

关于理论物理与理论生命科学的一些思考

Thoughts on Theoretic Physics and Theoretic Life Science

郝柏林

(中国科学院理论物理研究所, 院士 北京 100080)

自从1985年11月出席中国科学院生物学部召开的生物科学发展战略座谈会以来,参加了多次有关生命科学的讨论会。在这期间我一直还在研究“死物”的运动,同时在学习一些生物的基本知识。现在把学习中的一些零星思考,写在下面。

普遍与特殊

物理科学和生命科学的相互作用由来已久。电学和电生理学是同时诞生的(L. Galvani, 1791)。生物学为物理学启示了能量守恒定律,物理学为生命科学提供了从显微镜、X光到示踪原子、中子衍射、核磁共振谱仪、同步回旋辐射、扫描隧道显微镜等种种研究手段。然而,只有到了把生命物质和生命过程作为真正的物理对象加以研究的时候,理论物理的思维方法和数学工具才会在生命科学的研究中发挥更大作用。现在,生命科学所积累的大量事实和数据,物理科学在研究复杂系统方面的长足进展,已经把物理学和生命科学密切结合的任务提上日程。对于理论物理工作者而言,恰当地认识生物学和物理学思维方法的差异,会更有助于寻求共同语言。

物理学研究物质结构和运动的普遍规律。这些规律适用于人类观测所及的宇宙的任何部分,适用于它的过去与未来。物理规律可以由少量关键实验证实,物理规律没有例外。观测与理论的矛盾,通常表明理论还不够普遍,预示着理论的突破,而新的理论总把旧理论的正确部分作为特例包含在内。物理理论从数学框架到实验检验两方面受到严格制约,很难出现学派林立、假说纷纭的局面。物理规律本身是否也在演化,在更大程度上是个哲学问题,至少其变化的时间尺度远大于现代人类的观测时间,因而在当前的认识水平上,“永恒”的物理定律被认为是放之宇宙而皆准的。

生命科学研究的对象极为特殊。迄今还没有发现任何其它天体上有类似地球上的生命现象。就宇

宙尺度而言,我们目前所研究的对象,只存在于微乎其微的小小环球的“生物圈”的有限空间,存在于狭窄的温度范围,存在于出现生命后的有限的历史时期。为特殊现象建立理论是更困难的,但其内容却可能极其丰富。生命科学的理论和规律常有例外,它们既不能还原成简单的物理法则,也不会违反基本的物理规律。

遍历与进化

物理学处理“多体”、“多自由度”,甚至“无穷多自由度”的问题时,标准方法是统计平均。把对时间,对历史,对发展过程的平均,换成对一切可能的当前状态的平均。这是基于对经典力学的遍历假设,即认为一个力学系统或迟或早会经历能量上允许的一切状态。遍历假设早在物理学范围之内就出了问题:原来认为应当遍历的许多体系,例如偶合非简谐振子,已被严格证明并非如此。不过,确实还有许多遍历性成立的情况,例如耗散系统的混沌吸引子上的运动。

与遍历类似但更弱一些的要求,是时间变化的定态假设,即认为一个发展过程总处在同一机制、同一环境之中。其实,定态过程是无始无终、“始终如一”的。遍历和定态都是与进化或演化对立的概念。

虽然我们现在看到的宇宙,也处于演化之中,而且从对种种遥远距离的观测中不断带来不同演化时期的信息,但那时间尺度比之人类历史是极其漫长的。因此,物理学家们往往从现存的结构和相互作用来解释现象,更习惯于在“遍历”和“定态”的假设下思考问题。

进化是生命发展的基本特征,生命现象只有在进化背景中才能正确理解。进化乃是一种“遍历破缺”,即破坏了遍历性。这“破缺”在生物学中叫做“选择”。

另一方面,相对于单个生物体甚至物种的寿

命,生命进化过程也是极其漫长的。大自然有足够的时间来“择优”。彭桓武教授曾经引用老子的话“天地不仁,以万物为刍狗”,来概括生物进化过程的漫长和物竞天择的“无情”。

因果论和目的论

在生命科学文献中,拟人的、“目的论”的语言颇为常见,而因果论则是物理学的一条基本原则。从表面看,这两者似乎矛盾。细想起来,当物理学处理比较复杂的系统时,已经有转向“目的论”说法的迹象。举两个例子。

描述布朗运动的朗之万方程中有磨擦系数和随机力强度两个参量,两者不能任意给定,而必须满足涨落耗散定理。这是为了保证时间趋向无穷时布朗粒子要和周围环境达到热平衡,使得能量均分定理成立。实质上,这是从终值而非初值条件选取随机微分方程的解。统计物理学中许多关于趋近平衡条件的讨论都有这么一点“目的论”的味道。其实,趋近平衡只是最简单的一种非平衡过程。处理远离平衡现象时,可能更要注意“目的”,才能正确描述。

自旋玻璃的基态和蛋白质折叠问题,都涉及在高维的能量空间中寻求最小值。这是多参数的优化问题,其提法本身就是要求“目标函数”达到极值。一般说来,优化问题常与一定的识别问题相联系,但往往更困难。穷举一切可能的状态,有点“遍历”的精神,但不能解决规模较大的实际问题。近年发展起来的某些有效的近似方法,例如“模拟退火”,实质是选取合理有利的演化途径,尽量接近某种事先设定的目标。演化的每一步,应遵从因果律,但长期演化所达到的状态,看来象是有目的安排的结果。

目标函数的极值也有绝对、相对之差。没有理由认为复杂系统的演化一定沿最优方向,发展到最优状态。有许多足够好的演化方式,可以达到相当“优”的状态,后者在实践上与最优状态相差不远。不同演化道路的初等实例是优选法:按 0.618 作“黄金分割”,比使用 0.5 的“对分法”,在有限精度下快不了几步。复杂性不是随机性,但复杂性自然地包含着随机性,在“同样”复杂的条件下,一个系统如果有两次演化机会,不一定走同一条演化道路。

特殊中的普遍

生命现象涉及众多的层次,每个层次上都有空间结构和时间进化,层次之间并没有简单的“自相似”。理论物理工作者比较宜于从具有普遍性、可以

“言之及物”的那些层次入手,争取逐步由“物”及“生”。

十多年来,我国已有不少理论物理工作者在各种层次从事研究,例如:生物大分子的结构;它们之间、它们和中小分子以及环境(例如水)的相互作用;核酸碱基序列、蛋白质氨基酸序列的信息结构分析;蛋白质的变性、复性、折叠和高级结构;生物膜的连续介质模型,特别是液晶模型、膜的静态形状和动态行为,膜上的和跨膜的物理过程。

在组织、器官、个体乃至群体、种群、生态系统的诸层次上,“物”理逐渐成为并不直接考虑的因素。许多唯象的模型研究,更近乎数学生物学。但是,细胞内和细胞间的通讯、正常和病态的心律、脑电波分析、神经脉冲传播、免疫系统等方面的研究,不少可以借鉴非线性科学的方程和语言,同熟知的数学物理模型相联系,都还是理论物理工作者的用武之地。这里要注意普适性的理论,如相变、混沌、分形等,不能代替对特殊问题的具体研究。用新颖的名词术语包装,成不了科学成果。

有一些永恒的问题,如生命的起源和本质,意识的产生和发展,人类必将世代代思索和探求。对生命物质的结构、运动和相互作用已经作过系统研究的人,或许能对这类根本性问题悟出些深刻的道理。但这些问题看来不宜作为研究的起点。

在若干年内,我国的计算机仍将同美、日等国相当差距。我们不能也不应在拼机器能力的“从头计算”或大规模的分子动力学模拟方面与这些国家抗衡。而应当注重“巧算”,发展粗粒化和唯象的方法。用化学反应动力学语言,平衡和非平衡热力学概念和工具进行的研究,也属于粗粒化和唯象的理论,也应提倡。不过,下面只谈谈这几年用粗粒化方法研究非线性动力学^[1]所引发的一些思考。

粗粒化和符号描述

自然科学不能同时在一切实层和细节上把握客观现象的规律性,而必须聚焦到一定的观测精度和描述水平。当我们写下质子和中子的符号 p 和 n 时,通常不必关心它们由哪些夸克组成,而只须知道其电荷和质量。在化学中使用原子符号写出分子结构式时,H,C,O,N 等记号主要代表一定的原子量、离子半径、亲和力、化学价,而无须涉及原子核的构造。对于由几十个原子组成的嘌呤和嘧啶,化学结构式并非处处必需。人们引入 A,C,G,T 等等符号,更注意它们形成双螺旋时提供两个还是三个氢键。20 种符号所代表的不同氨基酸,更重要的不是“大同”的化学结构,而在于“小异”所导致的亲水、疏水、极性等种种性质的差别。

粗粒化不可避免地导致符号描述。在核酸和蛋

白质的情形下,粗粒化都导致了一维的符号序列。其实,对于有限长的符号序列,一维和高维并没有原则性的差异。高维序列可以归结为具有远邻关系的一维序列。这当然比只有最近邻关系的一维序列复杂些。但对于多数一维序列,也必须研究其中的长程关联。人们用符号序列来表示研究成果,更希望从符号序列中解读出自然界的规律。

面对大量符号序列,首先会想到统计分析,例如计算和比较各种符号串的出现频度与关联。事实上,人们用统计方法对核酸和蛋白质序列作了大量研究。罗辽复教授1995年4月在国家自然科学基金委召开的一次座谈会上曾说,多年来用统计分析核酸编码序列的经验表明,这些序列的统计性质往往偏离随机序列不远。这是否表示,传统的统计分析不足以把生命现象远离平衡的“远”充分放大,以供更细致的分类和刻划?这也提出了对符号序列进行确定论分析的必要性。其实,对于有限长的序列,确定论分析和统计分析都是近似的,但恰当的粗粒化可能导致某些严格的结论。

形式语言和语法复杂性

从一定的字母集中选取符号,形成一维的符号序列。这些符号序列的集合,构成某种形式语言。粗粒化和符号描述自然地导致形式语言。对形式语言和自然语言都可以进行统计分析。然而,形式语言也把我们带到因计算机科学发展而成熟起来的另一个领域,开辟了对符号序列进行确定论分析的可能性。

原来在50年代中期,乔姆斯基(N. Chomsky)^[2]就发展了一种形式语言的分类体系。他把语言分成四大类,每类语言可由一类计算机接受(或者说识别、执行):由不需存储器的有限自动机接受的正规语言,由带堆栈的下推自动机接受的上下文无关语

言,由存储器比例于输入量的计算机接受的上下文有关语言,由具备无穷存储的图灵机接受的递归可数语言。这同时给出复杂性的一种参考尺度,通常称为乔姆斯基阶梯。

乔姆斯基的阶梯,并不是对形式语言的唯一的分类体系。在符号动力学所导致的符号序列的分类上,已经感到乔氏阶梯的不足。例如,谢惠民等人的研究表明,一维单峰映射中的符号序列,从简单的正规语言一跃而至上下文有关语言,跨越了上下文无关语言的整个台阶。在图像表示领域,也发现许多直观上极其简单的图像,如直线段或圆,竟然导致复杂的上下文有关语言,使得研究上下文无关语言所得到的丰富成果无用武之地。

1968年荷兰生物学家林登梅耶(A. Lindenmayer)在研究多细胞植物的发育过程时,建议了一种符号描述方式。它后来发展成另一套形式语言的分类体系,现在称为L系统。乔姆斯基用串行的生成规则,L系统用并行生成,它所导致的分类比乔姆斯基更细致,但看来并未覆盖一切可能的形式语言。乔姆斯基分类中的某些复杂序列,在林登梅耶系统中处于简单层次,反之亦有实例。符号动力学中郑伟谋建议的广义生成法则,则自然地对应L系统的框架。L系统的一大优点,是它与分形几何有自然的联系。用基于L系统的图形生成程序,模拟草本植物的宏观形态发育过程,已经达到可以乱真的程度。对于核酸碱基序列或蛋白质氨基酸序列的分析,L系统或其变种,可能发挥作用。

参考文献

- [1]郑伟谋、郝柏林:《实用符号动力学》(非线性科学丛书)上海科技教育出版社,1994
- [2]本节提到的形式语言的乔姆斯基分类和L系统,可以参阅谢惠民著《复杂性与动力系统》(非线性科学丛书)上海科技教育出版社,1994

(责任编辑 蔡德诚)

(上接第54页)

前提下,形成从种子、化肥、农药等农用物资的供应,到产品收购、贮藏、加工、销售,到农业技术的指导、推广的系列化服务,使整个生产过程产业化、社会化。从根本上将农业的含义由农、林、牧、渔的生产转化为多行业、多产业构成的经济体系,传统农业只是这个体系中的一个生产步骤。也只有在这样的农业体系中,才可能实现大部分农业剩余劳力的转化,才可能提高农民素质,提高农业质量,才可能最终实现农业的高水平持续发展。

生态环境是支撑长江上游经济系统的另一个重要基础。生态环境也是一种资源,是构成长江上游地区资源优势的因素之一,同时也是构成该地区

投资发展环境的重要因素,良好的生态环境是经济持续发展的前提。长江上游地区应把生态建设作为经济建设的一个重要环节,把环境保护作为改善投资环境的一个内容,把两项工作与经济建设同步进行。

参考文献

- [1]王保奋:“略论建设长江经济带”,《长江流域资源与环境》,1992.1,P1
- [2]虞孝感、陈雯:“关于长江产业带建设总体布局的初步思考”,《长江流域资源与环境》,1993.1,P3
- [3]汪一鸣:“不发达地区区域开发的几个理论问题”,《地理学报》,1993.1,P48 (责任编辑 孙立明)