

*А. Д. Долгов  
Я. Б. Зельдович  
М. В. Сажин*

***КОСМОЛОГИЯ  
РАННЕЙ  
ВСЕЛЕННОЙ***

*ИЗДАТЕЛЬСТВО  
МОСКОВСКОГО УНИВЕРСИТЕТА  
1988*

Долгов А. Д. и др. Космология ранней Вселенной / А. Д. Долгов, Я. Б. Зельдович, М. В. Сажин. — М.: Изд-во Моск. ун-та, 1988. — 199 с. — ISBN 5—211—00108—7.

Первая в СССР монография, посвященная бурно развивающейся сейчас теории ранней Вселенной: рождению и развитию Вселенной с  $10^{-43}$  с до приблизительно  $10^{-33}$  с. Основные идеи, позволяющие рассматривать столь раннюю эпоху, лежат на стыке космологии и теории элементарных частиц высоких энергий (вплоть до  $10^{19}$  ГэВ). Впервые систематизированы достижения современной космологии, включая вопросы квантового рождения Вселенной, инфляционной стадии, бариосинтеза, теорий великого объединения и их связи с космологией. Избегаются выкладки, способные заслонить физическую суть рассматриваемых идей.

Для научных работников, аспирантов и студентов, интересующихся космологией.

Ил. 54. Библиогр.: 74 назв.

Рецензенты: канд. физ.-мат. наук В. А. Кузьмин,  
член-кор. АН СССР Р. А. Сюняев

*Печатается по постановлению  
Редакционно-издательского совета  
Московского университета*

Д 1705070000—058 108—88  
077(02)—88

ISBN 5—211—00108—7

© Издательство Московского университета, 1988 г.

## ОГЛАВЛЕНИЕ

Введение	5
<b>Глава 1. Стандартная космологическая модель</b>	8
§ 1. Расширение Вселенной. Постоянная Хаббла	8
§ 2. Реликтовое излучение	9
§ 3. Плотность материи во Вселенной	10
§ 4. Скрытая масса	11
§ 5. Неоднородности	12
§ 6. Тепловая история ранней Вселенной	13
§ 7. Закалка элементарных частиц. Реликтовые частицы	16
§ 8. Нуклеосинтез	19
<b>Глава 2. Трудности классической космологии</b>	23
§ 1. Закон расширения и уравнение состояния	23
§ 2. Проблема горизонта	25
§ 3. Проблема плоскостности	27
§ 4. Инфляционная теория. Первое знакомство	29
§ 5. Проблема космологической постоянной	30
§ 6. Барionная асимметрия Вселенной	31
<b>Глава 3. Теория поля. Краткое введение</b>	32
§ 1. Поля в специальной теории относительности	32
§ 2. Векторное поле. Электродинамика	37
§ 3. Тензорное поле. Теория гравитации	44
§ 4. Безмассовые и массивные поля. Дальнодействующие силы	50
<b>Глава 4. Элементарные частицы</b>	53
§ 1. Скалярные поля. $\pi$ -мезоны	53
§ 2. Кварки. Строение адронов	55
§ 3. Взаимодействие кварков. Квантовая хромодинамика	56
§ 4. Кварк-лептонная симметрия. Ограничение на число типов лептонов и кварков	61
§ 5. Электрослабые взаимодействия. Промежуточные бозоны	61
§ 6. Теории большого объединения. Несохранение барионов	70
§ 7. Высшие порядки теории возмущений. Перенормировка	72
§ 8. Спонтанное нарушение симметрии. Скалярное (хиггсовское) поле	74
<b>Глава 5. Космологическая постоянная</b>	78
§ 1. Определение и физический смысл	78
§ 2. Решение де Ситтера	80
§ 3. Замечания о соотношении $p = -\epsilon$	93
§ 4. Проблема $\Lambda$ -члена в поздней Вселенной	94
<b>Глава 6. Скалярные поля в космологии</b>	100
§ 1. Скалярное поле в плоском пространстве-времени	100
§ 2. Скалярное поле в расширяющейся Вселенной	106
§ 3. Хиггсово поле и инфляция	112
§ 4. Окончание инфляции. Повторный разогрев	116
<b>Глава 7. Поляризация вакуума и инфляция</b>	118
§ 1. Поляризация вакуума в электродинамике	118
§ 2. Поляризация вакуума в искривленном пространстве	123
<b>Глава 8. Бариосинтез во Вселенной</b>	126
§ 1. Несохранение барионов. Распад протона	127
§ 2. Нестационарность Вселенной и зарядовая асимметрия	130
§ 3. Бариосинтез и сохраняющиеся заряды	133
§ 4. Нарушение зарядовой инвариантности	135
§ 5. Барionная асимметрия в модели большого объединения. Топологическое несохранение барионов	136
§ 6. Барionная асимметрия и черные дыры	137

Глава 9. Квантовое рождение Вселенной	139
§ 1. Рождение из «ничего» и закон сохранения энергии	139
§ 2. Свойства полузамкнутого мира	143
§ 3. Масса вещества родившегося мира, ее увеличение в процессе инфляции	148
§ 4. Возникновение хэббловского расширения	151
§ 5. Квантовое описание рождения Вселенной	153
§ 6. Хаотическая инфляция и вечная Вселенная	157
Глава 10. Структура Вселенной	160
§ 1. Начальная стадия роста возмущений	161
§ 2. Возмущение в плазме	163
§ 3. Замечание о спектре флуктуаций	164
§ 4. Структура Вселенной в адиабатической теории	167
§ 5. Нейтринная Вселенная	171
§ 6. Блины, ячеистая структура	173
§ 7. Структура Вселенной в моделях со скрытой массой	176
§ 8. Изотермические флуктуации	179
§ 9. Флуктуации реликтового излучения	180
Глава 11. Квантовая теория и происхождение флуктуаций	187
§ 1. Флуктуации скалярного поля	187
§ 2. Гравитационные волны	190
§ 3. Изотермические возмущения	194
§ 4. Струны и первичные возмущения	195
Литература	198

Развитие современной космологии идет сейчас в основном в двух направлениях. Одно из них включает в себя детальное изучение структуры Вселенной, исследование реликтового излучения, попытки прямого обнаружения скрытой массы Вселенной. На этом пути используются как новые теоретические идеи, так и наблюдательные данные, особенно обильно полученные в течение последнего десятилетия.

Другое направление в космологии — изучение все более и более ранних моментов в истории развития нашей Вселенной. Эта область космологии популярна не только среди космологов, но и среди физиков, занимающихся проблемами элементарных частиц и фундаментальных взаимодействий.

Дело в том, что ранние этапы эволюции Вселенной — это та естественная лаборатория, в которой были реализованы экстремальные условия, недоступные сегодняшнему эксперименту. Отодвигаясь все дальше и дальше в прошлое, мы в соответствии со стандартной космологической моделью приходим к столь высоким значениям плотности, температуры, средней энергии частиц первичной плазмы, что нет никакой надежды достичь их в земных условиях в обозримом будущем. Поэтому нельзя считать, что теория ранней Вселенной опирается на хорошо обоснованный экспериментальный базис. Напротив, логика сегодня в существенной степени переменялась: современные теории элементарных частиц, модели объединения их взаимодействий обязательно проходят проверку на согласие с космологией. В этом смысле о ранней Вселенной иногда говорят как об «ускорителе бедного человека», имея в виду, что ее изучение позволяет получить информацию о процессах взаимодействия элементарных частиц при таких энергиях, что необходимые для их достижения ускорители находятся вне рамок материальных возможностей человечества.

В настоящее время возник весьма плодотворный симбиоз между космологией, с одной стороны, и физикой элементарных частиц и квантовой теорией поля — с другой.

Космологическая проверка теорий элементарных частиц при сверхвысоких энергиях является чуть ли не единственным способом сравнения их с физической реальностью, а использование результатов физики частиц и методов квантовой теории поля в космологии дало возможность ответить на ряд вопросов о Вселенной, которые совсем недавно казались находящимися вне пределов нашего познания. Речь идет о таких свойствах Вселенной в целом, как однородность и изотропия,

о величине средней плотности энергии во Вселенной и ее близости к критическому значению  $\rho_c$ , отвечающему трехмерно плоскому миру. Фундаментальные физические законы позволяют объяснить происхождение и спектр флуктуаций плотности на ранних этапах, которые за счет гравитационной неустойчивости развиваются впоследствии в галактики и их скопления. Еще несколько лет назад все эти свойства постулировались как начальные условия космологической модели, причем эти постулаты выглядели весьма неестественно в рамках тогдашнего понимания (или, вернее, непонимания). Теперь же мы видим, что они являются следствием весьма общей динамики, приводящей на каком-то этапе в прошлом к экспоненциальному закону расширения Вселенной, т. е. к модели де Ситтера. Космологическую модель такого рода называют инфляционной моделью, или моделью раздувающейся Вселенной.

Сегодня у нас даже есть надежда получить ответ на вопрос о сотворении мира, математически описать процесс его рождения и понять причину первоначального толчка, приведшего к наблюдаемому расширению Вселенной. Согласно современной теории этот процесс выглядит как квантовый скачок, аналогичный туннельному переходу в  $\alpha$ -распаде, от состояния с квантовым пространством-временем к классическому пространству-времени. Есть надежда на этом пути понять, почему наш мир четырехмерен. Для квантовой теории поля предпочтительной размерностью является  $D=10$ . Точнее говоря, это утверждение относится к теории суперструн, с помощью которой мы надеемся получить непротиворечивую единую теорию всех взаимодействий. Возможно, именно эта теория объясняет, почему дополнительные шесть измерений являются компактными и ненаблюдаемыми при достижимых в лаборатории энергиях. Пока еще не все проблемы решены, но открывающиеся перспективы вселяют самые радужные надежды.

Настоящая книга написана на основе лекций, прочитанных Я. Б. Зельдовичем на физическом факультете МГУ. Она предназначена главным образом для астрономов, в том числе и для астрономов-наблюдателей. Книга может быть интересна также физикам других специальностей, желающим познакомиться с современной космологией. Мы предполагаем, что читатель знаком с основами специальной теории относительности. Знания общей теории относительности при чтении этой книги не требуется, хотя и желательно. Мы считаем, что читатель обладает достаточным математическим багажом (в пределах трех курсов физического факультета) и знанием основ квантовой механики, однако нуждается в сведениях о физике элементарных частиц, которые здесь популярно приводятся. Прекрасным введением в теорию элементарных частиц могут служить книги Л. Б. Окуня « $\alpha, \beta, \gamma, \dots, Z$ » (1985) и «Физика элемен-

тарных частиц» (1984). Для желающих изучить так называемую стандартную космологию можно рекомендовать замечательную популярную книгу С. Вайнберга «Первые три минуты» (1981), а также более специальные книги Я. Б. Зельдовича и И. Д. Новикова «Строение и эволюция Вселенной» (1973) или С. Вайнберга «Гравитация и космология» (1975). Однако в них не отражено бурное развитие космологии в последние годы. Мы надеемся, что предлагаемая книга поможет получить представление об этой захватывающей области.

Помимо нескольких основополагающих работ, мы даем лишь указания на обзоры и книги, которые помогут читателю более глубоко изучить материал и ознакомиться с оригинальными статьями по данной теме. Особенно это относится к физике частиц. Всю необходимую литературу по этим вопросам можно найти в упомянутых выше книгах. Что касается космологии, которой главным образом посвящена наша книга, то мы приводим ссылки лишь на относительно недавние работы, которые не отражены в существующих стандартных учебниках и которые привели к возникновению новых представлений о самых ранних стадиях развития Вселенной.

Для читателя, совсем незнакомого с физикой частиц, некоторые термины в первых главах могут быть непонятны, но мы надеемся, что он получит ответы на свои вопросы, прочитав гл. 4. Вообще различные главы относительно независимы, но тем не менее читать книгу лучше последовательно. В книге используется, возможно непривычная для астронома система единиц, в которой  $\hbar=c=k=1$  (см., например, книгу [25]).

Поскольку предмет книги развивается очень быстро, следует указать, что работа над ней была завершена в июле 1986 г. (гл. 1–3, 5, 6, 9 написаны М. В. Сажиним, гл. 4, 7, 8 А. Д. Долговым, гл. 10, 11 — совместно). Поэтому в книге, в частности, не нашли отражения последние достижения теории суперструн.

## Глава 1.

## СТАНДАРТНАЯ КОСМОЛОГИЧЕСКАЯ МОДЕЛЬ

Под термином, использованным в заголовке, мы понимаем модель горячей Вселенной, или модель Большого взрыва, которая стала общепринятой после открытия реликтового электромагнитного фона. Динамика модели хорошо известна и описывается уравнениями общей теории относительности (ОТО), однако начальные условия, приводящие к наблюдаемому миру, представляются загадочными. Новейшее развитие космологии состояло как раз в выяснении смысла и физических причин этих начальных условий, но не отметало справедливости старой космологии на более позднем этапе.

## § 1. РАСШИРЕНИЕ ВСЕЛЕННОЙ. ПОСТОЯННАЯ ХЭББЛА

В основе современной космологии лежит решение уравнений общей теории относительности для случая изотропного и однородного распределения вещества, найденное Фридманом свыше 60 лет тому назад. Предсказанное в этом решении расширение Вселенной было обнаружено через несколько лет Хабблом (1929) и носит теперь его имя. Закон расширения имеет весьма простой вид:

$$v = Hr, \quad (1.1)$$

где  $v$  — скорость удаления от нас астрономического объекта, находящегося на расстоянии  $r$ . Коэффициент пропорциональности носит название постоянной Хаббла и согласно современным измерениям равен

$$H = (50 \div 100) \text{ км/с Мпк.}$$

Заметим, что название «постоянная Хаббла» не очень удачно. Оно подчеркивает тот факт, что  $H$  не зависит от расстояния и от направления на небесный объект, однако  $H$  является функцией времени. Грубо говоря, величина  $H^{-1} = (20 \div 10) \cdot 10^9$  лет равна возрасту Вселенной, отсчитываемому от момента сингулярности. В более ранний период «постоянная» Хаббла была соответственно больше и скорость расширения Вселенной выше.

Согласно наблюдениям самого Хаббла значение  $H$  составляет примерно 500 км/с Мпк, что отвечало бы слишком малому значению возраста Вселенной  $t_U \approx 2 \cdot 10^9$  лет. Изотопные методы, а также оценки возраста старых звездных скоплений дают

$t_U = (10 \div 20) \cdot 10^9$  лет. Однако за полвека, прошедшие с момента открытия Хаббла, точность астрономических наблюдений существенно возросла, и принятое сейчас значение  $H$  не противоречит другим данным о возрасте мира. Остающаяся довольно большая неопределенность в величине  $H$  связана с трудностями в определении шкалы галактических расстояний и едва ли может быть ликвидирована в ближайшее время.

Заметим, что закон расширения (1.1) не выделяет во Вселенной никакой особой точки. Перенос начала координат из одной точки в другую и одновременное галилеево преобразование оставляют закон Хаббла неизменным и в новой системе координат. Это находится в соответствии с фактом однородности Вселенной.

Расширение Вселенной сейчас уже не вызывает сомнений, тем более что оно описывается одним из возможных решений уравнений Эйнштейна. Однако физические причины, которые привели к реализации в природе именно этого решения, в рамках стандартной космологии оставались загадочными. Стандартная модель Фридмана не могла ответить на вопрос о механизме первоначального толчка, вызвавшего разлетание галактик, несмотря на противодействие сил притяжения. К ответу на этот вопрос удалось приблизиться лишь в последние годы благодаря применению результатов, полученных в физике элементарных частиц, к теории ранней Вселенной. Забегая вперед, отметим, что теория элементарных частиц предсказывает возможность уравнения состояния с отрицательным давлением, о котором мы много будем говорить ниже. В релятивистской теории тяготения вещество с таким уравнением состояния фактически антигравитирует, что и приводит к расширению мира.

## § 2. РЕЛИКТОВОЕ ИЗЛУЧЕНИЕ

Важнейшим космологическим фактом после обнаружения Хабблом расширения Вселенной явилось наблюдение в 1965 г. (Пензиас и Вильсон) реликтового электромагнитного излучения. Гипотезу о существовании электромагнитного фона, оставшегося от ранней горячей стадии развития Вселенной, высказал в 1946 г. Гамов. Точнее говоря, им двигало желание объяснить нуклеосинтез процессами в горячем веществе в первые минуты от начала. Побочным результатом предложенной модели было как раз реликтовое излучение. Хотя Гамов исходил из неверных предположений, считая, что все ядра, в том числе и тяжелые, рождаются при Большом взрыве\*, сделанная им

\* Сейчас установлено, что в первые минуты возникают практически лишь ядра не тяжелее  $\text{Li}^7$ , а более тяжелые элементы образуются в звездах.

оценка температуры реликтового излучения (6К) довольно близка к измеренной позднее величине. Наблюдения подтвердили планковскую форму спектра электромагнитного фона в диапазоне волн от десятков сантиметров ( $\lambda_{\max}=73$  см) до миллиметра, причем температура оказалась около 3 К в неплохом согласии с первой оценкой. Наблюдения показали также высокую степень изотропии реликтового излучения. За вычетом тривиальной дипольной анизотропии, связанной с движением наблюдателя относительно космологического фона, никаких других отклонений от изотропии на уровне порядка нескольких единиц на  $10^{-5}$  обнаружено не было.

### § 3. ПЛОТНОСТЬ МАТЕРИИ ВО ВСЕЛЕННОЙ

В отличие от очень хорошо изученного реликтового излучения важная космологическая величина — средняя плотность материи во Вселенной  $\rho$  известна плохо. Для характеристики величины  $\rho$  используют безразмерный параметр  $\Omega$ :

$$\Omega = \rho/\rho_c = \rho/(3H^2/8\pi G), \quad (1.2)$$

где  $G$  — гравитационная постоянная,  $H$  — постоянная Хаббла, а  $\rho_c$  — так называемая критическая плотность. При  $\Omega > 1$  ( $\rho > \rho_c$ ) Вселенная трехмерно замкнута и наблюдаемое сейчас расширение должно смениться сжатием, этот вывод несправедлив для Вселенной с отличным от нуля космологическим членом (см. гл. 5). При  $\Omega \leq 1$  Вселенная открыта и расширение будет вечным. Случай  $\rho = \rho_c$  отвечает Вселенной в среднем пространственно плоской, с обычной евклидовой геометрией. Подчеркнем, что речь идет о геометрии трехмерного пространства, определяемого сечением  $t = \text{const}$  4-мерного многообразия пространства-времени. Это четырехмерное многообразие отнюдь не является плоским. Заметим еще, что в выборе временной координаты существует известный произвол и в данном случае речь идет о так называемом сопутствующем времени, которое определяется условием  $\rho = \text{const}$  при постоянном  $t$  (т. е. как нормаль к гиперповерхности  $\rho = \text{const}$ ).

Согласно астрономическим наблюдениям величина  $\Omega$  заключена в пределах

$$0,03 < \Omega < 2.$$

Нижний предел отвечает видимой, непосредственно наблюдаемой материи. Для этой величины иногда используют обозначение  $\Omega_b$ , где индекс «b» указывает, что  $\Omega_b$  связана с барионной, т. е. с обычной, материей, состоящей из протонов и нейтронов, объединенных общим названием барионы (см. ниже).

Возможно, часть барионного вещества по той или иной

причине не видна. Например, она находится в форме ионизованного газа или темных звезд. Но даже с учетом этого  $\Omega_b$  едва ли превосходит 0,2, а аргументы, связанные с первичным нуклеосинтезом, приводят к ограничению на  $\Omega_b$  вблизи нижнего предела астрономических наблюдений:  $\Omega_b \approx 0,03$ .

### § 4. СКРЫТАЯ МАССА

Сейчас можно считать твердо установленным, что основная масса вещества во Вселенной находится в какой-то необычной, не барионной форме. Непосредственно наблюдать эти формы материи невозможно, она невидима, однако по ее гравитационным проявлениям можно заключить, что отвечающий ей параметр  $\Omega$  больше или порядка 0,3. Об этих формах материи говорят как о скрытой массе Вселенной. Различают две формы скоплений и ее пространственное распределение следует за распределением видимого вещества. Для нее  $\Omega \approx 0,3$ . Нельзя исключить, что помимо кластеризованной, неоднородной распределенной невидимой материи, во Вселенной существует однородный фон невидимого вещества. Последний весьма трудно изучать в астрономических наблюдениях, и точность здесь очень низка. Однако мы делаем гипотезу о его существовании, так как инфляционная модель, на которой основано решение важнейших космологических проблем (см. ниже), предсказывает  $\Omega = 1$ , и, таким образом, необходимо иметь  $\Omega_{\text{фон}} \approx 0,7$ .

Кажется совершенно удивительным, что Вселенная более чем на 90% по массе состоит из неизвестной нам формы материи. Однако этот вывод, по-видимому, неизбежен. Даже если бы удалось найти какой-то физический механизм (что весьма маловероятно), который бы объяснил, почему подавляющая часть обычного вещества невидима, имеются независимые аргументы, говорящие о том, что вся масса Вселенной не может быть связана с барионами. Во-первых, это первичный нуклеосинтез, который требует  $\Omega_b \approx 0,03$  (см. § 8), и, во-вторых, это теория образования структуры Вселенной. Мы обсудим ее в гл. 10, а сейчас отметим только, что высокая изотропия спектра реликтового излучения говорит о том, что неоднородности в плотности барионного вещества на ранней стадии должны быть весьма малы и поэтому галактики и их скопления не смогли бы за имеющееся время развиться из этих неоднородностей. Положение могла бы спасти гравитирующая материя, не взаимодействующая с электромагнитным излучением, которая и составляет скрытую массу.

### § 5. НЕОДНОРОДНОСТИ

Наблюдения показывают, что в малых (в космологическом смысле) масштабах видимое вещество распределено весьма неравномерно. Однако в больших масштабах Вселенная однородна. Линейный размер, характеризующий переход от неоднородного распределения вещества к однородному, составляет примерно 200 Мпк. Другими словами, при усреднении плотности вещества внутри куба с ребром меньше чем 200 Мпк, например 2 Мпк, мы будем получать контрасты плотности порядка единицы. При усреднении плотности внутри куба с ребром более 200 Мпк мы будем получать однородное распределение вещества, не зависящее от места нахождения куба. Подчеркнем, что линейные размеры наблюдаемой части Вселенной составляют несколько гигапарсек, так что кубов с ребром в 200 Мпк в части Вселенной, охваченной наблюдениями, будет несколько тысяч. Такое количество испытаний дает, разумеется, статистически значимую величину. Поэтому однородность вещества в масштабах 200 Мпк и больше не является случайным.

Изучение реликтового излучения показывает, что в прошлом Вселенная была еще более однородна. Согласно последним наблюдениям реликтовое излучение изотропно с точностью лучше, чем  $3 \cdot 10^{-5}$  для углов от  $90^\circ$  до  $10'$  и с точностью  $\approx 10^{-2}$  для малых углов. Отсюда следует, что в эпоху рекомбинации водорода степень однородности в масштабах сегодняшнего дня порядка 10 Мпк была выше, чем  $\sim 10^{-4}$ , а в масштабах  $\sim 10$  пк — лучше, чем  $10^{-1}$ .

Данные о степени однородности в более ранние эпохи скуднее. Наблюдения космического дейтерия позволяют сделать вывод, что барионы были достаточно однородно распределены в эпоху первичного нуклеосинтеза ( $t \approx 100$  с), так как выход дейтерия очень чувствителен к концентрации барионов.

Неоднородность в ранние эпохи на очень больших масштабах (сильно превышающих размер горизонта в то время) исключается современными данными по крупномасштабной анизотропии реликтового излучения, по измерениям постоянной Хаббла и т. п.

Исходя из сказанного, мы полагаем, что ранняя Вселенная была с высокой степенью однородна как в малых, так и в больших ( $R \gg H^{-1}$ ) масштабах. В теории космологической эволюции в качестве нулевого приближения используют идеализированное представление об однородной и изотропной Вселенной и на фоне этой однородной эволюции изучают развитие неоднородностей, связанных с гравитационной неустойчивостью. Сразу подчеркнем, что какие-то начальные неоднородности необходимы, так как на фоне абсолютно однородной Вселенной образо-

вание наблюдаемой крупномасштабной структуры (галактик, их скоплений), невозможно. Эти небольшие неоднородности мало влияют на космологическую эволюцию на ранних этапах, но позже служат зародышами гравитационной кластеризации.

Происхождение и спектр этих неоднородностей до недавнего времени оставались загадочными. Естественная физическая гипотеза, что неоднородности обязаны своим происхождением квантовым или тепловым флуктуациям, предсказывала для них слишком малую величину. Однако в инфляционной модели положение кардинальным образом изменилось. Оказалось, что при экспоненциальном расширении квантовые флуктуации очень быстро нарастают из-за превращения коротковолновых квантовых мод при расширении в длинноволновые. В итоге неоднородности становятся даже слишком большими. Однако от этой трудности можно избавиться, если потребовать специальной малости входящих в теорию постоянных. Заметим также, что инфляционная модель предсказывает так называемый плоский спектр флуктуаций в соответствии с наблюдениями.

Теория развития гравитационной неустойчивости в рамках ньютоновского подхода была построена в начале этого столетия Джинсом (1902). Ее обобщение на случай фридмановской космологии с учетом релятивистских эффектов было сделано 40 лет назад Е. М. Лифшицем (1946). Инфляционная модель позволяет предсказать форму исходных флуктуаций плотности, которые необходимы в теории развития гравитационной неустойчивости как начальные значения. На основе этого теория в принципе уже сейчас объясняет возникновение структуры Вселенной, однако окончательная количественная модель пока еще не построена, в частности потому, что неизвестна физическая природа скрытой массы, играющей, очевидно, доминирующую роль в развитии структуры. Напомним, что согласно современному взглядам параметр  $\Omega = \rho/\rho_c$  близок к единице, в то время как вклад барионов в эту величину составляет  $\Omega_b = \rho_b/\rho_c \approx 0,03$ , т. е. невидимое неизвестное вещество составляет основную часть массы Вселенной.

### § 6. ТЕПЛОВАЯ ИСТОРИЯ РАННЕЙ ВСЕЛЕННОЙ

Обратимся теперь к эволюции однородной Вселенной на ранних этапах еще до развития структуры, но после инфляционного периода. Если двигаться вспять по времени, то при возрасте меньше, чем  $\sim 10^4$  лет от начала, в средней плотности энергии Вселенной будет доминировать релятивистское вещество (т. е. электромагнитное излучение, нейтрино и т. п.), для которого справедливо уравнение состояния  $p = \varepsilon/3$ . Для этой стадии можно написать простое приближенное соотношение,

связывающее температуру первичной плазмы  $T$  с возрастом Вселенной  $t$ :

$$T \text{ (МэВ)} \approx t^{-1/2} \text{ (с)}. \quad (1.3)$$

Его можно получить, если сравнить два выражения для плотности энергии. Первое — выражение для критической плотности (как мы увидим ниже, на стадии релятивистской доминантности с очень высокой степенью точности  $\rho \approx \rho_c$ ):

$$\rho \approx \rho_c = \frac{3H^2}{8\pi G} = \frac{3}{32\pi G t^2}, \quad (1.4)$$

где  $H = 1/2t$  — «постоянная» Хаббла при релятивистском законе расширения. Второе выражение — это выражение для плотности энергии равновесного газа с температурой  $T$ :

$$\rho = \frac{\pi^2}{30} Q T^4, \quad (1.5)$$

где  $Q$  — число степеней свободы различных частиц в газе (для фотонного газа  $Q=2$ , так как фотон имеет два спиновых состояния: каждая бозонная степень свободы дает вклад в  $Q$ , равный 1, а каждая фермионная дает вклад  $7/8$ ).

Справедливость формулы (1.5) предполагает весьма важную вещь, что первичный газ (или, может быть лучше говорить, плазма) находится в термодинамически равновесном состоянии. Можно убедиться в том, что это действительно так, потому что в общем случае темп расширения мира ниже, чем скорость установления равновесия в плазме. Чем ближе к «началу»,  $t \rightarrow 0$ , тем темп расширения мира выше. Он определяется постоянной Хаббла  $H = 1/(2t)$  (на ранней стадии доминирует, как полагают, релятивистское вещество и закон расширения имеет вид  $a(t) \sim \sqrt{t}$ ). Однако плотность числа частиц растет еще быстрее,  $n \sim T^3 \sim t^{-3/2}$ . Именно это обеспечивает условие  $\sigma n > H$  (где  $\sigma$  — сечение взаимодействия), что и приводит к быстрому установлению равновесия на ранних стадиях.

При  $t \approx 1$  с температура первичной плазмы достигала  $10^{10} \text{ К} \approx 1 \text{ МэВ}$ . При этом в ее составе должны быть, наряду с фотонами, нейтрино и антинейтрино, упомянутым выше невидимым веществом (если это не массивные нейтрино), также пары электронов и позитронов. Если предположить, что в какое-то мгновение пар  $e^+e^-$  не было, то при  $T \approx 1 \text{ МэВ}$  они должны очень быстро возникнуть за счет процессов обратной аннигиляции:  $\gamma + \gamma \rightarrow e^+ + e^-$ . При остывании плазмы преимущественно идет прямой аннигиляционный процесс  $e^+ + e^- \rightarrow 2\gamma$  и концентрация  $e^+$ ,  $e^-$  убывает по экспоненциальному закону,  $n \sim \exp(-m/T)$ .

Нейтрино и антинейтрино находятся в равновесном взаимо-

действию с первичной плазмой лишь при  $T > 3 \div 5 \text{ МэВ}$  ( $t \lesssim 0,1$  с), а при более низкой температуре их взаимодействием можно пренебречь. При расширении Вселенной они адиабатически теряют энергию, и в настоящее время температура пар  $\nu$ ,  $\bar{\nu}$  равна примерно 2 К (если  $m_\nu = 0$ ), а концентрация  $n_\nu \approx n_{\bar{\nu}} = 75 \text{ см}^{-3}$ .

Роль протонов и нейтронов при  $T \approx 1 \text{ МэВ}$  невелика, так как их количество составляет примерно  $10^{-9} \div 10^{-10}$  от количества легких частиц ( $e^\pm$ ,  $\nu$ ,  $\bar{\nu}$ ,  $\gamma$ ). Еще ближе к «началу» при  $t < 10^{-2}$  с в первичной плазме находятся  $\pi$ -мезоны и протон-антипротонные пары (а также нейтрон-антинейтронные), которые при  $T > 10^{12} \text{ К}$  и соответственно  $t < 10^{-4}$  с диссоциируют на свободные кварки, образуя так называемую кварк-глюонную плазму. Точнее, при остывании кварки сливаются в мезоны, протоны, антипротоны и т. п. Подробнее об этих и других упомянутых ниже частицах пойдет речь в следующих разделах. Мы хотели бы только указать, что по мере приближения к «началу» с ростом температуры в первичной плазме должны быть все новые частицы, масса которых удовлетворяет соотношению  $mc^2 < kT$ .

В частности, к моменту  $t \lesssim 10^{-10}$  с при  $T \gtrsim 10^{15} \text{ К}$  во Вселенной в обильном количестве находились переносчики слабого взаимодействия  $W^\pm$ - и  $Z^0$ -бозоны (сейчас для них предлагается термин вионы), масса которых примерно в 100 раз больше массы протона.

Существенным обстоятельством является то, что доминирующую роль в космологической плазме на ранних этапах играют релятивистские частицы, и в силу этого ее уравнение состояния имеет вид  $p = \epsilon/3$ . Это так называемая радиационно-доминированная плазма (РД-плазма). В стандартной космологической модели доминантность релятивистских частиц имела место, как уже упоминалось, при  $t \lesssim 10^{3+4}$  лет, когда плотность энергии реликтового электромагнитного излучения превосходила плотность энергии тяжелых частиц — нуклонов, и, что более существенно, невидимого вещества. В классической космологии предполагалось, что уравнение состояния релятивистского идеального газа справедливо и при  $t < 10^{-10}$  с вплоть до  $t \approx 10^{-43}$  с и  $T \approx m_{\text{Pl}} \sim 10^{19} \text{ ГэВ}$ . Справедливость подобной экстраполяции, впрочем, вызывает законные сомнения, так как надежная, подтвержденная экспериментом информация о свойствах частиц имеется сейчас лишь до энергии  $\sim 100 \text{ ГэВ}$ . Теория, правда, говорит нам, что взаимодействия частиц являются асимптотически свободными, т. е. ослабляются с ростом энергии, и поэтому приближение идеального газа оказывается справедливым. Впрочем, даже если взаимодействие не исчезает с ростом энергии, характер расширения в термодинамически равновесной ситуации не должен качественно измениться.



### § 7. ЗАКАЛКА ЭЛЕМЕНТАРНЫХ ЧАСТИЦ. РЕЛИКТОВЫЕ ЧАСТИЦЫ

Как мы отмечали выше, для первичной плазмы типично термодинамически равновесное состояние. Это означает, что все частицы с  $m \lesssim T$  присутствуют в плазме в равном количестве на каждую спиновую степень свободы (с точностью до коэффициента  $3/4$ , отличающего фермионы от бозонов в плотности числа частиц, и  $7/8$  — в плотности энергии). Что произойдет с этими частицами при расширении и падении температуры? С уменьшением плотности частиц скорость реакций между ними падает и в какой-то момент равновесие нарушается. Здесь возможны два случая. Первый — когда присущее частицам взаимодействие является весьма слабым, и они эффективно перестают взаимодействовать при  $T \gtrsim m$ . Пример такого рода дают нейтрино, которые выключаются при  $T = 3 - 5$  МэВ, а их масса не превышает 100 эВ. В силу этого их концентрация в настоящее время близка к концентрации реликтовых фотонов:  $N_{\nu_i} \simeq N_{\nu_j} \simeq (3/22) N_\gamma$ . Величина  $N_\gamma$  известна из наблюдений  $N_\gamma = 550 (T_\gamma/3 \text{ К})^3 \text{ см}^{-3}$ , где  $T_\gamma = (2,7 - 3) \text{ К}$  — современное значение температуры реликтового излучения. В силу этого ожидаемая концентрация реликтовых нейтрино составляет примерно  $75 \text{ см}^{-3}$  для каждого типа нейтрино.

Поясним происхождение коэффициента  $3/22$ . В термодинамическом равновесии следовало бы ожидать  $N_\nu/N_\gamma = (3/8)$  (предполагается, что у нейтрино лишь одно спиновое состояние, см. § 5 гл. 4). Однако уже при  $T < 3$  МэВ равновесие не успевает поддерживаться, так как вероятность процессов с участием нейтрино становится слишком малой. При  $T < m_\nu \approx 0,5$  МэВ  $e^+e^-$ -пары аннигилируют, увеличивая тем самым  $N_\gamma$ , а  $N_\nu$  при этом, очевидно, не меняется. Процесс аннигиляции происходит адиабатически, с сохранением энтропии. В силу этого, если до аннигиляции число внутренних степеней свободы в плазме (т. е. коэффициент  $Q$ ) в выражении (1.5) составлял  $2 + 4 \cdot (7/8) = 11/2$ , то после аннигиляции он стал равен 2. Соответственно плотность числа фотонов возрасла в  $11/4$  раз. Это и приводит к отношению  $N_\nu/N_\gamma = 3/22$ .

Для более сильновзаимодействующих частиц положение иное. Они продолжают находиться в равновесии с плазмой и при  $T < m$ . Их концентрация следует при этом равновесному закону  $N \sim (mT)^{3/2} \exp(-m/T)$ . Если бы такое положение продолжалось до наших дней, никаких свидетелей ранней горячей стадии сейчас фактически бы не осталось. Однако это падение концентрации не вечно. В какой-то момент концентрация становится настолько низкой, что даже сильное взаимодействие не успевает поддерживать равновесие и падение концентрации за вычетом тривиального фактора, связанного с общим расши-

рением, прекращается. Концентрация частиц в элементе сопутствующего объема становится постоянной. Это явление носит название заалки концентрации (в английской литературе — freezing). В результате стабильные реликты Большого взрыва могут дожить до наших дней. Для нестабильных частиц (т. е. для частиц, время жизни которых меньше возраста Вселенной:  $\tau < 10^{10}$  лет) концентрация, если даже когда-то отклоняется от равновесной, в конечном счете должна следовать равновесному закону, так как скорость их распада и рождения постоянна, а темп расширения мира падает.

Как известно, равновесная концентрация частиц в идеальном газе определяется не только их массой и температурой, но и химическим потенциалом. Последний необходимо ввести, чтобы учесть следствия сохранения заряда (или зарядов) в реакциях между частицами. Например, в электронейтральной плазме, состоящей из электромагнитного излучения и  $e^+e^-$ -пар, распределение частиц по энергии дается выражением

$$n_{e^+} = n_{e^-} = n_\gamma = \exp(-E/T)$$

(без учета квантовых ферми- или бозе-эффектов).

Эта формула, очевидно, неприменима, если полный заряд плазмы отличен от нуля, т. е.  $n_{e^+} \neq n_{e^-}$ . Электрический заряд строго сохраняется, поэтому никакие процессы не могут привести к уравниванию концентраций  $e^+$  и  $e^-$ . В равновесных функциях распределения это учитываем введением дополнительного множителя, например:

$$n_{e^-} = \exp\left\{\frac{\mu_e - E}{T}\right\}.$$

Величина  $\mu_e$  называется химическим потенциалом (электронов). Заметим, что число фотонов не сохраняется, поэтому в равновесии  $\mu_\gamma = 0$ .

Согласно наблюдениям во Вселенной имеются барионы (протоны, нейтроны), но нет антибарионов. Это означает, что плотность барионного заряда отлична от нуля и, следовательно, на равновесной стадии распределение барионов описывается отличным от нуля химическим потенциалом  $\mu_b^*$ . Химический потенциал нейтрино, по-видимому, также отличен от нуля и близок по величине к  $\mu_b$ , но в отличие от барионов, концентрация которых низка при  $T \ll m_p \approx 1$  ГэВ и для которых роль химического потенциала весьма существенна, концентрация нейтрино не мала при  $T < m_\nu$ , и химический потенциал приводит лишь к малым поправкам.

Здесь речь идет об этапах, когда несохранение барионного заряда, предсказываемое современной теорией (см. ниже), невелико. Это заведомо верно при  $T \lesssim 1$  ГэВ.

Химические потенциалы являются дополнительными параметрами, от которых зависит концентрация находящихся во Вселенной реликтовых частиц. В принципе их можно вычислить (см. гл. 8 о бариосинтезе во Вселенной), однако нам неизвестны характеристики взаимодействия частиц в существенной области энергий (или температур) и поэтому конкретный численный ответ мы получить не можем.

Если предположить, что химический потенциал тех или иных стабильных частиц (обозначим их индексом  $a$ ) равен нулю, то нетрудно вычислить их концентрацию в современный период (или их «закаленную» концентрацию). Она целиком определяется термодинамическим равновесием. Если химический потенциал отличен от нуля, то ответ зависит от его величины, а полученное выражение при  $\mu=0$  можно рассматривать как нижнюю границу для концентрации.

Для последней справедлива оценка

$$N_a/N_\gamma = (v\sigma_{aa}m_a m_{p1})^{-1}, \quad (1.6)$$

где  $N_\gamma = 550 (T/3K)^3 \text{ см}^{-3}$  — плотность реликтовых фотонов,  $m_{p1} = 10^{19} \text{ ГэВ}$  — масса Планка,  $\sigma_{aa}$  — сечение аннигиляции  $a$  и  $\bar{a}$ , а  $v$  — их скорость при столкновении.

Эти соображения были использованы Я. Б. Зельдовичем, Л. Б. Окунем и С. Б. Пикельнером (1965) для расчета концентрации реликтовых кварков\*, если предполагать, что кварки могут существовать как свободные частицы. По их оценкам содержание свободных кварков в природе должно быть столь же высоким, как, скажем, содержание золота, что находится в резком противоречии с экспериментальными ограничениями, причем независимо от массы кварков. Полученный результат послужил важным аргументом в пользу того, что кварки не могут существовать в свободном виде, т. е. имеет место невыедание (или конфайнмент) кварков. Иными словами, кварки — объекты принципиально нового типа, которые существуют внутри элементарных частиц, но не могут быть оттуда вырваны (см. гл. 4).

Если применить формулу (1.6) для барионов, то получим результат:

$$N_b/N_\gamma \approx 10^{-19},$$

что на 9—10 порядков ниже наблюдаемой концентрации барионов\*\*. Это лишний раз демонстрирует, что  $\mu_b \neq 0$ .

\* Что такое кварки, объясняется в гл. 4.

\*\* В космических лучах антибарионы (на уровне порядка  $\sim 10^{-4}$  от барионов), но они имеют вторичное происхождение. Концентрация первичных антибарионов при наличии барионного избытка экспоненциально мала.

С помощью аналогичных рассуждений была вычислена концентрация реликтовых магнитных монополий, т. е. частиц, несущих элементарный магнитный заряд (Я. Б. Зельдович, М. Ю. Хлопов, 1978). Теория предсказывает, что монополии должны возникать при охлаждении Вселенной во время определенного фазового перехода. Концентрация этих монополий во Вселенной в настоящее время оказывается неприемлемо высокой. Эта трудность может быть разрешена в рамках инфляционной модели, если период экспоненциального расширения имел место после возникновения монополий. В этом случае концентрация монополий окажется экспоненциально подавленной.

Знание количества реликтовых частиц типа  $a$  в сегодняшней Вселенной позволяет получить ограничение на некоторые их свойства, в частности на массу. Дело в том, что плотность энергии, заключенная в этих частицах, не может быть слишком велика, так как в противном случае будет слишком мал возраст Вселенной, связанный с  $\rho_a$  соотношением

$$t_U \approx \frac{1}{H} \left(1 + \frac{1}{2} \sqrt{\Omega}\right)^{-1}, \quad (1.7)$$

где  $\Omega = \rho_{\text{tot}}/\rho_c > \rho_a/\rho_c$ . Таким образом, было получено ограничение на массу нейтрино (любого типа):  $m_\nu < 30 \text{ эВ}$  (С. С. Герштейн, Я. Б. Зельдович). Любопытно, что для тяжелого аналога нейтрино (если он существует) космология дает ограничение на массу снизу:  $m_L > 4 \text{ ГэВ}$  [3, 49, 54, 64]. Это связано с тем обстоятельством, что сечение аннигиляции  $L\bar{L}$  растет с ростом  $m_L$ ,  $\sigma \sim m_L^2$  и поэтому их плотность энергии в соответствии с формулой (1.6) падает.

Такого рода рассуждения сейчас широко используются в литературе, чтобы извлечь какую-то информацию о свойствах большого числа новых слабозаимодействующих частиц, предсказываемых современными теориями и недоступных пока ускорительным экспериментам.

## § 8. НУКЛЕОСИНТЕЗ

Период времени от 1 до 200 с от начала играет существенную роль в истории Вселенной. В это время образуются первичные легкие ядра:  $\text{He}^4$  (25%),  $\text{H}^2$  ( $3 \cdot 10^{-5}$ ),  $\text{He}^3$  ( $2 \cdot 10^{-5}$ ),  $\text{Li}^7$  ( $10^{-9}$ ), т. е. фактически начинает рождаться привычное нам вещество. Более тяжелые элементы на этом этапе не успевают образоваться, они возникают значительно позже в звездах.

Выход различных ядер за счет первичного нуклеосинтеза существенно зависит от величины  $(n/p)$ -отношения в этот период. При  $t \leq 1$  с и соответственно  $T \geq 1 \text{ МэВ}$  относительная концентрация нейтронов и протонов описывалась равновесной формулой

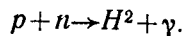
$$N_n/N_p = \exp(-\Delta m/T), \quad (1.8)$$

где  $\Delta m \approx 1,3$  МэВ — разность масс нейтрона и протона. Равновесие поддерживалось реакциями слабого взаимодействия.

При падении температуры до  $T_f = 0,7$  МэВ эти реакции эффективно прекращались и происходила закалка  $(n/p) =$  отношения, т. е.

$$(N_n/N_p)_f = \exp(-\Delta m/T_f) = \text{const.} \quad (1.9)$$

На этом этапе нейтроны и протоны существовали в свободном виде, не связываясь в ядра. В дальнейшем при  $T \lesssim 100$  кэВ большая часть нейтронов (кроме тех, что успеют распасться) связывалась в дейтерий за счет реакции



В свою очередь дейтерий, эффективно захватывая барионы в первичной плазме, рождал  $He^3$  и  $H^3$  и наконец с захватом еще одного протона или нейтрона образовывался  $He^4$ , в котором практически все нераспавшиеся нейтроны заканчивали свой путь. Отсутствие подходящих ядер с  $A=5$  тормозило дальнейшие реакции, и в силу малой плотности барионного вещества процессы  $He^3 + He^4 \rightarrow Be^7$ ,  $3He^4 \rightarrow C^{12}$  и т. п. почти не шли.

Относительный (по массе) выход  $He^4$  и  $H^2$  в зависимости от плотности барионов в первичной плазме изображен на рис. 1. Уменьшение выхода дейтерия с ростом  $\rho_b$  объясняется тем, что при увеличении плотности барионов растет вероятность столкновения дейтерия с ними и соответственно возрастает вероятность образования более тяжелых ядер. Лишь очень

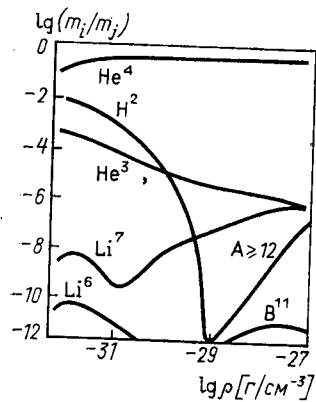


Рис. 1. Расчетные весовые концентрации легких элементов, образовавшихся при первичном нуклеосинтезе в зависимости от барион-фотонного отношения  $N_B/N_T$

малой части дейтерия удается избежать связывания в конечном счете в  $He^4$ , и эта часть тем меньше, чем выше плотность барионов. В силу сказанного количество дейтерия во Вселенной является чувствительным индикатором плотности барионной материи. Здесь имеется, однако, некоторая тонкость, состоящая в том, что за время эволюции Вселенной количество дейтерия могло измениться. Известно, что первичный дейтерий сгорает в звездах, образуя  $He^3$ . Поэтому для вывода ограничения на  $\rho_b$  следует использовать совместные данные по  $H^2$  и  $He^3$  (три-

тий радиоактивен и быстро распадается). На их основе можно сделать вывод, что доля первичных дейтерия и гелия-3 по массе составляет  $\sim 10^{-4}$ . Отсюда следует, что относительное количество барионов во Вселенной не может быть слишком велико:  $\Omega_b \lesssim 0,1$ .

Согласие предсказаний теории первичного нуклеосинтеза с наблюдениями (по обилиям  $H^2$ ,  $He^3$ ,  $He^4$ ,  $Li^7$ ) является одним из краеугольных камней, на которых покоится модель горячей Вселенной. Более того, это согласие позволяет сделать ряд нетривиальных выводов об элементарных частицах. В частности, можно получить ограничение сверху на число различных типов нейтрино. Идея этого ограничения принадлежит В. Ф. Шварцману. Сейчас известны три типа нейтрино: электронное, мюонное и тау. Космология позволяет заключить, что могут существовать еще два-три типа нейтрино. Рассуждения выглядят следующим образом. Как мы уже упоминали, содержание  $He^4$  определяется количеством нейтронов в первичной плазме, т. е. отношением  $N_n/N_p$  (1.9). Температура закалки  $T_f$  зависит от соотношения между скоростью реакций (1.8)  $\sigma N_v$  и темпом расширения Вселенной  $H(t) = 1/2t$ . Величина  $\sigma N_v$  пропорциональна пятой степени температуры, так как  $N_v \sim T^3$ , а сечение процессов (1.8) при малых энергиях пропорционально квадрату энергии, что при статистическом усреднении дает  $T^2$ .

Согласно соотношениям (1.4) и (1.5) скорость падения температуры со временем зависит от числа типов частиц в первичной плазме  $Q$  следующим образом:

$$T = \left( \frac{4}{16\pi^3} \frac{1}{Qg t^2} \right)^{1/4}. \quad (1.10)$$

Каждый тип нейтрино вносит в  $Q$  вклад, равный 7/8. Отсюда видно, что чем больше количество типов нейтрино  $k_\nu$ , тем выше темп остывания и тем выше  $T_f$ . Поэтому с ростом  $k_\nu$  закаленное значение  $(n/p)$ -отношения становится больше и больше образуется первичного  $He^4$ . В процессе эволюции Вселенной  $He^4$  не разрушается, а может только создаваться, поэтому на основании данных по сегодняшнему обилию  $He^4$  можно получить ограничение сверху на количество первичного  $He^4$  и, следовательно, ограничение на  $k_\nu$ . К сожалению, содержание  $He^4$  во Вселенной известно не очень хорошо. Имеются области, где оно достигает 30% по массе, а кое-где лишь 22—23%. В первом случае, помимо  $He^4$ , наблюдается обогащение межзвездного газа тяжелыми элементами, поэтому полагают, что столь большая концентрация  $He^4$  связана с дополнительным поздним рождением в звездах и последующим выбросом. С другой стороны, оценки количества невидимого  $He^4$  не очень надежны, поэтому (22—23%) едва ли является строгим верхним пределом. Сейчас в мировой литературе полагают, что достаточно

надежная верхняя граница весового содержания первичного  $\text{He}^4$  составляет  $R(\text{He}^4) < 25\%$ .

Величина расчетной концентрации  $\text{He}^4$  зависит, хотя и очень сильно (см. рис. 1), от общей концентрации нуклонов (протонов и нейтронов), или, точнее, от отношения  $N_b/N_p$ , так как образование  $\text{He}^4$  идет в двухчастичных столкновениях нуклонов или ядер. Эта величина известна из наблюдений довольно плохо. Однако на нее можно получить ограничения как сверху, так и снизу, если вспомнить, что концентрация первичного дейтерия сильно зависит от плотности барионов.

На этом основании можно сделать вывод, что  $k_v < 4$ . Более осторожные авторы считают, что данные о содержании легких элементов имеют большие неопределенности и что, кроме того, соотношение между обилиями, наблюдаемыми сегодня и в период нуклеосинтеза, также содержит заметную неопределенность. На этом основании они увеличивают верхний предел:  $k_v < 6$ .

Таким же образом можно получить ограничение на другие возможные типы частиц, если в период нуклеосинтеза ( $t=1-100$  с,  $T=1$  МэВ — 100 кэВ) они присутствуют в плазме. В ча-

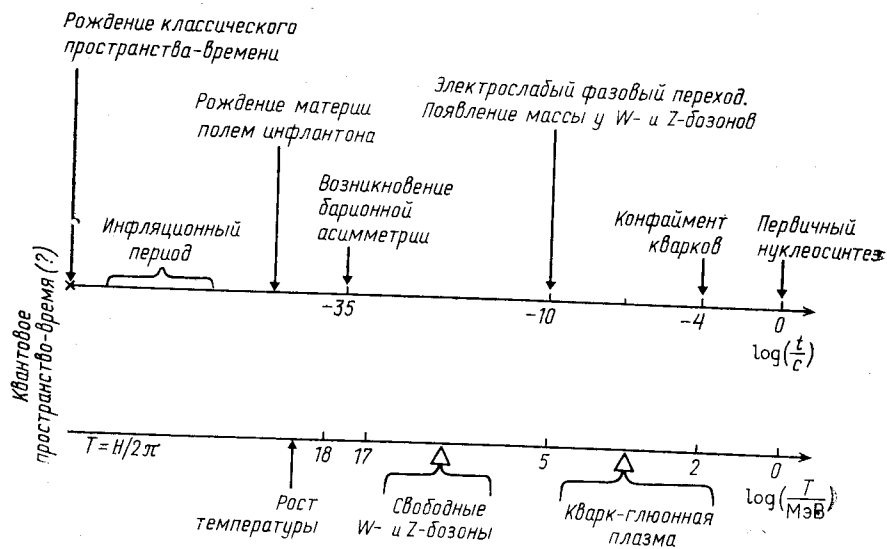


Рис. 2. Основные этапы эволюции ранней Вселенной

стности, на этом основании можно исключить существование зеркального, или теневого, мира, полностью симметричного нашему, но связанного с нашим только гравитационно. Существование такого мира до инфляционной стадии, впрочем, иск-

лючить такими соображениями нельзя, так как возможна ситуация, когда концентрация зеркальных, или тневых, частиц сильно разбавлена по сравнению с нашими вследствие инфляции и их роль в нуклеосинтезе ничтожна. Существование на ранних этапах теневого мира предсказывает активно развиваемая в последнее время теория суперструн.

Другим примером использования первичного нуклеосинтеза в физике частиц является ограничение сверху на плотность долгоживущих частиц, распадающихся на пару  $p\bar{p}$ , в период нуклеосинтеза и позже, так как возникающие при этом антипротоны, взаимодействуя с  $\text{He}^4$ , будут давать аномально высокое содержание дейтерия. Именно так были получены наиболее сильные ограничения на новые частицы, предсказываемые суперсимметрией.

Материал этой главы иллюстрируется рис. 2.

## Глава 2.

### ТРУДНОСТИ КЛАССИЧЕСКОЙ КОСМОЛОГИИ

Космологическая модель, описанная в предыдущей главе, выдержала проверку на согласие с наблюдениями, она подтверждена реликтовым электромагнитным фоном, наблюдениями за обилиями легких элементов и наконец теорией — ОТО.

В этой главе мы укажем на некоторые ее концептуальные трудности и на пути их разрешения. Подчеркнем, что модификация модели, необходимая для устранения этих трудностей, не должна менять уже апробированную стандартную космологическую модель, которая в этом смысле уже никогда не исчезнет.

#### § 1. ЗАКОН РАСШИРЕНИЯ И УРАВНЕНИЕ СОСТОЯНИЯ

Уравнения ОТО однозначно предсказывают характер расширения Вселенной, если известны плотность энергии  $\rho$  и давление  $p$  вещества (в однородной и изотропной космологической модели). Плотность энергии можно задать с помощью параметра  $\Omega = \rho/\rho_c$ , а давление — уравнением состояния  $p = p(\rho)$ . Сразу подчеркнем, что понятие уравнения состояния можно применять только к релаксированной системе, когда давление однозначно определяется плотностью энергии. Ниже у нас будут примеры, когда это не так.

Расширение Вселенной описывают с помощью масштабного фактора  $a(t)$ , который характеризует изменение со временем расстояния между космологическими объектами. Пространственно-временной интервал можно записать в виде

$$ds^2 = dt^2 - a^2(t) [dr^2 + f(r)(d\theta^2 + \sin^2\theta d\varphi^2)], \quad (2.1)$$

где функция  $f(r)$  зависит от топологических свойств Вселенной в целом. Для пространственно плоской Вселенной, в которой геометрия в среднем евклидова,  $f(r) = r^2$ , для замкнутой Вселенной  $f(r) = \sin^2 r$  и для открытой  $f(r) = \sinh^2 r$ . Здесь  $r$  — безразмерная величина, а масштабный фактор  $a(t)$  имеет размерность длины.

Зависимость  $a$  от  $t$  определяется уравнениями Эйнштейна, которые в однородном и изотропном случае имеют вид

$$\frac{1}{2} \left( \frac{da}{dt} \right)^2 - \frac{4\pi G \rho a^2}{3} = -k/2, \quad (2.2)$$

$$\frac{d^2 a}{dt^2} = -\frac{4\pi G}{3} (\rho + 3p) a, \quad (2.3)$$

где постоянная  $k$  принимает значения  $+1$  для замкнутого мира,  $-1$  для открытого и  $0$  для плоского.

Если кроме этого задана зависимость  $p = p(\rho)$ , то все три неизвестные функции будут определены.

Заметим, что из уравнений (2.2) и (2.3) следует закон изменения энергии материи в расширяющемся мире:

$$\frac{d\rho}{dt} = -3H(\rho + p), \quad (2.4)$$

где  $H = \dot{a}/a$  — постоянная Хаббла. Изменение плотности энергии связано с двумя факторами: с общим расширением пространства и с работой сил давления.

Рассмотрим различные законы расширения, отвечающие различным формам уравнения состояния при  $\Omega = 1$  ( $k = 0$  в уравнении (2.2)). Как мы уже отмечали, для идеального релятивистского газа  $p = \rho/3$ . В этом случае масштабный фактор зависит от  $t$  степенным образом:

$$a(t) = a_0 (t/t_0)^{1/2}. \quad (2.5)$$

В стандартной фридмановской космологии предполагается, что такой закон расширения справедлив от «начала» до  $t \approx 10^{11}$  с, когда в плотности энергии начинают доминировать нерелятивистские частицы. Как легко видеть, плотность энергии релятивистского вещества падает как  $a^{-4}$ . Это связано с расширением мира (что дает  $a^{-3}$  для числа частиц в единице объема и с потерей энергии частиц из-за красного смещения ( $a^{-1}$ ). Плотность энергии нерелятивистского газа ведет себя как  $a^{-3}$ . Поэтому в какой-то момент  $t$  нерелятивистские частицы должны начать доминировать. Нетрудно показать, что при  $\Omega = 1$  и при  $H = 50$  км/с Мпк это происходит при  $z$ -факторе, равном по порядку величины:

$$z = a_0/a(t) - 1 = \rho_c/\rho_r - 1 \approx 10^4,$$

где  $\rho_r \approx 4,5 \cdot 10^{-34}$  г/см<sup>3</sup> — плотность энергии реликтовых фотонов\*, а критическая плотность  $\rho_c = 5 \cdot 10^{-30}$  г/см<sup>3</sup>. Как легко видеть, в момент перехода температура фотонного газа должна составлять  $T = z \cdot 3K \approx 30\,000$  К  $\approx 3$  эВ. Согласно формуле (1.3) это отвечает  $t \approx 10^{11}$  с от начала.

После этого момента расширение определяется уже уравнением нерелятивистского газа  $p \ll \rho$ ; с хорошей точностью можно положить  $p = 0$ . В этом случае

$$a(t) = a_1 (t/t_1)^{2/3}. \quad (2.6)$$

Отметим еще, что экспоненциальное расширение отвечает случаю, когда  $p = -\rho$ :

$$a(t) = a_0 \exp \left[ \sqrt{\frac{8\pi G \rho}{3}} t \right]. \quad (2.7)$$

Заметим, что при этом согласно (2.4) плотность энергии остается постоянной, т. е. работа сил давления в точности компенсирует убывание энергии из-за расширения.

Основные трудности старой классической космологии как раз связаны с медленным ростом масштабного фактора (как  $t^{1/2}$  или  $t^{2/3}$ ). В двух словах их можно охарактеризовать так: малым  $t$  соответствуют слишком большие  $a$  и слишком большой размер Вселенной. Ниже мы подробно обсудим эти проблемы.

В заключение этого раздела подчеркнем, что инфляционная модель, успешно решающая упомянутые космологические проблемы, отнюдь не отменяет классическую космологию. Она лишь задает начальные условия стандартной космологической модели. После окончания периода экспоненциального расширения, которое имело место во времена, близкие к планковским,  $t \approx 10^{-34}$  с, режим расширения становится степенным и мы возвращаемся к стандартному сценарию, подробно описанному, например, в книгах Зельдовича и Новикова (1975) и Вайнберга (1975).

## § 2. ПРОБЛЕМА ГОРИЗОНТА

Как отмечалось выше, реликтовое излучение, приходящее с различных участков неба, является абсолютно одинаковым. В настоящее время оно практически не взаимодействует с веществом во Вселенной. Вселенная стала прозрачна для реликтового излучения после рекомбинации водорода, когда заря-

\* Безмассовые нейтрино удваивают плотность энергии релятивистских частиц и уменьшают  $z$  вдвое.

женные электроны и протоны объединились в нейтральные атомы водорода, с которыми длинноволновые фотоны взаимодействуют очень слабо. Температура рекомбинации водорода равна 3000 К, значит это произошло при  $z \approx 10^3$ , что отвечает моменту времени  $t_r = 10^{12} - 10^{13}$  с. После этого момента реликтовое излучение уже ни с чем не взаимодействовало. Размер причинно связанной области в это время (или размер горизонта) по порядку величины составляет  $c \cdot t_r$  (более точные формулы будут приведены ниже). Поэтому участки на небесной сфере с угловыми размерами  $\theta = (1+z_r) (t_r/t_0) \approx 10^{-2}$  ничего не должны «знать» друг о друге. Тем не менее излучение всюду одинаково. Этот загадочный для фридмановской космологии факт называют проблемой горизонта (рис. 3).

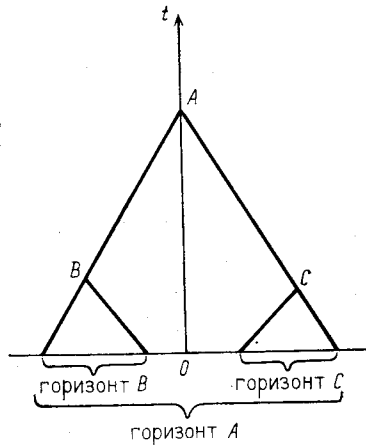


Рис. 3. Горизонты событий во фридмановской космологии. События B и C причинно не связаны

Вселенной (см. гл. 10) в литературе обсуждались модели с вторичной ионизацией на достаточно раннем этапе. В этом случае первоначальные малые флуктуации реликтового излучения могут несколько уменьшиться.

Точные формулы для размера горизонта получатся, если учесть расширение Вселенной. Рассмотрим уравнение для луча света, идущего от какого-то объекта к наблюдателю. Луч света распространяется по геодезической, для которой можно положить  $d\phi = d\theta = 0$  в выражении (2.1). Тогда уравнение, описывающее движение луча, будет иметь вид

$$dr = -dt/a(t),$$

а расстояние между источником и наблюдателем в момент  $t$  равно

$$l = a(t) \int_0^t dt' / a(t'). \quad (2.8)$$

Отсюда  $l = 2 \cdot t = H^{-1}$  для закона расширения (2.5) (т. е. для релятивистского газа) и  $l = 3 \cdot t = 2 \cdot H^{-1}$  для нерелятивистского вещества (2.6).

Заметим, что  $l \sim t$  и, следовательно, растет быстрее, чем масштабный фактор  $a(t) \sim t^{2/3}$  (или  $t^{1/2}$ ), поэтому в пределах горизонта в далеком будущем окажется любая частица. При экспоненциальном законе расширения  $a \sim \exp[H_0 t]$  это не так: с частицей, исходно находящейся на расстоянии большем, чем  $H_0^{-1}$ , невозможно обменяться никакими сигналами, она всегда будет находиться за горизонтом. Однако если по какой-то причине область размера  $l_0$  была причинно связанной, то через время  $t$  ее размер сильно возрастает:  $l(t) = l_0 \exp(H_0 t)$ . И хотя различные участки этой области перестают взаимодействовать, как только окажутся на расстоянии  $H_0^{-1}$ , они «помнят» о своей связи в прошлом. На этом основано решение проблемы горизонта в инфляционной модели (см. рис. 4).

### § 3. ПРОБЛЕМА ПЛОСКОСТНОСТИ

Эта проблема состоит в необходимости чрезвычайно точной близости  $\rho$  к  $\rho_c$  на ранних стадиях эволюции Вселенной. Если в планковский момент  $(\Omega - 1) \sim o(1)$ , то характерное время для такой Вселенной  $\sim 10^{-33}$  с, т. е. такая Вселенная либо схлопывается за  $10^{-33}$  с, если она открыта, либо расширяется столь быстро, что ни галактики, ни звезды не успели бы образоваться, а плотность вещества к моменту  $10^{10}$  лет была бы намного меньше наблюдаемой сейчас ( $\rho \approx 10^{-29}$  г/см<sup>3</sup>).

Для количественной оценки степени необходимой тонкой настройки параметра  $\Omega$  обратимся к уравнению (2.2). Вспомогательная определения  $H = \dot{a}/a$  и  $\rho_c = 3H^2/8\pi G$ , перепишем его в виде:

$$H^2(1 - \Omega) = -k/a^2. \quad (2.9)$$

Пользуясь этим уравнением, можно выразить величину  $\Omega$  в более раннюю эпоху, которой соответствует красное смещение  $z$  через  $\Omega_0$ , наблюдаемую сегодня:

$$\frac{1 - \Omega}{\Omega} = \frac{1 - \Omega_0}{\Omega_0} \cdot \frac{1}{(1+z)^n},$$

где  $n=1$ , если доминирует нерелятивистское вещество (т. е.  $\epsilon \propto a^{-3}$ ), и  $n=2$ , если доминирует излучение ( $\epsilon \propto a^{-4}$ ).

Полагая для простоты, что  $(1 - \Omega_0)/\Omega_0 \approx 1$ , оценим, насколько близко к единице должна быть  $\Omega$  на ранних этапах. Из таблицы видно, что чем раньше по времени задавались начальные условия для параметра  $\Omega$ , тем точнее нужно было «попасть» в единицу, чтобы получить наблюдаемую сегодня Вселенную.

Те же рассуждения можно проиллюстрировать соотношением между размерами горизонта в момент рождения мира и

радиусом кривизны модели Фридмана. Радиус кривизны  $R$  в однородной и изотропной модели Фридмана определяется через инвариант тензора кривизны

$$C_G = |k|/a^2 = 1/R^2$$

и совпадает с масштабным фактором  $R=a$ . Поскольку  $a \sim \sqrt{t/t_0}$ , а размер горизонта  $h=ct$ , то  $h \ll R$  при  $t \ll t_0$ . Это значит, что в момент рождения мир был почти плоским.

Проведем дополнительные вычисления, чтобы понять, каких начальных условий требует классическая космологическая

Возраст Вселенной	Период	$\Omega$
$t = 2 \cdot 10^{10}$ лет	современный	$1 \pm 0,5$
$t = 10^5$ лет	рекомбинация	$1 \pm 10^{-3}$
$t = 1$ с	начало нуклеосинтеза	$1 \pm 10^{-16}$
$t \approx 10^{-5}$ с	кварк-глюонная плазма	$1 \pm 10^{-21}$
$t \approx 10^{-10}$ с	свободные	$1 \pm 10^{-28}$
$t \approx t_{\text{pl}} \approx 10^{-43}$ с	«сотворение» мира	$1 \pm 10^{-60}$

модель при рождении мира. В качестве момента рождения возьмем не  $t=0$ , а  $t \approx t_{\text{pl}} = 5 \cdot 10^{-43}$  с. При этом плотность энергии имеет величину  $\sim m^4_{\text{pl}}$ . Мы не имеем теории, которая позволяла бы описывать физические процессы в этой области. Здесь наверняка существуют квантово-гравитационные эффекты, надежное описание которых пока отсутствует. Таким образом, нам трудно пока делать какие-либо высказывания относительно моментов времени  $t < t_{\text{pl}}$ , но мы надеемся, что при  $t \geq t_{\text{pl}}$  понимаем ситуацию, хотя бы качественно.

Для целей этого параграфа вполне достаточно будет отнести рождение мира к моменту  $t_{\text{pl}}$ . Будем считать, что радиус кривизны нашей Вселенной сейчас совпадает с размерами горизонта  $R \approx 10^{28}$  см (точнее  $R > 10^{28}$  см). Полагая для простоты, что закон расширения имел вид  $a(t) \sim \sqrt{t}$ , вплоть до настоящего времени, получим, что в планковский момент радиус кривизны был порядка

$$R \approx 10^{28} \text{ см } \sqrt{10^{-43} \text{ с} / 10^{17} \text{ с}} \approx 10^{-2} \text{ см},$$

а размер горизонта в то время  $l_{\text{pl}} \approx 10^{-33}$  см. Это значит, что от плоского этот мир отличается только на величину  $t_0 = 10^{-31}$ ! Прделаем обратные вычисления. Пусть в момент  $t_{\text{pl}}$  рождается мир с естественно ожидаемым радиусом кривизны  $l_{\text{pl}}$ . Каким он будет сейчас к моменту  $t_0 = 10^{10}$  лет? Простые вычис-

ления дают, что сейчас радиус кривизны должен быть  $R \approx 10^{-2}$  см.

Может быть что-то не в порядке с нашими рассуждениями о радиусе кривизны, размерах мира и размерах горизонта? Нет, здесь все правильно. Те же самые выводы получаются при рассмотрении значений плотности сейчас и в планковскую эпоху. Как известно, плотность вещества с уравнением состояния  $p = \epsilon/3$  при расширении изменяется как  $a^{-4}$ . Если мир родился с  $a \approx 10^{-33}$  см и  $\rho = \rho_{\text{pl}}$ , то к настоящему моменту времени его плотность должна была уменьшиться в  $(10^{-33}/10^{28})^4 = 10^{-244}$  раз и стать  $\rho = 10^{-150}$  г/см<sup>3</sup>. Аналогичное положение имеет место и в случае  $p=0$ .

Основной вывод из всех этих рассуждений заключается в том, что классическую модель Фридмана нельзя экстраполировать на слишком ранние эпохи. Закон изменения масштабного фактора  $a(t)$  при малых  $t$  был иным, отвечающим значительно более быстрому росту  $a(t)$  со временем. При этом удается ответить на вопросы: почему  $\Omega$  столь близка 1? почему Вселенная однородна и изотропна? почему она расширяется, а закон расширения имеет вид  $v = Hr$ ?

#### § 4. ИНФЛЯЦИОННАЯ ТЕОРИЯ. ПЕРВОЕ ЗНАКОМСТВО

Существенно новая черта, позволяющая решить отмеченные выше проблемы, состоит во введении отрицательного давления и изменении соотношения между давлением и плотностью энергии с  $p = \epsilon/3$  на

$$p = -\epsilon \quad (2.10)$$

(достаточно, чтобы это соотношение было приблизительно справедливым в течение некоторого промежутка времени). Не будем пока обсуждать физических причин для этого закона, а лишь рассмотрим его следствия. Как мы уже отмечали, при  $p = -\epsilon$  расширение является экспоненциальным (см. (2.7)). Величина  $H = \dot{a}/a$  при этом не зависит от времени, поэтому из соотношения (2.10) следует, что  $|1 - \Omega| \sim \exp(-2 \cdot H \cdot t)$  и поэтому, начав с  $|1 - \Omega| = o(1)$ , за 70 хэббловских времен  $t = 70H^{-1}$  придем к плоской Вселенной с необходимой степенью точности  $\sim 10^{-60}$ . Отметим разницу между законами расширения (2.7) и (2.5) или (2.6). В первом случае  $\Omega \rightarrow 1$  при  $t \rightarrow \infty$ , а во втором  $\Omega \rightarrow 0$  при  $t \rightarrow \infty$ .

Очевидно, что при этом решается и проблема горизонта, так как причинно связанная область, существующая в момент рождения Вселенной и имеющая размер  $l \sim t_{\text{pl}}^{-1} \sim 10^{-33}$  см, за время инфляции достигнет нужной величины  $\geq 10^{-3}$  см (рис. 4). В дальнейшем эта область расширится до размеров сегодняшнего горизонта. Неоднородности, существовавшие до

экспоненциальной стадии, за время инфляции сильно разглаживаются, и видимая часть Вселенной становится в высокой степени однородной (при подходящих параметрах модели).

Согласно уравнению (2.3) источником гравитационного поля является не только энергия, но и давление вещества. Это связано с релятивистскими эффектами и для соотношения (2.10) приводит к качественно новому результату — к гравитационному отталкиванию. Именно это отталкивание и могло быть причиной первоначального толчка, вызвавшего расширение Вселенной, если в родившейся в результате квантового скачка микровселенной с размером  $m_{\text{Pl}}^{-1}$  давление и плотность энергии вещества удовлетворяли, хотя бы приближенно, условию (2.10).

Весьма важным является вопрос об окончании инфляционной стадии, о переходе к обычному уравнению состояния  $p = \epsilon/3$ . Ответ на него зависит от конкретных физических условий, за счет которых реализуется соотношение  $p = -\epsilon$ . Мы заметим только, что, несмотря на экзотический вид этого соотношения, оно естественно возникает в целом ряде моделей. В частности, к нему приводят фазовые переходы от симметричной к несимметричной фазе в первичной плазме, динамика скалярного поля при весьма естественных предположениях или теории с высшими измерениями ( $D > 4$ ). На всех этих вопросах мы остановимся ниже.

### § 5. ПРОБЛЕМА КОСМОЛОГИЧЕСКОЙ ПОСТОЯННОЙ

Соотношение  $p = -\epsilon$  фактически было введено в космологию Эйнштейном. Именно, обнаружив, что сформулированные им уравнения релятивистской теории гравитации в космологической ситуации не имеют стационарных решений, Эйнштейн предложил добавить к источнику гравитационного поля, каковым является тензор энергии-импульса (подробнее см. ниже), слагаемое  $\Lambda g_{\mu\nu}$ , где  $g_{\mu\nu}$  — метрический тензор. Величину  $\Lambda$  называют космологической постоянной. Можно показать, что такой член как раз отвечает  $p = -\epsilon$ . Впоследствии выяснилось, что Вселенная нестационарна, она расширяется и надобность в космологической постоянной исчезла. Более того, из астрономических наблюдений мы знаем, что в настоящее время она чрез-

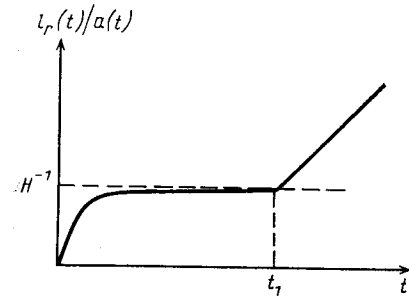


Рис. 4. Отношение размера горизонта  $l_r$  к масштабному фактору при различных режимах расширения

вычайно мала,  $|\Lambda| < 10^{-56} \text{ см}^{-2}$ . Сейчас, однако, мы вновь приходим к необходимости введения космологической постоянной, или, точнее, какого-то ее аналога, в ранней Вселенной, так, чтобы получить условие  $p = -\epsilon$ . Однако это уже не космологическая постоянная как свойство вакуума, а метастабильное состояние поля, заполняющего расширяющееся пространство. Заметим, что физика элементарных частиц предсказывает, что космологическая постоянная должна быть отлична от нуля, причем ее величина, говоря грубо, на 100 порядков (!) превосходит существующее ограничение. Естественный механизм, который мог бы привести к сокращению космологической постоянной с такой степенью точности, в настоящее время неизвестен, и его поиски представляют сейчас одну из наиболее увлекательных задач, как в космологии, так и в физике элементарных частиц. В виду важности вопроса мы посвятим ему отдельную главу.

### § 6. БАРИОННАЯ АСИММЕТРИЯ ВСЕЛЕННОЙ

В заключение отметим еще одну проблему космологии современной стадии (кроме горизонта и плоскостности), которая нашла свое решение при применении идей физики элементарных частиц к ранней Вселенной. Речь идет о так называемой барионной асимметрии Вселенной, о факте наличия во Вселенной в подавляющем количестве вещества в форме протонов и нейтронов, объединенных общим названием барионы, и практически полного отсутствия антивещества — антибарионов. Заметим, что антивещество в форме антинейтрино во Вселенной должно присутствовать, причем количество  $\bar{\nu}$  примерно равно количеству  $\nu$  и количеству реликтовых фотонов и очень велико по сравнению с количеством барионов (1 барион приходится на 10 миллиардов реликтовых фотонов).

Для решения проблемы барионного избытка нет необходимости в модификации фридмановского закона расширения, необходим лишь сам факт нестационарности (расширения) Вселенной, различие во взаимодействиях частиц и античастиц (установленное экспериментально) и несохранение количества барионов, что в настоящее время предсказывает теория элементарных частиц. При наличии этих трех составляющих первичная плазма независимо от начальных условий при  $T > m_B \approx 1 \text{ ГэВ}$  придет в состояние с небольшим избытком барионов над антибарионами  $(N_B - N_{\bar{B}})/(N_B + N_{\bar{B}}) \sim 10^{-9} - 10^{-10}$  и после аннигиляции почти всех барионов и антибарионов при  $T < m_B$  упомянутый избыток приведет к возникновению Вселенной со 100%-ной доминантностью барионов (А. Д. Сахаров, 1967). В настоящее время барионная асимметрия Вселенной является единственным «экспериментальным» доказательством несохранения барионов.



Этому вопросу также посвящена отдельная глава. Предварительно нам, однако, нужно будет сделать краткое введение в физику элементарных частиц, а затем перейти к собственно космологической тематике.

### Глава 3.

## ТЕОРИЯ ПОЛЯ. КРАТКОЕ ВВЕДЕНИЕ

В этой главе приведены некоторые сведения из теории поля, необходимые для дальнейшего. Находящийся здесь материал подходит для первоначального поверхностного ознакомления с предметом, но для более глубокого изучения следует обратиться к специальным курсам по теории поля, таким, как, например, книга Ландау и Лифшица (1973) или первая глава книги Боголюбова и Ширкова (1976). Наше изложение построено не в порядке усложнения от скалярного поля к векторному и тензорному, а в историческом: от электродинамики к теории гравитации. К скалярному полю, которое, как показало недавнее развитие теории, играет важную роль в физике частиц и в космологии, мы обратимся в гл. 4 и будем возвращаться позднее при обсуждении механизма инфляции. Такого порядка изложения пока разумно придерживаться, поскольку электромагнитное и гравитационное поля наблюдаются в природе, а фундаментальные скалярные поля еще не открыты.

### § 1. ПОЛЯ В СПЕЦИАЛЬНОЙ ТЕОРИИ ОТНОСИТЕЛЬНОСТИ

Исторически первой теорией поля, по-видимому, можно считать ньютоновскую теорию гравитационного взаимодействия, согласно которой гравитацию можно полностью описать, задав в 3-мерном пространстве\* скалярную функцию (скалярное поле)  $\phi(\mathbf{r})$ , называемую гравитационным потенциалом. Эта теория правильно описывает гравитационные силы в природе при малых относительных скоростях взаимодействующих тел ( $v \ll c$ ). Принципиальным недостатком этой теории является встроенный в нее принцип мгновенной передачи взаимодействия.

Кроме скалярных, в природе могут существовать и более сложные поля — векторные, тензорные и т. п. В случае векторного поля в каждой точке трехмерного пространства задаются три функции  $V_i$ , которые определенным образом преобразуются при переходе от одной трехмерной системы координат к другой. Мы не будем останавливаться на этом подробнее, так как перейдем к специальной теории относительности (СТО), в которой классификация полей производится относительно 4-мерного

\* Смысл этого ограничения 3-мерием будет понятен ниже.

пространства-времени. Сразу же отметим, что та же классификация справедлива и в ОТО.

В физике до появления СТО время и пространственные координаты рассматривались отдельно. При переходе от одной системы координат к другой предполагалось, что время не меняется, оно абсолютно:  $t=t'$ , а пространственные координаты определенным образом преобразовывались друг через друга:  $x \leftrightarrow x'$ . Принцип относительности Галлилея утверждал, что при переходе от одной координатной системы к другой, движущейся относительно первой равномерно и прямолинейно, физические законы не изменяются. При движении вдоль оси  $z$  преобразование координат определялось формулами  $t=t'$ ,  $x=x'$ ,  $y=y'$ ,  $z=z'+vt'$ . С этим была связана классификация полей в терминах трехмерного пространства.

Эта инвариантность была лишь приближенной, отвечающей случаю  $v \ll c$ . При скоростях, приближающихся к скорости света, законы физики по-прежнему не изменяются при переходе от одной инерциальной системы к другой, но преобразования координат изменяют свой вид, так что временные и пространственные координаты выражаются друг через друга:  $t'=f_1(t, x)$ ;  $x'=f_2(t, x)$ . Поэтому мы приходим к 4-мерному пространству-времени.

Преобразования СТО, называемые преобразованиями Лоренца, оставляют неизменным интервал

$$ds^2 = dt^2 - dx_1^2 - dx_2^2 - dx_3^2. \quad (3.1)$$

Их можно формально рассматривать как вращения в 4-мерном пространстве-времени, для которых генераторы поворота в плоскостях  $tx_1$ ,  $tx_2$  и  $tx_3$  являются чисто мнимыми. Физически они отвечают переходу к движущейся вдоль какой-либо из осей  $x_j$  системе координат. Преобразования в плоскостях  $x_i x_j$  являются обычными вращениями. Итак, преобразования Лоренца состоят из трех независимых вращений в плоскостях  $x_i x_j$  и трех независимых движений вдоль осей  $x_j$ , определяемых шестью независимыми параметрами. Кроме этих шести преобразований симметрии в пространстве Минковского допустимы еще четыре, отвечающие произволу в выборе начала системы координат:  $x_\mu \rightarrow x_\mu + a_\mu$ .

При переходе к движущейся системе координат ход времени, как мы знаем, изменяется, поэтому и говорят о пространственно-временных преобразованиях. Эти преобразования обладают важными свойствами инвариантности. В частности, световой конус  $t^2 - \mathbf{r}^2 = 0$  переходит при преобразованиях Лоренца сам в себя. Мировые линии физических частиц (т. е. набор координат  $t, x_j$ ) при любых их движениях остаются внутри светового конуса. Это отвечает невозможности движения со скоростью больше скорости света. При любых преобразованиях

Лоренца точки, находящиеся внутри конуса, продолжают оставаться там. В частности, при этом не может измениться знак времени — преобразования Лоренца не могут заставить идти время вспять (их называют ортохронными преобразованиями; иногда рассматривают расширенную группу преобразований с изменением знака времени, но эти преобразования невозможно реализовать физически).

Вернемся к классификации полей в СТО. Так как при движении временная ось «перепутывается» с пространственными, то классификация полей в 4-мерном пространстве-времени отличается от их классификации в 3-мерном. Например, скаляр в размерности  $D=3$  может быть компонентой вектора или тензора в  $D=4$ . Однако скаляр в  $D=4$  также является скаляром в  $D=3$ . Скаляр можно назвать тензором нулевого ранга. Вектор  $V_\mu$  (или тензор первого ранга) ( $\mu=0, 1, 2, 3$ ) в  $D=4$  распадается на вектор  $V_i$  ( $i=1, 2, 3$ ) и скаляр  $V_0$  в  $D=3$ . Численные значения компонент  $V_0$  и  $V_i$ , очевидно, зависят от выбора системы отсчета. Инвариантной величиной является лишь квадрат 4-вектора  $V_\mu^2 = V_0^2 - V_i^2$ . Симметричный тензор второго ранга  $T_{\mu\nu}$  в  $D=4$  распадается на симметричный тензор  $T_{ij}$ , вектор  $T_{0j}$  и скаляр  $T_{00}$  в  $D=3$ . Заметим, что симметричный тензор  $T_{ij}$  в  $D=3$  (или  $T_{\mu\nu}$  в  $D=4$ ) не является, как говорят, неприводимым: из него можно выделить скалярную часть — след  $T_i^i$  ( $T_{\mu}^{\mu}$ ) и собственно тензорную  $T_{ij} - \frac{1}{3} \delta_{ij} T_k^k$ . Антисимметричный тензор  $F_\mu$  в  $D=4$  распадается на два вектора:  $H_k = \varepsilon_{kij} F_{ij}$  и  $E_k = F_{0k}$  в  $D=3$ .

Тензорные характеристики поля в трехмерном пространстве определяют, как говорят, спин поля  $S$ . Для квантов этого поля — частиц — спин представляет собой собственный момент вращения данной частицы. Очевидно, спин скалярного поля равен нулю. В векторном поле, как мы видели, имеются две части: скаляр с  $S=0$  и вектор с  $S=1$ ; в последнем случае возможны три проекции спина на ось  $z$ :  $S_z=1, 0, -1$ . Из опыта мы знаем, что физические частицы обладают определенным спином, а не смесью их. Поэтому на поле накладывают некоторое дополнительное условие, уничтожающее низшие спины. В квантовой теории поля вероятность процессов, например, с участием векторного поля  $V_\mu$  может оказаться отрицательной, если не устранить состояние с нулевым спином. Так, например, в теории векторного поля требуют  $\partial_\mu V^\mu=0$ ; аналогичные условия можно выписать и для тензорного поля. Однако в теории безмассовых полей, таких, как электромагнитное или гравитационное, имеется дополнительная симметрия, которая сама «следит», чтобы ненужные низшие спины не появлялись. В электродинамике такой симметрией является калибровочная инвариантность, позволяющая добавлять к электромагнитным потен-

циалам градиент скалярной функции при одновременном фазовом преобразовании полей заряженных частиц, а в гравитации — общая ковариантность, т. е. возможность перехода к произвольно движущейся системе координат.

Заметим, что одновременно с уничтожением низших спинов эти симметрии уничтожают и низшие проекции высшего спина. Удобно выбрать направление движения частицы в качестве оси  $z$ , на которую проектируем спин. При таком выборе оси проекция  $S_z$  во время движения не меняется. Массивное векторное поле имеет три степени свободы  $S_z=0, \pm 1$ ; а безмассовое электромагнитное лишь две:  $S_z=\pm 1$ . Аналогично тензорное гравитационное поле тоже имеет лишь два состояния  $S_z=\pm 2$ , а низшие проекции  $S_z=\pm 1, 0$  отсутствуют. Так как кванты этих полей движутся со скоростью света, утверждения об отсутствии низших проекций инвариантны относительно преобразований Лоренца. Очевидно также, что состояния с  $S_z+S$  и  $S_z-S$  не перепутываются при преобразованиях Лоренца, не включающих пространственные отражения.

Общий вид уравнения, описывающего свободное поле, установить весьма просто. Мы знаем из общих соображений, что поля должны удовлетворять некоторому уравнению II порядка, а единственным инвариантным оператором II порядка является оператор Даламбера:

$$\square = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} - \frac{\partial^2}{\partial x_3^2}. \quad (3.2)$$

Отсюда безмассовое скалярное поле должно удовлетворять уравнению

$$\square \varphi = J, \quad (3.3)$$

где  $J$  — некоторый источник этого поля. Не нарушая общих требований инвариантности, в это уравнение можно ввести дополнительный член

$$(\square + m^2) \varphi = J. \quad (3.4)$$

Это уравнение описывает массивное поле, или массивные частицы — кванты этого поля с массой  $m$ . Нетрудно убедиться, что уравнения (3.3) и (3.4) имеют решение вида

$$\exp(-iEt + ikr),$$

только в первом случае  $E^2=k^2$ , а во втором  $E^2=k^2+m^2$ , т. е. как раз релятивистское соотношение между энергией, импульсом и массой частицы.

Обычно теорию поля формулируют на лагранжевом языке, т. е. строят некоторую функцию поля и его первых производных — лагранжиан  $\mathcal{L}$  (точнее, плотность лагранжиана), явля-

ющуюся скаляром относительно преобразований Лоренца. Это автоматически обеспечивает лоренц-инвариантность теории. Далее с помощью  $\mathcal{L}$  определяют функционал действия

$$S = \int dt d^3x \mathcal{L} \equiv \int d^4x \mathcal{L}.$$

Уравнения движения определяются требованием экстремальности действия, т. е. условием того, что функциональная производная  $S$  по полю обращается в ноль:

$$\frac{\delta S}{\delta \varphi} \equiv \frac{\partial \mathcal{L}}{\partial \varphi} - \frac{\partial}{\partial x^\mu} \frac{\partial \mathcal{L}}{\partial \frac{\partial \varphi}{\partial x^\mu}} = 0. \quad (3.5)$$

Легко видеть, что лагранжиан скалярного поля имеет вид

$$\mathcal{L}_\varphi = \frac{1}{2} \left( \frac{\partial \varphi}{\partial x^\mu} \right)^2 - \frac{1}{2} m^2 \varphi^2. \quad (3.6)$$

Источник поля  $J$  можно ввести, добавив в лагранжиан член  $J\varphi$ ; так как лагранжиана плотность  $\mathcal{L}$  по определению является скалярной величиной, то ток должен быть скаляром. Аналогично источник векторных полей, добавляемый в лагранжиан в виде  $A_\mu J^\mu$ , должен быть вектором, а тензорных — тензором.

Для полей с высшими спинами в уравнениях движения могут возникать некоторые дополнительные члены, хотя в конечном счете все фактически сводится к уравнениям типа (3.3) или (3.4).

Пока мы говорили лишь о полях с целыми спинами  $S=0, 1, 2$ . Это так называемые бозонные поля. Теория допускает существование полей с полуцелым спином  $S=1/2, 3/2$ . Такие поля носят название фермионных. Между квантами бозонных и фермионных полей есть принципиальное отличие: две частицы с полуцелым спином — фермионы не могут находиться в одном и том же состоянии. Для бозонов же ситуация прямо противоположна: если в каком-то состоянии уже находится много бозонов, вероятность рождения новых бозонов в этом же состоянии существенно возрастает. На этом основан принцип действия лазера. Упомянутое явление носит название бозе-конденсации. Благодаря этому обстоятельству возможно существование классических бозонных полей, которые с квантовой точки зрения как раз представляют собой конденсат соответствующих частиц. Поэтому очевидно, что классического аналога фермионных полей не существует.

Более аккуратная формулировка отличия фермионов от бозонов состоит в том, что полная волновая функция системы должна быть антисимметричной функцией при перестановке

любой пары одинаковых фермионов (принцип Паули) и симметрична при перестановке бозонов. Отсюда, в частности, следует невозможность пребывания двух тождественных фермионов в одинаковом состоянии и тот факт, что связанное состояние четного числа фермионов является бозоном. Заметим еще, что принцип Паули и бозе-конденсация не могут быть получены за счет какого-то взаимодействия потенциального типа.

Классические электромагнитное и гравитационное поля известны давно, лишь позже было осознано, что существуют отдельные кванты этих полей — фотоны, гравитоны. Что касается фермионов, то они были обнаружены сразу в квантовой форме отдельных частиц.

В заключение этого раздела заметим, что в настоящее время в физике рассматриваются теории поля не только в 4-мерном пространстве, но и в более высоких размерностях  $D=10, 26$  (об этом пойдет речь ниже). В полной аналогии с описанным выше одно поле в размерности 10 или 26 разбивается на большую совокупность различных полей в  $D=4$ . Нужно отметить, кроме того, что лишние  $(D-4)$  размерности компакты и поэтому при малых энергиях ( $E \ll m_p \approx 10^{19}$  ГэВ) не наблюдаемы.

## § 2. ВЕКТОРНОЕ ПОЛЕ. ЭЛЕКТРОДИНАМИКА

Классическая теория электромагнитных взаимодействий, в окончательном виде сформулированная Максвеллом, была первой релятивистски инвариантной теорией поля, хотя и появилась задолго до возникновения теории относительности. Основными объектами теории являются два трехмерно векторных поля (или, как говорят для краткости, 3-векторы) — электрическое ( $E_i$ ) и магнитное ( $H_i$ ) поля. Хотя техника 4-мерных преобразований тогда не применялась в физике, тем не менее было известно, что поля  $E$  и  $H$  взаимосвязаны. Так, например, если в одной системе отсчета  $H \neq 0$ , а  $E=0$ , то при переходе к другой системе, движущейся относительно первой со скоростью  $v$ , появлялось электрическое поле  $E=[v \times H]$  и изменялось магнитное поле. Верно и обратное утверждение: если однородные поля  $E$  и  $H$  перпендикулярны по направлению, то существует система координат, равномерно движущаяся относительно исходной, в которой либо  $E=0$  (если  $|E| < |H|$ ), либо  $H=0$  (если  $|H| < |E|$ ). Эта взаимосвязь  $\mathcal{E}$  и  $\mathcal{H}$  при преобразованиях координат показывает, что они входят в состав некоторого общего 4-мерного тензора. Сказанное в предыдущем параграфе наводит на мысль, что естественным кандидатом на эту роль является антисимметричный тензор  $F_{\mu\nu}$ :

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -H_z & H_y \\ -E_y & H_z & 0 & -H_x \\ -E_z & H_y & H_x & 0 \end{pmatrix}. \quad (3.7)$$

Эта догадка обратится в уверенность, если проверить, что законы преобразования  $\mathbf{E}$  и  $\mathbf{H}$  при переходе от одной системы координат к другой выглядят как результат «вращения» тензора  $F_{ik}$  в четырехмерном пространстве-времени.

Возможность перехода от одной системы координат к другой нередко позволяет добиться существенных технических упрощений в расчетах. Рассмотрим, например, движение электрона в перпендикулярных электрическом и магнитном полях. Если  $|\mathbf{H}| > |\mathbf{E}|$ , то, как уже отмечалось, можно перейти в такую систему, где  $\mathbf{E} = 0$  и имеется лишь одно магнитное поле  $\mathbf{H}'$ , определенным образом связанное с исходными полями. Как известно, в чистом магнитном поле электрон движется по спирали. Сделав преобразование к первоначальной системе координат, можно легко установить форму траектории движения электрона и в ней. Аналогично можно рассмотреть случай  $|\mathbf{E}| > |\mathbf{H}|$ . Случай  $|\mathbf{E}| = |\mathbf{H}|$  и  $\mathbf{E}\mathbf{H} = 0$  отвечает полям, которые одинаково выглядят во всех системах координат, т. е. электромагнитной волне.

Хотя теперь известно, что  $\mathbf{E}$  и  $\mathbf{H}$  — две части тензора второго ранга, мы продолжаем называть электромагнитное поле векторным полем. Это не просто дань истории. Уравнения, которым подчиняется  $F_{ik}$  или эквивалентно  $\mathbf{E}$  и  $\mathbf{H}$ , суть

$$F_{\mu\nu, \lambda} + F_{\lambda\mu, \nu} + F_{\nu\lambda, \mu} = 0, \quad F_{,\nu}^{\mu\nu} = -4\pi j^{\mu} \quad (3.8)$$

в четырехмерных обозначениях. Причем здесь и ниже мы иногда пользуемся известной сокращенной формой записи  $f_{\mu} \equiv \equiv \partial f / \partial x^{\mu}$ . В трехмерных обозначениях эти две пары уравнений Максвелла записываются в виде

$$\operatorname{div} \mathbf{H} = 0, \quad \operatorname{rot} \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, \quad \operatorname{rot} \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \mathbf{j}, \quad \operatorname{div} \mathbf{E} = 4\pi\rho,$$

здесь  $\rho$  — плотность электрического заряда, а  $\mathbf{j}$  — вектор электрического тока.

Мы видим, что имеются шесть величин  $\mathbf{E}$ ,  $\mathbf{H}$ , но не все они независимы. Четыре уравнения  $\operatorname{div} \mathbf{H} = 0$  и  $\operatorname{rot} \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}$  накладывают четыре связи, так что независимыми остаются две компоненты. Введем четырехмерный вектор  $A_{\mu}$  и выразим через него тензор  $F_{\mu\nu}$  согласно

$$F_{\mu\nu} = \frac{\partial A_{\mu}}{\partial x^{\nu}} - \frac{\partial A_{\nu}}{\partial x^{\mu}}. \quad (3.9)$$

Этим представлением мы автоматически удовлетворим условию антисимметрии  $F_{\mu\nu} = -F_{\nu\mu}$  и уравнению  $F_{\mu\nu, \lambda} + F_{\lambda\mu, \nu} + F_{\nu\lambda, \mu} = 0$ . Поэтому шесть величин  $\mathbf{E}$  и  $\mathbf{H}$  можно выразить через четыре компоненты 4-мерного вектора. Именно поэтому электромагнитное поле называется векторным полем. Однако условие (3.9) не определяет  $A_{\mu}$  однозначно. Можно сделать так называемое калибровочное преобразование, т. е. добавить к  $A_{\mu}$  градиент скалярного поля:

$$A_{\mu} \rightarrow A_{\mu} + \frac{\partial f}{\partial x^{\mu}}. \quad (3.10)$$

Преобразование такого вида не изменит тензор  $F_{\mu\nu}$  и, следовательно, сохранит вид уравнений поля и уравнений движения электрона в этом поле. На компоненты вектора  $A_{\mu}$  можно наложить еще одно условие, так что из четырех компонент  $A_{\mu}$  независимыми будут только две.

Электромагнитное поле может быть как связанным с зарядами, например кулоновским, так и свободным полем, полем электромагнитной волны. Мы знаем, что есть радиоволны, есть свет — электромагнитные волны очень высокой частоты. Электромагнитные волны в теории Максвелла могут иметь любую частоту. Например, есть решение уравнений поля вида  $E_x = H_y = u(x-t)$ , где  $u$  — произвольная функция. Такая волна распространяется в вакууме, не искажаясь. В частности, есть решение  $u = \cos[\omega(x-t)]$  при сколь угодно малой частоте  $\omega$ , а следовательно, возможна сколь угодно малая энергия кванта поля с этой частотой. Исчезающе малая частота возможна только у поля, масса покоя которого равна нулю. По этой же причине кулоновское взаимодействие является дальнедействующим,  $E \sim 1/r^2$ . Соответственно поток напряженности поля через удаленную поверхность не падает с расстоянием. Заметим, что поле статических источников убывает все же быстрее ( $\sim 1/r^2$ ), чем поле электромагнитной волны, убывающей как  $1/r$ . Именно поэтому при астрономических наблюдениях (т. е. наблюдениях очень удаленных объектов) мы всегда регистрируем электромагнитные волны (радиоволны, свет, рентгеновское излучение), а не статическое электрическое или магнитное поле, которые вблизи звезды могут заметно превышать поле излучения. Нужно заметить, впрочем, что электрические заряды астрономических объектов незначительны, поэтому кулоновское поле мало не только из-за фактора  $r^{-2}$ , но и из-за малости суммарного заряда. Вблизи пульсаров, впрочем, электрическое поле может быть весьма значительным, до  $10^{14}$  В/см. Оно порождается магнитным полем пульсара, достигающим  $10^{13}$  Гс, при

быстром вращении последнего. Однако, так как пульсар не обладает ни существенным электрическим, ни магнитным зарядом, эти поля убывают на больших расстояниях быстрее, чем  $r^{-2}$ , и вдали от пульсара незаметны. Наличие таких полей приводит к ускорению заряженных частиц до энергий  $10^{20}$  эВ (при характерном размере порядка  $10^6$  см).

Следует отметить еще одно чрезвычайно важное следствие уравнений Максвелла. Сами уравнения накладывают определенные условия на источники поля. Рассмотрим пару уравнений Максвелла:

$$\operatorname{div} \mathbf{E} = 4\pi\rho, \quad \operatorname{rot} \mathbf{H} = \frac{4\pi}{c} \mathbf{j} + \frac{4\pi}{c} \frac{\partial \mathbf{E}}{\partial t}.$$

Применив оператор  $\operatorname{div}$  ко всем членам второго уравнения и подставив первое, получим связь между плотностью заряда и током.

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{j} = 0. \quad (3.11)$$

Поэтому токи и заряды нельзя задавать произвольно. Они должны удовлетворять соотношению (3.11), которое называется законом сохранения заряда, или уравнением неразрывности для зарядов.

Подчеркнем, что закон сохранения электрического заряда является следствием уравнений электродинамики. Поскольку до сих пор ни в многочисленных практических применениях электромагнетизма, ни в специально поставленных опытах не обнаружено отклонений от теории Максвелла, мы твердо уверены в том, что электрический заряд сохраняется. Или, иными словами, ответ на вопрос о несохранении электрического заряда в рамках электродинамики Максвелла является однозначно отрицательным.

Закон сохранения (3.11) можно записать в 4-мерной форме, если учесть, что две величины: 3-вектор  $j_k$  и 3-скаляр  $\rho$  составляют 4-вектор  $j_\mu$ :

$$\partial_\mu j^\mu = 0. \quad (3.12)$$

Тот факт, что  $j_\mu$  является 4-вектором, можно увидеть из второго уравнения (3.8), так как 4-дивергенция тензора второго ранга  $F_{\mu\nu}$  превращает его в 4-вектор.

Часто бывает удобно записывать уравнения электродинамики не в терминах напряженности поля  $F_{\mu\nu}$ , а через вектор-потенциал  $A_\mu$ . Они имеют особенно простой вид, если за счет калибровочной свободы (3.10) наложить на  $A_\mu$  условие  $\partial_\mu A^\mu = 0$ . В этом случае одно векторное уравнение

$$\square A_\mu = 4\pi j_\mu \quad (3.13)$$

эквивалентно всей системе уравнений Максвелла. Заметим, что оно имеет формально такой же вид, как уравнение для скалярного поля (3.3).

Уравнения (3.8) или эквивалентные им (3.13), очевидно, не полностью определяют теорию. С их помощью можно вычислить поле  $F_{\mu\nu}$  или потенциалы  $A_\mu$  по заданному току  $j_\mu$ . Кроме того, необходимо уметь находить движение частиц в электромагнитном поле, или, иными словами, нужно знать уравнения этих частиц с учетом электромагнитного взаимодействия. Эти взаимодействия однозначно фиксируются принципом калибровочной инвариантности. Именно мы требуем, чтобы теория была инвариантной при преобразованиях (3.10), сопровождающихся изменением фазы заряженного поля согласно  $\psi \rightarrow \exp(ief)\psi$ , где  $e$  — заряд поля  $\psi$ . Отсюда следует, что электромагнитные взаимодействия вводятся с помощью простого «удлинения» производных: все производные в уравнениях свободного поля нужно заменить согласно

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - ieA_\mu. \quad (3.14)$$

Величина  $D_\mu$  называется ковариантной производной.

Такой способ построения взаимодействий лежит в основе почти всей современной физики элементарных частиц. Аналогичная идея применяется и в гравитации. Имея в виду приложения к гравитации, которая, как мы увидим ниже, описывается тензорным полем второго ранга, приведем выражение для тензора энергии-импульса электромагнитного поля, который является источником гравитационного поля. Процедура построения тензора энергии-импульса описана в цитированных в начале этой главы книгах. Мы приведем лишь окончательный результат:

$$T_{\mu\nu} = \frac{1}{4} g_{\mu\nu} F_{\alpha\beta} F^{\alpha\beta} - F_{\mu\alpha} F^\alpha_\nu.$$

Заметим, что величина  $T_{\mu\nu}$  сохраняется:

$$\frac{\partial T^\nu_\mu}{\partial x^\nu} = T^\nu_{\mu,\nu} = 0,$$

в чем нетрудно убедиться, используя уравнение Максвелла (3.8). Таково общее свойство тензора энергии-импульса. Помимо этого тензор энергии-импульса электромагнитного поля обладает еще одним замечательным свойством:

$$T^\mu_\mu = 0.$$

Это условие отражает конформную инвариантность теории электромагнитного поля. Из него, в частности, следует, что для фотонного газа

$$p = e/c.$$

Упомянем об очень красивом эффекте, связанном с калибровочной инвариантностью электромагнитного взаимодействия. При движении заряженной частицы в электромагнитном поле фаза ее волновой функции изменяется на величину

$$\Delta\varphi = e \int A_\mu dx^\mu,$$

где интеграл берется вдоль траектории частицы. Заметим, что это выражение согласовано с изменением волновой функции при калибровочном преобразовании, указанном выше. Если рассматривать интерференционную картину при распространении частиц в магнитном поле ( $A_0 = 0$ ,  $\mathbf{A} \neq 0$ ) по двум возможным путям (рис. 5), то она определяется разностью фаз:

$$\Delta\varphi_{12} = e \oint \mathbf{A} dx,$$

где интеграл берется по замкнутому контуру  $A1B2A$ . Таким образом, наблюдаемый эффект явно зависит от, казалось бы, фиктивной величины вектор-потенциала  $A_\mu$ . Впрочем, величина  $\Delta\varphi_{12}$  явно калибровочно инвариантна: при преобразовании (3.10) она, очевидно, не меняется. Это можно увидеть и по-другому, воспользовавшись известной теоремой векторного анализа:

$$\oint \mathbf{A} dx = \int_S ds \operatorname{rot} \mathbf{A} = \int_S d\mathbf{s} \mathbf{H},$$

которая позволяет выразить интеграл от циркуляции вектор-потенциала по замкнутой кривой через поток магнитного поля сквозь охватываемую этой кривой поверхность. Казалось бы, удивляться нечему: все в конечном счете свелось к напряженности поля. Однако мы можем поставить опыт так, что магнитное поле будет находиться внутри ограниченной области, например в длинном тонком соленоиде, а траектории частиц будут проходить вдали от этого соленоида, но тем не менее дифракционная картина будет зависеть от поля в соленоиде. Этот эффект, называемый эффектом Бома—Аронова, наблюдался на эксперименте в полном соответствии с теорией.

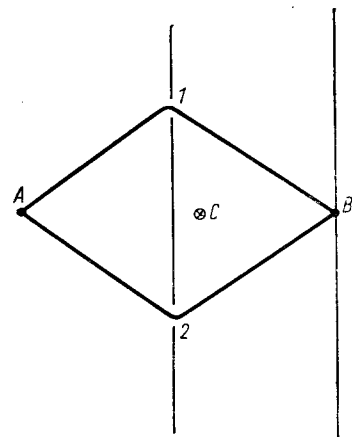


Рис. 5. Интерференционный опыт, иллюстрирующий эффект Бома—Аронова,  $A1B$  и  $A2B$  — два возможных пути движения частицы (электрона);  $C$  — соленоид, создающий внутри магнитное поле

Уравнения Максвелла в форме (3.8) или (3.13) на микроскопическом уровне справедливы всегда, однако в макроскопической электродинамике материальной среды вид уравнений Максвелла меняется. К «видимым» зарядам добавляются заряды от поляризации среды электромагнитным полем. В принципе эти уравнения могут быть выведены, как говорится, из первых принципов, т. е. из исходных уравнений (3.8) (или (3.13)) и из фундаментальных уравнений, описывающих свойства среды. На практике же обычно поступают иначе, задавая свойства среды феноменологическими параметрами, такими, как диэлектрическая или магнитная проницаемость, и т. д. Предметом макроскопической электродинамики являются вопросы преломления и дисперсии света, черенковского излучения и многое другое, что здесь просто невозможно перечислить.

Оказывается, однако, что и вакуум поляризуется под воздействием внешних полей. Это связано с рождением из вакуума и последующим исчезновением пар заряженных частиц, взаимодействующих с внешним полем. Правда, такое наглядное описание является не вполне точным. Рождение и исчезновение виртуальных пар в вакууме не означает его нестационарности. Роль внешнего поля состоит в модификации стационарных волновых функций. Это можно понять на примере атома водорода, который может находиться в каком-либо из стационарных состояний

$$\psi_k e^{-i\omega_k t}.$$

Суперпозиция таких состояний нестационарна, происходит излучение. Однако атом во внешнем поле может иметь стационарную волновую функцию вида

$$(C_1\psi_1 + C_2\psi_2 + \dots) e^{-i\omega' t}.$$

Это, конечно, не означает, что атом «перепрыгивает» из состояния  $\psi_1$  в  $\psi_2$  и т. п. Тем не менее при быстром снятии поля атом окажется в нестационарном состоянии

$$C_1\psi_1 e^{-i\omega_1 t} + C_2\psi_2 e^{-i\omega_2 t} + \dots$$

Иными словами, возникла суперпозиция основного и возбужденного состояний. Это возбуждение возникает из-за нестационарности (быстрого выключения) поля.

В силу этого возникают (небольшие) поправки к уравнениям Максвелла (3.8), не меняющие, однако, их основных свойств, таких, как сохранение тока или калибровочная инвариантность. При этом уравнения становятся нелинейными, возникает, например, эффект двойного лучепреломления — коэффициент преломления при распространении света в магнитном поле оказывается различным для разных линейных поляризаций света. И все богатство этих явлений возникает из уравне-

ний (3.8) при усреднении либо по материальной среде, либо по вакууму. (В последнем случае уравнения (3.8) следует понимать, как операторные, описывающие квантованные электромагнитные и материальные поля. С этой оговоркой уравнения (3.8) верны всегда.) Разумеется, для полного описания системы, кроме уравнений электромагнитного поля, необходимы, как мы уже говорили, уравнения движения материи в этом поле.

Мы завершим этот параграф небольшим замечанием о том, что в однородной и изотропной Вселенной средние значения  $\Delta$  и  $H$  должны быть равны нулю, что, очевидно, верно вследствие симметрии. Несколько менее тривиально то, что для замкнутой Вселенной полный заряд (или, что то же, средняя плотность заряда  $\rho$ ) должен быть равен нулю. Это утверждение следует из теоремы Гаусса, если ее применить к двум частям Вселенной, на которые ее делит произвольная замкнутая поверхность.

### § 3. ТЕНЗОРНОЕ ПОЛЕ. ТЕОРИЯ ГРАВИТАЦИИ

Успех максвелловской электродинамики побудил строить релятивистские теории и других полей по аналогии с электромагнетизмом. В начале века было известно еще одно поле — гравитационное. Хотя гравитационное взаимодействие было открыто раньше электромагнитного, а закон всемирного притяжения был сформулирован раньше закона Кулона, релятивистской теории гравитации не существовало. Была теория гравитационного потенциала, описываемая уравнением:

$$\Delta\varphi_{\text{ньют}} = -4\pi G\rho, \quad (3.15)$$

где  $\rho$  — плотность массы. Это уравнение очень похоже на уравнение для электростатического потенциала

$$\Delta\varphi_{\text{эл}} = 4\pi\rho_e$$

с той только разницей, что вместо плотности массы с правой части стоит плотность электрического заряда.

Однако между электрическим и гравитационным полем имеется важное различие, состоящее в том, что два одноименных электрических заряда отталкиваются, а два одноименных гравитационных заряда (две массы) притягиваются. Это означает, что теории гравитационного и электромагнитного полей должны быть существенно различны. Требование положительной определенности энергии в формализме релятивистской теории поля позволяет заключить, что к притяжению одноименных зарядов приводят поля с четным спином, т. е. скалярное или тензорное II ранга, а к отталкиванию — поля с нечетным спином, например, векторное (электромагнитное). В этом месте мы просим читателя поверить нам на слово. Ниже мы увидим, что

релятивистская формулировка теории гравитации требует, чтобы гравитационное поле описывалось либо скалярной функцией, либо тензором II ранга. Релятивистское обобщение уравнения (3.15) состоит, во-первых, в замене оператора Лапласа на релятивистски инвариантный оператор Даламбера (3.2) и, во-вторых, в установлении тензорных свойств стоящего в правой части источника, а следовательно, тензорных свойств и самого поля.

При релятивистском обобщении ньютоновской теории гравитации, определяемой уравнением (3.11), необходимо конкретизировать тензорные свойства источника  $4\pi G\rho$  при преобразованиях Лоренца. Величина  $\rho$  может быть нерелятивистским пределом либо тензора энергии-импульса  $T_{\mu\nu}$ , либо скаляра  $T^{\nu}_{\nu}$ . В первом случае гравитационное поле должно быть тензорным, во втором — скалярным. Эксперимент решает вопрос в пользу тензорного случая.

Рассмотрим, как преобразуется плотность энергии при переходе от покоящейся системы отсчета к равномерно движущейся, и сравним закон ее преобразования с законом преобразования четырехмерного вектора ( $\rho$ ,  $\mathbf{j}$ ) — плотности зарядов и тока.

Как известно, объем тела в движущейся системе отсчета сокращается:

$$V = V_0 \sqrt{1-v^2},$$

где  $V_0$  — собственный объем тела. Заряд при переходе в движущуюся систему отсчета остается неизменным, а масса увеличивается согласно

$$m = m_0 / \sqrt{1-v^2}.$$

Поэтому плотность энергии изменяется пропорционально второй степени  $\gamma = (1-v^2)^{-1/2}$ , т. е.

$$\rho_m = \rho_{0m} \gamma^2,$$

в отличие от плотности электрического заряда, пропорциональной первой степени

$$\rho_e = \rho_{0e} \gamma.$$

По сравнению с преобразованием векторной величины (плотности заряда) в преобразовании плотности массы появляется лишняя степень лоренц-фактора. С формальной точки зрения это значит, что  $\rho$  ведет себя как 00-компонента тензора второго ранга.

Мы знаем такой тензор. Это четырехмерный тензор энергии-

импульса-натяжений. Он обозначается,  $T_{\mu\nu}$  и является симметричным тензором второго ранга. Компонента  $T_0^0$  представляет собой плотность энергии,  $T_i^0$  ( $i=1, 2, 3$ ) — поток энергии или плотность импульса. Компоненты с двумя пространственными индексами являются трехмерным тензором натяжений второго ранга. Весьма простой вид тензор энергии-импульса-натяжений имеет, например, для невязкой жидкости. В этом случае его матрица представляет собой

$$\begin{aligned} T_0^0 &= \varepsilon, \\ T_i^0 &= j_i = (\rho + p) v_i, \\ T_j^i &= (\rho + p) v_i v^j - p \delta_j^i. \end{aligned}$$

Для покоящейся жидкости ( $v_i=0$ )  $\rho$  — плотность энергии, а  $p$  — давление.

Если теперь мы хотим строить релятивистскую теорию гравитации, то в качестве источника поля в уравнение Даламбера (3.3) нужно подставить тензор  $T_{\nu}^{\mu}$ . В силу релятивистской ковариантности гравитационное поле должно описываться, очевидно, также тензором второго ранга  $\psi_{\nu}^{\mu}$ :

$$\square \psi_{\nu}^{\mu} = 4\pi G T_{\nu}^{\mu}. \quad (3.16)$$

Для полного определения гравитационного поля необходимо знание функций  $\psi_{\nu}^{\mu}$ . Однако не все 10 функций  $\psi_{\nu}^{\mu}$  независимы. Условие сохранения  $T_{\nu}^{\mu}$ , записываемое в виде

$$\partial_{\mu} T_{\nu}^{\mu} = 0, \quad (3.17)$$

налагает аналогичное условие на  $\psi_{\nu}^{\mu}$ :

$$\partial_{\mu} \psi_{\nu}^{\mu} = 0.$$

Обычно гравитационное поле выражают не в терминах  $\psi_{\nu}^{\mu}$ , а посредством величин  $h_{\nu}^{\mu}$

$$h_{\nu}^{\mu} = \psi_{\nu}^{\mu} - \frac{1}{2} \delta_{\nu}^{\mu} \psi_{\alpha}^{\alpha}, \quad (3.18)$$

где  $\delta_{\nu}^{\mu} = \text{diag}(1, 1, 1, 1)$  — единичный тензор. Отметим симметрию  $T_{\mu\nu} \equiv g_{\mu\alpha} T_{\nu}^{\alpha}$  и, следовательно, симметрию  $\psi_{\mu\nu}$ . Гравитационный потенциал ньютоновской теории  $\varphi = 2h_0^0$  удовлетворяет уравнению

$$\square \varphi = 4\pi G T_0^0. \quad (3.19)$$

Другая возможность при построении релятивистской теории гравитации состоит в выборе в качестве источника поля скалярной величины  $T_{\alpha}^{\alpha}$  — следа тензора энергии-импульса.

В этом случае величина  $\varphi$  также должна быть скаляром. Эта так называемая скалярная теория гравитации должна быть, однако, отброшена, так как она противоречит наблюдениям, в частности тому, что электромагнитные волны отклоняются в гравитационном поле, несмотря на то что для них  $T_{\alpha}^{\alpha} = 0$ .

Заметим, что для стационарных нерелятивистских систем отличить скалярную теорию тяготения от тензорной очень трудно. Различие проявляется лишь в весьма слабых, трудно поддающихся измерению эффектах. Близкое совпадение двух теорий обязано соотношению

$$\int T_{\alpha}^{\alpha} dV = m_{\text{tot}}, \quad (3.20)$$

справедливому для стационарной системы. Если, например, рассмотреть сферическую полость, заполненную фотонным газом, то для внутренней части полости  $T_{\alpha}^{\alpha} = 0$ , однако благодаря давлению газа в стенках полости возникают напряжения, которые дают ненулевой вклад в  $T_{\alpha}^{\alpha}$ , обеспечивающие выполнение соотношения (3.20).

Хотя скалярная теория гравитации логически допустима, опыт показывает, что справедлива именно тензорная теория. В настоящее время она с высокой точностью,  $\sim 0,1\%$ , подтверждена наблюдениями. Здесь речь идет о точности в отклонениях от предсказаний ньютоновской теории, которые сами по себе чрезвычайно малы.

Сейчас тензорная теория является общепризнанной теорией гравитации, она даже стала необходимым инструментом при расчетах в космической навигации.

Чтобы не вводить читателя в заблуждение, нужно сразу же отметить, что уравнения (3.16) никоим образом не могут быть точными уравнениями гравитации по следующей очевидной причине. Источником гравитационного поля  $\psi_{\nu}^{\mu}$  служит сохраняющийся тензор энергии-импульса материи. Однако поле  $\psi_{\nu}^{\mu}$  само обладает энергией, которая может переходить от поля  $\psi_{\nu}^{\mu}$  к материи и обратно. С учетом этого  $T_{\nu}^{\mu}$  уже не будет сохраняться; в него нужно добавить члены, отвечающие энергии гравитационного поля. В итоге уравнения (3.16) нужно будет заменить на некоторые более сложные нелинейные уравнения, учитывающие самодействие гравитационного поля и изменение  $T_{\nu}^{\mu}$  в гравитационном поле. Уравнения (3.16) можно рассматривать как предел точной теории в случае слабого гравитационного поля.

В этом пункте имеется глубокая аналогия между гравитацией и теорией векторных полей, если сами эти поля являются заряженными, обладают в силу этого самодействием и описываются нелинейными уравнениями. Теории такого рода носят название неабелевых калибровочных теорий (в отличие от абелевой теории — электродинамики) и с ними связан замечатель-



ный прогресс последних лет в физике элементарных частиц (см. гл. 4).

Взаимодействие материи с гравитационным полем определяется принципом эквивалентности, который фактически гласит, что ускорения всех материальных тел в гравитационном поле одинаковы. Несколькими иначе его можно сформулировать как возможность выбора в любой точке пространства-времени в произвольном гравитационном поле такой системы координат, что гравитационное поле в достаточно малой окрестности этой точки будет отсутствовать. Исторически Эйнштейн пришел к релятивистской теории гравитации как раз в такой геометрической формулировке. Наше упрощенное изложение, пригодное для описания слабых полей, опиралось на понятие плоского псевдоевклидова пространства-времени (пространства Минковского), в котором гравитация рассматривалась, как внешнее поле. В принципе можно обобщить этот подход на случай произвольно сильного поля. Однако плоское пространство-время оказывается ненаблюдаемым, фиктивным. Так как все материальные поля и частицы взаимодействуют с гравитацией универсальным образом и нет частиц, которые были бы нейтральны по отношению к гравитации, то не существует физических объектов, которые могли бы «начертить» нам геометрию мира Минковского. В таком подходе не виден принцип эквивалентности, локальная скорость света не постоянна, а меняется с изменением величины гравитационного поля, также изменяется частота света, испускаемого возбужденными атомами, и даже их размер. Тем не менее адекватная и непротиворечивая теория гравитации на фоне плоского пространства-времени может быть построена и не исключено, что такой формализм окажется полезным, например, для супергравитации. В настоящее время такую теорию в противовес ОТО развивает А. А. Логунов с сотрудниками. В статье Я. Б. Зельдовича и Л. П. Грищука (1986) подробно обсуждаются принципиальные трудности такого подхода и показана правильность и необходимость геометризованной ОТО. Еще раньше Л. Д. Фадеев (1982) показал отсутствие противоречий в вопросе об энергии и импульсе гравитационного поля в ОТО.

Вернемся к общей теории относительности. Хотя мы не ставим себе здесь задачу изложения деталей геометрической теории тяготения, тем не менее приведем точные уравнения гравитационного поля в силу исключительной красоты последних:

$$R_{\mu\nu} - 1/2 g_{\mu\nu} R = 8\pi G T_{\mu\nu}, \quad (3.21)$$

где  $G = m_{\text{Pl}}^{-2}$  — постоянная гравитационного взаимодействия;  $T_{\mu\nu}$  — тензор энергии-импульса, источник гравитационного поля. В левой части уравнений стоят чисто геометрические величины:  $g_{\mu\nu}$  — метрический тензор пространства-времени, определяющий

интервал согласно

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu;$$

$R_{\mu\nu}$  — тензор Риччи, определенным образом построенный из  $g_{\mu\nu}$  и его первых и вторых производных;  $R = g^{\mu\nu} R_{\mu\nu}$  — скаляр кривизны.

Введенная выше величина  $h_{\mu\nu}$  представляет собой отклонение метрики в присутствии материальных гравитирующих тел от метрики плоского пространства:  $h_{\mu\nu} = g_{\mu\nu} - g_{\mu\nu}^{(0)}$ . Если разложить уравнения (3.21), удерживая лишь члены первого порядка по  $h_{\mu\nu}$ , то мы получим уравнения (3.16).

Аналогом калибровочной инвариантности электродинамики в гравитации является принцип общей ковариантности, отражающий свободу в выборе различных координатных систем. Этот принцип фактически лежит в основе ОТО, и вся теория может быть построена с помощью этого принципа.

Перечислим принципиальные свойства тензорной теории гравитационного поля и полностью эквивалентной ей ОТО. Тензорная теория дает универсальное притяжение. Она автоматически приводит к ковариантному (вспомним об удлинении производной) закону сохранения тензора энергии-импульса.

Если источником гравитационного поля являются частицы одного сорта или одноединственное поле (например, электромагнитное), то из законов сохранения энергии-импульса следуют уравнения движения этих частиц или поля. Заметим, что в отсутствие гравитации можно сконструировать такой сохраняющийся тензор второго ранга, закон сохранения которого не налагает никаких условий на исходное поле.

Таким образом, ОТО является самосогласованной теорией: источники порождают гравитационное поле, которое в свою очередь определяет движение источников. Этот чрезвычайно важный вывод породил идею свести все взаимодействия к геометрии и создать геометрическую единую теорию поля. Однако в такой простой форме это не удалось сделать. Если есть несколько сортов частиц или полей, то одного только закона сохранения энергии-импульса недостаточно для вывода уравнений движения всех полей и частиц.

Поясним, что имеется в виду. Например, во время ядерной реакции в куске урана давление сильно увеличивается. Ядра урана распадаются на более легкие ядра, при этом плотность остается неизменной. Благодаря давлению появляется сила и частицы куска урана разлетаются, образование давления в куске урана не определяется законами ОТО. Поэтому уравнения для различных реакций: химических, ядерных, с элементарными частицами, не могут быть получены из ОТО, а являются дополнительными уравнениями, описывающими взаимодействия различных сортов частиц.

Впрочем, нельзя исключить того, что все эти внешние уравнения, которые мы вынуждены вводить, являются следствием нашего незнания природы единого фундаментального поля, уравнение движения которого (согласованное с ОТО) содержит всю информацию о нашем мире, так что, возможно, единая геометрическая теория еще возродится, хотя наблюдаемое сейчас обилие типов частиц приводит к значительным трудностям при попытках реализации этой идеи.

#### § 4. БЕЗМАССОВЫЕ И МАССИВНЫЕ ПОЛЯ. ДАЛЬНОДЕЙСТВУЮЩИЕ СИЛЫ

Несмотря на отмеченные выше существенные различия, гравитационные и электромагнитные взаимодействия обладают одним важным общим свойством: и те и другие силы являются дальнедействующими; при удалении от источника они убывают по закону  $r^{-2}$  (а соответствующие потенциалы — как  $r^{-1}$ ). Благодаря этому свойству возможно существование атомных и планетарных систем и в конечном счете существование самой жизни. Гравитационное и электромагнитное дальнедействие непосредственно связано с тем, что кванты этих полей, т. е. гравитоны и фотоны, являются безмассовыми. Действительно, уравнение (3.3), описывающее безмассовое поле, имеет решение  $\varphi \sim r^{-1}$ . А при  $m \neq 0$  (см. (3.4)) решение превращается в

$$\varphi \sim \exp(-mr)/r, \quad (3.22)$$

т. е. при  $r > m^{-1}$  поле практически исчезает.

Квантовая теория обучила нас тому, что безмассовость какого-то поля является весьма нетривиальным свойством. Дело в том, что, начав со свободного безмассового поля, необходимо наложить на теорию весьма жесткие требования, которые бы гарантировали, что взаимодействие не приводит к возникновению массы. Если об этом специально не позаботиться, то кванты первоначально безмассового поля при своем распространении, испуская и вновь поглощая (или превращаясь в) кванты других полей обязательно станут массивными.

И в электродинамике, и в гравитации в теорию исходно, т. е. еще на классическом, доквантовом уровне, был встроены как раз такой принцип, который запрещал появление массы и на квантовом уровне (т. е. появление членов  $m^2 A_\mu^2$  в лагранжиане). Это были уже упомянутые принцип калибровочной инвариантности в электродинамике и принцип общей ковариантности в теории тяготения. Отметим глубокую связь между этими принципами и сохранением электрического тока в первом случае и сохранением тензора энергии-импульса — во втором. Оба закона сохранения являются следствиями упомянутых принципов. Обратное утверждение несправедливо, и, в

частности, можно построить теорию массивного векторного поля, взаимодействующего с сохраняющимся током.

Возникает естественный вопрос: исчерпываются ли все дальнедействующие силы в природе электромагнитными и гравитационными или существуют еще какие-то, пока необнаруженные поля и соответствующие им безмассовые частицы?

Космологам хорошо известны теории, в которых предполагается существование еще одного (на сей раз скалярного) безмассового поля. Это так называемые скалярно-тензорные теории гравитации (как, например, теория Бранса-Дике), в которых к тензорному гравитационному  $h_{\mu\nu}$  полю добавляется еще небольшая примесь скалярного  $\varphi$ . Однако в соответствии со сказанным выше учет квантовых эффектов с необходимостью должен привести к тому, что поле  $\varphi$  приобретает массу, поэтому создаваемый им потенциал уже не будет дальнедействующим и на макроскопических расстояниях не проявится. Оценки величины возникающей при этом массы поля  $\varphi$  весьма неопределенны и варьируются от  $m_\varphi = m_{\text{Pl}}$ , что отвечает радиусу взаимодействия  $l = 10^{-34}$  см, до  $m_\varphi = m_0/m_{\text{Pl}}$ , где  $m_0$  — некоторая характерная для используемой модели масса. Если  $m_0$  связана с характерным масштабом нарушения суперсимметрии (см. гл. 4), т. е.  $m_0 = 10^2 - 10^3$  ГэВ, то  $l = 0,1 - 10$  см; если же  $m_0$  связана с характерными масштабами сильного взаимодействия, т. е.  $m_0 \approx 0,1$  ГэВ, то  $l = 10^7$  см. Наличие такого поля не было заметно в астрономических масштабах и поэтому от астрономических проверок теории Бранса — Дике и других подобных теорий следует ждать отрицательного результата, как мы это и наблюдаем. Существование поля  $\varphi$  можно было бы заметить по вариации гравитационной константы с расстоянием с характерным масштабом  $l = m_\varphi^{-1}$ .

Существование еще одного безмассового тензорного поля, помимо гравитационного, исключено, так как безмассовое тензорное поле должно иметь сохраняющийся источник, а мы знаем (на этот счет существует специальная теорема), что имеется лишь один сохраняющийся тензор  $\Pi$  ранга — тензор энергии-импульса.

Для векторных полей свободы заметно больше. Помимо электромагнитного, можно сконструировать множество других сохраняющихся векторных токов и задаться вопросом, не существует ли связанного с этими токами дальнедействия. Свыше 30 лет назад Т. Д. Ли и Ч. Н. Янг (1955) проанализировали вопрос о наличии дальнедействия, связанного с барионным зарядом  $B$ . Это заряд, который приписывается протонам, нейтронам и некоторым другим частицам, чтобы феноменологически описать наблюдаемое на опыте их сохранение (см. гл. 4).

Заметим, что если бы существовало дальнедействие, связанное, например, с барионным зарядом, то мы бы могли узнать,

сколько барионов находится в теле, не пересчитывая их непосредственно, а по теореме Гаусса.

Было, однако, показано, что даже если барионное дальноедействие существует, его сила должна быть ничтожно малой. Этот вывод основан на экспериментах по проверке принципа эквивалентности. Согласно этому принципу гравитационная масса равна инертной, т. е. гравитационное ускорение всех тел должно быть одинаково независимо от вида вещества. Это равенство ускорений для притяжения к Солнцу было проверено с точностью  $10^{-12}$  В. Б. Брагинским и В. И. Пановым (1971). Если бы существовали дальнедействующие (т. е. падающие как  $1/r^2$ ) силы взаимодействия между барионными зарядами, то утверждение равенства ускорений различных тел было бы несправедливо. Последнее связано с тем обстоятельством, что масса ядра не строго пропорциональна его барионному заряду, причем отклонение от пропорциональности может достигать нескольких тысячных. Отрицательные результаты этих опытов говорят, что барионное поле (если оно существует) взаимодействует слабее, чем гравитационное поле в  $10^{-3}/10^{-12}=10^{-9}$  раз, слабее электрического в  $10^{45}$  раз, так что по всей видимости такого взаимодействия нет.

Нужно, однако, помнить, что в физике могут возникать очень малые (или очень большие) числа. Мы знаем примеры, когда в квантовой теории поля возникают факторы типа  $\exp(-1/\alpha)$ , где  $\alpha \approx 10^{-2}$  — величина характерной константы связи в теории. Поэтому, строго говоря, мы никогда не можем абсолютно исключить существование того или иного взаимодействия, а можем говорить лишь об ограничении на его силу, основываясь на данных эксперимента.

Заметим, что если бы векторное поле обладало массой, то оно приводило бы к потенциалу взаимодействия заряженных частиц типа (3.22) вместо  $u \sim r^{-1}$ . Ограничения на силу взаимодействия в этом случае, т. е. на величину константы связи  $e_b$ , зависят от  $m$  и для больших  $m$  оказываются весьма слабыми.

Если в соответствии со сказанным выше не существует барионного дальнегодействия, то ничего не мешает предположить, что барионный заряд не является строго сохраняющимся, что, как мы уже отмечали, необходимо для объяснения наблюдаемой барионной асимметрии Вселенной.

В заключение подчеркнем, что хотя утверждение о гравитационном и электромагнитном дальнедействии покоятся на прочном теоретическом фундаменте, последнее слово должно оставаться за экспериментом. Закон обратных квадратов (т. е.  $F \sim r^{-2}$ ) в гравитации заведомо не проверен на галактических расстояниях. В этой связи естественно возникает мысль, нельзя ли объяснить наблюдаемую динамически скрытую массу Вселенной модификацией законов гравитационного взаимодей-

ствия. Материал этого параграфа показывает, что это едва ли возможно с помощью дополнительных безмассовых полей. Если изменение гравитации на малых расстояниях  $r \sim m_{\pi}^{-1}$  не встречается с затруднениями ни с теоретической, ни тем более с экспериментальной точки зрения, то изменение гравитационного взаимодействия на большом расстоянии если и возможно, то с помощью какой-то весьма нетривиальной модификации существующей теории.

## \*Глава 4.

### ЭЛЕМЕНТАРНЫЕ ЧАСТИЦЫ

Эта глава представляет собой краткое введение в физику элементарных частиц. Она знакомит читателя-астронома с основными понятиями теории, с ее словарем, с тем, чтобы читатель, не имеющий специального образования в этой области физики, мог легче ориентироваться в последующих главах. В качестве дополнительного чтения к материалу этой главы можно рекомендовать три книги Окуня (1985, 1983 и 1981), расположенные в порядке нарастания степени сложности. Первая из них представляет превосходное популярное введение в физику частиц, доступное школьнику, в то время как последняя ориентирована уже на профессионала или на готовящегося им стать.

### § 1. СКАЛЯРНЫЕ ПОЛЯ. $\pi$ -МЕЗОНЫ

Хотя идея о скалярном поле не нашла своего применения в гравитации, она с успехом была использована в 1937 г. Юкавой для описания ядерных сил. Он записал уравнения поля в виде

$$\square \phi = 4\pi g n_{\text{бар}} + \mu^2 \phi, \quad (4.1)$$

здесь  $g$  — константа взаимодействия ядерных сил. В качестве источника поля стоит скалярная величина  $n_{\text{бар}}$ , которая в нерелятивистском пределе совпадает с плотностью числа барионов. Добавка  $\mu^2 \phi$  в уравнении приводит к тому, что кванты  $\phi$ -поля оказываются массивными, а ядерные силы короткодействующими. Действительно, для статического, сферически-симметричного случая уравнение (4.1) имеет решение вида

$$\phi \sim e^{-\mu r}/r.$$

Благодаря экспоненциальному фактору потенциал очень быстро убывает при  $r > \mu^{-1}$  и на расстояниях  $r > (2 \div 3) \mu^{-1}$  от бариона ядерные силы фактически уже не действуют. Для того чтобы получить согласие с экспериментальными данными о ядре, ве-

личину  $\mu$  нужно было выбрать порядка обратного размера ядра, т. е.  $10^{13}$  см<sup>-1</sup>, что отвечает массе квантов поля  $\phi$  около 100 МэВ.

Уравнение (4.1) имеет также волновые решения, аналогичные электромагнитным волнам:

$$\phi = \phi_0 e^{i\omega t - i \mathbf{k} \cdot \mathbf{r}}.$$

Связь между частотой  $\omega$  и волновым вектором  $\mathbf{k}$  для волны называется дисперсионным соотношением. Для безмассового поля, например электромагнитного, дисперсионное соотношение имеет вид

$$\omega^2 = k^2, \quad (4.2)$$

что означает возможность любой сколь угодно малой частоты  $\omega$ . Дисперсионное уравнение для массивного поля имеет вид

$$\omega^2 = \mu^2 + k^2. \quad (4.3)$$

Частота у такого поля не может быть меньше, чем  $\mu$ .

Скорость переноса энергии у этого поля, которая тождественно равна групповой скорости,

$$V_{\text{гр}} = \frac{k}{\sqrt{\mu^2 + k^2}}$$

изменяется от нуля (при  $k=0$ ) до скорости света ( $k \rightarrow \infty$ ). Заметим, что волновой пакет с дисперсионным уравнением (4.3), в отличие от (4.2), при распространении не сохраняет своей формы, он деформируется.

Частицы с массой, которую предсказал Х. Юкава (1935), вскоре были найдены. Их масса оказалась 105 МэВ, и эти частицы назвали  $\mu$ -мезонами. Однако немедленно был преподнесен сюрприз, когда выяснилось, что  $\mu$ -мезоны практически не взаимодействуют с ядрами. В дальнейшем стало известно, что поле  $\mu$ -мезонов не является скалярным, а имеет спин 1/2. Их сейчас не называют мезонами, оставляя этот термин для другого семейства частиц. Установившимся наименованием для них сейчас является «мюон».

Впрочем вскоре нашли скалярные  $\pi$ -мезоны с массой 135 МэВ, которые уже участвовали в ядерных взаимодействиях.

Позже выяснилось, что есть три вида  $\pi$ -мезонов: два заряженных —  $\pi^+$ ,  $\pi^-$  и один нейтральный —  $\pi^0$ . Этими частицами переносятся ядерные силы. Когда это стало понятно, наступил период расцвета теории скалярного поля. Была открыта изотопическая инвариантность, и стало ясно, что все три типа  $\pi$ -мезонов можно рассматривать как разные состояния одного и того же  $\pi$ -мезонного поля. Аналогично протон и нейт-

рон можно рассматривать как одно нуклонное поле, в точности так же, как состояние электрона со спином вверх и вниз мы рассматриваем как одну и ту же частицу в разных состояниях.

По аналогии с обычным спином разные состояния нуклонного и  $\pi$ -мезонного полей стали характеризовать так называемым изотопическим спином. Протон отвечает состоянию с изотопическим спином «вверх», а нейтрон — «вниз»,  $p$  и  $n$  являются компонентами изотопического спинора со спином 1/2.  $\pi$ -мезоны представляют собой состояние с изотопическим спином 1, т. е. изотопический вектор. Аналогично тому как физические законы не зависят от поворота системы координат, оказалось, что ядерные силы не зависят от вращений в этом внутреннем изотопическом пространстве. Это свойство получило название изотопической инвариантности. Изотопические вращения переводят  $p \leftrightarrow n$  и  $\pi^+ \leftrightarrow \pi^0 \leftrightarrow \pi^-$ . Разные электрические заряды приводят к разным электромагнитным взаимодействиям этих частиц, но на фоне ядерных сил этот эффект незначителен.

В заключение отметим, что  $\pi$ -мезонное поле является не скалярным, а псевдоскалярным. Его волновая функция изменяет знак при пространственном отражении.

## § 2. КВАРКИ. СТРОЕНИЕ АДРОНОВ

Позднее стало ясно, что теория ядерных сил, основанная на  $\pi$ -мезонном обмене и изотопической инвариантности, применима лишь в ограниченном круге явлений, что  $\pi$ -мезонное поле является не фундаментальным, а составным, и о скалярном поле как о фундаментальном объекте теории забыли\*. Его место уверенно заняли векторные поля. Опыт показал, что и барионы (протон, нейтрон и т. д.), и мезоны являются составными частицами, построенными из более элементарных составляющих, названных кварками. Последние являются фермионами и имеют спин 1/2. Для описания нуклонов и  $\pi$ -мезонов достаточно двух сортов кварков, так называемых  $u$ - и  $d$ -кварков. Электрический заряд кварков меньше заряда электрона ( $-e$ ), а именно

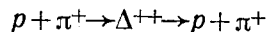
$$q_u = 2/3e, \quad q_d = -1/3e.$$

Из кварков построен большой класс частиц, называемых адронами. К ним принадлежат  $\pi$ -мезоны, нуклоны и множество других. Лептоны, т. е. электрон, мюон, нейтрино, не состоят из кварков. Кварковый состав адронов определяется тем, чтобы суммарный электрический заряд адрона был целочисленным:  $p = (uud)$  (т. е. протон состоит из двух  $u$ -кварков и одного  $d$ -кварка),  $n = (udd)$ ,  $\pi^+ = (u\bar{d})$ ,  $\pi^- = (\bar{u}d)$ ,  $\pi^0 = (u\bar{u} - d\bar{d})/\sqrt{2}$ .

\* Ниже мы увидим, как скалярные поля в другом облике появились в физике элементарных частиц, а вскоре и в космологии.

Кварки обладают еще одним весьма важным квантовым числом, называемым цветом. Чтобы пояснить, как пришли к этому, отвлечемся немного в сторону.

При изучении реакций рассеяния  $\pi$ -мезонов на нуклонах был открыт так называемый  $\Delta$ -резонанс (изобара):



(были открыты и изотопические партнеры  $\Delta^{++}$ :  $\Delta^+$ ,  $\Delta^0$  и  $\Delta^-$ ). Резонанс при взаимодействии элементарных частиц во многом аналогичен резонансной линии поглощения света в веществе. Когда свет проходит через вещество, а атомы имеют переход между уровнями с частотой, равной частоте света, происходит интенсивное поглощение квантов света атомами. Атомы возбуждаются. Возбужденное состояние в атоме нестабильно, и атом возвращается в стабильное состояние, либо испуская несколько фотонов с меньшей энергией, либо один фотон с первоначальной энергией. В обоих случаях образуется промежуточное состояние — возбужденный атом. Такой возбужденный атом аналогичен  $\Delta^{++}$  — возбужденному состоянию из кварков.

По энергии  $p$  и  $\pi^+$  была определена масса  $\Delta^{++}$ , которая оказалась равной 1240 МэВ, что больше массы протона. Заряд  $\Delta^{++}$  оказался равен удвоенному заряду протона, а спин  $3/2$ . Но  $\Delta^{++}$  состоит из трех  $u$ -кварков ( $d$  и  $\bar{d}$ , входящие в состав  $\pi$  и  $p$ , аннигилируют) и значит их спины должны быть параллельными, чтобы сформировать спин  $3/2$ . Как известно, согласно знаменитому принципу Паули фермионы не могут существовать в одинаковых квантовых состояниях. Поэтому кварки должны как-то различаться. Простейшая возможность, состоящая в том, что кварки различаются орбитальным моментом, приводит к антисимметричной по координатам кварков волновой функции  $\Delta$ -изобары. С такой волновой функцией не удастся описать низкоэнергетические параметры этой частицы. Гораздо менее тривиальным и, как выяснилось, весьма плодотворным выходом явилось предположение, что кварки нетождественны. Иными словами, существует не один сорт  $u$ - или  $d$ -кварков, а три, и каждый сорт стали именовать цветом. Сейчас различают кварки  $u_r, d_r; u_b, d_b; u_y, d_y$  (индексы — первые буквы английских слов red — красный, blue — голубой, yellow — желтый). В силу этого, например,  $\Delta^{++}$  можно представить как связанное состояние трех кварков разного цвета  $u_r, u_b, u_y$ , но в одном и том же низшем энергетическом состоянии.

### § 3. ВЗАИМОДЕЙСТВИЕ КВАРКОВ. КВАНТОВАЯ ХРОМОДИНАМИКА

Введение цвета как внутренней характеристики кварков было первым важным шагом в построении теории сильного (ядер-

ного) взаимодействия элементарных частиц. Следующий шаг состоял в гипотезе, что цвет является зарядом, с которым взаимодействуют векторные поля, названные глюонами (символ  $g$ ), аналогично тому как фотоны взаимодействуют с электрическим зарядом\*. При взаимодействии с глюоном может измениться цвет кварка, например

$$u_r \rightarrow u_b + g_{rb}.$$

Идея изменения сорта частиц пришла из квантовой электродинамики. Отталкивание двух электронов можно записать в виде двухступенчатого процесса. Во-первых,

$$e \rightarrow e' + \gamma,$$

электрон  $e$  «распадается» на электрон  $e'$  (с другим импульсом) и фотон  $\gamma$  (виртуальный фотон). Далее:

$$\gamma + e'' \rightarrow e''',$$

т. е. этот фотон поглощает второй электрон  $e''$ , превращаясь в электрон  $e'''$ . Хотя в квантовой электродинамике менялся только импульс и поляризация электрона, такая схема подвела к идее об изменении типа частицы при взаимодействии. Испускание или поглощение глюона меняет цвет кварка, оставляя другие внутренние квантовые числа без изменения, т. е.  $u$ -кварк переходит в  $u$ -кварк,  $d$ -кварк — в  $d$ -кварк и т. д. Как говорят, глюон меняет цвет кварка, не меняя его аромата. Глюоны, очевидно, не несут электрического заряда, но сами обладают цветом (это нечто вроде заряженного фотона). В противном случае цвет не сохранялся бы. Из-за наличия цвета возникает самодействие глюонов, что приводит к нелинейности их уравнений движения. (Вспомним о нелинейности уравнений ОТО из-за того, что гравитоны обладают энергией и, значит, гравитируют).

В отличие от фотона, который единствен, глюонов много и различаются они цветовыми зарядами. Глюон  $g_{ij}$  переводит кварк цвета  $i$  в кварк цвета  $j$ . Матрица  $g_{ij}$  размерности  $3 \times 3$  имеет 9 элементов. Однако из нее исключают единичный оператор  $\delta_{ij} Tg$ , который «ничего не делает» с кварками. В итоге приходим к восьми глюонам.

Взаимодействие между кварками можно наглядно представить в виде силовых линий глюонного поля, аналогичных силовым линиям электромагнитного поля. Но у глюонов есть од-

\* Эта теория напоминает теорию Юкавы с той, однако, разницей, что вместо обмена (псевдо) скалярными  $\pi$ -мезонами между нуклонами рассматривается обмен векторными глюонами между кварками. Это малозначимое на первый взгляд обстоятельство приводит к глубокому качественному различию теорий.

но чрезвычайно важное отличие от электрического поля. При удалении двух электронов друг от друга силовые линии электрического поля вблизи каждого из зарядов выпрямляются и стремятся к радиальным линиям. Энергия электростатического взаимодействия убывает как  $1/r$  и соответственно сила между зарядами — как  $1/r^2$ . Глюонное поле, в отличие от электромагнитного, имеет заряд (цветовой), с которым оно само взаимодействует, т. е. является нелинейным. Поэтому, как полагают, при удалении кварков силовые линии глюонов не расходятся, а собираются в струну.

Энергия такой конфигурации поля пропорциональна расстоянию между кварками, т. е. длине струны. Соответственно потенциал взаимодействия кварков ведет себя подобно кулоновскому на малых расстояниях и линейно растет при больших  $r$  (рис. 6). Это приводит к принципиально новому явлению —

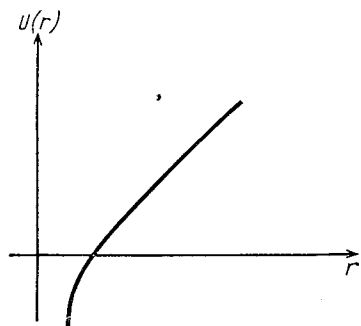


Рис. 6. Потенциал взаимодействия кварков как функция расстояния

вылетанию кварков (по английски confinement). В отличие от протонов, электронов и т. п. кварки не могут существовать обособленно, как свободные частицы. При попытках развести кварки на большое расстояние, например «разбивая» протоны при столкновениях их друг с другом, кварк может получить большую порцию энергии и сильно «натянуть» глюонную струну. Напряженность глюонного поля при этом может стать настолько большой, что из вакуума рождается пара кварк-антикварк. И этот антикварк вместе с исходным энергичным кварком уже может уйти на большое расстояние, т. е. рождается, например,  $\pi$ -мезон, а не свободный кварк. То же самое происходит и при попытке выбить глюон. Вообще на макроскопическое расстояние может уйти только состояние с нулевым цветовым зарядом, или, как говорят, белое состояние. Такими являются все известные элементарные частицы. Заметим, что нулевой цветовой заряд может иметь не только система из кварка и антикварка, но и система из трех кварков, как, например, протон или нейтрон. В этом важное отличие хромодинамики от электродинамики.

Отметим, что все бесцветные состояния имеют целочисленный электрический заряд, и наоборот все построенные из кварков частицы с дробным электрическим зарядом обязательно имеют ненулевой цветовой заряд. Возможны, впрочем, и

состояния с целым электрическим зарядом и ненулевым цветовым. Они также не могут существовать в свободном виде.

Явление невылетания или экранировки цвета теоретически на основе первых принципов пока не доказано, но такая картина хорошо подтверждается опытом. В частности, при разрыве глюонной струны должна возникнуть адронная струя, т. е. поток сильновзаимодействующих частиц (адронов), летящих в одном направлении. Существование таких адронных струй надежно установлено в многочисленных экспериментах. Адронная струя представляет собой след кварка в вакууме, подобный следу электрона в камере Вильсона.

Взаимодействие адронов на малых расстояниях ( $< 10^{-13}$  см) полностью описывается довольно простой динамикой кварков и весьма похоже на электромагнитное. Известные более тяжелые, чем  $u$  и  $d$ , кварки, и наблюдаемые на опыте их связанные состояния напоминают позитроний. Сейчас твердо установлено существование пяти сортов кварков со следующими массами:

Сорт кварка	$u$	$d$	$s$	$c$	$b$	$t$
Масса	5 МэВ	7 МэВ	150 МэВ	1,2 ГэВ	5 ГэВ	40 ГэВ*

Масса протона, как известно, равна 940 МэВ и существенно превосходит сумму масс составляющих его кварков ( $uud$ ). Это объясняется тем, что в массу протона вносит вклад энергия глюонного поля и кинетическая энергия кварков, движущихся в области размером около  $10^{-13}$  см. Значения, приведенные в таблице, отвечают массе «голового» кварка, когда измерение производится с большой передачей импульса (или на малых расстояниях), так что влияние глюонного поля несущественно.

Потенциальное описание адронов неприменимо к легким  $u$ -,  $d$ -,  $s$ -кваркам, так как они в существенной области являются релятивистскими. Однако для новых тяжелых кварков потенциал (рис. 6)

$$U = \frac{\alpha_c}{r} + br$$

дает хорошее описание спектра адронов и вероятностей переходов с одного уровня на другой. Сейчас известно большое семейство частиц, состоящих из  $c\bar{c}$  и из  $b\bar{b}$ , и согласие предсказаний теории с опытом является серьезным аргументом в пользу кварковой структуры адронов.

\* Надежных экспериментальных доказательств обнаружения шестого кварка пока нет, но сейчас мало кто сомневается в его существовании, так как теоретические аргументы в его пользу очень сильны.

Хромодинамика становится чрезвычайно сложной наукой на больших, т. е.  $\sim 10^{-13}$  см, расстояниях. Дело в том, что эффективный цветовой заряд, определяющий силу взаимодействия глюонов с кварками и между собой, растет с увеличением расстояния (и соответственно падает с уменьшением) и становится порядка 1 при  $r \approx 10^{-13}$  см. Это явление получило название асимптотической свободы. Большой эффективный заряд является причиной интересных физических явлений при  $r \approx 10^{-13}$  см и приводит к весьма сложному адронному миру.

Ядерные силы в современном понимании возникают в результате заэкранированного цветового взаимодействия адронов аналогично электрическим ван-дер-ваальсовым силам между нейтральными атомами. Они относительно слабы, приводя к энергии связи ядер порядка 10 МэВ, а характерная энергия для цветового взаимодействия кварков на порядок или даже на два выше.

Из-за того, что цветовой заряд не мал на больших расстояниях, описание ядерных сил в рамках теории возмущений невозможно. Однако наиболее дальнедействующую часть потенциала, связанную с обменом легчайшим из адронов —  $\pi$ -мезоном, мы можем определить.

Разумеется, кварковая картина не привела к отмене наших взглядов на структуру атомного ядра как состоящего из протонов и нейтронов. В нулевом приближении кварки не важны в ядре, как строение ядра не важно в химии. Однако мы уже сейчас понимаем, что в некоторых тонких ядерных явлениях кварковая структура адронов становится существенной.

В настоящее время не ясно, являются ли кварки последним этапом на пути дробления материи или они в свою очередь имеют внутреннюю структуру, являются составными. Однако независимо от того, как будет решен этот вопрос, кварковая теория адронов сохранится, во всяком случае для описания физических явлений в области энергии от 1 до 100 ГэВ.

В настоящее время, как мы уже отмечали, известно 6 сортов кварков (твердо установлено 5). Три из них  $u$ ,  $c$ ,  $t$  имеют заряд  $+2/3e$ , а три других  $d$ ,  $s$ ,  $b$  —  $-1/3e$ . Каждый кварк может быть в трех цветовых состояниях, так что мы имеем всего 18 состояний, но ни заряд, ни масса кварка не зависят от его цвета, поэтому говорят, что мы имеем всего 6 сортов кварков, или 6 кварковых ароматов. Напомним, что взаимодействие глюонов со всеми кварковыми ароматами одинаково, как одинаково взаимодействие фотонов, например, с электроном и мюоном.

#### § 4. КВАРК-ЛЕПТОННАЯ СИММЕТРИЯ. ОГРАНИЧЕНИЕ НА ЧИСЛО ТИПОВ ЛЕПТОНОВ И КВАРКОВ

В природе наблюдается любопытная симметрия: каждой паре кварков отвечает пара лептонов:

$$\begin{pmatrix} u \\ d \end{pmatrix} \leftrightarrow \begin{pmatrix} \nu_e \\ e \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix} \leftrightarrow \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix} \leftrightarrow \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}.$$

Такая кварк-лептонная симметрия, вероятно, не является просто случайным совпадением. В единой теории электрослабых взаимодействий она необходима для того, чтобы исчезала так называемая аномалия в дивергенции аксиального тока, наличие которой препятствует перенормируемости теории.

Из опыта известны три семейства кварков и лептонов, но мы не можем исключить, что существуют и другие более тяжелые, для рождения которых пока не хватает энергии. Существование кварк-лептонной симметрии позволяет получить ограничение сверху на число кварковых ароматов. Здесь нам на помощь приходит космология (гл. 1 § 8). Дело в том, что сопутствующие каждой паре кварков нейтрино, в отличие от тяжелых лептонов и кварков, обильно содержатся в первичной плазме во время нуклеосинтеза легких элементов (характерное время здесь около  $1 \div 100$  с от «начала»), вносят заметный вклад в общую плотность материи, в силу этого влияют на темпы расширения и охлаждения Вселенной. С другой стороны, скорость расширения определяет выход легких элементов в результате первичного нуклеосинтеза. Поэтому по наблюдаемому сегодня обилию  $\text{He}^4$  и  $\text{H}^2$ , если брать центральные значения, мы можем сделать вывод, что  $k_\nu < 4$ , и, следовательно, новых кварковых ароматов не существует. Все типы кварков нам уже известны!\*

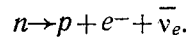
Этот вывод можно будет в ближайшем будущем проверить в непосредственном эксперименте, так как время жизни нейтрального промежуточного бозона слабых взаимодействий  $Z^0$ , о котором мы будем подробнее говорить ниже, зависит от числа типов нейтрино. Эта идея высказана в 1968 г. безвременно погибшим В. Ф. Шварцманом.

#### § 5. ЭЛЕКТРОСЛАБЫЕ ВЗАИМОДЕЙСТВИЯ. ПРОМЕЖУТОЧНЫЕ БОЗОНЫ

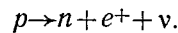
Еще один пример известного в физике векторного поля — это поле  $W^\pm$ ,  $Z^0$ -бозонов (их иногда называют «вионы»). Оно

\* Более осторожные физики призывают не переоценивать точность данных и дают менее сильное ограничение  $k_\nu \leq 6$ .

является переносчиком слабого взаимодействия. Впервые с этим взаимодействием физики столкнулись при наблюдении  $\beta$ -радиоактивности атомных ядер. В ее основе, как мы сейчас знаем, лежит распад нейтрона



Заметим, что, хотя свободный нейтрон нестабилен, он может оказаться стабильным в составе атомного ядра, если разность энергий связи нейтрона и образовавшегося протона превосходит энергосвязь при распаде. Этим объясняется существование стабильных атомных ядер. Наконец, если ядро таково, что энергия связи нейтрона в нем велика, а протона мала, возможен обратный процесс:



Период полураспада нейтрона составляет около 15 мин, что на много порядков превосходит естественную размерную оценку времени полураспада по дефекту массы при распаде:

$$t \approx 1/\Delta mc^2 \approx 10^{-21} \text{ с.}$$

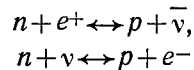
Это показывает, что взаимодействие, которое приводит к распаду нейтрона, является чрезвычайно слабым, с чем и связано его название.

Поскольку в этой реакции участвуют четыре фермиона, то физики стали говорить о 4-фермионном взаимодействии. Примерно 50 лет назад Ферми записал выражение для амплитуды такого перехода в виде:

$$g_F \psi_p^* \psi_n \psi_e^* \psi_\nu; \quad (4.4)$$

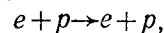
здесь  $\psi_n$  — волновая функция нейтрона,  $\psi_p$  — протона и т. п.,  $g_F$  — константа слабого взаимодействия.

Согласно квантовой теории поля выражение (4.4) описывает не только распад нейтрона, но и процессы



и тому подобное. Отметим, что при этом сохраняется количество барионов (т. е. барионный заряд) и количество лептонов (т. е. лептонный заряд) и, конечно, электрический заряд.

Существенной чертой выражения (4.4) является то, что все четыре величины  $\psi$  взяты в одной и той же пространственно-временной точке. Такое взаимодействие называют локальным. С другой стороны, похожий четырехфермионный процесс, связанный электромагнитному взаимодействию



является нелокальным. Его можно описать диаграммой рис. 7, что означает, что протон в точке  $x$  испускает виртуальный фотон, который поглощается электроном в точке  $y$ , и, следовательно, в амплитуду этого процесса входят  $\psi_e(y) \cdot \bar{\psi}_e(y) \cdot \psi_p(x) \cdot \bar{\psi}_p(x)$ . Локальной амплитуде (4.4), очевидно, отвечает диаграмма вида рис. 8.

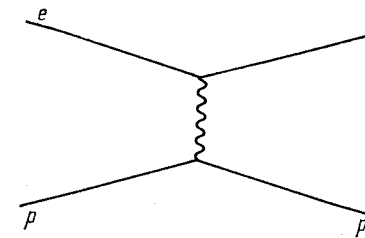


Рис. 7. Диаграмма, иллюстрирующая электрон-протонное рассеяние. Волнистая линия изображает фотон

Успех электродинамики, с одной стороны, и внутренние трудности теории локального четырехфермионного взаимодействия — с другой, связанные с так называемой неперенормируемостью, на которой мы остановимся ниже, привели к идее построения теории слабого взаимодействия по аналогии с электродинамикой.

Важным аргументом в пользу такого подхода явилось сходство тока слабого взаимодействия с электромагнитным. В обоих случаях амплитуду можно было записать как произведение векторных токов:  $J_\mu J^\mu$ , где  $J_\mu$  содержит произведение двух фермионных волновых функций. Аналогия усугуб-

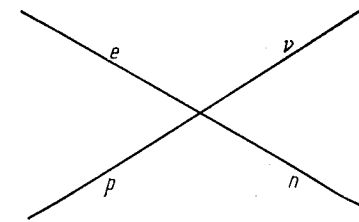


Рис. 8. Диаграмма, иллюстрирующая процесс  $e + p \rightarrow n + \nu$ , который при малых энергиях можно считать локальным

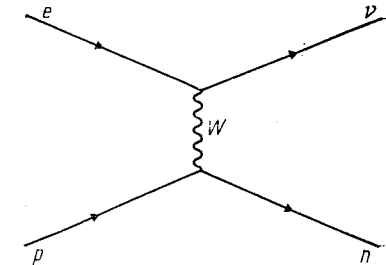


Рис. 9. Процесс  $e + p \rightarrow n + \nu$  в новой теории слабого взаимодействия; нуклонный и лептонный токи взаимодействуют не локально, а за счет обмена  $W$ -бозоном

ляется еще и тем, что векторный ток слабого взаимодействия так же, как электромагнитный, является сохраняющимся. В силу этого возникла естественная гипотеза, что слабый процесс также идет через промежуточную векторную частицу согласно диаграмме рис. 9.

Такая частица,  $W$ -бозон, действительно была обнаружена



на опыте с предсказанной теорией массой около 80 ГэВ. Столь высокая масса обеспечивает наблюдаемую на опыте приближенную локальность слабого взаимодействия, точнее его весьма короткий радиус действия  $r \sim 10^{-16}$  см. В этом состоит важное отличие слабых и электромагнитных взаимодействий, которые на расстояниях  $r < 10^{-16}$  см очень похожи друг на друга, а при  $r > 10^{-16}$  см имеют существенно разную силу, так как масса промежуточного бозона электромагнитных взаимодействий — фотона равна нулю и поэтому электрические силы дальнедействующие, а слабые силы резко падают при  $r > 10^{-16}$  см. Другое отличие состоит в том, что  $W$ -бозоны имеют электрический заряд, а фотоны нейтральны.

Большая масса  $W$ -бозона приводит к тому, что процесс  $n \rightarrow p + W$  энергетически невозможен. Но это, конечно, не означает, что такие процессы не идут совсем и схема распада нейтрона, изложенная выше, неверна. Дело в том, что когда мы говорим о сохранении энергии, то в соответствии с квантовомеханическим принципом неопределенности подразумеваем, что система находится в заданном состоянии достаточно большое время. В течение времени  $\Delta t$  можно говорить о несохранении энергии на величину  $\hbar/\Delta t$ . Поэтому на очень короткое время нейтрон может перейти в протон и  $W$ -бозон. Это, как говорят, виртуальный процесс, а  $W$ -бозон в данном случае называют виртуальной частицей. Зная массу  $W$ -бозона, можно вычислить время, в течение которого будет существовать виртуальный  $W$ -бозон:

$$t_W \sim \hbar/m_W c^2 \approx 10^{-26} \text{ с}$$

и расстояние, на котором разыгрывается процесс распада нейтрона,

$$l = ct_W \approx 10^{-16} \text{ см.}$$

На формальном языке квантовой механики процесс вида

$$n \rightarrow p + W^-$$

описывается как появление примеси состояния  $(p + W^-)$  к состоянию  $n$ .

Понятие о подобных смешанных состояниях возникло задолго до появления теории слабого взаимодействия. Например, в атоме водорода возможны  $S$ -состояние и  $P$ -состояние атома. Помещая атом в электрическое поле ( $E$ ), мы получаем смесь  $\psi_S$  и  $\psi_P$  состояний:

$$\psi = \psi_S + C\psi_P.$$

Амплитуда  $C$ , описывающая примесь  $P$ -состояния к  $S$ -состоянию, выражается согласно

$$C \sim \frac{eE \cdot d}{\mathcal{E}_S - \mathcal{E}_P}; \quad (4.5)$$

здесь  $(\mathcal{E}_S - \mathcal{E}_P)$  — разность энергии между  $S$ - и  $P$ -уровнями, а  $d$  — «размер» атома водорода.

Точно так же при описании распада нейтрона учет перехода  $n \leftrightarrow p + W^-$  приводит к примеси  $(p + W^-)$  к нейтрону с амплитудой  $C_1 \sim (e/\sqrt{\hbar c}) \Delta m / (m_W + m_p - m_n)$ . Знаменатель есть обычная разность энергий начального и конечного состояний, которые обычно возникают при расчетах по теории возмущений. Появление  $\Delta m$  в числителе отражает специфику матричного элемента  $\langle \psi_{\text{нач}} | H | \psi_{\text{кон}} \rangle$  перехода  $n$  в  $p + W$ . Поскольку такой процесс энергетически невозможен и идет только виртуально, то при вычислении вероятности полного процесса  $n \rightarrow p + e^- + \bar{\nu}$  надо учесть второй порядок теории возмущений, т. е. принять во внимание переход  $W^- \rightarrow e^- + \bar{\nu}$ . В полной аналогии с изложенным выше к пси-функции  $\psi_{W+p}$  появится добавка  $\psi_{e\bar{\nu}p}$  с амплитудой  $C_2$ , равной

$$C_2 \sim \frac{e}{\sqrt{\hbar c}} \frac{\Delta m}{m_W + m_p - m - (E_e + E_{\bar{\nu}})/c^2}. \quad (4.6)$$

Примесь состояния  $e\bar{\nu}p$  к начальному нейтрону, как легко видеть, равна  $C_1 \cdot C_2$ . Отсюда вероятность перехода в единицу времени можно оценить как

$$P = |C_1 C_2|^2 \frac{\Delta m c^2}{\hbar} \approx \frac{\Delta m c^2}{\hbar} \left( \frac{\Delta m}{m_W} \right)^4 \left( \frac{e^2}{\hbar c} \right)^2. \quad (4.7)$$

В этих оценках мы предположили, основываясь на аналогии между  $W$  и фотоном, что слабая константа связи близка по величине к электромагнитной  $e$ . Точная теория приводит именно к такому результату. В итоге вероятность распада содержит произведение двух малых множителей:  $(\Delta m/m_W)^4$  и квадрата постоянной тонкой структуры. Именно они и дают понижение от наивной оценки  $10^{21} \text{ с}^{-1}$  до наблюдаемой величины  $\sim 10^{-3} \text{ с}^{-1}$ .

Теория слабых взаимодействий на основе обмена промежуточными бозонами в своем окончательном варианте связана с работами Вайнберга и Салама 1967 г., а также с более ранними работами Глэшоу. В этой теории, кроме заряженных  $W^\pm$ -бозонов должен существовать и нейтральный бозон  $Z^0$ , несколько более тяжелый, чем  $W^\pm$ . Эта частица с массой около 90 ГэВ также обнаружена на опыте.  $Z^0$  напоминает тяжелый фотон, однако, в отличие от фотона, он взаимодействует не только с заряженными частицами, но и с нейтральными.

Если  $W$ -бозон обеспечивает переходы вида

$$n \rightarrow p + W^-, e^- \rightarrow \nu_e + W^-, \mu^- \rightarrow \nu_\mu + W^- \text{ и т. п.,}$$

то  $Z^0$  приводит к переходам без изменения заряда  
 $p \rightarrow p + Z^0, n \rightarrow n + Z^0, e^- \rightarrow e^- + Z^0, \nu \rightarrow \nu + Z^0$  и т. п.,

в результате которых могут идти реакции

$$N + \nu \rightarrow N + \nu, e^- + \nu \rightarrow e^- + \nu,$$

описываемые диаграммами рис. 10. Это так называемые процессы с нейтральными токами. Таким образом, существует универсальное рассеяние нейтрино на протонах и электронах.

Эти процессы очень важны для астрофизики. Раньше, до создания теории слабых взаимодействий, считалось, что энергия из внутренних слоев звезд может уноситься УРКА-процессами. Это процессы, в которых участвуют нейтрино. После рождения они свободно выходят из центра звезды и эффективно ее охлаждают. Теперь, после создания правильной теории слабых взаимодействий, мы знаем, что необходимо дополнительно

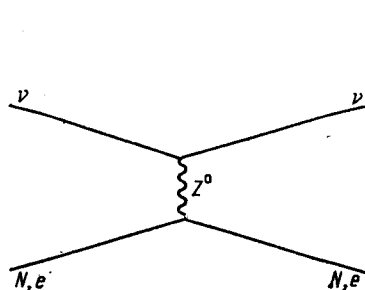


Рис. 10. Процесс  $\nu N \rightarrow \nu N$  или  $\nu e \rightarrow \nu e$ , связанный с взаимодействием слабых нейтральных токов, или, что то же, с обменом  $Z^0$ -бозоном

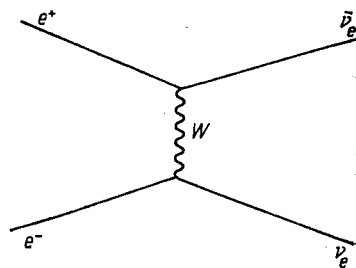


Рис. 11. Процесс  $e^+ + e^- \rightarrow \nu_e + \bar{\nu}_e$  идущий за счет заряженных токов (или  $W$ -обмена). В этом случае электрон взаимодействует только с электронным нейтрино

учитывать нейтральные токи. В результате сечение рассеяния для  $\nu$  возрастает и звезда охлаждается значительно медленнее.

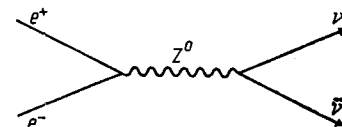
Выводы теории оказались важными не только для звездной астрофизики, но и для космологии. При наличии только заряженных токов, т. е. за счет обмена  $W$ -бозонами, электрон-позитронные пары могли аннигилировать лишь в нейтрино электронного типа рис. 11. Теперь, при наличии  $Z^0$  разрешены также реакции

$$e^+ + e^- \rightarrow \nu_\mu + \bar{\nu}_\mu, \nu_\tau + \bar{\nu}_\tau,$$

описываемые диаграммой рис. 12, которые поддерживают в равновесии  $\nu_\mu$  и  $\nu_\tau$  без участия тяжелых  $\mu$  и очень тяжелых  $\tau$ . Поэтому при температурах выше 5 МэВ во Вселенной в равновесии находились все три типа нейтрино, что имеет важное значение для первичного нуклеосинтеза.

Есть еще одно существенное отличие слабых взаимодействий от электромагнитных, о котором мы пока не говорили. Слабое взаимодействие нарушает четность. Это проявляется, в частности,

Рис. 12. Взаимодействие электронов с нейтрино за счет нейтральных токов ( $Z^0$ -обмен). В этом случае электрон взаимодействует со всеми типами нейтрино



в следующем: у безмассовых частиц спин может быть направлен либо по движению, либо против движения. Для фотонов возможны оба состояния поляризации, и при рождении фотонов в соударениях заряженных частиц половина фотонов рождается с правой круговой поляризацией, половина — с левой. При одинаковой частоте правые и левые состояния фотонов вырождены по энергии, поэтому имеет смысл рассматривать любую суперпозицию, в частности полусумму или полуразность, отвечающую линейной поляризации. У нейтрино возможно только одно состояние, при котором спин и импульс антипараллельны, у антинейтрино, наоборот, параллельны.

Хорошая классическая аналогия этой ситуации — штопор для вытаскивания пробок из бутылок. Штопор, будучи вставленным в пробку, может двигаться вперед, вращаясь только в вполне определенном направлении, — скажем слева направо. Штопоров, которые двигались бы вперед при вращении справа налево промышленность не выпускает. Это и есть нарушение четности. Четность сохранялась бы, если бы промышленность выпускала половину штопоров лево-правых и половину право-левых.

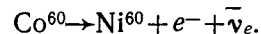
Разумеется, несохранение четности не связано с тем, что нейтрино безмассовое. Если  $m_\nu \neq 0$ , то поляризационное состояние зависит от системы отсчета. Действительно, для массивного нейтрино со спином, антипараллельным импульсу (так называемого левого нейтрино), существует система отсчета, двигающаяся быстрее, чем это нейтрино, в которой, очевидно, направление импульса противоположно исходному, а спина — тоже самое. Таким образом, в этой системе отсчета спин и импульс параллельны, т. е. нейтрино является правым. Следовательно, независимо от механизма рождения возможны оба поляризационных состояния. Несохранение четности же проявляется в том, что доля правых и левых нейтрино, рождающихся в каком-либо процессе, различна. В случае  $m_\nu = 0$  рождаются

только левые нейтрино, если же  $m_\nu \neq 0$ , то доля правых нейтрино составляет величину порядка

$$(1 - v/c) \approx (mc^2/E)^2.$$

Справедливость этого экспериментально проверена для электронов. При  $\beta$ -распаде электрон рождается левополяризованным со степенью поляризации  $v/c$ . Это значит, что  $W$ -бозон взаимодействует только с определенным поляризационным состоянием нейтрино, электрона, протона и т. д. А именно только с левыми частицами. Это фактически и было обнаружено опыте, в котором ниспровергли четность.

Для эксперимента по проверке закона сохранения четности в слабых взаимодействиях был выбран  $\beta$ -распад в ядрах радиоактивного кобальта —  $\text{Co}^{60}$ . В этом эксперименте  $\text{Co}^{60}$  охлаждали до температуры жидкого гелия, чтобы свести к минимуму флуктуации в направлении момента вращения ядра, и ориентировали момент вращения с помощью магнитного поля. Поляризованные таким образом ядра испытывали  $\beta$ -распад:



В опыте измерялось угловое распределение электронов относительно направления момента вращения ядер  $\text{Co}^{60}$ , которое мы будем считать для удобства изложения направленным вверх. Если бы четность сохранилась, то количество электронов, вылетевших в верхнюю и нижнюю полусферу, было бы одинаковым. Однако эксперимент показал резкую асимметрию: электроны преимущественно вылетали в нижнюю полусферу в направлении, противоположном моменту вращения ядра  $\text{Co}^{60}$ . При зеркальном отражении мы бы увидели, что электроны летят по направлению момента. Таким образом, зеркальный процесс физически невозможен и, следовательно, четность не сохраняется.

В силу сказанного теория слабых взаимодействий была построена таким образом, что  $W$ -бозоны взаимодействуют только с левыми частицами (лептонами и кварками) и с правыми античастицами. Можно было бы иметь теорию, в которой точно так же взаимодействовали и  $Z^0$ -бозоны, но теория строилась как единая теория слабых и электромагнитных взаимодействий, а мы знаем, что фотон одинаково взаимодействует с заряженными частицами независимо от их поляризации. Это привело к тому, что и  $Z^0$  взаимодействуют как с лево-, так и с правополяризованными частицами, но все же по-разному. Поэтому  $Z^0$ , в отличие от фотона, тоже нарушает четность. В частности, с правыми нейтрино, даже если бы они существовали,  $Z^0$  совсем не взаимодействует. Здесь нужно подчеркнуть разницу между кварками и заряженными лептонами, для которых существуют оба поляризационных состояния, и нейтрино, для

которого, возможно, есть только одно состояние поляризации (т. е.  $\nu_L$  и  $\bar{\nu}_R$ ). Причем этот факт не связан с безмассовостью нейтрино, так как для нейтрального фермиона разрешена так называемая майорановская масса, отвечающая переходу левой частицы в правую античастицу, в то время как обычная дираковская масса дает переход левой частицы в правую частицу.

В теории электрослабого взаимодействия фотон,  $W^\pm$  и  $Z^0$  рассматриваются как разные проявления единого векторного поля. Точнее говоря, имеются два фундаментальных векторных поля — одно нейтральное  $B^0_\gamma$ , взаимодействующее с гиперзарядом частицы  $Y$ , который мы определим ниже, а другое представляет собой триплет  $(W^+, W^-, B^0_I)$ , взаимодействующий с так называемым слабым изоспином  $I$ . Это поле в точности аналогично  $\pi$ -мезонному триплету  $(\pi^+, \pi^-, \pi^0)$ . Каждая пара лептонов также рассматривается как единое лептонное поле, аналогично тому как протон и нейтрон рассматриваются как нуклонное поле в соответствии со сказанным выше. Пары  $(\nu_e, e)$ ,  $(\nu_\mu, \mu)$  и  $(\nu_\tau, \tau)$  образуют дублеты в пространстве слабого изоспина с  $I=1/2$ . Аналогичные дублеты образуют кварки одного цвета, например  $(u_b, d_b)$  и т. д. Слабый гиперзаряд частицы определяется как удвоенный средний электрический заряд мультиплетта:  $Y=2(Q-T_3)$ . Только левые частицы (соответственно правые античастицы) образуют слабые дублеты и поэтому  $(W^+, W^-, B_I)$  взаимодействуют только с левыми частицами. Правые частицы являются изотопическими синглетами с гиперзарядом  $Y=2Q$ . Поэтому поле  $B_\gamma$  взаимодействует как с левыми, так и с правыми частицами. Спонтанное нарушение симметрии (см. § 8 этой главы) приводит к тому, что определенную массу получает линейная комбинация  $B_\gamma$  и  $B_I$  (ее отождествляют с  $Z^0$ ), а ортогональная ей комбинация остается безмассовой (это, очевидно, фотон).

Нужно отметить, что электрослабое взаимодействие не приводит к конфайнменту. Это отличие от хромодинамики связано с нарушением симметрии электрослабого взаимодействия. Образуя говоря, массивность  $W$  и  $Z$  спасает их от «заклечения». Безмассовый фотон вылетает, так как абелева  $U(1)$ -симметрия не приводит к конфайнменту.

Здесь уместно сказать несколько слов об осцилляциях нейтрино. Если  $m_\nu \neq 0$ , то может оказаться так, что состояния, входящие в слабый дублет, не имеют определенной массы, т. е., например,  $\nu_e = C_1 \nu_1 + C_2 \nu_2 + C_3 \nu_3$ , где  $\nu_1$  и  $\nu_2$  и  $\nu_3$  имеют массы  $m_1$ ,  $m_2$ ,  $m_3$  соответственно. В этом случае  $\nu_e$ , рожденное, скажем, в реакции  $e p \rightarrow \nu_e$ , при распространении в вакууме непрерывно переходит в  $\nu_\mu$  и  $\nu_\tau$  и обратно, так что вероятность образования  $e$  в обратном процессе  $\nu_e n \rightarrow p e$  является осциллирующей функцией расстояния. Осцилляции могли бы объяснить недостаток солнечных нейтрино в эксперименте Девиса.

Несохранение четности, о котором шла речь выше, можно обнаружить не только в реакциях с элементарными частицами. В обычном веществе тоже есть процессы с несохранением четности, хотя там они выражены очень слабо. Рассмотрим, например, атом водорода.

Электрон удерживается на орбите электромагнитными силами. Дополнительно к этому между протоном и электроном есть взаимодействие через нейтральный бозон  $Z^0$ . Непосредственно, по сдвигу уровней в атоме, обнаружить  $Z^0$ -взаимодействие очень трудно, так как оно дает малую поправку к большой величине. Но его можно обнаружить по несохранению четности, которое целиком обусловлено  $Z^0$ -обменом. С учетом этого взаимодействия атом водорода и вообще любой атом обладает малой нечетностью при взаимодействии со светом. Это проявляется в том, что при прохождении через вещество у света начинает вращаться плоскость поляризации.

Величина эффекта, очевидно, больше для тяжелых атомов, так как в них больше вероятность нахождения электрона вблизи ядра, т. е. больше  $|\psi(0)|^2$ , и поэтому взаимодействие за счет короткодействующего обмена  $Z^0$ -бозоном более эффективно. Несмотря на слабость, эффект вращения плоскости поляризации был впервые обнаружен в Новосибирском институте ядерной физики (1979). Таким образом, мы узнали о несохранении четности в атомной физике.

Независимый эксперимент, подтвердивший несохранение четности в электрон-протонных взаимодействиях, был поставлен в Стенфорде (1981). Там измерялась разность сечений лево- и правополяризованных электронов на дейтерии. Результаты находятся в прекрасном согласии с теорией Глэшоу — Вайнберга — Салама.

Промежуточные бозоны слабых взаимодействий  $W^\pm$  и  $Z^0$  были открыты совсем недавно, в 1984 г. на протон-антипротонном ускорителе. Непосредственно  $W^\pm$  и  $Z^0$  мы увидеть не можем, поскольку время жизни этих частиц весьма мало,  $t \approx 10^{-24}$  с, и за это время они проходят путь меньше размеров электрона  $\sim 10^{-14}$  см. Однако их существование доказывают характерные реакции распада  $W$  и  $Z$  на другие частицы. Измерение энергии продуктов распада позволяет определить массу  $W$  и  $Z$ . Зарегистрировано уже более 100 событий рождения этих бозонов.

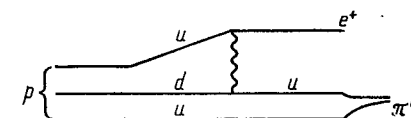
## § 6. ТЕОРИИ БОЛЬШОГО ОБЪЕДИНЕНИЯ. НЕСОХРАНЕНИЕ БАРИОНОВ

Успех в построении единой теории электромагнитных и слабых взаимодействий стимулировал следующий естественный шаг — поиски единой теории сильного и электрослабого вза-

имодельствий. Оказалось возможным построить модель (и даже не одну, а несколько), в которой промежуточные бозоны сильных (глюоны) и электрослабых взаимодействий ( $g$ ,  $\gamma$ ,  $W^\pm$ ,  $Z^0$ ) описывались как разные проявления единого векторного поля — фундаментального переносчика взаимодействий. Лептоны и кварки при этом также рассматривались единообразно как проявления этого фермионного поля. Во всех вариантах теории, кроме известных 12 векторных бозонов, приходилось вводить дополнительно столько же или более новых векторных частиц  $X$ - и  $Y$ -бозонов. Аналогично тому как глюоны меняют цвет кварка,  $X$ - и  $Y$ -бозоны меняют заряд кварка или лептона и, более того,  $X$ - и  $Y$ -бозоны превращают кварк в лептон. Другими словами, взаимодействие с этими бозонами не сохраняет барионный заряд. В частности, за счет диаграммы рис. 13 они должны приводить к распаду протона.

Теория предсказывает чрезвычайно большую величину массы  $X$ - и  $Y$ -бозонов — около  $10^{15}$  ГэВ. И это является причиной

Рис. 13. Диаграмма, иллюстрирующая распад протона. Волнистой линией обозначен калибровочный бозон модели большого объединения, переводящий кварки в лептоны



того, что распад протона, несмотря на интенсивные поиски, до сих пор не обнаружен. В действительности положение сейчас довольно напряженное, так как экспериментаторы получили ограничение снизу на время жизни протона, заметно превышающее предсказания первых простых моделей объединения, и пока единственный «экспериментальный» аргумент в пользу несохранения барионов дает космология, а именно барионная асимметрия Вселенной.

Появление в теории частиц с массами  $10^{15}$  ГэВ стимулировало изучение космологических следствий таких моделей, так как не видно никакого другого способа (кроме распада протона) узнать что-либо о физике на таких масштабах.

В настоящее время распад протона не обнаружен на уровне примерно  $10^{32}$  лет. В ближайшие годы можно рассчитывать на повышение точности примерно на порядок или, быть может, на два. После этого земные возможности доказательства несохранения бариозаряда будут исчерпаны. Тем более интересен тот факт, что инфляционная космология определенно указывает на несохранение барионов после окончания инфляции (см. § 3 гл. 8).

Заметим, что обсуждаемый здесь механизм несохранения барионов не является единственно возможным. Другие варианты обсуждаются в гл. 8.

### § 7. ВЫСШИЕ ПОРЯДКИ ТЕОРИИ ВОЗМУЩЕНИЙ. ПЕРЕНОРМИРОВКА

Наше изложение, казалось бы, показывает, что все взаимодействия, кроме гравитационного, определяются векторными полями и о скалярном поле можно спокойно забыть. Развитие физики за последние десятилетия, однако, изменило общий подход к теории. Если раньше считалось, что «все, что не доказано, не существует», то теперь — «все, что не опровергнуто, разрешено». Уже поэтому давно бы следовало рассмотреть роль (не обнаруженного на опыте) скалярного поля в космологии. Фактически, однако, рассмотрение скаляров стало популярно, лишь когда возникли единые теории взаимодействий частиц (типа теории электрослабого взаимодействия) и выяснилось, что без скалярных полей эти теории наталкиваются на серьезные трудности. Чтобы пояснить, о чем идет речь, начнем с электродинамики. Рассмотрим, например, рассеяние электронов, которое описывается диаграммой рис. 14, а. Однако кроме

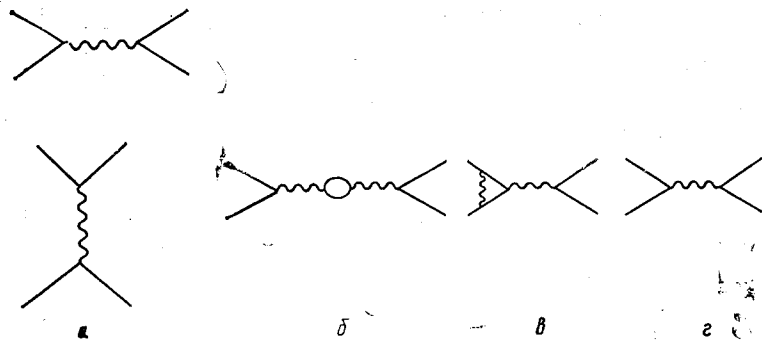


Рис. 14. Радиационные поправки к упругому  $ee$ -рассеянию

процесса, отвечающего этой диаграмме, возможны и другие процессы, как, например, изображенные на диаграмме б—г, которые также дают вклад в ее рассеяние. Каждое дополнительное испускание и поглощение фотона приводит к малому фактору  $\alpha = e^2/\hbar c = 1/137$ , и поэтому можно ожидать, что вклад диаграмм б—г в амплитуду ее рассеяния будет невелик. Однако расчет показывает, что каждая из диаграмм б—г дает бесконечно большую величину из-за расходимости возникающих при вычислении интегралов. Слегка перефразируя известное высказывание Фейнмана, можно сказать, что «поправки малы, но бесконечны».

Эта трудность при вычислении высших порядков теории возмущений по  $\alpha$  долгое время стояла перед теорией, пока не выяснилось, что ее можно обойти с помощью так называемой про-

цедуры перенормировок. Дело в том, что *все* бесконечности, возникающие в теории, можно «спрятать» всего в две определяемые из опыта величины — заряд и массу электрона. Масса фотона должна быть равна нулю вследствие калибровочной инвариантности. Иными словами, независимо от типа процесса и структуры диаграммы все расходящиеся интегралы дают вклад либо в массу электрона, либо в его заряд. А если так, то мы можем собрать все эти бесконечности воедино и вместе с затравочной массой электрона  $m_0$  объявить всю сумму физической массой электрона:

$$m_{\text{физ}} = m_0 + \alpha \delta m_1 + \alpha^2 \delta m_2 + \dots,$$

где постоянные  $\delta m_i$  формально бесконечны. После этого «замечания бесконечностей под ковер» все величины получаются конечными, хорошо определенными.

Разумеется, эта процедура оставляет чувство неудовлетворенности, однако до последнего времени мы не имели для нее никакой альтернативы\*, а, с другой стороны, вычисления с использованием процедуры перенормировок показали блестящее согласие с экспериментом. Например, можно вычислить поправки к дираковскому магнитному моменту электрона. Они связаны с диаграммами типа рис. 15, которые модифицируют взаимодействие электронов с внешним магнитным полем  $H$ . Ответ составляет

$$\mu_e = \mu_{\text{Dirac}} \left( 1 + \frac{\alpha}{2\pi} - 0,328 \frac{\alpha^2}{\pi^2} + \dots \right).$$

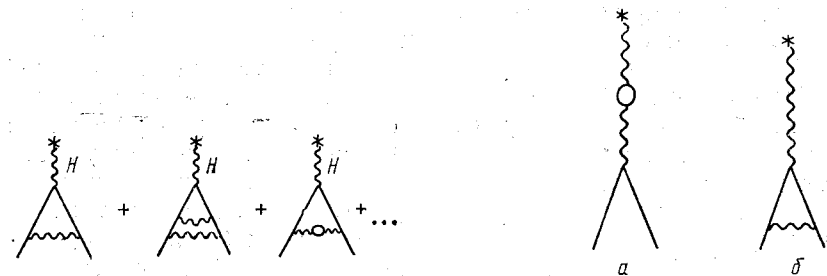


Рис. 15. Радиационные поправки к магнитному моменту электрона. Крестиком обозначен источник внешнего магнитного поля  $H$ . Волнистые линии представляют собой виртуальные фотоны

Рис. 16. Диаграммы, ответственные за расщепление  $S$ - и  $P$ -уровней в водороде

\* Суперсимметричные теории, развиваемые в последние годы, возможно, не страдают этим недостатком. Бесконечности в этих теориях могут и не возникать из-за того, что вклады фермионов и бозонов в расходящиеся интегралы взаимно сокращаются. Разумеется, полученные на основе метода перенормировок конечные результаты при этом не изменятся.

Он согласуется с экспериментом с относительной погрешностью лучше  $10^{-8}$ . Мало какая область физики может похвастать такой степенью согласия теории и эксперимента.

Другим ярким примером такого рода является расщепление  $2S_{1/2}$  и  $2P_{1/2}$  уровней атома водорода. Как известно, без учета квантовых поправок эти уровни вырождены, однако учет того, что поле отклоняется от кулоновского, и модификации взаимодействия электрона с фотоном (рис. 16, а и б соответственно) приводит к тому, что  $S$ -уровень оказывается выше  $P$ -уровня примерно на 1057 МГц, в прекрасном согласии с опытом.

Таким образом, электродинамика учит нас, что высшие поправки теории возмущений (так называемые радиационные поправки) существуют и приводят к наблюдаемому на опыте эффектам. Для слабых взаимодействий радиационные поправки пока не столь актуальны, так как их величина существенно меньше, чем в электродинамике, а точность опыта хуже. Однако логика теории требует возможности их вычисления. И тут локальная четырехфермионная теория, о которой шла речь выше, сталкивается с существенным затруднением — эта теория оказывается неперенормируемой. Иными словами, при вычислении все более высоких порядков теории возмущений возникают все новые типы бесконечностей, для устранения которых требуется вводить все новые константы перенормировки, а не удастся ограничиться лишь двумя (зарядом и массой), как в электродинамике. Корень зла лежит в том, что локальная амплитуда четырехфермионного процесса растет с ростом энергии сталкивающихся частиц. Поэтому при расчете высших порядков, когда приходится интегрировать по импульсам и энергиям виртуальных частиц, характер расходимости становится все более сильным по мере роста порядка теории возмущений.

Если взаимодействие происходит за счет обмена промежуточным бозоном, то четырехфермионная амплитуда уже не растет с ростом энергии и с этим связана перенормируемость электродинамики. Для перенормируемости, однако, существенно еще и то, что масса фотона равна нулю. Дело в том, что амплитуда испускания массивных векторных бозонов содержит фактор  $(E/m)^2$  и опять-таки растет с энергией. Для безмассовых частиц такие члены не возникают в силу калибровочной инвариантности. Но мы знаем, что промежуточные бозоны слабых взаимодействий должны быть массивны, и, следовательно, мы снова пришли к неперенормируемой теории. И тут на помощь приходит скалярное поле.

#### § 8. СПОНТАННОЕ НАРУШЕНИЕ СИММЕТРИИ. СКАЛЯРНОЕ (ХИГГСОВСКОЕ) ПОЛЕ

Идея построения состоит в следующем. Начнем с теории безмассовых полей  $\psi$ ,  $W^\pm$  и  $Z^0$ , а затем неким динамическим

механизмом, изменяющим свойства вакуума, дадим массу  $W^\pm$ - и  $Z^0$ -бозонам. Это так называемое мягкое включение массы, при котором исходная перенормируемость теории безмассовых векторных полей сохранится, так как поведение амплитуды при высоких энергиях, которое существенно для перенормируемости, не «чувствует» свойств вакуума.

Поясним этот механизм несколько более детально. Напомним сначала, что фаза волновой функции в квантовой механике является ненаблюдаемой величиной. Поэтому все величины не должны зависеть от выбора этой фазы. На формальном языке это звучит как требование инвариантности лагранжиана относительно преобразований вида

$$\psi \rightarrow e^{i\Phi} \psi. \quad (4.8)$$

Если сделать естественное предположение, что фазовые преобразования можно осуществлять независимо в разных точках пространства-времени, то мы приходим к ситуации, когда  $\Phi$  будет функцией координат  $\Phi = \Phi(x)$ . Однако наличие в лагранжиане производных поля  $\psi$  нарушает инвариантность относительно преобразований (4.8). Для того чтобы ее сохранить, лагранжиан нужно дополнить векторным полем  $A_\mu$ , преобразующимся по закону (3.10), а производные  $\partial_\mu$  заменить на ковариантные производные  $D_\mu$  (3.14).

Этот принцип калибровочной инвариантности, с успехом используемый в электродинамике, лежит в основе и более сложных теорий, таких, как теория электрослабого взаимодействия. Для спинорных полей, в лагранжиан которых производные входят в комбинации  $\bar{\psi} \gamma^\mu \partial_\mu \psi$ , процедура «удлинения» производной определяет вид тока фермионов  $e \bar{\psi} \gamma^\mu \psi A_\mu \equiv J^\mu \cdot A_\mu$  и тем самым полностью фиксирует взаимодействие с векторным полем. Вследствие калибровочной инвариантности теории ток  $J^\mu$  оказывается сохраняющимся.

В лагранжиан скалярного поля производные входят квадратично,  $(\partial_\mu \phi)(\partial^\mu \phi)$ , что дает при замене  $\partial_\mu$  на  $D_\mu$ :

$$\left| \left( \partial_\mu - \frac{ie}{c} W_\mu \right) \phi \right|^2.$$

После раскрытия этой скобки видно, что один из членов имеет вид

$$(e^2/c^2) W_\mu^2 \phi^2.$$

Если  $\phi \neq 0$ , то множитель  $e^2 \phi^2 / c^2$  эквивалентен массе поля  $W$ , в частности он входит как масса в уравнение поля  $W$ -бозонов\*:

\* В этом уравнении опущены члены, отвечающие взаимодействию с другими полями и, в частности, самодействие  $W$  (т. е. нелинейная часть).

$$\left( \square + \frac{e^2 \varphi^2}{c^2} \right) W_\mu = 0. \quad (4.9)$$

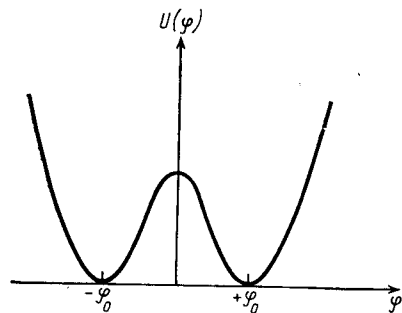
Таким образом, взаимодействие двух полей  $W$  и  $\varphi$  дает  $W$ -бозонам массу. При таком способе введения массы теория оказывается перенормируемой.

Физически картина напоминает появление массы фотона в плазме. В вакууме масса фотона равна нулю, а в плазме из-за взаимодействия со средой электромагнитное дальнее действие исчезает, что эквивалентно отличию массы фотона от нуля.

В нашей теории вакуум оказывается средой такого рода для  $W$ - и  $Z$ -бозонов.

Мы не ответили пока на вопрос, как можно добиться того, чтобы в вакууме, т. е. в низшем по энергии состоянии, поле  $\varphi$  имело бы всюду постоянное отличное от нуля значение. Предположим, что потенциальная энергия поля  $\varphi$  имеет форму, изображенную на рис. 17. Очевидно, что потенциал такого вида как раз приводит к тому, что нам необходимо. Низшему по энергии состоянию соответствует  $\varphi = \pm \varphi_0 \neq 0$ . В данной модели имеются два вакуумных состояния, поэтому говорят, что вакуум является вырожденным.

Рис. 17. Потенциал  $U$  поля  $\varphi$ , приводящий к спонтанному нарушению симметрии. Точка  $\varphi = 0$  отвечает неустойчивому положению равновесия, а точки  $\varphi = \pm \varphi_0$  — устойчивому. Когда  $\varphi = +\varphi_0$  или  $\varphi = -\varphi_0$ , симметрия состояния нарушена



Точка  $\varphi = 0$ , в которой потенциал  $u(\varphi)$  имеет локальный максимум, так же как и точки  $\varphi = \pm \varphi_0$  является точкой равновесия поля  $\varphi$ . Однако, в отличие от точек  $\varphi = \pm \varphi_0$ , в точке  $\varphi = 0$  равновесие поля  $\varphi$  неустойчиво и малые флуктуации приводят к тому, что поле начинает расти по амплитуде, «скатываясь» в один из минимумов  $\varphi = +\varphi_0$  или  $\varphi = -\varphi_0$ . По этой причине точку  $\varphi = 0$  называют ложным вакуумом теории. Поле  $\varphi_0$  называют хиггсовским, так как оно было введено в физику частиц в работах Хиггса (а также Энглерта и Браута и Гуральника, Хагена и Киббла — все в 1964 г.).

Потенциальная энергия вида рис. 17 уже давно известна в теории ферромагнетиков. Она описывает взаимодействие атомных магнитных моментов между собой. Точке  $\varphi = 0$  отвечает состояние, когда все магнитные моменты направлены хаотически, так что средний магнитный момент единицы объема вещества равен нулю. Однако энергетически более выгодно состоя-

ние, когда все магнитные моменты смотрят в одну произвольную сторону, что отвечает ненулевому среднему моменту единицы объема или на нашей картинке  $|\varphi| = \varphi_0$ . Направление среднего магнитного момента произвольно и это значит, что в данной модели вакуум бесконечно кратко выражен.

Теории такого типа называют теориями со спонтанно нарушенной симметрией. Точка  $\varphi = 0$  соответствует симметричному состоянию. В примере с ферромагнетиком это была симметрия относительно вращений. В примере со скалярным полем при  $\varphi = 0$  мы имели симметрию свойств фотона,  $W^\pm$ - и  $Z^0$ -бозонов. Заметим, что массы всех четырех бозонов в этом случае были равны нулю.

В точке  $\varphi \neq 0$  симметрия нарушена, причем система переходит в это состояние самопроизвольно, спонтанно. Массы векторных бозонов при этом становятся разными, а для ферромагнетика в этом состоянии появляется выделенное направление — направление среднего магнитного момента и симметрия относительно вращений нарушается. Таким образом, хотя лагранжиан инвариантен относительно каких-то преобразований симметрии во внутреннем изотопическом пространстве, основное состояние таковым не является.

Заметим (это нам понадобится в дальнейшем), что нагревание ферромагнетика разрушает выстроенность магнитных моментов и приводит систему в симметричное состояние  $\varphi = 0$ .

Введение скалярного поля со столь специфической потенциальной энергией является результатом требования перенормируемости теории, основанной на обмене массивными векторными бозонами. Потенциал для простейшего случая, изображенного на рис. 17, можно записать в виде

$$V(\varphi) = \frac{\lambda}{4} (\varphi^2 - \varphi_0^2)^2. \quad (4.10)$$

Полю с лагранжианом вида

$$\mathcal{L} = \frac{1}{2} \left( \frac{\partial \varphi}{\partial x^\mu} \right)^2 - V(\varphi)$$

соответствуют частицы с массой  $m = \sqrt{2\lambda} \varphi_0$ .

В объединенных теориях электрослабых взаимодействий их зовут хиггсовскими бозонами и обозначают  $H$  (или  $\chi$ ). Учет обмена хиггсовскими бозонами в точности сокращает упомянутый выше рост амплитуд с энергией, связанный с ненулевой массой  $W$ - и  $Z$ -бозонов, и, таким образом, хиггсовское поле обеспечивает перенормируемость теории.

В эксперименте  $H$ -бозоны до сих пор не обнаружены и их

\* Более подробно эти вопросы обсуждаются в гл. 6, § 1, посвященном скалярному полю в плоском пространстве-времени.

Экспериментально обнаружены

масса неизвестна. Экспериментальное доказательство существования хиггсовских бозонов является в настоящее время одним из центральных задач в физике элементарных частиц.

Комплексное хиггсовское поле может образовывать интересные космологически значимые объекты, так называемые вихревые нити. При спонтанном нарушении симметрии в теории с потенциалом  $V(|\phi|) = \lambda(|\phi|^2 - \phi_0^2)^2$  фаза поля  $\phi$  в причинно-несвязанных точках различна и может оказаться так, что набег фазы по замкнутому контуру будет отличен от нуля:  $\Delta\phi = 2\pi n$ . Это означает, что образовался топологически стабильный объект — вихревая нить (точнее,  $n$  нитей). В центре нити  $\langle\phi\rangle = 0$ , а на большом расстоянии  $|\langle\phi\rangle| = \phi_0$ . Эти объекты имеют микроскопические поперечные размеры  $L \sim (\lambda\phi_0)^{-1}$  и макроскопические продольные. Они могут играть важную роль в образовании крупномасштабной структуры Вселенной, являясь зародышами растущих неоднородностей плотности.

## Глава 5.

### КОСМОЛОГИЧЕСКАЯ ПОСТОЯННАЯ

В теории ранней Вселенной и решении проблем горизонта, плоскостности и других чрезвычайно важную роль играет метрика де Ситтера, отвечающая модели мира с экспоненциальным расширением. Метрика де Ситтера — это однородное и изотропное решение уравнений ОТО в пустоте при наличии космологической постоянной, называемой еще  $\Lambda$ -членом. Он был введен в уравнения гравитации Эйнштейном, чтобы получить стационарную космологическую модель за счет компенсации гравитационного воздействия материи и  $\Lambda$ -члена. Когда выяснилось, что Вселенная расширяется, Эйнштейн счел добавление  $\Lambda$ -члена излишним. Однако сейчас мы убеждены в том, что на ранней стадии расширение Вселенной все же определялось именно  $\Lambda$ -членом или каким-то имитирующим его состоянием материи.

Наблюдения показывают, что космологическая постоянная в современную эпоху либо очень мала, либо вовсе равна нулю. С другой стороны, и квантовая теория поля, и анализ фазовых переходов при остывании Вселенной предсказывают для нее очень большую величину. Мы приходим таким образом к проблеме космологической постоянной в теории поздней Вселенной.

#### § 1. ОПРЕДЕЛЕНИЕ И ФИЗИЧЕСКИЙ СМЫСЛ

Не нарушая ни принципа общей ковариантности, ни сохранения энергии-импульса,  $D_\mu T^{\mu\nu} = 0$ , в уравнения Эйнштейна

(3.21) можно ввести дополнительное слагаемое

$$R_{\mu\nu} - 1/2 g_{\mu\nu} R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (5.1)$$

где константа  $\Lambda$  называется космологической постоянной. В отличие от первых двух членов в левой части (5.1) слагаемое  $\Lambda g_{\mu\nu}$  не исчезает, когда  $g_{\mu\nu}$  совпадает с метрическим тензором плоского пространства-времени. Этот член выглядит как некая добавка к тензору энергии-импульса материи. Он ковариантно сохраняется, так как

$$D_\alpha g_{\mu\nu} = 0, \quad (5.2)$$

и обладает тем замечательным свойством, что одинаково выглядит во всех инерциальных системах координат. По этой причине величину  $\Lambda/8\pi G$  называют еще плотностью энергии вакуума, так как вакуум не изменяется при преобразованиях Лоренца. Квантовая теория дает глубокие физические основания для такого названия. Дело в том, что в квантовой теории энергия основного состояния — вакуума, вообще говоря, отлична от нуля. Среднее по вакууму значение тензора энергии-импульса в плоском пространстве-времени как раз пропорционально  $g_{\mu\nu}$ :

$$T_{\mu\nu}^{\text{вак}} \equiv \langle \text{вак} | T_{\mu\nu} | \text{вак} \rangle = \rho_{\text{вак}} g_{\mu\nu}. \quad (5.3)$$

Подробнее мы вернемся к этому вопросу в § 4.

Заметим, что тензор энергии-импульса идеальной жидкости

$$T_{\mu\nu} = (p + \rho) u_\mu u_\nu - p g_{\mu\nu} \quad (5.4)$$

при  $p = -\rho$  сводится к вакуумному тензору (5.3), т. е. вакуумный тензор энергии-импульса (5.3) обладает упоминавшимся выше свойством  $p = -\rho$ . Покажем теперь непосредственно, что соотношение  $p = -\rho$  не меняется при преобразованиях Лоренца. Это утверждение практически очевидно, но мы проведем эти простые выкладки из-за большой важности результата.

Компоненты тензора второго ранга, в частности  $T_{\mu\nu}$ , преобразуются согласно выражениям

$$T_{00} = (1 - v^2)^{-1} (T'_{00} - 2vT'_{01} + v^2T'_{11}),$$

$$T_{11} = (1 - v^2)^{-1} (T'_{11} - 2vT'_{01} + v^2T'_{00})$$

при движении штрихованной системы  $K'$  вдоль оси  $Ox_1$  нештрихованной системы  $K$ . Преобразования остальных компонент разбиваются на две группы: компоненты  $T_{22}, T_{23}, T_{33}$  не преобразуются:  $T_{23} = T'_{23}, T_{22} = T'_{22}, T_{33} = T'_{33}$ . К другой группе принадлежат компоненты  $T_{12}, T_{13}$ , а также  $T_{01}, T_{02}$  и  $T_{03}$ :



$$T_{12} = (1-v^2)^{-1/2} (T'_{12} - vT'_{02}), \quad T_{02} = (1-v^2)^{-1/2} (T'_{02} - vT'_{12}),$$

$$T_{01} = (1-v^2)^{-1/2} [(1+v^2)T'_{01} - v(T'_{00} + T'_{11})],$$

аналогично для  $T_{02}$  и  $T_{13}$ .

Пусть в штрихованной системе тензор  $T'_{\mu\nu}$  имеет вид (5.3). Отметим, что поскольку  $T'_{0i} = (\varepsilon + p)v'_i$ , то потоки энергии  $K'$ -системе отсутствуют и  $T'_{0i} = 0$ . Поскольку давление в  $K'$ -системе изотопно, «скальвающие» напряжения отсутствуют, т. е.  $T'_{12} = T'_{13} = T'_{23} = 0$ . Остаются нетривиальные преобразования

$$T_{00} = (1-v^2)^{-1} (T'_{00} + v^2 T'_{11}),$$

$$T_{11} = (1-v^2)^{-1} (T'_{11} + v^2 T'_{00}),$$

$$T_{01} = -v(1-v^2)^{-1/2} (T'_{00} + T'_{11}).$$

Подставляя значения  $T'_{00} = \varepsilon$  и  $T'_{11} = -\varepsilon$ , получаем равенства  $\varepsilon = \varepsilon'$  и  $p = -\varepsilon$ . Отсюда делаем вывод, что условие  $p = -\varepsilon$  является лоренц-инвариантным. Внимательный читатель мог бы увидеть это без выкладок из лоренц-инвариантности метрики Минковского.

Отметим, что решение Минковского не согласовано с  $\Lambda \neq 0$  (что, впрочем, не опровергает сказанного выше). Решение уравнений ОТО с  $\Lambda \neq 0$  отвечает искривленному пространству. Если при этом обычная материя отсутствует, т. е.  $T_{\mu\nu} = 0$ , то такое решение носит название решения де Ситтера.

## § 2. РЕШЕНИЕ ДЕ СИТТЕРА

Добавление  $\Lambda$ -члена не изменит общего вида пространственно-временного интервала в однородной и изотропной Вселенной (2.1), что заранее очевидно, так как  $\Lambda$ -член эквивалентен специальному выбору  $T_{\mu\nu}$ . Уравнения, определяющие эволюцию масштабного фактора  $a(t)$ , при этом примут вид

$$\frac{1}{2} \left( \frac{da}{dt} \right)^2 - \frac{4\pi G}{3} \rho a^2 = \frac{1}{6} \Lambda a^2 - \frac{k}{2};$$

это аналог уравнения (2.2) или закона сохранения энергии. Уравнение, связывающее «ускорение», т. е. величину  $\ddot{a}/a$  с полной массой, тоже изменится:

$$\ddot{a} = \frac{1}{3} \Lambda a - \frac{4\pi G}{3} (\rho + 3p) a,$$

здесь  $p$  и  $\rho$  — давление и плотность вещества, вакуумные давления и плотность записаны как  $\Lambda$ .

Рассмотрим теперь физический смысл  $\Lambda$ -члена. Для этого используем следующую простую модель. Выделим в пространстве шарик малого радиуса  $r_0$  и посмотрим, как будет выглядеть его уравнение движения. Расстояние между центром шарика и какой-либо точкой на его поверхности будет меняться согласно

$$\frac{d^2(ar_0)}{dt^2} = \frac{1}{3} \Lambda ar_0 - \frac{GM}{(ar_0)^2}. \quad (5.5)$$

В этом уравнении мы сделали замену обозначений, введя вместо величины  $(4\pi/3)(ar_0)^3(\rho + 3p)$  одну величину  $M$ , обозначающую полную массу. Теперь из (5.5) видно, что на пробную частицу на поверхности шарика, кроме стандартной ньютоновской силы

$$F_H = -GM/(ar_0)^2,$$

дополнительно действуют необычные, но тоже гравитационные силы, связанные с  $\Lambda$ -членом:

$$F_\Lambda = 1/3 \Lambda ar_0.$$

Поле таких сил является глобальным и однородным. Величина самой силы зависит от расстояния между взаимодействующими пробными частицами как  $(ar_0)$ . Чем дальше эти частицы друг от друга, тем больше сила. Отметим формальное сходство с потенциалом взаимодействия кварков, обсуждавшимся выше.

В современную эпоху  $\Lambda$ -член очень мал или равен нулю и силы  $\sim \Lambda ar_0$  в обычной астрономической ситуации не играют никакой роли. Они могут проявлять себя на расстояниях, сравнимых с горизонтом, и важны только для космологии, в связи с чем  $\Lambda$  иногда называют космологической постоянной.

При положительном  $\Lambda$  сила  $F_\Lambda$  соответствует отталкиванию тем более сильному, чем дальше друг от друга находятся частицы. Наоборот, при отрицательной величине  $\Lambda$  появляются дополнительные силы притяжения. Это является физической причиной, из-за которой масштабный фактор в мире де Ситтера с замкнутыми, плоскими и открытыми пространственными сечениями ведет себя не так, как в мире Фридмана. В частности, даже в открытом мире масштабный фактор не увеличивается бесконечно, а, переходя через максимум, вновь стремится к нулю.

Закон изменения плотности энергии со временем определяется уравнением  $\rho = \rho(\rho)$  и имеет в важных частных случаях следующий вид:

$$\frac{d\rho}{dt} = -3H\rho, \quad \rho = 0;$$

$$\frac{d\rho}{dt} = -4H\rho, \quad \rho = \rho/3;$$

$$\frac{d\rho}{dt} = 0, \quad \rho = -\varepsilon.$$

Именно то обстоятельство, что  $\rho_{\text{вак}}$  не убывает, изменяет предсказания относительно будущего Вселенной.

При отрицательном  $\Lambda$ , что соответствует наличию дополнительных сил притяжения, во Вселенной после некоторого периода расширения обязательно начнется сжатие. Это видно из того, что при росте масштабного фактора силы космологического притяжения тоже растут:  $F_{\Lambda} \sim \Lambda a$  (так как при  $\Lambda < 0$  сила направлена к центру). При положительном  $\Lambda$ , что соответствует наличию космологических сил отталкивания, поведение  $a(t)$  разнообразнее, чем в случае  $\Lambda < 0$ . В частности, приобретает условный смысл само понятие замкнутого, плоского или открытого пространства. Формально говоря, можно так провести разбиение четырехмерного пространства-времени на трехмерное пространство и время, что полученное трехмерное пространство при одном и том же наборе космологических параметров может быть замкнутым, плоским или открытым. Подробнее на этом удивительном свойстве пространства-времени де Ситтера мы остановимся ниже.

Топологическое отличие пространства де Ситтера от пространства Фридмана связано с тем, что при  $\Lambda = 0$  трехмерно замкнутый мир  $t = \text{const}$  замкнут и по времени. При  $\Lambda \neq 0$  он открыт по времени и, как мы увидим ниже, другое трехмерное сечение с  $t' = \text{const}$  становится открытым.

Рассмотрим решения уравнений тяготения с  $\Lambda$ -членом. Вспомним вначале более простой случай  $\rho = 0$ . Уравнения, определяющие эволюцию масштабного фактора, можно записать в виде

$$\ddot{a} = -\frac{GM_1}{a}, \quad \frac{1}{2} \dot{a}^2 = \frac{GM_2}{a} - \frac{k}{2}.$$

При  $\rho = 0$  в обоих уравнениях стоит одна и та же масса  $M_1 = M_2 = \rho V$ . Первое уравнение представляет собой ньютоновское уравнение движения, а второе — закон сохранения энергии. Первое уравнение можно получить путем дифференцирования второго, полагая  $M = \text{const}$ .

Если  $\rho \neq 0$ , положение меняется. Теперь уже  $M_1 = V(\rho + 3p)$ , а  $M_2 = \rho V$ . При этом с уравнениями согласован известный закон изменения полной энергии  $M_2 = -pdV$ .

В частном случае  $\rho = -p$  уравнение движения имеет вид

$$\ddot{a} = \frac{8\pi G\rho}{3} a. \quad (5.6)$$

Оно, очевидно, имеет решения вида  $\exp\left(\pm \sqrt{\frac{8\pi G\rho}{3}} t\right)$ . Кон-

кретная форма решения зависит от величины  $k$ , определяющей соотношение между кинетической и потенциальной энергией.

При положительном  $\Lambda$  и  $k=0$  уравнение имеет простое экспоненциальное решение

$$a(t) = a_0 \exp\{\sqrt{\Lambda/3} t\}. \quad (5.7)$$

Параметр Хаббла во Вселенной с такой метрикой является постоянным (здесь действительно можно говорить о постоянной Хаббла, в отличие от фридмановского расширения, когда  $H$  зависит от  $t$ ):

$$H = \frac{\dot{a}}{a} = \sqrt{\Lambda/3}. \quad (5.8)$$

Красное смещение в мире де Ситтера с  $k=0$  ведет себя как

$$Z = \exp\{\sqrt{\Lambda/3}(t_0 - t)\} - 1.$$

При  $k = \pm 1$  масштабный фактор будет вести себя уже как гиперболический косинус или синус. Конкретно, при  $k = +1$ , что в метрике Фридмана соответствует замкнутому миру, имеющему топологические свойства сферы, масштабный фактор имеет вид

$$a(t) = a_0 \text{ch}\{\sqrt{\Lambda/3} t\}. \quad (5.9)$$

В этом выражении  $a_0$  — произвольная постоянная. Она связана с  $\Lambda$  выражением

$$a_0 = \sqrt{\Lambda/3}.$$

Решением уравнений (5.6) при  $k = -1$  является функция

$$a(t) = \sqrt{\frac{3}{\Lambda}} \text{sh}(\sqrt{\Lambda/3} t). \quad (5.10)$$

Параметр Хаббла в метрике де Ситтера с  $k \neq 0$  не является постоянной величиной, а выражается через гиперболический тангенс или котангенс

$$H = \sqrt{\Lambda/3} \text{th}(\sqrt{\Lambda/3} t) \quad \text{при } k = +1, \quad (5.11)$$

$$H = \sqrt{\Lambda/3} \text{cth}(\sqrt{\Lambda/3} t) \quad \text{при } k = -1.$$

Заметим, что при  $\sqrt{\Lambda/3} t \gg 1$  различие между этими тремя типами решений экспоненциально быстро стирается.

Пространство де Ситтера обладает весьма интересным свойством: оно имеет очень высокую степень симметрии, такую же, как плоское пространство, и благодаря этому выбор направле-

ния оси времени в нем в значительной мере произволен. В мире Фридмана такого произвола нет. Мы можем разделить четырехмерное пространство-время на трехмерное пространство и ось времени. Для выделения трехмерной гиперповерхности у нас есть физический критерий. На всей гиперповерхности  $t = \text{const}$  и, следовательно,  $\rho = \text{const}$ . Во Вселенной, описываемой метрикой Фридмана, плотность со временем изменяется и поверхности равной плотности определяют направление оси времени — ось перпендикулярна гиперповерхности, на которой  $\rho = \text{const}$ . Для мира де Ситтера, в котором  $\rho = \text{const}$ , критерий для выбора оси  $t$  нет. По-разному направляя ось времени в одном и том же 4-мерном пространстве, мы можем получить все три решения, отвечающие  $k = +1, 0$  и  $-1$ . Заметим, что из них только замкнутое решение ( $k = +1$ ) покрывает все указанное 4-мерное пространство, а плоское и открытое решения являются некоторыми частями его. Ниже мы проиллюстрируем это в простом случае двумерного пространства, которое можно представить себе как поверхность в псевдоевклидовом трехмерном пространстве.

Для нахождения группы инвариантности и описания топологии мира де Ситтера воспользуемся следующим приемом. Представим себе решение де Ситтера как поверхность гиперболоида вращения, вложенного в абстрактное пятимерное пространство с интервалом вида

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 - d\mathcal{Y}^2;$$

здесь  $t, x, y, z$  — обычные четырехмерные координаты,  $\mathcal{Y}$  — пятая, абстрактная координата, которая и позволяет рассматривать четырехмерное пространство как вложенное в пятимерное.

Уравнение гиперболоида в 5-мерном пространстве имеет вид

$$t^2 - x^2 - y^2 - z^2 - \mathcal{Y}^2 = -r_0^2. \quad (5.12)$$

Вводя сферические координаты в этом пространстве, параметризуем поверхность гиперболоида новой переменной  $\tau$ , которая будет в дальнейшем играть роль времени в четырехмерном мире, являющемся поверхностью гиперболоида вращения в пятимерном пространстве:

$$\begin{aligned} t &= r_0 \operatorname{sh} H\tau, & x &= r_0 \operatorname{ch} H\tau \sin r \sin \vartheta \cos \varphi, \\ y &= r_0 \operatorname{ch} H\tau \sin r \sin \vartheta \sin \varphi, & z &= r_0 \operatorname{ch} H\tau \sin r \cos \vartheta, \\ \mathcal{Y} &= r_0 \operatorname{ch} H\tau \cos r. \end{aligned}$$

Введенные таким образом функции ( $t, x, y, z, \mathcal{Y}$ ) описывают поверхность гиперболоида, другими словами, удовлетворяют уравнению (5.12). Интервал, выражаемый уже через 4 координаты ( $\tau, r, \vartheta, \varphi$ ), описывается формой

$$ds^2 = d\tau^2 - H^{-2} \operatorname{ch}^2 H\tau [dr^2 + \sin^2 r (d\vartheta^2 + \sin^2 \vartheta d\varphi^2)],$$

если  $r_0^2 = H^{-2}$ , т. е. является решением де Ситтера при  $k = +1$ . Отметим сразу, что это самая полная координатная система, покрывающая весь мир де Ситтера целиком.

Представить себе 4-мерный гиперболоид, вложенный в пятимерное пространство, чрезвычайно трудно, поэтому для наглядности опишем это вложение на примере вложения двухмерного гиперболоида в трехмерный псевдоевклидов мир.

Интервал в трехмерном псевдоевклидовом мире возьмем в виде

$$ds^2 = dt^2 - dx^2 - d\mathcal{Y}^2;$$

здесь  $t$  играет роль «трехмерного» времени,  $\mathcal{Y}$  — дополнительная абстрактная координата. Уравнение, описывающее гиперболоид, вложенный в трехмерное пространство, возьмем аналогично (5.12):

$$t^2 - x^2 - \mathcal{Y}^2 = -r_0^2. \quad (5.13)$$

Так же, как и в пятимерном пространстве, введем координату  $r$  — аналог сферических координат и параметр  $\tau$  вдоль гиперболоида:

$$\begin{aligned} t &= r_0 \operatorname{sh} (\tau/\tau_0), & x &= r_0 \operatorname{ch} (\tau/\tau_0) \sin r, \\ \mathcal{Y} &= r_0 \operatorname{ch} (\tau/\tau_0) \cos r. \end{aligned}$$

Естественно, что введенные таким образом три функции  $t, x, \mathcal{Y}$  автоматически удовлетворяют уравнению (5.12) и дают двумерный интервал вдоль поверхности гиперболоида в виде

$$ds^2 = d\tau^2 - r_0^2 \operatorname{ch}^2 (\tau/\tau_0) dr^2.$$

Интервал напоминает метрику замкнутого мира. Действительно, хотя координата  $r$  изменяется от  $-\infty$  до  $+\infty$ , она является циклической и фактически меняется на интервале  $(0, 2\pi)$ .

Сечениям гиперболоида  $t = \text{const}$  отвечают пространственные сечения, а линии, образованные сечениями гиперболоида плоскостью  $t = \text{const}$ , являются одномерным пространством в выбранном нами мире. Эти сечения являются кругами с длиной  $l = 2\pi r_0 \operatorname{ch} (\tau/\tau_0)$ , другими словами, одномерной моделью замкнутого мира. Круг с минимальной длиной, или, что эквивалентно, минимальный объем трехмерного пространства, достигается при  $\tau = 0$  или на сечении  $t = 0$ . Следует особо отметить, что определение момента, когда объем минимальный, неинвариантно. Действительно, при переходе к другой системе координат, движущейся относительно исходной со скоростью  $\beta$ ,

$$t = \frac{t' - \beta x'}{\sqrt{1 - \beta^2}}, \quad x = \frac{x' - \beta t'}{\sqrt{1 - \beta^2}} \quad \text{и} \quad \mathcal{Y} = \mathcal{Y}'.$$

Уравнение, описывающее гиперboloид, сохраняет свой вид в новых координатах

$$t'^2 - x'^2 - \mathcal{V}'^2 = -r_0^2,$$

интервал тоже сохраняет свою форму:

$$ds^2 = d\tau'^2 - r_0^2 \operatorname{ch}^2(\tau'/\tau_0) dr'^2.$$

В новой системе координат круг с минимальной длиной  $l = 2\pi r_0$  тоже определяется уравнением  $\tau' = 0$ , но со старым кругом этот уже не совпадает. По отношению к прежней секущей плоскости с уравнением  $t = \operatorname{const}$  новая плоскость в координатах  $(t, x)$  будет иметь вид  $t + \beta x = \operatorname{const}$ . Плоскость, выделяющая круг минимального радиуса, будет иметь уравнение

$$t + \beta x = 0,$$

а следовательно, будет наклонена к плоскости  $t = \operatorname{const}$  на угол  $\beta$ , как показано на рис. 18. Заметим, что в центральной точке

можно провести бесконечно много окружностей с экстремальной длиной. Это связано со знаком неопределенности метрики. Аналогичное явление имеет место и в пространстве Минковского. Заметим еще, что линия  $x = \operatorname{const}$  отвечает максимуму  $\Delta t$ . В связи с этим вспомним парадокс близнецов: «кто в движении, тот моложе».

На этой наглядной модели мира де Ситтера можно показать, как из единого гиперboloида

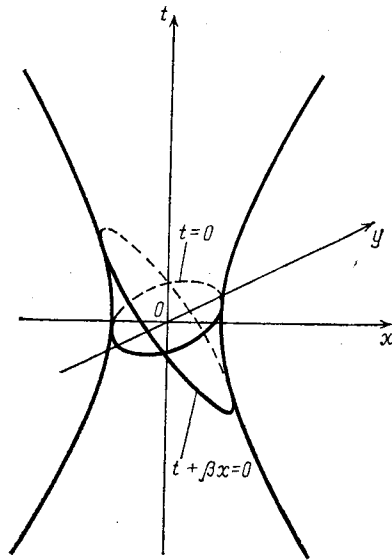


Рис. 18. Сечения двумерного пространства де Ситтера, отвечающие пространственно замкнутому миру. Все окружности с центром в точке 0, касающиеся поверхности гиперboloида, имеют одинаковую длину

да получаются различные миры: замкнутый, плоский и открытый.

Мы выделили замкнутый мир де Ситтера, но можем также получить метрику открытого мира. Правда, теперь придется ввести другое определение времени и другую параметризацию поверхности гиперboloида. Рассмотрим сечение гиперboloида плоскостью, перпендикулярной оси  $O\mathcal{U}$ , задаваемой уравнением

$\mathcal{V} = \operatorname{const}$ . Сечения представляют собой гиперболы, описываемые уравнениями  $t^2 - x^2 = \mathcal{V}^2 - r_0^2$ . Выберем параметризацию поверхности гиперboloида в виде

$$t = r_0 \operatorname{sh} \frac{\tau}{\tau_0} \operatorname{ch} \chi, \quad x = r_0 \operatorname{sh} \frac{\tau}{\tau_0} \operatorname{sh} \chi, \quad \mathcal{V} = r_0 \operatorname{ch} \frac{\tau}{\tau_0}.$$

Метрическая форма теперь примет вид

$$ds^2 = d\tau^2 - r_0^2 \operatorname{sh}^2 \frac{\tau}{\tau_0} d\chi^2,$$

что соответствует метрике открытого мира.

Пространство в этом мире представляет собой гиперболу, образованную пересечением гиперboloида с плоскостью, перпендикулярной оси  $O\mathcal{U}$ . Длина гиперболы неограниченна, что соответствует открытому миру. Время теперь отсчитывается

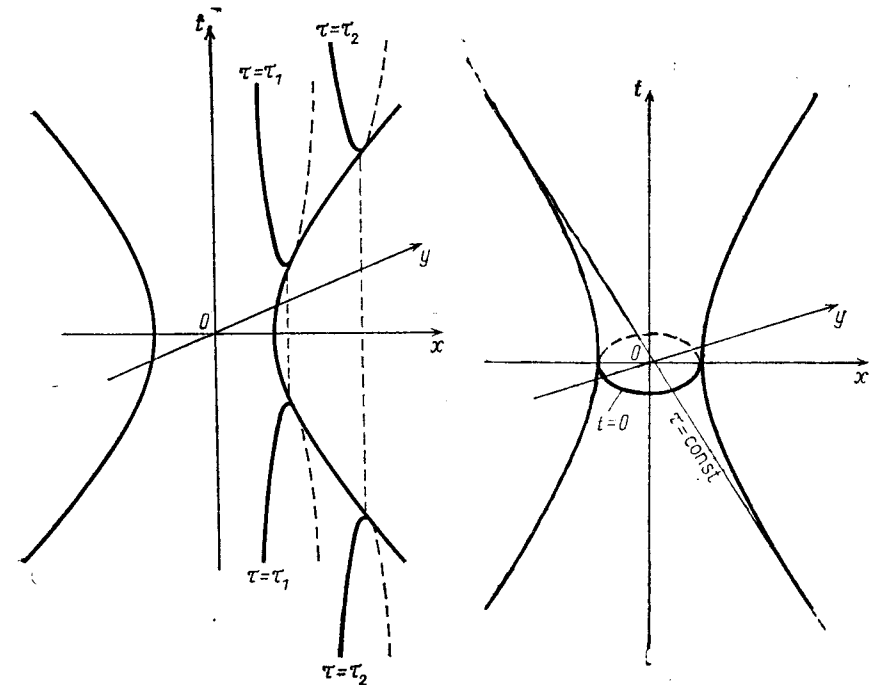


Рис. 19. Сечения двумерного пространства де Ситтера, отвечающие пространственно открытому, но искривленному миру

Рис. 20. Сечения двумерного пространства де Ситтера, отвечающие пространственно плоскому миру

вдоль оси  $O\mathcal{U}$ , как изображено на рис. 19. Вдоль гиперболы, разумеется,  $\tau = \operatorname{const}$ , поэтому разные гиперболы представляют

собой пространственноподобные гиперповерхности, относящиеся к разным моментам времени. Еще раз подчеркнем здесь условность выбора оси времени.

Наконец, не выписывая явных формул параметризации, нарисуем наглядный образ мира с плоскими пространственными сечениями. Как известно, гиперлоид можно реализовать прямолинейными образующими, подобно тому как это изображено на рис. 20. Такая прямая отвечает плоским пространственным сечениям в мире де Ситтера. Перпендикулярно к этим прямым отсчитывается время, выделяя тем самым метрику де Ситтера с плоскими пространственными сечениями.

Таким образом, разбивая различными способами четырехмерный мир де Ситтера на пространство и время, мы получаем различные миры де Ситтера: открытый, плоский или замкнутый. Следует отметить, что замкнутый мир покрывает весь гиперлоид, а открытый и плоский — не весь. В этом смысле говорят о геодезической неполноте последних.

К возможности преобразовать мир с плоскими пространственными сечениями к миру с замкнутыми сечениями вплотную примыкает вопрос о степени подвижности мира де Ситтера или, на формальном языке, о группе преобразований в мире де Ситтера. Группа преобразований в пространстве де Ситтера как и в пространстве Минковского, является 10-параметрической группой преобразований. Это означает, что существует 10 фундаментальных преобразований симметрии, с помощью которых можно генерировать любое преобразование. Перечислим их. Во-первых, из-за изотропии мира можно делать любые пространственные повороты. Группа вращений характеризуется тремя параметрами. Точно так же свойства пространства де Ситтера не изменяются при пространственных сдвигах вдоль трех независимых направлений. Помимо этих преобразований, определяемых шестью параметрами де Ситтера, есть еще четыре. Три из них — комплексные «повороты» с преобразованием оси времени. Физически эти преобразования соответствуют переходу в другую равномерно движущуюся систему координат. Наконец, еще один параметр — сдвиг вдоль временной оси. Все эти десять параметров характеризуют так называемую группу движений мира де Ситтера. Говоря точно, группой симметрии пространства де Ситтера является 10-параметрическая группа однородных преобразований Лоренца в 5-мерном пространстве  $SO(4,1)$ , которая как раз и генерирует упомянутые выше преобразования.

Рассмотрим теперь поведение основных кинематических величин в мире де Ситтера. А именно найдем, как ведет себя горизонт в метрике де Ситтера, в чем его отличие от горизонта в метрике Фридмана. Кинематические характеристики мира де Ситтера скажут нам, что такое проблема плоскостности

в мире де Ситтера и как ведут себя причинно-связанные области в этом мире.

Напомним определение горизонта, введенное в § 2 гл. 2. Расстояние, которое луч света проходит от момента излучения  $t_e=0$  до момента наблюдения  $t_0$ , называется горизонтом:

$$l_r = a(t) \int_{t_e}^{t_0} \frac{dt'}{a(t')}.$$

В мире с плоскими пространственными сечениями масштабный фактор равен

$$a(t) = a_0 e^{Ht},$$

а горизонт  $l_2$  зависит от момента  $t_0$ , как

$$l_2 = H^{-1} e^{Ht_0} (1 - e^{-Ht_0}) \quad \text{при } t_e = 0.$$

Хотя  $l_2$  и очень быстро возрастает, его рост все же медленнее, чем рост масштабного фактора. Поэтому в  $r$ -координатах,  $r = l/a(t)$ , расстояние до горизонта в мире де Ситтера при возрастании времени стремится к постоянной величине, равной

$$r_2 \rightarrow (a_0 H)^{-1}. \quad (5.14)$$

Это значит, что даже в бесконечно удаленном будущем световыми сигналами смогут обменяться только те точки, которые первоначально находились внутри сферы радиуса  $H^{-1}$ . В частности, если параметр Хаббла  $H \approx t_{Pl}^{-1}$ , то в момент «рождения» мира де Ситтера устанавливается причинная связь в сфере размером  $t_{Pl}$  и в  $r$ -координатах эта причинная связь не распространяется ни на какие точки! Точки, разделенные первоначально расстоянием  $l > H^{-1}$ , никогда не смогут обменяться сигналами. Отметим, что в современном мире эти точки находятся экспоненциально далеко друг от друга: при  $Ht_0 > 100$  причинно-связанная область превосходит размер современного горизонта.

Сравним теперь этот вывод с тем, который мы сделали при анализе горизонта в метрике Фридмана. Там расстояние до горизонта совсем по-другому зависело от момента наблюдения:

$$l_r = \text{const } t_0, \quad r_r = l_r/a(t) \sim \begin{cases} t^{1/2} & \text{RD-случай,} \\ t^{2/3} & \text{MD-случай.} \end{cases}$$

В бесконечном будущем устанавливалась связь между бесконечно удаленными точками (в  $r$ -координатах). В этом принципиальное отличие кинематики мира Фридмана от кинематики мира де Ситтера. Как это ни парадоксально, выяснится, что стадия де Ситтера, которая была в нашей Вселенной в ранние эпохи, помогает установить причинную связь во всем наблюдаемом сейчас мире Фридмана.

Перейдем теперь к рассмотрению того, как ведут себя с течением времени размеры причинно-связанных областей в мире де Ситтера с  $k=0$ . Для этого прежде всего надо заметить, что координата  $r$  является так называемой лагранжевой координатой частицы\*. Различные численные значения  $r$  характеризуют, грубо говоря, различные частицы. Расстояние между этими частицами в  $r$ -координатах, разумеется, остается постоянным, не зависит от времени. Физическое расстояние между ними, или расстояние, измеряемое лучом света, изменяется.

Вновь рассмотрим численный пример с поведением причинно-связанных областей в мире де Ситтера. Как и прежде, будем считать, что в момент  $t=t_{p1}$  у нас «родился» мир де Ситтера с параметром Хаббла  $H=t_{p1}^{-1}$  и с причинно-связанными областями размером  $l \approx l_{p1}$ . По прошествии приблизительно 70 планковских времен причинно-связанные области вырастут в  $e^{70} \approx 10^{30}$  раз и будут занимать размер  $\sim 10^{-3}$  см. Это как раз тот размер, который необходим для решения проблемы причинной связности нашего мира. Однако  $Ht=70$  отнюдь не предел для инфляции. Существующие модели дают  $Ht \gg 70$ , поэтому естественно ожидать, что размер области, возникшей из  $l=l_{p1}$ , много больше современного горизонта.

Величина горизонта в мире де Ситтера с замкнутыми пространственными сечениями отличается от (5.14) только численным множителем. Формально для масштабного фактора вида  $a(t) = a_0 \text{ch } Ht$  горизонт в  $r$ -координатах равен  $r_r = \pi / (2Ha_0)$  при  $Ht_0 \gg 1$ . (5.15)

Принципиальное отличие от этой ситуации возникает в мире де Ситтера с открытым пространством. В нем масштабный фактор ведет себя как

$$a(t) = a_0 \text{sh } Ht,$$

а интеграл, определяющий размер горизонта, логарифмически расходится на нижнем пределе. А именно:

$$a_0 r_r = \int_{t_e}^{t_0} \frac{dt}{\text{sh } Ht} = H^{-1} \left[ \ln \frac{e^{Ht_0} - 1}{e^{Ht_0} + 1} + \ln \frac{e^{Ht_e} + 1}{e^{Ht_e} - 1} \right] \approx \approx H^{-1} [\ln 2 + \ln(Ht_e)] \text{ при } t_e \rightarrow 0.$$

Рассуждая формально, мы должны прийти к выводу, что причинная связь устанавливается во всем пространстве. Действительно, для  $t_e \rightarrow 0$  размер горизонта неограниченно растет. Однако это нефизический результат. Решение парадокса заклю-

\* Это один из возможных, хотя и не единственный выбор. По определению лагранжевые координаты представляют собой совокупность траекторий пробных частиц.

чается в том, что деситтеровская стадия в реальном мире начинается с  $t_{p1}$  или даже с еще больших моментов времени и вместо  $t_e$  надо подставлять не нуль, а момент начала деситтеровской стадии.

Итак, мы обсудили вопрос о размерах горизонта в метрике де Ситтера с различными пространственными сечениями, а также эволюцию размеров причинно-связанных областей в мире де Ситтера. Заключая обсуждение кинематических характеристик метрики де Ситтера, посмотрим, как стадия де Ситтера помогает решить проблему плоскостности. Для этого рассмотрим, как изменяются кинетическая и потенциальная энергии Вселенной при расширении. В пустом мире с положительной космологической постоянной есть только силы отталкивания вида

$$F_A = H^2 a.$$

Напомним, что  $a$  имеет размерность длины. Поэтому потенциальная энергия такого мира, приходящаяся на единицу массы, есть

$$U = 1/2 H^2 a^2.$$

Кинетическая энергия единицы массы дается обычным выражением

$$E = 1/2 (da/dt)^2.$$

Обе величины  $U$  и  $E$  в мире де Ситтера экспоненциально растут:

$$E \sim \text{ch}^2 Ht, \quad U \sim \text{sh}^2 Ht,$$

но так, что их разность все время остается постоянной:

$$E - U = k/2.$$

В мире, прожившем достаточно долго, обе энергии могут быть значительно больше своей разности, оставаясь примерно равными друг другу:

$$E \approx U \gg |k/2|.$$

Именно в этих простых рассуждениях лежит решение проблемы близости к единице  $\Omega = \rho/\rho_c$ .

Параметр  $\Omega$  определяется как отношение плотности мира к критической плотности. Можно записать его в эквивалентной форме как отношение потенциальной энергии к кинетической. Теперь видно, что разность  $1 - \Omega$  представляет собой отношение постоянной разности двух величин  $(E - U)$  к величине  $E$ , сильно растущей с течением времени. Поэтому, какой бы ни была разность  $1 - \Omega$  в начале деситтеровской стадии, она будет

близка к нулю, если продолжительность этой стадии  $t_0$  достаточно велика. Как уже отмечалось, для решения известных космологических проблем необходимо  $Ht_0 > 70$ .

Это значит, что, несмотря на произвольное значение  $\Omega$  в начале деситтеровской стадии, в конце этой стадии к моменту перехода на фридмановский режим расширения мы будем иметь параметр  $\Omega$ , совпадающий с единицей с точностью  $\pm 10^{-60}$ . Именно таким способом решается вопрос близости ранней Вселенной к плоской\*.

Напишем теперь закон эволюции обычной материи с уравнением состояния  $p=0$  или  $p=\varepsilon/3$  в мире де Ситтера. В первом случае плотность убывает, как  $a^{-3} \sim a^{-3Ht}$ , во втором, как  $a^{-4} \sim a^{-4Ht}$ . После роста масштабов в  $10^{30}$  раз плотность обычного вещества падает в  $10^{120}$  раз или  $10^{90}$  раз соответственно. Другими словами, плотность обычного вещества быстро уменьшается и скоро становится динамически незначимой. Поэтому рассматривать свойства раздувающегося мира с  $a(t) \sim e^{Ht}$  при  $t > H^{-1}$  можно без учета обычного вещества.

Заметим еще, что решения де Ситтера с  $k=+1$  и  $k=0$  обладают замечательным свойством. Они не сингулярны, т. е. масштабный фактор  $a(t)$  не обращается в ноль\*\*. В этом состоит важное отличие от решения Фридмана, для которого  $a(t) \rightarrow 0$  при  $t \rightarrow 0$ . Эту точку называют начальной сингулярностью, в ней  $\rho \rightarrow \infty$  и  $T \rightarrow \infty$ . Реальная космология, однако не является деситтеровской. Точное решение де Ситтера с  $\Lambda \neq 0$  а  $T_{\mu\nu} = 0$  является устойчивым и никогда не переходит во фридмановское решение, которое, как мы знаем, справедливо, по крайней мере начиная с  $t \approx 1$  с. Рассматриваемые в инфляционных моделях деситтеровские режимы являются приближенными. С течением времени вакуумоподобная энергия скалярного поля переходит в энергию частиц и режим расширения сменяется на фридмановский.

С другой стороны, если  $\Lambda$  хоть сколько-нибудь отличается от нуля и положительно, то в обычном случае при достаточно больших  $t$  должен произойти переход от фридмановского режима к деситтеровскому. В этом случае, однако, хотя в целом Вселенная и будет расширяться экспоненциально, но связанные системы, такие, как, например, галактики или даже скопления, не будут разваливаться, так как силы гравитационного притяжения между ними превосходят «антигравитацию» вызываемую  $\Lambda$ -членом.

\* В видимой части Вселенной отличие  $\Omega$  от единицы может (должно) быть существенно больше из-за флуктуаций плотности и достигать величины  $\sim 10^{-4}$ .

\*\* Теоремы о неизбежности космологической сингулярности основывались на условии энергодоминантности, т. е. на неравенстве  $\varepsilon + 3p > 0$ , которое здесь явно не справедливо.

Возвращаясь к вопросу о сингулярности, мы хотели бы отметить, что принцип поиска несингулярных решений в современной теории заведомо должен быть отброшен. Возможно, как раз сингулярные решения дают адекватное приближение к действительности. В этой связи можно вспомнить об ударных волнах — сингулярных решениях в гидродинамике идеальной жидкости. Видимо, и в космологии мы должны искать не несингулярные решения, а физические механизмы отклонения от «идеальности».

### § 3. ЗАМЕЧАНИЯ О СООТНОШЕНИИ $p = -\varepsilon$

Обсудим, какими свойствами обладала бы среда, если бы в ней реализовывалось соотношение  $p = -\varepsilon$ . Пусть уравнение состояния записано в неявной форме:

$$p = p(n), \quad \varepsilon = \varepsilon(n),$$

где  $n = 1/V$  — плотность числа частиц в единице объема. С помощью известного термодинамического соотношения

$$dE = -pdV, \quad (5.16)$$

учитывая, что  $E = \varepsilon/n$ , нетрудно получить

$$\frac{d\varepsilon}{dn^{-1}} = -(\varepsilon + p)/n.$$

Отсюда, в частности, следует, что  $\varepsilon \sim n^{4/3}$  при  $p = \varepsilon/3$  и  $\varepsilon \sim n$  при  $p = 0$ .

Случай  $p = -\varepsilon$  является, очевидно, вырожденным. В этом случае ни  $\varepsilon$ , ни  $p$  от  $n$  не зависят. В системе нет никаких параметров, меняя которые можно изменить значение плотности энергии. Вместо целого набора значений  $p$  и  $\varepsilon$ , связанных каким-то функциональным соотношением  $p = p(\varepsilon)$ , мы получили одну-единственную точку  $p = -\varepsilon = \text{const}$ .

В случае безграничной среды такого рода градиент давления отсутствует и она была бы стационарной, если пренебречь гравитацией. При наличии границ для стабилизации среды необходима оболочка, компенсирующая давление.

Учет гравитационного взаимодействия приводит к тому, что при  $\varepsilon > 0$  среда антигравитирует, как уже отмечалось выше. Заметим, однако, что силы гравитационного отталкивания возникают только внутри среды с «уравнением состояния»  $p = -\varepsilon$ . Если в пустом пространстве с  $p = \varepsilon = 0$  находится ограниченный кусок вещества с отрицательным давлением, то по отношению к внешним объектам он будет вести себя, как обычное гравитирующее тело. Рассмотрим сначала более простой стационарный случай. Чтобы тело находилось в равновесии, должны существовать какие-то внешние силы, удерживающие его в этом

состоянии. При вычислении гравитационного потенциала, создаваемого этим объектом во внешнем пространстве, необходимо учитывать не только  $T_{\mu\nu}$ , определяемое выражением (5.3), но и добавки к нему от дополнительных сил, удерживающих рассматриваемый кусок вещества в равновесии. Нетрудно показать, что для находящегося в стационарном состоянии объекта интеграл по объему пространственных или смешанных компонент полного тензора энергии импульса должен обратиться в ноль. Поэтому гравитационное поле определяется, как и следовало ожидать, полной массой системы

$$m = \int T_{00}^{\text{tot}} dV > 0.$$

Таким образом, независимо от уравнения состояния любой стационарный объект будет притягивать внешнее пробное тело. Существенным обстоятельством является, впрочем, не стационарность, а ограниченность тела. При этом ненулевой вклад в тензор энергии-импульса дают пространственные и временные производные, такие, как  $(\partial\phi/\partial t)^2$ ,  $(\partial\phi/\partial r)^2$ , учет которых приводит к положительной определенности полной массы.

Заметим еще для примера, что отрицательное давление можно получить в твердом теле при его растяжении внешними силами, хотя, конечно, в реальной ситуации не столь огромное, как мы здесь обсуждаем, а, очевидно,  $p \ll \varepsilon$ .

Отметим еще раз, что при  $p = -\varepsilon$  плотность энергии остается постоянной при изменении объема системы. Докажем это на сей раз исходя из термодинамического соотношения (5.16). Вспомним, что  $E = V\varepsilon$  по определению. Продифференцировав обе части этого равенства, получим

$$dE = \varepsilon dV + Vd\varepsilon. \quad (5.17)$$

Приравняем правые части равенства (5.16) и (5.18) друг другу:

$$Vd\varepsilon = -(\varepsilon + p)dV.$$

Теперь видно, что только в случае  $p = -\varepsilon$  изменение плотности энергии равно нулю ( $d\varepsilon = 0$ ) при изменяющемся объеме ( $dV \neq 0$ ). Поэтому в расширяющейся Вселенной с уравнением состояния плотность энергии остается постоянной:  $\varepsilon = \text{const}$ .

#### § 4. ПРОБЛЕМА $\Lambda$ -ЧЛЕНА В ПОЗДНЕЙ ВСЕЛЕННОЙ

Как мы уже отмечали, факт расширения Вселенной избавил от необходимости вводить космологическую постоянную в уравнения тяготения. Более того, астрономические наблюдения показывают, что космологическая постоянная должна быть достаточно мала. Ограничение, записанное в терминах соответствующей ей плотности энергии вакуума, имеет вид

$$\rho_{\text{вак}} < 10^{-29} \text{ г/см}^3. \quad (5.18)$$

С точки зрения космологии это ограничение не столь сильное — верхняя граница по порядку величины совпадает с критической плотностью энергии во Вселенной. Однако с точки зрения физики элементарных частиц  $\rho_{\text{вак}}$  чудовищно мала. Для иллюстрации этого (5.18) удобно переписать в других единицах:

$$\rho_{\text{вак}} < 10^{-47} (\text{ГэВ})^4. \quad (5.18')$$

Казалось бы, какое отношение имеет космологическая постоянная к элементарным частицам? Ответ состоит в том, что в квантовой теории энергия наимизшего состояния — вакуума — отлична от нуля и составляет величину  $\sim m^4$ , где  $m$  — какая-то характерная для физики частиц масса. Численная величина  $m$  зависит от модели, но все же в любой модели  $m \geq 1$  ГэВ. Можно, конечно, думать, что происходит случайная компенсация различных вкладов в  $\rho_{\text{вак}}$ , но когда речь идет о компенсации с точностью по крайней мере 50 порядков, на случайность надеяться трудно.

Появление  $\Lambda$ -члена напоминает появление массы и первоначально безмассовых частиц из-за квантовых поправок — перенормировок. Калибровочная инвариантность, как отмечалось выше, требует исчезновения массы фотона не только на классическом уровне, но и на квантовом уровне, по крайней мере при достаточно малой константе связи. Однако для космологической постоянной аналогичного принципа пока не найдено. Опыт работы с квантовой теорией поля учит, что в ней «все, что не запрещено, разрешено». Космологическая постоянная выглядит на нашем сегодняшнем уровне понимания как исключение из этого правила.

Чтобы пояснить, как квантовая теория приводит к ненулевой энергии вакуума, вспомним, что наимизшая энергия квантово-механического осциллятора не равна нулю, а равна  $\omega/2$ , что непосредственно связано с принципом неопределенности Гайзенберга. В этом состоит важное отличие квантовой механики от классической. Согласно последней минимальная энергия частицы в осцилляторном потенциале  $U = m\omega^2 x^2/2$  равна нулю.

Если перейти от квантовой механики одной частицы к твердому телу, т. е. к квантовой механике большого числа частиц, то в пределе малых колебаний энергии этих частиц (электронов) можно представить в виде

$$U = \frac{1}{2} \sum m_j^2 x_j^2 + \frac{1}{2} k \sum (x_j - x_{j+1})^2, \quad (5.19)$$

где первый член отвечает взаимодействию электрона с ионом кристаллической решетки, а второй — взаимодействию сосед-



них электронов между собой. Для решения такой задачи известным удобным приемом является диагонализация матрицы взаимодействия, т. е. переход к нормальным (невзаимодействующим) модам колебаний, который осуществляется преобразованием Фурье. Основное состояние твердого тела, очевидно, имеет энергию, равную полусумме частот нормальных мод. Таким образом, даже при нулевой температуре энергия осцилляторов, составляющих твердое тело, отлична от нуля. Об этой энергии говорят, как об энергии нулевых колебаний. Заметим, что энергия нулевых колебаний прекрасно проявляется в эксперименте. В частности, для изотопов  $\text{Li}^6$  и  $\text{Li}^7$  энергия взаимодействия атомов одинакова, а энергия нулевых колебаний различна, что приводит к разной температуре испарения. Этот эффект может использоваться для разделения изотопов.

Квантовая теория поля строится аналогично теории твердого тела. В классической теории состояние поля в каждой точке пространства определяется функцией  $\varphi(x, t)$  для скалярного поля, или четырьмя функциями  $A_\mu$  для векторного и т. п. Переход к квантовой теории осуществляется, как и в случае квантовой механики одной частицы, заменой числовых функций на операторные. В этом смысле квантовая теория поля — это квантовая механика системы с бесконечным числом степеней свободы. Гамильтониан свободного скалярного поля, как мы знаем, можно записать в виде

$$H = 1/2 \int d^3x [\pi^2(x, t) + (\nabla\varphi(x, t))^2 + m^2\varphi^2(x, t)], \quad (5.20)$$

что весьма близко по форме к выражению (5.19) для энергии электронов в твердом теле. Можно, как и выше, перейти к нормальным координатам, сделав преобразование Фурье:

$$\varphi(x, t) = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} [a(k) e^{ikx - i\omega_k t} + a^+(k) e^{-ikx + i\omega_k t}],$$

где  $\omega_k = \sqrt{k^2 + m^2}$ , а величины  $a^+(k)$  и  $a(k)$  представляют собой операторы рождения и уничтожения частиц с импульсом  $k$ . В терминах этих величин — нормальных координат — гамильтониан будет диагональным:

$$H = 1/2 \int d^3k \omega_k [a^+(k) a(k) + a(k) a^+(k)].$$

Он описывает совокупность невзаимодействующих осцилляторов с частотами  $\omega_k$ . Низшее по энергии состояние — вакуум такой системы — имеет ненулевую (бесконечную) плотность энергии

$$\epsilon_0 = \frac{1}{2} \cdot \frac{4\pi}{(2\pi)^3} \int k^2 \omega_k dk.$$

Ее можно представить как энергию рождающихся и исчезающих в вакууме частиц, что можно изобразить диаграммой (рис. 21).

Немного более наглядным ответ будет выглядеть, если перейти от интегрирования к суммированию по  $k$ , сделав обычную замену

$$\int d^3k = \sum \Delta V_k, \quad \delta^3(k - k') \rightarrow \frac{\delta k k'}{\Delta V_k}, \quad a_k = \sqrt{\Delta V_k} a(k).$$

Тогда операторы  $a_k$  и  $a_{k'}^+$  коммутируют согласно

$$[a_k, a_{k'}^+] = \delta_{kk'}.$$

Учитывая это, можно переписать гамильтониан (5.20) в виде

$$H_b = \sum_k \omega_k \left( a_k^+ a_k + \frac{1}{2} \right). \quad (5.21)$$

Вспоминая, что вакуум определяется как состояние без частиц, т. е. условием  $a_k|\text{вак}\rangle = 0$ , получим, что вакуумная энергия равна  $\langle \text{вак} | H_b | \text{вак} \rangle = \omega_k/2$ . Обычно об этой бесконечной энергии попросту забывали, объявляя ее ненаблюдаемой и отсчитывая энергию засты от этого бесконечно высокого уровня. Такую точку зрения можно последовательно провести, если не учитывать гравитационного взаимодействия. Гравитация чувствительна

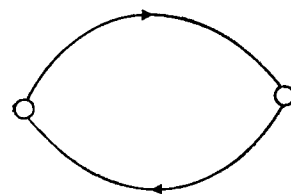


Рис. 21. Виртуальные частицы, рождающиеся и исчезающие в вакууме. Их существование приводит к изменению энергии вакуума

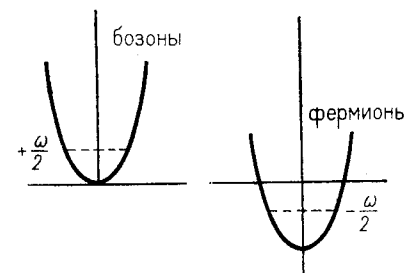


Рис. 22. Вакуумные энергии бозонов и фермионов. Энергия основного состояния фермионов отличается знаком от энергии основного состояния бозонов при равных массах

ко всем формам энергии, в том числе и к вакуумной. Тензор энергии-импульса последней имеет вид, как мы уже отмечали,  $T_{\mu\nu} = \rho_{\text{вак}} g_{\mu\nu}$ , и, таким образом, еще раз подчеркнем, вакуумная энергия эквивалентна космологической постоянной.

Выражение (5.21) справедливо для бозонов, что отражено

индексом  $b$ . Если обратиться к фермионам, то нужно вспомнить, что для них операторы рождения и уничтожения удовлетворяют соотношению антикоммутиации. Это приведет к изменению знака последнего слагаемого:

$$H_f = \sum_k \omega_k \left( b_k^+ b_k - \frac{1}{2} \right). \quad (5.22)$$

Это изменение знака можно интерпретировать как сдвиг фермионного потенциала нормальных колебаний на величину  $(-\omega_k)$  по сравнению с бозонным для той же массы (на рис. 22) схематически изображены потенциальные энергии бозонных и фермионных осцилляторов и основные уровни)\*. В настоящее время мы думаем, что существует глубокая симметрия между фермионами и бозонами (называемая суперсимметрией), так что каждому бозонному состоянию отвечает в точности одно фермионное. Однако опыт показывает, что эта симметрия неточна, так как массы бозонов и фермионов различны. Поэтому точного сокращения вакуумной энергии не происходит, даже при точном соответствии числа фермионных и бозонных состояний. Вклады, расходящиеся, как четвертая и даже как вторая степень верхнего предела интегрирования, сокращаются, а остаток пропорционален разности масс в четвертой степени, т. е. в конечном счете величине  $(m_{\text{SUSY}})^4$ , где  $m_{\text{SUSY}}$  — массовый параметр, определяющий масштаб нарушения суперсимметрии. Согласно современным данным  $m_{\text{SUSY}} \geq 10^3$  ГэВ, что даст  $\rho_{\text{вак}} \geq 10^{12}$  ГэВ<sup>4</sup> в «некотором» (!) противоречии с (5.18).

Кроме этого, при расширении и охлаждении Вселенной в первичной плазме, как учит нас теория, происходит целый ряд фазовых переходов, как, например, переход от кварк-глюонной плазмы к адронной, переход от симметричной электро-слабой фазы к несимметричной, когда  $m_{W,Z} \neq 0$ , и др. При каждом из этих фазовых переходов тензор энергии-импульса системы изменяется на величину  $\Delta\rho g_{\mu\nu}$ , т. е. по форме выглядит как скачок в плотности энергии вакуума.

Поясним это немножко подробнее. Пусть система описывается потенциалом, изображенным на рис. 23. И пусть в начальный момент волновая функция локализована вблизи  $\varphi=0$ . Если вероятность туннелирования сквозь потенциальный барьер мала, то в хорошем приближении точку  $\varphi=0$  можно рассматривать как вакуум такой модели. По очевидной причине он называется ложным вакуумом. Можно рассматривать возбужденные состояния (частицы) над этим вакуумом и построить обычную полевую модель, не учитывая истинного вакуума при  $\varphi=\varphi_0$ . При малой вероятности перехода к  $\varphi=\varphi_0$  можно долго

\* На гравитационную роль моря Дирака указывал еще Г. Гамов в письме к А. Ф. Иоффе в 1930 г.

находиться в этом мире над ложным вакуумом, не подозревая, что живешь на «пороховой бочке». Однако если вероятность туннелирования отлична от нуля, то рано или поздно произойдет взрыв ложного вакуума с выделением энергии с плотностью  $\Delta\rho = U(0) - U(\varphi_0)$  и система перейдет к состоянию  $\varphi = \varphi_0$ . Величина  $\Delta\rho$  при кварк-адронном фазовом переходе равна  $\sim 10^{-4}$  ГэВ<sup>4</sup>, при электрослабом  $\sim 10^8$  ГэВ<sup>4</sup>, а при аналогичном переходе на масштабе большого объединения  $\sim 10^{60}$  ГэВ<sup>4</sup> (об этих переходах будет идти речь в гл. 6). Если у людей осторожных еще могут быть какие-то сомнения относительно последних фазовых переходов, то в кварк-адронном переходе можно быть абсолютно уверенным. В частности, существование глюонного конденсата, обладающего тензором энергии-импульса  $T_{\mu\nu} \sim g_{\mu\nu} (0,1 \text{ ГэВ})^4$ , доказано фактически экспериментально путем сравнения, вычисляемых в квантовой хромодинамике характеристик адронов с наблюдаемыми на опыте.

Невозможно поверить, что в «начале» затравочная космологическая постоянная была подобрана так, чтобы полностью скомпенсировать все эти скачки в  $\rho_{\text{вак}}$  при последующих фазовых переходах, причем с фантастической степенью точности. Должна существовать какая-то динамическая модель, автоматически обеспечивающая необходимую компенсацию. Возможно, при этом компенсация будет происходить не мгновенно, а с каким-то запаздыванием, причем темп этой компенсации будет определяться теми же характерными гравитационными масштабами, что и, скажем, расширение Вселенной. По этой причине возможны некоторые отклонения от стандартного фридмановского сценария на всех стадиях (в том числе и сегодня) эволюции Вселенной. Нам трудно, однако, делать на этот счет сколько-нибудь детальные предсказания, так как ни одной удовлетворительной во всех отношениях модели компенсации  $\Lambda$ -члена в литературе не известно. В настоящее время эта проблема является центральной как в физике элементарных частиц, так и в космологии, представляя собой серьезный вызов теоретикам.

В заключение процитируем шутливое высказывание Я. И. Померанчука, что вся физика — это физика вакуума:

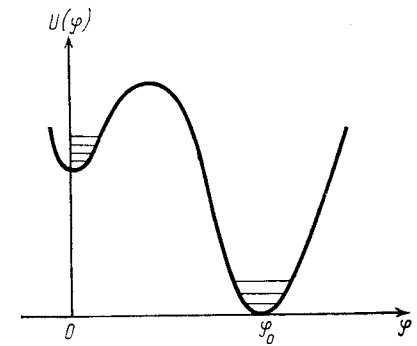


Рис. 23. Эффективный потенциал системы, обладающей метастабильным (или ложным) вакуумом при  $\varphi=0$  и истинным вакуумом при  $\varphi=\varphi_0$

часть I «Насосы и манометры» и ч. II «Квантовая теория вакуума».

## Глава 6.

### СКАЛЯРНЫЕ ПОЛЯ В КОСМОЛОГИИ

Скалярное поле было введено в физику Юкавой в 1935 г. Кванты этого поля —  $\pi$ -мезоны — были вскоре открыты, однако в 60-е годы выяснилось, что  $\pi$ -мезоны не фундаментальны, а состоят из кварков. В итоге скалярное поле было изгнано из избранного круга фундаментальных полей, но продолжалось это недолго. Калибровочные теории со спонтанным нарушением симметрии потребовали введения скалярных — хиггсовских полей, которые обеспечили бы перенормируемость теории.

Дополнительный аргумент в пользу существования скалярных полей дает суперсимметрия. В последнее время с ней связывают надежды на построение единой теории взаимодействий элементарных частиц. Суперсимметрия связывает между собой поля с разными спинами и таким образом представляет собой еще один шаг на пути унификации частиц и их взаимодействий. Частицы с разными спинами рассматриваются как разные состояния фундаментального поля. Из факта существования полей со спином  $1/2$ ,  $1$  и  $2$  суперсимметрия делает вывод о существовании полей с спином  $0$  и  $3/2$ . Таким образом, мы снова приходим к скалярному полю.

Интерес к скалярным полям в космологии вызван тем обстоятельством, что с их помощью можно естественным образом получить столь интересующее нас соотношение  $p = -\varepsilon$ . Ниже мы на различных примерах разберем, как это достигается.

#### § 1. СКАЛЯРНОЕ ПОЛЕ В ПЛОСКОМ ПРОСТРАНСТВЕ-ВРЕМЕНИ

Лагранжиан комплексного скалярного поля  $\varphi$  можно записать в виде

$$\mathcal{L} = |\partial_\mu \varphi|^2 - U(\varphi), \quad (6.1)$$

где потенциал  $U(\varphi)$  для свободных невзаимодействующих полей имеет вид

$$U_0(\varphi) = m^2 |\varphi|^2. \quad (6.2)$$

Отсутствие множителя  $1/2$  в этом лагранжиане по сравнению с лагранжианом (3.6) связано с тем, что последний был записан для действительного поля. Вводя действительные поля

$\varphi_1$  и  $\varphi_2$ :  $\varphi = (\varphi_1 + i\varphi_2)/\sqrt{2}$  мы перейдем от лагранжиана (6.1) к сумме двух лагранжианов (3.6). На рис. 24 мы изобразили вид потенциала  $U(\varphi) = m^2 |\varphi|^2$  (точнее, его «продольное» сечение, из которого полный потенциал в плоскости  $\varphi_1\varphi_2$  получается вращением вокруг оси  $U$ ). Заметим, что лагранжиан (6.1) с потенциалом (6.2) инвариантен относительно фазовых преобразований

$$\varphi \rightarrow \varphi e^{i\alpha}. \quad (6.3)$$

Потенциал  $U(\varphi)$ , отвечающий хиггсовским частицам (рис. 17), можно записать в виде

$$U_H(\varphi) = 1/4\lambda(\varphi_0^2 - |\varphi|^2)^2. \quad (6.4)$$

Этот потенциал также инвариантен относительно преобразований (6.3), как и любой потенциал  $U = U(|\varphi|)$ .

Как мы уже отмечали, точка  $\varphi = 0$  не является точкой устойчивого равновесия потенциала (6.4). Устойчивой является линия

$$\varphi = \varphi_0 e^{i\sigma_0} \quad (6.5)$$

с произвольной фазой  $\sigma_0$ . Разлагая лагранжиан вблизи этого положения устойчивого равновесия, т. е. полагая  $\varphi = (\varphi_0 + \chi) e^{i\sigma}$ , получим

$$\mathcal{L} = (\partial_\mu \chi)^2 + (\varphi_0 + \chi)^2 (\partial_\mu \sigma)^2 - \lambda(\varphi_0^2 \chi^2 + \varphi_0 \chi^3 + \chi^4/4). \quad (6.6)$$

Этот лагранжиан описывает два типа полей: массивное поле  $\chi$  с массой  $m^2 = \lambda\varphi_0^2$  и самодействием  $1/4\lambda(\chi^4 + 4\varphi_0\chi^3)$  и безмассовое поле  $\sigma$ . Инвариантность относительно преобразований (6.3) теперь отсутствует. Произошло спонтанное нарушение симметрии. Появление при этом безмассового поля  $\sigma$  является следствием общей теоремы Голдстоуна. Заметим, однако, что если бы мы добавили в лагранжиан векторные калибровочные поля, взаимодействующие с  $\varphi$ , то безмассовое поле не возникло бы, а соответствующая степень свободы превратилась бы в продольную компоненту векторного поля, как это и происходит в теории электрослабого взаимодействия (§ 8 гл. 4).

Специальный интерес для нас представляет выражение для тензора энергии-импульса поля  $\varphi$ , так как последний является источником гравитационного поля. Рассмотрим простейший случай свободного действительного поля, описываемого лагранжианом (3.6). Тензор энергии-импульса определяется как вариационная производная действия по метрике

$$\begin{aligned} T_{\mu\nu} &= \frac{\delta}{\delta g^{\mu\nu}} \int d^4x \sqrt{g} [g^{\alpha\beta} \partial_\alpha \varphi \partial_\beta \varphi - U(\varphi)] = \\ &= \partial_\mu \varphi \partial_\nu \varphi - \frac{1}{2} g_{\mu\nu} [g^{\alpha\beta} \partial_\alpha \varphi \partial_\beta \varphi - U(\varphi)]. \end{aligned} \quad (6.7)$$

В частном случае однородного поля  $\varphi$ , зависящего только от времени,  $T_{\mu\nu}$  сводится к

$$T_{00} = \varepsilon = 1/2(\dot{\varphi}^2 + m^2\varphi^2), \quad (6.8)$$

$$T_{ij} = p\delta_{ij} = 1/2\delta_{ij}(\dot{\varphi}^2 - m^2\varphi^2). \quad (6.9)$$

Отсюда видно, что при специальном выборе состояния однородного поля  $\varphi = \varphi(t)$ , таком, что  $\dot{\varphi} = 0$ , а  $\varphi \neq 0$ , возникает как раз чрезвычайно важное для космологии соотношение  $p = -\varepsilon$ .

Как мы увидим ниже, в случае свободного скалярного поля с потенциалом (6.2) это равенство может быть справедливым только в определенные дискретные моменты времени, так как в силу уравнений движения состояние с  $\dot{\varphi} = 0$  и  $\varphi \neq 0$  отвечает  $\ddot{\varphi} \neq 0$ . Если, однако, изменение  $\varphi$  медленное (т. е.  $\dot{\varphi}$  — достаточно мала), то соотношение  $p = -\varepsilon$  может приближенно выполняться в течение достаточно длительного времени. Легко видеть, что в случае хиггсовского поля условие  $\dot{\varphi} = 0$  и  $U(\varphi) \neq 0$  может быть справедливо при всех  $t$ .

Проиллюстрируем еще раз лоренц-инвариантность соотношения  $p = -\varepsilon$  (ср. § 1 гл. 5). Прежде всего отметим тот факт, что равенство нулю всех производных поля  $\varphi$  — это инвариантное равенство при координатных, в частности лоренцевых, преобразованиях. Действительно, поскольку  $\varphi$  — скалярное поле, то при лоренц-преобразованиях  $\varphi$  не изменяется. Градиент скалярного поля — четырехмерный вектор. При четырехмерных преобразованиях компоненты вектора преобразуются и величина каждой компоненты меняется, за исключением одного случая, когда все компоненты равны нулю. Это единственный случай лоренц-инвариантного вектора. Из инвариантности  $\varphi$  и его производных (равных нулю) относительно лоренц-преобразований немедленно следует, что соотношение  $p = -\varepsilon$  является лоренц-инвариантным.

Рассмотрим теперь физический смысл выбранных начальных условий  $\dot{\varphi} = 0$  и  $\varphi \neq 0$ . Уравнение движения для однородного поля  $\varphi$  является дифференциальным уравнением второго порядка:

$$\frac{d^2\varphi}{dt^2} = -\frac{\partial U(\varphi)}{\partial\varphi}; \quad (6.10)$$

оно допускает произвольное задание двух начальных условий, которые в нашем случае имеют вид  $\dot{\varphi} = 0$  и  $\varphi = \varphi_1 \neq 0$ .

Уравнение (6.10) эквивалентно уравнениям движения материальной точки в классической механике. При этом роль координаты точки играет величина скалярного поля  $\varphi(t)$ , а  $U(\varphi)$  —

силовой потенциал, в котором движется точка. Потенциал вида  $U = 1/2m^2\varphi^2$  описывает гармонический осциллятор с частотой  $m$ . Выбор условий  $\dot{\varphi} = 0$  и  $\varphi = \varphi_1$  в начальный момент времени соответствует в осцилляторе тому, что мы «отткнули» грузик на пружинке из положения равновесия на расстояние  $\varphi_1$  и зафиксировали его. Вернемся опять к графику, изображающему потенциал в зависимости от  $\varphi$  (рис. 24). Состояние осциллятора обозначено точкой  $A$ . Легко видеть, что состояние неустойчиво. Как только внешние силы перестают удерживать осциллятор в  $A$ , так сразу же точка  $A$  «покатится» вниз по пара-

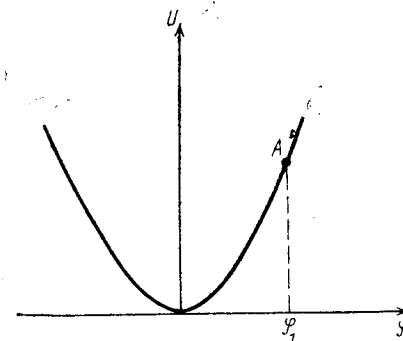


Рис. 24. Потенциал  $U(\varphi)$  свободного скалярного поля

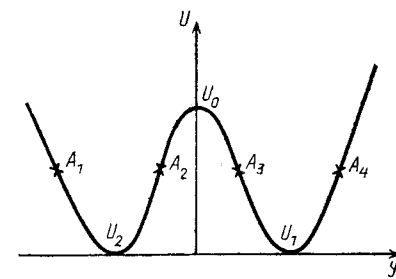


Рис. 25. Потенциал  $U(\varphi)$  в модели Хиггса

боле к положению равновесия. Движение точки, или зависимость  $\varphi(t)$  от времени, будет описываться обычным гармоническим законом вида

$$\varphi(t) = \varphi_0 \cos mt.$$

Отметим, что выбор начальных условий фиксирует фазу колебаний. Набор двух условий  $\dot{\varphi} = 0$  и  $\varphi = \varphi_0$  будет повторяться через период, а через полпериода  $\varphi(t)$  будет удовлетворять условию  $\dot{\varphi} = 0$ , т. е. тоже соотношению  $p = -\varepsilon$ .

Потенциал  $U(\varphi)$  более сложной формы, отвечающий хиггсовскому случаю (рис. 25), имеет два устойчивых минимума  $U_1$  и  $U_2$  и один неустойчивый  $U_0$ . Подобно предыдущему случаю мы можем выбрать неустойчивое начальное состояние в одной из точек  $A_1, A_2, A_3, A_4$ . Тогда ситуация будет аналогична разобранной выше. Выберем теперь в качестве начального условия  $\dot{\varphi} = 0$ , т. е. состояние неустойчивого равновесия. В таком состоянии поле не может оставаться неограниченно долго, так как квантовые флуктуации выведут его из положения равновесия. Кроме того, состояние  $\dot{\varphi} = 0$  на рис. 25 соответствует согласно (6.4) ненулевому значению потенциала

$$U(0) = \frac{\lambda}{4} \varphi_0^4, \quad (6.11)$$

а  $\varphi=0$ . Это значит, что состоянию поля в положении  $U_0$  отвечает соотношение

$$p = -\varepsilon.$$

Отметим еще одно любопытное свойство скалярного поля. В момент, когда  $\dot{\varphi}=0$ ,  $\varphi=0$ , что соответствует быстрому прохождению поля через точку  $\varphi=0$  на рис. 24, реализуется предельно жесткое уравнение состояния

$$p = +\varepsilon. \quad (6.12)$$

Для среды с таким уравнением состояния скорость звука принимает максимально возможное значение, равное скорости света. Это уравнение было предложено около 20 лет тому назад в теории векторных полей, однако, как мы только что видели, может быть реализовано для скалярного поля.

Заметим, что термин «уравнение состояния» мы здесь употребляем в несколько условном смысле, так как система не релаксировала и задание  $\varepsilon$  не определяет  $p$ . В рассматриваемом случае соотношение между  $p$  и  $\varepsilon$  изменяется со временем, причем у нас нет динамического уравнения, которое позволяло бы по заданным  $p$  и  $\varepsilon$  в начальный момент определить всю их дальнейшую эволюцию. Необходимо еще знание уравнения движения поля  $\varphi$ . В свободном случае скалярное поле массы  $m$  удовлетворяет, как мы знаем, уравнению Клейна—Гордона:

$$(\square + m^2)\varphi \equiv (\partial_t^2 - \Delta + m^2)\varphi = 0. \quad (6.13)$$

Это уравнение обладает решениями в виде плоских волн:

$$\varphi = \varphi_0 \exp(-i\omega t + ikr), \quad (6.14)$$

где

$$\omega^2 = k^2 + m^2. \quad (6.15)$$

В пространственно однородном случае (т. е. для нулевого импульса  $k$ ) решение принимает вид

$$\varphi(t) = \varphi_0 e^{\pm i(mt + \psi)}, \text{ или } \varphi(t) = \varphi_0 \cos(mt + \psi). \quad (6.16)$$

Мы видим, что для массивных частиц даже в состоянии покоя частота осцилляций волновой функции отлична от нуля. Это связано с тем, что энергия покоящейся частицы равна не нулю, а ее массе  $m$  (или  $mc^2$ , если не полагать  $c=1$ ). Вычислим плотность энергии и давление, отвечающие решению (6.16):

$$\varepsilon = \frac{1}{2} \dot{\varphi}^2 + \frac{1}{2} m^2 \varphi^2 = \frac{1}{2} m^2 \varphi_0^2 [\cos^2(mt + \psi) + \sin^2(mt + \psi)] =$$

$$= \frac{1}{2} m^2 \varphi_0^2, \quad (6.17)$$

$$p = \frac{1}{2} \dot{\varphi}^2 - \frac{1}{2} m^2 \varphi^2 = \frac{1}{2} m^2 \varphi_0^2 [\cos^2(mt + \psi) - \sin^2(mt + \psi)] = \\ = -\frac{1}{2} m^2 \varphi_0^2 \cos(2mt + 2\psi).$$

Плотность энергии, как и следовало ожидать, не зависит от времени, а давление оказалось отличным от нуля, несмотря на то, что частицы покоятся. Впрочем, усредненное по времени давление обращается в нуль. Если поле  $\varphi$  представляет собой суперпозицию состояний (6.16) со случайными фазами  $\psi$ , т. е. набор независимых частиц, то в результате усреднения по фазам давление обратится в нуль. Мы получаем в итоге естественный результат, что некогерентная суперпозиция покоящихся частиц имеет нулевое давление и, таким образом, справедливо пылеподобное уравнение состояния  $p=0$ .

Поле  $\varphi$  может представлять собой когерентную суперпозицию квантов. В этом случае можно говорить о классическом поле, давление которого отлично от нуля и осциллирует со временем, хотя, повторим, в среднем по времени  $\bar{p}=0$ . Заметим, что это утверждение справедливо лишь для свободного поля, когда  $U(\varphi) = m^2 \varphi^2 / 2$ . Для более сложных потенциалов  $U(\varphi)$  среднее давление не исчезает.

Если существенный интервал времени достаточно мал,  $\tau \ll m^{-1}$ , то для когерентного состояния поля  $\varphi$  соотношение между  $p$  и  $\varepsilon$  с течением времени может меняться от предельно жесткого,  $p = +\varepsilon$ , до вакуумного,  $p = -\varepsilon$ . Это удобное упрощение более сложной картины состояния вещества. Дело в том, что в рассматриваемом случае когерентного скалярного поля вообще не существует уравнения состояния, позволяющего выразить  $p(\varphi, \dot{\varphi})$  через  $\varepsilon(\varphi, \dot{\varphi})$ . Тем не менее при достаточно медленной эволюции поля имеет смысл говорить об уравнении состояния. То, что эффективное уравнение состояния меняется от  $p = +\varepsilon$  до  $p = -\varepsilon$ , наводит на мысль, что поле  $\varphi$  может вызвать экспоненциальное расширение Вселенной в течение того времени, пока справедливо уравнение состояния  $p = -\varepsilon$ . Чтобы более аккуратно получить этот результат, нам придется рассмотреть уравнение движения поля  $\varphi$  в расширяющейся Вселенной.

Для космологических приложений весьма существен вопрос, как долго система пребывает в состоянии, характеризуемом условием  $p = -\varepsilon$  (хотя бы приближенно). Ниже мы увидим, что в нестационарной Вселенной в уравнении движения (6.10) возникает дополнительная сила трения, тормозящая скатывание  $\varphi$  вниз к положению равновесия и обеспечивающая достаточную продолжительность инфляции.

Подчеркнем, забегая вперед, что само по себе расширение Вселенной не меняет соотношения  $p = -\varepsilon$ . Но если поле  $\phi$  не постоянно, то внутренние силы меняют  $p$ , а заодно меняются и  $\varepsilon$ .

## § 2. СКАЛЯРНОЕ ПОЛЕ В РАСШИРЯЮЩЕЙСЯ ВСЕЛЕННОЙ

Выше мы рассмотрели поля материи в плоском пространстве-времени, когда гравитационное поле было несущественно. При этом можно установить многие важные качественные свойства этих полей. В космологической ситуации, однако, необходимо учитывать взаимодействие полей материи с гравитацией, или, иными словами, описывать их на фоне нестационарной метрики, отвечающей искривленному пространству-времени. Метрика, как известно, определяется именно полями материи — их тензором энергии-импульса. Мы приходим, таким образом, к необходимости решения самосогласованной задачи о совместной эволюции гравитационного и материальных полей. В качестве предварительного этапа мы рассмотрим скалярное поле  $\phi$ , эволюционирующее на фоне заданной однородной метрики. Физически такая ситуация может реализоваться, когда гравитационное поле определяется какими-то другими, помимо  $\phi$ , полями, а вклад  $\phi$  в полный тензор энергии-импульса пренебрежимо мал.

Уравнения движения поля  $\phi$  в искривленном пространстве получаются из уравнения в плоском пространстве (6.13) стандартной процедурой замены обычных производных  $\partial_\mu \equiv \partial/\partial x^\mu$  на ковариантные  $D_\mu = \partial_\mu + \Gamma_\mu$ , где величина  $\Gamma_\mu$  зависит от того объекта, на который действует  $D_\mu$ . Смысл этой замены состоит в следующем. Мы знаем, что в плоском пространстве дифференцирование в декартовых координатах является тензорной операцией. Иными словами, если  $\phi$  — скалярное поле, то  $\partial_\mu \phi$  — вектор; для векторного поля  $A_\mu$  величина  $\partial_\mu A_\nu$  является тензором второго ранга и т. п. В искривленном пространстве (или даже в плоском, но в криволинейных координатах, например в сферических) это положение уже не справедливо. Взятие производной приводит к величине, не являющейся тензором. Дело в том, что в операцию дифференцирования включается вычисление разности функций в двух близких, но различных точках. В искривленном пространстве тензоры в разных точках преобразуются по-разному, поэтому их разность не обладает тензорным законом преобразования. Для компенсации возникающего отклонения и производят упомянутое выше «удлинение» производной. Из сказанного выше очевидно, что для скалярной величины ковариантная производная совпадает с обычной. Для вектора

$$D_\mu V_\nu = \partial_\mu V_\nu - \Gamma_{\mu\nu}^\lambda V_\lambda,$$

где  $\Gamma_{\mu\nu}^\lambda = \frac{1}{2} g^{\lambda\sigma} (\partial_\nu g_{\mu\sigma} + \partial_\mu g_{\nu\sigma} - \partial_\sigma g_{\mu\nu})$  — символы Кристоффеля. При ковариантном дифференцировании тензора произвольного ранга величины  $\Gamma_{\mu\nu}^\lambda$  действуют последовательно на все индексы по указанному выше закону:

$$D_\mu T_{\nu\dots} = \partial_\mu T_{\nu\dots} - \Gamma_{\mu\nu}^\lambda T_{\lambda\dots} - \Gamma_{\mu\gamma}^\lambda T_{\nu\lambda\dots} - \dots$$

Нетрудно проверить, что  $D_\mu g_{\alpha\beta} = 0$ .

На фоне однородной нестационарной метрики (2.1) уравнение для скалярного поля принимает вид

$$\frac{\partial^2 \phi}{\partial t^2} - \frac{1}{a^2} \Delta \phi + m^2 \phi + 3H\dot{\phi} = 0, \quad (6.18)$$

где  $H = \dot{a}/a$  — параметр Хаббла. Нас будут интересовать однородные (т. е. не зависящие от пространственных координат решения), когда  $\Delta \phi = 0$  и уравнение (6.18) сводится к уравнению колебаний с переменным трением

$$\ddot{\phi} + 3H\dot{\phi} + m^2 \phi = 0, \quad (6.19)$$

так как параметр Хаббла, вообще говоря, зависит от времени.

В действительности, как мы уже отмечали, надо решать самосогласованную задачу, когда эволюция  $H(t)$  определяется самим скалярным полем  $\phi$ . Ее мы рассмотрим ниже, а сейчас проанализируем простой случай  $H = \text{const}$ , отвечающий решению де Ситтера. Физически такой случай реализуется, если полный тензор энергии-импульса материи определяется какими-то другими полями (среди них, быть может, есть еще скалярные), которые устроены так, что для них справедливо соотношение  $p = -\varepsilon$ , приводящее к деситтеровскому режиму  $H = \text{const}$ .

Решение уравнения (6.19), как и прежде, будем искать в виде

$$\phi(t) = \phi_0 e^{\lambda t},$$

где  $\lambda$  — уже комплексное число, удовлетворяющее квадратному уравнению

$$\lambda^2 + 3H\lambda + m^2 = 0, \quad (6.20)$$

корни которого равны:

$$\lambda = -\frac{3}{2} H \pm \sqrt{\frac{9}{4} H^2 - m^2}. \quad (6.21)$$

Случай  $m > 3/2H$  сводится фактически к эволюции скалярного поля на стационарном фоне. Единственное отличие от ре-

шения вида  $e^{imt}$  заключается в адиабатическом уменьшении амплитуды с декрементом  $3/2H$ .

Значительно более интересная ситуация возникает при  $m \ll H$ . Физически это сводится к тому, что фон меняется быстрее самого поля  $\varphi$ . Тогда корни (6.21) приближенно равны

$$\lambda_1 = -3H, \quad \lambda_2 = -m^2/3H. \quad (6.22)$$

В соответствии с этим возникают два решения:

$$\varphi_1(t) = \varphi_1 e^{-3Ht} \quad \text{и} \quad \varphi_2(t) = \varphi_2 e^{-m^2 t/3H}.$$

Первое из этих двух решений с течением времени падает существенно быстрее второго, так как  $3H \gg m^2/3H$ . Поэтому из фактически произвольной суперпозиции

$$\varphi(t) = \varphi_1 e^{-3Ht} + \varphi_2 e^{-m^2 t/3H} \quad (6.23)$$

с течением времени выживает только второй член. Исключение составляет лишь вырожденный случай  $\varphi_2 = 0$ .

Плотность энергии и давление, отвечающие решению (6.23) соответственно равны:

$$\begin{aligned} \varepsilon &= \frac{1}{2} \left( 3H\varphi_1 e^{-3Ht} + \frac{m^2}{3H} \varphi_2 e^{-\frac{m^2}{3H}t} \right)^2 + \frac{m^2}{2} (\varphi_1 e^{-3Ht} + \varphi_2 e^{-\frac{m^2}{3H}t})^2 \rightarrow \\ &\rightarrow \frac{m^2}{2} \varphi_2^2 e^{-\frac{2m^2}{3H}t} \left( 1 + \frac{m^2}{9H^2} \right), \end{aligned}$$

$$\begin{aligned} p &= \frac{1}{2} \left( 3H\varphi_1 e^{-3Ht} + \frac{m^2}{3H} \varphi_2 e^{-\frac{m^2}{3H}t} \right)^2 - \frac{m^2}{2} (\varphi_1 e^{-3Ht} + \varphi_2 e^{-\frac{m^2}{3H}t})^2 \rightarrow \\ &\rightarrow -\frac{m^2}{2} \varphi_2^2 e^{-\frac{2m^2}{3H}t} \left( 1 - \frac{m^2}{9H^2} \right). \end{aligned}$$

Мы видим, таким образом, что соотношение между  $p$  и  $\varepsilon$  экспоненциально быстро (как  $e^{-3Ht}$ ) стремится к  $p = -\varepsilon$  с точностью до членов порядка  $(m/3H)^2$ . Наоборот, при сжатии Вселенной, что отвечает изменению знака  $\dot{H}$ , как легко видеть, выживает только первое слагаемое в (6.23) и «уравнение состояния» стремится к предельно жесткому  $p = +\varepsilon$ .

В вырожденных случаях  $\varphi_1 = 0$  или  $\varphi_2 = 0$  мы имеем соответственно  $p \approx -\varepsilon$  или  $p \approx +\varepsilon$  при любом  $t$ . В общем же случае соотношение между  $p$  и  $\varepsilon$  изменяется со временем, асимптотически выходя при расширении на  $p = -\varepsilon$  и на  $p = +\varepsilon$  при сжатии. Эта ситуация иллюстрируется рис. 26, где изображено отношение  $p/\varepsilon$  в зависимости от времени. Процесс сжатия можно получить из расширения заменой  $t \rightarrow -t$ .

Горизонтальные линии отвечают вырожденным случаям  $\varphi_1 = 0$  или  $\varphi_2 = 0$ . Сплошные кривые отвечают режиму расшире-

ния, пунктирные — сжатия. Положение и форма кривых зависят от начальных условий, но их асимптотический предел всегда одинаков. Он определяется только эволюцией Вселенной.

Иными словами, решение, отвечающее  $p = +\varepsilon$ , при расширении неустойчиво и наоборот при сжатии неустойчиво решение, отвечающее  $p = -\varepsilon$ .

Эта ситуация очень похожа на принцип Ле-Шателье—Брауна в термодинамике. В равновесии  $p = p(\rho)$  и энтропия постоянна. Поэтому независимо от пути по фазовой плоскости  $p$  всегда одно и то же (при замкнутой траектории). В случае неравновесных процессов газ при сжатии развивает давление больше равновесно-

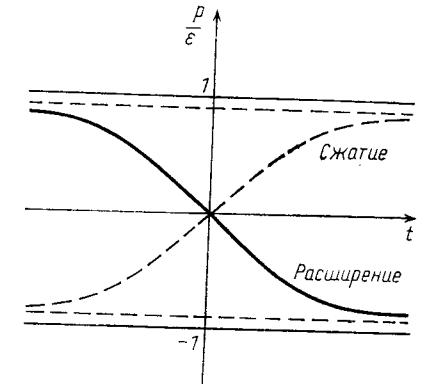


Рис. 26. Эволюция отношения  $p/\varepsilon$  для скалярного поля в расширяющейся (сплошная линия) и сжимающейся (пунктир) Вселенной. Асимптотические значения отношения  $p/\varepsilon$  отличаются от единицы (при  $m \ll H$ ) на величину  $2m^2/9H^2$

го, т. е. как бы дополнительно сопротивляется сжатию. Наоборот, при расширении газ сопротивляется расширению и давлению меньше. Поэтому в реальных газах после замкнутого цикла мы всегда имеем прирост энтропии  $\Delta S > 0$ . Но для этого нужны поправки типа

$$p = p(\rho, S) + \alpha \beta = p(\rho, S) - \beta \dot{V} \quad \text{с} \quad \alpha > 0 \quad \text{и} \quad \beta > 0,$$

т. е. как раз удовлетворяющие принципу Ле Шателье—Брауна. Изменение поля  $\varphi$  в рассматриваемой задаче происходит в полной аналогии с этим термодинамическим примером.

Перейдем теперь к анализу самосогласованной задачи. В начале рассмотрим ее упрощенный вариант. А именно, будем полагать, что вторая производная по времени  $\ddot{\varphi}$  значительно меньше, чем  $H\dot{\varphi}$  и чем  $m^2\varphi$ . Кроме того, будем предполагать, что  $\dot{\varphi} < m\varphi$  и что фон тоже меняется медленно, т. е.  $\dot{H} \ll H^2$ . Тогда самосогласованная система уравнения для поля  $\varphi$  и масштабного фактора принимает вид

$$\frac{1}{2} H^2 = \frac{4\pi G}{3} \varepsilon \quad \text{или} \quad H = \frac{m}{m_{Pl}} \varphi \sqrt{\frac{4\pi}{3}}. \quad (6.24)$$

В этом упрощенном случае решение нетрудно выписать:

$$\varphi = \varphi_0 - mm_{Pl}t / \sqrt{12\pi}. \quad (6.25)$$

Это решение справедливо на промежутке времени  $0 < t < \sqrt{12\pi} \varphi_0 / mm_{Pl}$ . На этом же промежутке времени справедливо и выражение для  $H$  вида

$$H = \sqrt{\frac{4\pi}{3}} \frac{m}{m_{Pl}} \left( \varphi_0 - \frac{mm_{Pl}}{\sqrt{12\pi}} t \right). \quad (6.26)$$

Зная выражение  $H(t)$ , можно найти основной инфляционный фактор — степень расширения в зависимости от параметров скалярного поля. Степень расширения для идеального решения де Ситтера дается выражением  $H\Delta t$ , где  $\Delta t$  — длительность деситтеровской стадии. В случае приближенного решения де Ситтера, когда  $H$  зависит от времени, степень расширения описывается выражением  $\int H dt$ . Для избранной модели

$$\int H dt = 2\pi (\varphi_0 / m_{Pl})^2.$$

Чтобы получить инфляцию с подходящим расширением, достаточно взять  $\varphi_0 \approx 4m_{Pl}$ , после чего показатель экспоненты, определяющий степень расширения, будет порядка 100, т. е.  $a_{\text{конец}} = a_0 \exp 100 = a_0 10^{43}$ . Это обстоятельство лежит в основе инфляции за счет медленного стремления поля  $\varphi$  к равновесию.

Теперь рассмотрим точную самосогласованную задачу. Будем считать, что темп расширения обеспечивается скалярным полем и как метрика, так и поле однородны. Вместе с тем мы рассмотрим вопрос о том, насколько вероятна инфляция в нашей Вселенной, и не составляют ли решения с инфляцией малую долю от всех возможных решений. Мы покажем, что это не так. Инфляция реализуется при очень широком классе начальных данных, и в этом смысле экспоненциальное расширение возникает при наличии скалярного поля весьма естественно, почти независимо от исходного состояния. Для доказательства этого факта рассмотрим уравнения движения массивного скалярного поля (6.18). Однако, в отличие от прежнего анализа, не будем делать никаких предположений о скорости изменения  $\varphi$  и  $a(t)$ .

Уравнения гравитационного поля с учетом обратного действия скалярного поля на метрику можно записать в виде

$$\ddot{\varphi} + 3H\dot{\varphi} + u'(\varphi) = 0, \quad (6.27)$$

$$H^2 + \frac{k}{a^2} = \frac{1}{6} G (m^2\varphi^2 + \dot{\varphi}^2). \quad (6.28)$$

Система уравнений (6.18), (6.27) и (6.28) не имеет точного решения де Ситтера. В единственной устойчивой точке потенциала  $\varphi=0$  потенциал обращается в нуль и соответственно со-

отношение  $p = -\varepsilon$  несправедливо. Однако при  $\varphi \neq 0$  движение к точке устойчивости может быть медленным, что приводит к приближенному решению де Ситтера, но приближение это настолько хорошее, что с физической точки зрения рассматриваемое решение несколько не хуже стандартного деситтеровского решения.

Наиболее наглядно свойства системы уравнений проявляются в фазовом пространстве, координатами которого являются  $x \sim \varphi$ ,  $y \sim \dot{\varphi}$  и  $z \sim H$ . Характеристики этой системы вырисовывают возможные траектории, которые описывают при эволюции  $\varphi(t)$  и  $a(t)$ . В случае плоского мира, т. е.  $k=0$ , диаграмма значительно упрощается. Она становится двумерной (рис. 27). Каждая кривая на рис. 27 представляет собой одну из возможных траекторий.

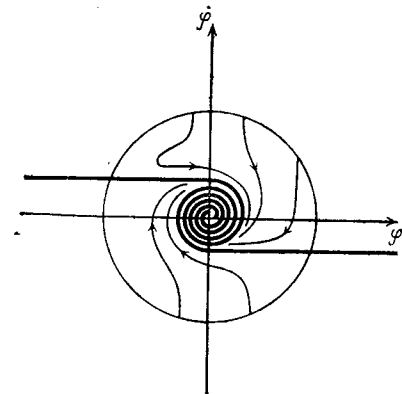


Рис. 27. Фазовые траектории поля  $\varphi$  в расширяющемся мире

Вдоль оси  $Ox$  идут две сепаратрисы, разделяющие два семейства решений. Сепаратрисы начинаются на круге единичного радиуса, который изображен пунктиром. Точки на этом круге соответствуют начальным значениям  $\varphi$  и  $\dot{\varphi}$ . (Переменные  $x$  и  $y$  как раз выбраны так, чтобы радиус круга был равен единице.) Бесконечному времени соответствует асимптотическое стремление  $\varphi$  и  $\dot{\varphi}$  к нулю. Хотя формально надо задавать значение  $\varphi$  и  $\dot{\varphi}$  при  $t=t_{Pl}$ , что соответствует положению точки на окружности единичного радиуса на плоскости  $x, y$ , надо осознавать, что при  $t=t_{Pl}$  требуется решать уже квантовые уравнения для гравитационного поля и система (6.18), (6.27), (6.28), вообще говоря, несправедлива. Отсутствие строгой квантовой теории гравитационного поля не дает возможности описывать состояние Вселенной при  $t=t_{Pl}$ . Эту трудность мы обойдем тем, что начальные условия будем задавать не на круге единичного радиуса, т. е. при  $t=t_{Pl}$ , а внутри этого круга, что отвечает  $t > t_{Pl}$ . Если начальные условия для  $\varphi(t)$  и  $\dot{\varphi}(t)$  таковы, что на фазовой диаграмме положение точки приходится на какую-либо из сепаратрис, то решение соответствующих уравнений приводит к эволюции, которая ближе к решению де Ситтера, чем все остальные решения этих уравнений. Другими словами, если в начале при  $t \approx t_{Pl}$  реализовались начальные условия для  $\varphi$  и



• такие, что  $\dot{\varphi} \approx 0$ ,  $\varphi \approx 1$ , то Вселенная испытывала сильную инфляционную стадию. Однако даже если начальная точка не находилась на сепаратрисе, то при довольно широком интервале начальных условий значительная инфляционная стадия все же возникала.

Вначале, стартуя с круга  $\varphi \leq 1$ , точка, описывающая состояние Вселенной, быстро проходит по плоскости  $x, y$ , подходя к одной из сепаратрис. В этот период Вселенная эволюционирует с эффективным (сверхжестким) уравнением состояния  $p = +\varepsilon$ . После того как точка подошла к одной из сепаратрис, она движется вдоль нее так, что  $\varphi = \varphi_{\text{сеп}}$  и  $\dot{\varphi} \approx 0$ . В этот период во Вселенной реализуется эффективное уравнение состояния  $p = -\varepsilon$  и Вселенная испытывает сильную инфляцию. Масштабный фактор в эту эпоху раздувается больше, чем в  $10^{30}$  раз, если  $H\tau > 70$ .

После конца инфляции траектория точки начинает «накручиваться» вокруг начала координат рядом с одной из сепаратрис. Траектории, которые описывают движение точки без инфляции или с малой ( $H\tau < 70$ ) инфляцией, соответствуют узкой области вблизи  $\varphi \approx 0$  вдоль оси  $Oy$ . Ширина области начальных условий решений без инфляции по сравнению с размерами круга составляет примерно  $\sim 10^{-6}$ . Таким образом, мы приходим к выводу, что скалярное поле естественно с «вероятностью» порядка  $(1-10^{-6})$  ведет к экспоненциальному расширению Вселенной.

В обсуждаемом здесь сценарии Вселенная начинает экспоненциально расширяться сразу после рождения (см. гл. 9), так что можно сказать, что в начальной стадии Вселенная была холодной, а разогрев ее происходит при окончании инфляции. Существуют и другие модели, в которых первоначально горячая Вселенная остывала в инфляционный период, а затем снова разогревалась. Модель такого типа мы рассмотрим в следующем параграфе.

### § 3. ХИГГСОВО ПОЛЕ И ИНФЛЯЦИЯ

Исторически первая инфляционная модель, использующая скалярное поле, была основана на ином механизме, чем обсуждался в § 2. Эта модель существенно опиралась на специфические свойства хиггсовского потенциала (6.4) и на его зависимость от температуры (разумеется, в формуле (6.4) не отраженную). Эту зависимость можно качественно установить с помощью следующих соображений.

Рассмотрим массивное вещественное скалярное поле с взаимодействием  $\lambda\varphi^4$ , т. е.

$$U = m_0^2\varphi^2 + \frac{\lambda}{4}\varphi^4. \quad (6.29)$$

Из-за симметрии теории относительно замены  $\varphi \rightarrow -\varphi$  среднее значение поля в основном состоянии равно нулю:  $\langle\varphi\rangle=0$ , аналогично тому как равно нулю среднее значение координаты в основном состоянии квантово-механического осциллятора:  $\langle x\rangle=0$ . Однако, как известно,  $\langle x^2\rangle \neq 0$ . Среднее значение квадрата поля из-за квантовых эффектов также отлично от нуля:

$$\langle\varphi^2\rangle \neq 0.$$

Фактически эта величина в рассматриваемой модели бесконечно велика, так как поле  $\varphi$  представляет собой систему с бесконечным числом степеней свободы. Однако от этой бесконечности можно избавиться с помощью упоминавшейся в § 7 гл. 4 процедуры перенормировок, «спрятав» ее в вакуумную энергию и в массу частицы:

$$\rho_{\text{вак}} = m^2 \langle\mu^2\rangle_0 + \frac{\lambda}{4} \langle\varphi^4\rangle_c.$$

Квантовая поправка к массе схематически возникает следующим образом:

$$\lambda\varphi^4 \rightarrow \lambda \langle\varphi^2\rangle_0 \varphi^2,$$

т. е.  $\delta m^2 = \lambda \langle\varphi^2\rangle_0$ , и мы называем физической массой величину  $m_0^2 + \lambda \langle\varphi^2\rangle_0$ , где индекс 0 указывает, что система рассматривается при нулевой температуре.

Что произойдет, если система находится в тепловом резервуаре с температурой  $T \neq 0$ ? Используя в качестве наводящих соображений аналогию с квантово-механическим осциллятором, мы можем заключить, что  $\delta m^2(T) = \lambda (\langle\varphi^2\rangle_0 + bT^2)$ , где  $b$  — некоторая постоянная; это значит, что эффективный потенциал\* поля  $\varphi$  при  $T \neq 0$  имеет вид

$$U(\varphi, T) = m^2\varphi^2 + bT^2\varphi^2 + \lambda\varphi^4/4. \quad (6.30)$$

В хиггсовском случае возникает аналогичная добавка к потенциалу и мы можем написать

$$U(\varphi, T) = \left(bT^2 - \frac{\lambda}{2}\varphi_0^2\right)\varphi^2 + \frac{\lambda}{4}\varphi^4 + \frac{\lambda}{4}\varphi_0^4. \quad (6.31)$$

Вид  $U(\varphi, T)$  для различных  $T$  изображен на рис. 28. При больших  $T$  потенциал имеет единственный минимум в точке  $\varphi=0$ . Конденсат поля  $\varphi$  отсутствует, т. е.  $\langle\varphi\rangle=0$ , и система находится в симметричном состоянии. Если в теории есть калибровочные поля, взаимодействующие с  $\varphi$ , то в этой фазе их масса равна нулю.

С падением температуры в  $U(\varphi, T)$  возникает дополнитель-

\* Более точно надо говорить не об эффективном потенциале, а о свободной энергии  $F$ , так как именно минимум  $F$  определяет положение равновесия системы.

ный минимум, который постепенно опускается ниже, чем значение  $U$  в точке  $\varphi=0$ . Точка  $\varphi=0$  оказывается точкой неустойчивого равновесия, и система переходит в состояние с ненулевым средним полем  $\langle\varphi\rangle\neq 0$ . Происходит спонтанное нарушение симметрии, и калибровочные поля приобретают массу в соответ-

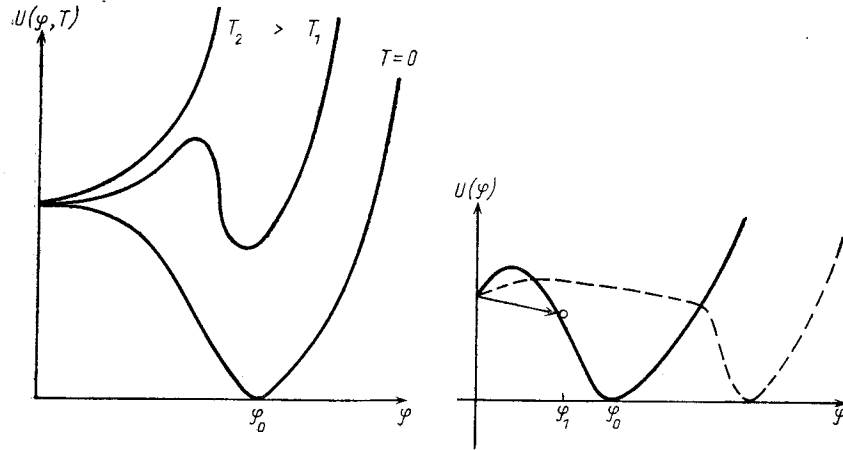


Рис. 28. Эффективный потенциал скалярного поля при разных температурах

Рис. 29. Туннелирование скалярного поля из  $\varphi=0$  в  $\varphi=\varphi_0$ , приводящее к спонтанному нарушению симметрии за счет фазового перехода I рода

вии со сказанным в § 8 гл. 4. Это явление было обнаружено в работах Д. А. Киржница (1972) и Д. А. Киржница, А. Д. Линде (1972).

Длительность пребывания в метастабильном состоянии  $\langle\varphi\rangle=0$  зависит от значений параметров  $\lambda$ ,  $\varphi_0$ . Если

$$-m_1^2 = -\left. \frac{\partial^2 U}{\partial \varphi^2} \right|_{\varphi=0, T=0} > H^2 \frac{8\pi}{3} \frac{\lambda \varphi_0^4}{4m_{Pl}^2},$$

то системы без задержки переходит от  $\langle\varphi\rangle=0$  к  $\langle\varphi\rangle=\varphi_0$ . Это быстрый фазовый переход II рода, и для нас он не интересен. Инфляция реализуется при  $m_1^2 < 0$ , когда  $|m_1^2| < H^2$ , либо, если  $m_1^2 > 0$ , т. е. в точке  $\varphi=0$  имеется локальный минимум. В этих случаях система долго (при  $m_1^2 > 0$  экспоненциально долго) находится в метастабильной фазе. Переходу  $\langle\varphi\rangle=\varphi_0$  препятствуют градиентные члены  $(\partial_\mu \varphi)^2$ , или, иными словами, поверхностная энергия пузырей новой фазы. При  $m_1^2 > 0$  фазовый переход от симметричной к асимметричной фазе напоминает образование пузырей пара в нагреваемой жидкости. Критический размер зародыша новой фазы должен быть достаточно велик,

чтобы его объемная энергия превышала поверхностную. Это и приводит к малой вероятности фазового перехода, в результате чего поле  $\varphi$  надолго «застревает» вблизи  $\varphi=0$  (рис. 29). При этом плотность энергии обычной материи падает как  $T^4$ , а плотность энергии поля  $\varphi$  остается постоянной:

$$\rho \equiv \rho(\varphi=0) = \lambda \varphi_0^4 / 4. \quad (6.32)$$

С помощью формул (6.7) и (6.11) можно убедиться, что  $T_{\mu\nu} = (\lambda \varphi_0^4 / 4) g_{\mu\nu}$ , т. е. ведет себя как эффективная космологическая постоянная. Если время задержки в точке  $\varphi=0$  достаточно велико, то  $\rho_0$  окажется больше, чем плотность энергии других форм материи, и Вселенная начнет расширяться экспоненциально:

$$a(t) \sim \exp\{H_0 t\}, \quad H_0 = \frac{\sqrt{\lambda} \varphi_0^2}{2m_{Pl}}. \quad (6.33)$$

При этом от всей остальной материи очень быстро ничего не остается (соответствующая плотность энергии ведет себя как  $\exp(-4Ht)$  для релятивистского вещества и  $\exp(-3Ht)$  для нерелятивистского) и Вселенная начинает выглядеть как расширяющаяся однородная и изотропная пустота, расширение которой определяется вакуумоподобной энергией (6.32). Заметим, что мы неявно предположили, что в симметричной фазе имеется отличная от нуля космологическая постоянная  $\Lambda = \lambda \varphi_0^4 / 32\pi m_{Pl}^2$  и что она точно сокращается энергией конденсата поля  $\varphi$ , когда  $\langle\varphi\rangle=\varphi_0$ . Это пример той самой тонкой настройки параметров, которая обсуждалась в § 4 гл. 5.

Как мы уже отметили, обсуждаемый процесс фазового перехода напоминает кипение воды с образованием пузырей пара внутри перегретой жидкости. В реальной жидкости сильной затяжки фазового перехода не происходит из-за наличия зародышей парообразования. Если для поля  $\varphi$  аналогичных зародышей нет, то затяжка перехода может быть достаточно большой и можно получить необходимое время инфляции  $H_0 t > 70$ .

Образование пузыря новой фазы выглядит как туннельный переход из точки  $\varphi_0$  в некоторую точку  $\varphi_1$  (рис. 29). Оценки вероятности этого перехода показывают, что время переохлаждения действительно может быть достаточно большим. После туннелирования поле  $\varphi$  стремится к точке устойчивого равновесия  $\varphi_0$ , подчиняясь уравнению

$$\ddot{\varphi} + 3H\dot{\varphi} + U'(\varphi) = 0. \quad (6.34)$$

По существу это уравнение близко к уравнению (6.19), отвечающему  $U = m^2 \varphi^2 / 2$ , так как форма  $U$  вблизи  $\varphi_0$  в хиггсовской модели близка к параболе.

Мы изложили сценарий раздувания Вселенной, предложенный в 1981 г. Гутом. Однако в модели Гута поле  $\phi$  после туннелирования быстро достигало своего равновесного значения  $\phi_0$ , т. е. размер пузыря был относительно мал, и наша Вселенная должна была состоять из множества пузырей. Такая модель приводит к сильным неоднородностям, так как контраст плотности энергии внутри пузыря и в его стенках очень велик. Этот недостаток был исправлен в модифицированной модели, предложенной Линде (1982) и Альбрехтом и Штейнхардтом (1982). Согласно этой модели потенциал  $U$  в окрестности точки  $\phi_1$  очень пологий, так что скорость изменения  $\phi$  мала по сравнению с темпом расширения мира (см. рис. 29, пунктир):

$$\dot{\phi}/\phi \ll H_0. \quad (6.35)$$

Мы фактически пришли к случаю, рассмотренному в § 2. Расширение Вселенной в точках  $\phi=0$  и  $\phi=\phi_1$  отличается не очень сильно, т. е. образовавшийся пузырь новой фазы с  $\langle\phi\rangle=\phi_1$  расширяется экспоненциально и его стенки (или неоднородности поля  $\phi$  в модели § 2) уходят далеко за пределы возможностей нашего наблюдения. В этой модели вся наша Вселенная представляет внутренность одного-единственного гигантски разросшегося пузыря, и мы можем в принципе понять, почему она во всех направлениях выглядит одинаковой.

#### § 4. ОКОНЧАНИЕ ИНФЛЯЦИИ. ПОВТОРНЫЙ РАЗОГРЕВ

На этом этапе у читателя может возникнуть резонный вопрос: чего мы добились с помощью этой инфляционной модели? Действительно, проблемы плоскостности, горизонта, однородности и изотропии решены, но какой ценой? Во Вселенной не осталось ничего кроме классического скалярного поля  $\phi(t)$ , или, как говорят, конденсата поля  $\phi(t)$ . Наш мир совершенно не похож на эту унылую картину. Нужно либо отказаться от обсуждавшейся модели, либо понять, как в ее рамках можно описать нашу реальную Вселенную. К счастью, второй вариант оказывается возможным.

Зависимость классического поля  $\phi$  от времени определяется уравнением (6.34). Качественно она изображена на рис. 30.

После образования пузыря новой фазы  $\phi$  попадает на пологий участок потенциала  $U$  (см. рис. 29) и растет очень медленно. По мере приближения к  $\phi_0$  потенциал становится круче и скорость роста  $\phi$  выше. Достигнув положения равновесия,  $\phi$ , вообще говоря, начинает осциллировать около него. Затухание осцилляций связано с эффективным трением для поля  $\phi$ , которое складывается из двух факторов: во-первых, из члена  $3H\dot{\phi}$  и, во-вторых, из-за рождения частиц, которое в уравнении (6.34) не учтено.

Мы знаем из электродинамики, что переменное электрическое поле  $E$  может рождать электрон-позитронные пары, причем вероятность рождения не мала при достаточно больших частотах  $\omega \gtrsim m_e$ . Если поле  $\phi$  обладает взаимодействиями с какими-то элементарными частицами, то его осцилляции будут в точности так же рождать эти частицы, как поле  $E$  рождает пары  $e^+e^-$ .

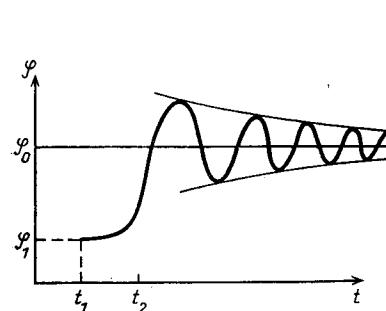


Рис. 30

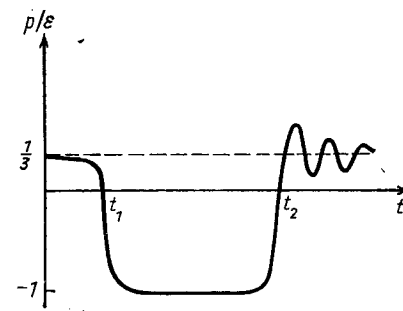


Рис. 31

Рис. 30. Зависимость скалярного поля, вызывающего инфляцию, от времени. Точка  $t_1$  отвечает моменту туннелирования. Между моментами времени  $t_1$  и  $t_2$  поле  $\phi$  медленно меняется и Вселенная продолжает экспоненциально раздуваться. После  $t_2$  эволюция  $\phi$  становится быстрой, возникают осцилляции вокруг положения равновесия, приводящие к рождению частиц и вторичному разогреву

Рис. 31. Эволюция отношения  $p/\epsilon$  в инфляционной модели. Момент  $t_2$  отвечает окончанию инфляции

Частота осцилляций поля в рассматриваемой модели равна его массе  $m_\phi$ . В моделях большого объединения  $m_\phi \gtrsim 10^{15} - 10^{14}$  ГэВ. Таким образом, поле  $\phi$  рождает элементарные частицы с  $m \lesssim m_\phi$ , перекачивая энергию своих колебаний в энергию частиц. Если темп расширения мира ниже скорости рождения частиц и скорости реакций между ними, то образовавшиеся частицы термализуются и мы приходим к стандартному космологическому сценарию. На начальном этапе, однако, возможны значительные отклонения от термодинамического равновесия, что помогает решить проблему барионной асимметрии Вселенной (Долгов, Линде, 1982).

Таким образом, скалярное поле не только является движущей силой инфляции, но и прародителем всей остальной материи во Вселенной.

В рассмотренном сценарии мы начали с горячей стадии, потом произошло экспоненциально сильное охлаждение, а затем повторный разогрев. В терминах соотношения между  $p$  и  $\epsilon$  история развития событий изображена на рис. 31.

## Глава 7.

## ПОЛЯРИЗАЦИЯ ВАКУУМА И ИНФЛЯЦИЯ

Квантовая теория поля принципиально изменила представления не только о взаимодействиях материальных частиц, но и о свойствах пустого пространства — вакуума. Казалось бы, что может быть проще пустоты! Однако сейчас мы знаем, что вакуум является весьма сложным объектом. Нетривиальные свойства вакуума согласно современным представлениям обеспечивают невылетание кварков; масса промежуточных бозонов  $W^+$  и  $Z^0$  обусловлена их взаимодействием с вакуумом. Такие примеры можно продолжать и далее. Мы подробнее остановимся на одном из них — явлении поляризации вакуума. Оказывается, в присутствии внешнего поля (электромагнитного, гравитационного и т. д.) вакуум изменяет свои свойства, поляризуется, аналогично тому как поляризуется в электрическом поле диэлектрик. Внешнее поле действует на виртуальные частицы и тем самым меняет свойства вакуума. Эти изменения могут быть наблюдаемыми и, в частности, в космологической ситуации могут привести к инфляции (в самосогласованном режиме при быстром расширении).

## § 1. ПОЛЯРИЗАЦИЯ ВАКУУМА В ЭЛЕКТРОДИНАМИКЕ

Мы начнем с более простого, лучше изученного и, главное, проверенного на эксперименте случая квантовой электродинамики. Как и диэлектрик, вакуум поляризуется в присутствии внешнего электрического поля, например в окрестности какого-то заряда  $Z$ . Поле этого заряда можно измерить, например, по рассеянию на нем пробных частиц с зарядом  $e$  (рис. 32).

Как мы уже отмечали, вакуумные поля в присутствии заряда  $Z$  изменяют свое поведение. Виртуальные электроны стремятся быть ближе к заряду, позитроны — дальше (если  $Z > 0$ ).

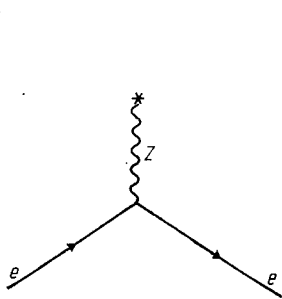


Рис. 32. Рассеяние электрона на тяжелом заряде  $Z$

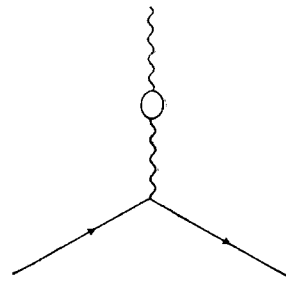


Рис. 33. Экранировка заряда  $Z$  виртуальными  $e^+e^-$ -парами

Соответствующая диаграмма изображена на рис. 33. В результате в полной аналогии с диэлектриком происходит экранировка заряда. Заряд на больших расстояниях оказывается меньше, чем на малых.

Обычно говорят не о зависимости заряда от расстояния, а о зависимости от переданного при рассеянии импульса. Между одним и другим языком существует очевидная связь; так как импульс и координата являются фурье-сопряженными переменными, то чем больше переданный импульс, тем меньше характерные расстояния в процессе и наоборот. Можно показать, что зависимость квадрата заряда от переданного импульса при больших  $q^2$  в электродинамике имеет вид

$$Z^2(q^2) = Z_0^2 \left[ 1 - \frac{\alpha}{2\pi} b \ln \frac{q^2}{q_0^2} \right]; \quad (7.1)$$

здесь  $\alpha = 1/137$  — постоянная тонкой структуры, а  $b$  — численный коэффициент, зависящий от виртуальных частиц (числа возможных различных петель на диаграмме). Такая зависимость как раз отвечает тому, что на малых расстояниях заряд окажется больше, чем на больших. Из-за взаимодействия с вакуумом частица перестает быть точечной, а приобретает некоторое пространственное распределение заряда.

С этим эффектом связана известная трудность квантовой электродинамики, получившая название нуля заряда. Дело в том, что при любом конечном значении заряда на предельно малых расстояниях (так называемом затравочном заряде) заряд на большом расстоянии (который измеряется, например, по закону Кулона) согласно теоретическим расчетам оказывался равным нулю. Это указывало на логическую незамкнутость, непоследовательность теории. Однако в последние десятилетия было понято, что существуют другие взаимодействия (кроме электромагнитного), учет которых приводит к коренной модификации поведения заряда на малых расстояниях, давая антиэкранировку заряда.

Эффект поляризации вакуума реально проявляется на опыте. Он с очень хорошей точностью измерен в атомной физике. Как известно, уровни  $^2S_{1/2}$  и  $^2P_{1/2}$  в атоме водорода даже с учетом релятивистских эффектов, являются вырожденными. Однако в квантовой теории поля это вырождение снимается. Снятие вырождения связано с двумя эффектами. Во-первых, электрон может испустить и поглотить виртуальный фотон (рис. 34). В результате этого меняется взаимодействие электрона с кулоновским полем ядра (на рис. 35 крестиком обозначено ядро) и, кроме того, при таком испускании виртуального кванта электрон как бы получает приращение импульса, в результате чего уменьшается локализация его волновой функции

вблизи ядра. Это заметно меняет уровень  ${}^2S_{1/2}$ , поднимая его вверх примерно на 1000 МГц, и почти не влияет на  ${}^2P_{1/2}$ . Это различие связано с тем, что волновая функция  $S$ -состояния конечна в нуле, а  $P$ -состояния — равна нулю; область же вблизи ядра наиболее существенна для энергии уровня, так как кулоновское поле там сильнее всего.

Второй эффект связан с модификацией закона Кулона из-за поляризации вакуума (см. рис. 32). Он приводит к смещению

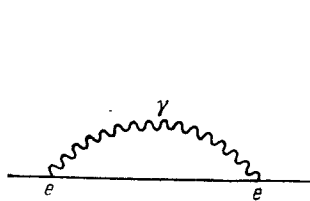


Рис. 34. Испускание и поглощение электроном виртуального фотона

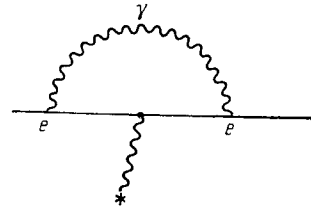


Рис. 35. Модификация взаимодействия электрона с внешним полем за счет испускания и поглощения виртуального кванта

${}^2S_{1/2}$ -уровня примерно на 25 МГц и по указанным выше причинам почти не влияет на  ${}^2P_{1/2}$  (последний из-за двух упомянутых эффектов вместе опускается примерно на 17 МГц).

Предсказываемое теорией расщепление уровней (называемое лэмбовским сдвигом) было с прекрасной точностью подтверждено опытом не только в обычном атоме водорода, но и в мюонном атоме, где вместо электрона находится отрицательно заряженный мюон. Последнее особенно интересно потому, что боровский радиус мюонного атома в  $m_e/m_\mu$  210 раз меньше боровского радиуса электронного атома и соответственно эффекты поляризации вакуума проявляются сильнее. Если в обычном атоме водорода поляризация вакуума дает около 3% от общего смещения, то в мюонном атоме ее вклад доминирует и приводит к смещению уровня вниз примерно на  $0,2 \text{ эВ} \approx 3 \cdot 10^8 \text{ МГц}$ .

Экспериментальное изучение сдвига уровней  ${}^2S_{1/2}$  и  ${}^2P_{1/2}$  особенно удобно для изучения специфических эффектов квантовой теории поля, так как этот сдвиг отсутствует в обычной квантовой механике. Эксперименты такого типа называются нулевыми, так как в них малый эффект не маскируется общим большим сигналом. Сегодня, однако, благодаря лазерным методам можно измерить добавку к сдвигу спектральной линии  $2P \rightarrow 1P$  или  $3P \rightarrow 1S$  и, таким образом, найти абсолютное смещение низшего стабильного  $1S$ -уровня, обусловленное виртуальными частицами. Это смещение в 8 раз больше смещения уровня  $2S$  в соответствии с величиной  $|\psi(0)|^2$ .

Поправки к закону Кулона, обусловленные поляризацией вакуума электрон-позитронными парами, имеют довольно простой вид в предельных случаях малых ( $r \leq m_e^{-1}$ ) и больших ( $r \gg m_e^{-1}$ ) расстояний:

$$\varphi(r) = \frac{e}{r} \left[ 1 - \frac{2\alpha}{3\pi} \left( \ln m_e r + C + \frac{5}{6} \right) \right], \quad (7.2)$$

$$\varphi(r) = \frac{e}{r} \left[ 1 + \frac{\alpha}{4\sqrt{\pi}} (m_e r)^{-3/2} \exp(-2m_e r) \right], \quad (7.3)$$

где  $C = 0,577$  — постоянная Эйлера. Для вывода этих выражений нужно уметь пользоваться аппаратом квантовой теории поля, но их физические основания весьма прозрачны.

Заметим, что согласно (7.3) закон обратных квадратов является лишь приближенным, но отклонения от него при  $r > m_e^{-1} \approx 10^{-11}$  см. фантастически малы.

Если в вакууме постоянно рождаются и исчезают виртуальные частицы, то возникает вопрос: нельзя ли их как-нибудь извлечь оттуда, сделать реальными? Ответ оказывается утвердительным: достаточно сильное электрическое поле рождает из вакуума пары заряженных частиц. Мы можем оценить вероятность перехода на основе следующих соображений. Пусть  $W(l)$  — вероятность того, что виртуальные электрон и позитрон возникли на расстоянии  $l$  друг от друга. Выражение (7.3) для поправки к закону Кулона дает основание полагать, что  $W(l) \sim \exp(-2ml)$ . Теперь, если  $l$  достаточно велико, так что разность потенциалов  $e^+$  и  $e^-$  в электрическом поле  $\mathcal{E}$  превышает сумму их масс, т. е.

$$e(\varphi_1 - \varphi_2) = e\mathcal{E}l \geq 2m,$$

то рождение реальных  $e^+$ ,  $e^-$  оказывается энергетически возможным. Отсюда для вероятности рождения можно получить оценку  $W \sim \exp(-\text{const} \cdot m^2/e\mathcal{E})$ . Разумеется, с помощью этих простых рассуждений мы не могли верно определить численный коэффициент перед фактором  $m^2/e\mathcal{E}$ . Точный ответ для малых  $\mathcal{E}$  имеет вид:

$$W = (\alpha \mathcal{E}^2 / \pi^2) \exp(-\pi \mathcal{E}_0 / \mathcal{E}), \quad (7.4)$$

где  $\alpha = e^2 = 1/137$  — постоянная тонкой структуры, а  $\mathcal{E}_0 = m^2/e \approx 1,3 \cdot 10^{16}$  В/см — так называемое критическое поле. Это выражение можно получить, рассматривая туннельный переход электрон-позитронной пары из области, где их нахождение по законам классической механики запрещено, в разрешенную область. Физический смысл экспоненциального подавления в формуле (7.4) тот же, что и в задаче о квантовом прохождении потенциального барьера, превосходящего по высоте энергию частицы.

Величина электрического поля порядка  $\mathcal{E}_0$  представляет собой абсолютный верхний предел для возможных будущих ускорителей традиционного типа, так как при полях больше этой величины происходит «пробой вакуума» и рождающиеся частицы экранируют поле. Заметим, что даже при таких гигантских напряженностях поля для рождения сверхтяжелых векторных бозонов моделей большого объединения с  $m \approx 10^{15}$  Гэв был бы необходим ускоритель размером около 1000 км.

Заметим, что ускорение частиц с помощью лазерных полей может оказаться более эффективным, чем с помощью статических (или, точнее, медленно меняющихся), так как статические поля максимальны на границах, а лазерное поле может быть сфокусировано прямо в зоне ускорения частиц.

Здесь уместно остановиться на следующем парадоксе. Скорость изменения плотности энергии образующихся частиц  $mc^2 dn/dt$  должна быть порядка работы, совершаемой над этими частицами электрическим полем

$$mc^2 \frac{dn}{dt} \approx j\mathcal{E},$$

где плотность тока  $j = env \ll enc$ . Отсюда нетрудно получить

$$\dot{n} \leq en\mathcal{E}/mc,$$

или

$$\ln \frac{n}{n_0} \leq \frac{e\mathcal{E}}{mc} dt.$$

Таким образом, если в начальный момент частицы отсутствуют, то они, казалось бы, должны отсутствовать всегда, т. е. рождения частиц не происходит. На этот парадокс впервые указал С. Хокинг в применении к гравитационному полю. Причиной отсутствия рождения является то обстоятельство, что в этом рассуждении скорость рождения частиц пропорциональна их плотности, т. е.  $\dot{n} \sim n$ . Очевидно, при нулевых начальных условиях это уравнение имеет решение  $n \equiv 0$ . Решение парадокса было указано Я. Б. Зельдовичем и Л. П. Питаевским, которые отметили, что частицы не возникают из ничего, они виртуально присутствуют в вакууме. Последний под действием электрического поля поляризуется, причем эффект поляризации пропорционален величине поля. В результате соотношения между энергией и плотностью изменяется и мы приходим к уравнению типа  $\dot{n} \sim \sqrt{n}$ , которое имеет ненулевые решения при нулевых начальных данных.

Чтобы проследить это, обратимся к более простому, но полностью аналогичному случаю поляризации атома водорода во внешнем электрическом поле. Как известно, атом водорода не имеет электрического дипольного момента:

$$d = \int \psi^* \psi x dx.$$

Однако под действием поля атом поляризуется, к основному состоянию  $1S$  примешивается  $2P$

$$\psi = \psi_{1S} + \delta \cdot \psi_{2P}$$

и возникает дипольный момент  $d \sim \delta$ . Зададимся вопросом, какова скорость образования примеси  $2P$ -состояния. Вероятность быть в  $2P$ -состоянии

$$W \sim \delta^2,$$

поправка к энергии уровня тоже пропорциональна  $\delta^2$ :

$$\langle E \rangle = E_{1S} + \delta^2 \cdot E_{2P},$$

но гамильтониан взаимодействия, перемешивающего уровни, пропорционален первой степени  $\delta$ :

$$H_{int} \sim \delta \cdot \mathcal{E}.$$

Отсюда видно, что вероятность образования  $2P$ -уровня описывается уравнением вида

$$\dot{W} = \text{const} \sqrt{W},$$

которое при нулевом начальном условии имеет ненулевое решение  $W \sim t^2$ .

## § 2. ПОЛЯРИЗАЦИЯ ВАКУУМА В ИСКРИВЛЕННОМ ПРОСТРАНСТВЕ

В полной аналогии с электродинамикой гравитационное поле также поляризует вакуум. При этом из-за нелинейности гравитационных уравнений могут возникнуть очень интересные эффекты. Классические уравнения Эйнштейна имеют вид

$$G_{\mu\nu} = \kappa T_{\mu\nu}, \quad (7.5)$$

где  $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$  (см. (3.21)). Квантовые поправки приводят к изменению вида тензора энергии-импульса материи. Мы уже отмечали, что квантовые поправки изменяют массы частиц, меняют значения констант взаимодействия, что приводит к перенормировке этих величин. Кроме того, квантовая энергия нулевых колебаний приводит к возникновению космологической постоянной, т. е. к появлению в  $T_{\mu\nu}$  дополнительного члена  $T_{\mu\nu}^{(1)} \simeq \rho_{\text{вак}} g_{\mu\nu}$  (гл. 5). Помимо этого относительно тривиального вклада в  $T_{\mu\nu}$  существуют и другие квантовые добавки, происхождение которых связано с изменением нулевых колебаний вакуума в присутствии гравитационного поля, т. е. фактически с поляризацией вакуума гравитационным полем. Очевидно, вклад этих добавок исчезает, когда кривизна простран-

ства-времени стремится к нулю. Модификация тензора — энергии-импульса за счет этих эффектов была рассмотрена в работе Гинзбурга, Киржница и Любошина (1971).

Простейшая добавка в  $T_{\mu\nu}$  такой природы имеет вид

$$T_{\mu\nu}^{(2)} = A G_{\mu\nu}, \quad (7.6)$$

где  $A$  — некоторая постоянная. Такая форма диктуется, во-первых, тензорной структурой  $T_{\mu\nu}$  и, во-вторых, ковариантным законом сохранения

$$T_{\mu;\nu}^{\nu} = 0.$$

Этот член просто дает перенормировку гравитационной константы за счет квантовых эффектов и в наблюдениях никак не проявится.

Нетривиальный вклад в  $T_{\mu\nu}$  за счет поляризации вакуума гравитационным полем имеет вид

$$T_{\mu\nu}^{(3)} = B_1 \left( 2R_{;\mu\nu} - 2g_{\mu\nu} \square R - \frac{1}{2} g_{\mu\nu} R^2 + 2RR_{\mu\nu} \right) + B_2 \left( 2R_{\mu;\nu\alpha}^{\alpha} - \square R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \square R + 2R_{\mu}^{\alpha} R_{\alpha\nu} - \frac{1}{2} g_{\mu\nu} R^{\alpha\beta} R_{\alpha\beta} \right). \quad (7.7)$$

Безразмерные коэффициенты  $B_1$  и  $B_2$  теорией не определяются; это перенормировочные константы, величина которых произвольна. Заметим, что оба слагаемых в выражениях (7.7) ковариантно сохраняются. Добавки (7.6) и (7.7) — локальны, т. е. зависят от геометрических характеристик пространства-времени в одной точке. Помимо них возможны и нелокальные члены, явный вид которых в общем случае установить не удастся.

Поправки к тензору энергии-импульса, обусловленные поляризацией вакуума в искривленном пространстве, имеют особенно простой вид в пространстве де Ситтера. Высокая симметрия этого пространства, о которой шла речь в § 2 гл. 5, позволяет однозначно установить тензорную структуру различных геометрических величин. В частности,

$$R_{\alpha\beta\gamma\delta} = H^2 (g_{\alpha\gamma} g_{\beta\delta} - g_{\alpha\delta} g_{\beta\gamma}), \quad (7.8)$$

$$R_{\alpha\beta} = 3H^2 g_{\alpha\beta},$$

и, наконец,

$$\Delta T_{\alpha\beta} = \text{const } H^4 g_{\alpha\beta}. \quad (7.9)$$

Если теперь мы учтем квантовые поправки в уравнениях ОТО, т. е. добавим в правую часть уравнений (7.5) дополнительное слагаемое (7.7), то придем к очень интересному выводу: помимо тривиального решения, отвечающего плоскому

пространству (т. е.  $R_{\mu\nu\alpha\beta} = 0$ ), в модифицированных уравнениях появятся решения с  $R_{\mu\nu\alpha\beta} \neq 0$ . Иными словами, даже в отсутствие материи, реальные частицы, пространство может быть искривленным, расширяющимся. Существует самосогласованный режим, когда кривизна пространства-времени создает поляризацию вакуума, которая в свою очередь вызывает искривление пространства.

Проиллюстрируем эти рассуждения примером пространства де Ситтера. Уравнение (7.5) без материи, но с поправкой (7.9) имеет вид

$$H^2 g_{\mu\nu} = A H^4 g_{\mu\nu} m_{\text{Pl}}^{-2}. \quad (7.10)$$

Важно отметить, что гипотеза о форме метрики оказалась самосогласованной: мы нашли поправки в  $T_{\mu\nu}$  в метрике де Ситтера, подставили их в уравнения ОТО и обнаружили, что решения этих уравнений приводят опять к метрике де Ситтера. Легко видно, что уравнение (7.10) помимо тривиального решения  $H=0$  имеет ненулевое решение  $H = m_{\text{Pl}} A^{-1/2}$ . Легко сообразить, что  $A$  пропорционально числу сортов элементарных частиц, дающих вклад в поляризацию вакуума.

Таким образом, мы пришли к модели, в которой экспоненциальное расширение обусловлено квантовыми поправками к тензору энергии — импульса. На возможность такого режима обратили внимание Гурович и Старобинский (1979). Впоследствии Старобинским (1980) на этом основании была разработана инфляционная модель, целиком основанная на эффектах поляризации вакуума. Можно проверить, что уравнения (7.5) с правой частью (7.7) действительно имеют (среди множества других) деситтеровское решение. Это решение является неустойчивым, и малые отклонения от него со временем нарастают, что приводит к переходу от экспоненциального режима к фридмановскому. Скорость роста этих возмущений зависит от упомянутого выше числа типов элементарных частиц  $N$ . Если  $N$  достаточно велико ( $N \approx 10^3$ ), то продолжительность такой квантовой инфляции оказывается достаточной для решения проблем плоскостности, горизонта и пр. Недавно, однако, удалось показать, что обсуждаемый механизм может работать и не при столь гигантских  $N$ , если учесть вклад в  $T_{\mu\nu}$  дополнительных слагаемых, опущенных в первоначальном варианте.

## Глава 8.

## БАРИОСИНТЕЗ ВО ВСЕЛЕННОЙ

Астрономические наблюдения показывают, что во Вселенной практически отсутствует антивещество (т. е. антипротоны, антинейтроны, позитроны). Мы уверенно можем говорить, что наша Галактика целиком состоит из вещества, так как в противном случае аннигиляция вещества и антивещества, сопровождающаяся колоссальным выделением энергии, несомненно, была бы замечена. Аналогично можно сделать вывод, что любая из наблюдаемых галактик состоит из вещества одного типа (т. е. либо только из вещества, либо только из антивещества). Кроме того, известны сталкивающиеся галактики, галактики, омываемые облаком межгалактического газа, и во всех этих случаях никакого всплеска аннигиляции не обнаружено. Эти факты ставят перед нами проблему зарядовой асимметрии Вселенной. Иногда ее называют барионной асимметрией, имея в виду, что во Вселенной присутствуют барионы (т. е. протоны, нейтроны) и отсутствуют антибарионы. Разумеется, мы видим небольшое количество антипротонов в космических лучах, но их доля мала, порядка  $10^{-4}$  от протонов. Интересно отметить, что количество реликтовых нейтрино и антинейтрино во Вселенной должно быть почти одинаковым, так что явление значительной зарядовой асимметрии в современную эпоху имеет место для достаточно сильно взаимодействующих частиц.

Явление барионной асимметрии в горячей Вселенной кажется особенно загадочным еще и потому, что количество протонов весьма мало по сравнению с количеством реликтовых фотонов:

$$N_B/N_\gamma = 10^{-9} - 10^{-10}. \quad (8.1)$$

Это значит, что на горячей стадии при температуре  $kT > m_p c^2$  в первичной термодинамически равновесной плазме существовало примерно равное количество  $p$  и  $\bar{p}$ :

$$(N_p - N_{\bar{p}})/N_p = 10^{-9} - 10^{-10}.$$

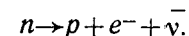
Точнее, при таких температурах мы должны говорить уже о кварках и антикварках, но факт не меняется: на ранней стадии должен существовать ничтожно малый избыток кварков над антикварками. Весь наш теперешний мир обязан своим существованием этому микроскопическому преобладанию частиц над античастицами.

Еще 15—20 лет назад происхождение этой небольшой вначале и практически полной сейчас асимметрии казалось чрезвычайно загадочным и рассматривалось как начальное условие, не определяемое законами физики. В частности, рассматривались модели с холодным вначале веществом, целиком сос-

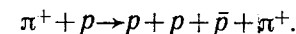
тоящим из барионов; последующий быстрый разогрев приводил к рождению пар. Все это далекое прошлое. Сейчас господствует противоположная точка зрения, позволяющая объяснить факт наличия этой асимметрии, но и оценить ее величину.

## § 1. НЕСОХРАНЕНИЕ БАРИОНОВ. РАСПАД ПРОТОНА

Критическим шагом здесь является предположение о несохранении количества барионов. Остановимся на этом подробнее. Известными барионами являются протоны и нейтроны, составляющие атомные ядра в окружающем нас веществе. Кроме них имеется еще ряд относительно короткоживущих барионов: странных (или гиперонов), очарованных и т. д. Эти частицы объединяются в одно семейство тем своим свойством, что их полное число во всех исследованных до настоящего времени процессах строго сохраняется. Например, если распадается нейтрон, то в результате распада появляется другой барион — протон:



Если в результате какой-либо реакции рождается дополнительный протон, то это обязательно сопровождается рождением какого-либо антибариона, например:



Для описания этого явления было введено понятие барионного заряда  $B$ , который положили равным  $+1$  для барионов и  $-1$  для антибарионов. Тогда закон сохранения барионов формулируется как закон сохранения барионного заряда в полной аналогии с сохранением электрического заряда.

Самым ярким свидетельством в пользу сохранения  $B$  является стабильность атомных ядер и стабильность легчайшего из барионов — протона. Эксперимент показывает, что его время жизни больше  $10^{31}$  лет.

Между электрическим и барионным зарядами, однако, имеется важное различие. Электрический заряд служит источником взаимодействия с безмассовым векторным полем — электромагнитным. В этом случае теория требует, чтобы заряд строго сохранялся. Несмотря на весьма тщательные поиски безмассового поля (и соответственно дальнего действия), связанного с барионным зарядом, оно не было обнаружено (см. § 4 гл. 3). Поэтому оснований для сохранения барионного заряда у нас нет. Более того, модели большого объединения, единым образом описывающие сильные и электрослабые взаимодействия, предсказывают, что барионный заряд не должен сохраняться. Как уже отмечалось, в этих моделях должны существовать векторные поля, обозначаемые  $X$  и  $Y$ , взаимодействующие с кварками ( $q$ ) и лептонами ( $l$ ) по схеме



$$X \leftrightarrow 2q, \quad X \leftrightarrow \bar{q}l$$

(и аналогично  $Y$ ). Электрический заряд квантов поля  $X$  равен  $\pm 4/3$  заряда электрона, а  $Y - \pm 1/3$ .

Легко видеть, что барионный заряд не сохраняется в цепочке процессов

$$q + q \leftrightarrow X \leftrightarrow \bar{q} + l,$$

так как  $B(q) = 1/3$ ,  $B(\bar{q}) = -1/3$  и  $B(l) = 0$ . Заметим, что  $\Delta B$  в этом процессе целое,  $\Delta B = -1$ .

Сразу возникает вопрос: как совместить несохранение  $B$  с наблюдаемой стабильностью (или точнее весьма большим временем жизни) протона? Оказывается это можно сделать, если масса квантов  $X(Y)$  — поля, или, как говорят,  $X(Y)$  — бозонов, достаточно велика. Теория именно это и предсказывает. В простейшей модели большого объединения, использующей  $SU(5)$ -симметрию, было получено  $m_x \approx m_y \approx 3 \cdot 10^{14}$  ГэВ (к сожалению, как мы увидим ниже, эта величина немного маловата).

Распад протона (состоящего из трех кварков) происходит согласно диаграмме рис. 36. Пара  $q\bar{q}$  в конечном состоянии образует  $\pi$ - или  $K$ -мезон, а  $l$  может быть, например, позитроном или мюоном. В итоге мы получаем распады типа  $p \rightarrow \pi^0 e^+$ ,  $K^0 \mu^+$  и т. п.

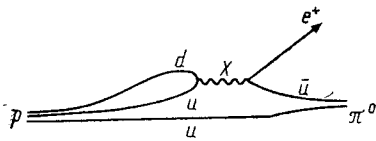


Рис. 36. Распад протона на  $\pi^0$  и  $e^+$

$$p = \begin{cases} d \\ u \\ u \end{cases}$$

Рис. 37. Кварковый состав протона

Точно в соответствии с квантово-механическим описанием распада нейтрона рассмотрим и распад протона. Протон, как мы знаем, представляет собой систему трех кварков, что изображено на рис. 37. Как отмечалось на предыдущих страницах, существует взаимодействие, переводящее два кварка ( $u, d$ ) в сверхтяжелый  $X$ -бозон. Амплитуда этого перехода  $C_1$  пропорциональна

$$C_1 \sim em_p / (m_x + m_u - m_p).$$

Появляется состояние, описываемое волновой функцией  $\psi_p + C_1 \psi_{Xu}$  или средней частью диаграммы на рис. 38. Поскольку образование  $X$ -бозона — процесс виртуальный, то следует учесть переход  $X$ -бозона в реальные частицы  $e^+$ ,  $\bar{u}$ . Существование таких процессов требует единая теория сильных, слабых и электромагнитных взаимодействий. Амплитуда соответствующего перехода пропорциональна

$$C_2 \sim em_p / (m_x + m_u - m_e - m_\pi).$$

Отличие от распада нейтрона заключается в большей массе  $X$ -бозона по сравнению с  $W$ -бозоном, а также в том, что дефект массы по порядку величины равен массе протона. Постоянная тонкой структуры  $e^2/4\pi$ , которая в электродинамике

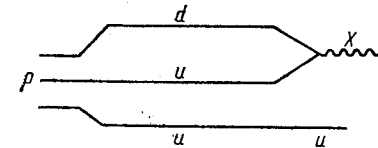


Рис. 38. Переход  $d$ - и  $u$ -кварков в виртуальный  $X$ -бозон

равна  $1/137$ , в единой теории сильных, слабых и электромагнитных взаимодействий равна примерно  $1/40$ . Это связано с перенормировкой заряда при высоких энергиях  $E \approx 10^{14}$  ГэВ.

Если бы масса  $X$ -бозона была бы меньше массы протона и он мог бы реально родиться в распаде протона, то распад был бы разрешен в первом порядке теории возмущений. Его вероятность была бы пропорциональна только  $|C_1|^2$ . Из-за того, что  $m_x > m_p$ , мы должны учитывать второй порядок теории возмущений, и вероятность распада протона в единицу времени по порядку величины описывается выражением

$$\Gamma \approx e^4 (m_p/m_x)^4 m_p,$$

а время жизни  $\tau \equiv \Gamma^{-1} \approx e^{-4} (m_x/m_p)^4 m_p^{-1} \approx 10^{-20} (m_x/m_p)^4$  с. Так как кварки в свободном состоянии не существуют, то они должны связаться в бесцветные адроны, такие, как  $\pi$ -,  $\eta$ - или  $K$ -мезоны, что отвечает третьей части диаграммы (рис. 39). Вероятность распада протона в тот или иной канал определяется вероятностью перехода  $X$ -бозона в кварк-лептонную пару и вероятностью образования того или иного адрона из пары кварк-антикварк. Неопределенность нашей размерной оценки величины  $\Gamma$  связана со структурой волновой функции кварков в протоне и парциальными вероятностями различных каналов распада. Учет этих факторов тем не менее может изменить нашу оценку не более чем на порядок.

Вероятность процессов с несохранением барионного заряда должна расти с ростом энергии сталкивающихся частиц. Разумеется, время жизни (период полураспада) отдельного протона не зависит от его энергии с точностью до известного релятивистского увеличения времени. Однако, например, при столкновении двух протонов возможна реакция

$$p + p \rightarrow \pi^0 + e^+ + p,$$

описываемая диаграммой рис. 40. Сечение рассеяния, а следовательно, вероятность реакции с нарушением барионного заря-

да растет пропорционально квадрату энергии (в системе центра) сталкивающихся протонов. Относительная вероятность таких процессов порядка 1 при энергиях  $E \approx m_X \approx 10^{14} - 10^{15}$  ГэВ. Для численных оценок времени жизни протона подставим

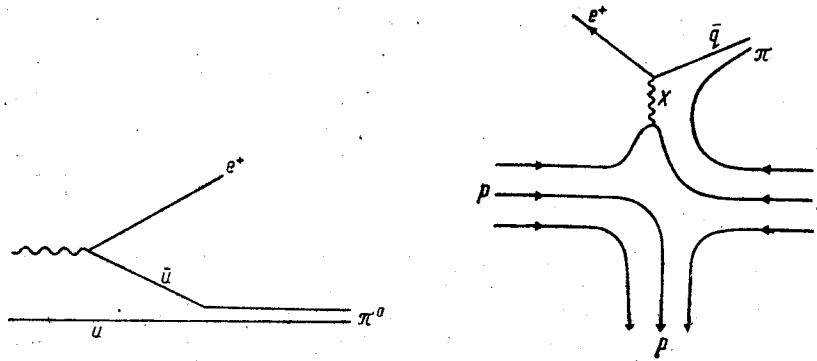


Рис. 39. Переход виртуального  $X$ -бозона в  $e^+e^-$  и слияние  $e^+e^-$  в  $\pi^0$ .

Рис. 40. Реакция  $p+p \rightarrow \pi^0 + e^+ + p$ , идущая с несохранением барионного заряда

$m_X = 3 \cdot 10^{14}$  ГэВ и получим  $\tau_p \approx 3 \cdot 10^{29}$  лет. Более точный расчет приводит к несколько более высокому результату, который, однако, все равно ниже достигнутого сейчас экспериментального предела. Это показывает, что простейший вариант теории несправедлив, но, разумеется, не уничтожает саму идею, так как можно, несколько усложнив схему, получить и более высокое значение  $\tau_p$ .

В настоящее время несколько групп экспериментаторов говорят о регистрации событий, которые можно интерпретировать как распад протона, однако решающей уверенности в этом пока нет, так как эти события могут быть вызваны фоном космического излучения, и сейчас самым веским аргументом в пользу нестабильности протона считается барионная асимметрия Вселенной. Развитие экспериментальной техники дает надежду продвинуться до  $\tau_p \sim 10^{34}$  лет, и если даже при этой чувствительности распад протона не будет обнаружен, космология останется нашим единственным орудием для исследования теорий большого объединения.

## § 2. НЕСТАЦИОНАРНОСТЬ ВСЕЛЕННОЙ И ЗАРЯДОВАЯ АСИММЕТРИЯ

Помимо несохранения барионов для возникновения барионной асимметрии Вселенной требуется еще нестационарность Вселенной и различие в свойствах частиц и античастиц. В от-

личие от гипотезы о несохранении  $B$  оба эти явления надежно установлены. Идея о возникновении барионного избытка за счет такого механизма была высказана более 15 лет тому назад А. Д. Сахаровым, а конкретные схемы ее воплощения широко обсуждаются в течение последних лет.

Очевидно, если барионный заряд не сохраняется, но взаимодействия частиц и античастиц абсолютно одинаковы, т. е., как говорят, имеет место зарядовая, или  $C$ -инвариантность, то никакого избытка барионов над антибарионами возникнуть не может. Более того, если плазма находится в термодинамически равновесном состоянии, то количество частиц и античастиц в ней должно быть одинаково независимо от  $C$ -инвариантности. Это утверждение является следствием  $CPT$ -теоремы, которая утверждает, что теория не меняется при совместном пространственном отражении, обращении времени и переходе от частиц к античастицам.

$CPT$ -инвариантность является следствием общих свойств теории, таких, как лоренц-инвариантность и аналитичность. В силу  $CPT$ -теоремы должны быть равны массы, времена жизни и статистические веса частицы и античастицы. Это и приводит к равному содержанию частиц и античастиц в термодинамически равновесной плазме, так как равновесная функция распределения определяется только массой и статистическим весом частицы. Нужно, впрочем, сделать одну оговорку: если частицы обладают каким-то сохраняющимся зарядом (например, электрическим) и средняя плотность этого заряда отлична от нуля, то в любом состоянии будет некоторый избыток (или недостаток) частиц над античастицами, отвечающий этому заряду\*.

Барионный заряд, мы полагаем, не сохраняется, поэтому при произвольном начальном его значении первичная плазма должна прийти в состояние с  $B=0$ , если в ней устанавливается термодинамическое равновесие. Однако Вселенная нестационарна: она расширяется, поэтому в ней возможно некоторое отклонение от равновесия, определяемое соотношением между темпом расширения мира  $H = \dot{a}/a$  и характерной скоростью реакций, за счет которых устанавливается равновесие.

Интересно отметить, что для частиц с массой, равной нулю, расширение Вселенной не приводит к отклонению от термодинамически равновесного состояния. Мы хорошо знаем это на примере реликтового электромагнитного излучения. Действительно, взаимодействие трехградусных фотонов между собой и с веществом во Вселенной пренебрежимо мало, а их спектр тем не менее имеет равновесный планковский вид с адиабатически

\* Для описания равновесных функций распределения в этом случае привлекают понятие химического потенциала.

надающей температурой. Это справедливо для любых безмассовых частиц. чтобы убедиться в этом, рассмотрим кинетические уравнения в расширяющемся мире. Они стандартным образом получаются из уравнений кинетики в плоском пространстве заменой обычных производных на ковариантные. Для метрики (2.1) нетрудно получить

$$\frac{\partial n_i}{\partial t} = H p_i \frac{\partial n_i}{\partial p_i} + S(n_i), \quad (8.2)$$

где  $n_i(p_i, t)$  — плотность частиц сорта  $i$  с импульсом  $p_i$ ,  $H$  — параметр Хаббла, а  $S$  — интеграл столкновений, зависящий от плотностей всех частиц, принимающих участие в реакциях с частицами  $i$ .

Нетрудно убедиться, что для безмассовых частиц уравнение удовлетворяется стационарными функциями распределения. Проверим это для простоты в случае статистики Больцмана, когда  $n_i = n_{eq} = \exp(-E/T)$ , оставляя случаи ферми- и бозе-статистик, а также случай ненулевого химического потенциала для самостоятельного упражнения. Проведя дифференцирование в левой части, получим

$$\frac{\partial n_i}{\partial t} = e^{-E/T} \frac{TE}{T^2} = -e^{-E/T} \frac{HE}{T}, \quad (8.3)$$

где мы воспользовались уравнением падения температуры при расширении:  $\dot{T} = -HT$ . Теперь, если учесть, что для безмассовых частиц  $p = E$  и что интеграл столкновений обращается в нуль при  $n = n_{eq}$ , то мы увидим, что равновесные функции распределения с зависящей от времени температурой удовлетворяют кинетическому уравнению (8.2) в нестационарном случае. Отсюда следует, что для безмассовых частиц отклонений от равновесия не возникает, несмотря на расширение Вселенной, а для массивных — отклонение пропорционально  $(m/T)^2$  (если  $m \ll T$ ). Если  $m > T$ , то отклонения от равновесия могут быть весьма велики в относительных единицах, но при этом концентрация самих массивных частиц будет мала (из-за фактора  $\exp(-m/T)$ ) и поэтому их роль будет малозаметна. Из сказанного следует, что существенным периодом для возникновения барионной асимметрии является период, когда температура первичной плазмы была порядка масс  $X$ -бозонов, т. е.  $10^{15}$  ГэВ, или  $10^{28}$  К.

Поэтому поведение барионного заряда в такой картине должно выглядеть следующим образом. В начальный момент при  $T = T_{p1} = 10^{32}$  К значение  $B$  могло быть произвольным, зависящим от механизма «сотворения мира». Затем по мере достижения термодинамического равновесия в первичной плазме это начальное значение должно исчезать с точностью до величины

$\sim (m_X/T)^2$ . На первый взгляд это уже довольно много, так как согласно наблюдениям отношение числа барионов к числу фотонов составляет  $N_B/N_\gamma = 10^{-9} - 10^{-10}$  (8.1), а отношение  $(m_X/T)^2 > 10^{-8}$ . Однако инфляция приводит к тому, что плотность барионного заряда перед началом бариосинтеза оказывается экспоненциально малой, если только не существует каких-то скалярных конденсатов с ненулевым барионным зарядом. Кроме того, не нужно забывать, что в начальный момент и почти вплоть до  $T \sim m_X$   $X$ -бозоны остаются безмассовыми и обретают массу лишь при температурах несколько выше  $m_X$ . Это явление тесно связано с восстановлением симметрии в теории поля при высоких температурах, отмеченным Киржницем и Линде (1972). Масса векторных бозонов, как мы обсуждали выше, появляется из-за выпадения конденсата хиггсовского поля, который может образоваться лишь при достаточно низких температурах. Это полностью аналогично восстановлению симметрии в ферромагнетике при его нагревании. Как мы знаем, в ферромагнетике при низких температурах образуется спонтанная намагниченность и, таким образом, симметрия относительно вращений теряется — появляется выделенное направление. При нагревании ферромагнетика выше точки Кюри намагниченность разрушается тепловой энергией, а симметрия восстанавливается. Точно такая же картина имеет место в теории поля, где роль намагниченности (параметра порядка) играет конденсат хиггсовского поля  $\langle \phi \rangle$ . При нулевой или достаточно малой температуре энергетически более выгодно состояние с  $\langle \phi \rangle \neq 0$ . При этом образуется выделенное направление в пространстве внутренних степеней свободы и симметрия в этом пространстве спонтанно нарушается, а масса векторных калибровочных бозонов пропорциональна  $\langle \phi \rangle$ .

В силу сказанного, даже если первоначальная зарядовая асимметрия существовала, она должна была уменьшиться практически до нуля в период, когда  $\langle \phi \rangle = 0$ . Снова возникнуть она может лишь ниже точки Кюри для хиггсовского поля после возникновения ненулевой массы  $X$ -бозонов, когда функция распределения последних будет отличаться от равновесной. Очевидно, степень отклонения от равновесия определяется соотношением между темпом расширения мира  $H = \dot{a}/a$  и скоростью реакций между частицами  $\Gamma$ . В соответствии с изложенным выше барионная асимметрия должна быть пропорциональна отношению  $H/\Gamma$  в момент, когда температура первичной плазмы  $T \sim m_X$ . В простейшей модели это отношение равно  $0,1 - 0,01$ .

### § 3. БАРИОСИНТЕЗ И СОХРАНЯЮЩИЕСЯ ЗАРЯДЫ

Прежде чем переходить к конкретной схеме, заметим, что в стандартной  $SU(5)$ -модели предсказывается несохранение ба-

рионного  $B$  и лептонного  $L$  зарядов. Однако их разность  $B-L$  строго сохраняется. Поэтому в такой теории величина  $B-L$  определяется начальными условиями.

В случае квантового рождения Вселенной (или, как говорят, рождения из ничего) образующиеся состояния должны иметь квантовые числа вакуума: нулевую энергию (как мы отмечали выше), нулевой электрический и другие сохраняющиеся заряды. Если  $(B-L)$  строго сохраняется, то начальное значение  $(B-L)$  должно быть равно нулю. Заметим, что сохранение  $(B-L)$  не обязательно и имеется ряд моделей, в которых оно не имеет места. В этом случае начальные значения  $B$  и  $L$  могли бы быть произвольны, но естественно ожидать, что они очень малы. Дело в том, что мы думаем, что за рождение Вселенной ответственно гравитационное взаимодействие (квантовые флуктуации метрики), которое по всей видимости сохраняет как  $B$ , так и  $L$ . Поэтому значения  $B$  и  $L$ , возникшие в результате квантового рождения мира, должны быть близки к значениям  $B$  и  $L$  в исходном состоянии, которое, как мы полагаем, имеет квантовые числа вакуума.

Однако реальная оценка начальных значений  $B$  и  $L$  (и других зарядов) в таком подходе, разумеется, не очень надежна и вообще говоря, они могли бы быть и велики. Но тут еще одну свою полезную роль могла сыграть инфляция. Очевидно, как-то бы ни были  $B$  и  $L$  в начале, за время инфляции  $t_i$  они уменьшатся в  $\exp(Ht_i) \geq 10^{30}$  раз, т. е. никакого следа от них не останется. Бариосинтез, приводящий к избытку барионов над антибарионами, о котором идет речь в этой главе, должен происходить после окончания инфляции.

Этот же факт можно понять и по-другому, экстраполируя вспять по времени современное барио-несимметричное состояние. Определенный бариозаряд после инфляции соответствует тому, что к концу инфляции во Вселенной был небольшой «дополнительный» обычный вещества  $10^{-9} \epsilon_{\text{век}}$  к общей плотности конденсата скалярного поля. Экстраполируя назад по времени, заметим, что этот небольшой избыток вещества очень скоро будет доминировать, определяя величину тензора энергии-импульса. Действительно, плотность вещества с уравнением состояния  $p = -\epsilon_{\text{век}}$  не зависит от времени  $\epsilon_{\text{век}} = \text{const}$ , в то время как плотность обычного вещества быстро нарастает с увеличением красного смещения  $\epsilon \propto (1+z)^4$ . Красное смещение во время инфляции нарастает при уменьшении времени экспоненциально:  $1+z \sim e^{-Ht}$ . Поэтому сдвинувшись назад на промежуток времени  $\Delta t = 7 \cdot H^{-1}$ , увидим, что плотность обычного вещества уже в  $10^3$  раз больше плотности скалярного поля. Другими словами, инфляции в этот момент времени быть не могло. Противоречие в рассуждениях указывает на то, что бариозаряд нарабатывался в нашей Вселенной уже после инфляции.

#### § 4. НАРУШЕНИЕ ЗАРЯДОВОЙ ИНВАРИАНТНОСТИ

Прежде чем перейти к рассмотрению конкретной модели, обсудим вкратце нарушение  $C$ -инвариантности, т. е. различие в свойствах частиц и античастиц. Как мы уже говорили, в квантовой теории поля строго доказывается  $CPT$ -теорема. Однако нигде не следует, что должна иметь место инвариантность при действии всех этих преобразований отдельно. Действительно, опыт показал, что зеркальная симметрия в слабых взаимодействиях максимально нарушается, — это известное явление несохранения четности. Сначала казалось, что хотя  $P$ -инвариантность нарушена, но если одновременно с отражением в зеркале перейти от частиц к античастицам (так называемое  $CP$ -преобразование), то симметрия сохранится. Однако выяснилось, что и эта симметрия является лишь приближенной, и именно одновременное нарушение  $C$  и  $CP$  делает возможным обсуждаемый ниже механизм.

$CPT$ -инвариантность говорит, что массы частиц и античастиц должны быть строго равны. Если частицы нестабильны, то должны совпадать также их полные времена жизни. Однако если имеется несколько способов (каналов) распада, то в силу нарушения  $CP$  вероятности распадов в соответствующие каналы для частиц и античастиц будут различны\*. В интересующем нас случае распада сверхтяжелых  $X$ - или  $Y$ -бозонов будут различаться вероятности распадов по каналам:

$$\Gamma(X \rightarrow qq) - \Gamma(\bar{X} \rightarrow \bar{q}\bar{q}) = \Delta\Gamma,$$

$$\Gamma(X \rightarrow \bar{q}l) - \Gamma(\bar{X} \rightarrow ql) = -\Delta\Gamma,$$

но  $\Gamma(X \rightarrow qq) + \Gamma(X \rightarrow \bar{q}l) = \Gamma(\bar{X} \rightarrow \bar{q}\bar{q}) + \Gamma(\bar{X} \rightarrow ql)$ . Так как  $\Delta\Gamma \neq 0$ , то в результате распада равного количества  $X$ - и  $\bar{X}$ -бозонов возникает неравное количество кварков и антикварков, т. е. ненулевая плотность барионного заряда.

Различие  $\Delta\Gamma$  мало, оно возникает, как можно показать, лишь во втором порядке теории возмущений, а в простейшей  $SU(5)$ -схеме имеется еще дополнительное подавление, связанное с высокой симметрией модели. Поэтому предсказываемая теорией величина  $N_B/N_T$  оказывается заметно меньше наблюдаемой. Здесь как и для распада протонов, предсказания простейшей  $SU(5)$ -модели расходятся с опытом. Однако эти расхождения можно ликвидировать за счет некоторого усложнения модели, грубо говоря, с помощью введения дополнительных частиц.

\* Явление близкого типа непосредственно наблюдалось на опыте. Вероятности распадов  $K_L^0$ -мезона по каналу  $K_L^0 \rightarrow \pi^- \mu^+ \nu$  и по зарядово-сопряженному каналу  $K_L^0 \rightarrow \pi^+ \mu^- \bar{\nu}$  оказались различны  $W(K_L^0 \rightarrow \pi^+ \mu^- \bar{\nu})/W(K_L^0 \rightarrow \pi^- \mu^+ \nu) \approx 0,997$ .

### § 5. БАРИОННАЯ АСИММЕТРИЯ В МОДЕЛИ БОЛЬШОГО ОБЪЕДИНЕНИЯ.

#### ТОПОЛОГИЧЕСКОЕ НЕСОХРАНЕНИЕ БАРИОНОВ

Возникновение барионной асимметрии в такой модели выглядит следующим образом. При  $T \approx m_X \approx 10^{14} - 10^{15}$  ГэВ распад  $X$ -бозонов будет несколько отличаться от равновесного. Согласно приведенным выше рассуждениям это отличие будет порядка  $H/\Gamma$ , где в данном случае  $\Gamma \approx \alpha m_X \approx m_X/40$  — распадная ширина  $X$ -бозона ( $\Gamma$  обратно пропорциональна периоду полураспада). Нетрудно оценить, что отношение  $H/\Gamma$  довольно мало, порядка  $3 \cdot 10^{-2}$ , что приводит к дополнительному (помимо малости нарушения  $C$  и  $CP$ ) подавлению рождения барионов. В силу (небольшой) неравновесности  $X$ -бозонов процессы рождения барионов при их распадах уже не будут компенсироваться другими реакциями, как это было бы в равновесии, и возникнет небольшой избыток кварков над антикварками. В дальнейшем при малых  $T$ , когда  $B$  будет практически сохраняться, почти все кварки и антикварки аннигилируют, кроме того нескомпенсированного избытка, который доживает до нашего времени, давая все наблюдаемое вещество.

В рассматриваемом здесь случае возникновение барионной асимметрии происходит при температурах  $T \approx 10^{14}$  ГэВ и за время порядка  $10^{-34}$  с. В последнее время в литературе обсуждаются модели, в которых барионная асимметрия возникает существенно позже при «низких» температурах  $\sim 10^3$  ГэВ. В этих моделях вероятность несохранения барионов не мала уже при  $T = T_0 \approx 10^3$  ГэВ, однако при малой температуре (или энергии  $E$ ) она подавлена экспоненциально  $W \sim \exp(-E/T_0)$ , в отличие от моделей большого объединения, где подавление степенное. Это несохранение связано с топологически нетривиальными решениями в теории поля или с конденсатом суперсимметричных частиц, и их обсуждение завело бы нас слишком далеко.

Только для ознакомления с терминологией укажем, что квантовые поправки приводят к несохранению некоторых токов, которые сохраняются на классическом уровне. Это явление называется квантовой аномалией. В частности, благодаря этому явлению не сохраняется барионный ток левых кварков (т. е. тех кварков, которые участвуют во взаимодействиях со слабыми  $W$ -бозонами). Аномальная часть в эффективном лагранжиане представляет собой полную дивергенцию и в обычной ситуации не должна проявляться. Однако в квантовой теории неабелевых калибровочных полей существуют решения, отвечающие не исчезающим на бесконечности потенциалам, но с нулевой напряженностью поля. Эти конфигурации полей нельзя устранить несингулярным калибровочным преобразованием.

так как они имеют нетривиальную топологическую структуру. Они не могут быть получены непрерывной деформацией близкого к 1 калибровочного преобразования. Такие решения получили название инстантонов [37]. Именно благодаря инстантонам проявляется аномальное несохранение барионного заряда. Однако это несохранение, на которое указал т'Офт, является экспоненциально малым ( $\sim \exp(-1/\alpha)$ ). Эффект может резко усиливаться в присутствии магнитного монополя (Рубаков, 1981; Каллан, 1982). Грубо говоря, протон вблизи магнитного монополя распадается с вероятностью единица. Монополю выступает как катализатор распада протона. Этот процесс, впрочем, по видимому, не влияет на возникновение барионной асимметрии, так как концентрация магнитных монополей должна быть невелика. Возможно, однако, что некоторые классические объекты типа монополя, возникающие в теории электрослабого взаимодействия, могут давать заметную барионную асимметрию (Кузьмин, Рубаков, Шапошников, 1985). Однако остается неясным вопрос о вероятности образования таких объектов при температуре  $T \approx T_0 \approx 10^3$  ГэВ.

Заметим еще, что в инфляционной модели можно получить примерно в 10—100 раз более высокое значение барионной асимметрии, чем в стандартной фридмановской космологии. Дело в том, что, как отмечалось выше, барионная асимметрия подавлена из-за малого отклонения  $X$ -бозонов от равновесия. В инфляционной модели исходное состояние может быть сильно неравновесным. Как мы знаем, к моменту завершения инфляции плотность частиц во Вселенной практически равна нулю, а вся энергия сосредоточена в когерентных колебаниях скалярного поля. Это осциллирующее скалярное поле рождает другие частицы, и в первую очередь сверхтяжелые бозоны большого объединения (векторные и скалярные). Плотность рожденных частиц далека от равновесной и поэтому термодинамического подавления барионной асимметрии не возникает (Долгов, Линде, 1981).

### § 6. БАРИОННАЯ АСИММЕТРИЯ И ЧЕРНЫЕ ДЫРЫ

В заключение рассмотрим еще одну возможность генерации барионной асимметрии, которая работает и в том случае, когда сохраняется барионный заряд (Хокинг, 1974; Зельдович, 1976; Долгов, 1980). Этот механизм связан с известным явлением квантового испарения черных дыр. Если пренебречь взаимодействием испаряемых частиц, то количество частиц и античастиц, рождаемых черной дырой, должно быть одинаково, но при учете взаимодействия и нарушения зарядовой инвариантности это не так. Рассмотрим простой пример. Пусть черная дыра испаряет тяжелые нестабильные частицы  $A$  и пусть относительные

вероятности сохраняющих  $B$  распадов  $A \rightarrow H + L$  и  $A \rightarrow \bar{H} + \bar{L}$  различны из-за наличия, как отмечалось, выше, других возможных каналов. Здесь  $H(H)$  — тяжелый барион, а  $L(L)$  — легкий барион (или соответственно антибарион). Вероятность обратного захвата черной дырой тяжелого бариона  $H$  больше, чем легкого  $L$ , поэтому в результате такого процесса во внешнем пространстве мог бы накапливаться барионный заряд, а антибарионы были бы законсервированы в (исчезающей) черной дыре. Иными словами, гравитация приводила бы к эффективному несохранению барионов при сохранении барионного заряда на микроскопическом уровне. В принципе при специальном подборе параметров таким образом можно получить необходимое значение  $N_B/N_T = 10^{-9} - 10^{-10}$ , однако модели с несохранением барионного заряда, несомненно, эстетически более привлекательны, и лишь отсутствие распада протона мешаем нам безоговорочно поверить в них.

Отметим сходство обсуждаемого здесь механизма бариосинтеза с тем, что рассматривался в конце предыдущего параграфа. Оба варианта связаны с классическими топологически нетривиальными решениями: инстантонами и монополями в первом случае и черными дырами во втором. И том и в другом случае барионный заряд сохраняется в рамках теории возмущений и все несохранение было связано с чисто непертурбативными эффектами (очевидно, выражение типа  $\exp(-1/\alpha)$  нельзя получить ни в каком порядке разложения по  $\alpha$ , и даже суммируя весь ряд теории возмущений).

Завершая эту главу, вспомним, как за последние 20 лет изменилась наша точка зрения на стабильность протона. Если раньше говорили: «наше существование убеждает нас в том, что протон стабилен», то теперь — прямо противоположное: «наше существование убеждает нас в том, что протон нестабилен», имея в виду в последнем случае, что при сохранении барионного заряда пригодная для жизни Вселенная могла и возникнуть. Добавим, впрочем, что вероятность несохранения  $B$ , значительная при ивысоких энергиях, должна быть, как видно из опыта, достаточно мала при «комнатной» температуре и низкой энергии.

## Глава 9.

### КВАНТОВОЕ РОЖДЕНИЕ ВСЕЛЕННОЙ

#### § 1. РОЖДЕНИЕ ИЗ «НИЧЕГО» И ЗАКОН СОХРАНЕНИЯ ЭНЕРГИИ

Рассматривая очень ранние эпохи эволюции Вселенной, вплоть до  $t = t_{PI} = 10^{-43}$  с, мы вправе задать вопрос о том, что было до этого и как родилась наша Вселенная в целом. Классическая космология при  $t \rightarrow 0$  приводила к сингулярностям, т. е. к бесконечным значениям плотности энергии, давления и нулевым значениям масштабного фактора и сопутствующего объема. Не имея количественной теории для описания процессов в сингулярности, космологи тем не менее понимали ограниченность классической космологии. Для описания Вселенной при  $t < t_{PI}$  употреблялся термин «сингулярность» зачастую без точной конкретизации его смысла. Иногда добавляли, что при  $t < t_{PI}$  классическая теория гравитации (ОТО) неприменима и возникающие проблемы надо решать в рамках квантовой теории гравитации.

Раньше наибольшей популярностью пользовалась точка зрения, согласно которой наш цикл расширения явился продолжением предыдущего цикла сжатия. Существовала модель пульсирующей Вселенной. Однако эта картина не избавляла от вопроса о происхождении Вселенной, она просто отодвигала его на несколько циклов раньше. Дело в том, что во время каждого цикла должна возрастать энтропия Вселенной и в результате увеличиваться ее радиус в точке максимального расширения. Поэтому, очевидно, бесконечного числа циклов быть не могло, если только не предполагать, что на стадии сжатия энтропия убывает или же, что при плотности порядка планковской происходит коренная ломка фундаментальных физических законов (М. А. Марков). С гипотезой об убывании энтропии недавно выступил С. Хокинг. Этот фантастический результат естественно возникает в развиваемой им модели, в которой время определено на компактном многообразии (т. е. не на бесконечной прямой, а на окружности), однако эта идея была подвергнута убедительной критике.

Теперь, в связи с развитием квантовой теории гравитационного поля, появилась возможность рассматривать, по крайней мере качественно, стадии эволюции Вселенной при  $t = t_{PI}$  и даже раньше. Некоторой модификацией модели пульсирующей Вселенной явилась модель «отскока» от сингулярности. Суть ее заключалась в том, что вблизи сингулярности фридмановский режим сжатия вида  $a(t) = a_0 t^\alpha$  менялся на деситтеровский режим вида  $a(t) = a_0 \operatorname{ch} Ht$ , и в момент  $t = 0$  масштабный фактор

становился хотя и очень малым,  $a_0 \sim a_{P1}$ , но не нулевым. Соответственно этому конечным оставались и все физические величины, плотность и давление. Формально такая модель получалась с помощью квантового обобщения уравнений Эйнштейна. К обычному действию ОТО

$$S = \int d^4x \sqrt{-g} R$$

добавлялись нелинейные по тензору кривизны дополнительные члены вида

$$\int d^4x \sqrt{-g} (a_1 R_{\mu\nu} R^{\mu\nu} + a_2 R^2 + a_3 \square R)$$

и т. п. Коэффициенты  $a_1, a_2, a_3$  все были порядка  $l_{P1}^2 = m_{P1}^{-2}$ . Это так называемые поляризационные добавки к гравитационному действию (см. § 2 гл. 7).

По физическому смыслу эти добавки аналогичны поляризации вакуума электрическим зарядом. Поляризация вакуума в атоме приводит к наблюдаемым явлениям — эффекту Лэмба—Резерфорда, обсуждавшемуся нами ранее. Наличие гравитационных поправок экспериментом не доказано, но урок электродинамики заставляет верить в их реальность.

Для нас сейчас важно то, что уравнения с поляризационными поправками могли дать решение де Ситтера  $a(t) = a_0 \operatorname{ch} Ht$  и, следовательно, помогали избежать сингулярности. Вопросы, связанные с поляризацией гравитационного вакуума, мы обсудили в предыдущем параграфе, модель с «отскоком» рассмотрим в конце этого.

Сейчас, однако, наиболее привлекательной выглядит идея квантового рождения мира, или рождения из «ничего». Под этим подразумевается, что наш мир в целом может возникнуть без нарушения основных физических законов, акт его рождения будет описываться законами квантовой гравитации, а для его рождения не потребуется никакой энергии. Вероятность рождения мира тоже должна быть достаточно высокой.

Прежде всего подчеркнем, что возникновение Вселенной не должно противоречить закону сохранения энергии. Закон сохранения энергии имеет два аспекта: локальный и глобальный. Локально закон сохранения энергии, в термодинамике например, записывается в виде

$$dE = -pdV$$

для каждого локального элемента объема  $dV$ . Каждый элемент объема обладает определенной энергией, которая больше нуля и которая может изменяться при взаимодействии с другими локальными объемами. Специальная теория относительности,

связав массу с энергией, научила нас, что энергия у вещества не может обратиться в нуль — останется по крайней мере энергия покоя. Ситуация изменилась после построения общей теории относительности. Было показано, что в ней энергия может уменьшаться на величину гравитационного дефекта масс. Идея о том, что определенное количество вещества может иметь нулевую энергию, была осознана только в 60-е годы, через полстолетия после создания ОТО. Точно так же может быть Вселенная с нулевой энергией. Она представляет собой замкнутый мир. При рождении такого мира из ничего не нарушается самый важный из физических законов — закон сохранения энергии. Энергия до рождения этого мира и после его рождения равна нулю. Но до рождения — вещества нет (как, впрочем, нет и пространства в нашем смысле понимания). После рождения замкнутого мира появляется вещество, но оно так сконцентрировано, что гравитационный дефект массы полностью уравнивает саму массу.

Рассмотрим подробнее глобальный закон сохранения энергии в ОТО и понятие гравитационного дефекта массы. Понятие дефекта массы существует не только для гравитационного поля, а для любых типов взаимодействия. Впервые оно появилось вместе с утверждением СТО об эквивалентности массы и энергии. Поэтому масса связанного состояния двух частиц должна быть меньше сумм масс этих частиц, так как она включает в себя отрицательную энергию связи. Это явление установлено экспериментально для атомных ядер. Например, известно, что ядро дейтона составлено из нейтрона и протона. Однако масса дейтона меньше, чем сумма масс одиночного протона и одиночного нейтрона:

$$m_D = m_N + m_p - \Delta E.$$

Ядерный дефект масс  $\Delta E = 2,2$  МэВ обусловлен ядерным взаимодействием между протоном и нейтроном.

Точно так же дефект масс появляется, например, в двойной звезде. Пусть у нас есть две звезды с массами  $M_1$  и  $M_2$ . Если они входят в двойную систему, то общая масса системы меньше суммы масс на гравитационный дефект массы.\*

$$M_{\text{дв.зв}} = M_1 + M_2 - \frac{GM_1 M_2}{r}.$$

В этой формуле  $U = GM_1 M_2 / r$  — потенциальная энергия взаимодействия двух звезд, которая и дает гравитационный дефект масс.

\* Надо учесть еще кинетическую энергию звезд:

$$E_{1,2} = \frac{1}{2} M_{1,2} / \sqrt{1 - v_{1,2}^2}.$$

Гравитационный дефект масс существует для многих наблюдаемых параметров нейтронных звезд. Масса звезды вычисляется согласно выражению

$$M = \int \rho(r) \cdot 4\pi r^2 dr.$$

На первый взгляд здесь нет гравитационного дефекта. Но дело в том, что в сильном гравитационном поле элемент объема изменяется. Объем сферического слоя толщиной  $dr$  в плоском пространстве записывается как произведение площади поверхности этого слоя  $4\pi r^2$  на его толщину, т. е.  $dV = 4\pi r^2 dr$ . В сильном гравитационном поле площадь поверхности остается прежней  $4\pi r^2$ , но толщина сферического слоя определяется уже по-другому:

$$dl = dr / \sqrt{1 - r_g/r},$$

и элемент объема будет иметь вид, отличный от  $dV$  в плоском пространстве

$$dV = \frac{4\pi r^2 dr}{\sqrt{1 - r_g/r}}.$$

Теперь, если выражение для массы нейтронной звезды записать через физические характеристики: плотность и элемент объема, то

$$M = \int \rho(r) \sqrt{1 - r_g/r} dV;$$

здесь  $r_g$  — гравитационный радиус нейтронной звезды. Множитель  $\sqrt{1 - r_g/r}$  и определяет гравитационный дефект масс. Таким образом, масса нейтронной звезды меньше суммы масс составляющих ее частиц.

Для нас во всех этих примерах главным является то, что масса тела не равна сумме масс составляющих его частей, а меньше на некоторую величину. Рассмотрим теперь шар с постоянной плотностью  $\rho$ . Его массу можно записать в виде

$$M = \rho V - \int \rho \phi dV.$$

Второй член в этом выражении — уже знакомый нам гравитационный дефект масс. Записывая массу как функцию радиуса тела, приходим к формуле

$$M = \frac{4\pi}{3} \rho r^3 - \frac{16}{15} \pi^2 G \rho^2 r^5.$$

Первый член в выражении для общей массы представляет собой сумму масс частиц, составляющих это тело. Сумма растет пропорционально плотности тела и пропорционально кубу его размеров. Второй член — это гравитационный дефект массы

тела. При увеличении плотности тела он растет как квадрат плотности; при увеличении размеров тела — пропорционально пятой степени его радиуса.

Важно то, что при постоянной плотности можно получить общую массу, равную нулю. Действительно, будем добавлять все больше и больше вещества к этому телу, сохраняя его плотность постоянной. Гравитационный дефект масс будет расти быстрее, чем сумма масс, составляющих тело, и в конце концов общую массу  $M$  можно обратить в нуль. Это происходит при достижении телом размеров

$$r = \sqrt{\frac{5}{4\pi G \rho}}.$$

Поскольку расчет велся в рамках ньютоновской теории гравитации и не было учтено изменение метрики при приближении массы к нулю, то конкретный коэффициент в выражении, связывающий  $r$  и  $\rho$ , неверен, но по порядку величины выражение справедливо. Точно так же справедлив и основной вывод: добавляя вещество можно обратить общую массу в ноль.

В другом случае можно взять некоторое определенное количество вещества и его сжимать — увеличивать плотность (увеличивать плотность надо так, чтобы произведение  $\rho r^2$  росло). Тогда мы сможем сжать его до такой степени, что  $M$  будет как угодно близка к нулю.

Разумеется, при этом нет никакого противоречия с законом сохранения энергии. Уменьшение массы тела сопровождается соответствующим дефекту массы излучением.

## § 2. СВОЙСТВА ПОЛУЗАМКНУТОГО МИРА

Рассмотренные примеры приводят к мысли о естественности того, что у замкнутого мира полная масса или эквивалентно энергия равна нулю. Интересно представить себе, как при добавлении массы может образоваться замкнутый мир. Поэтому будем рассматривать, как изменяется масса, измеряемая внешним наблюдателем, при добавлении к сферически-симметричному телу вещества при учете эффектов ОТО.

Прежде всего напомним некоторые свойства замкнутого мира и сферически-симметричного тела. Замкнутый мир Фридмана описывается метрикой

$$ds^2 = a^2(\eta) (d\eta^2 - d\chi^2 - \sin^2 \chi d\Omega^2),$$

$$d\Omega^2 = \sin^2 \theta d\varphi^2 + d\theta^2; \quad (9.1)$$

гиперсферический угол  $\chi$  описывает расстояние от центра системы координат до некоторой точки. Он изменяется от нуля



до  $\pi$ . Эта метрика описывает однородный и изотропный мир. За начало отсчета может быть принята любая точка.

Геометрия, отвечающая метрике (9.1), является геометрией на поверхности трехмерной сферы в некотором воображаемом 4-мерном пространстве. Введем фиктивную координату  $\mathcal{V}$ , как при описании топологических свойств мира де Ситтера. Поверхность сферы задается уравнением

$$\mathcal{V}^2 + x^2 + y^2 + z^2 = a^2. \quad (9.2)$$

Метрика в 4-мерном пространстве  $\mathcal{V}$ ,  $x$ ,  $y$ ,  $z$  имеет обычный евклидов вид

$$dl^2 = d\mathcal{V}^2 + dx^2 + dy^2 + dz^2.$$

На поверхности сферы (9.2)

$$dl^2 = a^2 (d\chi^2 + \sin^2 \chi d\Omega^2).$$

Если угол  $\chi$  меняется от нуля до  $\chi_0 < \pi$ , то метрика (9.1) описывает часть сферы или часть мира Фридмана. При этом если  $\pi/2 < \chi_0 < \pi$ , то такой случай обычно называют полузамкнутым миром Фридмана и соответствующая метрика описывает более половины замкнутого мира. Часть мира Фридмана уже, разумеется, не изотропна, а только сферически-симметрична с центром в точке  $\chi = 0$ .

Метрика вблизи сферически-симметричного тела имеет вид

$$ds^2 = \left(1 - \frac{r_g}{r}\right) dt^2 - \left(1 - \frac{r_g}{r}\right)^{-1} dr^2 - r^2 d\Omega^2; \quad (9.3)$$

здесь  $r_g$  — гравитационный радиус тела,

$$r_g = 2Gm = 2G \int_0^{r_0} \rho(r) \cdot 4\pi r^2 dr. \quad (9.4)$$

В формуле, выражающей связь гравитационного радиуса и массы тела, масса записана в виде объемного интеграла от плотности тела для того, чтобы естественным образом продолжить метрику внутрь тела. Для однородной плотности метрику внутри системы можно выразить как

$$ds^2 = (1 - 8\pi G \rho^2 r^2 / 3) dt^2 - (1 - 8\pi G \rho r^2 / 3)^{-1} dr^2 - r^2 d\Omega^2.$$

Будем рассматривать теперь только трехмерную часть полного интервала — часть, определяющую пространственную геометрию:

$$dl^2 = (1 - 8\pi G \rho r^2 / 3)^{-1} dr^2 + r^2 d\Omega^2. \quad (9.5)$$

Покажем теперь, что эта геометрия трехмерной гиперповерхности совпадает с геометрией замкнутого мира Фридмана, т. е. является сечением замкнутого мира в момент максимального

расширения. Введем величину  $a$ , имеющую размерность длины, согласно

$$a^2 = 3 / (8\pi G \rho),$$

а радиальную координату  $r$  выразим через гиперсферический угол  $\chi$  как

$$r = a \sin \chi.$$

Тогда метрику (9.5) можно свести к виду

$$dl^2 = a^2 (d\chi^2 + \sin^2 \chi d\Omega^2),$$

который очень похож на геометрию замкнутого трехмерного мира. Отличие его от геометрии замкнутого мира заключается в том, что у замкнутого мира угол  $\chi$  изменяется от нуля до  $\chi_0 = \pi$ , а в нашем случае  $\chi_0 < \pi$ .

Вернемся к наглядному изображению метрики как геометрии на некоторой поверхности, вложенной в фиктивное 4-мерное пространство  $\mathcal{V}$ ,  $x$ ,  $y$ ,  $z$ . В пустоте, в области решения Шварцшильда, поверхность описывается уравнением

$$1 + (d\mathcal{V}/dr)^2 = (1 - r_g/r)^{-1},$$

решение которого

$$\mathcal{V}(r) = 2\sqrt{r_g(r - r_g)}. \quad (9.6)$$

Внутри материи поверхность представляет собой часть сферы с уравнением

$$\mathcal{V}^2 + r^2 = a^2. \quad (9.7)$$

Представим себе полную геометрию как сшивку двух поверхностей вращения в евклидовом пространстве: одна поверхность (9.6) соответствует метрике в пустоте, другая (9.7) соответствует метрике внутри тела.

Кривая  $\mathcal{V}(r)$ , которая получается для малой части мира Фридмана  $\chi_0 < \pi/2$ , склеенная с решением Шварцшильда, представлена на рис. 41.

Отрезок  $OA$  соответствует части фридмановского решения. При вращении дуги  $OA$  вокруг оси  $O\mathcal{V}$  получается часть сферы. Метрике Шварцшильда соответствует кривая  $AB$ . Эта кривая, будучи продолжена вплоть до оси  $Or$  (пунктир), пересекает ось с вертикальной касательной при  $r = r_g$ . При склейке в  $A$ , однако, продолжение кривой  $AB$  не используется. Чем меньше угол  $\chi_0$ , тем меньше искривление поверхности вращения  $OAB$ .

Наоборот, чем больше  $\chi_0$ , тем более искривлена поверхность  $OAB$  и тем больше искривление реального пространства. При  $\chi_0 = \pi/2$  фридмановская часть полной метрики становится равной половине замкнутого мира. При этом к точке  $A$  кривые  $OA$  и  $AB$  подходят с вертикальными касательными. Для внешнего

наблюдателя это значит, что в центре системы координат — точке  $O$  — находится черная дыра с гравитационным радиусом, равным  $a$  — радиусу полусферы.

Поверхность, эквивалентная полузамкнутому миру, показана на рис. 42. На этом рисунке части мира Фрийдмана соответствует дуга  $OA$ . Обычное решение Шварцшильда с массой, рав-

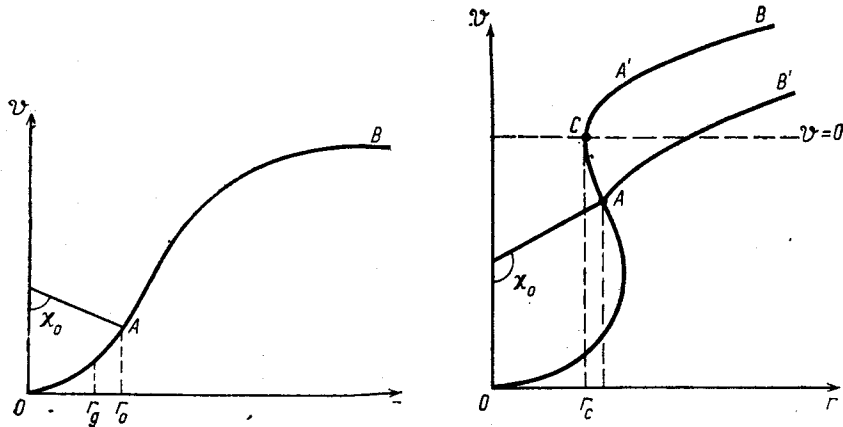


Рис. 41. Сшивка геометрии Фрийдмана с геометрией Шварцшильда

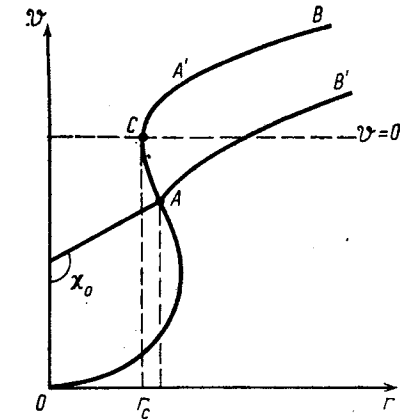


Рис. 42. То же, что на рис. 41, при увеличении кривизны мира Фрийдмана (полузамкнутый мир)

ной массе полузамкнутого мира, показано жирной линией  $AB$ . Однако сшить  $OA$  и  $AB$ , очевидно, невозможно, так как  $d\psi/dr$  будет неоднозначна в  $A$ . Правильный способ сшивки показан дугой  $OACAB$ .

Масса, измеряемая внешним наблюдателем по гравитационному полю в пустоте, определяется из выражения (9.4):

$$m = \int \rho(r) \cdot 4\pi r^2 dr.$$

Напомним, что здесь  $4\pi r^2 dr$  уже не есть элемент объема, который записывается в виде

$$dV = 4\pi r^2 (1 - r_g/r)^{-1/2} dr.$$

В полузамкнутом мире координата  $r$  с удалением от центра меняется немонотонно. На двумерном графике  $\psi(r)$  точка  $r$  — это некоторая точка на кривой  $OACA'B$ , описывающей поверхность вращения вокруг  $O\psi$ . На этой поверхности  $r = \text{const}$  уже окружность радиуса  $r = a \sin \chi$ . В трехмерной геометрии это — объем (трехмерная сфера).

Картина, показывающая  $r = \text{const}$  в виде кругов, изображена на рис. 43. При удалении от центральной точки

$x=y=0$  и увеличении гиперсферического угла  $\chi$  размер окружности сначала растет, затем достигает максимума при  $\chi = \pi/2$  и затем вновь уменьшается до значения  $r_c$  (рис. 42). Одновременно  $r_c$  совпадает с гравитационным радиусом тела, измеряемым внешним наблюдателем. Поскольку  $r_c < a$  при  $\chi_0 > \pi/2$ , то для внешнего наблюдателя полная масса полузамкнутого мира  $m$  меньше, чем сумма масс составляющих его элементов вещества.

Формально то же самое можно доказать, обратившись к выражению  $m$  через  $\rho$ :

$$m = \int \rho(r) \cdot 4\pi r^2 dr.$$

Поскольку  $r$  изменяется немонотонно, то удобно ввести параметр, обеспечивающий однозначность. Возьмем в качестве такого параметра расстояние  $q$ , отсчитанное от точки  $O$  по поверхности полузамкнутого мира, т. е. на рис. 42 длину кривой  $OA$ . В новых переменных масса есть

$$m = \int \rho(q) \cdot 4\pi r^2(q) \frac{dr}{dq} dq.$$

Добавляя вещество (строго говоря — пыль) во внешнее по отношению к полузамкнутому миру пространство, т. е.  $\chi > \chi_0$ , мы уменьшаем массу этого мира для внешнего наблюдателя, поскольку  $dr/dq < 0$  на дуге  $AC$ .

При постепенном увеличении объема полузамкнутого мира (увеличения  $\chi_0$  при данном  $a$ ) за счет добавлений пыли на границу  $A$  части мира Фрийдмана и Шварцшильда масса уменьшается: гравитационный дефект массы добавленной пыли больше ее массы покоя. По мере уменьшения массы  $m$  уменьшается и ее гравитационный радиус  $r_c$ , т. е. уменьшается поверхность, соединяющая полузамкнутый мир с внешним пространством. Напомним, что на этой поверхности есть шварцшильдовская особенность. Таким образом, можно непрерывно прийти к известному факту равенства нулю полной массы замкнутого мира.

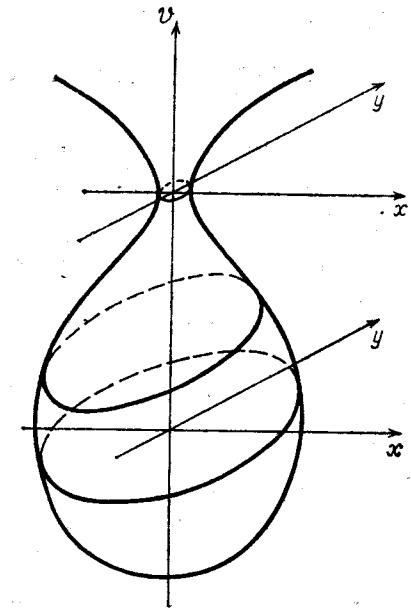


Рис. 43. Почти полное замыкание мира Фрийдмана. Его масса для внешнего наблюдателя может быть произвольно малой

Поскольку для замкнутого мира полная масса или эквивалентно, энергия равна нулю, то возможны переходы замкнутого мира с маленьким радиусом в замкнутый мир с большим радиусом. Самое главное, что такие переходы возможны без нарушения основного физического закона — закона сохранения энергии.

### § 3. МАССА ВЕЩЕСТВА РОДИВШЕГОСЯ МИРА, УВЕЛИЧЕНИЕ ЕЕ В ПРОЦЕССЕ ИНФЛЯЦИИ

Естественно ожидать, что Вселенная рождается со средним объемом порядка  $l_{Pl}^3$ . При этом радиус кривизны  $\sim l_{Pl}$  и все остальные параметры тоже имеют характерные планковские величины: плотность  $\rho_{Pl}$ , масса (без учета гравитационного дефекта масс)  $m_{Pl}$  и т. п. Поэтому, хотя полная масса родившегося замкнутого мира равна нулю, положительная часть этой полной массы — масса вещества — составляет  $\sim m_{Pl} = 10^{-5}$  г. С другой стороны, мы знаем, что сейчас масса вещества в наблюдаемой части Вселенной несравненно больше этой величины и составляет по порядку величины  $M \approx 10^{55}$  г  $\approx 10^{22} M_{\odot}$ . Естественно задаться вопросом, как могла увеличиться масса вещества.

Для этого вновь рассмотрим закон сохранения энергии, но записанный уже в локальной форме

$$dE = -pdV \quad (9.8)$$

или в тензорных обозначениях

$$T_{\mu,\nu}^{\nu} = 0.$$

Рассмотрим бесконечно малый элемент сопутствующего объема  $dV$ . В силу его малости можно пренебречь его внутренней гравитационной энергией и рассматривать только энергию вещества.

Закон сохранения энергии, записанный в дифференциальном виде

$$\frac{de}{dt} = -\text{div } \mathcal{J}, \quad (9.9)$$

где  $\mathcal{J}$  — плотность потока, описывает изменение плотности энергии в выбранном элементарном объеме. В ОТО дифференциальные законы сохранения имеют вид

$$T_{\mu;\nu}^{\nu} = 0. \quad (9.10)$$

Вместо плотности энергии, плотности импульса и потока им-

пульса в дифференциальном законе стоит единый комплекс — тензор энергии-импульса-натяжений. Этот тензор является источником в уравнениях гравитационного поля. Однако уравнения поля сами накладывают на него ограничения (9.10), которые, в частности, дают и закон сохранения энергии.

Напомним еще раз уравнения Максвелла:

$$\begin{aligned} \text{div } \mathbf{H} &= 0, & \text{div } \mathbf{E} &= 4\pi\rho, \\ \text{rot } \mathbf{E} &= -\frac{\partial \mathbf{H}}{\partial t}, & \text{rot } \mathbf{H} &= \frac{\partial \mathbf{E}}{\partial t} + 4\pi\mathbf{j}. \end{aligned}$$

Первая пара уравнения — уравнения связей, накладываемых на полевые переменные  $E$  и  $H$ . Вторая пара описывает связь  $E$  и  $H$  с источниками поля — плотностью зарядов и токов. Первая пара уравнений накладывает определенные ограничения на эволюцию  $\rho$  и  $\mathbf{j}$  — закон сохранения электрического заряда, который, будучи записанным в дифференциальной форме, имеет вид

$$\frac{\partial \rho}{\partial t} + \text{div } \mathbf{j} = 0.$$

Именно этот закон препятствует созданию электрического заряда из «ничего». Если мы захотим создать в какой-либо точке пространства плотность заряда, то должны подвести заряд к ней, т. е. создать ток в эту точку.

Точно такие же соображения препятствуют созданию массы из «ничего», и дифференциальный закон сохранения энергии в нерелятивистской физике (9.9) запрещает рождение массы из «ничего».

Однако в ОТО в случае замкнутой Вселенной гравитационный дефект полностью «съедает» массу вещества и полная масса тождественно равна нулю. При этом количество материи во Вселенной может меняться произвольно, не противореча ковариантному сохранению тензора энергии-импульса (9.10). Для нас важно, что в инфляционной модели это количество экспоненциально быстро растет.

Наиболее отчетливо можно проследить за изменением количества вещества, если перейти в систему координат, в которой  $\sqrt{-g}=1$ . В статическом гравитационном поле уравнение, выражающее рост плотности энергии в зависимости от притока импульса, останется таким же, как в нерелятивистской физике (9.9). Мы специально выбрали такую систему координат, чтобы тяготение в этом уравнении себя не проявляло. Это уравнение (9.10) при  $\mu=0$ . Для других уравнений сохранения этого уже сделать нельзя, и в правой части появится член, описывающий работу, совершаемую силами тяготения  $\nabla\phi$ :

$$\frac{\partial(\rho v)}{\partial t} = \nabla p + \rho \nabla \varphi.$$

Найдем в явном виде эту работу в однородной и изотропной Вселенной с метрикой Фридмана:

$$ds^2 = dt^2 - a^2(t)(dr^2 + \sin^2 r d\Omega^2).$$

Тензор энергии-импульса имеет вид

$$T_{\nu}^{\mu} = (\varepsilon + p)u^{\mu}u_{\nu} - p\delta_{\nu}^{\mu},$$

а уравнения сохранения

$$\frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^{\mu}} (\sqrt{-g} T_{\nu}^{\mu}) = \frac{1}{2} \frac{\partial g_{\alpha\beta}}{\partial x^{\nu}} T^{\alpha\beta}. \quad (9.11)$$

В однородной и изотропной Вселенной нет потоков импульса, а натяжения однородны.

Левая часть уравнения (9.11) представляет собой общерелятивистский аналог закона (9.9), а в правой части (9.11) стоит работа, совершаемая силами тяготения. Так как метрика не зависит от пространственных координат, то

$$\frac{1}{\sqrt{-g}} \frac{\partial(\sqrt{-g} T_0^{\mu})}{\partial x^{\mu}} = \frac{1}{\sqrt{-g}} \frac{\partial(\sqrt{-g} \varepsilon)}{\partial t} + \frac{\partial T_0^i}{\partial x^i}.$$

Запишем равенство (9.11) в интегральном виде, предварительно подставив в него явные выражения для  $T_{\mu}^{\nu}$  и метрики:

$$\frac{1}{a^3} \frac{\partial(a^3 \varepsilon)}{\partial t} + \operatorname{div} \mathcal{Y} = -3pH; \quad (9.12)$$

здесь  $H = \dot{a}/a$  — параметр Хаббла. Выделим элементарный лагранжев объем  $W$ , который позволяет следить за поведением отдельных частиц. Этот объем составлен из лагранжевых переменных, которые для каждой отдельной частицы неизменны со временем, т. е. нумеруют частицы. Малый элемент сопутствующего объема  $V$  связан с лагранжевым объемом  $W$  согласно

$$V = a^3 W,$$

где сам  $W$  равен

$$W = 4\pi \left( \frac{r}{2} - \frac{1}{4} \sin 2r \right).$$

При малых  $r$

$$W \approx \frac{4\pi}{3} r^3.$$

Количество энергии в объеме  $V$  есть

$$E = \varepsilon V = a^3 \varepsilon W, \quad (9.13)$$

а скорость изменения  $V$  со временем

$$\dot{V} = 3HV. \quad (9.14)$$

Умножим правую и левую части (9.12) на  $dW$  и проинтегрируем по лагранжеву объему. Первый член в левой части даст согласно (9.13),  $dE/dt$ , второй, согласно теореме Остроградского—Гаусса, будет иметь вид  $\oint \mathcal{Y} ds$  и обратится в нуль, поскольку нет потоков через границы объема. Член в правой части можно выразить через изменение физического объема согласно (9.14). Тогда уравнение сохранения (9.12) можно записать в эквивалентном виде  $\dot{E} = -p\dot{V}$  или в виде, совпадающем с (9.8), но теперь справедливость его мы проверили и для ОТО.

Отсюда видно, что, для того чтобы энергия вещества росла при расширении  $dV > 0$ , необходимо отрицательное давление  $p < 0$ . В случае уравнения состояния  $p = -\varepsilon$  плотность энергии постоянна и не зависит от изменения объема  $dV$ , как мы это уже видели в гл. 5.

Записывая (9.8) в виде

$$dE = \varepsilon dV,$$

видим, что  $E$  растет пропорционально  $V$ .

Таким образом, в период раздувания, т. е. когда  $a(t) = a_0 e^{Ht}$ ,  $\varepsilon = -p$ , стремительно нарастает масса вещества во Вселенной. Количество вещества растет так же, как растет объем мира.

#### § 4. ВОЗНИКНОВЕНИЕ ХЭББЛОВСКОГО РАСШИРЕНИЯ

Рассмотрим еще один важный вопрос. А именно, как получилось однородное и изотропное расширение вещества, которое мы сейчас называем хэббловским:

$$u = Hr.$$

Для этого рассмотрим шарик радиусом  $r$ , выделенный в расширяющемся мире. Уравнение сохранения энергии для него имеет вид

$$\frac{1}{2} \left( \frac{dr}{dt} \right)^2 = \frac{GM}{r} + \text{const}. \quad (9.15)$$

Уравнение движения получается дифференцированием этого уравнения по времени. В ньютоновской механике масса тела  $M$  остается постоянной и мы имеем

$$\ddot{r} = -\frac{GM}{r^2}.$$

В ОТО при переходе от закона сохранения энергии к уравнениям движения нельзя считать массу постоянной, нужно учесть

ее изменение в соответствии с  $dM = -pdV$ . Если есть давление, то оно совершает работу, а следовательно, изменяет массу.

Для того чтобы учесть это, подставим в (9.8)  $E \equiv M$ :

$$\frac{dM}{dt} = -p \frac{dV}{dt} = -4\pi r^2 p \frac{dr}{dt}.$$

После дифференцирования уравнения (9.15) по времени с учетом того, что масса  $M$  переменна, получаем уравнение движения в виде

$$\ddot{r} = -\frac{4\pi r^3}{3} (\rho + 3p) \frac{G}{r^2}.$$

Стоящая здесь масса  $M'$  уже отличается от  $M$ :

$$M' = M + 4\pi r^3 p, \text{ или}$$

$$M' = \frac{4\pi}{3} r^3 (\rho + 3p).$$

В среде с уравнением состояния  $p = -\epsilon$  масса  $M'$  становится отрицательной:  $M' = -8\pi r^3 \rho / 3$ , и гравитационное притяжение сменяется на отталкивание

$$\ddot{r} = \frac{8\pi G \rho}{3} r \equiv H_0^2 r.$$

Решение этого уравнения есть

$$r(t) = r_1 \operatorname{ch} H_0 t + r_2 \operatorname{sh} H_0 t,$$

где  $r_1$  и  $r_2$  — произвольные константы, определяющиеся разностью лагранжевых переменных двух частиц. Гравитационные силы отталкивания разгоняют частицы в инфляционный период, и дальше они движутся по инерции. Так получается хэббловское движение галактик с законом

$$u = Hr.$$

Следует четко понимать разницу между причиной взрыва в бомбе и причиной разлета вещества во время «Большого взрыва» в ранней Вселенной. В бомбе сила, ответственная за разлет частиц, вызвана градиентом давления внутри взрывчатого вещества. Во Вселенной с уравнением состояния  $p = -\epsilon$  вещество распределено однородно и градиента давления нет. Отрицательное давление эквивалентно силам натяжения в твердом теле. Поэтому если в обычном пространстве с уравнением  $p = 0$  образуется кусок вещества с отрицательным давлением, как на рис. 44, то на его границах возникает градиент давления. Появятся силы, стремящиеся схлопнуть этот кусок вещества. Силы отталкивания во Вселенной с уравнением  $p = -\epsilon$  возникают

из-за большой величины отрицательного давления, в результате чего меняется знак источника гравитационного поля ( $\rho + 3p$ ) и возникает эффективная антигравитация, т. е. разлетание вещества. Таким образом, толчком к расширению мира послужила антигравитация, вызванная отрицательным давлением, которое, как мы полагаем, существовало на очень ранних стадиях во Вселенной.

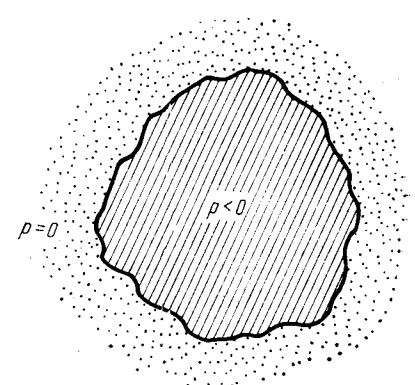


Рис. 44. Сжатие ограниченного куска вещества с отрицательным давлением

## § 5. КВАНТОВОЕ ОПИСАНИЕ РОЖДЕНИЯ ВСЕЛЕННОЙ

Вернемся к квантовому рождению мира. В исходном состоянии не существовало классического пространства-времени, а метрика  $g_{\mu\nu}$  была чисто квантовой величиной. Хорошую аналогию, позволяющую понять это явление, представляет процесс ядерного  $\alpha$ -распада. До распада понятие траектории  $\alpha$ -частицы не существовало, ее движение было квантовым. После туннельного перехода  $\alpha$ -частица вылетает из ядра, и с хорошей точностью можно говорить о ее классической траектории. Рождение мира мы понимаем сейчас в точности так же. Классическое пространство-время образуется в результате некоторого туннельного перехода, и только после этого можно говорить о времени и пространстве в нашем понимании этих слов и о возникновении мира. Разумеется, нужно помнить, что теория рождения мира из «ничего» настолько еще плохо разработана, что проводимые ниже соображения следует рассматривать лишь как иллюстративные.

Основная идея, заложенная в описании возникновения мира из ничего, заключается в том, что вероятность рождения именно замкнутого мира есть величина не нулевая. В традиционной квантовой механике вероятность классически запрещенного процесса (туннельного перехода) представляет собой величину

$$W = W_0 \exp(-S);$$

здесь  $S$  — действие, а  $W$  — некоторый предэкспоненциальный множитель, который получается из точного рассмотрения задачи. Основной фактор, качественно описывающий ситуацию,

есть  $\exp(-S)$ . Действие  $S$  вычисляется на решениях классических уравнений движения  $m\ddot{x}+u'(x)=0$ , в которых сделан переход к мнимому времени:  $t \rightarrow it$ . Можно показать, что именно таким образом получаются формулы квазиклассического приближения.

В квантовой теории поля квазиклассическая вероятность туннельного перехода вычисляется точно так же. Более того, квантовую теорию поля обычно с «самого начала» описывают в мнимом времени, так как при такой (так называемой эвклидовой) формулировке удается корректно добиться сходимости лежащих в основе теории функциональных интегралов вида

$$Z = \int D\varphi \exp(-S),$$

где  $D\varphi$  обозначает интегрирование по всем состояниям квантовых полей. Если в задаче рассматривается гравитационное поле, то интегрирование проводят по всем состояниям метрик с сигнатурой  $(+, +, +, +)$ . Вероятность того или иного процесса определяется функционалом  $Z$ , причем квазиклассическое приближение отвечает оценке этого интеграла «методом перевала», т. е. на экстремалах  $S$ , что и отвечает классическим уравнениям движения в мнимом времени.

Для гравитационного поля, взаимодействующего с полями материи  $\varphi$ , эвклидово действие имеет вид

$$S = -\frac{m_{\text{Pl}}^2}{16\pi} \int \left[ R - 2\Lambda + \frac{16\pi}{m_{\text{Pl}}^2} L(g_{\mu\nu}, \varphi) \right] \sqrt{g} d^4x,$$

где интегрирование ведется по всему 4-мерному многообразию. Действие оказывается конечной величиной только для замкнутого мира, в частности для мира де Ситтера с метрикой вида

$$ds^2 = -dt^2 + H^{-2} \text{ch}^2 Ht (dr^2 + \sin^2 r d\Omega^2).$$

Покажем это, предварительно заменив  $t \rightarrow it$ . Тогда получим метрику вида

$$ds^2 = dt^2 + H^{-2} \cos^2 Ht (dr^2 + \sin^2 r d\Omega^2). \quad (9.16)$$

Она представляет собой метрику 4-мерной сферы, вложенной в 5-мерное эвклидово пространство. Это утверждение легко проверить, написав метрику сферической поверхности в 5-мерном эвклидовом пространстве (ср. с § 2 гл. 5):

$$\begin{aligned} t &= r_0 \sin H\tau, & \mathcal{V} &= r_0 \cos H\tau \cos r, & r &= r_0 \cos H\tau \sin r \cos \theta, & y &= \\ & & & & & & & = r_0 \cos H\tau \sin r \sin \theta \sin \varphi, \\ x &= r_0 \cos H\tau \sin r \sin \theta \cos \varphi, & r_0 &= \text{const}; \end{aligned}$$

$$g_{00} = 1, \quad g_{11} = H^{-2} \cos^2 Ht, \quad g_{22} = H^{-2} \cos^2 Ht \sin^2 r,$$

$$g_{33} = H^{-2} \cos^2 Ht \sin^2 r \sin^2 \theta, \quad \sqrt{g} = H^{-3} \cos^2 Ht \sin^2 r \sin \theta.$$

4-мерный объем этой поверхности определяется выражением

$$V = H^{-4} \int_0^T d(Ht) \int_0^\pi dr \int_0^\pi d\theta \int_0^{2\pi} d\varphi \cos^2 Ht \sin^2 r \sin \theta = \pi^3 H^{-4}.$$

Для метрики (9.16) инвариант тензора кривизны, определяемый согласно

$$R = -6a^{-3} \left[ a - a \left( \frac{da}{dt} \right)^2 - a^2 \frac{d^2 a}{dt^2} \right],$$

есть  $R = -12H^2$ . Выражая космологическую постоянную  $\Lambda$  через  $H$ , получаем в случае пустого пространства ( $L(g_{\mu\nu}, \varphi) = 0$ ) действие в виде

$$S = \frac{18}{16\pi} H^2 m_{\text{Pl}}^2 \int \sqrt{g} d^4x = \frac{18\pi^2}{16} m_{\text{Pl}}^2 H^{-2}.$$

Теперь видно качественное поведение вероятности рождения мира с радиусом  $H^{-1}$ :

$$W \sim \exp \left\{ -\frac{18\pi^2}{16} \frac{m_{\text{Pl}}^2}{H^2} \right\}.$$

Если радиус кривизны велик  $H^{-1} \gg m_{\text{Pl}}^{-1}$ , то вероятность рождения такого мира мала. Наиболее вероятно рождение мира с радиусом кривизны порядка планковского. На рис. 45 рождение мира из ничего схематически представлено в виде графика  $a(t)$ . Часть кривой при  $t < 0$  представляет собой классически запрещенную область, которую мы условно называем «ничего». Эта часть графика описывается квантовыми уравнениями. В ней классическое пространство и время не существуют, аналогично тому как не существует классической траектории частицы в запрещенной области при  $\alpha$ -распаде ядра. Масштабный фактор в этой области, а точнее говоря метрика, испытывает сильные квантовые флуктуации.

Выпишем уравнение Шредингера, которое, как мы надеемся, качественно описывает рождение Вселенной. Для этого возьмем выражение для полной энергии мира в виде

$$\left( \frac{1}{a^2} \frac{da}{d\eta} \right)^2 - H_0^2 + \frac{k}{a^2};$$

здесь  $H_0^2 = 8\pi G\rho/3$ . Соответствующий гамильтониан есть

$$\mathcal{H} = \left( \frac{da}{d\eta} \right)^2 - H_0^2 a^4 + ka^2.$$

Так как полная энергия замкнутого мира равна нулю, то для него этот гамильтониан обращается в нуль. Обычный метод квантования заключается в том, что выбирается координата и

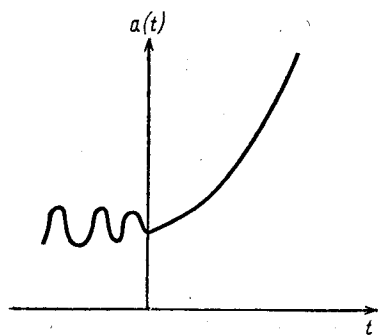


Рис. 45. Эволюция масштабного фактора при квантовом рождении Вселенной

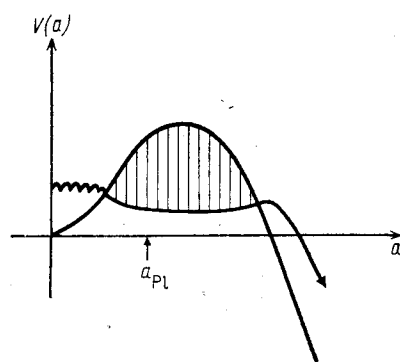


Рис. 46. Эффективный потенциал, определяющий динамику масштабного фактора

импульс и связываются коммутационным соотношением

$$[q, p] = i.$$

В качестве координаты можно выбрать масштабный фактор  $a$ . Сопреженный этой координате импульс равен

$$p = \frac{\partial \mathcal{H}}{\partial (da/d\eta)} = \frac{da}{d\eta}.$$

Тогда гамильтониан записывается в виде

$$\mathcal{H} = p^2 - H_0^2 a^4 + ka^2.$$

После введения операторов координаты и импульса  $\hat{a}$ ,  $\hat{p}$  и перехода в  $a$ -представление получаем уравнение Шредингера для волновой функции  $\psi(a)$  в виде

$$\mathcal{H}\psi = 0.$$

Особо подчеркнем, что  $\psi$  зависит только от масштабного фактора  $a$  и не зависит от времени. Зависимость от времени пропадает, так как энергия замкнутого мира равна нулю. Поэтому из стандартного разложения по собственным состояниям гамильтониана

$$\psi(a, t) = \sum_n \varphi_n(a) e^{iE_n t}$$

следует, что волновая функция от времени не зависит.

Как известно, гамильтониан представляет собой сумму двух

членов — кинетического  $\partial^2 \psi / \partial a^2$  и потенциального  $V(a)\psi(a)$ . В нашем случае потенциальный член есть  $V(a) = -H_0^2 a^4 + ka^2$ . Осцилляция  $\psi(a)$  вблизи нуля не приводит к расширению мира. Мир, родившийся с  $a \leq a_{p1}$ , вначале должен преодолеть классически запрещенную область, которая на рис. 46 заштрихована. Совершив подбарьерный переход (как в случае  $\alpha$ -распада), мир попадает на склон потенциала и начинает экспоненциально расширяться.

Найдем решение этого уравнения в области больших  $a$ . Решение будем искать в виде

$$\psi(a) = \exp\{if(a)\},$$

причем

$$f \frac{d^2 f}{da^2} \ll \left(\frac{df}{da}\right)^2.$$

Для  $f(a)$  получим уравнение

$$\left(\frac{df}{da}\right)^2 = H_0^2 a^4 - ka^2.$$

Для больших  $a$ , т. е. при  $a \gg a_{p1}$ , решение имеет вид

$$f(a) = \frac{1}{3} H_0 a^3.$$

В квазиклассической области волновая функция имеет вид

$$\psi(a) = \exp\{iH_0 a^3/3\}.$$

Проверим, как в этом случае масштабный фактор зависит от времени. Время вводится с помощью импульса  $\mathcal{P}$ :

$$\mathcal{P} = -\psi^* i \frac{\partial}{\partial a} \psi = A_0 a^2.$$

По определению  $\mathcal{P} \equiv da/d\eta$  и  $d\eta = dt/a(\eta)$ , или  $\frac{1}{a} \frac{da}{g\eta} = \frac{da}{dt} = H_0 a$ , откуда сразу же получаем

$$a(t) = a_0 \operatorname{ch} Ht,$$

т. е. после туннельного перехода Вселенная, рожденная из «ничего», приходит согласно нашей простой модели к классическому деситтеровскому закону расширения.

## § 6. ХАОТИЧЕСКАЯ ИНФЛЯЦИЯ И ВЕЧНАЯ ВСЕЛЕННАЯ

Пусть во Вселенной находится случайно распределенное скалярное поле  $\phi$ , на которое наложено только одно ограничение, а именно будем считать, что его плотность энергии  $\rho_\phi$  не превышает планковскую плотность энергии:

$$\rho_\phi < m_{\text{Pl}}^4. \quad (9.17)$$

Как заметил Линде (1983), если в распределении  $\phi$  будут области с достаточно однородным полем, то эти области будут претерпевать экспоненциальное расширение и в конечном счете они составят основную часть физического (в отличие от координатного) объема Вселенной.

Рассмотрим пока для простоты случай невзаимодействующего поля. В области однородного поля, где можно пренебречь пространственными производными  $\phi$ , оно удовлетворяет известному нам уравнению

$$\ddot{\phi} + 3H\dot{\phi} + m^2\phi = 0. \quad (9.18)$$

Параметр Хаббла  $H = \dot{a}/a$  связан с плотностью энергии поля:

$$H^2 = \frac{8\pi}{3m_{\text{Pl}}^2} \left( \frac{\dot{\phi}^2}{2} + \frac{m^2}{2} \phi^2 \right). \quad (9.19)$$

Если величина поля  $\phi^2$  достаточно велика:  $\phi > m_{\text{Pl}}$ , то решение уравнений (9.18) и (9.19) имеет вид

$$\phi(t) = \phi_0 - \frac{mm_{\text{Pl}}}{2\sqrt{3\pi}} t,$$

$$a(t) = a_0 \exp \left\{ -\frac{2\pi}{m_{\text{Pl}}^2} [\phi_0^2 - \phi^2(t)] \right\}.$$

Мы видим, таким образом, что в течение времени  $t \simeq 2\sqrt{3\pi} \phi_0 / mm_{\text{Pl}}$  масштабный фактор растет почти экспоненциально. Мы вновь встретились с механизмом инфляции за счет медленного «скатывания» поля  $\phi$  к положению равновесия. Так как исходное состояние предполагается хаотическим, эта реализация раздувания получила название хаотической инфляции.

Во время экспоненциального режима длинноволновые флуктуации поля  $\phi$  нарастают (см. гл. 12), достигая за время

$$H^{-1} \simeq \sqrt{3} m_{\text{Pl}} / 2\sqrt{\pi} m \text{ величины } |\delta\phi| \simeq \frac{H}{\sqrt{2\pi}} \simeq \sqrt{\frac{2}{3\pi}} \frac{m}{m_{\text{Pl}}}.$$

Для того чтобы флуктуации  $\delta\phi$  приводили к флуктуациям плотности, не противоречащим наблюдениям (см. гл. 11, 12), необходимо, чтобы  $m \lesssim 10^{-5} m_{\text{Pl}}$ . Это означает, что  $\phi$  может достигать величины  $10^4 m_{\text{Pl}}$ , не вступая в противоречие с условием (9.17).

Квантовые флуктуации  $\phi$  в экспоненциально расширяющемся мире приводят к неожиданным выводам. А именно оказывается, что величина  $\phi$  может расти, несмотря на то что по-

тенциальная энергия при этом нарастает. Необходимые для этого условия можно найти, сравнив уменьшение  $\phi$  за время  $H^{-1}$  в соответствии с формулой (9.20)

$$\Delta\phi \simeq \frac{1}{4\pi} m_{\text{Pl}}^2 / \phi$$

и величину флуктуаций (9.21). Отсюда нетрудно заключить, что при

$$\phi \gtrsim \frac{1}{3} m_{\text{Pl}} \left( \frac{m_{\text{Pl}}}{m} \right)^{1/2} \simeq 100 m_{\text{Pl}}$$

поле  $\phi$  нарастает, грубо говоря, с вероятностью  $1/2$ .

Мы пришли, таким образом, к очень интересной картине. Во Вселенной со случайно распределенным полем  $\phi$  области достаточной однородности  $\phi$  экспоненциально разрастаются. В течение характерного времени  $H^{-1}$  область размера  $H^{-1}$  вырастает в  $e$  раз, а ее объем в  $e^3$  раз. Таким образом, она превращается в  $e^3$  областей размера  $H^{-1}$ , каждая из которых продолжает экспоненциально расширяться. Однако лишь примерно в половине из этих областей поле  $\phi$  уменьшается на величину  $|\delta\phi| + \Delta\phi$ , а в другой половине увеличивается на  $|\delta\phi| - \Delta\phi$ . Хотя в среднем по координатному объему  $\phi$  падает в соответствии с выражением (9.20), физический объем областей, занятых растущим полем  $\phi$ , экспоненциально растет. Те области, где в конце концов поле  $\phi$  станет меньше  $100 m_{\text{Pl}}$ , должны будут прекратить экспоненциальное расширение, так как в них рост поля прекратится. Таким образом, мы приходим к процессу бесконечного возникновения Вселенных нашего типа из флуктуирующего поля  $\phi$  (Линде, 1986). Теперь нет надобности полагать, что существовал выделенный момент рождения Вселенной как целого. Этот процесс мог не иметь начала и не будет иметь конца.

Размер фридмановской области сейчас составляет величину порядка  $m_{\text{Pl}}^{-1} \exp(2\pi\phi_0^2/m_{\text{Pl}}^2) \sim 10^{10}$ , т. е. граница области находится далеко за горизонтом. Поэтому у нас нет никакой надежды когда-либо проверить эту картину, но философски она очень интересна.

Заметим, что в этой модели Вселенная в целом очень неоднородна и топология ее отличается от простой топологии фридмановской модели. Однако в доступных нам масштабах использованная выше классификация заведомо справедлива и, в частности, если  $\Omega > 1$ , может произойти коллапс нашей части Вселенной.



## Глава 10.

## СТРУКТУРА ВСЕЛЕННОЙ

Наша Вселенная однородна лишь в среднем, в очень больших масштабах  $L \geq 200$  Мпк. По мере уменьшения масштабов мы видим ярко выраженную иерархическую структуру в виде звезд и планет, шаровых звездных скоплений ( $L \sim 1$  пк), галактик ( $L \sim 10 \div 100$  кпк), их скоплений ( $L \sim 10$  Мпк) и сверхскоплений ( $L \sim 100$  Мпк). Последние образуют во Вселенной относительно тонкие стенки ( $L \sim 10-20$  Мпк), ограничивающие пустые области (точнее области с заметно более низкой концентрацией светящегося вещества) размера 100–200 Мпк, формируя, таким образом, ячеистую структуру Вселенной. В еще больших масштабах неоднородности уже не видны. Помимо непосредственного наблюдения светящегося вещества, об этом говорят изотропия реликтового (малость  $\delta T/T$ ) и рентгеновского фона. Из этих данных можно сделать вывод, что неоднородности в масштабах  $10^3-10^4$  Мпк не превышают 1%.

Как полагают, эта иерархическая структура возникла в результате обусловленного гравитационной неустойчивостью роста каких-то первоначально малых возмущений плотности. Вопрос о происхождении этих исходных неоднородностей рассматривается в следующей главе. Они обязаны своим происхождением физическим процессам на ранней (инфляционной) стадии развития Вселенной и в силу этого их величина и характерные размеры определяются фундаментальными постоянными физики микромира.

В этой главе мы будем считать начальные возмущения заданными и обсудим их развитие со временем, их амплитуду и спектральное распределение, необходимые для формирования наблюдаемой структуры. Обсуждаются также форма и физические свойства невидимой материи, которая согласно современным взглядам определяет образование структуры.

Мы начнем с более простого случая развития неоднородностей во Вселенной, заполненной лишь барионной материей, переходя ниже к более реалистическому случаю (по всей видимости) многокомпонентного — «заполнения» Вселенной.

Материал этой главы в значительной степени использует обзор Я. Б. Зельдовича [73], а также книгу Пиблса [28].

Современная структура Вселенной не является темой этой книги, поэтому наше изложение будет очень кратким и ориентированным на те выводы из наблюдений и теории развития структуры, которые касаются ранней Вселенной.

## § 1. НАЧАЛЬНАЯ СТАДИЯ РОСТА ВОЗМУЩЕНИЙ

Релятивистская теория развития возмущений в нестационарном мире была предложена в 1946 г. Е. М. Лифшицем [22]. Качественно результаты можно понять, однако, и в рамках ньютоновской гравитации. Физической причиной роста возмущений плотности является гравитационное притяжение. Пусть на фоне однородно распределенной материи в бесконечном пространстве с плотностью  $\rho_0$  возникает какая-то область с избытком плотности  $\delta\rho$ . Очевидно, она будет притягивать к себе окружающее вещество. И если силы давления не уравновесят гравитационных сил, эта неоднородность плотности будет нарастать. В случае, когда размер области с избытком вещества  $L$  достаточно велик, будут доминировать гравитационные силы. При малых  $L$  доминируют силы давления. Легко понять, что по порядку величины границе раздела между этими двумя возможностями отвечает равенство времени свободного падения к центру выделенной области  $t_g \approx (G\rho_0)^{-1/2}$  и времени распространения звуковой волны  $t_s \approx L/c_s$ . Если  $t_g < t_s$ , то флуктуации плотности нарастают. В противоположном случае  $t_g > t_s$  флуктуации плотности представляют собой звуковые волны. Отсюда видно, что нарастают возмущения лишь в достаточно больших областях с  $L > c_s(G\rho_0)^{-1/2}$ . Граничное значение размера флуктуации называют длиной волны Джинса (который в начале века исследовал задачу о гравитационной неустойчивости в ньютоновской теории на стационарном фоне) и обозначают  $\lambda_J$ . Ее точное значение отличается лишь численным множителем от приведенного выше выражения:

$$\lambda_J = \sqrt{\frac{\pi}{G\rho_0}} c_s. \quad (10.1)$$

Отсылая за выводом к [28, 73], укажем, что эволюция флуктуации плотности с длиной волны  $\lambda$  и амплитудой  $A_k$  определяется выражением

$$\frac{\delta\rho_k}{\rho_0} = A_k e^{\gamma t + i k x}, \quad (10.2)$$

где  $k = 2\pi/\lambda$ , а  $\gamma = \pm [4\pi G\rho_0 - k^2 c_s^2]^{1/2}$ .

В пределе малых  $\delta\rho$  растущее решение можно представить в факторизованном виде:

$$\delta\rho = \rho_0 e^{\gamma_0 t} \int A_k e^{i k x} d^3 k = \rho_0 e^{\gamma_0 t} \psi(x), \quad (10.3)$$

где  $\gamma_0 = \sqrt{4\pi G\rho_0}$ . Учет следующих членов в разложении по  $k$  приведет к замене  $\psi(x)$  на  $\psi(x, t)$ .

Итак, возмущения плотности на стационарном фоне растут

экспоненциально. Интуитивно ясно, что в расширяющейся Вселенной рост возмущений будет более медленным. Плотность  $\rho_0$  является убывающей функцией времени

$$\rho_0(t) = (6\pi G t^2)^{-1} \quad (10.4)$$

(это выражение справедливо при  $\Omega \equiv \rho/\rho_c = 1$ ), и можно наивно оценить степень роста возмущений как

$$\exp\left(\int^t \sqrt{4\pi G \rho_0(t)} dt\right) = \exp\left(\sqrt{\frac{2}{3}} \int \frac{dt}{t}\right) \sim t^{1/3}.$$

Вместо экспоненциального, мы пришли к степенному закону. В точной теории, которая строится на основе уравнений гидродинамики в расширяющемся мире, показатель степени оказывается слегка иным:

$$\frac{\delta\rho}{\rho_0(t)} = t^{2/3} \psi(x), \quad (10.5)$$

т. е. относительная величина флуктуации растет как масштабный фактор (при доминантности нерелятивистского вещества и в плоском мире, т. е.  $\Omega = 1$ ). Напомним, что это выражение справедливо для малых флуктуаций, в области применимости линейного приближения.

Строгая теория развития малых возмущений в релятивистской космологии была построена Лифшицем. Ее изложение можно найти в книге Ландау и Лифшица [20]. Чтобы не загромождать изложение, мы приведем здесь только результаты. Возмущение метрического тензора можно представить как суперпозицию возмущений следующих трех типов. Первый тип — это скалярные возмущения, о которых шла речь выше. В этом случае все компоненты добавки к метрическому тензору можно выразить через одну скалярную функцию. Скалярные возмущения метрики уже в первом порядке приводят к возмущению скорости и плотности материи, поэтому их роль в образовании структуры Вселенной наиболее существенна. Если рассмотреть поле скоростей материал  $v^a(x)$ , то скалярные возмущения связаны с дивергенцией  $\partial_a v^a$ .

Второй тип возмущений связан с антисимметричной частью тензора  $\partial_a v_b$ . Это векторное, или вихревое возмущение. Оно отвечает вращению вещества, когда испытывает возмущения скорость, но не плотность материи. Вихревые возмущения приводят к возмущению плотности лишь во втором порядке.

Симметричной части  $\partial_a v_b$  отвечают тензорные возмущения метрики. Плотность материи при этом, как и в предыдущем случае, в первом порядке не возмущается. Эти возмущения представляют собой гравитационные волны. В последнее время возрос интерес к поиску реликтовых гравитационных волн,

так как они существенно усиливаются на инфляционной стадии и их обнаружение дало бы уникальную возможность получить информацию об этом периоде. Мы вернемся к этому вопросу в гл. 11.

При малых  $t$  скалярные возмущения метрики с длиной волны меньше горизонта,  $\lambda < ct$ , на стадии релятивистской доминантности остаются постоянными, в то время как относительные флуктуации плотности линейно зависят от времени  $\delta\rho/\rho_0 \sim t$  (так как  $\rho_0 \sim t^{-2}$ , а  $\delta\rho \sim t$ ). По этой причине возмущения на начальной стадии удобно описывать в терминах возмущений метрики. При переходе к нерелятивистскому режиму расширения возмущение метрики начинают расти как первая степень масштабного фактора  $\delta\rho/\rho \sim a \sim t^{2/3}$  в соответствии с выражением (10.5).

Заметим в заключение, что, в отличие от скалярной и тензорной мод, амплитуда вихревого возмущения в линейном приближении сингулярна при малых  $t$ , т. е.  $\delta h_{uv} \rightarrow \infty$  при  $t \rightarrow 0$ . Это является следствием сохранения момента и означает, что слабовозмущенные вихревые моды должны отсутствовать.

## § 2. ВОЗМУЩЕНИЕ В ПЛАЗМЕ

Рассуждения предыдущего параграфа относились к однокомпонентному веществу. Однако реальная первичная плазма состоит из электронов, протонов, электромагнитного излучения, а также нейтрино или каких-то других частиц, которые, вероятно, образуют сейчас основную часть скрытой массы Вселенной. Это приводит к тому, что флуктуации плотности нельзя характеризовать одной величиной  $\delta\rho$ , а нужно учитывать распределение плотности различных компонент  $\delta\rho_i$ . Откладывая обсуждение невидимой компоненты до § 5, 7, мы рассмотрим здесь флуктуации плотности в плазме, состоящей из электромагнитного излучения и массивного ионизованного вещества. Их можно разбить на комбинацию двух независимых мод: адиабатическую и изотермическую. В первом случае отсутствуют пространственные флуктуации удельной (на один фотон) энтропии, или, что то же, флуктуации плотности энергии массивных частиц  $\rho_m$  и излучения  $\rho_r$  связаны соотношением

$$\frac{\delta\rho_m}{\rho_m} = \frac{3}{4} \frac{\delta\rho_r}{\rho_r}.$$

Это означает, что отношение плотности числа частиц постоянно  $N_m/N_r = \text{const.}$

Такие флуктуации возникают из исходно однородного распределения вещества при наличии скалярных возмущений метрики. Последние приводят к разному темпу расширения в разных точках пространства и, следовательно, к образованию флуктуа-

ций плотности. Химический состав, т. е. отношение  $N_m/N_r$  при этом не меняется из-за сильного взаимодействия материи и излучения.

Изотермические флуктуации определяются условием  $\delta\rho_r=0$ , т. е. излучение распределено однородно, флуктуации его температуры отсутствуют, а плотность вещества слегка меняется от точки к точке:  $\delta\rho_m \neq 0$ . Чтобы определить развитие изотермических флуктуаций, нужно задать в качестве начальных условий не только флуктуации метрики, но и распределение химического состава в первичной плазме.

Задачей теории, которая обсуждается в этой главе, является описание структуры Вселенной на основе начальных данных о спектре (т. е. об амплитуде и длине волны) и физических свойствах флуктуаций плотности. Мы увидим, что формирование структуры существенно отличается в моделях с адиабатическими и изотермическими флуктуациями из-за разных законов их развития в первичной плазме.

Удобство деления флуктуаций плотности на адиабатические и изотермические связано с тем обстоятельством, что они слабо взаимодействуют друг с другом и поэтому эволюционируют независимо в довольно широком диапазоне длин волн. Для достаточно длинных волн, когда можно пренебречь диффузией фотонов, энтропия сохраняется и адиабатические возмущения остаются адиабатическими. Для изотермических возмущений диффузия фотонов отсутствует и они могли бы привести к рождению адиабатических только гравитационно за счет взаимодействия с метрикой. Очевидно этот эффект мал для длин волн, меньших горизонта,  $\lambda < t$ , так как для релятивистских частиц джинсовская длина волны по порядку величины равна горизонту.

Ниже мы рассмотрим развитие структуры Вселенной в моделях с адиабатическими и изотермическими флуктуациями и увидим, что картины существенно отличаются.

Хотя адиабатические флуктуации физически кажутся более естественными, не нужно закрывать глаза и на возможность изотермических флуктуаций, и, возможно, в окончательной теории оба типа флуктуаций окажутся существенны.

### § 3. ЗАМЕЧАНИЕ О СПЕКТРЕ ФЛУКТУАЦИИ

Как мы уже отмечали, флуктуации удобно характеризовать их спектральным распределением. Пусть случайная функция представляет собой (скалярные) флуктуации метрики в синхронной системе  $ds^2 = dt^2 - a^2(t)(dx^2 + dy^2 + dz^2)$  с нулевым (по определению) средним значением  $\langle h(x) \rangle = 0$ . Пространственное распределение флуктуаций описывается их функцией корреляции:

$$G(x, y) = \langle h(x)h(y) \rangle$$

— средним значением произведения флуктуаций метрики в двух точках  $x$  и  $y$ . Обычно делают гипотезу о трансляционной инвариантности флуктуаций, связанную с однородностью Вселенной:

$$G(x, y) = G(x - y).$$

Преобразование Фурье функции  $G(x)$  определяет спектр флуктуаций метрики:

$$h_k^2 = \int d^3x e^{ikx} G(x). \quad (10.6)$$

Простейшее и наиболее интересное физически предположение о виде  $h_k$  состоит в том, что в теории отсутствует какая-либо выделенная длина, и тогда просто из размерных соображений

$$h_k = \delta \cdot k^{-3/2}, \quad (10.7)$$

где  $\delta$  — некоторая безразмерная постоянная (заметим, что размерность  $h_k$  иная, чем  $h(x)$ ).

Если ограничиться только лишь спектром степенной формы  $h_k \sim k^{-\nu}$ , то астрономические данные свидетельствуют в пользу  $\nu \approx 3/2$ . Случай  $\nu > 3/2$  означал бы, что Вселенная менее однородна на больших масштабах, чем на малых, в противоречии с наблюдениями. А если  $\nu < 3/2$ , то ему отвечали бы слишком сильные флуктуации плотности на малых масштабах и в результате возникло бы слишком большое количество черных дыр, которые наблюдались бы по рентгеновскому излучению. Величина постоянной  $\delta$  должна быть порядка  $10^{-4}$ , чтобы не возникло противоречий с наблюдаемой изотропией реликтового излучения.

Спектр возмущений  $h_k \sim k^{-3/2}$  относится к так называемым фрактальным спектрам. Важность понятия фрактальных спектров и фрактальных кривых была осознана сравнительно недавно. Решающую роль здесь сыграла книга Мандельброта (1977). Удобно пользоваться его терминологией, хотя спектр  $\sim k^{-1}$  и понятие подобия возникли ранее.

Физически понятие фрактала возникает, например, когда мы рассматриваем длину береговой линии в различных масштабах. На картах с грубым масштабом береговая линия кажется гладкой, переходя к более мелким масштабам мы вынуждены отмечать все более мелкие детали — заливы, заливчики, поймы речушек и ручьев и, наконец, размер камней. При повышении «разрешающей способности» карты появляются мелкие шероховатости и зазубрины, потом мы увидим кристаллическую структуру береговой линии, вступаем в мир квантовых представлений, где уже исчезает само понятие границы. Значит, при переходе от масштаба к масштабу длина береговой линии будет нарастать, в пределе устремляясь к бесконечности. По-

этому измерять длину береговой линии корректно лишь тогда, когда указано, в каком масштабе ведутся измерения.

Для физиков понятие фрактальной кривой тесно связано с понятием промежуточной асимптотики. Математическим примером фрактальной кривой является как раз кривая со спектром типа  $h_k \sim k^{-3/2}$  (Я. Б. Зельдович, Д. Д. Соколов, 1985).

Рассмотрим более наглядный пример, чтобы понять основные математические свойства фрактальных кривых. Пусть функция  $y=y(x)$  представляет из себя непрерывную функцию без производной вида

$$y(x) = \sum_{n=0}^{\infty} a(k_n) \cos(k_n x + \varphi_n)$$

и  $a(k_n) \sim k_n^{-\alpha}$ ,  $k_n \sim n$  при  $n \rightarrow \infty$ . Фазы  $\varphi_n$  — случайные числа, равномерно заполняющие промежуток  $(0, 2\pi)$ . Здесь следует подчеркнуть, что фазы должны быть случайными, как впрочем и то, что  $k_n \sim n$ , а не просто  $k_n = n$ . В последнем случае фрактальная кривая может вырождаться во всюду гладкую кривую, за исключением счетного множества точек, в которых  $y(x) = \infty$ . Другими словами, для нас важно, чтобы «сбивалась» не только фаза, но и частота.

Кривая подобного типа является непрерывной, но не дифференцируемой, если  $0 < \alpha < 1$ . Для того чтобы ввести понятие анализа на таких кривых, мы должны определить так называемые гёльдеровские производные вида

$$\Delta y = \mu (\Delta x)^\alpha.$$

Пример производных такого типа мы хорошо знаем из статистической физики. Это производные при броуновском движении, т. е.

$$\Delta l = \sqrt{D \Delta t},$$

у которых показатель Гёльдера  $\alpha = 1/2$ .

Посмотрим теперь, какова размерность получившейся кривой. По идее Хаусдорфа дробная размерность определяется следующим образом. Построим вокруг каждой точки кривой кружок радиусом  $\varepsilon \rightarrow 0$  и подсчитаем площадь всех кружков  $S(\varepsilon)$ . Причем площадь перекрытия нескольких кружков учитывается только один раз. Скорость убывания  $S(\varepsilon)$  с убыванием  $\varepsilon$  определяет размерность. Для гладкой кривой  $S(\varepsilon) \sim \varepsilon L$ , где  $L$  — длина кривой. Для одной точки  $S(\varepsilon) \sim \varepsilon^2$ . Для плоскости  $S(\varepsilon) \sim \varepsilon^0$ .

Определим размерность фрактальной кривой (10.7) аналогичным способом. Очевидно, характерная амплитуда осцилляций на масштабе  $\varepsilon$  равна  $a(\varepsilon^{-1})$ . Если  $a(\varepsilon^{-1})$  убывает медленнее, чем  $\varepsilon$  при  $\varepsilon \rightarrow 0$ , то площадь, покрытая кружками радиусом  $\varepsilon$ , будет не порядка  $\varepsilon$ , а порядка  $a(\varepsilon^{-1})$ . Если же  $a(\varepsilon^{-1}) \leq \varepsilon$ ,

то полоса кружков отслеживает все изгибы кривой и эта кривая является гладкой.

Для фрактальной кривой  $S(\varepsilon) \sim a(1/\varepsilon) \sim \varepsilon^\alpha$ . Если  $0 < \alpha < 1$ , то  $S(\varepsilon)$  убывает медленнее, чем для гладкой линии, и быстрее, чем для плоскости. Поэтому фрактальная кривая занимает промежуточное положение между гладкими кривыми и поверхностями, в связи с чем фрактальные кривые иногда называют толстыми линиями. Хаусдорф предложил определение, согласно которому размерность такого образования равна

$$\dim_{\text{ext}} \gamma = 2 - \alpha;$$

индекс ext (external — внешняя) указывает, что при построении этой величины пришлось выйти за пределы самой кривой.

Для физических приложений фрактальную кривую иногда берут с обрезанным с двух сторон спектром. Со стороны больших длин волн или низких частот — временем наблюдения или масштабами наблюдаемой области, а со стороны высоких частот — разрешающей способностью прибора.

Спектр (10.7) неплохо описывает наблюдаемую астрономами картину как в области малых, так и больших размеров. Заметим, что приближенно к такому спектру адиабатических флуктуаций приводит инфляционная модель в простейшем варианте (см. гл. 11). Однако более детальное описание структуры Вселенной, по-видимому, требует привлечения размерных параметров уже в исходном спектре флуктуаций. Заметим, впрочем, что даже безразмерный первичный спектр  $\nu = 3/2$  в ходе физических процессов при расширении и охлаждении Вселенной становится сложнее. В нем возникают характерные размеры, отвечающие затуханию волн определенной длины, длина свободного пробега частиц и т. п. (см. ниже).

Однако, возможно, привлечение только что упомянутых характерных масштабов окажется недостаточным, и необходимы какие-то дополнительные параметры. Физика микромира в принципе дает нам большое количество подобных параметров, к несчастью даже слишком большое, и проблема состоит не в том, чтобы найти какой-то физический механизм, приводящий к возникновению новых масштабов, а в том, чтобы сделать правильный выбор среди множества возможностей. Дело в том, что мы не знаем ни физической природы скрытой массы, ни деталей взаимодействия элементарных частиц при высоких энергиях, и не исключено, что теория образования структуры Вселенной явится для нас существенным источником информации обо всем этом.

#### § 4. СТРУКТУРА ВСЕЛЕННОЙ В АДИАБАТИЧЕСКОЙ ТЕОРИИ

При температуре плазмы выше температуры рекомбинации водорода  $T = 3000$  К вещество и излучение взаимодействуют

сильно и поэтому жестко связаны. Когда длина волны флуктуации больше горизонта  $\lambda > t$ , флуктуации метрики слабо зависят от времени:  $h \approx \text{const}$ , а флуктуации плотности растут как масштабный фактор:  $\delta\rho/\rho \sim (z+1)^{-1}$  на нерелятивистской стадии и как квадрат масштабного фактора на релятивистской. Этот рост можно понимать как более медленное расширение области с избытком плотности (и соответственно более быстрое — с недостатком) на фоне общего однородного расширения\*. Это иллюстрирует рис. 47. Когда размер флуктуаций становится меньше горизонта, они начинают осциллировать, образуя звуковые волны (длина волны Джинса для радиационно доминированной плазмы порядка горизонта). Из-за диффузии фотонов коротковолновые флуктуации затухают (затухание Силка). Мы можем оценить максимальную длину волны, для которой отсутствует затухание следующим образом. Длина свободного пробега фотонов в плазме равна  $l_{\text{св}} = (\sigma n_e)^{-1}$ , где  $n_e$  — плотность электронов,

а  $\sigma = 2\pi a^2/m_e^2$  — сечение рассеяния фотонов на электронах (томсоновское сечение),  $m_e$  — масса электрона,  $a = 1/137$  — постоянная тонкой структуры. Время, за которое фотон продифундирует на расстояние  $L$ , очевидно, равно  $t(L) = L^2/l_{\text{св}}$ . Это время не должно превосходить времени жизни Вселенной:  $t(L) < t$ . Отсюда минимальный размер незатухающей флуктуации в момент рекомбинации  $t_r$  равен

$$\lambda_{\text{min}} \approx (l_{\text{св}} t_r)^{1/2},$$

а минимальная масса образующихся объектов по порядку величины равна

$$M_{\text{min}} = \frac{4\pi}{3} \rho_m \lambda_{\text{min}}^3, \quad (10.8)$$

\* Для космологических приложений именно флуктуации с  $\lambda > t$  на раннем этапе особенно интересны, так как, возвращаясь под горизонт на стадии доминантности материи, они приводят к образованию космологически значимых структур. Образование флуктуаций с  $\lambda > t$  могло бы служить поводом для беспокойства, однако тут (в который раз!) приходит на помощь инфляционная модель. В этой модели флуктуации микроскопического размера экспоненциально растут и уходят далеко за горизонт.

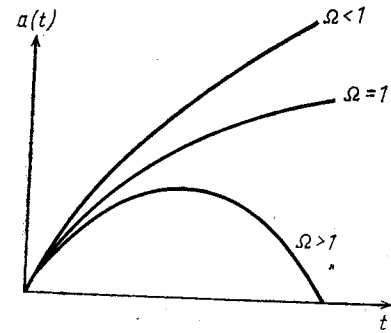


Рис. 47. Зависимость масштабного фактора от времени во Фридмановской Вселенной для разных значений параметра  $\Omega = \rho/\rho_c$ ; чем больше плотность, тем медленнее идет расширение

где  $\rho_m$  — плотность материи в момент рекомбинации. В результате получим  $M_{\text{min}} \approx 10^{14} M_{\odot}$ , где  $M_{\odot} = 2 \cdot 10^{33}$  г — масса Солнца.

После рекомбинации положение резко меняется: вещество становится нейтральным. Излучение практически перестает с ним взаимодействовать. Скорость звука в веществе падает до величины порядка  $(T/m_H)^{1/2}$ , где  $T$  — температура среды, а  $m_H$  — масса атома водорода. Масса Джинса, т. е. масса, сосредоточенная в объеме, ограниченном  $\lambda$ , (10.1), по порядку величины равна

$$M_J^B = \frac{4}{3} \pi \lambda^3 \rho \approx 10^6 M_{\odot}.$$

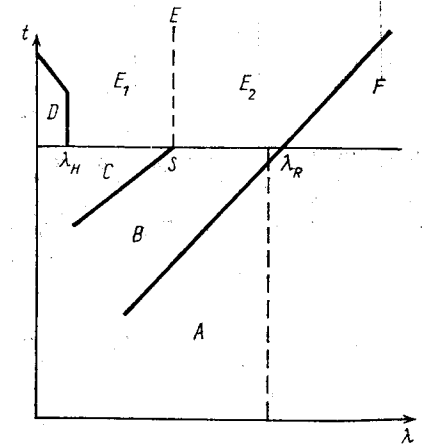
Возмущения большого масштаба должны после рекомбинации нарастать. Однако из-за диссипации коротких волн возмущения отсутствуют вплоть до  $M_{\text{min}}$  (10.8). Качественно эволюция возмущений изображена на рис. 48–50.

Для адиабатических возмущений флуктуации температуры излучения связаны с флуктуациями плотности соотношением:

$$\frac{\delta T}{T} = \frac{1}{3} \frac{\delta \rho}{\rho}. \quad (10.9)$$

Рис. 48. Области на плоскости «время—длина волны», в которых возмущения ведут себя по-разному. А — радиационно-доминированная (РД) плазма, длина волны возмущения больше горизонта. Возмущения метрики постоянны, а возмущения плотности растут. В — РД-плазма, возмущение короче горизонта. Возмущения плотности осциллируют, образуя звуковые волны. С — область затухания акустических колебаний. Горизонтальная линия при  $t_R \approx 3 \cdot 10^5$  лет отвечает моменту рекомбинации водорода, отключению радиации от вещества. Только возмущения, длина волны которых выше отвечающих точке S, выживают к моменту  $t_R$ .  $\lambda_H$  — длина волны Джинса нейтрального водорода;  $\lambda_R$  — длина волны Джинса ионизованной плазмы.

Д — область, в которой возмущения в нейтральном водороде не растут. E — область, в которой возмущения в нейтральном водороде возрастают. В области  $E_1$  возмущения отсутствуют из-за затухания в области С, однако если бы они там были, то возрастали бы. Наиболее существенные растущие возмущения находятся в области  $E_2$ . F — область, в которой длинноволновые возмущения в нейтральном газе возрастают. Длина волны возмущения в F больше горизонта. На линии между А и В и между  $E_2$  и F длина волны равна горизонту, а возмущения плотности равны начальным возмущениям метрики



Д — область, в которой возмущения в нейтральном водороде не растут. E — область, в которой возмущения в нейтральном водороде возрастают. В области  $E_1$  возмущения отсутствуют из-за затухания в области С, однако если бы они там были, то возрастали бы. Наиболее существенные растущие возмущения находятся в области  $E_2$ . F — область, в которой длинноволновые возмущения в нейтральном газе возрастают. Длина волны возмущения в F больше горизонта. На линии между А и В и между  $E_2$  и F длина волны равна горизонту, а возмущения плотности равны начальным возмущениям метрики

После рекомбинации флуктуации плотности растут как масштабный фактор (если  $\Omega \approx 1$ ), а флуктуации температуры не меняются. Поэтому измерения величины  $\delta T/T$  сегодня позволяют судить о величине флуктуаций плотности к моменту рекомбинации. Мы сразу сталкиваемся здесь с серьезной проблемой, так как флуктуации реликтового фона согласно современным

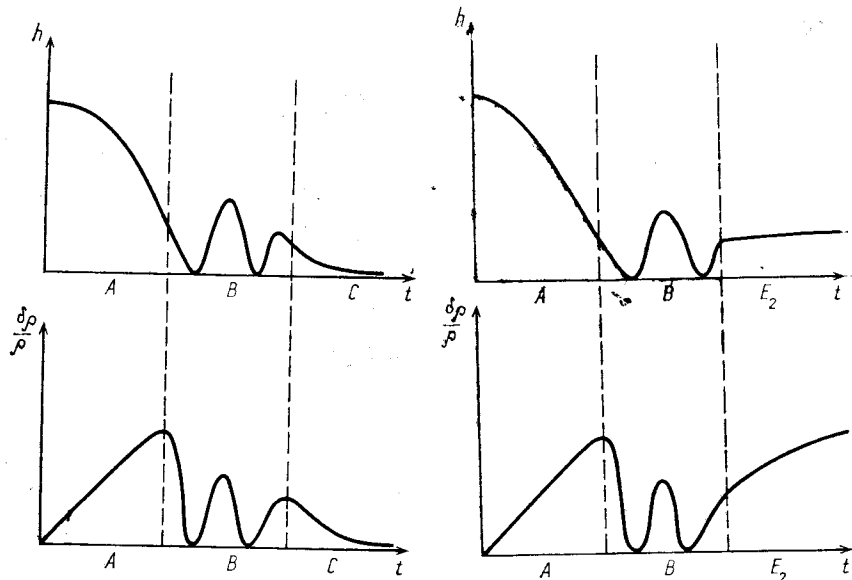


Рис. 49. Эволюция коротковолновых возмущений (короче  $S$  на рис. 48) метрики  $h$  и плотности  $\delta\rho/\rho$  в зависимости от времени. Буквы  $A$ ,  $B$  и  $C$  имеют то же значения, что и на рис. 48

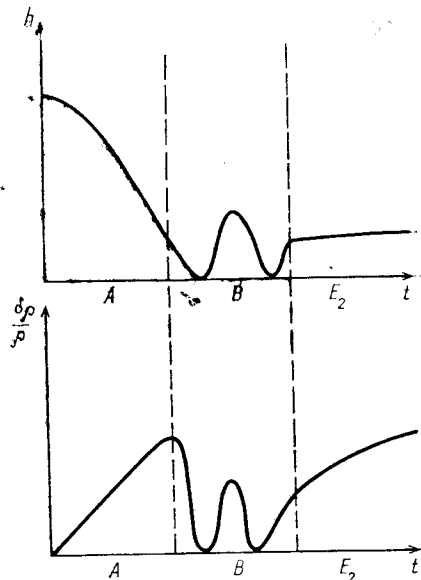


Рис. 50. Эволюция длинноволновых возмущений (длиннее  $S$  на рис. 48). Обозначения те же, что и на рис. 48.

49

наблюдениям столь малы, что соответствующие им  $\delta\rho/\rho$  не успевают дорости до требуемой величины  $(\delta\rho/\rho)_0 \approx 1$  от момента рекомбинации до наших дней (см. § 9).

Одним из способов решения этой проблемы является привлечение невидимой материи к формированию структуры Вселенной, тем более что на ее существование указывают и независимые соображения. Другая возможность состоит в учете изотермических флуктуаций, не связанных с излучением. Ниже мы перейдем к обсуждению этих вариантов.

### § 5. НЕЙТРИННАЯ ВСЕЛЕННАЯ

Нейтрино является замечательным кандидатом на роль носителя скрытой массы Вселенной. С одной стороны, реликтовые нейтрино весьма обильны во Вселенной, для каждого типа нейтрино и антинейтрино современная концентрация примерно равна

$$N_\nu = N_{\bar{\nu}} = 75 \text{ см}^{-3} (T_0/3\text{K})^3.$$

С другой стороны, теоретически очень естественно, чтобы масса нейтрино была отлична от нуля. Экспериментальная ситуация в настоящее время неопределенная: известно можно сказать, что масса электронного нейтрино меньше 30 эВ, но положительные результаты о массе нейтрино пока не подтверждены. Космология приводит к сравнимому ограничению на  $m_\nu$  и нуждается в массе стабильного нейтрино вблизи существующего верхнего предела. Перспективы непосредственного измерения в лабораториях масс  $\nu_\mu$  и  $\nu_\tau$  довольно плохие, так как характерная энергия  $\nu_\mu$  порядка 50 МэВ, а  $\nu_\tau$  — на порядок больше. Возможно, что детальное изучение спектра солнечных нейтрино с учетом резонансных осцилляций в веществе Солнца (Михеев, Смирнов, 1986) позволят определить массы всех нейтрино, но это дело не слишком близкого будущего.

В отличие от барионов нейтрино перестает взаимодействовать с излучением и вообще с первичной плазмой уже при температуре  $T = 3-5$  МэВ. Отсутствие взаимодействия, казалось бы, приводит к исчезновению сил давления, препятствующих гравитационному сжатию, и к возможности роста коротковолновых флуктуаций. Однако это неверно, так как именно в силу слабости взаимодействия нейтрино свободно уходят из области с большой концентрацией. Нетрудно оценить минимальную длину волны флуктуации, которая еще не успевает распасться. Она равна пути, которое проходит нейтрино, двигаясь со скоростью света, пока не станет нерелятивистским из-за красного смещения:

$$\lambda_J^\nu \approx t (T = m_\nu) \approx m_{\text{Pl}}/m_\nu^2. \quad (10.10)$$

Соответственно минимальная масса неразрушенной флуктуации равна

$$M_J^\nu \sim \rho_\nu \lambda_J^{\nu 3} \sim \lambda_J^{\nu 3} m_\nu^4 = M_{\text{Pl}}^3/m_\nu^2. \quad (10.11)$$

Численно это составляет  $M_J^\nu \approx 10^{16} M_\odot (10 \text{ эВ}/m_\nu)^2$ . Величину  $M_J^\nu$  также называют массой Джинса, хотя физика этого ограничения совершенно иная, чем в случае классической массы Джинса. Флуктуации указанного размера начинают расти раньше всех, быстрее всех достигают величины порядка единицы и являются в этой теории зародышами скоплений галактик. Эво-

люция флуктуаций в нейтринной модели иллюстрируется рис. 51.

В нейтринной Вселенной проблема с большими флуктуациями  $\delta T/T$  оказывается примерно на порядок слабее (если  $m_\nu =$

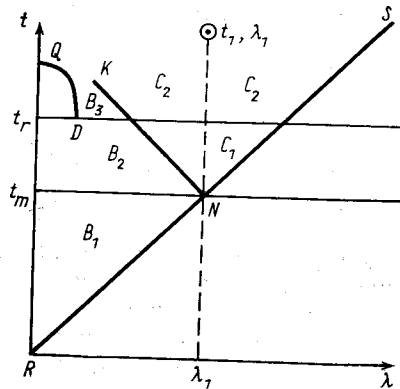


Рис. 51. Эволюция возмущений во Вселенной, в которой массивные нейтрино доминируют в плотности энергии.  $RNS$  — линия  $\lambda = t$  (длина волны равна горизонту);  $NK$  — длина волны Джинса тяжелых нейтрино;  $QD$  — длина волны Джинса обычного вещества.  $A$  — область  $\lambda > t$ , в которой возмущения метрики постоянны, возмущения плотности всех типов материи (нейтрино, обычное вещество, излучение) равномерно растут, достигая величины

$$\left(\frac{\delta\rho}{\rho}\right)_\nu \approx \left(\frac{\delta\rho}{\rho}\right)_m \approx \left(\frac{\delta\rho}{\rho}\right)_R \approx h$$

на границе, т. е. на линии  $RNS$ ;  $t_m$  — момент, когда нейтрино становятся нерелятивистскими;  $NK$  — линия  $\lambda = v_\nu t$  ( $v_\nu$  — скорость нейтрино);  $t_1$  — момент отключения вещества и излучения ( $p + e^+ \rightarrow H$ ). В областях  $B_1$ ,  $B_2$  и  $B_3$  возмущения в нейтринном газе затухают;  $B_1$  — период, когда нейтрино релятивистские;  $B_2$  — нерелятивистские нейтрино в ионизированной плазме;  $B_3$  — нерелятивистские нейтрино в нейтральном водородном газе и излучении.  $C_1$  и  $C_2$  — области, в которых возмущения плотности нейтрино нарастают. В областях  $B_2$  и  $C_1$  плазма неподвижна; в областях  $B_3$  и  $C_2$  нейтральное вещество увлекается гравитационным полем нейтрино.  $t_1$  — время, за которое наиболее долго возрастающее возмущение с длиной волны  $\lambda_{J\nu}$  (11.10) достигает величины порядка 1. Значения  $t$  и  $\lambda$  представлены в логарифмическом масштабе. Изображение, разумеется, схематично. После момента  $t_1$  в области  $C_2$  в веществе возникают возмущения за счет гравитационного воздействия нейтринных возмущений. Однако из-за поглощения в областях  $B_2$  и  $B_3$  к этому моменту выживают только возмущения в плотности нейтрино с  $\lambda > \lambda_{J\nu}$

$= 30$  эВ), чем в барионной Вселенной. Дело в том, что флуктуации плотности нейтрино начинают расти, когда нейтрино становятся нерелятивистскими, что происходит при  $z \approx 10^4$  (при  $m_\nu \approx 30$  эВ). Напомним, что для барионов рост флуктуаций начинается в момент рекомбинации, т. е. при  $z = 10^3$ .

Возникновение флуктуаций  $\delta T/T$  начинается только с момента просветления  $z = 10^3$  и продолжается вплоть до наших дней. Рост возмущений плотности в нейтринной компоненте в промежутке от  $z = 10^4$  до  $z = 10^3$  почти не оказывает влияния на анизотропию реликтового излучения (§ 9). В силу этого флуктуации плотности нейтрино могут стать в 10 раз больше, чем в чисто барионной Вселенной, при одном и том же  $\delta T/T$ .

Отметим, что наблюдаемая структура Вселенной проявляется только в форме обычной барионной материи, так как только

эта материя может образовывать звезды и другие небесные тела, которые испускают детектируемое нами электромагнитное излучение. Нейтрино, составляющие  $\sim 90\%$  по массе, невидимы, но следы скучивания (кластеризации) нейтрино мы видим в неоднородностях обычной материи, которая за счет гравитационного взаимодействия сваливается в потенциальную яму, уготованную ей нейтрино. В этом смысле наблюдаемая структура Вселенной является вторичным эффектом по отношению к движению и кластеризации нейтрино.

Если изложенная выше картина верна, то роль нейтрино в мироздании трудно переоценить. Долго ускользавшая от экспериментаторов частица, даже пошатнувшая веру в неизбежность закона сохранения энергии, возможно, составляет 90% (по массе) всего вещества во Вселенной, и без нее не могли бы образоваться ни звезды, ни галактики, ни планеты с жизнью на них.

## § 6. БЛИНЫ, ЯЧЕЙСТАЯ СТРУКТУРА

Замечательным свойством рассматриваемой модели является предсказание об образовании ячеистой структуры Вселенной, когда области с малой плотностью вещества разделены относительно тонкими стенками с повышенной плотностью. Этот результат основан на гипотезе о малости давления вещества и подавлении коротковолновых флуктуаций. Он находится в согласии с астрономическими наблюдениями, указывающими на то, что скопления галактик образуют поверхности толщиной порядка 10–20 Мпк, между которыми находятся пустые области с характерным размером 100–200 Мпк.

В космологической ситуации мы имеем дело со случайными начальными возмущениями. Они приводят к стохастическому поведению растущих флуктуаций плотности. Вероятность сжатия или расширения вещества относительно какой-то оси равна  $1/2$ . При подходящем выборе осей движение вдоль каждой оси можно считать независимым и соответственно эта вероятность не зависит от движения вещества вдоль других осей. Следовательно, вероятность одновременного сжатия относительно всех трех осей равна  $1/8$ . Разумеется, это только грубая оценка. Учет нелинейных эффектов уменьшает ее почти вдвое. Концентрация вещества в таких областях должна быть максимальной. Но даже и в этом случае скорость сжатия по различным осям различна. Таким образом, движение вдоль оси наиболее быстрого сжатия приводит к образованию двумерных структур-блинов. Позднее эти блины преобразуются в сверхскопления галактик.

Следующий вопрос: почему области повышенной плотности образуют ячеистую структуру? Вероятность сжатия относитель-

но какой-либо из осей равна  $7/8$ . В газе с нулевым давлением сжатие даже относительно только одной оси приводит к столкновению траекторий частиц, к образованию каустической поверхности с бесконечной плотностью. Следовательно, около 90% вещества окажется в области высокой плотности.

Рассмотрим поверхность начальных положений частиц для простоты в двумерном случае (рис. 52). Имеются три типа различных областей: (a) — области, в которых сжатие идет по двум направлениям; (b) — области, в которых по одному направлению идет сжатие, а по другому — расширение, и (c) — области, расширяющиеся во все стороны. Области (a) и (c) каждая заполняют по 25% всей поверхности. Очевидно, они представляют собой острова в море (b).

Рассмотрим, что происходит при движении частиц. Области (a) сжимаются во всех направлениях, области (b) сжимаются в одном направлении и расширяются в другом, т. е. море (b) превращается в систему каналов. Области (c) растут во всех направлениях; поверхность, занимаемая (c), возрастает.

Существенным пунктом является непрерывность движения. Две близкие точки остаются близкими. Замкнутая кривая оста-

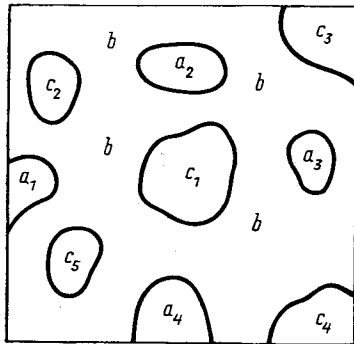


Рис. 52. Первоначальное распределение областей со сжатием в одном направлении:  $a_1, a_2, a_3, a_4$ ; сжатием в одном направлении и расширением в другом:  $b$ ; расширением в обоих направлениях:  $c_1, c_2, c_3$

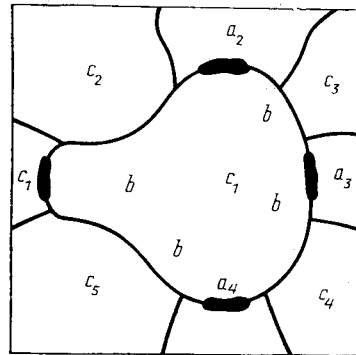


Рис. 53. Результат эволюции состояния, изображенного на рис. 52, приведшей к образованию каустик

ется замкнутой. В частности, замкнутые кривые, отделяющие (c) от (b), остаются замкнутыми. В итоге мы приходим к картине, изображенной на рис. 53, т. е. как раз к наблюдаемой клеточной структуре.

Возникновение блинов (или каустик) можно описать математически с помощью следующих простых рассуждений. Пространственная координата частицы в растущей моде возмущения может быть записана в виде

$$r(\xi, t) = a(t)\xi + b(t)\text{grad}_\xi \psi(\xi), \quad (10.12)$$

где  $\xi$  — лагранжева координата,  $a(t)$  — масштабный фактор,  $a(t) \sim t^{2/3}$  на нерелятивистской стадии расширения для  $\Omega=1$ . В растущей моде  $b(t)$  растет со временем быстрее, чем  $a(t)$ :  $b(t) \sim t^{4/3}$ . Первое слагаемое в выражении (10.12) отвечает невозмущенному движению частиц, а второй член описывает безвихревое растущее возмущение. Функция  $\psi$  является гладкой функцией координат, так как высокочастотные компоненты вымирают в релятивистской плазме.

Выражение для плотности материи  $\rho$  легко найти с помощью соотношения

$$dm = \rho d^3x = \rho_0 d^3\xi, \quad (10.13)$$

или  $\rho = \rho_0 / \det[\partial r / \partial \xi]$ , где  $\det[\partial r / \partial \xi]$  — якобиан перехода от переменных  $\xi$  к  $r$ .

Направление осей каждой частицы можно выбрать так, что смешанные производные исчезнут:

$$\frac{\partial^2 \psi}{\partial \xi_1 \partial \xi_2} = \frac{\partial^2 \psi}{\partial \xi_2 \partial \xi_3} = \frac{\partial^2 \psi}{\partial \xi_3 \partial \xi_1} = 0.$$

Вводя следующие обозначения для остающихся вторых производных

$$\frac{\partial^2 \psi}{\partial \xi_1^2} = -\alpha, \quad \frac{\partial^2 \psi}{\partial \xi_2^2} = -\beta, \quad \frac{\partial^2 \psi}{\partial \xi_3^2} = -\gamma,$$

получим

$$\rho(t) = \frac{\bar{\rho}(t)}{\left(1 - \frac{b}{a}\alpha\right)\left(1 - \frac{b}{a}\beta\right)\left(1 - \frac{b}{a}\gamma\right)}. \quad (10.14)$$

Пусть оси выбраны так, что  $\alpha > \beta > \gamma$ . Тогда плотность обращается в бесконечность в момент времени, когда  $ab(t)/a(t) = 1$ . Так как, вообще говоря, величины  $\alpha$ ,  $\beta$  и  $\gamma$  — различны, то только один из множителей в знаменателе выражения (10.14) обращается в ноль. Физически это означает, что сжатие происходит лишь вдоль одной оси, т. е. образуется тонкая поверхность с большой плотностью.

Разумеется, мы привели только качественные соображения. Изложение строгой математической теории этих явлений выходит за рамки данной книги. Упомянем только, что возник-



новение сеточной структуры при стохастическом процессе каждый из нас наблюдал, рассматривая световые блики на отраженной к воде стороне моста.

Изложенная здесь схема еще далека до своего полного количественного анализа. В частности, пока нельзя считать решенным вопрос перехода газообразной материи в звезды, галактики. В лучшем случае теория указывает на некоторые общие свойства распределения светящейся материи. Напомним еще раз, что эти свойства являются следствием движения с пренебрежимо малым давлением, что в свою очередь связано с подавлением коротких волн. Эта картина свидетельствует в пользу адиабатических флуктуаций в период радиационной доминантности. Очевидно, что все сказанное здесь справедливо и для нейтринной Вселенной.

Как мы уже отмечали, проблема флуктуаций температуры реликтового фона заметно смягчается в модели с массивными нейтрино, что в совокупности с развитием ячеистой структуры и нейтринной Вселенной является веским аргументом в ее пользу. Однако не следует делать все ставки на одну лошадь, поэтому ниже мы обсудим другие возможные формы скрытой массы, а также другой тип флуктуаций.

В заключение отметим, что рассмотренная здесь ячеистая структура не является последним этапом в эволюции. Сами «блины» гравитационно неустойчивы относительно движений в плоскости. Они распадаются на галактики, собираются в нити, в узлы. К моменту, когда границы областей на рис. 53 замкнутся, часть области уже может быть разрушена. Блины не являются конечной стадией эволюции. Позднее эволюция должна привести к большим несвязанным кластерам. То, о чем мы говорили, и то, что мы видим на небе, является промежуточной фазой, промежуточной асимптотикой. С этой точки зрения наша Вселенная еще очень молода.

### § 7. СТРУКТУРА ВСЕЛЕННОЙ В МОДЕЛЯХ СО СКРЫТОЙ МАССОЙ

Мы уже говорили, что наиболее естественным кандидатом на роль носителя скрытой массы Вселенной является нейтрино, и если данные о том, что масса нейтрино достигает величины  $\sim 20$  эВ (эксперимент ИТЭФ), подтвердятся, вопрос будет безоговорочно решен в пользу нейтринной модели. Однако мы пока не можем быть уверены в том, что исключены другие варианты. В отличие от нейтрино, которое заведомо существует, другие возможные носители скрытой массы являются чисто гипотетическими, их существование с той или иной степенью достоверности предсказывается (или, может быть, лучше сказать, разрешается) теорией, но прямые экспериментальные до-

казательства их существования отсутствуют. Общим свойством этих гипотетических частиц является либо очень слабое взаимодействие, либо довольно большая масса, что затрудняет их непосредственное наблюдение в лабораторных экспериментах.

Не вдаваясь в физическую природу носителей скрытой массы, в литературе ее делят на три типа: горячая, теплая и холодная.

Носителями *горячей* скрытой массы являются частицы, которые были в тепловом равновесии с первичной плазмой вплоть до  $T \lesssim 100$  МэВ, и поэтому их количество по порядку величины такое же, как количество реликтовых фотонов. В силу этого их масса ограничена величиной в несколько десятков электронвольт. Термин «горячая» связан с тем, что в момент выключения взаимодействия с плазмой эти частицы являются релятивистскими. Остывая при расширении Вселенной, они в какой-то момент становятся нерелятивистскими, и если их плотность энергии превосходит плотность энергии барионов (а именно этот случай и представляет интерес), то они начинают доминировать во Вселенной. Случай горячей скрытой массы мы уже рассмотрели на примере нейтрино и не будем к нему возвращаться. Напомним только, что в случае горячей скрытой массы подавлены флуктуации меньше размера скоплений галактик (см. (10.10) и (10.11)).

Для частиц *холодной* скрытой массы минимальный размер флуктуаций меньше размера галактик, и поэтому затухание возмущений за счет свободного разлетания этих частиц для космологии не существенно. Примером носителя холодной скрытой массы мог бы быть тяжелый нейтральный лептон с массой выше нескольких ГэВ (в соответствии с космологическим ограничением). В момент отключения эти частицы являются нерелятивистскими, и с этим связан термин «холодная». Концентрация реликтовых тяжелых лептонов должна быть, по крайней мере в  $10^8$  раз ниже концентрации реликтовых нейтрино.

Более популярным вариантом холодной скрытой массы является когерентно осциллирующее аксионное поле. Аксион-гипотетическая частица, существование которой обеспечило бы естественное сохранение  $CP$  ( $C$  — зарядовая, а  $P$  — пространственная четности) в сильных взаимодействиях. Первоначальная модель с относительно тяжелым аксионом ( $m_a = 0,1 - 1$  МэВ) была довольно быстро закрыта экспериментом, и в настоящее время рассматривается вариант почти стерильного легкого аксиона ( $m_a \approx 10^{-5}$  эВ). Как раз последний случай может быть существенным для космологии.

*Теплая* скрытая материя занимает промежуточное место между горячей и холодной; для нее затухание возмущений из-за свободного разлета начинается на галактическом масшта-

бе. Носителями теплой скрытой массы могут быть частицы с массой около 1 кэВ, взаимодействующие гораздо слабее, чем нейтрино. Из-за относительно большой массы их концентрация должна быть на несколько порядков ниже концентрации реликтовых фотонов.

К тому моменту, когда частицы теплой скрытой массы (как, впрочем, и холодной) становятся нерелятивистскими, их плотность энергии все еще относительно мала, и во Вселенной доминирует релятивистское вещество. В этом состоит важное отличие от случая горячей скрытой массы. Флуктуации плотности релятивистской материи с длиной волны меньше горизонта, как мы отмечали, на этом этапе не растут, а осциллируют. Так как релятивистская материя доминирует в общей плотности энергии, а вклад невидимой материи мал, то это приводит к замораживанию роста флуктуаций в скрытой массе, который возобновляется лишь на стадии нерелятивистской доминантности.

В модели с теплым скрытым веществом сначала образуются галактики и малые скопления галактик, а позже структуры больших размеров.

В принципе в моделях с теплой и/или холодной скрытой материей, по-видимому, можно получить удовлетворительное описание крупномасштабной структуры Вселенной, однако окончательного, полностью надежного расчета ни в одной из моделей (включая и горячую нейтринную) пока не проделано.

Заметим, что в сценарии инфляционной Вселенной предсказывается  $\Omega=1$ , однако астрономические наблюдения указывают на  $\Omega=0,2-0,3$ . Выход из этого противоречия состоит в том, что, помимо кластеризованной материи, имеется однородно распределенный фон с  $\Omega_{\phi}=0,7-0,8$ , к которому астрономические наблюдения не чувствительны. Физическим источником этого фона могли бы быть распады очень долгоживущих частиц скрытой массы ( $\tau=10^5-10^7$  лет), или космологическая постоянная. В многокомпонентной модели такого рода можно получить лучшее количественное описание структуры, чем, скажем, в простой нейтринной модели.

Говоря о количественном описании структуры, мы пока не упомянули о том, какими методами это делается. А эта задача не столь уж тривиальна. Одним из способов является использование корреляционных функций, т. е. пространственных средних произведения плотности материи в разных точках  $\bar{\rho}(x)\bar{\rho}(y)$ . Таким образом, можно количественно определить, каков характерный размер космических скоплений, но он ничего не говорит об их форме, в частности не позволяет различить клеточную структуру и отдельные сгустки. Красивый способ, не обладающий этим недостатком, был предложен С. Ф. Шандариным. Он основан на использовании понятия

перколяции и состоит в следующем. Рассмотрим как-то распределенные в пространстве объекты и окружим (мысленно) каждый объект сферой радиуса  $R$ . Объекты будем называть объединенными, если один из них находится внутри сферы, окружающей другой. Теперь будем увеличивать  $R$  до тех пор, пока объединенные объекты не образуют цепи, проходящие через все пространство. Соответствующее этому критическое значение радиуса обозначим  $R_c$ . Характерной безразмерной величиной, которая используется в нашем анализе, является среднее число объектов внутри сферы радиусом  $R_c$ :

$$N = \frac{4\pi}{3} R_c^3 n, \quad (10.15)$$

где  $n$  — средняя плотность объектов.

Если для однородно распределенных объектов  $R_c$  и  $N$  имеют какие-то значения  $R_{c0}$  и  $N_0$ , то для объектов, образующих отдельные сгустки,  $R_c$  и  $N$  окажутся больше  $R_{c0}$  и  $N_0$ . Напротив, в случае клеточной структуры  $R_c$  и  $N$  будут меньше. Таким образом,  $R_c$  и  $N$  отклоняются в разные стороны от однородного случая для кластеризованных объектов и для ячеистой структуры. Анализ распределения галактик с помощью этого критерия свидетельствует в пользу клеточной структуры Вселенной, что характерно для адиабатических флуктуаций и противоречит простейшему варианту изотермических.

## § 8. ИЗОТЕРМИЧЕСКИЕ ФЛУКТУАЦИИ

В изотермических возмущениях температура и плотность излучения однородны, а плотность жестко связанного с радиацией вещества (до момента рекомбинации) варьируется. Для таких возмущений все длины волн вплоть до горизонта не затухают на всей радиационно доминированной стадии. Более того, имеется специфический механизм роста флуктуаций, связанный со следующим процессом. По мере расширения Вселенной области с избытком барионов остывают несколько скорее, чем с недостатком, из-за передачи энергии от излучения к нерелятивистским частицам. Излучение диффундирует в область более низкой температуры, увлекая связанные с ним барионы, что приводит к дальнейшему росту  $\delta\rho_b$ .

После рекомбинации положение меняется. Флуктуации вещества с длиной волны больше, чем  $\lambda_J$  (10.1), продолжают нарастать из-за гравитационной неустойчивости, а флуктуации меньшего масштаба осциллируют, быстро затухая из-за диффузии фотонов. Возмущения достигают величины порядка единицы быстрее всего на шкале порядка  $\lambda_J$ , что отвечает массе около  $10^8 M_{\odot}$ . Это характерная величина массы шарового звездного скопления. Объекты большего размера образуются

в этой теории позже за счет сгущивания более мелких образований в противоположность обычной адиабатической теории, где процесс прямо обратный. Заметим, впрочем, что в адиабатической теории, когда масса гипотетических невидимых частиц равна 1–2 кэВ, спектр возмущений также обрезается на  $M \approx 10^6 M_\odot$ , и поэтому положение дел в такой модели сильно напоминает изотермический случай.

Завершим параграф замечанием, что теория образования структуры Вселенной переживает сейчас период бурного развития. Можно считать установленным, что простейшие однокомпонентная модель с плоским спектром флуктуаций не согласуется с наблюдениями. Исследуются более сложные модели с различными типами скрытой массы, с нетривиальным спектром флуктуаций, проводится численное моделирование систем с большим числом частиц. К сожалению, прогресс в этой области замедляется из-за того, что физика частиц дает слишком много вариантов скрытой массы, но ни одного твердо установленного. То же относится и к спектру флуктуаций. Как мы увидим ниже, в простейшем варианте инфляционного сценария и бариосинтеза возникают адиабатические флуктуации с плоским спектром. Однако имеются интересные модели, приводящие и к изотермическим флуктуациям с выделенным масштабом в спектре. Наличие таких флуктуаций серьезно повлияло бы на формирование структуры. Тем не менее, несмотря на все трудности, наше непонимание сейчас носит количественный, но не качественный характер, и, видимо, уже в ближайшем будущем грандиозная задача описания крупномасштабной структуры Вселенной будет решена.

### § 9. ФЛУКТУАЦИИ РЕЛИКТОВОГО ИЗЛУЧЕНИЯ

Стандартная модель горячей Вселенной предсказывала существование космического радиоизлучения, обладающего планковским спектром.

Любопытно, что Гамов (1946), сделавший это предсказание, получил оценку температуры этого излучения  $T=6\text{К}$ , довольно близкую к реальности, основываясь на представлениях, которые оказались ошибочными. При принятом в то время значении постоянной Хаббла  $H=564 \text{ км/с Мпк}$  (оценка самого Хаббла) возраст Вселенной оказывался слишком малым ( $t_0 \approx 2 \cdot 10^9$  лет), недостаточным для синтеза тяжелых элементов в звездах. Поэтому Гамов полагал, что тяжелые элементы также синтезируются во время первичного расширения и оценивал температуру, необходимую для сохранения изотопов с большим сечением захвата медленных нейтронов.

Предсказание радиоизлучения космологического происхождения было подтверждено в 1965 г. открытием изотропного из-

лучения с планковским спектром и температурой  $T \approx 3 \text{ К}$ . Оно было открыто при испытании новой приемной радиоаппаратуры космической связи Пензиасом и Вильсоном на волне 7,3 см.

Группа под руководством Дикке целенаправленно готовила эксперимент по поискам реликтового излучения на длине волны 3 см. Однако экспериментальные данные были ими получены несколько позже, чем предыдущей группой.

Еще раньше А. Г. Дорошкевич и И. Д. Новиков (1964) построили суммарный спектр всех источников в нескольких вариантах с реликтовым фоном той или иной температуры. Оказалось, что в области длин волн около 1 см наиболее целесообразны поиски реликтового излучения, здесь минимальна сумма излучения дискретных источников, обладающая максимумами в оптическом диапазоне и более длинноволновом радиодиапазоне.

Отметим также, что еще в 1940 г. было отмечено аномальное возбуждение молекул CN в межзвездном газе. Энергия возбужденного состояния больше энергии основного состояния на величину, равную энергии фотона с  $\lambda=0,264 \text{ см}$ , а степень возбуждения соответствует температуре около 3 К. Анализ показывает, что такое возбуждение возможно лишь при наличии изотропного фона излучения. В настоящее время спектр реликтового излучения промерен в широком интервале длин волн, отступлений от планковского спектра не найдено.

В первичной плазме при высокой температуре в равновесии находились все частицы, в частности фотоны. Эти фотоны сейчас мы воспринимаем как реликтовое излучение. Равновесие фотонов с остальными составляющими плазмы нарушается при  $z_{\text{рек}}=1000$ , и они доходят до нас, распространяясь практически свободно. Здесь следует подчеркнуть, что само расширение Вселенной не меняет спектр чернотельного излучения, плотность энергии фотонов адиабатически уменьшается с сохранением формы спектра.

Инфляция в ранней Вселенной приводит к высокой однородности в распределении вещества, а следовательно, и фотонов. Такой вывод из теории инфляции полностью подтверждается наблюдениями. Однородность в распределении фотонов в прошлом приводит к изотропии реликтового излучения сейчас. Поясним это утверждение рис. 54. Различные области A, B, C представляют собой разные области пространства, лежащие на одинаковом расстоянии от наблюдателя O. А именно сейчас они лежат на сфере с радиусом  $r=^3/2 H^{-1}$ , равным современному горизонту частиц. Лучи, которые доносят сведения о физических параметрах A, B, C, состоят из реликтовых фотонов. Наблюдатель увидит все три области лежащими в разных направлениях; другими словами, одинаковые пространственные свойства плазмы для наблюдателя O проявятся в свойстве изотропии излучения.

Любопытно отметить один важный наблюдательный аргумент в пользу теории инфляции.

В первой главе при рассмотрении свойств горизонта в решении Фридмана мы обнаружили, что различные области пространства, разделенные расстояниями большими, чем  $H^{-1}$ ,

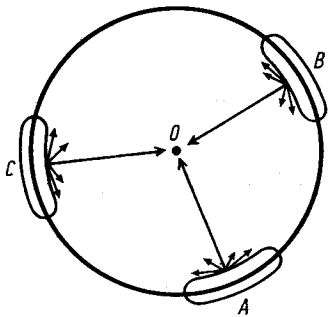


Рис. 54. Три различные области A, B, C, лежащие на горизонте наблюдателя O

никогда в прошлом не взаимодействовали. Значит к моменту рекомбинации в стандартной фридмановской модели причинно связанными между собой оказывались только те области, размер которых не превышал  $ct_{\text{рек}}$ . Внутри пространственных областей такого размера должна была установиться примерно одинаковая температура. Различные области могли бы сильно отличаться по свойствам, поскольку никогда в прошлом не взаимодействовали. В частности, они могли бы обладать различной температурой реликтового излучения. Угловой масштаб сильной анизотропии  $T$  на небе легко вычислить. Угол под которым сейчас видна область размером, равным размеру горизонта на стадии рекомбинации, сейчас составляет

$$\theta = t_{\text{рек}} H_0 \frac{(1 + z_{\text{рек}})^2}{2} (z_{\text{рек}} + 1 - \sqrt{z_{\text{рек}} + 1})^{-1} \approx 2^\circ.$$

Наблюдения показывают отсутствие сильной анизотропии на угловом масштабе в несколько градусов. Этот факт можно трактовать как прямое следствие инфляционной стадии.

Несмотря на высокую степень изотропии реликтового излучения, в распределении потока (или, что эквивалентно, температуры  $T$ ) по небу должны присутствовать угловые флуктуации величины потока, или, как их чаще называют, флуктуации температуры  $\delta T/T$ . Нам известны по крайней мере четыре основных механизма, которые должны вызывать угловые флуктуации температуры. На больших угловых масштабах основной причиной, вызывающей анизотропию, является эффект Сакса — Вольфа. Он заключается в изменении частоты фотонов в гравитационном поле.

Поясним подробнее. После того как фотон проходит разность гравитационных потенциалов  $\Delta\phi$ , его частота  $\nu$  меняется на величину

$$\frac{\delta\nu}{\nu} \approx \Delta\phi.$$

Для ансамбля фотонов, обладающих планковским спектром, сдвиг частоты каждого фотона эквивалентен изменению температуры спектра:

$$\frac{\delta T}{T} = \frac{\delta\nu}{\nu}.$$

Значит в неоднородном и переменном гравитационном поле температура излучения будет зависеть от направления.

На меньших угловых масштабах основной вклад в вариации температуры дает так называемый эффект Силка. Он возникает для адиабатических флуктуаций плотности. Если энтропия (т. е. отношение числа барионов к числу фотонов) постоянна по пространству, то флуктуации плотности приводят к флуктуациям числа фотонов по пространству. Из равенства

$$n_b/T^3 = \text{const}$$

следует

$$\frac{\delta T(r)}{T} = \frac{1}{3} \frac{\delta\rho(r)}{\rho}.$$

Неоднородность в распределении температуры в момент рекомбинации приведет к анизотропии  $\delta T/T$  сейчас.

Наряду с изменениями плотности важную роль играет и доплер-эффект пекулярного движения вещества, т. е. случайного движения, наложенного на среднее хэббовское расширение. Заметим, что в силу законов неразрывности рост возмущений не возможен без соответствующих пекулярных скоростей. Распределение галактик по пекулярным скоростям при разных  $z$  станет важным источником информации о первичных возмущениях.

Не связанным непосредственно с возмущениями метрики в эпоху рекомбинации, но важным для определения эволюции Вселенной при  $z=5-10$  является следующий эффект, который тоже приводит к изменению температуры (Зельдович, Сюняев, 1982). Он состоит в том, что горячий газ электронов на лучезрения из-за комптоновского рассеяния передает свою энергию реликтовым фотонам и изменяет их температуру. Угловая неравномерность в распределении горячего газа приводит к флуктуациям температуры реликтового излучения по небу.

Этот эффект позволяет сделать выбор при сравнении двух теорий образования крупномасштабной структуры Вселенной: теории блинов и теории взрывной неустойчивости. Теория взрывной неустойчивости описывает возникновение ячеистой структуры Вселенной из энтропийных возмущений. Основные этапы в ней таковы. Из энтропийных возмущений образуются первозвезды с чудовищной (до  $10^2 M_\odot$ ) массой. Взрываясь, эти звезды рожают сильную ударную волну плотности. Она ини-

цирует новые взрывы и порождает сетчатую структуру в крупномасштабном распределении материи. Как показывают расчеты (Хоган, 1984), образование сетчатой структуры серий последовательных взрывов приводит к слишком большим значениям  $\delta T/T$  именно из-за упомянутого эффекта. Поэтому сегодня предпочтительнее выглядит теория блинов.

Наконец, четвертым из основных физических механизмов, вызывающих анизотропию  $\delta T/T$ , является доплер-эффект. Электрон, переизлучающий фотон, меняет его частоту согласно  $\delta\nu/\nu \sim v$ , где  $v$  — величина, на которую отличается скорость электрона от хэббловской в данном месте. В направлении на облако горячих электронов мы увидим изменение температуры реликта на уровне

$$\frac{\delta T}{T} \sim v\tau_r;$$

здесь теперь  $v$  — пекулярная радиальная скорость облака (полная скорость, измеряемая по красному смещению, составляет  $H_0 R + v$ ), а  $\tau_r$  — оптическая толщина облака по томсоновскому рассеянию.

Для получения информации о параметрах ранней Вселенной значительно более важным выглядит образование  $\delta T/T$  в эпоху рекомбинации, когда угловой спектр  $\delta T/T$  адекватно отражает пространственный спектр возмущений плотности и гравитационных волн, порожденных в инфляционную эпоху (см. гл. 12).

Здесь механизмов образования  $\delta T/T$  всего два: эффект Сакса — Вольфа и эффект Силка. Остановимся на них подробнее.

Рассмотрим вновь однородную и изотропную метрику со слабыми отклонениями от фридмановского решения в виде

$$ds^2 = a^2(\eta) (\eta_{\alpha\beta} + h_{\alpha\beta}) dx^\alpha dx^\beta.$$

В эпоху просветления каждый объем первичной плазмы «вспыхивает» только один раз и на короткое время  $\Delta t$ , когда оптическая толщина плазмы по томсоновскому рассеянию меняется, грубо говоря, с  $\tau \approx 1$  до  $\tau \approx 0,1$ . Обозначим этот момент времени  $\eta_e$ . Момент наблюдения  $\eta_0$  выражается через  $\eta_e$  как  $\eta_0 = \sqrt{1 + z_{рек}} \eta_e \approx 30\eta_e$ . Наблюдаемая температура будет зависеть от температуры в момент излучения как

$$T_0 = T_e \left( \frac{\eta_e}{\eta_0} \right)^2 \left( 1 + \frac{\delta T}{T} \right),$$

где

$$\frac{\delta T}{T} = \frac{1}{2} \int \left( \frac{dh_{ij}}{d\eta} e^i e^j - 2 \frac{\partial h_{0i}}{\partial \eta} e^i \right) d\lambda, \quad (10.16)$$

интегрирование по параметру  $\lambda$  вдоль луча зрения производит-

ся от  $\eta_e$  до  $\eta_0$ . Вектор  $e^i = (\sin \rho \cos \varphi, \sin \rho \sin \varphi, \cos \rho)$  определяет направление наблюдения.

Выражение (10.16) является основным при вычислении эффекта Сакса — Вольфа. Растущая мода возмущений плотности выражается через  $\eta$  (ср. (10.5)) как

$$\frac{\delta \rho}{\rho} = \frac{1}{2} h(k\eta)^2 e^{i(kx + \xi)}; \quad (10.17)$$

здесь молчаливо предполагается зависимость амплитуды  $h$  от волнового вектора  $k$  и условие  $k\eta_{рек} < 1$ . Для  $k\eta_{рек} > 1$  связь  $\delta \rho$  с  $h$  другая. Но этот диапазон  $k$  не особенно интересен, поскольку скалярные возмущения с длиной волны меньше горизонта  $k\eta > 1$ , быстро затухают.

Подстановка (10.17) в (10.16) дает возможность найти  $\delta T/T$  в виде

$$\frac{\delta T}{T} = h e^{i(\xi + \frac{\pi}{2})} [k\eta_0 \cos \theta - k\eta_e \cos \theta e^{ik(\eta_0 - \eta_e) \cos \theta} + 1 - e^{ik(\eta_0 - \eta_e) \cos \theta}].$$

Наиболее прозрачна связь между флуктуациями плотности вещества и флуктуациями температуры в пределе очень длинных волн, т. е.  $k\eta_0 \ll 1$ . Тогда

$$\left( \frac{\delta T}{T} \right)_0 = \left( \frac{\delta \rho}{\rho} \right)_0 \cos^2 \theta;$$

Здесь  $\theta$  — угол между направлением наблюдения и волновым вектором волны плотности. Для возмущений плотности, чьи длины волн лежат в пределах  $\eta_e < k^{-1} < \eta_0$ , зависимость между  $\delta T$  и  $\delta \rho$  сложнее.

Угловой спектр  $\delta T/T$  при разложении по сферическим функциям, т. е. в ряд вида

$$\frac{\delta T(\theta, \varphi)}{T} = \sum_{n,m} C_{n,m} Y_n^m(\theta, \varphi),$$

отражает пространственный спектр  $h(k)$ . В предположении однородности и изотропности флуктуаций их пространственный спектр  $h(k)$  однозначно определяет угловой спектр  $\delta T/T$ . Для плоского спектра  $h(k) = Ak^{-3/2}$  эта зависимость выглядит как (Старобинский, 1983)

$$\langle C_{nm}^2 \rangle = \frac{A^2}{100\pi n(n+1)}.$$

Амплитуда спектра  $A$  определяет флуктуации плотности согласно  $\delta \rho/\rho = Ak^2 \eta^2 / (20k^{3/2})$  и момент формирования первых структур, т. е. достижения условия  $\delta \rho \approx \rho$ .

Гравитационные волны тоже посредством эффекта Сакса — Вольфа вызывают флуктуации температуры реликта. Возмущения метрики в поле плоской монохроматической волны имеют вид

$$h_{\mu\nu} = -\frac{1}{3} B_{\mu\nu} k^3 \left[ 1 - \frac{(k\eta)^2}{10} + \dots \right] e^{i(kr + \xi)}.$$

Здесь матрица  $B_{\mu\nu}$  является бесследовой и поперечной, она описывает тензорную структуру волны. Флуктуации температуры реликтовых фотонов для  $k\eta_0 < 1$  описываются выражением

$$\frac{\delta T}{T} = \frac{k^3}{60} \sin^2 \theta (C_1 \cos 2\varphi + C_2 \sin 2\varphi) (k\eta_0)^2 \times \\ \times \left[ \left( 1 - \frac{\eta_e^2}{\eta_0^2} \right) \cos \xi - k\eta_e \cos \theta \left( 1 - \frac{\eta_e}{\eta_0} \right)^2 \right].$$

Здесь  $C_1, C_2$  — амплитуды двух независимых состояний поляризации гравитационных волн.

Как возмущения плотности, так и гравитационные волны вызывают анизотропию реликтового излучения. Различить, что вызывает флуктуации  $\delta T/T$ , можно по поляризации реликтового излучения (Сажин, 1984).

Поиски анизотропии реликтового излучения позволили обнаружить дипольную составляющую в распределении интенсивности по небу (Смут, 1977). Сейчас дипольная гармоника интерпретируется как следствие движения Галактики, а вместе с ней и Солнечной системы сквозь реликтовое излучение внутри локального сверхскопления и движение его относительно ближайших сверхскоплений. Эта гармоника соответствует анизотропии на самых крупных угловых масштабах  $\sim 180^\circ$ . Попытки обнаружить анизотропию на других угловых масштабах к успеху не привели. Каждый раз результат был отрицательный, определялся только верхний предел. Сейчас наиболее сильный верхний предел на амплитуду низких гармоник наложен московской группой в эксперименте «Реликт» [18]. Так, для квадрупольной составляющей

$$(\delta T/T)_q \leq 3 \cdot 10^{-5}.$$

Такое малое значение коэффициентов анизотропии в совокупности с развитой структурой на больших масштабах является сильным наблюдательным аргументом в пользу наличия невидимого вещества, которое рассматривалось выше (§ 5).

## Глава 11.

### КВАНТОВАЯ ТЕОРИЯ И ПРОИСХОЖДЕНИЕ ФЛУКТУАЦИЙ

Наблюдаемая структура Вселенной возникла в результате развития первоначально небольших флуктуаций плотности. До последнего времени вопрос о происхождении этих флуктуаций оставался открытым. Было совершенно неясно, какие физические процессы могут привести к флуктуациям на гигантских, существенных для космологии расстояниях. И здесь важную роль сыграла инфляционная модель. Квантовые флуктуации, которые обычно разыгрываются на микроскопических масштабах, в экспоненциально расширяющейся вселенной гигантским образом увеличивают свою длину волны  $\lambda \rightarrow \lambda \exp(Ht)$ , а также (правда, существенно меньше) и амплитуду и становятся космологически значимыми. В силу этого галактики, их скопления и прочие структуры представляют собой макропроявления микроскопических, но сильно разросшихся квантовых флуктуаций.

#### § 1. ФЛУКТУАЦИИ СКАЛЯРНОГО ПОЛЯ

Как мы уже отмечали (§ 3 гл. 6), экспоненциальное расширение в инфляционной модели происходит во время медленного стремления пространственно однородного скалярного поля  $\varphi_0(t)$  к положению равновесия. Мы будем предполагать, что поле  $\varphi$  удовлетворяет уравнению Клейна — Гордона в метрике де Ситтера:

$$(\partial_t^2 + 3H\partial_t) \varphi - e^{-2Ht} \partial_i^2 \varphi = -\partial V / \partial \varphi, \quad (11.1)$$

где  $V(\varphi)$  — потенциал поля  $\varphi$  (в простейшем случае  $V = m^2 \varphi^2 - \lambda \varphi^4$ , а масштабный фактор имеет вид  $a = a_0 \exp(Ht)$ ;  $\varphi_0(t)$  является независимым от пространственных координат решением этого уравнения.

Как известно, на классическое поле  $\varphi_0(t)$  накладываются квантовые флуктуации  $\delta\varphi(\mathbf{r}, t)$ . Мы увидим, что из-за этих флуктуаций время «падения» поля  $\varphi_0(t)$  к положению равновесия изменяется на зависящую от координаты величину

$$\delta t(\mathbf{r}) = -\frac{\delta\varphi(\mathbf{r}, t)}{\dot{\varphi}_0(t)}. \quad (11.2)$$

Действительно, величина  $\delta\varphi(\mathbf{r}, t)$  удовлетворяет уравнению

$$\delta\ddot{\varphi} + 3H\delta\dot{\varphi} - e^{-2Ht} \partial_i^2 \delta\varphi = -\frac{\partial^2 V(\varphi_0)}{\partial \varphi^2} \delta\varphi. \quad (11.3)$$

При больших  $t$  роль члена, содержащего пространственные

производные, становится несущественной и  $\delta\varphi$  начинает удовлетворять тому же уравнению, что  $\varphi_0$ . Поэтому отношение двух решений  $\varphi_0$  и  $\delta\varphi$  при  $t \rightarrow \infty$  стремится к не зависящей от времени величине (11.2) или в первом порядке по  $\delta t$ :

$$\varphi(\mathbf{r}, t) = \varphi_0(t) + \delta\varphi(\mathbf{r}, t) \approx \varphi_0(t - \delta t(\mathbf{r})). \quad (11.4)$$

Из-за этого запаздывания переход от деситтеровского режима к фридмановскому в разных точках пространства происходит в разное время, что приводит к флуктуациям метрики

$$h \approx H \delta t(\mathbf{r}). \quad (11.5)$$

В действительности вопрос несколько сложнее из-за произвола в выборе временной координаты в пространстве де Ситтера. Однако, и здесь мы просим читателя поверить на слово или обратиться к упоминаемым ниже оригинальным работам, эта естественная размерная оценка величины флуктуаций метрики действительно справедлива. Физической причиной, выделяющей особую точку на оси времени, является изменение режима расширения.

В соответствии с общей теорией флуктуации метрики (11.5) приведут к таким же по величине относительным флуктуациям плотности в момент, когда длина волны флуктуации равна горизонту (см. гл. 10). Теперь вспомним, что  $\delta t$  связана с флуктуациями поля  $\varphi$  выражением (11.2). Квантовые флуктуации поля  $\varphi$  в пространстве де Ситтера (точнее, их отличие от квантовых флуктуаций в плоском пространстве-времени) имеют величину порядка  $H$ . Поэтому для флуктуаций плотности на горизонте получим оценку

$$\frac{\delta\rho}{\rho} \approx \frac{H^2}{\varphi_0(t)}. \quad (11.6)$$

Величину  $\varphi_0(t)$  можно найти с помощью уравнения (11.1). Так как мы знаем, что в существенный период поле  $\varphi_0(t)$  медленно меняется, то членом  $\varphi$  можно пренебречь, и для потенциала  $V = \lambda\varphi^4$ , который типичен для инфляционной модели при малых  $\varphi$ , получим

$$\varphi_0(t) = \left[ \frac{3H}{2\lambda(t_0 - t)} \right]^{1/2}, \quad (11.7)$$

здесь  $t_0$  — момент окончания инфляции. В действительности, когда  $\varphi_0(t)$  становится достаточно большим, поведение потенциала  $V(\varphi)$  меняется так, что  $\varphi_0(t)$  выходит на постоянное значение, однако для наших целей вполне достаточно приближения (11.7). Теперь, учитывая, что волна длиной  $H^{-1}$  вырастает в  $e^{Ht}$  раз и достигает длины  $k_0^{-1} = H^{-1}e^{Ht}$ , получим, что необходимое для этого время равно

$$\tau = H^{-1} \ln(H/k_0). \quad (11.8)$$

Отсюда нетрудно получить для амплитуды флуктуаций плотности на горизонте:

$$\left( \frac{\delta\rho}{\rho} \right)_H = \lambda^{1/2} \ln^{3/2}(H/k_0) \cdot \text{const}. \quad (11.9)$$

Постоянный множитель, который можно найти при точном расчете, составляет величину порядка 0,1. Отметим, что эти флуктуации чисто квантовые, они возникли в результате разрастания нулевых колебаний  $\hbar\omega/2$  (см. § 4 гл. 5).

В итоге мы получили приближенно плоский спектр адиабатических флуктуаций (с точностью до логарифмических членов), однако при естественных значениях  $\lambda$  величина флуктуаций оказывается слишком большой. Действительно, чтобы стать космологически значимой, длина волны должна вырасти в  $10^{20}$ — $10^{25}$  раз, что приводит к тому, что  $\ln(H/k) \approx 50$ , и если требовать  $(\delta\rho/\rho)_H < 10^{-4}$  в соответствии с изотропией реликтового излучения, то для  $\lambda$  получается весьма малое значение  $\lambda \leq 10^{-12}$ . Пока мы не имеем естественной модели для столь ничтожной величины  $\lambda$ .

Изложение здесь носит качественный полуинтуитивный характер. Желая познакомиться со строгим выводом этих результатов могут обратиться к оригинальным работам [44, 46, 67, 36, 57].

Итак, мы видим, что инфляционная модель может обеспечить (и даже с избытком) флуктуации, необходимые для образования структуры Вселенной. Приятно отметить, что при этом инфляция приводит к наблюдаемой однородности нашей Вселенной в большом масштабе, так как все первоначальные (доинфляционные неоднородности) во время деситтеровской стадии разглаждаются. Укажем, однако, что на очень больших масштабах, существенно превышающих сегодняшний горизонт, инфляционная модель (по крайней мере в варианте хаотической инфляции) предсказывает весьма существенные неоднородности. Случайные флуктуации скалярного поля над деситтеровском фоном приводят к его росту со временем [57, 70]:

$$\langle \varphi^2 \rangle = H^3 t / 4\pi^2, \quad (11.10)$$

если  $H \gg m$ , где  $m$  — масса поля  $\varphi$ . Этот рост можно рассматривать как случайное блуждание поля  $\varphi$  с шагом  $H$ . Как известно, проходимый при этом за  $N$  шагов путь пропорционален  $\sqrt{N}$ . По-другому это можно понять как результат расширения квантовых флуктуаций малого размера, когда все большее число мод вносит вклад в  $\langle \varphi^2 \rangle$ .

Если рассмотреть теперь сценарий хаотической инфляции с  $\varphi \gg m_{\text{Pl}}$ , то возникает парадоксальная ситуация. Классически

поле  $\phi$  стремится скатиться вниз по склону потенциала, т. е.  $\phi \rightarrow 0$ , однако квантовые флуктуации стремятся забросить его вверх. При этом увеличивается постоянная Хаббла  $H = (8\pi\lambda\phi^4/3m_{\text{Pl}}^2)^{1/2}$ , а это в свою очередь приводит к увеличению флуктуации и дальнейшему росту  $\phi$ . В результате поле  $\phi$  с большой вероятностью карабкается вверх, и лишь в относительно небольших областях  $\phi \rightarrow 0$  [58]. К сожалению, характерные масштабы при этом процессе очень велики  $L \gg H_0 \approx t_u \approx 10^{10}$  лет, и нет никакой надежды наблюдать какие-то следствия такой картины.

## § 2. ГРАВИТАЦИОННЫЕ ВОЛНЫ

Очень интересным способом, позволяющим сделать вывод о существовании во Вселенной деситтеровской фазы, явилось бы обнаружение реликтовых гравитационных волн.

Как мы уже говорили, флуктуации метрического тензора можно разбить на скалярную, векторную и тензорную части. Последняя как раз и отвечает гравитационным волнам. Существенно, что во время экспоненциального расширения амплитуда (и длина) гравитационных волн сильно возрастает, и с этим связаны надежды на их обнаружение в ближайшем будущем и на доказательство, таким образом, справедливости инфляционного сценария.

Формально поведение гравитационных волн подобно поведению скалярных возмущений, обсуждавшихся в предыдущем параграфе. Ниже мы относительно подробно рассмотрим генерацию гравитационных волн на инфляционной стадии. Аналогично можно рассмотреть и рождение других полей.

Метрический тензор, описывающий малые возмущения на однородном и изотропном фоне, можно представить в виде

$$g_{\mu\nu} = g_{\mu\nu}^{(\phi)} + h_{\mu\nu}; \quad (11.11)$$

здесь  $g_{\mu\nu}^{(\phi)}$  — метрический тензор фонового пространства,  $h_{\mu\nu}$  — малые возмущения. В соответствии с этим разбиением метрического тензора на фон плюс поправки уравнения Эйнштейна тоже можно разбить на уравнения для фонового метрического тензора и уравнение для его возмущений. Уравнения для метрического тензора фона  $g_{\mu\nu}^{(\phi)}$  совпадают с обычными уравнениями Фридмана, а уравнения для поправок имеют вид

$$\psi'_{\mu\nu;\alpha} + 2R_{\alpha\mu\nu\beta}^{(\phi)}\psi^{\alpha\beta} + R_{\mu\alpha}^{(\phi)}\psi_{\nu}^{\alpha} + R_{\nu\alpha}^{(\phi)}\psi_{\mu}^{\alpha} = 0; \quad (11.12)$$

здесь  $\psi_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}g_{\mu\nu}^{(\phi)}h_{\alpha}^{\alpha}$ , а индекс  $(\phi)$  указывает, что берутся величины, построенные только из фоновой части метрики.

Для анализа этих уравнений удобно перейти к так называемому

тому конформному времени  $\eta$ , когда метрику можно записать в виде

$$ds^2 = a^2(\eta)(d\eta^2 - dx^2 - dy^2 - dz^2)$$

с помощью очевидного преобразования координат. При этом уравнение (11.12) приобретает весьма простую форму

$$\ddot{h}_{\mu\nu} + \frac{\dot{a}}{a}\dot{h}_{\mu\nu} - \Delta h_{\mu\nu} = 0. \quad (11.13)$$

Для того чтобы проанализировать эти уравнения, подставим  $h_{\mu\nu}$  в виде плоской волны:

$$h_{\mu\nu}(\eta, \mathbf{r}) = \frac{\sigma(\eta)}{a(\eta)} G_{\mu\nu} e^{i(\mathbf{n}\mathbf{r})/a}; \quad (11.14)$$

здесь  $G_{\mu\nu}$  — постоянный тензор второго ранга, описывающий структуру волны. В случае волны, распространяющейся вдоль оси  $Ox$ , он имеет вид

$$G_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & G_1 & G_2 \\ 0 & 0 & G_2 & -G_1 \end{pmatrix},$$

$\mathbf{n}$  — волновой вектор, а  $\sigma(\eta)$  — переменная амплитуда волны. Масштабный фактор  $a(\eta)$  выделен в (11.14) специально, чтобы учесть адиабатическое затухание, связанное с расширением Вселенной.

Для амплитуды  $\sigma(\eta)$  получается уравнение, аналогичное уравнению для параметрически возбужденного осциллятора

$$\ddot{\sigma} + \left(n^2 - \frac{\ddot{a}}{a}\right)\sigma = 0. \quad (11.15)$$

Накачку энергии в этот осциллятор осуществляет работа сил гравитации. Вначале рассмотрим случай  $\ddot{a} = 0$ , т. е.  $a(\eta) = a_0\eta$ . Он соответствует так называемой радиационно-доминированной Вселенной или релятивистскому уравнению состояния  $p = \varepsilon/3$  фоновой материи. Решение (11.15) тогда тривиально и выглядит как сумма двух экспонент

$$\sigma(\eta) = A_1 e^{in\eta} + A_2 e^{-in\eta}. \quad (11.16)$$

Величины  $A_1$  и  $A_2$  отождествляются с амплитудами двух волн, из которых одна распространяется в  $(+\mathbf{n})$ -направлении, другая — в  $(-\mathbf{n})$ -направлении.

Это единственный случай, когда не происходит усиления волны. При всех остальных уравнениях состояния и другой зависимости  $a(\eta)$  усиление начальной волны будет происходить (Гришук, 1974). Отметим сразу, что для монохроматической



волны может быть как усиление, так и ослабление начальной амплитуды в зависимости от ее фазы. Для стохастического набора волн со случайными фазами всегда будет усиление (после усреднения по фазам).

Для  $a(\eta)$ , которые зависят от времени  $\eta$  степенным образом, величина  $\ddot{a}/a \sim \eta^{-2}$  с некоторым численным коэффициентом. При  $n\eta \ll 1$  в уравнении (11.15) выражение в скобках, являющееся квадратом переменной частоты, становится отрицательным, а мнимая частота соответствует уже нарастанию колебаний.

Уравнение (11.15) для произвольных  $a(\eta)$  не имеет решения в квадратурах, и анализ его в общем виде обременен большим количеством математических выкладок. Для того чтобы эти детали не заслоняли физической сути решения, рассмотрим случай таких волн, которые всегда остаются длинее горизонта, т. е. величины  $n$ . Это условие можно записать в виде (Старобинский, 1979)

$$n^2 \ll \ddot{a}/a.$$

Одно нетривиальное решение уравнения (11.15) легко угадывается:  $\sigma(\eta) = B_1 a(\eta)$ , а второе находится стандартным методом. Общее решение имеет вид

$$\sigma(\eta) = B_1 a(\eta) + B_2 a(\eta) \int \frac{d\eta}{a^2(\eta)}. \quad (11.17)$$

Первое слагаемое в (11.17) увеличивается с ростом масштабного фактора и потому соответствует так называемой растущей моде в гравитационных волнах. Второе уменьшается с увеличением  $a(\eta)$ . Это так называемая падающая мода. Начальные условия для  $\sigma(\eta)$  естественно задавать на стадии  $p = \varepsilon/3$ \*, когда мы хорошо определяем амплитуду волны и величины ее обеих мод. Они выражаются через амплитуду нулевых колебаний  $h\omega/2$ . После периода усиления, когда решение описывается выражением (11.17), вновь наступает эра радиационной доминантности  $p = \varepsilon/3$ . Решение для возмущений метрики вновь описывается выражением (11.16) с другими коэффициентами при экспонентах

$$\sigma(\eta) = C_1 e^{in\eta} + C_2 e^{-in\eta}.$$

При переходе от одного режима расширения к другому, что символически можно записать

$$(p = \varepsilon/3) \rightarrow (p = -\varepsilon) \rightarrow (p = \varepsilon/3),$$

поправки к метрике сшиваются так, чтобы ни величина  $\sigma$ , ни

\* Такая стадия могла иметь место до инфляционной стадии. Однако конкретный вид закона расширения до инфляционной стадии для нас не важен.

ее первая производная не терпели разрывов. Это позволяет выразить коэффициенты  $C_1, C_2$  через их значения на предыдущих стадиях  $B_1, B_2$  и окончательно  $A_1, A_2$ , когда задается амплитуда невозмущенной волны. Для волн, удовлетворяющих на инфляционной стадии условию  $n\eta \ll 1$ , эта зависимость имеет вид

$$C_1 = \frac{H}{in} Z^2 (A_1 + A_2) + \frac{1}{2} Z^2 (A_1 - A_2),$$

$$C_2 = -\frac{H}{in} Z^2 (A_1 + A_2) + \frac{1}{2} Z^2 (A_1 - A_2);$$

здесь  $H$  — параметр Хаббла на инфляционной стадии, связанной с плотностью энергии ложного вакуума  $\rho_v$ , как  $H^2 = 8\pi\rho_v/3m_{Pl}^2$  а  $Z$  — отношение масштабных факторов в конце и начале инфляции

$$Z = \exp\{H(t_f - t_s)\}.$$

После усреднения по фазам получаем среднеквадратичное усиление амплитуды возмущений

$$\langle (|C_1|^2) + \langle (|C_2|^2) \rangle \sim \frac{H^2}{n^2} Z^4 (\langle (|A_1|^2) + \langle (|A_2|^2) \rangle).$$

Окончательное выражение для современной плотности энергии гравитационных волн с частотой  $10^{10} \text{ Гц} > \nu > 10^{-14} \text{ Гц}$  есть

$$\varepsilon_g(\nu) = A\varepsilon_\gamma \frac{\varepsilon_v}{m_{Pl}^4} \frac{1}{\nu}, \quad (11.18)$$

где  $A$  — численный коэффициент порядка единицы, а  $\varepsilon_\gamma$  — плотность реликтового электромагнитного излучения. В более длинноволновом диапазоне

$$\varepsilon_g(\nu) = A\varepsilon_c \frac{\varepsilon_v}{m_{Pl}^4} \left(\frac{H_0}{\nu}\right)^2 \frac{1}{\nu}; \quad (11.19)$$

здесь  $\varepsilon_c = 3H_0^2/8\pi G$  — критическая плотность,  $H_0$  — значение параметра Хаббла в нашу эпоху [63].

Гравитационные волны, рожденные в деситтеровскую эпоху, обладают чрезвычайно широким спектром  $10^{10} \text{ Гц} > \nu > 10^{-17} \text{ Гц}$ . Наиболее сильно должны проявлять себя гравитационные волны с длиной порядка размера горизонта Вселенной. Фотоны реликтового излучения распространяются в переменном гравитационном поле этих волн. При движении в переменном поле фотоны изменяют свою частоту, что для наблюдателя эквивалентно угловым вариациям температуры реликтового электромагнитного излучения по небу.

Анизотропия реликтового излучения до сих пор не обнаружена. Однако на ее величину есть довольно сильные наблюда-

тельные ограничения. Из них можно найти ограничения на параметр Хаббла на инфляционной стадии. В терминах планковской массы ограничения на  $H_{\text{inf}}$  выглядят как

$$H_{\text{inf}} \leq 10^{-4} m_{\text{Pl}}. \quad (11.20)$$

### § 3. ИЗОТЕРМИЧЕСКИЕ ВОЗМУЩЕНИЯ

В § 1 мы рассмотрели, как возникают флуктуации плотности в инфляционной модели. Эти флуктуации, очевидно, являются адиабатическими. Далее, возникновение избытка барионов, описанное в § 5 гл. 8, также приводит к адиабатическим флуктуациям плотности барионов, так как в рассмотренной модели величина  $(N_B - N_{\bar{B}})/N_T$  определяется лишь фундаментальными постоянными теории и является пространственно однородной. Эти соображения привели к большой популярности адиабатических флуктуаций (тем более что предсказываемая ими картина качественно описывала наблюдения), а изотермические флуктуации рассматривались как некоторая экзотика. Однако дальнейшее развитие теории показало, что изотермические флуктуации могут быть столь же естественны, как и адиабатические. Ниже мы рассмотрим примеры возникновения изотермических флуктуаций.

Одна из таких моделей была рассмотрена А. Д. Линде и основана на флуктуациях плотности в скрытом секторе теории. В скрытом секторе находятся частицы (поля), которые взаимодействуют с остальным веществом только гравитационно. В современном теоретическом зоопарке элементарных частиц нетрудно отыскать такие объекты. Очевидно, что флуктуации плотности в скрытом секторе никак не связаны с температурой и поэтому являются изотермическими. В простейшей модели с двумя скалярными полями  $\phi_1$  и  $\phi_2$ , одно из которых ( $\phi_2$ ) принадлежит скрытому сектору и имеет большую константу самодействия  $\lambda_2 > \lambda_1$ , можно получить достаточно большие изотермические флуктуации, если плотность энергии, заключенная в поле  $\phi_2$ , падает при расширении Вселенной медленнее, чем плотность энергии остальных полей. Заметим, что в этой модели отношение  $N_B/N_T$  на ранней стадии постоянно, но имеет флуктуации плотности в скрытом секторе при постоянном  $N_T$  (или, что то же,  $T_T$ ). Заметим, впрочем, что стандартно используют термин изотермические флуктуации для переменного  $N_B/N_T$ .

В работе [41] возникновение больших флуктуаций  $N_B/N_T$  связано с малыми флуктуациями температуры. Авторы рассмотрели модель, в которой генерация барионного избытка происходит при относительно низких температурах. В этой модели выход барионов очень сильно (экспоненциально) зависит от температуры плазмы после некоторого фазового перехода и в

результате возникающие флуктуации температуры изменяют количество барионов на несколько порядков.

Рассмотренные примеры указывают на большое богатство возможностей и призывают не замыкаться только на адиабатических флуктуациях при построении теории возникновения крупномасштабной структуры Вселенной, а держать двери открытыми и для других вариантов, с помощью которых, быть может, удастся до конца решить проблему образования структуры и ее сопоставления с флуктуациями температуры реликтового излучения  $\delta T/T$ .

### § 4. СТРУНЫ И ПЕРВИЧНЫЕ ВОЗМУЩЕНИЯ

Рост квантовых флуктуаций скалярного поля на стадии инфляции не является единственным возможным источником возникновения возмущений плотности. В настоящее время широко обсуждается альтернативный механизм генерации возмущений за счет космических струн [72, 69]. Эти струны представляют собой сильно вытянутые конфигурации скалярного поля с микроскопическим поперечным размером. Как магнитные монополи и доменные стенки струны представляют собой топологически стабильные объекты, которые могут возникать во время охлаждения Вселенной при фазовых переходах, сопровождающихся развитием конденсата комплексного скалярного поля. Механизм их возникновения можно рассмотреть на следующем простом примере. Рассмотрим скалярное поле с уже знакомым нам потенциалом:

$$V(\varphi) = \lambda(\varphi^* \varphi - \eta^2)^2.$$

Как мы знаем, устойчивому состоянию такой системы отвечает постоянное во всем пространстве значение поля  $\varphi$ :

$$\langle \varphi \rangle = \eta e^{i\Phi}, \quad (11.21)$$

причем величина фазы  $\Phi$  произвольна. Она определяется случайными флуктуациями поля  $\varphi$  в процессе фазового перехода от  $\langle \varphi \rangle = 0$  к  $|\langle \varphi \rangle| = \eta$ . Если  $L$  — корреляционная длина этих флуктуаций (очевидно,  $L$  меньше или равна горизонту  $L_H$ ), то значения фазы на расстояниях, превышающих  $L$ , никак не связаны друг с другом. Теперь, если мысленно провести какой-то замкнутый контур в пространстве, то изменение фазы вдоль этого контура совсем не обязательно будет равно нулю. Однозначность  $\varphi$  требует лишь условия

$$\Delta\Phi = 2\pi n, \quad n = 0, 1, 2, \dots \quad (11.22)$$

Если, например,  $n=1$ , то это означает, что внутри контура находится точка, где  $\varphi=0$ , или, иными словами,  $\langle |\varphi| \rangle \neq \eta$  во всем пространстве, а при фазовом переходе возникает нетривиаль-

ная конфигурация поля  $\phi$  — струна. Очевидно, такие струны должны быть либо бесконечными, либо замкнутыми. Явное вычисление показывает, что поперечный размер струны  $r \sim \eta^{-1}$ , а масса единицы длины  $\mu \sim \eta^2$  (с точностью по множителю, зависящего от констант связи).

Замкнутые струны с характерной длиной меньше размера горизонта быстро схлопываются. Для космологических приложений представляют интерес струны с радиусом кривизны больше горизонта,  $L_h \approx t$ . Можно показать, что внутри объема, ограниченного  $L_h$ , должно находиться несколько таких струн, а их плотность энергии по порядку величины равна

$$\rho_s \sim \frac{\mu t}{t^3} = \frac{\mu}{t^2}. \quad (11.23)$$

Очевидно, возмущения плотности  $\delta\rho$  должны быть порядка  $\rho_s$ , и поэтому относительный контраст плотности для длин волн порядка  $L_h$  равен

$$\frac{\delta\rho}{\rho} = \frac{(\mu/t^2)}{(3m_{p1}^2/32\pi t^2)} = \frac{32\pi}{3} \frac{\mu}{m_{p1}^2}. \quad (11.24)$$

Численный коэффициент в этой формуле  $32\pi/3$  приведен для релятивистского режима расширения. Если доминирует нерелятивистское вещество, то его следует заменить на  $6\pi$ .

Таким образом, струны, как и скалярное поле приводят к безмасштабному спектру, обсуждавшемуся в гл. 10. Величина контраста плотности  $\delta\rho/\rho$  зависит от параметра  $\mu$ . Чтобы получить требуемое для образования галактик  $\delta\rho/\rho \approx 10^{-4} - 10^{-5}$ , необходимо иметь  $\mu = (10^{15} \text{ ГэВ})^2 \div (10^{16} \text{ ГэВ})^2$ , что очень близко к значениям, предсказываемым моделями большого объединения.

Более подробное изложение затронутых здесь вопросов, а также обсуждение возможностей астрономического наблюдения подобных струн проводятся в обзоре Виленкина [69].

Недостатком струнных моделей является то обстоятельство, что, в отличие от магнитных монополей, существование которых фактически однозначно предсказывается в современной теории, образование струн не является столь обязательным, а представляет собой одну, хотя и очень красивую, возможность среди множества остальных.

Читатель видит, что пока мы не имеем однозначной картины возникновения флуктуаций плотности в космологии. Однако, и в этом состоит большой прогресс, достигнутый за последние годы, физика микромира предлагает для этого ряд естест-

венных механизмов. В этом смысле ситуация радикально отличается от той, что была лишь несколько лет тому назад, когда не было известно ни одного подобного механизма и возникновение флуктуаций плотности во Вселенной выглядело совершенно загадочным. Хочется надеяться, что детальный анализ предсказаний различных моделей позволит сделать выбор между обсуждаемыми возможностями и построить, наконец, полную теорию образования крупномасштабной структуры Вселенной.

Со многими достижениями теории инфляции читатель может познакомиться по книге [75].

## ЛИТЕРАТУРА

1. Боголюбов Н. Н., Ширков Д. В. Введение в теорию квантованных полей. М.: Наука, 1976.
2. Брагинский В. Б., Панов В. И. //ЖЭТФ. 1971. Т. 61. С. 873.
3. Высоцкий М. И., Долгов А. Д., Зельдович Я. Б. //Письма в ЖЭТФ. 1977. Т. 26. С. 200.
4. Вейнберг С. Гравитация и космология. М.: Мир, 1975.
5. Вейнберг С. Первые три минуты. М.: Энергоиздат, 1981.
6. Герштейн С. С., Зельдович Я. Б. //Письма в ЖЭТФ. 1966. Т. 4. С. 174.
7. Гинзбург В. Л., Киржниц Д. А., Любошин А. А. //ЖЭТФ. 1971. Т. 60. С. 451.
8. Гришук Л. П. //ЖЭТФ. 1974. Т. 67. С. 825.
9. Гурович В. Ц., Старобинский А. А. //ЖЭТФ. 1979. Т. 77. С. 1699.
10. Долгов А. Д. //ЖЭТФ. 1980. Т. 79. С. 337.
11. Дорошкевич А. Г., Новиков И. Д. //ДАН СССР. 1964. Т. 154. С. 809.
12. Зельдович Я. Б. //Письма в ЖЭТФ. 1976. Т. 24. С. 29.
13. Зельдович Я. Б., Гришук Л. П. //УФН. 1986. Т. 149. С. 695.
14. Зельдович Я. Б., Новиков И. Д. Строение и эволюция Вселенной. М.: Наука, 1975.
15. Зельдович Я. Б., Окунь Л. Б., Пикельнер С. Б. //УФН. 1965. Т. 87. С. 115.
16. Зельдович Я. Б., Соколов Д. Д. //УФН. 1985. Т. 146. С. 493.
17. Зельдович Я. Б., Сюняев Р. А. //Астрофизика и космическая физика. М.: Наука, 1982. С. 9.
18. Кляпин А. А., Сажин М. В., Струков И. А., Скулачев Д. П. //Письма в АЖ. 1986. № 5.
19. Киржниц Д. А. //Письма в ЖЭТФ. 1972. Т. 15. С. 745.
20. Ландау Л. Д., Лифшиц Е. М. Теория поля. М.: Наука, 1973.
21. Логунов А. А. Новые представления о пространстве, времени и гравитации. М.: Изд-во МГУ, 1986.
22. Лифшиц Е. М. //ЖЭТФ. 1946. Т. 16. С. 587.
23. Любимов В. А., Новиков Е. Г., Нозик В. З., Третьяков Е. Ф., Козик В. О., Мясоедов Н. Ф. //ЖЭТФ. 1981. Т. 81. С. 1959.
24. Михеев С. П., Смирнов А. Ю. Ядерная физика. 1985. Т. 42. С. 1441.
25. Окунь Л. Б. Лептоны и кварки. М.: Наука, 1981.
26. Окунь Л. Б., афун... З. М.: Наука, 1982.
27. Окунь Л. Б. Физика элементарных частиц. М.: Наука, 1984.
28. Миблс П. Дж. Е. Структура Вселенной в больших масштабах. М.: Мир, 1983.
29. Рубаков В. А. //Письма в ЖЭТФ. 1981. Т. 33. С. 658; Nucl. Phys. 1982. V. B203. P. 311.
30. Сажин М. В. //Современные теоретические и экспериментальные проблемы теории относительности и гравитации. М.: ЛГПИ им. Ленина, 1984. С. 88.
31. Сахаров А. Д. //Письма в ЖЭТФ. 1967. Т. 5. С. 32.
32. Старобинский А. А. //Письма в ЖЭТФ. 1979. Т. 30. С. 719.
33. Фаддеев Л. Д. //УФН. 1982. Т. 136. С. 435.
34. Шварцман В. Ф. //Письма в ЖЭТФ. 1969. Т. 9. С. 315.
35. Albrecht A., Steinhardt P. //Phys. Rev. Lett. 1982. V. 48. P. 122.
36. Bardeen J., Steinhardt P., Turner M. //Phys. Rev. 1983. V. D28. P. 679.
37. Belavin A., Polyakov A., Schwarz A., Tyupkin Yu. //Phys. Lett. 1975. V. B59. P. 85.
38. Hlan C. G. Jr. //Phys. Rev. 1982. V. D25. P. 2141. V. D26. P. 2058.
39. Bogov A. D. Preprint ITEP-13. 1987.
40. Bogov A. D., Linde A. D. //Phys. Lett. 1982. V. 116B. P. 329.
41. Bogov A. D., Rubakov V. A. //Phys. Rev. Lett. 1986. V. 56. P. 988.
42. Bogov G. //Phys. Rev. 1946. V. 70. P. 572.
43. Guth A., Pi S.-Y. //Phys. Rev. Lett. 1982. V. 49. P. 1110.
44. Guth A., Pi S.-Y. //Phys. Rev. Lett. 1982. V. 49. P. 1110.
45. Hawking S. //Nature. 1974. V. 248. P. 331.
46. Hawking S. //Phys. Lett. 1982. V. 115B. P. 295.
47. Hogan C. //Ap. J. 1984. V. 284. P. L1.
48. Hubble E. P. //Proc. Nat. Acad. Sci. USA. 1929. V. 15. P. 168.
49. Hut P. //Phys. Lett. 1977. V. 69B. P. 85.
50. Jeans J. //Phil. Trans. 1902. V. 199A. P. 49.
51. Kirzhnits D. A., Linde A. D. //Phys. Lett. 1972. V. 42B. P. 471.
52. Kuzmin V. A., Rubakov V. A., Sharoshnikov M. E. //Phys. Rev. Lett. 1985. V. 155B. P. 36.
53. Lee T. D., Yang C. N. //Phys. Rev. 1955. V. 98. P. 1501.
54. Lee W., Weinberg S. //Phys. Rev. Lett. 1977. V. 39. P. 165.
55. Linde A. D. //Phys. Lett. 1983. V. 129B. P. 177.
56. Linde A. D. //Phys. Lett. 1982. V. 108B. P. 389.
57. Linde A. D. //Phys. Lett. 1982. V. 116B. P. 335.
58. Linde A. D. //Phys. Lett. 1986. V. 175B. P. 395.
59. Mandelbrot B. B. Fractals, form, chance and dimension. San-Francisco, 1977.
60. Markov M. A. //Very early universe. Proc. of the Nuffield Workshop. /Eds. Gibbons G. W., Hawking S. W., Siklos S. T. C. Cambridge, 1982. P. 353.
61. McKellar A. //Publ. Astron. Soc. Pacific. 1940. V. 52. P. 187.
62. Penzias A. A., Wilson R. W. //Astron. J. 1965. V. 142. P. 419.
63. Rubakov V. A., Sazhin M. V., Verjaskin A. V. //Phys. Lett. 1982. V. 115B. P. 189.
64. Sato K., Kobayashi H. //Progr. Theor. Phys. 1977. V. 58. P. 1775.
65. Smoot G. F., Gorenstein M. V., Muller R. A. //Phys. Rev. Lett. 1977. V. 39. P. 898.
66. Starobinsky A. A. //Phys. Lett. 1980. V. 91B. P. 99.
67. Starobinsky A. A. //Phys. Lett. 1982. V. 117B. P. 175.
68. Sunyaev R. A., Zeldovich Ya. B. //Astrophys. and Space Sci. 1969. V. 4. P. 301.
69. Vilenkin A. Phys. Reports. 1985. V. 121. P. 263.
70. Vilenkin A., Ford L. H. //Phys. Rev. 1982. V. D26. P. 1231.
71. Yukawa H. //Proc. Phys.-Math. Soc. Japan. 1935. V. 17. P. 48.
72. Zeldovich Ya. B. //Mon. Not. Roy. Astr. Soc. 1980. V. 192. P. 663.
73. Zeldovich Ya. B. //Sov. Sci. Rev. E. Astrophys. and Space Phys. 1984. V. 3. P. 1.
74. Zeldovich Ya. B., Pitajevsky L. P. //Commun. Math. Phys. 1971. V. 23. P. 185.
75. Inflationary cosmology. Singapore, 1986.

1. Бого полей.
2. Браг
3. Выс
4. ЖЭТС
5. Вейн
6. Герт С. 172
7. Гинз 1971.
8. Грин
9. Гур С. 16'
10. Дол
11. Дор С. 80
12. Зел
13. Зел
14. Зел М.: Г
15. Зел Т. 87
16. Зел
17. Зел знака
18. Клы //Пц
19. Киц
20. Ла
21. Ло виа
22. Ли
23. Лю Ко:
24. Ми С. 1
25. Ок
26. Ок
27. Ок Пи
28. Ми
29. Ру
30. Са лем.
31. Сз С.
32. С1
33. Ф.
34. Ш
35. А1

Долгов Александр Дмитриевич,  
Зельдович Яков Борисович,  
Сажин Михаил Васильевич

### КОСМОЛОГИЯ РАННЕЙ ВСЕЛЕННОЙ

Зав. редакцией С. И. Зеленский  
Редактор Г. Е. Горелик  
Художественный редактор Ю. М. Добрянская  
Технический редактор Н. И. Смирнова  
Корректоры М. И. Эльмус, Т. С. Миллякова

ИБ № 3102. Научное издание

Слано в набор 14.08.87. Подписано в печать 31.01.88.  
Лт-36537. Формат 60X90/16. Бумага тип. № 1.  
Гарнитура литературная. Печать высокая.  
Усл. печ. л. 12,5. Уч.-изд. л. 13,85. Тираж 3260 экз.  
Заказ 160. Изд. № 34. Цена 2 р. 70 к.

Ордена «Знак Почета» издательство  
Московского университета.  
108009, Москва, ул. Герцена, 5/7.  
Типография ордена «Знак Почета» изд-ва МГУ.  
119899, Москва, Ленинские горы



А. Д. Долгов  
Я. Б. Зельдович  
М. В. Сажин

# КОСМОЛОГИЯ РАННЕЙ ВСЕЛЕННОЙ

ИЗДАТЕЛЬСТВО  
МИР КОВСКОГО  
УНИВЕРСИТЕТА