

化学分子信息量的计算和可见宇宙信息量的估算

徐光宪

(北京分子科学国家实验室, 北京大学化学与分子工程学院稀土材料化学国家重点实验室, 北京 100871)

摘要 尝试计算化学分子的信息量,估算可见宇宙信息量: (1) 选定 4 种标准态:早期宇宙态 (大爆炸后 10^{-4} s, 10^{12} K, 0.1 GeV)、恒星标准态(10^{7} K)、地球标准态(298 K)、完美晶体态(0 K). (2) 计算了占可见宇宙总质量 75%的 H 和 23%的 He 在 4 种标准态的熵和信息量. (3) 将可见宇宙信息量分为: 微观结构信息量、人工信息量、人类自然信息量、地球生物信息量、宇观结构信息量等 5 个层次. (4) 估算了可见宇宙的总信息量、地球信息量、地球生物自然信息量、人类自然信息量和人工信息量的量级,依次为 10^{80} , 10^{53} , 10^{40} , 10^{35} 和 10^{20} bit.

关键词 分子信息量计算 微观结构信息量 可见宇宙信息量 地球信息量 地球生物自然信息量 人类自然信息量 人工信息量

1 引言

本文是前一篇文章 凹的继续,是从化学工作者的观点来讨论信息科学的基本问题.前文中提出: (1)信息的传递(通讯)必须有 4 个要素:信源、信宿、信道和信的. (2)任何物质粒子,如质子、中子、电子、原子、分子等,都有"互补配偶子"."互补配偶子"之间有信息传递,它们互为信源和信宿,它们之间都有一种或几种相互作用,这就是它们互相联系的信道.信息传递的结果就是互相吸引,组成高一级的粒子. 这就是"信的".例如质子和中子是一对互补配偶子,它们互相吸引组成原子核.原子核和电子也是一对互补配偶子,它们互相吸引组成原子。原子核和电子也是一对互补配偶子,它们互相吸引组成原子。具有自旋为α的未配对电子的原子和自旋为β的未配对电子的原子也是一对互补配偶子,它们互相吸引,以共价键结合组成分子. (3)讨论了信息的定义和分类.

本文将探索信息如何量化的问题, 计算由低级 粒子组成原子、分子时的信息量, 估算可见宇宙的信 息总量的量级.

2 信息如何量化计算

2.1 宇宙信息量是有限还是无限?如果有限,能 否估算?

一般认为宇宙信息量是无限的,至少是难以计算的.为了探讨这个问题,首先将信息量分为不同的层次.

2.2 信息量的不同层次

2.2.1 微观结构信息量

实物(substances)是指具有静质量的物质(matter). 实物都是由原子、分子组成的,而原子、分子是由质 子、中子、电子组成的.将后者组成原子、分子的结 构信息量称为微观结构信息量.

2.2.2 宇观结构信息量

由大量原子组成原子星云,慢慢诞生星系、恒星、行星(包括地球)等天体而增加的信息量,称为字观信息量.可见宇宙虽然有各种各样的天体和背景辐射等,但它们都是由早期宇宙的质子、中子、电子

和光子4种粒子进化发展而来的. 所以宇宙是高度统一的.

2.2.3 人类自然信息量(human natural information quantity)

2005 年全世界有 65 亿人口,每个人大约有 60 万亿个细胞,每个细胞核内有 23 对染色体,组成 DNA 双螺旋链,如果将其拉直,有 1.5 m 长,这就是人类的基因组,含有 30 亿对碱基. 碱基 ATGC 是 4 个选 1,信息量是 2 bit. 所以一个人的 DNA 基因信息量=6×10¹³×30×10⁸×2 = 3.6×10²³ bit. 人还有 RNA、蛋白质、糖类等的信息量,假定总数为 10^{25} bit 量级,乘以世界人口.得到人类自然信息量=65× 10^{8} × 10^{25} = 10^{35} bit 量级.

2.2.4 地球生物信息量

虽然全世界所有动物、植物,尤其是微生物个体的数量要比人类的数量大许多个量级,但每个微生物所含的细胞数少,最少的只有一个. 所以全世界所有生物的细胞总数的量级与人类细胞总数的量级相差并不多. 生物细胞和人类细胞也很相似,都含有DNA 基因、RNA、蛋白质等,所以生物世界也是高度统一的. 假定地球上的生物信息量比人类信息量多 5 个量级,那么全世界的生物信息总量大约为 10⁴⁰ bit 量级.

2.2.5 人工信息量(artificial information quantity)

人工信息分为系统人工信息和一般人工信息两类¹¹. 前者包括科学、文学、艺术、哲学、宗教等全部知识文化体系,后者是指日常人们交谈的一般人工信息.

2003 年报载美国加州大学伯克利分校信息管理及系统学院莱曼教授领导的小组统计了全球在 2002年,各种介质记录的人工信息生产量达到 5 万亿兆节,即 5×10¹⁸ bit,相当于 50 万座美国国会图书馆存储的信息量,是 1999 年全球生产的信息量的一倍.

按照 3 年信息量增加一倍的数据估算,信息量的年增长速率约为 30%,由此计算 2005 年增加的信息量为 5×[1.3]³×10¹⁸ bit =10¹⁹ bit. 假定过去的年增长速率也是 30%,则由等比级数加和规律,得到全球历年生产的人工信息总量为 10²⁰ bit 量级,这就是全球 65亿人在 2005 年底拥有的全部人工信息量.平均每人拥有 10¹⁰ bit =10 Gb.

由此可见, 全球人工信息总量 10²⁰ bit, 要比人类

自身的自然信息量 10³⁵ bit 小 15 个量级. 但各种信息 千差万别,有质的高低. 关于信息的质的高低,暂时 不列在本文探索的范围之内.

2.3 如何计算微观结构信息量,例如 1 mol 氢分子(H_2)在地球标准态的信息量?

计算分子的信息量要考虑它的平移运动、转动、振动、电子运动、核运动等信息,这些信息如何表达,如何量化呢?幸亏物理学早已建立了熵函数,它的数值可由统计力学来计算.计算分子的熵函数 S 是从分子的平移运动、转动、振动、电子运动、核自旋、电子自旋等各种运动的配分函数出发的.这些运动参量正是计算分子的信息量所需要的.

Stonier^[2]提出 1 mol某一纯物质的熵在不同状态的差值 ΔS ,与信息量的差值 ΔI 之间的关系:

$$-\Delta S = -1 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1} = \Delta I = 1.045 \times 10^{23} \text{ bit /mol}$$

= 0.1735 bit /粒子. (1)

由式(1)可以计算相对信息量 ΔI , 但要计算绝对信息量 I, 还必须假定标准状态. 详见下节讨论.

3 1 mol 的氢分子(H₂)在各种聚集态的信息量的计算

现代天文学从大量恒星的光谱分析得知,它们的化学元素丰度是高度一致的,其质量百分比是:75%的 H 元素和 23%的 He 及 2%的所有其他化学元素.在古老的第一代恒星中含75% H,25% He 和极少量 Li/Be,没有其他更重的元素.在第一代恒星作为超新星爆发后,才能通过核反应产生其他元素.所以在第二代恒星(例如太阳)才有 2%的其他元素.因为 H 是宇宙最主要的元素,为了估算宇宙的信息量,我们先计算 1 mol 的氢分子和 1 mol 的 He 原子在各种状态时的熵,把它们换算成信息量.

 $1 \text{ mol} \text{ bH}_2$ 在非常接近绝对零度(0 K)时,氢分子就完全有序地排列成为完美晶体. 按照热力学第三定律,它的熵为 $S_0 = 0$,有序程度最高,信息量 I_0 最大. 氢分子 H_2 在地球标准态的熵值可从常用的物理和化学数据手册 [31中查到.

从标准态 298 K 加热到 10^4 K, H_2 分子离解为 2 个 H 原子, 加热到 10^5 K, 电离成 2 个质子和 2 个电子, 再加热到恒星标准态 10^7 K, 最后加热到早期宇宙态 10^{12} K. 这些过程中熵的变化都可计算出来, 在计算中用到的 H_2 分子的离解能、H 原子的电离能、

及其 H_2 分子、H原子的热容 C_p 可从手册 ^[3]中查到. 我们近似假定 C_p 不随温度而变化,又假定质子加电子的 C_p 等于H原子的 C_p .

利用式(1)表示的熵的减少值与信息量的增加值 之间的换算关系,可得相应的信息量差值如表 1 所示.

表 1 中状态 7 已经是宇宙最初的 10^{-4} s, 那时宇宙处于超高温、超高密的质子、中子、电子、光子的无序混沌态,此时的信息量 I_7 非常小. 但严格说来,质子、中子也是有结构的,它们由夸克组成,因此 I_7 也有一定的信息量. 因为宇宙年龄在 10^{-4} s 以前的情况现在还不很清楚,所以假定 $I_7 = 0$ (将来有了确切的 I_7 数值,可以再作更正). 代入表 1 的最后一行,得

$$I_0 = 186 - I_7 = 186 \text{ bit } / \text{H}_2 \, \text{分子}.$$
 (2)

式(2)表示将宇宙最早期的 2 对配偶子[p^++e^-]组织成 2 个 H 原子, 再合成为 H_2 分子, 继续降温, 直到 0 K 时的完美 H_2 分子晶体, 获得的最大信息量为 186 bit.

将式(2)表示的 I_0 值代入表I的最后一列,并以信息量增加的次序排列,得到表I2的结果.

由表 2 可见, 从早期宇宙态(指宇宙大爆炸后 10^{-4} s 时的状态)的[p⁺e⁻, 10^{12} K]到恒星标准态[p⁺e⁻, 10^{7} K]的信息量为

$$[1/2] I_6 = 42 \text{ bit } / [p^+e^-].$$
 (3)

从早期宇宙态[p^+e^- , 10^{12} K]到地球标准态[1/2][H_2 , 气, 298 K]的信息量为

$$[1/2] I_1 = 82 \text{ bit } / [p^+e^-].$$
 (4)

4 1 mol 的 He 原子在各种聚集态的信息量的计算

1 mol He 原子在各种聚集态的熵和信息量的计算如表 3 所示.

同样假定 $I_7 = 0$, 代入表 3 的最后一行,

$$I_0 = 167 - I_7 = 167$$
 bit /He 原子. (5)

式(5)表示将早期宇宙态的[2p+2e+2n, 1012 K]组织

表 1	1 mol 的氢分子	H ₂ 在各种聚集态的熵和信息量差值的计算
-----	------------	----------------------------------

序号	状态	$S/J \cdot K^{-1} \cdot mol^{-1}$	I/ bit ⋅粒子 ⁻¹
0	完美晶体态 H ₂ , 固,0 K	$S_0 = 0$	<i>I</i> ₀ = 最大
1	地球标准态 H ₂ , 气, 298 K	$S_1 = 130.6$	$I_1 = I_0 - 22.7$
2	H_2 , \leq , 10^4 K	$S_2 = S_1 + 101 = 231.6$	$I_2 = I_0 - 40.2$
3	2H, 气, 10 ⁴ K	$S_3 = S_2 + 45 = 275.6$	$I_3 = I_0 - 48.0$
4	$2H, 10^5 K$	$S_4 = S_3 + 96 = 372$	$I_4 = I_0 - 64.5$
5	$2p^+ + 2e^-, 10^5 \text{ K}$	$S_5 = S_4 + 26 = 398$	$I_5 = I_0 - 69.0$
6	2p ⁺ +2e ⁻ , 10 ⁷ K, 恒星标准态	$S_6 = S_5 + 192 = 590$	$I_6 = I_0 - 102$
7	2p ⁺ +2e ⁻ , 10 ¹² K, 早期宇宙态	$S_7 = S_6 + 480 = 1070$	$I_7 = I_0 - 186$

表 2 1 mol H₂ 在不同标准态的熵和信息量

			-			
-	序号	状态	状态说明	$S/J \cdot K^{-1} \cdot mol^{-1}$	I/ bit⋅粒子-1	
-	7	早期宇宙态	$2p^{+}2e^{-}$, 10^{12} K	$S_7 = 1070$	$I_7 = 0$	
	6	恒星标准态	$2p^{+}2e^{-}$, 10^{7} K	$S_6 = 590$	$I_6 = 84/[2p^+2e^-]$	
	1	地球标准态	H ₂ , ≒, 298 K	$S_1 = 131$	$I_1 = 163/\text{H}_2 = 7$	
	0	完美晶体态	H ₂ , 固, 0 K	$S_0 = 0$	$I_0 = 186/\mathrm{H}_2$ 🛅	

表 3 1 mol 的 He 在各种聚集态的熵和信息量的计算

序号	状态	$S/J \cdot K^{-1} \cdot mol^{-1}$	I/ bit ⋅粒子 ⁻¹
0	He, 固,0 K	$S_0 = 0$	<i>I=I</i> ₀ =最大
1	He, 气, 298 K, 地球标准态	$S_1 = 126$	$I_1 = I_0 - 22$
2	He, 气, 8×10 ⁵ K	$S_2 = S_1 + 164 = 290$	$I_2 = I_1 - 28 = I_0 - 50$
3	$He^{2+}+2e^-, \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \$	$S_3 = S_2 + 12 = 300$	$I_3 = I_2 - 2 = I_0 - 52$
4	He ²⁺ +2e ⁻ , 10 ⁷ K, 恒星标准态	$S_4 = S_3 + 53 = 353$	$I_4 = I_3 - 9 = I_0 - 61$
5	He ²⁺ +2e ⁻ , 10 ¹⁰ K, 轻核合成	$S_5 = S_4 + 144 = 497$	$I_5 = I_4 - 25 = I_0 - 86$
6	$2p^+ + 2e^- + 2n$, 10^{10} K	$S_6 = S_5 + 273 = 770$	$I_6 = I_5 - 47 = I_0 - 133$
7	$2p^+ + 2e^- + 2n$, 10^{12} K	$S_7 = S_6 + 192 = 962$	$I_7 = I_6 - 4 = I_0 - 167$

成一个 He^{2+} 原子核,再与2个电子组成为He原子,继续降温,直到0 K时的完美He 晶体,获得的最大信息量为 167 bit. 将式(5)表示的 I_0 值代入表 3 的最后一列,并以信息量增加的次序排列,得到表 4 的结果.

由表 4 可见, 从早期宇宙态, 把 $2p^+2e^-2n$ 组成恒星标准态的 $He^{2+}2e^-$ 的信息量为

 I_6 = 106 bit /[2p⁺2e⁻2n] = 53 bit /[p⁺e⁻n]. (6) 将 2p⁺2e⁻2n 组成地球标准态的 He 原子的信息量为

 $I_1 = 145 \text{ bit } / [2p^+2e^-2n] = 72 \text{ bit } / [p^+e^-n].$ (7)

5 宇宙信息量的估算

5.1 宇宙的总质量 [4~9]

现代天文学准确测定太阳的质量为1.982×10³³ g. 银河系的总质量为太阳质量的 1000 亿倍(2×10⁴⁴ g). 银河系的形状像一个铁饼,直径8万光年,厚约6000光年. 星系团含有上千个像银河系那样的星系,尺度在1000万光年的量级. 超星系团的尺度在一亿光年量级.

可见宇宙的所有超星系团、星系团和类星体的总质量大约为银河系质量的 100 亿倍,即 $10^{10} \times 2 \times 10^{44} = 2 \times 10^{54}$ g.

宇宙广义物质的总质量为 5×10⁵⁵ g,由"物质"和"辐射能量"两大部分组成,其中电磁辐射和重子物质是可以观察到的,因此合称"可见宇宙",其质量为2×10⁵⁴ g,只占宇宙总质量的 4%.其余 26%为暗物质,70%为暗能量,不能用谱学方法观察,合称不可见宇宙或鬼宇宙(ghost universe).可见宇宙与鬼宇宙是互

相重叠而存在的. 暗物质是指有静质量,有引力作用,但没有强、弱和电磁相互作用,因而不能被各种波段的天文望远镜直接观察到的物质. 但它们具有引力质量,能够影响可观察到的恒星的运行轨道,或由于"引力透镜"作用使光线弯曲,因而证明其存在. 暗能量没有静质量,没有万有引力,而有万有斥力,能产生负压力,使超星系团互相远离,从而是导致现在宇宙正在加速膨胀的原因 [10]. 暗物质和暗能量究竟是什么? 现在不知道,是 21 世纪物理学要解决的世界难题之一. 李政道 [10.11]在北京 2005 世界物理年上报告,提到可见宇宙质量占总质量的不到 5%,暗物质占 25%以上,暗能量占 70%,也提到暗能量是导致宇宙膨胀的原因.

可见宇宙由4种粒子组成,每种粒子的数量和质量如表5所示.

表 5 中粒子数和粒子总质量的量级是可置信的. 误差主要来源于无量纲哈勃常数 h, 20世纪 90年代估计 h 在 0.5 与 0.8 之间,到了 21世纪已缩小到 0.66 与 0.71之间.另一误差来源是宇宙质量密度与理论临界密度之比 Ω (0),误差已缩小到 0.5%.表中保留小字号数字的原因,是因为质子和中子之比 7 是可靠的.如不保留小 5 号数字,这一比值将为 10,反而不可靠了.

4种粒子中 3种是物质粒子, 1种是辐射粒子. 3种物质粒子组成 9×10^{77} 个 $[p^+e^-]$ 对和 1.5×10^{77} 个 $[p^+ne^-]$ 三连体. 前者组成 H 原子, 后者组成 He 和其他化学元素. 由此计算 H 的丰度为 75%, He 和其他元素的丰度为 25%, 与天文观察数值符合.

M				
序号	状态	状态说明	$S/J \cdot K^{-1} \cdot mol^{-1}$	I/ bit⋅粒子 ⁻¹
7	早期宇宙态	2p ⁺ 2e ⁻ 2n, 10 ¹² K	$S_7 = 603$	$I_7 = 0$
4	恒星标准态	$He^{2+}2e^{-}$, 10^{7} K	$S_4 = 353$	$I_4 = 106/[2p^+2e^-2n]$
1	地球标准态	He, 气, 298 K	$S_1 = 126$	$I_1 = 145/\text{He} = 1$
0	完美晶体态	He, 固, 0 K	$S_0 = 0$	$I_0 = 167/\text{He}$ 固

表 4 1 mol He 在不同标准态的熵和信息量

表 5 可见宇宙由 4 种粒子组成

粒子	粒子数	粒子质量/g	粒子总质量/g	百分比/%
质子	1.05×10 ⁷⁸	1.67262×10 ⁻²⁴	1.75×10 ⁵⁴	87.4
中子	1.5×10 ⁷⁷	1.67492×10^{-24}	2.5×10 ⁵³	12.5
电子	1.05×10^{78}	0.91094×10^{-27}	9×10^{50}	0.05
光子	2×10 ⁸⁷	1×10^{-36}	2×10 ⁵¹	0.1
总计	2×10 ⁸⁷		2×10 ⁵⁴	100

5.2 可见宇宙的全部微观信息量等于 10⁸⁰ bit 量级

可见宇宙含有 75%的 H, 23%的 He, 2%的其他化等, 原子核的平均结合能也与 He 的平均结合能相当, 所以可近似地当作 He 计算. 可见宇宙共有 1.05×10⁷⁸ 质子-电子对, 和 1.5×10⁷⁷ 个中子, 后者与 1.5×10⁷⁷ 个质子-电子对组成 He 和其他原子核. 余下 9×10⁷⁷ 质子-电子对组成 H 原子. 由式(3)和(6), 可以计算出这一合成过程的信息量:

$$1.5 \times 10^{77} \times 53 + 9 \times 10^{77} \times 42 = 8 \times 10^{78} + 3.8 \times 10^{79}$$
$$= 4.6 \times 10^{79} \text{ bit}, \tag{8}$$

这就是可见宇宙的全部微观结构信息量,它等于 10⁸⁰ bit 量级.

可见宇宙还有 2.728 K 的背景辐射, 但它各向同性, 非常均匀地分布在整个宇宙空间, 没有像质子、中子、电子形成各种分层次的复杂结构. 所以背景辐射的信息量要比重子宇宙的信息量小许多个数量级,可以忽略不计.

5.3 微观结构信息量最大原理

在 2.2 节所述 5 个层次的信息量中,由于微观粒子的数目要比宏观物体和字观天体的数目大许多许多个量级,因此微观结构信息量也比宏观、字观和人工信息量大许多个量级.所以宇宙的总信息量等于它的微观结构信息量.我们将这个结论叫做微观结构信息量最大原理.

表 5 中列出可见宇宙共有 10 的 78 次方量级的质子和电子,77 次方量级的中子. 虽然一对[p+e]配偶子从早期宇宙到恒星态的信息量只有 42 bit. 但宇宙的总的微观信息量就达到 10 的 80 次方量级的比特.

另一方面,可见宇宙有 10²¹ 量级的恒星,即使每个恒星形成的信息量需要 10²⁰ 量级的比特,总的宇观结构信息量也只有 10⁴¹ bit,要比微观信息量小 39 个量级.因此在估算宇宙的总信息量时,只需计算其微观结构信息量,宇观结构信息量可以忽略不计.

又如下面将要估算地球的微观结构信息量是 10⁵³ bit 的量级, 生物信息量是 10⁴⁰ bit 量级, 人工信息量是 10²⁰ bit 量级. 因此微观结构信息量可以代表总信息量. 这是我们有可能估算可见宇宙总信息量的主要根据.

5.4 地球的全部信息量等于 10⁵³ bit 量级

地球的质量为

$$M(地球) = 6 \times 10^{27} \text{ g.}$$
 (9)

地球上的化学元素丰度和恒星不同,恒星以 H和 He 为主,地球因引力太小,大气中最轻的元素 H_2 和 He 都飞跑了. 所以包括大气、海洋的地壳中的元素丰度为 He 10^{-7} %,H 0.15%(主要在海水中),O 47.2%(主要在岩石、土壤、海水和大气中),Fe 5.1%,Ca 3.6%等. 地核则以 Fe 为主. 所以地球元素的质子和中子数是近似相等的,即地球含有(1/2)×6× 10^{27} ×6.023× 10^{23} = 2× 10^{51} 个中子、2× 10^{51} 个质子和 2× 10^{51} 个电子.

在 He 原子的信息量的计算中,采用 He 原子核的结合能为 7.05 MeV/核子. 这一数值和其他化学元素的原子核的平均结合能十分接近,例如 12 C 核的平均结合能为 7.64 MeV/核子, 14 N 为 7.47, O 为 7.97, U 为 7.58 等. 所以作为近似的数量级估算,可以采用由式 (8)表示的,从[p^+e^-n]组成 He 原子的地球标准态的信息量为 72 bit /[p^+e^-n].

$$I(地球)=2\times10^{51}\times72=1.4\times10^{53}$$
 bit. (10)

以上计算的是地球的微观信息量 10^{53} bit, 它等于地球总信息量,同样忽略了由原子、分子形成地球的宏观地质信息量,也忽略了全球生物信息量 10^{40} bit 和人工信息量 10^{20} bit.

参 考 文 献

- 1 徐光宪. 关于化学信息学的探索与思考. 中国科学 B 辑: 化学, 2007, 37(1): 6—11
- 2 Stonier T. Information and the Internal Structure of the Universe. London: Springer Verlag, 1990
- 3 Lide D R. Handbook of Chemistry and Physics. New York: CRC Press, 2002
- 4 Harrison E R. Cosmology. London: Cambridge University Press,
- 5 俞允强. 物理宇宙学讲义. 北京: 北京大学出版社, 2002
- 6 Harland D M. The Big Bang——A View from the 21st Century. London: Springer, 2003
- 7 Rainer M, Schmidt H J. ed. Current Topics in Mathematic Cosmology. Proceeding of the International Seminar, Postdam, Ger, 1998
- 8 徐光宪. 宇宙进化的八个层次结构. 科技导报, 2002, 9:8-13
- 9 Davies P. The Cosmic Blue Print. London: Heinemann, 1987
- 10 李政道. 在我的祖国纪念爱因斯坦. 科技导报, 2005, 23(5): 4-5
- 11 Lee T D. The strongly interacting quark-gluon plasma and future physics. Nucl Phys A, 2005, 750(1): 1—8[DOI]