



引力的本质

蔡荣根^{1*}, 王少江¹, 杨润秋², 张云龙³

1. 中国科学院理论物理研究所, 北京 100190;

2. Korea Institute for Advanced Study, Seoul 130-722, Korea;

3. Asia Pacific Center for Theoretical Physics, Pohang 790-784, Korea

* 联系人, E-mail: cairg@itp.ac.cn

2018-01-09收稿, 2018-01-22修回, 2018-01-23接受, 2018-06-15网络版发表

国家自然科学基金(11690022, 11375247, 11435006, 11647601)、中国科学院战略性先导科技专项(XDB23030100)和中国科学院前沿科学重点研究项目资助

摘要 回顾了关于引力本质的历史探索和最新进展. 从牛顿引力和爱因斯坦引力出发, 介绍了关于引力本质历史探索上的两次重大飞跃. 从修改引力、量子引力和全息引力三个方面, 介绍了关于引力本质的最新进展. 对于牛顿引力, 从开普勒行星运动定律出发, 介绍了牛顿万有引力定律. 介绍了最近关于修改牛顿力学和暗物质的进展; 对于爱因斯坦引力, 阐释了引力的几何化, 然后介绍了爱因斯坦引力在宇宙学和引力波方面的应用; 对于修改引力, 从额外的引力自由度、高阶导数引力和高维引力三个方面介绍; 对于量子引力, 从协变量子引力、正则量子引力和其他量子引力三个方面介绍; 对于全息引力, 介绍了它的全息图像、呈展性质以及它与量子信息之间的关系. 但是截至目前, 关于引力本质问题的答案依然是一个谜.

关键词 引力, 牛顿引力, 爱因斯坦引力, 修改引力, 量子引力, 全息引力

什么是引力? 引力虽然是我们最早就知道的, 但同时也是了解得最少的一种基本相互作用力. 对引力的认识在历史上经历了三次巨大的飞跃, 第一次飞跃是牛顿提出了万有引力定律, 第二次是爱因斯坦提出了广义相对论, 第三次是引力的全息原理的提出.

牛顿引力第一次从动力学上解释了开普勒行星运动定律, 阐明了引力的平方反比律, 认识到引力的源是质量, 并准确预言和发现了海王星轨道. 但是牛顿引力自身也有它的致命缺陷, 它暗示引力的传播不需要媒介, 也不用花费时间. 牛顿引力这种鬼魅的超距作用形式在爱因斯坦引力中得到了纠正, 使它作为爱因斯坦引力在弱场低速情况下的良好近似而存在. 在星系形成的大尺度模拟上, 假设暗物质粒子之间的相互作用形式为牛顿引力, 仍然能得到接近观测的大尺度结构. 但是在解释星系旋转曲线时, 如果不假设暗物质的存

在, 则需要在极小加速度的情况下修改牛顿力学.

爱因斯坦引力建立在对狭义相对论从惯性参考系到任意参考系的推广基础上, 此即广义相对性原理. 在等效原理的指引下, 任意加速参考系的效应可以完全等效为局部引力场效应(即曲率), 而马赫原理又认为局部惯性系会受到物质分布的影响, 因此物质分布会影响时空结构并由此产生引力效应, 影响物质运动从而进一步改变物质分布. 爱因斯坦引力将物质分布与时空的几何联系起来, 得到了迄今为止最好的描述引力相互作用的科学理论. 该理论不仅解释了水星近日点反常进动, 还正确预言了光线偏折角, 之后又成功经受住了引力红移和时间延迟的检验, 在天文学上还发现了显示引力透镜效应的爱因斯坦环, 最近引力波的直接探测更是对爱因斯坦引力的光辉验证. 爱因斯坦引力最成功的应用是在宇宙学上, 在此基础上物理学

家建立了被称为宇宙学标准模型的 Λ CDM模型, 解释了包括哈勃膨胀定律、原初核合成、宇宙微波背景辐射、大尺度结构、Ia型超新星加速退行、重子声学振荡、红移空间畸变、引力透镜效应等观测事实. 但是爱因斯坦引力对于宇宙学的应用只取得了部分的唯像上的成功, 它对于宇宙早期和晚期的加速膨胀的解释, 以及对主导星系动力学的经验规律的解释, 都还没有取得决定性的证据. 最后, 霍金和彭罗斯对奇点定理的证明, 宣告了爱因斯坦引力将在极小尺度极高能标下失效, 而霍金和贝肯斯坦对黑洞热力学特别是全息熵的发现, 宣告了爱因斯坦引力将在考虑量子效应后得到不同寻常的修正, 并且其他三种基本相互作用力的量子规范理论特别是粒子物理标准模型的成功, 都进一步要求引力的量子化与其他相互作用力的统一.

对爱因斯坦引力的修改主要分为三种方式, 第一种方式是添加新的与引力耦合的自由度, 如标量场、矢量场、张量场乃至旋量场, 这些新的引力自由度将额外诱导第五力, 因此需要合适的屏蔽机制以通过实验室和太阳系的检验; 第二种方式是修改爱因斯坦引力作用量从场方程中引入度规张量的高阶导数, 改良理论的紫外发散行为进而提高重整化特性; 第三种方式是添加额外的时空维度得到高维引力理论, 并且为了使额外维不至于显著修改太阳系内的引力的平方反比律, 额外维需要以紧致化的小额外维方案或者以嵌入高维空间的大额外维方案引入. 这些修改引力理论都受到了来自实验、观测以及模拟的众多限制, 未来还将受到来自包括引力波在内的新的探测手段的限制, 从而为更好地筛选出合适的引力理论铺平道路.

爱因斯坦引力及其各种修改引力理论的量子化方案主要分为协变量子化和正则量子化两种, 它们之间的主要区别在于是否保持爱因斯坦引力的背景无关性. 协变量子化方案借助弯曲时空量子场论, 将引力处理成经典时空背景上传播的自旋为2的引力子场. 此外通过对黑洞的研究, 发现了霍金辐射以及黑洞热力学, 暗示了引力、热力学和量子力学之间存在深刻联系, 为之后引力的全息性质和呈展性质的提出埋下了伏笔. 协变量子化方案的典型代表是有可能统一4种基本相互作用力的超弦/M理论, 但是目前还缺乏一种背景无关的非微扰表述. 正则量子化方案以圈量子引力为代表, 是目前最成功的保

持爱因斯坦引力背景无关性的量子引力理论, 但是它还无法为引力和物质的量子化提供一个统一的描述.

黑洞热力学的研究促成了对引力的全息性质和呈展性质认识上的飞跃. 引力的全息性质导致了所谓全息原理的提出, 其基本思想是认为一定时空区域内总自由度的数目与包围该区域的表面积成正比. 全息原理的一个具体实现是AdS/CFT对应, 它声称一个高维时空的弱耦合的引力理论对偶于比它低一维的边界上强耦合的量子场论, 这引发了全息字典的建立, 并迅速应用到包括核物理、凝聚态物理、流体力学和量子信息在内的其他学科. 引力的呈展性质则出于对引力起源的思考, 究竟引力是不是一种有着确定微观起源的基本相互作用力, 还是某些更基本的自由度的宏观热力学效应. 最近兴起的试图将量子引力与量子信息相结合的研究热潮认为时空结构起源于量子纠缠, 而物质的变化引起了它们之间量子纠缠的变化进而改变时空结构从而诱导出引力.

1 牛顿引力

牛顿引力是关于引力的第一个科学理论, 它不仅解释了开普勒关于行星运动的三大定律, 还首次预言了海王星的存在. 但它本身也有局限处. 例如, 理论上牛顿引力具有超距作用性质, 从现在的观点看, 实验观测上牛顿引力有一些无法解释的观测, 如水星近日点反常进动、光线偏折和星系旋转曲线.

1.1 开普勒行星运动定律

1609年开普勒(Kepler)出版了他的学术专著《新天文学》, 揭开了行星运动的规律. 在书中, 他分析了他的导师第谷(Tycho)毕生积累的天文观测资料, 论述了他著名的行星运动三定律中的前两条定律, 而第三定律则是在此书出版10年后也就是1619年在一篇文章中提出.

开普勒行星运动第一定律断言, 行星运动的轨道是椭圆形的, 而太阳则位于椭圆的两个焦点之上. 值得注意的是, 开普勒实际上笃信哥白尼体系, 但是发现哥白尼体系中所要求的圆形轨道, 无法拟合第谷高精度观测数据中的火星轨道. 这促成了他发现椭圆轨道

定律.

开普勒行星运动第二定律断言,从太阳到行星的连线在任意等时间间隔内扫过的面积相等.事实上,开普勒发现第二定律的时间早于第一定律.但是,开普勒的此条面积定律并没有使行星位置的计算变得简单.由此引出的开普勒问题促进了此后微积分技术的发展.后世根据轨道角动量守恒可以很轻易地推出该面积定律.

开普勒行星运动第三定律断言,行星运动的周期的平方正比于该行星运动轨道的半长轴的立方.根据该定律,周期的平方与半长轴的立方之比是一个对太阳系内所有行星普适的常数.事实上,考虑到太阳的质量远远大于行星,该定律在牛顿引力框架下是一个足够精确的近似,但是开普勒能从纷繁复杂的天文数据中归纳出如此美妙的规律实属天才.

开普勒行星运动定律属于运动学范畴内的规律,对于行星之所以会遵循这些规律运动的背后的动力学规律,在当时乃至其后的很长一段时间内都是不得而知的,直到牛顿的出现.牛顿借助他的牛顿运动学定律和微积分技术提出牛顿万有引力定律,才最终理解了开普勒行星运动定律的背后是一种与距离平方成反比的吸引力的作用.

1.2 牛顿万有引力定律

牛顿的万有引力定律,除了平方反比律以外,还有一个万有属性,即任何两个带有质量的质点1和2之间,都存在着一种与它们各自质量 m_1 和 m_2 成正比,与它们之间距离 $|\mathbf{r}_{12}|$ 平方成反比的吸引力,而不论它们各自的其他属性,用公式来表达即为

$$\mathbf{F}_{21} = -G_N \frac{m_1 m_2}{|\mathbf{r}_{12}|^2} \hat{\mathbf{r}}_{12}, \quad (1)$$

其中, $\mathbf{F}_{21}$ 是1对2施加的力;负号代表该力是吸引力,即从2指向1; G_N 是牛顿引力常数,它的数值为 $6.674 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$,直到1798年才被由英国科学家卡文迪许(Henry Cavendish)设计的扭秤实验精确测出,而这距离牛顿的《原理》一书的出版已经111年了.从现代物理的观点来看,质量(或者更准确地说是能量-动量-张量)是引力的源,而吸引力的特性以及严格的平方反比律,则都与引力子的自旋为2直接相关.最后,之所以说牛顿引力是关于引力的第一个科学理论,

不仅在于他提出了能够包含旧理论的新理论以解释旧理论所无法解释的事实,还在于新理论能够提出旧理论所没有的尚待检验的新预言.在这方面,海王星的发现可以说是牛顿万有引力定律的光辉胜利.

牛顿在提出牛顿万有引力定律的时候,对它的不足之处同样有着清晰的认识:在他的引力定律中,引力的作用既不需要传播媒介,也不需要传播时间.这令牛顿感到十分不安,但局限于牛顿所处的时代,他也确实对此无能为力.对于第1个问题,直到现在,由于还没有量子化引力,因此仍然没有办法确定引力相互作用的传播媒介是什么,是不是引力子或者其他媒介,抑或根本没有媒介.对于第2个问题,即所谓超距作用问题,与狭义相对论因果律相违背,肯定需要有所修正,这就引出了第三章将要介绍的广义相对论.

牛顿的万有引力定律除了以上两点理论上的不足以外,还有以下三点观测方面的不足:(1)水星近日点的反常进动,在扣除了所有其他行星对水星轨道的摄动以后,仍然有每一百年43''的差异是牛顿引力所无法解释的;(2)对于光线经过太阳时的偏折角,牛顿引力的计算结果只有实际观测值的1/2;(3)漩涡星系的外围星体的旋转速度随星体到星系中心的距离的分布显著偏离牛顿引力的预言.前两个问题的解答构成了对广义相对论早期的四大经典检验中的头两条检验,另外两条检验是引力红移和时间延迟现象.但是对于最后一个问题,直到现在也没有一个令人满意的答案,一种流行的看法是,星系旋转曲线(rotation curve)对牛顿引力的偏离,要么是因为有额外的不发光的暗物质,要么是因为牛顿引力在极小加速度情况下需要修正,即下一节将要介绍的修改牛顿力学(modified Newtonian dynamics, MOND).最近两年,由于对引力波直接观测的成功,使得广义相对论又多了一项经典检验,也将在第二章介绍引力波.

1.3 修改牛顿力学与暗物质

现代宇宙学的标准模型 Λ CDM模型在线性阶和宇宙尺度上取得了巨大的成功,得到了来自原初核合成、宇宙微波背景辐射、大尺度结构模拟、重子声学振荡、Ia型超新星测距、红移空间畸变、引力透镜效应以及子弹星云等观测的支持.但是包含冷暗物质的大尺度模拟在小尺度上也有一些问题,如core-

cusplike问题、missing satellite 问题和too-big-to-fail问题. 这些问题的产生可能来自于模拟中被忽略的重子物质的反馈机制, 或者是因为冷暗物质不是绝对的无碰撞的物质而是存在自相互作用或者量子相干效应, 即所谓自相互作用(self-interacting)暗物质和fuzzy暗物质. 另一种可能性是暗物质根本不存在, 而是由于某种修改引力的效应, 其中最著名的是MOND方案^[1~3]. MOND起源于对星系旋转曲线的解释, 并在其他主导星系动力学的经验规律的解释上显示出了比暗物质模型更加自然的成功, 如万有旋转曲线(universal rotation curves)、重子的Tully-Fisher 关系(baryonic Tully-Fisher relation)和质量差-加速度关系(mass discrepancy-acceleration relation). 但是MOND目前只有一些诸如AQUAL(a QUAdratic lagrangian), QUMOND(QUasi linear modified Newtonian dynamics), 张量-矢量-标量(tensor-vector-scalar, TeVeS)理论, bimetric MOND和爱因斯坦-以太(Einstein-æther)理论等唯像理论的表述, 直到最近才有一些对MOND的起源的研究, 如最近斯莫林(Lee Smolin) 和维林德(Erik Verlinde)关于MOND的量子引力起源的思考. 最后值得注意的是, 最近引力波天文学的发展无疑对检验MOND 模型打开了新的窗口. 例如, 根据LIGO(laser interferometer gravitational wave observatory)探测到的双黑洞合并的引力波事件宣称排除了Einstein-æther理论和bimetric MOND 理论^[4].

2 爱因斯坦引力

虽然上面提到了牛顿引力有这样和那样的不足, 但是爱因斯坦提出广义相对论的动机几乎纯粹是出于狭义相对论内部逻辑自洽性上的需要. 狭义相对论的两条基本假设告诉我们, 光速不变原理可以用来校准时钟, 从而定义惯性参考系; 而狭义不变性原理则保证了物理定律在所有惯性参考系里的形式相同. 但是爱因斯坦觉得惯性参考系太过特殊, 物理定律应当在任意参考系里的形式都相同. 沿着这样一条逻辑链条, 爱因斯坦成功地得到了广义相对论, 而爱因斯坦引力对于避免牛顿引力超距作用、解释水星近日点反常进动、正确预言光线偏折角, 以及其后对于引力红移、时间延迟、引力透镜乃至引力波的预言, 对于爱因斯坦来说都只是理论的副产品而非原始动机.

2.1 引力几何化

爱因斯坦引力中用时空度规来描述引力场. 场的概念最早由法拉第(Michael Faraday)引入, 并用力线(line of force)的方式将它图像化. 法拉第学徒出身, 所受教育十分有限, 但是他对经典物理的影响巨大. 仅举一例, 他的法拉第电磁感应定律使得之前研究的电场和磁场终于作为一个整体来对待, 为电磁场概念的建立打下了基础. 麦克斯韦(James Clerk Maxwell)在吸收了法拉第的场的概念后, 成功地总结了前人关于电与磁的研究成果, 并将法拉第的动磁生电的模型数学化, 提出了麦克斯韦电磁场方程. 之后赫兹(Heinrich Hertz)成功发现了由电磁场方程预言的电磁波的存在, 确认了光作为电磁波而电磁波以光速运动. 因此在麦克斯韦的电磁场理论中, 电磁力的传播需要时间, 传播速度为光速. 之后为了协调麦克斯韦电磁场理论与表征牛顿绝对时空观的伽利略变换之间的矛盾, 爱因斯坦于1905年建立了狭义相对论, 将时间和空间作为一个统一的整体来对待. 在认识到狭义相对论中惯性参考系的特殊性后, 爱因斯坦开始尝试将物理规律表述到任意参考系, 以遵循广义不变性原理, 最终经过8年的不懈努力终于在1915年建立了广义相对论. 在下文将会看到, 等效原理(equivalence principle)是如何更进一步将任意加速参考系与局部引力效应联系到一起, 而马赫原理(Mach's principle)又是怎样将引力效应与物质分布联系到一起. 至此, 广义相对论将引力几何化的整个逻辑链条初具雏形.

等效原理有很多版本. 早在牛顿时代, 人们就猜测, 出现在牛顿第二定律里的惯性质量与出现在牛顿引力定律里的引力质量其实是一回事, 这就是弱版本的等效原理(weak equivalence principle). 牛顿本人最早进行了相关实验以检验惯性质量是否精确等于引力质量, 其后贝塞尔(Friedrich Wilhelm Bessel)改进了实验, 更高精度的实验由厄特沃什(Loránd Eötvös) 于1885年至1909年完成. 惯性质量与引力质量等价的直接后果是, 引力是普适(universality)的, 在引力场中自由下落的物体的加速度是普适的, 与物体的内禀属性无关. 如果加速参考系的加速度等于局部引力场的引力加速度, 将无法区分在一个加速参考系内所进行的力学实验和静止在一个局部引力场内所进行的力学实验, 这就

是爱因斯坦著名的电梯理想实验. 这里需要强调的是电梯是处于引力场的局部, 否则电梯如果足够大, 是可以探测到由于局部引力场的变化所导致的潮汐力(tidal force). 弱等效原理因此暗示了在时空中存在一类特殊的轨迹, 即所谓惯性(自由落体)轨迹, 也即引力场中无加速度的粒子的轨迹, 这里说的无加速度是指仅受引力作用, 也就是说粒子走的是测地线运动. 弱等效原理还限制了原来在狭义相对论中定义全局惯性参考系的能力, 在引力场中只能定义局部惯性参考系. 这对我们来说, 也不失为一件好事, 即有引力场时的时空虽然全局的性质可以很复杂, 但是在局部还是可以近似为平直时空, 这为爱因斯坦找到合适的数学工具即黎曼几何来描述弯曲时空提供了依据. 如果将弱等效原理中的力学实验强化为局域非引力实验则得到爱因斯坦等效原理(Einstein equivalence principle), 若进一步强化为任意局域实验(引力的或者非引力的)则得到强等效原理(strong equivalence principle).

马赫原理本身的叙述比较模糊, 而且有很多并不完全等价的变种, 它也从来不是构建广义相对论的原理之一, 但是它在历史上对于启发爱因斯坦意识到物质分布会影响到物体运动起到过重要作用, 连“马赫原理”这个名称也是爱因斯坦首先引入的. 在牛顿力学中, 惯性定律叙述为: 当物体不受外力或所受合力为零的情况下总是保持静止或者匀速直线运动. 但是这仅仅只是叙述了什么是惯性, 并没有解释物体为什么具有惯性. 马赫原理试图理解惯性的起源, 认为局部惯性系会受到整个宇宙的运动和物质分布的影响. 爱因斯坦接受了马赫的观点, 认识到物质分布可以影响时空结构从而影响物体运动.

至此, 爱因斯坦建立广义相对论的逻辑链条已经清晰起来: 牛顿的绝对而割裂的时空在爱因斯坦狭义相对论中变成相对而统一的时空, 其中光速不变原理允许我们定义全局惯性参考系, 而狭义不变性原理要求物理规律在不同惯性参考系里的形式相同; 广义不变性原理则进一步将物理规律从惯性参考系推广到任意参考系, 等效原理保证了局部引力场与加速参考系的不可区分性, 马赫原理进一步认为局部惯性系将受到物质分布的影响, 因此局部引力场也将受到物质分布的影响, 数学化后即为爱因斯坦于1915年11月25日得到的引力场方程

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (2)$$

其中, 左边的爱因斯坦张量 $G_{\mu\nu}$ 描述了时空是如何弯曲的, 而右边的能量动量张量 $T_{\mu\nu}$ 则描述了物在时空中如何分布的, 这就是为什么惠勒(John Wheeler)对爱因斯坦引力场方程作出了如下描述: 物质告诉时空如何弯曲, 时空告诉物质如何运动.

爱因斯坦的广义相对论中所提出引力的几何化观点具有划时代的意义, 是人类历史上继牛顿引力之后对于引力本质认识的一大飞跃. 但也正是这种几何化观点使得引力相互作用的量子化之路和4种基本相互作用的统一之路充满艰辛, 原因有二: (1) 在爱因斯坦引力中, 引力作为时空弯曲的反映, 其本身可以说并不是一种基本相互作用力, 这与其他三种基本相互作用力的性质完全不同; (2) 引力场也与其他相互作用的媒介场不同, 其本身就是描述时空的度规张量, 打一个形象的比方, 引力场本身既是演员又是舞台.

2.2 宇宙学

在爱因斯坦完成广义相对论之后不久, 他开创性地首先将其运用到宇宙学中. 在当时没有观测数据直接支持的情况下, 爱因斯坦大胆地假设了所谓宇宙学原理: 宇宙在大尺度上是空间均匀且各向同性的. 由此爱因斯坦最初得到了收缩宇宙的结论, 但是基于某种信念, 1917年爱因斯坦在引力场方程中额外添加上了一项代表斥力的宇宙学常数项以平衡引力的作用, 从而得到了一个静态宇宙模型. 但是在1929年哈勃(Edwin Hubble)发表的哈勃定律即宇宙正在膨胀的观测事实之后, 爱因斯坦随即放弃了宇宙学常数项, 并称之为自己一生中最大的错误(the biggest blunder of life). 事实上, 在爱因斯坦引入宇宙学常数项的时候就被指出该静态宇宙解本身就是不稳定的, 偏离静态解的任何微扰都将使之更进一步收缩或膨胀. 在膨胀宇宙的观点下, 宇宙学原理应当重新叙述为: 宇宙在大尺度上是空间均匀且各向同性并随时间演化, 表达为数学形式即为FLRW(Friedmann-Lemaître-Robertson-Walker)度规. 但是宇宙学常数项在历史上的命运几经沉浮, 曾数次以不同原因被重新引入, 而又随即数次被抛弃, 最终在1998年由于对Ia型超新星的距离-红移观测, 人们才重新确认了宇宙学常数的存在, 而这一次它

的角色不再是维持静态宇宙而是推动宇宙加速膨胀,并得名暗能量(dark energy). 如果某种动力学场也能起到类似宇宙学常数那样的推动宇宙加速膨胀的作用,那么这种动力学场也被称为(动力学)暗能量. 最开始人们认为宇宙学常数来自真空能(vacuum energy),但是量子场论对于真空能的计算结果比实际所要求的大小大了上百个量级,即使引入超对称(supersymmetry)也不能将差异降到可接受的范围内,所以暗能量问题直到现在都是一个巨大的谜团. 某些修改爱因斯坦引力的模型也可以使宇宙加速膨胀,这里就不一一介绍了. 总之,宇宙学经历了从静态到膨胀,又从膨胀到加速膨胀的认知上的飞跃,对于推动当前宇宙加速膨胀的原因,究竟是修改引力还是暗能量,如果是暗能量,究竟是动力学暗能量还是宇宙学常数,是当前暗能量研究的两个主要议题,对它们的解答将有助于预测宇宙未来的命运.

回到哈勃发现宇宙正在膨胀的观测上来,由于当时对代表宇宙膨胀速度的哈勃常数的测量非常不准确,导致膨胀宇宙的观点也几经沉浮,甚至连哈勃本人也曾在1941年推翻自己所作出的膨胀宇宙的发现. 膨胀宇宙的一个令人不安的推论是,沿着时间回溯,宇宙越来越小且越来越热,宇宙似乎因此存在一个起点(也因此被戏称为热大爆炸模型),这种具有宗教意味的开端使膨胀宇宙模型饱受争议. 但是随后另外两项重大的天文观测无可辩驳地支持了热大爆炸模型(the hot big-bang model): 宇宙微波背景辐射(cosmic microwave background radiation)和大爆炸核合成(big-bang Nucleosynthesis). 但是热大爆炸模型也因此面临一些新的问题,如磁单极子问题、平坦性问题和视界问题. 为了解决热大爆炸模型的这些问题而提出的暴胀模型,在宇宙极早期引入一段指数(或者接近指数)式的加速膨胀时期. 暴胀模型中极早期的真空量子涨落(quantum fluctuation)在宇宙暴胀的作用下变成了原初扰动(primordial perturbation),这些原初扰动在宇宙微波背景辐射的最后散射面(last-scattering surface)上留下痕迹并被观测精确证实,而且这些原初扰动在引力坍缩的作用下进一步成长为现在的大尺度结构(large-scale structure). 这两项证据都强烈地支持了暴胀模型,但是对于暴胀模型的验证起决定性作用的原初引力波(原初扰动的B模式极化信号)还没

有在宇宙微波背景上找到,这也是中国的阿里计划正在努力的方向. 暴胀模型虽然一定程度上解决了热大爆炸的初值问题,但是它本身也具有其特殊的初值问题. 为了避免暴胀模型的初值问题,还提出了诸如反弹模型(bouncing model)、弦气体模型(string gas model)、火劫模型(ekpyrotic model)和anamorphic model等模型.

除了上文介绍的宇宙学常数和暴胀模型,现代标准宇宙学模型的另一个要素即所谓冷暗物质(cold dark matter). 暗物质是一种非重子物质(non-baryonic mater),一般认为没有或者具有很弱的电磁相互作用,从而在电磁辐射谱上是不可见的,只能通过引力相互作用探测,如引力透镜(gravitational lensing)效应. 暗物质的存在具有多方面的证据,学术界直到20世纪80年代才逐渐接受暗物质的存在,这其中最重要的贡献来自于鲁宾(Vera Rubin)于20世纪60~70年代所进行的星系旋转曲线的观测. 事实上,在此之前,就已经有人研究了星系旋转曲线并提出了暗物质的概念. 根据牛顿引力(牛顿力学在估计星体运动速度时依然是良好的近似),星系边缘的星体运动速度随距离的变化应当呈开普勒下降(Keplerian decline)的趋势,但是实际观测结果表明,星体的速度在一定半径范围外并没有预想中的下降而是继续保持常数. 由于爱因斯坦引力已经经受住了许多重大检验,因此最简单的假设就是星系中分布有不发光的暗物质,是它们维持住了涡旋星系悬臂外围星体的高速运动. 除了星系旋转曲线(galaxy rotation curves)对星体运动速度及其速度弥散的测量,还有其他相互独立的观测证据支持暗物质的存在. 例如,在星系团中,星系速度的测量、星系间热气体的X射线的观测以及星系团的引力透镜的成像;在星系碰撞后所形成的子弹星云(bullet cluster)、X射线所测量的发光物质的分布中心与引力透镜所测量的引力物质的分布中心的不一致性;结构形成的N体模拟里使用冷暗物质模型得到的大尺度结构与大尺度巡天观测的图像更加接近;还有一些来自诸如原初核合成、宇宙微波背景辐射、重子声学振荡(baryon acoustic oscillation)、拉曼阿尔法(Lyman-alpha)射线观测、红移空间畸变(redshift-space distortions)、Ia型超新星距离红移测量等观测限制都与假定了暗物质的宇宙学标准模型(Λ CDM模型)更加吻合.

2.3 引力波

在爱因斯坦1915年末发表广义相对论百年之际, LIGO(laser interferometer gravitational wave observatory)首次于2015年9月14日协调世界时9点50分45秒直接探测到距地球410百万秒差距外的一个双黑洞系统由于相互绕转并合所发出的引力波^[6], 由此开启了引力波天文学的新时代. 最近LIGO首次于2017年8月17日协调世界时12点41分04秒直接探测到距地球40百万秒差距外的NGC 4993星系中一个双中子星系统并合所发出的引力波和全波段的电磁波^[7], 由此开启了多信使引力波天文学的新时代. 未来引力波探测实验将为天文学、引力物理、宇宙学甚至粒子物理提供十分宝贵的信息^[8].

引力波探测始于韦伯(Joseph Weber)的直接探测实验, 他在1957年的Chapel Hill会议(后被追认为国际广义相对论大会)后, 立即开始进行引力波的探测实验, 虽然他在20世纪50~70年代进行的韦伯棒的探测实验后来被广泛认为是失败的, 但是他的行为在当时改变了人们认为因为引力波效应太小而无法探测的畏惧感, 激励了后世继续进行引力波探测. 引力波探测的第一次重大进展来自于豪斯(Russell Hulse)和泰勒(Joe Taylor)对双星系统绕转时间的长期观测. 由于引力波辐射带走能量, 双星系统的绕转时间越来越小, 而豪斯-泰勒双星系统中有一颗是脉冲星, 正好可以用来指示系统的绕转时间. 经过30多年的观测, 豪斯-泰勒双星系统的绕转时间的衰减与广义相对论的预言高度吻合, 从而间接证实了绕转双星会辐射引力波, 豪斯和泰勒也因为对引力波的间接探测而荣获1993年的诺贝尔奖. 但是, 对引力波的直接探测则非常困难, 除了早期的韦伯棒探测技术, 目前取得成功的只有LIGO的激光干涉技术, 此项技术未来还将应用到空间引力波项目中, 如欧洲主导的激光干涉空间天线(laser interferometer space antenna, LISA)项目、日本正在推进的DECIGO(deci-hertz interferometer gravitational wave observatory)项目和中国正在筹划的太极和天琴项目. 另一种引力波探测技术则是利用毫秒脉冲星计时阵列(pulsar timing array), 如国际上几个射电天文台联合进行的IPTA(international pulsar timing array)项目, 平方公里阵列SKA(square kilometre array)项

目, 以及中国的500 m口径球形射电望远镜FAST项目. 当引力波经过毫秒脉冲星计时阵列时, 时空的扰动将影响它们所发出的脉冲的到达时间, 然后根据不同脉冲星的到达时间的变化及相互关联来探测引力波. 这三项技术所探测的引力波源各不相同, 地面引力波探测器主要探测几十赫兹的引力波, 空间引力波探测器主要探测毫赫兹量级的引力波, 而脉冲星计时阵列技术主要探测纳赫兹频段的引力波. 原初引力波的探测与以上三种引力波的探测都不一样, 它主要依靠对宇宙微波背景辐射中的B模式极化信号的识别, 如作为下一代CMB(cosmic microwave background)探测卫星的Euclid卫星, 地面探测除了位于南极洲的BICEP项目, 还有我国位于青藏高原的阿里项目. 最后, 对于极高频引力波的探测技术目前还比较薄弱, 我国重庆大学在此方面亦有研究.

引力波之所以能引起广泛的关注, 除了作为人类工程技术的光辉胜利, 还因为引力波物理与其他领域的广泛联系. 首先, 整个引力波波谱的来源是十分丰富的. 高频引力波的波源可以是宇宙在重加热(reheating)时期所产生的背景引力波; 地面引力波探测器所探测的引力波波源大多是双黑洞系统、黑洞-中子星系统和双中子星系统; 空间引力波探测器所探测的引力波波源除了极端质量比(extreme mass ratio)的黑洞双星系统以外, 还有能够产生随机背景引力波的双白矮星系统、原初黑洞(primordial black hole)以及宇宙一阶相变(cosmic first-order phase transition)所产生的真空泡泡(vacuum bubbles); 脉冲星计时阵列所探测的极低频引力波则来自于超大质量的双黑洞系统; 最后还有存在于全频段的宇宙原初引力波. 其次, 引力波物理对基础物理也具有重要的现实意义. 双黑洞系统产生引力波的过程可以大致分为三个阶段, 即旋进(inspiring)、并合(merging)和铃荡(ringdown), 其对应的计算方法分别为后牛顿近似方法(post-Newtonian approximation)、广义相对论数值模拟(general relativity numerical simulation)和黑洞微扰理论(black hole perturbation theory), 近年来还发展了有效单体方法(effective-one-body method). 探测双黑洞并合的引力波可以研究黑洞无毛定理(no hair theorem)、黑洞视界附近的几何和物质的吸积、引力波传播的规律、引力波的记忆效应(memory effect)等. 黑洞-中子星系统或者双中子

星系统可以借由多信使天文学(multi-messenger astronomy)里的电磁对应体(electromagnetic counterparts)来研究引力波相对电磁波的一些特殊性质. 这些都使得对于黑洞物理和中子星相关的核物理不再局限于纯粹的理论猜测而进入到实际验证阶段. 引力波物理不仅对以上引力物理具有重要意义, 也对粒子物理和宇宙学具有不可替代的作用. 相变引力波可以对一些当前乃至未来很长一段时间内无法用现有对撞机技术探测的超出粒子物理标准模型的新物理进行有效限制. 引力波的标准汽笛(standard siren)效应还为宇宙测距提供了新思路, 并与原来的利用电磁波的宇宙测距方法形成互补优势. 重加热时期的背景引力波和暴胀时期的原初引力波也对探索早期宇宙的具有不可估量的作用.

3 修改引力

在爱因斯坦引力提出之后, 物理学家还提出了其他各式各样的引力理论(也有一些修改引力理论如诺斯特朗引力理论是在爱因斯坦引力之前提出的), 这里将它们粗略地统称为修改引力理论, 并且参照文献[9]中的分类方法分为三类. 它们有的是在爱因斯坦引力的基础上引入了新的自由度与引力耦合, 有的是修改爱因斯坦引力的作用量从而使得场方程含有度规张量的高阶导数, 还有的是4维爱因斯坦引力的基础上添加额外的时空维度.

3.1 额外的引力自由度

第一种修改引力的方法就是引入额外的自由度, 如额外的标量场、矢量场、张量场甚至旋量场等. 标量-张量理论的一个典型代表是Brans-Dicke理论, 它通过一个依赖于时空的标量场耦合引力场得到了一个随时空变化的引力常数, 并且诱导了由标量场传递的额外的引力相互作用. 一般认为拥有二阶运动方程的最一般的标量-张量理论是Horndeski理论. 矢量-张量理论的一个典型代表是Einstein-æther理论, 它所引入的额外的矢量场会破坏洛伦兹不变性. 同时引入标量和矢量的理论有标量-张量-矢量(scalar-tensor-vector, STVG)理论和张量-矢量-标量(TeVeS)理论, 它们提出的动机都与解释星系旋转曲线和MOND有关. 张量-张量理论也叫双度规(bimetric)理论, 它在背景度规 $g_{\mu\nu}$ 上

再引入一个参考度规 $f_{\mu\nu}$. MOND理论也有协变的双度规理论形式即Bimetric MOND. 另外双度规理论的一个典型代表是dRGT有质量引力理论, 它的参考度规 $f_{\mu\nu}$ 是非动力学的. 但是为了从dRGT有质量引力得到FLRW宇宙学, 需要对参考度规引入动力学, 从而得到双度规引力(bigravity)理论. 旋量-张量理论的一个典型代表是爱因斯坦-嘉当(Einstein-Cartan)理论, 它假定了一个非零的扭率张量(torsion tensor)即仿射扭率(affine torsion)的反对称部分从而可以使其耦合物质场的自旋角动量. 爱因斯坦-嘉当理论的一个突出特性是可以避免奇性(singularity), 这为之后的量子引力提供了新的思路. 引入额外的引力自由度一般也引入了额外的相互作用力即第五力(fifth force), 但是爱因斯坦引力在从实验室到太阳系的各个尺度下都被严格检验过, 因此这类修改引力理论需要合适的屏蔽机制(screening mechanism)以通过太阳系尺度内的检验. 常见的屏蔽机制有三种, 即变色龙(chameleon)机制、对称子(symmetron)机制和Vainshtein机制. 为了检验这些修改引力理论, 需要在更大尺度上限制爱因斯坦引力, 如大尺度结构、微波背景辐射和原初引力波等来自宇宙学方面的检验.

3.2 高阶导数引力

第二种修改引力的方案就是在场方程里引入度规张量的更高阶的导数, 因为爱因斯坦场方程中的度规张量最高只有二阶导数. 一个例子是 $f(R)$ 引力, 它的作用量是里奇(Ricci)曲率标量的任意函数, 它的场方程包含度规张量的四阶导数. 另一个例子是高斯-博内特(Gauss-Bonnet)引力, 它的作用量是黎曼(Riemann)曲率张量及其各阶缩并的特定的组合形式 $R^2 - 4R_{\mu\nu}R^{\mu\nu} + R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}$. 在4维或者低于4维的时空里高斯-博内特项是纯拓扑项, 只有在高于4维时空高斯-博内特引力才具有非平凡的动力学. 高斯-博内特项本身虽然含有度规张量二阶导数的四次项, 但是它的场方程却只含有度规张量的二阶导数项. 更一般地, 任意 D 维时空中保持引力场方程最多含有度规张量的二阶导数项的引力是Lovelock引力, 它在高于5维的情况下才具有非平凡的动力学. 其他引力作用量若包含黎曼曲率张量及其各阶缩并的一般形式 $f(R, R_{\mu\nu}R^{\mu\nu}, R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma})$, 那么它的场方程则一般是度

规张量的四阶导数. 在场方程中引入度规的高阶导数会使得引力的传播子在紫外部分下降得比较快, 从而可以提高重整化的特性. 但是引入度规的高阶导数也会伴随着一些问题, 如引入的类鬼(ghost-like)模式所导致的不稳定性. $f(R)$ 引力可以避免这些模式, 因为高阶导数只作用在不传播的非动力学的模式上, 而高斯-博内特引力在闵可夫斯基(Minkowski)背景下本身就是无鬼的(ghost-free).

3.3 高维引力理论

爱因斯坦引力的数学表述工具即黎曼几何并不局限于我们所处的1+3维时空, 因此原则上高维引力理论在数学表述上并不存在原则性的困难. 真正的难点在于, 现实世界的引力表现得十分接近平方反比律, 而高维引力将使得牛顿引力显著偏移平方反比律, 但是行星轨道只有在平方反比律下才能保持稳定. 现在大致有两种高维引力理论, 一种是添加小的额外维, 另一种是添加大的额外维. 小额外维的想法起源于卡鲁扎-克莱因(Kaluza-Klein)理论. 它试图在5维时空中统一电磁力与爱因斯坦引力, 虽然最后失败了, 但是其思想被之后的弦论所继承. 5种10维的超弦理论以及11维M理论需要紧致化掉额外维度得到4维的有效理论, 不同的紧致化方案得到不同的有效理论, 从而形成所谓的弦景观(string landscape). 大额外维的想法起源于所谓的膜世界(braneworld)方案, 它早在弦论中引入D膜时就已萌芽, 但真正引起广泛关注是在ADD(Arkani-Hamed-Dimopoulos-Dvali)模型^[10]和RS(Randall-Sundrum)模型^[11, 12]提出以后. 与Kaluza-Klein方案假设的紧致化的小额外维不同, 大额外维方案认为粒子物理的标准模型实际上是被束缚在一个1+3维的膜上, 而这张膜本身又被嵌入到更高维的bulk空间中, 这个bulk空间就是一个大的额外维. 现有的对粒子物理标准模型的粒子KK模式的探测实验只是限制了膜的厚度, 而并不能限制bulk空间本身的大小. 要限制大额外维的大小需要借助引力, 因为只有引力相互作用能够冲破膜的束缚在bulk空间里传播. 现有引力实验能将爱因斯坦引力限制到0.1 mm量级, 这在膜世界方案里其实是等价地限制了大额外维的大小, 该大小与被嵌入的bulk空间维数有关. 在ADD模型和RS模型之后, DGP(Dvali-Gabadadze-

Porrati)模型^[13]吸引了很大关注, 因为该模型自带的自加速解可以解释宇宙晚期的加速膨胀而不需要引入暗能量, 但是目前的观测对该模型给出了很强的限制^[14].

4 量子引力

广义相对论中的奇点定理^[15, 16]暗示了爱因斯坦引力不是引力理论的最终形式, 而霍金辐射与黑洞热力学的发现也暗示了引力最终需要量子化, 从而与其他三种基本相互作用力的量子规范理论相统一. 引力的量子化方案大体上可以分为协变量子化和正则量子化两大类, 协变量子化可以看作是对一个背景度规上自旋为2的引力子场的量子化, 而正则量子化则保持了爱因斯坦引力的背景无关性.

4.1 协变量子引力

协变量子引力的早期尝试是半经典引力或弯曲时空的量子场论, 它的背景度规是经典的而背景度规上的物质场是量子化的, 引力被看作是在这个经典时空上传播的自旋为2的引力子场. 从弯曲时空的量子场论得到的最重要的启示是, 原来平直时空的真空和粒子的概念在弯曲时空背景下不再是绝对的, 而是随观测者而定. 对于一个处在Rindler时空的加速观测者来说, 真空也可以辐射粒子从而具有温度, 此即安鲁(Unruh)效应. 将安鲁效应从Rindler时空推广到因果结构与之类似的一般黑洞时空则得到霍金(Hawking)辐射. Bekenstein^[17]和Hawking等人^[18~20]关于黑洞热力学的工作暗示了黑洞具有温度和全息熵, 这为之后引力的全息原理的提出埋下了伏笔.

协变量子引力最杰出的代表当属超弦/M理论, 并且由于它的迅猛成长使其已经成为当前最有可能统一4种基本相互作用力的万有理论, 这里将弦论归入协变量子引力仅仅是因为它目前还没有一个背景无关的非微扰表述. 弦论的5个基本特性是1维的弦、额外维、超对称、紧致化和对偶性. 1维的弦代替了场论中的点粒子, 最初是作为一个描述强相互作用力而引入的唯像模型, 但是对它进行量子化后意外发现它自动包含了一个自旋为2的粒子, 并被大胆地解释为引力子, 从而弦论走上了量子引力乃至统一场论的道路. 这样做的另一个重要意义在于, 时空坐标不再是场论中描述

点粒子的场的参数,而是被提升为可以被量子化的场,真正的参数是弦的世界面(world sheet)坐标.描述弦的作用量可以等效为一个2维西格玛模型,它的重整化流方程自然诱导出弦所在背景时空的真空爱因斯坦场方程.弦论出于自治性的考虑需要同时引入特定维数的额外维和超对称,但是为了使理论能描述真实的4维时空的物理现象,就必需假设一些额外维度被紧致化,从而不会被实验观测到,并与现有的粒子物理标准模型对接,但是这需要恰当地破缺超对称.弦论的最后一个特性即对偶性有望直接将5种超弦理论统一为一种理论即M理论,并且还具体实现引力的全息原理提供了技术准备.

4.2 正则量子引力

正则量子引力的代表是圈量子引力(loop quantum gravity theory),它是在Ashtekar^[21]的联络动力学表述的基础上发展起来的,以自对偶自旋联络(self-dual spin connection)1-形式在圈上的积分作为正则量子化中的位形变量,这种正则变量的选取完全避免了使用度规场,从而也不再引进所谓的背景度规,因此圈量子引力通常被称为是一种背景无关的非微扰的量子引力理论.另外它还在普朗克尺度上的空间量子化^[22]以及对黑洞熵的微观计算^[23]上都取得了重要的成功.虽然圈量子引力理论是一个非常漂亮的引力量子化方案,但是与一个完备的量子引力理论还相距甚远.圈量子引力理论的成果主要局限于理论的运动学方面,在动力学方面的研究却一直进展缓慢.直至今今,人们还不清楚圈量子引力理论是否以广义相对论为弱场极限,或者说圈量子引力理论对时空的描述在大尺度上能否过渡为人们熟悉的广义相对论时空.因此圈量子引力理论甚至无法证明其是一个量子引力理论.此外,人们也不知道如何在圈量子引力理论中考虑量子场论以及物质场的量子效应.圈量子引力理论是建立在将经典的广义相对论进行量子化的基础上的.然而,在后文中将看到,一些证据和分析表明引力相互作用也许并不是一种基本相互作用而是一种呈展现象.如果真的是这样,圈量子引力理论则完全忽略导致引力现象的微观自由度.

圈量子引力的独特之处在于它的背景无关性,在它之后另一个具有此特征的理论是CDT(causal dy-

namical triangulation),其基本思想是先将时空按时间分层,每个层上利用时空三角剖分(triangulation)后单形(simplex)的胶合(glue)来制造弯曲时空.与以往类似的三角剖分的尝试不同的是,CDT里的单形是特别构造的以保持因果性,因此它成功地在大尺度上再现了4维时空,而在小尺度即普朗克能标附近时空则表现为2维.与圈量子引力的自旋泡沫(spin foam)表述不同的是,CDT可以准确计算两点之间的间距.与CDT类似但走的更远的是因果集(causal set)理论,区别主要体现在CDT假设的离散的因果结构是更加特殊的单形.

4.3 其他量子引力

量子引力理论的困难促使物理学家在其他维度下考虑它可能具有良好表述的可能性.例如,在2维空间加1维时间的情况下,广义相对论没有可传播的引力自由度,因此2+1维引力是一种拓扑理论.Witten^[24]在1988年猜测2+1维拓扑引力的对偶理论是一个陈-西蒙斯(Chen-Simons)理论,而且它还是精确可解的,这使得它成为一个研究量子引力的玩具模型.但是Witten^[25]后来在2007年否定了他的这一猜测,并再次提出2+1维的拓扑量子引力的对偶理论应该是魔共形场论(monster conformal field theory).在3维引力的研究热潮褪去之后,从90年代以来,人们开始对2维引力感兴趣,因为2维引力的爱因斯坦方程即为刘维尔方程,所以刘维尔场论提供了2维量子引力的一个玩具模型.2维引力还可以由高维引力通过维度约化或者直接从非临界弦论得到,如CGHS(Callan-Giddings-Harvey-Strominger)模型^[26].由于纯粹的2维引力都是局部平坦的,所以一个非平凡的2维引力一般都要耦合一个被称为伸缩子(dilaton)的标量场,如JT(Jackiw-Teitelboim)模型.最后,近年来对于一个玩具模型即SYK(Sachdev-Ye-Kitaev)模型的研究^[27]重新复活了对2维引力研究的兴趣,因为SYK模型在强耦合下完全可解,而且它还具有最大混沌性,其低能有效理论是具有Schwarzian作用量的JT引力.

量子引力的另一种尝试是改变原来爱因斯坦引力的某些性质.例如,很早就知道爱因斯坦引力是不可重整化的,因为牛顿引力常数具有负的质量量纲.1976年,温伯格(Steven Weinberg)提出渐进安全引力,即假设在高能的紫外极限下重整化流存在一个不平凡的高斯型

不动点,也就是说牛顿引力常数在能区会趋于一个有限值,虽然已经积累了许多证据,但是这一点目前还没有证明.又如,爱因斯坦引力是一个洛伦兹协变的理论,时间和空间的地位是平等的且密不可分.2009年,Horava^[28]提出一种在能极限下破缺时间和空间之间的洛伦兹协变性,并在大尺度极限下近似恢复洛伦兹协变性的量子引力理论. Horava-Lifschitz 引力是一种可重整化的量子引力尝试.在能紫外区域,时间和空间有不同的缩放指标 $t \sim x^z$,其中 z 是Lifschitz指数.在低能红外区域, Horava-Lifschitz引力有望退回到广义相对论.在Horava-Lifschitz引力中,由于只引入了度规对空间的高阶导数,对时间仍然保持最高的二阶导数,因此可以避免不稳定的类鬼(ghost-like)模式.

5 全息引力

继牛顿引力和爱因斯坦引力之后,引力理论经历的第三次观念上的巨大飞跃来自于对引力的全息性质的认识,它起源于霍金和贝肯斯坦对黑洞熵及其热力学的研究,该研究后来也引发了对引力的呈展(emergent)性质的研究.最近一些年来,由于对全息纠缠的研究使得物理学家越来越多地将理解引力本质的希望寄托到它与量子信息的联系上来.

5.1 引力的全息图像

黑洞热力学的研究揭示了引力、量子力学和热力学之间的深刻关系.黑洞的熵与其视界面积成正比的性质使得't Hooft^[29]在1993年猜测任何一个系统在量子力学中可以由其边界上的理论完全描述.这一猜测被Susskind^[30]在1994年进一步发展并将其提升为一个原理,即任何含有引力的量子系统都满足全息原理.全息原理在其提出之后第一次受到广泛关注是在1997年到1998年之交,因为Maldacena第一次给出了全息原理的一个具体模型实现,即AdS/CFT对应:一个K-K约化到5维的 $AdS_5 \times S^5$ 时空中的Type IIB超弦理论和一个3+1维时空中 $N=4$ 的超Yang-Mills理论对偶.随后Witten^[31]和Gubser等人^[32]提出了以他们名字命名的GKPW公式,从AdS时空引力的角度出发计算得到了边界上共形场论的两点关联函数,从而给出了AdS/CFT对应一个严格的数学表述: $AdS_5 \times S^5$ 时空

中IIB超弦理论的离壳配分函数等价于边界上 $N=4$ 的超Yang-Mills共形场论的在壳配分函数,即

$$Z_{AdS_5 \times S^5} = Z_{N=4, CFT_4}. \quad (3)$$

AdS/CFT对应不仅为弦论的非微扰表述给出了一种定义方式,它本身所蕴含的全息原理可能比弦论更加基本,这得益于它的如下两个特点:(1)它是关于两个不同维度之间的理论的对偶,在 AdS_5 中多出来的一个径向维度实际上不是对偶于 $N=4$ 的超对称Yang-Mills理论的任何一个真实的时空维度,而是对偶于它的重整化能标,此即全息重整化的基本思想;(2)它是一个弱耦合的引力理论与一个大 N 极限下的强耦合的规范场论之间的对偶,此即规范/引力(gauge/gravity)对偶的基本思想.由于广义相对论被认为是任何一个自洽的量子引力的经典极限,因此借助这两条特性,物理学家第一次可以借助爱因斯坦引力来研究边界上各种强耦合场论,从而开启了引力的全息应用的大门.这些应用中最突出的几个成就包括:(1) AdS/CFT在核物理中的典型应用^[33]是从理论上推导了夸克胶子等离子体的黏滞系数与熵密度之比并与实验值十分接近;(2) AdS/CFT在凝聚态物理中的典型应用是提出了超导相变的全息模型^[34],以及其他各种强关联凝聚态体系,比如超流、费米/非费米液体、拓扑绝缘体、奇异金属、电荷密度波与金属/绝缘体相变、铁磁/反铁磁系统以及一些非平衡态系统;(3) AdS/CFT在流体力学中的典型应用是流体/引力(fluid/gravity)对偶^[35,36],它受到了爱因斯坦场方程与Navier-Stokes方程之间相似性^[37]的启发,即黑洞对外部的响应对于外部观测来说可以看作是黑洞外面有一层膜,黑洞的动力学行为可以用这个膜来模拟,并且这个假想的膜的动力学完全满足Navier-Stokes方程^[38].

5.2 引力的呈展性质

黑洞热力学除了启发引力的全息性以外,还启发了引力的呈展性,这两者之间的区别十分微妙,仅仅在于是否将引力作为一种基本相互作用力来看待.引力的呈展性认为,爱因斯坦场方程本身就不是描述一个基本相互作用的经典理论,而是描述时空热力学状态的态方程.这种思路否定了引力现象是来自于单一的基本相互作用的思路,也指出了为什么引力无法像其他基本相互作用那样量子化.

将引力现象归结为一种热力学效应的一个突破性进展发生在1995年, Jacobson^[39]首次利用热力学的基本关系 $\delta Q = TdS$ 以及黑洞熵正比于面积的假设推导出了作为热力学态方程而存在的爱因斯坦场方程. 不过Jacobson的推导需要假设系统处于一准静态, 因此无法适用于时空是远离平衡态的情况. 后来Cai和Kim^[40]将热力学第一定律应用到膨胀宇宙的表观视界上, 并且由此得到了宇宙学中描述宇宙动力学演化FLRW方程. 这一方法克服了Jacobson的方法不能应用于远离平衡态的缺陷, 并发现它也同样适用于一些修改引力理论, 如Gauss-Bonnet引力理论和Lovelock引力理论. Kothawala等人^[41]研究了一般的含时演化的视界上黑洞热力学和引力动力学的关系, 发现在这种情况下爱因斯坦场方程也可以写成一个热力学态方程的形式.

2011年, Verlinde^[42]通过对引力的全息原理中熵的思考提出了第一个引力的呈展理论: 引力是一种熵力, 或者更确切地说, 引力是一种绝热反作用力. 在这种观点下, 引力不再是一种基本的相互作用, 而是一种宏观力. 这种宏观力来自于熵力, 和压强、弹力等起源相似; 物体的质量和惯性也只是宏观表现. 引力现象是由于两个物体之间全息表面的熵发生改变导致系统能量发生变化而引起的. 在这种解释中全息原理才是根本的原理. 广义相对论是微观相互作用在宏观上所呈展出来的一个表现. 在Verlinde于2011年提出引力是一种熵力的观点之后, 由于没有具体理论模型的指引, 这个想法在经历了昙花一现的、如同超新星爆发般的关注之后, 逐渐趋于平淡乃至到了无人问津的地步. 但是在2017年, Verlinde^[43]再次发表文章进一步发展了他的想法, 并提出在星系尺度上的MOND也可以从更加基本的概念衍生呈展出来. Verlinde认为, 反德西特空间中的量子比特处于短程纠缠状态, 其纠缠熵满足面积律, 但是德西特空间中的量子比特除了满足面积律的短程纠缠外, 还有热化的长程纠缠, 这些长程纠缠的尺度在哈勃尺度上且对总的纠缠熵的贡献满足体积律. 在哈勃视界内, 纠缠熵是面积律占主导, 而在哈勃视界外, 纠缠熵则是体积律占主导. 当加入重子物质时, 德西特空间总的熵会因此减少, 而熵的变化对应自由能的变化, 即会引起对重子物质的反作用力, 其表现形式即为修改牛顿引力的形式.

5.3 引力与量子信息

前文介绍的全息原理在非引力系统的应用主要集中在系统的局域(local)性质上, 但是在强关联系统和量子信息论中, 非局域的量子纠缠(quantum entanglement)性质也扮演着重要的角色, 这其中纠缠熵(entanglement entropy)作为一个重要的非局域量在很多领域有着重要的应用. AdS/CFT对偶给出了一个利用渐进AdS时空几何来计算边界共形场论纠缠熵^[44, 45]的全息纠缠熵公式: 边界共形场论的一个子系统的纠缠熵对偶于AdS时空所对应的极小曲面的面积. 它的一个重要的意义在于暗示了对偶于一个共形场论的时空几何的许多性质被编码于边界态的纠缠结构之中. 原则上可以通过场论来计算边界许多不同区域所对偶的态的纠缠熵, 并寻找一个能够与这些纠缠熵相互匹配的时空几何来恢复出对偶时空内部的几何结构.

利用量子纠缠重构时空几何的这种观点实际上导致了另外一种引力的呈展理论, 即引力起源于量子纠缠. 这个想法来自于Maldacena和Susskind^[46]基于对黑洞火墙模型的研究而提出的ER=EPR猜想. ER原本指的是施瓦西时空中的爱因斯坦-罗森桥, 这里代指一切虫洞; 而EPR指的是量子力学中的EPR佯谬, 这里代指量子纠缠. Maldacena和Susskind认为是量子纠缠创造了虫洞而虫洞连接着两个处于纠缠态的粒子. 这个猜想的进一步研究导致了人们去猜测正是量子纠缠导致了时空及其动力学结构的出现.

在量子纠缠导致时空与引力出现的理论框架中, 时空的几何被认为是微观量子态的纠缠结构的表现, 物质引起了量子纠缠变化从而表现出引力效应. 这一个理论的有力证据来自于线性化的爱因斯坦引力方程, 可以从一般的量子信息原理得到^[47]. 另外真空的时空结构所给出的纠缠熵遵循面积律这一事实启发了物理学家尝试将时空表述为量子信息纠缠单元所组成的张量网络(tensor network), 这其中第1个方案叫做MERA方案^[48]: 利用多尺度纠缠重整化步骤构建边界的纠缠态. 边界内部的时空还可以作为一种全息纠错码^[49], 在这个方案中, 呈展出时空几何的张量网络因为量子纠缠引起了一个从内部到边界的幺正变换. 张量网络和量子纠错码这两个概念提供有用的视角去深

入思考时空的纠缠结构,特别地,呈展出的时空的微观结构应当被认为是量子信息的基本单元. 这些单元的短程纠缠应该给出全息熵的面积定律,并将整个时空连接起来. 最近有研究表明,黑洞的量子性质和量子信息论的计算复杂度有着密切的联系,并尝试利用复杂度这一概念来重新理解引力系统^[50,51].

6 结语

引力作为最早关注和最容易感知的相互作用力,

自近代物理学诞生开始、历经三百多年的研究,已经获得了非常深刻的认识. 人们已经能够从理论上解释绝大多数所能观测到的现象. 然而在理论上,它与量子理论的不兼容则预示着现有的理论还没有揭示出其本质. 宇宙加速膨胀和星系旋转曲线等观测结果更对现有的引力理论和基本物理学理论提出了现实的挑战. 在经历了引力本质认识上的三次飞跃后,我们正站在一个变革的时代的夜前,应该抓住机遇,迎难而上,努力做出属于新时代的成就.

参考文献

- Milgrom M. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. *Astrophys J*, 1983, 270: 365–370
- Milgrom M. A modification of the Newtonian dynamics: Implications for galaxies. *Astrophys J*, 1983, 270: 371–383
- Milgrom M. A modification of the Newtonian dynamics: Implications for galaxy systems. *Astrophys J*, 1983, 270: 384–389
- Chesler P M, Loeb A. Constraining relativistic generalizations of modified Newtonian dynamics with gravitational waves. *Phys Rev Lett*, 2017, 119: 031102
- Guth A H. The inflationary universe: A possible solution to the horizon and flatness problems. *Phys Rev*, 1981, D23: 347–356
- Abbott B P, Abbott R, Abbott T D, et al. Observation of gravitational waves from a binary black hole merger. *Phys Rev Lett*, 2016, 116: 061102
- Abbott B P, Abbott R, Abbott T D, et al. GW170817: Observation of gravitational waves from a binary Neutron star inspiral. *Phys Rev Lett*, 2017, 119: 161101
- Cai R G, Cao Z, Guo Z K, et al. The gravitational-wave physics. *Nat Sci Rev*, 2017, 4: 687–706
- Clifton T, Ferreira P G, Padilla A, et al. Modified gravity and cosmology. *Phys Rep*, 2012, 513: 1–189
- Arkani-Hamed N, Dimopoulos S, Dvali G R. The Hierarchy problem and new dimensions at a millimeter. *Phys Lett*, 1998, B429: 263–272
- Randall L, Sundrum R. A large mass hierarchy from a small extra dimension. *Phys Rev Lett*, 1999, 83: 3370–3373
- Randall L, Sundrum R. An alternative to compactification. *Phys Rev Lett*, 1999, 83: 4690–4693
- Dvali G R, Gabadadze G, Porrati M. 4-D gravity on a brane in 5-D Minkowski space. *Phys Lett*, 2000, B485: 208–214
- Fang W, Wang S, Hu W, et al. Challenges to the DGP model from Horizon-scale growth and geometry. *Phys Rev*, 2008, D78: 103509
- Hawking S W, Ellis G F R. *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press, 1975
- Penrose R. Gravitational collapse: The role of general relativity. *Riv Nuovo Cim*, 1969, 1: 252–276
- Bekenstein J D. Black holes and entropy. *Phys Rev D*, 1973, 7: 2333–2346
- Hawking S W. Black holes in general relativity. *Commun Math Phys*, 1972, 25: 152–166
- Bardeen J, Carter B, Hawking S W. The four laws of black hole mechanics. *Commun Math Phys*, 1973, 31: 161–170
- Hawking S W. Black hole explosions. *Nature*, 1974, 248: 30–31
- Ashtekar A. New variables for classical and quantum gravity. *Phys Rev Lett*, 1986, 57: 2244–2247
- Rovelli C, Smolin L. Discreteness of area and volume in quantum gravity. *Nucl Phys*, 1995, B442: 593–622
- Rovelli C. Black hole entropy from loop quantum gravity. *Phys Rev Lett*, 1996, 77: 3288–3291
- Witten E. (2+1)-dimensional gravity as an exactly soluble system. *Nucl Phys*, 1988, B311: 46
- Witten E. Three-dimensional gravity revisited. 2007, arXiv: 0706.3359
- Jr Callan C G, Giddings S B, Harvey J A, et al. Evanescent black holes. *Phys Rev*, 1992, D45: R1005
- Maldacena J, Stanford D. Remarks on the Sachdev-Ye-Kitaev model. *Phys Rev*, 2016, D94: 106002
- Horava P. Spectral dimension of the universe in quantum gravity at a lifshitz point. *Phys Rev Lett*, 2009, 102: 161301
- 't Hooft G. Dimensional reduction in quantum gravity. In: *Salam-festschrift*. Singapore: World Scientific, 1993. 284–296
- Susskind L. The world as a hologram. *J Math Phys*, 1995, 36: 6377–6396
- Witten E. Anti-de Sitter space and holography. *Adv Theor Math Phys*, 1998, 2: 253–291
- Gubser S S, Klebanov I R, Polyakov A M. Gauge theory correlators from noncritical string theory. *Phys Lett*, 1998, B428: 105–114

-
- 33 Kovtun P, Son D T, Starinets A O. Viscosity in strongly interacting quantum field theories from black hole physics. *Phys Rev Lett*, 2005, 94: 111601
- 34 Hartnoll S A, Herzog C P, Horowitz G T. Building a holographic superconductor. *Phys Rev Lett*, 2008, 101: 031601
- 35 Policastro G, Son D T, Starinets A O. From AdS/CFT correspondence to hydrodynamics. *J High Energy Phys*, 2002, 9: 43
- 36 Bhattacharyya S, Hubeny V E, Minwalla S, et al. Nonlinear fluid dynamics from gravity. *J High Energy Phys*, 2008, 2: 45
- 37 Bredberg I, Keeler C, Lysov V, et al. From Navier-Stokes to Einstein. *J High Energy Phys*, 2012, 7: 146
- 38 Price R H, Thorne K S. Membrane viewpoint on black holes: Properties and evolution of the stretched horizon. *Phys Rev*, 1986, D33: 915–941
- 39 Jacobson T. Thermodynamics of space-time: The Einstein equation of state. *Phys Rev Lett*, 1995, 75: 1260–1263
- 40 Cai R G, Kim S P. First law of thermodynamics and Friedmann equations of Friedmann-Robertson-Walker universe. *J High Energy Phys*, 2005, 2: 50
- 41 Kothawala D, Sarkar S, Padmanabhan T. Einstein's equations as a thermodynamic identity: The cases of stationary axisymmetric horizons and evolving spherically symmetric horizons. *Phys Lett*, 2007, B652: 338–342
- 42 Verlinde E P. On the origin of gravity and the laws of Newton. *J High Energy Phys*, 2011, 4: 29
- 43 Verlinde E P. Emergent gravity and the dark universe. *SciPost Phys*, 2017, 2: 16
- 44 Ryu S, Takayanagi T. Holographic derivation of entanglement entropy from AdS/CFT. *Phys Rev Lett*, 2006, 96: 181602
- 45 Nishioka T, Ryu S, Takayanagi T. Holographic entanglement entropy: An overview. *J Phys*, 2009, A42: 504008
- 46 Maldacena J, Susskind L. Cool horizons for entangled black holes. *Fortsch Phys*, 2013, 61: 781–811
- 47 Jacobson T. Entanglement equilibrium and the Einstein equation. *Phys Rev Lett*, 2016, 116: 201101
- 48 Evenbly G, Vidal G. Tensor network renormalization yields the multiscale entanglement renormalization ansatz. *Phys Rev Lett*, 2015, 115: 200401
- 49 Pastawski F, Yoshida B, Harlow D, et al. Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence. *J High Energy Phys*, 2015, 6: 149
- 50 Susskind L. Entanglement is not enough. *Fortsch Phys*, 2016, 64: 49–71
- 51 Brown A R, Roberts D A, Susskind L, et al. Holographic complexity equals bulk action? *Phys Rev Lett*, 2016, 116: 191301



蔡荣根

1995 年 6 月获得复旦大学理学博士学位。相继于中国科学院理论物理研究所、韩国汉城国立大学理论物理中心、日本大阪大学物理系从事理论物理研究。2000 年底入选中国科学院“引进国外杰出人才计划”。现任中国科学院理论物理研究所研究员, 博士生导师, 中国引力和相对论天体物理学会理事长, 国际广义相对论和引力学会理事, 亚太物理学会引力、天体物理和宇宙学分会副理事长。2011 年获得国家自然科学二等奖。长期从事引力理论、宇宙学、量子场论和超弦理论等方面的研究。

Summary for “引力的本质”

What is gravity?

Ronggen Cai^{1*}, Shaojiang Wang¹, Runqiu Yang² & Yunlong Zhang³¹ *Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China;*² *Korea Institute for Advanced Study, Seoul 130-722, Korea;*³ *Asia Pacific Center for Theoretical Physics, Pohang 790-784, Korea*

* Corresponding author, E-mail: cairg@itp.ac.cn

We review the historical pursuits and recent developments on the origin of gravity. We start with the Newtonian gravity and Einstein gravity to introduce the two major leaps of the historical pursuits on the origin of gravity. Then we introduce the recent developments on the origin of gravity from the point of view of modified gravity, quantum gravity and holographic gravity. For Newtonian gravity, we start with Kepler's law on planetary motion to introduce the Newton's universal law of gravitation, and recent developments on the modified Newtonian dynamics and dark matter; for Einstein gravity, we explain the geometrization of gravity, and introduce the application of Einstein gravity on cosmology and gravitational waves; for modified gravity, we introduce it from the point of view of extra gravitational degrees of freedom, higher derivative gravity and higher dimensional gravity; for quantum gravity, we introduce it from the point of view of covariant quantum gravity, canonical quantum gravity and other scenarios of quantum gravity; for holographic gravity, we introduce its holographic picture, emergent properties and its connection with quantum information. However, the answer to the question of gravity origin still remains unknown. We point out here that three possible roads to approach the nature of gravity can be taken in future. The first one is the coming observations of gravitational waves from both ground-borne and space-borne gravitational-wave detectors. A system involving either black hole or neutron star has certainly run into the regime of strong gravity, which is quite different from previous tests for general relativity that is in the regime of weak gravity. The gravitational-wave observation will most likely provide us with clues on classical gravity with any possible deviations if Einstein gravity does modify itself at classical level. The second one is the understanding of the cosmic acceleration at both early-time and late-time. The cosmic acceleration at early-time is supposed to be driven by cosmic inflation, and the cosmic acceleration at late-time is supposed to be driven by the cosmological constant or alternatively dark energy. Future precise observations of radiation fluctuations and matter fluctuations will show us the way approach the nature of quantum gravity. The third one is the cognition of the relation of quantum gravity and quantum information. The past fifty years or so of pursuit on quantum gravity has led us to the path on the conjecture that quantum gravity might be emerged from quantum information. Although the perspective of this idea is not clear, it might be our best shot for now.

gravity, Newtonian gravity, Einstein gravity, modified gravity, quantum gravity, holographic gravity

doi: 10.1360/N972018-00035