用于衡量高维系统复杂性的希尔伯特熵

Seong-Gyun Im [1,†]、Taewoo Kang [1,†]和 S. Joon Kwon [1,2,3,4*]

[1]成均馆大学化学工程学院（SKKU），大韩民国[2]成均馆大学能源科学与技术研究院（SIEST），成均馆大学，大韩民国[3]成均馆大学半导体融合工程系，大韩民国[4]成均馆大学未来能源工程系，大韩民国

* 通讯作者：sjoonkwon@skku.edu

† 这些作者对本研究的贡献相同。

摘要

衡量物理系统中高维数据的复杂性，是决定系统信息与质量的关键因素。然而，李雅普诺夫指数、分形维数和信息熵等传统度量方法，在衡量具有上下文的高维数据时存在局限，因为它们无法揭示物理系统的内在本质。在此，我们提出一种新方法：借助希尔伯特曲线等空间填充曲线，并结合广义熵测度，在降维的同时保留上下文，从而量化高维数据的复杂性。我们通过测量临界现象（包括自旋和渗流模型中的相变）验证了这一方法。研究结果表明，希尔伯特熵与理论相变点高度吻合。此外，我们进一步探究了希尔伯特熵与分形维数之间的潜在联系，例如，在尺度不变的二维/三维几何结构中，标度指数与欧几里得维数呈线性关系。本方法为理解和分析高维复杂系统提供了一个颇具前景的新框架，并有望应用于物理学的多个领域。

动力系统的复杂性已在统计物理学和非线性动力学领域得到深入研究[1–4]。具体而言，有研究提出，对于具有混沌特征的低维系统（例如时间序列数据），可以通过李雅普诺夫指数[5,6], 、分形维数[7,8], 以及 Rényi 熵的各种变体[9–14]等指标揭示其复杂本质。这些用于衡量动力系统复杂性的传统指标能够量化系统的部分特征；然而，在揭示高于一维系统的更高维度的内在特性方面，它们有所局限。已有若干尝试将现有的一维复杂度指标拓展至二维[15–18]。例如，可以采用基于图同构的方法衡量颗粒系统的拓扑复杂度[19,20]。还应注意的是，这种拓展尚未应用于三维或更高维度。然而，将复杂度指标简单拓展到更高维度的情境，尤其是涉及二维或更高维度的情境，会面临诸多挑战，例如数值伪影和局部信息丢失。这些挑战使得简单拓展不适用于衡量高维数据的复杂性。为解决这一问题，理想的做法是先将高维数据降至一维数组，再应用一维复杂度指标。遗憾的是，将高维数据简单转换为一维向量，并不能有效处理原始数据的局部性问题[21]。为解决局部性问题，可以采用基于空间填充曲线（SFC）的数据映射。SFC（或皮亚诺曲线）由皮亚诺于1890[22], 提出，是一种连续曲线，能够在离散格点或单元中逐一遍历每个点。顾名思义，SFC 使沿曲线连接的点数组在数学上等价于原始高维数据，同时保留局部性。在众多知名的 SFC 中，与其他 SFC 相比，希尔伯特曲线[23] 在大多数情况下被认为最能保留局部性，因为它在高维晶格中实现了最近邻之间的局部连通性[21,24]。因此，基于 SFC 的方法

提供了一种在数学上合理的降维方式。对于降维后的数据，可以根据其内在属性（如二元、连续、离散、逻辑等）通过熵来衡量其复杂性。

基于空间填充曲线的降维

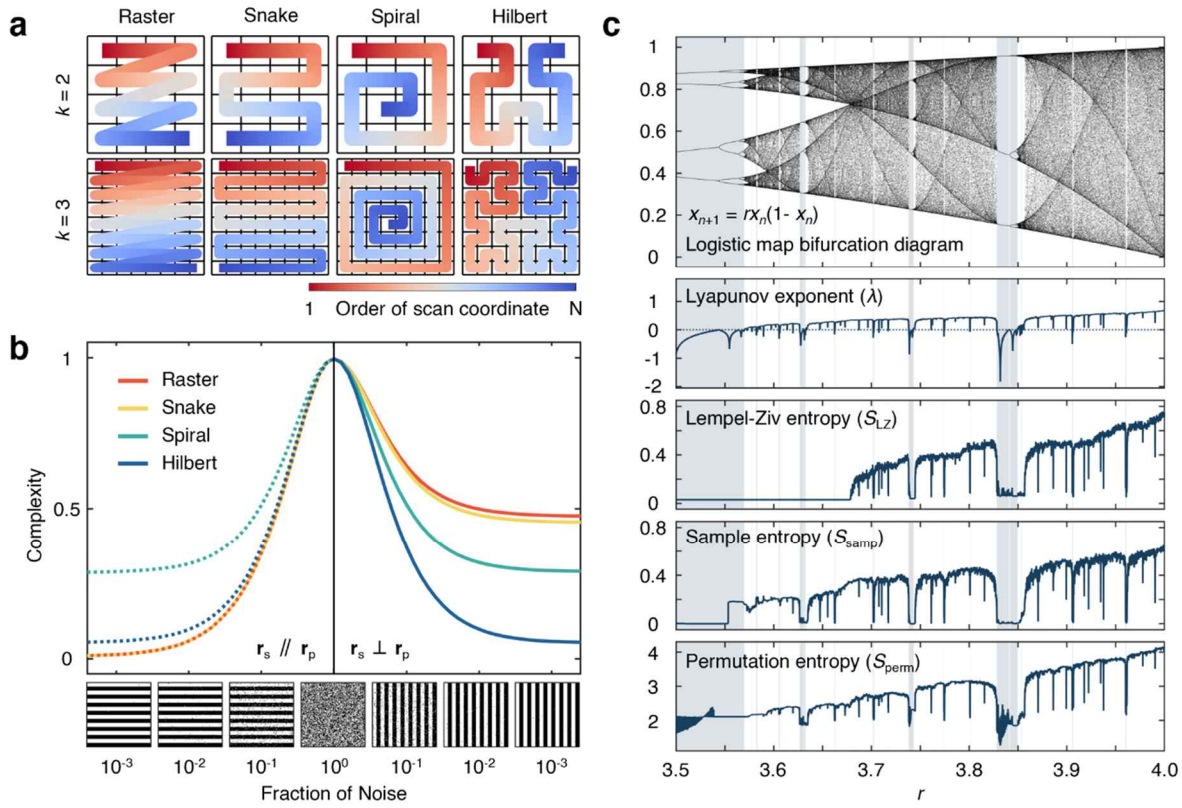


图 1：希尔伯特曲线与复杂度度量。 **a**，将二维数据展平为一维向量的扫描方法（从右向左依次为光栅、蛇形、螺旋和希尔伯特曲线）。红点表示在晶格上随机选取的两个点。底部还显示了映射函数的距离不确定性随递归阶数k 的变化。 **b**，不同扫描方法测得的复杂性随噪声比例的变化。虚线和实线分别表示扫描方向（$_s\mathbf{r}$）

分别平行或垂直于条纹图案的方向（$_p\mathbf{r}$）。**c**，熵的归一化值随李雅普诺夫指数λ 的变化，用于区分逻辑斯蒂映射的混沌与非混沌行为。

需要采用可靠的扫描方法，将高维数据降为一维向量，同时保留其局部性。需要指出的是，将高维数据转换为低维数据时，不可能完全保留其所有局部属性。此外，每次降维都会引入相距较远的数据点之间不自然的相关性，从而扭曲整体数据结构和信息分布。例如，按行或按列进行光栅扫描（图1a）会在系统边界处造成意料之外的不连续，导致高估系统的复杂性或产生边界效应。按行进行蛇形扫描可以保留行方向的信息（即相邻关系），但会损失列方向的信息（图1b）。对于螺旋扫描，可以保留靠近扫描中心的信息；然而，信息离中心越远，在特定方向上出现高估或低估的概率就越高。随着系统规模扩大，扫描偏差会加剧，导致不确定性相应增加，测量可靠性降低（补充图1）。

为解决偏差问题，我们设计了一种不同于现有方式的映射方法。空间填充曲线可以通过递归方式生成，适用于任意形状或形式的空间。特别是，希尔伯特曲线在填充高维空间以及各种形状各异的离散数据方面具有优势，包括各向异性或球形晶格（补充图2）。例如，k阶n维希尔伯特曲线可以完全填充由$2\prod_{}^{}{}_{k}$个单元组成的离散晶格。为测试基于希尔伯特曲线的扫描$_{=1i}$方法的性能，我们比较了不同扫描方法下复杂性（排列熵，后文将进一步详细阐述）随噪声比例的变化情况。我们向水平和垂直条纹图案中添加了椒盐噪声，其比例从

10^{-4} 到1不等。图1b显示，就复杂性而言，基于希尔伯特曲线的展平使垂直和水平条纹图案之间的差异微乎其微。还应注意，对于噪声极少的图案，希尔伯特曲线相对于最简单图案（即无噪声）对应的理想熵值0，仅表现出微小偏差。相比之下，其他扫描方法的复杂性会因扫描方向而出现显著变化（光栅扫描和蛇形扫描），或存在较大的偏差（螺旋扫描）。

接下来，我们为降维数据寻找相关的复杂度度量方法（展平方法见补充图3）。为此，我们测试了一个具有代表性的动力系统——逻辑斯蒂映射，它在递归函数中随参数r变化而产生众所周知的自相似分岔图案：$()_{+1} = \quad 1-_{n\,n\,n}\,x\,rx\,x\,\mathrm{for}[0\quad],_{n} \in \quad 1x$ 。对于逻辑斯蒂映射（图1c），李雅普诺夫指数（λ）可作为通用复杂度度量的参照。针对λ，我们测试了多种熵，例如 Lempel-Ziv熵（$_{\mathrm{LZ}}S$ ）[25], 样本熵（$_{\mathrm{samp}}S$ ）[26], 和排列熵（$_{\mathrm{perm}}S$ ）[27]。如图1c所示，每种熵都能捕捉映射中的大多数关键分岔点。然而，与λ 相比，每种熵在检测所有分岔点方面也都存在局限（补充图4）。遗憾的是，还应注意，大多数具有复杂结构的高维数据并不总是与逻辑斯蒂映射这类控制函数相关，因此λ 并不一定总能用于测量复杂性。因此，不应采用λ，而应谨慎地以互补方式使用基于熵的度量，针对特定数据类型组合或选择不同的熵。通过结合希尔伯特曲线与熵度量，我们可以使用希尔伯特熵，尽可能精确地测量高维数据的复杂性。

自旋模型

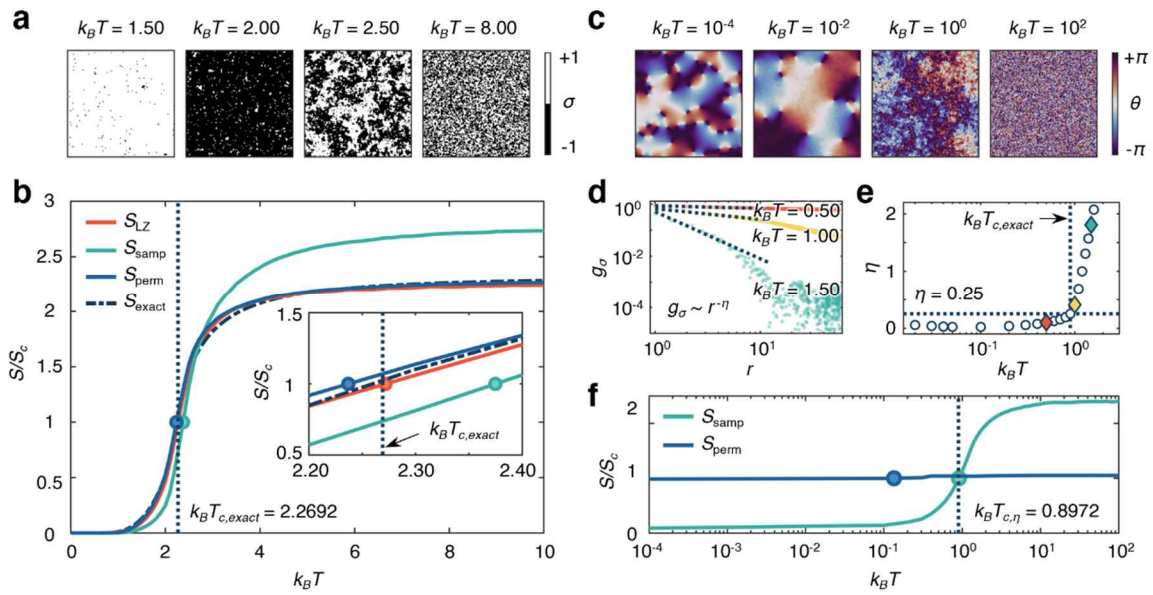


图2：自旋模型。 **a**，伊辛模型在不同温度下的二维方格晶格快照。**b**，熵随温度的增加。插图显示了图中的放大区域，展示了由不同熵度量检测到的拐点。实心圆和垂直虚线用于标示拐点。**c**，二维 XY 模型在不同温度下的快照。**d** 取向关联函数的幂律衰减行为，$g_\sigma(r)$，在$k_BT =$ 0.50、1.00和1.50（**d**）。**d**中的虚线斜率线表示各温度下$g_\sigma(r)$ 的包络函数。**e**，幂指数η 随温度升高的演化。**e**中的垂直线表示取向序减弱的点（η 3 0.25）。**f**，归一化熵随温度的变化。实心圆标记表示拐点，垂直虚线为取向序的临界温度，其中η =0.25（k_BT=0.8972）。

为验证希尔伯特熵，我们首先测试了基于二维自旋模型的知名物理模型。我们采用二维蒙特卡罗（MC）方法（$2^7 \times 2^7$ 正方晶格的维度）模拟二维自旋模型。通过测试希尔伯特熵能否检测模型有序-无序转变的临界点，可以验证其有效性。图2a展示了模拟的二维伊辛模型，其离散自旋态（σ ）为-1（自旋向下）和1

（自旋向上）。我们使用希尔伯特熵测量系统的复杂性，具体做法是对不同温度下每个模拟晶格中包含二元自旋排列的降维向量计算希尔伯特熵，并将其与伊辛模型的理论熵进行比较（精确S）[28]。如图 2b 所示，随着温度升高，三种熵（LZS 、sampS 和 permS ）都能定性地追踪二维伊辛模型从有序到无序的转变。然而，sampS会高估较高温度下的复杂性，而LZS 和permS 能够正确追踪熵的演变，并准确检测临界点（cT ）。例如，由LZS和permS得出的B ckT 分别为 2.2717 和 2.2370，接近理论精确值 2.2692[28]。特别是，与sampS 相比，LZS 在测量二元系统的内在复杂性方面具有优势，因为LZS 考虑了二元信号的所有可能组合，而sampS没有考虑信号排列的局部有序。对局部有序缺乏细致考量，会导致B ckT 处的熵值降低，进而使sampS 的归一化熵被高估。

与具有离散自旋的二维伊辛模型不同，二维 XY 模型考虑的是连续自旋取向（θ ），其范围从π – 到π ，因此，sampS 和permS 适用于测量无序系统的复杂性。我们对二维 XY 模型进行了蒙特卡罗模拟，以获得模拟样本（图 2c）。鉴于二维 XY 模型会发生 Kosterlitz-Thouless（KT）相变，我们追踪了取向关联（()g r ），据此可以通过相关性的衰减指数（η ）分析()g r 相对于距离矢量r 的幂律依赖（图 2d）。根据 Kosterlitz-Thouless-Halperin-Nelson-Young（KTHNY）模型，指数η 具有一个阈值

值为 1/4；超过该阈值后，系统因失去取向序而演变为无序系统[29]。通过追踪η 随温度的变化，可以提取临界温度，得到$_{B}{}_{C}kT$ = 0.8972，如图 2e 所示。有趣的是，利用希尔伯特熵和$_{samp}S$检测到的临界温度$_{B}{}_{C}kT$ = 0.9078，与蒙特卡罗模拟提取的数值接近（图 2f）。相比之下，$_{perm}S$ 在测量临界温度方面并不可靠，并且高估了低温下系统的复杂性。这种差异主要源于$_{perm}S$ 的测量本质：它会统计给定嵌入维数下所有可能的序数模式。因此，$_{perm}S$ 会在极低温下检测到不必要的详细微观自旋分布模式，容易错误地检测出虚假临界点（$_{B}kT$~10^{-1}）。根据有效性测试，我们可以得出结论：对于连续或二进制数据，希尔伯特熵结合$_{samp}S$ ；对于离散数据，希尔伯特熵结合$_{LZ}S$和$_{perm}S$，通常都可应用于具有临界现象的物理系统。

渗流模型

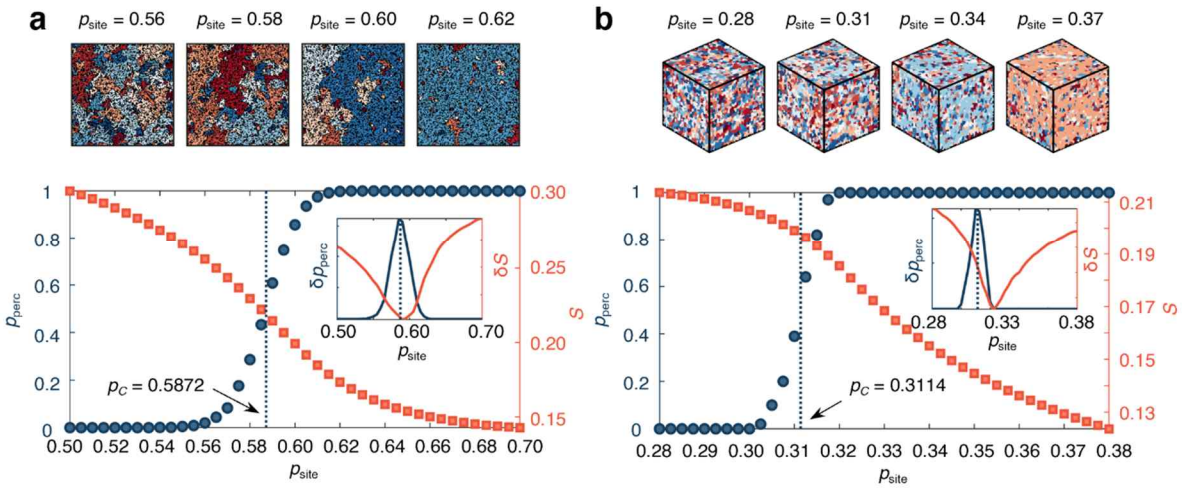


图3：渗流模型。 **a，** **b**，二维（**a**）与三维（**b**）系统的格点渗流模型。上方图幅展示了不同格点占据概率$_{site}p$ 取值下的团簇分布。下方图展示了渗流概率$_{perc}p$ （海军蓝色实心圆）以及用$_{LZ}S$ 测得的希尔伯特熵（红色实心方块）随$_{site}p$变化的行为（插图：$_{perc}p$ 与$_{LZ}S$ 关于$_{site}p$ 的一阶导数，分别记为$_{perc}\delta p$ 和δS）。竖直点线表示渗流阈值$_{c}p$，即$_{perc}p$ 的拐点。

接下来，我们将研究范围扩展至希尔伯特熵在更高维度和不同物理系统中的应用。为此，我们采用了离散晶格中的二维和三维渗流模型（格点渗流）。与二维伊辛模型或XY模型不同，渗流模型通过估算渗流概率$_{perc}p$ 来追踪渗流网络的结构关联。在格点渗流模型中，通过逐步提高晶格点被占据的概率（即格点占据概率；$_{site}p$ ），可以计算出使整个系统发生渗流的可能性骤增时的渗流阈值（$_{c}p$）。当$_{site}$ $_{c}p$ $p<$时，小团簇主导系统；而当$_{site}$ $_{c}p$ $p>$ 时，贯穿整个系统的团簇出现的概率会突然增加。因此，从多个小型、

从不连通的团簇转变为单个大型连通团簇，体现了二级相变。因此，二维或三维渗流晶格模型与检验希尔伯特熵在高维数据中的适用性和有效性相关。在建模中，我们使用蒙特卡罗模拟生成具有不同$_{位点}p$ 值的二维或三维晶格（二维的维度为 $2^7 \times 2^7$ ，三维的维度为 $2^6 \times 2^6 \times 2^6$ ）（图 3a 和 3b）。对于每个晶格，我们计算了$_{渗流}p$随$_{site}p$ 变化的函数关系。此外，考虑到晶格的离散性和非序数性，我们还使用希尔伯特熵以及$_{LZ}S$计算了晶格的复杂性。为检测$_{c}p$，我们计算了熵的一阶导数。如图 3a 所示，希尔伯特熵能够检测二维渗流晶格中的$_{c}p$，但存在轻微偏差（来自$_{LZ}S$的$_{c}p$为 = 0.5916；来自蒙特卡罗模拟的$_{c}p$为 = 0.5872）。值得注意的是，基于希尔伯特熵的$_{c}p$接近已报道的数值，例如 0.5927[30–32]。我们将分析扩展到三维晶格（图 3b），发现希尔伯特熵也能够识别$_{c}p$ ，且与参考值的偏差极小（来自$_{LZ}S$ 的$_{c}p$为 = 0.3225；来自蒙特卡罗模拟的$_{c}p$ 为 0.3114）。与三维渗流的理论值相比，该方法仍然有效，其中$_{c}p$为 = 0.3116[32,33]。这些偏差被认为是由模拟盒效应导致的。通过对二维和三维渗流晶格的分析，我们可以得出结论：将希尔伯特熵与信息熵（$_{LZ}S$ ）适当结合，能够测量高维数据的内在复杂性。

尺度不变系统内在复杂性的测量

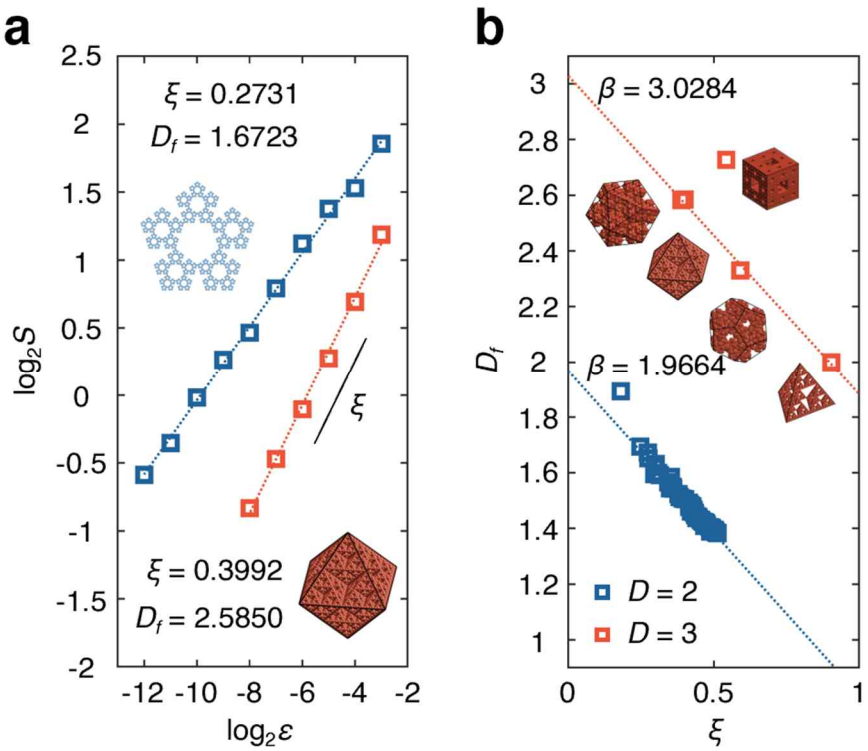


图 4：希尔伯特熵与尺度不变系统分形特性的关系。a，希尔伯特熵（采用$_{perm}S$ ）随盒子大小 ε 变化的幂律关系（相对于包围分形几何的圆形或球形容器的原始大小），适用于基于迭代函数系统（IFS）的二维分形（边的数量（E = 5），红色空心方块）和三维分形（面的数量（F = 8），蓝色空心方块），其分形维数为$_fD$。通过对每个 IFS 进行线性回归，以回归直线的斜率（幂指数ξ ）分析线性关系。**b**，二维（红色空心方块）和三维（蓝色空心方块）迭代函数系统的幂指数ξ 与分形维数$_fD$ 之间的线性关系。虚线表示回归线。对于三维迭代函数系统，相关的迭代函数系统示于相应数据点旁。

发生连续相变的动力系统的一项显著代表性特征，是伴随自相似性的尺度不变性。要定量测量自相似性，可以使用分形维数（$_fD$）。为了检验并发现分形性质与熵之间隐藏的相关性，我们计算了多种具有不同边数（E）或

面数（F）的迭代函数系统（IFS），例如五边形雪花（$E=5$）或八面体迭代函数系统分形（$F=8$），如图 4a 所示。原则上，希尔伯特曲线与由不同递归次数生成的分形几何相兼容，这些递归次数可对应于希尔伯特曲线的阶。因此，预计 ${}_fD$ 与熵之间存在很强的相关性。需要指出的是，传统的信息熵并未考虑分形几何的自相似性和递归性质。

我们首先研究了分形几何的分辨率与熵之间的关系。为量化分辨率，我们使用了不同的盒子尺寸ε 。相对于ε 的示例性量化过程如补充图5所示。熵应随着ε 值的增加而增加。换言之，ε 与希尔伯特曲线的阶成反比。具体而言，具有较高${}_fD$的系统中，随着ε 增加，熵会逐渐降低，因为相邻盒子对熵降低具有缓解作用。相反，在${}_fD$较低的系统中，相邻盒子的影响很小，从而导致熵更显著地下降。由于自相似性，预计基于 ${}_{\mathrm{perm}}S$ 的熵与ε 之间存在明确的幂律依赖。

有趣的是，如图4a所示，希尔伯特熵呈现出幂律关系，其特定指数ξ 对应于特定的迭代函数系统，可定义为

$$\xi = -\lim_{\varepsilon \to 0} \frac{\log S(\varepsilon)}{\log(1/\varepsilon)}. \tag{1}$$

如方程 (1) 所述，ξ 可被视为一个指标，用于推断各种自相似但复杂几何体的 ${}_fD$ 。这意味着 ξ 可用于揭示复杂高维系统的分形特性。为检验这一猜想，我们分析了

给定的二维和三维迭代函数系统中 $_fD$ 与 ξ 之间的关系。如图 4b 所示，我们观察到 ξ 与 $_fD$ 之间存在明显的线性关系。值得注意的是，这些线性关系普遍揭示了 $_fD$ 与 ξ 之间的一种简单关系，即 α ξ β = +–$_fD$。系数 α 的值–分别为 1.1637 和 1.1453，而常数 β 的值对于二维和三维迭代函数系统分别为 1.9664 和 3.0284。其他迭代函数系统的 ξ 与 $_fD$ 值见补充图 6。这意味着真实的 $_fD$ 与熵的标度指数之间可能存在理想且简单的线性关系，即

$$D_f = -\xi + D, \tag{2}$$

其中 D 表示欧几里得维数。为阐明这种线性背后的数学原理，我们可以参考示例性的一维分数布朗运动（分数布朗运动），$B(r)$ 在位置 r 满足如下自相关关系

$$\langle B(r+\varepsilon)B(r)\rangle = \frac{V_H}{2}\left(r^{2H} + |r+\varepsilon|^{2H} - |\varepsilon|^{2H}\right),\ V_H = \Gamma(1-2H)\frac{\cos(\pi H)}{\pi H} \tag{3}$$

其中，$\Gamma(\cdot)$ 表示伽马函数，H 表示一维向量的赫斯特指数。基于希尔伯特曲线，可将一维分数布朗运动视为降维后的高维数据。为了计算一维分数布朗运动的熵，我们可以计算$B(r)$之差的自相关，即$C(\varepsilon r) C(r)$，其中$C(r) B(r) B(r)$= +1–。可以证明，

〈 〉

$$\begin{aligned}\langle C(r+\varepsilon)C(r)\rangle &= \frac{V_H|\varepsilon|^{2H}}{2}\left(\left(1+\frac{1}{\varepsilon}\right)^{2H}+\left(1-\frac{1}{\varepsilon}\right)^{2H}-2\right)\\ &\approx \frac{V_H|\varepsilon|^{2H}}{2}\left(4H(2H-1)\varepsilon^{-2}\right)\\ &= 2H(2H-1)V_H\varepsilon^{2H-2}\sim\varepsilon^{2(H-1)}.\end{aligned} \tag{4}$$

自相关函数值应与熵的平方值成正比。还应注意，当 $0\varepsilon\to$ 时，熵值应趋近于零。然后，根据方程 (4)，我们可以得出

$$S\sim\varepsilon^{H-1}. \tag{5}$$

根据方程 (5)，我们可以得出 $1\xi=-H$ 。对于由给定H 的自相似序列生成的一维分数布朗运动，其分形维数D 满足关系 $=\quad 1-f\quad +DDH34$，由此得到我们此前发现的理想线性关系，即$\xi f=\quad +-DD$。

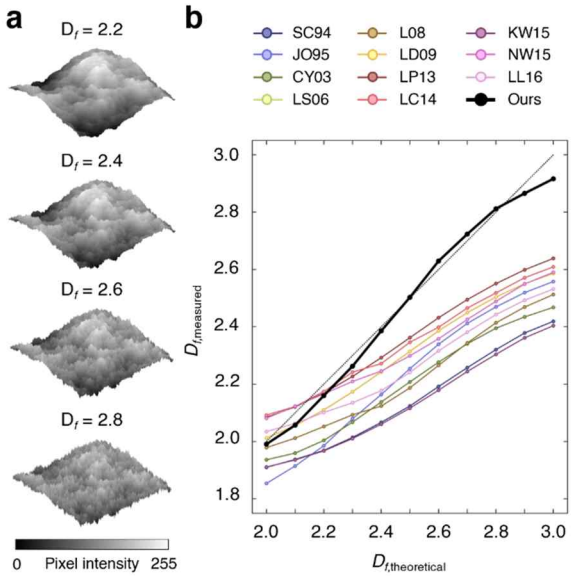


图 5：希尔伯特熵用于测量合成高维数据的分形维数。

a，不同分形维数（$_fD$ = 2.2、2.4、2.6 和 2.8，从上到下）的分数布朗表面 (fBs)，通过控制赫斯特指数（H ）生成。z 轴表示灰度图像的像素值（0 至 255）。**b**，基于幂指数ξ 的方法得到的分形维数（黑色）与基于盒计数法得到的分形维数（彩色）的比较，数据取自参考文献[35]。图例中列出了作者姓名首字母和发表日期。黑色虚线表示根据赫斯特指数关系计算出的理想分形维数：= 1– $_f$ ^+D DH 。曲线越接近虚线，该方法就越准确。

接下来，我们进一步探究所观察到的幂指数ξ 的普适性，并将其用作一般高维复杂数据的复杂性度量指标，这一应用并不局限于具有二元属性的数据。尽管已证明理想线性关系ξ =–$_f$$D$ D 适用于二元自相似系统，但更具一般性的线性关系α ξ β = +–$_f$$D$ 能够通过α 和β 的组合，为具有二元以外属性的更高维系统提供独特信息。如方程（2）所示，这一发现表明，即使对于以往难以确定$_fD$的非二元系统，也可以推断高维数据的内在维数。例如，众所周知，传统盒计数法无法可靠地评估灰度图像的$_fD$[35]。这主要是因为从纯数学角度而言，灰度图像并不对应于某个维数。因此，对于二维分数布朗运动，当分形维数趋近于3[35–37]时，计算得到的分形维数与理想分形维数之间的差异会逐渐增大。因此，与传统盒计数法相比，本文提出的基于希尔伯特熵的方法有助于获得更真实、准确的一般高维复杂数据（包括灰度图像）的分形维数。为测试基于希尔伯特熵的分形维数测量方法的性能，我们对具有理论定义分形维数的基于分数布朗运动的图像（即分数布朗表面（fBs）图像）进行了测试

（图5a）。计算像素值范围为0至255的灰度图像中的ξ 时，我们采用了 S_{samp} ，该方法已在二维 XY 模型中证明了其性能。灰度图像的维数应为3（即平面维数为2的欧几里得维数，再加上与像素值分布相对应的一个自由度；将图像视为亮度为高度的三维曲面）。如图5b所示，根据 $\xi \propto \beta^{-D_f}$ 的线性关系推断出的分形维数，比传统盒计数法的结果更好。特别是，即使理论分形维数趋近于3，基于 S_{samp} 的 D_f 测量结果也表现出较小的差异。这表明，本文提出的基于希尔伯特熵的方法能够有效评估一般高维复杂数据的分形特性。

总之，我们提出了一种用于评估高维物理系统复杂性的新方法——希尔伯特熵。先计算信息熵，再通过希尔伯特曲线进行降维，这种方法能够准确确定各种相变现象中的临界点。此外，希尔伯特熵具有幂律依赖的特征，在标度不变高维几何中，它与分形维数和欧几里得维数呈现出稳定的线性关系。特别是，我们证明，所提出的方法可用于准确评估一般系统的分形维数，例如灰度图像。这种方法为理解和分析复杂系统提供了新思路，并有望为高维物理系统复杂性的分析与测量奠定创新策略的基础。

方法

LZ熵

Lempel-Ziv（LZ）复杂度通过评估给定数据序列的模式和结构来定义复杂性。由于LZ 复杂度 $_{LZ}C$ 的内在性质决定了它与数据大小成正比，因此提出了归一化的$_{LZ}C$ ，即 LZ 熵（$_{LZ}S$ ），它是以给定序列的最大复杂度进行归一化得到的。当数据量（N）足够大时，$_{LZ}C$趋近于 Kolmogorov复杂度。$_{LZ}C$ 的详细计算方法见其他文献[25]。尽管 $_{LZ}C$ 被认为适合测量二进制数据的复杂性，但它也可以扩展并推广到n 位数据，如下所示：

$$S_{\mathrm{LZ}} = \frac{C_{\mathrm{LZ}} \log_n N}{N} \tag{6}$$

样本熵

样本熵（$_{samp}S$）基于相同长度采样数据之间的相似性来度量系统的复杂性。判定相似性时，通常使用Chebyshev距离。给定序列（u）被划分为嵌入维数为m 的子序列（$_m\mathbf{X}$），其定义为$()()\{\}=\quad+\quad 0\qquad 1\le\le-_m iuikkm\mathbf{X}$；若两个子序列之间的距离处于阈值距离$r$ 之内，则认为它们相似。通常，将r 设为$0.2\sigma=r$，其中σ 为给定序列的标准差。样本熵$_{samp}S$被定义为相似子-

序列相对于m的数量之比，如下：

$$S_{\text{samp}} = -\log \frac{(N-m)\sum_{j=1,j\neq i}^{N-m} \#\left\{i \middle| \mathbf{X}_{m+1}(i) \text{ s.t. } d\left[\mathbf{X}_{m+1}(i), \mathbf{X}_{m+1}(j)\right] < r\right\}}{(N-m)\sum_{j=1,j\neq i}^{N-m} \#\left\{i \middle| \mathbf{X}_{m}(i) \text{ s.t. } d\left[\mathbf{X}_{m}(i), \mathbf{X}_{m}(j)\right] < r\right\}}, \tag{7}$$

其中，$d(\mathbf{X}, {}_{*}\mathbf{X})$表示两个子序列$\mathbf{X}$和${}_{*}\mathbf{X}$之间的距离。

排列熵

排列熵（S_{perm}）可通过考虑序数模式出现的概率来测量。为此，将原始信号划分为大小为嵌入维数m的分区$\mathbf{X}_i$。要计算S_{perm}，特定序数模式

出现的概率（$p(\pi)$）定义为

$$p(\pi) = \frac{\#\left\{i \middle| \mathbf{X}_{m}(i) \text{ s.t. type } \pi\right\}}{N-m+1}. \tag{8}$$

随后，通过从采样数据中构造序数模式，并对$p(\pi)$应用香农信息熵来测量排列熵，即

$$S_{\text{perm}} = -\sum p(\pi)\log p(\pi). \tag{9}$$

自旋模型

我们采用周期性边界条件，在单位间距的N^2正方晶格（$=2^7N$）上，使用基于Metropolis算法的蒙特卡洛（MC）方法，对二维正方晶格Ising模型和XY模型进行模拟。最大迭代次数设为$10000N^2$。系统的能量

在0 K至10 K的温度范围内进行测量，间隔为0.1 K。

伊辛模型

为了计算二维正方晶格伊辛模型的理论临界温度，我们根据哈密顿量计算了自旋态的能量，其中

$$\mathcal{H}(\sigma)=-J\sum_{\langle i,j\rangle}\sigma_i\sigma_j\,, \tag{10}$$

其中，J表示相邻格点i和j之间的相互作用，σ为二元自旋态，满足$\sigma\in\{-1,+1\}$。为简化计算，将相邻自旋之间的相互作用J设为1。忽略外部磁场。每个自旋的自旋态由玻尔兹曼概率分布确定，如下所示：

$$P(\sigma)=\frac{e^{-\beta\Delta\mathcal{H}(\sigma)}}{Z}, \tag{11}$$

其中，$\beta=1/k_BT$为逆温度，Z为配分函数，其定义为

$$Z=\sum_{\sigma}e^{\beta\mathcal{H}(\sigma)}. \tag{12}$$

在初始阶段，将给定系统初始化为由-1和+1组成的随机自旋构型。每次迭代中，自旋翻转的概率由玻尔兹曼概率分布（方程(11)）决定。如果当前状态满足$\mathcal{H}(\sigma)\leq\mathcal{H}(\sigma-\Delta\sigma)$或$P\leq p$，则用新的自旋态替换当前自旋态，其中$p\in[0,1]$。

系统的熵S随后由自由能F推导得出，其中自由能遵循昂萨格的解法[28]：

$$S=-\frac{\partial F}{\partial T}$$
$$F=-\frac{1}{\beta}\left(\log 2+\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{2\pi}\log\left[\cosh^2 2\beta J-\sinh 2\beta J\left(\cos\omega_1+\cos\omega_2\right)\right]d\omega_1 d\omega_2\right). \tag{13}$$

结果如图 2b 所示。

XY模型

无外部磁场且相互作用强度为单位值的经典二维 XY 模型的能量为

$$\mathcal{H}(\theta)=-J\sum_{\langle i,j\rangle}\cos\left(\theta_i-\theta_j\right), \tag{14}$$

其中，θ_i 是第i个自旋的自旋角，该自旋与j个相邻自旋相伴，使得$\theta\in[-\pi,\pi]$[38]。

在蒙特卡罗过程中，自旋取向会随机变化 $\theta_i=\theta_i+k$，其中$k\in[-0.5,0.5]$。随后，根据玻尔兹曼概率分布（方程 (11)）确定自旋构型，其配分函数为[39]

$$Z=\int\prod_j e^{-\beta\mathcal{H}(\theta)}d\theta_j\,. \tag{15}$$

由于二维 XY 模型的解析解未知，因此可以通过引入序参量来分析 XY 模型的临界点，而非比较热力学熵。对于 XY 模型的连续自旋，取向关联$g_\sigma(r)$ 表示如下：

$$g_\sigma(r)=\left\langle\Psi(\mathbf{r}_i)\cdot\Psi^*(\mathbf{r}_j)\right\rangle \tag{16}$$

其中 = −i jr| r r, |对于第( ) θΨ = kk ier 个自旋。

根据关系式 ()η σ −~g r r ，临界指数 η 可通过各温度 Bk T 下包络函数相对于 r 的斜率求得（图 2d）。斜率（η ）在 0.8972 的转变温度处突然增大，这接近于已报道的临界点 0.8935[40]。

基于混沌游戏的分形几何

我们基于混沌游戏[41]在二维/三维空间中生成了 IFS 分形。为使系统的大小保持一致，将正多边形/正多面体内接于单位圆/球面中。我们通过调整边的数量（ {3, 4, , 50}∈ ⋯E）或面的数量（ {4, 6,8,12, 20}∈F）生成多边形。在迭代过程中，根据以下迭代函数，在初始多边形/多面体内部生成额外的点：

$$\mathbf{x_i} = r\mathbf{x_{i-1}} + (1-r)\mathbf{b_j}, \tag{17}$$

其中，$_i\mathbf{x}$ 是第 i 次迭代时点的坐标，$_j\mathbf{b}$ 是随机选取的第 j 个顶点的坐标，其中 {1,2, , } ∈ ⋯jn，顶点总数为 n；r 是由下式给出的缩放因子

$$r = \frac{1}{2\left(1+\sum_{k=1}^{\lfloor n/4 \rfloor}\cos\frac{2\pi k}{n}\right)}, \tag{18}$$

适用于二维几何[42]。

对于$n = 4$ 2维迭代函数系统，理论分形维数等于 2（即简单正方形）；因此，我们使用了分形维数为$\log 8 / \log 3 = 1.8928$ 的谢尔宾斯基地毯，用于$n = 4$ 2D 迭代函数系统。

参考文献

1. Grassberger, P. & Procaccia, I. 从混沌信号估计柯尔莫哥洛夫熵。《物理评论*A*》**28**, 2591 (1983). 2. Abarbanel, H. D. I., Brown, R., Sidorowich, J. J. & Tsimring, L. S. 物理系统中观测到的混沌数据分析。《现代物理评论》 **65**, 1331 (1993). 3. Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. 从时间序列确定李雅普诺夫指数。《物理*D*：非线性现象》 **16**, 285–317 (1985). 4. Grassberger, P. & Procaccia, I. 测量奇异吸引子的奇异性。《物理*D*：非线性现象》 **9**, 189–208 (1983). 5. Kaplan, J. L. & Yorke, J. A. 多维差分方程的混沌行为。《数学讲义》 **730**, 204–227 (1979). 6. Rosenstein, M. T., Collins, J. J. & De Luca, C. J. 一种利用小型数据集计算最大李雅普诺夫指数的实用方法。《物理*D*：非线性现象》 **65**, 117–134 (1993). 7. Politi, A. & Witt, A. 时空混沌的分形维数。《物理评论快报》 **82**, 3034 (1999). 8. Egolf, D. A. & Greenside, H. S. 广延混沌中分形维数与空间关联长度之间的关系。《自然》**369**, 129–131 (1994). 9. Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A. & Fuentes, M. A. 区分噪声与混沌。《物理评论快报》 **99**, 154102 (2007). 10. Shiozawa, K., Uemura, T. & Tokuda, I. T. 通过分区熵检测复杂时间序列的动力学不稳定性。《物理评论*E*》 **107**, 014207 (2023). 11. Politi, A. 量化混沌时间序列的动力学复杂性。《物理评论快报》 **118**, 144101 (2017). 12. Zunino, L., Soriano, M. C. & Rosso, O. A. 利用多尺度符号方法从时间序列中区分混沌与随机动力学。《物理评论*E*》**86**, 046210 (2012). 13. Lopes, S. R., Prado, T. L., Corso, G., dos, G. Z. & Kurths. 随机与混沌信号的无参数量化。《混沌、孤子与分形》 **133**, 109616 (2020). 14. Zanin, M. & Olivares, F. 基于序数模式的方法：在离散时间序列中区分混沌与噪声。《通讯·物理》 **4**, 1–14 (2021). 15. Morel, C. & Humeau-Heurtier, A. 二维模式的多尺度排列熵。《模式识别快报》 **150**, 139–146 (2021). 16. Ribeiro, H. V., Zunino, L., Lenzi, E. K., Santoro, P. A. & Mendes, R. S. 复杂性-熵因果平面：一种二维模式复杂度度量。《公共科学图书馆·综合》 **7**, e40689 (2012). 17. Humeau-Heurtier, A., Omoto, A. C. M. & Silva, L. E. V. 双维多尺度熵：与

离散傅里叶变换的关系及其生物医学应用。计算机生物学与医学 100, 36–40 (2018). 18. Silva, L. E. V., Duque, J. J., Felipe, J. C., Murta, L. O. & Humeau-Heurtier, A. 二维多尺度熵分析：在图像纹理评估中的应用。信号处理 147, 224–232 (2018). 19. Vink, R. L. C. & Barkema, G. T. 成网材料的构型熵。《物理评论快报》89, 076405 (2002). 20. Zhou, Y. & Milner, S. T. 基于图同构的玻璃态系统结构熵。《软物质》12, 7281 7288 2016– ( ). 21. Jagadish, H. V. 具有多个属性对象的线性聚类。《SIGMOD 记录》19, 332–342 (1990). 22. Peano, G. 关于一条填满整个平面区域的曲线。《数学年刊》36, 157–160 (1890). 23. Hilbert, D. 关于直线到一个面片的连续映射。《数学年刊》38, 459–460 (1891). 24. Moon, B., Jagadish, H. V., Faloutsos, C. & Saltz, J. H. 希尔伯特填充曲线的聚类性质分析。《IEEE 知识与数据工程汇刊》13, 124–141 (2001). 25. Lempel, A. & Ziv, J. 关于有限序列的复杂性。《IEEE 信息论汇刊》22, 75–81 (1976). 26. Richman, J. S. & Moorman, J. R. 使用近似熵和样本熵进行生理时间序列分析。《美国生理学杂志：心脏与循环生理学》278, (2000). 27. Bandt, C. & Pompe, B. 排列熵：一种自然的时间序列复杂度度量。《物理评论快报》88, 174102 (2002). 28. Onsager, L. 晶体统计学。一种具有有序-无序转变的二维模型。《物理评论》65, 117 (1944). 29. Nelson, D. R. & Halperin, B. I. 二维系统中的位错介导熔化。《物理评论B》19, 2457 (1979). 30. Newman, M. E. J. & Ziff, R. M. 用于渗流的高效蒙特卡罗算法与高精度结果。《物理评论快报》85, 4104 (2000). 31. Feng, X., Deng, Y. & Blöte, H. W. J. 二维中的渗流转变。《物理评论E》78, 031136 (2008). 32. Deng, Y. & Blöte, H. W. J. 二维和三维格点渗流模型的蒙特卡罗研究。《物理评论E》72, 016126 (2005). 33. Lorenz, C. D. & Ziff, R. M. 三维渗流中簇的超额数与穿越概率函数的普适性。J. Phys. A Math. Theor. 31, 8147 (1998). 34. Gneiting, T. & Schlather, M. 将分形维数与赫斯特效应分离的随机模型。《SIAM 综述》46, 269–282 (2006). 35. Panigrahy, C., Seal, A., Mahato, N. K. & Bhattacharjee, D. 灰度图像分形维数估计的差分盒计数方法：综述。《混沌、孤子与分形》126, 178–202 (2019). 36. Li, J., Du, Q. & Sun, C. 一种改进的图像分形维数估计盒计数法。《模式识别》42, 2460–2469 (2009). 37. Li, J., Sun, C. & Du, Q. 一种新的图像分形维数估计盒计数法。ICIP 会议 3029–3032 (2006). 38. Kosterlitz, J. M. & Thouless, D. J. 二维系统中的有序化、亚稳性与相变。J. Phys. C: 固态物理 6, 1181 (1973).

39. Kosterlitz, J. M. 二维 xy 模型的临界性质。*J. Phys. C:* 固态物理 **7**，1046 1974( )。40. Hsieh, Y. Da, Kao, Y. J. & Sandvik, A. W. Berezinskii–Kosterlitz– Thouless 相变的有限尺寸标度方法。*J. Stat. Mech.:* 理论与实验 **2013**，P09001 (2013)。41. Barnsley, M. F. (Michael F. & Rising, Hawley. 处处分形。(Academic Press Professional, 1993)。42. Abdulaziz, A. & Said, J. 关于吸引子为 Sierpinski n 边形的迭代函数系统的压缩比。《混沌、孤子与分形》 **150**，111140 (2021)。

致谢

本研究得到了三星研究资助中心为三星电子提供的资助，项目编号为 SRFC-MA2201-02。

作者信息

Seong-Gyun Im 和 Taewoo Kang 对本文贡献相同。

作者与所属机构

大韩民国成均馆大学（SKKU）化学工程学院

Seong-Gyun Im、Taewoo Kang 和 S. Joon Kwon

SKKU 大韩民国成均馆大学（SKKU）能源科学与技术研究院（SIEST）

S. Joon Kwon

系　　半导体融合工程系，成均馆大学，大韩民国　　a

S. Joon Kwon

先进能源工程系，成均馆大学，大韩民国

作者贡献

S.-G.I. 和 T.K. 完成了数值模拟与分析。S.-G.I.、T.K. 和 S.J.K. 撰写了稿件。所有作者均参与了论文的编辑与审阅。

通讯作者

请联系 S. Joon Kwon。

伦理声明

利益冲突

作者声明不存在利益冲突。

补充信息

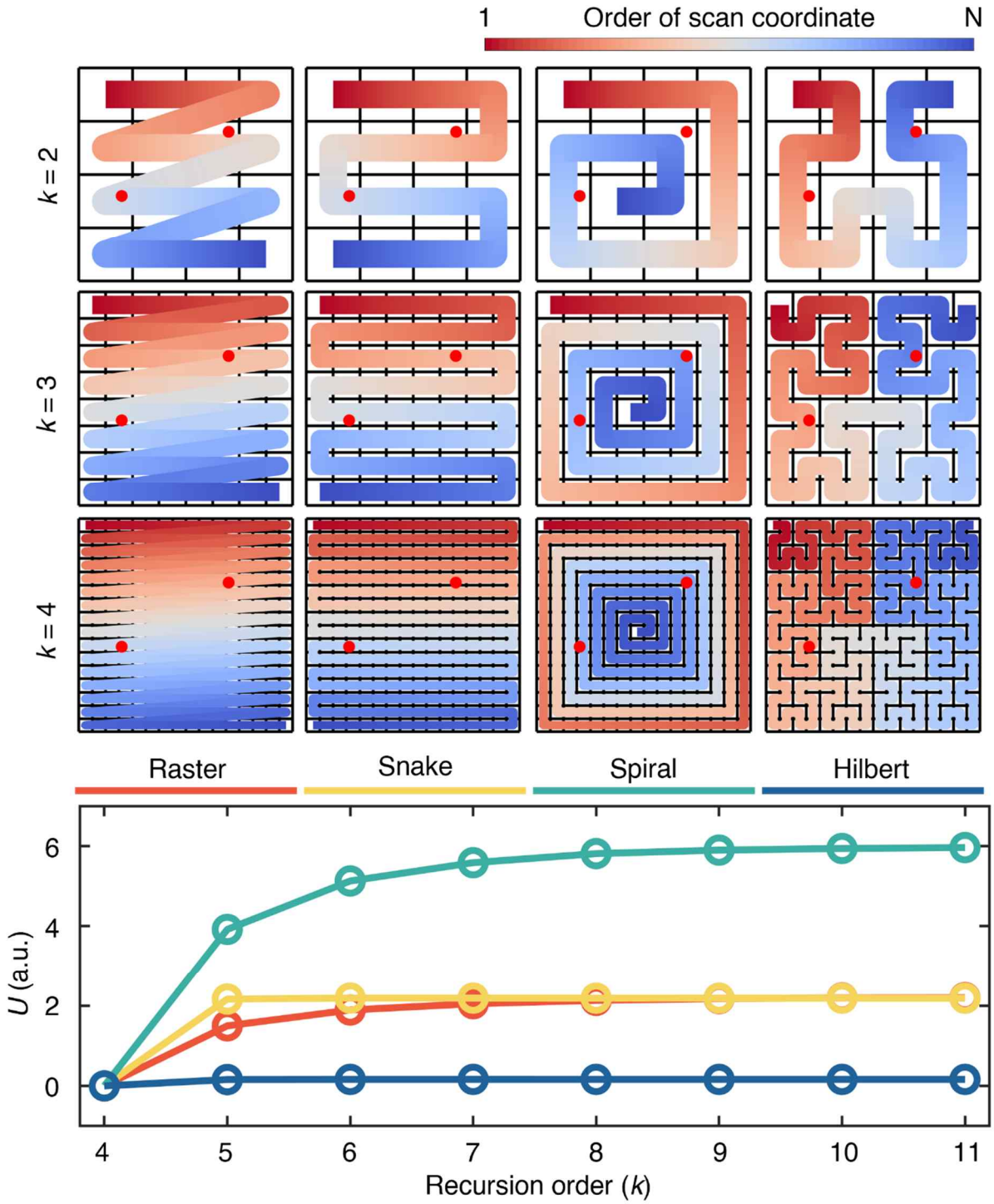
用于测量高维系统复杂性的希尔伯特熵

Seong-Gyun Im [1,†]、Taewoo Kang [1,†]和 S. Joon Kwon [1,2,3,4*]

[1]成均馆大学化学工程学院（SKKU），大韩民国[2]成均馆大学能源科学与技术研究院（SIEST），成均馆大学（SKKU），大韩民国[3]成均馆大学半导体融合工程系（SKKU），大韩民国[4]成均馆大学先进能源工程系（SKKU），大韩民国

* 通讯作者：sjoonkwon@skku.edu

† 这些作者对本研究的贡献相同。


1
Order of scan coordinate
N
k = 2
k = 3
k = 4
Raster
Snake
Spiral
Hilbert
U (a.u.)
0
2
4
6
4
5
6
7
8
9
10
11
Recursion order (k)

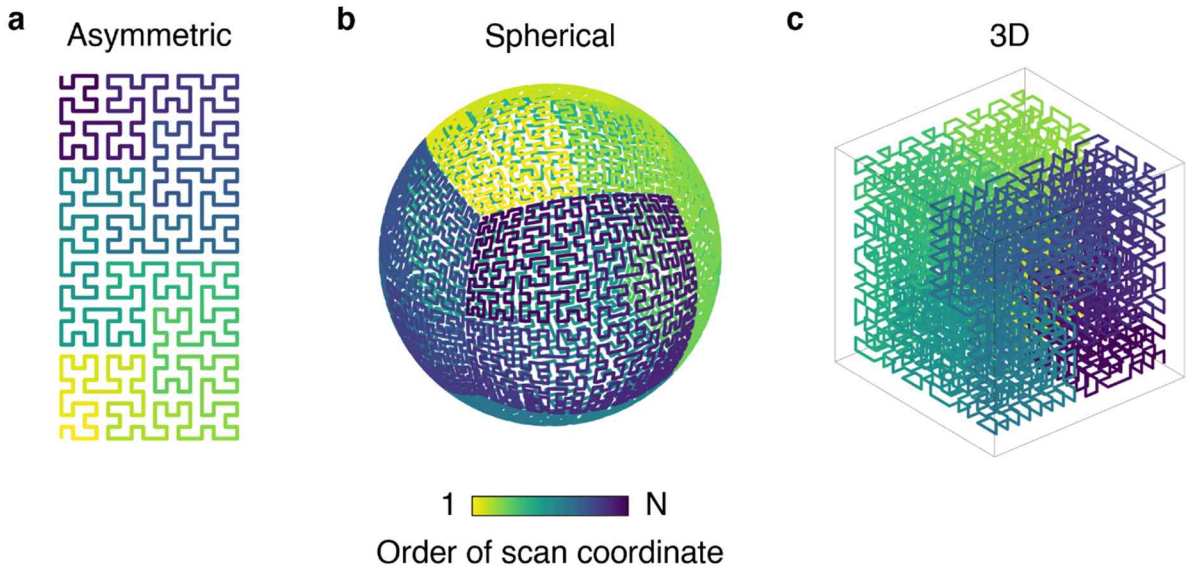
U

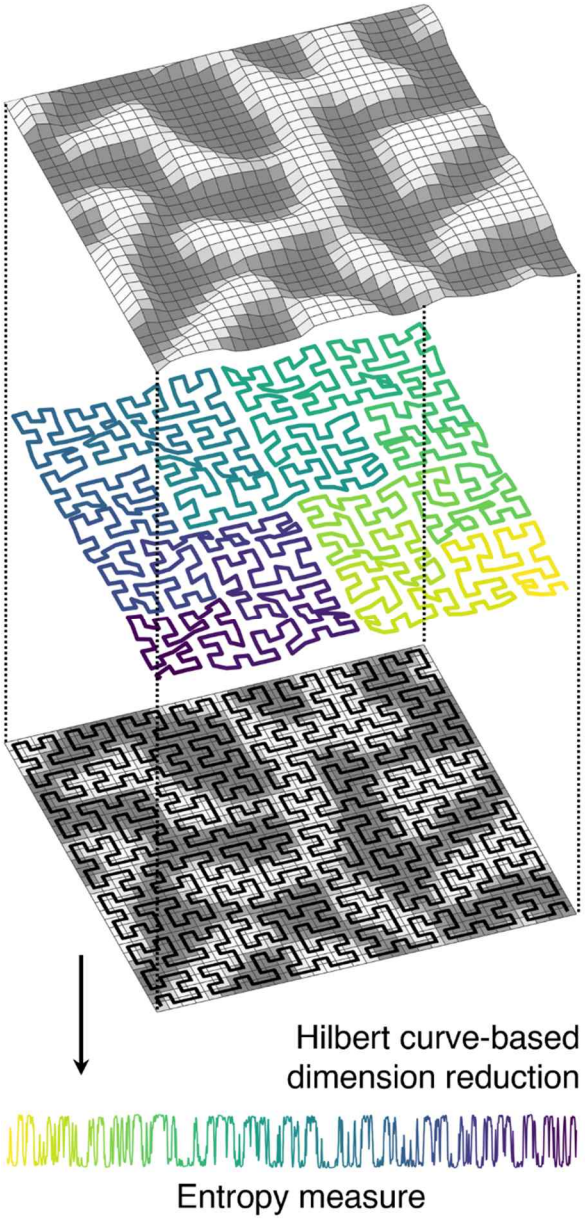
k

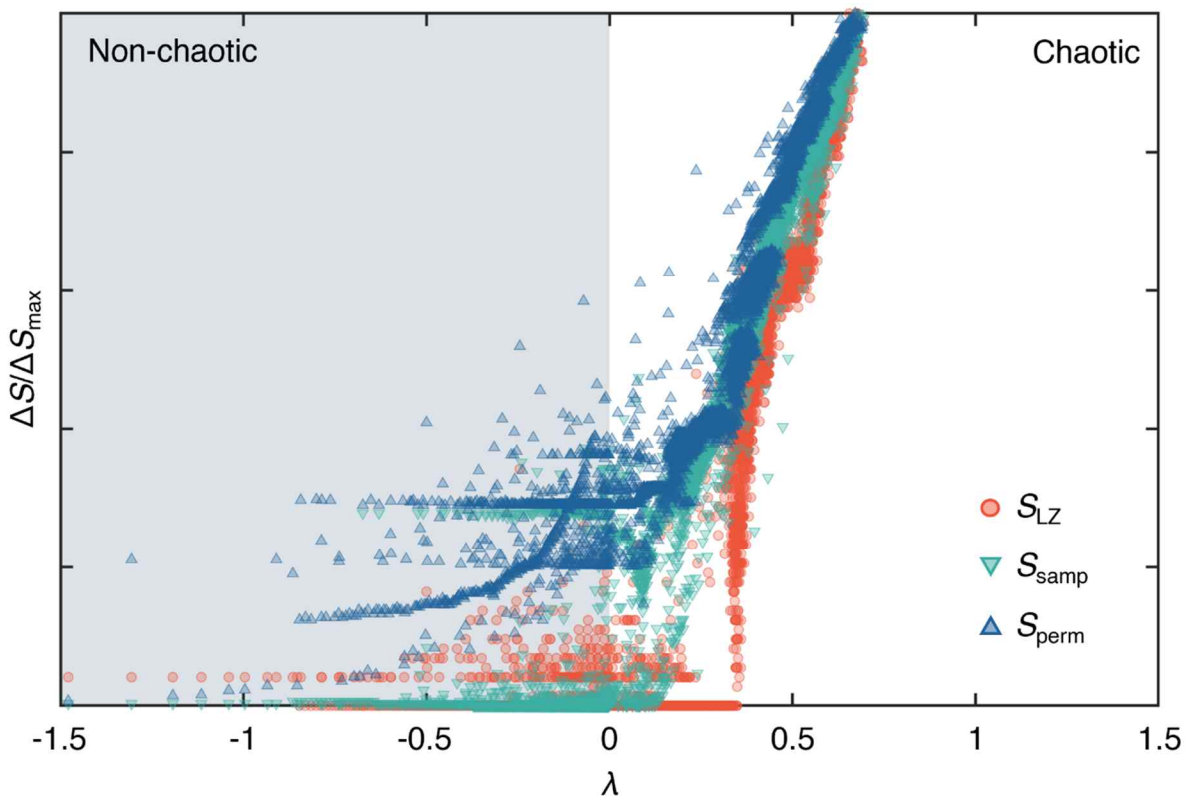
为了测试基于希尔伯特曲线的扫描方法的性能，我们比较了不同扫描方法下映射函数上随机选取的两个点之间的信息不确定性，

$$U = \mathrm{std}\left[\left|\frac{|q_{1,k} - q_{2,k}|}{2^{nk}} - \frac{|q_{1,\kappa} - q_{2,\kappa}|}{2^{m\kappa}}\right|\right],$$

其中$\min\{|\}\kappa = \in \leq \mathbb{N} k\, k\, M$，$q$ 表示这些点在展平数据中的位置，而M 表示晶格的尺寸。补充图1显示，基于希尔伯特曲线的降维对于任意随机选取的点对（测试点对的数量为 = 10^3）几乎不存在不确定性；而其他方法在晶格单元尺寸变化时会产生不确定性波动（即，随着递归阶数k 增加，波动也随之增大）。

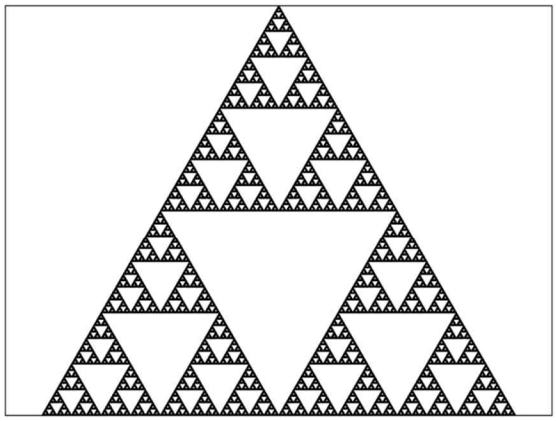

a
Asymmetric
b
Spherical
c
3D
1
N
Order of scan coordinate


Hilbert curve-based
dimension reduction
Entropy measure

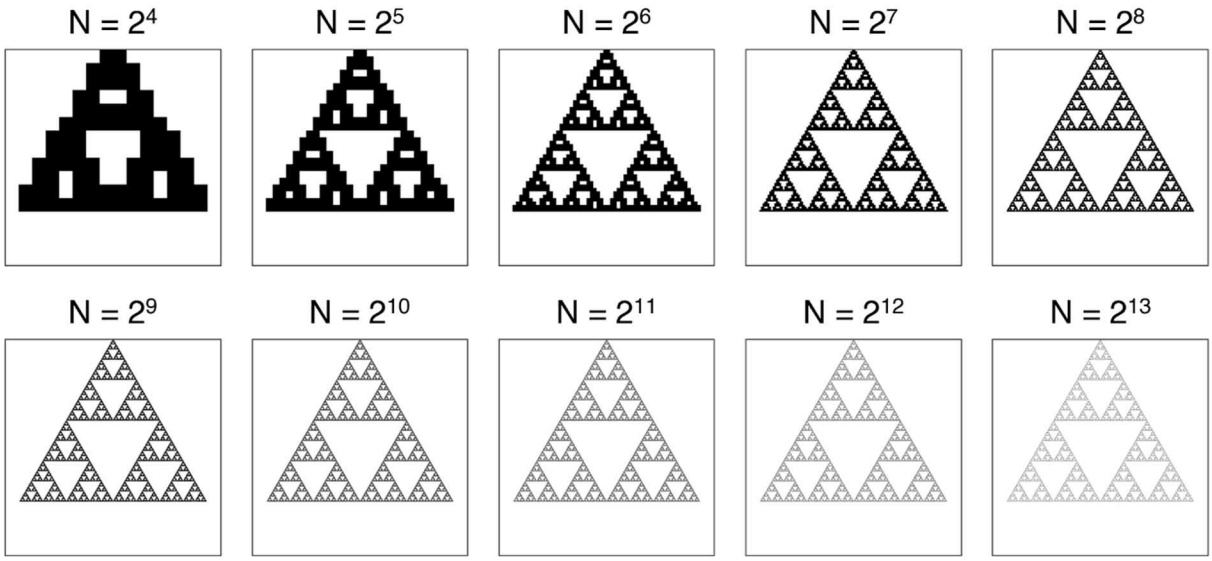

Non-chaotic
Chaotic
$\Delta S/\Delta S_{max}$
S_{LZ}
S_{samp}
S_{perm}
-1.5
-1
-0.5
0
0.5
1
1.5
λ

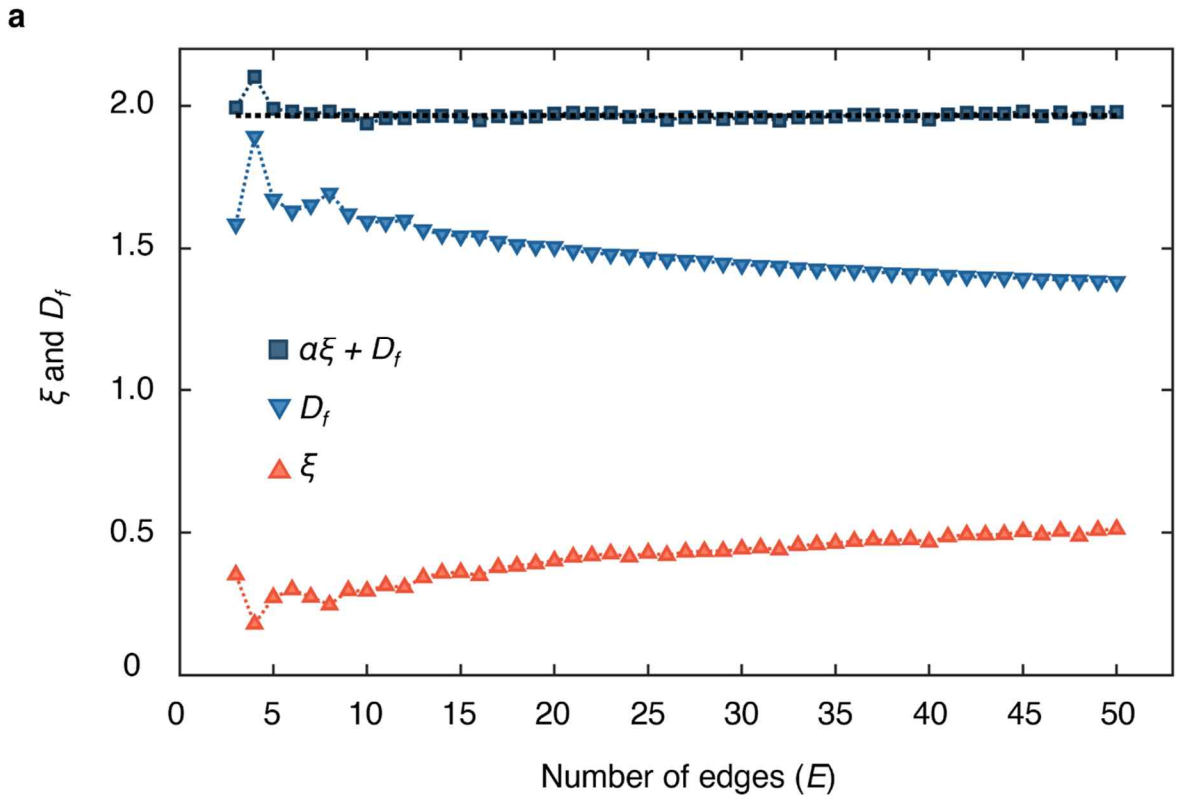
λ

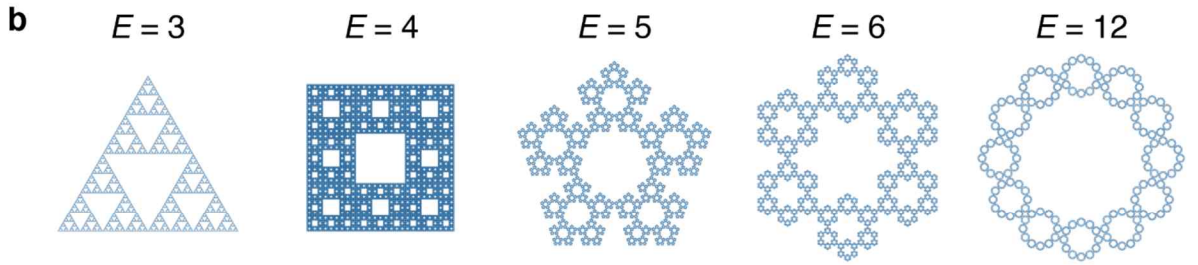
a

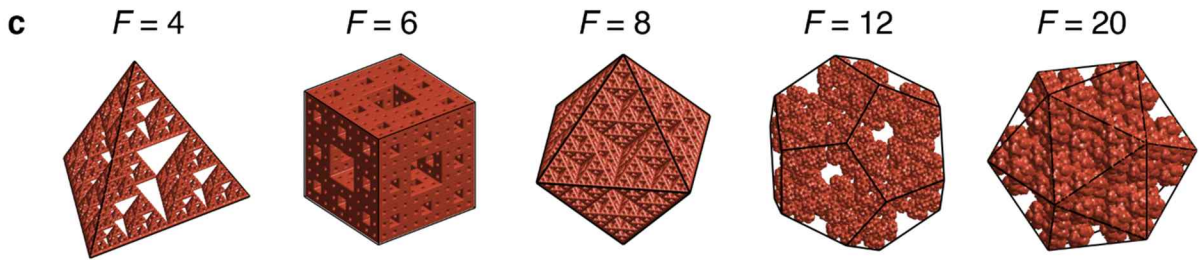


b









ξ D_f

E

D_f D_f ()()対数 / 対数 1/= n r n r

α

ξ = D

D D

二维（**b**）和三维（**c**）中边数或面数不同的分形。对于三维迭代函数系统，会调整面数（F），而二维迭代函数系统中的边数则受到控制。