【传教】《微分几何入门与广义相对论》梁灿彬、周彬

本帖子是我在阅读这本书过程中的一些摘抄，做了一些稍微亲民的翻译，内容文字和讲解顺序等与原文有出入.

阅读并理解本帖内容需要至少在高中数学和物理必修或以上的水平，小白勿入.

有问题请在楼中楼或者单开帖子提问，我会尽力解说.请尽量避免无意义插楼，谢谢合作.

粒子分为两类：第一类是有(静)质量的粒子，与质点同义；第二类是无(静)质量的粒子，为方便也常称为光子.狭义相对论的原始表述存在静质量m0，静能E0，运动质量m和总能E四个概念.然而，关系式E=mc^2和E0=m0c^2表明这四个概念中只有两个独立.事实上，近代文献(科普文献除外)中通常只保留质量和能量两个概念，分别为静质量和总能.

物理中经常提到守恒量和不变量，在这里做一个区分：守恒量是指在物理过程中不变的物理量，比如能量等是守恒量而非不变量；不变量是在参考系或坐标系变换等人为变动下不变的量，比如静质量等是不变量而非守恒量.粒子的电荷量等既是守恒量也是不变量.

“事件”在物理学中的概念是实际事件的模型化，不论是否发生了什么，空间的一点和时间的一瞬的结合就叫一个事件(event).全部事件的集合就是时空(spacetime)，故每个事件也叫一个时空点.将时空这种提法准确化的数学概念是“四维流形”.大多物理学都默认一个基本出发点(基本假设)，即把时空看作四维流形并配以适当的附加结构.非相对论和狭义相对论物理学都假定时空流形是平直的，区别仅在于附加结构的不同，而广义相对论物理学则允许时空为任意四维连通流形(可以发生所谓“时空扭曲”).

狭义相对论的时空是4维闵氏时空，对应数学中的4维闵氏空间.4维闵氏时空有四个相互正交的轴，分别为txyz轴，每一个轴和其他任意轴相互正交，和其他任意两个轴确定的“平面”正交，也和其他任意三个轴确定的“空间”正交.4维闵氏时空的度规为洛伦兹度规，矩阵[-1,1,1,1](为了简化说明，矩阵的性质等暂不深究).非相对论物理学中，一般认为时空为3维，或称(3+1)维，因为其时空中存在一个绝对的时间维度，与相对论的闵氏时空相比较，这里的时间维与其它三个维度没有那么紧密的联系.这个时空称为欧式时空或者欧式空间，度规为欧式度规，矩阵[1,1,1,1].当然，欧式空间这个名词在相对论中也可以使用.比如，在相对论的3维表述中，粒子所在的某时刻空间完全可以称为欧式空间.

注意！下述部分内容在原书中采用几何单位制，光速取c=1.读者请注意区分.

(本段见图①)相对论物理学中经常使用一种时空图(spacetime diagram)，类似空间坐标系.但时空图要求必须有一个时间轴t轴，通常画成竖直向上的.一般简单问题中，粒子在三个空间方向xyz上并非都有运动，比如匀速直线运动只涉及一个空间维，平原上的汽车只涉及两个空间维，所以可以把不涉及到的运动方向在画图时省略，这样画出的时空图就剩下两个或三个轴.一个粒子的全部历史由一系列时空点组成，因此对应时空中的一条曲线，称为该粒子的世界线(world line).将世界线画在时空图上(此处使用txy图)后，由于竖直向上代表时间的流逝，水平方向代表空间，每个水平面代表某一时刻的全部空间，从时空图底部向上看，就能看到运动(演化)的全过程.

狭义相对论中的时空图有一些特殊性质需要注意.

(本段见图②)因为光速取1，所以任何粒子的速率均为小于1的.(此处使用tx图)设有一过坐标原点的粒子M的世界线，其在时空图中的斜率应当介于1到无限大之间，即相对x轴更贴近t轴.由后文式①可知，该粒子的世界线斜率与其速率互为倒数，即k=1/u，此世界线的参数方程就是{t=a/u,x=a,y=0,z=0}(a为参数)，直线的切矢方向同(1/u,1,0,0).

现在，设有一直线L的参数方程是{t=b\*u,x=b,y=0,z=0}(b为参数)，切矢方向同(u,1,0,0).我们将两个切矢相乘.看似算式应为(1/u)\*u+1\*1=2，但注意到洛伦兹度规矩阵为[-1,1,1,1]，所以算式应为-(1/u)\*u+1\*1=0.如此可知，两切矢在洛伦兹度规下为正交切矢，这样，我们所列出方程的两个直线在洛伦兹度规下是正交的.

注意到在任何一个惯性坐标系中，时间轴总是和x轴正交，并且，每个做惯性运动的粒子，其世界线都是自己参考系中的时间轴.那么可以知道，上述两个方程所代表的直线中，第二个直线的方程应当代表粒子M的x轴.实际上，第二条直线在txy图中可以代表这样一个面：这个平面是粒子M参考系空间中的xy平面.因为时空图压缩掉了z轴，所以图上的这个平面还可以表示粒子M在该时刻所处的空间.由此，我们可以将L记做x'，将M记做t'，这便构成了另一个惯性参考系.

设某曲线L为粒子的世界线，p,q分别为L上两相邻点，(t1,x1,y1,z1),(t2,x2,y2,z2)分别为它们在某惯性坐标系中坐标.令dt=t1-t2，对dx,dy,dz作相同处理，则粒子在p点(或说q点也可)时相对该参考系的速率定义为

u=[(dx^2+dy^2+dz^2)^(1/2)]/dt ①

(接触过微积分知识的读者应能看出式中各符号的准确含义)根据狭义相对论，描述p,q联系的重要物理量是元间隔s，元间隔可借用任一惯性系定义为

ds^2=-dt^2+dx^2+dy^2+dz^2 ②

联立①②可得

ds^2=-(1-u^2)\*dt^2③

对于任一标准钟来说，在其世界线上任意两点间读数差等于该世界线在这两点间世界线长(这就是标准钟的定义：可以给出世界线长度的钟即为标准钟)，即dτ^2=(1-u^2)\*dt^2.注意！虽然τ和t看似都定义为“时间”，但实际上并不相同.

(本段见图③)为简化讨论，我们将世界线均取直线，这对应做匀速直线运动的粒子.非匀速运动中需对变量作些处理，但基本思路相同，此处不赘述.当我们选定一惯性系后，设有两粒子世界线A,B，其中A保持静止，B始终做匀速直线运动，则线A重合于t轴，线B与t轴有一夹角.对于粒子A来说，由于dx=dy=dz恒=0 ,所以对任意两点pA1,pA2，均有τA1-τA2=tA1-tA2.由于A是本参考系中的惯性观者，所以可以直接把t轴的读数(任意粒子的t都可作如此计算)差作为该参考系的时间间隔，相应地，τ的差则是该世界线对应粒子认为自己所经历的时间间隔.对于粒子B，τB1-τB2<tB1-tB2，图上还表示为|A1A2|>|B1B2|(注意洛伦兹度规下线长与线段直观长度的关系).这种现象就是钟慢效应，即B认为自己经历的时间τ是小于参考系/粒子A所经历的时间t的.我们将τ称为固有时(proper time)，将t称为坐标时(coordinate time)，惯性系的坐标时又可称为惯性坐标时.注意！固有时在世界线外无定义，而且当两世界线相交时，两固有时可以有不同读数.同时，坐标时的定义实际上不依赖于粒子世界线，唯一的要求是该点处于本时空之内.

注意！上段的钟慢效应有部分细节未交代，详细解说在后文中会出现.

(本段见图④)时空图中的所有曲线分成三类：类时(timelike)线，类光(lightlike)线，类空(spacelike)线.这三种线的区别在于斜率取值范围的不同.类光线的长度永远是0，斜率为1，画在图上是x轴和t轴(yt或zt也是如此)的平分线；类时线的斜率始终大于1，较xyz轴更贴近t轴，在式②中使得式右端为负值，实际上在计算线长的时候已默认式右端取绝对值；类空线的斜率始终小于1，较为贴近xyz轴，式②中使得右端为正值.由于类空线的斜率小于1，若某粒子沿其运动的速率将大于光速，所以类空线不是任何粒子的世界线，而类时线代表的速率小于1，可以代表质点的世界线，类光线则只能代表光子的世界线.当然在画图的时候也存在斜率在1上下不定的曲线，这种曲线不予考虑其性质(因其线长无法计算且无实际的物理意义).

就式②中的ds^2而言，人们在学习几何语言时往往遇到这样的问题，那就是：一条曲线的长度应是实数值，因为时空坐标都是实数，而根据式②所得出的ds对类时线必为虚数，这岂非有些不能令人接受？我们对此的回答是：其实式中的ds并非严格定义中的线长(或说其微分)，而仅仅是一个记号.在研究几何语言的历史上，数学家们最先接触的是欧氏几何，对于四维欧式几何来说线段长的定义自然是

dl^2=dt^2+dx^2+dy^2+dz^2

在将上式推广到四维闵氏几何时，最简单的做法无非是把l(称为线段长，或称欧式几何的元段长)换为s(元间隔，也可称为闵氏几何的元段长)，再把右边的dt^2项前添一负号(以保证式具有洛伦兹协变性).此时式左ds的意义已经不是元段长的微分，而仅是一个符号，写成ds^2只是为了形象化，方便同欧式几何线元表达式进行对比而已.

鉴于数学水平所限，这里只简单提及一下曲线的大致性质.虽然我在讲述三种曲线的时候用到了坐标系，但实际上在相对论时空中，三种曲线的定义并不依赖坐标甚至不依赖观者.相对论时空中，这三种曲线的分类是绝对的而非相对，线长也都是绝对量.

由于线长公式具有洛伦兹协变性，可知世界线上任意两点间线长在不同参考系看来有相同线长.这也例证了如下说法：相对论中并非所有物理量均为相对的.实际上，相对论中有很多量都是绝对量，与参考系的选择无关.当使用几何语言明确了有关相对与绝对的区别后，某些佯谬问题便迎刃而解.

式②在时空图中还有一个重要应用：将参考系中xyz空间中的任一点坐标带入式②，发现坐标的dt恒=0，又由于该点和坐标原点的连线为直线，式②退化为

s^2=x^2+y^2+z^2③

这正是空间中两点距离的公式(指所取点和原点的距离).所以可知，四维惯性系中的xyz空间实际上就是该时刻的欧式空间.

补充几条说明(注意！以下所有陈述在狭义相对论/平直时空中成立，有些在拓展到广义相对论/弯曲时空时仍可成立，暂不标注)：在所有观者中存在一类特殊观者，称为惯性观者(inertial observer)，他们的世界线是平行于其所在惯性系的t轴的，这些观者与其他(非惯性)观者有本质上的区别.且惯性观者之间是平权的，即不存在某惯性观者观测到的结果与其他惯性观者不同的情况(相对位置除外).

除引力外不受任何力的质点称为自由质点，其世界线必为测地线.在平直时空中(因为平直所以不存才引力，或者说引力被忽略)，测地线在直观上是直线，但不一定与t轴平行.由上，同参考系的所有惯性观者的世界线是一组平行测地线，不同参考系的惯性观者，世界线虽然不平行，但仍为测地线.

假设在平直时空中，我们取两个时空点p(0,0,0,0)和q(2,0,0,0)，这两点的直连线是测地线，线长是2.如果我们引入第三个时空点w(1,0.999,0,0)，那么可以构造一条接近折线pwq的光滑曲线，其长度接近0.可知，任意两个时空点之间的直连线是最长线而不是最短线，因为我们完全可以取得一段长度极接近0的曲线将两点联系起来(为了保证所考虑的类时线都有物理意义，它们都应是处处“光滑”的，以使它所代表的粒子世界线不会有无限大的加速度).

下面介绍一种在平直时空中的钟同步方法：大步雷达法.(本段见图⑤)

式②的退化公式式③表明，在同时面上的时空点处于该时刻的欧式空间中.为了研究物理问题方便，我们当然希望同一欧式空间中的钟是有相同读数的.实际上，在相对论中已经默认同一参考系中存在由同时时空点组成的欧式空间.该空间在图上被压缩掉一维，画出来为一个面，我们将其称为超曲面；又由于这个超曲面和系内惯性观者世界线正交，所以亦可称为正交超曲面.在平直时空中，只要空间中的全部观者在某一超曲面上的读数相等，则它们在任一平行的(即同一参考系内其他的)超曲面上的读数也将相等.因此同步操作是一劳永逸的，只要在某一时刻将所有观者钟进行同步即可.为了做到这点，我们需要用信号为系内所有观者进行沟通.

按照一般的思维进行考虑，钟同步有如下方法：观者A事先对观者B发出指令，令其在接收到A的信号时把钟调至A发出信号时的读数.实际上由于光速的限制，B在接收到A的信号时，A的钟已经比发出信号时走了一些.可见，这种方法不够完善，由这种方法获得的“钟同步”是不准的.如果存在速率无限大的信号，则可令其在零时间内完成从观者A出发到达观者B的要求，这样即可完成钟同步，依此类推，系内的所有钟都可以使用这种方法同步.可惜相对论不允许这样的信号存在，于是有必要设计一种(至少在理论上可行的)钟同步方法.

为了叙述方便，我们将“A发出信号”——事件a1、“B接收信号”——事件b、“A回收来自B的反射信号”——事件a2这三个事件记为一个“雷达三点组”.由于光速在平直时空中的绝对相等，又由于光速最大，所以我们仍然考虑使用光信号来进行同步.可以使用如下方法：观者A事先向B发出指令，“你在身上装一面反射镜，当反射镜接收到我的信号时(事件b)将你的钟调零”，这样，A发出的光信号在到达B时立即反射并最终被A回收.A的读数零点可以这样确定：发出信号的时刻(事件a1)和收回信号的时刻(事件a2)的中间时刻即为读数零点.不难看出，AB两钟的读数零点的连线与两钟世界线均正交.按照这种方法，我们可以在选定一领头观者后使其与系内的所有观者进行同步.将系内所有按这种方法获得的读数零点连线集合起来，组成的面就是该空间的超曲面，该超曲面上的所有观者钟有相同读数.这种方法的的特点是所有信号都要由领头观者发出并回收，称为大步雷达法.

对于非平直时空的钟同步可做如下考虑：因时空不平直，由大步雷达法获得的“同时面”未必与系内所有观者的世界线正交.但也不难看出，只要让每一观者都对紧邻观者执行雷达法，所得曲线就与每一观者世界线正交.准确做法如下：取一系列观者A,B,C,D,…先由A用大步雷达法在AB间确定一条起于A上点a止于B上点b的同时线，再用同样的方法确定一条起于b止于C上点c的同时线，以此类推，最后向A,B,C,D,…间添加更多观者，所得的极限曲线就是一条同时线.这一做法称为(无限)小步雷达法.

易知代表空间的超曲面上的两点之间的连线为类空线，其长度为两点对应空间点的空间距离.通过画图可以很容易看出，该类空线实际上就是a1a2中点事件和b事件的连线，其长度等于a1a2连线长度的一半.所以大步雷达法还可以起到标准尺的作用(参考标准钟的定义，标准尺的作用是可以给出类空线长度的尺).实际上月地距离就是用大步雷达法测得的.

经过上述讨论，我们应当树立起一个观念：相对论时空中存在(由参考系运动状况决定的)同时的空间.为了结合时空图时表示方便，可以将代表同时空间的超曲面称为同时面(它们都是和参考系运动状态对应的).

在牛顿力学中，时空流形被默认为四维，但其中附加的内禀结构使得牛顿力学中存在一个“绝对同时面”的概念.牛顿力学中的同时面将整个时空分为无数互相平行的层，这组同时面均正交于t轴，而相对论物理学中则有无数组同时面，每组同时面互相平行.实际上，二维时空图中的任意直类空线都确定了一组同时面，当然这组同时面是与类空线正交的类时线观者认为的同时面组.

现在我们以牛顿力学的思想，在时空中取任意一个时空点，这个点所在的同时面被认为和该时空点没有因果联系，而在同时面上方的时空则是“未来的”，下面的时空是“过去的”，全部时空的这两个部分和选定的时空点有因果关系.到了相对论中，我们可以选定时空点为顶点画出光锥(在txy图中就是两个对顶的圆锥，其任意一条母线是xy面与t轴的平分线，两个圆锥关于选定时空点所在的同时面对称)，上方的光锥内部的时空是“未来的”，下方光锥内部的时空是“过去的”，而除此之外的所有时空点(含光锥面上的所有点)与选定时空点无因果联系.可以看出，无因果联系的时空部分在相对论中比牛顿物理学中大得多.

典型效应分析：

尺缩效应：(本段见图⑥)

质点在三维语言中是空间的一个点，四维语言中是时空的一条类时线；同理，一把尺子在三维语言中是一条线，四维语言中就是由尺上各质点世界线组成的世界面.乍看在时空图中测量尺长的方法有些不明确：面上的哪条线长才是尺长？对于熟悉四维语言的人来说，尺子从来就不是一维的，它是个二维对象，这是一个绝对的对象，与参考系，坐标系及观者均无关.

为什么尺子在三维语言中却是个一维对象？因为人们习惯在自己的立场中看问题，也就是在自己的欧式空间中测量尺长，而欧式空间是随着观者不同而有所不同的，所以尺长对不同参考系的观者来说一般不等.现选定一基准坐标系，将某一尺子世界面在该坐标系下的方程表示如下(从现在开始，时空图中未涉及到的维度相关的方程将不再列出，在这些维度上的位移应默认为0)：t∈R,x∈[0,oa].显然这个尺子世界面在基准系中的长度为1.另设一运动观者，世界线参数方程为t=a/u,x=a(a为参数)，则其相对基准系速率为a，与其对应的同时面组方程之一为t=a\*u,x=a.

注意到运动观者的同时面是x'，所以其测得的尺长应为尺子世界面截得的x'上的线段ob.因为基准系认为尺子静止而运动观者认为尺子运动，则两者观测得到的尺子有不同长度，比较两尺长无非就是比较oa和ob的长度.直观看来ob>oa，似乎动尺较长！但注意到时空图使用洛伦兹度规，过点a参照点o做一校准曲线-t^2+x^2=oa^2[本例中使用到的校准曲线是人为划定的一条曲线，在洛伦兹度规下，其线上任意一点与参照点的直连线等长，尽管直观上看(即按照欧式度规来看)不等]，则发现oa=oc>ob.

为了定量计算尺缩的程度，我们用基准系中的oab三点的坐标来计算：o(0,0)，a(1,0)，b(1,u)，则可以看出ob^2=(-u^2+1)\*oa^2，开平方即得到ob=oa\*(1-u^2)^(1/2)，式子回到我们熟悉的尺缩公式.

由上述讨论我们可以得知，尺缩效应中没有什么“弹性”之类的物理机制在发生作用(根本没有东西实际发生了收缩)，其本质原因是：虽然尺子只有一把(尺子世界面只有一个)，但不同的参考系有不同的同时面，导致它们测到的是不同的一维尺，而不同的一维尺有不同的长度当然不足为怪.尺缩效应无非“盲人摸象”而已.

钟慢效应：

(本段见图⑦)与上例同样地，我们选定一基准系，其中放置两个钟A和B，AB之间距离为oa，两钟均相对基准系静止，实验前经过大步雷达对钟.另一钟C从A处运动到B，图上表示为从事件o到事件c，且C在o处也指零.现在若要观察钟慢效应，只需比较ob和oc的长度(世界线长即标准钟读数差).直观看来仍有oc>ob，但但过点b参照点o做一校准曲线，则容易看出ob=od>oc，即C钟在AB两钟所在参考系看来变慢.

钟慢的程度可以用如下方式计算：设C的世界线参数方程为t=a/u,x=a(a为参数)，容易看出c的时空坐标为(oa/u,oa)，则b坐标为(oa/u,0)，可以列出式ob^2\*(1-u^2)=oc^2，开方并移项可得oc=ob\*(1-u^2)^(1/2)，这就是我们熟悉的钟慢公式.

(本段见图⑧)因为相对论中结论的相对性，我们应当验证一下，在C钟看来AB两钟的走时情况.将上一时空图稍作修改，在其中添加上与C钟对应的两个同时面，观察此图可发现，在C钟看来AB两钟的读数零点并不同步，即在A中读数为0时B钟读数已是ab(B钟做了偷跑)，故比较BC读数时应当比较的是bd和od的长度.过点d参照点o做校准曲线交t轴于e，可以看出od=oe>oc=bd，得出的结论是相同的——AB两钟在C看来走时变慢.我们仍然可以使用AB参考系中的时空坐标来计算有关线长(线长是不变量)，得出的公式仍为oc=ob\*β(为简化文字表达，在之前文中有给出的公式中，有时会使用一些符号代替较长的表达式.读者应能自行判断).

以上讨论表明，同尺缩效应中没有任何东西发生了收缩一样，钟慢效应中也没有任何钟的走时率变慢(都坚持标准钟的走时率，即世界线长度等于读数差).

应该注意的是，不同惯性系的钟之间可以使用很多不同的比钟方式，上述方式虽然为人们所熟知，但并非是唯一的比钟方式.我们熟知的钟慢公式也只不过是上述比钟方式造成的必然结果而已.AB两钟是同一惯性系内事先经过同步的钟，如果缺少其中之一，虽然同样可以从图中明确世界线长关系，却无从得出“A觉得C钟慢”的结论，关键在于“觉得”两字无从谈起.对于图⑧，事件d不在A的世界线上，A钟对该处的C钟不能直接获得任何信息，唯一的方法只接收来自点d的光信号(或者其他信号)，而这就涉及到信号传递需要时间的问题(不是不能这样做，而是这样做时必须考虑这个问题).其实，利用B联合A得出“C钟较慢”的结论时就已经巧妙地发挥了光信号的信息传递作用，因为在把AB同步的过程中已经用到了光信号.总之，如果缺少了B钟，AC两钟就无法用上述方式比钟，或者说上述比钟方式没有意义.

(本段见图⑨)当全时空中只提供两个钟的时候也存在很有物理意义的比钟方式.取B钟所在惯性系为基准系，令AB在事件o处重合并在此同步调零，A钟相对B钟做匀速运动，图中可以清楚地看出oa=β\*ob，但B钟无法直接观测事件a，只能接收来自a处的光信号，则B在看到A读数为oa时，自身读数已经变成了oc，可知oc>ob=oa/β，即A钟在B钟看来比上一种比钟方式慢得更甚.

(本段见图⑩)我们甚至可以举出这样的比钟方式，它将导致“动钟较快”！令B钟向A钟匀速靠拢，并使两钟在b处指零，在时空图中过a参照b作校准曲线，发现曲线与B钟世界线交于点d，即ab=db<cb(对于校准曲线与B钟世界线交点d是否会处在x轴以下的疑问，注意到校准曲线的渐近线斜率为±1，不需实际写出各线的方程即可看出d点应该处在c的上方)，这代表a处的A钟通过光信号看到的B钟读数比自己的读数负值更负，A钟发现B钟比自己更快.这个看似荒谬的结论实际上无可非议，它不过是上述比钟方式的特定结果.因此，讨论比钟的问题是需要事先明确方案的每一细节，为此最好画时空图.

双子佯谬：

图中甲的世界线与t轴重合，表示甲呆在家中不动；乙的世界线为曲线/折线，它是非测地线表明其做非惯性运动.已知两人分手时年龄相等，问重逢时年龄还是否相等？若不等，孰大孰小？这无非是比较甲乙二人在“分手”“重逢”两个事件之间的世界线长度大小，因为世界线长代表固有时，也就是本人的时间流动(年龄增长).之前我们证明过闵氏时空中测地线为两点之间连线最长者，很容易得出“重逢时甲的年龄大于乙的”结论.

以上就是双子效应的实质内容.问题本来就如此简单，但是人们在相对论发展的过程中对有关问题缺乏深刻认识，双子效应在一段时间之内竟被视为悖论(paradox)，对此的争论竟迟至1957~1958年又掀高峰(虽然若干有识之士已经对此问题有清晰理解)，有关文章竟发表在《Nature》《Science》《Discovery》等重要刊物上.争论双方以物理学家McCrea和物理学家兼哲学家Dingle为代表人物.Dingle认为，相对论中一切都是相对的，因此双子重逢时应有相同年龄；McCrea针锋相对指出，相对论并不认为一切都是相对的，甲没有加速度而乙有，正是这一区别导致了重逢时年龄不同.随着研究的深入，特别是几何语言的引进，国际相对论界对双子效应已经有了如上段所述的共识.应该特别强调的是，许多人顾名思义地以为相对论中一切都是相对的，这是一种极为有害的误解.

双子效应已经在1971年被实验所证实，当然不是对人而是对铯原子钟.

下面回答几个有关双子效应的疑问：

问①：钟慢效应的结论对双方是平等的，你所举的钟慢效应前两图中得出的结论就是AB和C互相认为对方比自己慢，为什么双子效应对甲乙两人的结论就不平等？

答①：因为两种效应的前提不同.在钟慢效应中，ABC三钟都做惯性运动，由于惯性系平权，结论自然对双方平等，但在双子效应中乙不做惯性运动(世界线非测地线)，否则分手之后不会重逢.这个前提本身就确立了双方的不平等地位，结论自然是一边倒的.

问②：双子效应的结论有这样的表述“做加速运动的兄弟较为年轻”，但加速度是相对的，甲乙互相认为对方又加速度，这样乙是否应当认为甲更年轻？

答②：加速度有三维加速度(3加速)与四维加速度(4加速)之分，前者是相对的而后者是绝对的(与观者、参考系、坐标系等人为选择的因素无关).而惯性运动和非惯性运动的概念也都是绝对的：粒子做惯性运动，当且仅当其世界线为测地线(仍与人为因素无关).当把“加速运动”作为“非惯性运动”的同义词时，应当把“加速”理解为4加速.因此问②中的表述应当改为“有4加速的兄弟较为年轻”.无论什么观者看甲都不会说甲有4加速，因此不再出现问题.物理学家早已形成习惯，在3维语言中提到加速度而不指明相对的参考系时，都默认相对惯性系.在这种默契下，“做加速运动的兄弟较为年轻”和“电荷当且仅当在加速运动时产生辐射”等说法就都是正确的.

问③：常听说双子效应属于广义相对论范畴，只用狭义相对论讲不清楚，是吗？

答③：不是的，本段在一开始的时候就已经使用狭义相对论进行了清晰的讲述.认为双子效应涉及广相的一个原因是：有些人为了计算乙的固有时而选择了乙自己为观者的参考系，这个参考系是非惯性系，而他们以为只要涉及到非惯性系就属于广相范畴.物理学家对此的回答是：乙的固有时就是其世界线长，这是个与参考系无关的量，根本没必要自找麻烦地用非惯性系计算；退一万步说，就算有人愿意用非惯性系来计算，这同样与广相无关.应当明确广相与狭义论的划界标准：起初人们爱用参考系来划界，认为只要涉及非惯性系就涉及广相，后来认识到用绝对的(与人为因素无关的)时空几何来划界会自然(而且优雅)得多.现在国际物理学界的统一标准是：凡以闵氏时空为背景时空的物理问题都属于狭义论范畴，而广相的时空必然是存在弯曲的.讨论任何问题时，一个非常重要但却往往被人忽视的步骤是事先约定背景时空，双子效应的背景时空默认为闵氏时空，因此属于狭义论范畴而非广相(除非约定时空不是平直的，这等价于引力场不可忽略).不幸的是有些人甚至走得更远，他们在用几何语言分析问题时发现乙参考系的一个特征量(克氏符)非全为零，便认为加速运动可以造成时空弯曲，其实闵氏度规在非惯性系的克氏符非零很正常(当然惯性系中的克氏符必然为零).另一个类似的热门问题是“爱因斯坦转盘”，也常被误认为涉及广相，其实在讨论该问题时也已经默认整个现象发生在闵氏时空中，因此也属于狭相范畴.几何语言分析这个问题是十分简洁的，但它比双子效应复杂.后文会有解释.

车库佯谬：

现有静长均为l的汽车和车库，汽车匀速向车库中驶入.司机看来，动库变短，车放不下，司库看来，动车变短，放下有余.司机的想法对吗？司库的想法对吗？用时空图可以明确地表示出这种物理情形.

现在假设车库并无后墙，其“后墙”只是一条画在地上的直线.图中oa为车静长，ob为司机看到的库长；oc为库静长，od为司库看到的车长.很容易看出，司库的想法是对的，司机的想法也是对的(都以自己的参考系为参考)，关键是同时的相对性导致结论的相对性，正如尺缩效应中不允许问“到底哪一把尺子较长”一样(尺子都是同一把尺子/同一个尺子世界面)，车库佯谬中也不允许问这样的问题“到底放得下还是放不下”，结论的相对性导致这种绝对化的问题没有意义.

车库有坚硬后墙的情形则要麻烦一些，基本原则是(假设后墙和车本身都是绝对刚性的)：车头撞墙(因而停止前进)的信息传递到车尾需要时间，只有当车尾获悉此信息之后才能停止前进，因此汽车将被压缩到的确在库中装下有余的程度(谁看都装得下).有兴趣的读者不妨画出这种情形的时空图，试着求一下车被压缩后的新长.

鉴于狭义相对论中动量、能量和质量概念的重要性和微妙性，有必要在此明确一遍其定义过程.

相对性原理要求一切物理定律在不同惯性系中有相同形式.惯性系间的坐标变换在牛顿力学中是伽利略变换，在狭相中是洛伦兹变换，因此，相对性原理在牛顿力学中要求物理定律的数学表达式在伽利略变换下不变(即具有伽利略协变性)，在狭相中则要求表达式在洛伦兹变换下不变(即具有洛伦兹协变性).这是一个很强的“管定律的定律”，凡不具备洛伦兹协变性的物理定律在纳入狭相前都必须被修改，以使其具有洛伦兹协变性.动量守恒律就是典型一例.在牛顿力学中，质点的动量→p(注:今后凡矢量都以在字母前加注→箭头来表示)定义为其质量与速度的乘积，即→p=m\*→u，力则定义为质点的动量时变率，即→f=d→p/dt，两者结合即得→f=m\*d→u/dt=m\*→a.可见，→f=m\*→a虽被称为牛顿第二定律，其实只是力的定义式，它只有与每一具体的物理场合下力的表达式相结合才给出真正的物理定律(如弹簧体系下与→f=-k\*→x结合得d→p/dt=-k\*→x，这才是本质上的胡克定律).

现在从相对性原理的角度审查动量守恒律，在这里为简化讨论而使用两等质量小球发生完全非弹性正碰的情况.以→p1和→f1分别代表球1的动量和碰撞中受力(这一力来自球2)，则→f1=d→p1/dt，类似地有→f2=d→p2/dt.牛顿第三定律保证→f1=-→f2，于是d(→p1+→p2)/dt=0，即碰撞过程中动量守恒.可见动量守恒律是力的定义同牛顿第三定律结合的产物.如果从另一惯性系观测同一碰撞过程，则根据伽利略变换的公式仍可推出动量守恒表达式，所以动量守恒律具有伽利略协变性而满足相对性原理.然而若在狭相中仍使用动量的牛顿定义→p=m\*→u(牛顿力学认为质量守恒，因此此处m为一常数)，则下例足以说明动量守恒律不具洛伦兹协变性:设在参考系A中两球碰前速度等值反向(总动量为0)，由对称性可知碰后两球粘合体速度为0，这表明动量守恒在A系中成立；设在参考系B中碰前球2静止，则AB两系的相对速度即碰前球1和B系的相对速度→v，由相对论速度变换公式可得球1在B系的速率为(此处暂不令c=1而使表达式中仍含c):

u=(v+v)/[1+(v/c)^2]=2v/[1+(v/c)^2]④

因两球质量均为m，则B系中测得两球总动量在碰撞前后分别为mu+0=2mv/[1+(v/c)^2]，2mv(后者中体现了牛顿质量守恒律)，可见A中成立的动量守恒律在B中不成立，因而不是定律.这时候有两种做法：一是放弃动量守恒律，二是修改动量守恒律以使其具有洛伦兹协变性.鉴于守恒律对物理学的重要性，优先考虑第二条.不妨这样猜想：在物体由静止开始加速的过程中，物体的质量会逐渐增大，这样物体在恒力下的加速度将会越来越小，速度就有望永远达不到光速.这是一种出于物理理论自恰性的考虑.于是想到这样的修改方案：动量仍定义为质量乘以速度，只不过此时质量m是速率u的函数，记作mu(暂时称为运动质量).现在沿这一思路重新审查B系中的碰撞过程.既然碰前球2静止而球1以速率u运动，两者的质量就应分别为m0(暂时称为静质量)和mu，因此碰撞前后总动量分别为mu\*u+0=2mu\*v/[1+(v/c)^2]，Mv\*v，其中Mv代表碰后两球复合体的质量，默认Mv=mu+m0即碰撞前后总质量不变(这是很自然的默认，其含义将在后文阐明)，则碰后总动量可写作(mu+m0)\*v.对比此时的碰撞前后动量表达式可知，为使动量守恒律对B系成立，必须且只需使mu=m0\*{[1+(v/c)^2]/[1-(v/c)^2]}，将此表达式会同式④化简即可得出

mu=m0\*[1-(u/c)^2]^(-1/2)⑤

可见只有承认mu随速率u按式⑤变化才能保证B系中的动量守恒.所以狭相中质点的动量应定义为→p=mu\*→u，其中mu由式⑤给出.通常记

γ=[1-(u/c)^2]^(-1/2)

故动量也可表为

→p=γ\*m0\*→u⑥

有了动量的定义就可对力定义.狭相中仍把力→f定义为动量时变率，即

→f=d→p/dt⑦

相对性原理要求式⑦有洛伦兹协变性，依此可导出力在惯性系之间的变换式(此处略).

现在开始推导质点的能量定义.先仿照牛顿力学按以下两条标准定义质点的动能Ek：Ⅰ质点静止(u=0)时动能Ek=0；Ⅱ动能时变率等于力的功率→f\*→u，由此得dEk/dt=→f\*→u=(d→p/dt)\*→u=→u\*d(mu\*→u)/dt=mu\*→u\*(d→u/dt)+→u\*→u\*(dmu/dt)=mu\*u(du/dt)+u^2\*(dmu/dt)，其中dmu/dt可联立式⑤改写为dmu/dt=[mu\*u/(c^2-u^2)]\*(du/dt)，代回上式有dEk/dt=(c^2-u^2)\*(dmu/dt)+u^2\*(dmu/dt)=c^2\*(dmu/dt).注意到u=0时mu=m0及Ek=0，对上式积分便得到速率为u时的动能表达式：Ek(u)=mu\*c^2-m0\*c^2.爱因斯坦大胆地将上式中mu\*c^2解释为质点在速率u时的总能(E=m\*c^2，此处m是mu的简写)，于是m0\*c^2就是质点静止时的能量(E0=m0\*c^2，即质点静能)，而动能则是总能与静能之差.E=m\*c^2表明能量E和运动质量mu成正比(称为质能相当性).几何单位制中c=1，故E=m，即能量等于质量，而E0=m0则表明物体即使在静止时也有等于静质量的能量.这是一份不可思议的巨大能量.

牛顿力学既有质量守恒又有能量守恒，狭义相对论的情况如何？首先，按E=m\*c^2定义的能量满足能量守恒，这应看做理论假设，但这也取得了迄今为止所有实验的支持.至于质量守恒，则要看你谈的是运动质量m还是静质量m0.因为E=m\*c^2，所以能量E守恒就等价于m运动质量守恒，两者互不独立[但切勿以为这和牛顿力学中的质量守恒同义.前者是一个物理量“动质量”的守恒，后者则反映牛顿力学的如下信念：物质(matter)是永恒不灭的.今天看来这一信念并不正确，因物质(实物)可被转化为辐射，如正反物质发生湮灭以及核裂变放能，当然这些过程都是符合狭义相对论所认同的能量守恒的].至于静质量m0，则应强调它不符合守恒律(例如一静止核子分裂成两部分运动碎片的例子，读者可自己证明).

经过上面的讨论可知，在狭义相对论中，关于能量、动量、静质量和运动质量总共只有两个守恒律，即动量守恒和能量守恒.在前面的计算中曾默认有Mv=mu+m0，现在可以看到此式代表能量守恒.可见在证明动量表达式的洛伦兹协变性时需默认能量守恒(当然这一默认并无不妥).如本帖开头所说，只保留静质量(以后均写作m，单独提及质量二字时均指静质量)和总能(以后均写作E)两概念的做法其实很自然而不失科学性.经历了狭义相对论发展初期的一些曲折后，爱因斯坦在1948年的一次私人通信中曾写到：“为运动物体引入质量M=m[1-(v/c)^2]^(-1/2)的概念并无益处，……除了‘静质量’m外最好不要引入其他质量概念.”

由于数学水平所限，在此无法定量地又准确地为大家介绍广义相对论，以下只能为大家提一些思路，并在正式引入广相之前介绍部分基本概念.

当时当地观测、瞬时观者：

设想你自己是观者.你能直接观测发生在你身上的事件.如果事件发生在你身外，你当然也可以听见或看见(即用声光信号进行观测，这属于间接观测)，但间接观测都会涉及到信号传递需要时间的问题(如前文钟慢效应部分所说，若不考虑这一问题将会带来很多无法接受的结果)，讨论起来较为复杂.在实用方面，天文观测都属于间接观测，而间接观测的结果和直接观测结果可能有所不同(取决于观者和信号源的运动状态).理论上最简单明确又最基本的观测是直接观测，即对发生在观者世界线上的事件的观测，亦称当时当地观测(local measurement).而参考系由无处不在的观者组成，发生在其他观者身上的事件由该观者观测便是.你在观测发生在你世界线上p点的事件时，重要的往往不是你的整条世界线而是你在该点的运动状态(严格来说是四维速度)，这时便不需要给出你的世界线如何如何，而只需给出这一世界线在p点处的切矢(4速的定义就是世界线切矢).于是可提炼一个更为抽象的概念，称为瞬时观者(instantaneous observer)，它由两个要素——p点及p点的一个指向未来的类时单位矢量构成.

(惯性)观者、(惯性)参考系、(惯性)坐标系概念的明确：

进行物理观测的人叫观察者，物理学中将其概念模型化，即看成质点，简称观者(observer).为了观测，观者的手中应有一个标准钟和空间三维标架，这使得观者具有了时间感和空间方向感.为了对全时空(或其中一个开子集)的各处进行直接观测，需要处处设置观者(“流动哨”)，这种无处不在的观者们就构成一个参考系.准确地说，无数观者的集合叫一个参考系(reference frame)，若它满足以下条件：时空(或其中的一个开子集)中的任一点有且仅有这一集合内的唯一观者世界线经过.这一抽象定义就是我们常用的参考系定义的准确化和广义化.而按这个定义来说，系内的观者世界线便可以不平行，即距离可以随时间而变，甚至相对走时速度也可不同.为了在一定范围内简化讨论，我们可以取这样一类观者：处于惯性运动状态的观者称为惯性观者(inertial observer).惯性观者是最特殊而又最简单的一类观者，它们的世界线是类时测地线，每一惯性观者对应于洛伦兹坐标系中的一条t坐标线.因此，该系的全体t坐标线组成的参考系便称为惯性参考系，而该坐标系便称为该参考系内的一个惯性坐标系.在不必认真区分参考系和坐标系时，惯性参考系和惯性坐标系又统称惯性系.坐标系给它所从属于的参考系内的每一点赋予了确定数值组成的时空坐标.

在坐标系的概念形成后，为了研究方便，有时我们需要进行坐标变换，为此需要明确各种基本坐标变换(全部的坐标变换方式包括各种基本变换及它们的复合变换).同一参考系内的基本坐标变换有两种：平移和空间转动.其中平移又可分为时间平移(“统一调表”)和空间平移(“齐步走”)，空间转动则相当于把所有观者的三维空间标架统一转动一定角度(“全体向左转”).这些变换均不改变空间距离和时间长度(因为被测线段的起止点均未改变).不同参考系间的基本变换有一种，数学上称伪转动，就是我们熟知的洛伦兹变换.变换前后的两坐标系空间方向对应同向，空间坐标原点在坐标时t=t'=0处重合，而它们的t坐标线不同.这种变换可能会改变空间距离和时间长度(因被测线段的起止点发生变化.前述尺缩钟慢效应即是典例).可见同一惯性参考系内有无数不同的惯性坐标系，而由伪转动/洛伦兹变换联系起来的两坐标系必然属于不同参考系.这可以加深对参考系和坐标系概念的理解.

4速、4动量、物质场、能动张量：

质点世界线是一条以固有时τ为参数的参数曲线，在此基础上质点的四维速度(4速，4-velocity)定义为质点世界线(以固有时τ为参数)的切矢.质点4动量(4-momentum)的定义和3动量的定义在数学表达式上有一定的相似性，即“动量=质量\*速度”，前者的严格表达式为“4动量=m\*4速”.4速可借瞬时观者做3+1分解，其中空间分量是质点相对瞬时观者3速的γ倍；相应地，4动量也可借瞬时观者做3+1分解，其中空间分量是质点相对瞬时观者的3动量，时间分量是质点相对瞬时观者的能量.

质点的4动量把两个不同的概念——质点的能量和动量——有机地统一为一个物理量.4速和4动量均与观者无关(世界线的客观存在并不依赖于坐标系甚至参考系)，是绝对的物理量，但如何做3+1分解却与观者有关，分解是相对的.如果没有观者进行实际观测，4速和4动量均客观存在，但3速、3动量和能量则没有意义.现在可进一步理解近代文献只保留(静)质量m和能量E两个概念的原因：它们是不同类型的物理量.质点的质量m是不变量，从一个侧面反映质点的內禀性质，而能量E则还依赖于观者，所以它不是不变量.瞬时静止观者(4速和质点4速相等的瞬时观者)测得的能量就是质点静能，虽然和质量等值，但仍不是同类型的量.

在讨论连续介质(固液气流、等离子流等)时，我们注意的不是个别粒子的行为而是宏观的统计平均效应，关心的不是个别粒子的质量和动量而是空间每点(每一宏观小区域)的能量密度、动量密度、能流密度等.可见连续介质在很多方面类似于电磁场，于是两者可统称为物质场.设宏观小体积V内的静质量为m，相对某惯性系的速度为→u，则其3动量为→p=(E/c^2)\*→u，以V除以全式可得：3动量密度=(1/c^2)\*能量密度\*→u=(1/c^2)\*能流密度.电磁场的上述物理量定义可在电动力学教科书中找到.

正如质点的能量和3动量统一组成4动量这一矢量一样，电磁场的这些密度量统一组成一个张量，称为能动张量(energy-momentum tensor)，这是一个4维闵氏时空中的张量场，各种三维密度无非是它的(观者依赖的)不同分量.事实上所有物质场都有自己的能动张量.它有如下性质：Ⅰ任何封闭的(与外界无相互作用的)物质场的能动张量时变率为0，这体现着能量、3动量甚至角动量守恒律；Ⅱ借瞬时观者可将其分解出能量密度、3动量密度(包括其在空间各方向上的分量)、能流密度(同样包括分量)、应力张量(stress tensor)的各分量(应力的概念与流体力学和弹性力学有关，限于数学水平在此不深究)；Ⅲ能动张量是绝对的，而各种密度量则是相对的.

“抹匀”思想(统计平均思想)在经典物理学中的应用：

注意此处的经典物理学并非指牛顿力学，而是指不考虑量子效应的相对论物理学.研究连续介质时会遇到一个“理想与现实的差距”的问题：实际组成流体的微观粒子的运动是无规则的、杂乱无章的，它们具有近乎无限多的运动状态参数.如理想气体(ideal gas，理想流体的一例)中，由于频繁碰撞，各气体分子的世界线多处相交，分子的4速在碰撞时突变，所以各分子的世界线其实拧成一团乱麻.为了将这样的混乱体系理想化、简单化，我们在构造理想流体模型时就需要采取“抹匀”手法，即对微观粒子的运动做统计平均处理，这样我们才能得到明确的、描述整个流体(或其微观足够大的一部分)的张量场等物理量，由此又提炼出共动观者(comoving observer，4速和流体介质4速相等的瞬时观者)等概念.

注意我们上述中的“微观足够大”这一说法，它在不同的物理学分支中有不同的含义和表述：如定义物质密度时认为物质宏观上连续分布，宇宙学原理认为每一时刻的宇宙空间在大尺度(例如3\*10^8光年，也称宇观尺度)下均匀等实例.体系大多需同时满足“一定尺度下体现出各向同性”的要求才可做“抹匀”处理，目的就是在简化问题的同时不失科学性.

将实际的“光波”简化为光线(三维或四维均可)的做法称为几何光学(geometric optics)方法，这一做法的条件是光子世界线为直线，这在实际中当然是无法达到的要求.好在许多具有实用意义的电磁波都可以近似看做这样的光线.这样模型化的条件是所考虑的空间尺度很大于波长，突出了光的粒子性一面，使得用光子描述光的做法成为可能.其实这一描述也是量子电动力学的理论基础.

广义相对论的诞生：

相对性原理要求物理规律在所有惯性坐标系中具有相同数学表达式，用于狭义相对论就要求物理规律的数学表达式具有洛伦兹协变性.这是一个管定律的定律.因此，在建立狭义相对论物理学时，应该重新审查已有的物理定律，凡符合这个要求的就仍被视为定律，凡不符合这个要求的就要改造，直至符合要求才作为物理定律纳入狭义相对论物理学框架中.首先我们来审查麦氏电磁理论.麦氏方程天生就有洛伦兹协变性(改写为四维形式可以更加明确这一点)，因而可以不加改造地纳入狭义相对论物理学中.这其实并不奇怪，因为狭义相对论出现的重要原因之一就是麦氏理论与经典时空观有矛盾.再来审查牛顿运动定律，以动量守恒为例.正如前文所指出的，牛顿的动量定义→p=m\*→u使动量守恒不具有洛伦兹协变性，因而必须修改.只要把动量改定义为式⑥的→p=γ\*m\*→u，问题便迎刃而解，然后动量守恒就可作为定律而被纳入狭义相对论物理学.第三我们来审查牛顿引力论.牛顿引力论的基本方程是反映引力势和质量密度关系的泊松方程(这里不列出)，它有伽利略协变性而无洛伦兹协变性.从另一角度来看，该方程的一种形式的解表明某时空点的引力势仅由该时刻空间各处的质量密度决定，这意味着引力场以无限速度传播，显然与狭义相对论相悖.以上两点原因均表明牛顿引力论必须修改.牛顿的万有引力定律形式上和静电库仑定律很相似，既然麦克斯韦可以把库仑静电学推广改造为如此漂亮的麦氏电磁理论，看来也不难把牛顿引力论改造为狭相框架内的引力理论.但实际情况却远非如此简单.关键在，于万有引力定律和库仑定律虽很相似，却存在“符号差别”：电荷有两种，同吸异斥；质量则只正不负，同性却只吸不斥.仿照电磁理论，可以构建一个狭相框架内的引力理论，根据这个理论，引力场在变化时将出现类似电磁波的引力波，而且也以光速传播.不幸的是，由于上述符号差别，由引力波带走的能量竟是负的.这意味着系统在辐射引力波时自身能量增加，从而辐射强度变大，由此又会获得更多的能量.如此循环，必然导致物理理论上不可接受的后果.虽然可以通过修改理论克服这一困难，但又会出现新的困难.其实，狭相框架内的引力理论远非一个，但每个都有其自身问题.虽然无法绝对否定在狭相框架内建立满意的引力理论的可能性，爱因斯坦却独辟蹊径，于1915年成功地创建了革命化的、独立于狭义相对论的崭新引力理论——广义相对论.有趣的是，后人在克服狭相框架内的某种引力理论的困难的努力中，几经修改后得到的竟然也是和广义相对论完全一致的理论！

有两个重要原因促使爱因斯坦创立广义相对论，它们是引力的“普适性(universality)”和马赫原理，在此只介绍前者.牛顿引力的“普适性”包含两层含义：Ⅰ任何物体在引力场中都受引力(电中性的物体在静电场中不受电力，故电力无普适性)；Ⅱ引力场中的任何两个物体，不管其质量和组分如何，只要初始状态(质心的位置和其在该点的速度)相同，而且除引力外不受力，以后每一时刻的位置和速度就必然一样.这一结论已被许多越来越精确的实验所证实，它可以表述为：任意两个质点在引力场中同一点具有相同的引力加速度.这虽是司空见惯的结论，但为什么会这样？静电场中的两个电荷就不这样，它们在静电场中获得的加速度还取决于荷质比.在对引力做类似讨论时，不妨也把“荷“与”质”加以区分.质点的“荷”是其物质含量的量度，决定它在引力场中受多大的力，可称为引力质量，记作mG；质点的“质”则是其惯性的量度，决定它在力的作用下有怎样的加速度，可称为惯性质量，记作mI，不难导出质点在引力场中的引力加速度为→a=(mG/mI)\*→g，其中→g是该点的引力场强.若不同质点具有不同的引力荷质比mG/mI，它们在引力场中同一点就不能有相同的引力加速度.然而无数的、一个比一个精确的实验表明比值mG/mI对任何质点都相同，通过调整引力常量G还可使比值为1而简写为mG=mI.通常把这一事实称为等效原理.这是一个极其非同寻常的实验事实，应该引起深思.引力的“荷“与”质”本是两个完全不同的概念，它们为什么相等？牛顿引力论不能回答这个问题.在牛顿引力论中，这是作为实验事实(牛顿理论体系的一个公理)而被承认的.难道mG=mI仅是一种巧合吗？难道就没有更深刻的原因隐藏在这个事实的背后吗？难道一定不存在一个更优美的理论，其中mG=mI是可用推理证明的吗？对等效原理的思考，加上马赫原理的启发，促使爱因斯坦创立了他的广义相对论.

mG=mI的事实等价于初始位置和速度相同的、除引力外不受力的任何物体在引力场中都“齐步走”，这种毫无个性的集体行为强烈暗示着引力是整个时空背景的内禀性质，与其他力有本质区别.物理学是研究物理客体运动(演化)规律的学问.物理客体好比演员.正如演员的表演不能没有舞台一样，物理客体的演化也必须在某种舞台(或背景)上进行，这个舞台就是时空.在广相创立之前，人们默认相对论的背景时空是闵氏时空.闵氏时空是如此简单，以致人们往往不注意(忘记)它的存在.引力场中的“齐步走”现象引起了爱因斯坦对时空背景的注意.假如你看演出时发现某个演员的头顶突然下降了20cm，你会认为他蹲下了；然而假如台上所有演员的头顶及桌面、椅面都同时下降一致高度20cm，那么最大的可能性是舞台台面下降所致.类似地，在引力作用下的“齐步走”现象分明强烈地暗示着引力本身是一种纯时空效应.不妨这样猜测：引力可忽略时，时空是平直的；引力不可忽略时(例如必须考虑地球或太阳的引力场时)，时空变得弯曲，弯曲情况取决于产生引力场的物质分布.根据这一猜测，引力非常不同于其他力，它特殊到这样一个程度，以至在四维语言中它不再是力而是时空的弯曲！于是，除引力外不受力的质点应称为自由质点(已经自由到不能再自由的程度).注意到闵氏时空中自由质点世界线必为测地线的结论，自然进一步假定弯曲时空中自由质点世界线也为测地线(自由质点是最简单的质点，测地线是最简单的世界线，自由质点世界线是测地线的这一假定非常符合美学原则).引力的存在不表现为质点受到一个称为“引力”的四维力，而表现为时空的弯曲，它通过改变测地线来改变自由质点的运动方式.以上就是广义相对论最基本的假定.根据这一假定，可以把mG=mI作为逻辑结论来推出.(现在到了关键的、也可以说是水到渠成的一步.)设两个自由质点具有相同的初始位置和初始速度，即它们的世界线在相交且在交点处切矢相等由于自由质点的世界线为测地线，而测地线由初始条件(测地线出发点及其在该点的切矢)唯一决定，这两条世界线必然重合.翻译为物理语言，就是引力场中两个初始状态相同的自由质点在以后的各时刻状态必然相同，而这正是mG=mI的等价表述.可见，一旦认识到引力的实质是时空弯曲，mG=mI这一长期无法解释的实验事实就是十分自然的结论.广相就这样以其特有的优雅方式首次把引力解释为四维时空的几何效应(首次统一了引力与几何)，这一成功的关键就在于补上时间这一维，而直至目前单凭三维空间还无法把引力解释为几何效应.黎曼早在28岁(1854年)时就提出了内禀曲率的概念和计算方法，并在逝世(年仅40岁)前的一段时间内致力于寻求把电力和引力统一起来的某种理论.这一努力未获得成功的重要原因之一在于他专注于空间及其曲率而没有注意时空及其曲率.直至1905年狭相问世后，时间和空间才被同等看待；1908年Minkowski才明确提出时空这一绝对概念；再过不久，“引力的实质是时空曲率”这一划时代认识才伴随着爱因斯坦对广义相对论嗯构思而逐渐建立起来.把引力实质归结为时空弯曲的认识是人类智慧的一个伟大胜利.

以上讨论说明，广相是独立于狭相框架的物理理论，狭相框架中容不下广相，容不下体现为时空曲率的引力.

用现代语言来描述，广义相对论的最基本假设可归纳为以下三点(此处为简洁而形象地说明采取诸多归纳方式中的一种，这诸多归纳之间其实相互等价)：

Ⅰ三维空间中的引力实质是四维时空的弯曲.当引力存在时，时空背景不再是闵氏时空，而是某种弯曲时空，度规场也不再是闵氏度规的平直度规场，而是某种弯曲的洛伦兹度规场.这一假设大胆地把物理上的引力认定为纯粹的时空几何效应.据此，除引力外不受力的质点称为自由质点.

Ⅱ自由质点的世界线是它所在的弯曲时空的测地线.在假设Ⅰ的基础上，假设Ⅱ的提出是相当自然的.当引力不存在时，时空背景是闵氏时空，自由质点的运动方程是测地线方程，该方程右边含有一个与闵氏度规相适配的导数算符(左边为0代表质点所受四维力为0).当引力存在时，一个自然的假设是将该算符改为与弯曲度规相适配的导数算符，由此得出的运动方程恰是该弯曲度规下的测地线方程，可见自由质点的世界线为测地线.当质点所受四维力不为0时上述推理过程仍成立.这正是广相独立于狭相的一个体现.在广相中，引力不表现为一个四维力出现在方程的左边，它对质点运动的影响体现在把时空变得弯曲并要求质点沿弯曲了的测地线运动.或者说，引力对运动方程的影响就在于替换那个与度规适配的导数算符.

Ⅲ时空弯曲情况受物质分布影响，关系由爱因斯坦方程描述.在承认爱氏方程后假设Ⅱ不再是独立假设.

可以证明，当引力场足够弱，质点速度足够低时，牛顿力学的计算结果与广相的近似一致，可见牛顿力学可看做广相的弱场低速近似.然而应该说明，虽然结果近似，但两者看问题的观点却有明显区别.同是自由落体运动，牛顿力学认为落体受引力而获得加速度，属于非惯性运动，但按照广相，落体不受四维力，是自由质点，三维空间中落体的运动其实是四维时空测地线被引力场源弯曲了的结果.

广义相对论认为一切物理现象都是物理客体在弯曲时空中的演化.因此，为了用广相的观点研究物理，首先要找出各种物理客体在给定弯曲时空背景上的演化方程.由于实际生活中甚至是实验室中的引力场都很弱，广相和牛顿力学的差别很难体现出来，通过实验或观察来归纳出弯曲时空中物理定律的做法几乎不可行.因此，只能根据某些基本原则用假设的方法“猜出”这些物理定律，其正确性则有待于它们推出的各种结论的自洽性及其与实验(如果可能的话)结果的一致性来验证.当然这种猜测是必须有相当根据的，一个重要的根据就是广义协变性原理(principle of general covariance).爱因斯坦在创立广相时曾提出过这样的广义协变性原理（于1917年被证明有致命漏洞）：物理定律的数学表达式在任意坐标变换下形式不变.但后来发现这种表述对物理定律并无约束力，就连牛顿方程都可以通过非实质性的改写满足这一约束.这一问题引起了有关学者(包括爱因斯坦在内)的热烈讨论，从而出现了关于广义协变性原理的各种表述.此处介绍一种既能抓住实质又便于应用的表述：只有时空度规及其派生量才允许以背景几何量的身份出现在物理定律表达式中.可以看出，广义协变性原理的实质是要在物理定律表达式中排除一切与时空内禀几何无关的人为因素.

在以上讨论的基础上便可得到提出弯曲时空中物理定律必须服从的两个原则：Ⅰ广义协变性原理；Ⅱ在时空度规等于平直度规时应能回到狭相内的物理定律.这两个判据是必要而不一定充分的.虽然这并不能唯一决定弯曲时空的物理定律，但以它们为指导，加上物理和美学的考虑，在很多情况下可以自然地得出物理定律.广相和狭相的差别无非是背景时空的差别，因此狭相中用四维语言描述物理客体的方法可以自然推广到广相.

为了得出各种物理量服从的定律，在多数情况下只需要把狭相相应定律表达式中的平直度规及普通导数算符换成弯曲度规及其对应导数算符.这可被称为“最小替换法则”.但有时最小替换法则也可能推出不同的物理定律(比如含2阶或高阶导数算符的式子在变换时需要注意算符对易性等问题，限于数学水平不赘述).

任意观者的固有坐标系、惯性力与科氏力：

实际应用中我们不可能在全时空的任意位置都设置一个观者，而往往只能靠一个观者来进行观测.为了让这单一观者能起到理想的参考系作用，引入固有坐标系的概念.观者手中有自己的标准钟和三维空间标架，这可以统一整合为一个四维标架.而4标架只在观者世界线上有定义.为了记录发生在世界线之外的事件(实验结果)，就要设法把这标架向外延伸并形成一个坐标系，这一坐标系的坐标基底在世界线上应与观者的4标架重合.一种满足上述要求又十分简单的坐标系叫做固有(proper)坐标系，它由该观者的两个要素——世界线及线上的正交4标架场(每一时刻都要有确定的标架方向)决定.最简单的观者是自由下落(世界线为测地线)无自转观者，前文所述的惯性观者一般都是指这种观者(注意惯性观者可以有自转，这点在前文未做区分)，借它的固有坐标系就可准确定义质点的3速和3加速.但一般观者的运动状态可以有任意4加速，其空间3标架也可自转，牛顿力学中为使牛顿第二定律形式地成立，人们引入了假想力(fictitious force)的概念.即力学教材中常见的惯性力和科氏力，前者对应观者的轨道运动，后者则对应于观者的自转.如转动转盘边缘的坐转椅的观者，他有绕转盘中心的公转，其3标架也可发生自转.虽然实际生活中公转(拓展来说是轨道运动)和自转相互伴随，但注意它们在原则上是独立的，而分清这两种运动正是区别惯性力和科氏力的关键.在这个问题上，回转仪(一种类似陀螺的器材)起到了关键作用:转动回转仪飞轮的自转轴(简称回转仪轴)代表空间中不变的方向，而这一无自转判据又可通过角动量守恒的四维形式推广到广相中.坐过山车的观者必然有轨道运动，但若让他坐转椅(可向任意方向自由转动，宇航员训练中就有这种特殊的“转椅”)并时刻根据手中回转仪轴的指向调整角度，则他可以没有自转，他在观测质点时就可只有惯性加速度而无科氏加速度.当然在不少文献中，科氏力也被归结为惯性力的一种，这只是名称问题，并无不可，我们为了加以区分才采用分开提出的办法.

等效原理的推广内涵：

取闵氏时空任一惯性坐标系中任一条坐标线，在线上取一正交归一无自转4标架场，这一坐标线和4标架场便构成一个惯性观者，而该惯性坐标系正是他的固有坐标系.上述概念在一定程度上可推广至弯曲时空.首先，与闵氏时空惯性观者对应的是自由下落(世界线为测地线)无自转观者.为考察这种观者的观测结果，讨论著名的爱因斯坦电梯，设地面附近的电梯因缆绳断裂而自由下落，其内部的静止观者(自由下落无自转观者)便有失重感，这是牛顿力学就有的结果.当他放开手中的苹果，他将发现苹果不像平常那样离手下落，而是处于随遇平衡状态中.理由很简单：电梯观者相对惯性系(地球人参考系)有重力加速度→g，是非惯性观者(现在仍按牛顿力学思路考虑问题，广相则认为他没有4加速)，故他认为苹果受两个力，一是重力mG\*→g，二是惯性力-mI\*→g，由于mG=mI(这是关键)，合力为0，因此随遇平衡，或说处于失重状态.假如他做过宇航员，将觉得这苹果与远离各星球(因而时空平直或近似平直)的惯性飞船内的苹果有相同表现.推而广之，由于mG=mI，根据牛顿力学，爱因斯坦电梯内的一切非引力的(指实验室内物体间的引力被忽略，但室内可存在由室外物体引起的引力场)力学实验都与上述飞船中的相应实验有相同结果.这正是把mG=mI叫做等效原理的原因.

在构思广相嗯过程中，爱因斯坦又假设性地把这一原理从力学实验推广到一切物理实验，即假设电梯内的一切非引力的物理实验都与飞船内的相应实验结果一样，并由此推出光的引力红移，光在引力场中走曲线等结论.后人把同mG=mI相应的原理称为弱等效原理(weak equivalence principle，WEP)，把爱因斯坦推广后的原理称为爱因斯坦等效原理(Einstein equivalence principle，EEP).

对于广相的物理定律来说，自由下落无自转观者的固有坐标系与闵氏时空的整体惯性坐标系相类似，因此可称为局部惯性系或局部洛伦兹系(local Lorentz system，local Lorentz frame).人们常说物理定律在局部洛伦兹坐标系的形式与它在闵氏时空洛伦兹坐标系中的形式相同，从而自由下落无自转观者所做的一切物理实验与平直时空惯性观者的相应实验结果一样(等效).这正是EEP所要求的结论.不过上述提法不太准确，因为可以肯定的只有在观者世界线上才是如此，只要偏离世界线就无法保证完全等效.若要强行使世界线一个邻域内的物理定律形式同闵氏时空的完全一致，数学上可证明这需要使世界线各点曲率为0，与原来我们设定的讨论对象(弯曲时空)不符.其实选坐标系并不能消除其曲率(曲率与坐标系无关)，可见“物理定律在局部洛伦兹系的形式与它在闵氏时空洛伦兹系的形式相同”的说法对世界线邻域中的点未必成立.然而观者做实验时偏偏往往要涉及自己世界线的一个“小”时空邻域(例如电梯观者要涉及电梯空间)，于是问题不再如此简单.好在时空曲率只有在足够的时空范围内才会有明显体现(被实验所测出)，只要实验涉及的时空邻域足够小(对电梯而言只要其尺寸及下落时间足够小)，实验的结果便与平直时空的相应实验结果近似地不可区分.在任意弯曲时空中，如果只关心某点的一个足够小的邻域，就可以近似使用狭相的物理定律.

在牛顿力学中，设飞船在远离各星体的空间中做匀加速运动，宇航员将看到离手的苹果在惯性力的作用下做反向匀加速运动，就像在地球附近的自由落体一样.不难相信，飞船内的一切非引力的力学实验都与地面附近的对应实验结果近似相同，这可看做WEP的另一表述.据此，人们又常说“加速飞船内的宇航员发现自己置身于引力场中”，“加速度与引力场等效”.对这两句话应有正确理解.根据第一句话，初学者常提出这样的问题：既然加速飞船中的宇航员感受到引力，而引力就是时空曲率，宇航员岂非觉得自己置身于弯曲时空？答案是否定的，因为早已约定飞船远离各星体，它所在的时空必然近似平直(观者无关地平直).导致以上错误结论的关键在于他们推理过程中两次用到引力一词，而两次的含义不同.加速宇航员感受到的“引力”其实只是非真实的表观“引力”(惯性力)，它不由物质产生，不对应时空弯曲，只因宇航员的感觉而得名.

等效原理的含义、地位和作用，不同学者有不同看法，其中对等效原理作用的看法又因对其含义的理解不同而不同.Misner等学者认为“等效原理功能强大，用它可把所有狭相物理定律推广到弯曲时空.”，又说“把经典力学带进量子力学的运载工具是对应原理，类似地，平直时空与弯曲时空间的运载工具是等效原理.”.而Synge对等效原理则持极端相反观点，他在1960年的著作的序言中写到：“我从未懂过这一原理.……它意味着引力场的效应与一个观者的加速度的效应不可分辨吗？如果是这样，那它是错的.在爱因斯坦理论中，要么存在一个引力场，要么不存在，取决于黎曼张量是否为0(注：即时空曲率是否为0).这是一个绝对的性质，与任何观者的世界线都毫无关系.时空要么平直，要么弯曲.在本书的若干地方我都不得不煞费苦心地把由时空曲率导致的真实引力效应与那些由观者世界线的弯曲导致的效应区分开来(在多数情况下以后者为主).等效原理在广义相对论的诞生过程中实质上起到接生婆的作用，……我建议我们以适当的荣誉埋葬掉这位接生婆而正视绝对时空这一事实.”对等效原理的这一看法也许有些偏激，但上引段子中的一些说法不失为防止误解的清醒剂.例如，他关于分清真假引力效应的警告就是十分必要的.下面简述书的作者梁教授对等效原理的几点说明.第一，EEP是爱因斯坦在酝酿广相过程中对WEP的假设性推广，对广相的诞生起到过重要的“接生婆”作用，这是Synge也同意的.第二，弯曲时空的物理定律必须遵守两个原则：服从广义协变性原理，和在度规为平直度规时能回到狭相中相应的物理定律.更多教材对后者采用另一提法：服从EEP.结合广义协变性原理可得到他们的最小替代法则：把闵氏时空中洛伦兹坐标系的物理公式中的偏导数改为协变导数(逗号改为分号)便给出弯曲时空中局部洛伦兹系的相应物理公式.可见EEP配以广义协变性原理可从狭相的物理定律推广到广相的相应定律，因此可以说它是“狭相通往广相的桥梁”.不过如本书那样不提等效原理而只提度规(也要配以广义协变性原理)的做法也同样可得到弯曲时空的物理定律.无论采取哪种提法都可以得到最小替代法则.而一旦接受了这些物理定律从而建立起广相体系，讨论问题时原则上就可以完全不用等效原理(虽然有些作者讨论许多问题时都喜欢用等效原理).因此，从这个角度看，“埋葬掉接生婆”对广相似无影响.第三，对于某些较为复杂的情况如“弯曲时空中走测地线的带电粒子是否会发出电磁辐射”，“等效原理”是否被违反是长期以来有争论的问题.梁教授认为关键之一在于“等效原理”在这些情况下的准确含义尚待澄清(另一关键问题则是辐射的定义).在这个意义上，Synge说“我从未懂过这一原理”也许并不过分.第四，除广相外也存在许许多多互不相同的引力理论.所有引力论可以分为两大类，即度规理论和非度规理论.广相是当然的度规理论.还有一种著名的、很有竞争力的度规理论叫Brans-Dicke理论，其中描述引力的量除度规场外还有标量场.此外还有其他度规理论.判断哪个理论正确的标准当然是实验，为此也许要一个关于引力实验的理论.Dicke从上世纪60年代开始从事这种理论的研究，他的开创性成果使人们对等效原理及其意义的理解逐步加深，并终于意识到应把等效原理摆在考察所有引力理论(包括但不限于广相)的基础这一重要位置上.等效原理可分为三个层次，即弱等效原理WEP、爱因斯坦等效原理EEP和强等效原理(strong equivalence principle，SEP)，后两者的区别在于EEP只考虑系统所处的外引力场而不考虑系统内物体所激发的自引力场，即忽略主动方面而只考虑被动方面，而SEP则对主、被动方面都要考虑，讨论对象是“自引力系统”，大到恒星间引力，小到卡文迪许实验中两个铅球的引力都在考虑之列.EEP可看做SEP在自引力可忽略情况下的特例.这三个等效原理的实验检验对选择引力理论有重要意义.任何理论都满足WEP(因为有越来越精密的实验验证WEP，无人愿意创造一个不满足WEP的引力理论)，但对EEP和SEP则不然.讨论表明，如果EEP成立，则只有度规理论可能正确，这就表明若EEP被精密验证，则非度规理论将再无立锥之地.进一步讨论还表明，广相满足SEP而其他已知的度规理论(含Brans-Dicke理论)都不满足[可惜这一讨论还不等于严格的理论证明，因此上述讨论至今仍被称为猜想(conjecture)].所以，若SEP被精密地验证，则广相很可能就是唯一正确的引力理论.可见对三个等效原理的实验检验有着非常重要的理论意义，这些实验正以越来越高的精度在进行中.

潮汐力及其在广相中的体现：

因引力场空间分布的不均匀，一个有一定尺寸的物体在引力场中时其各部分受的引力有微小差别.比如我们取一串念珠(珠可沿绳自由滑动)，使其自由下垂并从静止释放，经一段时间后上端和下端的珠间的距离将产生微小变化.然而，若将同一实验拿到远离各星体的加速飞船中，虽然念珠也会“自由下落”，但无论经过多久，任意两颗珠之间的距离都不会有变化.可见即使就力学实验而言，飞船和电梯也并非完全等效.

以上是假想实验，但道理与此类似的现象也可在日常生活中找到.海水的潮汐现象就是一例.忽略实际起次要作用的太阳引力场而只考虑地球引力场，设地球是一个均匀的正球体，表面被一层海水覆盖，考虑水面上的AB两点，其连线经过地心.设某一时刻A离月球最近，B点则离月球最远，AB两点受到的来自月球的引力不同，两点就要相互远离，于是它们附近的海面将鼓起.地球观者看来，这两点鼓起的原因是两个力：月球引力和地球绕地月质心做圆周运动导致的惯性离心力.这两者的合力称为起潮力.随着地球自转，A点不再正对地球，海面高度降低，地球自转半周后，A点又离地球最远，海水再次鼓起.可见海水每天各有两次涨潮和退潮.这种“潮汐现象”其实也不仅局限于海水.对地面附近自由下落的人来说，头顶和脚底与地心距离不同，也存在把人拉长的力(只考虑地球引力场)，不过这种潮汐力很小，不会造成感觉.如果人在中子星表面自由下落，潮汐力可大到10^11牛顿的程度，人将会被撕裂而丧生.

上述讨论表明，牛顿引力论中的潮汐现象是是引力场的普遍性质.其实在狭相中也是如此.前文讲到过，只有观者世界线上的点处才是“完美的”等效参考系，由于测地线由起点(及其在该处的4速)唯一决定，而不同的珠当然有不同的世界线(要么不同时出发，要么不从同一地点出发)，它们之间距离发生变化的事实也就不难理解了.这种因引力场不均(时空弯曲情况不同)而引起的效应称为测地偏离效应，也称潮汐效应.

潮汐效应其实是任意弯曲时空的共有性质，这是一种“硬梆梆”的效应.它所代表的物理效应一般要通过两个质点的比对而得出，从而可称为相对引力效应.通过恰当选择坐标系可以消除引力的影响，但无论如何都消除不了潮汐效应.当谈到“弯曲时空必有引力”时，我们能明确的是“任意观者都可以观测到潮汐效应”.当时空曲率处处为0时，两种意义的引力效应都不存在，所以说“没有时空弯曲就没有引力”.

爱因斯坦场方程的提出：

既然物质分布产生引力，而引力表现为时空弯曲，一个自然的猜想就是时空曲率受物质分布的影响.物质分布可由能动张量描写，因此应存在一个把能动张量和时空曲率相联系的方程.考虑到牛顿引力论是广相的弱场低速近似，两个理论的对比就成为了寻求这一方程的一条重要线索.经一段时间的努力探索和纠错后，爱因斯坦于1915年11月发表了爱因斯坦场方程，并将其作为广相的一个基本假设.这是一个描述时空曲率与物质场关系的方程，方程左侧的项用来描述时空曲率也即度规场，右侧的项则用来描述物质的能动张量.

能动张量为0是方程的一类重要的特殊情况，叫做真空爱因斯坦方程.这一方程的适用条件是所求的时空范围内没有物质和能量(这也是“真空”二字的含义)，包括我们要研究的物理客体(类似静电场中的试探电荷).然而时空若真如这样没有物质，没有物质还有物理学吗？广相考虑的是自引力系统，因而物质场中必然要体现试探质点的存在.但若试探质点的质量(因而对时空曲率的影响)足够小，那么它对能动张量的贡献就可以忽略.因此真空爱因斯坦方程也是有足够的实践意义的.狭相所考虑的时空处处曲率为0，由爱氏方程导出的能动张量也为0，这又是真空爱因斯坦方程的一个特殊情形.平直时空中的能动张量必然为0，这就意味着狭相所考虑的系统之间不允许存在任何“引力”作用.可见狭相正是广相的弱场近似.当然，闵氏度规自然是真空爱因斯坦方程的一个解，但方程的解却可以是弯曲度规.一个重要例子就是施瓦西在爱氏方程发表不到一年后找到的真空度规.

能动张量非0的式子称为有源爱因斯坦方程，它很像闵氏时空中的麦氏方程.但两者存在一个重要区别：麦氏方程可以在事先指定的场源状态(4电流密度)下对电磁力进行求解.看来对爱氏方程也可在指定能动张量的情况下对时空曲率及度规场求解.然而在度规场不明确时能动张量也不明确(和能动张量直接挂钩的量是4速场和固有密度场，然而4速的定义又依赖于度规)，把度规和能动张量分别作为已知量和未知量的做法不妥.求解爱氏方程时应把描述物质场的量和时空度规一同作为未知量求解.

爱氏方程的非线性性使叠加原理不成立，这会导致诸多复杂后果.例如，方程的两个解之和并非方程的解.这是与麦氏方程的又一重大区别.可见方程的求解是相当复杂且繁琐的工作，这也给整个广相理论的发展带来诸多困难.然而在多数情况下引力场很弱，这意味着时空度规近似闵氏度规，因此爱氏方程在这类情况下便可采用一种较为方便的近似手法.这一手法就是线性引力论(线性近似，linearized theory of gravity)，与其对应的方程称为线性爱因斯坦方程(linearized Einstein equation).

广相的经典实验验证：

爱因斯坦创立广相的原始动机基本是纯理论的，然而，在任何物理理论问世后都要面对实验验证的问题.爱因斯坦很早就从广相出发做了三个有可能与实验对比的重要预言(后人称为三大经典验证)，最早的一个是1907年提出的光波引力红移，后两个分别是水星近日点进动和星光在太阳引力场中的偏折.近日点进动的计算结果与早已存在的观测数据吻合甚好，星光偏折的预言也很快取得观测的支持.然而由于缺乏足够精密的实验技术来测量极端微弱的广相效应(包括引力红移)，广相的实验研究从上世纪10年代起的45年内进展缓慢，几乎止步不前.从60年代开始，由于科技的进步及天文观测的新发现，广相的实验验证进入了全盛时期.既有对星光偏折和引力红移的更高精度的实验验证，又有一系列全新的实验.可以说，广相通过了迄今为止的所有实验检验，虽然精度和难度更高的许多实验还有待进行.本段只介绍三大经典验证.

Ⅰ引力红移：

当地面上的光源向上发射光波时，牛顿力学即可证光波会发生引力红移，红移量与引力场强和光程成正相关.广相的严格推导表明，红移量满足

λ'/λ=[(1-2M/r')\*(1-2M/r)]^(1/2)

其中λ和λ'分别是距离地心r处和r'处测得的光的波长.在上述物理情景中可以看出λ'>λ.红移的程度可以用相对红移量(波长增量与原波长的比值，简称红移)描述.计算表明，太阳发出的光达到地球时，红移只有约10^(-6)的数量级.为了加大红移量，可以测从白矮星表面发出的光，一来白矮星的M一般大于太阳，且r'的值也更大，白矮星来光的红移量可达太阳来光的红移量大几十倍.在广相发表后，人们曾多次测过白矮星来光的红移，但结果的精度还不足以确证理论预言.第一次高精度的引力红移实验是Pound和Rebka等在1960年利用穆斯堡尔效应完成的.穆斯堡尔在1960年发现，某些原子核(如Fe的质量数为57的同位素)在特定条件下可以发出谱线宽度很窄(很尖锐)的γ射线，含有这种原子核的晶体又能对这种频率的γ射线做选择性甚高的吸收.当发出的γ射线频率不论因什么原因发生微小变化，其被晶体吸收的程度就显著降低.这就为测出由地球引力场而引起的及其微弱的引力红移提供了强有力的手段.把这样的两块晶体分置于地球表面的不同高度处，低处一块做发射体，高处一块做吸收体.虽然根据两者高度差(当时实验为12.5m)算出的红移量只有1.36\*10^(-15)，但γ射线的吸收率仍因这一微弱引力红移而有所下降.为确认这一下降并测出下降的具体值，可让上块以某一定速率向下移动，利用多普勒效应的蓝移(波长变短)抵消引力红移.当时实验将这一速率调至3\*10^(-7)m/s时，吸收率达到了最大值.所得结果计算后与理论预测值高度吻合(相对不确定度仅1%，完全符合实验标准).这是第一次高精度的成功实验.

Ⅱ水星近日点进动：

按照牛顿力学，行星的轨道是以恒星为一个焦点的椭圆.然而观测结果与此略有歧离.以最靠近太阳的水星为例，虽然它在每一周期中的轨道很接近椭圆，但两个相邻周期的两个“椭圆”的长轴并不重合，这表现在它近日点(perihelion)的微小改变上.随着时间的推移，由于积累效应，“椭圆”的长轴(因而近日点)绕太阳的缓慢转动变得可以观测.这现象叫近日点的进动(precession).在广相问世之前，人们早已测得水星近日点的进动率约为每世纪5600弧秒.人们对此曾深入研究并找出过许多可能的原因(包括其他行星的影响等)，发现所有这些因素造成的进动率为每世纪5557弧秒，还有每世纪43弧秒无法解释.这就是著名的“43秒问题”.爱因斯坦根据广相认为水星是在由太阳造成的弯曲时空中的自由质点，通过对时空中类时测地线的近似计算自然导出水星的空间轨道不是闭合曲线、其近日点的进动率恰比牛顿引力论计算结果多每世纪43弧秒的结论.这一结果大大加强了人们对广相的信心.其实爱因斯坦的计算过程对任意行星都适用，由于水星的近日点进动最易精确观测，我们采用水星的运动状态计算得出的结果也最为可靠.

Ⅲ星光偏折：

远方恒星射到地面的光线经过太阳附近时要受太阳引力场的影响而弯曲，这是广相的一个重要预言.理论计算表明星光的偏转角近似值公式为

β=4M/l

(国际单位制中的公式为β=4G\*M/l\*c^2)，其中l是光子运动轨迹直线(指无限远处光子出发时的位置和速度所确定的空间直线)和引力质心的距离，M是引力质心的质量.式子表明偏折角随l的减小而增大.l的最小值为太阳半径，将具体数据带入式子可得β=1.75弧秒.这就是广相对星光偏折的定量预言.为了用观测验证这一预言，可设法拍摄当星光被太阳偏折时恒星的视位置，并与半年前(或后)当地球转到太阳另一侧时拍摄得的恒星真实位置相比较.然而观测恒星的视位置并不容易，因为太阳比恒星离地球的距离近得多，在太阳的光芒中根本看不到星光(“大白天看行星”的难度可想而知).于是人们想到利用日全食.日全食时，日光被日地之间的月球所挡住，但远方星体的光却可以“绕过”太阳而达到地球.

一战后不久，两支考察队从英国出发分别到巴西和非洲对1913年3月29日的日全食进行观测，两队的观测结果分别是理论预期值的1.13±0.07倍和0.92±0.17倍.这被认为是对广相的重要支持，它们在多数欧美报纸的公布引起了对战争厌倦的公众的注意，并使爱因斯坦声名显赫.但爱因斯坦本人对此反应平静.他对自己的理论是如此相信(基于其优雅性和内部自洽性)，以致曾说过“若观测结果与理论背离，我将对全能的上帝感到遗憾”一类的话.然而，后来对巴西考察队观测结果的独立分析认为观测值只是理论值的1.0至1.3倍，不如原来乐观.后来又有对多次日全食的观测，但仍不能认为是对广相完全肯定的支持.关键在于牛顿引力论也能预言星光偏折，只是偏折角约为广相预言值之半，因此还很难说观测结果对广相还是对牛顿引力论更有利.

广义相对论(结合其他现代物理理论)在天文学方面的应用：

引力辐射(引力波)：

引力场同电磁场的相似性使得人们期望广相存在和电磁辐射类似的引力辐射.爱因斯坦方程存在以光速传播的波动解一事在广相诞生后不久就已经为人所知.然而在相当一段时间内，引力波的真实性一直受到怀疑.Eddington曾在1922年提出如下疑问：引力波解可能只代表时空坐标的波动，因而没有观测效应(相当于它不携带能量和动量).情况在上世纪50年代开始出现转机.Bondi及其合作者们借用不依赖坐标系的手法证明了引力波的确携带能量和动量，以及引力波源在发射引力波时质量必然减小，这使得引力辐射的物理真实性和可观测性逐渐被普遍接受.

爱因斯坦方程是非线性方程，广相是非线性理论.虽然在很多情况下可以使用弱场近似，但在强引力场情况下必须对其非线性性给予充分注意，这是引力波与(闵氏时空的)电磁波的重要不同.麦氏方程是线性方程，叠加原理对电磁场适用，即同一空间传播的两列电磁波互不影响.反之，一般而言，两列引力波之间存在相互作用(散射).Penrose，Khan和Szekeres等对平面引力波(一种简化了的引力波模型)的碰撞问题曾有过开拓性的研究.

众所周知，带电粒子在做变速运动时发射电磁波.电磁波中占主要成分的是电偶极辐射，另外还有相比弱得多的磁偶极辐射和电四极辐射.类似地，质点做变速运动时就要发射引力波.然而引力波中不含对应于电磁偶极辐射的引力偶极辐射，研究其四极辐射才能得到非零结果.四极辐射从量级上小于偶极辐射，引力系统发射的引力波在量级上弱于条件类似的电磁系统发射的电磁波.

一般认为强引力波的发射都同剧烈变化的天体物理过程有关(但不包括球对称系统的球对称演化.Birkhoff定理说明这种演化无论多么剧烈都不会发射引力波.正如麦氏理论中不存在球对称电磁波一样.)，例如星体晚期的急剧的非球对称引力坍缩、超新星爆发以及活动星系核中的剧烈扰动等.这时引力场很强，线性近似不再适用，对这些过程的严格处理必然涉及在非球对称条件下处理非线性爱因斯坦方程这一艰巨课题.人们对强引力波的发射问题至今仍了解得不完善.

既然广相预言了引力波的物理存在性，引力波的探测就成了重要课题.由于达到地球的强引力波波源都很遥远，被探测的引力波完全可以看做平面波，并且弱得使线性近似适用.这使引力波的探测理论比发射理论简单得多.然而，引力波的直接探测在实验上仍有不小的难度.微弱的探测对象对探测仪器的灵敏度提出很高的要求，探测实验本身还带走“守株待兔”的味道(坐等较近的剧烈天体演化带来强引力波).Weber于1966年在马里兰大学开拓性地建立了世界上第一个引力波探测器，这是一台由一根巨大铝棒及其附加装置组成的复杂仪器.数年的不懈努力，Weber宣布在分装两处的两台引力波探测器上同时测得引力波脉冲.遗憾的是其他引力波探测者对此都未予认证.例如，Tyson的探测器比Weber的有更高灵敏度，却丝毫未接收到类似的脉冲.1987年2月，天文学家观测到一次距离地球很近的超新星爆发(SN1987A，相距地球16万和光年)，同年，国外一观测小组宣称接收到来自该超新星的引力波，但也未能取得世界上其他(为数不多的)引力波探测器的认证.然而，对脉冲星由于发射引力波对自身运动的影响的观测却取得了突破性的成果.脉冲星(pulsar)是一种快速旋转的中子星，由于某种机制而不断发射电磁波.若地球位于电磁波束扫描范围之内，就会按准确周期接收到射电脉冲(角动量守恒律保证其有相当稳定的自转周期).由两颗恒星组成的近似孤立的引力系统叫双星，这两颗恒星称为子星.子星围绕系统的公共质心公转.根据广相，这种加速运动将会发射引力波从而损失能量，后果是轨道半径变小和周期变短.与剧烈的天体演化过程不同，双星系统的引力场较弱，可用线性近似计算引力波带走的能量，从而得出由此导致的轨道和周期变化.要使这些效应被测量，应满足两个条件：Ⅰ轨道半径非常小(两子星足够接近)，以使广相效应足够明显；Ⅱ有一种精度很高的轨道周期测量方法.Hulse和Taylor在1974年发现的脉冲双星PSR1913+16正好满足这两个条件[脉冲双星(binary pulsar)是指有一子星为脉冲星的双星，PSR是脉冲星的识别符，1913和+16分别代表其赤经和赤纬(角度坐标).有兴趣的读者可以参考天文学有关教材].该双星两子星间最大距离只有10^9米的量级(约1个太阳半径)，一个子星为脉冲星则使条件Ⅱ得以满足：脉冲星所发的脉冲周期被誉为“钟一般地准确”，Taylor及其合作者们便能以异常高的精度进行观测，从而推算轨道周期变化率.经过4年来的上千次观测，他们于1978年宣布了对轨道周期变化率的观测结果，与线性近似计算的理论值吻合甚好.这是引力波提出60多年来关于引力波携带能量的第一个定量观测证据，虽然只是间接的证据.他们后来又继续对这一脉冲双星进行观测并取得进步，终于获得1993年诺贝尔物理学奖.

恒星演化：

恒星的前身是一团密度不均匀的气体(主要成分是氢)气团.密度较大处有较强的引力，吸引来更多的气体，逐渐形成一个近似球对称的气团.气团中任一部分气体受来自气团的引力都指向球心，因此整个气团将在自引力下收缩.这是引力势能转化为热能的过程，温度将不断升高.经典理想气体状态方程中压强和热力学温度成正比，可以看出温度足够高时气团内部的压强(气团的任意部分都受里侧和外侧气体所给的压力，此处的压强确切来说是指内外侧压强之差)可以顶住引力从而遏止收缩.然而在没有能源的情况下，维持这种状态是不可能的：由于气团温度比周围环境高，它不断向外辐射能量.如果收缩停止，温度就要下降，压强也就抗衡不了自引力.从能量的角度看它也必须不断收缩，以试引力势能的一部分不断补充辐射出去的能量.经过一段时间的缓慢收缩，气团中心的温度和密度终于大到可以引发氢核聚变反应的程度.星核处的氢核聚变释放出巨大能量以补充辐射散失的能量维持了温度，气团就不再收缩以消耗其引力势能补充热能.这时的气团开始成为一颗恒星.

总有一天星核的氢全部变为氦，只有周围一层薄薄的氢仍在燃烧.当星核温度尚未达到点燃氦的聚变反应时，情况与先前类似：氦球在自引力作用下再次收缩，同时变热.这使星核周围薄层的氢燃烧加剧，从而导致星球较外侧的气体膨胀和冷却，恒星变为红巨星(red giant).“红”是由于表面温度降低，“巨”则是因膨胀而得名.氦球收缩导致的高温高密可能达到点燃氦聚变的程度，所释放的能量再次使星核达到稳定平衡.而这种靠烧氦变碳或氧维持的平衡时间远短于烧氢变氦的.当氦全部烧成碳和氧时星核再度收缩.恒星晚年的命运因质量而异.对于质量较小的恒星(包括太阳)，星核的收缩不能提供足够温度以点燃碳，靠核能已经无法维持平衡.还有没有什么力量足以抗衡自引力？经典物理学中不存在这样的力量.

星体由氢、氦和其他元素组成，星内的高温使得这些元素的原子处于电离状态.按照经典物理学，这一离子和电子的组合可看做理想气体.以电子气为例，经典物理学中电子的平均动能和热力学温度成正比，热力学温度为0时平均动能(因而所有电子的动能)为0；然而根据量子物理学，即使绝对零度下的系统也可能存在可观的压强.量子物理学指出，电子服从泡利不相容原理，一个能级至多可被两个电子占据(这两个电子有相反的自旋)，因此，绝对零度时，电子一方面要挤进能量尽可能低的状态；另一方面，由于每一能量状态只允许存在两个电子，故所有电子必须占满能量从0开始直至能量为某确定值的所有状态，而大于这一能量的状态则全空(本段可参考无机化学课程中的能带理论).这一能量叫做费米能(fermi energy)，其值随密度的增大而增大.这就表明，即使绝对零度下，电子气中的电子也不像经典物理学中断言的那样完全不运动，它们具有起源于不相容原理而非热运动的动能，这种动能对压强和能量密度都有贡献.绝对零度下的电子气叫做(完全)简并电子气(degenerate electron gas)，由上述量子物理原因引起的压强叫做电子简并压(electron degenerate pressure).在普通密度下，费米能很小(例如一般金属中的费米能只有几个电子伏)，相应的电子简并压微不足道.但简并压在高密度情况下的作用却很可观.星核烧氦阶段过后的再次收缩造成的高密度使电子具有很高的费米能.虽然星核温度按一般标准来说很高，但因费米能很大，电子由热运动引起的压强贡献比由不相容原理和高费米能引起的压强贡献相比很小.在这个意义上与绝对零度下的情况无明显差别.所以这时星核内的电子气可看做简并电子气，其简并压有可能抵消自引力，使星体保持平衡而永不收缩.这种靠电子简并压支撑的星体就是白矮星(white dwarf).“矮”是指其大小比普通恒小得多，“白”则由表面温度很高得名.一个孤立的星体演化为白矮星后就不再有重要的演化过程.因温度比外界高，它将不断辐射能量.由于没有能源，辐射导致温度下降，直至与周围温度相等，因而再也不能通过它发出的光观测到它.许多文献也将此称为黑矮星(black dwarf).白矮星的存在已被天文观测所证实，天狼星B是人类发现的第一颗白矮星.直观来想，质量越大的星体自引力越强，只有质量不是很大的星体才有可能靠电子简并压支撑成为白矮星.钱德拉塞卡最先求得白矮星的质量上限约为1.3倍太阳质量.这一工作及他一生对天体物理学的贡献使他于1983年获诺贝尔物理学奖.

星体在演化过程中会因抛出物质而使质量减小，上段中的白矮星质量上限是指星体抛出物质后剩余质量必须小于这一数值.据估算，初始质量小于约6到8倍太阳质量的恒星都将经历红巨星阶段并抛出大量物质而成为质量约0.5到0.6倍太阳质量的白矮星.

若白矮星阶段的星体质量大于白矮星质量上限，则电子简并压不足以维持星体平衡，星核内部的核聚变反应将一级级继续，直至烧成铁和镍.这两种原子核是结合得最紧密的原子核(核子比结合能/平均结合能最大)，不可能再因核反应(包括裂变和聚变)而放能.于是星核在自引力作用下急剧收缩，密度和温度都急剧增大.这时的自引力已经强大到牛顿近似不再适用的程度，而当改用广相理论计算时发现维持平衡所需的压强比牛顿近似更大，平衡更难实现.在如此的高温高密下，高能光子可将铁镍原子核打碎成中子、质子或轻核(光分裂)，电子也将同质子发生反应(逆β衰变)成中子和中微子(后者溢出星体).于是中子在星核内占了绝大部分.中子也是费米子，也服从泡利不相容原理.在达到核密度(水密度的10^17倍以上)时，众多中子也成为近似的简并中子气，其简并压也有可能抵消自引力.星体再一次达到稳定平衡.这种靠中子简并压维持的星体称为中子星(neutron star).由于中子星内密度达到甚至超过核密度，而人们对此状态下的物态方程远不如低密度时的确切，中子星质量上限的计算极为困难.不同文献给出不同的计算结果，只能大概说中子星质量上限为2到3倍太阳质量.不妨认为中子星就是一个“超大型原子核”.中子星的尺寸要比白矮星小得多，典型中子星的半径不到100km，而白矮星半径通常在3000到20000km之间.中子星是一种极为特别且复杂的天体，它有各种“极端”(超常)表现：极高的密度、异乎寻常的超强磁场(可达10^12高斯)、甚高的自转速度(频率从1Hz到1000Hz)、离光速不远的超高声速、超流的内部、…….人们至今还很难对它了解得很透彻.

中子星的第一个理论模型是在1939年由Oppenheimer和Volkoff发表的.由于文章没有给出可观测的物理效应，人们对中子星的研究冷落了28年.中子星的存在从1967年发现脉冲星开始得以证实.脉冲星是一种在地球上测得的周期性电磁脉冲信号的信号源，周期约为1s甚至更小.其唯一可信的解释是：这是一颗旋转中的中子星，其表面的强磁场连同中子星的高速自转使得地球收到电磁脉冲信号(最开始发现的是射电脉冲).只有中子星程度的强大引力才能避免星体在如此高速的旋转中把自己“甩散”.

星核在形成中子星之前的收缩非常急剧，所以称为引力坍缩(gravitational collapse).正在坍缩的星核一旦达到足够的密度并被中子简并压所遏止，其强大的能量将表现为向外的冲击波将外层物质炸散，形成能量极大的超新星爆发(supernova explosion).著名的两个超新星遗址——蟹状星云和船帆状星云——中都发现了脉冲星，这是对上述理论的重要支持.我国古代文献对超新星爆发事件有过极其丰富的记录.其中宋史志卷九关于公元1054年(北宋期间)观测到的超新星SN1054的记录特别受到现代国际同行的重视.蟹状星云正是SN1054的遗迹.地球上最近一次观测到的肉眼可见的超新星爆发是在1987年的SN1987a.该超新星位于银河系附近的大麦哲伦云，距离地球约16万光年.超新星爆发的详细机制仍是一个正在深入研究的课题.

如果球对称恒星在抛出物质之后的剩余质量仍大于中子星质量上限，就没有任何力量可以阻止它的引力坍缩，它将无限制地缩为密度和曲率都无限大的“奇点”，并形成施瓦西黑洞.

施瓦西度规，奇性和奇点引言：

爱氏方程的求解是广相的重要问题.许多精确解对广相的研究和发展起到重要作用.由于爱氏方程是高度非线性的偏微分方程，一般情况下求精确解十分困难.然而在时空具有适当特性时，求解变得相对容易.方程的第一个精确解(也是物理上最重要的解之一)——施瓦西真空解——就是Schwarzschild在时空具有静态性和球对称性的前提下，在爱氏方程发表后不到一年内求得的.(注：静态时空是一种特殊的稳态时空，稳态时空的定义是时空中的引力场不随时间而变.为了照顾陈述的流畅性，对其具体性质不做解释.)

直观地说，施瓦西真空解描述的是一个演化已停止的孤立的球对称星体外部的时空度规.将施瓦西真空度规改写为线元表达式后可直观地看出度规场的几个奇特之处.以r代表所求空间点和引力质心的距离(距离的准确定义因时空扭曲而变得不甚明确，具体说明移至后文)，以M代表引力质心的质量(其实也是全施瓦西时空的总能).表达式中dt^2项前有一系数-(1-2M/r)，dr^2前有一系数1/(1-2M/r).当r=2M时后一系数变得无意义，当r=0时两个系数都变得无意义.这分别对应施瓦西度规的一个/两个分量变得无意义(或说退化).人们称这些能使度规分量退化的点为奇点(singularity)，并说度规在奇点处存在奇性.在英语文献中，奇点和奇性是同一个词singularity，但在汉语陈述中往往愿意把“奇性”这一性质和奇性出现的“地点”相区别，于是要用奇性和奇点两个词(然而奇点的“点”字有时会造成误解，比如r=0或r=2M在四维语言中代表的是一张超曲面而非仅仅一个时空点).奇点的出现有两个原因：Ⅰ度规张量在该点处表现良好，只是由于坐标系的选择不当而使得其某个分量在该点表现不好，这叫坐标奇点(coordinate singularity)，可通过选择适当的坐标系以消除；Ⅱ度规张量本身在该点就表现不好(是奇异的)，这叫真奇点(true singularity)或叫时空奇点(spacetime singularity)，是真正棘手的问题，也是广相的老大难问题.数学上可证r=2M处的奇性只是坐标奇性，时空奇性只有在r=0处才存在.令rs=2M(rs称施瓦西半径)并补上常数G和c可得rs=2GM/c^2，计算表明太阳的rs约为2千米，远小于太阳半径，而太阳内部的非真空度规表明太阳内部也不存在奇点，故对太阳(以及普通星体)来说根本没有奇性问题.然而，对于经历球对称坍缩后成为黑洞的球对称星体(Birkhoff定理保证其外部时空仍符合施瓦西度规)，奇性问题却有重要意义.

时空奇点(奇性)的定义及其性质：

奇点(奇性)的概念和物理量的发散性有密切联系，它在广相问世前早已存在.然而广相的时空奇性问题却比任何其他物理理论的奇点问题困难得多(从定义起就困难).关键在于，广相之外的物理理论中，背景时空是早已给定的(例如闵氏时空)，只要所关心的物理量在某点处发散(或无意义)，就说该点是该物理场的奇点.例如，在三维语言中，点电荷的静电场强→E=k\*Q\*→r/r^3在r=0处没有意义，或说r→0时E→∞，所以就说这个点是静电场的奇点，或说→E在r=0处有奇性.然而广相则有所不同.在这里我们关心的是时空奇性，也即度规场的奇性，于是度规场在这个问题中身兼背景场和物理场的双重身份(既是舞台又是演员).仿照静电场奇性的定义方法，似乎可对时空奇点做如下定义：“时空为奇异的，若时空中存在点p使度规在p点处无意义(或发散).p点叫时空奇点.”然而，时空本身就是一个配以洛伦兹度规的四维流形，度规在流形上的各点不但应有意义，而且要表现良好(如连续若干阶可微以保证引力场不突变)等.若流形中存在点p使该点的度规无意义，p点本就不应属于时空(不是合格的时空点)，所以真正的时空应是把p点从流形和度规场中开除后的结果.例如施瓦西时空不应包含所有r=0的点.看来时空奇性可改为：“时空为奇异的，若其中某些区域已被删去.”但问题在于如何才能准确判断“某些区域已被删去”.下面介绍一个巧法.以闵氏时空为例，我们应该能想象得出，任何一条不可延伸的测地线(已向过去和未来两端延伸到不可再延伸的程度)的仿射参数都可从-∞取到+∞.但若删去测地线上的某点，就会在时空中留下一个“洞”，这也使得该测地线分裂为两条测地线(注意仍是测地线，因其仍可以代表自由质点的世界线，哪怕仅仅是其中一段).我们说这两条测地线都是不完备测地线.一般地说，任意时空中的一条不可延伸的测地线称为不完备测地线，若其仿射参数的取值范围不是(-∞,+∞).不完备测地线的存在很大程度上可以看做是时空的某些区域被删去(因而有“洞”)的标志.于是可考虑这样的定义：“若时空存在一条或一条以上不完备测地线，则称它为奇异时空.”然而这个定义也有严重的漏洞，就是“打击面”过宽.一个本来不奇异的时空，若认为地挖去一个点(人为无视这个时空点的存在)，按上述定义成了奇异时空，这是不合理的.克服这个问题的办法是在定义中加上一个限制：所讨论的时空必须是不可延拓的，即不能通过添加某些点使得它变得更大.人为挖去若干点的时空不是不可延拓的，所以不满足这个定义.再从物理上是否奇异的角度审查上述定义.若不可延拓时空中存在一条不完备类时测地线，从物理上看它的确很奇异：该测地线代表的自由下落观者在有限时间内(根据他自己的标准钟)竟会在时空中消失，或在有限时间的过去不曾在时空中存在过！类似地，存在不完备类光测地线的时空也是物理上奇异的，因为类光测地线代表光子的世界线.然而类空测地线不是任何粒子的世界线.似乎没有理由认为只存在不完备类空测地线的时空是物理上奇异的.于是定义被细化为：

若不可延拓时空中存在一条或以上的不完备类时或类空测地线，就称它为奇异时空(singular spacetime)，或说它有时空奇点(奇性).

然而该定义仍有缺点.例如，存在这样的时空，其中没有任何不完备测地线，但却有一条奇怪的已做了最大延拓的非测地类时线，其线长有限，4加速的大小有界.这表明飞船沿此曲线旅行经有限时间后将在时空中消失！线长有限和4加速有界保证飞船只需有限燃料就可走完这条曲线，而这样的飞船原则上存在.这时空在物理上是奇怪得足以称为奇异时空的了，但按上述定义却不是.这说明该定义有“打击面”过窄的缺点.另一缺点是时空有“洞”的直观说法与存在不完备测地线并不总一致.例如存在这样的测地不完备时空(含有不完备类时、类光、类空测地线)，其背景流形是紧致的(紧致是数学上的提法)，因而流形无“洞”.虽然这定义存在这样那样的缺点，但仍不失为首选定义.Hawking和Penrose的奇性定理(1965～1970年)的证明中用到的就是这个定义.做了最大延拓的施瓦西时空仍含有很多不完备的类时和类光测地线(其存在性都与删去r=0有关)，因此施瓦西时空是奇异时空，r=0处存在时空奇性(因此r=0的点都不属于施瓦西时空).

许多奇异时空在沿不完备测地线趋于奇点时都有“曲率发散性”，比如施瓦西时空的度规场在奇点附近发散.当然上面提到的“曲率”还包括其他一些时空背景量.在肯定时空至少有一条不完备类时或类光测地线后，就可以说时空是奇异的.当沿着这些不完备测地线检查曲率发散性时可能有两种结果：有或者没有曲率奇性(曲率发散性).另一方面，也有些时空虽然存在曲率发散性，但只当“趋于无穷远”时才发散.“无限远”处的时空点一般没有实用性，因此若这种时空不含有不完备测地线，则它也不应称为奇异时空.可见，用测地不完备性定义的奇异时空比用曲率发散性定义的更恰当些.

Kruskal延拓下的施瓦西时空：

时空的延拓(extension)在数学上可以经过一系列的坐标变换来进行，Kruskal在他1960年的著作中就提出了一种对施瓦西时空的延拓方法——Kruskal延拓.这一延拓所得出的基本坐标系包含TX两轴(为了和原始施瓦西时空中的tx两轴相区分而使用大写字母)，其中T轴一般画成纵向，X轴画成横向.施瓦西时空的球对称性保证仅需两轴即可准确表示时空坐标而不丢失物理意义.在这一延拓中，T和X可以取遍r>0的一切值.画成二维时空图，径向(三维空间中体现为沿直线冲向r=0的粒子轨迹)类光曲线都是±45º的斜直线，任一双曲线上r为常数，任一过原点的直线上坐标时t为常数，…….这些性质都为讨论提供了很大方便.

以下介绍该延拓的坐标系中很重要的两个特例，读者可根据描述自行画图：

Ⅰ：r=0对应的X^2-T^2=-1

Kruskal延拓的范围限制是r>0，也可用坐标表示为X^2-T^2>-1.数学上可证任意r→0的径向类时或类光测地线都不完备，以及时空的一标量场在r→0时趋于∞，因而存在曲率奇性，这暗示施瓦西时空不能再延拓到r=0及其以外的r<0区域，说明r=0是时空奇点，从而Kruskal延拓是施瓦西时空的最大延拓(maximal extension).时空图中方程X^2-T^2=-1所代表的双曲线对应时空奇性所在处，一般把这双曲线画成虚线或锯齿状线，曲线上连同曲线内(指X^2-T^2<-1，即双曲线上下的区域)的点都不属于时空.

Ⅱ：r=2M对应的X^2-T^2=0，即T=±X

在画图中这代表两条过原点的45º斜直线，记为N1:T=X和N2:T=-X，若加+号或-号则代表其T>0或T<0的部分.两条直线又把全图又分成4个区域(不包含边界).分别将上下左右的区域称为B区、W区、A'区和A区.请注意B区和W区不包括双曲线X^2-T^2=1及以内的点.

A区相当于r>2M的区域，是施瓦西时空在做Kruskal延拓之前的{t,r}坐标的坐标域t∈(-∞,+∞),r>2M.B区相当于0<r<2M的区域.在从A区或A'区趋于N1时t→+∞,r→2M，趋于N2时t→-∞,r→2M.本帖在一开始就提到过，广相所允许的背景流形必须是连通流形，所以原始坐标域可以是r>2M，也可以是0<r<2M，但不可取两者之并(r=2M处的坐标奇性使得这两部分表观上不连通).但有了Kruskal延拓就不同，这一延拓成功地使得施瓦西度规在A、B区及其交界N1+上都有定义，它们的并集是一个连通流形.

在坐标取值上，A区和A'区都有r>2M，事实上这两区有完全相同的性质.A区或A'区出发的“内向”的(指r值不断减小的)、指向未来的类时过类光曲线都将不可避免地穿越N1+或N2+而进入B区.反之，从B区出发的任意一条指向未来的类时或类光曲线都不可能穿过N1+或N2+回到A区或A'区，它们的必然归宿是掉进奇点(奇点不属于时空，“掉进奇点”的准确含义是指类时线的r随对应观者的固有时增加而越来越小，无限逼近于0，这将使观者从固有时达某值时开始从时空中消失；或类光线的r也无限逼近于0，其对应的光子也将消失在时空中).这实在奇得不可思议.这表明N1+和N2+都是“有进无出”的单向膜，A区和A'区的任何物体或光子一旦穿过这单向膜进入B区就再也出不来，只能掉入奇点.因此B区叫黑洞(black hole)，N1+和N2+对A区和A'区分别是它们的事件视界(event horizon)，这两个超曲面在三维空间中都体现为球面.上述讨论还表明，A区和A'区之间没有任何因果联系，从A区出发的任一类时或类光曲线都将被B区拦住而无法进入A'区，反之亦然.在这个意义上也把A区和A'区称为两个(互不关联的)“宇宙”.与B区对应的W区也有类似的性质，它也有r<2M.但从W区出发的任一类时和类光曲线都将穿过N2-或N1-进入A区或A'区，N1-和N2-这两个在空间中体现为球面的超曲面都是“有出无进”的“单向膜”.既然B区叫黑洞，W区自然叫白洞(white hole).

关于白洞，有时我们会听到“白洞是由黑洞把t取相反数得到的物理模型，但这种取法是错误的，得到的模型都没有物理意义”这种说法.其实这种说法是有相当问题的.在做Kruskal延拓的过程中，B区和W区的出现是互相伴随缺一不可的，它们的地位近似等价.并且，虽然它们的T坐标的确互为相反数，但注意T和t之间有很复杂的函数关系(否则时空也不会在画图中被“扭曲”成这个样子)，只说把时间取相反数很容易让人误解为把t取相反数.并且注意到A区和A'区在X轴以下的部分t都小于0，引号中“取相反数所得的模型无意义”的说法岂非有同时否认黑洞外部分时空的存在性之嫌？

以上是在全时空为真空的前提下得到的施瓦西最大延拓.虽然这一延拓包含了黑洞、白洞、事件视界和两个全同“宇宙”等诱人术语，但其真实性的确需另做讨论.从物理演化初值的角度考虑，整个时空的存在可能性很小，但其中的一部分(包括A区、B区及其间事件视界的一部分)却很有意义.

在介绍黑洞的形成过程及黑洞的性质之前，补充说明一个概念：无限红移面.

如前述中引力红移所说，光子在远离引力质心运动时会发生红移.牛顿引力论给出高度类似的结论.(以下内容部分参考《高中物理竞赛培优教程》浙江大学出版社2007年4月第二版)牛顿力学中对任何星体都存在一个“逃逸速度”，这是运动质点“挣脱”星体引力场所需的最低速度，对于地球就是“第二宇宙速度”.它的计算思路是这样的：设星体半径为r质量为M，质点从星体表面出发沿远离星体质心的方向射出，若质点的初动能m\*c^2/2大于其在星体表面的引力势能G\*M\*m/r，则其可以在运动到“无限远”时仍剩余一定的动能，即质点可以“挣脱”星体引力场的束缚；若初动能小于引力势能，则质点不能运动到无限远处并终将落回星体表面.简单计算得出的星体半径临界值为r=2G\*M/c^2.这一数值和根据广相求得的数值相同，但两者却有不同的物理意义：经典力学无法断言半径小于这一数值的星体会将其表面的一切物质束缚在星体表面，比如(注意！以下是非广相物理学的看法)若采取“牺牲”打法，将粒子的一部分质量作为跳板，则其剩余部分可以强行挣脱引力场，作为跳板的部分则坠向星体.但这一数据仍可证明“无限红移面”的物理存在.光子也有E=m\*c^2，从而它也不能逃逸r=2G\*M/c^2的球体之内；或者说因从该半径处出发的光子逃逸到无限远处时动能(也即能量E)已经为0，从而波长变得无限大.无限红移面也因此得名.

广相中r=2M代表这一结论是通过对施瓦西时空的类光测地线的计算得出的.设一切实验均在r=2M之外进行，根据前述引力红移小节的红移公式，红移量z可表为外向光子出发点与引力质心距离r的函数(确切说和r及r'都有关).对于任意初始频率的光子来说，都有当r→2M时z→∞(对于任何r'来说都是如此).无限红移面在广相中的物理存在因此得证.然而不应说“无限红移面处发出的光在到达外部观者处时有无限红移”，因从r=2M处出发的一切外向类光测地线均被束缚在该超曲面上，其他方向的类光测地线则会被拉入超曲面内部，也即r越来越趋向0而根本不会被外部观者接收到.

读者还请额外注意一点，即无限红移面与事件视界两个词的不同意义：无限红移面定义的来源是“光的无限红移”，这和“单向膜”的定义其实不同.因无限红移面只对类光测地线进行限制，而事件视界则对一切类时和类光曲线(包括但不限于测地线)进行限制.施瓦西时空中这两个超曲面是重合的，但在一种更为贴近实际的时空(旋转黑洞所在的时空，后文对此有更为详细的介绍)中，无限红移面在空间中是一个椭球形，而事件视界仍是正球形，两者有且仅有两个空间交点.另一个类似的概念是静界，其定义和上述两者都不同，但巧合的是，它是必然与无限红移面重合的.

球对称星的引力坍缩与施瓦西黑洞：

如前文所述，演化后期的星体想要维持内部的压强平衡，其质量必须小于中子星质量上限.对于初始质量大于这一质量的恒星，若不能在演化过程中抛出足够多的质量而形成稳定的白矮星或中子星，就根本没有稳定状态可言，只能不断坍缩而形成黑洞.Birkhoff定理保证球对称星外必有施瓦西度规，但星体内部则要由另一度规(爱氏方程的非真空解)所描述.因此，球对称坍缩星的时空根本没有白洞区A'，也没有A'区.但黑洞区B及A区的一部分在此情况下却有重要意义.无论构成星体的物质如何坚固，只要星体表面越过事件视界，就只能不断收缩，直至整个星体被压为奇点.理由很简单，黑洞区内的所有类时线最终都要指向r=0的奇点处.原始施瓦西坐标只能覆盖r>2M或0<r<2M的时空域，因而不能表现恒星晚期坍缩为施瓦西黑洞的全过程，特别是就算把r的这两段取值拼起来，也无法表达全过程中最关键的一步——星体表面缩进事件视界以内，因施瓦西坐标在r=2M处无定义.

现在从外部观者的看法来讨论坍缩过程.乍一想由于视界以内的星体表面发出的光子不能越过视界，外部观者似乎会看到星体逐渐变小，在r=2M时突然消失.然而使用原始施瓦西坐标粗略画图时可知，对于视界外的外向光子世界线，越靠近视界的地方世界线越“陡”(即斜率越大)，在视界上完全竖直(躺在视界上).外部观者将永远可以收到视界外星体所发出光.他会感到星体的收缩越来越慢，星体无限趋于某一大小(稍后观者收到的光将存在越来越严重的红移，因此应假定观者对任何波长和强度的光都能进行观测)，即感到星体半径以越来越小的速率趋于2M并将“冻结”在这一大小.这种现象也称为在引力场中的“钟慢”效应.前文曾指出，比较钟速时首先要明确一可行的比钟方案，对于坍缩星的情况，光子就成为比钟的关键：设星体表面有一光源以固定频率(按它自己的固有时τ)向外不断发射光子，由画图可知外部观者收到相邻两光子的时间间隔(也依它自己的固有时τ')比光源发光的周期长，并且还越来越长.即dτ'/dτ和τ成正相关.所以外部观者将认为星体表面的标准钟比自己的钟慢，而且越走越慢.不过读者应明确这种“认为”是时空几何以及刚才所规定的比钟方式的共同结果.这种现象的另一个体现就是红移.把刚才实验中的“光子”换成“波峰”，对比易知光波存在红移且越来越甚(要注意的是这里不光包含引力红移效应，同时也包含由于光源远离观者运动的多普勒红移效应.因此这红移将比黑洞外稳定高度光源所发光的红移更甚.).

为便于理解，讨论以下假想实验(“假想”包括忽略潮汐力的作用).设某观者坐着燃料充足的飞船做黑洞探险.如果他不开动发动机，飞船将自由下落，必然穿过事件视界进入黑洞，并在有限的时间之内(对M约为三倍太阳质量的黑洞，这一掉落时间根据观者自己的标准钟约为2\*10^-5s)葬身奇点.若他在到达视界前“悬崖勒马”，调转船头，开足马力(即在r尚未小到2M时让r重新变大)，是可以安全返航并提交探险报告的.然而，如果他多走一步到达事件视界(须知他的世界线和世界相交时他并无特殊感觉)，就将“一失足成千古恨”，因为任何粒子的世界线一旦与视界相交就无法逆转.虽然外部观者会看到他的最后画面是几乎定格在黑洞之外，但他却永远不能再主动联系到外部观者，就连打个无线电话都打不出去，因为即使从视界上向外发射光子也只能使其“躺在”视界上而无法逃出视界.