l\sum_{\mu}\phi_{l-\hat{\mu}}\right)+\delta(\phi_{<k})\right)}$$

由于 $\phi_{<k}$ 中的值是固定的，不参与求和，因此仅包含这些值的项 $\delta(\phi_{<k})$ 会在分子和分母中相互抵消，得到：

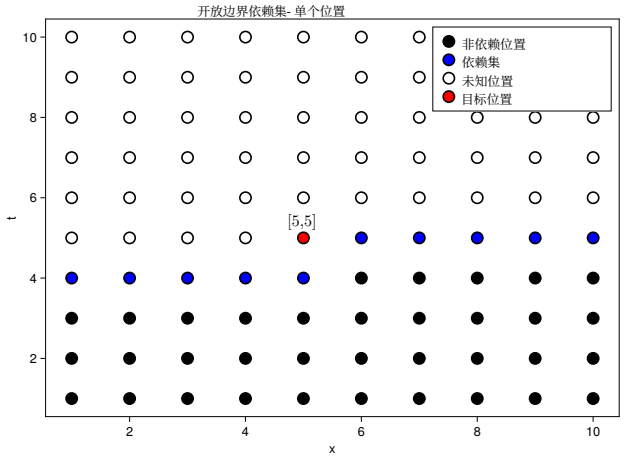
$$p(\phi_k|\phi_{<k})=\frac{\sum_{\phi_N,\dots\phi_{k+1}}\exp\left(-\beta J\sum_{l=k}^{N}\phi_l\sum_{\mu}\phi_{l-\hat{\mu}}\right)}{\sum_{\phi_N,\dots\phi_k}\exp\left(-\beta J\sum_{l=k}^{N}\phi_l\sum_{\mu}\phi_{l-\hat{\mu}}\right)} \tag{8}$$

尽管 $\phi_{<k}$ 包含 $k-1$ 个值，条件概率 $p(\phi_k|\phi_{<k})$ 仅取决于 $\phi_{<k}$ 中那些与 $\phi_{\geq k}$ 中位置互为最近邻的位置。对于标量格点场论，将求和替换为积分，并在上述表达式中加入 ϕ_l^2 和 ϕ_l^4 之类的项，也可以得出相同的结论。对于我们所选的排序，依赖集中的元素数量上限为 L^{d-1} 或 N/L（参见图 1a 中 10×10 2 维晶格上的示意图），比原先的上限 N 小多个数量级。事实上，我们可以将图 1a 中两条蓝色自旋条带合并为一条长度为 $L^{2-1}=10$ 的一维直线，而关于 ϕ_k 的条件分布仅取决于这条直线上的值。

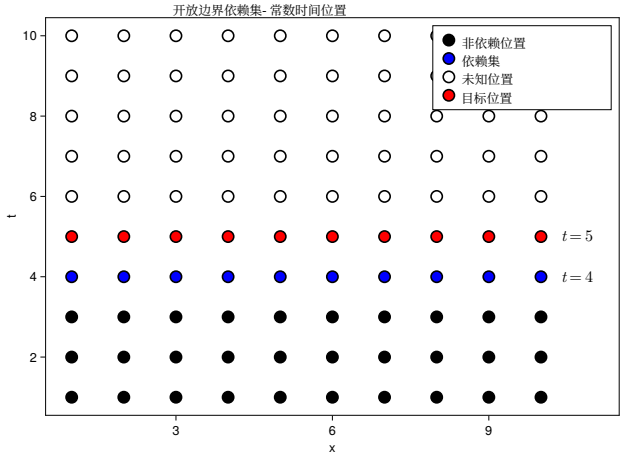
对于标量晶格场论，我们也可以得出相同结论：将求和替换为积分，并在 8 中加入诸如 ϕ_l^2和 ϕ_l^4这样的项。将晶格向量 $\boldsymbol{r}$ 分解为一个 $d-1$ 维的空间 向量 $\boldsymbol{x}=[x_1,\dots x_{d-1}]$ 和时间 t，使得 $\boldsymbol{r}=[r_1=x_1,\dots r_{d-1}=x_{d-1},r_d=t]$。在图 1a 中，对于一个 10×10 晶格，蓝色位置的依赖集实际上是一条沿空间维度 x延伸、长度为 L的直线。更一般地，$\phi_{k(\boldsymbol{r})}$的依赖集可以表示为一个 $d-1$ 维空间子晶格，并通过以下参数方式定义：

$$B_{\boldsymbol{r}}^{o}=\left\{\boldsymbol{r}'|\boldsymbol{r}'=\begin{cases}[\boldsymbol{x}',t] & \text{if } k([\boldsymbol{x}',t])<k(\boldsymbol{r})\\ [\boldsymbol{x}',t-1] & \text{if } k([\boldsymbol{x}',t])>k(\boldsymbol{r})\end{cases}\right\} \tag{9}$$

[2] $k-$ 应将 $\hat{}\mu$ 理解为晶格位置 $\boldsymbol{x}-\hat{}\mu$，其中 $\boldsymbol{x}$ 根据给定的排序映射到 k



(a) 红色位置处场的条件概率 $p(\phi_k|\phi_{<k})$仅取决于 $\phi_{<k}$内 $\phi_{\geq k}$ 中（红色/白色）的最近邻。依赖集仅由蓝色位置组成，而 $\phi_{<k}$ 内的黑色位置可以忽略。



(b) 在恒定时间 $p(\{\phi(x,t)|t = 5\}|\{\phi(x,t)|t < 5\})$，红色位置处场的条件概率仅取决于标记为蓝色的 $t = 4$ 位置 $\{\phi(x,t)|t = 4\}$，而不取决于其余标记为黑色的位置。

图 1：在开放边界条件下，单个位置的场值条件概率（1a）和恒定时间流形上各位置的联合条件概率（1b）的依赖集。

我们将其称为位于 $\boldsymbol{r}$ 的依赖面。根据 8，我们可以写出多个变量的联合条件概率。例如，对于伊辛模型，我们可以写出这个 m 变量的联合分布：

$$\begin{aligned} p\left(\{\phi_k, \dots \phi_{k+m-1}\} \,|\phi_{<k}\right) &= p(\phi_k|\phi_{<k}) \dots p(\phi_{k+m-1}|\phi_{<k+m-1}) \\ &= \frac{\sum\limits_{\phi_N, \dots \phi_{k+m}} \exp\left(-\beta J \sum\limits_{l=k}^{N} \left(\phi_l \sum\limits_{\mu} \phi_{l-\hat{\mu}}\right)\right)}{\sum\limits_{\phi_N, \dots \phi_k} \exp\left(-\beta J \sum\limits_{l=k}^{N} \left(\phi_l \sum\limits_{\mu} \phi_{l-\hat{\mu}}\right)\right)} \end{aligned} \tag{10}$$

对于 $\boldsymbol{r} = [1, \dots 1, t]$ 和 $m = L^{d-1}$ 的选择，集合 $\{\phi_k, \dots \phi_{k+m-1}\}$ 是恒定时间子晶格 $\{\boldsymbol{r}'|t' = t\}$，依赖面是前一个时间步的恒定时间流形：

$$B_t^o = \{\boldsymbol{r}'|t' = t - 1\} \tag{11}$$

图 1b 展示了一个 10×10 晶格中的示例。这实际上将整个自回归过程简化为时间维度上的一阶马尔可夫过程。

周期性边界条件一般而言，只要作用量是局域的，无论边界条件如何，$\phi_{k(\boldsymbol{r})}$ 的依赖面都只包含 $\phi_{<k}$ 内 $\phi_{\geq k}$ 的最近邻。不过，在周期性边界条件下，依赖面由两个 $n-1$ 维子晶格组成：

$$B_t^p = B_t^o \cup \{\boldsymbol{r}|r_d = 1\} \tag{12}$$

其中包括与初始时间 子晶格相对应的子晶格。请注意，周期性边界在拓扑上使晶格成为 d 维环面，因此，$B_t^o = \{r_d = t-1\}$ 在空间维度上将是一个 $d-1$ 环面。这个环面有效地"阻挡"了所有等时间 $r_d < t$ 环面对其的影响，唯独初始环面 $r_d = 1$ 除外，它与另一侧的 $r_d > t$ 相连。图2展示了 $d = 2$ 的这种情况。换言之，只有位于 $r_d = t - 1$ 和 $r_d = 1$ 的子晶格会与 子晶格 $r_d = t$ 相连，无论是直接相连，还是通过未知位置/未来位置间接相连。我们假设模型中采用周期性边界条件，因为这样可以利用作用量[3]的平移不变性。

2.4 自回归条件归一化流与 Metropolis 采样

2.4.1 条件掩码归一化流

条件归一化流[15] 对映射 f_θ 进行建模，将标准先验分布 $z \sim r(z)$˜ ˜ ˜中的样本映射为以 χ为条件的变分分布样本 $\phi \sim \tilde{q}_{f,\,\theta}(\phi|\chi)$，使得 $\phi = f_\theta(z)$。根据概率分布的变量变换公式，我们可以得到变分分布的对数概率密度函数：

$$\begin{aligned} \log r(z) &= \log \tilde{q}_{f,\theta}(f_\theta(z)|\chi) + \log \left|\det \frac{\partial f_\theta(z)}{\partial z}\right| \\ \implies \log \tilde{q}_{f,\theta}(f_\theta(z)|\chi) &= \log r(z) - \log \left|\det \frac{\partial f_\theta(z)}{\partial z}\right| \end{aligned} \tag{13}$$

映射 f_θ 由一系列 n 耦合层 $g^{(i)}$ 构成，其输入交替经过互补的二值掩码 m 和 $1 - m$[4]的作用：

$$\begin{aligned} f_\theta(z) &= \left(g_1 \circ \dots \circ g_n\right)(z) \\ g_i(z_i, \chi) &= m^{(i)}(z_i) + (1 - m^{(i)}) \left(\tilde{\phi} e^{s^{(i)}(m^{(i)}(z_i), \chi)} + t^{(i)}(m^{(i)}(z_i), \chi)\right) \\ m^{(i)} &= \begin{cases} m & i \text{ even} \\ 1 - m & i \text{ odd} \end{cases} \end{aligned} \tag{14}$$

[3]对于具有周期性边界条件的有限晶格，准确的描述是循环置换不变性；而真正的平移不变性严格来说只存在于无限晶格中。[4]除非另有说明，所有乘积均按元素逐项计算

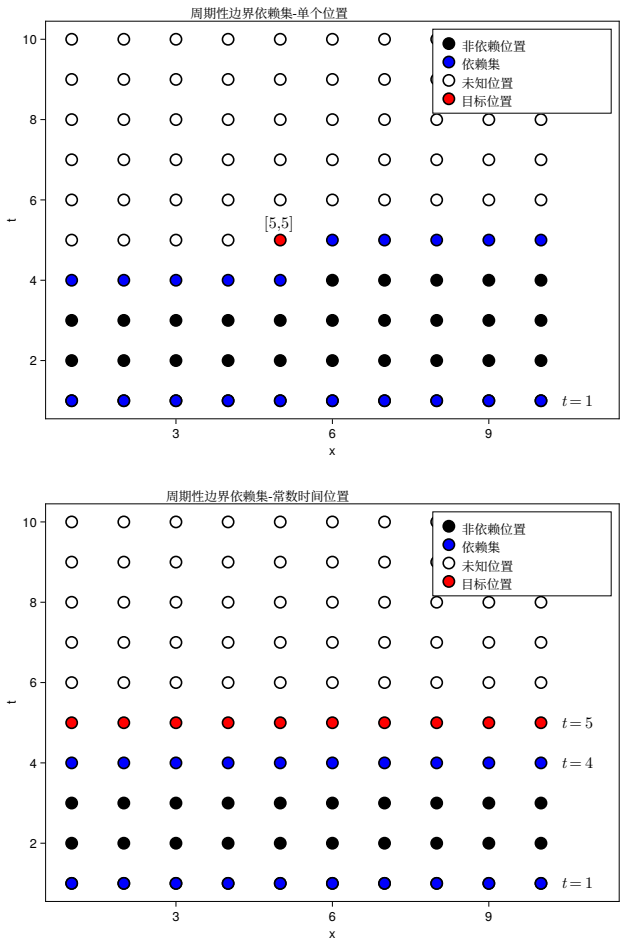


图2：对于周期性边界，单个红色位置（上）或恒定时间子晶格中的位置（下）的场值分布以依赖集（标记为蓝色）为条件。这些依赖集与开放边界的依赖集相似，但还包括初始时间子晶格 at $t = 1$。

其中 $z_1 = z$ 和 $z_i = g_{i-1}(z_{i-1},\ \chi)$ 对于 $i > 1$成立，因此 $z_{n+1} = \tilde{\phi}$。函数 $s^{(i)}$ 和 $t^{(i)}$ 则使用前馈神经网络建模。于是，式 13 中的对数行列式变为：

$$\log\left|\det\frac{\partial f_\theta(z)}{\partial z}\right| = \sum_{i=1}^{n}\log\left|\det\frac{\partial g_i(z_i)}{\partial z_i}\right| = \sum_{i=1}^{n}\text{sum}(s^{(i)}(m^{(i)}(z_i),\chi))$$

其中，求和函数将输入数组中的值相加。耦合层可以求逆并组合在一起，得到逆流：

$$\begin{aligned} g_i^{-1}(z_{i+1},\chi) &= m^{(i)}(z_{i+1}) + (1-m^{(i)})\left(z_{i+1} + t^{(i)}(m^{(i)}(z_{i+1}),\chi)\right)e^{-s^{(i)}(m^{(i)}(z_{i+1}),\chi)} \\ f_\theta^{-1}(\tilde{\phi}) &= \left(g_n^{-1}\circ\cdots\circ g_1^{-1}\right)(\tilde{\phi}) \end{aligned} \tag{15}$$

˜这有助于评估变分分布中已知样本的对数概率密度函数 ϕ：

$$\log\tilde{q}_{f,\theta}(\tilde{\phi}|\chi) = \log r(f_\theta^{-1}(\tilde{\phi})) - \sum_{i=1}^{n}\text{sum}(s^{(i)}(m^{(i)}(z_{i+1}),\chi)) \tag{16}$$

2.4.2 自回归过程

我们在 5 中将玻尔兹曼分布拆分为单个位置处场的条件分布之乘积。随后，我们将其推广，并评估了恒定

时间子晶格上场值的联合条件分布在 11 中。现在，让我们针对周期性边界条件重写这个联合条件分布，但这次还要以 2 中的耦合 λ 为条件：

$$p\left(\{\phi(\boldsymbol{r})|t\}|\phi_{<[1,\dots t]},\lambda\right) = p\left(\{\phi(\boldsymbol{r})|t\}|\{\phi(\boldsymbol{r}')|t'=t-1,1\}\cup\lambda\right)$$

$$\implies \log p(\phi|\lambda) = \sum_{t=1}^{L} \log p\left(\{\phi(\boldsymbol{r})|x_d=t\}|\{\phi(\boldsymbol{r})|x_d=t-1,1\}\cup\lambda\right)$$

其中第二行是完整玻尔兹曼分布的对数形式。我们使用条件归一化流 $\tilde{q}_\theta(\phi|\chi,\ t)$ 的输出变分分布$\tilde{}$来建模右侧的各个条件分布，其中$\chi = \{\phi(\boldsymbol{r})|x_d = t-1,\ 1\}\cup\lambda$。我们每次采样一个包含 $\phi(t) = \{\phi(\boldsymbol{r})|x_d=t\}$ 的恒定时间流形，并将其作为后续归一化流调用的输入：

$$\begin{aligned}
\tilde{\phi}(1) &\sim \tilde{q}_\theta(\tilde{\phi}|\Theta,\Theta,\lambda,1) = \tilde{q}_\theta(1)\\
&\vdots\\
\tilde{\phi}(k) &\sim \tilde{q}_\theta(\tilde{\phi}|\tilde{\phi}(k-1),\tilde{\phi}(1),\lambda,k) = \tilde{q}_\theta(k)\\
&\vdots\\
\tilde{\phi}(L) &\sim \tilde{q}_\theta(\tilde{\phi}|\tilde{\phi}(L-1),\tilde{\phi}(1),\lambda,L) = \tilde{q}_\theta(L)
\end{aligned} \tag{17}$$

其中，θ 是神经网络权重，Θ 是无用/缺失值[5]。请注意，我们明确将时间作为输入，因为条件分布不一定与时间无关，也不一定具有时间不变性。最后，我们可以沿时间方向拼接样本，并将对数概率密度函数相加，从而得到 l-ACNF 的最终模型分布 $q_\theta(\phi)$对应的样本及其对数概率密度函数：

$$\phi = \text{cat}_{t=1}^{L}\tilde{\phi}(t)$$

$$\log q_\theta(\phi) = \sum_{t=1}^{L}\log\tilde{q}_\theta(t)$$

我们优化权重 θ，使分布 q_θ 与 p 之间的 KL 散度最小化：

$$\begin{aligned}
\text{KL}(p\Big|\Big|q_\theta) &= \mathbb{E}_{q_\theta(\phi)}\left[\log q_\theta(\phi) - \log p(\phi)\right]\\
&= \frac{1}{M}\sum_{i=1}^{M}\left[\log q_\theta(\phi^{(i)}) + S[\phi^{(i)}]\right] \qquad \phi^{(i)}\in\{\phi\sim q_\theta\}^M
\end{aligned} \tag{18}$$

我们在每一步中对 ϕ 进行小批量采样（批量大小为 M），计算方括号内数量的均值，并使用该数量关于 θ 的梯度来更新权重。在展开 $\log p(\phi)$ 时，我们忽略配分函数的对数 $\log Z$，因为它是一个与 ϕ无关的常数。

2.4.3 棋盘格掩码与卷积网络

在标量格点场论的生成式建模中，针对 $s^{(i)}$和 $t^{(i)}$的特定神经网络架构选择，以及 14 中使用的二值掩码 m，具有算法上的优势。仅包含$d-1$ 维卷积层（以及激活函数和残差连接）并为 m 使用棋盘格掩码的神经网络，能够打破晶格[6]的 $d-1$ 个空间维度上的平移不变性（前提是这种对称性属于 13 中的先验分布 $r(z)$）。由于我们沿时间维度依次进行采样，因此该方向上不存在此类显式不变性，不过训练良好的模型预计能够近似满足这些不变性。全卷积网络还能够沿空间维度对晶格尺寸 L不同的构型进行采样。17 中的自回归采样具有沿时间维度且独立于 L 的依赖集，因此我们可以对尺寸不同于训练模型时所用 L 的晶格进行采样，使其具有可扩展性。为避免棋盘格掩码导致归一化流中出现“泄漏”，L 仍须为偶整数。

为了对我们的模型进行基准测试，我们还考虑了一种具有 d 维卷积层和循环填充的掩码归一化流，其中 NF 生成完整的晶格构型（而不是恒定时间子晶格），且仅以耦合 λ为条件，类似于 [6]中所述的实现方式。这使其能够在所有 d个方向上实现平移不变性，并且同样具有可扩展性。与后者（L^d）相比，l-ACNF 具有更好地应对 维数灾难 的（理论）优势，因为其有效输入规模（$O(L^{d-1})$）更小。

[5]我们在初始采样时提供零字段以及一个布尔标志。详情请参见代码实现。

2.4.4 Metropolis-Hastings 抽样

训练以最小化 18 中的 KL 散度后，变分分布中的任何偏差都会导致从该分布中采样后评估的各种可观测量/估计量产生偏差。为了从 q_θ中获得无偏的 p 样本，我们将其样本通过 Metropolis-Hastings 过程——一种马尔可夫链蒙特卡罗（MCMC）技术。从初始样本 $\phi^{(1)}$ 及其对应的对数概率密度函数 $\log q_\theta(\phi^{(1)})$开始，我们使用当前样本 $\phi^{(i)}$ 和提议样本 ϕ'，在第 i步更新链，具体如下：

$$\begin{aligned} A(\phi^{(i-1)}, \phi') &= \min\left(1, \exp\left(\log q_\theta(\phi^{(i-1)}) - \log p(\phi^{(i-1)}) - \log q_\theta(\phi') + \log p(\phi')\right)\right) \\ &= \min\left(1, \exp\left(\log q_\theta(\phi^{(i-1)}) - S[\phi^{(i-1)}] - \log q_\theta(\phi') + S[\phi']\right)\right) \\ \phi^{(i)} &= \begin{cases} \phi' & A < \pi \\ \phi & \text{else} \end{cases} \qquad \text{where } \pi \sim \text{Uniform}(0,1) \end{aligned} \tag{19}$$

Metropolis-Hastings 样本渐近收敛到 p 的必要条件是遍历性，这要求：

$$q_\theta(\phi) > 0 \quad \text{or} \quad \log q_\theta(\phi) > -\infty \quad \forall \phi \in \mathbb{R}^{L^d}$$

如果流的先验分布满足 $r(z) > 0\ \forall z \in \mathbb{R}$，则 q_θ 中的归一化流满足这一条件；例如，高斯先验就满足这一条件。

2.4.5 对称化与绝热再训练

在 Metropolis-Hastings 更新步骤 19 中，确保所提议构型的各种对称变换（这些变换使作用量 S 保持不变）不会改变接受概率 A，这很有用。这包括晶格的平移和反射、沿每个维度的平移，以及取反 $\phi \to -\phi$。具体而言，在应用变换 $\mathfrak{g}$ 后，对概率密度函数 $\text{Avg}_{\mathfrak{g}} q_\theta(\mathfrak{g}(\phi))$进行平均，并在每一步随机选择一个经变换的场 $\mathfrak{g}(\phi)$：

$$\begin{aligned} \log q_\theta(\phi) &\to \text{logsumexp}_{\mathfrak{g}\in G}\left(\log q_\theta\left(\mathfrak{g}(\phi)\right)\right) - \log|G| \\ \phi &\to \mathfrak{g}(\phi) \qquad \mathfrak{g} \sim G \end{aligned} \tag{20}$$

其中，G 是我们考虑的所有对称变换的集合。当分布 $p(\phi)$具有彼此分离的模态时，这是一种应对收敛不佳的方法，相关研究见 [6]。在我们的情形中，时间维度与空间维度之间的旋转还有助于解决空间维度与时间维度处理方式固有的不对称问题，以及 l-ACNF 沿时间维度缺乏不变性的问题，正如 2.4.3 节所述。模型能够扩展到不同于训练所用值的晶格尺寸，因此我们可以进行绝热再训练：先用较小的 L进行训练，再以该网络作为训练较大的 L时的初始状态，这与 [3]类似。

3 数值实验

3.1 网络架构与训练

本研究使用 Python 深度学习库 PyTorch[10] 构建所有模型。我们在长度为 16（$L = 16$）的二维晶格（$d = 2$）上对 ϕ^4 理论进行建模。对于模型1，条件归一化流网络包含 14 中所述的 $n = 8$ 个耦合层。如 2.4.3 节所述，我们使用一维棋盘格二值掩码 m 和一维门控卷积网络[14]实现函数 $s^{(i)}$ 和 $t^{(i)}$；条件变量 χ 则由一条包含 4 个门控卷积网络的公共链进行处理，并作为跳跃连接传递至另一个输入为 $m^{(i)}z_i$的门控卷积网络。设置 $m^2 = -4$ 和$\lambda \sim$后，模型使用 18 中的 KL 散度损失进行训练；每一步均从 Uniform(4.8,5.4) 中采样，使用 Adam 优化器和学习率 $\eta = 10^{-3}$，配合 Step 调度策略训练 ≈ 90 分钟，共 9000 步。模型2通过对模型1进行绝热再训练构建，使用 $\eta = 10^{-5}$ 以相同的优化器和调度策略训练 ≈ 40分钟，采用 $L = 20$。模型3采用 2.4.3 节第二段所述的二维掩码归一化流构建，用于与模型1进行基准比较。它使用二维棋盘格掩码，以及带循环填充的二维门控卷积网络，共包含 $n = 12$ 个耦合层。这确保所有模型的卷积运算和非线性激活数量相同，但由于使用二维卷积层，权重数量略多。模型3使用 $\eta = 10^{-3}$ 训练 ≈ 90 分钟，并采用 $L = 16$ 以保持与模型1等效。

所有训练和推理均使用 Nvidia RTX 2060 移动版 GPU 完成。源代码可在 https: //github.com/dinesh110598/l-ACNF-scalar 获取。所有图表的数据和代码均可在此交互式 Julia 笔记本中获取：https://github.com/dinesh110598/l-ACNF-scalar/blob/main/Figures/notebook.jl。

L	8	10	12	14	16	18	20
λ	6.008	5.550	5.276	5.113	4.99	4.89	4.82

表1：先前的一项研究使用了对应于 $L \leq 14$ 的耦合层 λ，使系统沿着 $m_p L \approx 4$ 临界线分布。我们使用模型对不同 λ 下的 $m_p L$ 进行试算，确定了 $L \geq 16$ 的取值。有关这些 λ 取值下 $m_p L$ 与 L 的关系图，请参见图3。

3.2 参数与可观测量

参数 m^2 和 λ 在 2 中的取值选在系统对称相中接近 $m_p L \approx 4$ 临界线的位置，与研究 [2]类似。不同之处在于，我们使用一个训练好的模型来评估不同晶格长度 L 和耦合层 λ下的所有可观测量。我们关注的可观测量包括二点格林函数：

$$G_c(\boldsymbol{r}) = \frac{1}{N} \sum_{\boldsymbol{r}'} \left(\langle \phi(\boldsymbol{r}') \phi(\boldsymbol{r}' + \boldsymbol{r}) \rangle - \langle \phi(\boldsymbol{r}') \rangle \langle \phi(\boldsymbol{r}' + \boldsymbol{r}) \rangle \right)$$

其空间傅里叶变换

$$\tilde{G}_c(\boldsymbol{p}, t) = \frac{1}{L^{d-1}} \sum_{\boldsymbol{x}} e^{i\boldsymbol{p}.\boldsymbol{x}} G_c(\boldsymbol{x}, t)$$

极点质量的估计量

$$m_p(t) = \text{arccosh} \left(\frac{\tilde{G}_c(0, t-1) + \tilde{G}_c(0, t+1)}{2\tilde{G}_c(0, t)} \right) \tag{21}$$

二点磁化率

$$\chi_2 = \sum_{\boldsymbol{r}} G_c(\boldsymbol{r})$$

以及平均伊辛能量密度

$$E = \frac{1}{d} \sum_{\mu=1}^{d} G_c(\hat{\mu})$$

我们还研究了一个基于长度为 T的 Metropolis-Hastings 链接受/拒绝统计量的自相关函数：

$$\widehat{\rho(\tau)/\rho(0)}_{acc} = \frac{1}{T-\tau} \sum_{j=1}^{T-\tau} \prod_{i=1}^{\tau} \mathbf{1}_{rej}(i+j) \tag{22}$$

其中， $\mathbf{1}_{rej}$ 是拒绝一次 MH 步骤时的恒等映射。另一组自相关估计量基于前文定义的 2 点可观测量

$$\widehat{\rho(\tau)/\rho(0)}_{\mathcal{O}} = \frac{\frac{1}{T-\tau} \sum_{i=1}^{T-\tau} (\mathcal{O}_i - \bar{\mathcal{O}})(\mathcal{O}_{i+\tau} - \bar{\mathcal{O}})}{\frac{1}{T} \sum_{i=1}^{T} (\mathcal{O}_i - \bar{\mathcal{O}})^2} \tag{23}$$

基于接受统计量的积分自相关时间定义为

$$\tau_{acc}^{int} = \frac{1}{2} + \lim_{\tau_{\max} \to \infty} \sum_{\tau=1}^{\tau_{\max}} \frac{\rho_{acc}(\tau)}{\rho_{acc}(0)} \approx \frac{1}{2} + \lim_{\tau_{\max} \to \infty} \sum_{\tau=1}^{\tau_{\max}} \widehat{\rho(\tau)/\rho(0)}_{acc} \tag{24}$$

上述量的动力学临界指数 z 通过以下方式拟合：

$$\tau_{acc}^{int}(L) \approx cL^z \tag{25}$$

我们设定 $T = 10^5$ 和 $L \in [8, 10, 12, 14, 16, 18, 20]$。在式 19 给出的每个 Metropolis-Hastings 步骤中，每经过（第i步）后，都进行式 20 所示的对称化，以变换 $\phi^{(i)}$ 和 $\log q_\theta(\phi^{(i)})$。虽然这会降低 MCMC 的计算速度，但接受率显著提高了，从 50–55% 提升至 65–70%。表 1 列出了 $L \leq 14$中用于 [2]的 λ 取值，其余拟合则基于模型1，通过试错法确定，见文献 21。图3展示了在指定 λ 下 $m_p L$ 相对于 L 的关系，这些点接近此处的红色虚线临界线。图4依据文献 22 和 23，绘出了基于接受/拒绝统计量（模型1）以及可观测量 G_c(0) χ_2, E 的自相关函数。与其他曲线相比，橙色和红色曲线（对应于 $L = 18$,20）似乎下降得较慢。在图5的积分自相关时间图中，我们使用模型1、2和3，将文献 24 中的 τ_{int}^{acc}与 L 进行比较。最显著的

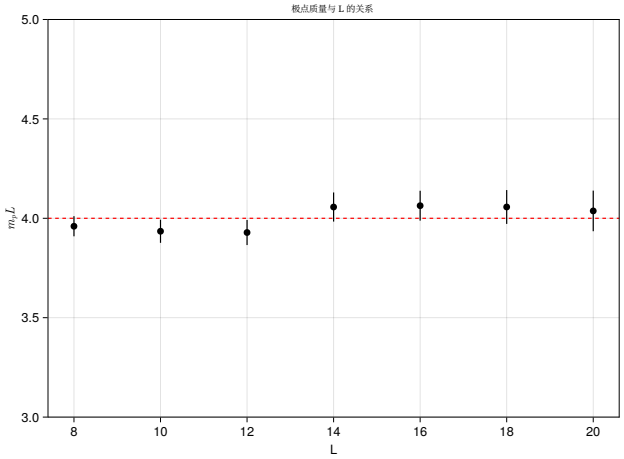


图3：极点质量与晶格长度的乘积 m_pL 相对于 L 作图，使用模型1计算。误差棒表示 68% 置信区间，置信区间通过区块大小为 100 的移动区块自助重采样获得

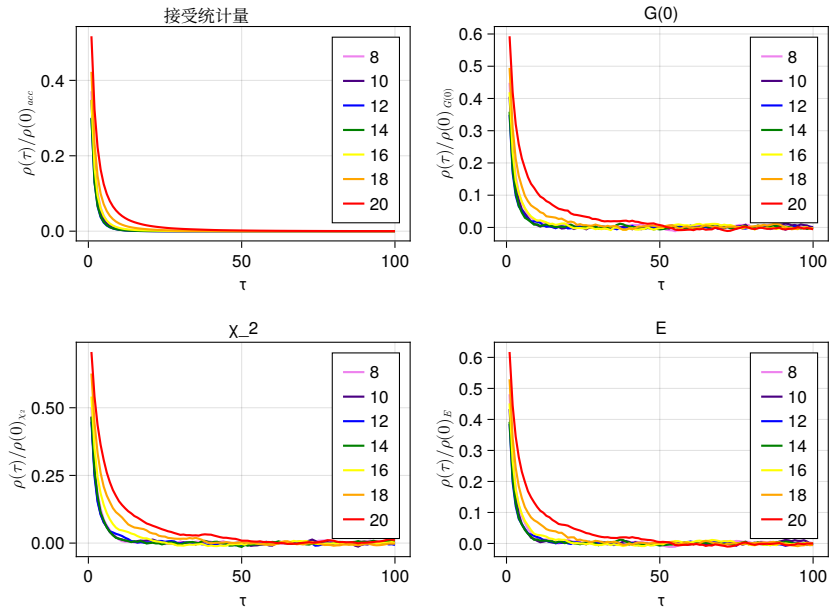


图4：使用模型1计算的不同晶格长度下，与接受统计量及不同可观测量相关的自相关函数图

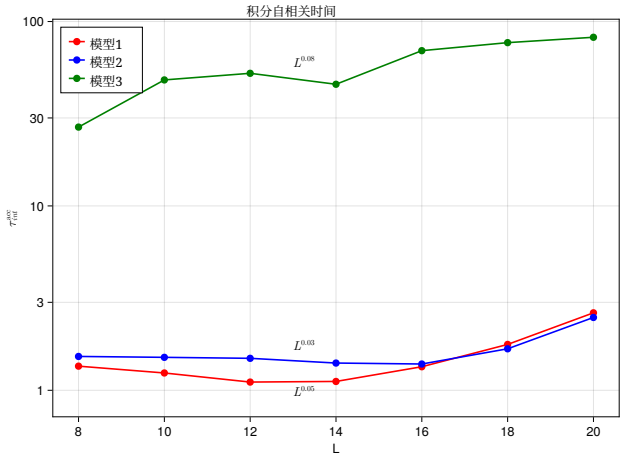


图5：使用模型1、2和3，在对数尺度上绘制接受统计量相对于 L 的积分自相关时间

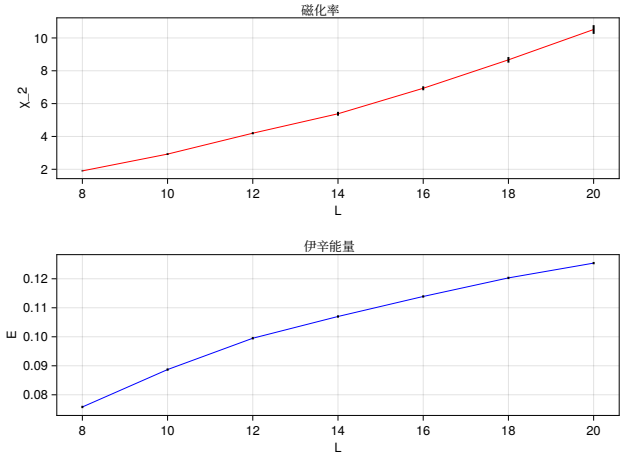


图6：使用模型1绘制可观测量 χ_2 和 E 相对于 L 的变化。误差棒表示通过移动区块自助重采样得到的95%置信区间，区块大小为100

该图的一个特点是，代表模型3自相关时间的绿色曲线，其数值比模型1、2高出数个数量级；模型1、2按照我们提出的方法，以自回归方式对恒定时间子晶格进行采样。蓝色曲线所示的经过绝热再训练的模型2，其数值变化更平缓，动力学临界指数也小于红色曲线所示的模型1。自相关时间在 $L = 8$ 和 $L = 14$ 之间总体上大致保持恒定，并在 $L = 16$ 和 $L = 20$ 之间增加，这表明当采样的晶格尺寸大于模型训练时使用的尺寸 $L = 16$ 时， q_θ 的质量（以接近实际玻尔兹曼分布 p 的程度衡量）会下降。这也可能是图6中磁化率图（模型1）的 $L = 16$ 以上以及图3中极点质量的误差棒增大的原因。尽管如此，通过对25中的曲线进行拟合得到的动力学临界指数接近于零，而传统方法通常并非如此 [6]。

4 结论性评述

在本工作中，我们提出了一种生成模型（称为 l-ACNF），该模型利用掩蔽条件归一化流，以自回归方式对标量格点场的“等时”子晶格进行采样。根据 ϕ^4 理论中作用量的局域性 S

[6]基于流的模型的标度行为通常取决于具体的实现细节，而曲线拟合到指数函数可能不准确，这也可从我们的对数尺度图中缺乏直线走势看出。更多详情请参见 [1]中的研究。

在 2 中，我们根据 12 确定，这些归一化流所依赖的条件是一个大小为 $O(L^{d-1})$ 的更小子晶格，而不是通用上界 L^d。换言之，作用量的局域性将玻尔兹曼分布的自回归展开简化为时间上的马尔可夫过程。与先前研究 [6] 中使用的、以等效方式构建的现有模型相比，按照我们方案构建的模型所报告的自相关时间缩短了数个数量级。这是因为我们模型中条件归一化流的输入/输出空间更小，从根本上缓解了现有方法中的维数灾难——避免训练轮次呈指数级增加和/或使用更庞大的神经网络。计算资源方面的限制使我们的研究仅限于能够在相对普通的硬件上于数小时内完成训练并收敛的模型。然而，我们的模型仍实现了较短的自相关时间和接近于零的动态临界指数，与 [2] 等大规模研究相近。如果我们采用 $d = 4$的真实时空规格，l-ACNF 会使用三维卷积而不是四维卷积来采样恒定时间子晶格。在图形处理器设备上运行时，CUDNN 库[5] （任何深度学习框架用于访问图形处理器优化函数/方法的库）由于其相对实用性，仅包含针对最多 3 个维度的优化卷积核，这使 l-ANF 相较于其他方案略占优势。

由于局域性是所有格点量子场论普遍具有的性质，我们的方法也可以扩展到更复杂的理论。例如，格点规范场的威尔逊作用量

$$S_W = -\sum_p \frac{2}{g^2}\mathrm{Re}(\mathrm{Tr}(U(p)))$$

是对基本方格 p 的求和，这些基本方格是由链路 $U_\mu(\boldsymbol{r})$ 沿每个位置 $\boldsymbol{r} = [\boldsymbol{x}, t]$ 及方向 μ 上的最近邻构成的回路。我们可以扩展第 2.3 节中的分析，并确定等时间子晶格 $p(\{U_\mu(\boldsymbol{r})|r_d = t\}|\{U_\mu(\boldsymbol{r})|r_d < t\})$ 上链路的条件分布，其依赖集如下（假设采用周期性边界条件）：

$$B(\boldsymbol{r}|r_d = t) = \{U_\mu(\boldsymbol{r})|r_d = t-1, \mu < d\} \cup \{U_\mu(\boldsymbol{r})|r_d = 1\}$$

其中包含 $r_d = t-1$ 子晶格内部的链路（时间维度方向上的链路除外，因为它们不属于包含沿 $t = t_o$ 的链路的基本方格，也不与之相连），以及 $r_d = 1$ 子晶格；这又将相应流网络的输入规模缩减至 $O(L^{d-1})$。将我们的方法与 [7] 和 [3] 提出的规范等变流相结合，是一个值得进一步研究的方向，但遗憾的是超出了本研究的范围。

参考文献

[1] Ryan Abbott 等. “格点量子色动力学流式采样的尺度与可扩展性问题”. In: *arXiv* 预印本 *arXiv:2211.07541* (2022).[2] Michael S Albergo、Gurtej Kanwar 和 Phiala E Shanahan. “格点场论中用于马尔可夫链蒙特卡罗的流式生成模型”. In:《物理评论 *D*》100.3 (2019), p. 034515.[3] Denis Boyda 等. “ 使用 SU(N) 规范等变流进行采样 ”. In:《物理评论 *D* 》 103.7 (2021), p. 074504.[4] Michael M Bronstein 等. “几何深度学习：网格、群、图、测地线与规范”. In: *arXiv* 预印本 *arXiv:2104.13478* (2021).[5] Sharan Chetlur 等. “cudnn：用于深度学习的高效原语”. In: *arXiv* 预印本 *arXiv:1410.0759* (2014).[6] Daniel C Hackett 等. “格点场论中多模态分布的流式采样”. In: *arXiv* 预印本 *arXiv:2107.00734* (2021).[7] Gurtej Kanwar 等. “格点规范理论的等变流式采样”. In:《物理评论快报》125.12 (2020), p. 121601.[8] Mario Köppen. “维数灾难”. In: 第五届工业应用软计算在线世界会议（ *WSC5*）. Vol. 1. 2000, pp. 4–8.[9] Colin Morningstar. “量子场论中的蒙特卡罗方法”. In: *arXiv* 预印本 *hep-lat/0702020*(2007).[10] Adam Paszke 等. “PyTorch：一种命令式风格的高性能深度学习库”. In: 神经信息处理系统进展32 (2019).[11] Tomaso Poggio 等. “深层网络何以能够以及何时能够避免维数灾难，而浅层网络则不能：综述”. In:《自动化与计算国际期刊》14.5 (2017), pp. 503–519.[12] Dinesh PR. 使用神经网络分析伊辛模型. 2021年7月. URL: http://dr.iiserpune.ac.in: 8080/xmlui/handle/123456789/6014.[13] Robert H Swendsen 和 Jian-Sheng Wang. “蒙特卡罗模拟中的非普适临界动力学”. In:《物理评论快报》58.2 (1987), p. 86.

[14] Aaron Van den Oord 等. “使用 PixelCNN 解码器进行条件图像生成”. In: 神经信息处理系统进展29 (2016).[15] Christina Winkler 等. “使用条件归一化流学习似然”. In: *arXiv* 预印本 *arXiv:1912.00042* (2019).[16] Ulli Wolff. “自旋系统的集体蒙特卡罗更新”. In: 《物理评论快报》62.4 (1989), p. 361.[17] Ulli Wolff. “临界减速”. In:《核物理 *B*：会议录增刊》17 (1990), pp. 93–102.[18] Dian Wu、Riccardo Rossi 和 Giuseppe Carleo. “利用自回归神经网络进行无偏蒙特卡罗簇更新”. In:《物理评论研究》3.4 (2021), p. L042024.[19] Dian Wu、Lei Wang 和 Pan Zhang. “使用变分自回归网络求解统计力学问题”. In: 《物理评论快报》122.8 (2019), p. 080602.