

R. PENROSE

Structure of space-time

BATTELLE RENCONTRES
1967 Lectures in
Mathematics and Physics
Chapter VII

Edited by
CECILE M. DEWITT
University of North Carolina at Chapel Hill
and
JOHN A. WHEELER
Princeton University

W. A. BENJAMIN, INC., NEW YORK-AMSTERDAM
1968

Р. ПЕНРОУЗ

Структура пространства-времени

ПЕРЕВОД С АНГЛИЙСКОГО

Л. П. ГРИЩУКА и Н. В. МИЦКЕВИЧА

ПОД РЕДАКЦИЕЙ

акад. Я. Б. ЗЕЛЬДОВИЧА и И. Д. НОВИКОВА

С ПОСЛЕСЛОВИЕМ

Р. ПЕНРОУЗА

Издательство «Мир»

МОСКВА 1972

Имя автора хорошо знакомо физикам-теоретикам и космологам. Именно Пенроузу принадлежит доказательство важной теоремы о неизбежности возникновения физической сингулярности пространства-времени в ходе релятивистского коллапса тел.

В этой небольшой книге изложены проблемы общей теории относительности, в том числе природа общей относительности, конформная бесконечность, горизонты событий и частиц, релятивистский гравитационный коллапс и роль сингулярностей в космологии, а также развитые автором методы анализа структуры пространства-времени.

Книга представляет большой интерес для физиков-теоретиков, астрофизиков, космологов — как специалистов, так и студентов старших курсов высших учебных заведений и аспирантов.

Редакция космических исследований, астрономии и геофизики

Оглавление

От редакторов перевода	5
1. Введение	11
2. Сущность общей теории относительности	18
3. Метод абстрактных индексов	33
4. Пространства-времени со спинорной структурой	42
5. Истолкование спин-вектора	56
6. Явные формулы для кривизны	66
7. Уравнения Эйнштейна и фокусирование	72
8. Конформная бесконечность	89
9. Горизонты	111
10. Гравитационный коллапс	130
11. Сингулярности в космологии	161
Л и т е р а т у р а	176
Послесловие автора к рускому изданию	181

От редакторов перевода

Созданная Эйнштейном общая теория относительности легла в основу научного подхода к изучению геометрических свойств пространства и времени физического мира. Уравнения Эйнштейна описывают локальные искривления пространства-времени; решения же этих уравнений определяют структуру пространства-времени. Уже в первых космологических работах самого Эйнштейна возник вопрос о геометрических свойствах пространства-времени в целом, т. е. о его топологии. В этих работах (выполненных задолго до открытия Хабблом разбегания галактик) 3-мерное пространство рассматривалось как 3-сфера неизменного радиуса, вечно существующая во времени. Так был впервые поставлен вопрос о геометрических свойствах 3-мерного пространства: бесконечно ли оно, бесконечно ли число небесных тел во вселенной или же пространство замкнуто и конечно по объему?

Дальнейшее развитие космологии (построение моделей расширяющейся вселенной Фридманом, работы Эддингтона, Леметра и др.) привело к постановке новой проблемы — проблемы сингулярности в космологии. В расширяющихся однородных изотропных моделях вселенной (справедливость которых подтверждается наблюдениями) расширение начиналось с сингулярного состояния — бесконечно большой плотности материи и бесконечно большой кривизны пространства-времени. Естественно возникает вопрос, является ли эта особенность свойством лишь однородных изотропных моделей. Быть может, в общем случае, когда рассматриваются отклонения от строгой изотропии и однородности, бесконечности кривизны

и плотности материи не появляются? Не оказывается ли сингулярность специальным случаем «фокусировки» в вырожденном случае строго однородных и изотропных моделей?

Определенную надежду на последнюю возможность дает рассмотрение в ньютоновской теории кинематической задачи разлета сферического облака частиц. Если частицы движутся точно по радиусам, то неизбежно пересечение их траекторий в прошлом в центре шара и возникновение сингулярности. Однако если придать частицам небольшие случайные тангенциальные скорости, то частицы проскакивают друг мимо друга вблизи центра, и сингулярность не возникает. Может быть, аналогичная ситуация имеет место и в космологической проблеме общей теории относительности!

Развитие общей теории относительности, а затем и развитие астрофизики поставило вопрос о структуре пространства-времени, о сингулярностях, о топологических свойствах пространства применительно к проблеме эволюции не только всей вселенной, но и отдельного тела. Исследование гравитационного коллапса отдельного тела показало, что когда размер тела приближается к гравитационному радиусу $r_g = = 2GH/c^2$, гравитационное поле становится настолько сильным, а искривление пространства-времени столь велико, что даже лучи света и нейтрино, испускаемые телом, не могут уйти к далекому наблюдателю.

Что произойдет после сжатия тела до радиуса r_g , внешний наблюдатель никогда не узнает. Но каков же исход сжатия для наблюдателя на сжимающемся теле? В простейшем случае сферической симметрии здесь также возникает сингулярность: бесконечная кривизна, бесконечная плотность материи. Возможно ли избежать сингулярности в общем случае? Что будет с коллапсирующей материей после сжатия, сменится ли сжатие расширением, какова структура пространства-времени в этом случае?

Помимо этих вопросов, связанных со структурой пространства-времени в больших масштабах — так сказать, в целом. — стоит задача изучения локальных

свойств пространства-времени, т. е. локальных инвариантов, которыми оно характеризуется в разных физических процессах.

Уравнения общей теории относительности, описывающие структуру пространства-времени, суть нелинейные уравнения в частных производных второго порядка. Их анализ в сколько-нибудь сложном случае — задача необычайно трудная. В последнее время для решения подобных проблем развиваются тонкие математические методы. Автор предлагаемой книги, Роджер Пенроуз, является одним из создателей таких методов. До сих пор имя Пенроуза было хорошо известно у нас лишь сравнительно малому кругу лиц — математикам и физикам. Мы надеемся, что эта книга будет способствовать тому, что с методами Пенроуза познакомятся многие. Применение их привело Пенроуза и его последователей к глубоким и важным результатам, имеющим принципиальное значение. Главными из них являются доказательства неизбежности сингулярности в космологии и гравитационном коллапсе.

В предлагаемой книге излагаются основы строгих современных методов исследования структуры пространства-времени. Книга является переводом лекций Пенроуза, прочитанных им на I Баттельской конференции математиков и физиков. Эта конференция состоялась в 1967 г. в Сиэтле (США) и была посвящена топологическим проблемам в математике, квантовой физике и общей теории относительности. Лекции вошли в сборник, изданный под редакцией Дж. Уилера и Сесиль Де-Витт.

Следует указать, что данная книга не может быть первой книгой для тех, кто намерен изучить общую теорию относительности. Для чтения ее необходимо предварительное изучение учебника: Л. Д. Ландау и Е. М. Лифшиц, «Теория поля», а также книги: П. К. Рашевский, «Риманова геометрия и тензорный анализ».

Методы, излагаемые в книге, являются типично математическими и отличаются необходимой строгостью. Однако Пенроуз обладает прекрасной

физической интуицией, и это делает изложение живым и наглядным. Он не прячет способ получения результатов за формальными математическими доказательствами, а делится с читателями теми идеями, которые помогают ему находить общие и глубокие решения проблем.

Строгое математическое рассмотрение гравитационного коллапса реальных массивных звезд в конце эволюции, с учетом вращения и отклонения от сферической симметрии, показывает, что для внешнего наблюдателя картина коллапса стремится к застыванию, когда размер звезды приближается к гравитационному радиусу r_g . Что будет после этого момента, внешний наблюдатель никогда не узнает. В настоящее время доказано, что внешнее поле при этом стремится к метрике Керра (рассматриваемой в данной книге), зависящей только от массы, углового момента и заряда коллапсирующего тела. На фоне этих строгих результатов особенно неубедительно выглядят предпринимаемые иногда еще и сейчас попытки доказать, что гравитационный коллапс тела после сжатия до размеров меньше r_g может смениться расширением к тому же внешнему наблюдателю, так что наблюдатель будет видеть периодические колебания. Явно несостоятельны попытки объяснить на этой основе природу таких объектов, как пульсары или квазары.

Излагая вопросы общей теории относительности, Пенроуз часто обращается к квантовой теории и теории частиц. Хотя общеизвестно, что теория относительности может быть построена без всякого обращения к квантовой теории и является в этом смысле «классической», многие аспекты теории квантов и общей теории относительности оказываются тесно переплетенными, что специально подчеркивается автором.

Другие аспекты общей теории относительности, не освещенные в книге Пенроуза, читатель может найти в книгах: В. А. Фок, «Теория пространства, времени и тяготения», Физматгиз, М., 1961 и А. З. Петров, «Новые методы в общей теории относительности», изд-во «Наука», М., 1966; астрофи-

зические аспекты — в книгах: Я. Б. Зельдович и И. Д. Новиков, «Релятивистская астрофизика», изд-во «Наука», М., 1967, и «Теория тяготения и эволюция звезд», изд-во «Наука», М., 1971.

Часть вопросов, связанных со структурой пространства-времени, не вошла в данную книгу. Это главным образом результаты, полученные после 1967 г. Мы укажем здесь некоторые из них, дав необходимые ссылки.

Последние теоремы о неизбежности сингулярностей в космологии и коллапсе см. в обзоре: S. Hawking, R. Penrose, Proc. Roy. Soc., A314, 529 (1970).

Вопросы о структуре пространства-времени вблизи сингулярности изложены в статье: В. А. Белинский, Е. М. Лифшиц, И. М. Халатников, УФН, 102, 463 (1970).

В последнее время возникают идеи о рождении пар частиц-античастиц из вакуума в сильном и переменном гравитационном поле вблизи сингулярностей, как космологической, так и возникающих при коллапсе (Я. Б. Зельдович, А. А. Старобинский, ЖЭТФ, 61, № 6, 1971). Это одно из тех направлений, где квантовая теория непосредственно применяется в общей теории относительности. Заметим, что для этих вопросов оказалось важным понятие конформной инвариантности, широко обсуждаемое Пенроузом.

Структура пространства-времени, возникающая при коллапсе небесных тел (так называемые «черные дыры»), вопрос об эволюции материи после коллапса, поиски сколлапсировавших тел во вселенной, новый подход к тензору энергии-импульса гравитационных волн (тензор Айзексона), нелокальные инварианты для гравитационных волн обсуждаются в книге: Я. Б. Зельдович, И. Д. Новиков, «Теория тяготения и эволюция звезд», 1971.

Перевод книги выполнен канд. физ.-мат. наук Л. П. Грищуком и док. физ.-мат. наук Н. В. Мицкевичем, которые сами являются известными специалистами по общей теории относительности. Переводчи-

ками исправлен ряд опечаток английского издания и сделан ряд примечаний к тексту (помимо примечаний редакторов).

Редакторы и переводчики благодарят Р. Пенроуза за присылку послесловия, а также дополнений и исправлений для русского издания.

Данную книгу следует рекомендовать в первую очередь астрономам, физикам и математикам, интересующимся проблемами общей теории относительности. Ее можно также настоятельно рекомендовать студентам физико-математических специальностей и аспирантам, ибо знание современных математических методов совершенно необходимо для успешной работы в области теоретической физики.

*Я. Б. Зельдович
И. Д. Новиков*

1.

Введение

Согласно современной теории, все физические явления совершаются в рамках некоторого дифференцируемого многообразия, именуемого *пространственно-временным континуумом*. Мы настолько свыклись с этой мыслью, что такая структура пространства и времени кажется нам сейчас почти «очевидной». Однако прежде чем приступить к ее анализу, стоит разобраться, на чем основывается эта вера. Конечно, совсем не исключено, что когда-нибудь будет создана теория, которая сможет описывать природу лучше, чем теперешняя, и вместе с тем эта новая теория окажется несовместимой с представлением о пространстве-времени как о дифференцируемом многообразии. На такую возможность не следует закрывать глаза, но в то же время полезно подумать и о том, почему современный подход является таким прекрасным приближением при описании удивительно широкого круга явлений.

Весьма близкий к «локально евклидову» характер пространства вместе с непрерывностью времени, конечно, дает нам главное основание для строгого развития представлений о континууме. Во времена Зенона такого точного представления о континууме не было и возникали затруднения с понятием предельного перехода в пространстве или во времени. Сейчас мы этих затруднений не испытываем, но, может быть, в этом и состоит наш просчет! Стандартное разрешение парадоксов Зенона основывается больше на *математическом* понятии континуума, чем на природе самого пространства-времени. Утверждение о том, что пространство-время образует континуум, подразумевает сохранение его непрерывной природы

независимо от того, с каким «увеличением» мы его рассматриваем. Но ведь отнюдь не очевидно, что непрерывное описание соответствует действительности в достаточно малых масштабах, где существенную роль начинают играть квантовые эффекты. Возьмем, например, масштабы порядка 10^{-13} см (примерный радиус элементарной частицы). При любой попытке определить положение частицы с такой степенью точности становится вероятным (ввиду принципа неопределенности) возникновение чрезвычайно большого импульса. Тогда должны рождаться *новые* частицы, и некоторые из них могут оказаться не отличимыми от первоначальной¹⁾, так что понятие «положения» первоначальной частицы становится неопределенным [22]. Но еще более угрожающая картина вырисовывается, когда мы осмеливаемся перейти к явлениям, протекающим в масштабах порядка 10^{-33} см. Здесь квантовые флуктуации кривизны пространства-времени²⁾ становятся достаточно сильными, чтобы изменять топологию, и пространство-время должно оказаться каким-то беспорядочным наложением разнообразных топологий [112], а это уж никак не похоже на гладкое многообразие.

Абсолютно не ясно, имеет ли вообще смысл говорить о природе пространства-времени в таких масштабах, и если это *бессмысленно*, то мы *заведомо* не можем со всей строгостью описывать пространство-время как гладкое многообразие. Но, с другой стороны, можно утверждать, что представление о гладком многообразии соответствует подходу ко всем «разумным» физическим процессам. Лично я думаю, что *в конечном счете* это окажется не так — я не верю

¹⁾ В квантовой физике *неразличимость* частиц — это не просто тождественность их количественных характеристик, а *полное* отсутствие индивидуальности (для частиц с одинаковыми основными характеристиками — например, для всех электронов), так что невозможно даже просто перенумеровать эти частицы. — *Прим. перев.*

²⁾ Если, конечно, в эту область удастся корректно экстраполировать современные представления как квантовой теории, так и теории гравитации. — *Прим. перев.*

в то, что действительное понимание природы элементарных частиц вообще может быть достигнуто без более глубокого понимания природы самого пространства-времени. Но если мы ограничимся рамками того уровня явлений, при котором это понимание не обязательно (а ведь это почти вся физика наших дней), то представление о гладком многообразии окажется превосходным орудием для нашего анализа.

Оставим пока в стороне вопрос о субмикроскопической структуре пространства-времени и вместо этого сосредоточим внимание на его крупномасштабных свойствах. Тогда мы можем принять адекватность представления о гладком многообразии и положить, что его структура в большом может быть получена из сложения меньших «локально евклидовых» кусочков подобно тому, как в дифференциальной геометрии берутся перекрывающиеся координатные окрестности (карты). Этот путь может привести нас к топологии пространства-времени в большом, отличной от евклидовой. К сожалению, о крупномасштабной структуре вселенной известно слишком мало, и мы не можем с уверенностью говорить о ее глобальной топологии (разве что возможно сформулировать утверждения о ее ориентируемости). Итак, может оказаться, что крупномасштабная топология пространства-времени не представляет никакого интереса.

И все же, *несмотря на это*, не мешает с достаточной общностью рассмотреть вопросы топологии пространственно-временных многообразий. Для этого, в частности, имеются две различные, хотя и до некоторой степени взаимосвязанные, причины. Чтобы четко выразить первую из них, следует представить вселенную в ее четырехмерной целостности, а не как некоторое трехмерное пространственноподобное сечение («теперь»). Одной из сколько-нибудь надежно *установленных* наблюдательных черт четырехмерной крупномасштабной структуры вселенной является тот факт, что в некий «момент» в прошлом (порядка 10^{10} лет тому назад) все вещество вселенной, по-видимому, находилось в крайне сжатом и хаотическом (горячем) состоянии. Это вытекает из наблюдаемого

расширения вселенной и из уравнений общей теории относительности, если сделать определенные предположения о крупномасштабной однородности распределения вещества во вселенной. По-видимому, еще более непосредственно это вытекает из наблюдений недавно обнаруженного [25, 80a] фонового электромагнитного излучения, заполняющего все пространство. Теперешняя температура этого излучения равна примерно 3 К, чего можно было ожидать, отправляясь от чрезвычайно сжатого общерелятивистского «начального состояния» [33, 2]. (Охлаждение излучения до его теперешней температуры следует тогда из расширения вселенной¹⁾.) Если мы теперь безоговорочно примем те сглаженные космологические модели, которые обычно используются при этих расчетах, то мы и в самом деле придем к *сингулярному* начальному состоянию, для которого кривизна пространства-времени бесконечна. Вблизи этой сингулярности радиус кривизны может быть сколь угодно малым — меньше 10^{-13} см и даже меньше 10^{-33} см. Но в таких масштабах мы не можем считать картину, описываемую нашей моделью, адекватной хотя бы по указанным выше причинам, дающим основание сомневаться в правомерности описания пространства-времени как гладкого многообразия именно в этих масштабах.

Следует ли полагаться на такие модели в области, где радиус кривизны хотя бы немного приближается к таким значениям? Можно ожидать, что имеющиеся сейчас отклонения от однородности в кривизне пространства-времени (вызванные неоднородным распределением вещества) при экстраполяции в прошлое на сильно искривленные области пространства-времени дадут картину, резко отличающуюся от

¹⁾ Наличие фонового электромагнитного излучения горячей вселенной свидетельствует о большой плотности материи в прошлом и, следовательно, о ее расширении от сверхплотного состояния (что предсказывалось общей теорией относительности), но из нее никак не следует конкретное значение 3 К. Температура порядка нескольких градусов была предсказана из сопоставления наблюдений химического состава небесных тел с теорией синтеза элементов в начале космологического расширения. — *Прим. ред.*

сглаженной модели¹⁾. Есть ли тогда основания вообще говорить о какой-либо сингулярности? (Будем, например, исходить из наглядного определения «сингулярности» как области, в которой кривизна столь возрастает, что локальные физические законы катастрофически изменяются, возможно, вследствие нарушения гладкого характера пространственно-временного многообразия.) Не может ли оказаться, что, когда кривизна становится уже «умеренно» большой, картина существенно отклоняется от рассматриваемой модели, возможно даже приводя к совершенно иной топологической структуре? Одной из главных моих целей является изложение некоторых строгих результатов, убедительно указывающих (хотя, может быть, и не вполне доказывающих это) на существование сингулярностей пространства-времени, следующих из законов общей теории относительности. Чтобы прийти к этим результатам, потребуется рассмотреть довольно-таки сложные с топологической точки зрения случаи, даже если на самом деле эти случаи в природе и не реализуются!

Выше мы изложили первую из двух причин, убеждающих в необходимости исследования топологии пространства-времени. Вторая причина (если оставить в стороне вопрос о связи с субмикроскопическими свойствами пространства-времени) связана с проблемой гравитационного коллапса. Как следует из общерелятивистских соображений, внутренняя неустойчивость, присущая гравитационному взаимодействию при наличии чрезмерно больших концен-

¹⁾ Надо иметь в виду, что имеющиеся неоднородности в распределении материи (наличие отдельных небесных тел и их систем) являются, вероятно, следствием *роста* неоднородностей, т. е. в прошлом неоднородности вблизи сингулярности были меньше. Это так называемая гипотеза «минимальных» возмущений в начале расширения, необходимых для объяснения окружающего мира. Имеется, однако, и другая возможность: начало космологического расширения было в некотором смысле «максимально» неоднородным, и в ходе расширения неоднородности «сгладились». Подробнее об этих вопросах см. книгу Я. Б. Зельдовича и И. Д. Новикова «Строение и эволюция вселенной», изд-во «Наука», М., готовится к печати. — *Прим. ред.*

траций масс, отражена в существовании начальной сингулярности, которой обладают космологические модели. Эта неустойчивость вновь проявляется в фазе коллапса тех моделей, расширение которых не продолжается беспрестанно, и вселенная возвращается к состоянию с бесконечной кривизной в своем окончательном сингулярном состоянии. Однако для того, чтобы гравитационная неустойчивость проявила себя, необязательно брать вселенную *в целом*. Действительно, даже тела, масса которых не очень сильно превышает солнечную, могут оказаться способными к катастрофическому коллапсу, когда истощаются запасы их внутренней энергии. При определенных условиях такие тела, коллапсируя, уйдут за «безвозвратную точку», где, грубо говоря, гравитационное взаимодействие становится таким сильным, что даже свет, испускаемый телом, «затягивается» назад, вовнутрь, и никакой сигнал не может выйти в окружающее пространство. За этим пределом поведение тела во многом напоминает конечный этап коллапса вселенной (или, как можно думать, обращенный во времени начальный этап ее расширения). При этом следует ожидать возникновения сингулярностей пространства-времени, хотя (в рассматриваемом случае) эти сингулярности не могут быть обнаружены внешним наблюдателем. Такой наблюдатель видит тогда нечто вроде «дыры» в пространстве, в которую могут падать объекты, но из которой не может вырваться наружу ни объект, ни световой сигнал. Взяв правдоподобный диапазон возможных масс, можно оценить размеры таких «дыр» от нескольких километров до нескольких диаметров солнечной системы. Эти «дыры», по-видимому, было бы трудно обнаружить внешнему наблюдателю, хотя в принципе это и осуществимо. В настоящее время возможность их существования следует скорее из теории, чем из наблюдательных данных, но тем не менее их изучение сопряжено со многими загадочными вопросами топологии пространства-времени.

Природа пространства-времени не совсем такова, какой она нам «представляется», и в понимании ее

мы пока не достигли ясности. В будущей теории мы, несомненно, столкнемся со множеством сюрпризов и важных новых идей. Однако даже существующая теперь теория пространства-времени (а под ней я понимаю эйнштейновскую общую теорию относительности 1916 г.) таит в себе неожиданности и предвидения, которые мы только еще начинаем сколько-нибудь подробно анализировать. Из нее наверняка еще можно без особого труда получить уйму новых результатов, стоит лишь применить соответствующий математический аппарат. Я приведу здесь некоторые полученные результаты и укажу аппарат, действенность которого была недавно продемонстрирована. Тем самым я надеюсь возбудить к этому предмету интерес (может быть, активный) у специалистов в смежных областях науки.

2.

Сущность общей теории относительности

Самое главное, чему учит теория относительности, состоит, по-видимому, в том, что понятия пространства и времени не могут рассматриваться независимо друг от друга, но лишь в сочетании, дающем *четырёхмерное* описание явлений, т. е. описание на языке *пространства-времени*. При этом динамика становится предметом геометрии. Поучительно проследить, как этот подход применим к дорелятивистским динамическим теориям. Так, мы можем сравнить динамику Аристотеля, Галилея или Ньютона с частной или общей теорией относительности, выражая все эти пять вариантов теории на языке пространства-времени.

Рассмотрим пять типов пространства-времени, обозначив их следующим образом:

- | | |
|---------------------------------|-------|
| Пространство-время Аристотеля, | (2.1) |
| Пространство-время Галилея, | (2.2) |
| Пространство-время Ньютона, | (2.3) |
| Пространство-время Минковского, | (2.4) |
| Пространство-время Эйнштейна. | (2.5) |

В каждом случае пространство-время будет гладким четырёхмерным многообразием, однако ему будет приписываться некоторая дополнительная геометрическая структура, отражающая характерный аспект динамики. Каждая *точка* пространства-времени в действительности является «событием», т. е. чем-то таким, что изображается как точка в пространстве, существование во времени которой ограничено, однако, лишь одним моментом. История частицы — это *кри-*

вая¹⁾ в пространстве-времени, именуемая мировой линией частицы.

Пространство-время Аристотеля является просто произведением $E^3 \times E^1$, где через E^n обозначено n -мерное евклидово пространство, несущее обычную евклидову метрику и обладающее обычной, $\frac{1}{2}n(n+1)$ -мерной группой движений. Метрика E^3 описывает пространственные расстояния, а метрика E^1 — отрезки времени. Тогда в динамике Аристотеля имеет смысл говорить об абсолютном расстоянии между двумя событиями в пространстве, даже если разность времен между ними не равна нулю. В частности, состояние покоя частицы должно здесь выделяться из всех других состояний движения тем, что пространственные расстояния между любыми двумя точками на мировой линии покоящейся частицы равны нулю. Семипараметрическая транзитивная группа движений пространства-времени Аристотеля является прямым произведением евклидовых групп для E^3 и E^1 .

Пространства-времени Галилея и Ньютона отличаются от пространства-времени Аристотеля в том отношении, что расстояние между двумя точками в пространстве определено для них лишь при обращении в нуль разности времен для этих точек. Напротив, разность времен здесь всегда определяется однозначно. Структуру геометрии тогда можно сравнить с расслоенным по E^1 пространством со слоями E^3 , так что «время» E^1 можно понимать как *фактор-пространство* полного пространства относительно слоев E^3 . (Топология здесь, конечно, все еще остается той же, что и в случае Аристотеля, хотя структура объединения слоев уже иная²⁾.) Различие между пространствами-временами Галилея и Ньютона проявляется лишь после дальнейшего уточнения их структуры. Для этого мы выделим специальное (шестипараметрическое) семейство кривых в пространстве-времени, которые будем называть «геодезическими». (Здесь

¹⁾ Если не оговорено противное, «кривая» не будет параметризована (ср. разд. 9).

²⁾ Я благодарен А. Траутману, который разъяснил мне этот вопрос.

не делается никаких предположений об экстремальных свойствах этих кривых — мы просто требуем, чтобы через каждую точку пространства-времени в каждом направлении проходила лишь одна такая геодезическая.) Эти геодезические (за исключением лежащих в слоях E^3) будут тогда мировыми линиями частиц, *движущихся по инерции*. Чтобы найти геодезические в пространстве-времени Галилея, достаточно постулировать их как систему прямых в некотором E^4 , причем слои E^3 отождествляются с максимальной системой взаимно параллельных плоскостей в E^4 , а фактор-пространство E^1 определяется очевидным образом. Группой движений пространства-времени Галилея, сохраняющей эту структуру, является десятипараметрическая группа Галилея.

Более сложно определение геодезических в пространстве-времени Ньютона. Идея здесь состоит в том [20, 108, 109], чтобы рассматривать ньютоновскую теорию тяготения с точки зрения общей теории относительности Эйнштейна. Гравитации тогда ставятся в соответствие *не истинные силы*, а силы инерциальные, или «фиктивные» — такие, как силы инерции поступательно ускоренного движения, центробежные силы и силы Кориолиса. Это возможно в силу принципа эквивалентности Галилея — Эйнштейна¹⁾, провозглашающего равенство инертной и пассивной гра-

¹⁾ Фактически принцип эквивалентности сам *требует* от нас принятия такого взгляда на ньютоновскую теорию тяготения. Постоянное во всем пространстве гравитационное поле (в ньютоновском смысле) не поддается вообще наблюдению, так как все тела приобретают в нем одно и то же ускорение, и внутренние физические отношения оказываются такими, как если бы поля не было. В этой ситуации есть нечто общее со случаем электрического потенциала: если мы, например, добавим к нему постоянное во всей вселенной слагаемое, внутренние физические отношения никак не изменятся. Поэтому-то мы и понимаем ньютоновское гравитационное «поле» в сущности как род потенциала. Его нужно *продифференцировать*, чтобы найти настоящее поле, а именно кривизну. Можно представить себе каноническую калибровку потенциалов, требующую обращения их в нуль на бесконечности, но трудно видеть, как было бы возможно приложить ее к реальному миру, так как ничто не свидетельствует об убывании плотности гравитирующих тел при уходе на бесконечность.

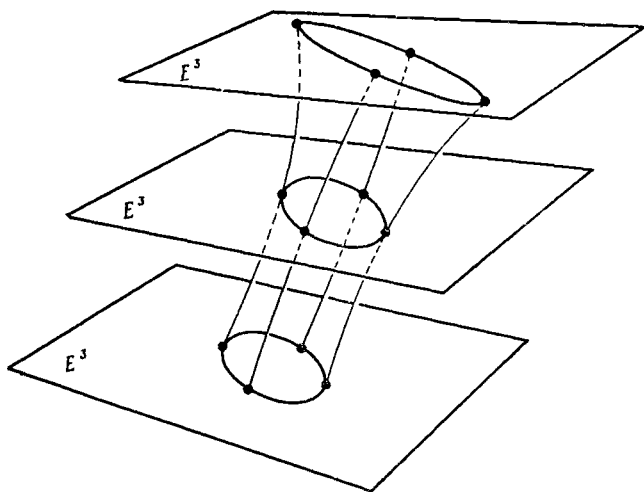


Рис. 1. Естественной мерой кривизны пространства-времени Ньютона является отклонение геодезических, т. е. приливные силы.

витационных масс. При этом считается, что частица, движущаяся под воздействием силы тяжести, но в отсутствие всяких других сил, движется по инерции, и ее мировая линия должна быть геодезической. Различных пространств-времен Ньютона может быть много в соответствии с возможными различными не эквивалентными друг другу гравитационными полями. Пространство-время Галилея — это частный случай, соответствующий нулевому гравитационному полю, либо, что эквивалентно, полю *постоянного* во всем пространстве-времени гравитационного ускорения, поскольку здесь, как и в общей теории относительности, физически содержательное понятие гравитационного поля дают лишь «приливные силы», возникающие в *неоднородном* поле гравитационного ускорения¹⁾. Эти приливные силы находят выражение

¹⁾ Эта идея подробно обсуждается в недавно вышедшей новом издании в русском переводе книге Э. Тейлора, Дж. Уилера «Физика пространства-времени», изд-во «Мир», М., 1971. — Прим. перев.

в «кривизне» пространства-времени Ньютона, которая характеризует степень отклонения внутренней структуры геодезических от внутренней структуры пространства-времени Галилея (рис. 1). «Уравнения поля» пространства-времени Ньютона связывают кривизну с распределением гравитирующего вещества; их можно получить, переходя к пределу при $c \rightarrow \infty$ (c — скорость света) в уравнениях общей теории относительности, но подробнее об этом мы здесь говорить не будем¹⁾.

Коренное отличие пространств-времен Минковского и Эйнштейна от трех предыдущих состоит в том, что в них не вводится *никакого* дополнительного понятия разности времен между событиями. Вместо этого в пространстве-времени определена псевдориманова²⁾ метрическая форма ds^2 с гиперболической нормальной сигнатурой $(+, -, -, -)$. Тогда разность во времени между двумя точками пространства-времени A и B *зависит от выбора мировой линии, соединяющей эти точки*, и дается интегралом вдоль мировой линии:

$$\tau = \int_A^B ds. \quad (2.6)$$

Везде вдоль допустимой (т. е. временноподобной или световой) мировой линии $ds^2 \geq 0$, так что τ оказывается действительным параметром. Величина τ определяет интервал времени (*собственного времени*) между событиями A и B по измерению (идеальными) часами, мировая линия которых совпадает с данной кривой. Так как собственное время стало теперь понятием, зависящим от пути, мы можем вернуться

¹⁾ Подробнее этот вопрос освещен у Траутмана [108, 109]. Траутман рассмотрел также более общий случай, когда ньютоновские пространства постоянного времени не обязательно представляют собой евклидовы E^3 . Это позволяет говорить о «ньютоновской космологии».

²⁾ Термин «риманов» относится вообще к случаю положительно определенной метрики, а «псевдориманов» — к случаю неопределенной метрики. Иногда псевдориманово многообразие сигнатуры $(+, -, \dots, -)$ называется «лоренцовым».

к определению «геодезической» как *экстремального* пути. Получаемая таким образом система (временно-подобных) геодезических определяет (согласно теории) движение частиц по инерции. В отличие от только что рассмотренных галилеева и ньютонова случаев теперь движение по инерции определяется сразу, как только конкретизировано поведение (идеальных) часов.

Пространства-времени Минковского и Эйнштейна соотносятся друг с другом так же, как пространства-времени Галилея и Ньютона. Так, пространство-время Минковского описывается единственным способом (например, структура его геодезических та же, что у E^4 , причем система временноподобных геодезических соответствует геодезическим в E^4 , наклоненным под углом менее 45° к некоторому фиксированному направлению), и гравитация в нем не определяется. Оно обладает десятипараметрической транзитивной группой движений — группой Пуанкаре (т. е. неоднородной группой Лоренца). Напротив, пространства-времен Эйнштейна может быть много, в соответствии с различными, не эквивалентными друг другу гравитационными полями. Здесь, как и в случае пространства-времени Ньютона, физический смысл имеют лишь *приливные* гравитационные силы, существующие при наличии градиента «поля гравитационного ускорения». Такие приливные силы описываются через отклонение внутренней структуры геодезических от аналогичной структуры пространства-времени Минковского, иначе говоря, *кривизной* пространства-времени Эйнштейна [ср. с (7.8)]. Эйнштейновские уравнения гравитационного поля описывают связь этой кривизны пространства-времени с плотностью материи (т. е. с тензором энергии-импульса-натяжений).

Прежде чем согласиться с точкой зрения Эйнштейна, было бы уместно спросить, обладает ли на самом деле пространство-время однозначно и последовательно определенной псевдоримановой структурой. Это зависит от существования в природе точных часов, от того, ведут ли себя такие часы локально по законам частной теории относительности, и от того

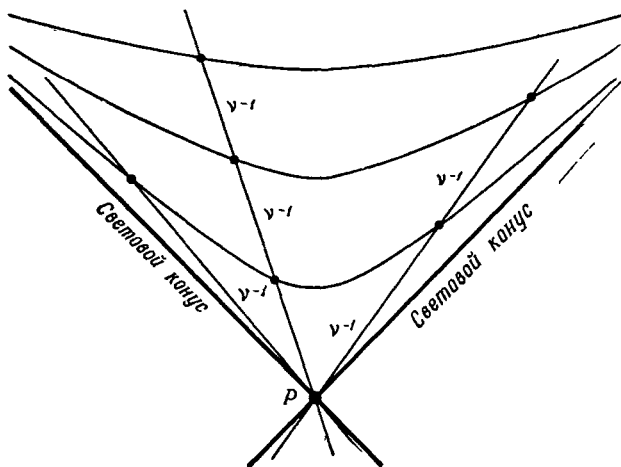
будут ли двое совмещенных друг с другом часов иметь одинаковый ход, независимо от выбора места и от их прошлой истории. По-видимому, факт существования точных часов теснейшим образом связан с квантовой природой материи¹⁾. В конечном счете это сводится к тому, что любой массе m соответствует (через постоянную Планка h) некоторая естественная частота ν , поскольку из эйнштейновского закона для массы и энергии $E = mc^2$ и формулы Планка $E = h\nu$ (если взять систему единиц, в которой скорость света $c = 1$, как это будет принято везде в дальнейшем), получим

$$\nu = m/h. \quad (2.7)$$

Поэтому каждая фундаментальная частица задает некую шкалу времени, связанную с ее собственной массой покоя m . Эту шкалу можно представить себе как последовательность «меток» времени вдоль мировой линии частицы²⁾, которые, по определению, отстоят друг от друга на ν^{-1} . Так определяется интервал ds вдоль мировой линии частицы, а если допустить изменение мировой линии, то мы получим ds для *любого* временноподобного интервала. Отсюда следует вывод о псевдоримановой структуре пространства-времени, поскольку частная теория относительности локально справедлива с чрезвычайно высокой степенью точности (рис. 2). Наконец, представляется, что получаемая псевдориманова структура не зависит от выбора частицы, используемой для ее определения. Это объясняется тем, что, по-видимому, отношения масс различных элементарных частиц являются строго определенными и не зависят от положения частиц в пространстве-времени или от их историй. (Если бы оказалось, что это не так, одно-

¹⁾ См. предисловие редакторов к русскому изданию. — *Прим. ред.*

²⁾ Планковские частоты элементарных частиц чрезвычайно велики и практически не могут быть непосредственно использованы в качестве часов. Соответствующие планковские частоты составных систем оказываются еще больше! По сути дела, те частоты, которые используются в атомных или ядерных часах, получаются из *разностей* масс (это как бы «биения»).



Р и с. 2. Для частиц, мировые линии которых, выходя из точки P , имеют различные временноподобные направления, интервал ds соответствует структуре псевдоримановой геометрии. Это следствие локальной применимости частной теории относительности.

значно определенной была бы только *конформная* геометрия, т. е. метрика была бы задана с точностью до множителя, зависящего от мировой точки.)

До сих пор я даже не упоминал о пространственноподобных интервалах, хотя риманову геометрию обычно излагают на языке расстояний, а не отрезков времени. Дело в том, что в действительности измерение расстояний — процесс более сложный, чем измерение времени. При задании двух близких точек P и Q , разделенных пространственноподобным отрезком, мы не можем просто приложить к ним линейку, чтобы измерить расстояние между ними. На самом деле линейка отнюдь не пригодна для измерения расстояний между событиями: в пространстве-времени «линейкой» служит двумерная полоса, протянутая во временноподобном направлении. Если не оговорено, что две точки P и Q на противоположных краях полосы «одновременны в системе отсчета, в которой

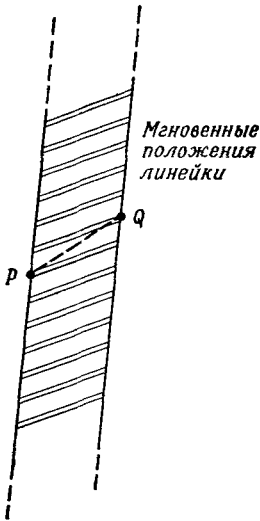


Рис. 3. Некорректное измерение пространственноподобного интервала PQ с помощью линейки: события P и Q не одновременны в системе отсчета, в которой линейка покоится.

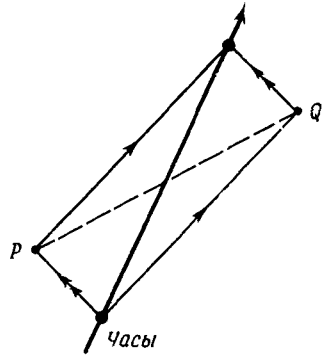


Рис. 4. Корректное измерение пространственноподобного интервала PQ с помощью часов и отраженных световых сигналов. (Весь рисунок следует представлять себе лежащим в одной плоскости.)

линейка покоится», то мы не определим правильной величины интервала PQ (рис. 3). К тому же положение осложняется вопросом о жесткости линейки, которая в конечном итоге зависит от пересечений «часов», представленных составляющими линейку атомами. Чтобы измерять пространственные интервалы более непосредственно, лучше пользоваться *часами*¹⁾ и отраженными световыми сигналами подобно тому, как показано на рис. 4. Тогда пространственноподобный интервал PQ будет равен временноподобному интервалу между событиями отправления исходного сигнала и приема отраженного (сигналы отправляют-

¹⁾ Как подчеркивал Синг [103], анализ геометрии пространства-времени по существу сводится к «хронометрии».

ся от часов во взаимно противоположных направлениях, и наш рисунок лежит во временноподобной плоскости).

Примем теперь, что пространство-время (при измерении правильно идущими часами) обладает однозначно определенной естественной псевдоримановой метрикой (с сигнатурой $+$, $-$, $-$, $-$). Тогда еще остается в силе вопрос, будут ли определенные с помощью этой метрики геодезические иметь какое-либо отношение к движению частиц по инерции. Когда общая теория относительности была впервые сформулирована, отождествление мировых линий частиц, движущихся по инерции, с временноподобными геодезическими было принято в качестве *постулата*. Однако позднее Эйнштейн и Громмер [30] сумели показать¹⁾, что в действительности это свойство является *следствием уравнений гравитационного поля Эйнштейна* [ср. с (7.1)] при некоторых разумных предположениях о том, как следует описывать (пробные) частицы в виде предельно малых распределений энергии-импульса. Вопрос этот довольно тонкий, но в конечном счете он сводится к выполнению ковариантного закона сохранения тензора энергии-импульса [ср. с (8.1)]. В остальном же конкретный вид уравнений гравитационного поля существенной роли не играет.

Возможен и другой ход рассуждений, когда движение частиц по инерции рассматривается как первичное, а метрика строится как вторичное понятие. Согласно рассуждениям Вейля [111] и Марцке и Уилера [56], если известны как временноподобные, так и световые геодезические (соответственно — мировые линии движущихся по инерции частиц и лучи света, не претерпевающие рассеяния), то метрику пространства-времени можно построить однозначно, с точностью до универсального множителя. В определенном смысле такой подход предпочтительнее, чем определение метрики с помощью часов, так

¹⁾ Более поздняя важная работа — [31]; более свежий обзор — [36]. [Современное состояние этой проблемы изложено в монографии В. А. Фока «Теория пространства, времени и тяготения», Физматгиз, М., 1961. — *Ред.*].

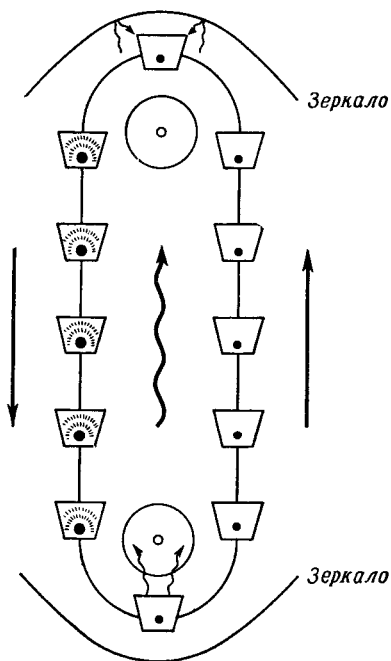
как при этом первичные понятия не зависят от тех разделов физики (например, квантовой теории), которые можно считать чуждыми духу общей теории относительности. Лично я придаю более фундаментальное значение часам, так как надо принять (без особой нужды) ряд предположений о геодезической структуре пространства-времени, чтобы она вообще соответствовала хоть какой-нибудь метрике, но тогда псевдориманова природа пространства-времени становится неясной. К тому же лично я считаю, что между квантовой теорией и общей теорией относительности *существует* глубокая связь, и было бы ошибкой пытаться строить эти теории в отрыве друг от друга.

Чтобы общая теория относительности имела отношение к реальной действительности, необходимо не только корректное определение псевдоримановой структуры пространства-времени, но и *нетривиальность* последней (т. е. чтобы кривизна хоть где-нибудь была отлична от нуля). Тот факт, что метрика пространства-времени *действительно* описывает искривленное пространство¹⁾, на самом деле следует из крайне элементарных рассуждений [96], касающихся

¹⁾ Последующие рассуждения не доказывают, что пространство-время не является *конформно-плоским*, т. е. не является таким, что его метрика может быть приведена к виду $ds^2 = f(x, y, z, t)\{dt^2 - dx^2 - dy^2 - dz^2\}$. Свидетельством в пользу искривленности пространства-времени является отклонение лучей света в гравитационном поле Солнца. По-видимому, единственный механизм, приводящий к одинаковому отклонению лучей света разных частот (отсутствию хроматической дисперсии), — это действие кривизны пространства-времени. В конформно-плоском мире (например, в теории тяготения Нордстрема [66]) гравитация *не приводит* к отклонению лучей света (главным образом потому, что световые геодезические конформно инвариантны). В то же время можно считать экспериментально доказанным существование *некоторого* ахроматического отклонения лучей.

Кривизна пространства-времени проявляется наиболее очевидным образом в существовании планетных орбит. Однако я не опираюсь на этот факт как на основной довод в пользу искривленности пространства-времени, потому что взаимосвязь между искривленностью орбит и поведением часов нельзя назвать очень прямой — ведь уравнения движения выводятся из уравнений гравитационного поля, являющихся дифференциальными уравнениями для метрического тензора (см. выше).

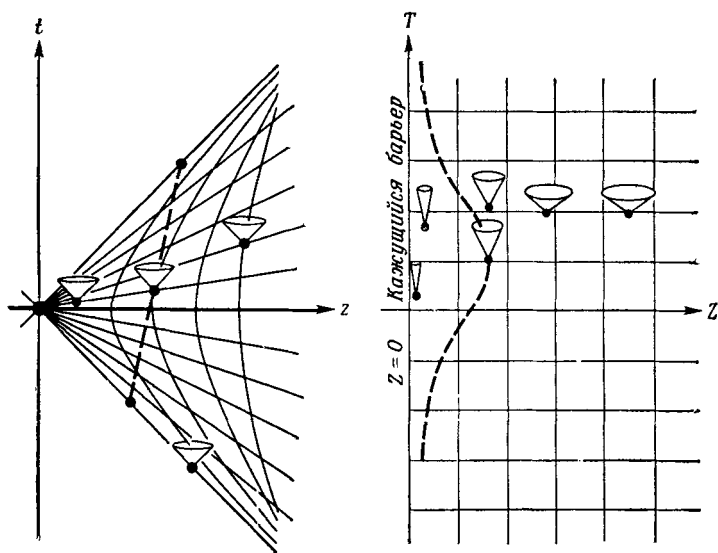
Рис. 5. Бесконечная цепь ковшей по Бонди. Вес избыточной энергии возбужденных атомов приводит цепочку во вращение.



невозможности построить вечный двигатель. Рассмотрим замкнутую бесконечную цепь ковшей, натянутую на два шкива, оси которых закреплены относительно поверхности земли в точках с разными значениями гравитационного потенциала (рис. 5; см. [9, стр. 418]). В каждом ковше находится атом, обладающий двумя состояниями — основным и возбужденным. Пусть все атомы в ковшах слева возбуждены, а атомы в ковшах справа находятся в основном состоянии. Возбужденные атомы, энергия которых превышает энергию атомов в основном состоянии, обладают при этом и большими массами. Значит, согласно самому слабому варианту принципа эквивалентности, и *весить* эти атомы должны больше. Допустим, что в нашей системе нет трения; тогда цепь начнет вращаться против часовой стрелки. Когда любой из ковшей достигнет своего нижнего положения, пусть находящийся в нем атом будет вынужден отдать свою избыточную энергию в виде фотона, направляемого системой зеркал к атому, находящемуся в верхнем положении. Если верхний атом *может* поглотить этот фотон, возбудившись

при этом, мы получим *перпетуум мобиле*, ибо такой процесс может продолжаться неограниченно долго, а энергию можно в принципе отвести от непрерывно вращающейся цепи. Нам остается поэтому предположить, что верхний атом *не может* поглотить фотон. Если бы мы могли спросить мнение Ньютона об этой системе, он, наверное, дал бы простой ответ: «Фотон «слабеет», пока он достигает верхнего шкива, и его энергии недостаточно, чтобы привести верхний атом в возбужденное состояние». Но из квантовой теории мы знаем, что фотон не может *просто* «ослабеть» — согласно формуле Планка, должна *понизиться* и его частота, когда он достигает верхнего атома, чтобы она оказалась ниже частоты возбуждения последнего. Поэтому в нижней части цепи часы должны идти во вполне определенном смысле медленнее, чем в верхней, а метрика пространства-времени, как мы ее определили, должна при наличии гравитационного поля быть иной, чем стандартная метрика плоского пространства-времени. (Описанный «прибор» является в высшей степени идеализированным, однако интересно отметить, что такой эффект замедления хода часов действительно удалось наблюдать непосредственно на Земле; это сделали Паунд и Ребка [85], используя эффект Мёссбауэра.) Заметим, что здесь использовался лишь чрезвычайно слабый вариант принципа эквивалентности, а именно утверждение, что энергия обладает *хоть каким-нибудь* весом. (Предельный вариант эксперимента состоял бы в использовании π^0 -мезонов, а не атомов. При распаде эти мезоны без остатка превращаются в фотоны, так что ковши справа будут совсем пустыми. Поэтому наши рассуждения справедливы, если π^0 -мезоны обладают хоть каким-нибудь весом!)

Однако этого рассуждения еще недостаточно для доказательства искривленности пространства-времени — метрика все же могла бы соответствовать плоскому миру и лишь записываться в необычной форме. Факт сферической симметрии гравитационного поля Земли, впрочем, сразу же показывает, что это невозможно, так как потребовалась бы связь метрики это-



Р и с. 6. Преобразование к равноускоренной системе отсчета. Кажущийся барьер при $Z=0$ не имеет физического смысла.

го поля со стандартной метрикой пространства-времени Минковского путем преобразования, включающего ускорение, направленное от центра Земли наружу симметрично во всех направлениях. Мы не будем разбираться более подробно в этих вопросах, и мне лишь кажется поучительным ограничиться анализом преобразования в пространстве-времени Минковского к «равномерно ускоренной системе». Такой пример ясно показывает, как статическая система *может* обладать свойствами системы, в которой присутствует гравитационное поле, хотя в действительности пространство-время остается плоским. Этот пример в некоторых отношениях сходен с представлениями Шварцшильда и Крускала сферически симметричного истинного гравитационного поля, как это будет видно позднее {см. (10.1) и (10.4), а также работу Бергмана [6a].}

Пусть x, y, z, t — стандартные координаты Минковского, в которых метрика имеет вид

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2. \quad (2.8)$$

Введем новые координаты X, Y, Z, T , связанные с предыдущими в области $z > |t|$ ($Z > 0$) преобразованием

$$x = X, \quad y = Y, \quad z = Z \operatorname{ch} T, \quad t = Z \operatorname{sh} T \quad (2.9)$$

(рис. 6). Теперь метрика принимает вид

$$ds^2 = Z^2 dT^2 - dX^2 - dY^2 - dZ^2. \quad (2.10)$$

Метрика (2.10) имеет такой вид, как будто она описывает систему со статическим гравитационным полем, роль потенциала которого играет Z . Член $(ZdT)^2$ в (2.10), стоящий на месте dt^2 , как будто говорит о «более медленном течении времени» вблизи $Z = 0$. Любая частица, неподвижная в системе X, Y, Z, T (т. е. обладающая постоянными координатами X, Y, Z), будет испытывать действие постоянной силы в направлении убывания Z , так как мировая линия частицы есть «на самом деле» гипербола постоянного ускорения в системе x, y, z, t . В системе X, Y, Z, T свободно падающая частица падает в направлении $Z = 0$ с замедлением, так что к $Z = 0$ она приближается лишь асимптотически (штриховая линия на рис. 6). При $Z = 0$ имеется как бы «барьер», который частицы неспособны преодолеть. Если произвести обратное преобразование к метрике (2.8), становится очевидным, что все это — только лишь проявление неполноты координатной системы (X, Y, Z, T) . В дальнейшем будет полезно помнить об этом обстоятельстве, когда мы перейдем к обсуждению проблемы гравитационного коллапса в разд. 10.

3.

Метод абстрактных индексов

В ходе вычислений в общей теории относительности часто приходится работать с тензорами довольно высокой валентности¹⁾. Даже такая фундаментальная величина, как тензор кривизны, уже имеет валентность, равную 4, и обладает известными довольно сложными свойствами симметрии. Отсюда следует необходимость в индексных обозначениях, позволяющих следить за различными используемыми величинами. Многие математики стремятся избегать таких обозначений, вероятно, потому, что эти обозначения подразумевают конкретную используемую систему координат. Но когда величины « g_{ab} » или « R^a_{bcd} » используются физиком, я не думаю, чтобы он часто подумывал при этом набор компонент, зависящих от выбора системы; скорее он имеет в виду не зависящий от системы координат физический объект, который олицетворяют эти компоненты. К тому же метод индексов позволяет очень удобно проводить ряд алгебраических операций, приводящих к новым объектам, и эти операции действительно никак не зависят от выбора системы. По сути эти алгебраические операции предельно просты, но вместе с тем гибки, и с их помощью можно производить более сложные операции. Было бы весьма досадно отказаться от такого мощного и гибкого метода всего лишь из-за какого-то чувства неловкости, связанного с условием суммирования и с зависимостью от конкретного выбора векторного базиса. Здесь я изложу алгебру,

¹⁾ Мы предпочитаем пользоваться здесь термином «валентность», а не «ранг», так как это более образный термин, а слово «ранг» несет иную смысловую нагрузку в приложении к матрицам.

совершенно не зависящую от выбора системы координат, но вместе с тем позволяющую, точно так же как и прежняя, проводить (но уже с чистой совестью!) вычисления с величинами, снабженными индексами. Эта новая алгебра (со своим методом абстрактных индексов) предоставляет даже большую свободу, чем старая, когда в ней вводят системы координат и векторный базис (ср. с [96a]). Преимущества нового подхода становятся особенно очевидны при рассмотрении спиноров, к которым мы перейдем в следующем разделе.

Мы сэкономим время и усилия, воздержавшись от чрезмерной формализации. Я надеюсь, что основные идеи при этом будут ясны. Рассмотрим векторное пространство V^* над полем F ; или, вообще говоря, допустим, что группа V^* — модуль¹⁾, εF — соответствующего вида кольцо (например, элементами V^* могут быть векторные поля²⁾, элементами же F — C^∞ -функции на многообразии). Идея состоит в том, чтобы построить величину, являющуюся по существу обычным тензорным произведением, кратным по V^* и кратным по дуальному³⁾ ему V_* , т. е. величину, в которой мы, однако, с помощью индексов смогли бы без затруднений контролировать следствия свойств симметрии и свертывания. Этого можно достичь, просто копируя обычные индексные обозначения (вместе с условием о суммировании и пр.), хотя теперь индексы a, b, c, \dots уже не *следует* понимать как просто однородные по своему происхождению символы, заменяющие, скажем, цифры $0, 1, 2, \dots, N$, а как *абстрактные значки*. Нам потребуется бесконечный запас абстрактных индексов

$$a, b, c, \dots, a_0, b_0, \dots, a_1, b_1, \dots, i_2, \dots, \quad (3.1)$$

¹⁾ Отличие модуля от векторного пространства состоит в том, что скаляры образуют не поле, а кольцо с тождеством. Кольцо отличается от поля тем, что в нем не всегда возможно деление на ненулевой элемент.

²⁾ Здесь под «полем» следует понимать сечение соответствующего векторного пучка над многообразием \mathcal{M} .

³⁾ Дуальное по отношению к V^* -пространству — это пространство всех линейных отображений модуля V^* на кольцо F .

чтобы можно было построить выражения произвольной протяженности. Обозначим через \mathbf{L} систему значков (3.1). Из любого элемента ξ из \mathbf{V}^* и любого значка $x \in \mathbf{L}$ мы теперь можем составить символ ξ^x . При пробегании символом ξ элементов из \mathbf{V}^* соответствующий объект ξ^x пробегает свое множество \mathbf{V}^x . Здесь следует подчеркнуть, что ξ^x — совершенно самостоятельная величина, а *вовсе не* набор компонент ξ в какой-то системе координат. Так как теперь мы хотим перенести обычные тензорные правила на действия над снабженными индексами величинами, следует запретить писать $\xi^a + \eta^b$ и разрешить только суммы $\xi^a + \eta^a$ и $\xi^b + \eta^b$. (Нужно считать ξ^a и ξ^b *разными* объектами.) Для любого $\lambda \in \mathbf{F}$ следует также разрешить писать $\lambda \xi^a$. Таким образом, любое из $\mathbf{V}^a, \mathbf{V}^b, \dots, \mathbf{V}^{a_0}, \dots$ есть векторное пространство или модуль, канонически изоморфные \mathbf{V}^* .

Может показаться противоестественным вводить бесконечное число изоморфных пространств, хотя в действительности у нас лишь одно пространство. Однако эту ситуацию можно рассматривать и с другой точки зрения. Каждый элемент из \mathbf{L} есть просто разновидность метки, выделяющей конкретный вектор (и т. д.) вне зависимости от того, где он может оказаться в конструкции. Так что ξ^x есть всего-навсего пара, составленная из ξ и из метки x . Иначе говоря, это элемент из $\mathbf{V}^* \times \mathbf{L}$. Мы имеем тогда $\mathbf{V}^a = \mathbf{V}^* \times \times(a)$, $\mathbf{V}^b = \mathbf{V}^* \times (b)$ и т. д. Аксиомы векторного пространства или модуля, конечно, применимы к любому \mathbf{V}^x :

$$\begin{aligned}
 \xi^x + (\eta^x + \zeta^x) &= (\xi^x + \eta^x) + \zeta^x, \\
 \lambda (\xi^x + \eta^x) &= \lambda \xi^x + \lambda \eta^x, \\
 (\lambda + \mu) \xi^x &= \lambda \xi^x + \mu \xi^x, \\
 \lambda (\mu \xi^x) &= (\lambda \mu) \xi^x, \\
 1 \xi^x &= \xi^x, \\
 0 \xi^x &= 0 \eta^x.
 \end{aligned}
 \tag{3.2}$$

Здесь $\lambda, \mu, 1, 0 \in \mathbf{F}$, причем 1 и 0 соответственно суть мультипликативная и аддитивная единицы. Кроме

того, $\xi^x + \eta^x = \eta^x + \xi^x$ [это следует из расшифровки $(1 + 1)(\xi^x + \eta^x)$] и $\xi^x + (-\xi^x) = 0$ [так как вместо $(-1)\xi^x$ можно записать $-\xi^x$, а вместо $0\eta^x$ — записать 0].

У *дуального* пространства V , также имеется бесконечное число канонически изоморфных копий: $V_a, V_b, \dots, V_{a_0}, \dots$. Пространство V_x можно понимать как дуальное по отношению к пространству V^x для каждого $x \in L$. Элементы V_x являются линейными отображениями V^x на F . Тогда для $\theta_x \in V_x$ имеем

$$\theta_x(\xi^x + \eta^x) = \theta_x \xi^x + \theta_x \eta^x, \quad (3.3)$$

$$\theta_x(\lambda \xi^x) = \lambda(\theta_x \xi^x), \quad (3.4)$$

где результат отображения θ_x на ξ^x записан просто как $\theta_x \xi^x$. Разрешим записывать то же самое и в обратном порядке: $\theta_x \xi^x = \xi^x \theta_x$. Потребуем, чтобы

$$\theta_a \xi^a = \theta_b \xi^b = \dots = \theta_x \xi^x = \dots \quad (3.5)$$

Теперь каждое V_a, V_b, \dots будет векторным пространством или модулем, для которого $\lambda \theta_x$ и $\theta_x + \varphi_x$ определяются через

$$\begin{aligned} (\lambda \theta_x) \xi^x &= \lambda(\theta_x \xi^x); \\ (\theta_x + \varphi_x) \xi^x &= \theta_x \xi^x + \varphi_x \xi^x. \end{aligned} \quad (3.6)$$

Идея состоит в том, чтобы задать нашу алгебру элементами из $F, V^a, V^b, \dots, V_a, V_b, \dots$. Чтобы сформулировать правила этой алгебры, следует вспомнить правила использования обычных тензорных индексных обозначений. Вспомним, например, что допустимы произведения вида $\xi^a \eta^b$, но не вида $\xi^a \eta^a$. Далее, допустимые произведения должны быть *коммутативны*:

$$\xi^a \eta^b = \eta^b \xi^a, \quad (3.7)$$

хотя в общем случае $\xi^a \eta^b \neq \eta^a \xi^b$. Из требования (3.7) видно, что, хотя *по существу* $\xi^a \eta^b$ есть тензорное произведение элементов $\xi^a \otimes \eta^b$, величины $\xi^a \eta^b$ и $\xi^a \otimes \eta^b$ нельзя попросту отождествлять, поскольку тензорные произведения, согласно строгому формальному опре-

делению, некоммутативны. Здесь же мы вправе определить коммутативный вариант тензорного произведения, потому что $\xi^a \eta^a$ в сущности не определено. В произведении $\xi^a \eta^b$ *значки* a и b указывают принадлежность множителей, но не их порядок следования. Один из способов точного определения типа произведения, использованного здесь (и предложенного С. Мак-Лейном), состоит в применении *симметричной алгебры* [54а] к прямой сумме $\mathbf{V}^a \oplus \mathbf{V}_a \oplus \mathbf{V}^b \oplus \mathbf{V}_b \oplus \dots$, после чего для каждой пары отдельных (конечных) систем элементов \mathbf{L} (скажем, a, p, r и b, m) мы выбираем соответствующее подпространство $\mathbf{V}_{bm}^{a,pr}$, натянутое на элементы типа

$$\xi^a \eta^p \zeta^r \theta_b \varphi_m. \quad (3.8)$$

Общий элемент $\mathbf{V}_{bm}^{a,pr}$ представляет собой линейную комбинацию выражений вида (3.8). Образованная таким путем конструкция гарантирует полную коммутативность каждого произведения (3.8) и выполнение ряда дистрибутивных законов [например, $\psi_m^{a,pr} (\theta_b + \chi_b) = \psi_m^{a,pr} \theta_b + \psi_m^{a,pr} \chi_b$ для $\psi_m^{a,pr} \in \mathbf{V}_m^{a,pr}$]. Порядок следования a, p, r или b, m в $\mathbf{V}_{bm}^{a,pr}$ не имеет значения, так что $\mathbf{V}_{bm}^{a,pr} = \mathbf{V}_{bm}^{p,ar} = \mathbf{V}_{mb}^{a,rp}$ и т. д. Однако для элемента $\rho_{bm}^{a,pr}$ порядок следования индексов не безразличен.

Каждый элемент $\rho_{bm}^{a,pr} \in \mathbf{V}_{bm}^{a,pr}$ является линейной комбинацией коммутативных произведений типа (3.8):

$$\rho_{bm}^{a,pr} = \sum_{i=1}^M \lambda^{(i)} \xi^a \eta^p \zeta^r \theta_b \varphi_m, \quad (3.9)$$

однако существует много способов выражения каждого $\rho_{bm}^{a,pr}$. Для интересующих нас модулей удобным критерием равенства двух выражений типа (3.9) является равенство соответствующих скаляров

$$\rho_{bm}^{a,pr} \alpha_a \beta_p \gamma_r \sigma^b \tau^m = \sum_{i=1}^M \lambda^{(i)} (\xi^a \alpha_a) (\eta^p \beta_p) (\zeta^r \gamma_r) (\theta_b \sigma^b) (\varphi_m \tau^m) \quad (3.10)$$

при любом выборе $\alpha_a \in V_a$, $\beta_p \in V_p$, $\gamma_r \in V_r$, $\sigma^b \in V^b$, $\tau^m \in V^m$. Отсюда вытекают и все алгебраические свойства. (Отметим, что в общем случае $\rho_{bm}^{arp} \neq \rho_{bm}^{par} \neq \rho_{mb}^{apr}$ и т. д.)

Полная тензорная система $\{V\}$ состоит из всех $V_u^{x \dots z}$, включая и $V = F$:

$$\{V\} = (V, V^a, V^b, \dots, V_a, V_b, \dots, V^{ab}, \dots, V_u^{x \dots z}, \dots).$$

На $\{V\}$ определены четыре основные операции, а именно:

СЛОЖЕНИЕ:
$$V_u^{x \dots z} + V_u^{x \dots z} \rightarrow V_u^{x \dots z}, \quad (3.11)$$

УМНОЖЕНИЕ:
$$V_p^{a \dots d} \times V_u^{x \dots z} \rightarrow V_p^{a \dots dx \dots z}, \quad (3.12)$$

ЗАМЕНА ИНДЕКСОВ:
$$V_u^{x \dots z} \rightarrow V_k^{f \dots h}, \quad (3.13)$$

СВЕРТЫВАНИЕ ПО a И b :
$$V_{bu}^{ax \dots z} \rightarrow V_u^{x \dots z}. \quad (3.14)$$

Фигурирующие в (3.11), (3.12) и (3.14) различные значки считаются различными элементами из L . В (3.13) различными элементами L являются как $x, \dots, z, u, \dots, \omega$, так и f, \dots, h, k, \dots, m . В остальном на них не наложено никаких ограничений кроме того, что число элементов x, \dots, z и f, \dots, h одинаково (то же самое касается u, \dots, ω и k, \dots, m). Сложение и умножение определяются очевидным образом. Замена индексов осуществляется просто перестановкой элементов в L (а такая перестановка не влияет на справедливость каких-либо уравнений). Чтобы определить свертывание, рассмотрим, например, свертывание по p и b ($V_{bm}^{par} \rightarrow V_m^{ar}$) в применении к элементу $\rho_{bm}^{apr} \in V_{bm}^{par}$, заданному в (3.9). В результате имеем

$$\rho_{xm}^{axr} = \sum_{i=1}^M \lambda^{(i)} (\eta^x \theta_x) \xi^a \zeta^r \Phi_m. \quad (3.15)$$

Отсюда $\rho_{xm}^{axr} \in V_m^{ar}$, так что индексы x — „немые“; они не дают вклада в общую валентность.

Алгебраическим путем можно проверить, что из таких построений следуют все обычные тензорные

правила¹⁾). Так, из сложения вытекает структура абелевой группы для каждого $V_u^x \dots z$. Умножение коммутативно и дистрибутивно по сложению. Свертывание должным образом коммутативно со сложением, умножением и с другим свертыванием. Свертывание нулевого элемента дает снова нулевой элемент. [При использовании равенства скаляров (3.10) в качестве определения равенства формальных выражений (3.9) это свойство нулевого элемента следует из того, что любая матрица над F , квадрат которой равен нулю, имеет и нулевой след. Это свойство выполняется в интересующих нас здесь случаях, хотя и теряет силу для некоторых колец с конечной характеристикой.]

До сих пор еще не возник вопрос о системе базисных векторов для V^* . Все же часто оказывается удобным работать с базисом, и нам понадобятся обозначения, позволяющие отличать базисные индексы от абстрактных значков. Я предлагаю использовать для описания элементов базиса *готические индексы* a, b, \dots, a_0, \dots , т. е. каждый индекс из a, b, \dots обозначает одно из целых чисел: $0, 1, \dots, N$ [рассматривается $(N+1)$ -мерное пространство]. Использование готических индексов должно напоминать нам о двух обстоятельствах: во-первых, что любое выражение, содержащее эти индексы, предполагает выбор (возможно, произвольный) базиса, из чего проистекает нарушение ковариантности; во-вторых, что при повторении в одном и том же члене индексов по ним всегда *подразумевается суммирование* согласно правилу Эйнштейна. Пусть теперь $\delta_0, \delta_1, \dots, \delta_N \in V^*$ будет базисом для V^* (мы предполагаем, что пространство конечномерно), а $\delta^0, \delta^1, \dots, \delta^N \in V_x$ — соответствующим дуальным базисом. При $x \in L$ в качестве канонических отображений в V^x и V_x имеем

$$\delta_0^x, \dots, \delta_N^x \in V^x, \quad \delta_x^0, \dots, \delta_x^N \in V_x. \quad (3.16)$$

Если воспользоваться собирательными символами $\delta_x^x \in V^x$ и $\delta_x^x \in V_x$, то условие ортогональности базис-

¹⁾ См. любой курс классического тензорного анализа, например [100].

ных векторов примет вид

$$\delta_y^x \delta_x^y = \delta_y^y, \quad (3.17)$$

где δ_y^y — обычный дельта-символ Кронекера (в этих обозначениях не предполагается никакой связи между x и y). С помощью равенства

$$\delta_y^y \delta_y^x = \delta_y^x \quad (3.18)$$

можно также определить элемент δ_y^x из V_y^x . Величина δ_y^x обладает обычными свойствами:

$$\xi \dots y \dots \delta_y^x = \xi \dots x \dots; \quad \eta \dots x \dots \delta_y^x = \eta \dots y \dots, \quad (3.19)$$

и отсюда видно, что она действительно не зависит от выбора базиса. Компоненты любого элемента $\zeta_u^x \dots^z \in V_{u \dots w}^x \dots^z$ относительно базиса задаются как

$$\zeta_u^y \dots^i = \zeta_u^x \dots^z \delta_u^x \dots \delta_w^z \delta_x^y \dots \delta_z^i. \quad (3.20)$$

И наоборот, чтобы выразить $\zeta_u^x \dots^z$ через их компоненты, следует просто написать

$$\zeta_u^x \dots^z = \zeta_u^y \dots^i \delta_u^y \dots \delta_w^z \delta_x^y \dots \delta_z^i. \quad (3.21)$$

Преимуществом таких обозначений является возможность в определенных случаях переводить отдельные индексы на язык компонент, сохраняя другие в виде абстрактных значков:

$$\rho_{bt}^{a pr} = \rho_{bt}^{a pr} \delta_p^b \delta_b^t \in V_m^{ar}. \quad (3.22)$$

Вообще все алгебраические соотношения никак не меняются при переходе к отдельным готическим индексам. Однако, когда в следующем разделе мы обратимся к ковариантному дифференцированию, обнаружится важное формальное различие в том, как следует работать с этими двумя типами индексов (плюс к уже имеющейся разнице в их смысле).

Алгебраическая структура типа построенной нами здесь обладает тем элементарным, но важным свойством, что иногда можно вложить две такие структуры одну в другую путем группирования индексов.

Тогда можно рассматривать новую систему значков, скажем \hat{L} , элементами которой служат (непересекающиеся) подсистемы элементов из L . Например, можно взять $\alpha = abc$, $\beta = def$, $\gamma = ghi$ и т. д., причем L разбивается целиком на непересекающиеся тройки и эти тройки уже суть элементы \hat{L} . Ясно, что при этом наборы $V^\alpha = V^{abc}$, $V^\beta = V^{def}$, ..., $V_\alpha = V_{abc}$, ..., $V_\gamma^{\alpha\beta} = V_{ghi}^{abc}$, ... будут подчиняться прежним правилам. Можно группировать векторные пространства (модули) и более сложным образом.

Кроме того, можно рассматривать системы, у которых «индексы» содержат значки разных шрифтов. В нашей схеме это привело бы к изменению лишь операций замены индексов (индексы могут заменяться лишь на *однотипные* с ними) и свертывания (свертывание разрешается только для верхнего и нижнего *однотипных* индексов). При анализе спиноров в следующем разделе мы рассмотрим пример системы такого рода со значками двух разных типов (связанных друг с другом операцией комплексного сопряжения).

4.

Пространства-времени со спинорной структурой

Для более подробного анализа структуры пространства-времени нам понадобится определение дифференцируемого многообразия. Я буду использовать определение, данное Шевалье (ср. с [65]), в котором дифференцируемая структура многообразия \mathcal{M} задается полностью вещественными функциями C^∞ на \mathcal{M} . Я выбрал это определение не из-за каких-то его особых внутренних математических достоинств, а потому, что, по-моему, оно лучше всего отражает ту роль, которую системы координат играют в физике. В самом деле, система координат в пространстве-времени есть набор четырех вещественных функций в том же пространстве-времени, таких, что в некоторой окрестности знание значений этих функций «гладким» образом задает точку.

Начнем с того, что назовем пространство-время \mathcal{M} топологическим пространством. Из числа подмножеств \mathcal{M} выделяются *открытые* множества; открытым называется множество, удовлетворяющее требованиям:

Пересечение любых двух открытых множеств есть открытое множество, (4.1)

Объединение любого множества открытых множеств есть открытое множество, (4.2)

Выполняется условие Хаусдорфа (т. е. для любых двух точек из \mathcal{M} существует два непересекающихся открытых множества, каждое из которых содержит одну из этих точек), (4.3)

Множество \mathcal{M} связное (т. е. это не объединение двух непересекающихся непустых открытых множеств), (4.4)

Имеется счетное количество открытых множеств, объединением которых является любое открытое множество на \mathcal{M} . (4.5)

Понятие открытого множества весьма нефизично, что существенно для введения математического многообразия. Обойтись без понятия открытого множества можно было бы лишь ценой изменения теории. Вопрос об определении открытых множеств с помощью более физического понятия *причинности* будет затронут в разд. 11.

Выделим теперь некоторое множество \mathbf{T} вещественных функций на \mathcal{M} , которые окажутся не чем иным, как вещественными C^∞ -функциями в пространстве-времени \mathcal{M} . Дифференцируемую структуру \mathcal{M} можно охарактеризовать следующими аксиомами для \mathbf{T} :

Если $f_1, f_2, \dots, f_n \in \mathbf{T}$, а φ — любая вещественная C^∞ -функция n переменных (в обычном смысле), то $\varphi(f_1, f_2, \dots, f_n) \in \mathbf{T}$, (4.6)

Если элемент f таков, что для каждого $P \in \mathcal{M}$ существует элемент $f_{(P)} \in \mathbf{T}$, совпадающий с f в открытом множестве, содержащем P , то $f \in \mathbf{T}$, (4.7)

Для каждого $P \in \mathcal{M}$ существует открытое множество \mathcal{P} , содержащее P , и четыре элемента $x^0, x^1, x^2, x^3 \in \mathbf{T}$, такие, что карта, переводящая $Q \in \mathcal{P}$ в $[x^0(Q), x^1(Q), x^2(Q), x^3(Q)]$, представляет собой гомеоморфизм на открытое подмножество из \mathbf{R}^4 и что любой элемент $f \in \mathbf{T}$ совпадает на \mathcal{P} с некоторой C^∞ -функцией переменных x^0, x^1, x^2, x^3 . (4.8)

Свойство (4.8) выражает то требование, которое предъявляется к функциям x^0, x^1, x^2, x^3 на \mathcal{M} , чтобы они могли играть роль *локальных координат* вблизи P .

Контравариантное векторное C^∞ -поле ξ на \mathcal{M} (т. е. сечение класса C^∞ касательного пучка к \mathcal{M}) можно определить как *производное* алгебры \mathbf{T} (над константами), т. е. таким отображением $\xi: \mathbf{T} \rightarrow \mathbf{T}$, что

$$\xi(f + g) = \xi(f) + \xi(g), \quad (4.9)$$

$$\xi(fg) = f\xi(g) + g\xi(f), \quad (4.10)$$

$$\xi(k) = 0, \text{ если } k \text{ постоянно на } \mathcal{M}. \quad (4.11)$$

Очевидно, что множество \mathbf{T}^* на таких ξ является модулем над \mathbf{T} , причем $(\xi + \eta)(f) = \xi(f) + \eta(f)$, $(\lambda\xi)(f) = \lambda\xi(f)$, $\lambda \in \mathbf{T}$, $\xi, \eta \in \mathbf{T}^*$. Можно показать, что

на языке локальных координат x^0, x^1, x^2, x^3 в открытом множестве $\mathcal{P} \in \mathcal{M}$ любое поле $\xi \in \mathbf{T}^*$ имеет вид

$$\xi = \xi^a \frac{\partial}{\partial x^a} \quad (4.12)$$

во всем \mathcal{P} . Поэтому ξ^a можно понимать как компоненты ξ в системе координат x^a на \mathcal{P} . (При переходе к другой локальной координатной окрестности компоненты ξ^a преобразуются обычным образом в области, где старая и новая окрестности перекрываются.)

Теперь, положив $\mathbf{V}^* = \mathbf{T}^*$ и $\mathbf{F} = \mathbf{T}$, мы могли бы ввести систему тензоров со значками (как в разд. 3) и пришли бы тогда к бескоординатному варианту обычного тензорного исчисления с «абстрактными индексами». Однако я предложу здесь несколько иной путь, приводящий к довольно необычному построению теории пространства-времени. Причина этого затрагивает, вероятно, сущность основного отличия математики от физики. В математике мы стремимся к максимально общим формулировкам, чтобы любой полученный результат имел наиболее широкое применение. Иногда его применения выходят далеко за рамки первоначальной задачи, а это позволяет объединять широчайший круг идей. В физике наши цели несколько иные. Нам дана чрезвычайно специфическая (?) структура, которую мы недостаточно хорошо понимаем, а именно — вселенная. Определенные аспекты этой структуры весьма точно (как это ни странно!) отображаются определенными математическими моделями.

Однако из разных источников нам известно, что эти модели не совсем (?) верны. Поэтому мы всегда заинтересованы в изменении математической теории. В физике формализм, содержащий хотя бы один произвольный параметр (такой, например, как *размерность* дифференцируемого многообразия), в определенном смысле уже *слишком* общ. Нам нужен скорее не общий, а *частный* формализм. При изложении старой теории на необычном (хотя с математической точки зрения и эквивалентном) языке иногда могут представиться новые, прежде не ожидавшиеся, но

с математической точки зрения естественные возможности видоизменения теории. Примером может служить теория тяготения Ньютона. В обычной ее формулировке можно было бы, например, рассматривать вопрос о замене показателя степени (-2) при радиальной координате на какую-либо другую степень. (Между прочим, такие попытки действительно предпринимались для объяснения движения Меркурия, но безрезультатно!) Если, однако, переформулировать теорию Ньютона как теорию пространства-времени (как это обсуждалось в разд. 2), то окажется математически естественным совершенно иное видоизменение теории, а именно переход к некоторой форме общей теории относительности.

Я намереваюсь ввести такое формальное описание пространства-времени, при котором и его *спинорная структура*¹⁾ будет более фундаментальной, чем псевдориманова. Конкретные размерность и сигнатура $(+, -, -, -)$ должны будут закладываться в теорию изначально. Таким образом, если бы возникло изменение сегодняшней картины пространства-времени как дифференцируемого многообразия, существенно зависящее от наличия такой спинорной структуры, то размерность и сигнатура нашего пространства-времени были бы одним из следствий теории.

Наше построение будет исходить из изоморфизма²⁾ между группой $SL(2, \mathbb{C})$ унимодулярных комплексных 2×2 -матриц и двукратным (универсальным) покрытием связной компоненты группы Лоренца $O(1, 3)$. Чтобы выразить этот изоморфизм в наиболее явном виде, рассмотрим общие эрмитовы 2×2 -матрицы, которые я запишу здесь в виде

$$\begin{pmatrix} u^{00'} & u^{01'} \\ u^{10'} & u^{11'} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} u^0 + u^1 & u^2 + iu^3 \\ u^2 - iu^3 & u^0 - u^1 \end{pmatrix}. \quad (4.13)$$

¹⁾ Я говорю «спинорная структура», а не «спиновая структура», чтобы подчеркнуть, что \mathcal{M} не просто спиновое многообразие, но что его структура должна определяться *конкретным* типом используемой здесь спинорной системы.

²⁾ Для анализа структуры пространства-времени может быть также использован $(2-1)$ -изоморфизм между $SU(2, 2)$ и связной компонентой $O(2, 4)$ (ср. с [79, 80]).

При умножении матрицы (4.13) слева на унимодулярную комплексную 2×2 -матрицу ($t_{\mathfrak{B}}^{\mathfrak{A}}$), а справа — на комплексно-сопряженную ей сохраняется как эрмитовость, так и величина детерминанта (4.13). Мы получаем таким образом линейное преобразование (u^0, u^1, u^2, u^3), сохраняющее как вещественность u^a , так и форму

$$g_{ab} u^a u^b = (u^0)^2 - (u^1)^2 - (u^2)^2 - (u^3)^2, \quad (4.14)$$

где

$$(g_{ab}) = \text{diag}(1, -1, -1, -1). \quad (4.15)$$

Следовательно, для u^a мы получили преобразование Лоренца

$$u^a \rightarrow L_{\mathfrak{B}}^{\mathfrak{A}} u^b, \quad (4.16)$$

непрерывно переходящее в тождество (так как $t_{\mathfrak{B}}^{\mathfrak{A}}$ непрерывно переходит в $\delta_{\mathfrak{B}}^{\mathfrak{A}}$). И наоборот, из ряда соображений следует, что любое преобразование типа (4.16) следует из *двух и только двух* унимодулярных матриц ($t_{\mathfrak{B}}^{\mathfrak{A}}$), а именно из ($t_{\mathfrak{B}}^{\mathfrak{A}}$) и ($\bar{t}_{\mathfrak{B}}^{\mathfrak{A}}$). Поэтому преобразование (4.16) эквивалентно преобразованию

$$u^{\mathfrak{A}\mathfrak{A}'} \rightarrow t_{\mathfrak{B}}^{\mathfrak{A}} \bar{t}_{\mathfrak{B}}^{\mathfrak{A}'} u^{\mathfrak{B}\mathfrak{B}'}. \quad (4.17)$$

[В этих обозначениях \mathfrak{A} и \mathfrak{A}' считаются *разными* буквами, так что, в частности, по ним не подразумевается суммирование. Следовательно, (4.17) — это *четыре* соотношения. При комплексном сопряжении нештрихованные индексы становятся штрихованными, а штрихованные — нештрихованными. Так, например, величина $\bar{t}^{0'}$, является комплексно-сопряженной по отношению к t^0 . Прописные готические индексы принимают два значения: 0, 1 или 0', 1'. Строчные готические индексы принимают значения 0, 1, 2, 3.]

Обычно компоненты мирового вектора в псевдоортогональной системе координат представляют в виде линейно упорядоченной последовательности (u^0, u^1, u^2, u^3). Однако матричная запись (4.13) несколько не хуже, и мы намереваемся рассматривать ее как *более* фундаментальную. С этой точки зрения

мировой вектор трактуется уже не как самая элементарная «векторная» величина в пространстве-времени, поскольку это теперь «бивалентный» объект и можно ожидать обнаружения неких «унивалентных» (двух-компонентных) объектов, имеющих какое-то пространственно-временное значение. Так оно и есть на самом деле. Назовем эти «унивалентные» объекты *спин-векторами*. Идея состоит тогда в том, чтобы взять в качестве основного модуля (V^\bullet в разд. 2) не множество полей мировых векторов T^\bullet , а *спин-векторные* поля S^\bullet класса C^∞ ¹⁾ на \mathcal{M} . Тот факт, что я еще не дал определения спин-вектора на языке пространства-времени, не играет роли. Процедура будет скорее обратной, и алгебра, определяемая спин-векторами, послужит для определения структуры пространства-времени. Найдя эту структуру, мы сможем затем вернуться обратно и дать трактовку спин-векторов на более привычном языке. В конце концов мы сможем даже дать на редкость полное «физическое» истолкование спин-вектора (см. разд. 5).

Какой же внутренней структурой должен обладать наш модуль S^\bullet спин-векторных полей? Мы требуем, чтобы около каждой точки $P \in \mathcal{M}$ слой $S^\bullet(P)$ над P был двумерным комплексным векторным пространством («спиновым пространством»). Симметрия этого пространства задается группой $SL(2, \mathbb{C})$, так что в этой симметрии отражается симметрия локальной группы Лоренца пространства-времени. Если ω^0, ω^1 — компоненты $\omega \in S^\bullet(P)$ в некотором базисе, то (активное) преобразование симметрии для $S^\bullet(P)$ дается в виде

$$\omega^{\mathfrak{A}} \rightarrow t_{\mathfrak{B}\mathfrak{A}}^{\mathfrak{A}} \omega^{\mathfrak{B}} \quad (4.18)$$

(при этом базис остается фиксированным); здесь матрица $(t_{\mathfrak{B}\mathfrak{A}}^{\mathfrak{A}})$ — комплексная и унимодулярная. При сравнении (4.18) с (4.17) и (4.16) через (4.13) заметим, что если рассматривать пространство $T^\bullet(P)$ мировых векторов как тензорное произведение $S^\bullet(P)$

¹⁾ Окажется, что система S^\bullet есть сечение класса C^∞ пучка спин-векторов над \mathcal{M} , рассматриваемое как модуль над комплексными C^∞ -функциями на \mathcal{M} .

и комплексно-сопряженного ему, то допустимые (активные) симметрии в $\mathbf{T}^*(P)$ оказываются именно преобразованиями Лоренца, не включающими пространственных отражений и обращения времени. Каждый элемент модуля \mathbf{S}^* будет непересекающимся объединением всех $P \in \mathcal{M}$ — по одному элементу от каждого $\mathbf{S}^*(P)$. При этом накладывается и некоторая дифференциальная структура. Мы дадим краткое аксиоматическое определение. Назовем объект, построенный из тензорных произведений \mathbf{S}^* , дуального ему \mathbf{S} , и комплексно-сопряженных им полей, *спинорным полем* на \mathcal{M} . Как видно, частным случаем здесь будет поле мирового вектора. При этом «тензорное произведение» следует понимать в смысле, указанном в разд. 3. Принятая система значков окажется весьма ценной при контроле за различными операциями.

Нам понадобятся спинорные значки двух типов

$$A, B, C, \dots, A_0, \dots, A_1, \dots, \quad (4.19)$$

$$A', B', C', \dots, A'_0, \dots, A'_1, \dots \quad (4.20)$$

Каждому $\omega \in \mathbf{S}^*$ будут соответствовать элементы $\omega^A \in \mathbf{S}^A$, $\omega^B \in \mathbf{S}^B$, ..., $\omega^X \in \mathbf{S}^X$, ..., а также элементы $\bar{\omega}^{A'} \in \mathbf{S}^{A'}$, $\bar{\omega}^{B'} \in \mathbf{S}^{B'}$, ..., $\bar{\omega}^{X'} \in \mathbf{S}^{X'}$, Как и прежде, мы при желании можем представлять себе ω^X просто как *пару* (ω, X) и т. д. Кольцо \mathbf{S} будет трактоваться как *комплексные* C^∞ -функции на \mathcal{M} :

$$\mathbf{S} = \mathbf{T} \oplus i\mathbf{T}. \quad (4.21)$$

Различие между штрихованными и нештрихованными индексами возникает из следующего требования: если каждый модуль \mathbf{S}^A , \mathbf{S}^B , ... над \mathbf{S} канонически изоморфен \mathbf{S}^* , то каждый модуль $\mathbf{S}^{A'}$, $\mathbf{S}^{B'}$, ... над \mathbf{S} канонически *антиизоморфен* \mathbf{S}^* . Иначе говоря, если $\lambda, \mu \in \mathbf{S}$, а $\alpha, \beta, \gamma \in \mathbf{S}^*$, причем $\gamma = \lambda\alpha + \mu\beta$, то

$$\gamma^X = \lambda\alpha^X + \mu\beta^X \quad \text{и} \quad \bar{\gamma}^{X'} = \bar{\lambda}\bar{\alpha}^{X'} + \bar{\mu}\bar{\beta}^{X'}. \quad (4.22)$$

Таким образом, мы будем рассматривать каждый элемент $\bar{\omega}^X$ как *комплексно-сопряженный* по отношению

к соответствующему ω^X и будем иногда писать

$$\overline{\omega^X} = \bar{\omega}^{X'}, \quad \overline{\pi^{X'}} = \bar{\pi}^X. \quad (4.23)$$

[Можно представлять себе операцию комплексного сопряжения как замену списка (4.19) на список (4.20), когда эта операция распространяется и на индексы.] Спинорная система строится тогда в точности, как в разд. 3, но с тем условием, что свертывание и замена индексов могут касаться лишь индексов *одного и того же* типа (т. е. лишь штрихованных или лишь нештрихованных). Так как замена индексов не приводит к перестановкам штрихованных индексов с нештрихованными, относительную расстановку штрихованных и нештрихованных индексов мы будем считать несущественной. Напротив, будет важно (в большинстве случаев) следить за порядком индексов *одного и того же* типа, *даже если* одни из них верхние, а другие — нижние. Так, например,

$$\rho^{AA'}_{B'B}{}^Q = \rho^{A'A}{}^Q_{B'B'} = \rho^{AA'Q}_{B'B'} \neq \rho^{AA'Q}_{BB'}. \quad (4.24)$$

Комплексное сопряжение — это пятая операция, в дополнение к (3.11)—(3.14), действующая на спинорную систему $\{S\}$:

комплексное сопряжение:

$$S^X_P \dots Z^{W'} \dots Y'_{U'} \rightarrow S^W_Q \dots Y_{X'} \dots Z'_{S'}. \quad (4.25)$$

Эта операция применима очевидным образом [ср. с (4.23)] к каждому элементу $\rho^X \dots Z^{W'} \dots Y'_{U'}$, представленному в виде разложения типа (3.9).

Мы еще не отразили в аппарате спинорной алгебры того факта, что матрицы $(t^X_{\mathfrak{B}})$ должны быть унимодулярными. Условие унимодулярности можно записать либо в виде

$$\epsilon^{\mathfrak{A}\mathfrak{C}} = t^{\mathfrak{A}}_{\mathfrak{B}} t^{\mathfrak{C}}_{\mathfrak{D}} \epsilon^{\mathfrak{B}\mathfrak{D}}, \quad (4.26a)$$

либо в виде

$$\epsilon_{\mathfrak{B}\mathfrak{D}} = t^{\mathfrak{A}}_{\mathfrak{B}} t^{\mathfrak{C}}_{\mathfrak{D}} \epsilon_{\mathfrak{A}\mathfrak{C}}. \quad (4.26b)$$

Здесь использован символ Леви-Чивиты

$$(\epsilon_{\mathbb{R}}) = (\epsilon^{\mathbb{R}}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (4.27)$$

Обозначим элементы S_{AB} и S^{AB} , компоненты которых в некотором базисе имеют вид (4.27), соответственно как ϵ_{AB} и ϵ^{AB} . Тогда из (4.26) видно, что ϵ_{AB} и ϵ^{AB} инвариантны относительно активного преобразования симметрии $S^*(P)$, имеющего вид (4.18). Мы будем рассматривать ϵ_{AB} , ϵ^{AB} и комплексно-опряженные им $\epsilon_{A'B'}$, $\epsilon^{A'B'}$ как основной элемент внутренней структуры спинорной системы. (Более последовательно было бы писать $\bar{\epsilon}_{A'B'}$ вместо $\epsilon_{A'B'}$ и $\bar{\epsilon}^{A'B'}$ вместо $\epsilon^{A'B'}$, но обычно черточки опускают.) Любой базис, в котором компоненты $\epsilon_{\mathbb{R}}$ имеют вид (4.27) называется нормированным базисом, или *спиновой системой отсчета*.

Символ Леви-Чивиты, в частности, обеспечивает установление канонического изоморфизма между S^* и дуальным ему S_* . Это достигается путем поднятия и опускания спинорных значков:

$$\xi^{\dots A \dots} \epsilon_{AB} = \xi^{\dots B \dots}; \quad \epsilon^{AB} \xi_{\dots B \dots} = \xi^{\dots A \dots}; \quad (4.28)$$

$$\eta^{\dots A' \dots} \epsilon_{A'B'} = \eta^{\dots B' \dots}; \quad \epsilon^{A'B'} \eta_{\dots B' \dots} = \eta^{\dots A' \dots}. \quad (4.29)$$

(В связи с этим возникает ощущение, что использование сразу и верхних и нижних значков излишне. Сохранение различий между ними, в сущности, вопрос «бухгалтерского учета». На самом деле ведь можно пользоваться обозначениями, *стицающими* эти различия, но тогда мы столкнемся с непривычной ситуацией, когда спин-векторы антикоммутируют! Я слеую в этом отношении традициям.) Замечим, что

$$\xi^{\dots X \dots} \epsilon_{\dots X \dots} = - \xi^{\dots X \dots} \epsilon_{\dots X \dots} \quad (4.30)$$

(и аналогично для штрихованных индексов) вследствие антисимметрии

$$\epsilon_{AB} = - \epsilon_{BA}, \quad \epsilon^{AB} = - \epsilon^{BA}. \quad (4.31)$$

Поэтому нужно внимательно следить за порядком индексов, поднимая и опуская их. По этой же причине, а также во избежание недоразумений в обозначениях на дальнейших этапах, мы не будем пользоваться символами δ_B^A и $\delta_{B'}^{A'}$ типа (3.18). Вместо них введем эквивалентные символы ϵ_B^A и $\epsilon_{B'}^{A'}$, для которых

$$\epsilon_{AB}\epsilon^{CB} = \epsilon_A^C = -\epsilon^C_A; \quad \epsilon_A^A = 2; \quad (4.32)$$

$$\xi_{\dots B \dots} = \xi_{\dots A \dots} \epsilon_B^A; \quad \xi_{\dots}^{B \dots} = \xi_{\dots}^{A \dots} \epsilon_A^B. \quad (4.33)$$

Итак, ϵ_B^A обладает свойствами (3.19), которых мы требовали бы от δ_B^A . (Отметим, что операция поднятия и опускания индексов здесь согласуются друг с другом.) При любом выборе спиновой системы отсчета $\epsilon_{\mathfrak{A}}^A, \epsilon_A^{\mathfrak{A}}$ имеем

$$(\epsilon_{\mathfrak{A}}^{\mathfrak{B}}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (4.34)$$

Если спинор антисимметричен по трем или более индексам, он тождественно равен нулю (в силу двумерности). Поэтому

$$\epsilon_{AB}\epsilon_{CD} + \epsilon_{AC}\epsilon_{DB} + \epsilon_{AD}\epsilon_{BC} = 0. \quad (4.35)$$

Одно из следствий (4.35) состоит в том, что если $\xi_{\dots A \dots B \dots}$ антисимметричен по A и B , т. е.

$$\xi_{\dots A \dots B \dots} = -\xi_{\dots B \dots A \dots}, \quad (4.36)$$

то

$$\xi_{\dots A \dots B \dots} = \frac{1}{2} \epsilon_{AB} \xi_{\dots X \dots}^X. \quad (4.37)$$

Это свойство тесно связано с представлением группы $SL(2, \mathbf{C})$, так как указывает на тот факт, что лишь спиноры, полностью симметричные по всем нештрихованным индексам и по всем штрихованным индексам, не могут быть сведены к более простым. Говорят, что компоненты спинора с r симметричными нижними нештрихованными индексами и с s симметричными нижними штрихованными индексами преобразуются по неприводимому представлению $D(r/2, s/2)$ группы Лоренца.

Включим теперь в нашу спинорную схему и мировые тензоры, введя тензорные значки (3.1) как *парные комбинации* спинорных значков:

$$a = AA', \quad b = BB', \quad c = CC', \quad \dots, \quad a_0 = A_0 A'_0, \quad \dots \quad (4.38)$$

Таким образом получим подсистему $(\mathfrak{S}, \mathbf{S}^a, \mathbf{S}^b, \dots, \mathbf{S}_a, \dots, \mathbf{S}_{u \dots w}^{x \dots z}, \dots)$ комплексных мировых тензоров, замкнутую относительно операций (3.11) — (3.14), а также относительно операции комплексного сопряжения (4.25), действие которой здесь сводится к $\mathbf{S}_{u \dots w}^{x \dots z} \rightarrow \overline{\mathbf{S}_{u \dots w}^{x \dots z}}$. Инвариантными при комплексном сопряжении элементами являются *вещественные мировые тензоры*: $\overline{\alpha_{u \dots w}^{x \dots z}} = \alpha_{u \dots w}^{x \dots z}$. Поля вещественных мировых тензоров образуют подсистему $\{\mathbf{T}\}$ полной системы $\{\mathbf{S}\}$:

$$\{\mathbf{T}\} = (\mathbf{T}, \mathbf{T}^a, \mathbf{T}^b, \dots, \mathbf{T}_a, \dots, \mathbf{T}_{u \dots w}^{x \dots z}, \dots). \quad (4.39)$$

Покажем, что этот процесс — именно то, что содержится в сопоставлении (4.13); для этого рассмотрим спиновую систему отсчета $\epsilon_{\mathfrak{A}}^A, \epsilon_{\mathfrak{I}}^{\mathfrak{A}}$ для \mathbf{S}^\bullet и определим (при $\epsilon_{A'}^{\mathfrak{A}'} = \overline{\epsilon_A^{\mathfrak{A}'}}$)

$$\begin{aligned} \delta_a^0 &= \frac{1}{\sqrt{2}} (\epsilon_A^0 \epsilon_{A'}^{0'} + \epsilon_A^1 \epsilon_{A'}^{1'}), \\ \delta_a^1 &= \frac{1}{\sqrt{2}} (\epsilon_A^0 \epsilon_{A'}^{0'} - \epsilon_A^1 \epsilon_{A'}^{1'}), \\ \delta_a^2 &= \frac{1}{\sqrt{2}} (\epsilon_A^1 \epsilon_{A'}^{0'} + \epsilon_A^0 \epsilon_{A'}^{1'}), \\ \delta_a^3 &= \frac{i}{\sqrt{2}} (\epsilon_A^1 \epsilon_{A'}^{0'} - \epsilon_A^0 \epsilon_{A'}^{1'}). \end{aligned} \quad (4.40)$$

Отсюда непосредственно следует (4.13), причем $u^a = u^{AA'}$ и

$$u^a = u^a \delta_a^a, \quad u^{\mathfrak{A}\mathfrak{A}'} = u^{AA'} \epsilon_A^{\mathfrak{A}} \epsilon_{A'}^{\mathfrak{A}'} \quad (4.41)$$

в согласии с условием (3.20). Формулы (4.40) дают нам стандартный способ определения мирового векторного базиса для любой данной спиновой системы

отсчета. Важно, однако, что мы не обязаны всегда пользоваться этим конкретным методом сопоставления. Иногда мы будем отдавать предпочтение мировому векторному базису, полученному естественным образом с помощью *системы координат*. В других случаях удобный базис задается *световой тетрадой*¹⁾

$$\begin{aligned}\delta_0^a &= \varepsilon_0^A \varepsilon_{0'}^{A'}, & \delta_1^a &= \varepsilon_1^A \varepsilon_{1'}^{A'}, \\ \delta_2^a &= \varepsilon_0^A \varepsilon_{1'}^{A'}, & \delta_3^a &= \varepsilon_1^A \varepsilon_{0'}^{A'}.\end{aligned}\quad (4.42)$$

В любом случае можно определить переводящие символы Инфельда — ван-дер-Вердена [46, 110]:

$$\sigma_a^{\mathfrak{A}\mathfrak{A}'} = \delta_a^A \varepsilon_A^{\mathfrak{A}} \varepsilon_{A'}^{\mathfrak{A}'}, \quad \sigma_a^{\mathfrak{A}} = \delta_a^A \varepsilon_A^{\mathfrak{A}} \varepsilon_{A'}^{A'}. \quad (4.43)$$

(Вспомним, что $a = AA'$; таким образом, свертывание имеет место для *абстрактных* значков. С другой стороны, не вводится никакой связи между a и \mathfrak{A} , \mathfrak{A}' , так что по ним нет суммирования. Каждое из соотношений (4.43) — это 16 уравнений.) Символы Инфельда — ван-дер-Вердена — это в действительности просто δ_a^b , только в них верхние и нижние индексы выражены на языке разных базисов. Пример связи между спинорными и мировыми тензорными компонентами величин можно записать как

$$\begin{aligned}\psi_a^b &= \psi_{\mathfrak{A}\mathfrak{A}'}^{\mathfrak{B}\mathfrak{B}'} \sigma_a^{\mathfrak{A}\mathfrak{A}'} \sigma_b^{\mathfrak{B}\mathfrak{B}'}, \\ \psi_{\mathfrak{A}\mathfrak{A}'}^{\mathfrak{B}\mathfrak{B}'} &= \psi_a^b \sigma_{\mathfrak{A}\mathfrak{A}'}^a \sigma_b^{\mathfrak{B}\mathfrak{B}'}.\end{aligned}\quad (4.44)$$

Рассмотрим теперь удвоенный детерминант матрицы (4.13):

$$u^{\mathfrak{A}\mathfrak{A}'} u^{\mathfrak{B}\mathfrak{B}'} \varepsilon_{\mathfrak{A}\mathfrak{B}} \varepsilon_{\mathfrak{A}'\mathfrak{B}'} = u^a u^b g_{ab}. \quad (4.45)$$

При обратном преобразовании к абстрактным значкам это дает

$$u^a u^b (\varepsilon_{AB} \varepsilon_{A'B'} - g_{ab}) = 0 \quad (4.46)$$

¹⁾ Автор пользуется термином «*null*», что означает изотропный, светоподобный. Вместе с тем, часто встречается термин «*null cone*» — *световой конус*. Для унификации в русском переводе делается попытка писать «световой», хотя, конечно, комплексные изотропные векторы типа (5.3), составленные из пространственноподобных частей, «световыми» называть не следует. — *Прим. перев.*

[ср. с (3.21)], и мы получаем [ввиду симметрии g_{ab} и $\epsilon_{AB}\epsilon_{A'B'}$ при перестановке a и b и ввиду справедливости (4.46) для всех u^a]

$$g_{ab} = \epsilon_{AB}\epsilon_{A'B'}. \quad (4.47)$$

Уравнение (4.47) представляет собой фундаментальное соотношение, с помощью которого метрика пространства-времени определяется его спинорной структурой. Отметим, в частности, что *сигнатура* $(+, -, -, -)$ *автоматически включается в формализм* через соотношение (4.47), так как выбор мирового векторного базиса (4.40) дает нам значение (4.15). Вследствие (4.47), (4.32) и (4.31) получим

$$\begin{aligned} g^{ab} &= \epsilon^{AB}\epsilon^{A'B'}, & \delta_a^b &= \epsilon_A^B\epsilon_{A'}^{B'}, \\ g_{ab} &= g_{ba}, & g_{ab}g^{bc} &= \delta_a^c, \end{aligned} \quad (4.48)$$

так что правила поднятия и опускания спинорных индексов приводят к тензорным правилам

$$\xi_{\dots a \dots} \dots = \xi_{\dots}^{\dots b \dots} g_{ab}; \quad \xi_{\dots}^{\dots b \dots} = \xi_{\dots a \dots} \dots g^{ab}. \quad (4.49)$$

Итак, наш метод согласуется с обычными определениями.

Однако мы еще не связали локальную спинорную структуру с многообразием \mathcal{M} в целом. Чтобы связать различные $\mathbf{S}^\bullet(P)$ друг с другом, необходимо рассмотреть операцию дифференцирования. Дадим аксиоматическое определение ряда правил, которые нам понадобятся. Нам потребуется ввести операцию *ковариантного дифференцирования*

$$\nabla_{XY'} : \mathbf{S}_{F \dots MH' \dots K'}^{P \dots SR' \dots T'} \rightarrow \mathbf{S}_{F \dots MXH' \dots KY'}^{P \dots SR' \dots T'}, \quad (4.50)$$

обладающую свойствами

$$\nabla_{XY'}(\psi_{\dots} + \chi_{\dots}) = \nabla_{XY'}\psi_{\dots} + \nabla_{XY'}\chi_{\dots}; \quad (4.51)$$

$$\nabla_{XY'}(\psi_{\dots}\chi_{\dots}) = \chi_{\dots}\nabla_{XY'}\psi_{\dots} + \psi_{\dots}\nabla_{XY'}\chi_{\dots}; \quad (4.52)$$

$$\nabla_{XY'}\epsilon_{AB} = 0, \quad \nabla_{XY'}\epsilon^{AB} = 0; \quad (4.53)$$

$$\text{Если } \psi_{\dots} = \nabla_{XY} \chi_{\dots}, \text{ то } \overline{\psi_{\dots}} = \nabla_{YX} \overline{\chi_{\dots}}; \quad (4.54)$$

$$\text{Если } \psi_{\dots A \dots B \dots} = \nabla_{XY} \chi_{\dots A \dots B \dots}, \text{ то } \psi_{\dots A \dots B \dots} = \\ = \nabla_{XY} \chi_{\dots A \dots}; \quad (4.55)$$

Операция ∇_{XY} коммутрует с любой заменой индексов (не включающих X или Y); (4.56)

$$\nabla_x \nabla_z \varphi = \nabla_z \nabla_x \varphi; \quad (4.57)$$

Для любого производного ξ на \mathbf{T} существует единственный элемент $\xi^a \in \mathbf{T}^a$, такой, что $\xi(\varphi) = \xi^a \nabla_a \varphi$ для любого $\varphi \in \mathbf{T}$. (4.58)

В действительности аксиомы (4.51) — (4.58) не все друг от друга независимы, однако в них выражены все свойства, которые нам нужны. Отметим, что свойство (4.58) связывает $\{\mathbf{S}\}$ с пучком, касательным к \mathcal{M} . Мировые векторы оказываются векторами, касательными к \mathcal{M} . Все эти свойства однозначно определяют операцию ∇_{XY} . (Ясно, что для любой другой пары значков, скажем M, N' , мы получим $\nabla_{MN'}$ путем замены индексов.) Наоборот, можно было бы задаться вопросом, не накладывает ли факт существования системы $\{\mathbf{S}\}$, удовлетворяющей всем сформулированным условиям, каких-либо ограничений на \mathcal{M} как на дифференцируемое многообразие или как на псевдориманово многообразие с данной сигнатурой метрики $(+, -, -, -)$ [с которой по предположению согласуется g_{ab} , данное как (4.47)]. На это можно ответить, что локально никаких ограничений не возникает (т. е. $\{\mathbf{S}\}$ существует для некоторого открытого множества, содержащего любое данное $P \in \mathcal{M}$, однако глобальные ограничения имеются. Аналогичное замечание можно сделать и относительно единственности $\{\mathbf{S}\}$. Мы перейдем к обсуждению этих глобальных ограничений на \mathcal{M} в следующем разделе.

5.

Истолкование спин-вектора

В предыдущем разделе спиноры были введены совершенно формально. Непосредственную интерпретацию на языке пространства-времени получили лишь элементы $\{\mathbf{T}\}$. Можно было бы думать, что любая геометрическая интерпретация остальных элементов $\{\mathbf{S}\}$ должна быть по необходимости лишь сугубо косвенной. Оказывается, однако, что это не так. В действительности любой спин-вектор может быть интерпретирован в пространстве-времени даже графически (с точностью до знака). И даже знаку можно приписать свой смысл на языке физических построений.

Перейдем к истолкованию спин-вектора

$$\omega^A \in \mathbf{S}^A(P).$$

Самым простым мировым вектором, который можно построить из ω^A , является мировой вектор

$$u^a = \omega^A \bar{\omega}^{A'}. \quad (5.1)$$

Тогда $u^a u_a = |\omega^A \omega_A|^2 = 0$ (так как $\omega^A \omega_A = = -\omega_A \omega^A = 0$); следовательно, вектор u^a является образующей светового конуса с вершиной в точке P . При этом *всякий* световой мировой вектор имеет либо вид (5.1), либо вид

$$v^a = -\omega^A \bar{\omega}^{A'}, \quad (5.2)$$

тогда как всякий комплексный мировой вектор, изотропный в смысле $\omega^a \omega_a = 0$, имеет вид

$$\omega^a = \omega^A \pi^{A'}. \quad (5.3)$$

Эти выводы следуют непосредственно из представления (4.13), так как ранг 2×2 -матрицы, детерминант которой равен нулю, не может превышать единицы,

и поэтому такая матрица является внешним произведением двух векторов. Отметим, что наличие спинорной системы $\{\mathbf{S}\}$ приводит к абсолютному различию между двумя световыми полуконусами в точке P : можно *определить* как *направленные в будущее* те световые векторы, для которых верно (5.1) (при $\omega^A \neq 0$), и как *направленные в прошлое* — те световые векторы, для которых верно (5.2) (при $\omega^A \neq 0$). [Эти определения согласуются с (4.13), если принять, что положительность u^0 соответствует возрастанию времени.] Отсюда следует первое глобальное ограничение на \mathcal{M} как псевдориманово многообразие, обусловленное наличием $\{\mathbf{S}\}$: многообразие \mathcal{M} должно быть *ориентируемым во времени*. Иначе говоря, деление световых полуконусов в \mathcal{M} на два класса — полуконусы будущего и прошлого — может быть осуществлено непрерывным образом на всем многообразии.

Итак, из уравнения (5.1) следует, что направленный в будущее вещественный световой вектор однозначно определяется любым ненулевым спин-вектором. Однако одному и тому же световому мировому вектору соответствует множество разных спин-векторов, так как равенство (5.1) инвариантно относительно преобразования

$$\omega^A \rightarrow e^{i\theta} \omega^A, \quad (5.4)$$

где θ вещественно. Чтобы истолковать «фазу» ω^A , необходимо пойти иным путем. Рассмотрим для этого «квадрат» $\omega^A \omega^B$ и, чтобы ввести такое же количество штрихованных индексов, умножим его на $\epsilon^{A'B'}$. Наконец, добавив сюда комплексно-сопряженное выражение, получим вещественный мировой тензор

$$p^{ab} = \omega^A \omega^B \epsilon^{A'B'} + \epsilon^{AB} \bar{\omega}^{A'} \bar{\omega}^{B'}. \quad (5.5)$$

Теперь уже видно, что p^{ab} антисимметричен и прост:

$$p^{ab} = u^a k^b - k^a u^b, \quad (5.6)$$

где

$$k^a = \omega^A \bar{\kappa}^{A'} + \kappa^A \bar{\omega}^{A'}. \quad (5.7)$$

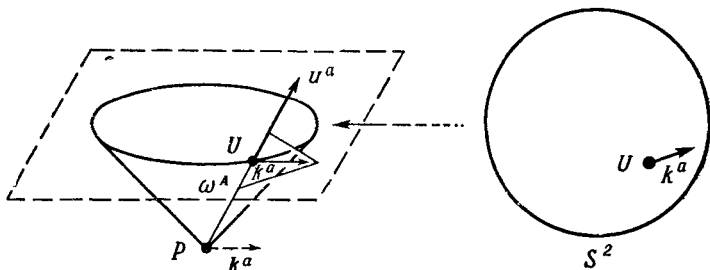


Рис. 7. Спин-вектор ω^A в точке P определяет световой флаг. Это можно изобразить, построив вектор, касательный к небесной сфере в точке U .

Здесь κ^A — произвольный элемент $\mathbf{S}^A(P)$, для которого $\omega_A \kappa^A = 1$, причем из (4.37) следует $\omega^A \kappa^B = \kappa^A \omega^B = \varepsilon^{AB}$. Вектор k^a является вещественным и пространственноподобным, его модуль равен $\sqrt{2}$, и он ортогонален вектору u^a :

$$\bar{k}^a = k^a, \quad k^a k_a = -2, \quad k^a u_a = 0. \quad (5.8)$$

Этот вектор определяется бивектором p^{ab} с точностью до (вещественного) слагаемого, пропорционального u^a . Поэтому возможные направления k^a определяют элемент двумерной полуплощадки в P , касательный к световому конусу вдоль светового направления u^a . Можно сказать, что ω^A определяет *световой флаг* [70, 73, 114] в касательном пространстве при P , причем флагштоком служит направленный в будущее световой вектор u^a , а полотнище флага есть световая 2-площадка, в которой лежит u^a (рис. 7).

Чтобы полнее обрисовать природу площадки этого флага, рассмотрим сечение касательного пространства в P пространственноподобной гиперплоскостью, не проходящей через P , вырезав тем самым конус в сфере S^2 . Флагшток пересекает S^2 в точке U , а площадка флага дает в U *единичный касательный вектор* к S^2 (его можно отождествить с вектором $k^a/\sqrt{2}$ в том случае, если он выбран параллельным нашей гиперповерхности). Сфере S^2 можно дать не-

посредственное физическое истолкование: это «небесная сфера» (или поле зрения) для наблюдателя в P . (Пространственноподобная гиперплоскость может быть выбрана ортогональной мировой линии наблюдателя в P .) Фотон, приходящий в P , обладает мировой линией (световой геодезической) со световым касанием в P , так что может быть однозначно связан с точкой U на S^2 . Можно также описывать поляризацию этого фотона с помощью площадки флага, т. е. вектора, касательного к небесной сфере в U . Следовательно, спин-вектор оказывается естественным средством для описания поляризованного фотона. В разд. 8 эта связь между спинорами и полями частиц нулевой массы покоя, обладающими различными значениями спина, будет рассмотрена более подробно.

Здесь уместно сделать одно заключительное замечание относительно структуры небесной сферы. Хотя риманова метрика, введенная на S^2 описанным выше построением, зависит от нашего выбора пространственноподобной гиперплоскости, конформная структура S^2 , напротив, определена жестко. Это можно увидеть, например, вернувшись к (4.18), откуда при $\xi = \omega^0/\omega^1$ следует дробно-линейное преобразование $\xi \rightarrow (t^0_0\xi + t^0_1)/(t^1_0\xi + t^1_1)$. Комплексное число ξ определяет точку U , а связь между комплексной ξ -плоскостью и S^2 задается стереографической проекцией [71]. Оба эти преобразования — конформные (так, они переводят окружности в окружности). Поскольку преобразование (4.18) соответствует преобразованию Лоренца, небесные сферы двух наблюдателей в P оказываются связанными друг с другом конформным преобразованием [71, 106] (так что окружности, наблюдаемые одним наблюдателем, останутся окружностями и для другого). Теперь мы переходим к другому способу инвариантного введения конформной структуры S^2 .

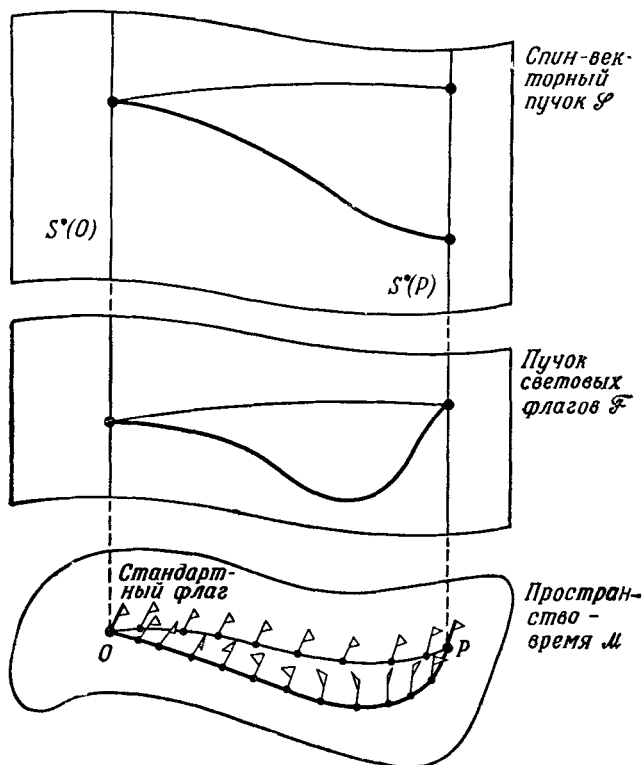
При умножении ω^A на отличное от нуля комплексное число $\lambda = re^{i\theta}$ (где r и θ вещественны) вектор u^a умножается на r^2 , а k^a (при $\kappa^A \rightarrow \lambda^{-1}\kappa^A$) поворачивается на угол 2θ в плоскости, натянутой на k^a и $k^a = i(\omega^A \bar{\kappa}^{A'} - \kappa^A \bar{\omega}^{A'})$. Таким образом, мы видим, что

преобразование $\omega^A \rightarrow \lambda \omega^A$ приводит к умножению длины флаштока на $\lambda \bar{\lambda}$ и к повороту площадки флага на угол $2 \arg \lambda$. Отметим, что тем самым задается прямой метод для определения конформной структуры сферы S^2 через спиноры. Если U является точкой на S^2 , то угол между двумя векторами, касательными к S^2 в U и заданными соответственно как ω^A и $e^{i\theta} \omega^A$, равен просто 2θ . Получив инвариантное определение углов на S^2 , мы тем самым нашли и инвариантное определение конформной структуры S^2 . К тому же имеем инвариантно определенную ориентацию для S^2 , а следовательно, и для \mathcal{M} . Можно просто определить понятие правого поворота как такой поворот, при котором площадка флага $e^{i\theta} \omega^A$ поворачивается в направлении возрастания θ . Это определение согласуется с понятием правой системы координат, если связать ее с (4.13). Отсюда следует второе глобальное ограничение на \mathcal{M} как псевдориманово многообразие: многообразии \mathcal{M} должно быть не только ориентируемым во времени, но и ориентируемым в пространстве. Поэтому на языке топологических многообразий \mathcal{M} можно просто назвать ориентируемым.

Третье и последнее глобальное ограничение на \mathcal{M} как псевдориманово (или топологическое) многообразие, обусловленное наличием $\{\mathbf{S}\}$, проявляется в связи с другим аспектом данного выше представления спин-вектора как светового флага. Заметим, что как u^a в (5.1), так и r^{ab} в (5.6) инвариантны относительно преобразования

$$\omega^A \rightarrow -\omega^A. \quad (5.9)$$

Если мы теперь рассмотрим площадку флага, заданную через $e^{i\theta} \omega^A$, то при изменении θ от 0 до 2π мы обнаружим, что она поворачивается при этом на 2π (полный оборот), возвращаясь в исходное положение, тогда как ω^A меняет знак. Спин-вектор вновь достигает своего первоначального значения лишь после еще одного полного поворота площадки флага на 2π . При этом ясно, однако, что никаким добавлением к световому флагу каких-либо новых локальных геометрических структур мы уже не сможем отличить ω^A от $-\omega^A$ (если понимать под «локальными геометриче-

Рис. 8. Флаг-пути на \mathcal{M} и их отображения на \mathcal{F} и \mathcal{S} .

скими структурами» построения в пространстве, касательном к \mathcal{M} в P). Дело в том, что поворот любой такой структуры на 2π вернет ее в первоначальное положение, тогда как ω^A преобразуется по закону (5.9).

Напротив, можно отличить ω^A от $-\omega^A$ с помощью *нелокального* геометрического построения. Для этого рассмотрим конкретный световой флаг в конкретной точке 0 из \mathcal{M} и назовем его *стандартным флагом*. Для простоты ограничимся анализом односвязной области в \mathcal{M} (либо будем работать с универсальным накрывающим \mathcal{M}). Идея тогда состоит в отождествлении

спин-вектора уже не со световым флагом, а с *классом эквивалентности флаг-путей*¹⁾. Флаг-путем назовем непрерывную последовательность световых флагов на \mathcal{M} , начинающуюся от стандартного флага и кончающуюся некоторым световым флагом в точке P из \mathcal{M} (рис. 8). Два флаг-пути будут эквивалентными, если их можно перевести друг в друга с помощью непрерывной деформации при сохранении конечных точек и световых флагов в этих точках фиксированными. Существование спин-векторов связывается теперь с существованием *двух* классов эквивалентности, соответствующих каждому световому флагу в P . Если рассматривать флаг-пути, основанные на *фиксированной* кривой на \mathcal{M} и кончающиеся фиксированным световым флагом, то из топологии группы Лоренца следует, что существует два и только два класса эквивалентности флаг-путей, связанных с этой кривой, различающихся результирующим относительным поворотом на 2π между O и P . Если же допустить на многообразии изменение кривой, на которой основывается флаг-путь, с ее дальнейшим возвращением в исходное положение, может оказаться, что эти классы эквивалентности объединятся в один класс эквивалентности. Тогда на многообразии нельзя будет определить спин-векторы. Если допустить, что P и O совпадают, то можно сформулировать это свойство другим способом. Пусть (замкнутый) флаг-путь дает просто поворот стандартного флага на угол 2π вокруг его флагштока; если теперь этот флаг-путь можно так непрерывно деформировать (смещая кривую в многообразии), что он станет постоянным (т. е. световой флаг не будет поворачиваться вообще), то спинорной структуры не существует.

Проделанное построение соответствует пучку световых флагов \mathcal{F} для \mathcal{M} (8-мерное пространство всех световых флагов во всех точках \mathcal{M}). Флаг-путь

¹⁾ Я сознательно допустил некоторую нечеткость, не описав геометрической структуры, соответствующей *равному нулю* спин-вектору в P , которую можно было бы назвать «световым флагом». Однако в случае *флаг-путей* «нулевые световые флаги» безусловно следует исключить.

соответствует кривой в \mathcal{F} . Когда \mathcal{F} односвязно, наше построение дает пучок спин-векторов \mathcal{P} (пространство всех спин-векторов во всех точках \mathcal{M}) в качестве универсального (двулистного) накрывающего для \mathcal{F} , если только данное \mathcal{F} обладает двулистным покрытием. Если же \mathcal{F} не обладает двулистным покрытием (является односвязным), то для \mathcal{M} не существует $\{S\}$. В случае, когда $\{S\}$ для \mathcal{M} существует, элементы S можно истолковать как сечения \mathcal{P} класса C^∞ .

Если \mathcal{M} не является односвязным, положение усложняется ввиду несравнимости флаг-путей, принадлежащих гомотопически неэквивалентным кривым на \mathcal{M} . При этом, если для \mathcal{M} спинорная структура $\{S\}$ и существует, то она оказывается не единственной. Но во всяком случае для установления существования спинорной структуры мы можем апеллировать к универсальному накрывающему для \mathcal{M} , и положение останется прежним.

Мы вели рассуждения в терминах пучка \mathcal{F} световых флагов, поскольку это представляется наиболее непосредственным геометрическим путем к пространственно-временному описанию спин-вектора. Чаше, однако, пользуются пучком всех ортонормированных систем на \mathcal{M} , и это лучше подходит для общего n -мерного анализа. Легко видеть, что обе процедуры равноценны в случае пространственно-временного многообразия. Кроме того, ввиду ориентируемости \mathcal{M} (что означает равенство нулю класса ω_1 Штейфеля — Уитни) нетрудно использовать приведенные выше аргументы и показать, что другое топологическое условие на \mathcal{M} , следующее из существования $\{S\}$, состоит в обращении в нуль класса ω_2 Штейфеля — Уитни [38, 57]. Это означает, что для каждой точки на этой 2-поверхности в \mathcal{M} можно задать непрерывную систему трех линейно независимых векторов. (Обсуждение этого вопроса см. в [113а, 116, 117].) Любопытно, что существуют «правдоподобные» пространственно-временные многообразия \mathcal{M}^4 , ориентируемые

¹⁾ Возьмите комплексную проективную плоскость \mathcal{P} (вещественное 4-многообразие) и задайте на ней (положительно определенную) риманову метрику класса C^∞ . Пусть t — вещественная

в пространстве и во времени и не содержащие замкнутых (или почти замкнутых) временноподобных линий, для которых тем не менее $\omega_2 \neq 0$. В таких \mathcal{M} система $\{S\}$ существовать не может.

Можно спросить, какие физические соображения свидетельствуют в пользу существования у пространства-времени спинорной структуры. В самом деле, какую степень физической реальности можно приписать объекту, не возвращающемуся в свое исходное состояние при повороте на 360° ? Эта картина не имеет ровно ничего общего с нашим опытом! Тем не менее волновые функции электронов, протонов, нейтронов, нейтрино и многих других частиц ведут себя именно так. Конечно, можно утверждать, что это ничего не доказывает: ведь волновая функция — понятие довольно туманное. И уж во всяком случае при повороте на угол 2π меняется-то лишь знак волновой функции в целом — а разве этот знак не считается как раз ненаблюдаемым? В недавно опубликованной интересной статье Ааронов и Сусскинд [1] утверждают, что эта точка зрения должна быть отброшена. Они пишут, как можно в принципе построить прибор, который именно в силу спинорной природы волнового уравнения для электрона ведет себя при поворотах как спинор в *макроскопических* масштабах. Этот прибор состоит из двух частей, при соединении которых возникает наблюдаемый электрический ток из одной части

C^∞ -функция на \mathcal{P} , имеющая лишь изолированные особые точки. Устраните эти точки из многообразия и с помощью единичного вектора, соответствующего градиенту t , постройте лоренцеву (+, —, —, —) псевдориманову метрику класса C^∞ . «Пространство-время», которое вы получите, будет ориентируемым во времени и в пространстве и в нем не будет замкнутых (т. е. $\approx S^2$) временноподобных кривых [фактически оно будет удовлетворять сильному требованию причинности (11.1)], но тем не менее у него не будет спиновой структуры ($\omega_2 \neq 0$). Я благодарен Р. Ботту и С. Смейлу за этот пример.

Примечание в корректуре. В недавно написанной работе (посланной в Journ. of Math. Phys.) Героч показал, что в случае некомпактного пространства-времени существование спинорной структуры эквивалентно глобальному существованию непрерывного поля ортонормированных тетрад. Он показал также, что наше предположение (4.5) фактически излишне.

в другую. Части прибора затем разделяются и одна часть поворачивается на 2π относительно другой. Затем их снова соединяют, и снова из одной части в другую течет наблюдаемый ток, только теперь его направление стало противоположным! Первоначальное направление тока восстанавливается, лишь если части прибора вновь разделить, повернуть их на угол 2π и присоединить друг к другу. Мы приходим, таким образом, к ситуации, когда геометрия прибора (в обычном смысле) еще не определяет его поведения. И лишь расширив понятие геометрии путем включения спинорных величин, мы сможем здесь снова утверждать, что поведение системы определяется ее геометрической конфигурацией.

Прибор Ааронова — Сусскинда дает «физический» способ изображения спин-вектора. Можно представить себе, что одна часть прибора скреплена с нашим стандартным флагом, тогда как другая движется в пространстве-времени вместе с другим световым флагом. Такой прибор при этом «следит» за четностью числа полных относительных оборотов — так, в сущности, мы получаем спин-вектор. Если бы удалось убедительно доказать, что такие части прибора могут (в принципе) сохранять свою «память» на больших (скажем, космологических) расстояниях и интервалах времени, мы получили бы решающий довод в пользу того, что ω_2 действительно равняется нулю для вселенной!

6.

Явные формулы для кривизны

Метод, изложенный в разд. 3, автоматически дает все общеизвестные основные формулы, включающие кривизну, символы Кристоффеля и коэффициенты вращения Риччи, как только будут введены известные основные определения. Мы перечислим их здесь вместе с параллельной записью различных спинорных формул, вытекающих из аппарата разд. 4.

Рассмотрим базис δ_a^a для T^a и (возможно, не связанный с ним) базис $\epsilon_{\mathfrak{A}}^A$ для S^A . Пусть δ_a^a и $\epsilon_A^{\mathfrak{A}}$ будут соответствующими дуальными базисами. Введем определения

$$\Gamma_{bc}^a = \delta_a^a \nabla_c \delta_b^a = -\delta_b^a \nabla_c \delta_a^a \quad (6.1)$$

и

$$\gamma_{\mathfrak{A}\mathfrak{B}\mathfrak{C}\mathfrak{D}} = \epsilon_{A\mathfrak{B}} \nabla_{\mathfrak{C}\mathfrak{D}} \epsilon_{\mathfrak{A}}^A = \epsilon_{\mathfrak{A}}^A \nabla_{\mathfrak{C}\mathfrak{D}} \epsilon_{\mathfrak{B}}^A \quad (6.2)$$

Здесь, в согласии с (3.20), $\nabla_c = \delta_c^c \nabla_c$ и

$$\nabla_{\mathfrak{C}\mathfrak{D}} = \epsilon_{\mathfrak{C}}^C \epsilon_{\mathfrak{D}}^D \nabla_{CD}.$$

Эквивалентность определений (6.1) и (6.2) непосредственно следует из правил (4.51)–(4.56), а также (3.17), (4.30) и (4.31). Величины Γ_{bc}^a конкретизируются либо как *коэффициенты вращения Риччи* [в том случае, когда базис выбран так, чтобы давать g_{ab} в специальном виде, например (4.15)], либо как *символы Кристоффеля* (при взятии координатного естественного базиса). Величины $\gamma_{\mathfrak{A}\mathfrak{B}\mathfrak{C}\mathfrak{D}}$ называют *спинорными коэффициентами* [64]. В качестве примера выведем обычные свойства симметрии символов Кристоффеля. Пусть x^0, x^1, x^2, x^3 — локальная система

координат на \mathcal{M} . Тогда связанный с ней координатный базис определяется через

$$\delta_a^a = \nabla_a x^a \quad (6.3)$$

и через дуальные им компоненты δ_a^a . Тогда мы имеем

$$\Gamma_{bc}^a = -\delta_b^a \nabla_c \nabla_a x^a = -\delta_c^a \delta_b^a \nabla_d \nabla_a x^a \quad (6.4)$$

величину, явно симметричную по c и b ввиду условия обращения в нуль кручения (4.57). Из (6.2) столь же непосредственно следует симметрия спиновых коэффициентов

$$\gamma_{\mathfrak{A}\mathfrak{B}\mathfrak{C}\mathfrak{D}} = \gamma_{\mathfrak{B}\mathfrak{A}\mathfrak{C}\mathfrak{D}}. \quad (6.5)$$

Определения (6.1) и (6.2) вместе с (3.20) и (3.21) дают обычные формулы ковариантного дифференцирования в компонентах, например

$$\begin{aligned} \delta_c^c \delta_b^b \delta_a^a (\nabla_c \xi^a_b) &= \delta_b^b \delta_a^a \nabla_c (\xi^b \delta_b^a \delta_c^a) = \\ &= \nabla_c \xi^a_b + \xi^b \Gamma_{bc}^a - \xi^a \Gamma_{bc}^a, \end{aligned} \quad (6.6)$$

$$\begin{aligned} \varepsilon_{\mathfrak{C}}^{\mathfrak{C}} \varepsilon_{\mathfrak{D}'}^{\mathfrak{D}' } \varepsilon_{\mathfrak{A}}^{\mathfrak{A}} \varepsilon_{\mathfrak{B}'}^{\mathfrak{B}' } (\nabla_{\mathfrak{C}\mathfrak{D}'} \theta_{\mathfrak{A}'}^{\mathfrak{B}'}) &= \varepsilon_{\mathfrak{A}}^{\mathfrak{A}} \varepsilon_{\mathfrak{B}'}^{\mathfrak{B}' } \nabla_{\mathfrak{C}\mathfrak{D}'} (\theta_{\mathfrak{C}}^{\mathfrak{C}} \varepsilon_{\mathfrak{A}'}^{\mathfrak{A}} \varepsilon_{\mathfrak{B}' }^{\mathfrak{B}'}) = \\ &= \nabla_{\mathfrak{C}\mathfrak{D}'} \theta_{\mathfrak{A}'}^{\mathfrak{B}' } - \theta_{\mathfrak{C}}^{\mathfrak{C}} \gamma_{\mathfrak{A}\mathfrak{C}\mathfrak{D}'}^{\mathfrak{B}' } + \theta_{\mathfrak{A}'}^{\mathfrak{B}' } \bar{\gamma}_{\mathfrak{C}\mathfrak{D}'}^{\mathfrak{B}' } \varepsilon_{\mathfrak{C}}^{\mathfrak{C}}. \end{aligned} \quad (6.7)$$

[Вспомним, что $\nabla_{..}$ — это просто обычный оператор частного дифференцирования (взятия градиента), когда он действует на величины без абстрактных значков.]

Определим теперь *тензор кривизны* и *спиноры кривизны*. Введем

$$R^a_{bcd} = 2\delta_a^a \nabla_{[c} \nabla_{d]} \delta_b^a \quad (6.8)$$

и, полагая,

$$\square_{AB} \equiv \nabla_{Y'} (A \nabla_{B'}^Y), \quad \square_{A'B'} \equiv \nabla_Y (A' \nabla_{B'}^Y), \quad (6.9)$$

определим

$$\Psi_{ABCD} = \varepsilon_{D\mathfrak{D}} \square_{(AB} \varepsilon_{C)}^{\mathfrak{D}}, \quad (6.10)$$

$$\Phi_{CDA'B'} = \varepsilon_{D\mathfrak{D}} \square_{A'B'} \varepsilon_C^{\mathfrak{D}}, \quad (6.11)$$

$$\Lambda = \frac{1}{6} \varepsilon_{A\mathfrak{D}} \square^{AB} \varepsilon_B^{\mathfrak{D}}. \quad (6.12)$$

Здесь принято условие, согласно которому квадратные скобки означают антисимметризацию по заключенным в них индексам, а круглые скобки — симметризацию, например

$$\alpha_{[ab]} = \frac{1}{2} (\alpha_{ab} - \alpha_{ba}), \quad b_{[abc]} = \frac{1}{6} (\beta_{abc} + \beta_{bca} + \beta_{cab} - \\ - \beta_{acb} - \beta_{cba} - \beta_{bac}), \quad \gamma_{(AB)} = \frac{1}{2} (\gamma_{AB} + \gamma_{BA}), \quad \text{и т. д.}$$

Так как в определениях (6.8) и (6.10) — (6.12) справа содержатся готические индексы, требуется еще доказать, что они фактически не зависят от выбора базиса. Это действительно так, ибо каждый из операторов $\nabla_{[c}\nabla_{d]}$, \square_{AB} и $\square_{A'B'}$ удовлетворяет соотношениям

$$\begin{aligned} \mathcal{D}(\xi_{\dots} + \eta_{\dots}) &= \mathcal{D}\xi_{\dots} + \mathcal{D}\eta_{\dots}, \\ \mathcal{D}(\xi_{\dots}\xi_{\dots}) &= \xi_{\dots}\mathcal{D}\xi_{\dots} + \xi_{\dots}\mathcal{D}\xi_{\dots}, \\ \mathcal{D}\varphi &= 0 \end{aligned} \quad (6.13)$$

[ср. с (6.16)], где вместо \mathcal{D} можно подставить любой из этих операторов. Чтобы произвести замену базиса в (6.8) и (6.10) — (6.12), возьмем

$$\delta_b^a \rightarrow T^a_{\tilde{x}} \delta_b^{\tilde{x}}, \quad \delta_a^a \rightarrow T^{(-1)}_{\tilde{z}} \delta_a^{\tilde{z}}, \quad (6.14)$$

$$\varepsilon_C^{\mathfrak{D}} \rightarrow t^{\mathfrak{D}}_{\tilde{x}} \varepsilon_C^{\tilde{x}}, \quad \varepsilon_{D\mathfrak{D}} \rightarrow t^{(-1)\mathfrak{G}}_{\mathfrak{D}} \varepsilon_{D\mathfrak{D}}, \quad (6.15)$$

где степень (-1) означает обратную матрицу. В силу (6.13) матрица преобразования всякий раз переносится через операцию дифференцирования, сокращаясь с обратной ей матрицей, так что выражения не изменяют своего вида. (Следует помнить, что абстрактные значки не преобразуются!)

Рассмотренные операторы связаны между собой соотношением

$$2\nabla_{[a}\nabla_{b]} \equiv \varepsilon_{A'B'} \square_{AB} + \varepsilon_{AB} \square_{A'B'}, \quad (6.16)$$

так что, например, $\square_{AB} \equiv \varepsilon^{A'B'} \nabla_{[a}\nabla_{b]}$. Определения (6.10) и (6.12) можно тогда объединить в виде

$$\varepsilon_{D\mathfrak{D}} \square_{AB} \varepsilon_C^{\mathfrak{D}} = \Psi_{ABCD} - 2\Lambda \varepsilon_D (A^E B)_C. \quad (6.17)$$

Теперь (6.17) в комбинации с (6.11) и при учете (6.16) дает

$$2\varepsilon_{D\mathcal{D}}\nabla_{[a}\nabla_{b]}\varepsilon_C^{\mathcal{D}} = \Psi_{ABCD}\varepsilon_{A'B'} - 2\Lambda\varepsilon_D(A\varepsilon_B)C\varepsilon_{A'B'} + \\ + \Phi_{CDA'B'}\varepsilon_{AB}. \quad (6.18)$$

В самом деле, величины Ψ_{ABCD} , Λ и $\Phi_{ABC'D'}$ являются [72, 115] *неприводимыми* частями R_{abcd} относительно локальных (лоренцевых) преобразований $SL(2, \mathbb{C})$:

$$R_{abcd} = \Psi_{ABCD}\varepsilon_{A'B'}\varepsilon_{C'D'} + \varepsilon_{AB}\varepsilon_{CD}\bar{\Psi}_{A'B'C'D'} + \\ + 2\Lambda\{\varepsilon_{AC}\varepsilon_{BD}\varepsilon_{A'B'}\varepsilon_{C'D'} + \varepsilon_{AB}\varepsilon_{CD}\varepsilon_{A'D'}\varepsilon_{B'C'}\} + \\ + \varepsilon_{AB}\Phi_{CDA'B'}\varepsilon_{C'D'} + \varepsilon_{CD}\Phi_{ABC'D'}\varepsilon_{A'B'}. \quad (6.19)$$

Разложение (6.19) может быть получено из двукратного применения операции (6.18) с учетом (4.35) в определении (6.8). Свойства симметрии

$$R_{abcd} = R_{cdab} = R_{[ab][cd]}, \quad R_a[bcd] = 0 \quad (6.20)$$

находят отражение в равенствах

$$\Psi_{ABCD} = \Psi_{(ABCD)}, \quad \Lambda = \bar{\Lambda}, \\ \Phi_{ABC'D'} = \Phi_{(AB)(C'D')} = \bar{\Phi}_{ABC'D'}, \quad (6.21)$$

так что Ψ_{ABCD} , Λ и $\Phi_{ABC'D'}$ принадлежат соответственно к неприводимым представлениям $D(2, 0)$, $D(0, 0)$ и $D(1, 1)$ локальной группы Лоренца.

На тензорном языке выразить разложение R_{abcd} на неприводимые части оказывается несколько сложнее. Можно определить *тензор Вейля* (тензор конформной кривизны) как

$$C_{abcd} = \Psi_{ABCD}\varepsilon_{A'B'}\varepsilon_{C'D'} + \varepsilon_{AB}\varepsilon_{CD}\bar{\Psi}_{A'B'C'D'}. \quad (6.22)$$

В чисто тензорных обозначениях это записывается как

$$C^{ab}_{cd} = R^{ab}_{cd} - 2R^{[a}_{[c}\delta_{d]}^{b]} + \frac{1}{3}R\delta^c_{[a}\delta_{b]}^d; \quad (6.23)$$

$$R_{ab} = R^x_{axb}; \quad R = R^x_x. \quad (6.24)$$

Тогда, наряду с обычными свойствами симметрии тензора Римана (6.20), тензор Вейля обладает еще дополнительным свойством

$$C^x_{bxd} = 0. \quad (6.25)$$

Тензор C_{abcd} еще не является неприводимым (если допустить рассмотрение комплексных тензоров) и распадается далее на так называемые самодуальную и антисамодуальную части:

$$C_{abcd} = C_{abcd}^+ + C_{abcd}^-,$$

где $C_{abcd}^+ = \Psi_{ABCD} \varepsilon_{A'B'} \varepsilon_{C'D'}$. (Термин «самодуальный» означает здесь, что $\frac{1}{2} C_{abxy} i e_{cd}^{xy} = C_{abcd}^+$, причем $-e_{abcb}$ — полностью антисимметричный символ Леви-Чивиты в правом ортонормированном базисе; на языке спиноров $e_{abcd} = i \varepsilon_{AC} \varepsilon_{BD} \varepsilon_{A'D'} \varepsilon_{B'C'} - i \varepsilon_{AD} \varepsilon_{BC} \varepsilon_{A'C'} \varepsilon_{B'D'}$. Остальные неприводимые части R_{abcd} суть

$$R_{ab} - \frac{1}{4} R g_{ab} = -2\Phi_{AB A'B'} \quad (6.26)$$

и

$$R = 24\Lambda. \quad (6.27)$$

Чтобы проверить совпадение определения (6.8) тензора Римана с его обычным определением, можно вывести *тождества Риччи*. Возьмем, например, ξ^a_b :

$$2\nabla_{[a} \nabla_{b]} \xi^c_d = 2\nabla_{[a} \nabla_{b]} (\xi^c_d \delta^b_a) = \xi^c_x R^x_{dab} - \xi^x_d R^c_{xab}. \quad (6.28)$$

Это видно из (6.13) и из свойства $R^a_{bcd} = -2\delta^a_b \nabla_{[c} \nabla_{d]} \delta^a$, следующего также из применения (6.13) к (6.8). В точности по тому же рецепту получим и соответствующие спинорные тождества Риччи, взяв, например,

$$\begin{aligned} \square_{A'B'} \theta_C^{D'} &= \square_{A'B'} (\theta_{\mathfrak{G}}^{D'} \varepsilon_{\mathfrak{D}'}^D \varepsilon_C^{\mathfrak{G}}) = \\ &= \theta_C^{X'} (\bar{\Psi}^{D'}_{X'A'B'} - 2\Lambda \varepsilon_{X'(A} \varepsilon_{B')}^{D'}) - \theta_{X'}^{D'} \Phi^X_{CA'B'}. \end{aligned} \quad (6.29)$$

Можно также непосредственно получить выражение для компонент R_{abcb} через Γ^a_{bc} , используя (6.1) в (6.8) при учете (3.20). Это дает формулы, выражающие тензор Римана обычным образом как через символы Кристоффеля, так и через коэффициенты вращения Риччи. Несколько менее привычны аналогичные выражения, связывающие спиновые коэффициенты и спинорные компоненты кривизны. Мы приведем их здесь

в явном виде, так как они нам дважды понадобятся в дальнейшем. Воспользуемся соотношением (6.18), так как в нем фигурируют сразу все компоненты кривизны. Получим [64]

$$\begin{aligned}
 \nabla_{\xi^{\alpha}} \gamma_{\alpha\beta\gamma} - \nabla_{\xi^{\beta}} \gamma_{\alpha\gamma\xi} = & \\
 = \varepsilon^{\beta\Omega} \{ & \gamma_{\beta\Omega\alpha} \gamma_{\alpha\gamma\xi} + \gamma_{\alpha\Omega\beta} \gamma_{\alpha\gamma\xi} - \\
 - \gamma_{\beta\Omega\xi} \gamma_{\alpha\gamma\xi} - & \gamma_{\alpha\Omega\xi} \gamma_{\alpha\gamma\xi} \} + \\
 + \varepsilon^{\eta\zeta} \{ & \gamma_{\alpha\zeta\eta} \gamma_{\xi\beta\zeta} - \gamma_{\alpha\zeta\eta} \gamma_{\xi\beta\zeta} \} + \\
 + \Psi_{\alpha\beta\gamma} \varepsilon_{\xi\alpha} + & \Lambda \varepsilon_{\xi\beta} (\varepsilon_{\alpha\gamma} \varepsilon_{\alpha\beta} + \varepsilon_{\alpha\gamma} \varepsilon_{\alpha\beta}) + \\
 + \Phi_{\alpha\beta\gamma} \varepsilon_{\xi\alpha}. & \tag{6.30}
 \end{aligned}$$

Уместно также записать здесь коммутатор для двух «внутренних производных»:

$$\begin{aligned}
 \{\nabla_{\xi^{\alpha}} \nabla_{\xi^{\beta}} - \nabla_{\xi^{\beta}} \nabla_{\xi^{\alpha}}\} \phi = & \\
 = \{ \varepsilon^{\beta\Omega} (\gamma_{\beta\Omega\alpha} \nabla_{\xi^{\alpha}} - & \gamma_{\beta\Omega\alpha} \nabla_{\xi^{\alpha}}) + \\
 + \varepsilon^{\eta\zeta} (\gamma_{\alpha\zeta\eta} \nabla_{\xi^{\alpha}} - & \gamma_{\alpha\zeta\eta} \nabla_{\xi^{\alpha}}) \} \phi. \tag{6.31}
 \end{aligned}$$

7.

Уравнения Эйнштейна и фокусирование

Разложение (6.19) тензора Римана на его неприводимые спинорные части позволяет нам детальнее обсудить структуру кривизны пространства-времени, которая определяется уравнениями поля Эйнштейна. Уравнения Эйнштейна суть

$$R_{ab} - \frac{1}{2} R g_{ab} + \lambda g_{ab} = -2K T_{ab}, \quad (7.1)$$

где λ и K — действительные постоянные, причем $K > 0$, а T_{ab} — локальный тензор натяжений-энергии-импульса материи («тензор энергии»). Здесь термин «материя» относится к любому полю за исключением гравитационного. В этом смысле, например, свободное электромагнитное поле считается «материей». Вопрос о том, какая энергия должна быть приписана *собственно* гравитационному полю, является более тонким. Он будет обсужден несколько позднее. Гравитационная энергия не может быть адекватно определена *локальным* образом; наоборот, она проявляется как своего рода нелокальная величина. Локальную гравитационную энергию надо, очевидно, представлять себе равной нулю, что согласуется с уравнениями (7.1). Космологический член λg_{ab} обычно отбрасывают, так как нет теоретических или наблюдательных доводов, заставляющих верить в его существование. Во всяком случае λ должно быть чрезвычайно мало, порядка единицы, деленной на квадрат радиуса вселенной¹⁾.

¹⁾ Под радиусом вселенной автор подразумевает величину, равную произведению скорости света на время, прошедшее с начала расширения вселенной. Это величина порядка 10^{28} см. — *Прим. ред.*

Уравнения (7.1) ничего не говорят о кривизне пространства, так как тензор T_{ab} не связан еще никакими ограничениями. Но мы можем рассматривать (7.1) для трех различных ситуаций. Во-первых, мы можем интересоваться теорией свободного гравитационного поля, для которого $T_{ab} = 0$. Тогда уравнения (7.1), записанные в координатной системе x^a , становятся хорошо определенной системой нелинейных дифференциальных уравнений в частных производных для функций g_{ab} . Эти уравнения трудно решить точно, и известно очень мало решений, имеющих непосредственное физическое истолкование. (Некоторые из них будут приведены в последующих разделах.) С другой стороны, можно сделать некоторые общие утверждения о поведении решений, в частности, об их асимптотическом поведении, если наложены подходящие граничные условия.

Во-вторых, нас могут интересовать решения уравнений (7.1), для которых T_{ab} хотя и не равно нулю, но подчинено некоторым условиям («уравнениям состояния»), которые обычно выражаются дифференциальными уравнениями в частных производных, но относятся теперь к полевым переменным, входящим в T_{ab} . (Например, в случае уравнений Эйнштейна — Максвелла в качестве T_{ab} выбирают тензор энергии свободного электромагнитного поля, который подчинен ковариантным уравнениям Максвелла без источников.) Опять-таки, в общем случае уравнения практически невозможно решить в точном виде для уравнений состояния, имеющих физический смысл.

В-третьих, мы можем довольствоваться некоторыми общими утверждениями о решениях уравнений (7.1), в которых мы просто ограничили T_{ab} неравенствами. Например, такое неравенство может устанавливать положительность локальной плотности энергии или удовлетворять некоторым другим физически разумным требованиям. Такой подход наиболее удобен во многих отношениях; поэтому последние два раздела будут посвящены главным образом рассмотрению пространств-времен (миров. — *Перев.*) при таких условиях.

Наконец, можно отметить, что один из подходов к уравнениям (7.1), который, строго говоря, является *безосновательным*, состоит в том, чтобы рассматривать T_{ab} как некоторое *заданное* распределение источников, для которого мы пытаемся построить гравитационное поле. Такой подход часто применяется в других областях физики, но здесь, *до тех пор*, пока не определена метрика g_{ab} , даже не ясно, что *подразумевать* под заданным распределением источников.

Если разложить R_{abcd} на его спинорные части и использовать (6.26) и (6.27), то мы получим

$$KT_{ab} = \Phi_{ABA'B'} + \left(3\Lambda - \frac{1}{2}\lambda\right)\epsilon_{AB}\epsilon_{A'B'}, \quad (7.2)$$

так что Φ_{ab} представляет собой бесследовую часть T_{ab} , а Λ — его след:

$$\Phi_{ab} = K\left(T_{ab} - \frac{1}{4}T_x^x g_{ab}\right), \quad \Lambda = \frac{1}{12}(KT_x^x + 2\lambda). \quad (7.3)$$

Та часть R_{abcd} , которая (локально) совершенно не зависит от T_{ab} , есть тензор Вейля C_{abcd} или, что эквивалентно, спинор Ψ_{ABCD} . Таким образом, мы можем говорить о Ψ_{ABCD} как о *свободной гравитационной части кривизны*. Области, где нет материи, все-таки могут иметь кривизну. Эта кривизна, согласно уравнениям Эйнштейна, будет определенного рода, и ее структура полностью определяется совершенно симметричным спинором Ψ_{ABCD} . Если $T_{ab} = 0$ и $\lambda = 0$, то получим $R_{abcd} = C_{abcd}$, где C_{abcd} определяется по (6.22). Исследуем природу Ψ_{ABCD} . Для этой цели особенно хорошо подходят спинорные методы, в то время как аналогичное исследование тензора Вейля C_{abcd} было бы значительно менее наглядным.

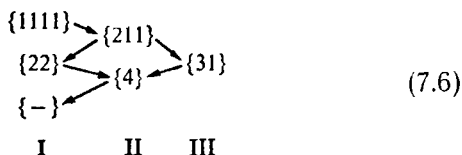
Рассмотрим форму $\Psi_{\alpha\beta\gamma\delta}\xi^\alpha\xi^\beta\xi^\gamma\xi^\delta$, где для простоты мы выбираем $\xi^0 = 1$, $\xi^1 = z$. Это полином 4-й степени по z с комплексными коэффициентами, и потому он факторизуется на четыре линейных множителя:

$$\Psi_{\alpha\beta\gamma\delta}\xi^\alpha\xi^\beta\xi^\gamma\xi^\delta = (\alpha_\mu\xi^\mu)(\beta_\nu\xi^\nu)(\gamma_\sigma\xi^\sigma)(\delta_\tau\xi^\tau). \quad (7.4)$$

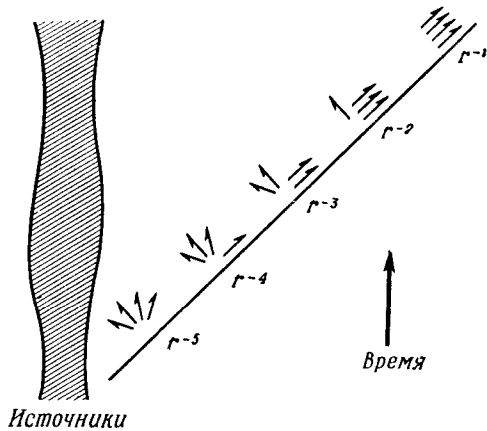
Сравнивая коэффициенты и исключая базис [ср. с (3.21)], мы получим

$$\Psi_{ABCD} = \alpha_{(A} \beta_B \gamma_C \delta_{D)}. \quad (7.5)$$

Факторизация однозначна с точностью до комплексного множителя в каждом из $\alpha_A, \dots, \delta_D$. Таким образом, световые направления, задаваемые каждым из $\alpha_A, \dots, \delta_D$, определяются однозначно, и мы видим, что Ψ_{ABCD} представляет неупорядоченное множество четырех (возможно, совпадающих) направлений вдоль светового конуса в каждой точке P на M (в которой $\Psi_{ABCD} \neq 0$). Эти направления называются *главными гравитационными световыми направлениями в P* [23, 72, 91]. Схема совпадений для главных световых направлений в P может быть представлена в следующем виде:



где стрелки указывают направление увеличивающегося вырождения. Здесь $\{111\}$ представляет общий случай, когда световые направления различны; $\{211\}$ — случай, когда есть одно сдвоенное световое направление и два единичных несовпадающих, и т. д. Символ $\{-\}$ означает случай $\Psi_{ABCD} = 0$ в P ; римские цифры I, II, III, относящиеся к столбикам в (7.6), это типы Петрова, которые представляют собой более прямую классификацию [80b, 81] величин C_{abcd} в соответствии с размерностью пространства, натянутого на собственные векторы матрицы C^{ab}_{cd} (индексы a, b и c, d попарно собираются вместе). Случай $\{4\}$, в котором все главные световые направления совпадают, называют *световым*; случай $\{22\}$ называют *вырожденным*; все случаи, за исключением $\{111\}$, называются алгебраически специальными. Используя тензор C_{abcd} , мы можем написать необходимое и достаточное условие [23, 93] того, чтобы



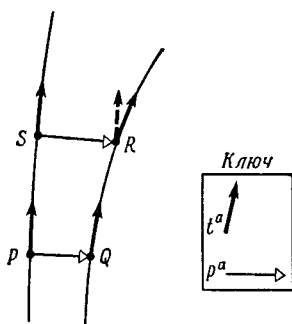
Р и с. 9. Свойство Сакса — свойство последовательного вырождения.

световой вектор l^a указывал главное световое направление:

$$l_{[a} C_{b] cd} l_{\{e} l_{f\}} l^c l^d = 0. \quad (7.7)$$

Классификационная схема (7.6) играет существенную роль для понимания геометрической структуры гравитационных полей. По-видимому, физический смысл главных световых направлений проявляется наиболее ясно в свойстве Сакса — свойстве последовательного вырождения (peeling-off) [64, 76, 93—95]. Оно гласит, что для асимптотически плоского пространства-времени, пустого на бесконечности (см. определение в разд. 8), вдоль любой световой геодезической тензор кривизны обнаруживает определенное характерное асимптотическое поведение. Пусть r — аффинный параметр на γ . Тогда кривизна уменьшается вдоль γ таким образом, что с точностью до величин порядка r^{-1} тензор кривизны является «световым», причем счетверенное главное световое направление указывает вдоль γ ; для величин порядка r^{-2} имеется строенное главное световое направление вдоль γ ; для величин порядка r^{-3} — двоекное; для величин порядка r^{-4} — единичное, а для величин по-

Рис. 10. Относительное ускорение соседних геодезических при наличии кривизны.



рядка r^{-5} кривизна не связана с γ (рис. 9). Это общее положение, но возможны также и более специальные случаи. В разд. 8 приведена схема доказательства этого результата.

Разложение (6.19) было сделано чисто алгебраически. Можно спросить, существует ли более непосредственный геометрический (или «физический») способ разделения эффектов, вызываемых тензором Вейля и другими частями кривизны, а также каков физический смысл главных световых направлений? Наиболее прямое физическое проявление кривизны пространства-времени обнаруживается в «приливных силах» — следствие эффекта геодезического отклонения на временноподобных геодезических [82, стр. 266]. Соответствующее уравнение может быть записано так:

$$D^2 p^a - R^a{}_{bcd} t^b p^c t^d = 0, \quad (7.8)$$

где t^a — касательный вектор к конгруэнции временноподобных геодезических, гладко параметризованный в соответствии с собственным временем s , скажем, $t^a t_a = 1$. Вектор p^a связывает точки с одинаковым значением s на двух соседних геодезических (рис. 10). Оператор D обозначает ковариантное дифференцирование в направлении t^a , т. е.

$$D \equiv t^a \nabla_a. \quad (7.9)$$

Уравнение (7.8) известно математикам как *уравнение Якоби*. [В действительности (7.8) выполняется при более общих условиях, чем приведенные, а именно требуется только, чтобы параметр s был аффинным на

каждой геодезической.] Если мы выберем p^a ортогональным к t^a (что можно сделать), то $D^2 p^a$ измеряет *относительное ускорение* соседних инерциальных пробных частиц. [Простой «интуитивный вывод» (7.8), когда две соседние геодезические первоначально параллельны, можно получить просто путем переноса вектора t^a вдоль контура $RQPSR$ на рис. 10. Изменение t^a дает нам изменение относительной скорости соседних частиц.]

Уравнение (7.8), примененное к временноподобным геодезическим, не позволяет нам разделить каким-нибудь простым способом эффекты, вызванные разными неприводимыми частями тензора кривизны. Но если мы используем световые геодезические, то эффекты тензора Вейля и (бесследового) тензора Риччи становятся резко разделенными [83, 102]. Так как случай световых геодезических имеет некоторые специфические свойства, подходящие для спинорной трактовки, я приведу независимый спинорный вывод необходимых уравнений, вместо того чтобы пытаться вывести их из (7.8). Это дает нам также интерпретацию некоторых спиновых коэффициентов, которая будет полезна в дальнейшем.

Выберем спиновую систему отсчета ε_{α}^A и положим

$$\varepsilon_0^A = o^A, \quad \varepsilon_1^A = \iota^A. \quad (7.10)$$

Из условия нормировки (4.27) получим

$$o_A \iota^A = 1. \quad (7.11)$$

Попытаемся дать интерпретацию следующим спиновым коэффициентам [ср. (6.2)]:

$$\kappa = \gamma_{0000}', \quad \rho = \gamma_{0010}', \quad \sigma = \gamma_{0001}'. \quad (7.12)$$

Имеются все формы $o^A \nabla_{\mathbb{C}\mathbb{D}'} o_A$, где \mathbb{C} , \mathbb{D}' равны соответственно 0, 0'; 1, 0'; 0, 1'. При

$$o^A \rightarrow \xi o^A, \quad \iota^A \rightarrow \xi^{-1} \iota^A \quad (7.13)$$

мы имеем

$$\kappa \rightarrow \xi^3 \bar{\xi} \kappa, \quad \rho \rightarrow \xi \bar{\xi} \rho, \quad \sigma \rightarrow \xi^3 \bar{\xi}^{-1} \sigma, \quad (7.14)$$

а члены с производными пропадают в силу $o^A o_A = 0$. Теперь выясним смысл условия $\kappa = 0$:

$$o^A o^B \bar{o}^C \nabla_{BC} o_A = 0. \quad (7.15)$$

Это условие имеет вид $o^A \theta_A = 0$, откуда можно получить, что $\theta_A = \chi o_A$ для некоторого χ . Таким образом, $\kappa = 0$ означает

$$o^B \bar{o}^C \nabla_{BC} o_A = \chi o_A. \quad (7.16)$$

Полагая

$$l^a = o^A \bar{o}^{A'}, \quad m^a = o^A \dot{i}^{A'}, \quad (7.17)$$

перепишем (7.16) как $l^a \nabla_a o_A = \chi o_A$. Это равенство означает, что o_A переносится параллельно самому себе в l^a -направлении (т. е. в o^A -направлении). Другими словами, условие $\kappa = 0$ является необходимым и достаточным для того, чтобы o^A -направления были касательными к *световым геодезическим*. Следовательно, κ есть своего рода мера *кривизны* o^A -кривых. (Она является комплексной, так что указывает также *направление* этой кривизны относительно флаговой площадки o^A .)

Теперь предположим, что $\kappa = 0$, так что мы имеем конгруэнцию световых геодезических o^A -кривых, и прочтем (7.15) двумя различными способами, а именно как $o^B \phi_B = 0$ и как $\bar{o}^{C'} \psi_{C'} = 0$. По аналогии с тем, что было выше, имеем соотношения следующей формы:

$$o^A \bar{o}^{C'} \nabla_{BC} o_A = \rho o_B, \quad (7.18)$$

$$o^A o^B \nabla_{BC} o_A = \sigma \bar{o}_{C'} \quad (7.19)$$

для некоторых ρ, σ . В действительности ρ и σ суть величины, введенные в (7.12), что следует из умножения (7.18) на i^B и (7.19) на $i^{C'}$. Значение (7.18) и (7.19) состоит в том, что они показывают, что при $\kappa = 0$ величины ρ и σ не зависят от i^A . А поскольку они просто изменяют масштаб согласно (7.14), если для o^A справедливо (7.13), то следует ожидать простого геометрического смысла для ρ и σ как некоторого свойства конгруэнции световых геодезических. Чтобы убедиться в этом, рассмотрим элемент π двумерной плоскости в некоторой точке P на одной из

геодезических γ . Выберем π пространственноподобным и ортогональным к направлению γ . Мы можем также выбрать l^A ортогональным к π , так что π будет натянуто на m^a и \bar{m}^a , где m^a определяется по (7.17). Тогда любой (действительный) вектор, лежащий в π , имеет форму

$$\bar{z}m^a + zm^a, \quad (7.20)$$

и, следовательно, мы можем рассматривать π как аргандову плоскость $\sqrt{2}z$. В окрестности γ двумерное множество световых геодезических «протыкает» π . Эти геодезические, с точностью до величин первого порядка, лежат на световой гиперповерхности, проходящей через γ . Есть, разумеется, и другие световые геодезические в окрестности γ , но мы их не рассматриваем. Посмотрим, как изменяется z , если мы будем следить за одной из световых геодезических, протыкающих π . (Отметим, что z в точке P для определенной световой геодезической из конгруэнции вблизи γ не зависит от выбора π в P . Изменение π в точке P сводится просто к замене m^a на $m^a + kl^a$ в (7.20) с комплексным k .)

Положим

$$D \equiv l^a \nabla_a = \nabla_{00'}. \quad (7.21)$$

Для удобства будем считать, что o^A , l^A параллельно переносятся вдоль γ :

$$D o_A = 0; \quad D l_A = 0. \quad (7.22)$$

Тогда m^a , \bar{m}^a и, следовательно, π также переносятся параллельно вдоль γ . Пусть ϕ — некоторая функция, постоянная вдоль световых геодезических

$$D\phi = 0. \quad (7.23)$$

Требование неизменности значения z , выделяющего определенную геодезическую γ нашей системы, мы характеризуем условием [см. (7.20)]

$$D \{(\bar{z}m^a + zm^a) \nabla_a \phi\} = 0. \quad (7.24)$$

Это дает

$$D\bar{z}\nabla_{01'}\phi + Dz\nabla_{10'}\phi + \bar{z}(\nabla_{00'}\nabla_{01'} - \nabla_{01'}\nabla_{00'})\phi + z(\nabla_{00'}\nabla_{10'} - \nabla_{10'}\nabla_{00'})\phi = 0, \quad (7.25)$$

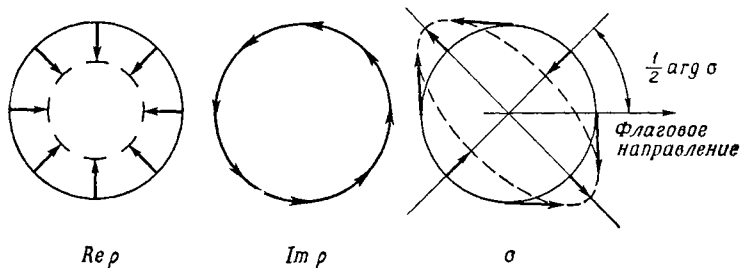


Рис. 11. Интерпретация ρ и σ в терминах их влияния на первоначально круглый пучок световых лучей.

т. е., используя (6.31) (или непосредственно вычисляя эти коммутаторы), получим:

$$(D\bar{z} + \bar{\rho}\bar{z} + \bar{\sigma}\bar{z})\nabla_{01'}\phi + (Dz + \rho z + \sigma\bar{z})\nabla_{01'}\phi = 0. \quad (7.26)$$

Таким образом

$$Dz = -\rho z - \sigma\bar{z}. \quad (7.27)$$

Теперь интерпретация ρ и σ становится очевидной. Действительная часть ρ измеряет сходимости световых геодезических, а мнимая часть ρ измеряет их вращение вокруг γ . Тот факт, что σ не обращается в нуль, указывает на присутствие дисторсии, или сдвига (рис. 11). Таким образом, если $\sigma \neq 0$, то малый круг превратится в эллипс по мере продвижения вдоль световых геодезических. Ориентация осей этого эллипса относительно флаговой плоскости o^A определяется как $\frac{1}{2} \arg \sigma$. Если световые геодезические являются образующими световых гиперповерхностей, то мы можем выбрать $l_a = \nabla_a u$, откуда следует, что

$$o^A \bar{o}^{B'} (\nabla_a l_b - \nabla_b l_a) = 0,$$

т. е. вращение отсутствует:

$$\rho = \bar{\rho}. \quad (7.28)$$

Условие (7.28) является также достаточным для того, чтобы световые геодезические образовывали световые гиперповерхности. Если (7.28) выполняется,

то ρ и σ относятся к геометрии каждой световой гиперповерхности в отдельности, а не к связи между различными световыми гиперповерхностями.

Рассмотрим далее D -производные от ρ и σ . Мы можем получить их из (6.30) [используя (7.22): $\gamma_{\alpha\beta\gamma\delta} = 0$] или непосредственно из определения ρ , σ , $\Phi_{ABC'D'}$ и Ψ_{ABCD} . Тогда

$$D\rho = \rho^2 + \sigma\bar{\sigma} + \Phi, \quad (7.29)$$

$$D\sigma = \sigma(\rho + \bar{\rho}) + \Psi, \quad (7.30)$$

где

$$\Psi = \Psi_{0000}, \quad \Phi = \Phi_{000'0'}. \quad (7.31)$$

Уравнения (7.29) и (7.30) есть *уравнения Сакса* [93] для «оптических скаляров» ρ , σ . Комбинируя их с (7.27), получим

$$D^2z = -\Phi z - \Psi \bar{z}. \quad (7.32)$$

Фактически (7.32) можно получить также непосредственно из (7.8), если мы положим $t^a = l^a$ и $\rho^a = \bar{z}m^a + z\bar{m}^a$. В этой связи приведем формулы:

$$\begin{aligned} \Psi &= R_{abcd}l^a m^b l^c m^d, & \Phi &= R_{abcd}l^a m^b l^c \bar{m}^d, \\ \rho &= m^a \bar{m}^b \nabla_b l_a, & \sigma &= m^a m^b \nabla_b l_a. \end{aligned} \quad (7.33)$$

Из (7.29) и (7.30) видно, что изменения в ρ и σ определяются главным образом величинами Φ и Ψ соответственно, но в обоих случаях есть дополнительные нелинейные члены. Чтобы выделить эффекты, связанные с Φ и Ψ , в наиболее чистом виде, рассмотрим частный случай, в котором Φ и Ψ локализованы в очень малой области на γ . [Фактически возможен также идеализированный случай, когда Φ и Ψ являются дельта-функциями Дирака. См. пример с плоскими волнами (9.23)–(9.26).]

Предположим, что имеется импульс света, представляющий собой первоначально пучок параллельных лучей. Световые геодезические, образующие пучок, составляют небольшую часть (окрестность γ) световой гиперповерхности, проходящей через γ , и для них $\rho = \sigma = 0$. Тогда из (7.29) мы видим, что часть кривизны Φ (при $\Psi = 0$) действует как *линза*

без астигматизма (поскольку в пучок вносится сходимос-
ть ρ , но сдвиг σ остается нулевым). В то же время
часть кривизны Ψ (при $\Phi = 0$) действует как *чисто*
астигматическая линза, т. е. в одну плоскость вносятся
точно такая же положительная сходимос-
ть, какая отрицательная сходимос-
ть вносится в плоскость, пер-
пендикулярную к первой (поскольку для результи-
рующего пучка $\rho = 0$, а $\sigma \neq 0$). Интерпретация ча-
стей кривизны на языке линз остается в силе и для
произвольного пучка, в котором ρ и σ не обязательно
равны нулю вначале. Нелинейные члены в (7.29) и
(7.30) также в точности соответствуют эффектам, ко-
торые производятся разделенными линзами [78]. Та-
ким образом мы можем представлять себе действие
кривизны пространства-времени на пучок света ана-
логичным эффектам, производимым серией линз, рас-
положенных вдоль пучка. Степень сходимости описы-
вается локально величиной Φ , а астигматизм — вели-
чиной Ψ . Вследствие симметрии (6.21) компоненты
(7.31) дают *всю* информацию, содержащуюся в Ψ_{ABCD}
и $\Phi_{ABC'D'}$ при изменении γ .

Одно важное свойство кривизны состоит в том, что
всегда имеется *положительное* фокусирование вдоль
любой световой геодезической, если мы предполагаем,
что локальная плотность энергии положительна. Плотность энергии, измеренная наблюдателем, миро-
вая линия которого имеет касательный вектор t^a ,
равна

$$T_{ab}t^at^b, \quad (7.34)$$

если еще предположить, что $t^at_a = 1$. Если мы тре-
буем, чтобы (7.34) было неотрицательно, то $T_{ab}t^at^b \geq 0$
 ≥ 0 для любого временноподобного вектора t^a (без-
относительно к условию нормировки на t^a). Таким об-
разом, если мы возьмем предел вектора t^a при его
стремлении к световому вектору l^a , мы получим [со-
гласно (7.3)]

$$\Phi = KT_{ab}l^al^b \geq 0 \quad (7.35)$$

(неважно, равно λ нулю или нет, поскольку $g_{ab}l^al^b = 0$). Отсюда следует, что в отсутствие Ψ кривизна

Φ действует вдоль пучка как серия положительных (собирающих) линз. Присутствие Ψ проявляется в уравнении (7.29) посредством члена $\sigma\bar{\sigma}$, который опять-таки действует как положительная линза. Грубо говоря, мы можем считать σ интегралом от Ψ вдоль γ . Таким образом, Ψ действует *нелокально*, способствуя фокусировке вдоль γ . Своим суммарным эффектом осциллирующее Ψ сильно напоминает Φ . Это тесно связано с вопросом о гравитационной энергии. Хотя локально гравитационное поле не дает вклада в плотность энергии, но гравитационная волна переносит нелокальную энергию, причем полная энергия, переносимая волной, всегда *положительна*. (Поразительным примером того, как гравитационная «рябь» создает эффективную положительную энергию, является гравитационный «геон» — теоретическая конструкция Уилера [112]. Он представляет собой область пустого пространства, состоящую из «ряби», которая остается ограниченной длительное время. На достаточно большом расстоянии геон проявляется как «частица» с положительной массой.)

Рассмотрим положительный фокусирующий эффект несколько подробнее. Если в элементарной площадке π мы выберем малый треугольник с вершинами в точках $0, z_1, z_2$, то его площадь

$$a = \frac{i}{4} (z_1 \bar{z}_2 - z_2 \bar{z}_1) \quad (7.36)$$

удовлетворяет уравнению [согласно (7.27)].

$$Da = -(\rho + \bar{\rho})a. \quad (7.37)$$

Так как наш пучок лучей выбран как часть световой гиперповерхности, то $\rho = \bar{\rho}$. Следовательно,

$$D(a^{1/2}) = -\rho a^{1/2}. \quad (7.38)$$

Из (7.29) получаем

$$D^2(a^{1/2}) = D(-\rho a^{1/2}) = -(\sigma\bar{\sigma} + \Phi)a^{1/2} \leq 0. \quad (7.39)$$

Таким образом, если $\rho > 0$ в некоторой точке γ , то по (7.38) $a^{1/2}$ уменьшается. Из (7.39) следует, что $a^{1/2}$

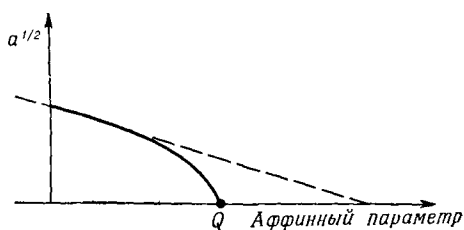


Рис. 12. Как только площадь поперечного сечения пучка световых лучей начинает убывать, она неизбежно падает до нуля.

уменьшается до нуля¹⁾ (рис. 12). Таким образом, пучок неизбежно достигает фокальной точки Q . (В общем случае фокальная точка будет астигматической, т. е. пучок собирается в линию, а не в точку.) Отметим, что ρ становится неограниченным вблизи Q , так как $\rho = -\frac{1}{2} D(\ln a)$ согласно (7.38).

Такое поведение является предельным случаем эффекта Райчаудури [49, 86], который относится к конгруэнции временноподобных геодезических, ортогональных к гиперповерхности. Пусть t^a — касательный вектор к геодезическим данной конгруэнции, параметризованный в соответствии с собственным временем s (так что $t^a t_a = 1$, а t^a направлено в будущее) и ортогональный к пространственноподобной гиперповерхности $s = 0$. Определим дивергенцию геодезических как

$$\theta = \nabla_a t^a. \quad (7.40)$$

¹⁾ Это свойство имеет одно простое применение, которое, видимо, не отмечалось прежде, а именно: может ли сферически-симметричное тело в асимптотически плоском пространстве-времени не иметь центра (см. [9, стр. 436]). Если мы рассмотрим световые лучи (световые геодезические), сходящиеся внутрь сферы, расположенной симметрично вокруг тела на некотором расстоянии от него, то мы видим, что лучи первоначально сходятся, так что $\rho > 0$. Они должны достичь (симметричной) фокальной точки [при условии, что выполняется (7.35)], и эта фокальная точка определяет центр (сравните со свойствами ловушечной поверхности, описанной в разд. 10).

Из тождества Риччи мы получим уравнение Рай-чаудури

$$D\theta = -\nabla_a t^b \nabla_b t^a + R_{ab} t^a t^b, \quad (7.41)$$

где $D = t^a \nabla_a$, как в (7.9).

Из условия ортогональности к гиперповерхности получим $\nabla_a t_b = \nabla_b t_a$, а из неравенства Шварца в 3-мерном пространстве, ортогональном к t^a , имеем

$$\nabla_a t^b \nabla_b t^a = \nabla_a t_b \nabla^a t^b \geq \frac{1}{3} \theta^2 \quad (7.42)$$

[из условия геодезичности $t^a \nabla_a t_b = 0$ и $t^a \nabla_b t_a = \frac{1}{2} \nabla_b (t^a t_a) = 0$]. Таким образом,

$$D\theta \leq -\frac{1}{3} \theta^2 + R_{ab} t^a t^b. \quad (7.43)$$

Пусть V — элемент 3-мерного объема, ортогональный к t^a . Тогда по (7.40)

$$DV = \theta V \quad (7.44)$$

и

$$D(V^{1/3}) = \frac{1}{3} \theta V^{1/3}. \quad (7.45)$$

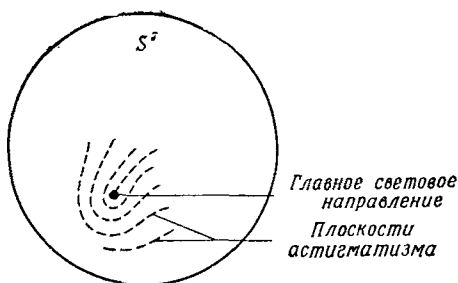
Теперь (7.43) дает

$$D^2(V^{1/3}) = \frac{1}{3} D(\theta V^{1/3}) \leq \frac{1}{3} V^{1/3} R_{ab} t^a t^b \leq 0, \quad (7.46)$$

если выполняется *энергетическое условие* [ср. с (7.1)]

$$T_{ab} t^a t^b \geq \frac{\lambda}{2K} + T^c_c. \quad (7.47)$$

[При $\lambda = 0$ неравенство (7.47) выполняется, если в «собственной тетраде» тензора T_{ab} плотность энергии не меньше каждого из главных давлений, взятых с *отрицательным знаком*, и не меньше суммы главных давлений с отрицательным знаком. Это справедливо для всех видов обычной материи.] С этого места доказательство проводится так же, как и прежде, с заменой (7.38) на (7.45) и (7.39) на (7.46). Видно, что



Р и с. 13. Главное световое направление проявляется как сингулярность в поле направлений астигматизма на S^2 , представляющей все световые направления в точке P .

если в некоторой точке на одной из геодезических $\theta < 0$, то где-то в будущем на этой геодезической достигается фокальная точка ($V = 0$, $\theta \rightarrow -\infty$). Аналогично, фокальная точка ($V = 0$, $\theta \rightarrow +\infty$) достигается в прошлом на геодезической, если $\theta > 0$.

Здесь уместно сделать еще одно заключительное замечание относительно (7.30). Согласно (7.31),

$$\Psi = \Psi_{ABCD} o^A o^B o^C o^D, \quad (7.48)$$

так что, сравнивая (7.4) и (7.5), мы видим, что Ψ обращается в нуль, если o^A ориентировано вдоль *главного светового направления*. Таким образом, мы можем дать физическую интерпретацию главных световых направлений. Это те световые направления (в точке P), вдоль которых *нет астигматического фокусирования*. Поскольку астигматизм зависит от направления, путем топологического рассмотрения можно показать, что в точке P должно быть по крайней мере *четыре* световых направления, где астигматизм обращается в нуль. Рассмотрим сферу S^2 , представляющую собой световые направления в P , на которой мы введем линейный элемент, представляющий астигматизм для каждого светового направления в P . Главные световые направления соответствуют четырем точкам, в которых линейный элемент обращается

в нуль (рис. 13). Аргументация этого типа применима к полям других спинов (например, к полю спина 1, а именно к электромагнетизму, когда аналогичный линейный элемент является *ориентированным*, так что случай, представленный на рис. 13, не осуществляется, и мы получаем только два главных световых направления). Этот подход может быть также применен к *асимптотическим* полям, для которых S^2 будет на бесконечности, а не в точке P .

8.

Конформная бесконечность

Вопрос о смысле гравитационной энергии был кратко рассмотрен в предыдущем разделе. Полезность понятия энергии обусловлена, вообще говоря, фактом ее *сохранения*. Теория Эйнштейна содержит «локальный закон сохранения» для энергии, а именно

$$\nabla_a T^{ab} = 0, \quad (8.1)$$

который в силу уравнений поля (7.1) является двойной сверткой тождеств Бианки:

$$\nabla_{[a} R_{bc]de} = 0. \quad (8.2)$$

Однако (8.1) *не приводит* к *интегральному* закону сохранения обычного типа, т. е. к закону, утверждающему, что интеграл по границе компактного 4-мерного объема от «потока» некоторой величины через эту границу равен нулю.

Прототипом такого интегрального закона может служить закон сохранения электрического заряда. Если J^a обозначает вектор тока, то

$$\nabla_a J^a = 0. \quad (8.3)$$

Так как J^a является «вектором» или, точнее, элементом, «дуальным к 3-форме», то закон сохранения (8.3) подразумевает существование также интегрального закона указанного выше типа, где поток заряда через элемент поверхности dS_a равен $J^a dS_a$. Однако (8.1) имеет другой вид. Мы можем получить *приближенный* интегральный закон сохранения, если проинтегрируем (8.1) по границе некоторой области, размеры которой малы по сравнению с радиусами кривизны, входящими в R_{abcd} . Для этой цели можно ввести базисный репер δ_a^a , который приближенно будет

ковариантно постоянным во всей области, и записать (8.1) как $\nabla_a T^{ab} \approx 0$. Это равенство имеет вид (8.3), так что возникает приближенный интегральный закон сохранения энергии-импульса с потоком энергии-импульса $T^{ab} dS_a$.

Мы можем рассматривать кривизну пространства-времени как заданный (нелокальный) *гравитационный* вклад в энергию-импульс, который должен учитываться для получения точного интегрального закона сохранения. Этот подход может быть формализован различными способами (используя «псевдотензоры» и т. д.), однако еще нет метода, который позволил бы приписать в общем случае какую-нибудь осмысленную однозначность, скажем, потоку гравитационной энергии-импульсу или полной энергии-импульсу в области. Я думаю, что мы можем считать вопрос о гравитационной энергии одной из важных нерешенных проблем в общей теории относительности¹⁾.

Но имеем ли мы право надеяться на решение этой проблемы вообще? Не может ли оказаться, что энергия-импульс просто не являются вполне сохраняющимися величинами в общей теории относительности или, другими словами, понятие энергии-импульса не подходит для этой цели (кроме, как в локальном смысле)? Возможно, что это так и есть. Я не знаю. Однако, по-моему, очень существенным является тот факт, что разумное (т. е. «ковариантно» или «геометрически» определенное) понятие энергии-импульса

¹⁾ К настоящему времени этот вопрос можно считать в значительной степени выясненным. Его решение шло по пути, описанному далее автором. С одной стороны, Айзексон [120*] построил тензор энергии-импульса для пакетов гравитационных волн в общем случае искривленного пространства-времени. С другой стороны, Торн [121*], Берк [122*] и Чандрасекар и Эспозито [123*] прямыми расчетами показали, что торможение гравитационным излучением движений в излучающей системе в точности соответствует энергии, уносимой волнами и рассчитанной стандартными методами. Вычислено гравитационное излучение нейтронных звезд. Указано также, что по аналогии с электромагнитным случаем гравитационная волна характеризуется нелокальным инвариантом, имеющим смысл числа гравитонов в волновом пакете [124*]. — *Прим. ред.*

действительно существует для интересного подкласса миров, а именно для тех из них, которые являются *асимптотически плоскими*.

Изучение асимптотически плоских миров составляет важную часть общей теории относительности. Это происходит не потому, что вселенную считают почти асимптотически плоской, а потому, что в любой ситуации, в которой следует ожидать существенных общерелятивистских эффектов (за исключением случая изучения вселенной в целом), значения кривизны, входящие в локальный процесс, будут превосходить на много порядков значения сглаженной кривизны общего фона. (Например, если бы даже сколлапсировала целая галактика, размер результирующего объекта был бы меньше радиуса фоновой кривизны вселенной примерно в 10^{11} раз. См. обсуждение в гл. 10.) Таким образом, асимптотическая евклидовость (flatness) является превосходным приближением во многих случаях. Это — счастливое обстоятельство, так как только в асимптотически плоских случаях общая теория относительности начинает напоминать остальную физику. Теперь мы можем обсуждать вопрос об опережающих и запаздывающих волнах, гравитационных или негравитационных. Мы можем говорить об энергии, переносимой этими волнами, о проблемах рассеяния и, возможно, даже о квантовании. Я не буду здесь углубляться в детали. Я просто изложу математически строгую схему для обсуждения этих вопросов и укажу на некоторые ее применения.

Чтобы сформулировать основные принципы, рассмотрим с физической точки зрения вопрос о том, переносят ли энергию гравитационные волны, излученные изолированной системой, и если переносят, то является ли эта энергия с необходимостью положительной. (Существование гравитационных волн, излученных системой с переменной асимметрией, следует прежде всего из линейного приближения эйнштейновой теории.) Чтобы выяснить, какую энергию мы должны приписать волнам, измерим массу m_1 системы до излучения волны и массу m_2 — после излучения. Если сохранение энергии (т. е. массы) продолжает иметь

какой-либо смысл, потребуем, чтобы полная энергия, перенесенная волнами, была в точности равна $m_1 - m_2$ (предполагая, что отсутствуют приходящие волны, которые могли бы быть поглощены системой).

Но как мы должны измерять массу? Один способ состоит в интегрировании некоторого выражения для плотности массы по пространственноподобной гиперповерхности \mathcal{P} . Однако мы должны были бы принять во внимание «нелокальную плотность массы» гравитационного поля самого по себе, что усложняет дело. В качестве альтернативы мы могли бы измерять массу, исследуя поле *только* в окрестности *бесконечности* на \mathcal{P} , так как характер падения кривизны на больших расстояниях на \mathcal{P} должен определять полную массу. Вместо интеграла по 3-объему мы используем тогда просто интеграл по 2-поверхности, взятой на бесконечности. Любой из этих методов *может* быть использован для получения удовлетворительного определения. Но в обоих случаях пространственноподобная гиперповерхность \mathcal{P} должна быть продолжена до бесконечности, и мы сталкиваемся с трудностью при измерении массы $m_1 - m_2$, переносимой волнами. Мы могли бы наложить «естественное» требование, чтобы \mathcal{P} надлежащим образом стремилась к пространственноподобной гиперплоскости по мере того, как пространство-время приближается к плоскому на бесконечности. Это требование законно для гиперповерхности \mathcal{P}_1 , которая пересекает мировую линию источника до излучения и дает массу m_1 (рис. 14). Но в случае гиперповерхности \mathcal{P}_2 , которая пересекает источник *после* излучения, в результате измерения массы мы получим не m_2 , а опять-таки m_1 . Это происходит потому, что \mathcal{P}_2 пересекает все волны, и какую бы энергию они ни переносили, мы получаем $m_1 = m_2 + (m_1 - m_2)$.

Можно попытаться осуществить некоторый сложный процесс, позволяющий гиперповерхности двигаться «вверх» по времени с постепенным расширением области интегрирования до бесконечности, так, чтобы волны оставались вне этой области. Применим его ко второй из упомянутых возможностей для измерения массы. Тогда мы придем к рассмотрению интегралов

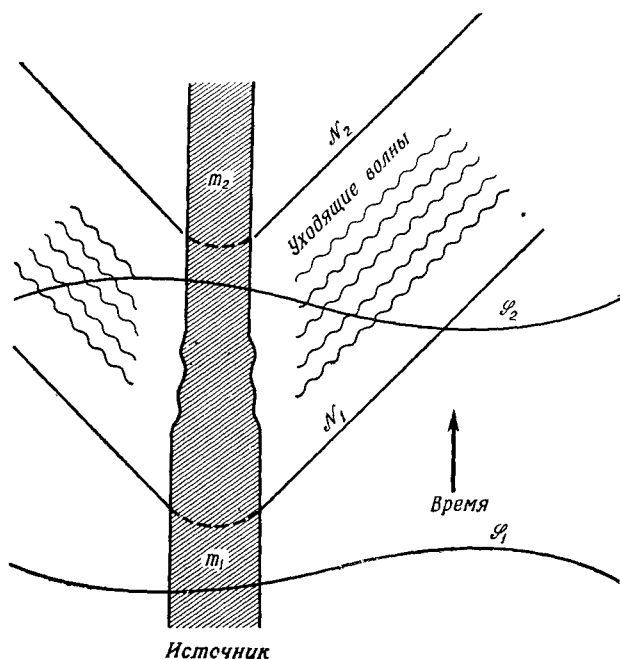


Рис. 14. Для измерения потери массы вследствие излучения \mathcal{N}_1 и \mathcal{N}_2 более удобны, чем \mathcal{S}_1 и \mathcal{S}_2 .

по 2-поверхности на бесконечности, но не в пространственноподобных, а в *световых* направлениях. Это предполагает использование световых (или, по крайней мере, асимптотически световых) гиперповерхностей \mathcal{N}_1 , \mathcal{N}_2 вместо \mathcal{S}_1 , \mathcal{S}_2 . При этом \mathcal{N}_1 полностью лежит вне конуса излучения, который, в свою очередь, полностью лежит вне \mathcal{N}_2 . Тогда интеграл по 2-поверхности на бесконечности \mathcal{N}_i можно использовать для измерения m_i (где $i = 1, 2$), а разность $m_1 - m_2$ дает массу волн.

Введя такое определение массы и показав, что разность $m_1 - m_2$ положительна в присутствии излучения, Бонди (с сотрудниками) [8, 11] и затем Сакс [95] существенно продвинулись в понимании гравитационной энергии. Их определение зависит от выбора

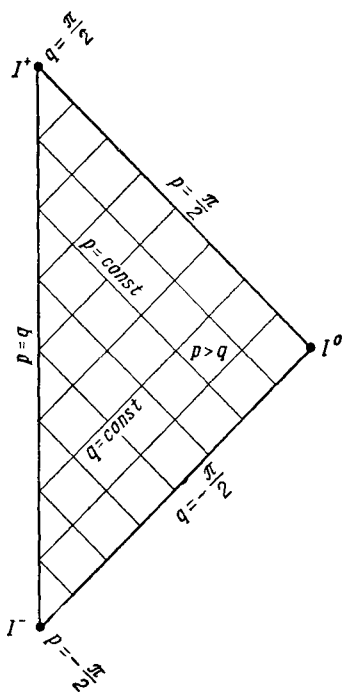


Рис. 15. Область изменения p и q . Чтобы построить \mathcal{M} , надо вращать рисунок вокруг $p = q$, так что каждая точка $p > q$ описывает S^2 (но I^0 должно оставаться фиксированным!).

специальной координатной системы, основанной на уходящих световых гиперповерхностях. Существование координатной системы требуемого типа взято в качестве определения необходимой степени асимптотической евклидовости. Я предлагаю дать другое определение асимптотической евклидовости, которое является более геометрическим, но по существу эквивалент-

но определению Бонди — Сакса. Идея состоит во введении *границы* пространства-времени \mathcal{M} , которая образована начальными и конечными точками каждой световой геодезической в \mathcal{M} . (Мы только что видели, что масса по Бонди — Саксу представляет собой интеграл по 2-поверхности «конечных точек» световых геодезических, а именно тех, которые образуют \mathcal{N}_i .) Чудесным образом оказывается, что асимптотическая евклидовость Бонди — Сакса находит свое выражение в существовании *гладкой конформной структуры* [74, 76] пространства-времени с границей. Имея на бесконечности такую хорошо определенную структуру, мы можем теперь выполнять вычисления, рассматривая бесконечность так, как если бы она была *локальной* структурой, что позволяет избежать неудобных асимптотических пределов.

Начнем с рассмотрения природы конформной бесконечности для пространства-времени Минковского \mathcal{M} . Выберем световые полярные координаты u, v, θ, φ , связанные с обычными координатами x^0, x^1, x^2, x^3 посредством

$$u = x^0 - r, \quad v = x^0 + r, \quad (8.4)$$

где $r = [(x^1)^2 + (x^2)^2 + (x^3)^2]^{1/2}$ и

$$x^1 = r \sin \theta \cos \varphi, \quad x^2 = r \sin \theta \sin \varphi, \quad x^3 = r \cos \theta. \quad (8.5)$$

Тогда

$$ds^2 = du dv - \frac{1}{4} (u - v)^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad u \leq v. \quad (8.6)$$

Выберем новые координаты p, q , такие, что

$$v = \operatorname{tg} p, \quad u = \operatorname{tg} q \quad (-\pi/2 \leq q \leq p \leq \pi/2), \quad (8.7)$$

тогда точки на бесконечности имеют конечные значения координат p, q (рис. 15). Метрика (8.6) теряет смысл при этих значениях ($p = \pi/2$ или $q = -\pi/2$), но если мы перейдем к конформной метрике

$$d\hat{s} = \Omega ds, \quad (8.8)$$

где

$$\Omega = (1 + u^2)^{-1/2} (1 + v^2)^{-1/2}, \quad (8.9)$$

то (8.6) превращается в

$$d\hat{s}^2 = dp dq - \frac{1}{4} \sin^2 (p - q) (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (8.10)$$

Метрика (8.10) вполне регулярна при $p = \pi/2$ и $q = -\pi/2$ (за исключением $p = q$ или $p - q = \pi$, где имеются устранимые координатные сингулярности). Таким образом, мы имеем хорошо определенную конформную структуру на многообразии $(-\pi/2 \leq q \leq p \leq \pi/2)$ с границей¹⁾ ($q = -\pi/2$ или

¹⁾ Строго говоря, здесь нет многообразия с границей, поскольку граница обладает «углами». Они будут устранены в строгом определении, которое дано несколько ниже.

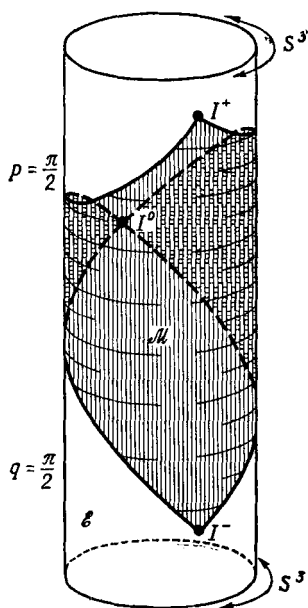
Есть также другой вариант компактного конформного мира Минковского, в котором отождествлены граничные гиперповерхности в будущем и в прошлом. Тогда имеет место компактное конформное многообразие с границей ($\approx S^1 \times S^3$) (см. [51, 92], а также [76, 79]).

$p = \pi/2$), внутренность которого ($-\pi/2 < q \leq p < \pi/2$) идентична по своей конформной структуре пространству-времени Минковского. Отметим, что $q = \text{const}$ и $p = \text{const}$ являются световыми гиперповерхностями (конусами) и образованы световыми геодезическими. Это же относится, в частности, и к граничным гиперповерхностям.

Метрика (8.10) является, по существу, метрикой статической вселенной Эйнштейна \mathcal{E} . Если полагать $p + q$ неограниченным, а $0 \leq p - q \leq \pi$ (с обычной областью изменения для сферических полярных координат θ, φ), то можно считать (8.10) уравнением «цилиндра» $\mathcal{E} \approx S^3 \times E^1$, если просто переписать (8.10) в форме $d\hat{s}^2 = dT^2 - d\Sigma^2$, где $d\Sigma^2$ — обычная метрика на 3-сфере. [Координатную сингулярность в (8.10) можно теперь покрыть, применяя к координатам p, q, θ, φ вращение в S^3 .] Часть \mathcal{E} , соответствующую \mathcal{M} , можно изобразить геометрически следующим образом (рис. 16). Выберем на \mathcal{E} точку I^- (это будет точка с координатами $p = q = -\pi/2$). Световые геодезические на \mathcal{E} , которые начинаются в I^- , образуют световой конус будущего для точки I^- . Эти геодезические собираются в первой фокальной точке, которая диаметрально противоположна (относительно S^3) точке I^- . Назовем ее I^0 (это будет точка $p = -q = \pi/2$).

Открытые отрезки световых геодезических от I^- до I^0 образуют гиперповерхность \mathcal{I}^- ($-\pi/2 < p < \pi/2, q = -\pi/2$), которая ограничена в прошлом точкой I^- , а в будущем — точкой I^0 . (\mathcal{I}^- имеет топологию $S^2 \times E^1$.) Гиперповерхность \mathcal{I}^- определяет световой конус прошлого для I^0 . Продолжим эти геодезические в будущее за точкой I^0 . Они достигнут второй фокальной точки I^+ ($p = q = \pi/2$), диаметрально противоположной I^0 (по отношению к S^3) и лежащей поэтому непосредственно «над» I^- в \mathcal{E} . Открытые сегменты световых геодезических от I^0 до I^+ образуют световую гиперповерхность \mathcal{I}^+ ($p = \pi/2, -\pi/2 < q < \pi/2$) опять-таки с топологией $S^2 \times E^1$, ограниченную в прошлом и будущем точками I^0 и I^+ соответственно. Часть \mathcal{E} , которая может быть отождествлена с \mathcal{M} (по конформной структуре), есть как раз множество точек \mathcal{E} , ле-

Рис. 16. Конформный мир Минковского как часть вселенной Эйнштейна.



жащих «между» \mathcal{I}^- и \mathcal{I}^+ (т.е. множество точек, которые лежат между I^- и I^+ на временноподобных кривых от I^- до I^+). Таким образом, \mathcal{M} является открытым подмножеством \mathcal{E} , и его граница есть

$$I^- \cup \mathcal{I}^- \cup I^0 \cup \mathcal{I}^+ \cup I^+.$$

Удобно изображать \mathcal{M} как внутренность двух соединенных по основаниям конусов (рис. 17). Тогда

разные части \mathcal{M} будут представлены частями этих ограниченных обрезанных конусов. Надо иметь в виду, однако, что эта картина не точна с точки зрения *конформности*. Наибольшее несоответствие имеет место вблизи I^0 , поскольку I^0 выглядит как экваториальная область, в то время как это должна быть точка. (Разумеется, мы должны также представлять себе картину четырехмерной.)

Точки I^- , I^0 , I^+ интерпретируются как «временная бесконечность прошлого», «пространственная бесконечность» и «временная бесконечность будущего». Гиперповерхности \mathcal{I}^- и \mathcal{I}^+ суть, соответственно, «световая бесконечность прошлого» и «световая бесконечность будущего». Происхождение таких названий станет очевидным, если мы исследуем прямые (согласно метрике Минковского ds) линии в \mathcal{M} . Временноподобная прямая линия становится кривой, соединяющей I^- и I^+ ; пространственноподобная линия становится замкнутой кривой, проходящей через I^0 ; световая линия

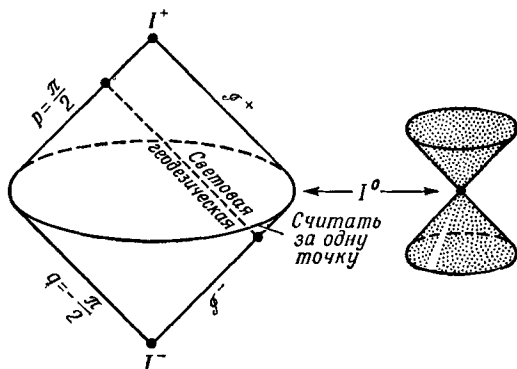


Рис. 17. Каждая световая геодезическая в \mathcal{M} достигает двух конечных точек в $\bar{\mathcal{M}}$; одна из них лежит на \mathcal{I}^- , другая — на \mathcal{I}^+ .

является световой геодезической, начинающейся на \mathcal{I}^- и заканчивающейся на \mathcal{I}^+ . Во всех случаях кривая становится «компактной» после добавления соответствующих точек на бесконечности.

Продолжение \mathcal{M} как конформного многообразия, конечно, не приводит однозначно ко *всей* конформной вселенной Эйнштейна, поскольку вне границы \mathcal{M} мы можем изгибать продолженное пространство произвольно. Однако замыкание $\bar{\mathcal{M}}$ пространства \mathcal{M} в \mathcal{E} однозначно определяется как конформное многообразие с границей. Для полной строгости было бы более логичным *удалить* точки I^- , I^0 , I^+ из определения $\bar{\mathcal{M}}$ и также из $\bar{\mathcal{M}} = \mathcal{I}$, так как в противном случае $\bar{\mathcal{M}}$ не было бы многообразием с границей в этих трех точках. Это будет сделано более аккуратно позднее в связи с обсуждением асимптотически плоских миров. (Однако я не всегда буду последователен в этом отношении, так как иногда бывает полезно говорить об I^- , I^0 , I^+ .) Есть несколько способов убедиться в том, что $\bar{\mathcal{M}}$ определяется конформно однозначно. По-видимому, наиболее прямым из них является построение точек \mathcal{I} в терминах классов эквивалентности световых геодезических в \mathcal{M} . [Точки $\bar{\mathcal{M}}$ можно определить другим путем — через горизонты события и частицы

(разд. 9). Нетривиальные горизонты можно использовать для определения точек \mathcal{M} .] Понятие *световой геодезической* является *конформно инвариантным*, поэтому здесь такое построение уместно. Мы должны решить, когда две световые геодезические в \mathcal{M} надо рассматривать как пересекающиеся на \mathcal{M} . Метод, развитый Герочем [34, 35, 118*], можно применять к произвольному пространству-времени. В данном случае, когда \mathcal{M} является миром Минковского, две световые геодезические встречаются в одной и той же точке на \mathcal{I}^+ тогда и только тогда, когда они принадлежат (согласно метрике ds) к одной и той же *световой гиперплоскости* в \mathcal{M} . Если же они просто параллельны в \mathcal{M} , они пересекают одну и ту же образующую точки I^{+1}). Для рассматриваемого частного случая мира \mathcal{M} две световые геодезические, встречающиеся на \mathcal{I}^+ , должны также встречаться на \mathcal{I}^- , и *наоборот*. Однако это очень специфическое свойство мира Минковского, и оно не имеет места в общем случае.

Перейдем к более общему миру \mathcal{M} , а именно к решению Шварцшильда для метрики внешнего поля сферически симметричного массивного тела. Я буду использовать метрику Эддингтона — Финкельштейна [27, 32] с одной запаздывающей световой координатой u :

$$ds^2 = du^2[1 - 2m/r] + 2drdu - r^2(d\theta^2 - \sin^2\theta d\varphi^2) \quad (r > 0). \quad (8.11)$$

(Эта метрика обсуждается более подробно в разд. 10.)

Если выбрать $d\hat{s} = \Omega ds$, где

$$\Omega = r^{-1} = l, \quad (8.12)$$

¹⁾ *Примечание к русскому изданию.* Например, для мира Минковского в обычных координатах световая геодезическая $x = t, y = z = 0$ параллельна геодезической $x = t + 1, y = z = 0$, но всякая гиперплоскость, содержащая обе эти линии, должна быть временноподобной. Следовательно, эти геодезические пересекают \mathcal{I}^+ в различных точках некоторой образующей точки I^+ . С другой стороны, $x = t, y = z = 0$ лежит в той же световой гиперплоскости, что и $x = t, y = 1, z = 0$, а именно в гиперплоскости $x = t$, так что эти линии встречаются на \mathcal{I}^+ в одной и той же точке.

а l — новая координата, конечная при $r = \infty$, то мы получим

$$d\hat{s}^2 = du^2 (l^2 - 2ml^3) - 2 dl du - d\theta^2 - \sin^2 \theta d\varphi^2 \quad (l \geq 0). \quad (8.13)$$

Эта метрика регулярна на \mathcal{Y}^+ , определенном как $l=0$, u конечно. Как и раньше, \mathcal{Y}^+ есть $S^2 \times E^1$. Чтобы найти \mathcal{Y}^- , перепишем (8.11) в терминах опережающей световой координаты $v = u + 2r + 4m \ln(r - 2m)$:

$$ds^2 = dv^2 \left(1 - \frac{2m}{r}\right) - 2 dr dv - r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \quad (r > 0) \quad (8.14)$$

и опять применим (8.12). Это дает метрику

$$d\hat{s}^2 = dv^2 (l^2 - 2ml^3) + 2 dl dv - d\theta^2 - \sin^2 \theta d\varphi^2 \quad (l \geq 0), \quad (8.14a)$$

которая регулярна на $\mathcal{Y}^- \approx S^2 \times E^1$, определенном посредством $l=0$. Единственное реальное различие между этим случаем и рассмотренным ранее, с точки зрения структуры бесконечности, состоит в том, что здесь мы не получаем *регулярных* точек I^- , I^0 , I^+ . Не удивительно, что I^- и I^+ оказываются сингулярными, так как источник сконцентрирован именно в этих точках — на двух концах его истории. Но I^0 — также сингулярная точка, в которой конформная кривизна бесконечна (хотя I^0 может быть присвоена конформная метрика класса C^0). Таким образом мы исключаем I^0 и I^\pm из определения $\bar{\mathcal{M}}$ (и из $\mathcal{Y} = \mathcal{Y}^- \cup \mathcal{Y}^+$).

Запись шварцшильдовской метрики в виде (8.11), (8.14) является частным случаем

$$ds^2 = r^{-2} A dr^2 - 2B_i dx^i dr + r^2 G_{ij} dx^i dx^j \quad (i, j = 1, 2, 3) \quad (8.15)$$

где A , B_i , C_{ij} — функции x^1 , x^2 , x^3 и r , достаточное число раз дифференцируемые (скажем, класса C^3) по x^a ($a=0, 1, 2, 3$) на и в окрестности гиперповерхности \mathcal{Y} , определенной как $x^0=0$, где $x^0=r^{-1}$.

Если предположить, что детерминант из метрических коэффициентов A , B_i , C_{ij} не обращается в нуль, то станет ясно, что метрика $d\hat{s} = \Omega ds$, где $\Omega = r^{-1}$ регулярна (C^3) на \mathcal{Y} . Метрика (8.15) включает все

метрики типа Бонди — Сакса, так что предположение о регулярности \mathcal{U} не кажется неприемлемым, если мы хотим изучать асимптотически плоские миры и допустить возможность гравитационного излучения. В действительности (8.15) является более общим, чем метрика Бонди — Сакса в предположении, что не налагаются уравнения поля, ограничивающие R_{ab} . В частности (8.15) включает мир де Ситтера и асимптотические миры де Ситтера. Эти случаи имеют место при условии, что в эйнштейновых уравнениях свободного поля присутствует космологический член.

Выбор Ω для (8.15), который требуется произвести, чтобы сделать метрику ds регулярной на \mathcal{U} , имеет то важное свойство, что градиент Ω на \mathcal{U} не равен нулю [$\partial\Omega/\partial x^a = (1, 0, 0, 0)$ на \mathcal{U}] и таким образом определяет направление нормали к \mathcal{U} . Предположение регулярности \mathcal{U} удобно дополнять условием нетривиальности градиента Ω на \mathcal{U} , но при некоторых обстоятельствах это условие можно вывести из предположения регулярности.

Сформулируем более точно условия, которые должны быть удовлетворены. Мы требуем, чтобы пространство-время \mathcal{M} с метрикой ds^2 было продолжимо до конформного многообразия с границей $\bar{\mathcal{M}} \supset \mathcal{M}$ ($\text{int } \bar{\mathcal{M}} = \mathcal{M}$; $\dot{\mathcal{M}} = \bar{\mathcal{M}} - \mathcal{M}$), причем так, чтобы выполнялись условия:

Существует гладкая (скажем, класса C^3 , по крайней мере) вещественная функция $\Omega (\geq 0)$ на $\bar{\mathcal{M}}$ и гладкая псевдориманова метрика $d\hat{s}^2$ на $\bar{\mathcal{M}}$ (совместимая с его конформной структурой), такая, что $ds^2 = \Omega^2 d\hat{s}^2$ на \mathcal{M} . (8.16)

На $\dot{\mathcal{M}}$ имеем $\Omega = 0$, $\hat{\nabla}_a \Omega \neq 0$. (8.17)

Каждая световая геодезическая в \mathcal{M} имеет две концевые точки на $\dot{\mathcal{M}}$. (8.18)

Если такое $\bar{\mathcal{M}}$ существует, мы будем называть \mathcal{M} *асимптотически простым*. Тогда $\dot{\mathcal{M}}$ определяется однозначно (как следует, например, из построения $\dot{\mathcal{M}}$ в терминах световых геодезических [34]). Мы пишем $\dot{\mathcal{M}} = \mathcal{U}$ для точек, находящихся на бесконечности \mathcal{M} .

Условие (8.18) обеспечивает включение всей бесконечности \mathcal{M} . Это условие обычно трудно проверить практически, и оно даже не выполняется в некоторых мирах, которые можно было бы трактовать как асимптотически *плоские*, но которые содержат ограниченные световые орбиты (т. е. световые геодезические, которые не уходят на бесконечность, как это имеет место при $r = 3t$ в решении Шварцшильда). Чтобы учесть такую возможность, будем называть пространство-время \mathcal{M} *асимптотически простым в слабом смысле*, если существует асимптотически простое \mathcal{M}_0 , такое, что для некоторого открытого подмножества \mathcal{K} множества $\bar{\mathcal{M}}_0$ (с $\dot{\mathcal{M}}_0 \subset \mathcal{K}$) область $\mathcal{M}_0 \cap \mathcal{K}$ (с метрикой, индуцированной из \mathcal{M}_0) изометрична подмножеству \mathcal{M} . Другими словами, асимптотически простое в слабом смысле пространство-время содержит конформную бесконечность $\mathcal{I} = \mathcal{M}_0$ асимптотически простого пространства-времени, но может также содержать и другие «бесконечности».

Чтобы получить связь с асимптотической евклидовостью Бонди — Сакса, предположим, что \mathcal{M} является асимптотически простым в слабом смысле с конформной бесконечностью \mathcal{I} . Используя уравнения Эйнштейна (7.1), можно показать, что если вблизи \mathcal{I} тензор T_{ab} не становится пропорциональным g_{ab} с ненулевым коэффициентом, то

\mathcal{I} является пространственноподобным, временно-подобным или световым в зависимости от того, является ли космологическая постоянная λ положительной, отрицательной или нулевой. (8.19)

(Если $\lambda \neq 0$, нет нужды предполагать условие $\hat{\nabla}_a \Omega \neq 0$ из (8.17), так как его можно вывести.) Если \mathcal{I} является пространственноподобным или световым, то оно представляет собой непересекающееся объединение двух гиперповерхностей \mathcal{I}^- , \mathcal{I}^{+1}). Точки

¹⁾ *Примечание к русскому изданию.* Только в том случае, когда \mathcal{I} является временноподобным, его нельзя разделить на \mathcal{I}^- и \mathcal{I}^+ . Если же \mathcal{I} является пространственноподобным или световым, ни одна точка \mathcal{I} не может принадлежать одновременно и \mathcal{I}^- и \mathcal{I}^+ .

\mathcal{I}^- отличаются от точек \mathcal{I}^+ тем, что их световые конуса будущего, а не прошлого, лежат в $\bar{\mathcal{M}}$. Поскольку мы интересуемся здесь больше асимптотической евклидовостью, чем космологией, предположим, что \mathcal{I} является световым. Отсюда следует¹⁾ (при условии, что \mathcal{M}_0 временно- и пространственноориентируемо и кроме того не содержит замкнутых временноподобных кривых, но это предположение, возможно, является излишним), что

$$\mathcal{I}^- \approx \mathcal{I}^+ \approx S^2 \times E^1, \quad (8.20)$$

где E^1 — световые образующие \mathcal{I}^\pm . Результат (8.20) означает, что, когда \mathcal{I} является световым, оно сильно напоминает бесконечность мира Минковского.

Чтобы двигаться дальше, нам нужно более строго представлять себе поведение T_{ab} вблизи \mathcal{I} . Для простоты вначале предположим, что $T_{ab} = 0$ (и $\lambda = 0$) в окрестности \mathcal{I} . Тогда мы можем сказать больше. Действительно,

$$\hat{\nabla}_{A(A'} \hat{\nabla}_{B')}^B \Omega = 0 \text{ на } \mathcal{I} \quad (8.21)$$

и вследствие (8.21) и (8.20) (ср. с [76]) получим

$$\hat{\Psi}_{ABCD} = 0 \text{ на } \mathcal{I}. \quad (8.22)$$

[Мы также можем получить $\hat{\nabla} \Omega \neq 0$ на \mathcal{I} , а не предполагать это в (8.17).] На основании (8.22) для главных световых направлений можно вывести *свойство последовательного вырождения*, которое было упомянуто в разд. 7. Аргументация будет приведена позже.

Тождества Бианки (8.2) в спинорной форме имеют вид

$$\nabla_{P'}^A \Psi_{ABCD} = \nabla_B^{Q'} \Phi_{CDQ'P'} - 2\epsilon_{B(C} \nabla_{D)P'} \Lambda. \quad (8.23)$$

Тогда при $T_{ab} = 0$ получаем

$$\nabla^{AP'} \Psi_{ABCD} = 0. \quad (8.24)$$

¹⁾ План доказательства этого результата приведен в приложении к [76]. Мне кажется, однако, что должен существовать другой, более простой вывод результата (8.20).

Уравнение (8.24) интересно тем, что оно обобщает уравнение свободного поля частиц нулевой массы в частном случае спина $s = 2$ на случай искривленного пространства. Для общего случая ненулевых спинов $s = 1/2, 1, 1 1/2, 2, \dots$ мы бы имели

$$\nabla^{AP'} \Phi_{AB \dots L} = 0, \quad (8.25)$$

где

$$\Phi_{AB \dots L} = \Phi_{(AB \dots L)} \quad (8.26)$$

есть спинор с $2s$ индексами. Для $s = 1/2$ уравнение (8.25) есть не что иное, как уравнение Вейля для нейтрино (уравнение Дирака для частиц нулевой массы). Для $s = 1$ выражение (8.25) есть спинорный вариант уравнений Максвелла в свободном пространстве

$$\nabla_{[a} F_{bc]} = 0, \quad \nabla^a F_{ab} = 0 \quad (8.27)$$

с максвелловским тензором

$$F_{ab} = -F_{ba} = \Phi_{AB} \epsilon_{A'B'} + \epsilon_{AB} \bar{\Phi}_{A'B'}. \quad (8.28)$$

В плоском пространстве-времени (8.25) является вполне удовлетворительным уравнением, решения которого имеют для *любого* значения s то же самое число «степеней свободы» (а именно, две), что и в двух случаях $s = 1/2, 1$. Однако в искривленном пространстве-времени мы имеем условие совместности [16, 84]

$$\begin{aligned} (2 - 2s) \Psi^{ABM} {}_{(C} \Phi_{D \dots L) ABM} &= \nabla_{P'}^{(B} \nabla^{A) P'} \Phi_{ABCD \dots L} = \\ &= \nabla_{P'}^B \nabla^{AP'} \Phi_{ABCD \dots L} = 0, \end{aligned} \quad (8.29)$$

так что при $s \geq 3/2$ уравнение (8.25) определяет связь между $\Phi_{AB \dots L}$ и спинором конформной кривизны Ψ_{ABCD} . Но в случае $\Phi_{ABCD} = \Psi_{ABCD}$ уравнение (8.29) удовлетворяется автоматически в силу условий симметрии. Мы оставим в стороне вопрос об уравнениях поля для высших спинов и займемся интерпретацией (8.25) только в случаях $s = 1/2$ (нейтрино), $s = 1$ (электромагнетизм) и $s = 2$ (гравитация).

Важным свойством (8.25) является его *конформная инвариантность*, если принять

$$\hat{\Phi}_{AB \dots L} = \Omega^{-1} \Phi_{AB \dots L} \quad (8.30)$$

Нам потребуются различные формулы конформных преобразований. Из $ds = \Omega ds$ мы имеем

$$\hat{g}_{ab} = \Omega^2 g_{ab}, \quad \hat{g}^{ab} = \Omega^{-2} g^{ab} \quad (8.31)$$

и можем вывести ¹⁾

$$\begin{aligned} \hat{\varepsilon}_{AB} &= \Omega \varepsilon_{AB}, & \hat{\varepsilon}_{A'B'} &= \Omega \varepsilon_{A'B'}, \\ \hat{\varepsilon}^{AB} &= \Omega^{-1} \varepsilon^{AB}, & \hat{\varepsilon}^{A'B'} &= \Omega^{-1} \varepsilon^{A'B'}. \end{aligned} \quad (8.32)$$

Для скаляра χ получаем

$$\hat{\nabla}_a \chi = \nabla_a \chi. \quad (8.33)$$

Для величин с индексами имеем, например,

$$\begin{aligned} \hat{\nabla}_{AA'} \xi_B &= \nabla_{AA'} \xi_B - v_{BA'} \xi_A, & \hat{\nabla}_{AA'} \eta_{B'} &= \nabla_{AA'} \eta_{B'} - v_{AB'} \eta_{A'}, \\ \hat{\nabla}_{AA'} \xi^B &= \nabla_{AA'} \xi^B + \varepsilon_A{}^B v_{CA'} \xi^C, & & \\ \hat{\nabla}_{AA'} \eta^{B'} &= \nabla_{AA'} \eta^{B'} + \varepsilon_{A'}{}^{B'} v_{AC} \eta^C, & & \end{aligned} \quad (8.34)$$

где

$$v_{AA'} = \Omega^{-1} \nabla_{AA'} \Omega. \quad (8.35)$$

При обращении с величинами, имеющими более одного индекса, мы просто рассматриваем их как произведение, согласно правилу Лейбница. Другими словами, мы получаем по одному члену $v_{\chi\gamma'}$ для каждого индекса. Для проверки правильности (8.34) надо лишь показать, что удовлетворяются аксиомы (4.50) — (4.58).

После подстановки (8.30) в (8.25) с учетом (8.34) получаем выражение

$$\hat{\nabla}^{AP'} \hat{\phi}_{AB \dots L} = 0, \quad (8.36)$$

которое и доказывает конформную инвариантность. Однако применять это уравнение к гравитационному

¹⁾ Несколько иное, по сравнению с [76], определение конформного преобразования спиноров, принятое здесь, приводит к отличающимся формулам преобразования. Это различие обусловлено постановкой вопроса, принятой в разд. 3. В действительности формулы, полученные здесь, несколько проще тех, которые фигурируют в [76]; кроме того, устранены дробные степени Ω .

полю надо с осторожностью. Тензор Вейля конформно инвариантен, т. е. $\hat{C}^a{}_{bcd} = C^a{}_{bcd}$ и $\hat{C}_{abcd} = \Omega^2 C_{abcd}$. Следовательно,

$$\hat{\Psi}_{ABCD} = \Psi_{ABCD}. \quad (8.37)$$

Таким образом, если в соответствии с физической метрикой ds мы полагаем

$$\Phi_{ABCD} = \Psi_{ABCD}, \quad (8.38)$$

то, согласно (8.30) и (8.37), имеем

$$\hat{\Phi}_{ABCD} = \Omega^{-1} \hat{\Psi}_{ABCD}. \quad (8.39)$$

Именно поле $\hat{\Phi}_{ABCD}$ мы отождествляем с гравитационным полем, когда исследуем $\bar{\mathcal{M}}$, соответствующее метрике $d\hat{s}$.

Значение этого состоит в поведении поля на \mathcal{I} и в связи с этим поведением со свойством последовательного вырождения. Напомним, что согласно (8.22) Ψ_{ABCD} обращается в нуль на \mathcal{I} . Из (8.39) и условия гладкости получаем

$$\hat{\Phi}_{ABCD} \text{ непрерывно на } \mathcal{I}. \quad (8.40)$$

Теперь можно без всяких усилий вывести свойство последовательного вырождения для произвольного спина s , предполагая непрерывность $\hat{\Phi}_{AB\dots L}$ на \mathcal{I} .

Введем спиновую систему отсчета $o^A = \epsilon_0^A$, $\iota^A = \epsilon_1^A$. Для метрики $d\hat{s}$ положим

$$\hat{o}_A = o_A, \quad \hat{\iota}_A = \Omega \iota_A, \quad \hat{\delta}^A = \Omega^{-1} o^A, \quad \hat{\iota}^A = \iota^A; \quad (8.41)$$

это согласуется с (8.32) и (4.27). Мы сделали асимметричный выбор (8.41) для того, чтобы направить o_A вдоль световой геодезической η , причем o_A , ι_A переносятся параллельно (относительно ds) вдоль η :

$$\nabla_{00'} o^A = 0, \quad \nabla_{00'} \iota^A = 0. \quad (8.42)$$

Тогда из (8.34) и (8.41) следует

$$\hat{\nabla}_{00'} \hat{\delta}^A = 0, \quad \hat{\nabla}_{00'} \hat{\iota}^A = \Omega^{-1} \hat{\delta}^A \hat{\nabla}_{10'} \Omega. \quad (8.43)$$

Таким образом, o^A переносится параллельно вдоль η относительно метрики ds и достигает

определенного значения в точке $Q^\pm = \eta \cap \mathcal{I}^\pm$. На самом деле \hat{t}^A также имеет определенное предельное значение в Q^\pm и направлено вдоль \mathcal{I}^\pm в точке Q^\pm (так что $\hat{\nabla}_{10} \Omega = 0$ на \mathcal{I} ; ср. с [76]). Аффинный (относительно ds) параметр r на η , определенный как

$$\nabla_{00} r = 1, \quad (8.44)$$

имеет асимптотическое поведение

$$r \sim \Omega^{-1}. \quad (8.45)$$

Далее, по предположению, все компоненты $\hat{\Phi}_{\text{AB} \dots \text{L}}$ относительно спиновой системы \hat{o}^A, \hat{i}^A непрерывны в Q^\pm . Используя (8.30), (8.41), (8.45), мы сформулируем это на языке физических величин следующим образом:

$$\text{Существует } \lim_{r \rightarrow \pm \infty} \{r^{2s+1-i} \Phi_{(i)}\} \quad (i = 0, 1, \dots, 2s), \quad (8.46)$$

где $\Phi_{(i)}$ обозначает вырождение $\Phi_{00 \dots 01 \dots 1}$, содержащее i «единиц» и $(2s - i)$ «нулей». Уравнение (8.46) гласит, что r^{-k} (где $k = 1, 2, \dots, 2s$) часть $\Phi_{AB \dots L}$ вдоль η удовлетворяет соотношению

$$\Phi_{AB \dots DE \dots L} o^A o^B \dots o^D = 0, \quad (8.47)$$

содержащему k спиноров o^A . Условие (8.47) есть как раз условие того, что по крайней мере $2s - k + 1$ главных световых направлений $\Phi_{A \dots L}$ совпадают в направлении η . Это и есть свойство последовательного вырождения.

В случае гравитации свойство последовательного вырождения является характерной чертой асимптотической евклидовости Бонди—Сакса. Можно связать более непосредственно точный тип координат, используемых в подходе Бонди—Сакса, с конформной структурой, возникающей в рассматриваемом формализме. Это строго сделали Тамбурино и Виникур [104]. Поэтому в данном подходе мы можем определить асимптотическую евклидовость \mathcal{M} как асимптотическую простоту в слабом смысле в случае, когда $\lambda = 0$ и $T_{ab} = 0$ вблизи \mathcal{I} . Если же T_{ab} только стре-

мится к нулю некоторым образом при приближении к бесконечности, нам могут потребоваться добавочные условия типа (8.21) и (8.22), в дополнение к асимптотической простоте в слабом смысле. Если вблизи \mathcal{I} выполняются уравнения Эйнштейна — Максвелла, такие добавочные условия не нужны (за исключением, возможно, некоторых особых случаев), и свойство последовательного вырождения можно вывести для обоих полей — гравитационного и электромагнитного.

Возвратимся на короткое время к вопросу об энергии и импульсе, с которого мы начали. Если уходящая в будущее световая гиперповерхность \mathcal{N} пересекается с \mathcal{I}^+ по 2-поверхности \mathcal{R} (рис. 18), то мы

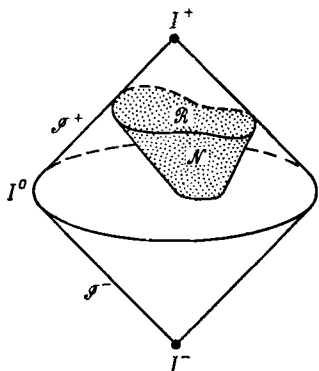


Рис. 18. Энергия-импульс по Бонди — Саксу является интегралом по $\mathcal{R} = \mathcal{N} \cap \mathcal{I}^+$.

можем вычислить интеграл по \mathcal{R} , чтобы получить полную энергию и импульс, содержащиеся на произвольной гиперповерхности, натянутой на \mathcal{R} . Если выбрать Ω так, чтобы метрика на \mathcal{R} была метрикой единичной сферы, то, интегрируя

$$\hat{\sigma}N - \Psi_2 \quad (8.48)$$

по \mathcal{R} , мы получим энергию (связанную с выбором Ω), а интегрируя (8.48) с надлежащим весовым фактором, получим импульсные компоненты. Здесь

$$\Psi_2 = \hat{\phi}_{0011} \quad (8.49)$$

на \mathcal{R} , так что Ψ_2 происходит от r^{-3} -части тензора Вейля. Величина $\hat{\sigma}$ является сдвигом $\hat{\Gamma}_{0001}$ гиперповерхности \mathcal{N} , вычисленным на \mathcal{R} (с δ^A и i^A , опре-

деляющими световые направления в \mathcal{N} и \mathcal{I} соответственно). Величина N является «функцией новостей» Бонди — Сакса, определенной здесь как

$$N = \hat{\Phi}_{110'0'}. \quad (8.50)$$

Хотя N представляет собой компоненту тензора Риччи, она сохраняет определенную конформную ковариантность благодаря указанному выше ограничению на Ω . Оказывается, полная конструкция энергии-импульса имеет правильные трансформационные свойства как «асимптотический 4-вектор».

По существу формула Бонди — Сакса для потери массы выражает тот факт, что если мы повторим интегрирование с новой световой гиперповерхностью, лежащей в будущем по отношению к предыдущей, то полученная масса всегда будет меньше или равна первоначальной. Масса, уносимая гравитационными волнами, равна интегралу от $N\bar{N}$ по части \mathcal{I}^+ , лежащей между двумя световыми гиперповерхностями; \bar{N} можно описать как интеграл от $\Psi_4 = \hat{\Phi}_{1111}$ вдоль образующих \mathcal{I}^+ . [Фактически это следует из (8.50).] В разложении по r величина Ψ_4 является r^{-1} -частью гравитационного поля; следовательно, эту величину можно рассматривать как уходящее поле излучения. По существу «функция новостей» является «интегралом по времени от уходящего гравитационного поля излучения». Оказывается, формула Бонди — Сакса для потери массы тесно связана с положительным фокусированием (7.39) на световой линии γ , которая лежит строго внутри \mathcal{I}^+ . При этом N играет роль σ в (7.39), а Ψ_4 — роль Ψ (см. [78]).

Можно отметить, что потеря массы имеет место *безотносительно* к наличию *приходящего* излучения (аналог Ψ_4 на \mathcal{I}^-). Предположение об асимптотической евклидовости на \mathcal{I}^+ *не исключает* приходящих волн (как когда-то считалось), хотя оно запрещает слишком большую концентрацию приходящих волн внутри \mathcal{I}^+ . Чтобы получить *прирост* массы вследствие приходящего излучения, мы должны обратить рассмотренную выше конструкцию и проделать вычисления на \mathcal{I}^- .

И, наконец, последнее замечание об энергии в общей теории относительности. Мы видим, что последовательно измеряемая масса всегда убывает и нет никакой гарантии, что масса не станет отрицательной! В том случае, когда все источники в конце концов уйдут на \mathcal{I}^+ в виде поля излучения нулевой массы покоя, можно сказать на основании формулы Бонди — Сакса, что полная масса всегда *была* положительной¹⁾. Насколько я знаю, еще не решен вопрос о том, является ли полная масса положительной в *общем* случае, когда мы предполагаем лишь подходящие, физически разумные неравенства на T_{ab} .

¹⁾ Брилл [14] получил некоторые результаты, относящиеся к положительной определенности полной массы системы, измеренной на пространственноподобной гиперповерхности. Брилл и Дезер [14а] недавно предложили новый перспективный подход к этой проблеме.

9.

Горизонты

Рассматривая частные космологические модели, иногда полезно сделать конформное представление бесконечности, даже в тех случаях, когда пространство-время не является асимптотически простым в слабом смысле. Например, в обычных моделях Робертсона — Уолкера мы можем представить начальную сингулярность с бесконечной кривизной как несингулярную гиперповерхность, изображающую границу в прошлом. Это можно сделать, если предположить, что $\Omega = \infty$ в этой области. Природа таких граничных гиперповерхностей тесно связана с четырьмя типами горизонтов, которые могут встретиться в космологии: горизонт событий, горизонт частицы [88] и горизонты Коши [40, 42], будущий и прошедший. Ниже все они будут определены точно.

Прежде всего дадим некоторые определения. Под *кривой* я подразумеваю *образ*, получаемый при помощи отображения f класса C^1 (с необращающейся в нуль производной) действительного интервала ненулевой длины, причем я требую, чтобы каждая точка многообразия M имела окрестность \mathcal{L} , для которой $f^{-1} \bar{\mathcal{L}}$ не имеет некомпактной связной компоненты. Таким образом, предполагается существование *концевых точек* (в которых кривая должна быть гладкой и поэтому продолжимой), за исключением случаев, когда кривая продолжается неограниченно в том или в ином направлении. (В противном случае мы бы имели дело с открыто-замкнутой кривой.) Если рассматривается само отображение f , то кривая является *параметризованной*. Кривая называется *временноподобной* (соответственно — *непространственноподобной*), если все касательные векторы отображения f яв-

ляются временноподобными (соответственно — временноподобными или световыми). Непространственноподобная кривая имеет естественную ориентацию, называемую *ориентацией в будущее*, которая индуцируется временной ориентацией \mathcal{M} .

Если P и Q — две точки \mathcal{M} , то (см. [50]) обозначение $P \ll Q$ используется тогда, когда существует временноподобная кривая с начальной точкой P и конечной точкой Q . Обозначение $P < Q$ используется тогда, когда либо $P = Q$, либо существует непространственноподобная кривая от P до Q . Если $P < Q$ и $Q \ll R$ или $P \ll Q$ и $Q < R$, то $P \ll R$. Если $P < Q$ (и $P \neq Q$), тогда либо $P \ll Q$, либо существует световая геодезическая от P до Q (либо и то, и другое). Хронологическое *будущее* и *прошлое* точки P обозначаются соответственно как

$$I_+(P) = \{X : P \ll X\}, \quad I_-(P) = \{X : X \ll P\}. \quad (9.1)$$

Как легко видеть, они являются *открытыми* множествами. С другой стороны, множества

$$J_+(P) = \{X : P < X\}, \quad J_-(P) = \{X : X < P\} \quad (9.2)$$

не обязательно замкнуты. Примером может служить мир Минковского с выброшенной точкой, в котором P расположено на световом конусе выброшенной точки. В качестве другого примера можно рассмотреть плоскую волну [см. (9.23) — (9.31)]. Если \mathcal{H} — произвольное подмножество \mathcal{M} , то его хронологическое будущее и прошлое обозначаются соответственно как

$$I_+[\mathcal{H}] = \bigcup_{P \in \mathcal{H}} I_+(P), \quad I_-[\mathcal{H}] = \bigcup_{P \in \mathcal{H}} I_-(P). \quad (9.3)$$

Эти множества также являются открытыми. Их *замыканиями* являются:

$$\begin{aligned} \bar{I}_+[\mathcal{H}] &= \{X : I_+(X) \subset I_+[\mathcal{H}]\}, \\ \bar{I}_-[\mathcal{H}] &= \{X : I_-(X) \subset I_-[\mathcal{H}]\}. \end{aligned} \quad (9.4)$$

Доказательство (9.4) тривиально, и мы его не приводим. Из (9.4) получаем¹⁾ границы $I_{\pm}[\mathcal{K}]$

$$\begin{aligned} I_+[\mathcal{K}] &= \{X : I_+(X) \subset I_+[\mathcal{K}], X \notin I_+[\mathcal{K}]\}, \\ I_-[\mathcal{K}] &= \{X : I_-(X) \subset I_-[\mathcal{K}], X \notin I_-[\mathcal{K}]\}. \end{aligned} \quad (9.5)$$

Множество вида $I_+[\mathcal{K}]$ будем называть *полупространственноподобной границей* (сокращенно ППГ). Оно является *симметричным по времени*, так как

$$I_+[\mathcal{K}] = I_-[\mathcal{L}], \quad (9.6)$$

где \mathcal{L} есть дополнение $I_+[\mathcal{K}]$ в \mathcal{M} . И наоборот, (9.6) выполняется, если \mathcal{L} — произвольное подмножество \mathcal{M} , а \mathcal{K} является дополнением к $I_-[\mathcal{L}]$. Вообще, *полупространственноподобное множество* — это такое подмножество множества \mathcal{M} , у которого никакие две точки X, Y не удовлетворяют соотношению $X \ll Y$. Ясно, что множество ППГ — полупространственноподобное. Хотя множества ППГ не обязательно должны быть гладкими подмножествами \mathcal{M} , они обладают некоторыми привлекательными свойствами, которые будут играть важную роль в «сингулярных теоремах» в разд. 10 и 11. Пока мы просто отметим, что

Любое ППГ является трехмерным, класса C^0 , подмногообразием \mathcal{M} (без границы) и представляет собой замкнутое множество. (9.7)

Чтобы убедиться в том, что ППГ локально гомеоморфно E^3 , рассмотрим малую окрестность в нормальных координатах x^a произвольной точки P , принадлежащей ППГ, причем $\nabla_a x^0$ временноподобно и $\nabla_a x^1$ пространственноподобно, $i \neq 0$. Любая кривая $x^1 = \text{const}$, $i \neq 0$, достаточно близкая к P , пересекает $I_+(P)$ и $I_-(P)$. Следовательно, она пересекает ППГ в определенной точке. Это дает локальный гомеоморфизм на открытое множество \mathbf{R}^3 чисел (x^1, x^2, x^3) (рис. 19).

¹⁾ Обозначение $\mathcal{A} \subset \mathcal{B}$ подразумевает только, что \mathcal{A} является подмножеством \mathcal{B} , но не обязательно собственным подмножеством (так что $\mathcal{A} \subset \mathcal{A}$).

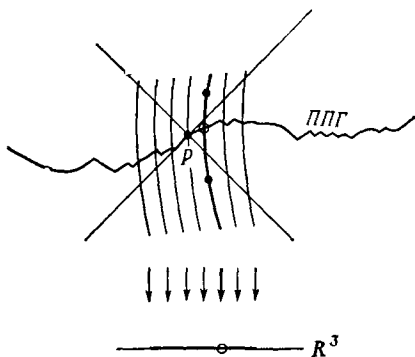


Рис. 19. Любая полупространственноподобная граница ППГ локально гомеоморфна E^3 .

Теперь мы можем определить четыре типа горизонтов. Пусть γ — произвольная временноподобная кривая в \mathcal{M} . Тогда (если соответствующее множество не пустое)

$$\text{Горизонт событий для } \gamma \text{ есть } \dot{I}_-[\gamma], \quad (9.7a)$$

$$\text{Горизонт частицы для } \gamma \text{ есть } \dot{I}_+[\gamma].$$

Если кривая γ имеет конечную точку F , то горизонт событий для нее есть просто $\dot{I}_-(F)$, и если γ имеет начальную точку P , горизонт частицы для нее есть $\dot{I}_+(P)$. Мы будем называть эти случаи *тривиальными*. Примером пространства-времени, *не содержащего* нетривиальных горизонтов события и частицы, является статический мир Эйнштейна, рассмотренный в разд. 8. Однако мир Минковского *обладает* нетривиальными горизонтами события и частицы, например, в случае, когда γ изображает мировую линию тела, движущегося с постоянным ускорением. [См. (2.8)–(2.10) и рис. 6. Если γ есть $Z = \text{const} > 0$, $X = Y = 0$, то $z = t$ обозначает горизонт событий для γ , а $z = -t$ есть горизонт частицы для γ .] С другой стороны, ни одна из временноподобных *геодезических* в пространстве-времени Минковского не имеет горизонта событий или частицы. Физически горизонт событий кривой γ разделяет наблюдаемые и ненаблюдаемые события для наблюдателя, мировая линия которого есть γ . Аналогично горизонт частицы разделяет области

Рис. 20. Если \mathcal{I}^+ пространственноподобно, всякая временноподобная кривая имеет горизонт событий.

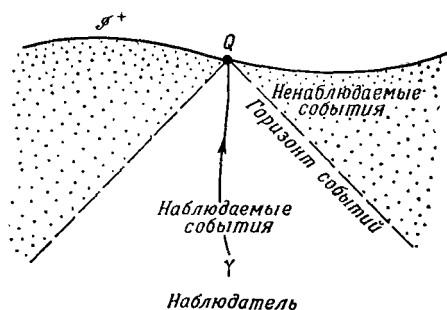
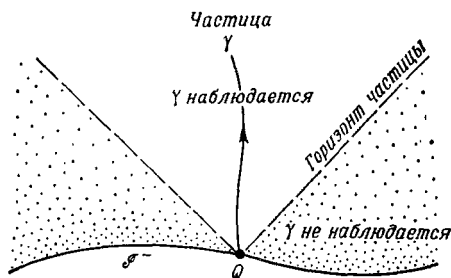


Рис. 21. Если \mathcal{I}^- пространственноподобно, всякая временноподобная кривая имеет горизонт частицы.



мира, из которых частица с мировой линией γ может или не может в принципе наблюдаться.

Примером пространства-времени с нетривиальным горизонтом событий или частицы для геодезической γ может служить \mathcal{M} , содержащее пространственноподобную конформную бесконечность \mathcal{I}^+ или \mathcal{I}^- . (Допускается $\Omega \neq 0$ или $\Omega = \infty$ на \mathcal{I} .) Если \mathcal{I}^+ — пространственноподобно, то всякая временноподобная геодезическая γ (в действительности всякая продолжаемая временноподобная кривая в \mathcal{M}), проходящая через некоторую точку \mathcal{M} , достаточно близкую к \mathcal{I}^+ , пересечет \mathcal{I}^+ в точке Q в $\bar{\mathcal{M}}$. Мы можем определить $\dot{I}_-(Q)$ (используя конформную структуру $\bar{\mathcal{M}}$); это и будет искомым горизонт событий $\dot{I}_-[\gamma]$ (рис. 20). Совершенно аналогично, если \mathcal{I}^- — пространственноподобно, мы можем найти временноподобную геодези-

ческую γ с нетривиальным горизонтом частицы (рис. 21).

Чтобы определить горизонт Коши, предположим, что \mathcal{P} есть некоторое замкнутое полупространственноподобное множество. Определим *область зависимости* $D_+(\mathcal{P})$, $D_-(\mathcal{P})$ следующим образом:

$D_+(\mathcal{P})$ есть дополнение к объединению всех временноподобных кривых, продолжимых в прошлое и не пересекающих \mathcal{P} . (9.8)

$D_-(\mathcal{P})$ определяется просто как обращение во времени формулировки (9.8). Другими словами:

$X \in D_+(\mathcal{P})$ тогда и только тогда, когда всякая временноподобная кривая, проходящая через X , может быть продолжена в прошлое до пересечения с \mathcal{P} , (9.9)

и аналогичным образом для $D_-(\mathcal{P})$. [Принимаем $\mathcal{P} \subset D_+(\mathcal{P})$.] Легко видеть, что продолжимые в прошлое кривые, не пересекающие \mathcal{P} , охватывают открытое подмножество множества \mathcal{M} . Следовательно, из (9.8)

$D_+(\mathcal{P})$ и $D_-(\mathcal{P})$ есть замкнутые множества. (9.10)

Горизонтами Коши (рис. 22) являются «граница в будущем» множества $D_+(\mathcal{P})$ и «граница в прошлом» множества $D_-(\mathcal{P})$:

$H_+(\mathcal{P}) = \{X : X \in D_+(\mathcal{P}), I_+(X) \cap D_+(\mathcal{P}) = \emptyset\},$
 $H_-(\mathcal{P}) = \{X : X \in D_-(\mathcal{P}), I_-(X) \cap D_-(\mathcal{P}) = \emptyset\}.$ (9.11)

Легко получить, что

$H_+(\mathcal{P})$ и $H_-(\mathcal{P})$ есть полупространственноподобные замкнутые множества. (9.12)

Мы можем определить ¹⁾ *край* замкнутого полупространственноподобного множества \mathcal{P} как множе-

¹⁾ *Примечание к русскому изданию.* Это определение сформулировано не вполне корректно. Правильная формулировка содержится в [131]: край ахронального (полупространственноподобного) замкнутого множества \mathcal{P} определяется как множество точек $P \in \mathcal{P}$, таких, что если $R \ll P \ll Q$ и γ есть временноподобная кривая от R до Q , проходящая через P , то всякая окрестность γ содержит временноподобную кривую от R до Q , не пересекающую \mathcal{P} .

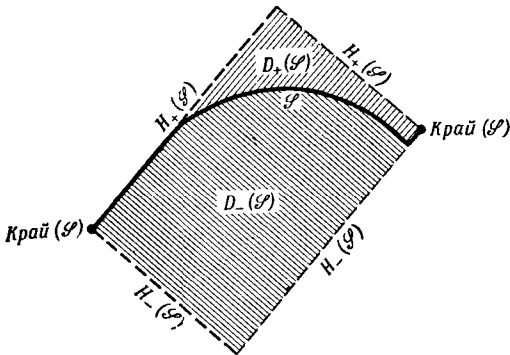


Рис. 22. Области зависимости и горизонты Коши для полупространственноподобного множества \mathcal{S} (в «тривиальном» случае).

ство точек $P \in \mathcal{S}$, таких, что для любого открытого множества \mathcal{A} , содержащего P , и для точек $R, Q \subset \mathcal{A}$, $R \ll P \ll Q$ существует временноподобная кривая в \mathcal{A} от R до Q , не пересекающая \mathcal{S} . Имеем:

$$\text{Край}(\mathcal{S}) \subset H_+(\mathcal{S}) \cap H_-(\mathcal{S}). \quad (9.13)$$

Множество «край (\mathcal{S})» всегда является замкнутым. Оно состоит из тех точек \mathcal{S} , в которых \mathcal{S} не локально гомеоморфно E^3 . Если $\text{край}(\mathcal{S}) = \emptyset$, мы можем назвать \mathcal{S} *бескраевым*. Если \mathcal{S} — гладкое полупространственноподобное множество, являющееся *гиперповерхностью* (без границы) в \mathcal{M} (так что касательные гиперплоскости всюду пространственноподобные или световые), то $\overline{\mathcal{S}}$ бескраевое тогда и только тогда, когда \mathcal{S} есть *замкнутое* подмножество в \mathcal{M} (т. е. \mathcal{S} является собственно вложенным.) И другой пример: всякое множество ППГ — бескраевое.

Горизонт Коши (если он существует) будем называть *нетривиальным*, если он представляет собой $H_+(\mathcal{S})$ или $H_-(\mathcal{S})$ некоторого замкнутого, бескраевого полупространственноподобного множества \mathcal{S} . Тривиальные горизонты Коши легко построить (в мирах без замкнутых временноподобных кривых), принимая

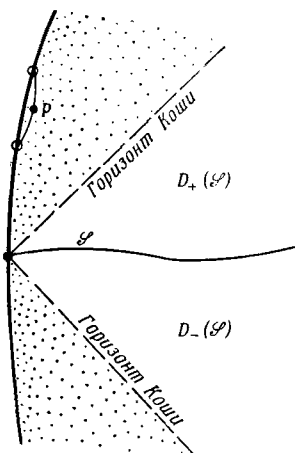


Рис. 23. Если \mathcal{U} временноподобно, то каждое полупространственноподобное множество имеет горизонт Коши.

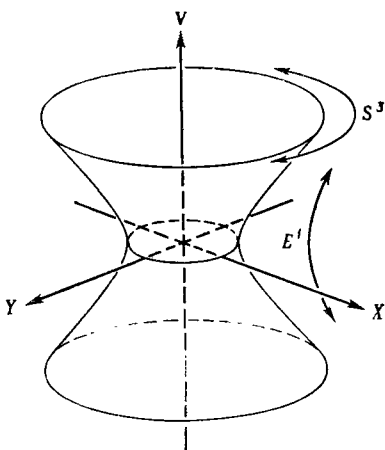


Рис. 24. Модель де Ситтера.

в качестве \mathcal{S} малый замкнутый кусок пространственноподобной гиперповерхности. Статический мир Эйнштейна не содержит нетривиальных горизонтов Коши. Однако мир Минковского содержит нетривиальные горизонты Коши. Например, если \mathcal{S} есть $x^0 = x^1$ в обычных координатах Минковского, то $H_{\pm}(\mathcal{S}) = \mathcal{S}$; если \mathcal{S} есть $x^0 = -[1 + (x^1)^2 + (x^2)^2 + (x^3)^2]^{1/2}$, то $H_{-}(\mathcal{S}) = \emptyset$, но $H_{+}(\mathcal{S})$ определяется как $x^0 = -[(x^1)^2 + (x^2)^2 + (x^3)^2]^{1/2}$.

Глобальной гиперповерхностью Коши (или ГГК) для \mathcal{M} называется полупространственноподобное множество \mathcal{S} (обычно — гладкая гиперповерхность), для которого $H_{+}(\mathcal{S})$ и $H_{-}(\mathcal{S})$ пусты. Другими словами, \mathcal{S} пересекает каждую продолжимую временноподобную кривую в \mathcal{M} . Эквивалентное условие того, чтобы полупространственноподобное множество \mathcal{S} было ГГК, состоит в том, что \mathcal{S} должно пересекать каждую продолжимую световую геодезическую в \mathcal{M} по непустому компактному множеству. Это вытекает из последующего рассмотрения структуры горизонтов Коши. Нетрудно видеть, что пересечение световой геодезиче-

ской с ППГ должно быть *связным*. Легко установить одно свойство пространства-времени, содержащего ГГК \mathcal{S} : топологически мир является произведением $\mathcal{S} \times E^1$, где E^1 могут быть представлены временноподобными кривыми. Труднее доказать¹⁾, что все \mathcal{S} топологического произведения также могут служить ГГК для \mathcal{M} . Оказывается также, что существование ГГК эквивалентно условию «глобальной гиперболичности» Лере²⁾.

Ясно, что обычные сечения «постоянного времени» статического мира Эйнштейна и мира Минковского являются ГГК. Все миры, которые обладают *временноподобной* конформной бесконечностью \mathcal{S} , не содержат ни одной ГГК (рис. 23). Всякая точка P вблизи \mathcal{S} лежит на временноподобных кривых, которые и в прошлом и в будущем достигают \mathcal{S} в некоторой окрестности P (с точки зрения $\bar{\mathcal{M}}$). Таким образом, любое \mathcal{S} , претендующее на роль ГГК, должно было бы проходить вблизи P . Поскольку это рассуждение применимо ко *всем* точкам вблизи \mathcal{S} , причем некоторые из них имеют между собой конечный временноподобный интервал, то отсюда видно, что искомого полупространственноподобного \mathcal{S} не существует.

Деситтеровский и антидеситтеровский миры являются точными примерами, в которых \mathcal{S} , соответственно, пространственноподобно и временноподобно. Чтобы описать эти модели, рассмотрим гиперболы в пятимерном псевдоевклидовом пространстве с двумя типами сигнатуры. Мир де Ситтера (рис. 24) определяется как

$$C^2 = -V^2 + W^2 + X^1 + Y^2 + Z^2, \quad (9.14)$$

где метрика есть

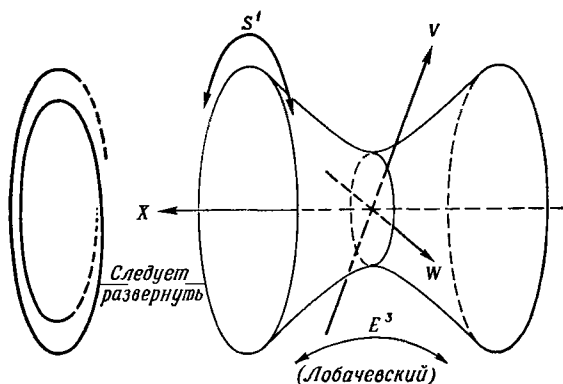
$$ds^2 = dV^2 - dW^2 - dX^2 - dY^2 - dZ^2. \quad (9.15)$$

Антидеситтеровский мир (рис. 25) является универсальным накрывающим многообразием

$$C^2 = V^2 + W^2 - X^2 - Y^2 - Z^2 \quad (9.16)$$

¹⁾ Одно из доказательств этого факта дано Зейфергом [98].

²⁾ Это будет показано в другой работе.



Р и с. 25. Антидеситтеровская модель.

с метрикой

$$ds^2 = dV^2 + dW^2 - dX^2 - dY^2 - dZ^2. \quad (9.17)$$

Топологически мир де Ситтера есть $S^3 \times E^1$ и обладает сечениями S^3 , которые являются ГГК. Но каждая временноподобная кривая имеет горизонт частицы и горизонт событий. Мы можем представить пространство-время конформно на статической вселенной Эйнштейна в виде области между двумя параллельными пространственноподобными сечениями (рис. 26). Очевидным образом \mathcal{Y}^+ и \mathcal{Y}^- являются пространственноподобными. Модель стационарной вселенной есть просто «верхняя половина» модели де Ситтера (см., например, [97]), которую мы отрезаем, скажем, гиперплоскостью $V = W$ в (9.14) (рис. 27). При этом световая граница остается в прошлом, так что некоторые временноподобные кривые (например, мировые линии частиц среды) не обладают горизонтом частицы. Сечения $V - W = \text{const} > 0$ в (9.14) дают (плоские) гиперповерхности постоянного времени в стационарной модели. В отличие от деситтеровского и антидеситтеровского миров эта модель не является геодезически полной.

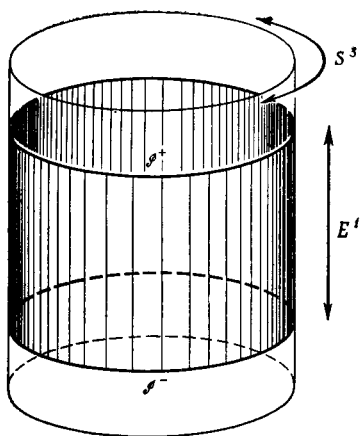


Рис. 26. Модель де Ситтера как конформная часть вселенной Эйнштейна. Модель Фридмана с $\lambda=0$, $k>0$ представляется аналогичным образом.

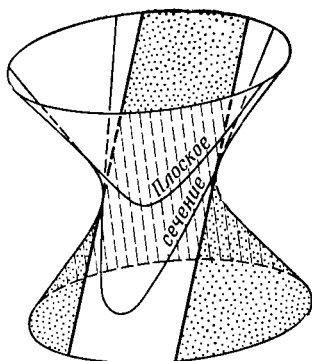


Рис. 27. Модель стационарной вселенной.

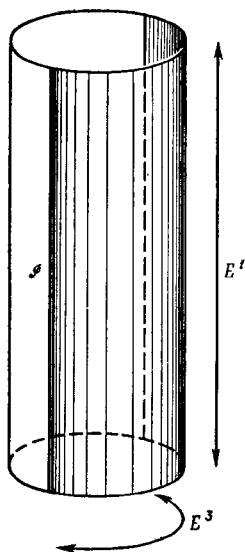


Рис. 28. Антидезиттеровская модель как конформная часть вселенной Эйнштейна.

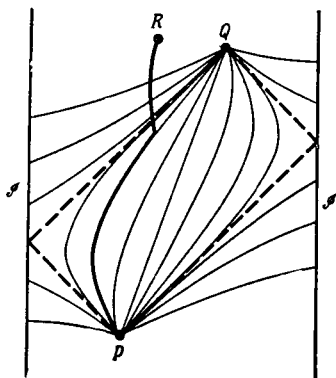
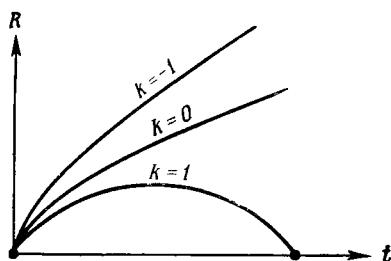


Рис. 29. Точки P и R нельзя соединить геодезической.

Гиперсфера (9.16) имеет топологию $S^1 \times E^3$ и поэтому она не односвязна. Она обладает замкнутыми временноподобными кривыми (S^1), которые могут быть разомкнуты в универсальном накрывающем многообразии. Тогда «развернутая» антидеситтеровская модель, представленная конформно на статической вселенной Эйнштейна $S^3 \times E^1$, становится просто полусферой S^3 , умноженной на E^1 (рис. 28). Ее топология есть E^4 , но *конформная* структура глобально отличается от пространства-времени Минковского. В данном случае мы имеем *временноподобное* \mathcal{P} , и поэтому каждое полупространственноподобное \mathcal{P} имеет горизонт Коши. С другой стороны, ни одна из продолжимых временноподобных геодезических не обладает горизонтом частицы или событий (хотя некоторые продолжимые временноподобные кривые обладают им). Существование горизонтов Коши связано с интересной чертой антидеситтеровской модели. Пусть P — произвольная точка на гиперсфере (9.16), а Q — диаметрально ей противоположная. Любая плоскость, проходящая через PQ , пересекает гиперсферу по «большому кругу», который является геодезической. Нетрудно видеть, что всякая временноподобная геодезическая на гиперсфере, проходящая через P , должна также проходить и через Q , но пространственноподобные и световые геодезические уходят через P на бесконечность и не попадают в Q . Мы можем перейти к «развернутому» антидеситтеровскому пространству (универсальное накрывающее пространство). Для простоты выберем P и Q как можно ближе друг к другу, причем так, чтобы $P \ll Q$ (рис. 29). Пусть точка R отделена от Q пространственноподобной геодезической. Тогда легко убедиться, что *не существует геодезической, соединяющей P и R* [17]. (От P до R можно дойти за два шага, используя ломаную временноподобную геодезическую.) Антидеситтеровская модель геодезически полна, так что ситуация здесь сильно отличается от случая положительно определенных метрик. Отсутствие геодезической, соединяющей P и R , тесно связано с тем фактом, что R лежит вне $H_+[I_+(P)]$, что будет видно из разд. 11.

Рис. 30. Различные типы поведения моделей Фрийдмана во времени при $\lambda = 0$.



В космологических приложениях наиболее часто изучаются модели Фрийдмана с метрикой

$$ds^2 = dt^2 - R^2 d\Sigma^2. \quad (9.18)$$

Здесь $d\Sigma^2$ обозначает метрику единичной 3-сферы ($k = 1$), или евклидова 3-пространства ($k = 0$), или единичного гиперболического 3-пространства (Лобачевского) ($k = -1$). Величина $R = R(t)$ удовлетворяет уравнениям

$$\frac{4}{3} \pi \rho R^3 = M = \text{const} > 0, \quad (9.19)$$

$$\dot{R}^2 = \frac{KM}{2\pi R} + \frac{1}{3} \lambda R^2 - k, \quad (9.20)$$

которые соответствуют тензору энергии-импульса «пыли» (т. е. идеальной среды без давления):

$$T_{ab} = \rho(t) t_a t_b, \quad (9.21)$$

где

$$t_a = \nabla_a t. \quad (9.22)$$

Решения уравнения (9.20) при $\lambda = 0$ изображены на рис. 30 (их обращения во времени также являются решениями, но они не согласуются при $k \leq 0$ с наблюдаемым расширением вселенной). Отметим, что только в том случае, когда пространственные сечения обладают положительной кривизной (и, следовательно, они компактны), модель коллапсирует ко второй сингулярности $R = 0$. (Посредством отождествлений пространственные сечения можно сделать компактными и в случаях $k \leq 0$, однако это достигается ценой уменьшения глобальной изотропии.)

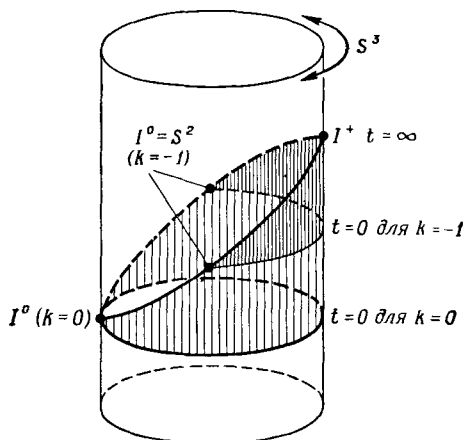


Рис. 31. Модели Фрийдмана с $k \leq 0$, $\lambda = 0$ как конформные части вселенной Эйнштейна.

Фрийдмановские модели также могут быть представлены как конформные подмножества статической вселенной Эйнштейна. Примем $\lambda = 0$. Случай $k > 0$ напоминает ситуацию для мира де Ситтера (рис. 26); отличие состоит лишь в том, что $\Omega = \infty$ на \mathcal{I}^\pm вместо $\Omega = 0$. Случаи $k \leq 0$ изображены на рис. 31. В обоих случаях \mathcal{I}^- пространственноподобно с $\Omega = \infty$, и \mathcal{I}^+ — световое с $\Omega = 0$. При $k = 0$ имеем $\bar{\nabla}_a \Omega = 0$ на \mathcal{I}^+ и I^0 является точкой. Но $\bar{\nabla}_a \Omega \neq 0$ при $k < 0$, и I^0 есть S^2 . Все фрийдмановские модели содержат ГГК и обладают горизонтами частицы для всех временно-подобных кривых.

Наконец, рассмотрим два примера, в которых отсутствуют ГГК, хотя и по разным причинам. Во-первых, рассмотрим волну с плоским фронтом, метрика которой [15, 28, 77]

$$ds^2 = 2\{du + H(v, x, y)dv\}dv - dx^2 - dy^2. \quad (9.23)$$

Область изменения действительных переменных u , v , x , y неограничена. Неисчезающие компоненты тензора кривизны для (9.23) определяются посредством

$$\frac{\partial^2 H}{\partial x^2}, \quad \frac{\partial^2 H}{\partial x \partial y}, \quad \frac{\partial^2 H}{\partial y^2}. \quad (9.24)$$

Метрика удовлетворяет уравнениям Эйнштейна в вакууме с $\lambda=0$ (и, таким образом, представляет собой чистую гравитационную волну), если

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)H = 0. \quad (9.25)$$

Но если условие (9.25) не налагается, то (9.23) описывает более общую ситуацию комбинированной гравитационно-электромагнитно-нейтринной волны. Все семь главных световых направлений при этом совпадают. Если H квадратично по x , y , то мы получаем частный случай *плоской волны*. Многие плоские волны геодезически полны [28, стр. 96].

Если $H = 0$ в некоторой области, то мы имеем там мир Минковского. Фактически мы можем выбрать H нулевым всюду кроме некоторого ограниченного интервала $v_0 < v < v_1$. Тогда мы получим волну типа «сэндвича» [10]. Экстремальным случаем является идеализированная ситуация, когда волна становится бесконечно малой по длительности (скажем, $v_0, v_1 \rightarrow 0$) и производит ненулевой результирующий эффект. Функция H становится дельта-функцией по v :

$$H(v, x, y) = \delta(v)h(x, y). \quad (9.26)$$

Подставляя (9.26) в (9.23), мы получаем метрику, которая не удовлетворяет условиям, обычно накладываемым на пространство-время с дельта-функцией в *кривизне* [как это получается в (9.24)], поскольку в данном случае компоненты *метрического* тензора включают дельта-функции. Можно преобразовать координаты так, чтобы новые компоненты метрического тензора были функциями класса C^0 от координат. Но приведенная форма удобней, так как ведет к конструкции многообразия, получаемой с помощью «ножниц и клея» [80]. Разделим мир Минковского с метрикой

$$ds^2 = 2 du dv - dx^2 - dy^2 \quad (9.27)$$

на $\mathcal{M}^- (v < 0)$ и $\mathcal{M}^+ (v > 0)$, выбрасывая световую гиперплоскость $v = 0$. Присоединим теперь границу $v = 0$ к каждой половине и отождествим две половины так, чтобы каждая из них выглядела «перевер-

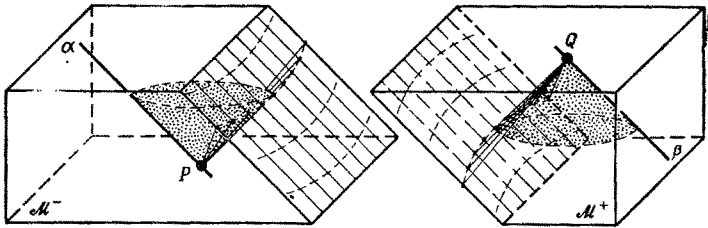


Рис. 32. Плоская волна с амплитудой в виде дельта-функции. Обе половины пространства плоские, но присоединены с перерыванием.

нутой» с точки зрения другой (рис. 32). Это достигается переходом

$$x \rightarrow x, \quad y \rightarrow y, \quad u \rightarrow u + h(x, y) \quad (9.28)$$

от \mathcal{M}^- к \mathcal{M}^+ на их общей границе $v = 0$. Скачок координат в (9.28) в точности соответствует (9.26), подставленному в (9.23).

Для простоты рассмотрим случай электромагнитной плоской волны. Тогда мы можем принять

$$h(x, y) = a(x^2 + y^2), \quad (9.29)$$

где $a = \text{const} > 0$. Световой конус точки $P \in \mathcal{M}^-$ с координатами $u = x = y = 0$, $v = -a^{-1}$ определяется уравнением

$$2u(v + a^{-1}) = x^2 + y^2. \quad (9.30)$$

Вследствие (9.28) и (9.29) это согласуется с

$$2u(v - a^{-1}) = x^2 + y^2 \quad (9.31)$$

при $v = 0$. Уравнение (9.31) дает световой конус точки $Q \in \mathcal{M}^+$ с координатами $u = x = y = 0$, $v = a^{-1}$. Это показывает, что световой конус точки P (в этом примере) опять фокусируется после прохождения через волну в точке Q . Однако имеется одна исключительная световая геодезическая α , проходящая через P , а именно $x = y = 0$, $v = -a^{-1}$, которая никогда не пересекает волну вообще. Соответственно, световая геодезическая β , проходящая через Q и определяемая как $x = y = 0$, $v = a^{-1}$, не переходит в другую половину пространства. Таким образом, две продолжимые световые геодезические α и β являются

предельной конфигурацией для последовательности *одиночных* световых геодезических, проходящих через P и Q . (Таким образом, пространство световых геодезических здесь не является хаусдорфовым.)

Пространственноподобная ГГК должна пересекать каждую световую геодезическую последовательности только один раз. Следовательно, она должна была бы встретить предельную пару α, β только в *одной* точке. Однако обе геодезические α, β не могут быть пересечены, как это требуется для ГГК. В терминах рис. 32 можно сказать, что любая бескраевая пространственноподобная гиперповерхность, проходящая через P , захватывается внизу светового конуса прошлого для точки Q и никогда не достигает β . Легко также установить отсутствие полупространственноподобной ГГК. (Рассмотренный пример можно конформно изобразить как часть статической вселенной Эйнштейна. В этом случае отсутствие ГГК вытекает не из пространственноподобности \mathcal{I} , а из того факта, что \mathcal{M} содержит световую геодезическую во *внутренней* области $\bar{\mathcal{M}}$.) В более общем случае конформных плоских волн, таких, как чисто гравитационные плоские волны, ГГК отсутствует в сущности по той же причине, что и в рассмотренном примере [77], но ситуация становится несколько сложнее из-за астигматизма при фокусировании.

В качестве последнего примера в этом разделе рассмотрим пустой мир Тауба — НУТ ¹⁾, описанный Мизнером [59, 61, 63, 105, 118*, 119*]. В данном случае пространство-время \mathcal{M} есть топологически $S^3 \times E^1$ (но конформно отличается от вселенной Эйнштейна). Имеется выделенное пространственноподобное сечение \mathcal{S} с максимальным собственным объемом, метрика которого есть метрика 3-сферы (рис. 33). Однако 4-мерная геометрия на \mathcal{S} не обладает полной вращательной симметрией S^3 . Тензор Вейля отличен от нуля, главные световые направления попарно совпадают {22} [ср. с (7.6)]. Ортогональное проектирова-

¹⁾ Мир НУТ — мир Ньюмана — Унти — Тамбурино. — *Прим. перев.*

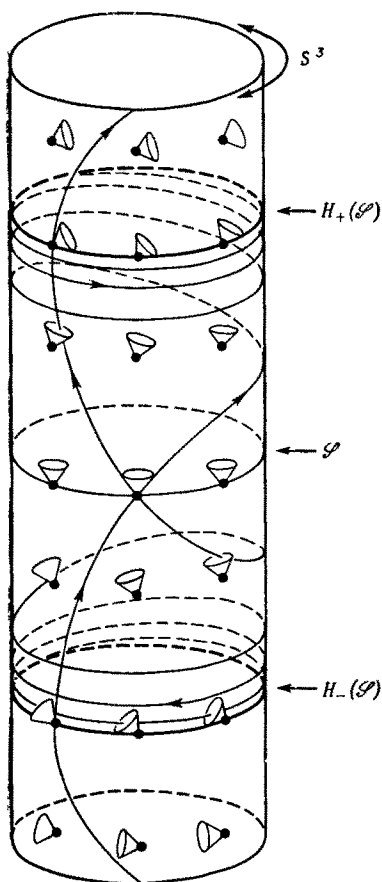


Рис. 33. Модель Тауба — НУТ.

ние на \mathcal{S} этих световых направлений дает нам нигде не исчезающее на \mathcal{S} поле линейного элемента. (В каждой точке \mathcal{S} два световых направления проектируются как диаметрально противоположные.) Эти линейные элементы тангенциальны к конгруэнции больших кругов на S^3 , которые образуют множество параллелей Клиффорда. Следовательно, эти круги образуют расслоение Хопфа S^3 . Остающаяся симметрия представляет собой транзитивную 4-параметрическую группу.

Будем передвигать наше сечение вверх параллельно самому себе на рис. 33, сохраняя эту симметрию. Мы получим похожие пространственноподобные сечения, но с меньшим объемом S^3 . Наконец, этот объем уменьшается до нуля [в соответствии с эффектом Райчаудури; ср. с (7.40) — (7.47)], но вместо сжатия в точку мы просто получим $H_+(\mathcal{S})$, которое является здесь компактной световой гиперповерхностью, порожденной световыми геодезическими, образующими множество параллелей Клиффорда на $H_+(\mathcal{S}) \approx S^3$.

Позади $H_+(\mathcal{S})$ сечения становятся временноподобными¹⁾ и поэтому содержат замкнутые временноподобные кривые, явно нарушая тем самым причинность.

Световые и временноподобные геодезические, направленные в будущее и берущие свое начало на \mathcal{S} , распадаются на два класса. Есть такие, которые попадают в часть \mathcal{M} , находящуюся за $H_+(\mathcal{S})$ [и, следовательно, пересекают $H_+(\mathcal{S})$ в определенной точке], и такие, которые навиваются спирально, асимптотически приближаясь к $H_+(\mathcal{S})$. Кривые второго класса имеют конечную аффинную длину в направлении будущего, так что модель не является геодезически полной. Мизнер показал, что довольно удивительным образом может быть реализовано альтернативное продолжение на $H_+(\mathcal{S})$ и за ним, отличное от изображенного на рис. 33. Кривые, которые прежде пересекали $H_+(\mathcal{S})$, теперь навиваются спирально, в то время как большинство геодезических, которые прежде навивались, теперь пересекают новое $H_+(\mathcal{S})$ и попадают в новое продолжение. Оба продолжения являются аналитическими, но могут быть осуществлены одновременно лишь ценой отказа от требования, чтобы \mathcal{M} было хаусдорфовым многообразием [59, 61, 118*]. Поведение модели на $H_-(\mathcal{S})$ в точности аналогично поведению на $H_+(\mathcal{S})$ (рис. 33). Модель оказывается чрезвычайно неустойчивой в окрестности $H_{\pm}(\mathcal{S})$ относительно малых возмущений в начальных условиях на \mathcal{S} . Однако поучительно то, что она проявляет специфические свойства, которые должны приниматься во внимание при изучении более общих ситуаций.

¹⁾ Эти сечения являются компактными лоренцовыми 3-многообразиями рассмотренного Авезом типа [3]. Всякая замкнутая временноподобная кривая γ этого 3-многообразия обладает любопытным свойством, а именно: хотя γ (что очевидно) не может быть непрерывно деформирована в точку, оставаясь временноподобной, для каждого сечения существует некоторое целое n , обладающее тем свойством, что кривая $n\gamma$ (т. е. γ , описанная умножением на n — для чего необходимо, чтобы γ была параметризована собственным параметром) может быть деформирована в $(n+1)\gamma$, оставаясь временноподобной. Следовательно, она может быть деформирована в кривую произвольно большой длины. Это является контрпримером результату Авеза [4]. Я благодарен Кундту и Зейферту, которые обратили на это мое внимание.

Гравитационный коллапс

Очень интересным в теории относительности и космологии является вопрос о существовании сингулярностей. Обычные модели «большого взрыва» в релятивистской космологии характеризуются наличием начальных состояний бесконечной кривизны и плотности (рис. 30). В такой начальной сингулярности неприменимо обычное представление о пространстве-времени как о гладком многообразии. Существование горизонта частиц для каждой мировой линии также приводит нас к серьезной проблеме. Разные, сильно искривленные части вселенной, сообщение между которыми было прежде невозможным, должны объединиться таким образом, чтобы их стыковка была непротиворечивой. Единственная предшествующая «причинная» связь между этими частями осуществлялась в самой сингулярности. В дополнение ко всем грандиозным, но по существу техническим проблемам, относящимся к физике ранних состояний высокой плотности, мы сталкиваемся с проблемой *первопричины*. Это вопрос о *начальных условиях* в существенно сингулярном состоянии [60, 119*].

Однако уместно спросить, не является ли начальная сингулярность в этих моделях скорее следствием использованных математических упрощений, чем действительной физической ситуацией, которую описывает модель. В частности, на практике бывает важно постулировать высокую степень симметрии (например, пространственную однородность) для возможности детального математического анализа модели. А при наличии такой симметрии почти неизбежен вывод о том, что в какой-то момент в прошлом вся материя во вселенной должна была быть одновременно

сжата в точку или по крайней мере в линию или в двумерную поверхность [6, 43, 47а, 99]. Возможные пути выхода из указанных затруднений состоят во введении достаточно большой космологической постоянной, в наличии «непонятной» материи (например, с отрицательной плотностью энергии: сюда относятся модели с непрерывным творением вещества в пространстве), в допущении достаточно большого вращения или в изменении законов физики, которые могли действовать на очень ранней стадии развития вселенной. Кроме того, можно принять во внимание квантовые эффекты гравитации, хотя они могли быть важны только при «нелепо» высоких плотностях (например, 10^{93} г/см³ или по крайней мере 10^{53} г/см³). Но гораздо более существенной представляется роль локальных неоднородностей материи и кривизны. Действительно, нет оснований ожидать, что вблизи сингулярности модель должна напоминать современный мир.

Возможно, ситуация будет более ясной для процесса, обращенного во времени. Представим себе, что вся материя во вселенной одновременно сжата в единственную «центральную точку» (или «линию» и т. д.). Тогда обращение плотности в бесконечность в этой точке (или линии, и т. д.) едва ли удивительно. Если же мы немного возмутим движение концентрирующегося вещества, так чтобы оно не фокусировалось строго в одну точку (линию и т. д.), то можно ожидать, что хотя плотность и становится очень высокой, но она не обращается в бесконечность¹⁾. Можно рассмотреть возможность того, что вселенная испытывает

¹⁾ В действительности может реализоваться иная ситуация, что особенно хорошо видно на примере пылевидной среды (уравнение состояния $p = 0$). В этом случае одновременная «фокусировка» (сингулярность) собираемого вещества действительно требует специального задания одной из произвольных функций, определяющих поле скоростей вещества. Однако отказ от этого специального выбора (произвольное возмущение первоначального поля скоростей) не устраняет сингулярность, а просто приводит к существенно неодновременной «фокусировке» (т. е. к такой «фокусировке», которая не может быть сделана одновременной никаким выбором системы отсчета) [130*]. — *Прим. перев.*

эффективный «отскок», т. е. наблюдаемое в настоящее время расширение вселенной можно было бы рассматривать как результат предшествующей фазы сжатия, причем переход от сжатия к расширению осуществляется через сконденсированную, чрезвычайно асимметричную и сложную промежуточную фазу.

Картина такого типа предлагалась многими авторами [54]. Детальное рассмотрение, проведенное Лифшицем и Халатниковым [53], также предполагает, что устранение сингулярности малыми возмущениями движения действительно может иметь место. Однако, как мы увидим в разд. 11, недавние исследования исключили большинство возможностей такого рода¹⁾. В настоящем разделе я остановлюсь на тесно связанном с этой проблемой явлении сингулярностей при гравитационном коллапсе. Существенное сходство между явлением гравитационного коллапса и обращенными во времени ранними стадиями космологических моделей типа «большого взрыва» особенно сильно подчеркивалось Уилером [113]. Есть, по-видимому, определенные преимущества в том, чтобы сосредоточить внимание на более локальном явлении коллапса, а не на глобальных космологических вопросах, поскольку для изучения локального явления нам не требуется знания всей крупномасштабной структуры вселенной (или требуется?). Если локальные законы физики сильно подвержены влиянию структуры вселенной как целого («принцип Маха»), мы по крайней мере могли бы ожидать, что локальное явление *менее чувствительно* к таким изменениям законов; и наконец, в отличие от вселенной как целого коллапс *не есть* нечто, данное нам только *однажды*, так что мы в состоянии эксплуатировать огромные теоретические возможности «мысленного эксперимента». По моему

¹⁾ Недавно Лифшиц, Халатников и Белинский [125*, 126*], а затем Мизнер [127*] открыли «колебательный» характер приближения к сингулярности в однородных космологических моделях IX типа Бианки. В дальнейших исследованиях Лифшиц, Халатников и Белинский [128*] приводят аргументы в пользу того, что решение будет иметь такой характер и в общем случае вблизи сингулярности. — *Прим. ред*

мнению, более глубокие аспекты космологической проблемы не могут быть решены без одновременного более глубокого понимания того, что представляет собой коллапс.

Но что же строго понимается под «гравитационным коллапсом?» Идея гравитационного коллапса берет свое начало из классического изучения больших, сферически симметричных тел. Согласно работе Чандрасекара [21, 39], холодное, сферически симметричное (невращающееся) тело, достигшее конца термоядерной активности, не может сопротивляться действию гравитационных сил, если его масса намного больше солнечной¹⁾. Поэтому такое тело будет катастрофически коллапсировать. Его коллапс может быть остановлен только каким-либо механизмом, способным выбросить достаточно большую массу, чтобы тело попало ниже чандрасекаровского предела. Согласно некоторым подробным вычислениям (для ссылок см. [107]), для начальных масс между 1, 2 и 15 масс Солнца высокая интенсивность нейтринного излучения из коллапсирующего ядра звезды может быть достаточной для того, чтобы сбросить внешние слои звезды и оставить центральную *нейтронную* звезду. Холодная нейтронная звезда (которая по существу является огромным атомным ядром, удерживаемым гравитационными, а не ядерными силами) должна быть несколько *менее* массивной, чем Солнце (предел Оппенгеймера — Волкова [67]), хотя и во много раз меньшей, с диаметром всего лишь около 10 км. Полагают, что гигантский взрыв, сопровождающий этот процесс, объясняет по крайней мере один тип сверхновых [107].

Нейтронная звезда может существовать неограниченно долго, если не происходит аккреции. Но что произойдет при больших начальных массах? Например, наблюдаются звезды с массами порядка 60 сол-

¹⁾ Проблемы коллапса и его астрофизическое значение подробно описаны в книгах Я. Б. Зельдовича и И. Д. Новикова «Релятивистская астрофизика», изд-во «Наука», 1967 и «Теория тяготения и эволюция звезд», изд-во «Наука», 1971. — *Прим. ред.*

нечных. Согласно упомянутым вычислениям, есть указания на то, что ядро остается *слишком* массивным, чтобы образовать нейтронную звезду, и поэтому продолжает коллапсировать. Если предполагать, что поддерживается *точная* сферическая симметрия, то ядро будет сжиматься, проходя *через* радиус Шварцшильда $r = 2m$, до тех пор пока не достигнет сингулярности пространства-времени в центре, $r = 0$. Остальное вещество звезды, не выброшенное слишком далеко, будет следовать за ядром. Однако наблюдатель, расположенный на больших расстояниях от звезды, никогда не увидит коллапс внутрь сферы $r = 2m$. Коллапс звезды будет казаться ему замедленным, а радиус звезды будет достигать значения $r = 2m$ асимптотически.

Прежде чем обсуждать физическую реалистичность этой картины, рассмотрим решение уравнений Эйнштейна, на котором она основана, а именно решение Шварцшильда. Оно определяет сферически симметричное гравитационное поле вне коллапсирующей звезды. В обычных шварцшильдовских координатах имеем

$$ds^2 = \left(1 - \frac{2m}{r}\right) dt^2 - \left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta d\varphi^2). \quad (10.1)$$

Пусть граница звезды определяется как $r = f(t)$. Тогда метрика (10.1) применима только вне этой границы, т. е. при $r \geq f(t)$ (с обычными неравенствами и отождествлениями для θ и φ). Граница звезды должна быть временноподобной: $|f'(t)| < 1 - 2m/r$, откуда следует, что она никогда не пересекает $r = 2m$ в (10.1) (рис. 34). На этом основании можно было бы подумать, что звезда всегда остается вне $r = 2m$. Но если мы вычислим полное собственное время частицы, находящейся на поверхности звезды, причем для простоты предположим, что звезда коллапсирует свободно, то мы обнаружим, что собственное время является *конечным*. Таким образом, наблюдатель, падающий вместе со звездой, по истечении этого

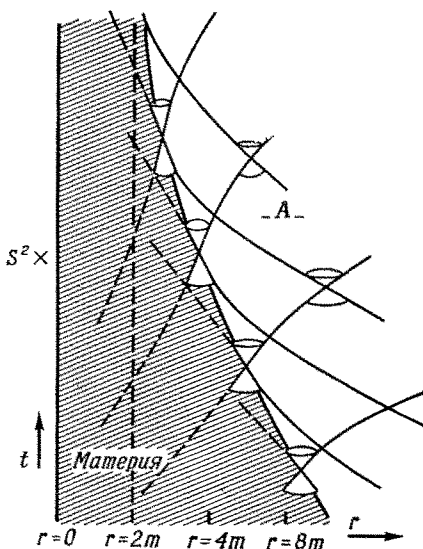
Рис. 34. Картина Шварцшильда.

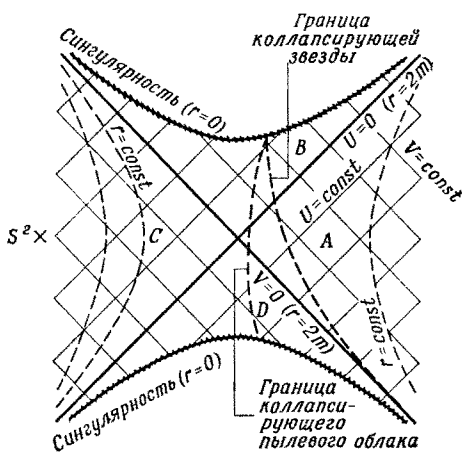
конечного времени должен быть либо уничтожен (например, вследствие бесконечных приливных сил или других форм сингулярности), либо он обнаружит, что находится в другой части вселенной, не покрываемой координатами (10.1).

В данном случае реализуется вторая возможность. В этом можно убедиться с помощью преобразования координат в (10.1), если ввести опережающий временной параметр $v = t + r + 2m \ln(r - 2m)$. Тогда метрика принимает форму Эддингтона — Финкельштейна [27, 32] [ср. с (8.14)]:

$$ds^2 = (1 - 2m/r) dv^2 - 2 dv dr - r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (10.2)$$

При $r > 2m$ формы (10.1) и (10.2) в точности эквивалентны, но (10.2) обладает тем преимуществом, что метрика покрывает большую область, причем без всяких особенностей. В окрестности (световой) гиперповерхности $r = 2m$ метрика является совершенно регулярной. Эту гиперповерхность иногда называют сингулярностью Шварцшильда только потому, что более привычная форма (10.1) обладает координатной сингулярностью на $r = 2m$. Поверхность звезды пересекает $r = 2m$ при определенном значении v (хотя, конечно, неподходящий параметр t становится здесь бесконечным) и движется дальше в направлении $r = 0$.





Р и с. 36. Картина Крускала.

продолжить решение дальше. Если вычислить скаляры кривизны, построенные из тензора Вейля, то мы найдем, что они стремятся к бесконечности при $r \rightarrow 0$. Таким образом наш наблюдатель, успешно сопровождавший звезду через $r = 2m$, теперь должен быть разорван бесконечными приливными силами в истинной сингулярности $r = 0$.

Эта схема была нарисована Оппенгеймером и Снайдером [68], когда они рассматривали динамику коллапсирующего однородного облака пылевых частиц. Пространство-время *внутри* такого пылевого облака также может быть точно и просто описано и оказывается не чем иным, как (частью) вселенной Фридмана (разд. 9). Это опять подчеркивает тесную связь между конечной сингулярностью при гравитационном коллапсе и конечной (или начальной) сингулярностью в релятивистской космологии.

Если мы хотим рассмотреть всю историю гравитационно связанного газового облака Оппенгеймера—Снайдера, а не только заключительные стадии гравитационного коллапса, то для этой цели даже метрическая форма (10.2) не годится, так как она не покрывает

вает всей пустой области. Поведение пылевого облака симметрично во времени. Так же как и вселенная Фридмана при $k = 1$ ($\lambda = 0$), облако расширяется из начальной сингулярности до максимального объема и затем сжимается до конечной сингулярности. Таким образом, следует ожидать, что возможна сшивка области, заполненной материей, с наружной пустой сферически-симметричной областью, которая также симметрична во времени. Мы можем изобразить эту область, используя U, V -координаты Крускала [50а], которые связаны с координатами Шварцшильда r, t (в области $r > 2m$) посредством соотношений

$$\frac{V}{U} = -e^{t/2m}, \quad VU = e^{r/2m} \left(1 - \frac{r}{2m}\right). \quad (10.3)$$

Тогда метрика принимает вид

$$ds^2 = f^2 dU dV - r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (10.4)$$

где

$$f^2 = \left(\frac{32m^3}{r}\right) e^{-r/2m}. \quad (10.5)$$

Область, занятая материей, соединяется с областью Крускала вдоль временноподобной гиперповерхности. Эта гиперповерхность изображается временноподобной геодезической на (U, V) -плоскости Крускала (рис. 36). Другими словами, в этой модели рассматривается только часть диаграммы, лежащая вправо от этой геодезической. Левая часть рис. 36 должна быть заменена сферически симметричной частью ($\simeq D^3 \times E^1$) вселенной Фридмана (9.18), граница которой является опять-таки гиперповерхностью ($\simeq S^2 \times E^1$), образованной временноподобными геодезическими. Эти две части сшиваются достаточно гладко (ср. с [107]).

Один недостаток формы Крускала метрики Шварцшильда состоит в том, что (10.4) не дано «точно», а зависит от решения уравнения (10.3) для r . Форма Эддингтона — Финкельштейна (10.2) значительно

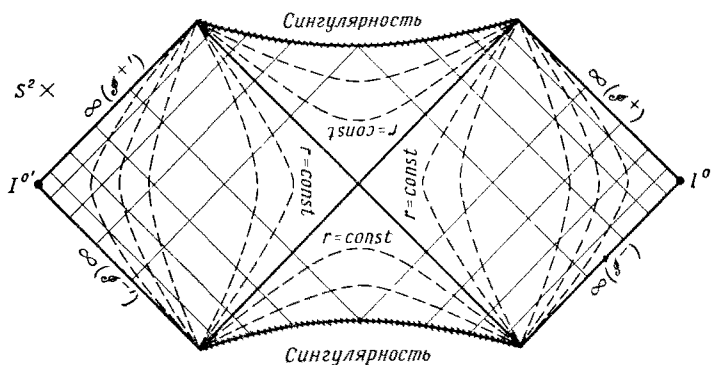


Рис. 37. Картина Крускала с кочформной бесконечностью.

проще и интересна тем, что из нее можно легко получить главные черты крускаловского продолжения. Чтобы достичь этого, заметим, что (10.2) покрывает две части A, B диаграммы Крускала (рис. 36) (с $2mV^2 = = e^{v/2m}$). Используя запаздывающий параметр u [ср. с (8.11)] вместо опережающего параметра v , мы можем аналогичным образом покрыть части A, D рис. 36 другой «координатной картой», причем A будет областью перекрывания. Совершенно аналогичным способом можно покрыть части C, B или C, D «координатной картой», имеющей форму (10.2). При желании в областях перекрывания B и D можно получить формулы преобразования. Для этого надо вернуться к форме (10.1) (при $0 < r < 2m$) как к промежуточной стадии. Этим методом нельзя получить одно свойство крускаловского продолжения, а именно регулярность пространства-времени на 2-сфере $U = V = 0$. Однако метод имеет большую эвристическую ценность в применении к другим, более сложным продолжениям, из которых наиболее известны решения Керра и Рейснера — Нордстрема [ср. с (10.18), (10.20)] [12, 18, 37]. В этой связи полезно также иметь вариант диаграммы Крускала, в котором представлена конформная бесконечность (например, с помощью координат $p = \text{arctg} [\text{Arsh } V]$, $q = \text{arctg} [\text{Arsh } U]$ (рис. 37).

Теперь вернемся к случаю коллапсирующей звезды. Может ли вообще описание, которое я дал, считаться физически реалистичным? Например, можно ли даже при сохранении точной сферической симметрии сказать, что мы знаем достаточно много о свойствах материи при плотностях (для рассмотренных случаев — несколько больше ядерной), при которых звезда, как предполагается, падает сквозь ее «шварцшильдовскую горловину»? Не может ли оказаться, что по некоторым причинам массивная звезда неизбежно сбрасывает достаточно много вещества при приближении к $r = 2m$, так что ее масса неизменно уменьшается ниже предела Чандрасекара или Оппенгеймера — Волкова, создавая тем самым условия для устойчивого, самоподдерживающегося состояния? Повидимому, в течение многих лет преобладала точка зрения, согласно которой звезда всегда в состоянии предохранить себя от коллапса каким-нибудь способом. Возможно, это происходило потому, что астрономы не чувствовали необходимости рассматривать самоподдерживающиеся тела с массами, во много раз превосходящими солнечную. Но открытие квазизвездных объектов (см., например, [90]) стимулировало возобновление интереса к гравитационному коллапсу.

Дело в том, что фантастическое количество энергии, излучаемое объектами, и их очень малые размеры предполагают, что массы отдельных объектов достигают порядка 10^6 или 10^8 солнечных масс. Далее, шварцшильдовский радиус сферически симметричного тела пропорционален его массе. Таким образом, характерная плотность, при которой тело пересекает $r = 2m$, должна быть обратно пропорциональна квадрату его массы. Для объекта в 10^6 — 10^8 солнечных масс эта плотность не является чрезмерно высокой. В самом деле, Фаулер подчеркивал, что для объекта в 10^{11} солнечных масс (масса галактики средних размеров) эта характерная плотность меньше плотности воздуха! И нет никаких оснований думать, что такой объект обязательно выбросит практически все вещество перед достижением $r = 2m$.

Далее, что можно сказать о роли отклонений от сферической симметрии? Например, можно ли *определить* аналог $r = 2m$ для асимметричного тела? И не может ли случиться, что наличие *вращения* всегда предотвращает наступление конечных стадий коллапса? Вопрос о вращении особенно важен. Например, если присутствует дифференциальное вращение, то существуют равновесные состояния для объектов с массами, значительно превосходящими чандрасекаровский предел [69]. Однако при наличии вязкости возможному конечному состоянию может соответствовать только однородное вращение. Устойчивость таких тел является решающим соображением [21a], и оказывается, что при определенных обстоятельствах вращающееся тело может коллапсировать. К счастью, известно решение уравнений Эйнштейна в вакууме, а именно решение Керра [12, 48], обобщающее решение Шварцшильда путем включения *углового момента*. Решение Керра содержит два произвольных параметра m и a , где m определяет массу, как и в решении Шварцшильда, а ma определяет угловой момент. Решение [приведенное точно в (10.18)] носит довольно специальный характер (например, квадрупольный момент должен быть равен $2ma^2$), но кажется вероятным, что по крайней мере главные черты геометрии пространства-времени, окружающего вращающееся коллапсирующее тело, можно получить из анализа этого решения. Одно из наиболее значительных свойств этого решения состоит в том, что при $a > m$ нет аналога «шварцшильдовской горловины», т. е. для внешних наблюдателей нет горизонта событий, который мог бы помешать получению сигналов из внутренних областей. С другой стороны, при $a < m$ такой горизонт *существует*, и решение качественно напоминает решение Шварцшильда в том отношении, что оно обладает похожей «горловиной», которая в состоянии в конце концов «проглотить» весь материал (сверх-)звезды. Есть, однако, и другие свойства решения, качественно отличающиеся от случая решения Шварцшильда. Некоторые из них будут ниже обсуждены, а пока я остановлюсь на вопросе «горловины» самой

по себе и дам определение, которое можно будет использовать как критерий ее существования.

Вернемся к решению Шварцшильда. Для удобства рассмотрим метрику в форме (10.2) и исследуем рис. 35. Мы хотим выделить свойство области V , которое характеризовало бы ее как «особую» в некотором смысле. Надо подчеркнуть, однако, что область V не имеет ничего особенного *локально*. Малая окрестность любой точки в области V является таким же хорошим решением вакуумных уравнений Эйнштейна, как и любая другая, и в ней нет ничего «сингулярного». Но создается впечатление, что вся область V с неизбежностью «схлопывается» некоторым образом к сингулярности пространства-времени в $r = 0$. Чтобы описать это схлопывание, нам надо подобрать частично *глобальный* аспект области V , но он не должен быть слишком глобальным, поскольку нам требуется свойство, которое не будет разрушаться малыми возмущениями метрики во «время» коллапса.

Выберем точку T в области V на рис. 35, которая лежит внутри $r = 2m$ (но вне области, занятой материей). Такая точка изображает пространственноподобную 2-сферу \mathcal{F} с площадью поверхности $4\pi r^2$ ($r < 2m$). Всякая система материальных частиц, мировые линии которых пересекают \mathcal{F} , должна двигаться дальше со скоростями, ограниченными скоростью света. Другими словами, если мы рассмотрим предельную ситуацию, когда вспышка света излучается на \mathcal{F} , то будут «уходящий» и «входящий» сигналы, изображаемые на рис. 35 двумя отрезками световых геодезических с начальной точкой в \mathcal{F} . Эти отрезки описывают световые гиперповерхности, проходящие через \mathcal{F} , которые образуют $\dot{I}_+[\mathcal{F}]$. Существенно то обстоятельство, что *обе* части границы «сжимаются». Уменьшается не только площадь поверхности сферы, которая изображается точкой, движущейся по «входящей» световой геодезической, но уменьшается также площадь поверхности сферы, соответствующей «уходящему» сигналу. Именно это свойство поверхности \mathcal{F} я буду использовать, чтобы характеризовать ее как *ловушечную поверхность* [75].

Наличие ловушечной поверхности в области будет служить для нас указанием того, что в последующей истории этой области может случиться нечто «особенное». Точнее:

Ловушечная поверхность \mathcal{F} — это гладкая, компактная, пространственноподобная 2-поверхность, обладающая тем свойством, что световые геодезические, которые пересекают \mathcal{F} ортогонально, локально сходятся в направлении будущего¹⁾. (10.6)

Эту сходимость удобно измерять величиной ρ из разд. 7. Световые геодезические, ортогонально пересекающие пространственноподобную 2-поверхность, всегда образуют световую гиперповерхность, так что по (7.28) мы имеем $\rho = \bar{\rho}$. Характерной чертой ловушечной поверхности является то, что

$$\rho > 0 \text{ на } \mathcal{F} \quad (10.7)$$

для обеих световых гиперповерхностей, проходящих через \mathcal{F} . По мере продвижения в будущее по каждой из этих гиперповерхностей площадь поверхности вначале уменьшается в каждой точке \mathcal{F} .

Теперь идея состоит в том, чтобы установить связь между существованием в пространстве-времени ловушечной поверхности и последующим развитием либо сингулярности [такой, как $r = 0$ в (10.2)], которая может включать бесконечную кривизну, либо каких-нибудь других свойств, весьма необычных с физической точки зрения. Прежде чем рассматривать строгие математические результаты, выясним, имеются ли вообще основания ожидать, что в реальной вселенной может образоваться ловушечная поверхность. В каждом конкретном случае коллапсирующей звезды или сверхзвезды это, очевидно, вопрос детального астрофизического рассмотрения. Скорость вращения, потеря массы вследствие гравитационного излучения и ухода нейтрино, точный характер асимметрии, тип уравнения состояния, магнитные поля и т. д. — все это

¹⁾ Удобно также предполагать, что \mathcal{F} есть полупространственноподобное множество, однако это не строго необходимо для доказательства теоремы 1.

может иметь отношение к делу. Но существенное обстоятельство, которое я хочу подчеркнуть, состоит в том, что не может быть *принципиальных* возражений против развития ловушечной поверхности. Чтобы проиллюстрировать это, рассмотрим следующий *мысленный эксперимент*.

Технически развитые (но, по-видимому, безрассудно храбрые) существа населяют галактику (предпочтительно эллиптическую), содержащую около 10^{11} звезд. С помощью ракет эти существа ухитрились изменить скорости звезд, причем так, что практически не осталось тангенциальной компоненты. Более того, радиальные скорости были подобраны так, что все звезды падают в направлении центра и должны достичь его окрестности почти в одно и то же время. (Масса-энергия, израсходованная в этой операции, была бы мала по сравнению с массой звезд.) Размер области, в которой должны собраться звезды, приблизительно в 50 раз больше размера солнечной системы. В таком объеме достаточно места для всех звезд, и они могут там собраться раньше, чем возникнет опасность столкновений. (Если угодно, они могут двигаться так, чтобы избежать столкновений вообще!) Но приведет ли это к развитию ловушечной поверхности? Неизбежность этого доказывается простыми рассуждениями. Рассмотрим вспышку (идеализированного) света, излученного в центре примерно в тот момент времени, когда звезды достигают границы критической области. Лучи, первоначально расходившиеся из центра, будут проходить вблизи достаточно большого числа звезд и вследствие фокусирования идеализированных световых лучей [7, 87] (т. е. световых геодезических) гравитационным полем каждой звезды (этот эффект является *наблюдаемым* эффектом в общей теории относительности) начнут сходиться опять. Теперь можно показать, что топологическая сфера, лежащая внутри наиболее удаленной границы световых лучей, будет ловушечной поверхностью.

Этот пример показывает также, что нет причин, по которым наблюдатель должен быть «уничтожен» после

того, как он попал в критическую область. В окрестности ловушечной поверхности кривизна все еще чрезвычайно мала, а пространство-время совершенно регулярно¹⁾. Действительно, локально нет ничего peculiarного в самой ловушечной поверхности. Даже в мире Минковского существует много поверхностей, которые являются «локально ловушечными». Другими словами, они удовлетворяют условиям (10.6), за исключением требования компактности. Чтобы построить такую поверхность, достаточно рассмотреть пересечение световых конусов прошлого двух пространственно разделенных точек.

Но каковы следствия существования ловушечных поверхностей? Я покажу, что если мы (времененно) предположим, что вселенная является *открытой* (т. е. пространственно некомпактной), то наших знаний законов физики недостаточно, чтобы вычислить, хотя бы *в принципе*, будущую историю коллапсирующей системы.

Чтобы предсказать будущее системы, мы обычно требуем, чтобы на пространственноподобной гиперповерхности были определены соответствующие данные Коши. Мы можем также принять, что локальные законы физики таковы, что сигналы распространяются вдоль временноподобных или световых кривых, но не вдоль пространственноподобных. Таким образом, если данные Коши известны на некотором пространственноподобном замкнутом множестве \mathcal{S} , то они могут определять поведение системы только в $D_+(\mathcal{S})$ и $D_-(\mathcal{S})$. Во внешних же областях информация может распространяться вдоль непространственноподобных кривых и не «регистрироваться» на \mathcal{S} . Теперь можно устроить состояние коллапса, исходя из совершенно разумных данных на пространственноподобной гиперповерхности \mathcal{S} , которые (поскольку мы рассматриваем локальную физическую ситуацию, а не космологическую) можно считать *эффективно* продолженными до бесконечности, безотносительно к тому, является

¹⁾ В [26] рассматривается слабо несферический коллапс и подтверждается вывод об образовании ловушечной поверхности.

ли вселенная действительно пространственно некомпактной. Таким образом, для *предсказуемости* мы требуем существования некомпактной пространственноподобной гиперповерхности \mathcal{C} , которая является глобальной гиперповерхностью Коши для \mathcal{M} , или точнее, для которой $D_+(\mathcal{C}) = I_+(\mathcal{C})$. Как мы увидим позже, вопрос о том, действительно ли космологическая природа пространства-времени подходит для локального процесса коллапса, отнюдь не тривиален.

Теорема I. Следующие требования к пространству-времени ¹⁾ \mathcal{M} несовместимы:

- (Ia) Существует некомпактная пространственноподобная гиперповерхность \mathcal{C} , для которой $D_+(\mathcal{C}) = I_+(\mathcal{C})$.
- (Iб) Существует ловушечная поверхность $\mathcal{T} \subset I_+(\mathcal{C})$.
- (Iв) $R_{ab}l^a l^b \leq 0$ для каждого светового вектора l^a .
- (Iг) \mathcal{M} полно относительно световых геодезических, направленных в будущее.

Требование (Iг) означает, что всякая световая геодезическая в \mathcal{M} может быть продолжена в будущее до сколь угодно больших значений (заданного) аффинного параметра. Физически мы можем представлять себе это как утверждение: «фотоны (или нейтрино, или гравитоны) не могут исчезать». Это представляется весьма разумным физическим требованием на пространство-время. В разд. II мы должны будем рассмотреть иной тип полноты, а именно полноту относительно временноподобных геодезических («частицы, находящиеся в инерциальном движении, не могут исчезать»). Возможны также и другие типы, такие, как «полнота относительно кривых с ограниченным ускорением». Это означает, что все продолжимые

¹⁾ Рассматриваемые здесь миры предполагаются ориентируемыми во времени (разд. 5). Однако все результаты (например, теоремы I, II, III) будут также применимы в соответствующей форме и без этого предположения. Для всякого неориентируемого во времени лоренцовского многообразия всегда существует дважды накрывающее многообразие, которое является ориентируемым во времени. (Рассмотрите пространство всех световых полуконусов на многообразии. Это дает искомое двойное накрывающее; ср. с [55].)

временноподобные кривые с ограниченной кривизной имеют бесконечную полную длину¹⁾ (т. е. «частицы, испытывающие ограниченные силы, не могут исчезать»). Все эти типы полноты *неэквивалентны*, что можно видеть на специально подобранных примерах [35, 52].

Однако, по-видимому, нет большой разницы в том, какое из определений применять. Возможно, определения неэквивалентны только в присутствии областей бесконечно большой кривизны. В любом случае доказательство существования таких областей приближает нас к цели — теоремам I, II, III. К сожалению, еще не установлены результаты, которые позволили бы *непосредственно* предсказать существование таких областей на основании «разумных» или «общих» физических предположений, хотя теореме I и особенно теореме II и III можно трактовать как сильное косвенное указание на это [118*].

Условие $R_{ab}l^a l^b \leq 0$ есть просто (7.35) и является следствием уравнений Эйнштейна (7.1) (с λ -членом или без него) и положительности плотности энергии (7.34). (В «собственной тетраде» тензора T_{ab} положительная плотность означает, что $T_{00} \geq 0$, $T_{00} + T_{11} \geq 0$, $T_{00} + T_{22} \geq 0$, $T_{00} + T_{33} \geq 0$.) Это неравенство является необходимым с физической точки зрения, особенно потому, что его нарушение, по-видимому, приводит к серьезной проблеме, связанной с квантовой теорией поля, а именно к явной катастрофической неустойчивости вакуума. Но этот вопрос не вполне ясен.

Прежде чем приступать к доказательству теоремы I, будет полезно установить некоторые леммы относительно структуры полупространственноподобной границы. Напомним [ср. с (9.4)–(9.6)], что любая ППГ является подмножеством \mathcal{P} множества \mathcal{M} вида

$$\mathcal{P} = i_+[\mathcal{K}] \text{ или, что эквивалентно, } \mathcal{P} = i_-[\mathcal{L}], \quad (10.8)$$

где \mathcal{K} , $\mathcal{L} \subset \mathcal{M}$. Разделим \mathcal{P} на четыре части (некоторые из них могут быть пустыми): \mathcal{P}_N , \mathcal{P}_+ , \mathcal{P}_- и \mathcal{P}_0 ,

¹⁾ Термин «длина» будет часто использоваться здесь для обозначения «собственного времени» на временноподобной кривой.

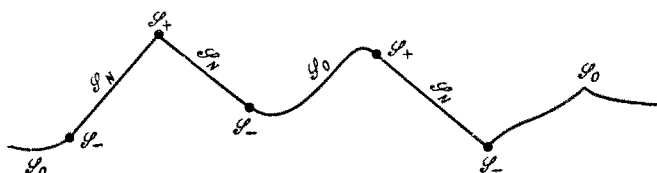


Рис. 38. Разные части полупространственноподобной границы. (Световые направления наклонены под углом 45° .)

определенные следующим образом. Если $X \in \mathcal{P}$, то могут существовать или не существовать точки $Y, Z \in \mathcal{P}$, отличные от X , для которых

$$I_+(X) \subset I_+(Y), \quad I_-(X) \subset I_-(Z). \quad (10.9)$$

Разные возможности определяют подмножества \mathcal{P}_N , \mathcal{P}_+ , \mathcal{P}_- , \mathcal{P}_0 согласно схеме

$$x \in \begin{array}{|c|c|} \hline \begin{array}{c} (\exists Z) \\ \mathcal{P}_N \end{array} & \begin{array}{c} (\exists Z) \\ \mathcal{P}_+ \end{array} \\ \hline \mathcal{P}_- & \mathcal{P}_0 \\ \hline \end{array} \begin{array}{l} (\exists Y) \\ (\exists Y) \end{array} \quad (10.10)$$

Интуитивный смысл этих подмножеств состоит в том, что \mathcal{P}_N представляет собой световую часть \mathcal{P} , причем \mathcal{P}_+ и \mathcal{P}_- — образованы соответственно конечными и начальными токами \mathcal{P}_N , а \mathcal{P}_0 представляет собой пространственноподобную часть \mathcal{P} (рис. 38). Более точное утверждение содержится в следующей лемме, которая дает также несколько более удобное условие того, что точка лежит в \mathcal{P}_N , \mathcal{P}_+ или \mathcal{P}_- .

Лемма I. Пусть $X \in \mathcal{P}$, где \mathcal{P} есть ППГ, определенная по (10.8). Пусть \mathcal{A} — открытое множество, содержащее X . Тогда условие

а) $I_+(X) \subset I_+[\mathcal{X} - \mathcal{A}]$ влечет за собой $X \in \mathcal{P}_N \cup \mathcal{P}_+$,

б) $I_-(X) \subset I_-[\mathcal{P} - \mathcal{A}]$ влечет за собой $X \in \mathcal{P}_N \cup \mathcal{P}_-$.

Кроме того, отрезок световой геодезической на \mathcal{P}

в) проходит через X , если $X \in \mathcal{P}_N$,

г) имеет X в качестве конечной точки, если $X \in \mathcal{P}_+$,

д) имеет X в качестве начальной точки, если $X \in \mathcal{P}_-$.

Чтобы доказать лемму I, предположим, во-первых, что $I_+(X) \subset I_+[\mathcal{K} - \mathcal{L}]$. Пусть \mathcal{B} — малый шар в нормальных координатах с центром в X и границей \mathcal{B} , где $\bar{\mathcal{B}} \subset \mathcal{L}$. Рассмотрим последовательность (X_n) точек с $X_n \in I_+(X) \cap \mathcal{B}$, которая сходится к X . Для каждого X_n существует $W_n \in \mathcal{K} - \mathcal{L}$ с $W_n \ll X_n$. В \mathcal{B} содержится связная часть световой кривой, соединяющей W_n и X_n с начальной точкой $Y_n \in \mathcal{B}$ и конечной точкой X_n . Так как \mathcal{B} компактно, существует точка Y на \mathcal{B} , такая, что последовательность (Y_n) сходится к Y . В нормальной координатной системе Y и X соединяются геодезической k , проходящей через X . Легко видеть, что k не может быть пространственно-подобной (для малого \mathcal{B}). Она не может быть и временноподобной, так как тогда мы имели бы $Y \ll X$, так что $W_n \ll Y_n \ll X$ для некоторого n , что противоречит $X \notin I_+[\mathcal{K}]$. Итак, k является световой геодезической. Следовательно, $I_+(X) \subset I_+(Y)$. Далее, $Y_n \in I_+[\mathcal{K}]$ и поэтому $Y \in \bar{I}_+[\mathcal{K}]$. Так как $X \notin I_+[\mathcal{K}]$, то $Y \notin I_+[\mathcal{K}]$ и $Y \in \mathcal{P}$. Это доказывает (а), (б) и половину (в). Доказательство (г), (д) и другой половины (в) совершенно аналогично. [Совпадение двух геодезических, полученных для (в), следует из того, что \mathcal{P} полупространственноподобно.]

Два отрезка световой геодезической на \mathcal{P} могут иметь общую точку X только в том случае, когда $X \in \mathcal{P}_+$ или $X \in \mathcal{P}_-$. Этот вывод верен также в случае сколь угодно близких отрезков световой геодезической на \mathcal{P} . Следующий результат делает это утверждение более точным.

Лемма II. Пусть \mathcal{P} есть ППГ, а k — отрезок световой геодезической, лежащий на \mathcal{P} , с конечной (соответственно — начальной) точкой P . Предположим, что некоторое открытое множество \mathcal{L} , содержащее $k - P$, пересекает \mathcal{P} по гладкой световой гиперповерхности \mathcal{N} , для которой сходимость ρ неограниченна вблизи P (мы выбираем касательный вектор l^a к k гладким в P). Тогда $P \in \mathcal{P}_+$ (соответственно \mathcal{P}_-).

Доказательство леммы II зависит от другого результата (изложенного, впрочем, более подробно, чем необходимо).

Лемма III. Пусть k — отрезок световой геодезической в \mathcal{M} и пусть \mathcal{N}_1 и \mathcal{N}_2 — гладкие световые гиперповерхности, проходящие через k , для которых $\mathcal{N}_1 \subset I_+[\mathcal{N}_2] \cup \mathcal{N}_2$. Тогда $\rho_2 - \rho_1 \geq |\sigma_2 - \sigma_1|$ на k (ρ_i, σ_i относятся к \mathcal{N}_i ; l^a одно и то же для каждого \mathcal{N}_i на k ; $i = 1, 2$).

Для доказательства леммы III¹⁾ рассмотрим две гладкие скалярные функции u_i ($i = 1, 2$) на \mathcal{M} , где $u_i = 0$ определяет \mathcal{N}_i (в некоторой окрестности k), причем u_i увеличивается в направлении будущего с неисчезающим градиентом на \mathcal{N}_i . Мы можем выбрать u_i так, чтобы $\nabla_a u_1 = \nabla_a u_2 = l_a$ на k . В достаточно малой окрестности k мы имеем $u_2 \geq u_1$. Следовательно, $X^a \nabla_a (u_2 - u_1) + \frac{1}{2} X^a X^b \nabla_a \nabla_b (u_2 - u_1) \geq 0$ на k для каждого множества компонент (X^a) с достаточно малым $|X^a|$. Пусть $\nabla_a u_i = l_{(i)a}$, так что $l_{(1)a} = l_{(2)a} = l_a$ на k . Выберем m_a на k комплексным, световым и ортогональным к l_a , причем $m_a \bar{m}^a = -1$. Полагая $X^a = \lambda m^a + \bar{\lambda} \bar{m}^a$, получим $\lambda^2 (\sigma_2 - \sigma_1) + 2\lambda \bar{\lambda} (\rho_2 - \rho_1) + \bar{\lambda}^2 (\bar{\sigma}_2 - \bar{\sigma}_1) \geq 0$ [ср. с (7.33)]. Это должно выполняться при любых λ и, следовательно, $\rho_2 - \rho_1 \geq |\sigma_2 - \sigma_1|$.

Для доказательства леммы II предположим, что $P \in \mathcal{S}_N$ (где P — конечная точка k). Тогда k можно продолжить в будущее на \mathcal{S} до точки $R \in \mathcal{S}$. Световой конус прошлого $I_-(R)$ несингулярен в некотором открытом множестве \mathcal{G} , где $k - P \subset \mathcal{G} \subset \mathcal{M}$ (для R , достаточно близкого к P). Теперь применяем лемму III, где $\mathcal{N}_2 = \mathcal{G} \cap I_-(R)$ и $\mathcal{N}_1 = \mathcal{G} \cap \mathcal{N}$. Но если $\rho = \rho_1$ неограниченно вблизи P (и положительно, так

¹⁾ *Примечание к русскому изданию.* Изложенное здесь доказательство не вполне строго. Исправленный вариант доказательства будет опубликован в [132*].

как Φ не является неограниченным в P), то $\rho_2 - \rho_1$ должно стать отрицательным вблизи P , нарушая неравенство из леммы III. Таким образом, лемма II доказана.

Из леммы I сразу вытекает следствие: если \mathcal{T} есть некоторое замкнутое подмножество \mathcal{M} и

$$\mathcal{P} = \dot{I}_+[\mathcal{T}], \quad (10.11)$$

то

$$\mathcal{P}_- \cup \mathcal{P}_0 \subset \mathcal{T}. \quad (10.12)$$

Это следует из части (а) леммы I, поскольку при $X \in \mathcal{P} - \mathcal{T}$ некоторая окрестность \mathcal{L} точки X не будет пересекать \mathcal{T} , так что $\mathcal{K} - \mathcal{L} = \mathcal{K}$, где $\mathcal{K} = \mathcal{T}$. Из (10.12) и частей (в) и (г) следует, что каждая точка множества $\mathcal{P} - \mathcal{T}$ является конечной точкой отрезка световой геодезической на \mathcal{P} . В частности, этот результат применим к $\dot{I}_+(P)$ или к любому горизонту частицы. Этот результат можно также обратить во времени, так что при $\mathcal{P} = \dot{I}_-[\mathcal{T}]$ каждая точка множества $\mathcal{P} - \mathcal{T}$ является начальной точкой отрезка световой геодезической на \mathcal{P} . Это применимо, в частности, к $\dot{I}_-(P)$ или к любому горизонту событий.

Другое следствие леммы I относится к структуре горизонтов Коши. Этот результат не потребуется при доказательстве теоремы I, но будет использован в следующем разделе. Есть два простых пути для получения горизонта Коши как части ППГ. Рассмотрим первый из них. Пусть \mathcal{H} замкнуто и полупространственноподобно. Тогда множество $\text{int} \{D_-(\mathcal{H}) \cup I_+[\mathcal{H}]\}$ является объединением $I_+(P)$, где $P \in D_-[\mathcal{H}]$. Граница этого множества есть ППГ, содержащая горизонт Коши $H_-(\mathcal{H})$ в качестве части [т. е. как пересечение ППГ с $D_-(\mathcal{H})$]. Второй путь можно описать следующим образом. Пусть, как и прежде, \mathcal{H} замкнуто и полупространственноподобно. Определим [42]

$$W_-(\mathcal{H}) = I_-(\mathcal{H}) - D_-(\mathcal{H}). \quad (10.13)$$

[$W_+(\mathcal{H})$ определяется аналогично.] Тогда $P \in W_-(\mathcal{H})$ тогда и только тогда, когда P является одновременно начальной точкой для продолжимой в будущее вре

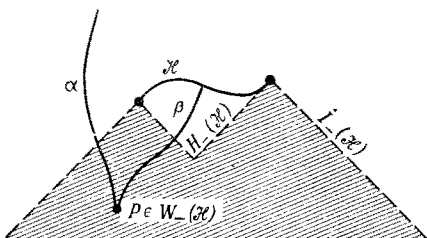


Рис. 39. Полупространственноподобная граница, содержащая горизонт Коши прошлого для \mathcal{H} .

менноподобной кривой α , не пересекающей \mathcal{H} , и начальной точкой для временноподобной кривой β , имеющей конечную точку на \mathcal{H} (рис. 39). Ясно, что $W_-(\mathcal{H})$ является открытым. А также, если $Q \ll P$ и $P \in W_-(\mathcal{H})$, то $Q \in W_-(\mathcal{H})$, так как \mathcal{H} полупространственноподобно. Следовательно,

$$W_-(\mathcal{H}) = I_-[W_-(\mathcal{H})] \quad (10.14)$$

и $\dot{W}_-(\mathcal{H})$ является ППГ. Легко проверить, что

$$H_-(\mathcal{H}) = \dot{W}_-(\mathcal{H}) \cap D_-(\mathcal{H}) \quad (10.15)$$

[из (9.5)], и опять горизонт Коши получен как часть ППГ.

Предположим, что $X \in \dot{W}_-(\mathcal{H})$, но $X \notin$ край (\mathcal{H}) [ср. с (9.13) и т. д.]. Тогда должна существовать либо α -кривая, либо β -кривая с начальной точкой X , но не обе, так как $X \notin W_-(\mathcal{H})$. (Это выполняется тривиальным образом, если мы определим существование вырожденной β -кривой, как только $X \in \mathcal{H}$.) Если не существует α -кривой, то $X \in H_-(\mathcal{H})$; если не существует β -кривой, то $X \in I_-(\mathcal{H}) - \mathcal{H}$. Предположим, что $X \in H_-(\mathcal{H})$, так что существует (возможно, вырожденная) β -кривая с началом в X . Тогда мы можем выбрать малую открытую окрестность \mathcal{A} точки X так, чтобы каждая точка из $\bar{\mathcal{A}} \cap I_-(\mathcal{H})$ была концевой точкой β -кривой (т. е. $I_-(\mathcal{H}) \cap \bar{\mathcal{A}} = \mathcal{H} \cap \bar{\mathcal{A}}$). Действительно, если $X \in \mathcal{H}$, то в силу $X \notin$ край (\mathcal{H}) существуют точки A, B вблизи X ($A \ll X \ll B$), такие, что всякая временноподобная кривая от A до B в окрестности X пересекает \mathcal{H} . В этом случае мы выбираем

$\bar{\mathcal{L}} \subset I_-(B) \cap I_+(A)$. Если $X \notin \mathcal{H}$, то выбираем $\bar{\mathcal{L}}$ внутри области прошлого для β -кривой, идущей из X . Любая точка U из $\mathcal{L} \cap I_-(X)$ лежит в $W_-(\mathcal{H})$, так что U является начальной точкой α -кривой, которая пересекает $\mathcal{L} \cap \bar{I}_-(\mathcal{H})$, скажем, в точке V . Теперь V является начальной точкой для β -кривой, и поэтому $V \in \mathcal{L} \cap W_-(\mathcal{H}) \subset W_-(\mathcal{H}) - \mathcal{L}$. Таким образом, условие (б) леммы I выполнено; следовательно:

Каждая точка множества $H_-(\mathcal{H}) - \text{край}(\mathcal{H})$ является начальной точкой отрезка световой геодезической на $H_-(\mathcal{H})$. (10.16)

Разумеется, то же самое относится к $\dot{I}_-[\mathcal{H}] - \mathcal{H}$. Заметим, что если $H_-(\mathcal{H})$ не пересекает край (\mathcal{H}) , то всякая точка $H_-(\mathcal{H})$ является начальной точкой для *продолжимой в будущее* световой геодезической на $H_-(\mathcal{H})$ [так как продолжая световую геодезическую в будущее, мы никогда не достигнем конечной точки на $H_-(\mathcal{H})$]. Это имеет место [ср. с (9.13)], когда \mathcal{H} есть *бескраевое полупространственноподобное замкнутое множество*. Ясно, что эти результаты применимы также и при их обращении во времени.

Другое следствие (10.16) состоит в том, что *если \mathcal{H} есть полупространственноподобное множество, которое пересекает каждую максимальную световую геодезическую в \mathcal{M} по непустому компактному множеству, то \mathcal{H} есть ГГК для \mathcal{M}* . В самом деле, если бы \mathcal{H} не было ГГК, то оно обладало бы непустым горизонтом Коши, образующие которого (световые геодезические) должны пересекать \mathcal{H} по компактному множеству. Согласно (10.16), концевая точка этого множества должна была бы принадлежать множеству край (\mathcal{H}) . Однако множество край (\mathcal{H}) пустое. Чтобы убедиться в этом, рассмотрим $\dot{I}_-[\mathcal{H}]$. Имеем $\mathcal{H} \subset \dot{I}_-[\mathcal{H}]$, так как \mathcal{H} полупространственноподобно. Положим, что существует $P \in \dot{I}_-[\mathcal{H}]$, но $P \notin \mathcal{H}$. Световая геодезическая γ , которая является образующей для $\dot{I}_-[\mathcal{H}]$, проходит через P и содержит компактную часть \mathcal{H} . Пусть Q лежит на γ между P и этой частью.

Любая световая геодезическая η , проходящая через Q и отличная от γ , может пересечь $I_-[\mathcal{H}]$ только в Q . Таким образом, η не может пересечь \mathcal{H} и, следовательно, точки P не существует. Мы имеем $\mathcal{H} = I_-[\mathcal{H}]$, так что \mathcal{H} есть ППГ; итак, \mathcal{H} является бескраевым, что и требовалось доказать.

В качестве обратного утверждения отметим, что если \mathcal{H} есть некоторое полупространственноподобное множество и $P \in \text{int } D_-(\mathcal{H})$, то всякая продолжимая в будущее световая геодезическая γ , проходящая через P , должна встретить \mathcal{H} . Это следует из существования точки $Q \in D_-(\mathcal{H})$ с $Q \ll P$. Если бы γ не пересекала \mathcal{H} , мы могли бы двигаться строго ниже γ по продолжимой в будущее временноподобной кривой, что противоречит условию $Q \in D_-(\mathcal{H})$. Верен и обратный во времени результат, согласно которому всякая продолжимая в прошлое световая геодезическая, проходящая через произвольную точку $P \in \text{int } D_+(\mathcal{H})$, должна пересекать \mathcal{H} . Это нам вскоре потребуется.

Доказательство теоремы I. Положим $\mathcal{S} = I_+[\mathcal{T}]$, как в (10.11). Тогда, если $P \in \mathcal{S} - \mathcal{T}$, то существует световая геодезическая k , отрезок которой лежит на \mathcal{S} и имеет P в качестве конечной точки. Максимально продолжим геодезическую в прошлое. Она «проткнет» \mathcal{C} , поскольку $P \in \text{int } D_+(\mathcal{C})$. Так как $\mathcal{S} (\subset I_+[\mathcal{C}])$ замкнуто, существует точка $R \in \mathcal{S}$, являющаяся начальной точкой для $k \cap \mathcal{S}$. Она должна лежать на \mathcal{S}_- , поэтому $R \in \mathcal{T}$. Кроме того, k должна пересечь пространственноподобную 2-поверхность \mathcal{T} ортогонально в R , т. е. k лежит на одной из двух световых гиперповерхностей, пересекающихся локально на \mathcal{T} , поскольку они представляют собой локальную границу $I_+[\mathcal{T}]$. Таким образом, \mathcal{S} образовано отрезками \hat{k} световых геодезических, которые ортогональны \mathcal{T} в начальной точке и которые могут иметь конечную точку (где они встречаются с \mathcal{S}_+). В действительности всякий такой отрезок \hat{k} должен иметь конечную точку по лемме II, поскольку условие ловушечной поверхности (10.7) вместе со свойствами фо-

кусирования [ср. с (7.39)] и полноты относительно световых геодезических (Iг) означает, что ρ становится неограниченным в некоторой первой точке Q на \hat{k} или на продолжении k отрезка \hat{k} в будущее. (Фактически вторая возможность является более распространенной. Она соответствует появлению области пересечения, которая обычно возникает прежде, чем достигается *каустика*, где $\rho = \infty$.) Q меняется непрерывно при изменении k [ср. с (7.29), (7.33) и т. д.], так что конечные отрезки RQ образуют компактное множество. Таким образом, \mathcal{S} , будучи замкнутым подмножеством компактного множества, само является компактным. Итак, мы имеем [ср. с (9.7)]:

$$\mathcal{S} \text{ есть полупространственноподобное компактное топологическое 3-многообразие в } \mathcal{M}. \quad (10.17)$$

Далее, хорошо известна теорема (см., например, [101, стр. 201]), согласно которой любое пространство-время допускает гладкое временноподобное единичное векторное поле. Мы можем использовать интегральные кривые этого векторного поля, чтобы попытаться осуществить одно-однозначное отображение \mathcal{S} в \mathcal{C} (оба множества полупространственноподобные), так как, согласно (Iа), $\mathcal{S} \subset D_+(\mathcal{C})$. Однако это невозможно, хотя их размерности одинаковы, но \mathcal{S} компактно, а \mathcal{C} некомпактно. Тем самым (Iа), (Iб), (Iв) (Iг) несовместимы, и теорема I доказана.

В свете теоремы I уместно спросить, какое из ее условий наиболее вероятно нарушается в реальной вселенной. Первым кандидатом является, по-видимому, условие (Iа). Есть две причины, по которым (Iа) может не выполняться. Во-первых, вселенная может быть «замкнутой» (другими словами, пространственнокомпактной). Тогда, казалось бы, мы могли бы произвести замену \mathcal{C} компактной пространственноподобной гиперповерхностью, являющейся ГГК [т. е. с $I_+(\mathcal{C}) = D_+(\mathcal{C})$]. Однако доказательство теоремы I нарушится на конечной стадии, и результат не будет больше верен. Например, все условия выполняются в пространстве де Ситтера (9.14), (9.15), где \mathcal{C} есть се-

чение (9.14) посредством $V = -2a < 0$ ($a = \text{const}$) и \mathcal{T} есть $V = -a$, $W = 0$. С другой стороны, трудно поверить, что вопрос о том, является ли вселенная как целое открытой или замкнутой, существенно влияет на «локальную» коллапсирующую систему (размеры которой меньше по крайней мере в 10^{11} раз). В самом деле, исследование доказательства теоремы I показывает, что оно может быть доведено до конца и в случае компактного \mathcal{V} , но *при условии*, что (в хорошо определенном смысле) коллапсирующий объект не «проглотит» всю вселенную! Замкнутая вселенная типа модели Фридмана с $k = 1$, которая в конце концов сжимается до состояния высокой плотности, может, в принципе, удовлетворять этому условию. Решение проблемы сингулярностей указанным выше способом, по-видимому, исключается теоремой Хокинга [41], которая по существу утверждает, что «почти все» миры с *компактной* ГГК должны быть неполны, если они удовлетворяют более сильному, но «разумному» требованию к R_{ab} , вытекающему из (7.47).

Вторая причина, по которой (1a) может нарушаться, является более серьезной. Лапласовская идея о том, что будущее вселенной должно полностью определяться ее поведением в один момент «времени», предшествовала как теории относительности, так и квантовой теории. Мы уже привыкли к индетерминизму второй теории¹⁾, но почему же мы должны требовать детерминизма во всех случаях в общей теории относительности? Возможность того, что вселенная не должна обладать ГГК, весьма заманчива. Известен ряд физически интересных миров, не содержащих ГГК. Два примера были рассмотрены в разд. 9. Решения Керра [48] и Рейснера — Нордстрема (см., например, [62]), аналитические продолжения которых даны соответственно Бойером и Линдквистом [12] (см. также [18, 19]) и Грейвсом и Бриллом [37], больше

¹⁾ Слово «индетерминизм» здесь вряд ли уместно. В случае квантовой теории до тех пор, пока мы не переходим к «локализации» частицы и тому подобным операциям, а работаем с самой волновой функцией, все причинно определено. — *Прим. ред.*

подходят к обсуждению вопроса о коллапсе (хотя и содержат сингулярности).

Метрика Керра есть

$$ds^2 = dv^2 - \frac{2mr}{h} (dv + a \sin^2 \theta d\varphi)^2 - 2 dv dr + \\ + h d\theta^2 + 2a \sin^2 \theta dr d\varphi + (r^2 + a^2) \sin^2 \theta d\varphi^2, \quad (10.18)$$

где m и a константы (m — масса, ma — угловой момент) и

$$h = r^2 + a^2 \cos^2 \theta. \quad (10.19)$$

Метрика Рейснера — Нордстрема есть

$$ds^2 = \left(1 - \frac{2m}{r} + \frac{e^2}{r^2}\right) dv^2 - 2 dv dr - \\ - r^2(d\theta^2 + \sin^2 \theta d\varphi^2), \quad (10.20)$$

где v — опережающее время, e — электрический заряд. Метрика (10.18) с $m > |a|$ и метрика (10.20) с $m > |e|$ отчасти подобны в том отношении, что они обладают не только горизонтом событий для внешних геодезических («шварцшильдовская горловина»), который определяется соответственно условием

$$r_+ = m + (m^2 - a^2)^{1/2}, \quad r_+ = m + (m^2 - e^2)^{1/2}, \quad (10.21)$$

но также вторым «горизонтом», на котором световые конусы наклоняются вторично:

$$r_- = m - (m^2 - a^2)^{1/2}, \quad r_- = m - (m^2 - e^2)^{1/2}. \quad (10.22)$$

Обе метрики имеют истинную сингулярность при $r=0$, но природа их несколько различна. В случае метрики Керра сингулярность имеет кольцевую структуру.

После прохождения кольца мы попадаем в область $r < 0$. (Как показал Картер [19], вблизи кольца встречаются замкнутые временноподобные кривые.) Отвлекаясь от поведения метрики вблизи (и на) сингулярности, мы можем изобразить общую картину, как на рис. 40. С помощью процедуры сшивания, аналогичной той, которая применяется в шварцшильдовском случае, можно получить «максимальное» аналитическое продолжение (рис. 41). Заштрихованная часть рис. 41 соответствует всему рис. 40. Вся

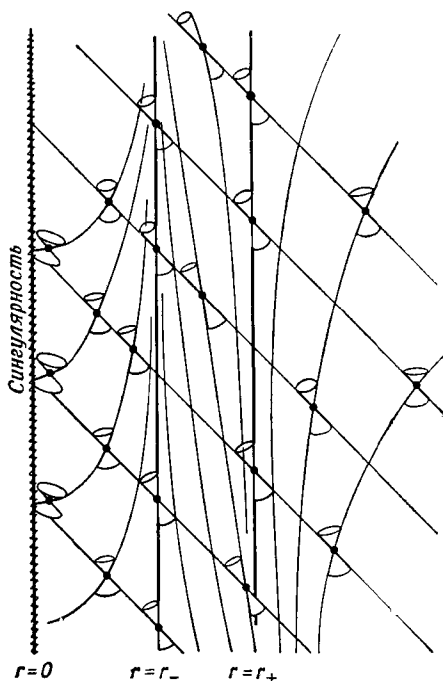
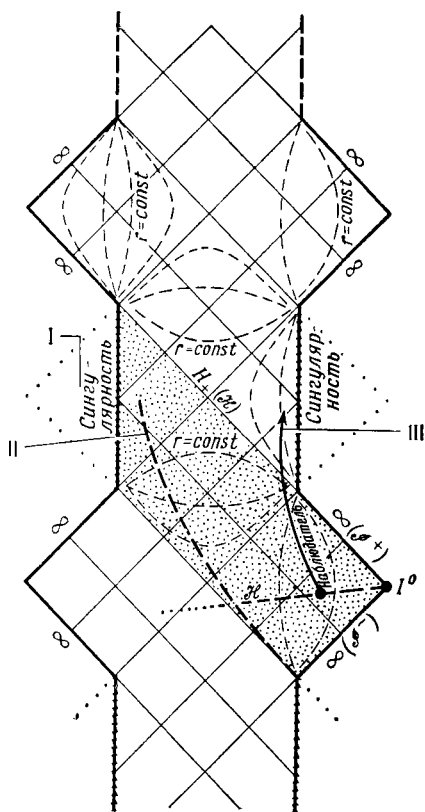


Рис. 40. Картина Керра, применимая также к метрике Рейснера — Нордстрема.

диаграмма рис. 41 может быть склеена (за исключением некоторых центральных точек) из перекрывающихся кусков, изометричных заштрихованной области. (Для этой цели необходимо отдавать себе отчет в существовании дискретной симметрии метрики, которая не видна из записей (10.18), (10.20); ср. с [12].)

Значение всего этого состоит в том, что наблюдатель, следующий внутрь со звездой, а после пересечения $r=r_+$, движущийся с ускорением наружу, может оказаться в части вселенной, которая причинно не определяется информацией, заданной на естественной начальной гиперповерхности \mathcal{H} . Действительно, наблюдатель пересекает $H_+(\mathcal{H})$. Область, в которую он попадает, содержит сингулярность, которую он может «видеть». Но поскольку он пересекает $H_+(\mathcal{H})$, мы, разумеется, не можем гарантировать, что пространство-время, в которое он попадает, является тем, которое мы для него построили. По существу, «другая вселенная» присоединяется к той, с которой он начал. То частное продолжение, которое мы выбрали, было получено путем *аналитического продолжения*, а не с помощью уравнений поля. (Но даже аналитическое

Рис. 41. Картина Картера—Бойера—Линдквиста, включающая картину Грейвса—Брилла. I — область, находящаяся «по ту сторону» кольца в решении Керра; II — граница звезды (ниже этой линии решение неприменимо); III — наблюдатель попадает в новую часть вселенной, не связанную причинно с \mathcal{H} .



продолжение не обеспечивает глобальной однозначности. Поразительным примером является рассмотренное в разд. 9 пространство-время Тауба — НУТ. Но в плоском пространстве-времени аналитическое продолжение также не подразумевает однозначности. Трудность в том, что в общем случае неизвестно, когда надо отождествлять события, достигнутые по разным путям.) Даже если бы мы избежали сингулярности в «новой вселенной», остается принципиальная проблема типа той, которая имеется на ранних стадиях космологических моделей «большого взрыва». Причинно не связанные области должны объединиться таким образом, чтобы они «стыковались». Впечатляющим примером является решение Рейснера — Нордстрема, в котором «новая вселенная» должна содержать новые заряженные частицы

в точности того же полного заряда. (Это следует даже независимо от симметрии.)

Есть еще одна трудность, с которой сталкивается наш наблюдатель, пересекая $H_+(\mathcal{H})$. Когда он смотрит на вселенную, которую он «покидает», то, бросив на нее последний взгляд при пересечении $H_+(\mathcal{H})$, наблюдатель видит всю последующую историю остальной части его «прежней вселенной». Именно здесь космологические вопросы могут быть очень важны для наблюдателя. Если, например, вместе со звездой будет сжиматься неограниченное количество материи, то, вероятно, наблюдатель столкнется с бесконечной плотностью материи вдоль « $H_+(\mathcal{H})$ ». Даже если сжимается только конечное количество материи, может оказаться невозможным в общем случае избежать сингулярности кривизны на $H_+(\mathcal{H})$. В настоящее время этот вопрос не решен, но не исключено, что именно в этой области надо скорее (или с той же вероятностью?) искать сингулярность, чем в «центре». В этой связи интересен пример коллапсирующего заряженного облака пыли, предложенный Бардином [5] (ср. с [23а, 66а]). В данном случае сингулярность *не наступает* внутри $D_+(\mathcal{H})$, хотя сингулярность заряда должна быть создана в «новой вселенной» позади $H_+(\mathcal{H})$ (рис. 41). Волнующим вопросом является возможность приписать какую-либо реальность той ситуации, когда «новая вселенная» создается в результате коллапса. Мы можем рассматривать (10.17) как указание на то, что ловушечная поверхность приводит к появлению пространственнокомпактной вселенной, которая «отпочковывается» от старой. Однако все эти результаты покоятся на предположении, что $H_+(\mathcal{H})$ может сохраниться без появления сингулярностей кривизны при произвольном возмущении на \mathcal{H} . Мы должны также иметь в виду, что возможность избежать сингулярностей кривизны кажется маловероятной в свете теоремы III из следующего раздела.

11.

Сингулярности в космологии

Этот раздел посвящен двум важным теоремам Хокинга [40, 42]. В некоторых местах аргументация будет дана только в общих чертах; за подробностями мы отсылаем читателя к оригинальным статьям Хокинга. Цель теорем состоит в том, чтобы показать, что с точки зрения современных физических представлений в пространстве-времени существуют (или существовали) особые области, вероятно, включающие огромные кривизны, где, как следует ожидать, локальные физические законы сильно отличаются от тех, к которым мы привыкли. Теорема II интересна главным образом тем, что в ней нигде не сделано предположений, связанных с причинностью. Она применима только к пространственнокомпактной вселенной, удовлетворяющей условию (неравенству), которое может выполняться и в реальной вселенной, но которое невозможно проверить путем наблюдений. С другой стороны, теорема III содержит аналогичное условие, которое, согласно Хокингу и Эллису [44], вероятно, можно удовлетворить даже на основании имеющихся в настоящее время астрономических наблюдений. В обращенной во времени форме теорема III будет также применима к соответствующей ситуации гравитационного коллапса (такой, как слабо возмущенное коллапсирующее пылевое облако Оппенгеймера — Снайдера; см. разд. 10). В теореме III используется предположение причинности, но гораздо менее жесткого типа, чем условие (Ia) существования глобальной гиперповерхности Коши в теореме I. Фактически должны быть исключены замкнутые временноподобные кривые (т. е. $\approx S^1$), но мы будем использовать более сильный вариант этого требования.

Пусть $P \in M$. Тогда на M эквивалентны следующие условия относительно P [42, 50]:

Существуют сколь угодно малые окрестности точки P , каждая из которых не пересекает ни одной временноподобной кривой по несвязному множеству, (11.1)

т. е., грубо говоря, временноподобные кривые из окрестности P не могут покинуть и затем снова вернуться в окрестности P ;

Если $P < Q$ и каждая точка из $I_-(Q)$ хронологически предшествует каждой точке из $I_+(P)$, то $P = Q$ (11.2)

(ср. разд. 9) и

Утверждение (11.2) обратимо во времени. (11.3)

Если выполняется (11.1), (11.2) или (11.3), то мы говорим, что в P выполняется сильная причинность. Если это условие выполняется для всех $P \in M$, то мы говорим, что сильная причинность выполняется в M . Эквивалентность (11.1) — (11.3) установить нетрудно, поэтому мы опускаем доказательство. Приведенный Хокингом пример, для которого условие не выполняется, изображен на рис. 42. [Используя метрику и координаты Минковского, удаляем области $(x^0 > 1, x^1 \leq -1)$, $(x^0 < -1, x^1 \geq 1)$, $(x^0 = -x^1 = 1)$ и $(x^0 = -x^1 = -1)$; затем отождествляем $(1, x^1, x^2, x^3)$ с $(-1, -x^1, x^2, -x^3)$, $x^1 < -1$; P — начало координат. Если в P выполняется сильная причинность, мы называем окрестность \mathcal{A} точки P причинно выпуклой, при условии, что ни одна из временноподобных кривых не пересекает \mathcal{A} по несвязному множеству [ср. с (11.1)].

Если в P нарушается сильная причинность, произвольно малое возмущение метрики в окрестности P может привести к появлению замкнутых временноподобных кривых, которые в силу ряда причин являются крайне нежелательными с физической точки зрения. Таким образом, представляется весьма разумным предположение о том, что сильная причинность должна выполняться в каждой точке пространства-времени. Другим условием, более слабым, чем (11.1) —

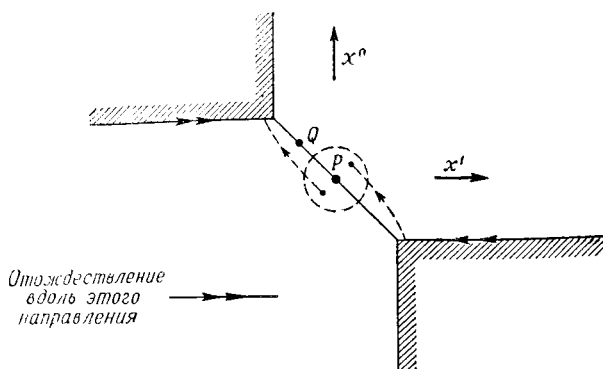


Рис. 42. Пространство, в котором отделимость прошлого и будущего выполняется, а сильная причинность не выполняется.

(11.3), но более сильным, чем отсутствие замкнутых временноподобных кривых, является условие *отделимости будущего* [50] в P

$$I_+(P) = I_+(Q) \text{ влечет за собой } P = Q. \quad (11.4)$$

Оно утверждает, что временноподобная кривая, проходящая через P в будущее, не может затем вернуться сколь угодно близко к точке P . Мы говорим, что \mathcal{M} удовлетворяет условию *отделимости прошлого* в P тогда и только тогда, когда для всех Q

$$I_-(P) = I_-(Q) \text{ влечет за собой } P = Q. \quad (11.5)$$

Картер¹⁾ предложил иерархию условий причинности, более сильных, чем (11.1). Опять-таки, нарушение любого из условий Картера могло бы привести к появлению замкнутых временноподобных кривых при слабом возмущении \mathcal{M} .

¹⁾ Одно из таких условий было бы нарушено, если бы существовали две точки $P, R \in \mathcal{M}$, такие, что для любой окрестности \mathcal{P} точки P и любой окрестности \mathcal{R} точки R всегда нашлась бы временноподобная кривая с начальной точкой в \mathcal{P} и конечной точкой в \mathcal{R} и другая временноподобная кривая с конечной точкой в \mathcal{P} и начальной точкой в \mathcal{R} . Такое поведение может иметь место без нарушения сильной причинности, в чем можно убедиться, если взять в качестве \mathcal{M} двулистное накрывающее многообразие для того многообразия, которое изображено на рис. 42 (Картер).

Условие того, чтобы сильная причинность выполнялась в каждой точке \mathcal{M} , фактически эквивалентно утверждению, что «топология Александрова» \mathcal{T}^* является хаусдорфовой [50], где открытые множества \mathcal{T}^* имеют в качестве базы множества $I_+(X) \cap I_-(Y)$. Это опять-таки эквивалентно утверждению, что \mathcal{T}^* согласуется с обычной топологией многообразия для \mathcal{M} . На рис. 42 приведен случай, когда упомянутое условие нарушается. В этом примере любое открытое в \mathcal{T}^* множество, содержащее P , пересекает любое открытое в \mathcal{T}^* множество, содержащее Q , так что свойство Хаусдорфа нарушается.

Теперь сформулируем теоремы.

Теорема II (теорема Хокинга). Следующие требования на пространство-время \mathcal{M} несовместимы:

- (IIa) Существует компактная пространственно-подобная гиперповерхность (без границы) \mathcal{H} .
- (IIб) Дивергенция θ единичной нормали к \mathcal{H} положительна в каждой точке \mathcal{H} .
- (IIв) $R_{ab}t^at^b \leq 0$ для каждого временноподобного вектора t^a .
- (IIг) \mathcal{M} полно относительно временноподобных геодезических, направленных в прошлое.

Можно рассматривать (IIa) и (IIб) как утверждения о том, что «вселенная является (или была) пространственнокомпактной и расширяющейся». Однако такое «расширение», какое подразумевается в (IIб), должно иметь место в каждой точке \mathcal{H} . Таким образом, в присутствии коллапсирующего объекта мы не можем обеспечить (IIб) для \mathcal{H} «в настоящее время». Поэтому, если бы это условие было выполненным, вероятно, \mathcal{H} следовало бы относить к ранним фазам расширения.

С учетом уравнений Эйнштейна (7.1) требование (IIв) является энергетическим условием (7.47), которое (если еще $\lambda = 0$) очень разумно с физической точки зрения.

Теорема III (теорема Хоккинга). Следующие требования на пространство-время \mathcal{M} несовместимы:

- (IIIa) Сильная причинность выполняется в каждой точке из $\bar{I}_-(P)$ для некоторой точки $P \in M$.
- (IIIб) Дивергенция всех временноподобных и световых геодезических, проходящих через P , изменяет знак где-то в прошлом от точки P .
- (IIIв) $R_{ab}t^at^b \leq 0$ для каждого временноподобного вектора t^a .
- (IIIг) \mathcal{M} полно относительно временноподобных и световых геодезических, направленных в прошлое.

Условие (IIIб) сформулировано несколько нестрого, поскольку дивергенция временноподобных геодезических измеряется величиной θ из (7.40), а дивергенция световых геодезических измеряется с помощью ρ из (7.18), (7.33). Область, в которой ρ изменяет знак, не есть в действительности предел области, в которой θ изменяет знак. Другая формулировка (IIIб) состоит в упоминании только временноподобных геодезических и в утверждении, что (соответствующее) изменение знака происходит внутри *компактной области*.

Я дам набросок доказательств теорем II и III, указывая также главные леммы. (Эти леммы и их доказательства по существу заимствованы у Хоккинга [40, 42], хотя приведенное здесь обсуждение несколько отличается от оригинального.)

Лемма IV. Если существует пространственноподобная гиперповерхность \mathcal{H} (без границы), являющаяся замкнутым подмножеством \mathcal{M} (т. е. собственно вложенным), то существует накрывающее многообразие \mathcal{M}^* для \mathcal{M} , где прообраз \mathcal{H} под накрывающей картой состоит из множества дискретных изометрических копий \mathcal{H} , каждая из которых является полупространственноподобной в \mathcal{M}^* .

Идея состоит в том, чтобы превратить \mathcal{H} [в (IIIa)] в полупространственноподобное множество, «развора-

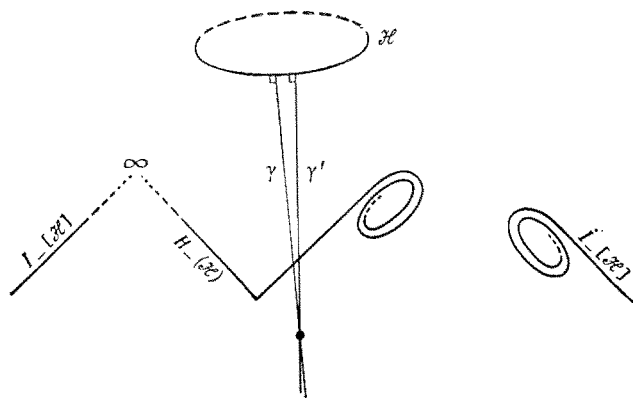


Рис. 43. Тип ситуаций в теореме II, который следует иметь в виду.

чивая» \mathcal{M} , но не настолько сильно, чтобы новое \mathcal{H} стало некомпактным. Для этой цели выбираем точку O в \mathcal{M} , затем каждому $P \in \mathcal{M}$ и каждому классу эквивалентности кривых, соединяющих O и P , число пересечений которых с \mathcal{H} есть заданное целое число, ставим в соответствие точку \mathcal{M}^* . (Число пересечений является гомотопическим инвариантом, поскольку \mathcal{H} не имеет краев.) Легко проверить, что \mathcal{M}^* обладает свойствами, требуемыми в лемме. [Эта конструкция позволяет эффективно получить \mathcal{M}^* из универсального накрывающего многообразия для \mathcal{M} , устанавливая отношение эквивалентности между элементами фундаментальной группы \mathcal{M} , а именно называя два таких элемента эквивалентными, если они дают равные числа пересечений с \mathcal{H} (см. также [34]).]

Без уменьшения общности мы можем теперь рассматривать \mathcal{M}^* как \mathcal{M} в теореме II, так что мы можем считать \mathcal{H} полупространственноподобным множеством. Чтобы мы могли рассматривать теоремы II и III параллельно, определим для теоремы III

$$\mathcal{H} = \dot{I}_{-}(P). \quad (11.6)$$

(Выполнение сильной причинности в P означает, что \mathcal{H} есть непустое множество, напоминающее световой

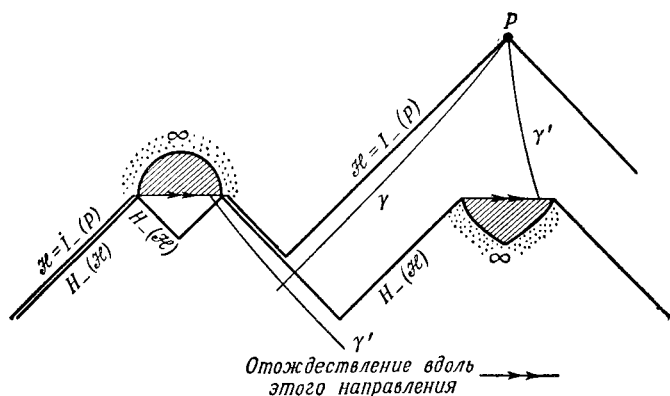


Рис. 44. Неприятный тип ситуации в теореме III, который следует иметь в виду.

конус в P .) Тогда в обоих случаях мы имеем $\mathcal{H} = \overline{\mathcal{H}}$ как бескрайнее полупространственноподобное множество.

Полезно иметь в виду некоторые типы ситуаций, которые могут возникнуть. В теореме II мы имеем дело с ситуацией типа той, которая изображена на рис. 43. Ни $H_-(\mathcal{H})$, ни $I_-(\mathcal{H})$ не могут встретить \mathcal{H} , поскольку \mathcal{H} является бескраевой и пространственноподобной гиперповерхностью. Каждая точка $H_-(\mathcal{H})$ лежит на продолжимой в будущее световой геодезической на $H_-(\mathcal{H})$. Геодезическая может уходить в бесконечность или (поскольку предположений причинности мы не делали) просто по винтовой линии в некоторой компактной области (ср. с рис. 33). Неприятный тип ситуаций, который может встречаться в теореме III, изображен на рис. 44. Здесь $H_-(\mathcal{H})$ может пересекать \mathcal{H} . [Именно эта возможность вынуждает нас ввести предположение причинности (IIIa).]

Лемма V. Если $\mathcal{H} = \overline{\mathcal{H}}$ является бескраевым и полупространственноподобным¹⁾, то $\text{int } D_-(\mathcal{H})$ состоит

¹⁾ Примечание к русскому изданию. Условия на \mathcal{H} необходимо дополнить требованием, чтобы в каждой точке \mathcal{H} выполнялась сильная причинность (см. [131*]).

из всех точек $X \in I_-[\mathcal{H}]$, для которых множество $J_+(X) \cap \bar{I}_-[\mathcal{H}]$ является компактным и не содержит точек, в которых нарушается сильная причинность.

Обозначим множество таких точек X буквой \mathcal{X} [относительно $J_+(X)$ см. (9.2)]. Пусть $X \in \mathcal{X}$. Рассмотрим продолжимую в будущее временноподобную кривую с X как начальной точкой. Если кривая не пересекает \mathcal{H} , то в силу компактности $\mathcal{Y} = J_+(X) \cap \bar{I}_-[\mathcal{H}]$ она должна иметь точку накопления в \mathcal{Y} . Это невозможно, поскольку в \mathcal{Y} выполняется сильная причинность. Следовательно, $\mathcal{X} \subset D_-(\mathcal{H})$. Кроме того, $X \notin H_-(\mathcal{H})$, так как в противном случае, согласно (10.16), продолжимая в будущее световая геодезическая с начальной точкой X будет лежать в \mathcal{Y} , причем \mathcal{H} — бескраевое. (Такая световая геодезическая не может содержаться в компактном множестве, если во множестве выполняется сильная причинность.) Отсюда $\mathcal{X} \subset \text{int } D_-(\mathcal{H})$. Для доказательства обратного утверждения прежде всего напомним, что если $Z \in \text{int } D_-(\mathcal{H})$, то любая продолжимая в будущее *непространственноподобная* кривая, проходящая через Z , должна пересечь \mathcal{H} . Далее, сильная причинность не может нарушаться в любой точке $Z \in \text{int } D_-(\mathcal{H})$. [В противном случае мы могли бы использовать Q из (11.2), чтобы получить ZQ как световую геодезическую, вдоль которой нарушается сильная причинность и максимальное продолжение которой не пересекает \mathcal{H} .] Предположим теперь, что $\mathcal{L}_0 = J_+(Z_0) \cap \bar{I}_-[\mathcal{H}]$ не является компактным для некоторого $Z_0 \in \text{int } D_-(\mathcal{H})$. Покроем \mathcal{L}_0 локально конечной системой причинно выпуклых открытых множеств \mathcal{U}_i , которые достаточно малы, чтобы выполнялось следующее требование. Любая точка U в \mathcal{U}_i должна быть центром открытого шара \mathcal{B}_U в нормальных координатах, содержащего $\bar{\mathcal{U}}_i$. Пусть $Z_0 \in \mathcal{U}_{i_0}$. Так как \mathcal{L}_0 некомпактно, оно содержит последовательность точек W_j без точки накопления в \mathcal{L}_0 . Существует непространственноподобная кривая от Z_0 до W_j , которая пересекает \mathcal{B}_Z , скажем, в V_j . Пусть V — точка накопления для V_j . Отрезок непространственноподобной гео-

геодезической Z_0V пересекает \mathcal{U}_{i_0} в единственной точке $Z_1 \in \mathcal{U}_{i_0}$. Тогда $\mathcal{Z}_1 = J_+(Z_1) \cap \bar{I}_-[\mathcal{H}]$ содержит последовательность точек W_j без точки накопления в \mathcal{Z}_1 и, следовательно, также некомпактно. Повторяя это рассуждение, получим последовательность $Z_0 < Z_1 < Z_2 < \dots$, которая лежит на непространственноподобной кривой γ . Из свойства локальной конечности системы \mathcal{U}_i следует, что Z_i не имеет точки накопления и поэтому γ есть продолжимая в будущее непространственноподобная кривая, проходящая через Z_0 и не пересекающая \mathcal{H} . Таким образом, $Z_0 \notin \text{int } D_-(\mathcal{H})$, т. е. $\text{int } D_-(\mathcal{H}) \subset \mathcal{R}$. Этим завершается доказательство леммы V.

Если \mathcal{A} и \mathcal{R} — подмножества или точки \mathcal{M} , определим $d(\mathcal{A}, \mathcal{R})$ как наименьшую верхнюю грань длин всех временноподобных кривых с начальной точкой (в) \mathcal{A} и конечной точкой (в) \mathcal{R} . Полагаем $d(\mathcal{A}, \mathcal{R}) = 0$, если таких кривых нет. В некоторых случаях мы будем иметь $d(\mathcal{A}, \mathcal{R}) = \infty$ (но, например, для ограниченных подмножеств мира Минковского d всегда конечно: временноподобная прямая линия, соединяющая две точки в мире Минковского, имеет максимальную длину по сравнению со всеми временноподобными кривыми, соединяющими эти точки).

Лемма VI. Если $\mathcal{H} = \bar{\mathcal{H}}$ — бескраевое и полупространственноподобное, то $d(X, \mathcal{H})$, где $X \in \text{int } D_-(\mathcal{H})$ ограничено и достигается на некоторой геодезической γ , соединяющей X и \mathcal{H} .

По лемме V мы знаем, что $\mathcal{U} = J_+(X) \cap \bar{I}_-[\mathcal{H}]$ компактно и поэтому может быть покрыто конечным числом малых открытых окрестностей \mathcal{B}_i в нормальных координатах. Так как на $\text{int } D_-(\mathcal{H})$ выполняется сильная причинность, мы можем сделать границы \mathcal{B}_i причинно выпуклыми. Обозначим через d_i наименьшую верхнюю грань длин временноподобных кривых в \mathcal{B}_i . Тогда из локального рассмотрения следует, что d_i конечно и достигается (фактически на геодезической).

Если γ — временноподобная кривая от X до \mathcal{H} , то ее длина не может превышать $d = \sum d_i$. Следова-

но, $d(X, \mathcal{H})$ конечно. Далее, из компактности следует, что максимум $d(X, Y)$ для $Y \in \mathcal{H}$ будет достигаться, скажем, при $Y = Y_0$. Чтобы построить γ максимальной длины от X до \mathcal{H} так, чтобы длина γ действительно равнялась $d(X, \mathcal{H})$, выберем нормальный координатный шар \mathcal{B} с центром Y_0 и выберем $Z = Z_0$ так, чтобы максимизировать $d(X, Z) = d(Z, Y_0)$, где $Z \in \mathcal{U}$ перемещается по компактной границе \mathcal{B} шара \mathcal{B} . Пусть γ — геодезическая с конечной точкой Y_0 , которая проходит через Z_0 и продолжается в прошлое до длины d . Повторим приведенную выше аргументацию, применяя ее к Z_0 вместо Y_0 и максимизируя $d(X, V) + d(V, Z_0)$. Мы получим V_0 , которое должно лежать на γ , так как временноподобная кривая с изломом всегда увеличивает свою длину при сглаживании излома. Продолжая этот процесс, мы видим, что γ должна заканчиваться в начальной точке X , что и доказывает лемму VI.

Лемма VII. Если \mathcal{H} — пространственноподобная гиперповерхность и γ — отрезок временноподобной геодезической от X до \mathcal{H} , максимизирующий $d(X, \mathcal{H})$, то γ пересекает \mathcal{H} ортогонально и не содержит точек (разве что за исключением X), сопряженных к \mathcal{H} .

Доказательство не отличается существенно от случая многообразия с положительно определенной метрикой [58]. Точка, сопряженная к \mathcal{H} на γ , есть фокальная точка конгруэнции Γ временноподобных геодезических, ортогональных к \mathcal{H} [т. е. точка, где θ из (7.40) становится бесконечным].

Мы можем также применять лемму VII в предельном случае к гиперповерхности \mathcal{H} из (11.6), требуемой в теореме III. В этом случае, если γ — отрезок временноподобной геодезической от X до \mathcal{H} , который максимизирует $d(X, \mathcal{H})$, то, чтобы быть «ортогональным» к \mathcal{H} , отрезок γ должен проходить через P . Всякая другая точка \mathcal{H} лежит на световой геодезической на \mathcal{H} , которая простирается в будущее до более далеких от X точек. Тогда мы интерпретируем лемму VII как утверждение о том, что γ проходит через

P и не содержит точек (разве что за исключением X), сопряженных к P (т. е. фокальных точек конгруэнции Γ временноподобных или световых геодезических, проходящих через P).

Теперь мы можем применить эти результаты к теоремам II и III. Пусть t^a обозначает поле единичных (направленных в будущее) касательных векторов к временноподобным γ из Γ . По (IIб) или (IIIб) мы имеем $\theta > 0$ в некоторой точке из $\bar{I}_-[\mathcal{H}]$ на каждой $\gamma \in \Gamma$. Тогда в силу (IIв) или (IIIв) и эффекта Райчаудури [ср. с (7.40) — (7.47), а также с (7.39)] мы получим фокальную точку $G(\theta = \infty, -\rho = \infty)$ где-то на каждой из $\gamma \in \Gamma$ в прошлом от \mathcal{H} . [В этом месте используется полнота относительно временноподобных (IIг), (IIIг) и световых (IIIг) геодезических, направленных в прошлое.] Если фокальных точек несколько, мы выбираем ближайшую к \mathcal{H} (и $\neq P$). Точка G будет изменяться непрерывно при переходе от одной геодезической γ к другой. В силу компактности \mathcal{H} (теорема II) или того факта, что временноподобные и световые направления в P образуют компактную систему (теорема III), мы получаем, что область \mathcal{G} , выметаемая сегментами γ от G до \mathcal{H} , будет компактной.

Согласно леммам VI и VII, мы имеем: $\text{int } D_-(\mathcal{H}) \subset \subset \mathcal{G}$ и, следовательно, $D_-(\mathcal{H}) \subset \mathcal{G} \cup \mathcal{H}$. В частности, $H_-(\mathcal{H}) \subset \mathcal{G} \cup \mathcal{H}$. В теореме II \mathcal{H} пространственно-подобно, поэтому $H_-(\mathcal{H}) \cap \mathcal{H} = \emptyset$. Горизонт Коши $H_-(\mathcal{H})$ как замкнутое подмножество компактного множества \mathcal{G} является компактным. Согласно (10.15) и (9.7), $H_-(\mathcal{H})$ есть, таким образом, компактное многообразие класса C^0 без границы. Согласно (IIг), мы можем продолжить каждую γ в прошлое до длин, больших, чем $\max d(G, \mathcal{H})$. Тогда, согласно леммам VI и VII, каждая γ пересекает $H_-(\mathcal{H})$. Положим $F = \gamma \cap H_-(\mathcal{H})$ и определим $p(F)$ как максимум длин сегментов от F до \mathcal{H} . (Для фиксированного F максимум достигается в силу компактности \mathcal{H} и системы Γ .) Нетрудно видеть, что $p(F)$ [с $F \in H_-(\mathcal{H})$] достигает своей минимальной величины. Пусть $F = F_0$ минимизирует $p(F)$. Световая геодезическая η на $H_-(\mathcal{H})$

имеет F_0 в качестве начальной точки. Пусть F_1 лежит строго в будущем относительно F_0 на η . Длина ломаной кривой от F_0 до \mathcal{H} , которая состоит из малого отрезка F_0F_1 геодезической η и максимальной кривой γ от F_1 до \mathcal{H} , не может быть меньше $\rho(F_0)$. Такая γ существует, поскольку $D_-(\mathcal{H}) \subset \mathcal{G}$. Сглаживая излом и сдвигая кривую с $H_-(\mathcal{H})$, мы получим временноподобную кривую ξ от F_0 до \mathcal{H} с длиной k , большей чем $\rho(F_0)$. Взяв B на ξ очень близко к F_0 , чтобы $d(B, \mathcal{H}) > \rho(F_0)$, и устремляя B к F_0 , получим предельное положение (соответствующей) γ , проходящей через B (при этом используем компактность \mathcal{H}). Это даст нам $\rho(F_0) \geq k > \rho(F_0)$, чем устанавливается требуемое в теореме II противоречие.

В теореме III \mathcal{H} не является пространственноподобным и поэтому $H_-(\mathcal{H})$ может пересечь \mathcal{H} (рис. 44)¹⁾. Таким образом, мы не можем предполагать, что $H_-(\mathcal{H})$ компактно, и изложенная выше аргументация неприменима. Взамен этого мы привлечем сильную причинность. Определим отношение \langle на $H_-(\mathcal{H})$ условием:

$$U \langle V \text{ тогда и только тогда, когда } I_-(U) \subset I_-(V), \quad (11.7)$$

[т. е. $U \in \bar{I}_-(V)$]. Тогда \langle является отношением порядка на $H_-(\mathcal{H})$ (более слабым, чем $<$). Положим $\mathcal{A} = \mathcal{G} \cap H_-(\mathcal{H})$. Тогда \mathcal{A} компактно, поскольку $H_-(\mathcal{H})$ замкнуто и \mathcal{G} компактно. Множество $\mathcal{K} = \mathcal{A} \cap \dot{\mathcal{G}} \subset \subset H_-(\mathcal{H}) - \text{int } \mathcal{G} \subset \mathcal{H}$. Пусть $U \subset \mathcal{A}$. Тогда продолжимая в будущее световая геодезическая α на $H_-(\mathcal{H})$ имеет начальную точку U . В силу компактности и сильной причинности α должна покидать \mathcal{A} , скажем, в $U_0 \in \mathcal{K}$, и мы будем иметь $U \langle U_0$. Но в окрестности $\alpha - \mathcal{A}$ есть временноподобные кривые, ведущие в P [так как $\alpha - \mathcal{A} \subset \mathcal{H} = \dot{I}_-(P)$]. Эти временноподобные кривые должны пересечь $H_-(\mathcal{H}) - \mathcal{H} \subset \mathcal{A}$ на их пути в P . В пределе по мере приближения этих кривых к α множество точек пересечения имеет точку накопления V на \mathcal{A} , и легко проверить, что $U_0 \langle V$

¹⁾ Примечание к русскому изданию. Более строгое доказательство изложенных в этом разделе утверждений см. в [131*].

(причем $U_0 \neq V$ вследствие сильной причинности). Подобным же образом V порождает V_0 на \mathcal{H} с $V \langle V_0$ и W с $V_0 \langle W$ ($V_0 \neq W$) и т. д. Последовательность точек $U_0 \langle V_0 \langle W_0 \langle \dots$ на \mathcal{H} должна иметь точку накопления в силу компактности \mathcal{H} . Мы должны были бы иметь нарушение сильной причинности в этой точке. Это доказывает теорему III.

Одно из замечательных свойств теоремы III состоит в том, что нам нет нужды делать какие-либо предположения, кроме сильной причинности, о природе вселенной как целого. Теорема относится только к световому конусу прошлого точки P и к его внутренности, а эти области в принципе могут наблюдаться из точки P . В обращенной по времени форме теорема применима к достаточно однородным моделям коллапса (хотя и не к столь общим, как те, которые укладываются в теорему I) и не требует информации о глобальной космологической картине. (Мы выбираем P в качестве центра и предполагаем, что коллапс подобен последней стадии решения Оппенгеймера — Снайдера.) Поскольку также нет нужды в наличии гиперповерхности Коши, это наталкивает на мысль, что даже если в общих ситуациях горизонт Коши (рис. 41) и существует, это, вероятно, не позволяет избежать сингулярностей где-либо в другом месте.

Когда речь идет о световом конусе прошлого точки P , мы, разумеется, ожидаем, что изменение знака дивергенции геодезических произойдет не раньше, чем будет охвачена существенная часть всей вселенной. Чтобы определить, где находится эта точка, необходимо знать плотность материи во вселенной, которая в настоящее время известна очень неуверенно. Кроме того, большая часть материи, сосредоточенной в галактиках, распределена очень неоднородно, что усложняет общий эффект. Не ясно, можно ли удовлетворить условию (IIIб) вообще на основании одних только наблюдений матеральных объектов. Идея Хокинга и Эллиса [44] состоит в том, чтобы использовать плотность *излучения* во вселенной, которое в настоящее время имеет чрезвычайно изотропное рас-

пределение со спектром, близким к чернотельному при абсолютной температуре 3 К. Они утверждают, что этого излучения, благодаря его однородности, должно быть в конечном счете достаточно *самого по себе*, чтобы обеспечить выполнение условия (IIIб), если исключить маловероятный вариант, что присутствует большое количество ионизованного газа. Если этот газ присутствует, тогда *его* должно быть достаточно, чтобы вызвать требуемый эффект фокусирования.

Отсюда следует, что сингулярности¹⁾ (кривизны?) являются реальным свойством нашей вселенной при условии, что выполняются уравнения Эйнштейна. (В действительности уравнения Эйнштейна не так уж сильно используются. Например, если бы мы заменили теорию Эйнштейна теорией Бранса — Дикке [13], то выводы теорем I, II и III остались бы по существу неизменными.) Однако требуется больше результатов, если мы хотим получить какое-нибудь представление о структуре этих особых (вероятно, чрезвычайно искривленных) областей пространства-времени²⁾.

1) Под «сингулярностью» я понимаю особую (пекулярную) область, в которой локальная физика должна быть существенно изменена, возможно, вплоть до того, что представление о пространстве-времени как о гладком многообразии не будет больше адекватным (разд. I).

2) Следует подчеркнуть, что теоремы I, II, III позволяют доказать неполноту пространства-времени [разумеется, при условии, что требования (а), (б), (в) каждой из теорем выполнены.] Очень важно установить природу этой неполноты, поскольку в каждом конкретном случае она может быть разной. Как следует из текста, неполнота пространства-времени не всегда связана с сингулярностью кривизны (скажем, с обращением в бесконечность инвариантов тензора кривизны). Примером может служить упоминавшееся решение Тауба — НУТ. Инварианты тензора кривизны для этого решения всюду имеют определенные конечные значения, в том числе и на горизонтах $H_+(\mathcal{P})$ и $H_-(\mathcal{P})$, за которые не могут быть продолжены некоторые из геодезических. Неполнота имеет здесь, так сказать, метрический, а не топологический характер, т. е. она связана с неопределенностью метрики, а не с тем, что геодезические «натываются» на выброшенные точки. Иная ситуация имеет место для решения в пустоте, обобщающего решение Тауба (так называемая «модель перемешанного мира»; первоначально ее исследовали В. А. Белинский и И. М. Халатников [125*] и Ч. Мизнер [127*]). Этот мир также обладает неполнотой, однако можно показать, что в этом случае инварианты

И чтобы выяснить, какая физика может действовать в такой области, нам требуется гораздо более глубокое, чем в настоящее время, понимание взаимоотношений между материей, квантовой теорией и структурой пространства-времени.

В обычных обстоятельствах при решении практических задач общая теория относительности может стоять в стороне (почти в отрыве) от остальной физики. Но в пространственно-временной сингулярности дело обстоит совсем наоборот.

тензора кривизны вообще не стремятся к какому-либо определенному пределу при приближении к горизонту.

Неполнота пространства-времени может быть вызвана и совершенно «безобидными» причинами. Рассмотрим простейший пример. Метрика $ds^2 = dt^2 - t^2 dz^2$ (остальные два измерения нас сейчас не интересуют) удовлетворяет всем условиям одной из теорем Хокинга [40]. (Эта же метрика удовлетворяет первым трем условиям теоремы II, если гиперповерхность $t = \text{const}$ является компактной, для чего достаточно рассмотреть $-z_0 \leq z \leq z_0$ и отождествить линии $z = -z_0$ и $z = z_0$.) Следовательно, рассматриваемый мир неполон. Однако природа этой неполноты очень проста; здесь мы имеем дело с частью (полного) мира Минковского, а именно с внутренностью светового конуса точки $\tau = 0$, $\xi = 0$, где $ds^2 = d\tau^2 - d\xi^2$; $t \operatorname{ch} z = \tau$, $t \operatorname{sh} z = \xi$. Кривизна не только не имеет сингулярностей, но и вообще тождественно равна нулю. Интересное решение уравнений Эйнштейна с материей получил Шепли [129*]. Рассмотренный им мир также неполон, однако плотность и скорость вещества не имеют на горизонте никаких сингулярностей. По-видимому, дальнейшие исследования позволят доказать, что в «достаточно общих» решениях причиной неполноты является все-таки бесконечная кривизна (или ее неаналитичность). — *Прим. перев.*

Литература¹⁾

1. Y. Aharonov, L. Susskind, *Phys. Rev.*, **158**, 1237 (1967)
2. R. A. Alpher, R. C. Herman, *Nature*, **162**, 774 (1948)
3. A. Avez, *Compt. Rend.*, **254**, 3984 (1962).
4. A. Avez, *Inst. Fourier*, **105** (1963).
5. J. Bardeen, Тезисы Международной конференции по гравитации и общей теории относительности, Тбилиси, 1968.
6. C. Behr, *Z. Astrophys.*, **60**, 286 (1965).
- 6a. P. G. Bergmann, *Phys. Rev. Lett.*, **12**, 139 (1964).
7. B. Bertotti, *Proc. Roy. Soc. (Lond.)*, **A 294**, 195 (1966).
8. H. Bondi, *Nature*, **186**, 535 (1960).
9. H. Bondi, in «Lectures on General Relativity; 1964 Brandeis Summer Institute in Theoretical Physics», Vol. I, Prentice-Hall, Englewood Cliffs, 1965.
10. H. Bondi, F. A. E. Pirani, I. Robinson, *Proc. Roy. Soc. (Lond.)*, **A251**, 519 (1959).
11. H. Bondi, M. G. J. van der Burg, A. W. K. Metzner, *Proc. Roy. Soc. (Lond.)*, **A269**, 21 (1962).
12. R. H. Boyer, R. W. Lindquist, *J. Math. Phys.*, **8**, 265 (1967).
13. C. Brans, R. H. Dicke, *Phys. Rev.*, **124**, 925 (1961).
14. D. Brill, *Annals of Phys.*, **7**, 466 (1959).
- 14a. D. Brill, S. Deser, *Phys. Rev. Lett.*, **20**, 75 (1968).
15. H. W. Brinkmann, *Nat. Acad. Sci. (U. S.)*, **9**, 1 (1923).
16. H. A. Buchdahl, *Nuovo Cimento*, **10**, 96 (1958); **25**, 486 (1962).
17. E. Calabi, L. Markus, *Ann. Math.*, **75**, 63 (1962).
18. B. Carter, *Phys. Rev.*, **141**, 1242 (1966).
19. B. Carter, *Stationary Axi-symmetric Systems in General Relativity*, Ph. D. Dissertation, Cambridge University, 1967.
20. E. Cartan, *Ann. Ecole Norm, Sup.*, **40**, 325 (1923); **41**, 1 (1924).
21. S. Chandrasekhar, *Mon. Not. Roy. Astron. Soc.*, **95**, 207 (1935).
- 21a. S. Chandrasekhar, *Proc. Nat. Acad. Sci.*, **57** (1967).
22. G. F. Chew, *Sci. Prog.*, **51**, 529 (1963).
23. R. Debever, *Compt. Rend.*, **249**, 1324, 1744 (1959).
- 23a. V. de la Cruz, W. Israel, *Nuovo Cimento*, **51A**, 745 (1967).
24. R. H. Dicke, *Phys. Rev.*, **125**, 2163 (1962).
25. R. H. Dicke, P. J. E. Peebles, P. G. Roll, D. T. Wilkinson, *Astrophys. J.*, **142**, 414 (1965).
26. А. Г. Дорошкевич, Я. Б. Зельдович, И. Д. Новиков, *ЖЭТФ*, **49**, 170 (1965).

¹⁾ Ссылки со звездочкой добавлены переводчиками и редакторами. — *Прим. ред.*

27. A. S. Eddington, *Nature*, **113**, 192 (1924).
28. J. Ehlers, W. Kundt, in «Gravitation», ed. L. Witten, John Wiley and Sons, New York, 1962.
29. A. Einstein, *Ann. Phys.*, **94**, 769 (1916) (русский перевод: А. Эйнштейн, Собрание научных трудов, т. I, изд-во «Наука», М., 1965, стр. 452).
30. A. Einstein, J. Grommer, *Sitzungsber. Preuss. Akad. Wiss.*, **2** (1927) (русский перевод: А. Эйнштейн, Собрание научных трудов, т. II, изд-во «Наука», М., 1966, стр. 198).
31. A. Einstein, L. Infeld, B. Hoffmann, *Ann. Math.*, **39**, 65 (1938) (русский перевод: А. Эйнштейн, Собрание научных трудов, т. II, изд-во «Наука», М., 1966, стр. 450).
32. D. Finkelstein, *Phys. Rev.*, **110**, 965 (1958).
33. G. Gamow, *Nature*, **162**, 680 (1948).
34. R. P. Geroch, *J. Math. Phys.*, **8**, 782 (1967).
35. R. P. Geroch, *Singularities in the Spacetime of General Relativity*, Ph. D. Dissertation, Princeton University, 1967.
36. J. N. Goldberg, in «Gravitation», ed. L. Witten, John Wiley and Sons, New York, 1962.
37. J. C. Graves, D. R. Brill, *Phys. Rev.*, **120**, 1507 (1960).
38. Haefliger, см. A. Lichnerowicz, *Topics on Space-Times in «Battelle Rencontres, 1967 Lectures in Mathematics and Physics»*, ed. M. DeWitt and J. A. Wheeler, W. A. Benjamin, New York — Amsterdam, 1968, Ch. V, p. 115.
39. B. K. Harrison, K. S. Thorne, M. Wakano, J. A. Wheeler, *Gravitation Theory and Gravitational Collapse*, University of Chicago Press, Chicago, 1965 (русский перевод: Дж. Уилер, Б. Гаррисон, М. Вакано, К. Торн, Теория гравитации и гравитационный коллапс, изд-во «Мир», М., (1967).
40. S. W. Hawking, *Proc. Roy. Soc. (Lond.)*, **A294**, 511 (1966).
41. S. W. Hawking, *Proc. Roy. Soc. (Lond.)*, **A295**, 490 (1966).
42. S. W. Hawking, *Proc. Roy. Soc. (Lond.)*, **A300**, 187 (1967).
43. S. W. Hawking, G. F. R. Ellis, *Phys. Lett.*, **17**, 246 (1965).
44. S. W. Hawking, G. F. R. Ellis, *Astrophys. J.*, **152**, 25 (1968).
45. S. Helgason, *Differential Geometry and Symmetric Spaces*, Academic Press, New York, 1962 (русский перевод: С. Хельгасон, Дифференциальная геометрия и симметрические пространства, изд-во «Мир», М., 1964).
46. L. Infeld, B. L. van der Waerden, *Sitzungsber. Preuss. Akad. Wiss.*, **9**, 380 (1933).
47. P. Jordan, J. Ehlers, R. K. Sachs, *Akad. Wiss.*, Mainz, **1** (1961).
- 47a. R. Kantowski, R. K. Sachs, *J. Math. Phys.*, **7**, 443 (1966).
48. R. P. Kerr, *Phys. Rev. Lett.*, **11**, 237 (1963).
49. A. Komar, *Phys. Rev.*, **104**, 544 (1963).
50. E. Kronheimer, R. Penrose, *Proc. Camb. Phil. Soc.*, **63**, 481 (1967).

- 50а. M. D. Kruskal, Phys. Rev., **119**, 1943 (1960).
51. N. H. Kuiper, Ann. Math., **50**, 916 (1949).
52. W. Kundt, Z. Phys., **172**, 488 (1963).
53. Е. М. Лифшиц, И. М. Халатников, УФН, **30**, 391 (1963).
54. R. W. Lindquist, J. A. Wheeler, Rev. Mod. Phys., **29**, 432 (1957).
- 54а. S. MacLane, G. Birkhoff, Algebra, Macmillan, New York, 1967.
55. L. Markus, Ann. Math., **62**, 411 (1955).
56. R. F. Marzke, J. A. Wheeler, in «Gravitation and Relativity», ed. H. Y. Chiu and W. F. Hoffmann, Benjamin, New York, 1964 (русский перевод: Р. Марцке, Дж. Уиллер, в сб. «Гравитация и относительность», изд-во «Мир», М., 1965, стр. 107.)
57. J. W. Milnor, Enseignement Math., (2) **9**, 198 (1963).
58. J. W. Milnor, Morse Theory, Annals of Math. Studies, Princeton University Press, Princeton, 1963 (русский перевод: Дж. Милнор, Теория Морса, изд-во «Мир», М., 1965).
59. C. W. Misner, J. Math. Phys., **4**, 924 (1963).
60. C. W. Misner, Astrophys. J., **151**, 431 (1968).
61. Ч. В. Мизнер, А. Х. Тауб, ЖЭТФ, **55**, 233 (1968).
62. C. Møller, The Theory of Relativity, Oxford University Press, Oxford, 1952 (русский перевод: Кр. Мёллер, Теория относительности, изд-во «Наука», М., 1972).
63. E. T. Newman, L. Tamburino, T. Unti, J. Math. Phys., **4**, 915 (1963).
64. E. T. Newman, R. Penrose, J. Math. Phys., **3**, 566 (1962); **4**, 998 (1963).
65. K. Nomizu, Lie Groups and Differential Geometry, Herald Printing Co. Ltd., Tokyo, 1956 (русский перевод: К. Номидзу, Группы Ли и дифференциальная геометрия, ИЛ, М., 1960).
66. G. Nordström, Ann. Phys., **42**, 533 (1913).
- 66а. И. Д. Новиков, Письма ЖЭТФ, **3**, 223 (1966).
67. J. R. Oppenheimer, G. Volkoff, Phys. Rev., **55**, 374 (1939).
68. J. R. Oppenheimer, H. Snyder, Phys. Rev., **56**, 455 (1939).
69. J. P. Ostriker, P. Bodenheimer, D. Lynden-Bell, Phys. Rev. Lett., **17**, 816 (1966).
70. W. T. Payne, Am. J. Phys., **20**, 253 (1952).
71. R. Penrose, Proc. Camb. Phil. Soc., **55**, 137 (1959).
72. R. Penrose, Ann. Phys., **10**, 171 (1960).
73. R. Penrose, in P. G. Bergmann's Aeronautical Research Lab. Tech. Documentary Rept. 63-56: Quantization of Generally Covariant Fields, Office of Aerospace Research, U. S. Air Force, 1963.
74. R. Penrose, in «Relativity, Groups, and Topology», ed. C. M. DeWitt and B. S. DeWitt, Gordon and Breach, New York, 1964.

75. R. Penrose, *Phys. Rev. Lett.*, **14**, 57 (1965).
76. R. Penrose, *Proc. Roy. Soc. (Lond.)*, **A284**, 159 (1965).
77. R. Penrose, *Rev. Mod. Phys.*, **37**, 215 (1965).
78. R. Penrose, in «*Perspectives in Geometry and Relativity*», Indiana University Press, Bloomington, 1966.
79. R. Penrose, *J. Math. Phys.*, **8**, 345 (1967).
80. R. Penrose, *Intern. J. Theor. Phys.*, **1**, 61 (1968).
- 80a. A. Penzias, R. Wilson, *Astrophys. J.*, **142**, 419 (1965).
- 80b. А. З. Петров, Уч. зап. Каз. ун-та, **114**, кн. 8, 55 (1954).
81. F. A. E. Pirani, *Phys. Rev.*, **105**, 1089 (1957) (русский перевод: Ф. Пирани, в сб. «Новейшие проблемы гравитации», ИЛ, М., 1961, стр. 257).
82. F. A. E. Pirani, in «*Lectures on General Relativity: 1964 Brandeis Summer Institute in Theoretical Physics*», Vol. I, Prentice-Hall, Englewood Cliffs, 1965.
83. F. A. E. Pirani, A. Schild, in «*Perspectives in Geometry and Relativity*», Indiana University Press, Bloomington, 1966.
84. J. Plebanski, *Acta Phys. Polon.*, **27**, 361 (1965).
85. R. V. Pound, G. A. Rebka, *Phys. Rev. Lett.*, **4**, 337 (1960) (русский перевод: Р. Паунд, Г. Ребка, в сб. «Новейшие проблемы гравитации», ИЛ, М., 1961, стр. 474).
86. A. K. Raychaudhuri, *Phys. Rev.*, **98**, 1123 (1955).
87. S. Refsdal, *Mon. Not. Roy. Astron. Soc.*, **128**, 295 (1964).
88. W. Rindler, *Mon. Not. Roy. Astron. Soc.*, **116**, 6 (1956).
89. I. Robinson, *J. Math. Phys.*, **2**, 290 (1961).
90. I. Robinson, A. Schild, E. L. Schucking, eds., *Quasi-Stellar Sources and Gravitational Collapse*, University of Chicago Press, Chicago, 1965.
91. H. S. Ruse, *Proc. Lond. Math. Soc.*, **50**, 75 (1948).
92. H. Rudberg, *The Compactification of a Lorentz Space*, Thesis, University of Uppsala, Sweden, 1958.
93. R. K. Sachs, *Proc. Roy. Soc. (Lond.)*, **A264**, 309 (1961).
94. R. K. Sachs, in «*Recent Developments in General Relativity*», PWN/Pergamon, Warszawa/New York, 1962.
95. R. K. Sachs, *Proc. Roy. Soc. (Lond.)*, **A270**, 103 (1962).
96. A. Schild, *Am. J. Phys.*, **28**, 778 (1960).
- 96a. J. A. Schouten, *Ricci-Calculus*, Springer, Berlin, 1954.
97. E. Schrodinger, *Expanding Universes*, Cambridge University Press, Cambridge, 1956.
98. H. J. Seifert, *Z. Naturforsch.*, **22a**, 1356 (1967).
99. L. C. Shepley, *Proc. Nat. Acad. Sci.*, **52**, 1403 (1964).
100. B. Spain, *Tensor Calculus*, Oliver and Boyd, London, 1960.
101. N. Steenrod, *The Topology of Fibre Bundles*, Princeton Univ. Press, Princeton, 1951 (русский перевод: Н. Стинрод, Топология косых произведений, ИЛ, М., 1953).
102. K. Stellmacher, *Math. Ann.*, **123**, 34 (1951).
103. J. L. Synge, *Relativity: the General Theory*, North-Holland, Amsterdam, 1960 (русский перевод: Дж. Синг, Общая теория относительности, ИЛ, М., 1963).
104. L. A. Tamburino, J. Winicour, *Phys. Rev.*, **150**, 1039 (1966).

105. A. H. Taub, *Ann. Math.*, **53**, 472 (1951).
106. J. Terrel, *Phys. Rev.*, **116**, 1041 (1959).
107. K. S. Thorne, *Relativistic Stellar Structure and Dynamics* in «1966, Les Houches Lectures», ed. C. M. DeWitt, Gordon and Breach, New York, 1967.
108. A. Trautman, in «Lectures on General Relativity; 1964 Brandeis Summer Institute in Theoretical Physics», Vol. I, Prentice-Hall, Englewood Cliffs, 1965.
109. A. Trautman, in «Perspectives in Geometry and Relativity», Indiana University Press, Bloomington, 1966.
110. B. L. van der Waerden, *Nachr. Ges. Wiss. Goettingen*, **100** (1929).
111. H. Weyl, *Nachr. Ges. Wiss. Goettingen*, **99** (1921).
112. J. A. Wheeler, *Geometrodynamics*, Academic Press, New York, 1962 (русский перевод: Дж Уилер, Гравитация, нейтринно и Вселенная, ИЛ, М., 1962).
113. J. A. Wheeler, in «Relativity, Groups, and Topology», ed. C. M. DeWitt and B. S. DeWitt, Gordon and Breach, New York, 1964.
- 113a. H. Whitney, *Bull. Am. Math. Soc.*, **43**, 785 (1937).
114. E. T. Whittaker, *Proc. Roy. Soc. (Lond.)*, **A158**, 38 (1937).
115. L. Witten, *Phys. Rev.*, **113**, 357 (1959).
- 116*. A. Lichnerowicz, *Topics on Space-Time*, in «Battelle Rencontres; 1967 Lectures in Mathematics and Physics», ed. C. M. DeWitt and J. A. Wheeler, W. A. Benjamin, Inc., New York — Amsterdam, 1968, Ch. V, p. 115.
- 117*. R. Bott, J. Mather, *Topics in Topology and Differential Geometry*, там же, Ch. XVIII, p. 460.
- 118*. R. Geroch, *The Structure of Singularities*, там же, Ch. VIII.
- 119*. Ch. Misner, *Relativistic Fluids in Cosmology*, там же, Ch. VI.
- 120*. R. A. Isaacson, *Phys. Rev.*, **166**, 1263, 1272 (1968).
- 121*. K. S. Thorne, *Astrophys. J.*, **158**, 1; 997 (1969).
- 122*. W. L. Burke, *Phys. Rev.*, **A2**, 1501 (1970).
- 123*. S. Chandrasekhar, F. P. Esposito, *Astrophys. J.*, **160**, 153 (1970).
- 124*. Я. Б. Зельдович, И. Д. Новиков, *Теория тяготения и эволюция звезд*, изд-во «Наука», М., 1971.
- 125*. В. А. Белинский, И. М. Халатников, *ЖЭТФ*, **56**, 1700 (1969).
- 126*. И. М. Халатников, Е. М. Лифшиц, *Phys. Rev. Lett.*, **24**, 76 (1970).
- 127*. C. W. Misner, *Phys. Rev. Lett.*, **22**, 1071 (1969).
- 128*. В. А. Белинский, Е. М. Лифшиц, И. М. Халатников, *УФН*, **102**, 463 (1970).
- 129*. L. C. Shepley, *Phys. Lett.*, **A28**, 695 (1969).
- 130*. Л. П. Гришук, *ЖЭТФ*, **51**, 475 (1966); **53**, 1699 (1967).
- 131*. S. W. Hawking, R. Penrose, *Proc. Roy. Soc. London*, **A34**, 29 (1970).
- 132*. R. Penrose, *Techniques of differential topology in general relativity*, Pittsburgh notes, S. I. A. M., 1972.

Послесловие автора к русскому изданию

За пять лет, прошедших после написания этих заметок, проведены дальнейшие исследования по теории сингулярностей и достигнут существенный прогресс в понимании «черных дыр» (объектов, находящихся на поздних стадиях катастрофического коллапса). Эти исследования подтверждают уже известные результаты и опираются на них, давая нам значительно более полные представления по некоторым вопросам. Следует упомянуть, в частности, работы Лифшица, Халатникова и Белинского [Phys. Rev., **24**, 76 (1970); Успехи физ. наук, **102**, 463 (1970)], которые впервые построили общее решение с пространственно-временной сингулярностью и устранили тем самым все существовавшие противоречия между аналитическим и топологическим подходами к проблеме. Топологические методы, изложенные в этой книге, привели Хокинга и автора к другой теореме сингулярности [S. Hawking, R. Penrose, Proc. Roy. Soc. London, A314, 529 (1970); см. также R. Penrose, Techniques of Differential Topology in Relativity, S. I. A. M., 1972], которая оказалась весьма эффективной для объединения и улучшения описанных здесь теорем. Особого внимания заслуживает также метод В-границ Шмидта для изучения структуры сингулярностей (B. G. Schmidt, A new definition of singularities in general relativity, Universität Hamburg, 1970, preprint).

На развитие теории черных дыр в последние пять лет важное влияние оказали результаты, полученные Израелем [W. Israel, Phys. Rev., **164**, 1776 (1967)], Картером [B. Carter, Phys. Rev. Lett., **26**, 331 (1971)], Хокингом [S. Hawking, Comm. Math. Phys., **25**, 546 (1972)] и другими, вследствие того, что, как оказалось,

в наиболее общем случае черная дыра описывается метрикой Керра (приведенной в этих заметках), а это ведет к значительно более полной картине явления, чем мы вправе были ожидать.

Рассматриваемая тема все еще быстро развивается, так как она стимулируется замечательными астрономическими наблюдениями таких явлений, как квазары, пульсары, галактические взрывы и реликтовое излучение. Но главным стимулом является возможное открытие Вебером гравитационных волн, идущих из галактического центра. Я почти не сомневаюсь, что методы, аналогичные тем, которые описаны в этих лекциях, будут тем более полезны, чем яснее мы будем сознавать, что сильные гравитационные поля играют важную роль в астрофизических процессах.

Поэтому мне очень приятно, что Н. Мицкевич и Л. Грищук, переведя эти заметки на русский язык, тем самым сделали их более доступными советским физикам и математикам, работающим в этой области.

Р. Пенроуз

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и др. просим присылать по адресу:

129820, Москва, ГСП, И-110, 1-й Рижский пер., д. 2,
изд-во «Мир».

Р. ПЕНРОУЗ

Структура пространства-времени

Редактор *Р. Золина и М. Рутковская*

Художник *С. Бычков*. Художественный редактор *В. Варлашин*
Технический редактор *Е. Потопенкова*. Корректор *Е. Кочегарова*

Сдано в набор 13/XII 1971 г. Подписано к печати 4/VII 1972 г.
Бумага № 2 84×108¹/₃₂—2,88 бум. л. Печ. усл. л. 9,66. Уч.-изд. л. 8,40
Изд. № 27/6429. Цена 84 коп. Зак. № 1379

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени
Ленинградская типография № 2 имени Евгении Соколовой
Главполиграфпрома Комитета по печати при Совете Министров
СССР. Измайловский проспект, 29

