arXiv:2305.05396v2 [cond-mat.stat-mech] 4 Mar 2024

# 拓扑柯尔莫哥洛夫复杂度与别列津斯基-科斯特利茨-索利斯机制

Vittorio Vitale，[1, 2, 3, *] Tiago Mendes-Santos，[4] Alex Rodriguez，[1, 5] 和 Marcello Dalmonte[1, 2]1国际理论物理中心（*ICTP*），*Strada Costiera 11*，*34151* 的里雅斯特，意大利 [2]*SISSA*，*via Bonomea 265*，*34136* 的里雅斯特，意大利[3]格勒诺布尔-阿尔卑斯大学，法国国家科学研究中心，凝聚态物理与建模实验室（*LPMMC*），格勒诺布尔 *38000*，法国[4]理论物理 *III*，电子关联与磁性中心，奥格斯堡大学物理研究所，*86135* 奥格斯堡，德国[5]数学与地球科学系，的里雅斯特大学，*via Alfonso Valerio 12/1*，*34127*，的里雅斯特，意大利（日期：2024年3月5日）

拓扑在我们理解多体物理的过程中发挥着根本作用，从经典场论中的涡旋和孤子，到量子物质中的相与激发。拓扑现象与信息内容的分布密切相关；不同于普通物质，这种分布如今由非局域自由度所支配。然而，拓扑效应如何支配多体态的复杂度——即其配分函数的复杂度——目前尚不清楚。在这项工作中，我们展示了在经典统计力学中，拓扑与复杂度是直接交织在一起的概念。具体而言，我们提出一种理论，说明经典配分函数采样的柯尔莫哥洛夫复杂度 会因系统中是否存在拓扑激发而呈现独特且鲜明的特征。我们研究了定义在多种拓扑类型上的二维伊辛模型、海森堡模型和 XY模型，并将相应的采样视为构型空间中的高维流形，通过内在维数来量化其复杂度。对于伊辛模型和海森堡模型，内在维数与实空间拓扑无关；而对于 XY模型，它则强烈依赖于温度：跨越 Berezinskii-Kosterlitz-Thouless（BKT）相变时，复杂度开始依赖于拓扑。在 BKT相中，复杂度呈现出与实空间流形的同调相关的特征性依赖关系；对于 $g$-环面，其标度关系仅取决于亏格。我们认为，这种行为与数据空间中序参量——条件连通性——的涌现密切相关，而后者呈现标度行为。我们的方法为从柯尔莫哥洛夫复杂度的视角理解多体相互作用所涌现的拓扑现象铺平了道路。

复杂度和拓扑是我们理解和诠释科学各个领域中各种现象与机制的两大基石 [1]。多体物理是二者发挥作用的一个典型领域：在这一领域中，人们关注大量微观组分之间的相互作用如何涌现出集体行为，从而使得仅使用有限数量的变量就能实现有效描述。从信息论的角度来看，这意味着对给定物理系统描述所对应的柯尔莫哥洛夫复杂度进行了大幅压缩：只需少数几个参数，就足以捕捉相应物理现象的基本特征 [2]。这种复杂度降低的典型例子是临界行为，此时，关联函数完全由少数几个参数——临界指数 [3]——精确确定。

近年来，得益于非参数无监督学习方法领域的显著进展 [4–6], ，如今我们已能够在分子科学以及广义而言的量子化学和物理学背景下，对多体过程的复杂度作出定量论述 [7–9]。在这些领域中，估算复杂度对于理解众多自由度之间的相互作用尤为重要，并最终有助于识别哪些自由度对于某种物理现象的发生最为关键。这与“特征选择” [10–13]或“降维” [14, 15], 密切相关，即剔除看似无关或冗余的特征，或者寻找一种数据表示方式，以少量变量呈现数据，而这些变量被视为原始变量的复杂函数。近期，同样的工具也被应用于临界现象 [16–18], ，并已证明能够定性揭示复杂度与临界行为之间的关联 [19]。相比之下，人们对复杂度如何由拓扑现象决定知之甚少：给定一个配分函数，能否仅通过考察其复杂度推断其中存在非局域自由度？除了理论层面的意义，探讨这一问题也可能对多种系统的实验研究具有重要价值，包括冷原子气体 [20], 、约瑟夫森结阵列 [21],和量子退火架构 [22], 。这些系统中近期已证实存在拓扑现象，而且通常能够实现单点分辨率。

在此，我们表明，在经典统计力学中，Berezinskii-Kosterlitz-Thouless [23] 机制的拓扑起源与表征配分函数的柯尔莫哥洛夫复杂度中拓扑贡献的出现有关。这种关系完全由定义统计力学问题所在流形的性质决定，并且

* vittorio.vitale@lpmmc.cnrs.fr

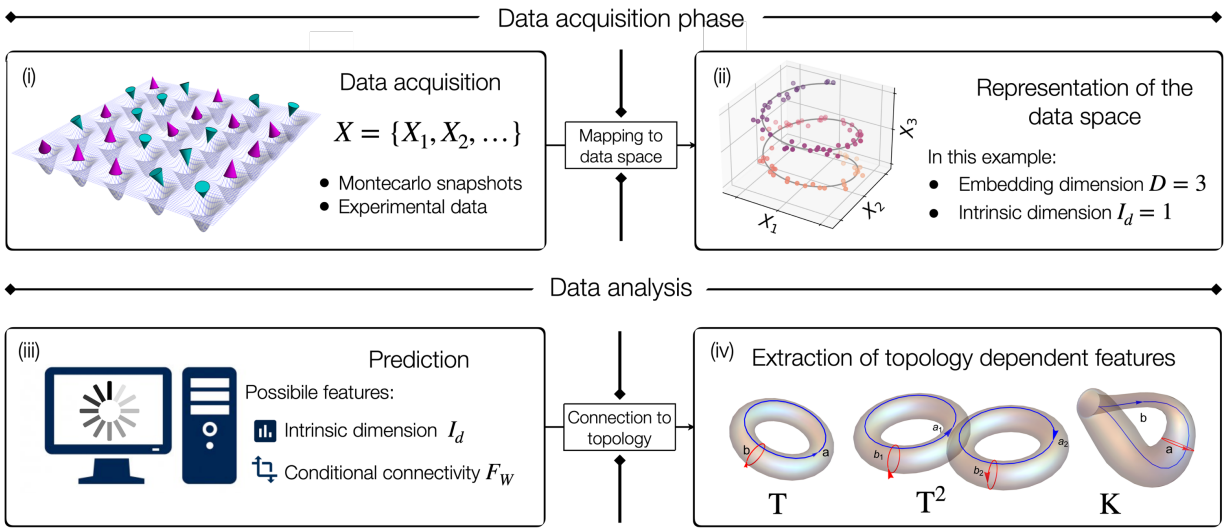


图 1. **将经典配分函数的复杂度与拓扑现象关联起来的指南。** (i) 我们收集物理系统的快照——蒙特卡洛构型或实验数据。(ii) 我们将构型映射到一个 $D$ 维（嵌入维数）数据空间。(iii) 我们提取数据集的特征，例如内在维数（$I_d$）、条件连通性（$F_W$）。(iv) 我们将发现与系统所处的流形联系起来。这里我们表示一个环面（T）、一个二维环面（$T^2$）和一个克莱因瓶（K）。

这与局域序参量支配的模型截然不同——这类模型的复杂度与流形拓扑无关。

首先，我们提出一个理论框架，阐明随着实空间流形的同调群维数增加，复杂度如何降低：这反映出实空间流形越复杂，系统的低能激发受到的约束就越多。随后，我们分析了定义在一系列二维流形（包括可定向和不可定向流形）上的 XY模型、伊辛模型和海森堡模型的配分函数。对于每个流形，我们利用内在维数估算柯尔莫哥洛夫复杂度；内在维数是描述某种数据结构所需的最少自由度数量。接着，我们将流形的拓扑柯尔莫哥洛夫复杂度定义为该流形上的配分函数复杂度与一维环面上的配分函数复杂度之差（$T^1$）。

在 XY模型的 BKT相中，我们观察到，复杂度强烈依赖于模型所定义其上的流形。与我们的理论一致，这种复杂度由第一同调群决定，并满足现象学上的有限尺寸标度。我们的结果表明，尽管信息在实空间中以非局域方式编码（即编码在与涡旋激发相对应的绕数中），但在数据空间中却显著地局域：这一点体现在一种数据空间中的局域序参量的出现，我们将其称为条件连通性，它在 BKT相变两侧均表现出标度行为。对此类序参量的识别，为同调与复杂度之间建立了关键且直观的联系。相比之下，对于伊辛模型和海森堡模型，我们发现二者之间不存在任何关系

复杂度与实空间拓扑之间的关系，这同样与我们的理论一致。

我们的研究结果首次揭示了拓扑与多体描述复杂度之间的直接联系，这与近期在经典 [24, 25] 和量子 [26–29]统计力学中从熵含量角度开展的研究相呼应。鉴于复杂度在其他学科中的影响，这一新的联系为不同领域之间的概念与方法迁移打开了大门。

## I. 柯尔莫哥洛夫复杂度：配分函数样本

柯尔莫哥洛夫复杂度是衡量算法复杂度的基础指标，已广泛应用于科学的各个领域（综述请参见文献 [2]）。它对应于能够生成给定输出的最短计算机程序的长度；在我们的研究中，该输出对应于通过蒙特卡洛（MC）模拟采样得到的构型所构成的数据集。换一种视角来看，它刻画了输出（数据集）的可压缩性。接下来，我们旨在将这一信息论视角与拓扑现象联系起来。为建立拓扑特征与复杂度之间的关系，我们提出一个四步流程（见图1），包括*(i)* 数据采集、*(ii)*适当的表示，以及*(iii)* 在构型空间中进行数据挖掘分析，*(iv)* 并在不同流形上重复上述过程，以筛选出拓扑特征。

*(i)* 配分函数的数据结构。- 我们讨论的第一个要素涉及对.....的识别

我们旨在描述某种物理现象的数据集。对于平衡经典统计力学模型，起点非常自然：配分函数。虽然配分函数无法通过实验测量（而且在许多情况下也无法直接计算），但我们关注的是对其进行采样。具体而言，将构型空间中的一个元素记作 $X_i$，我们的目标数据集就是由 $N_r$ 个这样的元素组成的集合：

$$X = \{X_1, X_2, \ldots, X_{N_r}\}, \tag{1}$$

其中 $N_r \ll D_C$，$D_C$ 是构型空间的总维数（例如，对于由 $N$ 个伊辛自旋组成的集合，$2^N$）。我们将讨论 $X_i \neq X_j$ 的情况，正如我们将在下文论证的，这对于大型多体系统而言是典型的。数据空间 $X$ 包含关于任意阶关联函数的信息，因此预计能够完整刻画物理性质。重要的是，它可以直接从数值实验（例如蒙特卡洛）和真实实验中获得。

*(ii)* 内在维数与复杂度。- 定义数据空间后，重要的是找到一种合适的信息压缩度量，用作柯尔莫哥洛夫复杂度的估计量，$K(X)$。通常，数据集中的数据点表示为空间中的点，该空间的维数等于描述每个样本所需特征的数量，这称为嵌入维数。数据点之间的关联决定了数据所在的流形，从而形成一种结构，其维数通常远小于嵌入维数。这称为内在维数，记作 $I_d$。对于局部密度在最近邻尺度上近似恒定的数据集，$I_d$ 大致对应于描述数据集所需的最少变量数，因此是柯尔莫哥洛夫复杂度的直接估计量。利用图论中的一种可能估计量，已经证明了这种联系：$I_d$豪斯多夫维数 [30, 31]。

内在维数与复杂度之间的关系已在多个领域得到应用。在分子科学中，$I_d$ 用于分析大型分子结构数据集（如蛋白质结构）的复杂度，以深入了解决定分子功能的底层结构基序 [32, 33]。在计算机视觉中，它用于分析图像和视频，识别与图像分类或目标检测等任务相关的最重要特征和模式 [34, 35]。在物理学中，$I_d$ 用于表征量子系统和经典系统 [19, 36–38], ，估算描述给定物理系统所需的自由度数量，并研究其在相变过程中的普适性；其普适性已通过数值方法得到证明。

*(iii)* 内在维数的估计方法。-已有不同的方法被提出，用于估计 $I_d$[14, 39–42]。本文采用的技术——TWO-NN 方法 [43], 依赖于数据集中近邻元素间距离的统计特性。此类方法假设，最近邻点可视为从 $I_d$-维数超球面中均匀抽取。因此，可以建立 $I_d$ 与邻近距离统计特性之间的关系，这种方法尤其适用于非线性流形。在 TWO-NN 方法中，对数据集中的每个点 $X_i$ 计算比值 $\mu_i = r_i^{(2)}/r_i^{(1)}$，其中 $r_i^{(1)}$ 和 $r_i^{(2)}$ 分别是最近邻距离和次近邻距离。在上述假设下，可以证明 $\mu_i$ 服从

$$f(\mu) = I_d \mu^{-I_d - 1}. \tag{2}$$

因此，可以根据累积分布函数 $P(\mu)$ 估算 $I_d$：

$$I_d = -\frac{\ln\left[1 - P(\mu)\right]}{\ln \mu}. \tag{3}$$

图中展示了一个示例。7 针对三种情况：对 $I_d$ 的估计对应于直线的斜率。值得一提的是，$I_d$ 是一个依赖于尺度的量。特别是，改变 $N_r$的值，即数据集中构型的数量，相当于改变 $I_d$ 的值，但不会改变其函数依赖关系。此外，如上所述，$I_d$ 已被用于研究经典相变和量子相变，因为由于数据集中各点之间的关联发生变化，它在相变附近表现出奇异行为。本质上，某些相（例如伊辛模型中的铁磁相或 XY 模型中的 BKT相 [19]）的特征是其构型之间具有很强的关联，因为它们表现出相同的物理性质。相反，在这些有序相之外，随着我们逐渐远离临界点，数据之间的关联会越来越弱（从网络理论的角度来看，这些数据可以从 Erdős–Rényi网络中抽取）。这些效应反映在参数空间中 $I_d$ 的行为上。

最后，值得讨论一下 TWO-NN 方法中使用的距离定义。选择适当度量的依据是，它应当非负、仅当构型相同时才等于零、具有对称性，并满足三角不等式。本文将使用欧几里得度量来估算点与点之间的距离，因为这是在数据空间中引入几何学最自然的方式。在本例中，也可以等效地使用其他合适的距离。

*(iv)* 拓扑柯尔莫哥洛夫复杂度。现在，我们需要一个判据来判断复杂度与拓扑是否“相互关联”——也就是说，配分函数的复杂度能否揭示信息是否由拓扑激发承载。受拓扑场论基础中类似推理的启发，我们采用以下方法：给定一个模型，我们输入两种不同的拓扑类型（具体区别将在下文说明），并考察配分函数采样的复杂度是否发生变化，以及如何变化。

为简便起见，我们考虑 BKT 相（我们的论证可直接推广到具有不同拓扑缺陷的情形，例如液晶 [44]）。我们的出发点是配分函数

它对应于定义在具有 1-环面拓扑的流形上的 XY 模型，其采样和柯尔莫哥洛夫复杂度分别记为 $X^0$ 和 $K(X^0)$。接着，我们将后者与从同一模型定义在具有不同拓扑的流形上所得配分函数的采样 $X$ 中得到的复杂度进行比较。

其背后的依据如下 [45]。当拓扑自由度承载系统中的信息时，也就是说，在 XY模型的低温下，会出现一组新的约束，例如这些拓扑激发的绕数（w）。可以计算独立绕数的路径，同时也代表着对配分函数施加的不同且彼此独立的约束；这些路径与所考虑的特定流形严格相关。图 1iv) 展示了这些路径，其中不同的绕数表示在三维流形上：一维环面（$T^1$）、二维环面（$T^2$）和克莱因瓶。不同约束的存在必然会影响配分函数采样的复杂度。因此，预期可以通过研究其 $I_d$来观察到这一效应。

为了识别与拓扑相关的信息含量，我们将一维环面设为参考流形，并将问题在流形 $\mathcal{M}$ 上定义时的$_{\text{topo}}$ 拓扑柯尔莫哥洛夫复杂度 定义为由流形 $\mathcal{M}$上的配分函数计算得到的内在维数与一维环面上的内在维数之差:

$$K_{\text{topo}}^{(\mathcal{M})} = K^{(\mathcal{M})} - K^{(T)}. \tag{4}$$

下文中，我们通过数值结果表明，这一量能够显著区分具有拓扑起源的现象与不具有拓扑起源的现象。最重要的是，对于 $g$-环面，前者呈现出普适标度行为；对于更复杂的闭流形，它还与第一同调群具有特征性的联系，这证明了拓扑与复杂度之间存在明确无疑的关联。

## II. 模型

为阐释上述思想，我们采用了三种经典统计力学模型，正如前文所预期：伊辛模型、海森堡模型和 XY 模型。二维 XY 模型在温度 $T_{BKT} \simeq 0.8933$ [46–48] 下会发生一种（BKT）拓扑相变，并伴随着涡旋等拓扑缺陷的出现。物理自由度由平面上的转子表示，系统的能量可写为

$$E(\vec{\theta}) = -\sum_{\langle i,j \rangle} \vec{S}_i \vec{S}_j, \tag{5}$$

其中 $\vec{S}_i = (\cos\theta_i,\ \sin\theta_i)$ 和 $\theta \in [0, 2\pi[$。构型由以下方式表示

$$\vec{\theta} = \{\cos\theta_1, \sin\theta_1, \ldots, \cos\theta_N, \sin\theta_N\}, \tag{6}$$

其中 $N$ 是系统的总大小。

为了比较 XY模型激发的拓扑特征与缺乏此类现象的情形，我们研究了二维经典伊辛模型和经典海森堡模型。

伊辛模型可描述为

$$E(\vec{s}) = -\sum_{\langle i,j \rangle} s_i s_j \tag{7}$$

其中自由度为自旋 $s_i = \pm 1$ ，求和遍历晶格上的最近邻键。构型态定义为

$$\vec{s} = \{s_1, \ldots, s_N\}, \tag{8}$$

其中 N 是晶格中的自旋数。该模型在临界温度$T_c = 2/\ln(1+\sqrt{2})$处表现出二阶相变，其特征是 $\mathbb{Z}_2$ 对称性的破缺。

海森堡模型可描述为

$$E(\vec{s}) = -\sum_{\langle i,j \rangle} \vec{S}_i \vec{S}_j \tag{9}$$

其中求和遍历最近邻， $S_i = (\sin\theta_i \cos\phi_i,\ \sin\theta_i \sin\phi_i,\ \cos\theta_i)$ 是一个三分量自旋，其中 $\theta \in [0, \pi]$ 和 $\phi \in [0, 2\pi]$。该统计力学模型不会发生温度相变。

我们将伊辛模型和海森堡模型作为基准：由于它们不会表现出拓扑相变，因此可用来验证二维 XY 模型中柯尔莫哥洛夫复杂度的行为是由该模型特有的拓扑特征所致，而非其他不希望出现的效应（例如有限体积效应）。

## III. 复杂度与同调werhu GY

### A. 理论：绕数与信息压缩

在对自旋模型中的内在维数进行定量确定之前，先给出一个预示拓扑与复杂度之间联系的启发式论证是有益的。

考虑数据集 $X$ ，该数据集在方程（1）中引入。数据集 $X$ 的复杂度，即 $K(X)$，指出了数据中存储着什么信息，也就是系统能够编码多少信息内容。当对流形施加约束时，我们预期复杂度会随之降低，因为可以为快照赋予非平凡标签，我们将这些标签记作 $W_j$，$j = 1, \ldots\ldots, N_r$。用公式表示， $K(X)$可以写成裸数据集的复杂度与以下贡献之差:

裸数据集 $X^0$ 的复杂度，以及与每个 $X_j$ 相关联的标签 $W_j$ 所带来的贡献：

$$K(X) = K(X^0) - K(W|X^0). \tag{10}$$

这一观察与数据集所在流形的拓扑密切相关。首先，我们用一个简单的例子来说明这一点。考虑一组自旋变量，它们描述方格晶格上一个定义在 1-环面上的强有序伊辛铁磁态。自旋的构型在各处都与采用开放边界条件（OBC）的晶格所得构型类似，区别仅在于平面边界处的约束。这意味着，增加周期性边界条件会降低数据集的复杂度（从物理上说，这是一种几何效应，与开放边界条件会削弱有限空间区域内的磁有序有关 [49]）。然而，考虑闭流形时，我们预期在热力学极限下不会有任何差异：从伊辛铁磁体获得的样本，无论其底层流形如何，复杂度都相同。

如果磁态起源于 BKT 机制，情况预计会截然不同。在这种情况下，低能激发是涡旋：局域算符无法区分它们，而它们由绕数标记。这意味着，在低温下，配分函数所张成的构型空间受到强烈约束，因为涡旋必须具有特定的绕数。这与较高温度下的情况形成鲜明对比：此时涡旋不再束缚，可以自由传播（而且，随着具有极强局域自旋涨落的构型开始出现，绕数这一概念最终也失去了意义）。

低温展开。- 在低温相中，可以针对上述图景做出具体的定量预测。考虑对 $T \ll T_c$ 的 XY配分函数进行采样所得到的数据集，其中 $N_r \gg 1$，且定义在具有同调群 $\mathcal{G} = \bigoplus_{j=1}^{k} \mathbb{Z}_{\alpha_j}$ 的闭流形上。在这种情况下，大量构型的所有自旋都会沿着某个给定方向排列——在极端情况下，它们全部如此 $T = 0$。这样的簇的内在维数为 $\sim 1$，状态之间的平均距离为 $\propto \sqrt{N/N_r}$ [50]。我们将这个簇称为 $T = 0$ 簇，它具有恒定曲率：事实上，由于所有自旋都可以用一个角度来描述 $\theta$，因此它在局部可视为数据空间中的一条直线。

重要的是，这些并非被抽样到的唯一构型。其中也有相当一部分代表低能激发。它们可分为两类：(*i*)包含局域自旋激发的构型，其能量差为 $\delta\epsilon$，量级为 1（因此，相对于最低能态具有趋于零的能量密度）；以及 (*ii*)拓扑激发，其与最低能态之间的能量差同样有限，并具有确定的一组绕数。值得注意的是，拓扑激发尽管相较于固定方向的情形具有截然不同的自旋构型，却仍然具有趋于零的能量密度，这一点在此尤为重要。

整体数据结构及其复杂度会受到激发采样的显著影响。我们通过分析数据空间中起作用的各种长度尺度来仔细研究这些影响。为此，我们先聚焦于仅考虑构型局部变化的激发情形（情况（$i$））。由于这些变化只涉及少数几个自旋的改变，因此它们描述的是在$T = 0$处发现的一维流形的局部变形。根据最大似然推理，文献 [19]，估算了相应数据集的内在维数，其随 $I_d \propto \sqrt[N_r]{\langle N \rangle}$ 变化。该数据集不呈现聚类，可视为一个长度为 $\propto \sqrt{N}$ 的（高度各向异性的）多维圆柱体（其长度即 $\overline{T} = 0$ 簇的长度）。因此，我们将该数据集称为圆柱簇。圆柱体的宽度可按如下方式计算。给定两个构型 $\vec{\theta}^0$ 和 $\vec{\theta}^1$，前者表示一个所有自旋完全极化、且 $\theta_k^0 = \theta^0$ 的状态，后者表示在该极化态基础上叠加了热激发的状态，可得：

$$\begin{aligned} d(\vec{\theta}^0, \vec{\theta}^1) &= \sqrt{2\sum_{k-1}^{N}(1 - \cos(\theta^0)\cos(\theta_k^1) - \sin(\theta^0)\sin(\theta_k^1))} = \\ &= \left\{ 2 \left[ N - \cos(\theta^0) \sum_k \cos(\theta_k^1) - \right.\right. \\ &- \left.\left. \sin(\theta^0) \sum_k \sin(\theta_k^1) \right] \right\}^{1/2} \end{aligned} \tag{11}$$

鉴于 $\vec{\theta}^1$ 仅在少数几个被沿着 $\theta^0$方向排列的自旋所包围的自旋处与 $\vec{\theta}^0$ 不同，可以将上面的表达式改写为：

$$d(\vec{\theta}^0, \vec{\theta}^1) = \sqrt{2\sum_k (E^0 - E_k^1)} \tag{12}$$

其中，$E_0$ 是 $\vec{\theta}^0$的平均能量，而 $E_k$ 是 $\vec{\theta}^1$中与 $k$顶点相关联的能量。上述公式直接建立了激发能量与圆柱簇内平均距离之间的联系，即 $d_{\text{cyl}} \simeq \delta\epsilon =$ O(1)。

在没有拓扑激发的情况下，上述描述足以刻画完整的数据结构。拓扑激发会相当显著地改变这一情形（情形（$ii$））。如下文所示，通过分析构型之间的平均距离，可以发现它们会产生一些簇，这些簇在几何上与圆柱面簇彼此断开，其数量由系统所定义其上的流形的同调群维数决定。这种数据结构使我们得以刻画内在

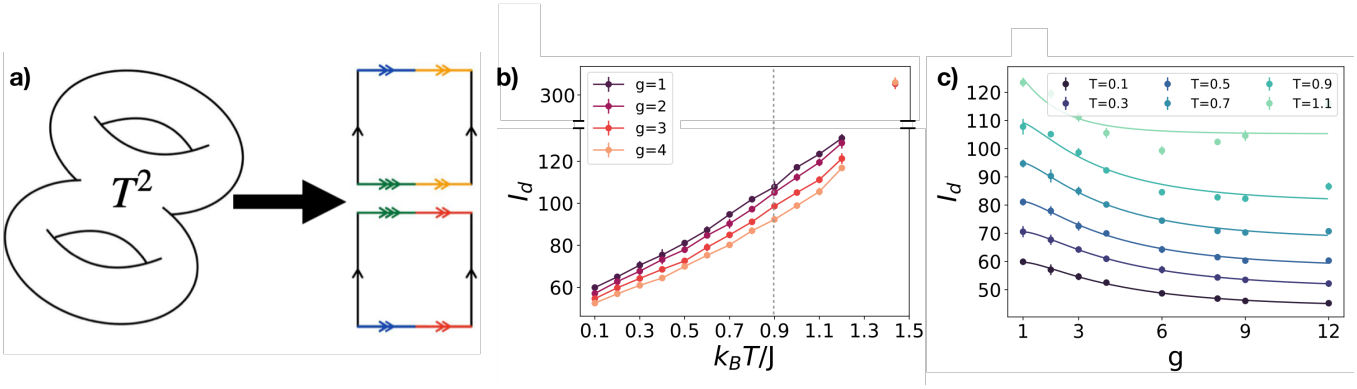


图2。 **绘制** $I_d$ **对于** $g$**个环面**。我们考虑方格晶格上的XY模型。a）二维环面（$T^2$）的构造示意图：颜色相同且带有箭头的边必须通过周期性边界条件连接。b）二维XY模型的 $I_d$ 随温度 $T$变化的曲线，亏格值逐渐增大 $g$ （从上到下）；c） $I_d$ 随亏格 $g$变化的标度关系，曲线对应于方程（14），温度从低到高依次排列（从下到上）。两个面板中，我们均考虑线性尺寸为 $L = 72$的方格晶格。我们对10次模拟的结果取平均，每次模拟包含 $N_r = 5000$ 种不同构型。误差条估计为均值标准差。竖直虚线标示临界点。

维数的拓扑依赖性，我们将在本节末对此作进一步讨论。

关键观察是， $d_{\text{topo}}$趋于零的能量密度下，圆柱簇与拓扑激发之间的平均距离远大于估算的 $d_{\text{cyl}}$。这是因为不能像此前那样从方程（11）推导到方程（12）：也就是说，对于拓扑激发，每个自旋构型都会随位置平滑变化，因此不能假设大多数邻居具有固定的磁化强度。相反，在这种情况下，由方程（11）可得

$$d_{\text{topo}} \simeq \sqrt{2\sum_k [1-\cos(\theta_k)-\sin(\theta_k)]} \simeq \sqrt{N} \qquad (13)$$

因为 $\cos(\theta_k)$ 和 $\sin(\theta_k)$的平均值相对于常数项可以忽略。从另一个角度来看，这可以理解为缺乏长程有序的结果：长程有序的缺失使$\cos(\theta_k)$ 的平均值在有限温度下为零。当然，对于任意低能态而言，这并非严格成立，但这为我们粗略估计式（13）中的标度不匹配提供了依据。

上述论证表明，拓扑激发会在数据空间中形成与主圆柱簇不同的“拓扑”簇。这类额外簇的数量也很重要：具体而言，该数量通常随理论中可能的低能激发数目而变化（这是因为这些簇之间的距离通常也为 order$L$）。对于闭流形，这一数量就是该流形的第一同调群的维数。

拓扑簇对流形内在维数的影响可以通过以下方式理解。内在维数估计仍然主要由簇圆柱面决定，因为后者包含了大部分的数据点，同时，该簇中点之间的最近邻距离小于拓扑数据点中的最近邻距离 [51]。

综上所述，我们的理论预测如下：(*i*) 在不存在拓扑激发的情况下，复杂度与拓扑无关；(*ii*) 在存在拓扑激发的情况下，复杂度会随着实空间流形的第一同调群的维数增加而成比例降低；(*iii*) 这种降低与聚类现象以及样本中最近邻距离分布的变化有关。

让我们基于这些预测，展望一些具体的预期。对于 $g$-环面情形，我们预期复杂度 $K(X)$ 将与 $2g$成正比地降低。值得注意的是，类似的标度关系也适用于量子统计力学中的拓扑相（例如，二维格点规范理论 [26]）。对于闭流形情形，我们则预期其复杂度取决于可能的环路数量，即第一同伦群的阿贝尔化维数（而非其贝蒂数）。例如，尽管克莱因瓶与1-环面的亏格不同，我们仍预期二者具有相同的复杂度。

总之，我们的理论仍有一些目前尚不知如何处理的方面。首先，尚不清楚这种图景在多高的温度范围内成立。第二个非平凡的问题是 $N_r$ 的作用：这里的主要挑战在于，最大似然论证（圆柱簇标度的基础）并不总能准确反映数据集大小所发挥的作用。为澄清这些问题，并佐证我们完整的理论图景，我们采用数值实验。

### B. 数值实验

现在，我们通过数值模拟来支持我们的理论。我们考虑一个二维方格晶格，并控制其边界条件，以实现拓扑上不同的流形。在图2a) 中，我们展示如何在两个方格晶格的边界上施加适当的约束来构造二维环面（$T^2$）。同样，也可以从一个八边形出发，在其相对边上施加周期性边界条件。推广到一般的 $g$维环面也很直接：可以从具有 $4g$ 条边的正多边形出发，也可以使用 $g$ 方格晶格。实际操作中，我们从一个方格晶格出发，将其划分为垂直条带，而不是采用 $g$ 个方格晶格。在补充材料第 C 节中，我们给出了数值证据，表明这两种方法等价。克莱因瓶、莫比乌斯带和实射影平面（RPP）等不可定向流形，可以通过在单个方格晶格上施加适当的边界条件来实现（见图3a）。

为了按照配分函数采样构型，我们使用 Wolff 算法 [52]。数据集定义为由以下内容组成的集合 $N_r$ 个点，如式（1）所示，每个点对应于上一节中指定的一个自旋构型。我们特别注意生成不相关的构型，以避免采样可能产生的任何偏差。为此，我们在蒙特卡罗步中保存的构型之间选择了合适的间隔。值得注意的是，据我们所知，此前尚未报告过其中一些流形的数值模拟，因此我们验证了自相关不会因流形拓扑而显著改变。

#### *1.* 可定向闭流形

我们分析的第一步是比较不同的 $g$-环面。亏格为 $g$ 的环面（$\mathrm{T}^g$）的非平凡回路数量为 $2g$，因为其第一同伦群为 $\pi_1(\mathrm{T}^g) = \bigoplus_{1=1}^{2g} \mathbb{Z}$。我们预期，在一个不具有拓扑特征的相中，复杂度不应表现出对 $g$ 的任何依赖。相反，在 BKT 相中，我们预期复杂度将取决于所考虑的特定流形。我们的理论预测，数据集 $X$ 的复杂度 $K(X, g)$ 随亏格 $g$ 的增加而单调递减，但会受到有限体积效应的影响。为了处理后者，我们采用如下现象学有限体积拟设：

$$K(X, g) = K_0 - a_0 \cdot g\mathrm{e}^{-a_1\sqrt{g}} \tag{14}$$

这会平滑掉较大 $g$ 处的贡献，因为这些贡献对有限体积效应更加敏感，其标度关系近似为 $V \simeq g^2$ [53]。在右侧，我们纳入一个与拓扑无关的项 $K_0$，以及一个与拓扑相关的项，后者将用于刻画 TKC。

在图 2 中，我们展示了估算得到的内在维数 $I_d$ 随不同温度变化的函数关系

亏格 $g$ (b) 的不同取值对应的数值，以及内在维数随不同亏格$g$ 变化的函数关系（c）。所考虑的系统为线性维数为 $L = 72$ 的方格晶格，数据集中的构型数为 $N_r = 5000$。如预期，曲线随温度单调递增。与文献 [19]，不同，内在维数无法检测到相变，这是因为所考虑的构型数较少；不过，正如我们下文所示，这一点与我们的分析无关。

图 2b) 中最显著的信号是，在$T/J \lesssim 1.2/k_B$以下，内在维数随亏格增加而降低，并且曲线之间的距离保持不变，至少在拓扑相深处是如此，此处我们的推理适用。我们在图 2c) 中仔细考察这种对亏格的依赖关系。低温下，数据点与拟设（通过基于式（14）进行最佳拟合得到的曲线）吻合极佳。相反，随着 $T$ 增加，吻合度会变差，并且在相变以上无法描述 $I_d$；此时拓扑性质已不再重要。我们认为，在热力学极限下，对于高于临界温度 $T > T_{BKT}$的所有温度，不同亏格对应的曲线都应重合。为检验这一假设，我们计算了 $I_d$ 在 $T = 1.4 > T_{BKT}$ 时的值，并观察到不同曲线的数值在误差条范围内彼此相容，如图2b) 所示。

#### *2.* 不可定向流形与开流形

对于可定向流形，我们根据非平凡回路的独立数量给出了预测；而对于不可定向流形，则需要适当地扩展我们的诊断方法。具体而言，可以考察流形的第一同调群，即其第一同伦群的阿贝尔化。我们预期：*i)* 定义在同调群维数相同的空间上的数据集，无论其亏格和可定向性如何，都具有相同的柯尔莫哥洛夫复杂度；*ii)* 在有限尺寸情形下，开放边界的数量可能会显著影响复杂度；*iii)* 在不具有拓扑性质的相中，复杂度不应随流形的拓扑特征而变化。

同调与复杂度。- 在图 3b) 中，我们绘制了 $I_d$，其数据来自线性尺寸为 $L = 72$的方格晶格上的 $N_r = 5000$ XY模型自旋构型，如式（1）所示。我们考虑自旋分别排列在实射影平面（RPP）、圆柱面、环面、克莱因瓶和莫比乌斯带上的情况（示意图见图 3a)）。相应的同调群列于表 I。这种拓扑类型的选择使我们能够分别考察同伦群和边界的影响。根据 TKC 的定义，我们采用一维环面作为参照：圆柱面有助于我们研究边界在估算 $I_d$中的作用，而其余对象则是非定向闭流形

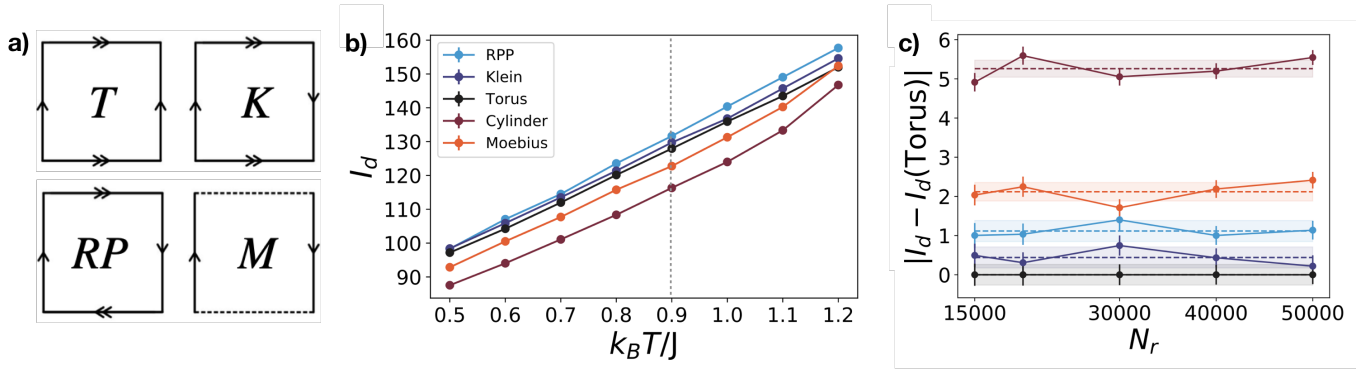


图 3. **RPP、克莱因瓶、环面、圆柱面和莫比乌斯拓扑的比较**。我们考虑方格晶格上的 XY模型。a) 环面（T）、克莱因瓶（$K$）、实射影平面（$RP$）和莫比乌斯带（$M$）的构造示意图。箭头方向相同的边界须通过周期性边界条件连接。箭头方向相反的边界表示反周期性边界条件。虚线表示开放边界条件。b) 在方格晶格上线性尺寸为 $L$ 时，$I_d$ 随温度变化并跨越 BKT 相变的结果；从上到下依次为 RPP、克莱因瓶 $\sim$ 环面、莫比乌斯带和圆柱面的结果。我们对 10 次模拟的结果取平均，$N_r = 5000$。误差条估算为均值标准差。竖直虚线标示临界点。c) 在 $K_bT/J = 0.1$ 处，$|I_d - I_d(\text{环面})|$ 随 $N_r$ 变化的结果；方格晶格的线性尺寸为 $L = 48$。虚线表示给定流形在 $N_r$ 上的 $I_d$ 平均值。误差条以相同颜色的阴影区域表示。从上到下依次为圆柱面、莫比乌斯带、RPP、克莱因瓶 $\sim$ 环面。

| 闭流形 | | | 开流形 | |
|---|---|---|---|---|
| RPP | 环面 | 克莱因瓶 | 圆柱面 | 莫比乌斯带 |
| $\mathbb{Z}$ | $\mathbb{Z} \oplus \mathbb{Z}$ | $\mathbb{Z} \oplus \mathbb{Z}_2$ | $\mathbb{Z}$ | $\mathbb{Z}$ |

表 I. 所考虑的各种拓扑类型的同调群

本节主要关注的正是这些流形。

从图 3b) 可看出，能够观察到的定性特征不多。首先，流形中存在边界会降低 $I_d$ 的值。这一点从莫比乌斯带（一个边界）和圆柱面（两个边界）的结果中可以清楚地看出，它们具有相同的同调群。这可以解释为，边缘的存在施加了约束，相当于向数据中加入更多标签（$W_j$）。其次，我们看到，在所考察的整个温度范围内，环面和克莱因瓶的 $I_d$ 相符，这与我们在上一节（第IIIA节）中的推理一致。为了量化这种联系，我们在图 3c) 中绘出了方格晶格在低温 $k_bT/J = 0.1$下，拓扑柯尔莫哥洛夫复杂度绝对值随 $N_r$ 变化的关系，其中 $L = 48$ 。对于每种拓扑，我们还估算了整个 $N_r$范围内 TKC 的平均值：图中以虚线表示，相应的估算误差则以阴影区域表示。

有两项主要观察。第一项涉及闭流形之间的相对表现：RPP 与一维环面之间存在显著差异。具体而言，无论采样方式如何，它们的 TKC 差异都与 1 相容。这一点值得注意，例如对于 $N_r = 50000$，整体内在维数超过 100。相反，克莱因瓶与一维环面之间的差异始终很小，且与 0 相容；在 3 种采样方式下皆是如此。因此，我们可以指出，支持拓扑激发的配分函数的 TKC 取决于流形的同调（例如，并非仅取决于其亏格）。第二项观察是，如预期所示，边缘的存在会显著影响 TKC。这一点显而易见：尽管莫比乌斯带、圆柱面和 RPP 的同调群维数相同或完全相同，它们的 TKC 却截然不同。

现在，我们将 XY模型的结果与海森堡模型的情况进行对比，后者如图所示。4。众所周知，在这种情况下，系统不会随温度变化发生相变。我们发现，对于任何拓扑类型，$I_d$ 的取值都彼此相容。我们认为，这是因为该模型在拓扑上是平凡的，其行为与 XY模型明显不同。在图4 的插图中，我们绘出了几个温度下的 $|I_d - I_d(\text{环面})|$，以表明这些曲线在误差条范围内彼此相容。在补充材料（第A节）中，我们还展示了伊辛模型的结果。我们可以得出类似的结论：内在维数的取值没有随拓扑类型变化而呈现明确趋势。事实上，不同拓扑类型下的所有取值在误差条范围内都彼此相容，因为伊辛模型的两个相在拓扑上都平凡。

### *3.* 条件连通性：数据空间中的局域序参量

为了验证上述论证，我们需要找出可检验的预测，将实空间中的拓扑激发与数据结构联系起来，并且

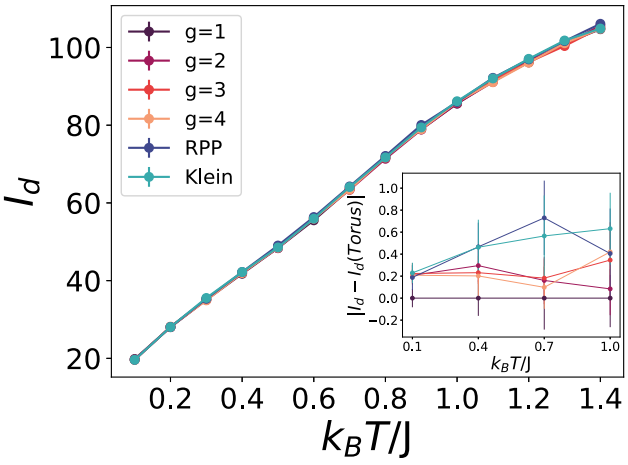


图 4. **绘制 $I_d$ 海森堡模型的结果**。我们考虑线性尺寸为 $L = 72$ 的方格晶格。我们考虑 $N_r = 5000$，并将 $I_d$ 估算为 10 次模拟的平均值。误差条估算为均值标准差。显然不存在任何拓扑特征。插图中，我们展示了所有拓扑类型的 $I_d$ 与环面这一参考情形之间的差异。我们观察到，各个数值在 1 个均值标准差的范围内相容。

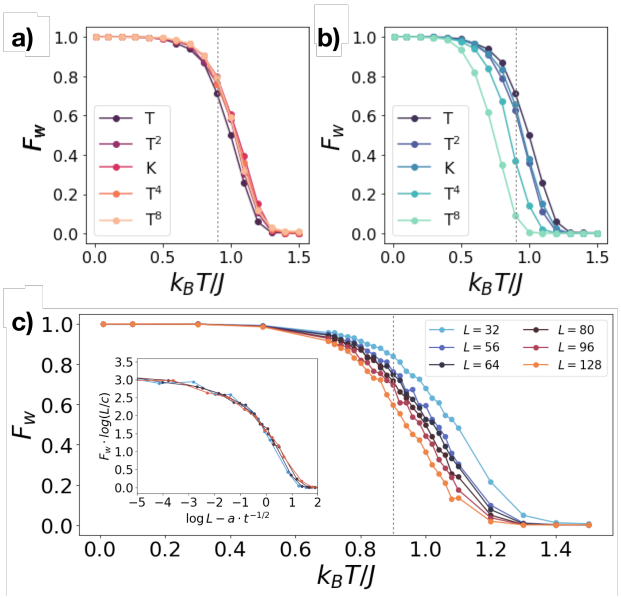


图5. **数据空间中前两个最近邻具有相同绕数的点所占比例 $F_{\mathbf{w}}$**。在面板 a) 中，我们只考虑两个非平凡回路。在面板 b) 中，我们考虑流形上的所有可能的非平凡回路。交叉点会根据图例所示的拓扑从右向左移动。方格晶格的线性尺寸为 $L = 72$。在面板 c) 中，我们考虑$F_{\mathbf{w}}$ 在不同系统尺寸 $L^2$下的一维环面情形，曲线从$L = 32$ （上方曲线）到 $L = 128$ （下方曲线）。这些结果表明，在低于临界温度时（由竖直虚线标出），所采样构型的局域 （连通性）性质会受到其全局 性质（绕数）的严重约束。这是数据空间中的一个局域序参量，在相变处表现出普适行为，如插图所示。这里 $t = K_b(T - T_c)/J$，其中 $T_c$ 的值设为 $K_bT_c/J = 0.8933$，而 $a = 1.767 \pm 0.003$ 和 $c = 0.721 \pm 0.003$则通过数据塌缩估算得到。

能够在研究配分函数样本时独立验证。根据上述分析，我们着重考察绕数在数据结构中的作用。

绕数 $w_\Gamma$ 是研究拓扑系统的关键量。在自旋系统中，它用于估计沿流形上的给定路径 $\Gamma$ 前进时，自旋绕行的次数。对于 XY模型，其中第 $i$ 个自旋由单个角度 $\theta_i$ 表示，绕数可写为：

$$w_\Gamma = \frac{1}{2\pi}\sum_{i=1}^{L_\Gamma} \Delta\theta_{i,i+1}, \tag{15}$$

其中，$\Gamma$ 由 $L_\Gamma$ 个格点组成，且 $\Delta\theta_{i,i+1} = \theta_{i+1} - \theta_i$ 被重新缩放到范围 $(-\pi, \pi]$。如图 1(iv) 所示，可用于计算绕数的独立路径取决于流形本身。我们将 $\mathbf{w} = \{w_{\Gamma_1}, \ldots, w_{\Gamma_k}\}$ 定义为在路径 $\Gamma_1, \ldots, \Gamma_k$ 上计算得到的 $w_\Gamma$ 的集合。例如，对于环面（$T$）和二维环面（$T^2$），分别有 $\mathbf{w}(T) = \{w_a, w_b\}$ 和 $\mathbf{w}(T^2) = \{w_{a_1}, w_{b_1}, w_{a_2}, w_{b_2}\}$（见图 1(iv)）。值得注意的是，$\mathbf{w}$ 是理解相空间中构型聚类结构的关键指标，因为绕数能够捕捉激发的拓扑性质。也就是说，在 BKT相变之上，涡旋-反涡旋对不受束缚，导致采样构型之间的 $\mathbf{w}$ 值分布范围很广。然而，在准长程有序区域（低于 $T_{BKT}$），大多数构型具有相同的 $\mathbf{w}$；也就是说，对于环面拓扑，$\mathbf{w} = (0, 0)$。

根据上述理由，BKT 相变的物理性质会强烈影响相邻构型之间的连通性。特别是，第一和第二近邻具有相同$\mathbf{w}$的点所占比例，即我们称为条件连通性 $F_{\mathbf{w}} = N_{\mathbf{w}}/N_r$的量，必然能够反映相变情况。这里，$N_r$ 是采样得到的构型总数，$N_{\mathbf{w}}$ 是前两个近邻具有相同$\mathbf{w}$的构型数。显然，在 BKT 相变之上（构型呈现出广泛的$\mathbf{w}$取值），$F_{\mathbf{w}}$ 接近于零；而在拓扑相中（由于激发的拓扑性质，$\mathbf{w}$ 的取值受到限制，构型在数据空间中呈现为簇），$F_{\mathbf{w}}$ 则等于 1。

在图5a) 中，我们展示了 $F_{\mathbf{w}}$ 每个流形仅考虑晶格上的两条路径的情况，这与图5b) 形成对比，后者考虑了所有可能的非平凡路径。我们观察到，这两幅图在定性方面呈现出相同的

行为呈现出类似的特征，在临界温度附近发生转变，但在定量上有所不同。这是可以预期的，因为在 BKT相中，可能存在的独立绕行数量越多，系统受到的约束就越强。在图 5c) 中，我们展示了 $F_{\mathbf{w}}$ ，对应于 a) 中讨论的情况，但采用了不同的系统尺寸。在插图中，我们使用估计参数 $c = 0.72(1)$ 和 $a = 1.76(7)$ 对曲线进行塌缩。为此，我们为这些参数选取了合适的网格，并固定临界温度 $k_B T_c/J = 0.8933$ [48]。随后，我们计算标度变量

$$x = \ln L - at^{-1/2}, \quad y = F_{\mathbf{w}} \ln(L/c) \tag{16}$$

并选择参数函数假设 $f(x,\ a,\ c)$，与 $y$进行比较。这里 $t = (T - T_c)/T_c$。我们使用 Levenberg–Marquardt算法，为每种 $a$，$c$ 和 $f$的取值计算最佳拟合，其中 $f$被视为次数可变的多项式，次数为 $k = 5$，6，7，8。我们对 $a$ 和 $c$的估计值取所有最佳拟合的平均值，其误差则为这些拟合值的标准差。我们观察到，条件连通性 $F_{\mathbf{w}}$呈现出普适行为。它可被视为数据空间中的一个序参量，并且正如前文所述，与柯尔莫哥洛夫复杂度在 BKT相中呈现拓扑贡献这一事实密切相关。事实上，局部连通性（由$I_d$估算）受到流形全局性质的强烈约束，而这些性质体现于 $F_{\mathbf{w}}$的行为。上述讨论使我们得以明确提出两点，并在此总结：i）相邻构型之间的连通性与 BKT相变的物理性质密切相关。值得注意的是，当系统处于有序（拓扑）相时，最近邻构型具有相同的物理性质（绕数）。相反，在无序相中，第一邻居和第二邻居的物理性质完全随机，这一点可由 $F_{\mathbf{w}}$趋近于零看出；ii）改变所计算并比较的 $w$的数量后，随着对构型施加更多约束，转变点（$F_{\mathbf{w}}$ 趋近于 1）会发生移动。因此，不同流形上的柯尔莫哥洛夫复杂度 $K(X)$ 应当呈现不同的行为，这一点此前已得到明确展示。

## IV. 讨论与结论

我们提出了一种理论，将经典统计力学的拓扑性质与根据问题配分函数对相应多体态进行统计采样的柯尔莫哥洛夫复杂度联系起来。我们提出的原始方法受到拓扑场论的启发：取一个流形，在该流形本身上定义的模型中对配分函数进行采样，并得到一个不变量。通过在一维环面上规范化这一过程，我们定义了与给定流形上给定问题相关联的拓扑柯尔莫哥洛夫复杂度。基于我们所建立的联系实空间中的拓扑性质与构型空间中的数据结构之间，我们的理论总体结论是：多体系统受到拓扑效应的约束越强，其采样的柯尔莫哥洛夫复杂度就越低。

随后，我们在定义于方格晶格上的经典统计力学模型——伊辛模型、海森堡模型和 XY模型——框架下探讨了我们的理论。在 XY模型的贝列津斯基-科斯特利茨-索利斯相中，配分函数的复杂度强烈依赖于目标流形的同调，展现出拓扑与复杂度之间强有力的定性联系。最重要的是，拓扑柯尔莫哥洛夫复杂度在 BKT相中取得了普适值（即与采样无关），这表明它可能与流形中的自旋绕数存在直接联系。尽管其中一些数值似乎与整数拓扑柯尔莫哥洛夫复杂度相符，但我们认为仍需开展更深入的研究，以处理比本文所考虑数据集更大的数据集。这将需要发展新的内在维数估计方法，尤其要考虑误差问题。在数值实验中，我们还观察到复杂度对边界具有强烈依赖性。最后，在高温下，以及在伊辛模型和海森堡模型中，我们没有观察到拓扑与复杂度之间的联系，这与预期一致。

在 XY模型中，我们的理论通过分析度量距离与绕数之间的相互作用，使我们能够追溯复杂度与拓扑类型之间关系的起源。具体而言，我们确定了一种序参量——条件连通性。尽管它在数据空间中完全是局部的，却能正确预测发生在实空间中的拓扑相变。该序参量符合唯象标度关系，这进一步证明了它与 BKT相变之间的联系。这种解释或许有助于理解为何诸如动力学映射和持久同调等学习方法，在 XY模型 [54–57], 中通过专门利用绕数信息，取得了非常令人信服的应用成果。事实上，这些方法可能间接检测到了我们提出的简单序参量。值得一提的是，尽管奇异拓扑类型通常似乎难以实现，但它们在量子模拟器中已成为实验现实。事实上，人们可以通过多种方式实现它们，包括合成维度 [58, 59], 原子构型，以及利用光镊进行原子重排；后一种方式已得到实验验证 [60]。

总而言之，我们的理论和数值实验支持这样一个观点：实空间中的拓扑信息在数据空间中会变得“局域”。这一概念性洞见可作为起点，帮助我们探索复杂度与拓扑之间的相互作用如何超越平衡态经典统计力学。一个明确的研究方向是非平衡物理，原则上，我们在此使用的同一套工具可以直接应用于这一领域。另一个研究前沿是量子平衡系统。在这种

情况下，信息与拓扑之间的联系已在拓扑物质中得到深入应用，例如拓扑纠缠熵的构想 [26]。然而，这类量通常难以在大规模实验中检验，因为它们需要掌握整个系统的波函数。像本文所采用的统计方法，或许能够揭示拓扑序和/或拓扑效应的迹象，即使采样有限但具有统计显著性，只要实验能够动态调控拓扑（例如，参见文献 [61, 62]）。不过，必须强调的是，尽管我们的工作为这一方向提供了基础工具（例如，我们研究的配分函数可以视为具有分数化激发的一维流体的路径积分），但要进军量子领域，必然需要采用更复杂的方法。这不仅是因为拓扑在该领域有不同的表现形式（例如拓扑序、能带结构和临界现象），更重要的是，量子系统会在多个基底中提供数据集。目前尚待解决的一个问题是，如何巧妙地结合这些数据，并可能辅以随机采样 [63]，，从而建立拓扑与复杂度之间的直接联系。

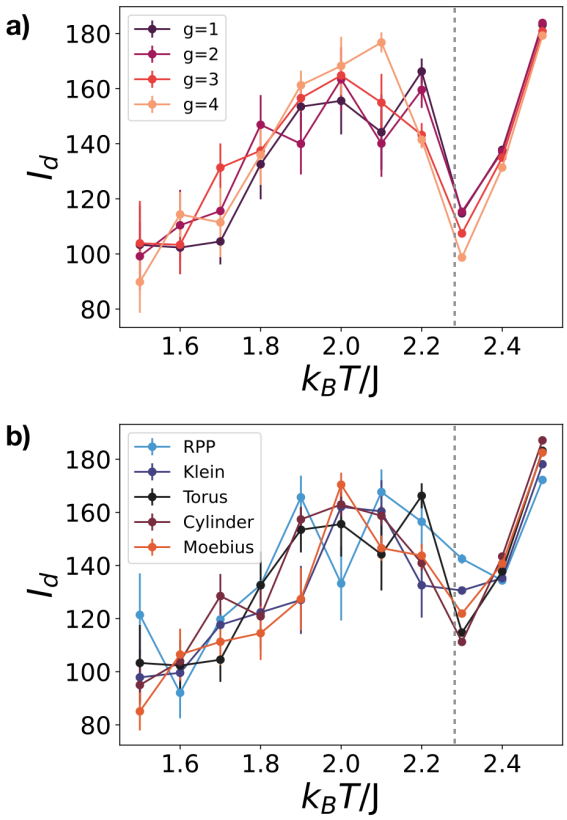


图 6. **的图伊辛模型**。我们考虑线性尺寸为 $L = 72$ 的方格晶格。我们考虑 $N_r = 10^4$，并将 $I_d$ 估算为 10 次模拟的平均值。不同拓扑类型下的 $I_d$ 值。对于 a) 和 b)，我们观察到所有数值在误差条范围内均相容，因为伊辛模型的两个相都属于拓扑平凡相。临界点以一条竖直虚线标出。

我们采集了 $N_r = 10^4$ 个快照，并将 $I_d$ 估算为 10 次模拟的平均值。我们观察到，$I_d$ 如预期般在临界温度附近呈现明显的局部极小值，但并不取决于系统所处流形的拓扑类型。我们发现，所有数值在误差条范围内均相容，因为伊辛模型的两个相在拓扑上都是平凡的。此外，在转变点处有限尺寸效应起着重要作用，此时大多数构型之间的差异与 XY模型中的情况截然不同。在 XY模型中，克莱因瓶和环面的复杂度在误差条范围内相当，这与它们的同伦群有关；但在此处，两者的复杂度却相差甚远。


### 致谢

我们感谢 R. Panda、R. Verdel 和 E. Vicari 提供的有益讨论，并感谢 A. Angelone、M. Heyl、S. Pedrielli 和 M. Schmitt 在相关主题上的合作。V.V. 感谢 P. Pegolo 和 G. Giachetti 的讨论。M. D. 和 V. V. 的工作部分由 ERC 资助，资助编号为 758329（AGEnTh）；同时也得到 MIUR FARE 计划（MEPH）以及 PNRR MUR 项目 PE0000023-NQSTI 的支持。V.V. 感谢通过意大利超级计算资源分配计划的 ISCRA 项目 ICTP21_CMSP 和 ICTP22_CMSP 获得的 Cineca 超级计算中心计算资源。M.D. 的工作部分得到慕尼黑天体物理、粒子物理与生物物理研究所（MIAPbP）的支持，该研究所由德国研究联合会（DFG）资助，属于德国卓越战略项目 EXC-2094（项目编号 390783311）。TMS 的工作由欧洲研究理事会（ERC）通过欧盟“地平线 2020”研究与创新计划资助（资助协议编号 853443）。


### 附录 A：伊辛模型结果

在此，我们将 XY模型的结果与伊辛模型的情况进行对比。具体而言，我们展示了$I_d$ 不同亏格的环面 $g$ （图6a））以及其他拓扑类型（图6b））上的结果随温度的变化。我们研究了方格晶格上的伊辛模型，其中 L=72。

### 附录 B：帕累托分布检验

TWO-NN 方法的基本假设是，$\mu_i = r_i^{(2)}/r_i^{(1)}$ （次近邻距离与最近邻距离之比）服从帕累托分布，即 $f(\mu) = I_d \mu^{-I_d-1}$。在此，我们针对几种情形检验这一假设

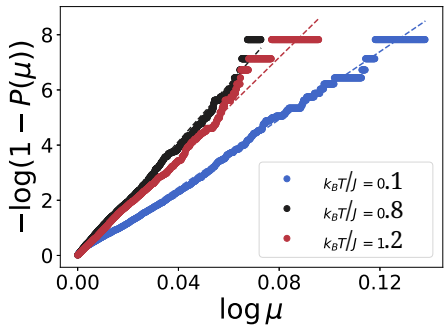


图 7. **帕累托分布检验**。这里我们考虑一个具有 $L = 72$ 和 $N_r = 2500$ 的方格晶格。这里 $t = k_bT/J$。

XY模型的情形，我们固定温度 $k_bT/J = 0.1$, $k_bT/J = 0.8$ 并取 $k_bT/J = 1.2$，然后计算 $I_d$。接着，我们将与估计得到的 $I_d$ 相对应的累积分布，与根据数据集中的距离分布计算得到的累积分布进行比较。我们得到图7中的曲线。可以明显看出，曲线（精确的累积分布函数）与点（根据样本构型得到的累积分布）完全吻合。

### 附录 C：字符串与正方形

在正文中，我们描述了用于计算 $I_d$的方法。简而言之，我们通过 Wolff 算法对构型进行采样，并通过适当调整边界条件来实现不同的流形。

我们发现，对于 $g$ 环面，从计算角度来看，最简便的方法是考虑一个方格晶格，将其划分为 $g$ 条带，并施加必要的周期性边界条件，而不是考虑 $g$ 个方格晶格。这样，我们便能轻松地为每个亏格选取相同数量的晶格点，即相同的晶格体积，$g$ 并且也能构造奇数$g$ 环面。本节中，我们比较 $g = 4$ 时，采用 4 个方格晶格计算得到的 $I_d$ 与将一个方格晶格划分为 4 条带后计算得到的结果。图 8 上图比较了 $48^2$个晶格点和 $72^2$ 个晶格点在两种不同几何结构下的结果（针对 XY模型）。我们发现，尽管二者有所不同

在临界温度附近（由于$k_bT/J = 1$ 有限尺寸修正，临界温度发生了偏移），随着系统尺寸增大，两种几何结构之间的一致性也随之提高。这是因为在转变点处，关联长度会增大，因此有限尺寸效应的影响要显著得多。不过，转变点处的这些效应并不影响正文讨论的目的，因为我们的理解适用于拓扑相。在图 8 下图中，我们明确展示了两个 $I_d$ 的差异

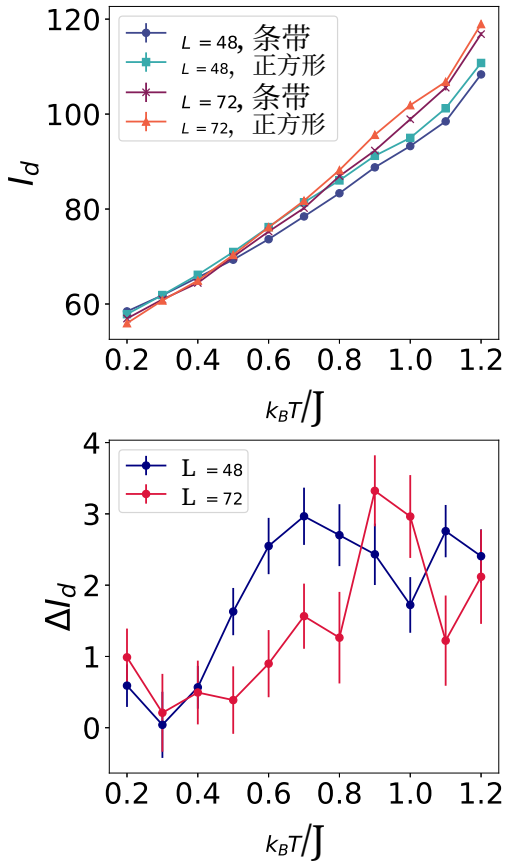


图 8. **获取一个4维环面**的不同方法之间的比较。比较带有 $L = 24$ 的 4 个正方形晶格与线性尺寸为 48 和 12 的矩形晶格；以及带有 $L = 32$ 的方格晶格与线性尺寸为 72 和 18 的矩形晶格。

趋近于 $\Delta I_{d0}$。我们观察到，在系统尺寸较大的情况下，$\Delta I_d$ 开始增长的时间更晚，这证明 $\Delta I_d > 0$是有限尺寸效应。

[1]J. Ladyman、J. Lambert 和 K. Wiesner，《什么是复杂系统？》，《欧洲科学哲学杂志》**3**，33（2013 年）。[2] M. Li、P. Vitányi，等，《柯尔莫哥洛夫复杂度及其应用导论》，第 3 卷（施普林格，2008 年）。

[3] S. Carroll，宇宙中最伟大的思想 | 23。临界性与复杂度。

[4] 升格于3/4 级临界性：重正化群与相变的基础理论的拓扑科尔莫哥洛夫复杂性及别列津斯基–科斯特利茨–索利斯机制

[4] 参金昌期、J. Healy、以及降维的均匀流形近似与投影，arXiv:1802.03426 (2018)。

[5] M. Kaya 和 H. Ş. Bilge，《深度度量学习：综述》，Symmetry 11, 1066 (2019)。[6] B. Kulis 等，《度量学习：综述》，Foundations and Trends in Machine Learning 5, 287 (2013)。[7] G. Carleo、I. Cirac、K. Cranmer、L. Daudet、M. Schuld、N. Tishby、L. Vogt-Maranto 和 L. Zdeborová，《机器学习与物理科学》，Review Mod. Physics 91, 045002 (2019)。[8] J. Carrasquilla，《量子物质的机器学习》，Advances in Physics : X 5, 1797528 (2020)。[9] P. Mehta、M. Bukov、C.-H. Wang、A. G. Day、C. Richard-son、C. K. Fisher 和 D. J. Schwab，《面向物理学家的机器学习：高偏差、低方差入门》，Physics Reports 810, 1 (2019)，《面向物理学家的机器学习：高偏差、低方差入门》。[10] A. Jović 、 K. Brkić 和 N. Bogunović，《特征选择方法及其应用综述》，载于2015年第38届信息与通信技术、电子与微电子国际会议（MIPRO）（Ieee，2015），第1200–1205页。[11] J. Cai、J. Luo、S. Wang 和 S. Yang，《机器学习中的特征选择：新视角》，Neurocomputing 300, 70 (2018)。[12] X. Deng、Y. Li、J. Weng 和 J. Zhang，《文本分类的特征选择：综述》，Multimedia Tools and Applications 78, 3797 (2019)。[13] A. Glielmo、C. Zeni、B. Cheng、G. Csányi 和 A. Laio，《距离度量信息内容的排序》，PNAS Nexus 1, pgac039 (2022)。[14] L. Van der Maaten 和 G. Hinton，《使用 t-SNE 可视化数据》，Journal of Machine Learning Research 9 (2008)。[15] Y. Bengio、A. Courville 和 P. Vincent，《表征学习：综述与新视角》，IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 1798 (2013)。[16] L. Wang，《利用无监督学习发现相变》，Physical Review B 94, 195105 (2016)。[17] W. Hu、R. R. P. Singh 和 R. T. Scalettar，《通过无监督机器学习发现相、相变和交叉转变：批判性考察》，Physical Review E 95, 062122 (2017)。[18] S. J. Wetzel，《相变的无监督学习：从主成分分析到变分自编码器》，Physical Review E 96, 022140 (2017)。[19] T. Mendes-Santos、X. Turkeshi、M. Dalmonte 和 A. Rodriguez，《利用内在维数进行无监督学习，揭示普适临界行为》，Physical Review X 11, 011040 (2021)。[20] Z. Hadzibabic、P. Krüger、M. Cheneau、B. Battelier 和 J. Dalibard，《囚禁原子气体中的别列津斯基–科斯特利茨–索利斯交叉》，Nature 441, 1118 (2006)。[21] Z. Han、A. Allain、H. Arjmandi-Tash、K. Tikhonov、M. Feigel'Man、B. Sacépé 和 V. Bouchiat，《锡–石墨烯混合约瑟夫森结阵列中超导性的崩塌》，Nature Physics 10, 380 (2014)。[22] A. D. King、J. Carrasquilla、J. Raymond、I. Ozfidan、E. Andriyash、A. Berkley、M. Reis、T. Lanting、R. Harris、F. Altomare 等，《在含1800个量子比特的可编程晶格中观测到拓扑现象》，Nature 560, 456 (2018)。[23] V. Drouin-Touchette，《科斯特利茨–索利斯相变：勇敢学生入门》，arXiv:2207.13748 (2022)。[24] H.-H. Tu，《克莱因瓶上共形临界理论的普适熵》，《物理评论快报》119，261603（2017）。[25] W. Tang、L. Chen、W. Li、X. Xie、H.-H. Tu 和 L. Wang，《共形场论中的普适边界熵：一项量子蒙特卡洛研究》，Physical Review B 96, 115136（2017）。[26] A. Kitaev 和 J. Preskill，《拓扑纠缠熵》，《物理评论快报》96，110404（2006）。[27] M. Levin 和 X.-G. Wen，《在基态波函数中检测拓扑序》，《物理评论快报》96，110405（2006）。[28] P. Fromholz、G. Magnifico、V. Vitale、T. Mendes-Santos 和 M. Dalmonte，《一维拓扑超导体的纠缠拓扑不变量》，Physical Review B 101，085136（2020）。[29] T. Micallo、V. Vitale、M. Dalmonte 和 P. Fromholz，《Su-Schrieffer-Heeger模型中非连通分区的拓扑纠缠性质》，SciPost Phys. Core 3，012（2020）。[30] T. Mendes-Santos、M. Schmitt、A. Angelone、A. Rodriguez、P. Scholl、H. Williams、D. Barredo、T. Lahaye、A. Browaeys、M. Heyl 等，《量子多体系统的波函数网络描述与柯尔莫哥洛夫复杂度》，arXiv:2301.13216（2023）。[31] L. Staiger，《柯尔莫哥洛夫复杂度与豪斯多夫维数》，《信息与计算》103，159（1993）。[32] K. T. Butler、D. W. Davies、H. Cartwright、O. Isayev 和 A. Walsh，《用于分子与材料科学的机器学习》，Nature 559，547（2018）。[33] E. Facco、A. Pagnani、E. T. Russo 和 A. Laio，《蛋白质序列演化的内在维数》，《公共科学图书馆·计算生物学》15，e1006767（2019）。[34] S. Gong、V. N. Boddeti 和 A. K. Jain，《关于图像表征的内在维数》，载于《IEEE/CVF计算机视觉与模式识别会议论文集》（2019），第3987–3996页。[35] P. Pope、C. Zhu、A. Abdelkader、M. Goldblum 和 T. Goldstein，《图像的内在维数及其对学习的影响》，arXiv:2104.08894（2021）。[36] R. Verdel、V. Vitale、R. K. Panda、E. D. Donkor、A. Rodriguez、S. Lannig、Y. Deller、H. Strobel、M. K. Oberthaler 和 M. Dalmonte，《量子模拟器中相关信息的数据驱动发现》，（2023），arXiv:2307.10040。[37] T. Mendes-Santos、A. Angelone、A. Rodriguez、R. Fazio 和 M. Dalmonte，《路径积分的内在维数：数据挖掘量子临界性与涌现的简化》，PRX Quantum 2，030332（2021）。[38] X. Turkeshi，《测量诱导临界性作为一种数据结构转变》，Physical Review B 106，144313（2022）。[39] F. Ca mastra 和 A. Staiano，《内在维数估计：进展与未解问题》，《信息科学》328，26（2016）。[40] P. Campadelli、E. Casiraghi、C. Ceruti 和 A. Rozza，《内在维数估计：相关技术与基准测试框架》，《数学问题工程》2015（2015）。[41] M. A. Kramer，《使用自联想神经网络进行非线性主成分分析》，《AIChE期刊》37，233（1991）。[42] S. Wold、K. Esbensen 和 P. Geladi，《主成分分析》，《化学计量学与智能实验室系统》2

' 37（1987）。

[43]E. Facco、M. d'Errico、A. Rodriguez 和 A. Laio，《通过最小邻域信息估计数据集的内在维数》，《科学报告》7，1（2017）。[44] M. Nakahara，《几何学、拓扑与物理学》（CRC出版社，2018）。[45] 顺便提一下，这种推理在一定程度上受到参考文献 [27] 所用方法的启发，该方法旨在识别量子力学系统中熵的拓扑贡献。关键区别在于，这里讨论的不是划分，而是不同的流形。[46] V. Berezinsky，《具有连续对称群的一维和二维系统中长程有序的破坏，1. 经典系统》（1971），《苏联物理学：JETP》32，493。[47] J. M. Kosterlitz 和 D. J. Thouless，《二维系统中的有序、亚稳态与相变》，《物理学杂志 C：固体物理学》6，1181（1973）。[48] A. W. Sandvik，《量子自旋系统的计算研究》，载于《AIP会议论文集》，第1297卷（美国物理联合会，2010），第135–338页。[49] M. Henkel，《共形不变性与临界现象》（施普林格科学与商业媒体，1999）。[50]请读者注意，两个构型 $\vec{\theta}^i$ 与 $\vec{\theta}^j$ 之间的欧几里得距离 [19]为：

$$d(\vec{\theta}^i,\vec{\theta}^j)=\sqrt{2\sum_{k=1}^{N}(1-\vec{S}_k^i\vec{S}_k^j)}. \tag{C1}$$

.

[51] 我们注意到，这一图景或许是一个有趣的出发点，可用于探究我们所用的 2-NN估计器之外的流形，例如考虑固定长度尺度估计器 [64]。

[52] U . Wolff，《用于自旋系统的集体蒙特卡罗更新》，《物理评论快报》**62**，361（1989）。

[53] 我们或许可以通过考虑系统几何中的各向异性，进一步改进这一拟设。然而，对于我们感兴趣的系统尺寸，要对此作出具体判断（因为我们的 $I_d$ 估计量随嵌入维数增长而面临严峻的尺度问题）颇具挑战，因此我们将这一方面留待未来研究。

[54] J. F. Rodriguez-Nieva 和 M. S. Scheurer，《通过无监督机器学习识别拓扑序》，Nature Physics 15, 790 (2019). [55] Q. H. Tran、M. Chen 和 Y. Hasegawa，《用于相变的拓扑持久性机器》，Physical Review E 103, 052127 (2021). [56] N. Sale、J. Giansiracusa 和 B. Lucini，《利用持久同调定量分析二维 $xy$ 模型中的相变》，Physical Review E 105, 024121 (2022). [57] H. Otsuka、K. Shiina 和 Y. Okabe，《关于 BKT 相变普适性的综合研究——机器学习研究、蒙特卡洛模拟与能级谱方法》，Journal of Physics A: Mathematical and Theoretical (2023). [58] O. Boada、A. Celi、J. Latorre 和 M. Lewenstein，《额外维度的量子模拟》，Physical Review Letters 108, 133001 (2012). [59] O. Boada、A. Celi、J. Rodríguez-Laguna、J. I. Latorre 和 M. Lewenstein，《非平凡拓扑的量子模拟》，New Journal of Physics 17, 045007 (2015). [60] D. Bluvstein、H. Levine、G. Semeghini、T. T. Wang、S. Ebadi、M. Kalinowski、A. Keesling、N. Maskara、H. Pichler、M. Greiner 等，《基于纠缠原子阵列相干输运的量子处理器》，Nature 604, 451 (2022). [61] F. Nogrette、H. Labuhn、S. Ravets、D. Barredo、L. Béguin、A. Vernier、T. Lahaye 和 A. Browaeys，《在具有任意几何形状的全息二维微型光阱阵列中俘获单原子》，Physical Review X 4, 021034 (2014). [62] S. Ebadi、A. Keesling、M. Cain、T. T. Wang、H. Levine、D. Bluvstein、G. Semeghini、A. Omran、J.-G. Liu、R. Samajdar 等，《利用里德伯原子阵列对最大独立集进行量子优化》，Science 376, 1209 (2022). [63] A. Elben、S. T. Flammia、H.-Y. Huang、R. Kueng、J. Preskill、B. Vermersch 和 P. Zoller，《随机测量工具箱》，Nature Review Physics 5, 9 (2023). [64] I. Macocco、A. Glielmo、J. Grilli 和 A. Laio，《离散度量的内在维数估计》，Physical Review Letters 130, 067401 (2023).