

# 熵的前世今生

陈关荣

(香港城市大学)

熵 (Entropy) 这个名词最早由德国物理学家鲁道夫·克劳修斯 (Rudolf J. E. Clausius, 1822-1888) 于 1865 年在热力学的研究中提出。词源来自希腊文  $\epsilon\eta + \tau\rho\omicron\pi\eta$ , 意为“内在的变动”。当时克劳修斯用它来形式化热力学的第二定律, 并以简洁的形式描述了热物理过程的不可逆性。他说:“宇宙的熵趋于极大”。在克劳修斯之前, 法国数学和物理学家拉扎尔·卡诺 (Lazare N. M. C. Carnot, 1753-1823) 在 1803 年以及他的儿子、工程师萨迪·卡诺 (N. L. Sadi Carnot, 1796-1832) 在 1824 年就研究过热机的效率和损耗, 其工作为克劳修斯的理论奠定了基础, 只是他们没有使用“熵”一词。

后来, 到了 1870 年代, 德国物理学家路德维希·玻尔兹曼 (Ludwig Boltzmann, 1844-1906) 和美国机械工程师威拉德·吉布斯 (J. Willard Gibbs, 1839-1903) 在数学物理学中赋予熵以统计力学的表达和解释。

1923 年 5 月 15 日, 德国物理学家鲁道夫·普朗克 (Rudolf A. V. Plank, 1885-1973) 在中国讲学, 时任南京高等师范学校和国立东南大学教授兼物理系主任的胡刚复 (1892-1966) 为其担任翻译。胡刚复是江苏无锡人, 早年考取首批庚款留美资格, 1918 年获哈佛大学物理博士学位, 回国后成为中国近代物理学事业奠基人之一。当普朗克讲到“Entropy”这一概念特别是其物理定义  $dS = dQ/T$  时, 胡刚复遇到了难题: 中文里没有现成的对应词语。诗书满腹的胡刚复灵机一动造了个新字——熵。他的构思极为精妙: 新字取义于商, 因为根据普朗克的数学定义, Entropy (S) 在可逆过程中的微小变化是热量 Q 的微小变化除以温度 T 的商; 加火字旁, 是因为这一物理概念源于热力学; 不仅如此, 熵字的读音也与商相同, 达到了形音义的完美统一。

1927 年, 匈牙利裔美国数学和物理学家约翰·冯·诺依曼 (John von Neumann, 1903-1957) 引入了“冯·诺依曼熵”。它是量子信息论中的一个核心概念, 用于量化一个量子态所包含的“不确定性”或“混合程度”。冯·诺依曼熵不仅是量子统计力学的基础, 也是量子信息、量子计算、量子热力学和量子引力等领域的主要数学工具。

1948 年, 美国数学家和电气工程师克劳德·香农 (Claude E. Shannon, 1916-2001) 在他奠基性论文《通信的数学理论》(A mathematical theory of communication) 中将熵这一概念引入到信息论, 提出了“香农熵”的概念。它用于量化通信系统中一个离散随机变量的“不确定性”或它所包含的“平均信息量”: (1) 不确定性度量: 熵越大, 变量的结果越难预测 (例如, 均匀分布时熵最大, 而熵为 0 意味着变量是确定性的); (2) 信息量度量: 熵是多次观察该变量后获得的平均信息量 (即不确定性消除的程

度)；(3) 编码长度下限：在无损数据压缩中，香农第一定理（信源编码定理）表明，平均码长不可能低于信源的熵。香农熵是信息论、统计学、计算机科学等领域的一个基本概念。在可比较的条件下，香农熵一般都大于或等于冯·诺依曼熵，并且依赖于测量基的选择。

到了近代，熵的概念在理论物理学和天体物理学均有长足的发展。在天文学中，以前黑洞被视为仅由质量、电荷和角动量三个参数决定的天体，结构非常简单。不过该理论有一个严重问题：如果把一个携带大量物理熵的物体（如一杯热水）或一个大量信息熵的物体（如一本百科全书）扔进黑洞，物体的熵似乎就在宇宙中消失了，但这直接违反热力学第二定律，即孤立系统的熵永远不会减少。

1972年，以色列裔美国物理学家雅各布·贝肯斯坦（Jacob D. Bekenstein, 1947-2015）提出了革命性的猜想——黑洞本身必须具有熵。他认为，当一个物体落入黑洞时，黑洞的熵会增加，且增加量至少等于物体消失的熵，从而使得“外部世界熵+黑洞熵”的总和不会减少。他进一步猜想，黑洞的熵与其事件视界的面积成正比。1974年，起初并不赞同贝肯斯坦猜想的英国物理学家史蒂芬·霍金（Stephen W. Hawking, 1942-2018）运用弯曲时空量子场论进行分析之后，得出一个惊人的结论：黑洞并非全“黑”，它会因为量子效应在视界附近辐射粒子。这就是著名的“霍金辐射”。这一发现不仅证明了黑洞具有温度，也为贝肯斯坦的熵公式提供了严格的理论基础，因为一个具有温度的物体必然有熵。后来，该熵公式被命名为“贝肯斯坦-霍金熵”。贝肯斯坦-霍金熵指出，一个黑洞的熵极其巨大，它与其事件视界的面积而不是与黑洞的体积成正比。后来，贝肯斯坦-霍金熵成为连接宏观与微观、时空与量子、信息与引力的一个关键，是通往更深层物理理论即量子引力理论道路上最重要的路标。然而，贝肯斯坦-霍金熵引发了著名的“黑洞信息悖论”：如果黑洞通过辐射完全蒸发，那么最初形成它的物质所携带的信息就永远消失了，这违反量子力学信息守恒的基本原理。

上面提及的都是物理学背景引进的熵。在数学领域则有“度量熵”，特别是由苏联数学家安德雷·柯尔莫哥洛夫（Andrey N. Kolmogorov, 1903-1987）和俄裔美国数学家雅科夫·西奈（Yakov G. Sinai, 1935-）引进的“柯尔莫哥洛夫-西奈熵”，简称KS熵。它用以刻画给定测度下的平均信息产生速率，可以视为香农信息熵在动力系统中的推广。KS熵可以用于刻画一个确定性动力系统在演化过程中表现出的混乱程度或信息产生速率。两个共轭或同构的动力系统的KS熵相同。因此，KS及其他度量熵可用于区分不同类型的动力系统。

数学领域中的另一个重要概念是“拓扑熵”。在动力系统理论中，它用来量化一个动力系统在迭代过程中其轨道随时间演化的复杂性或混乱程度。与度量熵不同，拓扑熵只依赖于系统的拓扑结构，而不依赖于任何特定的度量或概率测度。最早的拓扑熵由美国数学家罗伊·阿德勒（Roy L. Adler, 1931-2016）、美国计算机科学家艾伦·康海姆（Alan G. Konheim, 1934-2019）和英国数学家麦克安德鲁（M. H. McAndrew, 剑桥大学1960年数学博士）于1965年引入，称为“AKM拓扑熵”，后来经其他一些数学家用开覆盖的语言给出更通用的定义。拓扑熵的基本思想是动力系统中分离轨道的增

长速度。具体地说，考虑一个动力系统在相空间中两条从不同起点出发的轨道，它们在初始时刻可以如此接近以至难以区分，但如果经过  $n$  次迭代后，其轨迹会逐渐清晰地分离开来，则可认为它们携带有不同的信息。拓扑熵粗略地表示为随着  $n$  的增长，长度为  $n$  的可分辨轨道的指数增长率。

拓扑熵在现代数学中有很多形式，也有很多应用。在混沌理论中，正拓扑熵是 Li-Yorke 混沌的充分但非必要条件。托扑熵可以用来区分不同复杂度的动力系统，特别是符号动力系统和双曲型耗散系统。在分形几何与热力学中，通过拓扑熵引入压力概念，可以刻画分形维数与测度的复杂性。在组合优化与元胞自动机理论中，有限自动机生成的子移位熵由其语言增长率的对数来量化表达。

今天，各种熵用作复杂系统里的无序性、不确定性和不可逆性的统一概念和量度。物理和数学中常用的熵综合在表 1 里。

名称	应用领域	核心表达式 (符号的定义从略)	主要性质
克劳修斯熵	热力学 (宏观)	$dS = \delta Q_{\text{rev}}/T$	状态函数，可逆热与温度之比
玻尔兹曼熵	统计物理	$S = k_B \ln W$	微观状态数的对数
吉布斯熵	统计物理	$S = -k_B \sum p_i \ln p_i$	概率分布的泛函
香农熵	信息论	$H = -\sum p_i \log p_i$	平均信息量
冯·诺依曼熵	量子信息	$S = -\text{Tr}(\rho \ln \rho)$	量子态的混合度
贝肯斯坦-霍金熵	引力/黑洞物理	$S_{\text{BH}} = k_B A / (4\ell_P^2)$	正比于黑洞视界面积
KS 熵	动力系统	$H_{\text{KS}} = \sup \lim_{n \rightarrow \infty} H(V_{k=0}^{n-1} T^{-k} \xi)$	用香农熵 $H$ 和测度定义的复杂度
拓扑熵	动力系统	由开覆盖或轨道增长定义	与测度无关的复杂度

表 1 各种物理熵和数学熵的简要比较

下面，我们来谈谈香农的信息熵，它是现代计算技术、生物物理和社会科学领域中使用得最多的熵。

前面说过，1948年，香农把熵这一概念引入到信息论中。后人把他定义的信息熵称为“香农熵”。香农研究了“一句话最少需要多少个符号来表达”这类信息理论问题。他在通信领域的研究中，建立了包括信源、发送器（编码器）、信道、接收器（解码器）、信宿以及噪声源组成的通信系统模型。其中，信息量就是解除信源不确定性程度所需信息的度量。

香农以概率论和数理统计为工具，提出了一个计算信息不确定性的数学公式（见表1）。该数学公式应用于由随机事件组成的离散型信源时，随机事件出现的信息不确定性程度用其出现的概率来描述：事件出现的可能性越小，概率就越小，而所含信息不确定性则越大；相反，事件出现的可能性越大，概率就越大，而所含信息不确定性则越小。因此，香农的信息熵被用于量化通信系统中一个离散随机变量所包含的“信息不确定性”：熵值越大，信息越不确定。例如，当我们对某事件一无所知时，信息熵最大；当我们已完全知晓事件内容和规律时，信息熵为零。

具体地说，香农熵定义为一个随机变量所有可能取值的信息量的期望值（见表1）：

$$H(x) = -\sum_i p_i \log p_i$$

其中，单个随机事件的信息量（自信息量）用公式

$$S(x) = \log_2(1/p(x)) = -\log_2 p(x)$$

来计算，单位为“比特”（bit），满足：

概率  $p(x) = 1$ （事件必然发生） $\rightarrow$  自信息量  $S(x) = 0$ （信息完全可以确定）

概率  $p(x) = 1/2$ （抛硬币） $\rightarrow$  自信息量  $S(x) = 1$ （信息在一半程度上可以确定）

概率  $p(x) = 0$ （事件必不发生） $\rightarrow$  自信息量  $S(x) = \infty$ （对应的熵没有意义）

这里，对数的使用为表达和计算自信息量带来了方便：

$$\log_2(1/p(x)) = \log_2(1/1/2) = \log_2(2) = 1 \text{ 比特}$$

其中1比特 = 一次公平的“是/否”判断所代表的信息量。

值得注意的是，香农熵用来衡量随机变量整体的不确定性，因此“单个事件的自信息量”（例如上面说的“单个事件必然发生”和“单个事件必不发生”的信息量）可以计算，但“单个事件的熵”是没有严格数学意义的。

香农的信息熵也可以看成是广义熵，或称为泛熵，用来广泛描述物质运动（包括生命现象）的混乱度或无序度。其对立面信息量（俗称为“负熵”），则是有序度或组织结构复杂程度的表示：生物系统的复杂化或有序化，对应信息论的熵减。今天，香农熵被广泛应用于数据压缩、密码学、机器学习、统计推断、生物信息学、经济学以及复杂网络等诸多领域。

下面，我们来细说“熵增”和“负熵”这两个饶有趣味也至关重要的论题。

## (1) 宇宙熵增定律

热力学有三条基本定律。热力学第一定律是能量守恒定律在热力学系统中的具体表述。它指出能量既不会凭空产生，也不会凭空消失，它只会从一种形式转化为另一种形式，或者从一个物体转移到另一个物体，而在转化或转移的过程中，能量的总量保持不变。

热力学第二定律，又称“熵增定律”，是一条与能量守恒有同等重要地位而又十分有趣的基本定律。能量守恒定律是描述自然界普遍适用的定律，而熵增原理仅适合于热力学孤立体系。熵增原理指出，在孤立系统中，如果变化过程是可逆的，则熵不变；如果过程是不可逆的，则熵增加，并且当熵增加到最大时，系统处于一种平衡态。熵增刻画了系统从有序向无序的演化：熵越高，系统越无序；熵越低，系统越有序。

生活经验告诉我们，热不可能自发地、毫无损失地从低温物体转移到高温物体，也不可能从单一热源取热而把它全部变为功却不产生任何其他影响。早在1850年，物理家克劳修斯就指出：在自然条件下热量只能从高温物体向低温物体转移，而不能由低温物体自动向高温物体转移；也就是说，在自然条件下，这个转变过程是不可逆的。要使热传递方向逆转过来，只有靠消耗功来实现。克劳修斯指出：“宇宙的熵趋于极大”。具体地说，在一个孤立系统里，熵总是在增大的，直到不能再增大为止。这时，系统内部达到一种完全均匀的热动平衡状态，不会再发生任何变化，除非外界对系统提供新的能量。因此，如果认为宇宙是自我封闭而不存在“外界”的话，那么宇宙一旦到达热动平衡态，就会完全死亡。这种情景称为“热寂”。

生活经验也告诉我们，正如房间里的空气分子，常态下它们是均匀地分布在整个空间里的，不会全部一起挤逼在某个角落而在其他地方留下真空。科学测量表明，空气分子的数量越多，这种散乱分布的可能性或者说概率就越大，大到几乎是必然的，这时它们的熵达到最大值。这个过程就是熵增过程，期间空气分子的运动是一种“混沌”状态，它具有正的熵。熵值越大，运动状态就越混沌。在极限情形，熵值最大，对应着完全的无序、均匀、死寂——没有温度差、没有密度差、没有任何可做功的梯度；一切可能的结构、信息、模式都彻底消失，那是一种完全混沌的状态。打趣地说，在这个极限状态下，熵变成了“筋”。

不过要注意，热力学第二定律本质上是统计规律，适用于由很大但有限数目分子所构成的独立系统及有限范围内的宏观过程，因此不能简单地应用到无限的宇宙，而需要通过所谓的“视界熵”和整体演化才能作适当的推广。

热力学第二定律还有一个重要的不可逆性。1851年，英国数学物理学家、工程师、第一代开尔文男爵（Lord Kelvin）威廉·汤姆森（William Thomson, 1824-1907）指出，在有限空间和时间内，一切和热运动有关的物理及化学过程都具有不可逆性。他论证了，自然界中任何形式的能都会很容易地变成热，而反过来热却不能在不产生其他影响的条件下完全变成另一种形式的能。他还说明了，这种转变在自然条件下也是不可逆的。今天，从统计力学的角度来看，这种不可逆性源于系统微观状态数目的差异。

熵增的方向是系统从概率较小的有序状态向概率更大的无序状态发展的方向。逆过程原则上虽然概率不为零（如涨落），但随着系统粒子数增多，发生的概率极低，宏观上可以忽略不计。

在 1890 年代，德国逻辑学家和数学家恩斯特·策梅洛（Ernst F. F. Zermelo, 1871-1953）与奥地利物理学家路德维希·玻尔兹曼（Ludwig E. Boltzmann, 1844-1906）爆发过一场关于热力学第二定律的持久公开的大辩论。当年还是德国物理学家马克斯·普朗克（Max K. E. R. Planck, 1858-1947）博士研究生的策梅洛在《物理学纪事》（*Annalen der Physik*）杂志发表了一篇论文，向资深的物理学家玻尔兹曼关于热力学第二定律的统计理论提出了挑战，引起了一场旷世辩论。当时争论的核心是能否从经典力学推导出上面所说的不可逆性。辩论的结果澄清了热力学不可逆性的统计本质，促进了热力学从经典力学到统计力学的范式转变，成为热力学与统计物理学发展史上的一个里程碑。

关于热力学第二定律，爱因斯坦在 1949 年出版的《自述》一书中说：“这是唯一具有普遍内容的物理理论。我确信，在其基本概念的适用框架内，它将永远不会被推翻。”

热力学还有第三定律，它指出不可能通过有限步骤将任何系统的温度降低到绝对零度。因为通过测量物质从接近绝对零度到有限目标温度的热容量便可算得相应的熵值，所以热力学第三定律的各种表述均与绝对零度及其熵的不可达性有关。

## （2）生命以负熵为生

奥地利物理学家埃尔温·薛定谔（Erwin Schrödinger, 1887-1961）于 1944 年出版了《生命是什么》（*What is life?*）一书，那是一部从物理学视角审视生物学的经典著作。该书基于他本人 1943 年的一次演讲，运用热力学和量子力学理论，提出了“非周期性晶体”（遗传物质基础）和“负熵”（生命维持机制）等核心概念，预言了遗传密码的存在，对分子生物学的诞生和 DNA 的发现产生了深远影响。

薛定谔在《生命是什么》中说：“人活着就是在对抗熵增定律，生命以负熵为生。”

从物理学的角度来说，对抗熵增的有效途径是通过各种“耗散结构”进行的。这是比利时裔美国物理化学家伊利亚·普里高津（Ilya Prigogine, 1917-2003）提出的理论。他说的耗散结构是一个“不封闭、能自己保持有序”的系统，它有四个主要特征：开放性、非线性、远离平衡态、涨落触发。前面提到，熵增是在封闭（或说孤立）系统内部发生和进行的。开放系统允许与周边环境沟通交互，使得它有机会吸收负熵。此外，封闭系统内部熵增的结果是熵趋于极大从而系统趋于平衡态。因此，对抗熵增就要远离平衡态。此外，远离平衡态的耗散结构总是通过某种非线性的“突变”现象而发生的。事实上，突变临界值的存在是耗散系统的主要特征。最后，随机的微小扰动（涨落）在合适条件下会被放大，从而触发系统跃迁到新的有序态。

普里高津把经典热力学平衡态推广到非平衡态，将封闭系统熵的概念扩展到开放系统，从而建立了非平衡态热力学。他指出：“非平衡是有序之源”，即非平衡的不可逆性是

形成有序结构过程中不可或缺的因素。普里高津的新理论包括熵的平衡方程、非平衡线性区的最小熵产生定理以及远离平衡区域的耗散结构。他的理论超越了经典热力学范畴，在其他学科领域中也有许多应用。

地球上的生物系统是一个开放系统，它通过从环境中摄取低熵物质（有序高分子）和向环境释放高熵物质（无序小分子）来维持自身处于低熵有序的状态。上面提到，每个孤立系统和封闭过程，都在朝着“熵增”的方向发展，一步步走向“热力学平衡”的“衰退”直至“死亡”。那么，生命有机体是如何避免或延迟衰退的呢？答案是让自身系统开放，与环境交互，通过吃、喝、呼吸，即新陈代谢，来对抗熵增。通过物质交换和化学反应，生物体产生能量，并不断与环境进行交流互换，源源不断地获取负熵。这里，负熵的定义并非正熵的简单直接的相反量，而是泛指生命系统通过代谢、摄取能量等方式“输入秩序”，即减少系统内部的混乱程度，借以维持自身结构的有序性。这一过程可理解为“汲取负熵”，是一个熵减的过程，它消耗能量或资源，即进行减熵，让生命体从无序变得有序，并通过这个过程来维持生命。

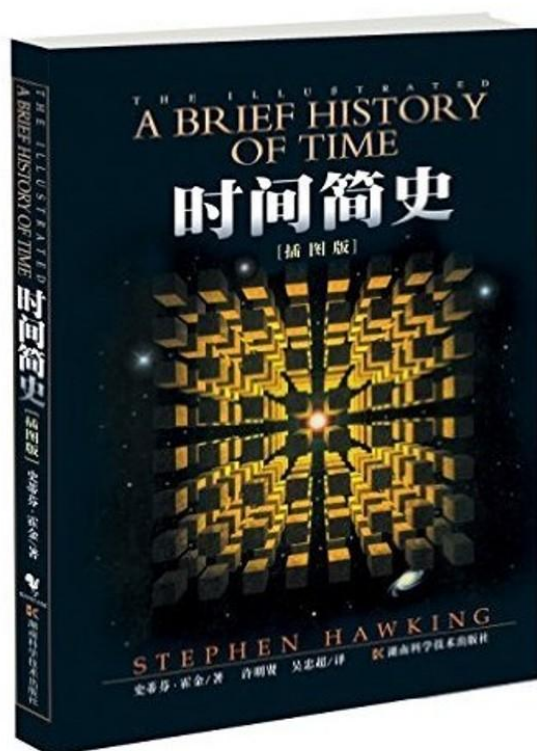


图1 霍金《时间简史：从大爆炸到黑洞》

霍金在其畅销书《时间简史：从大爆炸到黑洞》（A brief history of time: From the big bang to black holes, 图1）中说：“混乱度或熵的增加，正是区分过去与未来的所在，它为时间赋予了方向。”或许我们可以说，熵也为复杂系统的演化赋予了方向。

(笔者感谢中央财经大学梁超教授对文稿的细致评阅和修改意见)