

# 计算之物理：灵魂耗能吗？

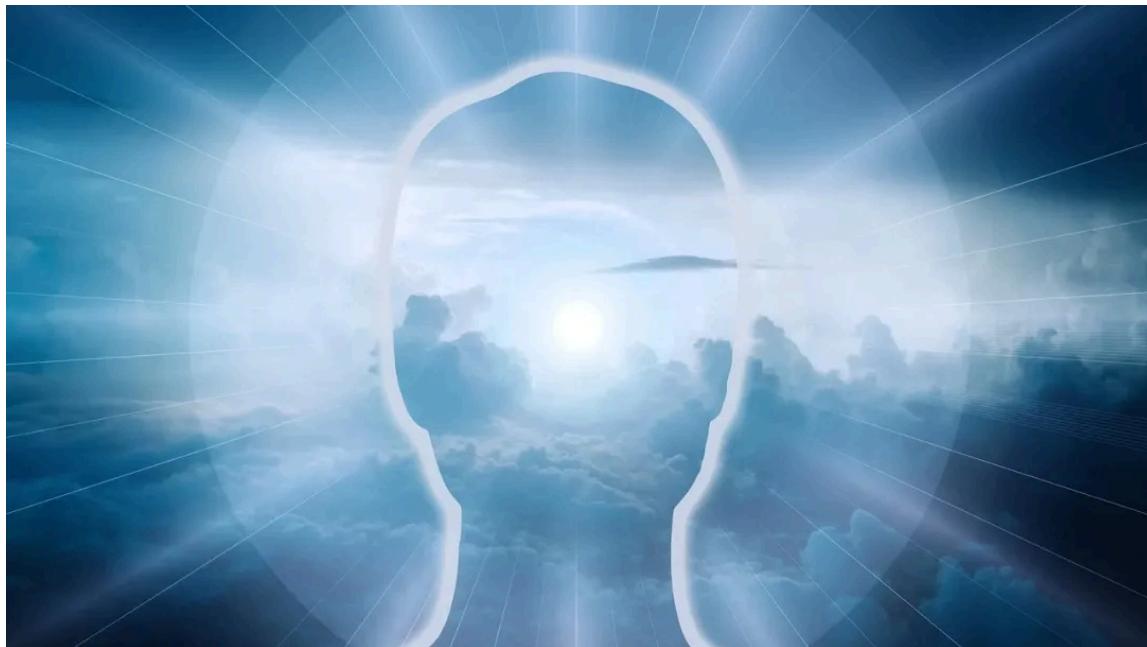
中国物理学会期刊网 2025年01月11日 10:03 北京

以下文章来源于赛先生，作者尼克

赛先生  
百家·科学·社会

## 赛先生

赛先生由非营利公益组织北京市海淀区智识前沿科技促进中心主办，中国科学院院士潘...



图源：Pixabay

“

### 导读：

从麦克斯韦妖到弗雷德金门，可以体会科学史和哲学史深刻且又老生常谈的概念对，如信息vs物理，唯心vs唯物，bit vs it，灵魂vs肉体等。人工智能对算力的剧烈需求，迫使我们思考新的概念对：计算vs能耗。对历史的小心审视能为我们展望未来提供可靠的基础。

尼克 | 撰文

只有一件事要紧，那就是遗忘。

There is only one thing. It is oblivion.

——博尔赫斯Jorge Luis Borges

## CAIXIANSHENG

### 1、引言

大语言模型强化了我们的一个直觉：计算需要能量。“奇点”派及“强人工智能派”都预言奇点，即人工智能会全面超越人类的时刻，会在2030-2040年间来临。有人甚至更为具体地宣称：2028年最大的大模型（极有可能是OpenAI的某个版本）需要相当于一千万张英伟达H100 算力的GPU集群，消耗10G瓦的电力，这个电量差不多是10个胡佛大坝的发电量，可以供美国一个中小规模的州的能耗。而当2030年奇点真正来临时（差不多2028年2到5年后），最大的大模型需要消耗美国20%的电力（见Aschenbrenner-2024）。

算力和能耗不仅是工程上的燃眉之急，它们之间的关系也是最深刻的科学议题之一，且有着悠久的智识 (intellectual) 传统。计算或者信息在哲学话语中对应着唯心的东西，而能量对应着唯物的东西。过去，我们很容易偷懒地把“心”最终归结为“灵魂”，从而把“意识”这类难以定义的词汇用“碳基”去消解 (explain away)。但我们很少去问：是否唯心的东西必须要具备物理的体现；此外，灵魂是否也要制约于物理定律？图灵之后对此有敏锐洞察的是布里渊和冯诺依曼。

冯诺依曼1949年在伊利诺伊大学做了题为“复杂自动机的理论和组织” (Theory and Organization of Complicated Automata) 的讲座，讲座分5次，讲稿经他的助手亚瑟·伯克斯 (Arthur Burks) 1966年整理并做注后编入文集《可再生自动机的理论》 (Theory of self-reproducing automata) 中（见 von Neumann-1966）。这些文献的价值并没有在后来计算机科学的发展中得到重视。布里渊在1956年出版了《科学与信息论》 (Science and Information Theory)。他与冯诺依曼都不约而同地意识到计算会受到能耗的约束，认为最小的计算（从0变为1或反之）会消耗 $k \cdot T \cdot \ln 2$ 的能量。“熵”原来是一个神奇的热力学概念，但冯诺依曼在帮香农给信息论的新概念也起名“熵”时，除了它们字面上的相似，是否想到它们其实是一回事呢？

## CAIXIANSHENG

### 2、麦克斯韦妖与熵

最早系统地提出热力学理论的是苏格兰启蒙运动的重要人物约瑟夫·布莱克 (Joseph Black, 1728年—1799年)。他是瓦特的朋友和资助者。他建立了“热质理论” (caloric theory, 或译“热素说”)，此理论认为热是一种称为“热质”的物质，热质可以从温度高的物体传到温度低的物体。拉瓦锡推翻燃素说之后，热素说开始流行。布莱克提出热质守恒定律，即热质不能产生也不能消失，只能从一个物体传到另一个物体。卡诺之后提出基于“卡诺循环”的理论热机“卡诺热机”，以及卡诺定理。

克劳修斯和开尔文勋爵由卡诺定理发展出各自对于热力学第二定律的表述。克劳修斯断言热不能从低温物体传到高温物体 (Heat can never pass from a colder to a warmer body without some other change...)。开尔文勋爵已经意识到热力学第二定律有可能和智能生命有关系，他的一种表述是：只靠无生命的物质媒介，不可能通过把物质的任何部分冷却到低于周围最冷物体的温度来获得机械效应 (It is impossible, by means of *inanimate material agency*, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects)。此时物理学家们认识到热是能量，但热本身并不能做功，只有热的流动才能做功。只有存在温度差的情况下，才能有热流。能转化为有效机械功的热能是有限的。物理学家诉诸“熵”这个物理量来刻画热力学第二定律。“熵”表示系统的杂乱无章的程度。

詹姆斯·克拉克·麦克斯韦 (James Clerk Maxwell, 1831-1879, 父亲出生名约翰·克拉克，但为保护所剩不多的遗产，在原姓克拉克后加了个远祖的姓麦克斯韦) 是可比肩牛顿和爱因斯坦的大智之人。现代物理学的很多成就源自于他。爱因斯坦在他的办公室里常年挂着麦克斯韦的照相。麦克斯韦从小就兴趣广泛且博学，他博学的一面被他学问的深刻所掩盖。麦克斯韦的传记作者猜测他受到了当时其他博学者的启发，例如数学家巴贝奇。



詹姆斯·克拉克·麦克斯韦 (James Clerk Maxwell, 1831-1879)

麦克斯韦在1860年代建立了基于分子运动的热力学理论，他指出分子运动的速度是温度的函数；分子的动能与温度成正比，温度越高，分子的无序运动就越快。他1867年12月给他的终身挚友Peter Tait写过一封信，信中说“设想一个有限生灵，能通过简单的观察知道所有分子的路径和速度……” (Now conceive a finite being who knows the paths and velocities of all molecules by simple inspection but who can do no work except open and close a hole in the diaphragm by means...) 此处finite being汉语常译为“有限生灵”，其实being并无“生灵”之意。1871年麦克斯韦出版《热理论》 (*Theory of Heat*) —

书，书中最后一章倒数第二节“热力学第二定律的限制”(Limitations of the Second Law of Thermodynamics) 中用有生命的he而不是无生命的it来指代 being。

Tatlock

O.T. Ranking has returned  
the proof, <sup>corrected</sup> somewhat changed.  
Do you now adopt it, and  
may I print off with a  
clear conscience? If not  
pray Criticize fully (in writing)  
his remarks on Clausius,  
that U & the Jewel may get  
justice. Also say what  
occurs to you about the  
enclosed which return speedily.  
The mixing of the gases which is  
an objection to the test may easily  
be got over by waiting longer till  
two (one from each vessel) impinge  
centrically. What then? P.S. 21

All very well now - But  
where is there?

Now let A & B be two  
vessels divided by a  
diaphragm and let them contain elastic  
molecules in a state of agitation which  
strike each other and the sides.  
Let the number of particles be equal in A & B  
but those in A have the greatest energy of motion.  
Then even if all the molecules in A have equal  
velocities, if collisions occur between them  
their velocities will become unequal & I have  
shown that there will be velocities of all magnitudes  
in A and the same in B only the sum of the  
squares of the velocities is greater in A than  
in B.

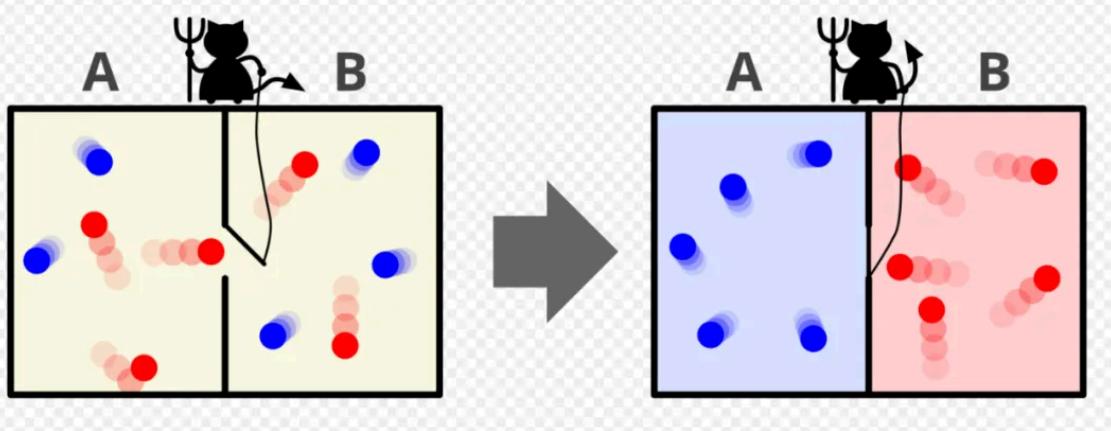
When a molecule is reflected from the <sup>fixed</sup> diaphragm  
CD no work is lost or gained.

If the molecule instead of being reflected were  
allowed to go through a hole in CD no work  
would be lost or gained, only its energy would  
be transferred from the one vessel to the other.

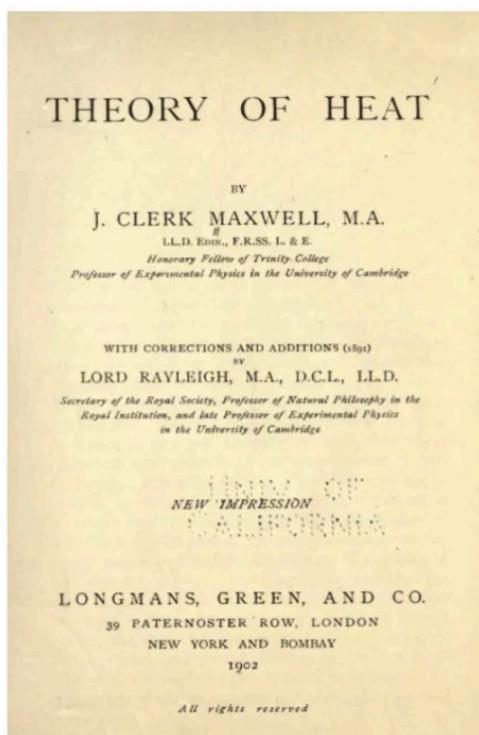
Now conceive a finite being who knows  
the paths and velocities of all the molecules  
by simple inspection but who can do no work  
except to open and close a hole in the diaphragm  
by means of a slide without mass.

麦克斯韦1867年12月给挚友泰特的信。右下角那段话提出“麦克斯韦妖”

麦克斯韦设想了一个思想实验，一个容器包含很多分子，能量是随机分布而不是均匀分布的，因而有些分子运动快而有些分子运动慢。把一个容器通过隔板一分为二，隔板上有个小洞，这个有限生灵可以把观察的知识应用于容器，即：让运动较快的分子通过小洞，而较慢的分子则不通过小洞。结果便是：容器的一侧是运动较快的分子，而另一侧是运动较慢的分子。这明显违背了热力学第二定律。



“麦克斯韦妖”示意图。妖 (demon) 决定哪些分子通过小洞。



#### LIMITATION OF THE SECOND LAW OF THERMODYNAMICS.

Before I conclude, I wish to direct attention to an aspect of the molecular theory which deserves consideration.

One of the best established facts in thermodynamics is that it is impossible in a system enclosed in an envelope which permits neither change of volume nor passage of heat, and in which both the temperature and the pressure are everywhere the same, to produce any inequality of temperature or of pressure without the expenditure of work. This is the second law of thermodynamics, and it is undoubtedly true as long as we can deal with bodies only in mass, and have no power of perceiving or handling the separate molecules of which they are made up. But if we conceive a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are still as essentially finite as our own, would be able to do what is at present impossible to us. For we have seen that the molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will thus, without expenditure of work, raise the tem-

《热理论》 *Theory of Heat*, 1871。最后一章最后两节，用代词he而不是it指代being

就像图灵机是图灵的老师丘奇给起的名，“麦克斯韦妖” (Maxwell's Demon) 的说法并非源自麦克斯韦本人，而是源自开尔文勋爵。他1879年给 *Nature* 写了篇短文“麦克斯韦的排序妖”，提到一个妖 (demon) 可以给原子按照各种性质为原子的位置排序，如果按照原子的运动速度排序，就会导致温度失衡，违反热力学第二定律。

pteruges, and a Macroglossus, also a species of *Scotophilus*, probably *S. nigrogriseus*, Gould.

The fishes were not numerous, but some interesting species were obtained, respecting which the Hon. Wm. Macleay read some remarks which will be published in due time.

The insects, about fifty species, will also we hope be taken up by Mr. Macleay. Among the spiders are some very interesting forms, apparently quite new.

The collection of birds numbers about 350 specimens, the mammals about 120. Spirit specimens about 100.

We believe a large portion of this fine collection has been secured by the curator, Mr. E. P. Ramsay, F.L.S., for the Australian Museum.

#### THE SORTING DEMON OF MAXWELL<sup>1</sup>

THE word "demon," which originally in Greek meant a supernatural being, has never been properly used to signify a real or ideal personification of malignity.

Clerk Maxwell's "demon" is a creature of imagination having certain perfectly well-defined powers of action, purely mechanical in their character, invented to help us to understand the "Dissipation of Energy" in nature.

He is a being with no preternatural qualities, and differs from real living animals only in extreme smallness and agility. He can at pleasure stop, or strike, or push, or pull any single atom of matter, and so moderate its natural course of motion. Endowed ideally with arms and hands and fingers—two hands and ten fingers suffice—he can do as much for atoms as a pianoforte player can do for the keys of the piano—just a little more, he can push or pull each atom in any direction.

He cannot create or annul energy; but just as a living animal does, he can store up limited quantities of energy, and reproduce them at will. By operating selectively on individual atoms he can reverse the natural dissipation of energy, can cause one-half of a closed jar of air or of a bar of iron to become glowingly hot and the other ice-cold; can direct the energy of the moving molecules of a basin of water to throw the water up to a height and leave it there proportionately cooled (1° Fahr. for 772 feet of ascent); can "sort" the molecules in a solution of salt or in a mixture of two gases, so as to reverse the natural process of diffusion, and produce concentration of the solution in one portion of the water, leaving pure water in the remainder of the space occupied; or, in the other case, separate the gases into different parts of the containing vessel.

"Dissipation of energy" follows in nature from the fortuitous concourse of atoms. The lost motivity is essentially not restorable otherwise than by an agency dealing with individual atoms; and the mode of dealing with the atoms to restore motivity is essentially a process of assortment, sending this way all of one kind or class, that way all of another kind or class.

The classification, according to which the ideal demon is to sort them, may be according to the essential character of the atom; for instance, all atoms of hydrogen to be let go to the left, or stopped from crossing to the right, across an ideal boundary; or it may be according to the velocity each atom chances to have when it approaches the boundary: if greater than a certain stated amount, it is to go to the right; if less, to the left. This latter rule of assortment, carried into execution by the demon, disequalises temperature, and undoes the natural diffusion of heat; the former undoes the natural diffusion of matter.

By a combination of the two processes, the demon can decompose water or carbonic acid, first raising a portion of the compound to dissociational temperature (that is, temperature so high that collisions shatter the compound

<sup>1</sup> Abstract of Lecture at the Royal Institution, Friday, February 28, 1879, by Sir William Thomson, LL.D., F.R.S.

molecules to atoms), and then sending the oxygen atoms this way, and the hydrogen or carbon atoms that way; or he may effect decomposition against chemical affinity otherwise, thus: Let him take in a small store of energy by resisting the mutual approach of two compound molecules, letting them press, as it were, on his two hands, and store up energy as in a bent spring; then let him apply the two hands between the oxygen and the double hydrogen constituents of a compound molecule of vapour of water, and tear them asunder. He may repeat this process until a considerable proportion of the whole number of compound molecules in a given quantity of vapour of water, given in a fixed closed vessel, are separated into oxygen and hydrogen at the expense of energy taken from translational motions. The motivity (or energy for motive power) in the explosive mixture of oxygen and hydrogen of the one case, and the separated mutual combustibles, carbon and oxygen, of the other case, thus obtained, is a transformation of the energy found in the substance in the form of kinetic energy of the thermal motions of the compound molecules. Essentially different is the decomposition of carbonic acid and water in the natural growth of plants, the resulting motivity of which is taken from the undulations of light or radiant heat, emanating from the intensely hot matter of the sun.

The conception of the "sorting demon" is purely mechanical, and is of great value in purely physical science. It was not, invented to help us to deal with questions regarding the influence of life and of mind on the motions of matter, questions essentially beyond the range of mere dynamics.

The discourse was illustrated by a series of experiments.

#### PAOLO VOLPICELLI

THIS eminent Italian physicist, whose death we recently recorded, was born at Rome on January 8, 1804. He lost his mother a few days after his birth; his father was Prof. Alexander Volpicelli, a member of the Medical College of the Roman University. Paolo was educated at the college of Veroli and the University of Rome, where, in accordance with the wish of his father, he commenced the study of medicine, but abandoned it after the first year, declaring that medicine was not a science. Of his own accord he applied himself seriously to the course of mathematical philosophy, and four years later received the degree of doctor *ad honorem* in that faculty. It should be mentioned that doctorates *ad honorem* are given to only two students each year, and Volpicelli's fellow-doctor was the eminent Professor Tortolini, who followed the same course. Before leaving the University his professors recommended him to the Government for a scientific position. In fact, Prof. Morichini wished to name him his successor to the Chair of Chemistry in the University of Rome, but Volpicelli preferred to succeed Dr. Barlocchi, Professor of Experimental Physics in the same University, and in 1845, on the death of Barlocchi, became titular professor. Volpicelli occupied this chair till 1873, when he was appointed Professor of Mathematical Physics in the same University. In 1851 he was made a member of the Philosophical College, an honour accorded to only twelve professors of the University of Rome. Besides his position at this University, Volpicelli also filled that of Professor of Mathematical Physics at the Roman Seminary, taught geometry to the pupils of St. Michael's Hospital, and founded at Rome the special School of Artillery, of which he was director for thirty years.

When Pope Pius IX. revived the celebrated and historical Lincei Academy in 1847, Volpicelli was appointed secretary, a post which he held for thirty years, when, as his health was failing, the academicians made him secre-

开尔文勋爵1879年在Nature上刊文第一次叫出"麦克斯韦妖"的说法

此文的最后一句话：“麦克斯韦妖的概念是纯机械的，在纯粹物理科学中有用，但不宜用来研究生命和心灵对物质运动的影响。” (The conception of the "sorting demon" is purely mechanical, and is of great value in purely physical science, It was not, invented to help us to deal with questions regarding the influence of life and of mind on the motions of matter...) 意思是唯物和唯心不是一回事。可见麦克斯韦和开尔文勋

爵的小心。冯诺依曼肯定知道开尔文勋爵的这个说法，估计也会猜测麦克斯韦和开尔文勋爵对物质和心灵的看法，否则他不会贸然为香农的信息熵冠名。

## CAIXIANGCHENG

### 3、西拉德与智能妖

费米曾经有个半玩笑的“费米悖论”：为什么地球之外没有智能生命的证据？如果地球上也有智能生命，其他地方出现生命的概率也应该不小啊。当时在场的匈牙利物理学家西拉德（Leo Szilard）回答说：“火星上原来就有啊，只不过都移民到地球上去了，他们现在被称为匈牙利人。”于是在二十世纪上半叶，科学界把那时匈牙利出生的犹太裔数学家和物理学家群体统称为“火星人”。他们中除了西拉德本人之外，还有终身流浪的数学家爱多什（Paul Erdos），逻辑学家哈尔莫斯（Paul Halmos），数论家和教育家乔治·波利亚（George Polya），氢弹之父爱德华·泰勒（Edward Teller），数学物理学家尤金·维格纳，力学家冯卡门等。当然，他们都心服口服地认为最聪明的“火星人”是冯诺依曼。

西拉德最出名的事是1939年给罗斯福总统写了封信，主张美国尽早开发原子武器，并拉着爱因斯坦签了名，这封被称为“爱因斯坦-西拉德信函”催生了曼哈顿计划。1945年7月他又联合155名科学家给杜鲁门总统写请愿信，呼吁小心使用原子弹，但杜鲁门没收到。为了写这封信，西拉德在曼哈顿计划的芝加哥实验室做了调查：46%的科学家认为应该在日本作为展示性使用，希望日本投降，否则就真扔原子弹；还有2%的科学家认为不应该被使用，事后应该完全保密。他的一些控核思路后来被国际原子能委员会采纳，在面临人工智能威胁的今天，仍有启发。



西拉德 (Leo Szilard, 1898-1964)

相比之下，西拉德的一些重要的物理学贡献却显得不那么出名。他本科时的专业是电机工程，一战结束后在柏林认识了爱因斯坦，并和另一位火星人尤金·维格纳一起参加了爱因斯坦在柏林大学主持的统计物理讨论班。他1922年的博士论文试图建立热力学和信息的联系，并得到爱因斯坦的赞许。西拉德一直认为自己数学不好，不适合干理论物理。而爱因斯坦则认为西拉德总有原创的思想，更适合做一个发明家，爱因斯坦还以自己的经历启发西拉德：“你为什么不到专利局找份差事呢？做一个别人等着你下金蛋的科学家并不是一份好差事。我在专利局工作的时候是我一生最好的时光。”西拉德和爱因斯坦合写了5项冰箱的专利。西拉德还写过加速器的专

利。后来劳伦斯因为回旋加速器得了诺贝尔奖，未能得奖的西拉德一直耿耿于怀。西拉德最早想到了链式反应，居然还把这个主意也申请了专利，1936年他在得到不制造武器的保证后，把这项专利移交给了英国海军。后来他还和费米合写过一个核反应堆的专利。

西拉德1929年用德文写成的文章“论在智能体干预下，热力学系统的熵减”（On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings），发表在《物理学杂志》（*Zeitschrift für Physik*）上。直到1964年他临死才被朋友翻译为英文。

In memory of Leo Szilard, who passed away on May 30, 1964, we present an English translation of his classical paper *Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen*, which appeared in the *Zeitschrift für Physik*, 1929, 53, 840–856. The publication in this journal of this translation was approved by Dr. Szilard before he died, but he never saw the copy. At Mrs. Szilard's request, Dr. Carl Eckart revised the translation.

This is one of the earliest, if not the earliest paper, in which the relations of physical entropy to information (in the sense of modern mathematical theory of communication) were rigorously demonstrated and in which Maxwell's famous demon was successfully exorcised: a milestone in the integration of physical and cognitive concepts.

## ON THE DECREASE OF ENTROPY IN A THERMODYNAMIC SYSTEM BY THE INTERVENTION OF INTELLIGENT BEINGS

by Leo Szilard

Translated by Anatol Rapoport and Mechthilde Knoller from the original article "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen." *Zeitschrift für Physik*, 1929, 53, 840–856.

~~~

The objective of the investigation is to find the conditions which apparently allow the construction of a perpetual-motion machine of the second kind, if one permits an intelligent being to intervene in a thermodynamic system. When such beings make measurements, they make the system behave in a manner distinctly different from the way a mechanical system behaves when left to itself. We show that it is a sort of a memory faculty, manifested by a system where measurements occur, that might cause a permanent decrease of entropy and thus a violation of the Second Law of Thermodynamics, were it not for the fact that the measurements themselves are necessarily accompanied by a production of entropy. At first we calculate this production of entropy quite generally from the postulate that full compensation is made in the sense of the Second Law (Equation [1]). Second, by using an inanimate device able to make measurements—however under continual entropy production—we shall calculate the

Requests for reprints of this paper should be sent to:

Mrs. Leo Szilard  
2380 Torrey Pines Road  
La Jolla, California 92038

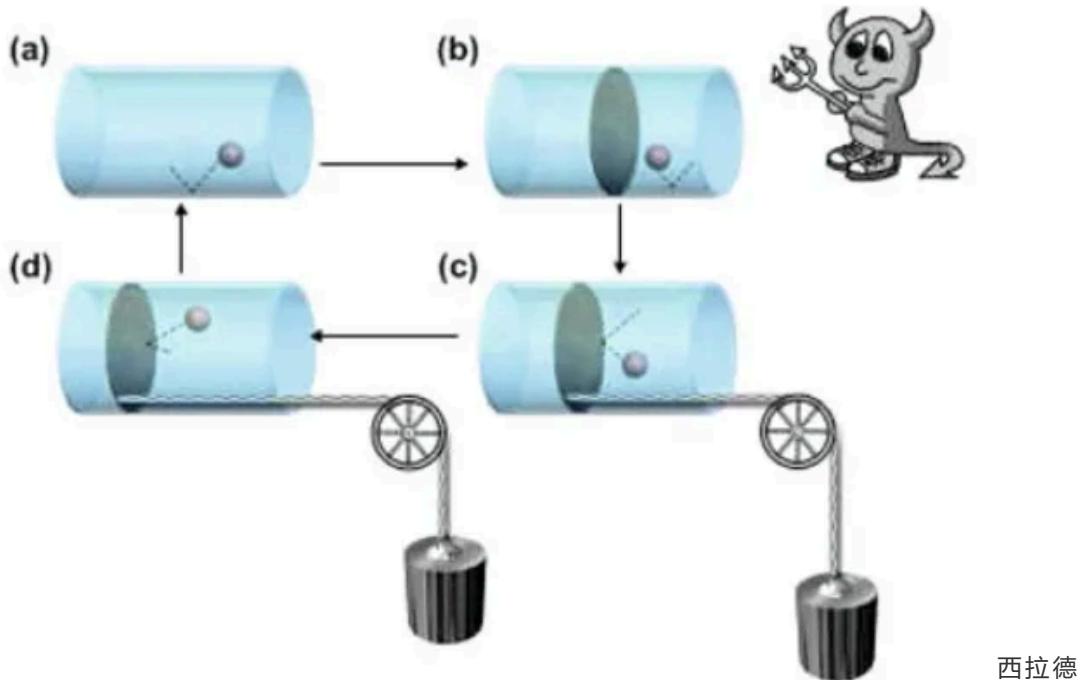
resulting quantity of entropy. We find that it is exactly as great as is necessary for full compensation. The actual production of entropy in connection with the measurement, therefore, need not be greater than Equation (1) requires.

~~~

**T**HREE is an objection, already historical, against the universal validity of the Second Law of Thermodynamics, which indeed looks rather ominous. The objection is embodied in the notion of Maxwell's demon, who in a different form appears even nowadays again and again; perhaps not unreasonably, inasmuch as behind the precisely formulated question quantitative connections seem to be hidden which to date have not been clarified. The objection in its original formulation concerns a demon who catches the fast molecules and lets the slow ones pass. To be sure, the objection can be met with the reply that man cannot in principle foresee the value of a thermally fluctuating parameter. However, one cannot deny that we can very well measure the value of such a fluctuating parameter and therefore could certainly gain energy at the expense of heat by arranging our interven-

只从题目我们就能领略西拉德是多么超前，他跨越了麦克斯韦和开尔文勋爵不以生命或智能论熵的约定。西拉德认为开尔文勋爵称之为“妖”(demon)的东西是某种智能生命(intelligent living being)。二十多年后冯诺依曼建议香农使用“熵”来命名信息量，估计是受到西拉德的启发。也许冯诺依曼并不仅仅是看到热力学熵和信息熵只是公式的形式上长得像，而是相信它们本质上就有某种关联。

西拉德的智能体 (intelligent being) 到底是什么？那时还没有图灵机。西拉德把麦克斯韦的多分子简化为单个分子。西拉德的思想实验可以把信息转化为能量。西拉德认为信息的获取，即测量，需要付出代价，即消耗能量，从而使得周边环境的熵增加。



西拉德

引擎。通过观察把信息转变为能量。整个操作分为4步，a) 单粒子引擎；b) 插入隔板；c) 在活塞上挂砝码；d) 单粒子运动使得活塞做功。结论：观察需要能量。

智能体的一次测量可获取一个比特的信息。热力学第二定律将系统能够做的最小的功限定为  $k_B T \cdot \ln(2)$ 。这里， $k_B$  是玻尔兹曼常数， $T$  是开尔文温度。“2”是因为西拉德引擎中有两个室。因此， $k_B T \cdot \ln(2)$  是可以从一个比特的信息中所能获得的最大能量。西拉德通过“智能体”把信息与能耗联系起来：信息熵和热力学熵是等价的。后来，杰恩斯 (Edwin Jaynes) 指出，熵本质上刻画化了对系统的无知程度，即在宏观监控系统时丢失的微观状态的信息量。只有在知道所有的微观状态时，能量才能完全转化为功。杰恩斯的这个观察成为信息与能耗关系的理论基础之一。孤立系统的热力学熵只能增加，其信息量只能减少。

法国生物学家雅克·莫诺曾经问西拉德为什么晚年的兴趣转往生物学——他一直认为西拉德所做的关于麦克斯韦妖的工作是生物学的而不是物理学的。一点也不奇怪，在相关领域工作过的薛定谔、冯诺依曼、图灵，香农，以及后来的蔡廷和李明等都转向生物学。事实上，麦克斯韦1871年的著作《热理论》的结尾中，他也考虑过生物物种的进化论与当时所知的分子理论之间的关系。

中国老话说：人往高处走，水向低处流。向低处流是自然的熵增过程，往高处走则是反抗的熵减过程。显然，向高处走更难，需要能量。调动能量的过程就是智能。这也许可以作为，除了图灵机之外，智能的另一种定义。这两个定义应该是等价的。

CAIXIANSHENG

## 4、兰道尔与不可逆计算

苏联数学家柯尔莫哥洛夫 (Kolmogorov) 利用图灵机重新定义了信息量，于是信息处理和计算本质上就是一回事。信息熵和热力学熵等价的话，一个自然的问题就是：信息处理过程，或者说计算，需要消耗能量吗？最早对此进行深刻分析的是 IBM 的物理学家兰道尔 (Rolf Landauer) ，他 1961 年发表文章“Irreversibility and Heat Generation in the Computing Process”，讨论了不可逆计算，以及信息擦除，所需要的最小能耗。此后计算及信息的热力学 (thermodynamics of computation and information) 作为统计力学和计算理论的交叉学科，一直是一小群理论物理学家和理论计算机科学家聚焦的研究课题。今天看起来，这也是当代最重要的智力问题之一。计算遇到了物理学的墙，2000 多年哲学传统中的唯心与唯物、经验与理性真要碰到一起了。



兰道尔Rolf Landauer (1927-1999)

一般认为计算是不可逆的，就像西拉德认为测量是不可逆的。数据从一个设备copy到另一个设备，就是测量。兰道尔认为copy不受热力学第二定律约束。兰道尔假设计算机中不同的逻辑状态必须由不同的物理状态表示。一个寄存器的清零操作，就是压缩，即把 $2^n$ 个状态压缩成一个状态。压缩物理状态必然伴随外部环境的熵增。清零必然产生热，是热力学不可逆过程。

兰道尔在IBM的同事和晚辈本内特 (Charles Bennett) 在计算理论，量子计算和通讯等领域都有杰出贡献。他1973年提出逻辑可逆 (logical reversibility) 的概念，并讨论了物理可逆的可能性。逻辑上不可逆的，热力学上肯定不可逆。就像数学上不可能的，物理上肯定也不可能。逻辑上不可逆的运算应用于已知数据，则该运算在热力学上是不可逆的，因为环境熵的增加不会被信息熵的减少所完全补偿。

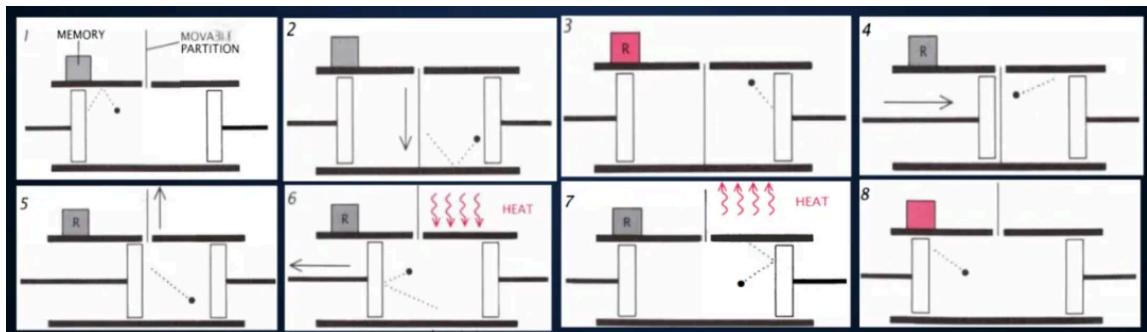


本内特 (Charles Bennett, 1943-) , 1973年论证可逆性与物理可逆性

本内特进一步强化了兰道尔原则，即摆脱旧信息才是要付出能量代价的。本内特细化了西拉德引擎，观测被分成多个步骤，本内特引入了记录装置。本内特引擎分八步。真正消耗能量的是第八步，即擦除记录设备的信息。

- 1) 一个两边有活塞的气缸；一个观察设备，一个记录设备
- 2) 有个隔板可以把气缸中的粒子隔离在一边，
- 3) 记录结果
- 4) 没有粒子一边的活塞被推向隔板，并不需要做功
- 5) 隔板被抽出
- 6) 粒子推动左边活塞，粒子的能量损失由环境的热补偿
- 7) 气缸恢复到初始状态

## 8) 记录设备的信息被擦除



本内特引擎 (Bennett Engine) 细化了西拉德引擎

本内特修正了西拉德之后物理学家之间的共识，即测量消耗能量；沿着兰道尔的思路进一步阐明了是信息擦除（即遗忘）消耗能量。

## CAIXIANSHENG 5、弗雷德金与可逆计算

弗雷德金 (Ed Fredkin, 1934-2023) 是天才式的人物，他只在加州理工学院读过一年书。弗雷德金1960年代末就在波士顿在计算机行业创业，成功上市后，在加勒比海买了蚊子岛 (Mosquito Island) 自娱自乐，后来把这个岛卖给维珍航空的布兰森，但弗雷德金一直没有放弃对学问的追求。1968年明斯基把没有本科学位的弗雷德金招到麻省理工学院做正教授。也许明斯基看中了弗雷德金的管理才华，明斯基本人不喜欢干脏活累活，他后来把麻省理工AI实验室主任的职务交给他刚毕业的学生温斯顿，是另一个例子。弗雷德金1971-1974做过麻省理工Project MAC的主任。MAC是明斯基和麦卡锡在计算机科学和人工智能的形成期创办的靠政府资助的研发机构，后分出AI Lab 和CS Lab，现在又合并成为CSAIL，回归到1960年代的MAC的架构了。

费曼欣赏弗雷德金的才华，并受弗雷德金的影响开始研究计算理论和理论物理，特别是量子计算。费曼认为学习一个新领域最好的办法就是开个课，于是联合同校的生物物理学家霍普菲尔德和集成电路专家米德以及校外的外援明斯基，本内特和弗雷德金等一起开设一门新式课程：“计算之物理学” (The Physics of Computation)。参与这门课的老师和学生后来多成为计算物理学尤其是量子计算领域的开拓者。汉语里“计算物理学”有两个不同的意思，一个是Computational Physics，即用计算手段研究物理；另一个是Physics of Computation，即用物

理学的视角审视计算。无疑，费曼、霍普菲尔德和米德属义后者。费曼此时身体已每况愈下。这门课成了绝响，但幸运的是它被录了音，后来由听过课的两位英国博士后整理成文字出版，书名《费曼计算机科学讲义》（*Feynman Lectures on Computation*, 见Feynman-1996），书中包罗了费曼的讲课内容。可惜费曼没等到书面世就病逝了。几位费曼的和计算机科学沾边的老朋友在费曼死后写了纪念文章，这些文章编辑成书《费曼与计算》（*Feynman and Computation*, 见Hey-1999）。

EDITED BY TONY HEY AND ROBIN W. ALLEN

 CRC Press  
Taylor & Francis Group

RICHARD P. FEYNMAN

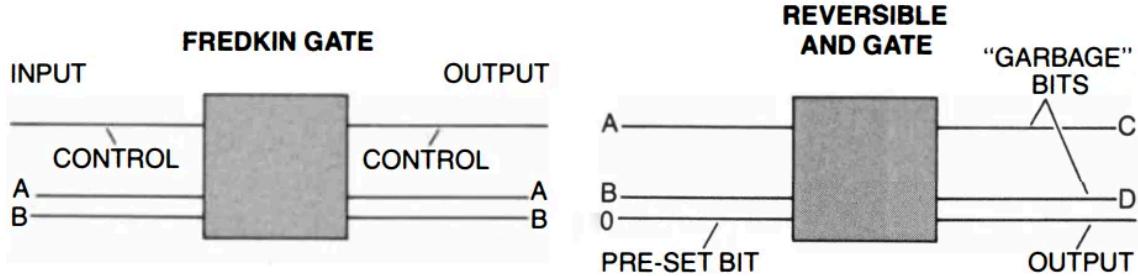
# FEYNMAN LECTURES ON COMPUTATION

费曼计算机科学讲义，一本被忽视的好书



弗雷德金 (Ed Fredkin, 1934-2023)

弗雷德金1982年提出Conservative Gate（守恒门，也被称为弗雷德金门）。弗雷德金门是可逆的，即根据输出，还能还原到输入。弗雷德金门的输入有一条线是“控制线”（control），当control=1时，A和B两个输入信号在输出时交换；而当control=0时，则直接通过。很明显，弗雷德金门是有冗余的（即垃圾位）。如图所示，可以用弗雷德金门实现“与”门（AND）。而“非”门本身就是可逆的。于是所有布尔电路都是可逆的，因为“与”与“非”两个操作可以构成所有布尔代数。所以理论上，任意电路都可以利用弗雷德金门改造成可逆实现，从而不消耗能量。这是令人吃惊的结果。



INPUT			OUTPUT		
CONTROL	A	B	CONTROL	A	B
1	1	1	1	1	1
1	1	0	1	0	1
1	0	1	1	1	0
1	0	0	1	0	0
0	1	1	0	1	1
0	1	0	0	1	0
0	0	1	0	0	1
0	0	0	0	0	0

			"GARBAGE"	
A	B	OUTPUT	C	D
1	1	1	1	0
1	0	0	1	0
0	1	0	0	1
0	0	0	0	0

弗雷德金门 (Fredkin Gate)

## CAIXIANGHENG 6、计算与能量

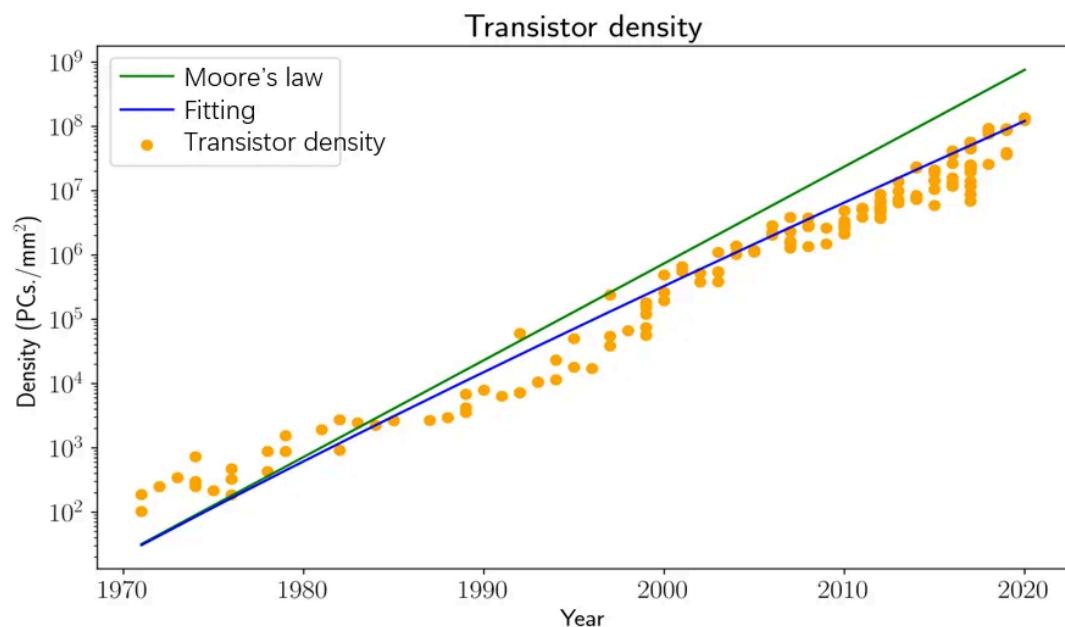
把兰道尔和本内特的不可逆计算需要消耗能量的结论，以及弗雷德金的所有计算都可归约成可逆计算的结论，组合起来看，我们可以得出另一个结论：计算不需要能量。这会让我们盲目乐观。

但正如费曼指出的：兰道尔-本内特-弗雷德金框架中，计算的速度是无穷小的。这就像在理想世界中，可以不用考虑摩擦。而在现实世界中，计算需要速度。费曼设想一个计算步骤需要消耗的最小能量和计算速度相关。这之后，人们的研究开始兵分两路：一条路，物理学家们仍然专注微观的科学考察，研究是否可以在物理世界中存在着符合兰道尔-本内特原则的现象（见 Ciliberto&Lutz-2015, Lutz&Ciliberto-2019, 以及 Lloyd-2000）。另一条路，工程师们开始认真分析宏观的工程实现，探讨改进的手段和未来进展的预测（见 Wolf-2018）。

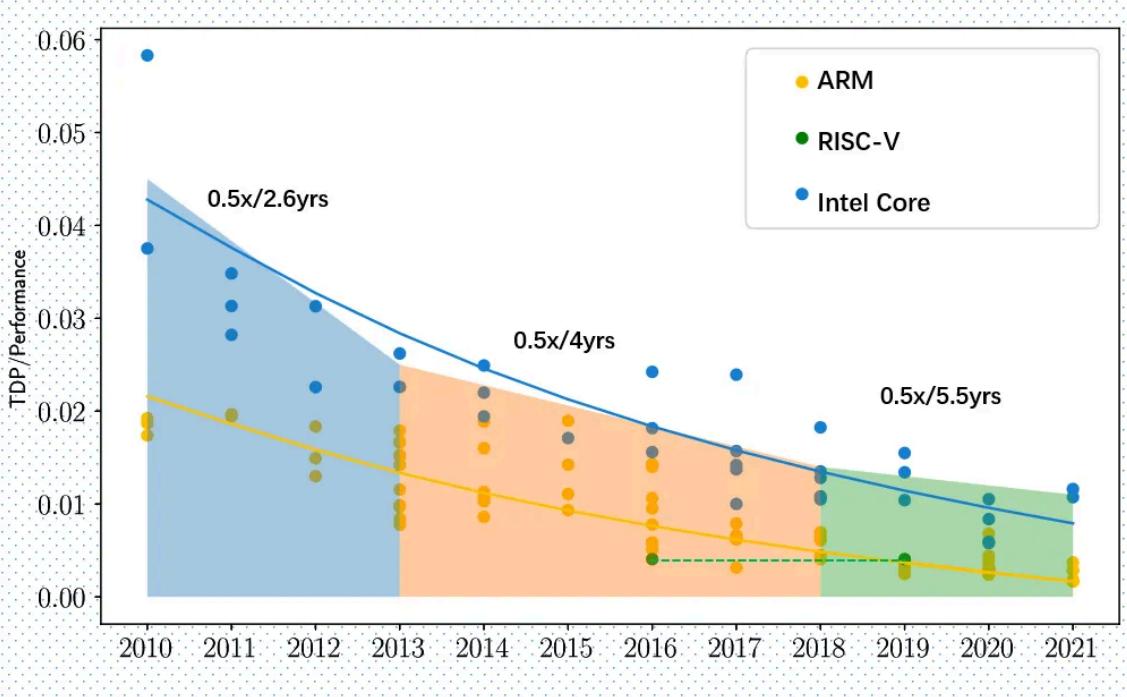
摩尔定律的原始版本（1965年）预测芯片的性能每隔年提升一倍，但到1975年摩尔修正为每隔两年提升一倍。而当下（2024年）最新的观察芯片性能差不多需要4年才能提升一倍，也就是说每年只能提升不到20%。

集成电路的功耗可用经验公式表达为： $P = fCV_{dd}^2$ ， $f$ 是频率， $C$ 是电容， $V$ 是电压，这三个参数又进一步互相关联。邓纳德缩放率 (Dennard Scaling) 断言：每代芯片的频率提升带来了40%的改进。但2006起，邓纳德缩放率失效，热失控 (thermal runaway) 导致不能依靠单纯提升频率来提升性能，于是走向多核并行。但多核冰心并不能线性地提升性能。

对应于摩尔定律，还有一个库米定律 (Koomey)，即在同等功耗下，芯片的性能每隔固定时间提升一倍。最近的观察表明无论是什么架构的CPU (如X86, ARM, 还是RISC-V)，无论是什么规模的计算装置 (手机, PC还是超级计算)，性能 (摩尔定律) 和能耗 (库米定律) 缩放都在放缓。最早的库米定律预测每隔1.57年提升一倍，但现在差不多要3到4年。摩尔定律和库米定律这种逐年放缓的现象，可以统一用“尼克定律”刻画 (见Zhang-2022)。



摩尔定律及其修正



库米 (Koomey) 定律及其修正

“擦除”与一般的信息update操作不一样，擦除消灭了历史。我们借用阿根廷盲诗人博尔赫斯诗句“只有一件事要紧，那就是遗忘。” (There is only one thing. It is oblivion.)，姑且把这个点称为“遗忘点” (Point of Oblivion)，或干脆O点，即碰到兰道尔极限的时刻。Oblivion含多重词义，它不仅捕捉到兰道尔和本内特的“擦除”，还有“遗忘”，“无意识”，“湮灭”和“宽恕”等意，它表达了严谨的科学本意，也表达了我们期望表达的所有文学衍生。

如果，把库米定律和兰道尔极限一起考虑，库米定律，是一条直线下降，总有一天库米定律会交到X轴，即碰到兰道尔极限。2018的研究表明当时离兰道尔极限还有差不多6个数量级（见Chen-Dongarra-Xu-2018），差不多等于原始摩尔定律的20年。而库米最初预测差不多2040就会遇到兰道尔极限。尼克定律放缓了原始摩尔定律和原始库米定律，即便如此，O点也差不多在2080年左右来临。

CAIXIANGCHENG

## 7、结论

计算与能耗的关系可以溯源到热力学第二定律，麦克斯韦和西拉德的研究的结论是测量需要消耗能量。兰道尔和本内特的研究更令人信服地指出：不是测量，而是擦除是不可逆计算的核心，擦除需要消耗能量。弗雷德金的研究证明所有计算都可用可逆的弗雷德金门表示，目前尚无弗雷德金门的可靠的物理实现。费曼指出不消耗

能量的可逆计算的速度必须是无限小的，如果我们需要计算速度，就得消耗能量，但兰道尔极限是最小的计算能量消耗。

压缩就是把唯物的东西变成唯心的东西的过程（见Zhang-2024），这个过程需要能量。信息与物理被庸人如维纳认为分别是不同的东西，没有关联。但现在看信息与物理密不可分，信息是物理的（本内特的说法），物理也是信息的（it from bit，惠勒的说法）。从麦克斯韦和开尔文（用词inanimate，和demon）到西拉德（用词intelligent being）可看出物理学家们的心路历程。某种意义上，计算理论是比理论物理学更加本质的学问。

“智能”或者“意识”困惑我们的一个原因是这些词的词义是个移动的目标，考虑到热力学第二定律，我们效仿西拉德，用“熵减”来定义智能。兰道尔在计算机科学尚未成熟时就意识到这一点。按照直觉，获取新信息是要付出代价的，而兰道尔和本内特的结论是摆脱旧信息才是要付出代价的。于是让我再套用老话：我擦除故我在。  
(I erase, therefore, I am.)

奇点是一个工程的经验的概念。与之不同，“遗忘点”或“O点”是科学的理性的概念。也许奇点的来临并不可怕，而当“O点”来临时，我们会困惑：这是否是文明的停滞点。那时我们（不仅仅是人类，也包括上帝，如果老人家也服从物理定律的话），将造不出更精美的东西。

一个迷茫的问题是：不知道是奇点先来还是O点先来。也许O点先来是更人道的。

## 致谢

鸣谢。本文写作得到诸位师友的启发和指正，特此鸣谢，他们是：白硕、洪涛、马少平、毛德操、李德毅、李明、刘江、戚肖宁、王怀民、徐志伟、张永刚、翟忠武、赵伟等。

### 作者简介：

尼克，乌镇智库理事长。曾获吴文俊人工智能科技进步奖。中文著作包括《人工智能简史》《理解图灵》《UNIX内核剖析》和《哲学评书》等。

参考文献：（上下滑动可浏览）

- [1] Aschenbrenner, Leopold (2024), SITUATIONAL AWARENESS: The Decade Ahead, July, 2024.
- [2] Bennett, C.H. (1982), The thermodynamics of computation-a review. *Int. J. Theor. Phys.* Vol 21, 905.
- [3] Bennett, C.H. (1987), Demons, Engines and the Second Law, *SCIENTIFIC AMERICAN*.
- [4] Bennett, C.H. (2003), Notes on Landauer' s principle, reversible computation, and Maxwell' s Demon, *Studies in History and Philosophy of Modern Physics* 34, 501–510.

本文转载自《赛先生》微信公众号

---

20:05

## // 《物理》50年精选文章

- 中子弹是怎么一回事? | 《物理》50年精选文章
- 晶体缺陷研究的历史回顾 | 《物理》50年精选文章
- 相变和临界现象(I) | 《物理》50年精选文章
- 相变和临界现象(II) | 《物理》50年精选文章
- 相变和临界现象(III) | 《物理》50年精选文章
- 凝聚态物理的回顾与展望 | 《物理》50年精选文章
- 声学与海洋开发 | 《物理》50年精选文章
- 模型在物理学发展中的作用 | 《物理》50年精选文章
- 我对吴有训、叶企孙、萨本栋先生的点滴回忆 | 《物理》50年精选文章
- 国立西南联合大学物理系——抗日战争时期中国物理学界的一支奇葩(I) | 《物理》50年精选文章
- 国立西南联合大学物理系——抗日战争时期中国物理学界的一支奇葩(II) | 《物理》50年精选文章

原子核裂变的发现：历史与教训——纪念原子核裂变现象发现60周年 | 《物理》50年精选

文章

回顾与展望——纪念量子论诞生100周年 | 《物理》50年精选文章

我的研究生涯——黄昆 | 《物理》50年精选文章

中国理论物理学家与生物学家结合的典范——回顾汤佩松和王竹溪先生对植物细胞水分关系研究的历史性贡献(上) | 《物理》50年精选文章

中国理论物理学家与生物学家结合的典范——回顾汤佩松和王竹溪先生对植物细胞水分关系研究的历史性贡献(下) | 《物理》50年精选文章

为了忘却的怀念——回忆晚年的叶企孙 | 《物理》50年精选文章

从分子生物学的历程看学科交叉——纪念金螺旋论文发表50周年 | 《物理》50年精选文章

美丽是可以表述的——描述花卉形态的数理方程 | 《物理》50年精选文章

爱因斯坦：邮票上的画传 | 《物理》50年精选文章

趣谈球类运动的物理 | 《物理》50年精选文章

转瞬九十载 | 《物理》50年精选文章

一本培养了几代物理学家的经典著作 ——评《晶格动力学理论》 | 《物理》50年精选文章

朗道百年 | 《物理》50年精选文章

以天之语，解物之道 | 《物理》50年精选文章

软物质物理——物理学的新学科 | 《物理》50年精选文章

宇宙学这80年 | 《物理》50年精选文章

熵非商——the Myth of Entropy | 《物理》50年精选文章

物理学中的演生现象 | 《物理》50年精选文章

普渡琐记——从2010年诺贝尔化学奖谈起 | 《物理》50年精选文章

我的学习与研究经历 | 《物理》50年精选文章

天气预报——由经验到物理数学理论和超级计算 | 《物理》50年精选文章

纪念Bohr的《伟大的三部曲》发表100周年暨北京大学物理专业建系100周年 | 《物理》

50年精选文章

同步辐射历史及现状 | 《物理》50年精选文章

麦克斯韦方程和规范理论的观念起源 | 《物理》50年精选文章

空间科学——探索与发现之源 | 《物理》50年精选文章

麦克斯韦方程组的建立及其作用 | 《物理》50年精选文章

凝聚态材料中的拓扑相与拓扑相变——2016年诺贝尔物理学奖解读 | 《物理》50年精选文章

我所熟悉的几位中国物理学大师 | 《物理》50年精选文章

量子力学诠释问题 | 《物理》50年精选文章

高温超导研究面临的挑战 | 《物理》50年精选文章

非常规超导体及其物性 | 《物理》50年精选文章

真空不空 | 《物理》50年精选文章

通用量子计算机和容错量子计算——概念、现状和展望 | 《物理》50年精选文章

谈书说人之一：《理论物理学教程》是怎样写成的？ | 《物理》50年精选文章

奋斗 机遇 物理 | 《物理》50年精选文章

关于量子力学的基本原理 | 《物理》50年精选文章

时空奇点和黑洞 ——2020年诺贝尔物理学奖解读 | 《物理》50年精选文章

凝聚态物理学的新篇章——超越朗道范式的拓扑量子物态 | 《物理》50年精选文章

物理学思维的艺术 | 《物理》50年精选文章

对于麦克斯韦方程组，洛伦兹变换的低速极限是伽利略变换吗？ | 《物理》50年精选文章

杨振宁先生的研究品味和风格及其对培育杰出人才的启示 | 《物理》50年精选文章

庞加莱的狭义相对论之一：洛伦兹群的发现 | 《物理》50年精选文章

[阅读原文](#)