

Л. Э. ГУРЕВИЧ,
А. Д. ЧЕРНИН

ВВЕДЕНИЕ В КОСМОГОНИЮ

ПРОИСХОЖДЕНИЕ
КРУПНОМАСШТАБНОЙ
СТРУКТУРЫ ВСЕЛЕННОЙ



МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1978

528

Г 95

УДК 523.12]

Г $\frac{20604-076}{053(2)-78}$ 126-78

© Главная редакция
Физико-математической литературы
Издательства «Наука», 1978

ОГЛАВЛЕНИЕ

Предисловие	5
Глава 1. Изотропная Вселенная	9
§ 1.1. Метагалактика	9
§ 1.2. Динамика космологического расширения	16
§ 1.3. Геометрические свойства изотропной Вселенной	26
§ 1.4. Космология Фридмана	40
Глава 2. Реликтовое излучение	59
§ 2.1. Излучение в горячей Вселенной	59
§ 2.2. Взаимодействие излучения и метагалактической плазмы	79
§ 2.3. Влияние горячей плазмы на спектр реликтового излучения	93
Глава 3. Малые возмущения в изотропной Вселенной	101
§ 3.1. Гравитационная неустойчивость	101
§ 3.2. Вихревые движения	119
§ 3.3. Диссипативные процессы	129
§ 3.4. Возмущения и реликтовый фон	144
Глава 4. Неоднородность и анизотропия в ранней Вселенной	153
§ 4.1. Пространственная симметрия Метагалактики и обобщения модели Фридмана	153
§ 4.2. Некоторые особенности физических процессов в анизотропном мире	169
Глава 5. Гидродинамические процессы в метагалактической среде и образование скоплений галактик	183
§ 5.1. Активная фаза космогонического процесса	183
§ 5.2. Сильные возмущения	197
§ 5.3. Ударные волны в метагалактической среде	211
§ 5.4. Генерация вихрей в ударных волнах	233
§ 5.5. Формирование скоплений галактик	244

§ 5.6. Неправильные скопления и происхождение вращения галактик	260
§ 5.7. Скопления галактик: распад или скрытые массы?	285
Глава 6. Динамическая эволюция звездных систем . . .	294
§ 6.1. От протогалактики к звездной системе	294
§ 6.2. Звездные сближения	307
§ 6.3. Испарение звезд из звездной системы	314
§ 6.4. Равновесие двойных звезд, их образование и разрушение	322
Глава 7. Проблемы космогонии звезд и планетных систем	332
§ 7.1. Образование звезд	332
§ 7.2. Происхождение Солнечной системы	349
Дополнение. Некоторые сведения из общей теории относительности	363
Литература	372

ПРЕДИСЛОВИЕ

Геологический возраст Земли приближается к пяти миллиардам лет; таков или несколько больше возраст Солнца, а самым старым звездам Галактики более 10 миллиардов лет. История Вселенной в целом насчитывает 15—18 миллиардов лет, и до образования современных планет, звезд и галактик все ее вещество представляло собой, по-видимому, почти однородную среду. Накопленные за многие десятилетия знания о строении и эволюции небесных тел, наблюдательные открытия последнего полувека и особенно открытие расширения Вселенной и существования в ней изотропного реликтового излучения позволяют составить сейчас определенное представление о свойствах космической среды в дозвездную, догалактическую эпоху, о физических процессах, приведших к формированию из однородного вещества наблюдаемых структур Вселенной. Это и составляет содержание современной космогонии.

Исторически первой проблемой космогонии, сформулированной (в 18 веке) как проблема науки, было происхождение Солнечной системы. С начала 20-го века ее предметом стали также звезды, а затем и наиболее крупные структурные единицы Вселенной — галактики и скопления галактик, природа которых была выяснена лишь в 20 — 30-годы. Большая Советская Энциклопедия определяет космогонию как науку о «происхождении и развитии космических тел и их систем: звезд и звездных скоплений, галактик, туманностей, Солнечной системы...».

И в традиционных, и в новых своих областях космогония ни в какой степени не превратилась пока в законченную и полную теорию; она находится еще в стадии предварительного собирания и сопоставления идей. Современные космогонические гипотезы содержательны и разнообразны, некоторые из них интенсивно развиваются многими

Мы хотели бы поблагодарить наших ленинградских коллег В. А. Антонова, Д. А. Варшаловича, Э. Б. Глинера, Э. М. Дробышевского, В. А. Рубана и Э. А. Троппа, прочитавших отдельные части рукописи и сделавших полезные замечания.

Пользуемся случаем выразить признательность Т. А. Агекяну, Г. С. Бисноватому-Когану, Ю. Н. Гнедину, А. З. Долгинову, А. Г. Дорошкевичу, Я. Б. Зельдовичу, А. С. Зильберглейту, А. Ф. Илларионову, И. Д. Караченцеву, С. А. Каплану, Б. В. Комбергу, Л. С. Марочнику, И. Д. Новикову, К. Ф. Огородникову, Л. М. Озерному, Ю. Н. Парийскому, Р. А. Сюняеву, И. С. Шикину, И. С. Шкловскому, Я. Э. Эйнасто и многим другим физикам и астрономам за интересные обсуждения затронутых в книге проблем.

Мы благодарны также редактору издательства И. Г. Вирко за помощь при подготовке рукописи к печати.

ИЗОТРОПНАЯ ВСЕЛЕННАЯ

§ 1.1. Метагалактика

Метагалактика — это вся наблюдаемая область Вселенной. Основными элементами ее крупномасштабной структуры являются галактики и скопления галактик. Галактики представляют собой стационарные гравитационно-связанные звездные системы. Звездная система, в которую входит наше Солнце, Галактика, содержит более ста миллиардов звезд; ее масса приблизительно $2 \cdot 10^{44} \text{ г}$, т. е. сто миллиардов солнечных масс, а полная излучаемая звездами энергия — светимость — около $3 \cdot 10^{43} \text{ эрг/сек.}$

Схематически Галактику можно подразделить на быстро вращающуюся плоскую и медленно вращающуюся сферическую подсистемы. Их массы, по-видимому, близки друг к другу. Радиус плоской подсистемы, диска, составляет приблизительно 15 килопарсек, а толщина его — около 2,5 килопарсек ($1 \text{ парсек (пс)} = 3 \cdot 10^{18} \text{ см}$). Радиус сферической подсистемы, вероятно, близок к радиусу диска. В диске, кроме звезд, имеются еще межзвездный газ и пыль, масса которых составляет несколько процентов массы звезд; в сферической подсистеме диффузного вещества практически нет. Среди звезд диска имеется заметное количество молодых и горячих звезд, тогда как в сферической подсистеме они почти полностью отсутствуют.

Угловая скорость вращения диска — разная на разных расстояниях; в области, где находится Солнце, — примерно 10 килопарсек (кпс) от центра, — ей отвечает линейная скорость вращения, равная 220—250 км/сек. Звезды диска обращаются вокруг центра по почти круговым орбитам; отклонения от кругового движения характеризуются скоростями порядка десяти километров в секунду. У звезд сферической подсистемы, находящихся поблизости

от Солнца, скорость регулярного вращения раз в пять меньше, чем у звезд диска, а их полные скорости измеряются двумя-тремя сотнями километров в секунду.

Значительная часть звезд плоской подсистемы входит в различного типа группы. Большая их часть входит в пары, а наиболее крупными образованиями являются так называемые рассеянные скопления, содержащие до тысячи звезд, связанных взаимным тяготением. Самые молодые звезды диска вместе с облаками газа и пыли располагаются широкими полосами, называемыми спиральными рукавами, которые яркими дугами выходят из внутренней, центральной области Галактики.

Распределение звезд в сферической подсистеме более или менее сферически-симметрично. Приблизительно тысячная их доля входит в большие скопления, содержащие до миллиона звезд, которые называют шаровыми скоплениями.

Концентрация звезд повышается к центральной области Галактики — ее ядру, которое проявляет себя как источник повышенного радио- и инфракрасного излучения. Из ядра происходит, по-видимому, также истечение газа.

В Метагалактике имеется большое число галактик, подобных нашей. Такие галактики называют спиральными. Хорошо известной гигантской спиралью является туманность Андромеды. Ее масса и светимость раза в два больше, чем у Галактики. Другие спиральные галактики не так массивны; чаще всего их массы составляют 10^9 — $10^{10} M_{\odot}$, скорости вращения — 100—300 км/сек, а светимости соответственно в 10—100 раз ниже.

Кроме спиральных, существуют эллиптические галактики, подобные сферическим подсистемам спиралей. В них практически нет диффузного вещества и ярких молодых звезд. Гигантские эллиптические галактики имеют массу и светимость раз в 10 больше, чем у Галактики; карликовые эллиптические галактики имеют массы 10^7 — $10^8 M_{\odot}$ и соответственно меньшие светимости. Чаще всего в эллиптических галактиках имеются плотные ядра, которые обычно больше и активнее ядер нормальных спиральных галактик.

Третий тип галактики — неправильные. Их массы и светимости не превышают десятой доли этих величин

для Галактики. Звездный состав неправильных галактик подобен составу дисков спиральных галактик. Но эти звезды, а также и значительные диффузные массы, не образуют регулярной структуры и не обладают выраженным общим вращением. Кроме ярких молодых звезд, в неправильных галактиках имеются еще и звезды более старые, подобные звездам сферических подсистем спиралей, также образующие более или менее сферический остов.

Три указанных морфологических типа галактик были впервые обнаружены и изучены Э. Хабблом и другими астрономами в двадцатые-тридцатые годы. С тех пор стали известны и иные типы галактик, не всегда укладывающиеся в первоначальную классификацию. Это относится в первую очередь к галактикам с активными ядрами и с значительным радиоизлучением. Экстремальными объектами такого рода являются открытые в шестидесятые годы квазары. В них звездная составляющая не обнаруживается; она либо вообще отсутствует, либо, что более вероятно, имеется, но дает лишь очень малый вклад в полную светимость, достигающую до 10^{46} — 10^{47} эрг/сек, что на 3—4 порядка больше светимости Галактики. Эта энергия исходит из областей с размером 10^{16} — 10^{18} см, что на 4—6 порядков меньше размеров Галактики. Радиоизлучение квазаров сравнимо с оптическим, а их инфракрасное излучение в ряде случаев в несколько раз больше того и другого. Имеется распространенная разновидность квазаров с низким радиоизлучением; такие объекты называют квазагами.

Вследствие гигантской светимости квазары видны на очень больших расстояниях. Среди них имеются самые удаленные объекты, доступные наблюдению на современных астрономических инструментах. Поэтому именно квазары очерчивают границы той области Вселенной, которую мы называем Метагалактикой. Расстояние до далеких квазаров составляет $2 \cdot 10^{26}$ см.

Большая часть галактик входит в системы, называемые группами и скоплениями галактик. Первые содержат десятки, а вторые сотни и даже тысячи членов. Среди них имеются скопления относительно правильной сферической формы. Таково, например, скопление в созвездии Волосы Вероники, или Кома, имеющее радиус ~ 4 Мпс

и содержащее около десяти тысяч галактик, среди которых преобладают эллиптические. Собственные скорости движения галактик в этом скоплении достигают тысячи километров в секунду. Кома — одно из самых больших известных скоплений.

Как выяснено в последние годы, многие богатые скопления галактик содержат значительные количества горячего ионизованного газа, проявляющего себя в рентгеновском излучении. Масса такого газа в ряде скоплений сравнима с суммарной массой галактик. Судя по динамике галактик в скоплениях, можно ожидать, что в них содержатся еще бóльшие (в 3—10 раз) количества другого вещества, которое проявляет себя только гравитационно («скрытые массы»). В противном случае скопления галактик нужно считать нестационарными, что, по-видимому, трудно согласовать с другими наблюдательными данными. Проблему «скрытых масс» мы обсудим в дальнейшем подробнее в гл. 5.

Скопления и группы галактик распределены в пространстве, по-видимому, не вполне случайным образом. Как было обнаружено впервые де Вокулером (1953, 1959), Местная Группа галактик, в которую входит наша Галактика, туманность Андромеды и еще три десятка менее крупных объектов, образует вместе с другими близкими группами галактик систему, названную местным сверхскоплением. Это — уплотненное образование размером до 50 *Mpc*, плоскость которого перпендикулярна к плоскости Галактики, с центром в крупном скоплении Дева на расстоянии 20 *Mpc* от нас.

По данным Караченцева (1966), в Метагалактике имеются и другие сверхскопления; их средний размер 20—30 *Mpc*, а число членов — крупных групп или скоплений — в среднем не больше трех-четырех.

На карте звездного неба скопления кажутся собранными иногда в протяженные цепочки с собственным размером 20—50 *Mpc*; цепочки изгибаются, соединяются и пересекаются, складываясь как бы в кружевной узор (Петерсон (1974), Чинкарини и Руд (1976), Селднер и Пиблс (1977), Ёйевер и др. (1977)). Реальность такой «кружевной сверхструктуры» еще предстоит проверить наблюдениями; возможно, зрительное впечатление обманчиво и связано просто, как отмечают Селднер и Пиблс (1977), со свой-

ством глаза намечать линейные последовательности даже в поле случайно рассеянных точек. Более надежные сведения удастся, вероятно, получить с помощью электронных вычислительных машин из статистического анализа современных каталогов галактик (работы Пиблса и его сотрудников — см. Дэвис и др. (1977) и цитируемые там статьи).

Более подробные данные о галактиках и скоплениях читатель найдет в учебнике Мартынова (1971), книге Воронцова-Вельяминова (1972) и популярной книге Агекияна (1973). Некоторые дополнительные сведения мы будем также приводить в нашей книге.

Иерархия космических структур обрывается на скоплениях и сверхскоплениях. В различных областях Метагалактики, имеющих размер 100—300 Mpc и более и содержащих много галактик и скоплений, средняя плотность видимого вещества галактик оказывается одинаковой, где бы эти области не находились. Эта плотность составляет

$$\rho \approx (1 \div 5) \cdot 10^{-31} \text{ г/см}^3. \quad (1.1.1)$$

С учетом «скрытых масс» эта величина возрастает в 3—10 раз.

Одинаковость средней плотности в различных областях пространства означает, что Метагалактика является однородной, если рассматривать ее в большом масштабе, превосходящем «размер ячейки однородности» 100—300 Mpc . Это одно из фундаментальных свойств окружающей нас Вселенной.

Другим ее фундаментальным свойством является нестационарность. Наблюдения показывают, что галактики и скопления галактик, разделенные расстояниями, превосходящими размер ячейки однородности, удаляются друг от друга.

Скорость взаимного удаления галактик пропорциональна расстоянию между ними:

$$V = HR. \quad (1.1.2)$$

Этот закон установлен Хабблом в конце двадцатых годов (рис. 1). Коэффициент пропорциональности, называемый фактором Хаббла, измерен, как и все величины, о которых

здесь идет речь, с не очень большой точностью:

$$H = 50 - 100 \text{ км/сек} \cdot \text{Мпс} \approx (5 \div 10) \cdot 10^{-11} \text{ лет}^{-1}. \quad (1.1.3)$$

Беличина H не зависит от направления, а это означает, что Метагалактика не только однородна, но и изотропна.

Данные о распределении и движении галактик были до недавнего времени единственным источником сведений о Метагалактике. В 1965 г. было открыто существование электромагнитного излучения, однородно заполняющего

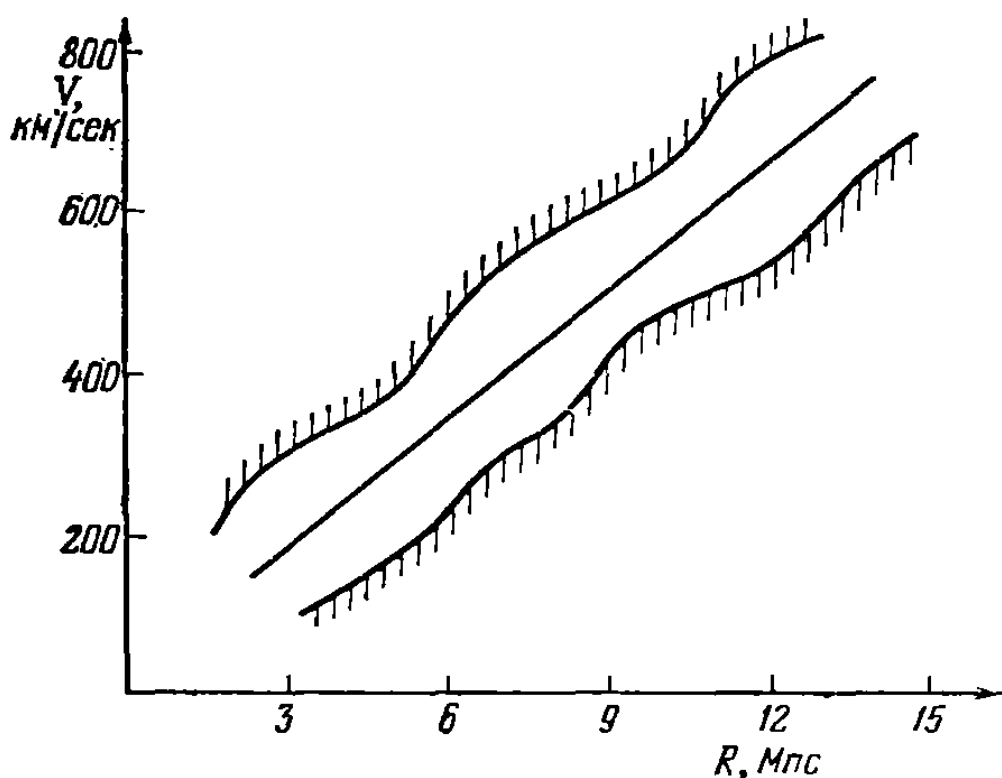


Рис. 1. Закон Хаббла. Полоса соответствует разбросу наблюдательных данных о движениях галактик; на прямой $H = 55$ км/сек·Мпс — значение, считающееся сейчас наиболее вероятным.

Метагалактику и приходящего равномерно со всех сторон (Пензиас и Вильсон (1965), Дикке и др. (1965)). Измерения его интенсивности в диапазоне длин волн от 20 до 0,3 см показали, что это излучение равномерно и имеет температуру $2,7^\circ \text{ К}$ (рис. 2). В указанной области длин волн изотропия этого излучения установлена с точностью до десятой доли процента. Это уже совсем другая точность, значительно превышающая точность, с которой установлена, например, изотропия фактора Хаббла (не лучше 20%). Плотность энергии излучения при указанной

температуре составляет (см. гл. 2)

$$\epsilon_r \approx 4 \cdot 10^{-13} \text{ эрг/см}^3. \quad (1.1.4)$$

Соответствующая массовая плотность

$$\rho_r = \epsilon_r/c^2 \approx 4 \cdot 10^{-34} \text{ г/см}^3 \quad (1.1.5)$$

значительно меньше средней плотности вещества галактик.

Однако плотность энергии ϵ_r больше суммарной плотности энергии, испущенной звездами в межзвездное и

Интенсивность,
эрг/см²·сек·стер·Гц

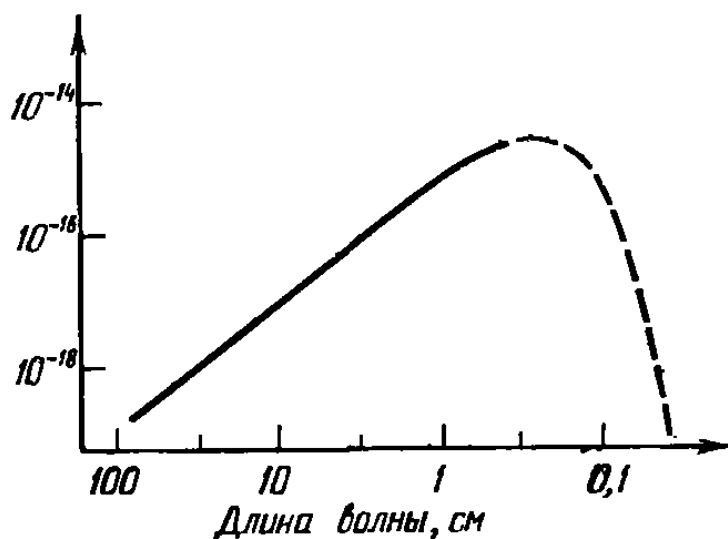


Рис. 2. Спектр реликтового излучения. Интенсивность — энергия, приходящаяся в единицу времени на единицу площади антенны из единичного телесного угла в единичном интервале частот; она пропорциональна спектральной плотности $\epsilon(\omega)$ (пунктир — область, где нет прямых измерений).

межгалактическое пространство за все время их существования (и принадлежит совсем другому участку спектра электромагнитного излучения). Это излучение — вслед за И. С. Шкловским его называют реликтовым, т. е. остаточным — является следствием горячего и плотного состояния Метагалактики в ее далеком прошлом (гл. 2).

Открытие и исследование реликтового излучения позволяет делать гораздо более уверенные утверждения об однородности и изотропии окружающей нас Вселенной. С его помощью можно получить, как мы увидим, также и важные для космогонии сведения о самых ранних стадиях эволюции крупномасштабных космических структур (гл. 3).

§ 1.2. Динамика космологического расширения

Изотропия Метагалактики в больших масштабах прежде всего означает, что относительная скорость двух тел V_{12} , находящихся друг от друга на расстоянии R_{12} (превосходящем размер ячейки однородности), должна иметь направление соединяющего их отрезка, так как нет никаких других выделенных направлений. Величина относительной скорости должна зависеть лишь от расстояния R_{12} , так как никаких других выделенных длин также нет.

Для трех тел, движущихся друг относительно друга со скоростями V_{12} , V_{23} , V_{13} , имеет место векторное равенство:

$$V_{12} + V_{23} = V_{13}. \quad (1.2.1)$$

Это обычный галилеев закон сложения скоростей. С другой стороны, векторы отрезков, соединяющих эти три тела, удовлетворяют аналогичному векторному равенству:

$$R_{12} + R_{23} = R_{13}. \quad (1.2.2)$$

Так как относительные скорости зависят лишь от расстояний, а оба эти равенства имеют место для любых трех тел, то они совместны, только если относительные скорости пропорциональны соответствующим расстояниям:

$$V = HR. \quad (1.2.3)$$

Это и есть закон Хаббла, который, таким образом, неизбежно вытекает из изотропии Метагалактики и служит ее динамическим проявлением.

Закон сложения скоростей (1.2.1) и закон сложения векторов расстояний (1.2.2) справедливы в условиях, когда применима классическая физика, т. е. скорости малы по сравнению со скоростью света c , а расстояния малы по сравнению с теми, на которых становится заметна кривизна пространства. Оба эти условия выполняются в Метагалактике для объемов с размерами

$$R < \frac{c}{H} \approx 10^{28} \text{ см}, \quad (1.2.4)$$

сравнимыми с расстояниями до наиболее удаленных астрономических объектов. В этом параграфе мы будем предполагать выполненными условия применимости классической теории. Такое приближение позволит нам рассмо-

треть динамику однородной гравитирующей среды на простом примере однородной сферы конечного размера, состоящей из вещества, давлением которого можно пренебречь. Результаты, касающиеся изменения со временем плотности вещества, окажутся при этом в согласии с точной релятивистской теорией. Совпадение такого рода не случайно, оно представляет собою следствие локальной евклидовости и локальной ньютоновости изотропного мира.

Итак, рассмотрим однородную расширяющуюся сферу, которая в некоторый момент времени t имеет плотность ρ и радиус R . Вне сферы плотность $\rho = 0$. Скорость изменения радиуса равна $\partial R/\partial t$, а скорость, с которой изменяется со временем расстояние r какой-либо частицы, находящейся внутри сферы, от ее центра, пропорциональна этому расстоянию согласно сказанному выше:

$$\frac{\partial r}{\partial t} \equiv \dot{r} = \frac{\partial R}{\partial t} \frac{1}{R} r = H(t) r. \quad (1.2.5)$$

Как известно, теорема Гаусса для ньютоновского потенциала приводит к выводу о том, что при сферически-симметричном распределении массы на каждую частицу действует лишь масса, паходящаяся внутри сферы радиуса r , и притом так, как если бы эта масса была сосредоточена в центре сферы:

$$F = - \frac{G m_1 M(r)}{r^2}, \quad (1.2.6)$$

где $G = 6,7 \cdot 10^{-8} \text{ г}^{-1} \text{ сек}^{-2} \text{ см}^3$ — постоянная тяготения, $M = \frac{4\pi}{3} \rho r^3$, m_1 — масса частицы. Ее потенциальная энергия

$$m_1 \Phi = - \frac{G m_1 M(r)}{r}, \quad (1.2.7)$$

где Φ — гравитационный потенциал, причем на бесконечном расстоянии $\Phi = 0$. Каждая частица движется, таким образом, в стационарном гравитационном поле и потому ее полная энергия, равная сумме кинетической энергии $\frac{1}{2} m_1 \dot{r}^2$ и потенциальной энергии $m_1 \Phi$, не меняется со временем, т. е. представляет собой интеграл движения. ■

Это относится и к частицам, находящимся на поверхности сферы, для которых $r = R$. Легко видеть, что такие

частицы всегда находились на поверхности и будут на ней оставаться. Это следует просто из того, что скорости частиц, находящихся под поверх остью, $\dot{R}r/R$, меньше \dot{R} согласно (1.2.5), и потому эти частицы никогда не смогут догнать частицы, находящиеся на поверхности. Для последних интеграл энергии в расчете на единичную массу, имеет вид

$$E = \frac{1}{2} \dot{R}^2 - \frac{GM}{R}, \quad (1.2.8)$$

где M — полная масса сферы. Это есть дифференциальное уравнение для функции $R(t)$. Оно имеет три различных решения в случаях $E > 0$, $E = 0$, $E < 0$.

Рассмотрим сначала случай $E = 0$, отвечающий, как говорят по аналогии с кеплеровской задачей, параболическому движению. Уравнение (1.2.8) при $E = 0$ легко интегрируется, и для случая расширения, т. е. для $\dot{R} > 0$, получаем

$$\frac{2}{3} R^{3/2} = (2GM)^{1/2} t + \text{const.} \quad (1.2.9)$$

Выберем в качестве начала отсчета времени $t = 0$ момент, когда $R = 0$. Тогда константа интегрирования исчезает:

$$R = \left(\frac{9}{2} GM \right)^{1/2} t^{2/3}. \quad (1.2.10)$$

Плотность вещества

$$\rho(t) = \frac{M}{\frac{4\pi}{3} R^3} = \frac{1}{6\pi G t^2}. \quad (1.2.11)$$

Отметим сразу же то обстоятельство, к которому нам еще придется вернуться при рассмотрении эволюции Метагалактики, что в начальный момент плотность равна бесконечности и, следовательно, состояние, выражаясь математически, сингулярно.

На основании (1.2.10) фактор Хаббла

$$H(t) = \frac{\dot{R}}{R} = \frac{2}{3} \frac{1}{t}, \quad (1.2.12)$$

совпадает, с точностью до множителя $2/3$, с обратным временем, протекшим от начала расширения, т. е. от сингулярного состояния.

Рассмотрим теперь случай $E > 0$, отвечающий, как говорят, гиперболическому движению. Заранее ясно, что в этом случае частицы, составляющие сферу, не являются гравитационно-связанными и должны неограниченно удаляться друг от друга. Прежде чем интегрировать общее уравнение (1.2.8), исследуем два его предела: при $R \rightarrow 0$ и при $R \rightarrow \infty$. В первом из них $E \ll GM/R$, и решение совпадает с (1.2.9), так что начальная стадия расширения имеет вид, близкий к параболическому. Иначе обстоит дело на поздней стадии расширения, когда $E \gg GM/R$, $R \rightarrow \infty$. Пренебрегая в этом случае потенциальной энергией в уравнении (1.2.8), видим, что оно превращается в уравнение инерциального движения, при котором скорости наружных частиц (а следовательно, и всех частиц) не меняются со временем, а расстояния между частицами возрастают пропорционально времени:

$$R = (2E)^{1/2}t. \quad (1.2.13)$$

Для точного решения уравнения (1.2.8) полезно ввести новую независимую переменную η по правилу

$$R d\eta = (2E)^{1/2} dt, \quad R' \equiv \frac{dR}{d\eta}. \quad (1.2.14)$$

Уравнение (1.2.8) принимает тогда вид:

$$(R')^2 = R^2 + \frac{GM}{E} R. \quad (1.2.15)$$

Его решение для расширения, т. е. для $\dot{R} > 0$, $R' > 0$, есть

$$R = \frac{GM}{2E} (\operatorname{ch} \eta - 1). \quad (1.2.16)$$

Далее, связь времени t с η определяется уравнением

$$t = \frac{GM}{(2E)^{3/2}} (\operatorname{sh} \eta - \eta). \quad (1.2.17)$$

Постоянные интегрирования в (1.2.16), (1.2.17) выбраны так, чтобы $t = 0$, $\eta = 0$ при $R = 0$. Формулы (1.2.16), (1.2.17) дают точное решение уравнения (1.2.8) при $E > 0$, записанное в параметрическом виде. Нетрудно проверить, что при $t \rightarrow 0$ (чему отвечает $\eta \ll 1$) и при $t \rightarrow \infty$ ($\eta \gg 1$) это решение переходит в найденные выше асимптотические решения.

Остается рассмотреть случай $E < 0$, который соответствует эллиптическому движению. В этом случае расширение не может быть, очевидно, неограниченным, т. е. радиус сферы $R(t)$ изменяется лишь в ограниченных пределах. Максимальное значение R определяется условием $\dot{R} = 0$, откуда согласно (1.2.8)

$$R_{\max} = \frac{GM}{|E|}. \quad (1.2.18)$$

На начальной фазе расширения, когда $t \rightarrow 0$, $R/R_{\max} \ll 1$,

$$|E| \ll \frac{GM}{R},$$

и уравнение (1.2.8) вновь показывает, что эта фаза имеет характер, близкий к параболическому. Параболический режим расширения является, таким образом, общей асимптотикой всех решений уравнения (1.2.8) при $t \rightarrow 0$.

Точное решение, справедливое для любых t , получим, применяя тот же метод, что и в предыдущем случае:

$$R = \frac{GM}{2|E|} (1 - \cos \eta), \quad (1.2.19)$$

$$t = \frac{GM}{(2|E|)^{3/2}} (\eta - \sin \eta). \quad (1.2.20)$$

Из этого решения видно, что расширение продолжается, пока радиус сферы не достигает найденного выше значения R_{\max} , что происходит при

$$t = t_{\max} = \frac{\pi GM}{(2|E|)^{3/2}}. \quad (1.2.21)$$

Затем расширение сменяется сжатием, продолжающимся столько же времени до наступления второй сингулярности, когда $R \rightarrow 0$.

Характер расширения в трех рассмотренных случаях показан на рис. 3.

Расширяющиеся сферы, к которым применимы полученные нами решения, можно характеризовать экстенсивными параметрами, такими, как полная масса M и радиус R , и интенсивными параметрами, такими, как плотность ρ и коэффициент Хаббла H (другими интенсивными свойствами — давлением, а следовательно, и температурой, — мы, как уже говорилось, пренебрегли). Особое

свойство решений состоит в том, что зависимость интенсивных параметров ρ и H от времени t не содержит никаких экстенсивных параметров; мы можем установить связь между ρ , H и t в каждой точке сферы, не зная ее полной массы или размера.

В случае параболического движения это свойство сразу можно видеть из выражений для плотности (1.2.11) и фактора Хаббла (1.2.12), которые действительно содержат только время. При $E \geq 0$ выражение для плотности вещества

$$\rho(t) = \frac{M}{\frac{4\pi}{3} R^3}$$

в начале расширения, когда $t \rightarrow 0$, принимает вид (1.2.11); оно не содержит размера или массы сферы, т. е. экстенсивных ее параметров, и потому справедливо для однородной сферы любых размеров и масс. Для произвольных t из (1.2.16), (1.2.17) имеем

$$G\rho t^2 = \frac{3}{4\pi} \frac{(\text{sh } \eta - \eta)^2}{(\text{ch } \eta - 1)^3}, \quad E > 0.$$

Правая часть этого выражения содержит только η . Вспомогательное время η связано с временем t соотношением

$$t = \tau_0 (\text{sh } \eta - \eta), \quad (1.2.22)$$

в котором коэффициент пропорциональности

$$\tau_0 = \frac{GM}{(2E)^{3/2}} \quad (1.2.23)$$

не зависит от экстенсивных параметров сферы и выражается через интенсивные локальные ее параметры. Действительно, как легко видеть из (1.2.8),

$$\tau_0 = \left(\frac{3}{32\pi G} \right)^{1/2} \frac{\rho}{(\rho_c - \rho)^{3/2}}, \quad E > 0, \quad (1.2.24)$$

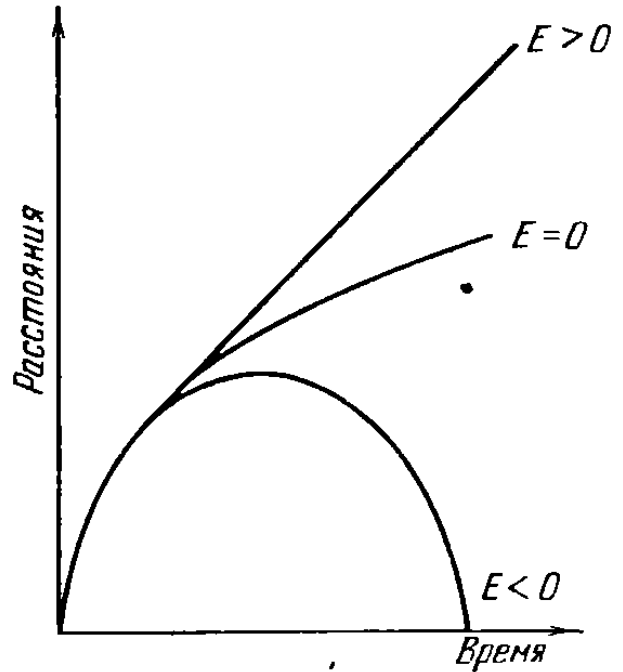


Рис. 3. Три типа изотропного расширения.

где

$$\rho_c = \frac{3}{8\pi G} H^2 \quad (1.2.25)$$

есть так называемая критическая плотность, важный смысл которой мы ниже выясним. Тот же результат получается и для $E < 0$; только вместо (1.2.24) имеем

$$\tau_0 = \left(\frac{3}{32\pi G} \right)^{1/2} \frac{\rho}{(\rho - \rho_c)^{3/2}}, \quad E < 0. \quad (1.2.26)$$

Если, как обнаруживается, связь между локальными величинами и временем не зависит от размеров рассматриваемой сферы и, следовательно, остается одной и той же для сфер любых размеров, то естественно предположить, что, увеличивая массы этих сфер, мы придем в конце концов к целой Вселенной, для которой, таким образом, эта связь сохранит тот же вид. Хотя это предположение, строго говоря, незаконно, так как, делая его, мы выходим за пределы применимости ньютоновской теории (нарушается ограничение (1.2.4)), тем не менее оно в действительности оказывается верным, как показывает релятивистская космология Фридмана (§ 1.4). Возможность такой экстраполяции означает, что динамика расширения любой сферы конечных размеров, представляющей собой часть общего, — конечного или бесконечного, — однородного распределения вещества во Вселенной, не зависит от внешнего распределения; расширение сферы происходит в точности так, как если бы вне ее не было никакого вещества.

Имея это в виду, представим экстенсивную величину — радиус сферы R — в виде произведения некоторой функции времени на функцию массы: $R = a(t)\psi(M)$. Удобно ввести a и ψ на основании (1.2.16), (1.2.19) следующим образом:

$$\left. \begin{aligned} a &= a_m (\operatorname{ch} \eta - 1), \\ \psi &= (2E/c^2)^{1/2}, \end{aligned} \right\} E > 0; \quad (1.2.27)$$

$$\left. \begin{aligned} a &= a_m (1 - \cos \eta), \\ \psi &= (2|E|/c^2)^{1/2}, \end{aligned} \right\} E < 0. \quad (1.2.28)$$

Константа размерности длины a_m связана с τ_0 :

$$a_m = c\tau_0.$$

Функция времени $a(t)$, пропорционально которой изменяются все расстояния между частицами и в том числе полный радиус сферы, называется масштабным фактором. Уравнение для $a(t)$ имеет вид

$$(\dot{a})^2 = \frac{8\pi G}{3} \rho a^2 \pm c^2 \quad (1.2.29)$$

соответственно для $E > 0 (+)$ и $E < 0 (-)$.

Перейдем теперь к сопоставлению полученных нами решений с наблюдаемыми свойствами реальной Вселенной. Эти свойства характеризуются в первую очередь значениями фактора Хаббла и плотностью вещества ρ . Вероятные значения H лежат, как указывалось в § 1.1, в интервале 50—100 км/сек·Мпс. Что же касается средней плотности, то вклад в нее видимых галактик составляет $(1 \div 5) \cdot 10^{-31}$ г/см³; если, кроме светящегося вещества, имеются еще и «скрытые массы», то это значение может повыситься примерно в 3—10 раз.

Наблюдаемые величины H и ρ , если они измерены достаточно точно, позволяют сделать заключение о том, какой из трех возможных динамических типов расширения действительно осуществляется во Вселенной. Воспользуемся снова законом сохранения (1.2.8) и перепишем его в виде

$$\rho_c - \rho = \frac{3}{4\pi G} \frac{E}{R^2}, \quad (1.2.30)$$

где ρ_c — критическая плотность, даваемая формулой (1.2.25). Знак E и, следовательно, тип расширения определяется, как мы видим, соотношением между ρ и ρ_c : $E > 0$ при $\rho < \rho_c$, $E = 0$ при $\rho = \rho_c$ и $E < 0$ при $\rho > \rho_c$.

Значение критической плотности

$$\rho_c = \frac{3}{4\pi G} H^2 \approx 5 \cdot 10^{-30} \left(\frac{H}{50 \text{ км/сек} \cdot \text{Мпс}} \right)^2 \text{ г/см}^3. \quad (1.2.31)$$

Эта величина превышает плотность вещества (вероятно, даже с учетом возможных «скрытых масс»). Можно поэтому полагать, что наблюдаемое космологическое расширение является, скорее всего, гиперболическим и потому будет продолжаться неограниченно. Правда, расхождение между ρ_c и ρ при наличии «скрытых масс» невелико; кроме того, мы не можем быть уверены в том, что нам известны все формы вещества, заполняющего

Вселенную (см. ниже), и поэтому сделанный вывод нельзя считать окончательным.

Важной характеристикой Вселенной является ее возраст, т. е. время, прошедшее от начала расширения. Пользуясь формулами (1.2.19), (1.2.20) для гиперболического расширения, найдем, что

$$t = \frac{1}{H} \frac{\text{sh } \eta (\text{sh } \eta - \eta)}{(\text{ch } \eta - 1)^2}. \quad (1.2.32)$$

Дробь, на которую умножается здесь $1/H$, изменяется при изменении η от 0 до ∞ в весьма ограниченных пределах: от $2/3$ до 1, так что знание величины H позволяет получить непосредственную оценку возраста мира:

$$t_0 = \left(\frac{2}{3} \div 1 \right) \frac{1}{H} \approx (13 \div 20) \left(\frac{H}{50 \text{ км/сек} \cdot \text{Мпс}} \right)^{-1} \text{ млрд лет.} \quad (1.2.33)$$

Малые значения η соответствуют, как мы знаем, режиму расширения, близкому к параболическому, который служит при этом общей асимптотикой расширения; поэтому оценка (1.2.33) справедлива для всех трех динамических типов космологического расширения.

Величина (1.2.33) хорошо согласуется с данными о возрастах звезд и галактик ($\lesssim 12$ млрд. лет), полученными на основании совершенно других соображений. Она согласуется также с геологическим возрастом Земли (4—5 млрд. лет), который служит нижней границей возраста Солнца. Имеется согласие и с оценкой возраста радиоактивных элементов, которая делается на основании сравнения распространенностей материнского элемента и дочерних продуктов (при известном периоде полураспада); самая высокая оценка получена для распада рения в осмий: 18 млрд. лет (Броунс и Берман (1976)).

Как видно из всего сказанного, динамика расширения существенно зависит от отношения

$$\Omega = \rho/\rho_c, \quad (1.2.34)$$

где значения ρ , ρ_c берутся в современную эпоху. Космологический параметр Ω определяет, в частности, момент перехода к инерциальной фазе гиперболического расширения. При инерциальном расширении

$$\rho = \rho(t_0) \left(\frac{t_0}{t} \right)^3, \quad H = \frac{1}{t}. \quad (1.2.35)$$

Из последних формул видно, что ρ заметно меньше ρ_c при

$$t > t_\Omega = t_0\Omega. \quad (1.2.36)$$

Момент t_Ω отделяет начальную фазу расширения, для которой приближенно справедливы формулы параболического расширения (1.2.10)—(1.2.12), от поздней фазы инерциального расширения, которая, по-видимому, и реализуется в настоящую эпоху.

Величине Ω можно придать наглядный динамический смысл, связав ее с так называемым параметром замедления, определяемым формулой:

$$q = - \frac{\ddot{a}a}{(\dot{a})^2}. \quad (1.2.37)$$

Используя уравнение движения (1.2.6) и закон сохранения (1.2.8) найдем

$$q = 1/2\Omega. \quad (1.2.38)$$

Это соотношение справедливо для всех трех типов расширения.

Мы уже упоминали о том, что в принципе не исключена возможность существования во Вселенной трудно обнаружимых форм вещества. Верхний предел суммарной плотности, включая и эти формы вещества, можно найти, воспользовавшись значением геологического возраста Земли как надежной нижней границей возраста Вселенной. Действительно, убывание плотности со временем означает, что нижняя граница времени, протекшего от начала расширения, соответствует верхней границе плотности. Мы заранее должны допустить значения ρ , превосходящие ρ_c , и потому нужно взять решение для эллиптического расширения. В этом решении по мере приближения к состоянию максимального расширения превышение плотности над критической становится сколь угодно большим. Для грубой оценки найдем (по формулам (1.2.19)—(1.2.21)) связь между ρ и t вблизи состояния максимального расширения:

$$t \approx \pi\tau_0 \approx \left(\frac{3}{32\pi G\rho} \right)^{1/2}. \quad (1.2.39)$$

Если $t > 5$ млрд. лет $\simeq 1,5 \cdot 10^{17}$ сек, то из последней формулы следует оценка максимально допустимой плотности

Вселенной, совместимой с ограничением на ее возраст:

$$\rho < 2 \cdot 10^{-28} \text{ г/см}^3. \quad (1.2.40)$$

Этот результат показывает, что теория космологического расширения дает возможность довольно резко ограничить плотность еще не обнаруженных форм вещества, если они вообще существуют (Зельдович и Смородинский (1961)). Если в качестве нижнего предела возраста Вселенной принять возраст, определяемый по распространенности радиоактивных элементов, эта оценка станет еще более жесткой.

§ 1.3. Геометрические свойства изотропной Вселенной

Общая теория относительности и космология. Строение и эволюция Метагалактики определяются в первую очередь тяготением. Теория тяготения Ньютона достаточно точна для описания движения планет, звезд и даже целых галактик, когда их относительные скорости много меньше скорости света c , а разности гравитационных потенциалов много меньше (по модулю) c^2 . Она становится неприменимой, когда эти сильные неравенства нарушаются; это как раз и имеет место для самых далеких квазаров, находящихся на границе Метагалактики. Здесь требуется релятивистская теория тяготения, разработанная Эйнштейном.

Основные идеи этой теории, называемой также общей теорией относительности, настолько красивы и прозрачны, что мы сможем, ни в какой степени не претендуя на полноту изложения, дать здесь о них представление, достаточное для понимания последующего материала.

Общая теория относительности изменила теорию Ньютона прежде всего в двух отношениях. Во-первых, она внесла в теорию тяготения один из самых общих законов природы, сформулированный специальной теорией относительности и заключающийся в том, что никакие взаимодействия не могут распространяться со скоростью, превышающей скорость света. Во-вторых, фундаментальное значение было придано установленному Галилеем закону, согласно которому все тела, независимо от их массы, материала, из которого они сделаны, и т. п., па-

дают в данном поле тяготения с одинаковым ускорением. Эйнштейн интерпретировал эту универсальность ускорения как проявление изменений общих свойств пространства и времени, вызванных гравитирующими массами. Движение в поле тяготения в этом смысле представляет собой свободное движение, заменяющее равномерное прямолинейное движение в отсутствие гравитационного поля. Такое свободное движение играет в общей теории относительности ту же роль, какую в механике Ньютона и в специальной теории относительности играло инерциальное движение.

В классической теории пространство считалось абсолютным, т. е. одинаковым всегда и везде и никак не зависящим от физических явлений, которые в нем происходят. Точно так же и время считалось абсолютным, т. е. неизменным по темпу, так сказать, протекания, одинаково для всех точек пространства, вне зависимости от каких-либо физических явлений. Два сформулированных выше положения приводят к представлению о том, что пространство и время не независимы друг от друга и от физических явлений. Они образуют единое, как принято говорить, четырехмерное многообразие, для описания свойств которого, меняющихся от одной его «точки» к другой, требуется геометрическая теория, отличающаяся от геометрии Евклида. В математике такая теория искривленных неевклидовых пространств была разработана Гауссом, Лобачевским, Больяйи и в наиболее общей форме Риманом.

Проще всего представить себе неевклидову геометрию в двумерном случае, т. е. на поверхности. Например, на сфере кратчайшими линиями между двумя точками, или «прямейшими», заменяющими здесь прямые евклидовой геометрии (их называют геодезическими), являются дуги больших кругов. Любые два больших круга пересекаются в двух противоположных точках, так что непересекающихся «прямейших» вообще нет. Сумма углов треугольника, построенного из отрезков «прямейших», при этом больше π .

Положение любой точки на сфере задается полярным углом ϑ и азимутальным углом φ . Квадрат отрезка dl «прямейшей», или геодезической, соединяющего две очень близкие точки на сфере, координаты которых отли-

чаются на очень малые (математически говоря, бесконечно малые) углы $d\vartheta$ и $d\varphi$, представляется в виде суммы (см. ниже)

$$dl^2 = a^2 \sin^2 \vartheta d\varphi^2 + a^2 d\vartheta^2, \quad (1.3.1)$$

где a — радиус сферы. Если сравнить это выражение с аналогичным выражением на плоскости в прямоугольных, или декартовых, координатах,

$$dl^2 = dx^2 + dy^2, \quad (1.3.2)$$

то можно сказать, что выражение (1.3.1) представляет собой теорему Пифагора для сферы.

Сфера — это поверхность, геометрические свойства которой во всех точках одинаковы. В общем же случае геометрические свойства пространства могут, очевидно, меняться от точки к точке. Его свойства в каждой точке определяются не одним только радиусом кривизны a , но целым набором величин. Происхождение последних можно проследить на простых примерах, относящихся к плоскости. Вместо формулы (1.3.2) можно выразить dl^2 в косоугольных координатах с углом α :

$$dl^2 = dx^2 + dy^2 + 2 \cos \alpha dx dy, \quad (1.3.3)$$

или в цилиндрических координатах r и φ :

$$dl^2 = dr^2 + r^2 d\varphi^2. \quad (1.3.4)$$

Сама величина dl остается, очевидно, одной и той же, в какой бы системе координат она ни представлялась. При любом, произвольном выборе координатной системы x_1, x_2 на евклидовой плоскости это выражение имеет вид

$$dl^2 = \gamma_{11} dx_1^2 + \gamma_{22} dx_2^2 + 2\gamma_{12} dx_1 dx_2, \quad (1.3.5)$$

где коэффициенты γ_{ik} могут, вообще говоря, зависеть от обеих координат x_1, x_2 . Сокращенно эту формулу пишут так:

$$dl^2 = \gamma_{ik} dx_i dx_k, \quad i, k = 1, 2.$$

По повторяющимся индексам подразумевается суммирование, при этом, разумеется, $\gamma_{ik} = \gamma_{ki}$.

Пусть теперь нам заданы коэффициенты γ_{ik} как какие-то функции координат x_1, x_2 . Поставим вопрос: всегда ли возможно преобразованием координат привести их на всей поверхности к виду $\gamma_{11} = \gamma_{22} = 1, \gamma_{12} = \gamma_{21} = 0,$

соответствующему евклидовой плоскости? Ответ на этот вопрос отрицательный. Действительно, в геометрии на сфере dl^2 дается формулой (1.3.1) и никаким преобразованием координат ее нельзя свести к (1.3.2), так как (1.3.1) есть теорема Пифагора для треугольника, сумма углов которого больше π .

Таким образом, набором величин γ_{ik} , зависящих от координат, можно в общем случае характеризовать геометрические свойства поверхности, отличной от евклидовой. Совокупность величин γ_{ik} называется метрическим тензором, а неевклидова геометрия, в которой свойства пространства меняются от точки к точке и характеризуются этим тензором, называется римановой.

В случае трехмерного риманова пространства общее выражение для квадрата расстояния между двумя бесконечно близкими точками имеет вид

$$dl^2 = \gamma_{ik} dx_i dx_k, \quad (1.3.6)$$

где значки i, k пробегают теперь значения 1, 2, 3, а по повторяющимся значкам по-прежнему подразумевается суммирование. Величины x_1, x_2, x_3 не являются обязательно декартовыми координатами, а могут быть набором любых величин, однозначно определяющих положение точки в пространстве.

Расстояние dl можно понимать, в частности, как путь, проходимый некоторой частицей за малый (бесконечно малый) промежуток времени dt . Положение точки в каждый момент времени задается значениями трех координат и соответствующим значением времени. Эти четыре величины определяют, как говорят, событие, или мировую точку. Совокупность мировых точек, по которым происходит движение частицы или распространение светового луча, называется мировой линией. Эти понятия вводятся специальной теорией относительности.

Пусть два события, рассматриваемые в рамках специальной теории относительности, связаны друг с другом распространением света, т. е. лежат на мировой линии светового луча. Тогда их расстояние в пространстве dl связано с промежутком времени между ними dt соотношением

$$dl = c dt. \quad (1.3.7)$$

Эти величины измерены в некоторой инерциальной системе отсчета. Если тот же процесс рассмотреть в другой системе отсчета, движущейся инерциально относительно первой, то расстояние dl' и промежуток времени dt' должны быть связаны тем же соотношением

$$dl' = c dt',$$

так как скорость света одинакова во всех системах отсчета. Таким образом, величина

$$ds^2 = c^2 dt^2 - dl^2, \quad (1.3.8)$$

которая называется квадратом интервала, оказывается равной нулю в обеих системах отсчета. Формально выражение (1.3.8) похоже на выражение для расстояния в некотором четырехмерном пространстве, с той только разницей, что квадраты разностей координат входят в него с разными знаками. Однако, как и в обычной геометрии, значение величины ds одинаково во всех системах координат; она является инвариантной. Интервал между двумя событиями, связанными движением частицы, имеющей скорость $v < c$, не равен нулю; он положителен, так как в этом случае

$$dl = v dt < c dt.$$

Интервал между двумя событиями может быть определен и в том случае, когда эти события не связаны распространением света или движением частиц. Два события, для которых $dl > c dt$, не могут вообще быть причинно связаны; квадрат интервала в этом случае отрицателен. Такой интервал называется пространственноподобным, в отличие от времени подобного, квадрат которого положителен.

Представление о едином пространстве-времени специальной теории относительности вместе с идеей неевклидовой римановой геометрии преобразуются в общей теории относительности в представление о едином четырехмерном римановом многообразии — искривленном пространстве-времени. Это выражается математически в новом обобщении формулы для интервала:

$$ds^2 = g_{ik} dx_i dx_k, \quad i, k = 0, 1, 2, 3. \quad (1.39)$$

Здесь $x_0 = ct$, а величины g_{ik} , являющиеся в общем случае функциями всех четырех координат x_i , играют ту же

роль, что и величины γ_{ik} в наших примерах римановых пространств, — они составляют четырехмерный метрический тензор пространства-времени, причем $g_{ik} = g_{ki}$. Составленный таким образом интервал является по-прежнему инвариантной величиной — он остается неизменным при любом выборе системы четырех координат x_i . Если не существует такого преобразования координат, которое свело бы совокупность величин g_{ik} к их значениям для интервала специальной теории относительности, то пространство-время существенно искривлено. Как мы уже говорили, такое изменение свойств пространства-времени связывается в общей теории относительности с присутствием в пространстве гравитирующих масс.

Совокупность величин g_{ik} (независимы из них 10, так как $g_{ik} = g_{ki}$) играет в этой теории ту же роль, что и гравитационный потенциал в теории тяготения Ньютона. Гравитационный потенциал φ связывается с плотностью массы ρ уравнением Пуассона

$$\Delta\varphi = 4\pi G\rho, \quad (1.3.10)$$

где G — гравитационная постоянная, Δ — оператор Лапласа. Величины g_{ik} определяются не только плотностью ρ или плотностью энергии $\epsilon = \rho c^2$, но также и плотностью импульса, плотностью потока энергии и плотностью потока импульса. Эта совокупность величин, характеризующих распределение и движение материи, образует тензор энергии-импульса. Подобно тому как в уравнение (1.3.10) входит не сам потенциал, а его производные, в соответствующие уравнения общей теории относительности, называемые уравнениями Эйнштейна, входит нелинейная комбинация g_{ik} и их производных первого и второго порядка.

Идеи общей теории относительности составляют основу современной космологии; мы будем следовать им в дальнейшем, а пока ограничимся этим кратким их изложением. Некоторые дополнительные сведения об общей теории относительности вынесены в Дополнение. Там же мы рассматриваем важные частные решения уравнений Эйнштейна. Систематическое изложение общей теории относительности дается в книге Эйнштейна (1955), в известных учебниках Эддингтона (1934), Толмена (1974)

и Ландау и Лифшица (1973), релятивистские эффекты в астрофизике исследуются в книгах Зельдовича и Новикова (1967, 1971, 1975).

Геометрия изотропного пространства. При анализе геометрических свойств Вселенной, так же как и при анализе ее динамики, мы исходим из того фундаментального факта, что в Метагалактике распределение вещества однородно, а его движение изотропно. Эти свойства симметрии должны, согласно принципам общей теории относительности, найти отражение в геометрических свойствах пространства, которое само должно быть однородно (в смысле равноправия всех его точек) и изотропно (в смысле равноправия всех направлений в каждой точке). В ньютоновской механике, которой мы пользовались в предыдущем параграфе, эти геометрические свойства с самого начала присущи абсолютному евклидову пространству классической физики. В общей теории относительности требования симметрии уже не выполняются автоматически, так как пространство релятивистской физики неевклидово и его геометрия зависит от распределения и движения материи. Евклидова геометрия является лишь одним частным случаем изотропного пространства.

Удобно начать с аналогии с двумерными изотропными пространствами (поверхностями). Плоскость является евклидовым двумерным изотропным пространством. Другим двумерным пространством, обладающим той же симметрией, является поверхность сферы. Это пространство, в отличие от евклидова, искривлено. Геометрия на этой поверхности описывается формулой для расстояния между бесконечно близкими точками. Чтобы найти эту формулу, примем во внимание, что такая поверхность находится в трехмерном евклидовом пространстве, в котором ее уравнение имеет вид:

$$x_1^2 + x_2^2 + x_3^2 = a^2. \quad (1.3.11)$$

Здесь x_1, x_2, x_3 — декартовы координаты с началом в центре сферы, a — ее радиус. Решение этого уравнения можно представить в виде

$$\left. \begin{aligned} x_1 &= a \sin \vartheta \cos \varphi, \\ x_2 &= a \sin \vartheta \sin \varphi, \\ x_3 &= a \cos \vartheta, \end{aligned} \right\} \quad (1.3.12)$$

где $0 \leq \vartheta \leq \pi$, $0 \leq \varphi \leq 2\pi$, ϑ имеет смысл полярного угла, т. е. угла между радиусом-вектором точки на сфере и полярной осью x_3 , а φ — азимутального угла, т. е. угла между проекцией радиуса-вектора на экваториальную плоскость и некоторым направлением, выбранным в качестве начального.

Для бесконечно близких точек с координатами, отличающимися на dx_1 , dx_2 , dx_3 , квадрат расстояния — по теореме Пифагора — есть

$$dl^2 = dx_1^2 + dx_2^2 + dx_3^2. \quad (1.3.13)$$

Для точек, лежащих на поверхности сферы, координаты x_1 , x_2 , x_3 связаны уравнением (1.3.11), а в его решении для бесконечно близких точек углы ϑ , φ различаются на $d\vartheta$, $d\varphi$. Дифференцируя (1.3.12) и подставляя в (1.3.13) находим теорему Пифагора для поверхности сферы (1.3.1).

Обратим внимание на одно свойство этой геометрии, которое окажется полезным в дальнейшем. Пусть некоторая точка перемещается по сфере вдоль меридиана, удаляясь от другой точки, которая остается неподвижной и с которой мы свяжем полюс, т. е. начало отсчета ϑ . При этом перемещении изменяется только координата ϑ , каждому значению которой соответствует расстояние $a\vartheta$ от начальной точки. Координата ϑ возрастает от нуля до π при движении к противоположному полюсу. Когда точка перейдет через второй полюс и начнет снова приближаться к первому, проекция ее радиуса-вектора на экваториальную плоскость изменит знак на обратный, что означает изменение азимутального угла φ скачком на величину π . Угол же ϑ при дальнейшем движении точки уменьшается от π до нуля. Пройдя по большому кругу путь $2\pi a$, точка возвращается в первый полюс. Дуги больших кругов являются «прямыми», т. е. кратчайшими расстояниями на сфере, и расстояние между точками на них не превосходит πa . Эти обстоятельства, тривиальные для сферы, оказываются совсем не тривиальными для трехмерного пространства.

По аналогии с предыдущими соображениями представим себе гиперсферу (т. е. трехмерный аналог поверхности сферы), «вложенную» в четырехмерное евклидово

пространство. Это воображаемое пространство, разумеется, не имеет ничего общего с реальным четырехмерным пространством-временем. Оно нужно лишь для того, чтобы простым способом выяснить геометрические свойства неевклидова трехмерного пространства гиперсферы.

Уравнение гиперсферы

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = a^2 \quad (1.3.14)$$

получается простым обобщением уравнения (1.3.11) и допускает решение, которое, подобно (1.3.12), можно представить в виде

$$\left. \begin{aligned} x_1 &= a \sin \chi \sin \vartheta \cos \varphi, \\ x_2 &= a \sin \chi \sin \vartheta \sin \varphi, \\ x_3 &= a \sin \chi \cos \vartheta, \\ x_4 &= a \cos \chi. \end{aligned} \right\} \quad (1.3.15)$$

Углы ϑ и φ , аналогичные полярному и азимутальному углам, имеют те же пределы изменения. Что же касается нового угла χ , то его пределы изменения,

$$0 \leq \chi \leq \pi, \quad (1.3.16)$$

обеспечивают полное и однократное покрытие поверхности гиперсферы, так что каждой точке на ней соответствует одно и только одно значение трех координат: χ , ϑ , φ .

Расстояние между бесконечно близкими точками в четырехмерном евклидовом пространстве, координаты которых отличаются на dx_1 , dx_2 , dx_3 , dx_4 , определяются выражением

$$dl^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2, \quad (1.3.17)$$

представляющим собою теорему Пифагора для четырехмерного пространства. Декартовы координаты этих точек гиперсферы связаны соотношением (1.3.14), а координаты χ , ϑ , φ отличаются на $d\chi$, $d\vartheta$, $d\varphi$. Дифференцируя (1.3.15) и подставляя в (1.3.17), находим

$$dl^2 = a^2 [d\chi^2 + \sin^2 \chi (d\vartheta^2 + \sin^2 \vartheta d\varphi^2)]. \quad (1.3.18)$$

Это — теорема Пифагора для искривленного пространства гиперсферы. Сравним (1.3.18) с выражением для квадрата расстояния в обычном трехмерном евклидовом

пространстве, записанным в сферических координатах r, ϑ, φ :

$$dl^2 = dr^2 + r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2), \quad 0 \leq r < \infty. \quad (1.3.19)$$

Если положить $r/a = \chi$, где χ есть безразмерная радиальная координата, то последнее выражение примет вид

$$dl^2 = a^2 [d\chi^2 + \chi^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2)], \quad 0 \leq \chi < \infty. \quad (1.3.20)$$

Искривленное пространство с выражением (1.3.18) для квадрата расстояния отличается от евклидова прежде всего тем, что в последнем случае радиальная координата (r или χ) может изменяться неограниченно, тогда как в искривленном пространстве (1.3.18) все три координаты изменяются в конечных пределах. Поэтому в евклидовом пространстве расстояние между двумя точками может быть сколь угодно велико, а в искривленном пространстве гиперсферы оно всегда ограничено. Действительно, для точек, углы ϑ и φ которых совпадают, имеем

$$dl = a d\chi. \quad (1.3.21)$$

При конечных расстояниях между точками

$$l = a (\chi_2 - \chi_1). \quad (1.3.22)$$

Такое соотношение получается как для (1.3.18), так и для (1.3.19), но в искривленном пространстве

$$\chi_2 - \chi_1 \leq \pi, \quad l \leq a\pi. \quad (1.3.23)$$

Другая особенность рассматриваемого неевклидова пространства связана с тем, что в выражении (1.3.18) перед угловой частью $(d\vartheta^2 + \sin^2 \vartheta d\varphi^2)$ стоит множитель $\sin^2 \chi$, а не χ^2 , как в (1.3.20). Рассмотрим точки, радиальные координаты которых χ совпадают, т. е. точки на двумерной поверхности, которая в евклидовом случае есть просто поверхность сферы. В неевклидовом пространстве такая «сфера» обладает некоторыми особыми свойствами. Радиус поверхности, т. е. расстояние ее точек от центра, в обоих случаях есть $a\chi$ (если принять, что в центре $\chi = 0$). Площадь бесконечно малого квадрата на ней есть $a^2 \chi^2 \sin \vartheta d\vartheta d\varphi$ в трехмерном евклидовом пространстве и $a^2 \sin^2 \chi \sin \vartheta d\vartheta d\varphi$ в нашем случае, когда

трехмерное пространство неевклидово. Поэтому полные площади поверхностей равны соответственно $4\pi a^2 \chi^2$ и $4\pi a^2 \sin \chi$. Во втором случае отношение площади «сферы» к квадрату радиуса меньше, чем 4π . Это отношение зависит от величины радиуса: при малых χ оно близко к 4π , а с ростом χ уменьшается, достигая нуля при $\chi = \pi$. Удивительным образом площадь «сферы» сначала возрастает с увеличением радиуса, а затем уменьшается, и при максимальном радиусе «сфера» имеет площадь, равную нулю. Однако этот вывод покажется уже не столь неожиданным, если вспомнить свойство длины параллелей на обычной сфере; эта длина сначала возрастает, а затем убывает по мере удаления от начальной точки, пока не обратится в нуль для параллели, максимально удаленной от начальной точки.

На «сфере» в нашем неевклидовом пространстве длина параллели, т. е. линии, вдоль которой $\chi = \text{const}$, $\vartheta = \text{const}$, есть $2\pi a \sin \chi \sin \vartheta$; она ведет себя немонотонно по ϑ и по χ : при увеличении ϑ и χ от нуля она сначала возрастает, а потом уменьшается и при достижении одной из этих координат максимального значения обращается в нуль. Отношение длины параллели к ее радиусу меньше 2π :

$$\frac{2\pi a \sin \chi \sin \vartheta}{a\chi \sin \vartheta} < 2\pi.$$

Мы рассмотрели свойства поверхностей и линий в нашем неевклидовом пространстве — гиперсфере. Рассмотрим теперь свойства трехмерных объемов в ней. Объем сферического слоя с радиусом $a\chi$ и толщиной $a d\chi$ есть $4\pi a^2 \sin^2 \chi a d\chi$. Суммируя объемы этих слоев с радиусами от нуля до $a\pi$, получим полный объем гиперсферы:

$$V = 4\pi \int_0^{\pi} \sin^2 \chi d\chi = 2\pi^2 a^3. \quad (1.3.24)$$

Объем гиперсферы оказался конечным, так же как конечна площадь поверхности обычной сферы в трехмерном евклидовом пространстве.

Итак, в отличие от евклидова трехмерного пространства, изученное нами неевклидово трехмерное пространство, гиперсфера, имеет конечный объем; площади и

длины в этом пространстве также конечны. Но несмотря на конечность, это пространство не имеет границ: все его точки равноправны и ни одна из них не может считаться граничной, совершенно так же, как нет границ у поверхности обычной сферы. Такое конечное, но не имеющее границ пространство называют замкнутым, или закрытым.

Это пространство, как мы видим, отличается от евклидова трехмерного пространства качественным образом. Однако в малых областях это различие оказывается малым. Действительно, так как все точки нашего неевклидова пространства равноправны, то любую из них можно принять за начало координат. Вблизи этого начала $\chi \ll 1$ и потому

$$\sin \chi \approx \chi.$$

Расстояние между двумя близкими точками вблизи начала координат определяется выражением

$$dl^2 = a^2 [d\chi^2 + \chi^2 (\sin^2 \vartheta d\varphi^2 + d\vartheta^2)]. \quad (1.3.25)$$

Достаточно произвести замену $a\chi = r$ (см. выше), и мы получим для dl^2 ту самую формулу, которая имеет место в евклидовом пространстве — (1.3.19). Таким образом, в бесконечно малой области, вблизи любой точки нашего пространства оно не отличается от евклидова. Эта, как говорят, локальная евклидовость пространства имеет прямую аналогию в случае обычной сферы: малый отрезок дуги большого круга мало отличается от касательной к ней, и чем меньше этот отрезок, тем меньше отличие. Точно так же небольшой участок поверхности сферы мало отклоняется от касательной к ней плоскости, и геометрия в бесконечно малой области сферы не отличается от евклидовой геометрии на плоскости.

Но если пространство локально везде евклидово, это еще не характеризует, как мы видим, его полностью. Кроме этого, пространство характеризуется кривизной. Кривизной поверхности сферы называют величину

$$K = \frac{1}{a^2}. \quad (1.3.26)$$

Для любой другой искривленной поверхности определение кривизны требует проведения через какую-либо ее

точку двух взаимно перпендикулярных разрезов, или сечений, которые образуют на поверхности две взаимно перпендикулярные линии. Вблизи точки их пересечения маленькие отрезки этих линий мало отклоняются от прямых, но теперь нас будет интересовать как раз это отклонение. Его наличие приводит к тому, что в следующем приближении эти маленькие отрезки почти совпадают с отрезками некоторых окружностей. Если обозначить их радиусы через a_1 и a_2 , то, как показывается в геометрии поверхностей (впервые это доказал Гаусс), произведение $a_1 a_2$ не зависит от направлений, в которых проведены разрезы. Поэтому такое произведение характеризует свойства самой поверхности, и величину

$$K = \frac{1}{a_1 a_2} \quad (1.3.27)$$

называют гауссовой кривизной. При этом, если центры обеих окружностей находятся по одну сторону от поверхности, то обоим радиусам приписывают одинаковые знаки, и кривизна оказывается положительной. Положительной кривизной, одинаковой во всех ее точках, обладает, очевидно, поверхность сферы. Если же центры окружностей находятся по разные стороны от поверхности, то знаки радиусов a_1 и a_2 противоположны, и кривизна поверхности отрицательна. Примером поверхности отрицательной кривизны является однополостный гиперболоид, описываемый уравнением

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \quad (1.3.28)$$

Гиперсфере — трехмерному аналогу сферы — также можно приписать кривизну, и притом положительную. Можно показать, что и в этом случае кривизна пропорциональна a^{-2} (Ландау и Лифшиц (1973)). Величину a называют при этом радиусом кривизны.

Аналогично двумерным пространствам отрицательной кривизны существует и трехмерное пространство отрицательной кривизны. Если представить его «вложенным» в четырехмерное евклидово пространство, то соответствующее уравнение для такого пространства можно получить из (1.3.14) заменой a^2 на $-a^2$ (что эквивалентно замене a на ia). Одновременно нужно, конечно, изменить

и левую часть, которая в (1.3.14) положительна; именно нужно заменить x_4^2 на $-x_4^2$, оставляя то, что относится к трем координатам x_1, x_2, x_3 , неизменным. Тогда уравнение примет вид

$$x_1^2 + x_2^2 + x_3^2 - x_4^2 = -a^2, \quad (1.3.29)$$

а его решение есть

$$\left. \begin{aligned} x_1 &= a \operatorname{sh} \chi \cos \vartheta \cos \varphi, \\ x_2 &= a \operatorname{sh} \chi \cos \vartheta \sin \varphi, \\ x_3 &= a \operatorname{sh} \chi \sin \vartheta, \\ x_4 &= a \operatorname{ch} \chi. \end{aligned} \right\} \quad (1.3.30)$$

Квадрат расстояния между бесконечно близкими точками в этом случае есть

$$dl^2 = a^2 [d\chi^2 + \operatorname{sh}^2 \chi (\sin^2 \vartheta d\varphi^2 + d\vartheta^2)]. \quad (1.3.31)$$

Координата χ теперь меняется от нуля до бесконечности. Это выражение определяет геометрию трехмерного пространства всюду одинаковой отрицательной кривизны. Его радиус кривизны есть a . Сравним свойства этого пространства со свойствами пространства положительной кривизны (1.3.18). Расстояние между двумя точками с одинаковыми угловыми координатами имеет вид (1.3.22), однако координата χ теперь изменяется от нуля до бесконечности, и это расстояние может иметь сколь угодно большие значения, так же как в евклидовом пространстве. Далее, поверхность «сферы» радиуса $a\chi$, по аналогии с предыдущим случаем, можно записать в виде $4\pi a^2 \operatorname{sh}^2 \chi$. Отсюда видно, что, во-первых, отношение площади поверхности «сферы» к квадрату ее радиуса

$$\frac{4\pi \operatorname{sh}^2 \chi}{\chi^2}, \quad (1.3.32)$$

больше 4π . Во-вторых, при увеличении радиуса «сферы» ее площадь возрастает неограниченно. Наконец, полный объем пространства отрицательной кривизны,

$$V = 4\pi a^3 \int_0^{\infty} \operatorname{sh}^2 \chi d\chi, \quad (1.3.33)$$

бесконечен, так как координата χ изменяется до бесконеч-

ности. Такое пространство называют открытым. Оно однородно и изотропно, как и закрытое пространство, и также обладает свойством локальной евклидовости: на достаточно малых расстояниях от начала координат (которым может служить любая из равноправных точек этого пространства) выражение (1.3.31) сколь угодно мало отличается от евклидова выражения (1.3.20), так как $\text{sh } \chi \approx \chi$ при $\chi \ll 1$.

Итак, существуют три геометрических типа однородных изотропных трехмерных пространств: пространство нулевой кривизны, или евклидово пространство, а также пространства постоянной положительной и отрицательной кривизны. Как мы увидим в следующем параграфе, этим трем типам геометрии отвечают взаимно однозначным образом три динамических типа расширения однородной гравитирующей среды, рассмотренные в § 1.2.

§ 1.4. Космология Фридмана

В общей теории относительности пространство и время составляют, как мы говорили, единое многообразие, которое является римановым, и свойства его описываются интервалом ds . Выделение собственно трехмерного пространства из этого единого многообразия предполагает определенный выбор системы отсчета. Та геометрия изотропных пространств, которую мы изучали в предыдущем параграфе, должна пониматься как геометрия в системе отсчета, сопутствующей движению заполняющей пространство материи, т. е. в такой системе, в которой макроскопические скорости материи равны везде нулю. Действительно, в противном случае, если бы система отсчета двигалась относительно материи (так что скорость материи в ней была бы отлична от нуля), имелось бы выделенное направление в пространстве — направление этой скорости, и возникала бы кажущаяся анизотропия, неэквивалентность различных направлений в пространстве.

В сопутствующей системе отсчета, опирающейся на саму материю, время является собственным временем, одинаковым во всех точках пространства. По этой причине в общем выражении для квадрата интервала

$$ds^2 = g_{00}dt^2 + g_{0\alpha} dt dx_\alpha - dl^2 \quad (1.4.1)$$

коэффициент перед dt^2 не должен зависеть от координат. При этом время можно выбрать так, чтобы g_{00} было константой, которую удобно взять равной c^2 . Кроме того, совокупность величин $g_{0\alpha}$ ($\alpha = 1, 2, 3$) можно рассматривать как компоненты трехмерного вектора, наличие которого создавало бы выделенное направление в пространстве. Поэтому указанные величины должны быть равны нулю, и выражение для квадрата интервала примет тогда вид

$$ds^2 = c^2 dt^2 - a^2 [d\chi^2 + f^2(\chi) (\sin^2 \vartheta d\varphi^2 + d\vartheta^2)]. \quad (1.4.2)$$

Мы воспользовались здесь выражениями для трех типов пространственной геометрии, выведенными в предыдущем параграфе, для которых

$$f = \begin{cases} \sin \chi, & K > 0; \\ \chi, & K = 0; \\ \text{sh } \chi, & K < 0, \end{cases} \quad (1.4.3)$$

соответственно для пространств положительной, нулевой и отрицательной кривизны. Выражение (1.4.2) является в этом смысле самым общим; в реальной Вселенной осуществляется, разумеется, только одна из этих трех возможностей (см. ниже).

Координаты χ , ϑ , φ в выражениях (1.4.2), (1.4.3) — это сопутствующие координаты частиц вещества; они, так сказать, маркируют, нумеруют частицы, и значения этих координат для каждой данной частицы остаются неизменными во времени. В механике такие координаты называют лагранжевыми.

При таком описании пространственные расстояния между частицами пропорциональны величине a . Но, как мы знаем, однородная гравитирующая материя не может находиться в покое, и эти расстояния увеличиваются в ходе космологического расширения. Значит, величина a приобретает, кроме геометрического смысла, еще и динамический смысл: она должна зависеть от времени и служить масштабным фактором.

Динамику космологического расширения (в ньютоновском приближении) мы рассмотрели в § 1.2. Там мы выяснили, что существует три типа расширения однородной гравитирующей среды. С другой стороны, в § 1.3

было показано, что допустимы три типа однородных изотропных пространств. Возникает вопрос: какова связь между динамикой расширения геометрией пространства?

Естественно ожидать, что финитность движения при эллиптическом расширении имеет геометрическое соответствие в конечности объема замкнутого пространства постоянной положительной кривизны, а неограниченность движения при гиперболическом расширении — в бесконечности объема открытого пространства постоянной отрицательной кривизны. Промежуточный случай — параболическое расширение — следует тогда отождествить с пространством нулевой кривизны. Эти подсказываемые интуицией соответствия находят, как оказывается, строгое формальное подтверждение в общей теории относительности; имеется три и только три возможности:

$$\left. \begin{array}{l} 1) \quad E < 0, \quad K > 0; \\ 2) \quad E = 0, \quad K = 0; \\ 3) \quad E > 0, \quad K < 0. \end{array} \right\} \quad (1.4.4)$$

В § 1.2 мы ввели масштабный фактор, который там, как и в метрике (1.4.2), был обозначен через a . Обе величины, обозначенные этой буквой, по своему смыслу могут отличаться только постоянным множителем. Действительно, в малых областях пространства, где применимо (локально) ньютоновское описание, релятивистское выражение для расстояний между частицами $R = a(t)\chi$, $\chi \ll 1$, должно совпадать с ньютоновским выражением вида (см. § 1.2): $R = a(t)\psi(M)$. Так как зависимость от времени целиком содержится в первых множителях этих выражений, то они действительно должны быть пропорциональны друг другу.

Специально выбранный в § 1.2 вид функций $a(t)$, $\psi(M)$ как раз таков, что оба масштабных фактора, классический и релятивистский, просто совпадают. Масштабный фактор — это, вообще говоря, величина, которая определена лишь с точностью до постоянного множителя. Радиус же кривизны — это величина (пропорциональная масштабному фактору в силу общего расширения), имеющая вполне определенное значение. Уравнение для радиуса кривизны дается общей теорией относительности; оно

определяет его зависимость от времени и имеет, как мы и ожидали на основании локальной ньютоновости, ту же структуру, что и уравнение для расстояния $R(t)$ в ньютоновской теории:

$$(\dot{a})^2 = \frac{8\pi}{3} G \rho a^2 + e c^2, \quad e = -1, 0, 1. \quad (1.4.5)$$

Оно отличается от (1.2.29) лишь заменой величины E на $\frac{1}{2} e c^2$; в § 1.2 эта замена формально достигалась нужной нормировкой a . Согласно сказанному выше (см. (1.4.4)) значения $e = -1, 0, +1$ в соотношении (1.4.5) отвечают соответственно положительной, нулевой и отрицательной кривизне пространства.

Сравнивая снова релятивистское и классическое выражения для расстояния R , мы можем — в пределе малых расстояний — установить механический смысл лагранжевой координаты χ :

$$\chi = \psi(M) = \left(\frac{2|E|}{c^2} \right)^{1/2} = \left(\frac{GM}{\tau_0 c^2} \right)^{1/2}, \quad E \neq 0,$$

где, как и в § 1.2,

$$\tau_0 = \left(\frac{3}{32\pi G} \right)^{1/2} \frac{\rho}{(|\rho - \rho_c|)^{3/2}}.$$

Благодаря этому разность значений координаты χ для двух частиц, находящихся на расстоянии R , можно найти (в пределе $\chi \ll 1$), если известны интенсивные величины ρ и ρ_c .

Таким образом, мы получили довольно полное представление о структуре и динамике пространства-времени; мы сконструировали выражение для четырехмерного интервала ds и нашли функциональный вид входящей в него (переменной во времени) величины радиуса кривизны. При этом мы исходили лишь из самых общих представлений релятивистской теории и из свойств симметрии пространства, следующих из наблюдений.

Современному состоянию Вселенной, как мы видели в § 1.2, вероятнее всего отвечает гиперболическое расширение, близкое к своему асимптотическому, инерциальному, режиму. Последнее означает, что гравитационное взаимодействие между частицами среды несущественно. Раз

так, то обязанное гравитации искривление пространства-времени должно быть в силу тех же общих соображений незначительным. Пренебрегая им, мы приходим, таким образом, к галилееву пространству-времени. В таком случае должна существовать система отсчета, уже не обязательно сопутствующая материи, в которой выражение для четырехмерного интервала имеет вид

$$ds^2 = c^2 d\tau^2 - r^2 (\sin^2 \vartheta d\varphi^2 + d\vartheta^2) - dr^2. \quad (1.4.6)$$

Эти качественные соображения подтверждаются простым преобразованием выражения для метрики (1.4.2) при $E > 0$ в пределе $t \rightarrow \infty$, когда согласно (1.2.13)

$$ds^2 = c^2 dt^2 - (ct)^2 [(\sin^2 \vartheta d\varphi^2 + d\vartheta^2) + d\chi^2]. \quad (1.4.7)$$

Вводя вместо t и χ переменные τ и r ,

$$\tau = t \operatorname{ch} \chi, \quad r = ct \operatorname{sh} \chi,$$

находим, что такое преобразование действительно переводит метрику открытой Вселенной на поздней, инерциальной стадии расширения в галилееву метрику. Хотя метрики (1.4.6) и (1.4.7) для четырехмерного пространства-времени эквивалентны в пределе $t \rightarrow \infty$, трехмерные пространства в них совершенно различны: одно из них евклидово, а другое — неевклидово пространство отрицательной кривизны. Это обстоятельство демонстрирует ту идею релятивистской теории, что лишь четырехмерное пространство-время представляет собой физическую реальность, независимую от системы отсчета. Выделение же трехмерного пространства связано с выбором той или иной системы отсчета и потому в различных случаях приводит, вообще говоря, к различной трехмерной геометрии.

До сих пор во всех рассуждениях и количественных выводах мы обходились без прямого привлечения уравнений общей теории относительности. Формальное интегрирование этих уравнений подтверждает сделанные выше заключения. Для метрики (1.4.2) уравнения Эйнштейна (см. Дополнение) сводятся к двум уравнениям, одно из которых есть уравнение (1.4.5), а второе представляет собой термодинамическое тождество:

$$d\mathcal{E} = -p dV, \quad (1.4.8)$$

где \mathcal{E} — внутренняя энергия (включая энергию покоя), p — давление, V — объем, занятый данными частицами. Уравнение (1.4.8) записано для постоянной энтропии, что с большой точностью выполняется в космологических условиях. Так как $V \propto a^3$, это уравнение можно переписать в виде

$$\frac{d\varepsilon}{p + \varepsilon} = -3 \frac{da}{a}, \quad (1.4.9)$$

где ε — плотность (внутренней) энергии, которая связана с величиной ρ из формулы (1.4.5) соотношением $\varepsilon = \rho c^2$.

Если давление равно нулю, то из (1.4.9) следует

$$\varepsilon \propto \rho \propto a^{-3}, \quad (1.4.10)$$

так что масса в расширяющемся объеме, занятом данными частицами, не меняется со временем. Именно этот случай и рассматривался в § 1.2, так как для нерелятивистской среды (которую только и можно описывать классическими формулами механики) давление всегда много меньше плотности энергии (в которую включается и энергия покоя): $p \ll \rho c^2$. По этой причине функциональный вид зависимости a от t , найденный ранее, соответствует случаю, когда давлением среды можно пренебречь по сравнению с плотностью энергии. В дальнейшем (гл. 2) мы увидим, когда это условие выполняется в реальной Вселенной.

В случае релятивистской среды, когда давление может быть сравнимо с плотностью энергии, уравнения (1.4.5), (1.4.9) позволяют найти — путем дифференцирования (1.4.5) по времени и подстановки $\dot{\varepsilon}$ из (1.4.9) — соотношение, которое является релятивистским аналогом уравнения движения:

$$\ddot{a} = -\frac{4\pi}{3} G \left(\rho + 3 \frac{p}{c^2} \right) a. \quad (1.4.11)$$

Мы видим, что в то время как релятивистское уравнение (1.4.5) полностью аналогично закону сохранения механической энергии в классической теории, уравнение (1.4.11) существенно отличается от классического уравнения движения. В этом уравнении роль плотности гравитирующей массы играет величина

$$\rho_{\text{eff}} = \rho + 3 \frac{p}{c^2}. \quad (1.4.12)$$

Давление среды создает, таким образом, дополнительный вклад в гравитационную силу. В этом проявляется общее соотношение между гравитацией и геометрией, с одной стороны, и материей, — с другой, даваемое уравнениями общей теории относительности. В этих уравнениях «источником» тяготения является не только плотность массы, как в классическом уравнении Пуассона, а и тензор энергии-импульса, включающий в себя, в частности, и давление (см. Дополнение). Уравнения (1.4.5), (1.4.9) дают возможность полностью решить задачу о поведении радиуса кривизны (и масштабного фактора) в метрике (1.4.2), если известно уравнение состояния, т. е. зависимость давления от плотности энергии.

Для ультрарелятивистской среды,

$$p_r = \frac{1}{3} \varepsilon_r,$$

уравнения космологии (1.4.5), (1.4.9) приводят к решению, в котором, как и в решении для нерелятивистской среды, имеется три динамических варианта: гиперболическое расширение

$$a = a_r \operatorname{sh} \eta, \quad ct = a_r (\operatorname{ch} \eta - 1); \quad (1.4.13)$$

эллиптическое расширение

$$a = a_r \sin \eta, \quad ct = a_r (1 - \cos \eta); \quad (1.4.14)$$

параболическое расширение

$$a = a(t_0) \left(\frac{t}{t_0} \right)^{1/2}. \quad (1.4.15)$$

В первых двух случаях величина $a(t)$ имеет, согласно уже сказанному, смысл не только масштабного фактора, но и радиуса кривизны. При этом плотность среды, $\rho_r = \varepsilon_r/c^2$, дается соотношением

$$\rho_r = \frac{c^2}{\pi G} \frac{a_r^2}{a^4}. \quad (1.4.16)$$

Константа размерности длины a_r выражается через плотность среды ρ_r и критическую плотность, отвечающую параболическому расширению,

$$\rho_c = \frac{3}{32\pi G t^2}, \quad (1.4.17)$$

следующим образом:

$$a_r = \left(\frac{3c^2}{8\pi G} \right)^{1/2} \frac{\rho_r^{1/2}}{|\rho - \rho_c|}. \quad (1.4.18)$$

Как и для нерелятивистской среды с малым давлением, динамика гиперболического и эллиптического расширения, даваемая формулами (1.4.13) и (1.4.14), асимптотически при $t \rightarrow 0$, $\eta \rightarrow 0$ не отличается от параболической, так что при малых временах для всех трех вариантов имеем

$$a \propto t^{1/2}. \quad (1.4.19)$$

Для решения (1.4.13) — (1.4.18) сохраняются общие соображения о связи между динамическим типом расширения и кривизной сопутствующего материи трехмерного пространства.

В гл. 2 будет показано, что метagalактическая среда является ультрарелятивистской в первый миллион лет космологического расширения.

Одним из наиболее непосредственных следствий космологического расширения является красное смещение спектральных линий в излучении отдаленных источников, увеличивающееся с возрастанием расстояния до них. Рассмотрим теорию распространения света (или вообще любого электромагнитного излучения) в изотропной расширяющейся Вселенной, откуда будет следовать, в частности, и соотношение для длин волн и частот света.

Интервал ds между точками на мировой линии светового луча равен нулю (см. § 1.3). В изотропной Вселенной свет распространяется, очевидно, по прямой линии, и потому условие $ds = 0$ принимает для метрики (1.4.2) вид:

$$\pm c dt = a d\chi. \quad (1.4.20)$$

Введем вместо времени t вспомогательное время η , которое уже фигурировало прежде:

$$a d\eta = c dt. \quad (1.4.21)$$

Тогда уравнение (1.4.20) можно проинтегрировать:

$$\chi = \pm \eta + \text{const.} \quad (1.4.22)$$

В обеих формулах знаки \pm отвечают случаям распрост-

ранения света соответственно от начала координат $\chi = 0$ и к началу координат.

Чтобы выяснить смысл уравнения (1.4.22), вспомним, что возраст Вселенной конечен; в начальный момент $t = 0$ имела место сингулярность, при которой $a = 0$, $\rho = \infty$, а в дальнейшем расширение происходило с замедлением ($\ddot{a} < 0$, см. (1.4.11)), так что величина a возрастала медленнее, чем t . Вследствие этого вспомогательное время

$$\eta = \int_0^t \frac{c dt'}{a(t')},$$

в любой момент t имеет конечное значение. Согласно (1.4.22) это означает, что к данному моменту t все световые лучи распространились лишь на конечные расстояния по χ . Расстояние, на которое распространились лучи, испущенные в момент сингулярности,

$$L \approx ct,$$

представляет собой наибольшее расстояние, которое может разделять два причинно связанные события. Для наблюдателя, находящегося на одном из разбегающихся тел Вселенной, это есть максимальное расстояние, допускающее принципиальную возможность получения информации, т. е. принципиальный предел «видимости» в расширяющейся Вселенной. Расстояние L называют расстоянием до горизонта событий. Вследствие равноправия точек изотропного пространства это расстояние одинаково для всех наблюдателей, движущихся вместе с расширяющейся материей.

Расстояние до горизонта событий конечно в каждый момент времени как в закрытом, так и в открытом пространстве. Поэтому область пространства, доступная наблюдению, в обоих случаях ограничена, и в этом смысле различие между обоими типами пространств оказывается не столь значительным.

С геометрической точки зрения область, доступная наблюдению, представляет собой сечение четырехмерного пространства-времени световым конусом. Между тем то, что мы называем трехмерным пространством в сопутствующей системе отсчета, есть сечение пространства-времени

гиперплоскостью $t = \text{const}$, что означает «рассмотрение» всей Вселенной в один и тот же момент времени. И в том, и в другом случае это есть, так сказать, моментальная фотография Вселенной; в первом случае она «снята» при помощи реальных световых лучей, распространяющихся вдоль светового конуса (с конечной скоростью), а во втором — при помощи воображаемых лучей, распространяющихся с бесконечной скоростью. Различие этих трехмерных образов Вселенной иллюстрирует относительность разделения четырехмерного пространства-времени на пространство и время, а тем самым и относительность представления о конечности или бесконечности трехмерного пространства.

Вернемся к анализу уравнения (1.4.22). В открытом пространстве световой луч, испущенный источником, распространяется, неограниченно удаляясь от него. В замкнутом пространстве мы встречаемся с некоторыми особенностями в распространении света.

Одна из них аналогична той, которая имеет место при распространении (длинных) радиоволн вдоль земной поверхности. Волна, исходящая из некоторой точки A , может достигнуть другой точки на земной поверхности B как коротким путем, так и длинным путем, обойдя Землю и пройдя с другой стороны (оба пути равны только если точки A и B лежат на разных концах диаметра). Поэтому сигнал от радиостанции может быть принят дважды. Точно так же в замкнутом пространстве свет, исходящий из источника A , в принципе мог бы достигнуть другой точки B , пройдя как коротким, так и длинным путем. Первый путь соответствует «расстоянию» $\Delta\chi$, меньшему π , а второй — «расстоянию» $\Delta\chi$, большему π (последнее значит, что при распространении света χ сначала возрастает на π , а затем убывает). В силу уравнения (1.4.22) время η изменяется на коротком пути на величину, меньшую π , а на длинном — на величину, бóльшую π . Но мы знаем, что на фазе расширения η меняется в пределах от 0 до π , и потому на этой фазе световой луч успеет пройти от источника к приемнику только коротким путем. Двукратное изображение одного и того же источника на этой фазе невозможно и каждый источник виден лишь один раз. На фазе сжатия дело обстоит бы иначе и второе изображение («призрак») возможно.

Другое следствие конечности изменения координат в замкнутой Вселенной состоит в том, что свет, вышедший из некоторой точки, может снова вернуться в нее. При удалении от источника, находящегося, скажем, в начале координат, координата χ возрастает до величины π . Дальнейшее удаление от источника уже невозможно, и дойдя до «противоположного полюса» пространства, свет начнет затем снова приближаться к этому источнику и достигнет его, пройдя «расстояние» $\Delta\chi = \pi$ (от $\chi = \pi$ до $\chi = 0$).

Полное «расстояние», пройденное светом на всем замкнутом пути, равно 2π . В силу уравнения (1.4.22), на этот путь должен быть затрачен промежуток времени $\Delta\eta = 2\pi$. Но $\Delta\eta = 2\pi$ есть промежуток времени от одной сингулярности (начало расширения) до другой (конец сжатия).

На фазе расширения замкнутой Вселенной свет может распространиться к данному моменту времени на расстояние, не превосходящее ct , т. е. расстояние до горизонта; это расстояние в свою очередь не больше максимального расстояния $la(t)$. В последнем можно убедиться, если сравнить эти две величины в момент остановки расширения. Первое равно в этот момент ct_0 , а второе $2\pi ct_0$. При удалении в прошлое отношение t/a убывает, стремясь к нулю при $t \rightarrow 0$.

Перейдем теперь к тому явлению, которое представляет собою основное наблюдательное доказательство расширения Вселенной, а именно к красному смещению спектральных линий света, испускаемого удаленными телами.

Пусть в момент t из точки, координату χ которой мы выберем равной нулю, исходит световой сигнал, достигающий другой точки с координатой $\chi \neq 0$ в момент t' . Промежуток вспомогательного времени $\Delta\eta$ между испусканием и приемом сигнала согласно (1.4.22) равен «расстоянию» χ . Пусть теперь в момент $t + dt$ из начальной точки испускается второй сигнал, который достигает точки с координатой χ в момент $t' + dt'$. Разность значений вспомогательного времени, соответствующих испусканию обоих сигналов, есть

$$d\eta = \frac{c dt}{a(t)}.$$

Так как промежуток $\Delta\eta$ для обоих сигналов один и тот же и равен χ , то разность значений η , соответствующих приему обоих сигналов, есть тоже $d\eta$:

$$d\eta' = \frac{c dt'}{a(t')} = d\eta.$$

Отсюда следует, что

$$\frac{dt'}{dt} = \frac{a(t')}{a(t)}.$$

Значит, если промежуток времени между отправлением обоих сигналов и промежуток времени между их приемом невелики по сравнению с временем распространения сигналов от источника к приемнику, то мы можем сказать, что второй промежуток больше первого в таком же отношении, в каком за время распространения увеличился масштабный фактор. В частности, если два рассматриваемых сигнала представляют собой последовательные амплитудные значения испускаемого источником волнового поля, так что промежуток времени между ними равен периоду колебания, то период колебания, принятого источником, больше периода колебания, испущенного источником, в отношении соответствующих значений масштабного фактора $a(t)$. Возрастание периода при распространении света означает, очевидно, соответствующее возрастание⁷ длин волн, так как скорость распространения света в нашей метрике постоянна *) и равна c . Длины волн света возрастают в расширяющейся Вселенной по тому же закону, что и вообще все длины: $\lambda \propto a$. Наблюдатель воспринимает спектральные линии излучения сдвинутыми в красную сторону по сравнению с их стандартным положением, известным по лабораторным измерениям. Это красное смещение характеризуется величиной

$$z = \frac{\lambda' - \lambda}{\lambda} = \frac{a(t')}{a(t)} - 1. \quad (1.4.23)$$

*) Напомним, что скорость света равна фундаментальной константе c лишь в пространстве-времени с интервалом вида

$$ds^2 = c^2 dt^2 - dl^2$$

(так называемая синхронная метрика). В общем случае скорость света зависит от координат и времени.

Если за промежуток времени $t' - t$ масштабный фактор изменился мало (а это имеет место для не слишком больших расстояний между источником и приемником), то, разлагая $a(t')$ в ряд по степеням разности $t' - t$ и ограничиваясь первой степенью, мы получим

$$z \approx \frac{a(t) + \frac{da}{dt}(t' - t)}{a(t)} - 1 = \frac{\dot{a}}{a}(t' - t).$$

Расстояние между источником и приемником $R \simeq a\chi$ в то же время есть

$$R \simeq c(t' - t).$$

Скорость их взаимного удаления вследствие общего космологического расширения есть

$$V = \dot{a}\chi = \frac{\dot{a}}{a}R.$$

Подставляя это в формулу для z , найдем

$$z \approx V \frac{t' - t}{R} \approx \frac{V}{c}. \quad (1.4.24)$$

В этом приближении космологическое красное смещение описывается хорошо известной формулой для эффекта Доплера.

Так как согласно формуле (1.4.23) красное смещение z определяется моментом испускания света, то вместо того, чтобы характеризовать эти моменты значениями космологического времени t , их можно характеризовать значениями красного смещения z . Этот способ часто бывает удобен. При этом современной эпохе соответствует значение $z = 0$, а моменту сингулярности — значение $z = \infty$. Некоторой эпохе t в прошлом отвечает красное смещение

$$z = \frac{a_0}{a(t)} - 1, \quad (1.4.25)$$

где $a_0 = a(t_0)$ — современное значение величины a .

Красное смещение может играть роль времени и при описании таких космологических величин, как, например, плотность вещества:

$$\rho = \rho(t_0) \left(\frac{a_0}{a(t)} \right)^3 = \rho(t_0) (1 + z)^3. \quad (1.4.26)$$

В простейшем случае параболического расширения красное смещение связано с возрастом мира соотношениями:

$$t \propto (1+z)^{-3/2}, \quad p \ll \varepsilon; \quad (1.4.27)$$

$$t \propto (1+z)^{-2}, \quad p_r = \frac{1}{3} \varepsilon_r. \quad (1.4.28)$$

На инерциальной фазе расширения (которая, как говорилось, вероятно, реализуется в современную эпоху)

$$t \propto (1+z)^{-1}, \quad t > t_\Omega. \quad (1.4.29)$$

Моменту t_Ω перехода к инерциальному расширению отвечает, очевидно, красное смещение

$$z(t_\Omega) = z_\Omega = \Omega^{-1} - 1. \quad (1.4.30)$$

Так как параболические (степенные) формулы являются хорошим приближением для расширения двух других типов на ранних фазах (см. § 1.2), соотношения (1.4.27), (1.4.28) имеют более общий смысл и в дальнейшем мы часто будем пользоваться ими.

Изменение частоты при распространении света в расширяющемся мире представляет собой следствие некоторого более общего закона, относящегося не только к ультрарелятивистским частицам — фотонам, но и вообще к любым частицам, совершающим, как говорят, собственные движения. Под этим понимаются движения, дополнительные к общему регулярному космологическому расширению. При исследовании распространения света мы рассматривали фотоны как пробные частицы, т. е. как такие частицы, которые движутся в заданном внешнем поле, или в заданной метрике. Можно, очевидно, рассмотреть и пробные движения нерелятивистских частиц с отличной от нуля массой покоя.

Пусть, например, такая частица движется в нерелятивистской изотропно расширяющейся среде с малым давлением. В силу пространственной симметрии ее траектория является прямой линией,

$$r = r(t), \quad \vartheta = \text{const}, \quad \varphi = \text{const},$$

в системе отсчета, начало которой связано с одной из частиц регулярного фона, лежащей на траектории рассматриваемой частицы. Уравнение движения имеет при этом

тот же вид, что и для частиц фона:

$$\ddot{r}(t) = - \frac{GM}{r^2} = - \frac{\dot{a}}{a} r. \quad (1.4.31)$$

Его первый интеграл

$$\dot{r}a - \dot{a}r = \text{const.} \quad (1.4.32)$$

Собственная скорость частицы

$$v_p = \dot{r} - Hr = \dot{r} - \frac{\dot{a}}{a} r \quad (1.4.33)$$

удовлетворяет, согласно (1.4.32), соотношению:

$$v_p a = \text{const.} \quad (1.4.34)$$

При движении частиц с массой покоя на фоне ультра-релятивистской ($p_r = 1/3 \epsilon_r$) расширяющейся среды уравнение (1.4.31) сохранит свой вид, но под массой и плотностью надо в этом случае понимать эффективную массу и плотность: $M_{\text{eff}} = \frac{4\pi}{3} \rho_{\text{eff}} r^3$, $\rho_{\text{eff}} = \rho + 3 \frac{p}{c^2}$. Вид первого интеграла (1.4.32), а с ним и соотношение (1.4.34), останутся при этом неизменными.

Если вместо скорости частицы написать ее импульс, $P_p = mv_p$, то это соотношение примет вид

$$P_p a = \text{const.}, \quad (1.4.35)$$

который справедлив и для фотонов. Действительно, импульс фотона, P_{ph} , связан с его энергией, $E_{\text{ph}} = cP_{\text{ph}}$, а энергия E_{ph} — с частотой и длиной волны: $E_{\text{ph}} = \hbar\omega = \hbar c \frac{2\pi}{\lambda}$. Поэтому зависимость $\lambda \propto a$, описывающая красное смещение, равносильна выражению (1.4.35). (Здесь $\hbar = 1,1 \cdot 10^{-27} \text{г} \cdot \text{см}^2 \cdot \text{сек}^{-1}$ — постоянная Планка.)

В качестве собственных движений можно рассматривать, в частности, хаотические тепловые движения частиц среды. Для нерелятивистской среды с показателем адиабаты $\gamma = 5/3$ средняя скорость таких движений изменяется при расширении по закону:

$$\langle u \rangle \propto T^{1/2} \propto \left(\frac{p}{n} \right)^{1/2} \propto n^{\frac{\gamma-1}{2}} \propto a^{-1}, \quad (1.4.36)$$

что совпадает с (1.4.34) или (1.4.35). (Здесь T , p , n — температура, давление и концентрация частиц среды.)

Для тепловых движений в ультрарелятивистской среде с показателем адиабаты $\gamma = 4/3$ вместо (1.4.36) имеем

$$\langle P_p \rangle \propto T \propto \frac{P}{n} \propto n^{\gamma-1} \propto a^{-1}, \quad (1.4.37)$$

т. е. опять (1.4.35). Можно сказать, что красное смещение квантов, уменьшение собственных скоростей частиц и адиабатическое охлаждение среды при космологическом расширении — явления одной природы.

Теория изотропной Вселенной создана А. А. Фридманом в 1922—1924 гг. *). Предсказанное теорией явление нестационарности Метагалактики было вскоре открыто Э. Хабблом, обнаружившим красное смещение в спектрах далеких галактик. Это и послужило первым непосредственным наблюдательным подтверждением космологии Фридмана. На этом этапе развития релятивистской космологии значительный вклад в понимание связи между теорией и наблюдениями внес Ж. Леметр. Далеко идущая аналогия между релятивистской теорией и ньютоновской моделью однородного гравитирующего шара (из которой исходило в основном наше изложение) была замечена в 1934 г. У. Х. Мак-Кри и Э. А. Милном (см. также Зельдович (1965), Гинзбург (1970)).

Теория Фридмана дает полное описание динамических и геометрических свойств изотропной Вселенной. Она допускает три геометрических типа пространств в системе отсчета, сопутствующей материи, — закрытое, плоское и открытое, — взаимно однозначно связанные с тремя динамическими типами расширения. Мы видели в § 1.2, что выбор между последними может быть в принципе сделан на основе астрономических наблюдений. Тем самым дается возможность для наблюдательного решения вопроса о конечности или бесконечности пространства, сопутствующего движению вещества. Дальнейшее развитие теории связано с трудными проблемами, многие из которых до сих пор остаются нерешенными.

А. С и н г у л я р н о с т ь. Согласно сказанному в предыдущих параграфах Вселенная «родилась» конечное время назад в состоянии, выражаясь математически, бесконечной плотности и температуры. Очевидно, что та-

*) Главные космологические работы Фридмана перепечатаны в посвященном ему специальном выпуске УФН 80, № 3 (1963).

кое представление может быть только математической идеализацией некоего не известного нам состояния. До сих пор не ясно, что скрывается за этой идеализацией.

Б. Причинная несвязность событий. Если Вселенная была однородной и изотропной «с самого начала», т. е. начиная от сингулярности, то изменение во времени всех размеров и длин в ней определялось масштабным фактором $a(t)$. На ранней стадии расширения, когда преобладало излучение (см. гл. 2), масштабный фактор $a \propto t^{1/2}$. С другой стороны, расстояние, на которое за время t могла распространиться причинная связь событий, осуществляемая определенными физическими агентами (например, электромагнитными волнами), не превышало расстояния до горизонта событий ct . Отношение этих двух величин $ct/a \propto t^{1/2}$, т. е. оно стремилось к нулю по мере приближения к сингулярности. Это значит, что чем ближе к сингулярности, тем на меньших расстояниях могла быть осуществлена причинная связь, и в непосредственной близости от сингулярности эти расстояния бесконечно малы. Но при таких условиях совершенно непонятно, каким образом во всем объеме Вселенной однородность и изотропия могли быть обеспечены, каким образом удаленные друг от друга области Вселенной могли получить информацию о «нужном» значении плотности, температуры и т. д.

Ситуация была бы иной, если бы масштабный фактор возрастал быстрее, чем космологическое время t : тогда отношение ct/a расходило бы при стремлении t к нулю. Это было бы расширение с ускорением («антигравитация»). Согласно формулам (1.4.11), (1.4.12) сумма $\epsilon + 3p$ должна быть при этом отрицательной, что возможно лишь при отрицательном давлении. Такое состояние среды было бы механически неустойчивым, так как скорость звука, $c \cdot (\partial p / \partial \epsilon)^{1/2}$, была бы мнимой и любые малые возмущения неограниченно возрастали бы со временем. Но во всех реальных физических средах давление положительно и, значит, их расширение в собственном поле тяжести не может происходить с ускорением.

С течением времени размер области Вселенной, охватываемой единой причинной связью, возрастает пропорционально t . В настоящую эпоху он заведомо превосходит дальность действия астрономических инструментов,

т. е. размер наблюдаемой области Вселенной. Однородность и изотропия в этой области могли поэтому установиться, во всяком случае в принципе, благодаря физическим процессам.

Тем не менее проблема остается. Дело в том, что реликтовое излучение, которое, как мы знаем, исключительно однородно и изотропно (§ 1.1), перестало взаимодействовать с веществом не позже, чем при красном смещении $z = 6-8$, а, может быть, и значительно раньше (см. гл. 2). Так как фотоны друг с другом не взаимодействуют, то установление между фотонами реликтового излучения причинной связи могло иметь место только посредством их взаимодействия с веществом при $z > 6-8$. Расстояние до горизонта событий при указанных значениях z раз в 10 меньше современного расстояния до горизонта. Из области такого размера фотоны приходят к современному наблюдателю по направлениям, угол между которыми *)

$$\theta \approx \frac{t(z)}{t_0} \frac{1+z}{1 - \frac{1}{(1+z)^{1/2}}},$$

не превосходит 30° при $z > 7$; между тем наблюдаемая изотропия реликтового фона имеет место и в существенно больших угловых масштабах.

В рамках модели Фридмана изотропия излучения, а с ней и изотропия Вселенной, не может быть объяснена процессами не только на ранних стадиях расширения вблизи сингулярности, но и на более поздних стадиях, о которых нас информирует реликтовое излучение.

В. Начальные отклонения от изотропии и однородности. Космология Фридмана предполагает наивысшую симметрию распределения и движения вещества. В современной Вселенной она заведомо не существует в масштабе звезд, галактик, скоплений галактик. Для объяснения такой сравнительно мелкомасштабной неоднородности и анизотропии необходимо, как мы увидим в дальнейшем (гл. 3, 5), предположить, что Вселенная «с самого начала» не была идеально сим-

*) Эта формула легко получается с помощью соотношений, имеющих в настоящем параграфе.

метричной во всех масштабах. В ней существовали различные отклонения от изотропии и однородности: неравномерности плотности, собственные движения отдельных масс вещества и т. п. Эволюция этих возмущений должна привести в конце концов к формированию наблюдаемых космических структур. Это обстоятельство определенно указывает на необходимость выхода за рамки простой модели Фридмана. Хотя, возможно, проблема начальных возмущений и не имеет столь фундаментального характера, как две рассмотренные выше, она исключительно важна с космогонической точки зрения.

В свое время Л. Д. Ландау высказывал предположение о том, что сингулярность в модели Фридмана связана с далеко идущей идеализацией, с исключительно высокой симметрией, и что в космологических решениях более общего вида, анизотропных и неоднородных, сингулярность отсутствует. Если бы это было так, то переход к обобщенной модели устранил бы все три обсуждаемые здесь проблемы. Действительно, в несингулярной Вселенной нет проблемы причинности, так как время ее существования неограничено и все ее области имели возможность «обмениваться информацией». Наблюдаемая мелкомасштабная асимметрия была бы тогда остатком прежней неоднородности и анизотропии.

Однако позднее (в 1970 г.) С. Хоукинг и Р. Пенроуз показали, что сингулярность есть неустранимая черта космологических решений в общей теории относительности, сколь бы несимметричными по распределению и движению вещества эти решения ни были. Это означает, что первая из указанных проблем, проблема сингулярности, не есть исключительно «дефект» космологии Фридмана. Что же касается двух других проблем, то они, вероятно, не являются безнадежными и в современной космологии исследуются возможные пути их решения. Они связаны, например, с анализом анизотропных космологических моделей, которые мы рассмотрим в гл. 4. Анизотропные модели могут быть применимы лишь к достаточно ранним фазам космологического расширения. Их обязательным свойством должно быть приближение со временем к изотропной модели Фридмана, правильно описывающей современное состояние Вселенной, рассматриваемой в самых больших масштабах,

РЕЛИКТОВОЕ ИЗЛУЧЕНИЕ

§ 2.1. Излучение в горячей Вселенной

Измерения интенсивности реликтового излучения в различных длинах волн показывают, как мы говорили в § 1.1, что оно является равновесным с температурой $T_0 = 2,7$ °К. Согласно общим формулам статистической физики (см., например, Ландау и Лифшиц (1964)), число фотонов равновесного излучения с данной температурой в единице объема и в единичном интервале частот ω

$$n(\omega) = \frac{1}{\pi^2 c^3} \frac{\omega^2}{e^{\hbar\omega/kT} - 1}. \quad (2.1.1)$$

Это — распределение Планка и здесь \hbar — постоянная Планка, $k = 1,4 \cdot 10^{-16}$ эрг·град⁻¹ — постоянная Больцмана. Простой вывод формулы (2.1.1) мы дадим ниже, в § 2.3.

Плотность энергии (также на единичный интервал частот) получается отсюда умножением на $\hbar\omega$:

$$\epsilon(\omega) = \frac{1}{\pi^2 c^3} \frac{\hbar\omega^3}{e^{\hbar\omega/kT} - 1}. \quad (2.1.2)$$

В спектральном распределении энергии равновесного излучения (2.1.2) максимум энергии приходится на частоту

$$\omega_{\max} = 2,8 \frac{kT}{\hbar}. \quad (2.1.3)$$

Если выразить (2.1.2) через длины волн, $\lambda = 2\pi c/\omega$, то плотность энергии, приходящаяся на единичный интервал длин волн, запишется в виде

$$\epsilon(\lambda) = \frac{16\pi^2 c \hbar}{\lambda^5} \frac{1}{e^{2\pi\hbar c/kT\lambda} - 1}. \quad (2.1.4)$$

Максимум энергии отвечает длине волны

$$\lambda_{\max} = 1,3 \frac{\hbar c}{kT}. \quad (2.1.5)$$

Для современной температуры реликтового излучения находим

$$\omega_{\max} = 10^{12} \text{ сек}^{-1}, \quad (2.1.6)$$

$$\lambda_{\max} = \frac{2\pi c}{\omega_{\max}} = 0,2 \text{ см.} \quad (2.1.7)$$

В той области спектра, на которую приходится бóльшая часть измерений реликтового излучения, т. е. в области $\hbar\omega < kT$, называемой областью Рэлея — Джинса,

$$\varepsilon(\omega) \approx \frac{kT}{\pi^2 c^3} \omega^2. \quad (2.1.8)$$

В области $\hbar\omega > kT$, называемой областью Вина,

$$\varepsilon(\omega) \approx \frac{\hbar\omega^3}{\pi^2 c^3} e^{-\hbar\omega/kT}. \quad (2.1.9)$$

Полная плотность энергии излучения, получаемая интегрированием по всем частотам, дается формулой Стеффана — Больцмана:

$$\varepsilon_r = \alpha T^4. \quad (2.1.10)$$

Если температуру измерять в градусах, а ε_r в эрг/см^3 , то

$$\alpha = 7,5 \cdot 10^{-15} \text{ эрг/см}^3 \cdot \text{град}^4. \quad (2.1.11)$$

Вычисление по этой формуле дает для современной плотности реликтового излучения величину $\varepsilon_r(T_0) = 4 \cdot 10^{-13} \text{ эрг/см}^3$, которая и была приведена в § 1.1.

Плотность энтропии равновесного излучения есть

$$s = \frac{4}{3} \frac{\varepsilon_r}{T} = \frac{4}{3} \alpha T^3. \quad (2.1.12)$$

Эта величина пропорциональна концентрации фотонов,

$$n_r = 0,24 \left(\frac{kT}{\hbar c} \right)^3. \quad (2.1.13)$$

Для $T = T_0$ найдем

$$s = 2 \cdot 10^{-13} \text{ эрг/см}^3 \cdot \text{град}, \quad n_r = 4 \cdot 10^2 \text{ см}^{-3}. \quad (2.1.14)$$

Реликтовое излучение и космологическое расширение. Равновесный характер и высокая степень изотропии реликтового излучения указывают на то, что оно не является продуктом множества беспорядочных актов излучения каких-либо современных источников, а представляет собой скорее закономерное следствие предшествующей эволюции Вселенной. Если рассматривать все более ранние эпохи космологического расширения, то мы найдем, что температура, а с ней и плотность энергии излучения, нарастают в прошлое.

Последнее легко видеть из закона красного смещения, который мы получили в § 1.4: распространение волн в расширяющейся Вселенной сопровождается возрастанием длин волн пропорционально масштабному фактору. Но равновесное излучение в таком случае остается равновесным, и только положение максимума интенсивности сдвигается в сторону больших длин волн. Так как в равновесном излучении длина волны, соответствующая максимуму интенсивности, обратно пропорциональна температуре, то, следовательно, температура равновесного излучения уменьшается при расширении обратно пропорционально масштабному фактору:

$$T_r(t) = T_0 \frac{a(t_0)}{a(t)} = T_0(1+z). \quad (2.1.15)$$

Плотность энергии

$$\varepsilon_r = \varepsilon_r(t_0)(1+z)^4. \quad (2.1.16)$$

Так как плотность энергии вещества обратно пропорциональна кубу масштабного фактора,

$$\varepsilon = \varepsilon(t_0)(1+z)^3, \quad (2.1.17)$$

то отношение плотностей,

$$\frac{\varepsilon_r}{\varepsilon} = \frac{\varepsilon_r(t_0)}{\varepsilon(t_0)}(1+z),$$

нарастает в прошлое. Плотности вещества и излучения сравниваются при

$$1+z_* = \frac{\varepsilon(t_0)}{\varepsilon_r(t_0)} \approx 10^4 \Omega. \quad (2.1.18)$$

Красному смещению z_* отвечают следующие значения плотности и температуры реликтового излучения:

$$\left. \begin{aligned} \rho_r(z_*) = \rho(z_*) \equiv \rho_* \approx 5 \cdot 10^{-18} \Omega^4 \text{ г/см}^3, \\ T_r(z_*) \equiv T_* \approx 3 \cdot 10^4 \Omega \text{ }^\circ\text{К.} \end{aligned} \right\} \quad (2.1.19)$$

Из соотношений (2.1.16) — (2.1.18) следует, что в раннюю эпоху расширения ($z > z_*$) энергия излучения больше энергии вещества; поэтому, в частности, динамика расширения определялась в ту эпоху именно излучением. Излучение — газ фотонов, частиц, движущихся со скоростью света, которому отвечает ультрарелятивистское давление $p_r = 1/3 \epsilon_r$. Два предельных случая состояния среды, $p = 0$ и $p_r = 1/3 \epsilon_r$, рассмотренные в § 1.4, осуществлялись, следовательно, в эпоху преобладания вещества, $z < z_*$, и в эпоху преобладания излучения, $z > z_*$, соответственно.

Возраст мира на рубеже этих эпох, т. е. время, прошедшее от начала расширения $z = \infty$ до $z = z_*$, можно оценить, воспользовавшись связью между космологическим временем и красным смещением, указанной в § 1.4. Кроме того, учтем, что при значениях красного смещения, даваемых формулой (2.1.18), динамика расширения является с большой точностью параболической, так что

$$2\rho_* \approx \frac{1}{6\pi G t_*^2}.$$

Отсюда, с учетом (2.1.18), (2.1.19), найдем

$$t_* \approx 3 \cdot 10^{11} \Omega^{-2} \text{ сек.} \quad (2.1.20)$$

Эта оценка является приближенной; точные формулы можно получить из космологического решения, учитывающего как вещество, так и излучение (см. ниже).

Для ранней эпохи, когда преобладает излучение, связь между возрастом мира и красным смещением дается, в соответствии с общими соотношениями § 1.4, формулой

$$t = t_* \left(\frac{1 + z_*}{1 + z} \right)^2 \approx 3 \cdot 10^{19} z^{-2} \text{ сек.} \quad (2.1.21)$$

Для более поздней эпохи, когда преобладает вещество (но в приближении параболичности расширения),

$$t = t_* \left(\frac{1 + z_*}{1 + z} \right)^{3/2} \approx 3 \cdot 10^{17} \Omega^{-1/2} (1 + z)^{-3/2} \text{ сек.} \quad (2.1.22)$$

Космологическая среда в обе эпохи своей эволюции состоит из двух компонент — излучения и вещества. Решение уравнений фридмановской космологии (§ 1.4) для масштабного фактора $a(t)$, справедливое как в упомянутых предельных случаях, так и во всей области $0 < t < t_0$, имеет вид (Чернин (1965)):

$$\left. \begin{aligned} a &= a_m (\operatorname{ch} \eta - 1) + a_r \operatorname{sh} \eta, \\ ct &= a_m (\operatorname{sh} \eta - \eta) + a_r (\operatorname{ch} \eta - 1). \end{aligned} \right\} \quad (2.1.23)$$

Формулы (2.1.23) относятся к гиперболическому типу расширения, или открытому миру Фридмана. Параметр η имеет тот же смысл, что и в §§ 1.2, 1.4. Выражения для a и t представляют собой суммы решений для вещества и излучения по отдельности*). Постоянные коэффициенты a_m и a_r определяются вкладами вещества и излучения:

$$\left. \begin{aligned} a_m &= \frac{1}{2} \kappa^{-1/2} \rho (\rho_c - \bar{\rho})^{-3/2}, \\ a_r &= \kappa^{-1/2} \rho_r^{1/2} (\rho_c - \bar{\rho})^{-1}. \end{aligned} \right\} \quad (2.1.24)$$

Здесь $\kappa = 8\pi G/3c^2$, $\bar{\rho} = \rho + \rho_r$ ($\bar{\rho} < \rho_c$). Текущая критическая плотность

$$\rho_c(t) = \frac{3}{8\pi G} H^2, \quad H \equiv \frac{1}{a} \frac{da}{dt}.$$

Формулы для эллиптического расширения, или для закрытого варианта модели Фридмана, имеют ту же структуру:

$$\left. \begin{aligned} a &= a_m (1 - \cos \eta) + a_r \sin \eta, \\ ct &= a_m (\eta - \sin \eta) + a_r (1 - \cos \eta). \end{aligned} \right\} \quad (2.1.25)$$

Постоянные a_m и a_r даются формулами вида (2.1.24), в которых нужно только заменить теперь $(\rho_c - \bar{\rho})$ на $(\bar{\rho} - \rho_c)$.

Для параболического расширения (ему отвечает плоское сопутствующее пространство) решение можно записать также в симметричном параметрическом виде:

$$\left. \begin{aligned} a &= \frac{1}{2} a_m \eta^2 + a_r \eta, \\ ct &= \frac{1}{6} a_m \eta^3 + \frac{1}{2} a_r \eta^2, \end{aligned} \right\} \quad (2.1.26)$$

*) Что нетривиально, ибо сами уравнения нелинейны.

где в этом случае константы a_m и a_r , как и сам масштабный фактор a , определены с точностью до постоянного множителя *):

$$a_m = \frac{1}{6} \kappa \rho a^3, \quad (2.1.27)$$

$$a_r^2 = \frac{1}{3} \kappa \rho_r a^4. \quad (2.1.28)$$

Две последние формулы, дающие связь между плотностями и масштабным фактором, справедливы для всех трех изотропных моделей.

При малых временах $t \rightarrow 0$, $\eta \rightarrow 0$, выражения (2.1.23) и (2.1.25) переходят, в соответствии с общими динамическими особенностями изотропного расширения (§ 1.2), в соотношения вида (2.1.26), сохраняя значения величин a_m и a_r , выраженных, скажем, через современные плотности $\rho(t_0)$, $\rho_r(t_0)$, $\rho_c(t_0)$ с помощью формул (2.1.24).

Согласно формулам (2.1.27), (2.1.28) радиус кривизны в момент равенства плотностей вещества и излучения

$$a(z_*) \equiv a_* = \frac{1}{2} \frac{a_r^2}{a_m}. \quad (2.1.29)$$

Из (2.1.26) находим соответствующие значения параметра η и космологического времени:

$$\eta_* = \frac{a_r}{a_m} (\sqrt{2} - 1), \quad (2.1.30)$$

$$t_* = \frac{1}{6} \frac{a_m}{c} \left(\eta_*^3 + 3 \frac{a_r}{a_m} \eta_* \right). \quad (2.1.31)$$

Динамическая роль давления, устанавливаемая релятивистской теорией тяготения (§ 1.4), проявляется в решении (2.1.23) — (2.1.28) в том, что ускорение \ddot{a} определяется не суммой плотностей вещества и излучения, как это можно было бы ожидать по ньютоновской теории,

*) Напомним, что в моделях с искривленным сопутствующим пространством величина a , даваемая формулами (2.1.23) и (2.1.25), имеет смысл радиуса кривизны. Она имеет также смысл и масштабного фактора, но выбранного так, что он всегда совпадает по величине с радиусом кривизны. В плоском пространстве эта величина имеет только смысл масштабного фактора.

а эффективной гравитирующей плотностью, включающей давление *):

$$\rho_{\text{eff}} = \rho + \rho_r + 3 \frac{P_r}{c^2} = \rho + 2\rho_r.$$

Это проявляется также и в том, что формулы для наблюдаемых величин в изотропной модели, учитывающей обе компоненты среды, содержат не полную, а эффективную плотность. Так, соотношение, связывающее красное смещение спектральных линий света, испущенного источником, с расстоянием r до источника, имеет вид **):

$$\frac{\omega_0 - \omega}{\omega_0} = \frac{H}{c} \left[1 - \frac{1}{2} \left(1 - \frac{1}{2} \frac{\rho_{\text{eff}}}{\rho_c} \right) \frac{H}{c} r \right]. \quad (2.1.32)$$

Расстояние r до источника с данной сопутствующей координатой χ определяется формулой

$$r = \int_0^\chi \sqrt{g_{11}} d\chi' = a\chi,$$

в которой сопутствующая координата наблюдателя принята равной нулю, а компонента метрического тензора g_{11} (коэффициент перед квадратом дифференциала радиальной координаты) определяется метрикой Фридмана. Соотношение (2.1.32) является точным для модели с плоским сопутствующим пространством, а для моделей с искривленным пространством — приближенным с точностью до первых двух членов разложения по отношению расстояния до объекта к расстоянию до горизонта, $Hr/c \sim \sim r/ct \ll 1$.

В решении (2.1.23) — (2.1.28), как и в космологических решениях, рассмотренных в гл. 1, расширение метагалактической среды происходит адиабатически. В изотропной Вселенной условие адиабатичности с хорошей точностью выполнялось по крайней мере до эпохи образования звезд и галактик (подробный анализ имеется в кни-

*) Но тип расширения (и знак кривизны) зависит от соотношения полной и критической плотностей.

***) Общие формулы для других наблюдаемых величин, обладающие той же особенностью, даются в работе Чернина (1965); подробные вычисления и таблицы имеются в работах Сапара (1964, 1965, 1970).

ге Вейнберга (1972)). Поэтому энтропия некоторого расширяющегося объема, приходящегося на данное число нуклонов (энтропия сопутствующего объема), не меняется со временем. Энтропия излучения в каждом таком объеме много больше энтропии вещества (когда температуры нерелятивистские; см. ниже), и потому с хорошей точностью можно говорить о сохранении энтропии излучения.

Как мы видели в начале параграфа, энтропия единицы объема равновесного излучения равна с точностью до численного множителя концентрации фотонов. Поэтому сохранение энтропии эквивалентно сохранению числа фотонов в сопутствующем объеме:

$$n_r(t) a^3(t) \propto n_r(t) (1+z)^{-3} = \text{const.}$$

Именно такое соотношение и следует из формул (2.1.12) — (2.1.15). Так как нуклоны тоже сохраняются, их концентрация изменяется со временем по тому же закону, что и концентрация фотонов, и отношение числа фотонов к числу нуклонов есть величина, не меняющаяся при космологическом расширении:

$$\frac{n_r}{n} \approx 10^8 \Omega^{-1}. \quad (2.1.33)$$

Кроме реликтовых фотонов во Вселенной должны иметься реликтовые нейтрино (см. следующий раздел), а, возможно, и гравитоны; те и другие являются релятивистскими частицами и должны быть учтены в оценке плотности релятивистской компоненты метagalктической среды. Вводя поправочный множитель ξ , равный отношению полной плотности энергии ультрарелятивистского фона к измеренной плотности энергии реликтового электромагнитного излучения, найдем оценку констант (2.1.27), (2.1.28):

$$a_m \approx 10^{28} \Omega (|1 - \Omega|)^{-3/2} \text{ см.},$$

$$a_r \approx 2 \cdot 10^{26} \xi (|1 - \Omega|)^{-1} \text{ см.}$$

Значение величины ξ остается не вполне определенным, главным образом, из-за того, что вклад гравитационных волн не поддается точному вычислению без привлечения дополнительных предположений. При вероятном значе-

нии $\xi = 1,5-3$ постоянные a_m и a_r оказываются близкими:

$$a_m \sim a_r \approx a_* \approx 10^{26} \text{ см},$$

а само решение (2.1.23) при $a_m = a_r = A$ принимает простой и удобный для вычислений вид (Чернин (1968б)):

$$a = A (e^n - 1),$$

$$ct = A (e^n - \eta - 1).$$

Интересно и другое совпадение: константы $a_m \sim a_r$ близки по порядку величины к численному значению выражения $h^2/Gm^2 \approx 10^{26} \text{ см}$, составленного из универсальных констант и, так же как и a_m, a_r , имеющего размерность длины ($h = 2\pi\hbar$). При этом отношение n_r/n , являющееся мерой зарядовой асимметрии Вселенной (см. следующий раздел), также выражается через одни только фундаментальные константы (Зельдович и Новиков (1967)).

Это один из примеров численной связи между космологическими и микроскопическими константами. Поиск и анализ такого рода совпадений, в надежде, что они неслучайны, — восходят к П. Дираку, заметившему сорок лет назад, что имеет место близость двух больших чисел: отношения электрической силы, действующей между электроном и протоном, к гравитационной силе между ними

$$e^2/Gm_e m_p \sim 10^{40},$$

и отношения современного радиуса горизонта $\sim ct_0$ к классическому радиусу электрона

$$ct_0 \left/ \frac{e^2}{m_e c^2} \right. \sim 10^{40}.$$

Дирак предположил, что в силу некоего неизвестного физического принципа эти две величины равны не случайно и их совпадение отражает глубинную связь микромира с Вселенной в целом. Отсюда следовало, в частности, что так как радиус горизонта увеличивается со временем, гравитационная константа должна в том же темпе уменьшаться со временем, чтобы обе величины оставались близкими друг к другу. Из этой идеи родилась впоследствии новая теория гравитации — теория Бранса и Дикке (1961). Другие численные совпадения в космоло-

гии обсуждались Эддингтоном (1934), Гамовым (1967), Зельдовичем (1968), Крамаровским и Чечевым (1970), Харрисоном (1972).

Первые пятнадцать минут. Экстраполяция теории Фридмана до космологической особенности, когда $t \rightarrow 0$, $z \rightarrow \infty$, приводит к бесконечным значениям плотности и температуры. Это означает, что в этот момент и в непосредственной близости от него известные законы физики неприменимы. Однако через некоторое время после начала расширения плотность вещества уменьшается до значения, соответствующего плотности атомных ядер. После этого можно пользоваться «обычной» физикой. Ядерной плотности $\rho_N \approx 10^{15} \text{ г/см}^3$ отвечает красное смещение $z_N = (\rho_N / \rho_c \Omega)^{1/3} \approx 3 \cdot 10^{14} - 10^{15}$, и температура $T_N \approx (1 \div 3) \cdot 10^{15} \text{ }^\circ\text{К}$. Энергия, соответствующая этой температуре, превышает энергию покоя нуклонов, $mc^2 \approx 10^{-3} \text{ эрг}$. При этой температуре в среде имелись нуклон-антинуклонные, электрон-позитронные, нейтрино-антинейтринные пары, мезоны и, разумеется, электромагнитное излучение*). Все эти частицы взаимодействовали друг с другом и с излучением, причем характерные времена столкновений были меньше возраста мира и потому метagalактическая среда находилась в состоянии статистического равновесия (подробнее об этой эпохе см. в книгах Зельдовича и Новикова (1975) и Вейнберга (1972)). Это важное обстоятельство означает, что физические явления и состав среды на предшествующей стадии, когда плотность вещества была сверхъядерной, «забываются» и никак не влияют на состояние при $z < z_N$, которое, как и всякое равновесное состояние, однозначно определяется температурой и значениями сохраняющихся физических величин.

Одной из таких величин является барионный заряд, т. е. избыток барионов над антибарионами. Если этот избыток действительно имеется, он характеризует зарядовую асимметрию Вселенной. Другая такая величина — электрический заряд. Значения этих величин не следуют из уравнений космологии и служат для них, так сказать, начальными условиями; их не удастся вывести из «первых принципов». Предположение о равенстве нулю элект-

*) При достаточно высоких температурах должны были иметься пары гиперонов, а еще раньше и свободные кварки, если последние действительно существуют.

рического заряда ничему не противоречит и его мы молчаливо принимали в гл. 1, когда в качестве единственной силы, действующей на частицы, рассматривали гравитационную силу.

Что же касается барионного заряда, то его, по-видимому, следует считать отличным от нуля. Барионный заряд — это число барионов, которые остаются после аннигиляции барион-антибарионных пар, когда температура падает в ходе космологического расширения ниже значения $kT \approx mc^2$ (такой температуре отвечает красное смещение $z = z_N (mc^2/kT_N)$). Эти барионы и составляют наблюдаемые космические тела в современную эпоху. Барионный заряд мог бы быть равным нулю, если в современной Вселенной вещество и антивещество имелись бы в равных количествах; но тогда в силу каких-то причин они должны были оказаться пространственно разделенными, начиная с эпохи аннигиляции. Возможные механизмы разделения вещества и антивещества исследовались в ряде работ, обзор и критический анализ которых дается, например, в книге Зельдовича и Новикова (1975). Эти механизмы, по-видимому, недостаточно эффективны, и, скорее всего, следует считать зарядовую асимметрию Вселенной одним из ее фундаментальных изначальных свойств.

Мерой зарядовой асимметрии является отношение числа избыточных барионов B к числу барион-антибарионных пар $B + \bar{B}$ при высоких ($kT > mc^2$) температурах. Концентрации пар всех сортов примерно равны концентрации фотонов в состоянии равновесия. Поэтому такая мера зарядовой асимметрии почти в точности равна величине, обратной большому космологическому числу (2.1.33):

$$\frac{B}{B + \bar{B}} \sim 10^{-8} \Omega.$$

Падение плотности и температуры при космологическом расширении приводит к ряду изменений в составе и состоянии метagalактической среды. Прежде всего происходит аннигиляция нуклон-антинуклонных пар, а после этого нейтрино перестают взаимодействовать с другими компонентами среды, так как длина их свободного пробега становится сравнимой с расстоянием до горизонта и в дальнейшем превышает его. Это происходит сначала

с мюонными, а потом и с электронными нейтрино (рис. 4). После этого газ нейтрино обоих сортов адиабатически охлаждается в процессе космологического расширения, сохраняя равновесное (фермиевское) распределение по энергии. Нейтринный фон должен существовать в метagalактике и в современную эпоху. Эти реликтовые нейтрино должны иметь температуру, близкую к температуре

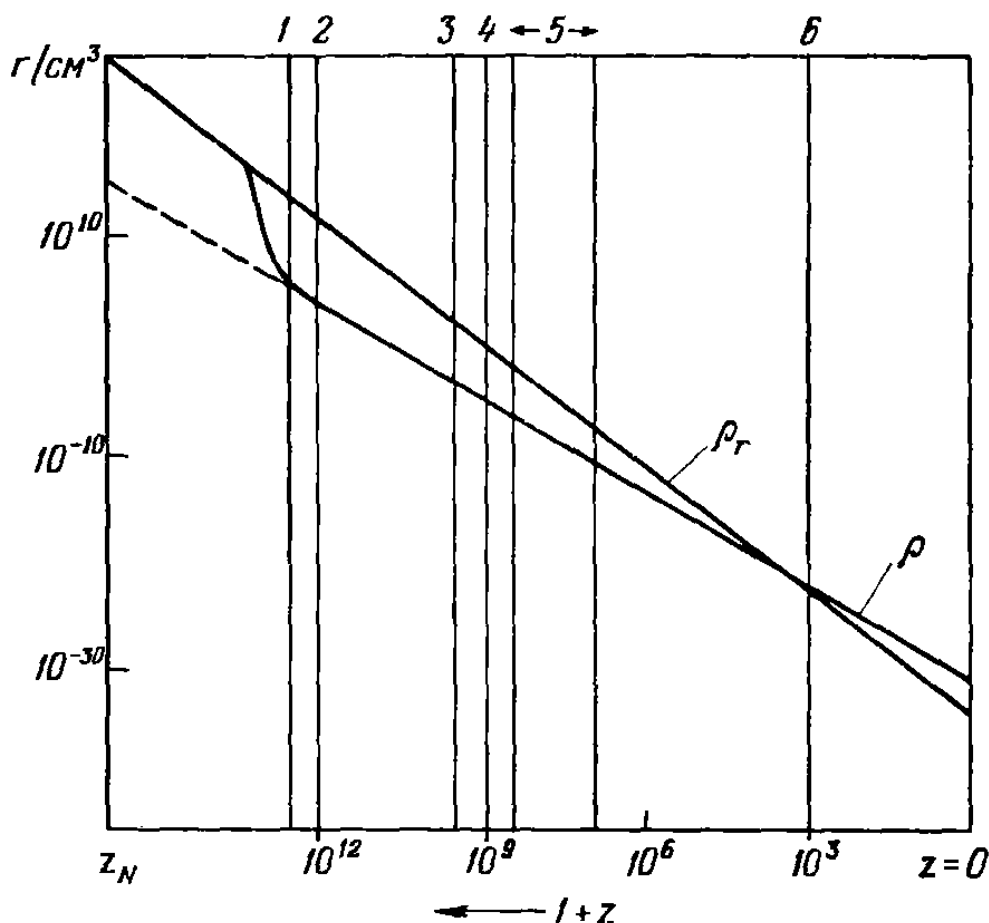


Рис. 4. Последовательность событий в ранней Вселенной: 1 — аннигиляция антинуклонов; 2, 3 — отделение мюонных и электронных нейтрино; 4 — аннигиляция позитронов; 5 — нуклеосинтез; 6 — рекомбинация плазмы.

реликтового электромагнитного излучения, так как плотности обеих этих компонент были близки друг к другу в состоянии равновесия (скажем, при z_N), а закон их изменения при адиабатическом расширении один и тот же, ибо обе компоненты представляют собой ультрарелятивистский газ. Температура нейтрино в современную эпоху около 2°K (Пиблс (1966)), а их плотность энергии вдвое меньше плотности энергии реликтового излучения. Экспериментальное обнаружение реликтовых нейтрино остается пока вне возможностей современной техники наблюдений.

В соответствии с падением температуры аннигилируют пары мезонов, и, наконец, при температуре $T \approx m_e c^2/k \approx 5 \cdot 10^9$ °К и красном смещении $z = z_N (m_e c^2/kT_N)$ происходит аннигиляция электрон-позитронных пар, причем остаются избыточные электроны в числе, равном числу избыточных протонов (если, как мы считали выше, общий электрический заряд среды равен нулю).

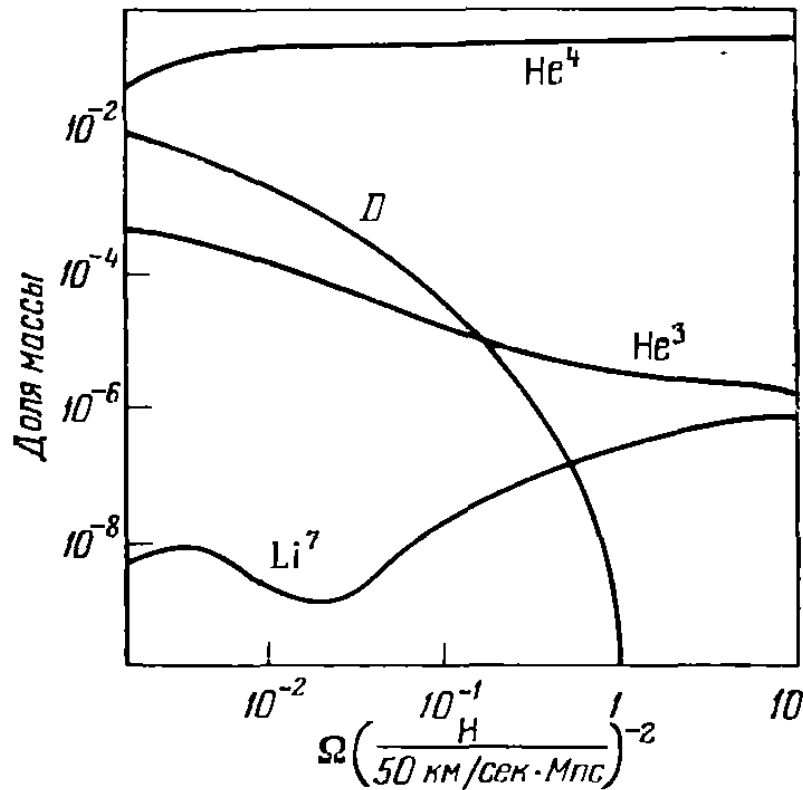


Рис. 5. Образование элементов в первичном нуклеосинтезе.

Приблизительно в ту же эпоху происходит еще одно важное изменение состава среды — синтез атомных ядер. Тепловые движения протонов и нейтронов становятся достаточно слабыми и последние могут объединяться и связываться ядерными силами, но остаются еще достаточно сильными для того, чтобы протоны могли преодолевать при сближениях порог кулоновского отталкивания. Это продолжается приблизительно пятнадцать минут, и ядерные реакции не успевают зайти слишком далеко: образуется главным образом гелий-4 (два протона и два нейтрона) в количестве 25—30 процентов по весу*). Синтез гелия не очень чувствителен к космологическим параметрам (прежде всего, к современной плотности вещества),

*) Детальное обсуждение ядерных реакций в космологии имеется в книгах Зельдовича и Новикова (1975) и Пиблса (1975).

которые известны недостаточно точно. Другие ядра — дейтерий, гелий-3, литий-7 — образуются в очень малых количествах (меньше процента), причем их доля довольно сильно зависит от космологических параметров (рис. 5). Содержание гелия-4, образующегося в космологическом нуклеосинтезе, в общем хорошо согласуется с данными о химическом составе атмосфер наиболее старых звезд,

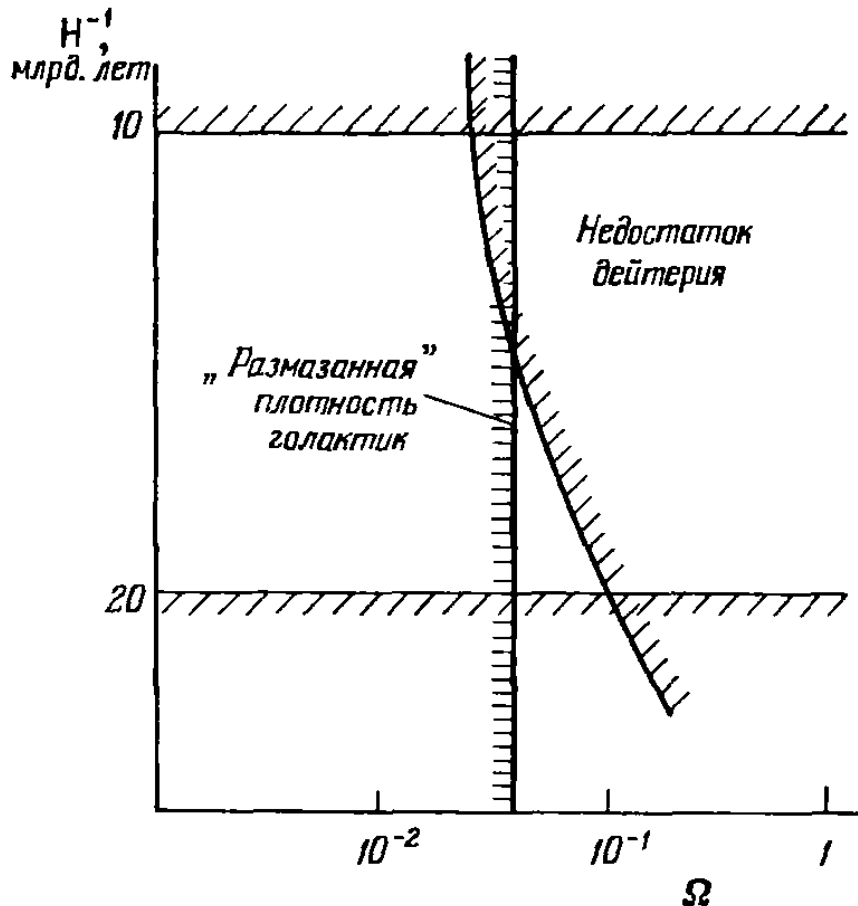


Рис. 6. Ограничения на космологический параметр Ω в современную эпоху по распространенности дейтерия (две горизонтальные прямые ограничивают область вероятных значений H^{-1}).

где, как можно полагать, преобладает «первичное» дозвездное вещество. Что же касается других ядер, то наблюдательные данные о них менее определены. Их уточнение позволило бы, возможно, дать независимую оценку современной средней плотности вещества в Метагалактике. Если окажется, например, что наблюдаемый дейтерий имеет целиком космологическое происхождение, то можно будет утверждать, что эта плотность не превосходит десятой доли критической плотности (см. рис. 6, построенный по данным Ривса и др. (1973)).

Представление о ранней стадии эволюции Вселенной в предположении ее высокой температуры и энтропии

было разработано в конце 40-х годов Г. Гамовым и его сотрудниками на основе космологии Фридмана. Теория «горячей» Вселенной Гамова включала в себя, в частности, идею первичного нуклеосинтеза; она позволила сделать одно из наиболее важных предсказаний в современной астрофизике — предсказание реликтового излучения, которое действительно было открыто в 1965 г.

Эпоха рекомбинации. После нуклеосинтеза эволюция метagalактической среды происходит длительное время спокойно, без изменения ее состава, и только соотношение между энергиями излучения и вещества постепенно меняется в пользу последнего (см. рис. 4).

События, важные для формирования наблюдаемых космических структур, разыгрываются во Вселенной главным образом в более позднюю эпоху, и наиболее существенным из них является рекомбинация метagalактической плазмы, служащая переломным пунктом эволюции метagalактической среды. К этой эпохе температура среды падает настолько, что электроны и ионы начинают объединяться в нейтральные атомы; взаимодействие вещества и излучения сначала ослабевает, а потом и практически прекращается.

Мы можем оценить параметры метagalактической среды в эпоху рекомбинации на основании термодинамических соображений. Напомним, однако, сначала некоторые общие сведения, относящиеся к термодинамике процесса ионизации и рекомбинации, которые пригодятся нам и в дальнейшем — в звездной динамике (!); см. § 6.4.

Равновесие системы при заданных давлении и температуре соответствует минимуму ее термодинамического потенциала $\Phi(p, T)$. Так как последний есть экстенсивная величина, т. е. он пропорционален объему системы или числу ее частиц, а давление и температура — величины интенсивные, т. е. характеризующие свойства системы локально, то термодинамический потенциал можно представить в виде

$$\Phi = \mu N,$$

где коэффициент μ , называемый химическим потенциалом, зависит лишь от давления и температуры. Если система состоит из нескольких компонент, представляющих собой идеальные газы (только такие системы мы и рассматри-

ваем), то ее термодинамический потенциал есть сумма соответствующих слагаемых:

$$\Phi = \sum_k \mu_k N_k.$$

В частности, если какой-либо газ с числом частиц N_a и химическим потенциалом μ_a находится в равновесии со своими однократными ионами (соответствующие параметры μ_i , N_i) и электронами (параметры μ_e , N_e), то

$$\begin{aligned} \Phi &= \mu_a N_a + \mu_i N_i + \mu_e N_e = \\ &= \mu_a (N - N_i) + (\mu_i + \mu_e) N_i, \end{aligned}$$

где N — полное число атомов газа, как ионизованных, так и не ионизованных. Тогда минимум термодинамического потенциала определяется равенством:

$$\frac{\partial \Phi}{\partial N_i} = \mu_i + \mu_e - \mu_a = 0. \quad (2.1.34)$$

(Экстремум такого рода действительно является минимумом, как показывается в термодинамике.) Это означает, что сумма химических потенциалов частиц, появляющихся в данном процессе, должна в равновесии равняться химическому потенциалу исчезающих частиц.

Найдем химический потенциал частиц каждого из этих типов. Для этого вспомним, что равновесная функция распределения частиц одноатомного идеального газа, т. е. число частиц в объеме dV и в интервале импульсов частиц $d^3p \equiv dp_x dp_y dp_z$, имеет вид

$$f(p) \frac{dV d^3p}{(2\pi\hbar)^3} = F e^{-E(p)/kT} \frac{dV d^3p}{(2\pi\hbar)^3}, \quad E(p) = \frac{p^2}{2m}. \quad (2.1.35)$$

Если она нормирована на полное число частиц *) N ,

$$\int f(p) \frac{dV d^3p}{(2\pi\hbar)^3} = N, \quad (2.1.36)$$

то, как показывается в статистической физике,

$$F = e^{\mu/kT}.$$

*) Интегрирование по всему объему, занятому данными частицами, и от $-\infty$ до $+\infty$ по каждой компоненте импульса.

Деля на объем и вынося $e^{\mu/kT}$ за знак интеграла, получаем

$$\mu = -kT \ln(Z/n), \quad (2.1.37)$$

где

$$Z = \int_0^{\infty} e^{-p^2/2mkT} g \frac{4\pi p^2 dp}{(2\pi\hbar)^3} = g \left(\frac{mkT}{2\pi\hbar^2} \right)^{3/2}, \quad (2.1.38)$$

а n — концентрация частиц данного сорта.

Коэффициент g означает статистический вес каждого состояния; для электронов и протонов он обусловлен спином и равен 2, а для нейтрального атома водорода он равен произведению статистических весов ядра и электрона, т. е. 4. Мы ввели сферические координаты в пространстве импульсов; множитель 4π есть результат интегрирования по телесному углу. Величина Z называется, как известно, статистическим интегралом (на единицу объема). В (2.1.38) он записан для случая, когда у частиц имеется только кинетическая, но не потенциальная энергия. Если учесть дополнительно, что для ионизации атома необходима затрата энергии ионизации (для водорода $E_i = 13,6$ эв, что соответствует температуре $T_i = 158\,000$ °К), то из условия (2.1.34) найдем соотношение, называемое формулой Саха:

$$\frac{n_e n_i}{n_a} = \frac{Z_e Z_i}{Z_a} e^{-E_i/kT}, \quad (2.1.39)$$

где Z_e , Z_i , Z_a — выражения вида (2.1.38), отнесенные к электронам, ионам и атомам соответственно. Из (2.1.38) следует, что для водорода $Z_i/Z_a = 1/2$ с большой точностью, так как массы ядра и атома водорода отличаются лишь на массу электрона. Поэтому для равновесия ионизации водорода мы получаем из (2.1.39)

$$\frac{x^2}{1-x} = \frac{1}{n} \left(\frac{m_e kT}{2\pi\hbar^2} \right)^{3/2} e^{-E_i/kT}. \quad (2.1.40)$$

Здесь $x = n_e/n = n_i/n$ — степень ионизации, n — теперь означает полную концентрацию атомов, как ионизованных, так и нейтральных. Для равновесия ионизации других газов, например, гелия, мы получили бы формулу того же вида, отличающуюся лишь несколькими иными значениями статистических весов g_i и g .

Если бы ионизация определялась только энергетическим фактором $e^{-E_i/kT}$, то она возникла бы только при $T \approx \approx T_i$. Однако в действительности энтропия газа из электронов и ионов значительно больше энтропии газа из одних только нейтральных атомов, и это делает ионизованное состояние термодинамически более выгодным даже в том случае, когда средняя энергия теплового движения заметно меньше энергии ионизации. Вследствие этого водород может быть практически полностью ионизован уже при $T \gtrsim 10^4$ °К для довольно широкого интервала интересных для дальнейшего значений концентрации газа. Зависимость от концентрации, как видно, гораздо слабее зависимости от температуры. Действительно, при $T = = 10^4$ °К находим, что $n \frac{x^2}{1-x} \approx 10^{14} \text{ см}^{-3}$, т. е. ионизация является полной при всех $n < 10^{14} \text{ см}^{-3}$. При вдвое меньшей температуре, $T = 5 \cdot 10^3$ °К, имеем $n \frac{x^2}{1-x} \approx 10^7$, и ионизация наступает при $n < 10^7 \text{ см}^{-3}$. Но уже при $T = = 3 \cdot 10^3$ °К произведение $n \frac{x^2}{1-x}$ падает до 10^{-3} см^{-3} , так что ионизованным может быть только весьма разреженный газ, $n < 10^{-3} \text{ см}$.

Перейдем теперь от этих общих соображений к ситуации, которая имела место в метагалактической среде в интересующую нас эпоху. Вплоть до этой эпохи взаимодействие между веществом и излучением еще очень эффективно (оценки соответствующих времен даются в следующем параграфе), и между ними имелось полное термодинамическое равновесие. Характерная температура $T = = T_H \approx 10^4$ °К соответствует, по формуле (2.1.15), красному смещению $z = 3 \cdot 10^3$ и, следовательно, концентрации $n = (\rho/m_H) (1+z)^3 \approx 2 \cdot 10^5 \text{ } \Omega \text{ см}^{-3}$. Согласно сделанным выше оценкам можно утверждать, что при $z = 3 \cdot 10^3$ метагалактическая плазма была еще практически полностью ионизована при всех допустимых n . Однако при $T = 3 \cdot 10^3$ °К, чему отвечает $z \approx 10^3$, $n = 3 \cdot 10^3 \text{ } \Omega \text{ см}^{-3}$, степень ионизации падает до значения

$$x = \frac{1}{2\sqrt{\Omega}} - \frac{1}{8\Omega}. \quad (2.1.41)$$

(Это приближенное выражение оказывается справедливым

при $\Omega > 1/16$.) Для значения $\Omega = 1/3$ степень ионизации по (2.1.41) составляет $1/2$.

Мы оценили степень ионизации водорода. Что касается гелия, то для него энергия первой ионизации есть 24,6 эв и степень однократной ионизации равна половине уже при температуре 7000 °К (двукратно ионизованные атомы при этой температуре уже отсутствуют).

Итак, переломная эпоха, о которой здесь идет речь, имеет место при красном смещении порядка тысячи: $z_{\text{rec}} \approx \approx 1000$. Интересно, что эпоха рекомбинации оказывается довольно близкой к другой важной вехе в истории Вселенной — моменту перехода от стадии преобладания излучения к стадии преобладания вещества: $z_* \approx 10^4 \Omega$. Величины z_{rec} и z_* точно совпадают при $\Omega = 0,1$. В шкале космологического возраста красному смещению z_{rec} отвечает время

$$t_{\text{rec}} \approx 10^{13} \Omega^{-1/2} \text{ сек.} \quad (2.1.42)$$

При температуре ниже $3 \cdot 10^3$ °К статистическое равновесие вещества и излучения уже не может поддерживаться. Поэтому формула Саха оказывается неприменимой и приходится рассматривать неравновесную кинетику, приводящую к некоторым важным следствиям. Прежде всего, степень ионизации плазмы при уменьшении температуры не стремится к нулю, как это следует из формулы Саха, но уменьшается до конечного значения $\sim 10^{-4} - 10^{-5}$. Это явление называют закалкой ионизации. Оно имеет место вследствие космологического расширения, уменьшающего частоту столкновений электронов с ионами, а следовательно, и скорость рекомбинации. Другое обстоятельство, также связанное с неравновесностью и увеличивающее остаточную ионизацию водорода (Зельдович, Курт, Сюняев (1968), Пиблс (1968)), состоит в следующем. Возбуждение атома водорода из основного ($1s$) состояния на первый возбужденный ($2s$) уровень требует энергии, равной $3/4$ энергии ионизации. Наоборот, при обратном переходе излучается квант (называемый квантом Лайман- α), имеющий такую же энергию. Переходы между всеми уровнями, начиная с ($2s$) до самых высоких, требуют энергии, значительно меньшей этой. Поэтому энергичный квант может излучаться и поглощаться лишь при переходах ($1s$) \rightleftharpoons ($2s$). Можно сказать, что такие кванты не

могут расходоваться ни на что, кроме этого перехода. По мере рекомбинации водорода их число все больше превышает равновесное. Как только такой энергичный квант возбудит атом водорода на уровень ($2s$), другие, более «мелкие» кванты (которые имеются в большем количестве) уже сравнительно легко смогут ионизовать этот атом. «Накопление» этих квантов препятствует дальнейшему падению ионизации. Остаточная ионизация была бы еще большей, если не существовал бы некоторый процесс, который, хотя и является медленным, уменьшает число лаймановских квантов. Этот процесс состоит в переходе ($2s$) \rightarrow ($1s$) с испусканием не лаймановских квантов, а двух квантов, сумма энергии которых равна энергии перехода. Подробный анализ (Сюняев и Зельдович (1970), Лонгейр и Сюняев (1969, 1971)) показывает, что указанный выше закаленный уровень ионизации устанавливается при $z \approx 600$, $T \approx 1600$ °К, а степень ионизации падает (от $z = 1400$ до $z = 900$) по закону

$$x(z) = 6 \cdot 10^6 \Omega^{-1/2} z^{-1} \exp(-1,5 \cdot 10^4 z^{-1}). \quad (2.1.43)$$

Если до рекомбинации вещество и излучение были прочно сцеплены между собой электромагнитным взаимодействием (см. следующий параграф) и представляли собой единую «жидкость», то вскоре после эпохи $z = z_{\text{rec}}$ взаимодействие этих двух компонент среды прекращается, нарушается термодинамическое равновесие между ними, и излучение, «оторвавшись» от вещества, свободно распространяется в расширяющейся Вселенной. В соответствии с общими закономерностями, которые мы обсуждали в § 1.4, его температура и плотность энергии падают, приближаясь со временем к современным, измеренным в наблюдениях значениям.

Взаимодействие вещества и излучения может, однако, на какое-то время восстановиться, если внутренние процессы в веществе (например, возникновение в нем ударных волн или вспышки излучения формирующихся космических объектов) вновь разогреют и ионизуют его (см. §§ 2.2, 2.3 и гл. 5). Из-за быстрого теплоотвода непрерывное поддержание состояния ионизации вряд ли возможно — это требовало бы слишком большого энерговыделения (Сюняев (1968)).

§ 2.2. Взаимодействие излучения и метagalактической плазмы

В эпоху, предшествующую рекомбинации, метagalактическая среда была полностью ионизована и после аннигиляции позитронов и нуклеосинтеза состояла из трех взаимодействующих друг с другом компонент: излучения, электронов и ионов. Взаимодействие этих компонент обеспечивало термодинамическое равновесие среды как целого при космологическом расширении.

Мы рассмотрим сейчас процессы взаимодействия и сделаем оценки соответствующих характерных времен, считая температуру нерелятивистской.

Томсоновское рассеяние в плазме. Этот процесс представляет собой рассеяние фотонов на свободных электронах. Импульс рассеянного фотона отличается при этом от начального только направлением, а энергия не меняется, если пренебрегать энергией теплового фотона по сравнению с энергией покоя электрона при указанных температурах. Сечение этого процесса не зависит от частоты фотона и определяется по порядку величины квадратом электромагнитного радиуса электрона

$$r_e = \frac{e^2}{m_e c^2} \quad (2.2.1)$$

(m_e , e — масса и заряд электрона).

Точное значение сечения, которое называется томсоновским, есть

$$\sigma_T = \frac{8\pi}{3} r_e^2 \approx 0,665 \cdot 10^{-24} \text{ см}^2. \quad (2.2.2)$$

Длина свободного пробега фотона в плазме по отношению к рассеянию на электронах

$$l_{re} = \frac{1}{n\sigma_T}, \quad (2.2.3)$$

где $n = \rho/m$ — концентрация электронов, m — масса атома водорода.

Время свободного пробега фотона

$$\tau_{re} = l_{re}/c \approx 2 \cdot 10^{19} \Omega^{-1} (1+z)^{-3} \text{ сек.} \quad (2.2.4)$$

Сечение рассеяния фотонов на ионах плазмы меньше сечения (2.2.2) в отношении $(m_e/m_i)^2 \approx 3 \cdot 10^{-7}$, где

$m_i \approx m$ — масса иона (мы будем для простоты оценок считать ионы протонами, не учитывая наличие ядер гелия).

Торможение электрона равновесным излучением. В этом процессе (и в некоторых других, рассматриваемых в дальнейшем) существен обмен энергией при взаимодействии плазмы и излучения. При этом уже нельзя пренебрегать энергией фотонов (например, тепловых) по сравнению с энергией покоя электрона. Рассеяние с заметным изменением энергии называют комptonовским.

Если газ фотонов находится в равновесном состоянии и потому распределение фотонов в некоторой системе отсчета изотропно, то электрон, движущийся относительно этого распределения, испытывает тормозящее его трение. Качественно это трение можно понять следующим образом. Фотоны с импульсом $\hbar k_r$, налетающие на электрон спереди, имеют в его системе отсчета, вследствие доплеровского смещения, импульс $\hbar k_r \left(1 + \frac{v}{c} \cos \theta\right)$, где v — скорость электрона, θ — угол между ней и волновым вектором фотона k_r . Такие же фотоны, налетающие сзади, имеют меньший импульс, так как для них $\cos \theta < 0$. Фотоны, налетающие на электрон спереди, рассеиваясь на нем упруго, отбирают у него импульс, примерно равный $\hbar k_r (1 + v/c)$, и, таким образом, тормозят его. Фотоны же, налетающие сзади, хотя и передают электрону импульс, ускоряющий его, но этот импульс меньше импульса, отбираемого у электрона, так как он примерно равен $\hbar k_r (1 - v/c)$. Поэтому в среднем электрон при столкновениях с фотонами теряет импульс $\hbar k_r v/c$.

Число столкновений электрона с фотонами в единицу времени есть

$$\nu_{er} = n_r c \sigma_T, \quad (2.2.5)$$

где n_r — концентрация фотонов. (Мы пишем томсоновское сечение, пренебрегая здесь поправками порядка $\hbar \omega / m_e c^2$.)

Так как в каждом столкновении электрон в среднем теряет импульс, примерно равный $\hbar k_r m_e v / m_e c$, то изменение импульса электрона $m_e v$ в единицу времени при взаимодействии с изотропным излучением описывается уравнением

$$\frac{dm_e v}{dt} = - \frac{m_e v}{\tau_{er}} \quad (2.2.6)$$

с характерным временем комптоновского процесса, которое, согласно (2.2.5) и выражению для изменения импульса в одном столкновении, имеет вид

$$\tau_{er} \approx \frac{m_e}{\hbar k_r \sigma_T n_r} \approx \frac{m_e c}{\epsilon_r \sigma_T}.$$

Мы заменили здесь $\hbar c k_r = \hbar \omega$ на kT , имея в виду, что главную роль в торможении электрона играют тепловые фотоны. Точный вывод (Ландау и Лифшиц (1973)) дает

$$\tau_{er} = \frac{4}{3} \frac{m_e c}{\epsilon_r \sigma_T}.$$

Подставляя сюда плотность энергии излучения ϵ_r как функцию z , находим

$$\tau_{er} = 9 \cdot 10^{19} (1 + z)^{-4} \text{ сек.} \quad (2.2.7)$$

Решение уравнения (2.2.6) (в котором удобно разделить обе части на m_e , чтобы иметь дело не с импульсом электрона, а с его скоростью) есть

$$v(\Delta t) = v_1 e^{-\Delta t / \tau_{er}},$$

где Δt — время, протекшее от момента, когда скорость электрона имела значение v_1 . При этом считается, что за время Δt плотность энергии излучения не изменилась сколько-нибудь заметно. Путь, который пройдет электрон до полной остановки, есть

$$l_{er} = \int_{\Delta t=0}^{\infty} v(\Delta t) dt = v \tau_{er}. \quad (2.2.8)$$

Хотя интегрирование ведется формально до $\Delta t = \infty$, наличие в подынтегральной функции экспоненты привело, как мы видим, к тому, что фактически в нем играет роль лишь временной интервал $\sim \tau_{er}$, который также может быть мал по сравнению с характерным временем изменения плотности энергии излучения. Длина l_{er} есть длина свободного пробега электрона для торможения равновесным излучением.

Трение плазмы об излучение. Рассмотрим теперь движение относительно фона излучения макроскопического объема плазмы, состоящего из многих электронов и

ионов. Вследствие кулоновского взаимодействия разделение положительных и отрицательных зарядов на большие расстояния невозможно. При этом возникают сильные электрические поля, стремящиеся сблизить противоположные заряды. Действительно, если в каком-нибудь месте плазмы возникает (по каким-либо причинам, например, вследствие теплового движения) разность концентрации зарядов n , то она создает разность потенциалов, определяемую уравнением Пуассона:

$$\Delta\varphi = 4\pi en.$$

Если линейные размеры такой области возмущения $\sim r$, то разность потенциалов $\delta\varphi \sim 4\pi e\delta n r^2$, а потенциальная энергия частицы с зарядом e , связанная с этой разностью потенциалов, т. е. с силой, стремящейся сблизить противоположные заряды, $e\delta\varphi \sim 4\pi e^2\delta n r^2$. Очевидно, что эта энергия не может быть больше энергии теплового движения частицы kT . Отсюда следует, что относительная флуктуация $\frac{\delta n}{n} \lesssim \frac{kT}{4\pi e^2 r^2 n} = \left(\frac{r_D}{r}\right)^2$, где

$$r_D = \left(\frac{kT}{4\pi n e^2}\right)^{1/2} \quad (2.2.9)$$

— так называемый дебаевский радиус. Флуктуации заряда, связанные с тепловым движением частиц плазмы, очень малы, если рассматривать области с размерами, много большими дебаевского радиуса. В этом смысле плазму можно считать нейтральной, или, как говорят, квазинейтральной.

Электроны и ионы плазмы, движущиеся относительно фона излучения, взаимодействуют с ним. Сечение взаимодействия для ионов при упругом рассеянии на фотонах ничтожно мало по сравнению с сечением для электронов, как отмечалось выше, так что непосредственным взаимодействием ионов с излучением можно в этом случае пренебречь. Тем не менее их наличие существенно: когда электроны теряют свой импульс, отдавая его фону излучения, то одновременно в силу условия квазинейтральности должны терять импульс и ионы. Это значит, что через электроны излучение отбирает импульс у плазмы в целом. Для плазмы формула (2.2.6), описывающая потерю

импульса, должна быть поэтому изменена:

$$\frac{dP}{dt} = -\frac{P}{\tau_f}, \quad \tau_f = \frac{4/3 mc}{e_r \sigma_T} \approx 2 \cdot 10^{23} (1+z)^{-4} \text{ сек}, \quad (2.2.10)$$

где P — импульс некоторого движущегося объема плазмы, m — суммарная масса электрона и иона, связанных кулоновским взаимодействием. Процесс представляет собой трение плазмы об излучение, характеризуемое временем τ_f . Это время больше, чем τ_{er} в отношении масс атома и электрона. В соответствии с этим увеличивается и длина свободного пробега $l_f = v\tau_f$.

Взаимодействие плазмы и излучения, имеющих разные температуры. При вторичном разогреве и ионизации плазмы в послерекомбинационную эпоху (о чем мы упоминали в конце предыдущего параграфа) температура плазмы выше температуры излучения. В этом случае в среде происходили три процесса; во-первых, электроны отдавали энергию фотонам; во-вторых, они получали ее от ионов и, наконец, в-третьих, ионы могли непосредственно передавать энергию фотонам. В разные эпохи доминировали разные процессы, и состояние среды зависело от соотношения их скоростей.

Характерное время передачи энергии от электронов излучению в этом случае, очевидно, дается формулой (2.2.7). Время же, за которое электрон получает от иона энергию порядка kT , где T — температура плазмы, превосходящая температуру излучения T_r ,

$$\tau_{ei} = \frac{1}{\sigma_0 n_i u_e} \frac{m}{m_e}, \quad u_e \approx \left(\frac{kT}{m_e} \right)^{1/2}. \quad (2.2.11)$$

Здесь вновь появляется отношение масс сталкивающихся частиц: ион отдает электрону энергию $\sim kT$ не за одно столкновение, а за m/m_e столкновений. Величина σ_0 есть сечение кулоновского рассеяния, которое можно оценить следующим образом. Электрон с тепловой скоростью u_e испытывает существенное взаимодействие с ионом на таком расстоянии, на котором энергия электростатического кулоновского взаимодействия сравнима с его средней энергией, т. е. с температурой: $e^2/r_0 \approx kT$, откуда $r_0 \approx e^2/kT$. Это расстояние r_0 называется кулоновским радиусом; соответствующее ему кулоновское сечение $\sigma_0 \approx r_0^2$. Но так как кулоновское взаимодействие спадает мед-

ленно с расстоянием, в рассеяние вносят вклад и более далекие сближения. Теория показывает, что этот вклад может быть учтен некоторым множителем Λ , который слабо (логарифмически) зависит от температуры и концентрации плазмы и называется кулоновским логарифмом:

$$\sigma_0^2 = r_0^2 \Lambda, \quad \Lambda = 2\pi \ln \left(\frac{r_D}{r_0} \right),$$

где r_D — дебаевский радиус, т. е., как было объяснено выше, расстояние, на котором кулоновское взаимодействие частиц плазмы становится слабым и плазму можно считать квазинейтральной. Подставляя выражение для r_D , получим

$$\Lambda = \pi \ln \frac{1}{2\pi n r_0^3}. \quad (2.2.12)$$

Ионы передают электронам энергию, очевидно, также за время (2.2.11). Характерное время передачи энергии от ионов излучению другое:

$$\tau_{ir} = \frac{1}{n_r \sigma_{Tc}^i} \left(\frac{mc^2}{kT_r} \right). \quad (2.2.13)$$

Здесь томсоновское сечение рассеяния фотонов на ионах

$$\sigma_T^i = \frac{8\pi}{3} \left(\frac{e^2}{mc^2} \right)^2 = \sigma_T \left(\frac{m_e}{m} \right)^2. \quad (2.2.14)$$

Отношение времен τ_{er} и τ_{ei} есть

$$\frac{\tau_{er}}{\tau_{ei}} = \frac{n}{n_r} \frac{u_e}{c} \frac{m_e c^2}{kT_r} \frac{\sigma_0}{\sigma_T}. \quad (2.2.15)$$

Подставляя значения входящих сюда величин, получим:

$$\frac{\tau_{er}}{\tau_{ei}} \approx 10^{-8} \frac{\Omega}{\Lambda} \left(\frac{kT_r}{m_e c^2} \right)^{-5/2} \left(\frac{T_r}{T} \right)^{3/2}. \quad (2.2.16)$$

Если температуры вещества и излучения не слишком сильно различаются, то это отношение меньше единицы при температурах, превосходящих миллион градусов, или при красном смещении, большем, чем $3 \cdot 10^5$ (кулоновский логарифм $\Lambda \sim 10-30$). При этих условиях электроны

быстрее приходят в температурное равновесие с излучением, чем с ионами. В более позднюю эпоху (но еще до рекомбинации) соотношение времен обратное, и электроны быстрее приходят в равновесие с ионами, чем с излучением. Существенное различие температур плазмы и излучения (а именно — горячая плазма на фоне холодного излучения) могло быть возможным после рекомбинации. Отношение (2.2.16) может вновь оказаться меньшим единицы, но только при весьма высокой температуре плазмы:

$$\frac{T}{T_{\text{rec}}} > 3 \cdot 10^4 \left(\frac{1 + z_{\text{rec}}}{1 + z} \right)^{2/3}.$$

Сравним теперь времена τ_{ir} и τ_{ei} :

$$\frac{\tau_{ir}}{\tau_{ei}} \approx \left(\frac{m}{m_e} \right)^3 \frac{\tau_{er}}{\tau_{ei}} \approx 10^{14} \left(\frac{T_{\text{rec}}}{T} \right)^{3/2} \frac{T_{\text{rec}}}{T_r}.$$

Эта оценка показывает, что практически для всей области нерелятивистских температур ионы передавали энергию электронам быстрее, чем излучению, так что излучение получает энергию от плазмы при этих условиях только через посредство электронов. Скорость передачи энергии различна в зависимости от того, находятся ли электроны в температурном равновесии с излучением или с ионами. В первом случае, который, как мы видели, осуществляется при красных смещениях, превышающих $3 \cdot 10^5$, излучение вместе с электронами получает энергию от ионов (когда они более горячие); скорость передачи энергии, отнесенная к единице объема, есть

$$q_{ie} = \frac{3}{2} \frac{nk(T - T_r)}{\tau_{ie}}. \quad (2.2.17)$$

При меньших красных смещениях, которые и будут интересовать нас в дальнейшем, излучение получает энергию от всей плазмы, и скорость теплоотвода есть

$$q_{er} = \frac{3nk(T - T_r)}{\tau_{er}}. \quad (2.2.18)$$

Тормозные процессы в метagalacticкой плазме. При кулоновском рассеянии электрона на ионе электрон испытывает ускорение (или торможение) и потому излучает

электромагнитные волны. Процессы, при которых и до, и после столкновения электрон находится в свободном состоянии, представляют собой свободно-свободные переходы.

Если же после столкновения свободный электрон оказывается захваченным ионом, то говорят о свободно-связанных переходах. Аналогично можно говорить о связанно-свободных переходах. Потенциальная энергия e^2/r взаимодействия электрона и иона оказывается равной его кинетической энергии $1/2 m_e v^2$ на расстоянии $r_0 = e^2 / 1/2 m_e v^2$, которое, как уже говорилось, называется кулоновским радиусом, соответствующим заданной энергии. Для электронов средней тепловой энергии $v \approx u_e = (kT/m_e)^{1/2}$ кулоновский радиус $r_0 \approx e^2/kT$.

Это понятие, однако, имеет смысл лишь в том случае, если квантовая неопределенность положения электрона, равная длине волны де-Бройля $\hbar/m_e v$, меньше кулоновского радиуса, т. е. если энергия электрона меньше $m_e e^4/\hbar^2$. Но последнее представляет собою характерную квантовую энергию, определяющую энергию ионизации и энергии возбуждения атома водорода. Энергия ионизации водорода $E_i = 1/2 m_e e^4/\hbar^2$, а энергия перехода из основного в первое возбужденное состояние (квант Лайман- α) есть $3/4 E_i$. Следовательно, кулоновский радиус больше, чем дебройлевская длина, лишь при таких энергиях электрона, которые меньше этой характерной энергии. В тех случаях, когда нас интересует основная масса электронов в тепловом максвелловском распределении, т. е. электроны с энергией $\sim kT$, указанное неравенство означает, что $kT < E_i$. Но в этих условиях *) основную роль играют не свободно-свободные переходы, а свободно-связанные переходы, при которых электрон, испуская тормозной фотон, захватывается ионом на один из дискретных уровней. Поэтому свободно-свободные переходы будут интересовать нас лишь при температурах, превосходящих энергию ионизации; при этом расстояние, на котором электрон пролетает мимо иона, может быть определено с погрешностью, много меньшей этого рассеяния. Если учесть это квантовое ограничение, связанное с прин-

*) Предполагается также, что нет внешних источников излучения, ионизирующего вещество.

ципом неопределенности, то оказывается возможным классическое рассмотрение свободно-свободного излучения.

Согласно классической, электродинамике электрон, движущийся с ускорением \dot{v} , излучает каждую секунду электромагнитную энергию $\frac{2}{3} \frac{e^2}{c^3} (\dot{v})^2$. Здесь $\dot{v} = e^2/m_e r^2$ есть абсолютное значение ускорения электрона на расстоянии r от иона водорода. Оно довольно быстро убывает при удалении от иона; излучение максимально на минимальном прицельном расстоянии от иона и уже через время $\sim r/v$ уменьшается вчетверо. Поэтому можно считать, что основной вклад в излучаемую энергию создается за промежуток времени $\sim r/v$. Тогда полная энергия, излученная в единичном акте столкновения, есть приблизительно

$$\frac{e^2}{c^3} \dot{v}^2 \frac{r}{v} \approx \frac{e^2}{r} \left(\frac{r_e}{r} \right)^2 \frac{c}{v}.$$

В единицу времени электрон испытывает, пролетая мимо ионов в интервале прицельных расстояний от r до $r + dr$, число столкновений $2\pi n v r dr$, где n — концентрация ионов, которая равна концентрации электронов. Умножая это число на энергию, излучаемую при каждом столкновении, и на концентрацию n электронов и интегрируя по прицельному расстоянию от r_{\min} до r_{\max} , мы получим энергию, излучаемую каждую секунду в среднем из единицы объема при свободно-свободных переходах:

$$q_{\text{н}} = -n \frac{d\bar{E}}{dt} \approx \pi r n^2 r_e^2 c e^2 \left(\frac{1}{r_{\min}} - \frac{1}{r_{\max}} \right). \quad (2.2.19)$$

Здесь \bar{E} — средняя энергия электрона (которая для тепловых электронов порядка kT). Максимальным расстоянием r_{\max} является, очевидно, дебаевский радиус, или радиус экранирования (2.2.9), так как на расстояниях, превышающих этот радиус, кулоновское поле рассеивающего заряда убывает экспоненциально, экранируясь окружающими зарядами противоположного знака. Что касается минимального расстояния r_{\min} , то согласно сказанному выше, его следует принять равным длине волны де-Бройля. Таким образом, интенсивность излучения при свободно-свободных переходах зависит от соотношения между дебаевским радиусом и дебройлевской длиной волны. Для электронов тепловой энергии отношение второй

к первому, $\left(\frac{4\pi e^2 \hbar^2}{m_e (kT)^2}\right)^{1/2}$, много меньше единицы, если

$$n < \frac{1}{16\pi} \left(\frac{m_e e^2}{\hbar^2}\right)^3 \left(\frac{kT}{E_i}\right)^2 \approx 3 \cdot 10^{23} \left(\frac{kT}{E_i}\right)^2 \text{ см}^{-3}.$$

Это условие заведомо выполняется для нерелятивистских ($kT \ll m_e c^2$) температур (но бóльших температуры ионизации). Тогда в формуле (2.2.19) можно пренебречь вторым членом в скобках и она принимает вид

$$q_{ff} \approx 2\pi n \cdot n r_e^2 c \sqrt{E_i kT}.$$

Точная теория дает еще численный множитель ≈ 2 , так что

$$q_{ff} = 1,4 \cdot 10^{-27} n^2 (T \text{ }^\circ\text{K})^{1/2} \text{ эрг/см}^3 \cdot \text{сек.} \quad (2.2.20)$$

Эта формула относится к чисто водородной плазме. Реальная метagalacticкая плазма содержит еще примесь гелия, относительная концентрация которой $\text{He}/\text{H} \sim 0,1$ (см. § 2.1). Так как тормозное излучение пропорционально квадрату заряда иона, для такой смеси

$$q_{ff}(\text{H} + \text{He}) = \left(\frac{n_{\text{H}}}{n} \cdot \frac{4n_{\text{He}}}{n}\right) q_{ff}(\text{H}) \approx \frac{7}{6} q_{ff}(\text{H}),$$

$$n = n_{\text{H}} + 2n_{\text{He}},$$

причем в выражении для q_{ff} под n следует понимать полную концентрацию электронов и считать, что гелий ионизован двукратно.

Характерное время потери энергии вследствие свободно-свободных переходов есть

$$\tau_{ff} = \frac{3nkT}{q_{ff}} \approx \tau_{re} \left(\frac{kT}{E_i}\right)^{1/2} \approx 3,6 \cdot 10^{11} \frac{(T \text{ }^\circ\text{K})^{1/2}}{n} \text{ сек.} \quad (2.2.21)$$

Если конечное состояние электрона связанное, т. е. он оказывается захваченным ионом, то излучение электрона возрастает, так как в излучение переходит не только кинетическая энергия электрона, но и энергия его связи в атоме. Процессы такого рода, называемые свободно-связанными переходами, играют роль при температурах, при которых кинетическая энергия электрона сравнима с энергией ионизации атома. Для гелия это температура около 10^6 °К, а для водорода — раза в два-три меньше. В интервала температур от 10^6 до 10^4 °К существенны также свя-

зависимости от температуры. Вычисления в этом случае весьма громоздки (Вейман (1967), Дорошкевич и Сюняев (1969), Силк и Тартер (1973)).

Условия эффективности. Рассмотренные процессы могли эффективно осуществляться в расширяющейся Вселенной в ту или иную эпоху лишь в том случае, если их характерные времена меньше текущего космологического возраста. На ранних стадиях расширения, когда концентрации плазмы и фотонов были велики, эти времена удовлетворяли указанному неравенству. Однако в процессе расширения концентрации уменьшались настолько быстро, что рано или поздно такие неравенства должны были перестать выполняться.

Комптовское взаимодействие вещества и излучения сопровождается передачей импульса и энергии. Передача энергии и импульса макроскопического движения плазмы излучению характеризуется временем (2.2.10), а для передачи тепловой энергии характерно время (2.2.7); их отношение

$$\frac{\tau_f}{\tau_{er}} = \frac{m}{m_e},$$

и потому обмен импульсом должен перестать быть эффективным, вообще говоря, раньше, чем обмен тепловой энергией.

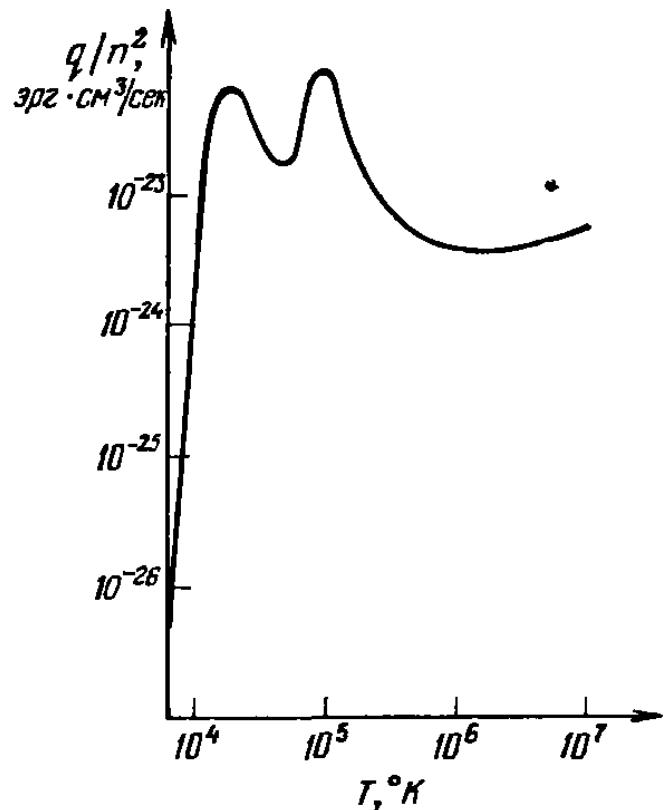


Рис. 7. Скорость теплоотвода в плазме, содержащей водород (70% по весу) и гелий (Вейман (1967)).

Равенство времени (2.2.10) и космологического возраста наступило, как мы увидим, позже эпохи преобладания излучения, но до перехода на инерциальную стадию расширения ($z > z_\Omega$), если она вообще осуществляется. Поэтому для этой эпохи можно написать

$$t \approx 3 \cdot 10^{17} \Omega^{-1/2} (1+z)^{-3/2} \text{ сек}, \quad (2.2.22)$$

$$\epsilon_r = \rho_* c^2 (t_*/t)^{3/2}, \quad (2.2.23)$$

и тогда

$$\frac{\tau_f}{t} = \frac{m}{\sigma_T \rho_* c t_*} \left(\frac{t}{t_*} \right)^{3/2}. \quad (2.2.24)$$

Это отношение равно единице в момент, когда

$$\left. \begin{aligned} t = t_f &\equiv t_* \left(\sigma_T \rho_* c t_* / m \right)^{2/3} \approx 6 \cdot 10^{13} \Omega^{-4/5} \text{ сек}, \\ z = z_f &\equiv z(t_f) \approx 200 \Omega^{1/5}, \\ \rho_f &\equiv \rho(t_f) \approx 4 \cdot 10^{-23} \Omega^{3/5} \text{ г/см}^3. \end{aligned} \right\} (2.2.25)$$

При всех вероятных значениях космологического параметра момент (2.2.25) действительно наступает в эпоху преобладания вещества и до перехода к инерциальной стадии

$$t_* < t_f < t_0 \Omega,$$

$$z_* < z_f < \Omega^{-1}.$$

Равенство времени (2.2.7) и космологического возраста наступает в момент, когда

$$\left. \begin{aligned} t = t_{er} &= \left(\frac{m}{m_e} \right)^{3/5} t_f \approx 100 t_f \approx 6 \cdot 10^{15} \Omega^{-4/5} \text{ сек}, \\ z = z_{er} &\equiv z(t_{er}) \approx 4 \cdot 10^{-2} z_f \approx 8 \Omega^{-4/5}, \\ \rho_{er} &\equiv \rho(t_{er}) \approx 2 \cdot 10^{-27} \Omega^{-7/5} \text{ г/см}^3. \end{aligned} \right\} (2.2.26)$$

Это время больше, чем t_* , но при $\Omega < 0,1$ оно может оказаться относящимся к инерциальной стадии. В этом случае $t = t_0 (1+z)^{-1}$, $t < t_\Omega$, и тогда

$$t_{er} = t_0 \left(\frac{m_e}{\sigma_T \rho_r(t_0) c t_0} \right)^{-1/3} \approx 0,1 t_0, \quad \Omega < 0,1, \quad z_{er} \approx 10. \quad (2.2.27)$$

Здесь $t_0 = 6 \cdot 10^{17} \left(\frac{H}{50 \text{ км/сек} \cdot \text{Мпс}} \right)^{-1}$ сек — современный возраст мира.

Время свободного пробега фотона по отношению к томсоновскому рассеянию (2.2.4),

$$\tau_{re} = \tau_f \frac{\rho_r}{\rho} \approx 10^{-4} \tau_f \Omega^{-1} (1+z), \quad (2.2.28)$$

меньше космологического возраста при

$$\left. \begin{aligned} t < t_{re} \approx 4 \cdot 10^{14} \text{ сек}, \\ z > z_{re} \equiv z(t_{re}) \approx 15 \Omega^{-1/3}, \\ \rho > \rho_{re} \equiv \rho(t_{re}) \approx 10^{-25} \text{ г/см}^3. \end{aligned} \right\} \quad (2.2.29)$$

Мы видим, что обмен энергией между веществом и реликтовым излучением посредством комптоновских процессов прекратился задолго до нашей эпохи, при возрасте

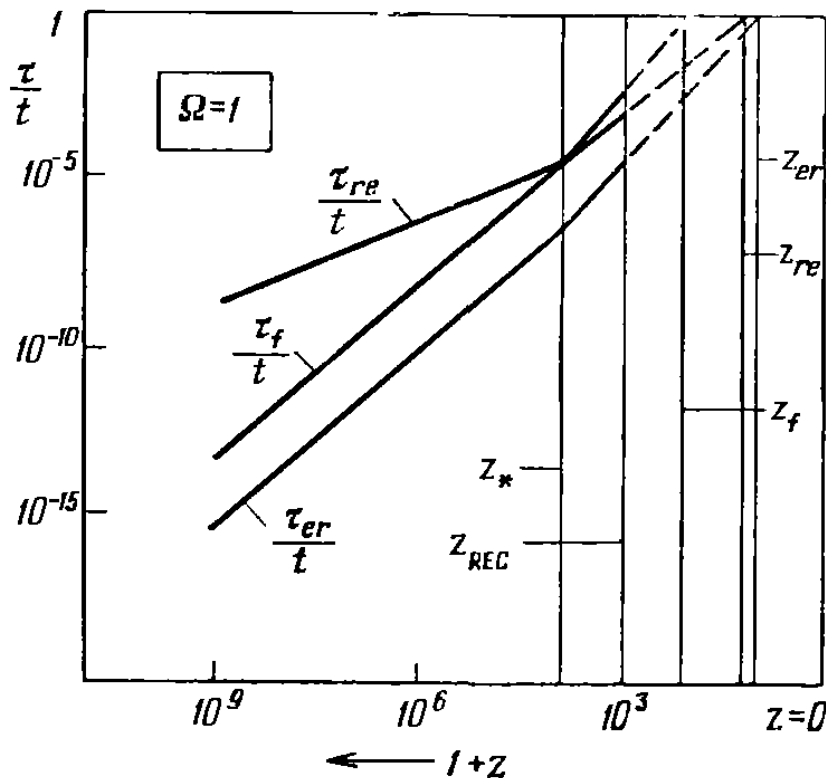


Рис. 8. Времена взаимодействия излучения и плазмы в горячей Вселенной.

Вселенной по крайней мере на порядок меньшем современного, а обмен импульсом прекратился еще ранее (рис. 8). Разумеется, выражения (2.2.21) — (2.2.26) имеют смысл лишь в том случае, если после эпохи рекомбинации газ действительно мог вновь оказаться ионизованным. Во

все более ранние времена до эпохи рекомбинации взаимодействие между веществом и излучением происходит весьма эффективно, с характерными временами, много меньшими текущего возраста мира. Оно не обрывается и сразу после эпохи рекомбинации, а продолжается даже без дополнительного нагрева и ионизации вещества, за счет малой остаточной ионизации, о которой говорилось в конце предыдущего параграфа. Даже при концентрации электронов $\sim (10^{-4} - 10^{-5})\rho/m$ комптоновский теплообмен (и столкновения электронов с ионами, а ионов с атомами) обеспечивает равенство температур вещества и излучения до красных смещений ≈ 150 (Сюннев и Зельдович (1970)). После этого вещество охлаждается из-за космологического расширения, и его температура падает в соответствии с показателем адиабаты $\gamma = 5/3$:

$$T \propto (1 + z)^2. \quad (2.2.30)$$

Вещество становится в эту эпоху холоднее излучения,

$$\frac{T}{T_r} \propto (1 + z), \quad z < 150, \quad (2.2.31)$$

и всякое взаимодействие между веществом и реликтовым фоном прекращается вплоть до вторичного разогрева и ионизации, если таковые происходят до момента (2.2.26).

Что же касается тормозных процессов, то их эффективность определяется не только универсальными космологическими характеристиками среды, но также и температурой плазмы в условиях вторичного разогрева, когда она гораздо выше температуры реликтового излучения. Отношение времени (2.2.21) к космологическому возрасту (в эпоху преобладания вещества, но до перехода на инерциальный режим расширения)

$$\frac{\tau_{ff}}{t} \approx \frac{1}{2} \Omega^{-2} (T \text{ } ^\circ\text{K})^{1/2} (1 + z)^{-3/2}, \quad (2.2.32)$$

больше единицы, так как соотношение (2.2.21) предполагает, что температура газа выше температуры ионизации.

В эпоху до рекомбинации, когда температуры излучения и вещества должны быть равны, отношение времени (2.2.21) к космологическому возрасту всегда меньше единицы, но его отношение ко времени комптоновского процесса (2.2.7) больше единицы и в эту эпоху и после нее.

Излучение при свободно-связанных переходах и в линиях при температурах среды $< 5 \cdot 10^5$ °К гораздо эффективнее излучения при свободно-свободных переходах; поэтому соответствующие им времена могут быть меньше космологического возраста в эпоху вторичного разогрева (см. гл. 5, 6).

Заметим, что сделанные здесь оценки относятся к однородной метagalактической плазме. При наличии существенных неоднородностей времена (2.2.4) и (2.2.21) изменятся и потому изменятся их отношения к другим характерным временам.

§ 2.3. Влияние горячей плазмы на спектр реликтового излучения

Теплоемкость единицы объема излучения есть

$$c_r = \frac{d\epsilon_r}{dT} \approx 10 kn_r.$$

Это теплоемкость при постоянном объеме; для излучения понятия о теплоемкости при постоянном давлении не имеет смысла, так как давление зависит только от температуры. С другой стороны, для плазмы (которую мы можем считать в этих оценках состоящей из полностью ионизованного водорода) теплоемкость при постоянном объеме $c_V = 3kn$, где $n = n_e = n_i$. Отношение теплоемкостей

$$\frac{c_r}{c_V} \approx 3 \frac{n_r}{n} \approx 3 \cdot 10^8 \Omega^{-1}$$

(см. соотношение (2.1.33)). Из-за столь большой разницы теплоемкостей излучение, получая энергию от плазмы, нагревается крайне незначительно. Если газ, нагретый до высокой температуры T , передает излучению всю свою энергию, то должно иметь место соотношение

$$nT \approx n_r \Delta T_r.$$

Следовательно, отношение изменения температуры излучения ΔT_r к температуре плазмы T есть

$$\frac{\Delta T_r}{T} \approx \frac{n}{n_r} \ll 1.$$

Даже в том случае, если после рекомбинации происходил вторичный разогрев газа и его температура достигала десятков миллионов градусов, температура излучения могла измениться менее, чем на градус, что гораздо меньше температуры реликтового излучения в эпоху, когда комптоновский процесс мог быть существенным.

Мы видели, что ведущим процессом взаимодействия между веществом и излучением в горячей Вселенной является комптоновское рассеяние. В этом процессе при передаче энергии фотонам в условиях, когда их температура ниже температуры плазмы, число фотонов не меняется, а энергия каждого фотона возрастает. Очевидно, что это несовместимо с распределением Планка, при котором, как мы знаем, плотность энергии излучения и концентрация фотонов однозначно связаны друг с другом и с температурой. Значит, распределение фотонов в присутствии горячей плазмы не может оставаться равновесным (Компанеец (1956), Гуревич и Владимиров (1964), Вейман (1965, 1966)).

Чтобы выяснить, как проявляются при этом отклонения от распределения Планка, полезно вспомнить сначала вывод (по Эйнштейну) равновесного распределения из принципа детального равновесия. Если между веществом и излучением существует взаимодействие, при котором фотоны излучаются и поглощаются, так что их число не постоянно, и излучающие атомы имеют равновесное максвелловское распределение, то равновесное распределение излучения устанавливается посредством трех процессов. Первый из них — спонтанное излучение атомов, находящихся в возбужденном состоянии. Если атомы имеют энергию E_{p+q} , где q — импульс излучаемого фотона, а p — импульс атома после излучения, то число актов спонтанного излучения фотонов с импульсом в интервале от q до $q + dq$ в единице объема в единицу времени может быть записано в виде $A_{p+q, p} n_{p+q} (dq)$. Здесь $A_{p+q, p}$ — коэффициент спонтанного излучения, n_{p+q} — число атомов в единице объема с энергией E_{p+q} , $(dq) = d^3q / (2\pi\hbar)^3$ — число квантовых состояний фотонов в интервале импульсов d^3q и в единице объема. Для краткости мы не указываем поляризацию испускаемого фотона; при последующих интегрированиях по импульсам фотонов будет всегда подразумеваться суммирование по их поляризациям.

Второй процесс — вынужденное излучение. Если в пространстве, окружающем атомы, уже имеются фотоны с импульсом q , которые могут сталкиваться с данным атомом, то эти столкновения вызывают усиление излучения квантов с тем же импульсом и той же поляризацией. Такое усиление тем больше, чем больше фотонов имеется в данном квантовом состоянии. Число актов вынужденного излучения в единице объема в единицу времени равно $B_{p+q,p}^{(1)} n_{p+q} N_q (dq)$. Здесь $B_{p+q,p}^{(1)}$ — коэффициент вынужденного, или индуцированного, излучения.

Наконец, в равновесном состоянии процессы излучения уравниваются обратным процессом поглощения, при котором атомы переходят с энергетического уровня E_p на уровень $E_{p+q} = E_p + cq = E_p + \hbar\omega_q$. Число процессов поглощения фотонов с импульсами в интервале от q до $q + dq$ в единице объема в единицу времени равно $B_{p,p+q}^{(2)} n_p N_q (dq)$, где $B_{p,p+q}^{(2)}$ — коэффициент поглощения.

При равновесии частоты прямых и обратных переходов $p \rightleftharpoons p + q$ одинаковы для всех значений импульсов p и q (принцип детального равновесия). Значит, имеет место равенство

$$A_{p+q,p} n_{p+q} + B_{p+q,p}^{(1)} n_{p+q} N_q = B_{p,p+q}^{(2)} n_p N_q. \quad (2.3.1)$$

Множитель (dq) в равенстве (2.3.1) сократился.

Решая (2.3.1) относительно N_q , мы получим

$$N_q = \frac{A_{p+q,p} n_{p+q}}{B_{p,p+q}^{(2)} n_p - B_{p+q,p}^{(1)} n_{p+q}}. \quad (2.3.2)$$

Так как для атомов имеет место распределение Максвелла

$$n_p = C e^{-E_p/kT}, \quad (2.3.3)$$

и $E_{p+q} = E_p + \hbar\omega_q$, то

$$N_q = \frac{A_{p+q,p}}{B_{p,p+q}^{(2)} e^{\hbar\omega_q/kT} - B_{p+q,p}^{(1)}}.$$

Это выражение эквивалентно распределению Планка (2.1.1), если все три коэффициента равны:

$$A_{p+q,p} = B_{p+q,p}^{(1)} = B_{p,p+q}^{(2)}$$

(напомним, что эти величины зависят также от поляризации фотона).

Для дальнейшего удобно соотношение (2.3.1) написать, используя распределение Максвелла, в несколько ином виде:

$$A_{p+q, p} n_{p+q} (1 + N_q \overset{\neq}{=} N_q e^{\hbar\omega_q/kT}) = 0,$$

откуда

$$\frac{1 + N_q}{N_q} e^{-\hbar\omega_q/kT} = 1. \quad (2.3.4)$$

Выражение, стоящее слева, есть отношение числа фотонов, излучаемых ежесекундно, к числу поглощаемых; это отношение при равновесии равно единице.

В реальной плазме состояние излучающих и поглощающих атомов определяется не только импульсом, представляющим собой непрерывную переменную, но и набором дискретных квантовых чисел, характеризующих внутреннее состояние атома. Мы обозначим этот набор квантовых чисел одной буквой l , так что энергия атома $E_{l,p}$ должна быть теперь снабжена двумя значками. Распределение Максвелла тогда принимает вид

$$n_{l,p}(d\mathbf{p}) = C e^{-E_{l,p}/kT} (d\mathbf{p}).$$

Выражение $(d\mathbf{p}) = d^3\mathbf{p}/(2\pi\hbar)^3$ представляет собой число квантовых состояний в интервале импульсов $d^3\mathbf{p}$ и в единице объема. Закон сохранения энергии для переходов, при которых набор l квантовых чисел, изменяясь, принимает значение l' , и для обратных переходов можно теперь написать так:

$$E_{l, p+q} \rightleftharpoons E_{l', p} + \hbar\omega_q.$$

С этими изменениями наш вывод можно провести совершенно аналогично.

Равенство коэффициентов спонтанного $A_{p+q, p}$ и вынужденного $B_{p+q, p}^{(1)}$ излучения имеет глубокие основания в микроскопической природе фотонов и не зависит от предположения о статистическом равновесии атомов. Вероятность рождения фотона с импульсом q и определенной поляризацией всегда пропорциональна $N_q + 1$, где N_q — число уже имеющихся фотонов в этом квантовом состоянии. Равенство $B_{p, p+q}^{(2)} = A_{p+q, p}$ также не связано с предположением о равновесном состоянии атомов; оно

универсально и выражает принцип обратимости, согласно которому вероятности прямых и обратных переходов одинаковы. Мы воспользуемся этими общими свойствами для последующих выводов.

Вывод Эйнштейна относится к тому случаю, когда вещество взаимодействует с излучением, излучая и поглощая фотоны. Теперь же нас будет интересовать стационарное распределение излучения в результате комптоновского взаимодействия фотонов с электронами, сопровождающегося обменом энергией и импульсом. В этом случае понятия испускания и поглощения фотона приобретают специфический смысл. Если фотон с импульсом q' после рассеяния на электроне имеет импульс q , то первоначальный фотон будет фигурировать в выводе, аналогичном приведенному выше, как поглощенный, а рассеянный фотон — как испущенный. Число актов рассеяния этого типа в единицу времени равно

$$A_{qq'} n_p N_{q'} (N_q + 1), \quad (2.3.5)$$

где p — импульс электрона до акта рассеяния.

Частота процессов обратного типа, когда фотон с импульсом q рассеивается, имея импульс q' ,

$$A_{qq'} n_{p'} N_{q'} (N_{q'} + 1). \quad (2.3.6)$$

Здесь $p' = p + q' - q$ — импульс электрона после акта рассеяния прямого типа; в обратном процессе он является начальным импульсом. Вероятности $A_{qq'}$ обоих процессов одинаковы в согласии с принципом обратимости. В стационарном состоянии частоты обоих этих процессов должны быть по принципу детального равновесия одинаковы. Следовательно,

$$\frac{N_{q'} + 1}{N_{q'}} = \frac{N_q + 1}{N_q} \cdot \frac{n_p}{n_{p'}}.$$

При максвелловском распределении атомов, используя закон сохранения энергии $E_{p+q'-q} = E_p + \hbar\omega_{q'}$, $-\hbar\omega_q$, мы получим

$$\frac{N_q + 1}{N_q} e^{-\hbar\omega_q/kT} = \frac{N_{q'} + 1}{N_{q'}} e^{-\hbar\omega_{q'}/kT}. \quad (2.3.7)$$

Полученное соотношение должно иметь место при всех значениях q и q' ; следовательно,

$$\frac{1 + N_q}{N_q} e^{-\hbar\omega_q/kT} = \text{const.} \quad (2.3.8)$$

Это соотношение отличается от (2.3.4) тем, что там справа стояла единица. Следовательно, комптоновское взаимодействие излучения с веществом не может, вообще говоря, привести к равновесному распределению излучения. Стационарное состояние, к которому приводят эти процессы, не описывается распределением Планка. Для достижения последнего, как мы уже говорили, необходимы тормозные процессы, которые рассмотрены в § 2.2.

Константу в выражении (2.3.8) удобно обозначить через $e^{-\mu/kT}$. Тогда из (2.3.8) следует распределение

$$N_q = \frac{1}{e^{(\hbar\omega_q - \mu)/kT} - 1}. \quad (2.3.9)$$

Распределение (2.3.9) отличается от функции распределения Планка членом, аналогичным химическому потенциалу в распределении Бозе—Эйнштейна для частиц с конечной массой покоя (см. Ландау и Лифшиц (1964)). Появление «химического потенциала» фотонов связано с тем уже упоминавшимся обстоятельством, что, несмотря на разогрев излучения энергией, передаваемой ему плазмой, число фотонов в рассмотренных нами процессах не меняется, и потому при новой, более высокой температуре уже не определяется планковской формулой. Число фотонов меньше, чем было бы при равновесии, и потому в формуле (2.3.9) должно быть $\mu < 0$.

Заметим, что стационарное состояние системы фотонов не является равновесным и потому величину T , строго говоря, нельзя называть температурой излучения. В действительности, это есть температура, которую получают электроны в результате взаимодействия с излучением и которая, как уже сказано, в космологических условиях мало отличается от начальной температуры последнего. Из-за малости изменения температуры излучения естественно ожидать, что $|\mu| \ll kT$.

Появление «химического потенциала» особенно существенно сказывается на распределении фотонов в области Рэлея — Джинса, где $\hbar\omega \ll kT$. Разлагая в этом случае выражение для N_q в ряд по степеням $\hbar\omega/kT$ и ограничиваясь первой степенью, получим

$$N_q \approx \frac{kT}{\hbar\omega_q + |\mu|}. \quad (2.3.10)$$

Спектральная плотность излучения в области Рэлея — Джинса, т. е. энергия единицы объема, приходящаяся в этой области на единичный интервал частот, есть

$$\varepsilon(\omega) \propto \frac{\omega^3}{\hbar\omega + |\mu|}. \quad (2.3.11)$$

В распределении Планка в той же области $\varepsilon(\omega) \propto \omega^2$; в выражении же (2.3.11) эта зависимость сохраняется лишь при $\hbar\omega \gg |\mu|$. Для самых низких частот ($\hbar\omega \ll |\mu|$) величина $\varepsilon(\omega)$ пропорциональна ω^3 , т. е. уменьшается при уменьшении частоты быстрее, чем в распределении Планка.

Наблюдения реликтового излучения в низкочастотной области спектра позволяют считать это излучение равновесным, планковским; если интенсивность излучения и отличается от планковской, то не более чем на 10%. Эти данные можно использовать для оценки верхних пределов тепловыделения в те или иные эпохи космологического расширения.

Вопрос об ограничениях такого рода на энерговыделение в метagalacticкой среде впервые поставлен и детально рассмотрен Зельдовичем и Сюняевым (1969), Сюняевым и Зельдовичем (1970), Илларионовым и Сюняевым (1974) (см. также книгу Зельдовича и Новикова (1975)). Оказывается, что выделение энергии в интервале красных смещений $10^8 > z > 10^4$ могло бы быть замеченным: оно приводит к появлению распределения фотонов с химическим потенциалом. Отсюда возникает ограничение на энергию $\Delta\varepsilon$, переданную излучению в единице объема: $\Delta\varepsilon/\varepsilon_r < 0,03$. Что же касается позднего энерговыделения после рекомбинации, то оно могло бы создать значительные отклонения от планковского спектра в его низкочастотной части. Поэтому исключается, например, воз-

возможность, того, что электроны были нагреты до температуры $\sim 10^6$ °К вплоть до $z \sim 700$. Но остается не исключенной возможность более сильного нагрева, до $\sim 10^7$ — 10^8 °К, вблизи момента рекомбинации, что допускает в этот момент диссипацию движений со скоростями $\sim 10^8$ см/сек. Разогрев электронов мог быть следствием диссипации возмущений — локальных отклонений от однородности и изотропии Вселенной, связанных с гидродинамическими движениями, паложенными на общее космологическое расширение (см. гл. 3 и 5).

МАЛЫЕ ВОЗМУЩЕНИЯ В ИЗОТРОПНОЙ ВСЕЛЕННОЙ

§ 3.1. Гравитационная неустойчивость

Догалактические возмущения. В двух первых главах мы исследовали общие свойства Вселенной, отвлекаясь от ее структуры, основными элементами которой являются галактики и скопления галактик. Это возможно благодаря тому, что иерархия космических структур не продолжается к неограниченно большим пространственным масштабам, а «обрезается» размером ячейки однородности в современном мире $\sim 100\text{—}300$ *Mpc*. В больших масштабах наблюдаемая Вселенная бесструктурна — однородна и изотропна. Для анализа ранних фаз космологического расширения, предыстории современной Вселенной важно и другое обстоятельство. При высоких температурах и плотностях, характерных для ранней Вселенной (см. § 2.1), галактики — эти гравитационно-связанные стационарные системы — не могли, очевидно, существовать в их современном виде. Достаточно сказать, что даже атомные ядра тяжелее протона не существовали во Вселенной изначально, а атомы возникли только в эпоху рекомбинации.

Однако полная бесструктурность — идеальная однородность и изотропия во всех масштабах — была невозможна. Мы увидим в дальнейшем, что изначальные догалактические структуры существовали во Вселенной по крайней мере после эпохи ядерной плотности $z < z_N \approx \approx 10^{15}$, и соответствующие им отклонения от однородности и изотропии должны были на много порядков превосходить уровень статистических флуктуаций в метagalacticкой среде. В настоящей главе мы рассмотрим эволюцию локальных отклонений от изотропии, считая, что отвечающая им относительная амплитуда, хотя и велика по сравнению с флуктуационной, но все же не превосходит единицы.

В условиях метагалактической среды возможны несколько различных типов малых возмущений. Это в первую очередь возмущения плотности и макроскопической скорости среды. В соответствии с принятой в гидродинамике терминологией, мы будем разделять такие возмущения на потенциальные и вихревые. При потенциальных возмущениях изменяется плотность среды ρ (отклонение плотности от средней обозначим через $\delta\rho$) и в ней возникает движение с некоторой скоростью v_p , являющейся потенциальным вектором,

$$v_p = \text{grad } \varphi,$$

поэтому

$$\text{div } v_p \neq 0, \quad \text{rot } v_p = 0, \quad \delta\rho \neq 0.$$

При вихревых возмущениях скорость

$$v_v = \text{rot } A,$$

так что

$$\text{rot } v_v \neq 0, \quad \text{div } v_v = 0, \quad \delta\rho = 0.$$

Метагалактическая среда представляет собой многокомпонентную систему; она характеризуется распределением излучения, распределением барионов. Поэтому, кроме гидродинамических возмущений указанного типа, возможны локальные возмущения, при которых имеются локальные неоднородности одной компоненты при неизменной плотности другой. Реальный интерес могут представить возмущения плотности барионов при неизменной плотности излучения, когда последняя преобладает. Вариации плотности плазмы на невозмущенном фоне излучения называются энтропийными возмущениями в соответствии с общей классификацией малых возмущений в гидродинамике (см., например, книгу Мони́на и Яглома (1966)); в астрофизике такие возмущения впервые рассматривались Гуревичем и Лебединским (1955), в космологии — Зельдовичем (1966).

Неоднородности плазмы, «вмороженные» в изотропное излучение, мало влияют на общую динамику расширения, когда плотность излучения много больше плотности вещества, и в процессе космологического расширения величина относительного возмущения плотности плазмы не меняется, но отношение избыточной плотности вещества в некоторой области к полной плотности возрастает со

временем:

$$\frac{\delta\rho}{\rho_r + \rho} \propto a \propto (1+z)^{-1}, \quad \rho_r \gg \rho.$$

Вблизи космологической особенности возмущения этого типа очень малы даже при $\delta\rho/\rho > 1$ и практически не искажают метрику фридмановской модели. Их роль возрастает со временем и может стать существенной в ту эпоху эволюции Вселенной, когда плотность излучения становится сравнимой с плотностью вещества. После рекомбинации, когда динамическое взаимодействие вещества и излучения прекращается (см. § 2.2), энтропийные возмущения превращаются в потенциальные возмущения.

Помимо гидродинамических возмущений, возможны также магнитные возмущения, при которых в метагалактической плазме возникают электрические токи и связанные с ними магнитные поля.

Возможны, наконец, возмущения метрического тензора, никак не связанные с гидродинамикой, — гравитационные волны.

Все перечисленные возмущения являются локальными. Кроме этого, допустимы возмущения совершенно иного типа, которые можно назвать глобальными. При таких возмущениях Вселенная, оставаясь даже однородной, оказывается в большей или меньшей степени анизотропной. Эти возмущения, которые могут быть и не малыми, рассматриваются в гл. 4.

Возмущения различных типов по-разному ведут себя в расширяющейся горячей Вселенной. Некоторые из них затухают со временем, тогда как другие могут усиливаться. С точки зрения космогонии возможность усиления слабых возмущений, их превращение со временем в немалые локальные отклонения от однородности и изотропии представляют принципиальный интерес. На этом основывается точка зрения, согласно которой наблюдаемые космические структуры сформировались благодаря усилению первоначальных возмущений, совместимых с почти полной однородностью и изотропией мира на всех стадиях космологического расширения, начиная с особенности. Другая, противоположная точка зрения, также допустимая на нынешнем уровне исследований, предполагает возможность сильных отклонений от изотропии, сильной нефридманов-

ности уже вблизи сингулярности. Анизотропные космологические модели мы рассмотрим в гл. 4. Теоретически не исключено, что дальнейшая эволюция такого, как говорят, первичного хаоса (противопоставляемого почти полной начальной регулярности в картине малых возмущений) способна дать начало наблюдаемой космической структуре.

Физические процессы в метagalактической среде, которые мы обсуждали в первых двух главах и будем обсуждать здесь, служат основой, на которой строятся современные космогонические представления. Среди этих процессов особое место занимает гравитационная неустойчивость. Для ее анализа требуется изучение динамики неоднородного распределения гравитирующей среды. Наиболее важные черты процесса могут быть выяснены, однако, на самом простом примере неоднородного, но сферически-симметричного распределения.

Ньютоновская теория сферически-симметричного движения гравитирующей среды. Пусть в некоторой области однородной среды имеется сферически-симметричное возмущение плотности. Примем следующие предположения: 1) давлением среды можно пренебречь, 2) масса, находящаяся внутри сферического слоя, образованного данными частицами, не изменяется со временем (это означает, что отсутствует пересечение слоев). Ограничимся сначала ньютоновским приближением.

Уравнение движения частицы при наличии сферической симметрии имеет вид

$$\frac{\partial^2 R}{\partial t^2} = - \frac{GM}{R^2}. \quad (3.1.1)$$

Здесь R — расстояние частицы от центра, M — масса внутри сферического слоя, к которому принадлежит частица. Величина R есть, очевидно, функция двух переменных — t и M . Первый интеграл уравнения (3.1.1) — это закон сохранения механической энергии частицы:

$$\frac{1}{2} \left(\frac{\partial R}{\partial t} \right)^2 = \frac{GM}{R} + E. \quad (3.1.2)$$

Не зависящая от времени величина $E(M)$ — сумма кинетической и потенциальной энергий данной частицы, приходящаяся на единицу ее массы. Второй интеграл

уравнения движения получим, разрешая (3.1.2) относительно dt :

$$t - t_i(M) = \int_0^R \frac{dR'}{\sqrt{2} \left(\frac{GM}{R'} + E \right)^{1/2}}. \quad (3.1.3)$$

Здесь $t_i(M)$ — вторая постоянная во времени величина. Из (3.1.3) видно, что $t_i(M)$ есть момент времени, когда данная частица находится в центре. При $t = t_i(M)$ плотность внутри данного слоя частиц бесконечна, т. е. имеет место сингулярность того же типа, что и в модели Фридмана, но она наступает, вообще говоря, в разное время для разных слоев.

Результат интегрирования (3.1.3) относится к разным частицам, характеризуемым неизменной во времени массой M . Величина M служит здесь сопутствующей, лагранжевой координатой частицы; расстояние частицы от центра $R(t, M)$ — это обычная эйлерова координата, определяющая положение частицы в пространстве. Две произвольные функции лагранжевой координаты M — $E(M)$ и $t_i(M)$, допускают возможность задания «начальных условий», т. е. распределений плотности $\rho(t_1, M)$ и скорости $v(t_1, M)$ частиц в некоторый момент времени $t = t_1$. Связь между величинами R и t удобнее всего найти из (3.1.3), вводя некоторую вспомогательную переменную η , через которую могут быть выражены обе интересующие нас величины:

$$\left. \begin{aligned} R &= R_m (1 - \cos \eta), \\ t - t_i &= \frac{R_m}{(2|E|)^{1/2}} (\eta - \sin \eta), \end{aligned} \right\} E < 0: \quad (3.1.4)$$

$$\left. \begin{aligned} R &= R_m (\operatorname{ch} \eta - 1), \\ t - t_i &= \frac{R_m}{(2E)^{1/2}} (\operatorname{sh} \eta - \eta), \end{aligned} \right\} E > 0. \quad (3.1.5)$$

Здесь

$$R_m = R_m(M) = \frac{GM}{2|E|}. \quad (3.1.6)$$

В случае $E = 0$ зависимость R от t легко получить в явном виде:

$$R = \left(\frac{3}{2} \right)^{2/3} (2GM)^{1/3} (t - t_i)^{2/3}. \quad (3.1.7)$$

Решение (3.1.3) — (3.1.7) представляет собой ньютоновский аналог (см., например, Рубан и Чернин (1969)) общерелятивистского решения, найденного в 30-е годы Толменом.

Решение (3.1.3) — (3.1.7) можно рассматривать как обобщение найденного ранее (§ 1.2) ньютоновского решения для однородного распределения вещества. Если плотность среды однородна в течение всей эволюции, она не зависит от расстояния и должна быть одинакова внутри любого слоя, характеризуемого данной массой M . Но так как в этом случае

$$M = \frac{4\pi}{3} \rho R^3,$$

эйлерова координата R должна быть связана с лагранжевой координатой M соотношением:

$$R \propto M^{1/3}. \quad (3.1.8)$$

Положим теперь в (3.1.2)

$$R = M^{1/3} F(t) \quad (3.1.9)$$

и разделим левую и правую стороны на $M^{2/3}$. Тогда из (3.1.2) следует, что в данном случае

$$E(M) \propto M^{2/3}. \quad (3.1.10)$$

Если мы теперь сделаем ту же подстановку (3.1.9) в (3.1.3), то интеграл справа примет вид

$$\int_0^F \frac{dF}{\left(\frac{G}{F} - \frac{E}{M^{2/3}}\right)^{1/2}},$$

т. е. окажется не зависящим от M . Поэтому и величина t_i слева в (3.1.3) также не должна зависеть от M . Последнее означает, что все частицы находятся в центре одновременно. Это — одновременная фридмановская сингулярность. В параметрических формулах (3.1.4) — (3.1.6) функция $R_m(M)$ принимает при этом вид

$$R_m(M) = 2\sqrt{|E|}.$$

Возвращаясь к общему решению (3.1.3) — (3.1.7), укажем на существенное значение знака удельной энергии $E(M)$. Если $E < 0$, то расстояние $R(t, M)$ не может

быть неограниченно велико. Правая сторона (3.1.2) должна быть положительна, и потому

$$R < R_{\max} = \frac{GM}{|E|}. \quad (3.1.11)$$

Если в начальный момент имело место удаление частиц от центра, т. е. происходило расширение, то оно должно смениться сжатием после достижения некоторого максимального радиуса. Если $E \geq 0$, то возможно неограниченное расширение. При $E > 0$ в процессе расширения будет достигнуто состояние, в котором $GM/R \ll E$; тогда действием гравитации вообще можно пренебречь и расширение происходит инерциально.

Эти особенности сходны с теми, с которыми мы встретились, рассматривая ньютоновский аналог космологии Фридмана. Однако в нашем случае произвольная зависимость $E(M)$ допускает и иные возможности. В частности, не только величина, но и знак E могут быть различны для разных сферических слоев. При этом все же должно выполняться ограничение, запрещающее пересечение слоев.

Насколько это ограничение существенно, видно из одного частного сферически-симметричного решения, в котором, как и в решении (3.1.3) — (3.1.7), частицы взаимодействуют только гравитационно, но допускается взаимное пересечение слоев (Чернин, Мороз и Вандакуров (1972)). Такие движения с пересечениями называют негомологическими. В самом простом случае это однородная, но двухкомпонентная среда, причем одна компонента свободно проходит через другую, взаимодействуя с ней только гравитационно. Хотя каждая из компонент расширяется по хаббловскому закону, движение в целом не однородно и анизотропно: имеется только центральная симметрия относительно точки, в которой частицы обеих компонент покоятся. Однако изотропия может достигаться с течением времени, когда, как оказывается, обе компоненты становятся асимптотически сопутствующими друг другу и пересечения прекращаются. Действительно, уравнения движения для такой двухкомпонентной среды можно записать в виде уравнений для двух масштабных факторов, характеризующих расширение каждой из

компонент:

$$\ddot{a}_i(t) = -\frac{4\pi}{3} G \rho(t) a_i(t), \quad (3.1.12)$$

где индекс i принимает значения 1, 2, а плотность

$$\rho = \rho_1 + \rho_2, \quad \rho_i = \left(\frac{4\pi}{3} G\right)^{-1} \frac{A_i}{a_i^3}. \quad (3.1.13)$$

Здесь A_1, A_2 — положительные константы; давление, как и в решении (3.1.3) — (3.1.7), равно нулю. Полная масса, находящаяся внутри сферы, сопутствующей данным частицам, изменяется со временем, так что каждая частица движется в переменном гравитационном поле и ее механическая энергия не сохраняется. Первый интеграл уравнений движения есть

$$\dot{a}_1 a_2 - \dot{a}_2 a_1 = C_0 = \text{const.}$$

Отсюда видно, что если, например, a_1 возрастает со временем быстрее, чем $t^{1/2}$, отношение a_2/a_1 стремится к постоянному пределу, когда время стремится к бесконечности. Второе интегрирование дает:

$$a_1 = \frac{2C_0}{I(\xi)} = a_2 \xi,$$

$$t - t_i = \int_0^{\xi} x^2 [I(x)]^{-2} dx,$$

где

$$I(\xi) = A_2 \xi^3 + 2C_2 \xi^2 + 2C_1 \xi + A_1.$$

Здесь C_1, C_2, t_i — константы, а параметр ξ есть монотонная функция времени. При малых временах расширение негомологично:

$$a_1 \propto t^{1/2}, \quad a_2 \propto t^{1/2}, \quad \rho_1/\rho_2 \rightarrow 0, \quad t \rightarrow t_i.$$

При больших временах расширение переходит в инерциальный режим:

$$a_1 \propto a_2 \propto t, \quad \rho_1/\rho_2 \propto \text{const}, \quad t \rightarrow \infty,$$

если положительные корни полинома $I(\xi)$ не равны друг другу. Если же эти корни равны, то асимптотикой при больших временах является параболическое расширение:

$$a_1 \propto a_2 \propto t^{2/3}, \quad \rho_1/\rho_2 \propto \text{const}, \quad t \rightarrow \infty.$$

Эволюция потенциальных возмущений. Рассмотрим развитие во времени гидродинамических возмущений в первоначально однородной и изотропной метагалактической среде. Нас будут интересовать возмущения плотности и скорости при слабых потенциальных движениях, наложенных на общее космологическое расширение. Для космогонического процесса существенны возрастающие возмущения, и мы рассмотрим условия, при которых этот рост происходит под действием гравитации. Собственное тяготение частиц среды стремится замедлить расширение вещества в областях повышенной плотности, а затем привести к смене расширения сжатием. Возможность таких возрастающих возмущений представляет собой проявление неустойчивости однородно распределенного гравитирующего вещества.

Как мы уже говорили в гл. 1, наблюдательные данные, вероятнее всего, указывают на то, что пространство реального мира — открытое, либо близкое к плоскому. Для открытого пространства гравитация не играет роли на поздних стадиях расширения, когда движение становится инерциальным; на этих стадиях малые возмущения не могут усиливаться. На ранних стадиях расширения открытой Вселенной гравитация существенна, и невозмущенное расширение можно считать параболическим с хорошей точностью для всех трех моделей Фридмана.

Начнем с простейшего случая сферически-симметричного возмущения в среде с пренебрежимо малым давлением и будем проводить его исследование на основе ньютоновской теории. Воспользуемся для этого полученным выше общим решением (3.1.2) — (3.1.7).

Пусть имеется сферическая область с начальным радиусом, который мал по сравнению с расстоянием до горизонта событий и с радиусом кривизны сопутствующего 3-пространства. Пусть, далее, в некоторый начальный момент времени распределение плотности и скорости частиц внутри области возмущения немного отличается от распределения в окружающем пространстве. Примем, согласно сказанному выше, что в окружающем пространстве (невозмущенная среда) удельная энергия частиц $E(M)$ равна нулю; в рассматриваемом же сферическом объеме будем считать ее отрицательной и малой (по модулю) по сравнению с кинетической энергией расширения. Оцени-

вая последнюю по порядку величины, можем написать:

$$|E| \ll \left(\frac{R}{t}\right)^2. \quad (3.1.14)$$

Этот критерий гарантирует малость возмущения поля скоростей; для того чтобы однородность плотности также нарушалась лишь слабо, необходимо, чтобы отношение $E/M^{2/3}$ слабо зависело от M (в однородном случае оно не зависит от M). Представляя это условие в безразмерном виде, напишем:

$$\frac{M}{E} \frac{\partial E}{\partial M} - \frac{2}{3} \ll 1. \quad (3.1.15)$$

Сферический объем с отрицательной полной энергией должен расширяться медленнее, чем окружающее вещество; определим относительное изменение радиуса $\delta R/R$ и относительное изменение плотности $\delta\rho/\rho$, которые возникнут в нем к некоторому моменту времени (рис. 9).

Для нахождения первой величины разложим тригонометрические функции в (3.1.4) в ряды по $\eta \ll 1$ и ограничимся в каждом из них двумя первыми членами. Тогда

$$R = R_F \left(1 - \frac{t_i(M)}{t}\right)^{2/3} \left[1 - \frac{1}{5} \frac{|E|}{GM} \left(1 - \frac{t_i(M)}{t}\right) R_F\right], \quad (3.1.16)$$

где

$$R_F = \left(\frac{3}{2}\right)^{2/3} (2GM)^{1/3} t^{2/3} \quad (3.1.17)$$

есть невозмущенное фридмановское параболическое решение (см. § 1.2). Если возмущение описывается также параболическим решением

$$R = R_F \left(1 - \frac{t_i(M)}{t}\right)^{2/3}, \quad E = 0, \quad (3.1.18)$$

то отклонение от фридмановского решения убывает и возмущение затухает со временем.

При $E \neq 0$ возмущения могут нарастать, и для самого простого случая, когда $t_i(M) = 0$, имеем

$$\frac{\delta R}{R} \equiv \frac{R - R_F}{R} \propto R_F \propto t^{2/3}. \quad (3.1.19)$$

Если определить среднюю плотность области возмущения

формулой

$$\tilde{\rho}(t) = M / \frac{4\pi}{3} R^3,$$

то относительное возмущение плотности в растущих возмущениях

$$\frac{\delta\rho}{\rho} \equiv \frac{\tilde{\rho} - \rho}{\rho} = 3 \frac{\delta R}{R} \sim t^{2/3}. \quad (3.1.20)$$

Возмущение скорости также может быть найдено из

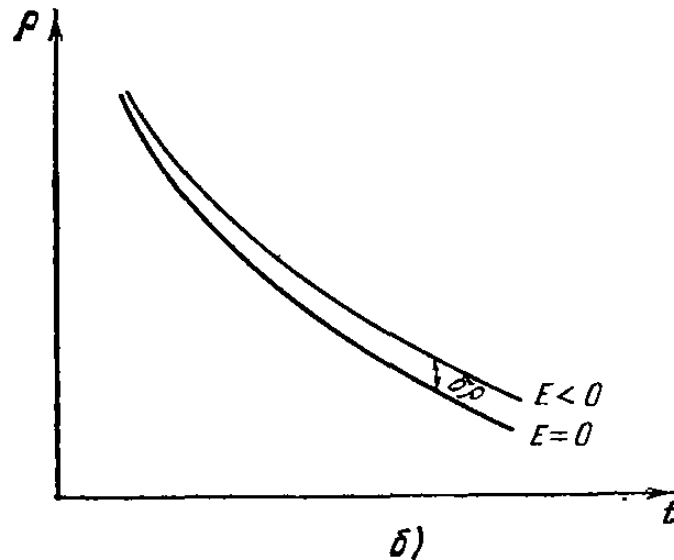
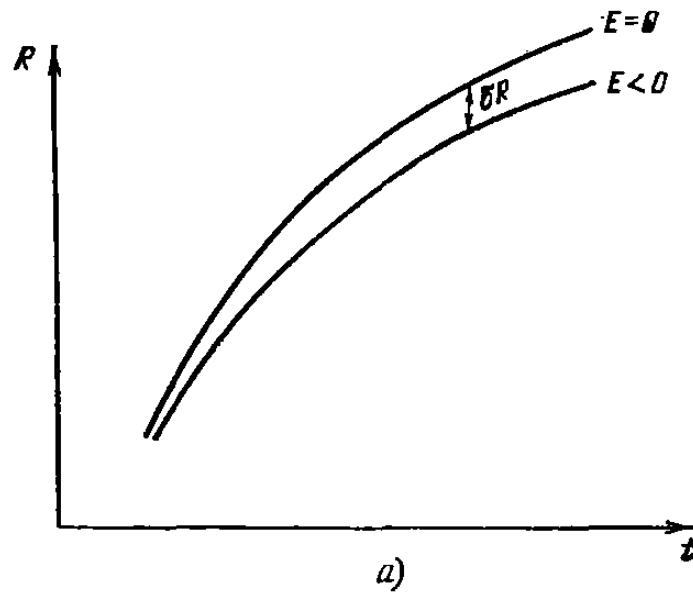


Рис. 9. Изменение со временем размеров (а) и плотности (б) в области возмущения ($E < 0$) и в невозмущенной среде ($E = 0$).

равложения (3.1.16). Для растущих возмущений

$$\frac{\partial R}{\partial t} = \frac{\partial R_F}{\partial t} \left(1 - \frac{2}{5} \frac{|E|}{GM} R_F \right).$$

Отсюда скорость дополнительных, собственных движений

среды, обязанных возмущению, есть

$$v \equiv \frac{\partial R}{\partial t} - \frac{\partial R_F}{\partial t} = \frac{\partial R}{\partial t} - \infty t^{1/2} \quad (3.1.21)$$

(V — хаббловская скорость регулярного космологического расширения.)

Ньютоновское и релятивистское решения для сферической симметрии обладают весьма далеко идущим подобием (см. Дополнение). Так, функции $R(t, M)$ и $\rho(t, M)$, играющие ключевую роль в общих решениях, имеют в них один и тот же вид. Естественно поэтому ожидать, что итоговые временные зависимости для растущих возмущений,

$$\frac{\delta \rho}{\rho} \propto t^{1/2}, \quad v \propto t^{1/2} \quad (p = 0), \quad (3.1.22)$$

должны сохранить силу и в релятивистской теории. Это очевидным образом подтверждается для сферически-симметричного случая, но, как оказывается, имеет и более общий смысл: соотношения (3.1.22) справедливы для произвольной формы возмущений. Это следует из теории гравитационной неустойчивости Лифшица (1946), разработанной для самого общего вида малых возмущений.

Действительно, в релятивистской постановке задачи уравнения для возмущений могут быть написаны в форме, в которой коэффициенты зависят лишь от космологического времени, что естественно, поскольку невозмущенное состояние является однородным. Для таких уравнений переменные разделяются, и после выделения множителя, зависящего от координат, остаются уравнения для временной части искомых функций, не зависящих от геометрии возмущений. Поэтому зависимость возмущений от времени и оказывается универсальной.

С другой стороны, переход от релятивистской теории к ньютоновской требует ограничений на скорости (они должны быть малы по сравнению со скоростью света) и на гравитационный потенциал (малость его перепадов по сравнению с квадратом скорости света); этим ограничениям можно удовлетворить с помощью координатной части возмущений, и в таком случае зависимость от времени ими не затрагивается. Но раз так, полученные нами эволюционные зависимости (3.1.22) должны быть дей-

ствительно справедливы и в релятивистской теории для возмущений произвольной формы.

Принятое выше условие равенства нулю давления, разумеется, не означает полного отсутствия давления в среде; оно означает лишь пренебрежение силами давления по сравнению с гравитационной силой (см. ниже). В среде, в которой давление p сравнимо с плотностью энергии, дело обстоит несколько иначе. В этом случае давление обуславливает не только силу, которая связана с его градиентом и которой при достаточно малом градиенте можно пренебречь; оно, кроме того, наряду с плотностью энергии создает гравитационное действие (ср. § 1.4). При $p = 1/3\varepsilon$ (ультрарелятивистский газ или излучение) этим гравитационным действием пренебречь нельзя; в задаче о малых возмущениях оно приводит к тому, что временные зависимости оказываются иными, хотя степенной их характер сохраняется. Согласно Лифшицу (1946)

$$\frac{\delta\rho}{\rho} \propto t, \quad v \propto t^{1/2} \quad \left(p = \frac{1}{3}\varepsilon\right). \quad (3.1.23)$$

Возмущения гидродинамических величин (плотности и скорости) связаны, как мы видели, с изменениями гравитационного потенциала. На языке общей теории относительности это означает изменения метрического тензора. Обозначим возмущения метрического тензора через $h_{ik} = g_{ik} - g_{ik}^F$, где g_{ik}^F — невозмущенный метрический тензор модели Фридмана. Для ультрарелятивистской среды теория малых возмущений (в линейном приближении) дает следующее соотношение, связывающее возмущения плотности и метрики:

$$\frac{\delta\rho}{\rho} \approx h \left(\frac{ct}{R}\right)^2, \quad p = \frac{1}{3}\varepsilon.$$

Здесь $h = h_i^i = g_F^{ik} h_{ik} < 1$ — след тензора h_{ik} , причем h , как показывает решение уравнений гравитации, не меняется в этом случае со временем. Возмущения плотности и скорости усиливаются, пока их пространственный масштаб больше расстояния до горизонта событий (см. ниже). Так как $R \propto t^{1/2}$, возмущения данного масштаба возрастают в линейной теории лишь ограниченное время и по-

тому остаются малыми:

$$\frac{\delta\rho}{\rho} < h < 1.$$

В нерелятивистской среде указанные ограничения для роста возмущений, как оказывается, отсутствуют. Последнее можно увидеть из точных решений для сферически-симметричного движения гравитирующей среды, которые мы использовали выше для анализа эволюции слабых потенциальных возмущений. Эти решения справедливы и тогда, когда описываемые ими отличия от изотропного решения не малы. Однако в этом случае частные решения уже не позволяют получать результаты, применимые и в более общем случае; нелинейная стадия развития гравитационной неустойчивости гораздо сложнее и для ее анализа такие решения недостаточны (см. гл. 5).

Критерий Джинса. Пренебрежение эффектами, происходящими от градиентов давления, означает рассмотрение достаточно крупных областей возмущений, для которых гравитационные силы преобладают.

Действительно, силу давления, действующую на единицу массы в объеме возмущения, можно оценить следующим образом:

$$-\frac{\nabla p}{\rho} \approx -\frac{p_1 - p}{\rho R} \sim \frac{p}{\rho R}.$$

Здесь p есть давление внутри рассматриваемого объема, а p_1 — давление на его границе. Для оценки можно пренебречь последним, поскольку мы рассматриваем область повышенной плотности, а потому и повышенного давления.

С другой стороны, гравитационная сила, действующая на единицу массы, по порядку величины $\sim -GM/R^2 \approx \approx -G\rho R$. Поэтому условие пренебрежения первой силой по сравнению со второй сводится к неравенству

$$R > R_J \approx \left(\frac{p}{G\rho^2}\right)^{1/2}. \quad (3.1.24)$$

Так как $p \approx \rho u^2$ (где $u = \left(\frac{\partial p}{\partial \rho}\right)_s$ — адиабатическая скорость звука, s — энтропия), то критическая длина есть

$$R_J \approx u/(G\rho)^{1/2} \approx ut. \quad (3.1.25)$$

Ее называют длиной Джинса, по имени английского астрофизика, установившего критерий (3.1.25) в начале века (Джинс (1919, 1929)).

В случае ультрарелятивистской среды, когда

$$p = \frac{1}{3} \varepsilon, \quad u = c/\sqrt{3}, \quad \text{[(3.1.26)}$$

джинсова длина

$$R_J \approx c/(G\rho)^{1/2} \approx ct, \quad (3.1.27)$$

сравнима с расстоянием до горизонта.

Для нерелятивистской среды джинсова длина много меньше расстояния до горизонта, так что размер возмущения может превышать R_J и в то же время быть меньше, чем ct .

В горячей Вселенной метагалактическая среда является ультрарелятивистской на ранней стадии, когда вещество и излучение представляют собой единую жидкость с преобладанием излучения ($z > z_{\text{rec}}, z > z_*$). Джинсова длина нарастает со временем вплоть до эпохи рекомбинации метагалактической плазмы, $z = z_{\text{rec}}$. После рекомбинации вещество отделяется от излучения, и упругость среды определяется только ее собственным давлением, которое при температурах $T < T_{\text{rec}} \approx 3 \cdot 10^3$ °К много меньше плотности энергии ρc^2 . Это приводит к резкому падению скорости звука, а с ней джинсовой длины и соответствующей ей массы при $z = z_{\text{rec}} \approx 10^3$ (рис. 10). Сразу после рекомбинации джинсова длина

$$R_J(z_{\text{rec}}) \approx \left(\frac{kT_{\text{rec}}}{mG\rho_{\text{rec}}} \right)^{1/2} \approx 3 \cdot 10^{19} \Omega^{-1/2} \text{см} \quad (3.1.28)$$

соответствует массе

$$M_J(z_{\text{rec}}) = \frac{4\pi}{3} \rho R_J^3 \approx 3 \cdot 10^5 \Omega^{-1/2} M_{\odot}. \quad (3.1.29)$$

В результате растущими оказываются возмущения тех масштабов, в которых до этого гравитационная неустойчивость не развивалась и потенциальные движения представляли собой колебания давления и плотности, распространявшиеся как звуковые волны. Самый переход к неустойчивости сопровождается для них быстрым возрастанием амплитуды плотности (Зельдович и Сюняев (1970а)).

Если до рекомбинации эта величина была связана с собственной скоростью движений обычным акустическим соотношением

$$\frac{\delta\rho}{\rho} \approx \frac{v}{u}, \quad u \approx \frac{c}{\sqrt{3}},$$

то после нее, когда возмущение становится неустойчивым,

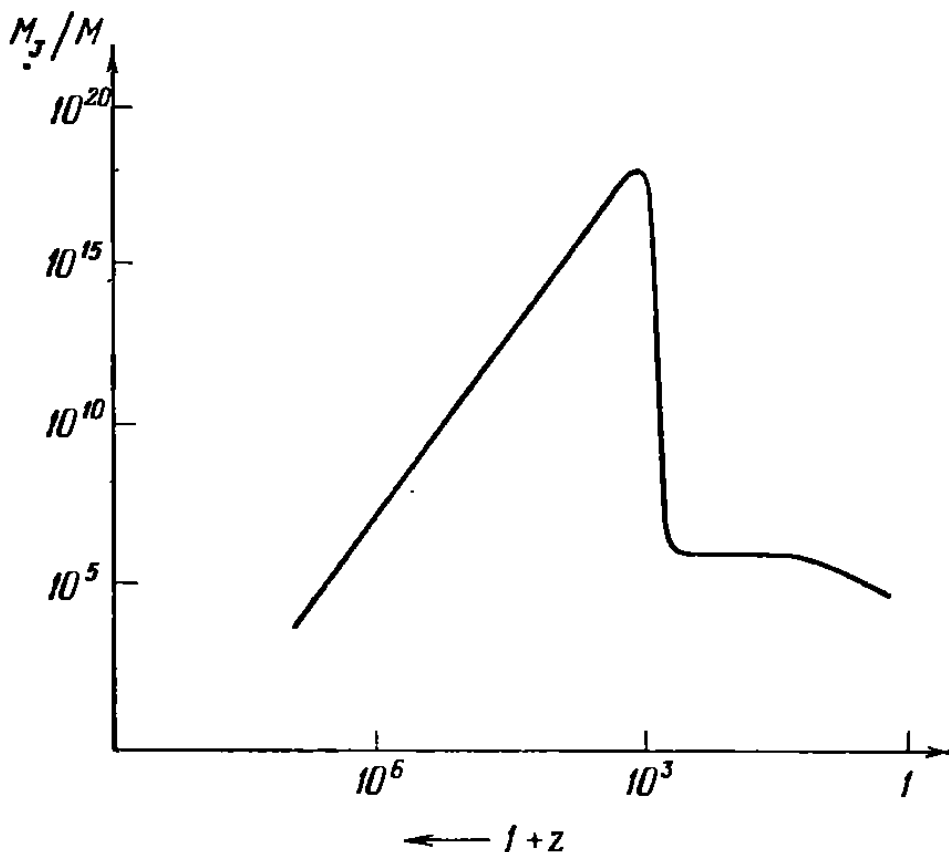


Рис. 10. Джинсова масса в горячей Вселенной.

связь $\delta\rho$ и v оказывается, как легко видеть на основании (3.1.20), иной:

$$\frac{\delta\rho}{\rho} \approx \frac{v}{V}, \quad V \approx \frac{R}{t}.$$

«Сшивка» обеих формул при $z = z_{\text{rec}}$ с учетом того, что скорость v остается практически неизменной, показывает, что возмущение плотности возрастает при такой смене режимов движения приблизительно в отношении ct/R .

Заметим, что в оригинальной работе Джинса (1929) характерная длина гравитационной неустойчивости находилась из анализа распространения звука в однородной стационарной среде. При наличии гравитации, однако, стационарность однородной среды невозможна, и чтобы

избежать нестационарности, Джинс учитывал гравитацию только для возмущения, но не для фона (что, конечно, не вполне корректно). Если так, то можно искать зависимость возмущений от времени и координат в виде $\sim \exp(i\omega t - ikx)$, где $k = 2\pi/R$ — волновое число для возмущений масштаба, R , ω — частота звука. Уравнения движения и неразрывности вместе с уравнением Пуассона для гравитационного потенциала приводят в этом случае к соотношению:

$$\omega^2 = u^2 k^2 - 4\pi G\rho, \quad (3.1.30)$$

откуда и следует неравенство (3.1.24) для растущих со временем возмущений, которым отвечает $\omega^2 < 0$. В обратном предельном случае гравитацией можно пренебречь, и возмущение есть обычный звук.

Малы ли малые возмущения? Итак, в однородной и изотропно расширяющейся метagalактической среде возможно развитие и усиление малых возмущений плотности и скорости. Усиление возмущений обязано собственной гравитации среды, которая является определяющим фактором эволюции для возмущений достаточно большого пространственного масштаба.

В открытом мире ($\Omega < 1$) развитие гравитационной неустойчивости возможно до наступления инерциальной фазы расширения, т. е. при $t < t_\Omega = t_0\Omega$, $z > z_\Omega = \Omega^{-1}$ (см. формулы (1.2.36) и (1.4.30)) *). Учитывая связь между космологическим временем и красным смещением (1.4.27), (1.4.28), рост возмущений в (3.1.22), (3.1.23) можно представить формулами

$$\left. \begin{aligned} \frac{\delta\rho}{\rho} \propto (1+z)^{-1}, \quad v \propto (1+z)^{-1}, \quad z > z_*; \\ \frac{\delta\rho}{\rho} \propto (1+z)^{-1}, \quad u \propto (1+z)^{-1/2}, \quad z < z_*. \end{aligned} \right\} \quad (3.1.31)$$

Примем мажорирующий закон $\delta\rho/\rho \propto (1+z)^{-2}$ и, считая возмущение данного масштаба непрерывно растущим (что также означает мажорирование), найдем, какова должна быть относительная амплитуда начального возмущения в некоторую отдаленную эпоху в прошлом, чтобы к

*) Точнее, неустойчивость прекращается при $z = 2/5\Omega^{-1}$ (Сюняев (1971), Зельдович и Новиков (1975)).

современной эпохе эта амплитуда достигла сравнимого с единицей значений. Выберем в качестве начальной эпохи, например, эпоху ядерной плотности, когда $\rho = \rho_N \approx \approx 10^{15} \text{ г/см}^3$, $z_N \approx 10^{15}$, и тогда получим для нее

$$\left(\frac{\delta\rho}{\rho}\right)_{z_N} \gtrsim (1 + z_N)^{-2} \approx 10^{-30}.$$

Возмущения с такой амплитудой являются весьма слабыми, но, однако, их можно считать и очень сильными, если сравнивать со статистическими флуктуациями в среде. Для последних, например, амплитуда возмущений плотности, захватывающая частиц газа,

$$\left(\frac{\delta\rho}{\rho}\right)_{st} \approx N^{-1/2}. \quad (3.1.32)$$

Подставим сюда в качестве N , скажем, число частиц в галактике ($\approx 10^{68}$); тогда

$$\left(\frac{\delta\rho}{\rho}\right)_{st} \approx 10^{-39},$$

что на много порядков меньше величины $(\delta\rho/\rho)_{z_N}$.

«Затравочные» возмущения для гравитационной неустойчивости должны быть, как мы видим, много выше уровня статистических флуктуаций в метагалактической среде на протяжении всей ее эволюции, которая вообще может изучаться на основании установленных физических законов. В этом смысле можно говорить о первичных возмущениях как о некоей структуре, «изначально» существующей во Вселенной.

Если экстраполировать формулы линейной теории возмущений на эпоху сверхъядерных плотностей, то можно, очевидно, определить момент, близкий к сингулярности, когда амплитуда необходимых потенциальных возмущений оказывается равной амплитуде статистических флуктуаций, вычисляемых по формуле (3.1.32). Но ни сама такая экстраполяция, ни результат, который при этом получается, не имеют, по-видимому, особого смысла. Остается лишь подозревать, что вопрос о природе исходных малых возмущений в изотропной Вселенной принадлежит к категории того же порядка, что и вопрос о природе самой космологической сингулярности.

§ 3.2. Вихревые движения

Вихри в расширяющейся Вселенной. Вихревые гидродинамические возмущения, не связанные с изменением плотности, можно понимать как локальные вращения вокруг осей различных направлений. Если пренебречь диссипативными процессами, то в изотропной среде момент количества движения для каждого из этих вращений сохраняется со временем:

$$I\omega = \text{const}, \quad (3.2.1)$$

где I — момент инерции, ω — угловая скорость. Для приближенной оценки можно положить

$$I \approx MR^2 \approx \rho R^5, \quad v \approx \omega R. \quad (3.2.2)$$

Здесь M — масса, охваченная вращательным движением масштаба R , а v — скорость движения.

В эпоху преобладания излучения

$$R \propto t^{1/2}, \quad \rho \propto t^{-2},$$

и потому масса, приходящаяся на данное число барионов, или масса в данном сопутствующем объеме, убывает со временем:

$$M \propto t^{-1/2}. \quad (3.2.3)$$

Тогда момент инерции

$$I \propto t^{1/2}, \quad (3.2.4)$$

а угловая и линейная скорости

$$\omega \propto t^{-1/2}, \quad v = \text{const}. \quad (3.2.5)$$

Хотя с течением времени возрастают размеры области вращения и убывает угловая скорость, линейная скорость остается неизменной.

При малом давлении в эпоху преобладания вещества

$$R \propto t^{2/3}, \quad \rho \propto t^{-2},$$

и потому

$$\left. \begin{aligned} M = \text{const}, \quad I \propto t^{1/3}, \\ \omega \propto t^{-4/3}, \quad v \propto t^{-2/3}. \end{aligned} \right\} \quad (3.2.6)$$

Значит, в этом случае при росте размеров области уменьшаются и угловая и линейная скорости.

Таким образом, в возмущениях вращательного типа скорости не нарастают. Следует, однако, упомянуть гипотетический случай так называемого «сверхжесткого» уравнения состояния (Зельдович (1961)), когда

$$\frac{1}{3} \varepsilon < p < \varepsilon. \quad (3.2.7)$$

Это уравнение состояния не исключено, по-видимому, общими принципами теории и, возможно, осуществляется для системы сильно взаимодействующих частиц. В предельном случае $p = \varepsilon$

$$R \propto t^{1/2}, \quad \rho \propto t^{-2},$$

и потому

$$\left. \begin{aligned} M \propto t^{-1}, \quad I \propto t^{-1/2} \\ \omega \propto t^{1/2}, \quad v \propto t^{1/2}, \end{aligned} \right\} \quad (3.2.8)$$

т. е. в «сверхжесткой» среде, если таковая возможна, угловая и линейная скорости вращения возрастали бы вместе с размерами области возмущения.

Вращательные возмущения обладают одной важной для космогонии особенностью. Хотя в этих возмущениях плотность и скорость не меняются со временем, они вызывают изменения метрики. Для ранней фазы расширения, когда $p = 1/3 \varepsilon$, теория Лифшица (1946) приводит к следующему соотношению для возмущения метрики:

$$h \approx \frac{v}{c} \frac{R}{ct}. \quad (3.2.9)$$

На этой фазе эволюции $R \propto t^{1/2}$, и для возмущенной метрики находим

$$h \propto t^{-1/2}, \quad (3.2.10)$$

так что $h \rightarrow \infty$ формально при $t \rightarrow 0$. Отсюда следует, что если на некоторой фазе расширения Метагалактики в ней существовали вращательные, вихревые возмущения, то при удалении в прошлое, т. е. при приближении к сингулярности, связанные с ними возмущения метрики должны были бы быть не малыми. В таком случае самые ранние фазы расширения — существенно не фридмановские; Метагалактика имела бы тогда более сложную метрическую структуру (Озерной и Чернин (1967)).

В гидродинамике вместо формулы (3.2.1) пользуются некоторым другим законом сохранения: сохранением циркуляции скорости, выражающейся через ее вихрь $\frac{1}{2} \text{rot } \mathbf{v}$. Рассмотрим сначала классическое уравнение движения Эйлера:

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \nabla \mathbf{v}) = - \frac{\nabla p}{\rho}. \quad (3.2.11)$$

Если давление зависит только от плотности (такое течение называют баротропным), то правая сторона этого уравнения равна градиенту некоторой функции, а уравнение для вихря имеет вид

$$\frac{d}{dt} \text{rot } \mathbf{v} = - [\mathbf{v} \text{rot rot } \mathbf{v}]. \quad (3.2.12)$$

В этом случае поток вихря через площадку, заполненную данными частицами («жидкая площадка»), не меняется при их движении:

$$\int \text{rot } \mathbf{v} d\sigma = \text{const.} \quad (3.2.13)$$

Действительно, по теореме Стокса

$$\int \text{rot } \mathbf{v} d\sigma = \frac{1}{2} \oint \mathbf{v} dl,$$

где стоящий справа интеграл по замкнутому «жидкому» контуру есть циркуляция скорости. Полная производная по времени от циркуляции есть

$$\frac{d}{dt} \oint \mathbf{v} dl = \oint \frac{d\mathbf{v}}{dt} dl + \oint \mathbf{v} \frac{d}{dt} (dl).$$

Второй интеграл справа обращается в нуль, так как под знаком интеграла стоит полная производная

$$\mathbf{v} \frac{d}{dt} (dl) = \mathbf{v} d\mathbf{v} = \frac{1}{2} d\mathbf{v}^2.$$

Первый же интеграл

$$\oint \frac{d\mathbf{v}}{dt} dl = - \oint \frac{\nabla p}{\rho} dl = - \oint \text{grad } W dl$$

равен нулю по теореме Стокса, с учетом того, что $\text{rot grad } W = 0$:

$$- \oint \text{grad } W dl = - \int \text{rot grad } W d\sigma = 0,$$

Следовательно, циркуляция скорости, а с ней и поток вихря не меняются вдоль траекторий частиц. Можно сказать, что вихрь вморожен в частицы и переносится вместе с ними.

Если $\text{rot } \mathbf{v}$ на данной площадке не зависит от координат (твердотельное вращение), то закон сохранения потока вихря можно записать в виде

$$\text{rot } \mathbf{v} \Sigma = \text{const}, \quad (3.2.14)$$

где Σ — площадь сечения, перпендикулярного к вихрю. Эта формула эквивалентна закону сохранения момента, с которого мы начинали, так как $\Sigma \propto R^2$. При растягивании площадки вихрь изменяется точно так же, как в формуле (3.2.6) для космологической среды с нулевым давлением.

При расширении среды с сохранением ее однородности предположение о независимости $\text{rot } \mathbf{v}$ от координат несущественно, так как зависимость всех фигурирующих здесь величин от времени одинакова во всех точках; в выражении $\int \text{rot } \mathbf{v} d\sigma$ оба множителя под знаком интеграла изменяются со временем только вследствие общего однородного расширения, так что временное поведение интеграла не зависит от того, изменяется ли $\text{rot } \mathbf{v}$ от точки к точке или нет (можно сказать, что дифференцирование по времени и интегрирование по площадке представляют собой коммутирующие операции).

Это классическое рассмотрение требует видоизменения в случае релятивистского уравнения состояния среды, когда ее давление не мало по сравнению с плотностью энергии, даже если макроскопические скорости по-прежнему малы по сравнению со скоростью света. Закон сохранения момента (3.2.1) остается при этом, разумеется, в силе, но из него уже следует не сохранение потока вихря, а сохранение величины $M \text{rot } \mathbf{v} \Sigma$. Другими словами, в случае ультрарелятивистской среды сохраняется поток величины

$$M \text{rot } \mathbf{v}, \quad (3.2.15)$$

называемой псевдовихрем (Франкль (1956)). Если масса не меняется со временем (что и имеет место при $p = 0$), то сохранение потока псевдовихря равносильно, очевидно, уравнению (3.2.14).

Рассмотрим еще одну особенность релятивистского вихревого движения. Пусть в однородном распределении

вещества имеется ограниченная область \mathcal{V} размера R , в которой помимо общего расширения существуют слабые вращательные, вихревые движения с полем скорости \mathbf{v}_K , не нарушающие однородности плотности. Этим движениям отвечает некоторый момент импульса, который мы обозначим K . Оценим возмущение, возникающее благодаря таким движениям, во внешней среде, окружающей указанную область.

Если разности гравитационного потенциала в интересующем нас объеме малы по сравнению с квадратом скорости света, то в течение ограниченного промежутка времени метрику можно считать галилеевой. Гравитационное поле описывается при этом ньютоновской теорией, так что в любой точке (или для любой частицы), находящейся на расстоянии r от центра, выбранного внутри рассматриваемой области, гравитационная сила, действующая на единичную массу, есть $(GM/r^2) \mathbf{n}$, где M — полная масса шара радиуса r , \mathbf{n} — орт радиус-вектора. В таком классическом приближении влияние вращательных движений на внешнюю среду отсутствует (если пренебрегать, как это предполагается, деформацией объема из-за вращения и вязкостью).

В следующем, постньютоновском приближении, которое мы и будем рассматривать, необходимо учитывать отклонения метрики от галилеевой, причем гравитационное силовое поле уже не будет центрально-симметрическим. Малые отклонения от центральной симметрии, обязанные слабым вихревым движениям, создают вне возмущающей области дополнительную силу, эквивалентную кориолисовой силе, появляющейся при вращении с угловой скоростью:

$$\boldsymbol{\omega}_K = \frac{G}{c^2 r^3} [K - 3\mathbf{n} (K\mathbf{n})]. \quad (3.2.16)$$

Это утверждение представляет собой простое обобщение результата, относящегося к слабому стационарному гравитационному полю в вакууме (см. Ландау и Лифшиц (1973)). Очевидно, что наличие вещества вне области движений изменяет в принятом приближении лишь гравитационную массу «источника»; именно, массой «источника» следует считать всю массу, заключенную внутри сферы радиуса r , как в решении для однородного расширения.

Неизменность во времени момента импульса K обеспечивает подобие рассматриваемой ситуации стационарной задаче. Что же касается массы M , то без ущерба для такого подобия она может и изменяться во времени (например, убывать при расширении, если среда является ультрарелятивистской).

Считая гравитационные релятивистские эффекты малыми, при вычислении эффекта, происходящего от вращательных движений, мы будем пренебрегать влиянием ньютоновости центрально-симметрического силового поля. Тогда уравнение движения некоторой частицы внешней среды есть

$$\frac{dv}{dt} = -\frac{GM}{r^2} n + 2[v\omega_K]. \quad (3.2.17)$$

Скорость v — это сумма хаббловской скорости общего расширения и искомой дополнительной скорости. Поскольку последняя мала по сравнению с первой, мы можем пренебрегать ею во втором слагаемом в правой части (3.2.17). Если выбрать сферическую систему координат и направить полярную ось вдоль момента K , то проекция этого уравнения на направление радиуса-вектора даст невозмущенное уравнение. Проекция на φ -направление есть

$$\frac{d}{dt} v_\varphi = -\frac{2G|K|}{c^2} \sin \vartheta \frac{HR}{r^3}.$$

Интегрируя, находим (Тиунов и Чернин (1971)):

$$v_\varphi \approx \frac{G|K|}{c^2} \frac{\sin \vartheta}{r^2} \quad (3.1.18)$$

(несущественную константу интегрирования полагаем равной нулю).

Дивергенция скорости (3.2.18) равна нулю и потому эта дополнительная скорость не вызывает нарушения однородности плотности.

Решение (3.2.18) показывает, что движения, создающие момент K , увлекают своим гравитационным полем во вращение внешние слои среды. Исходное вихревое движение, взаимодействуя гравитационно с потенциальным движением (расширением), порождает новое движение со скоростью (3.2.18). Величина эффекта гравитационного взаимодействия движения определяется в принятом приближении отношением ньютоновского потенциала GM/r

к квадрату скорости света. Например, в экваториальной плоскости вблизи границы вращающейся области ($r \sim R$) имеем по порядку величины:

$$K \approx MRv_K, \quad v_\varphi \approx v_K \frac{GM}{Rc^2} \approx v_K \left(\frac{R}{ct} \right)^2. \quad (3.2.19)$$

Как и в формуле для возмущений метрики, связанных с вихревыми движениями (3.2.9), в последнее выражение входит отношение размера возмущенной области к расстоянию до горизонта. Рассмотренный эффект квадратичен по этому (считающемуся малым) отношению. Так как расстояние до горизонта возрастает быстрее, чем все размеры и длины в расширяющемся мире, этот эффект ослабляется со временем. Он должен быть существен вблизи особенности (где для него требуется полное релятивистское, а не наше постньютоновское рассмотрение) и приводить вместе с другими эффектами такого рода к сильному нарушению однородности и изотропии, о котором говорилось выше.

Генерация магнитных полей. До рекомбинации метagalактической плазмы вещество и излучение совместно участвуют в гидродинамических движениях — в регулярном расширении и в наложенных на него возмущениях. Они представляют собой при этом как бы единую жидкость, и именно так мы и рассматривали их до сих пор. Это рассмотрение, очевидно, приближенно, но оно достаточно для анализа гидродинамических явлений. Гидродинамикой, однако, картина не исчерпывается полностью; как мы сейчас увидим, наличие в метagalактической среде нескольких компонент приводит к интересному электродинамическому эффекту — генерации магнитных полей вихрями. Для анализа этого эффекта требуется рассмотреть поведение различных компонент среды, охваченных вихревым движением, и оценить взаимодействие между ними в ходе общего космологического расширения.

Как мы уже знаем, вихревая скорость в «обычном» газе падает при расширении, а в ультрарелятивистской среде она остается неизменной. Если в какой-то момент смесь этих компонент имела в данном участке объема единую для них вихревую скорость, то затем по мере общего расширения скорости вращения компонент оказались бы различными.

В § 2.2 мы выяснили, что в дорекомбинационную эпоху электроны как бы заморожены в излучение, тогда как ионы связаны с излучением и электронами слабее. Поэтому можно говорить о двух компонентах среды — электрон-фотонной и ионной. Первая из них является ультрарелятивистской, когда излучение преобладает над веществом, а вторая — «обычной», когда температура среды ниже той, которая отвечает энергии покоя иона. Если бы компоненты вообще не взаимодействовали и момент вращения каждой из них сохранялся по отдельности, то согласно сказанному выше их вихревые вращательные движения привели бы к возникновению относительной скорости двух компонент.

Но обмен вращательным моментом в действительности происходит, и на ионы действует сила, стремящаяся увеличить их скорость до значения скорости релятивистской компоненты. Космологическое расширение уменьшает скорость ионной компоненты, а увлечение ионов электронами и излучением увеличивает ее. В нулевом приближении по отношению времени увлечения к космологическому возрасту обе скорости равны. В следующем приближении поведение обеих компонент среды не вполне совпадает, и скорость ионной компоненты всегда несколько меньше. Поэтому в метagalacticкой плазме электроны обгоняют ионы в направлении вращения, и появляется электрический ток в масштабе исходного гидродинамического движения. Этот ток генерирует магнитное поле \mathcal{H} , направленное по оси вращения (Харрисон (1971)).

Возникающее поле можно оценить из баланса сил, действующих на один протон:

$$m \frac{dv}{dt} = -mv \frac{dR}{dt} + e\mathcal{E}.$$

Здесь первый член описывает убывание вихревой скорости из-за расширения, а второй есть сила, обязанная электрическому полю \mathcal{E} , индуцируемому изменением магнитного потока:

$$\frac{d}{dt} \pi R^2 \mathcal{H} = -c 2\pi R \mathcal{E}.$$

Электрическое поле увлекает ионы, «подтягивая» их к электронам, и тем самым компенсирует убывание скорости

при расширении, так что полная сила в итоге равна нулю:

$$m\nu \frac{dR}{dt} = \frac{e}{Rc} \frac{d}{dt} R^2 \mathcal{H}.$$

Полагая для оценки $\frac{d}{dt} \sim \frac{1}{t}$, $\nu \sim \frac{R}{t}$, найдем

$$\mathcal{H} \sim \frac{mc}{et}. \quad (3.2.20)$$

Для времени рекомбинации $t \sim 10^{13}$ сек это дает поле $\sim 10^{-17}$ гс. Оно много меньше магнитного поля, наблюдаемого в плоской системе нашей Галактики ($\sim 10^{-6}$ гс), но такое поле могло бы, возможно, послужить «затравкой» и быть усиленным в процессах, действующих при образовании галактик или даже в современном их состоянии (Вайнштейн и Зельдович (1972)).

Рассмотренный электродинамический эффект представляет собою один конкретный механизм генерации магнитного поля. В общем случае магнитное поле, по-видимому, так или иначе всегда возникает при хаотических вихревых движениях проводящей среды, каковой является плазма. На это указывает в первую очередь (Ландау и Лифшиц (1957)) подобие уравнения для магнитного поля,

$$\frac{d\mathcal{H}}{dt} = - [\mathbf{v} \operatorname{rot} \mathcal{H}], \quad (3.2.21)$$

уравнению для вихря скорости (3.2.12). Магнитное поле, точно так же как и вихрь, заморожено в проводящую среду, и при деформации среды магнитный поток через данную материальную площадку сохраняется:

$$\mathcal{H}\Sigma = \text{const.} \quad (3.2.22)$$

При космологическом расширении такие замороженные магнитные поля убывают:

$$\mathcal{H} \propto R^{-2} \propto a^{-2} \propto (1+z)^2. \quad (3.2.23)$$

Соответствующая им плотность энергии,

$$\epsilon_{\mathcal{H}} = \frac{1}{8\pi} \mathcal{H}^2 \propto (1+z)^4, \quad (3.2.24)$$

изменяется при этом по тому же закону, что и плотность энергии излучения.

Магнитное поле создает в среде и давление, добавляющееся к давлению излучения и вещества. Благодаря этому магнитные поля могут влиять, в частности, на развитие в среде гравитационной неустойчивости. Если магнитное поле однородно, то среда, сжимающаяся или расширяющаяся вдоль его направления, не испытывает лоренцевой силы. Магнитное поле на такие движения среды не влияет, и потому джинсова длина гравитационной неустойчивости, определяемая балансом упругих и гравитационных сил, в этом случае не меняется. Иначе обстоит дело при поперечных к полю движениях среды. Силовые линии магнитного поля, которые вморожены в среду, сближаются или расходятся вместе с ней, и при этом проявляется магнитное давление, которое наряду с давлением газа и излучения препятствует сжатию. Это значит, что упругость среды возрастает, а вместе с ней возрастает и джинсова длина:

$$R_J^2 \approx \frac{p + p_{\mathcal{H}}}{G\rho^2} \approx \frac{p + \frac{1}{8\pi} \mathcal{H}_{\perp}^2}{G\rho^2} \quad (3.2.25)$$

(\mathcal{H}_{\perp} — поперечное движению вещества магнитное поле). Критические размеры, начиная с которых возмущения плотности могут возрасти со временем, оказываются в поперечном направлении большими, чем в продольном.

В случае хаотических магнитных полей такой анизотропии гравитационной неустойчивости нет. Если масштаб движений вещества много больше размера неоднородности поля, упругость среды увеличивается изотропным образом, и джинсова длина возрастает одинаково во всех направлениях (Чернин (1966)). Эффективное усредненное давление $p_{\mathcal{H}}$, создаваемое магнитным полем в больших масштабах, равно при этом $1/3 \epsilon_{\mathcal{H}}$. Действительно, магнитное поле создает положительное давление в двух поперечных к нему направлениях, равное $\frac{1}{8\pi} \mathcal{H}^2$ в каждом из них, и такое же по величине, но противоположное по знаку, натяжение силовых линий. Среднее по направлениям давление, т. е. среднее из этих трех величин, есть

$$\langle p_{\mathcal{H}} \rangle = \frac{1}{3} \frac{\mathcal{H}^2}{8\pi} = \frac{1}{3} \epsilon_{\mathcal{H}}. \quad (3.2.26)$$

Эффективное уравнение состояния для хаотического магнитного поля, рассматриваемого в масштабах, превосходящих размер его неоднородности, оказывается таким же, как уравнение состояния излучения. Другие вопросы, связанные с ролью магнитных полей в космологии, рассматриваются в работах Риса и Рейнгарда (1972), Рейнгарда (1972).

§ 3.3. Диссипативные процессы

Для всех рассмотренных нами возмущений существенным фактором развития являются диссипативные процессы — вязкость, теплопроводность, трение. В эпоху преобладания излучения ведущим механизмом взаимодействия между веществом и излучением является комптоновское рассеяние фотонов на свободных электронах. Оно и определяет в основном диссипативное затухание возмущений в горячей Вселенной. В этом параграфе мы оценим эффективность диссипации на различных фазах космологического расширения.

Лучистая вязкость. Наличие вязкости проявляется в возникновении дополнительной силы, вид которой может быть установлен из следующих простых соображений. Эта сила возникает, когда макроскопические скорости движения среды меняются от точки к точке. Вязкость стремится уничтожить различия в скорости и потому зависит от производных компонент скорости по координатам. В дальнейшем, как и во всей этой главе, нас будут интересовать малые возмущения, т. е. в термодинамическом смысле лишь малые отклонения от равновесия. Именно для таких малых возмущений применимо гидродинамическое приближение вязкости, теплопроводности и диффузии.

В случае вязкости это означает, что вязкая сила должна выражаться линейно через производные от скорости. Легко видеть, что эти производные должны быть второго порядка. Действительно, векторные дифференциальные операции первого порядка, примененные к скорости, а именно: $\text{rot } \boldsymbol{v}$ и $\text{div } \boldsymbol{v}$, приводят соответственно к псевдовектору (аксиальному вектору) и скаляру; между тем, сила есть полярный вектор, и потому не может выражаться только через эти величины. Двукратное дифференцирование приводит к двум полярным векторам: $\Delta \boldsymbol{v}$ и $\text{grad div } \boldsymbol{v}$.

Производными более высокого порядка мы должны пренебречь в силу сделанного предположения о малости возмущений; в данном случае это означает, что скорость изменяется медленно от точки к точке, т. е. ее производные по координатам малы, и чем выше порядок производных, тем они меньше. Таким образом, вязкая сила представляет собой в общем случае линейную комбинацию векторов Δv и $\text{grad div } v$:

$$F = \alpha \Delta v + \beta \text{grad div } v, \quad (3.3.1)$$

где α, β — некоторые постоянные, зависящие от свойств среды.

Физический смысл обоих слагаемых в выражении для вязкой силы станет ясным, если мы запишем (3.3.1) в виде

$$F_i = \frac{\partial}{\partial x_k} \sigma_{ik}, \quad (3.3.2)$$

где по повторяющимся индексам, как всегда, предполагается суммирование. Тензор σ_{ik} есть так называемый тензор вязких напряжений:

$$\sigma_{ik} = \eta \left(\frac{\partial v_i}{\partial x_k} + \frac{\partial v_k}{\partial x_i} - \frac{2}{3} \delta_{ik} \frac{\partial v_l}{\partial x_l} \right) + \xi \delta_{ik} \frac{\partial v_l}{\partial x_l}. \quad (3.3.3)$$

Формулы (3.3.2), (3.3.3) получаются тождественным преобразованием формулы (3.3.1); при этом мы учли, что

$$\frac{\partial v_l}{\partial x_l} \equiv \text{div } v, \quad \frac{\partial^2 v_i}{\partial x_k^2} \equiv \Delta v_i, \quad \delta_{ik} = \begin{cases} 1, & i = k, \\ 0, & i \neq k. \end{cases}$$

Коэффициенты η и ξ связаны с α и β соотношениями

$$\alpha = \eta, \quad \beta = \xi + \frac{1}{3} \eta.$$

Слагаемое в тензоре σ_{ik} , пропорциональное η , характеризует ту часть вязких напряжений, которая связана с трением при скольжении одних слоев жидкости относительно других; оно отлично от нуля лишь в том случае, когда производные $\partial v_i / \partial x_k$ при $i \neq k$ не равны нулю. Если же имеет место только общее расширение или сжатие жидкости, то отличны от нуля лишь производные вида $\partial x_i / \partial x_k$, при $i = k$; тогда слагаемое, пропорциональное η , равно

нулю, и в тензоре σ_i остается лишь второе слагаемое, пропорциональное $\xi + \frac{1}{3}\eta$. Эти два проявления вязкости носят название соответственно первой и второй вязкости. Наглядный смысл второй вязкости будет пояснен ниже.

Если вязкую силу относить к единице объема, то соответствующие коэффициенты η и ξ называют коэффициентами первой и второй вязкости; если же относить силу к единице массы, т. е. разделить ее на плотность ρ , то выражения

$$\nu = \frac{\eta}{\rho}, \quad \zeta = \frac{\xi}{\rho}$$

называют кинематическими коэффициентами первой и второй вязкости.

Величины η и ζ всегда положительны. Это следует из условия возрастания энтропии. Действительно, энергия макроскопического движения данного элемента объема уменьшается под влиянием вязкой силы; это уменьшение в единицу времени равно

$$\frac{\partial E}{\partial t} = \int F \mathbf{v} dV = \int \mathbf{v} \left[\eta \Delta \mathbf{v} + \left(\xi + \frac{1}{3} \eta \right) \text{grad div } \mathbf{v} \right] dV.$$

Интегрируя по частям, получим

$$\frac{\partial E}{\partial t} = - \int dV \left[\eta \left(\frac{\partial v_i}{\partial x_k} \right)^2 + \left(\xi + \frac{1}{3} \eta \right) (\text{div } \mathbf{v})^2 \right].$$

Отсюда видно, что $\partial E / \partial t < 0$, если η и ξ положительны. Энергия макроскопического движения превращается в тепловую энергию, а энтропия при этом возрастает.

Дадим порядковую оценку коэффициентов вязкости и начнем с первой вязкости. При наличии в среде макроскопических движений частицы среды имеют, кроме скорости беспорядочного теплового движения u , еще и направленную скорость \mathbf{v} . Эта последняя меняется от точки к точке и на протяжении длины свободного пробега частицы l изменяется на величину

$$\delta \mathbf{v} \approx l \frac{\partial \mathbf{v}}{\partial x}.$$

В соответствии с этим частица, перемещаясь в процессе теплового движения на длину свободного пробега, переносит с собой «избыточный» импульс направленного движения $m \delta \mathbf{v}$. Этот процесс происходит со средней скоростью

и, так что частицы единицы объема создают плотность потока импульса

$$-\rho_i \delta v \approx -\rho_i l \frac{\partial v}{\partial x}. \quad (3.3.4)$$

Перенос импульса направленного движения происходит в каждом единичном акте случайным образом; однако в среднем его тенденция такова, что различие направленных скоростей соседних элементов объема жидкости уменьшается. Поток импульса направлен в сторону меньших значений плотности импульса, и на это указывает знак минус в формуле (3.3.4). Такое сглаживание различий в направленных скоростях и есть вязкость.

В действительности плотность потока импульса есть тензор, так как вместо производной $\partial v / \partial x$, которую мы писали для простоты, надо писать различные комбинации величин $\partial v_i / \partial x_k$. Если поток импульса, входящий в некоторый объем, равен потоку, выходящему из него, то импульс этого объема остается неизменным, и, значит, на него не действует никакая сила. Сила появляется лишь в том случае, если входящий и выходящий потоки отличаются друг от друга, т. е. если плотность потока импульса изменяется от точки к точке. Сила, действующая на единицу объема, пропорциональна производной от плотности потока импульса по координатам, и, следовательно, связь между силой и плотностью потока импульса точно такая же, как связь между силой F и введенным выше тензором напряжений σ_{ik} . Это значит, что тензор σ_{ik} и есть не что иное, как тензор плотности потока импульса, взятый с обратным знаком. Сила, действующая на единицу объема, есть изменение импульса этого объема в единицу времени, и, следовательно, уравнение (3.3.3), как мы теперь видим, выражает то обстоятельство, что изменение со временем импульса данного объема определяется разностью входящего и выходящего потоков импульса. Уравнение (3.3.3) есть, по существу, закон сохранения импульса, записанный для единицы объема. Сравнивая (3.3.3) и (3.3.4), мы можем дать приближенное, с точностью до численного коэффициента порядка единицы, выражение для коэффициента вязкости (первой):

$$\eta \approx \rho_i l. \quad (3.3.5)$$

Вводя время свободного пробега для передачи импульса,

τ_p , запишем η в виде

$$\eta \approx \rho u^2 \tau_p.$$

Вторая вязкость имеет более сложную природу. Она возникает в том случае, когда газ, испытывающий изменение объема при макроскопических движениях, состоит из частиц с внутренними степенями свободы (например, колебаний или вращений в молекулах) или если он состоит из двух или более компонент с разными показателями адиабаты. Последнее и будет интересоваться нас в дальнейшем при рассмотрении смеси излучения и вещества в космологии. В этом случае при расширении среды излучение охлаждается меньше, чем газ, так как их показатели адиабаты равны соответственно $4/3$ и $5/3$. Лишь через некоторое время, которое есть время релаксации, их температура сравнивается.

Сила, обусловленная второй вязкостью, есть не что иное, как изменение упругой силы, связанное с отклонением давления от его равновесного значения при быстром изменении объема среды. Действительно, имея в виду выражение (3.3.3), мы можем написать полную силу в виде

$$-\nabla p_{\text{eff}} \approx -\nabla (p - \xi \operatorname{div} \mathbf{v}).$$

Так как по уравнению неразрывности

$$\operatorname{div} \mathbf{v} = -\frac{1}{\rho} \frac{\partial \rho}{\partial t}$$

(в линейном приближении), то «эффективное» давление

$$p_{\text{eff}} = p + \xi \frac{1}{\rho} \frac{\partial \rho}{\partial t}.$$

Эффективное давление p_{eff} отличается от «истинного» давления p , связанного термодинамически с мгновенными значениями плотности и температуры среды, тем больше, чем быстрее изменяется объем. То обстоятельство, что неравновесная добавка к давлению, $p_{\text{eff}} - p$, пропорциональна $\operatorname{div} \mathbf{v}$, естественно: единственный скаляр, выражающийся через первые производные от скорости, есть $\operatorname{div} \mathbf{v}$.

В эпоху преобладания излучения, когда $\rho \ll \rho_r$, вторая вязкость может приводить лишь к слабым эффектам. Действительно, для однокомпонентной жидкости коэффициент $\xi = 0$, а при малой примеси второй компоненты он

должен определяться малым отношением $\rho u^2/\rho_e c^2$ и потому много меньше коэффициента первой вязкости. Количественное исследование слабых эффектов второй вязкости в космологии читатель может найти в книге Вейнберга (1972).

Применим полученные оценки к условиям в метагалактической среде на начальных стадиях расширения, когда вещество ионизовано и взаимодействует с излучением, плотность энергии которого не меньше плотности вещества. Такая среда является релятивистской, так как для основной ее компоненты — излучения — скорость u микроскопического хаотического движения равна c . Уравнения гидродинамики в этом случае отличаются от классических уравнений Навье — Стокса (Ландау и Лифшиц (1953)). Однако при малых макроскопических скоростях направленных движений, а именно этот случай и будет нас интересовать, такое отличие не приводит к изменению порядковых оценок рассматриваемых нами величин.

Метагалактическая среда содержит, кроме излучения, еще две компоненты — электроны и ионы. В соответствии с этим выражения для коэффициентов вязкости должны представлять собой сумму трех членов. Для первой вязкости имеем

$$\eta = \rho_r c l_r + \rho_e u_e l_e + \rho_i u_i l_i,$$

где значки i , e , r относятся соответственно к ионам, электронам и излучению, а величины l_i , l_e , l_r — длины пробега для передачи импульса. Частица каждой компоненты может сталкиваться как с частицами того же типа, так и с частицами других компонент. Например, для электронов можно говорить о длинах

$$l_{ee} \approx u_{ee} \tau_{ee} \approx u_e \tau_{ee}, \quad (3.3.6)$$

$$l_{ei} \approx u_{ei} \tau_{ei} \approx u_e \tau_{ei}, \quad (3.3.7)$$

$$l_{er} \approx u_{er} \tau_{er} \approx c \tau_{er}. \quad (3.3.8)$$

Здесь величины u_{ee} , u_{ei} , u_{er} — относительные скорости беспорядочного движения электрона по отношению к другим электронам, ионам и фотонам, а τ_{ee} , τ_{ei} , τ_{er} — соответствующие времена релаксации по импульсу. Так как изменение импульса электрона зависит от полного числа столкновений, то очевидно, что для определения полного

изменения импульса надо складывать не времена τ , а обратные им частоты столкновений, т. е. величины, обратные длинам свободного пробега. Поэтому длина свободного пробега для изменения импульса электрона определяется выражением

$$\frac{1}{l_e} = \frac{1}{l_{ee}} + \frac{1}{l_{ei}} + \frac{1}{l_{er}}. \quad (3.3.9)$$

Длины l_{ee} , l_{ei} , l_{er} отличаются друг от друга значительно, и потому длина l_e определяется наименьшей из этих величин. Используя связь между длиной свободного пробега и временем релаксации и учитывая выражения для τ_{ee} , τ_{ei} , τ_{er} , указанные в § 2.2, найдем для интересующих нас условий:

$$l_{ee} \approx l_{ei} \approx \frac{1}{\sigma_0 n}, \quad (3.3.10)$$

$$l_{er} \approx \frac{m_e}{\sigma_T \rho_r}. \quad (3.3.11)$$

(Напомним, что σ_0 , σ_T — кулоновское и томсоновское сечения; в рассматриваемую эпоху тормозные процессы, как показано в гл. 2, были несущественны.) Отношение последней длины к первым двум,

$$\frac{l_{er}}{l_{ee}} \approx \frac{l_{er}}{l_{ei}} \approx \frac{\sigma_0}{\sigma_T} \frac{\rho}{\rho_r} \frac{m_e}{m} \approx 2 \cdot 10^{21} \Omega (1+z)^{-3}, \quad (3.3.12)$$

больше единицы при $z < 10^7$. Для таких красных смещений $l_{ee} \approx l_e$ и потому

$$\eta_e \approx \rho_e \mu_e l_{ee} \approx 10^{-17} (1+z)^{1/2} \text{ г/см} \cdot \text{сек}. \quad (3.3.13)$$

Оценим теперь вязкость ионов. Длина их свободного пробега определяется выражением

$$\frac{1}{l_i} = \frac{1}{l_{ii}} + \frac{1}{l_{ie}} + \frac{1}{l_{ir}}. \quad (3.3.14)$$

Очевидно, что $l_{ie} \gg l_{ii}$, так как массивные частицы мало меняют свой импульс при столкновении с легкими. Далее, если для электронов обмен импульсом с фотонами был менее существен, чем обмен друг с другом при $z < 10^7$, то для ионов это тем более выполняется, так как их собственное взаимодействие с фотонами слабее. Поэтому

$l_i \approx l_{ii} \approx l_{ee} u_i$ и тогда

$$\eta_i \approx \rho_i u_i l_{ii} \approx \left(\frac{m^3}{m_e} \right)^{1/2} \approx 4 \cdot 10^{-16} (1+z)^{3/2} \text{ г/см} \cdot \text{сек.} \quad (3.3.15)$$

Остается оценить вязкость излучения или, как говорят, лучистую вязкость. Фотоны практически не взаимодействуют друг с другом и рассеиваются преимущественно на электронах. Соответствующая вязкость

$$\eta_r \approx \rho_r c l_{re} \approx \frac{\rho_r}{\rho} \frac{m c}{\sigma_T} \approx 10^7 \Omega^{-1} (1+z) \text{ г/см} \cdot \text{сек.} \quad (3.3.16)$$

Отношение лучистой вязкости к ионной,

$$\frac{\eta_r}{\eta_i} \approx 10^{23} \Omega^{-1} (1+z)^{-3/2}, \quad (3.3.17)$$

больше единицы при всех красных смещениях, отвечающих нерелятивистской стадии расширения, когда $kT \ll \ll mc^2$.

Таким образом, наиболее эффективной оказывается вязкость, обусловленная фотонами. Точное, а не порядковое, выражение для коэффициента лучистой вязкости отличается от полученного нами только лишь численным коэффициентом:

$$\eta_r = \frac{4}{15} \rho_r c l_{re}. \quad (3.3.18)$$

Для оценки характерного времени действия вязкости сравним по порядку величины соответствующие члены в уравнениях гидродинамики. Ограничимся случаем первой вязкости и приближением, линейным по гидродинамической скорости. Тогда изменение скорости, связанное с действием вязкости, описывается уравнением

$$\frac{\partial v}{\partial t} = \nu \Delta v.$$

Если скорость изменяется на величину, сравнимую с ее начальным значением, за время τ_v , то $\partial v / \partial t \sim v / \tau_v$. Если это изменение затрагивает относительные скорости элементов среды, разделенных расстоянием, λ , то $\Delta v \sim v / \lambda^2$.

Тогда уравнение движения приводит к соотношению:

$$\tau_v \approx \frac{\lambda^2}{v},$$

т. е. время, за которое вязкость способна сгладить различие скоростей и погасить относительное движение в масштабе λ , пропорционально квадрату этого масштаба. В расширяющейся метagalактической среде изменение макроскопической скорости пекулярных движений вследствие вязкости происходит быстрее, чем ее изменение вследствие общего расширения среды (или же гравитационной неустойчивости), если

$$\tau_v < t. \quad (3.3.19)$$

Следовательно, к данному моменту времени t гидродинамические движения, наложенные на общее космологическое расширение, успевают затухнуть во всех масштабах, меньших некоторого характерного:

$$\lambda < \lambda_v \approx (vt)^{1/2}. \quad (3.3.20)$$

Подставляя выражение для $v = \frac{\eta_r}{\rho_r + \rho}$, находим для эпохи преобладания излучения:

$$\lambda_v \approx 10^{30} \Omega^{-3/4} (1+z)^{-5/2} \text{ см}. \quad (3.3.21)$$

В ходе космологического расширения эта характерная длина возрастает быстрее, чем масштабный фактор $a(t) \propto z^{-1}$, пропорционально которому изменяются масштабы гидродинамических движений. Поэтому действие вязкости становится все более значительным и охватывает со временем все бóльшие масштабы. В конце эпохи преобладания излучения — если до этого газ еще не рекомбинировал — наименьший из масштабов, в которых еще возможны пекулярные гидродинамические движения,

$$\lambda_v(t_*) \approx 10^{20} \Omega^{-13/4} \text{ см}, \quad (3.3.22)$$

содержит массу вещества

$$M_v(t_*) \approx 2 \cdot 10^9 \Omega^{-6} M_\odot. \quad (3.3.23)$$

Если рекомбинация происходит еще в эпоху преобладания излучения (что, согласно формулам § 1.2, возможно при $\Omega < 0,14$), то наименьший «выживающий» масштаб

движений,

$$\lambda_v(t_{\text{rec}}) \approx 10^{22} \Omega^{-3/4} \text{ см} \quad (3.3.24)$$

отвечает массе

$$M_v(t_{\text{rec}}) \approx 5 \cdot 10^{12} \Omega^{-5/4} M_{\odot} \quad (3.3.25)$$

Если $\Omega > 0,14$ и рекомбинация происходит уже в эпоху преобладания вещества, то для $z_* > z > z_{\text{rec}}$

$$v = \frac{\rho_r}{\rho} \frac{mc}{\sigma_{\text{TR}}} \approx 10^{36} \Omega^{-2} (1+z)^{-2} \text{ см}^2/\text{сек} \quad (3.3.26)$$

В соответствии с этим

$$\lambda_v(t_{\text{rec}}) \approx 6 \cdot 10^{21} \Omega^{-1} \text{ см}, \quad (3.3.27)$$

$$M_v(t_{\text{rec}}) \approx 10^{12} \Omega^{-2} M_{\odot} \quad (3.3.28)$$

Соотношения (3.3.20) — (3.3.28) относятся, строго говоря, только к вихревым движениям, так как при потенциальных движениях нужно учитывать еще и теплопроводность (см. ниже).

Приближенные оценки для затухания вихрей были впервые даны в работе Озерного и Чернина (1967). В позднейших более точных вычислениях были учтены, во-первых, численные коэффициенты, которые в порядковых оценках отбрасывались, а, во-вторых, то обстоятельство, что переход от ионизованного состояния к нейтральному происходит в веществе не скачком*). Эти вычисления, сделанные Пиблсом и Ю (1970) и Чибисовым (1972), качественно подтверждают указанные выводы и дают близкую величину массы в наименьшем «выживающем» масштабе вихревых возмущений:

$$M_v(t_{\text{rec}}) \approx 5 \cdot 10^{11} \Omega^{-7/4} M_{\odot} \quad (3.3.29)$$

Рассмотрение диссипативных процессов приводит к важному для космогонии выводу. Первичные возмущения плотности и скорости метагалактической среды (которые существовали в ней, так сказать, с «самого начала») имели, по-видимому, весьма широкий спектр масштабов, но даль-

*) Последнее приводит, в частности, к тому, что после «просветления» данного масштаба движения становится на некоторое время существенным трение об излучение.

нейшая судьба возмущений существенно зависела от масштаба. Диссипативные процессы быстрее всего действовали в наименьших масштабах; при этом эффективный масштаб затухания увеличивался со временем, и к моменту рекомбинации возмущения масштаба галактик $\approx (M \approx 10^8 - 10^{12} M_{\odot})$ оказались погашенными. Могли сохраниться лишь возмущения большего масштаба — масштаба групп галактик, их скоплений и сверхскоплений. Отдельные галактики должны были формироваться, таким образом, из иных возмущений, имевших первоначально иную природу (энтропийные возмущения).

Сделанные выше оценки относились к эпохе рекомбинации метагалактической плазмы. После рекомбинации взаимодействие газа с излучением прекратилось и в газе действовала уже не лучистая, а обычная корпускулярная вязкость. Но лучистая вязкость могла возникнуть вновь, если имел место разогрев газа, сопровождавшийся его ионизацией. Ее действие продолжалось до тех пор, пока в масштабе, превышающем длину свободного пробега фотонов l_{re} , характерное время вязкого затухания τ_v было меньше космологического времени t . Этому условию отвечает неравенство

$$z > z_f \approx 100$$

(см. рис. 8). Если вторичный разогрев приводил к ионизации достаточно больших областей метагалактической среды, то диссипативные процессы могли погасить возмущения в огромных масштабах, вплоть до масштаба ячейки однородности: $M_v(t_f) \sim 10^{16} - 10^{17} M_{\odot}$.

Вязкость же нейтрального газа в эпоху $z < z_{rec}$ не существенна для возмущений интересующих нас масштабов. Действительно, коэффициент вязкости в этом случае есть

$$\nu \approx ul,$$

где

$$l \approx \frac{1}{n\sigma_H} \approx \frac{m}{\rho\sigma_H} \approx 10^{20}\Omega^{-1}(1+z)^{-3} \text{ см}; \quad (3.3.30)$$

$\sigma_H \approx 10^{-15} \text{ см}^2$ — газокинетическое сечение для столкновений нейтральных атомов. Характерный масштаб затухания,

$$\lambda_v \approx (\nu t)^{1/2} \approx 5 \cdot 10^{20}\Omega^{-1/2}(1+z)^{-5/4} \text{ см}, \quad (3.3.31)$$

отвечает очень малым массам

$$M_v \approx \Omega^{-1/2} (1 + z)^{-3/2} M_\odot, \quad (3.3.32)$$

и уже для масс, превышающих массу Солнца, газовая вязкость не существенна. Эта оценка сделана для случая, когда температура газа уменьшалась после рекомбинации в соответствии с показателем адиабаты $\gamma = 5/3$, т. е. $T \propto (1 + z)^2$.

Лучистая теплопроводность. В эпоху преобладания излучения перенос тепла, как и перенос импульса, осуществляется в подавляющей степени фотонами и потому говорят о лучистой теплопроводности. Перенос тепловой энергии в явлении теплопроводности и перенос импульса в явлении вязкости представляют собой процессы диффузионного характера: они являются результатом случайных и беспорядочных перемещений носителей энергии и импульса, — в нашем случае фотонов. При наличии градиента температуры ∇T средняя энергия фотонов, находящихся на расстоянии l_{re} длины свободного пробега, различается на величину

$$\delta T \approx l_{re} \nabla T. \quad (3.3.33)$$

Энергия переносится при столкновении фотонов с электронами в случайных направлениях. Однако в результате этих случайных событий возникает направленный поток тепловой энергии, стремящийся выровнять температуры. Энтропия при этом должна, очевидно, возрасти. Плотность потока энергии

$$w \approx n_r c k \delta T, \quad (3.3.34)$$

где n_r — концентрация фотонов. Величина n_r для термодинамически равновесного излучения,

$$n_r \approx \frac{\epsilon_r}{kT},$$

пропорциональна (см. § 2.3) теплоемкости единицы объема излучения

$$c_r \approx 10 k n_r. \quad (3.3.35)$$

Тогда, подставляя n_r и δT в соотношение для w , получаем

$$w \approx -n_r c l_{re} k \nabla T = -\kappa \nabla T,$$

где

$$\kappa \approx k n_r c l_r$$

есть коэффициент теплопроводности. Как и все потоки случайного теплового характера, поток тепла пропорционален градиенту температуры, т. е. той величины, различие которой в разных точках он стремится уничтожить. Эта пропорциональность имеет место при не слишком больших градиентах; когда же градиент настолько велик, что изменение температуры на длине свободного пробега сравнимо с самой температурой, т. е. $\delta T \sim T$, это простое соотношение уже неприменимо и в выражении для w должна появиться, кроме первой степени ∇T , еще $(\nabla T)^2$ — величина, представляющая собой вектор, содержащий следующую низшую степень градиента температуры.

Уравнение для изменения температуры со временем, — уравнение теплопроводности, — следует из закона сохранения энергии:

$$\frac{\partial \varepsilon}{\partial t} = - \operatorname{div} w, \quad (3.3.36)$$

где ε — плотность внутренней энергии, или, в другом виде,

$$c_r \frac{\partial T}{\partial t} = \operatorname{div} \kappa \nabla T.$$

В линейном по ∇T приближении мы должны пренебречь зависимостью κ от температуры. Тогда, вынося κ за знак дифференцирования, получим

$$\frac{\partial T}{\partial t} = \frac{\kappa}{c_r} \Delta T. \quad (3.3.37)$$

Это и есть уравнение теплопроводности, а величина κ/c_r называется температуропроводностью. По форме это уравнение аналогично уравнению Навье — Стокса (в линейном по скорости приближении) при отсутствии изменения плотности. Оценка характерного времени выравнивания температуры делается поэтому точно так же, как и в случае первой вязкости, и приводит к выражению

$$\tau_\kappa \approx \frac{\lambda^2}{\kappa/c_r} \approx \frac{\lambda^2}{cl_{re}}. \quad (3.3.38)$$

Таким образом, время выравнивания температур посредством лучистой теплопроводности по порядку величины совпадает с характерным временем лучистой вязкости τ_ν в том же масштабе.

Теплопроводность не действует на вихревые движения; она существенна только для потенциальных движений, которые в метagalacticкой среде с преобладанием излучения непосредственно связаны с возмущениями температуры. Действительно, если в некоторой области плотность энергии оказывается повышенной, то, согласно общим соотношениям для термодинамически равновесного (в данном случае локально) излучения, в ней также повышена температура и концентрация фотонов. Диффузия фотонов из области их повышенной концентрации в окружающую среду приводит к теплопроводности, к уменьшению плотности в области возмущения. По этой причине потенциальные движения затухают быстрее, чем вихревые. Точный расчет, сделанный в работах Силка (1968), Пиблса и Ю (1970) и Чибисова (1972), приводит к следующей оценке для наименьшего «выживающего» масштаба потенциальных движений:

$$M_x (t_{\text{rec}}) \approx 5 \cdot 10^{13} \Omega^{-1/2} M_{\odot}. \quad (3.3.39)$$

Заметим, что к затуханию возмущений, связанных с изменением плотности среды, ведет также и вторая вязкость. Однако вследствие малости коэффициента второй вязкости ξ по сравнению с коэффициентом первой вязкости η , а следовательно, и с величиной $\rho_r \kappa / c_r$, мы можем здесь не учитывать вторую вязкость.

Диффузия плазмы на фоне изотропного излучения. Помимо возмущений, в которых вещество и излучение ведут себя как единая жидкость, в метagalacticкой среде возможны, как мы говорили в § 3.1, возмущения состава. Нас будет здесь интересовать случай, когда плазма распределена неравномерно на фоне полностью однородного излучения. Согласно общим законам термодинамики, в таком состоянии среды должен появиться поток плазмы, стремящийся выровнять различия в ее плотности и концентрации. Как и в картине, рассмотренной в двух предыдущих разделах, можно ожидать, что этот поток пропорционален градиенту концентрации частиц плазмы, т. е. плотность потока

$$j = - D \nabla n. \quad (3.3.40)$$

Знак минус, как в формуле (3.3.36), указывает на то, что поток направлен в ту сторону, куда концентрация умень-

шается; величина D есть коэффициент диффузии. Концентрация n , вследствие квазинейтральности плазмы, одинакова для электронов и ионов. Коэффициент D можно оценить примерно так же, как мы оценивали диссипативные коэффициенты в предыдущих параграфах.

При наличии градиента концентрации ∇n на расстоянии длины свободного пробега l имеется различие в концентрациях δn , равное $l\nabla n$. Эти «избыточные» (или «недостающие») частицы испытывают случайные столкновения, приводящие к их беспорядочному движению со скоростями, соответствующими их температуре. Однако результатом беспорядочного движения должен быть, как мы уже знаем, направленный процесс: перенос избытка концентрации со средней тепловой скоростью u в том направлении, где концентрация меньше. Значит, плотность потока

$$j = -u\delta n = -ul\nabla n. \quad (3.3.41)$$

Сравнивая (3.3.40) и (3.3.41), находим

$$D \approx ul, \quad (3.3.42)$$

и остается определить среднюю длину свободного пробега l .

Наибольшее сопротивление при движении на фоне излучения испытывают электроны, и, значит, их столкновения с фотонами, сопровождающиеся обменом импульсом, и ограничивают перемещения плазмы. Отличие рассматриваемой сейчас ситуации от той, которая рассматривалась при анализе трения плазмы об излучение (§ 2.2), заключается в том, что там мы рассматривали гидродинамические движения плазмы в присутствии излучения, а здесь речь идет о беспорядочном микроскопическом движении частиц плазмы. Однако как там, так и здесь, элементарный акт, ограничивающий движение частиц плазмы, представляет собой их столкновение с фотонами, сопровождающееся обменом импульсом. Поэтому средняя длина пробега имеет в обоих случаях одинаковое значение:

$$l \approx u\tau_f \approx u \frac{mc}{\sigma_T \epsilon_r}. \quad (3.3.43)$$

Напомним, что в силу квазинейтральности u есть тепловая скорость ионов, а не электронов. Уравнение диффузии

имеет вид

$$\frac{\partial n}{\partial t} = - \operatorname{div} j = - D \Delta n. \quad (3.3.44)$$

Мы вынесли D за знак дифференцирования, так как считаем ∇n величиной малой, и потому величину D следует вычислять по невозмущенным значениям u , ε_r . Полученное уравнение аналогично уравнениям вязкого переноса и теплопроводности, что дает нам возможность написать сразу выражение для характерного времени выравнивания концентрации в масштабе λ :

$$\tau_D \approx \frac{\lambda^2}{D}. \quad (3.3.45)$$

Значит, за космологическое время t концентрация плазмы сглаживается лишь в масштабах, не превышающих величины

$$\lambda_D \approx (Dt)^{1/2}. \quad (3.3.46)$$

Выражая D и t через космологические параметры, находим

$$\lambda_D \approx 6 \cdot 10^{21} (1+z)^{-2} \text{ см}; z > z_*. \quad (3.3.47)$$

В масштабе λ_D содержится масса

$$M_D \approx 2 \cdot 10^{19} \Omega (1+z)^{-3} M_\odot; z > z_*. \quad (3.3.48)$$

К концу эпохи преобладания излучения возмущения состава затухают лишь в ничтожно малых масштабах, не превышающих по массе

$$M_D(z_*) \approx 10\Omega M_\odot. \quad (3.3.49)$$

Мы приходим, таким образом, к выводу, что возмущения состава метагалактической среды — это единственный тип возмущений, который мог противостоять диссипативным процессам в масштабах галактик (и даже отдельных звезд) в течение всей эпохи преобладания излучения.

§ 3.4. Возмущения и реликтовый фон

Если в прошлом Метагалактики имелись совместные движения вещества и излучения, наложенные на общее космологическое расширение, то фон реликтового излучения в современную эпоху не может быть совершенно однородным и изотропным. Он должен хранить какие-то следы

этих возмущений в виде вариаций температуры и интенсивности излучения в различных направлениях. Наблюдения, как мы говорили в § 1.1, не обнаруживают никаких вариаций такого рода с точностью, достижимой современными инструментами. Эта точность служит верхним наблюдательным пределом амплитуды вариаций температуры фона в угловых масштабах, соответствующих галактикам и их скоплениям (Карпентер, Гулкис и Сато (1973), Парийский (1973), Станкевич (1974)):

$$\frac{\delta T}{T} < (2 \div 7) \cdot 10^{-4}. \quad (3.4.1)$$

Ограничение является весьма сильным; оно дает исключительно важную информацию о возможных отклонениях от полной однородности и изотропии в ту эпоху, когда излучение перестало взаимодействовать с веществом. После этого излучение свободно распространяется и приходит к наблюдателю, сохраняя картину, существовавшую на «стенке последнего рассеяния». В отсутствие вторичной ионизации плазмы ограничение (3.4.1) переносится, таким образом, на амплитуду относительных возмущений температуры излучения в эпоху рекомбинации.

Возмущения температуры связаны с совместными гидродинамическими движениями плазмы и излучения — потенциальными и вихревыми возмущениями. Рассмотрим сначала потенциальные движения. В этом случае возмущения температуры определяются тремя эффектами.

Во-первых, это эффект Доплера, происходящий от движения данного макроскопического объема плазмы и излучения относительно сопутствующей регулярному расширению системы отсчета, т. е. относительно изотропного фона излучения. Внутри этого объема в системе отсчета, движущейся вместе с ним, имеется термодинамическое равновесие, так что в нем имеет место распределение Планка для фотонов. Но в системе изотропного фона это будет уже распределение Планка с дрейфом:

$$\left[\exp \left(\frac{cq + qv \cos \vartheta}{kT} \right) - 1 \right]^{-1}, \quad (3.4.2)$$

где ϑ — угол между направлением волнового вектора фотона q и скоростью движения объема v относительно системы отсчета, связанной с фоном. Распределение излучения по

углу должно быть анизотропно. Так как $\frac{v}{c} \cos \vartheta \ll 1$, то с достаточной точностью мы можем перенести множитель $\left(1 + \frac{v}{c} \cos \vartheta\right)$ из числителя в экспоненте в знаменатель, написав его в виде:

$$\frac{cq}{kT \left(1 - \frac{v}{c} \cos \vartheta\right)} = \frac{cq}{kT \left(1 + \frac{\delta T}{T}\right)},$$

где относительное возмущение температуры, связанное с дрейфом, есть

$$\left(\frac{\delta T}{T}\right)_1 = \frac{v}{c} \cos \vartheta. \quad (3.4.3)$$

В звуковой волне, как известно, относительное изменение плотности связано со скоростью движения частиц u соотношением

$$\frac{\delta \rho}{\rho} = \frac{v}{u}, \quad (3.4.4)$$

где скорость (адиабатическая) звука

$$u = \left(\frac{\partial p}{\partial (\rho + \rho_r)}\right)^{1/2} \approx \frac{c}{\sqrt{3}} \left(\frac{\partial \rho_r}{\partial (\rho + \rho_r)}\right)^{1/2}. \quad (3.4.5)$$

Тогда (3.4.3) можно переписать в виде

$$\left(\frac{\delta T}{T}\right)_1 \approx \frac{u}{c} \frac{\delta \rho}{\rho} \cos \vartheta. \quad (3.4.6)$$

Во-вторых, возмущения температуры создаются сжатиями и расширениями объемов в звуковой волне. Для адиабатических потенциальных возмущений связь между $\delta T/T$ и относительным возмущением плотности вещества имеет вид

$$\left(\frac{\delta T}{T}\right)_2 = \frac{1}{3} \frac{\delta \rho}{\rho}. \quad (3.4.7)$$

Эта формула применима, когда вещество и излучение находятся в тепловом равновесии. Теплоемкость и энтропия излучения, как мы видели в §§ 2.2 и 2.3, много больше теплоемкости и энтропии вещества; при совместном адиабатическом изменении объема вещества и излучения сохранение энтропии

$$sV \propto s\rho^{-1} = \text{const},$$

приводит к соотношению

$$T \propto V^{-1/3} \propto \rho^{1/3},$$

откуда и следует (3.4.7).

Такая полная адиабатичность предполагает, в частности, что длина свободного пробега фотона,

$$l_{re} = \frac{m}{\sigma_T \rho} \approx 4 \cdot 10^{29} \Omega^{-1} (1+z)^{-3} \text{ см}, \quad (3.4.8)$$

меньше характерного размера возмущений (длины волны потенциального движения). В момент рекомбинации

$$l_{re}(z_{rec}) \approx 10^{20} \Omega^{-1} \text{ см}, \quad (3.4.9)$$

и в объеме с размером $\sim l_{re}$ содержится масса

$$M_r(z_*) = \frac{4\pi}{3} \rho l_{re}^3 \approx 3 \cdot 10^7 \Omega^{-2} M_\odot. \quad (3.4.10)$$

Адиабатические возмущения, сохраняющиеся к эпохе рекомбинации, характеризуются значительно большими массами (см. предыдущий параграф) и для них указанное условие выполняется.

Наконец, в-третьих, изменения плотности $\delta\rho$ в масштабе возмущения λ создают дополнительные перепады гравитационного потенциала в объеме возмущения:

$$\delta\phi \approx -G\delta\rho\lambda^2.$$

Возмущения потенциала, или, на релятивистском языке, возмущения метрики, влияют на распространение квантов, изменяя прежде всего их частоту: из области, где гравитационное поле сильнее, кванты приходят с несколько большим красным смещением. Это означает, что из области повышенной плотности кванты выйдут более красными, а из области пониженной плотности — более синими, чем в отсутствие возмущений. Этот релятивистский эффект, очевидно, тем существеннее, чем больше размер λ , чем ближе он к расстоянию до горизонта ($\sim ct$) в соответствующий момент времени. Для относительно малых масштабов ($\lambda \ll ct$) эффект слаб и потому должна иметь место линейная зависимость дополнительного сдвига частоты квантов от дополнительного перепада гравитационного потенциала:

$$\frac{\delta\omega}{\omega_0} = \frac{\omega - \omega_0}{\omega_0} = \frac{\delta\phi}{c^2}. \quad (3.4.11)$$

Это изменение частоты действует одинаково на весь спектр излучения из данной области, оставляя распределение планковским. Происходит лишь общий сдвиг распределения, так что частота, соответствующая максимуму интенсивности ω_{\max} , сдвигается в отношении (3.4.11). Но частота ω_{\max} пропорциональна температуре, и, значит, эффект сводится к дополнительному возмущению температуры:

$$\left(\frac{\delta T}{T}\right)_3 \approx \frac{\delta \varphi}{c^2} \approx -\frac{G\rho}{c^2} \lambda^2 \frac{\delta \rho}{\rho}. \quad (3.4.12)$$

По порядку величины

$$G\rho \approx t^{-2},$$

так что

$$\left(\frac{\delta T}{T}\right)_3 \approx -\left(\frac{\lambda}{ct}\right)^2 \frac{\delta \rho}{\rho}. \quad (3.4.13)$$

Для сравнительно мелкомасштабных возмущений ($\lambda \ll ct$) эта третья поправка к температуре много меньше первых двух. Лишь для масштабов, сравнимых с расстоянием до горизонта, она существенна и при этих условиях способна, по крайней мере частично, скомпенсировать поправку другого знака $(\delta T/T)_2$, происходящую от адиабатических изменений объема.

Наблюдения дают предел вариации температуры фонового излучения $\delta T/T$, которая не есть результат единичного возмущения в прошлом и единичного измерения в настоящем. Возмущения в прошлом представляли собой случайные явления, которые требуют статистического описания и которые можно характеризовать некоторыми средними величинами. Наблюдения также связаны с накоплением статистического (но уже иной природы) материала и выражаются в усредненных величинах. Таким образом, чтобы связать теоретические формулы для $\delta T/T$ с наблюдательной величиной (3.4.1), нужно произвести двойное усреднение.

Для этого возведем в квадрат сумму двух первых, наиболее существенных, вкладов в вариацию температуры:

$$\left[\left(\frac{\delta T}{T}\right)_1 + \left(\frac{\delta T}{T}\right)_2 \right]^2 = \left(\frac{\delta \rho}{\rho}\right)^2 \left(\frac{u^2}{c^2} \cos^2 \vartheta + \frac{2}{3} \frac{u}{c} \cos \vartheta + \frac{1}{9} \right).$$

Так как ориентация направления движения по отношению к лучу зрения случайна, можно усреднить по углу ϑ

между ними. Второй член в среднем равен нулю, и мы получаем

$$\left\langle \left(\frac{\delta T}{T} \right)^2 \right\rangle_{\vartheta} = \frac{1}{3} \left(\frac{\delta \rho}{\rho} \right)^2 \left(\frac{u^2}{c^2} + \frac{1}{3} \right).$$

Скорость звука u в среде есть $c/\sqrt{3}$ на стадии преобразования излучения; в эпоху рекомбинации $u/c \lesssim 1/\sqrt{3}$. Поэтому

$$\left\langle \left(\frac{\delta T}{T} \right)^2 \right\rangle_{\vartheta} \lesssim \frac{2}{9} \left(\frac{\delta \rho}{\rho} \right)^2. \quad (3.4.14)$$

Усредняя теперь по ансамблю возмущений, т. е. заменяя $(\delta \rho/\rho)^2$ на среднее значение этой величины в соответствующую эпоху, мы получим среднеквадратичную вариацию температуры:

$$\frac{\delta T}{T} = \left(1 \div \sqrt{\frac{2}{3}} \right) \frac{1}{\sqrt{3}} \frac{\delta \rho}{\rho}. \quad (3.4.15)$$

Что касается возмущений вихревого типа, то такие движения не создают возмущений плотности и потому из трех рассмотренных причин флуктуаций температуры реликтового излучения остаются доплеровское смещение и гравитационный эффект. Последний не описывается в этом случае формулой (3.4.12), так как в ньютоновском приближении он вообще отсутствует. Возмущения гравитационного поля, создаваемые вращательными движениями, возникают только в релятивистской теории (см. § 3.2). Для масштабов, много меньших расстояния до горизонта, этот эффект, очевидно, несуществен. Вариации температуры, обусловленные доплер-эффектом, описываются формулой (3.4.3), в которой под v можно понимать и вихревую скорость. После усреднения по углу зрения и ансамблю вихрей находим среднеквадратичную флуктуацию температуры:

$$\frac{\delta T}{T} = \frac{1}{\sqrt{3}} \frac{(\langle v^2 \rangle)^{1/2}}{c}. \quad (3.4.16)$$

Наблюдательные ограничения вариаций температуры реликтового фона позволяют, как мы видим, получить очень важные для космогонии сведения о догалактической структуре в Метагалактике в эпоху, когда излучение становится независимым от вещества. На эту возможность

указал впервые Силк (1968); в его работе были приведены простые формулы (подобные тем, которые мы получили в этом параграфе), связывающие амплитуды догалактических возмущений с вариациями температуры реликтового излучения в масштабах галактик и скоплений. Развитие этих представлений дано Зельдовичем и Сюняевым (1970а), Пиблсом и Ю (1970).

В «стандартной» горячей космологической модели излучение отрывается от вещества в эпоху рекомбинации $z = z_{\text{rec}} = 1000$. Детальный анализ физических процессов в эту эпоху привел к ряду существенных уточнений, связанных прежде всего с тем, что взаимодействие плазмы и излучения прекратилось не мгновенно. В действительности рекомбинация происходила в течение некоторого конечного промежутка времени (см. § 2.1) и в результате разные кванты испытали свое последнее рассеяние, после которого они дошли до нас по определенному лучу зрения, на разных, вообще говоря, возмущениях и в разные моменты времени. Если какой-либо луч зрения пересекал много мелкомасштабных возмущений, то создаваемые ими возмущения температуры должны частично или полностью погашать друг друга вследствие независимости, и итоговые, регистрируемые в наблюдениях вариации температуры окажутся сильно сглаженными, претерпев это новое — третье — усреднение. Зельдович и Сюняев (1970а) подробно рассмотрели этот эффект ослабления возмущений температуры и показали, что в масштабах, содержащих массы $< 10^{12} M_{\odot}$, флуктуации температуры практически полностью сглаживаются. Для больших масштабов ($M > 10^{15} M_{\odot}$) эффект сглаживания исчезает. Показано также, что вариации температуры реликтового излучения создаются главным образом благодаря эффекту Доплера при рассеянии фотонов на движущемся веществе вблизи момента $z = 1050$. Поэтому для движений обоих типов, охватывающих большие массы $M > 10^{15} M_{\odot}$, справедливо соотношение вида (3.4.16), где скорость v должна браться в указанный момент времени.

Кроме тех первичных возмущений при которых метагалактическая среда ведет себя как единая жидкость, допустимы уже упоминавшиеся выше статические возмущения плотности плазмы, замороженные в фон излучения, — энтропийные возмущения. В этом случае су-

существует гравитационное влияние возмущений на реликтовое излучение, причем применима формула (3.4.12) с тем, однако, отличием, что под $\delta\rho$ нужно понимать возмущения одного только вещества. Этот эффект мог бы быть замечен только для возмущений в масштабе горизонта. Более существенные вариации температуры создаются возмущениями этого типа по иной причине. Дело в том, что в ходе рекомбинации сгущения плазмы, ранее замороженные в фон излучения, начинают усиливаться гравитационной неустойчивостью. В результате появляются «вторичные» скорости, и фотоны рассеиваются на движущемся веществе. Соответствующие скорости могли создать в момент $z=1050$ определенные вариации температуры реликтового излучения, причем для их описания снова применимо соотношение вида (3.4.16). Вариации в масштабах скоплений галактик оказываются в этом случае несколько меньше тех, которые создавались бы первичными адиабатическими возмущениями с той же величиной возмущений плотности вещества.

Недавно Парийский (1977) сообщил о новых, более жестких, чем (3.4.1), наблюдательных ограничениях возможных вариаций температуры реликтового излучения:

$$\frac{\delta T}{T} < 1,6 \cdot 10^{-1} (\theta')^{-2/3}, \quad 5' < \theta' < 200'. \quad (3.4.17)$$

Здесь θ' — измеряемый в минутах дуги видимый угловой размер области возмущения. Он связан с массой вещества в этой области соотношением *)

$$\theta' = 10\Omega^{2/3} \left(\frac{M}{10^{16} M_{\odot}} \right)^{1/3}. \quad (3.4.18)$$

В масштабе самых крупных скоплений галактик ($M = (1 \div 3) \cdot 10^{16} M_{\odot}$) ограничению (3.4.17) легче, по-видимому, удовлетворяют первичные энтропийные возмущения **). На основании (3.4.17), (3.4.18) можно также заключить, что возмущения еще большего масштаба ($M > 10^{16} M_{\odot}$), скорее всего, не были в эпоху рекомбинации достаточно сильными, чтобы привести в более позднюю эпоху к сильным

*) Последнее можно получить на основании общих формул Фридмановской космологии (гл. 1).

***) Зенцова и Чернин (1977), Дорошкевич и др. (1977).

возмущениям, способным дать начало структурам соответствующих масс. Такой вывод согласуется с наблюдательными результатами внегалактической астрономии (о них мы говорили в первом параграфе книги), согласно которым иерархия космических структур не продолжается неограниченно к сколь угодно большим масштабам, а ограничена размером ячейки однородности, которой отвечают массы $\sim 10^{16} \div 10^{17} M_{\odot}$.

Характер вариаций температуры реликтового фона существенно зависит, очевидно, от того, в какую эпоху произошло в действительности «последнее рассеяние» реликтовых квантов. Выше предполагалось, что кванты свободно распространяются к наблюдателю от эпохи рекомбинации; именно для такой стандартной картины горячей Вселенной справедливо, в частности, соотношение (3.4.18). Если же в какую-то эпоху после рекомбинации произошел разогрев вещества и оно стало снова ионизованным, то дополнительное рассеяние квантов на электронах могло значительно сгладить следы первоначальных возмущений в угловом распределении излучения. В эпоху $z \gtrsim 10-30$ такое сглаживание распространилось бы на самые большие масштабы возмущений; это происходило бы по той же причине, по которой в эпоху рекомбинации ослабляются вариации температуры излучения в масштабах $M < 10^{12} M_{\odot}$. Возможность вторичной ионизации нельзя определенно исключить, и потому интерпретация наблюдательных данных об изотропии реликтового фона остается не вполне однозначной.

НЕОДНОРОДНОСТЬ И АНИЗОТРОПИЯ В РАННЕЙ ВСЕЛЕННОЙ

§ 4.1. Пространственная симметрия Метагалактики и обобщения модели Фридмана

Наблюдения галактик и скоплений галактик указывают на однородность распределения и изотропию движения вещества в Метагалактике; эта высокая пространственная симметрия проявляется в больших масштабах и относится к современному этапу эволюции Вселенной. Наиболее чувствительный индикатор — изотропный реликтовый фон — отражает состояние Вселенной в эпоху, когда излучение перестало взаимодействовать с веществом, что могло произойти при красном смещении ~ 1000 или позже (см. гл. 2). О более ранней эпохе, отвечающей бóльшим значениям красного смещения, прямых наблюдательных данных нет *).

Возможность и даже необходимость в прошлом каких-то отклонений от общей однородности и изотропии прямо связаны с современной структурой Вселенной. Действительно, наблюдаемые сейчас неоднородности (галактики и скопления) могли возникнуть, очевидно, только благодаря тому, что в прошлом имелись определенные «затравочные» отклонения от идеальной симметрии. Важно, что при этом масштаб галактики, например, не мог считаться малым в те времена, когда соответствующая ему масса была распределена в области, выходящей за пределы текущего горизонта событий. Данной массе вещества M отвечает пространственный масштаб

$$\lambda(t) = \left(M / \frac{4\pi}{3} \rho \right)^{1/3} \approx \frac{M^{1/3}}{1+z}, \quad (4.1.1)$$

который превосходит расстояние до горизонта событий

*) Такие данные мог бы сообщить фон реликтовых нейтрино, «отрывающихся» от вещества много раньше (§ 2.1), если бы он был доступен наблюдениям.

в изотропном мире,

$$ct \approx 10^{30} [(1+z)^{-2}], \quad z > z_*, \quad (4.1.2)$$

при красном смещении

$$z > z_M \approx 2 \cdot 10^4 \Omega^{1/2} \left(\frac{M}{10^{12} M_\odot} \right)^{1/2}. \quad (4.1.3)$$

Отсюда видно, что отклонения от однородности, охватывающие массы, сравнимые с массами галактик, нельзя считать мелкомасштабными практически в течение всей эпохи преобладания излучения.

Отклонения могли быть сначала малыми по амплитуде и тогда для их анализа было бы достаточно теории возмущений. Но, как мы уже упоминали (§ 3.1), возмущения могли быть в принципе и сильными. Например, существование первичных вихревых движений на некоторой фазе изотропного расширения предполагает, что самые первые фазы расширения были сильно анизотропными (§ 3.2).

В таком случае требуется анализ существенно нефридмановских космологических моделей для начальных фаз расширения.

Эти соображения служат мотивом изучения более общих, чем изотропные, космологических моделей. На этом пути, как предполагают, возможны также поиски подхода и к другой проблеме космологии — проблеме причинной связности Метагалактики (§ 1.4).

Вакуольный мир. Простым и красивым (и, по-видимому, первым по времени) обобщением модели Фридмана, допускающим локальные неоднородности в однородном и изотропном в среднем мире, является так называемая вакуольная модель, построенная Эйнштейном и Страуссом (1945).

Распределение вещества, давлением которого пренебрегается, в этой модели всюду однородно, кроме отдельных сферических полостей — вакуолей, которые между собой не перекрываются. В центре каждой из них имеется точечная масса, равная той, которая при однородном распределении была бы сосредоточена в том же объеме. Внутри каждой вакуоли реализуется гравитационное поле (или, на языке релятивистской теории, метрика) точечной массы в пустоте, а вне полостей ситуация такая,

как в однородном и изотропном мире Фридмана. Возможность «сшивки» на границе вакуоли определяется тем особым свойством центральной симметрии, которое мы уже не раз использовали, именно: движение материи в некотором сферическом слое зависит только от полной массы внутренней сферы, каково бы ни было ее распределение и движение. Это свойство очевидно в ньютоновской теории для конечных распределений вещества, оно справедливо и в общей теории относительности.

Чтобы получить полное представление о вакуольном мире, достаточно рассмотреть какую-то одну вакуоль в изотропном мире Фридмана. Удобно выбрать начало координат в центре вакуоли, и тогда картина станет полностью сферически-симметричной.

В ньютоновской теории для частиц на границе вакуоли $r = R$ имеет место уравнение движения

$$\ddot{R} = - \frac{GM_0}{R^2},$$

где M_0 — центральная масса. Для всех частиц вне вакуоли, $r > R$, уравнение движения по виду то же, но под гравитирующей массой нужно понимать, как и в §§ 1.2, 1.4, полную массу, лежащую внутри сферы данного радиуса. Такое сходство уравнений позволяет при должном выборе «начальных условий», т. е. постоянных интегрирования, согласовать движение частиц на границе вакуоли с общей динамикой внешней однородной области, чем и достигается упоминавшаяся выше «сшивка». При этом

$$M_0 = \frac{4\pi}{3} \rho(t) R^3(t),$$

где $\rho(t)$ — плотность вещества в соответствующем однородном решении.

Очевидно, что можно построить сколько угодно таких расширяющихся вакуолей в расширяющемся мире и ни одна из них не будет «знать» о других, если только вакуоли не перекрываются. Условие неперекрываемости при отсутствии каких-либо других ограничений на размеры вакуолей не мешает сосредоточить сколь угодно большую долю массы мира в центрах вакуолей. Так можно получить изотропную в большом масштабе модель с системой сильных локальных неоднородностей.

Конечно, это не самый общий случай локальной неоднородности, хотя бы потому, что центры вакуолей должны двигаться по строго хаббловским траекториям. Можно сказать, что это мир с неоднородностями, но без собственных движений материи.

Интересно в этом смысле дополнить такую модель анализом собственных движений в ней хотя бы в самом простом случае, когда частицы, имеющие собственные, не хаббловские скорости, не влияют на гравитационное поле вакуольного мира (Чернин (1974а)). В изотропном мире собственные движения таких пробных частиц происходят по прямым в силу равноправности всех направлений (см. § 1.4). То же относится и к пространству между вакуолями в вакуольном мире. В системе отсчета, началом которой служит одна из точек траектории частицы, траектория есть радиальная линия.

Но если частица при своем движении попадает в вакуоль, она оказывается в поле центральной точечной массы. При большой собственной скорости она может снова выйти в область однородной плотности, пройдя около центральной массы по гиперболической орбите. Если же скорость при входе в вакуоль не велика, частица переходит на эллиптическую кеплерову орбиту и становится связанной. Необходимое условие связанности — отрицательность энергии частицы в вакуоли. Это условие является достаточным для удержания частицы в вакуоли, если расширение последней происходит по параболическому или гиперболическому закону: вакуоль расширяется быстро и частица не может догнать ее границу. При эллиптическом расширении удержание частицы на неограниченное время вообще невозможно, так как на фазе сжатия, сменяющей через конечный период фазу расширения, орбита частицы рано или поздно пересечет границу вакуоли при любом значении энергии частицы.

Простая и наглядная вакуольная модель может послужить основой для обобщений, которые приблизили бы картину к реалистической. Так, с помощью решения Толмена или его классического аналога (см. Дополнение) нетрудно построить модель не с точечной массой в центре вакуоли, а с непрерывным неоднородным, но центрально-симметричным распределением материи во всем ее объеме. В этом случае «бродячая» пробная частица, вошедшая

в область неоднородности, окажется в поле меняющегося со временем потенциала; ее траектория уже не будет кеплеровым эллипсом, а условия захвата и удержания полем сгущения станут менее благоприятными.

Изотропизация. Обязательным свойством неоднородных анизотропных космологических моделей должно быть приближение в ходе расширения к состоянию крупномасштабной однородности и изотропии. Принципиальная возможность такой эволюции действительно допускается уравнениями гравитации как в ньютоновском приближении, так и в строгой общерелятивистской постановке задачи. Это легче всего показать на общем решении уравнений гравитации для сферической симметрии, которое мы уже изучали (с другой точки зрения) в гл. 3. Именно, рассмотрим динамику расширения неоднородного, но сферически-симметричного распределения вещества (без давления) и начнем с ньютоновской теории.

Будем считать, что сферически-симметричное расширение происходит без пересечения сферических слоев, так что масса вещества, заключенная в сфере под данными частицами, остается неизменной во времени и может служить лагранжевой координатой этих частиц.

Если расширение параболическое, т. е. механическая энергия каждого слоя равна нулю, то радиус слоя изменяется со временем по закону

$$R(t, M) = \left(\frac{3}{2}\right)^{2/3} (2GM)^{1/3} (t - t_i(M))^{2/3}. \quad (4.1.4)$$

В частном случае $t_i = \text{const}$ отсюда получается решение для однородного распределения, при котором сингулярность (в данном случае при $t = t_i$, а не при $t = 0$, как в § 1.2, что, разумеется, несущественно) является одновременной для всех слоев. Выделенность центра $R = 0$ при этом фиктивна и его можно связать с любой частицей среды. Когда величина t_i , т. е. момент выхода из сингулярности, зависит от M , распределение вещества неоднородно, и вблизи сингулярности отличия от фридмановской модели могут быть сколь угодно велики. Однако в ходе расширения различие в моменте начала расширения разных слоев постепенно становится все менее существенным для общего распределения и движения

среды:

$$R(t, M) = \left(\frac{3}{2}\right)^{2/3} (GM)^{1/3} \left(1 - \frac{t_i}{t}\right)^{2/3}. \quad (4.1.5)$$

Асимптотически, при $t \rightarrow \infty$, т. е. $t \gg t_i$, влияние функции $t_i(M)$ исчезает вовсе и отклонения от однородности, сколько бы значительными они ни были вначале, сглаживаются. Другими словами, любое сферически-симметричное параболическое распределение среды неизбежно сопровождается приближением к однородности распределения и изотропии движения. Таков самый простой пример обобщения модели Фридмана, демонстрирующий возможность изотропизации.

Другой пример — сферически-симметричное гиперболическое расширение, описываемое формулами

$$\left. \begin{aligned} R &= R_m(M) (\operatorname{ch} \eta - 1), \\ t - t_i(M) &= \frac{R_m(M)}{[2E(M)]^{1/2}} (\operatorname{sh} \eta - \eta), \end{aligned} \right\} \quad (4.1.6)$$

где $E(M) > 0$ — механическая энергия слоя, характеризуемого лагранжевой координатой M , которая, как и выше, имеет смысл внутренней массы,

$$R_m(M) = GM/2E(M). \quad (4.1.7)$$

В частном случае

$$t_i = \text{const}, \quad R_m/\sqrt{E} = \text{const}, \quad (4.1.8)$$

решение (4.1.6) совпадает с однородным изотропным решением.

В общем случае различия в моменте начала расширения для различных слоев, $t_i(M)$, сглаживаются со временем в ходе инфинитного разлета, но изотропизация при этом происходит не обязательно. Для изотропизации требуется еще, чтобы другая функция, величина (4.1.7), не зависела от массы. Тогда, при произвольной функции $t_i(M)$, переход к состоянию однородности и изотропии достигается асимптотически в пределе $t/t_i \rightarrow \infty$.

Этими двумя специальными типами движения исчерпываются сферически-симметричные обобщения (при нулевом давлении) модели Фридмана, отвечающие условию асимптотической изотропизации (Рубан и Чернин (1969)).

При эллиптическом движении сферически-симметричного распределения вещества фаза общего расширения имеет конечную длительность, и полная изотропизация недостижима. Однако приближение к изотропии происходит и здесь, если в соответствующем этому случаю решении,

$$\left. \begin{aligned} R(t, M) &= R_m(M)(1 - \cos \eta), \\ t - t_i(M) &= \frac{R_m}{(2|E(M)|)^{1/2}} (\eta - \sin \eta), \end{aligned} \right\} \quad (4.1.9)$$

величина $R_m/\sqrt{|E|} = M/2|E|^{3/2}$ не зависит от M , а продолжительность фазы расширения $0 < t < t_{\max} \approx \approx R_m/\sqrt{|E|}$ такова, что

$$t_{\max}(M) \gg t_i(M).$$

Смена расширения сжатием после $t = t_{\max}$ должна сопровождаться усилением неоднородностей, не исчезнувших окончательно на фазе расширения.

Далеко идущее подобие ньютоновской и релятивистской теории при сферической симметрии позволяет предполагать, что ньютоновские обобщения модели Фридмана, которые мы рассмотрели, должны в этом случае иметь релятивистские аналоги. Действительно, для трех динамических типов расширения, о которых мы говорили, существуют общерелятивистское решение Толмена (см. Дополнение):

$$\left. \begin{aligned} ds^2 &= c^2 dt^2 - R(t, \chi) d\Omega^2 - F(t, \chi) d\chi^2, \\ d\Omega^2 &= \sin^2 \vartheta d\varphi^2 + d\vartheta^2, \quad F = \frac{\left(\frac{\partial R}{\partial \chi}\right)^2}{1 + E/c^2}, \end{aligned} \right\} \quad (4.1.10)$$

содержащее в качестве частных случаев все три варианта изотропных фридмановских решений. Как и в ньютоновской теории, этим частным случаям отвечает условие

$$t_i = \text{const}, \quad R_m/\sqrt{|E|} = \text{const},$$

которое должно выполняться в формулах (4.1.4)—(4.1.9), приобретающих теперь смысл выражений для метрического тензора в интервале (4.1.10). Чтобы убедиться в этом, достаточно проверить, что при указанных условиях переменные в функциях $R(t, M)$ и $F(t, M)$ разделяются, как и в решении Фридмана.

Тем же способом легко установить, что менее жесткие условия

$$E = 0 \text{ или } E > 0, R_m/V\sqrt{|E|} = \text{const}$$

при произвольном виде второй функции $t_i(M)$ обеспечивают асимптотическую при $t/t_i \rightarrow \infty$ изотропизацию в релятивистском решении (4.1.10). Остаются в силе также и соображения о стремлении к изотропизации на фазе расширения при эллиптическом движении.

Итак, неограниченная экстраполяция в прошлом, вплоть до сингулярного состояния, свойств однородности и изотропии, характерных для современного состояния Вселенной, представляет собой только одну из ряда возможностей, допускаемых теоретически уравнениями гравитации и совместимых с известными наблюдательными данными. Мы увидим в дальнейшем, что анизотропные обобщения модели Фридмана для ранних фаз космологического расширения составляют довольно широкий класс, содержащий, помимо сферически-симметричных моделей, еще и модели совсем иного рода, не имеющие аналогов в ньютоновской теории. В таких моделях не только динамика расширения, но и сам характер начальной сингулярности приобретает существенно нефридмановский характер. Однако не в каждой из них изотропизация происходит сама собой при больших временах, как в рассмотренной нами сферически-симметричной модели и в анизотропной модели другого типа, которая обсуждается далее в этом параграфе.

Фактором изотропизации для них является само космологическое расширение (этот вопрос с большой полнотой исследован Зельмановым (1959)). Одна интересная с космогонической точки зрения особенность такой кинематической изотропизации состоит в том, что она развивается в больших пространственных масштабах быстрее, чем в малых (Чернин и Шварц (1972), Чернин (1974б)). Если понимать наблюдаемую космическую структуру как остаток, реликт анизотропной начальной фазы расширения, то это могло бы объяснить, почему иерархия космических структур не продолжается до сколь угодно больших пространственных масштабов. Анизотропия, сохраняющаяся в этом случае в относительно малых масштабах, стимулирует более быстрый рост мелкомасштабных неоднороднос-

тей, подобно тому как в глобально анизотропном мире возмущения плотности могут при определенных условиях расти быстрее, чем в изотропном мире Фридмана (последнее мы покажем в § 4.2).

Существенным фактором изотропизации совсем другой природы могли бы быть квантовые явления вблизи сингулярности (см. конец § 4.2).

Однородность без изотропии. Представим себе модель, в которой сопутствующее пространство является евклидовым, распределение вещества однородно, но расширение происходит с различными скоростями в различных направлениях. В каждом из трех взаимно перпендикулярных направлений скорость имеет хаббловскую линейную зависимость от расстояния

$$v_i = H_i R_i, \quad i = 1, 2, 3, \quad (4.1.11)$$

но три величины $H_i(t)$, вообще говоря, не равны друг другу. Различны и зависимости от времени масштабных факторов $a_i(t)$ вдоль этих направлений, так как $H_i = \dot{a}_i/a_i$. При таком движении кубик вещества испытывает деформацию, превращаясь в параллелепипед. Если в изотропной модели кубик оставался при расширении кубиком и его объем изменялся пропорционально кубу масштабного фактора, то теперь объем, занимаемый данными частицами, изменяется как произведение трех масштабных факторов:

$$V \propto a_1(t)a_2(t)a_3(t).$$

Указанному типу деформации отвечает квадрат метрического интервала

$$ds^2 = c^2 dt^2 - a_1^2(t) d\chi_1^2 - a_2^2(t) d\chi_2^2 - a_3^2(t) d\chi_3^2. \quad (4.1.12)$$

Система отсчета здесь сопутствующая, точно так же, например, как и в изотропной модели с евклидовым сопутствующим пространством, которая является частным случаем (4.1.12).

Конкретный вид зависимости $a_i(t)$ от времени определяется уравнениями общей теории относительности вместе с уравнением состояния вещества. В отличие от изотропной космологии, классическая аналогия здесь недостаточна, так как характер движения вещества в некоторой области определяется теперь, вообще говоря, не толь-

ко локальным распределением тяготеющих масс. Такая нелокальность объясняется тем, что в общей теории относительности, кроме гравитационного поля, создаваемого веществом, имеются еще и свободные гравитационные поля, никак не связанные с наличием или отсутствием вещества. Свободное гравитационное поле, как и свободное электромагнитное поле, не может обладать сферической симметрией и его нет в изотропной (без возмущений) модели, где имеется полная сферическая симметрия относительно любой точки пространства. В анизотропной космологии сферическая симметрия отсутствует, и свободное гравитационное поле возможно. Оно изменяется в ходе космологического расширения и само влияет на его характер. Мы увидим, что это влияние может быть особенно существенно на самых ранних фазах расширения.

Изучение анизотропных моделей на языке ньютоновской теории возможно только в тех случаях, когда свободное гравитационное поле не играет роли. Мы рассмотрим сначала простейший такой случай *). Прежде всего заметим, что, в отличие от модели изотропного расширения, здесь нужно рассматривать не сферы, а эллипсоиды, у которых длины полуосей пропорциональны соответствующим масштабным факторам. Можно ожидать, и это действительно подтверждается соответствующим точным решением уравнений общей теории относительности, что, как и в случае сферы, силы, действующие на частицы поверхности эллипсоида, определяются только внутренней массой. Наличие бесконечной и однородно распределенной массы вне эллипсоида, как и вне сферы в изотропном случае, не влияет в изучаемой модели на процессы внутри и на поверхности эллипсоида.

Так как анизотропные модели имеют смысл лишь для ранней фазы расширения, а на поздних фазах должна наступать изотропизация, то мы рассмотрим поведение эллипсоида непосредственно вблизи сингулярности. В сильно сплюсненном эллипсоиде вращения, полуось которого вдоль оси симметрии $a_3(t)$ много меньше равных полуосей $a_1(t) = a_2(t) = a(t)$, сила тяжести в направлении «3» (вертикальном) гораздо сильнее, чем в направлениях «1»

*) Анализ ньютоновских аналогов однородных анизотропных моделей дается в работах Шикина (1970, 1971); см. также книгу Зельдовича и Новикова (1975).

и «2» (горизонтальных). Если бы такой эллипсоид сжимался (из состояния покоя при бесконечных размерах), то его сжатие в вертикальном направлении происходило бы гораздо быстрее, чем в горизонтальных. В пределе, когда толщина стремится к нулю, а плотность к бесконечности, движением в горизонтальных направлениях можно вообще пренебречь. Обращая картину сжатия (изменяя знак времени), мы получим картину расширения, в которой последнее происходит преимущественно в направлении оси симметрии эллипсоида, а горизонтальные его размеры остаются почти неизменными. Чтобы найти зависимость от времени малой полуоси, т. е. функцию $a_3(t)$, заметим, что сила тяжести направлена почти строго вдоль вертикали и поле близко к полю на поверхности однородного бесконечного плоского слоя. Такое поле однородно, и уравнение движения имеет вид

$$\ddot{a}_3(t) = -4\pi G \rho a_3 = -A = \text{const.} \quad (4.1.13)$$

Интегрируя дважды, находим

$$\bullet \quad a_3(t) = -\frac{1}{2} At^2 + Bt + C. \quad (4.1.14)$$

Здесь $B > 0$ и C — произвольные константы.

Выбираем отсчет времени так, что $a_3 = 0$ при $t = 0$. Тогда $C = 0$, и можно написать

$$\left. \begin{aligned} a_3(t) &= \alpha t(t_a - t), \\ \alpha &= \frac{1}{2} A, \quad t_a = B/\alpha. \end{aligned} \right\} \quad (4.1.15)$$

Движение вдоль оси «3» оказывается финитным, что характерно для плоского слоя. Расширение сменяется сжатием при $t = \frac{1}{3}t_a$, $a_3 = \frac{1}{2}\alpha t_a^2$, а затем при $t = t_a$ вновь достигается бесконечная плотность. Но само это решение имеет смысл только при $t \rightarrow 0$. В этом пределе

$$a_3 \propto t, \quad (4.1.16)$$

т. е. движение является инерциальным. Плотность вещества изменяется обратно пропорционально толщине

$$\rho \propto a^{-1} \propto t^{-1}. \quad (4.1.17)$$

Инерциальность начала расширения резко отличает рассмотренную анизотропную модель от изотропной.

При изотропии расширения гравитационная сила стремится к бесконечности, когда плотность стремится к бесконечности, и движение с самого начала происходит с замедлением. При анизотропии рассматриваемого типа гравитационная сила остается конечной при стремлении плотности к бесконечности и потому вблизи сингулярности ее действием можно пренебречь: изменение скорости сколь угодно мало при достаточно малых временах от начала расширения.

Таким образом, анизотропное расширение на своей начальной фазе может происходить так, как если бы вообще никакого гравитационного поля — ни собственного, ни свободного — не было бы. Эта возможность действительно подтверждается точным решением уравнений общей теории относительности для выбранного нами типа деформации среды.

Зная теперь поведение масштабных факторов вблизи особенности, можно получить и асимптотическое выражение для метрики. Так как масштабные факторы определены лишь с точностью до постоянного множителя, то квадрат интервала можно представить в виде

$$ds^2 = c^2 dt^2 - d\chi_1^2 - d\chi_2^2 - t^2 d\chi_3^2. \quad (4.1.18)$$

(В этом выражении лагранжевы координаты χ_1, χ_2 имеют размерность длины, а χ_3 — скорости.) На первый взгляд, метрика (4.1.18) не похожа на галилееву, которую следовало бы ожидать на основании наших соображений о несущественности гравитационного поля вблизи сингулярности. В действительности она отличается от обычного выражения для галилеевой метрики,

$$ds^2 = c^2 d\tau^2 - dx_1^2 - dx_2^2 - dx_3^2, \quad (4.1.19)$$

лишь преобразованием координат:

$$\tau = t \operatorname{ch} \chi_3, \quad x_1 = \chi_1, \quad x_2 = \chi_2, \quad x_3 = t \operatorname{sh} \chi_3.$$

В обоих случаях (4.1.18) и (4.1.19) мы имеем галилееву метрику, но декартовой системе координат (4.1.19) отвечает система отсчета, которая не сопутствует анизотропно расширяющемуся веществу.

Вакуумная стадия. Рассмотрим теперь релятивистскую анизотропную модель, в которой метрика имеет вид

(4.1.12). Уравнения общей теории относительности (см. Дополнение) в сопутствующей системе отсчета сводятся к следующим уравнениям для масштабных факторов, плотности энергии и давления:

$$\left. \begin{aligned} \frac{\dot{a}_1}{a_1} + \frac{\dot{a}_1}{a_1} \left(\frac{\dot{a}_2}{a_2} + \frac{\dot{a}_3}{a_3} \right) &= -4\pi G (p - \epsilon), \\ \frac{\dot{a}_2}{a_2} + \frac{\dot{a}_2}{a_2} \left(\frac{\dot{a}_1}{a_1} + \frac{\dot{a}_3}{a_3} \right) &= -4\pi G (p - \epsilon), \\ \frac{\dot{a}_3}{a_3} + \frac{\dot{a}_3}{a_3} \left(\frac{\dot{a}_1}{a_1} + \frac{\dot{a}_2}{a_2} \right) &= -4\pi G (p - \epsilon), \\ \frac{\ddot{a}_1}{a_1} + \frac{\ddot{a}_2}{a_2} + \frac{\ddot{a}_3}{a_3} &= -4\pi G (3p - \epsilon). \end{aligned} \right\} \quad (4.1.20)$$

Масштабные факторы, определяемые отсюда, даются выражениями (см., например, книгу Зельдовича и Новикова (1975)):

$$a_i(t) = t^{q_i} (t + t_a)^{\alpha - q_i}, \quad i = 1, 2, 3. \quad (4.1.21)$$

Здесь t_a — константа, числа q_1, q_2, q_3 не превосходят по модулю единицу и связаны двумя соотношениями:

$$q_1 + q_2 + q_3 = q_1^2 + q_2^2 + q_3^2 = 1. \quad (4.1.22)$$

Например, в осесимметричных случаях

$$q_1 = q_2 = 0, \quad q_3 = 1; \quad (4.1.23)$$

$$q_1 = q_2 = 2/3, \quad q_3 = -1/3. \quad (4.1.24)$$

Соотношения (4.1.22) позволяют выразить числа q_i через один параметр, который мы обозначим через U . Решая для этого первое из соотношений относительно q_3 , подставляя во второе и производя простые сокращения, мы получим

$$q_1^2 + q_2^2 + q_1 q_2 = q_1 + q_2.$$

Разделим обе части равенства на q_1^2 и введем обозначение $U = q_2/q_1$. Тогда получим

$$q_1 = \frac{1 + U}{1 + U + U^2}, \quad q_2 = \frac{U + U^2}{1 + U + U^2}, \quad q_3 = -\frac{U}{1 + U + U^2}, \quad (4.1.25)$$

Осесимметричным случаям отвечают $U = -1$ и $U = 1$ соответственно.

В пределе больших времен решение становится изотропным:

$$a_1 \propto a_2 \propto a_3 \propto t^\alpha, \quad t/t_a \gg 1.$$

Очевидно, что число α совпадает с соответствующим показателем степени в параболическом изотропном решении, в котором трехмерное сопутствующее пространство также является евклидовым. Это число выражается через отношение давления к плотности энергии (когда это отношение есть константа):

$$\alpha = \frac{2}{3} \frac{1}{1 + p/\varepsilon}.$$

Для пылевидной и ультрарелятивистской среды соответственно находим $\alpha = 2/3$ и $\alpha = 1/2$. Решение (4.1.21) является точным в первом из этих случаев; во втором же — приближенным, интерполяционным; оно дает правильную асимптотику при очень малых и очень больших временах. Условие адиабатичности

$$\frac{d\varepsilon}{\varepsilon + p} = - \frac{dV}{V} = - \frac{d(a_1 a_2 a_3)}{a_1 a_2 a_3}, \quad (4.1.26)$$

содержащееся в исходной системе (4.1.20), вместе с (4.1.21) дает плотность энергии для этих двух случаев:

$$\varepsilon = \frac{c^2}{6\pi G t (t + t_a)}, \quad p = 0; \quad (4.1.27)$$

$$\varepsilon = \frac{3c^2}{32\pi G t^{4/3} (t + t_a)^{2/3}}, \quad p = \frac{1}{3} \varepsilon. \quad (4.1.28)$$

Решение (4.1.21) — (4.1.28) радикально отличается от изотропного вблизи особенности. Из формул (4.1.21) — (4.1.25) видно, что при любых, кроме (4.1.23), значениях q_i два из чисел q_i положительны, а третье отрицательно, т. е. в двух направлениях происходит расширение, а в третьем — сжатие. Только при больших временах сжатие по третьему направлению сменяется расширением. Объем, сопутствующий данным частицам, всегда увеличивается со временем. Изотропное решение с евклидовым трехмерным пространством содержится в данном решении только как предельный, но не как частный случай.

Ньютоновское решение, построенное нами (для малых времен) в предыдущем разделе, есть частный случай

релятивистского решения, получающийся при давлении, равном нулю, если показатели соответствуют осесимметричному набору (4.1.23).

Однако при любых значениях показателей релятивистское решение обладает тем же общим свойством, которое мы уже отметили в ньютоновском случае. Именно, вблизи состояния бесконечной плотности влияние собственного гравитационного поля вещества является пренебрежимым (как это ни странно на первый взгляд). В ньютоновском случае это следовало из уравнений движения вещества, которые указывали на инерциальность этого движения вблизи особенности.

В релятивистском случае этот вывод можно получить из следующих соображений. Левая, «геометрическая», сторона релятивистских уравнений (4.1.20) содержит вторые производные от масштабных факторов и квадратичные комбинации их первых производных. При малых временах, когда $t \ll t_a$, те и другие пропорциональны, согласно (4.1.21), определенным степеням времени. Поэтому члены вида \ddot{a}_i/a_i и $\dot{a}_i\dot{a}_k/a_ia_k$ ведут себя в указанном пределе как обратные квадраты времени. Величину $\epsilon_a \approx \approx c^2/Gt^2$, являющуюся мерой этих геометрических членов уравнений гравитации, назовем вслед за Мизнером и др. (1973) «плотностью энергии анизотропии». При этом правая, — «материальная», — сторона уравнений (4.1.20) пропорциональна плотности энергии материи, которая согласно (4.1.27) — (4.1.28) в том же пределе ведет себя либо как t^{-1} , либо как $t^{-4/3}$.

При малых временах каждое из слагаемых слева в (4.1.20) возрастает гораздо быстрее, чем правая сторона, так что последней можно пренебречь по сравнению с первыми (конечно, алгебраическая сумма всех членов слева остается тождественно равной правой стороне!). Это и означает, что влияние материи на метрику становится пренебрежимо малым вблизи состояния бесконечной плотности *).

*) При особом, предельно жестком уравнении состояния, $p = \epsilon$ (Зельдович (1961)), которое в принципе допустимо для среды из взаимодействующих частиц, плотность вблизи особенности изменяется со временем точно так же, как и геометрические члены уравнений (4.1.20).

Такого рода соображения предполагают, что, по крайней мере для одного из масштабных факторов, вторая производная по времени вблизи особенности отлична от нуля. Это действительно имеет место всегда, кроме единственного, ньютоновского случая, рассмотренного в предыдущем разделе этого параграфа. Во всех других случаях расширение вблизи особенности происходит не инерциально, а с ускорением. Точнее, расширение вдоль двух направлений происходит с отрицательным ускорением, а сжатие по третьему — с положительным, т. е. скорость обоих движений уменьшается со временем. Но если имеет место ускорение, то на классическом языке можно говорить о производящей его силе. А так как влиянием материи вблизи особенности мы пренебрегаем, то эта сила и есть результат действия свободного гравитационного поля, о котором мы говорили выше.

Таким образом, мы встречаемся здесь с новой ситуацией, которой не было и не могло быть в изотропной космологии. Там геометрия целиком определялась распределением и движением вещества. Здесь же метрика пространства-времени является вакуумной вблизи сингулярности: это решение уравнений гравитации при нулевой правой, «материальной» части. Такой вакуум принципиально отличается от галилеева вакуума (в котором отсутствует гравитационное поле). Соответствующая ему метрика — она называется метрикой Казнера — получается из (4.1.21) — (4.1.25) в пределе малых времен и имеет вид:

$$ds^2 = c^2 dt^2 - t^{2q_1} d\chi_1^2 - t^{2q_2} d\chi_2^2 - t^{2q_3} d\chi_3^2. \quad (4.1.29)$$

Мы опустили здесь, как и прежде, несущественные постоянные множители в масштабных факторах. Интервал (4.1.29) никаким преобразованием координат не может быть сведён к галилееву интервалу (4.1.18) или (4.1.19).

Итак начальная фаза анизотропного расширения оказывается вакуумной: метрика, динамика расширения вещества и сам характер «выхода» из сингулярного состояния бесконечной плотности определяются исключительно свободным гравитационным полем. В дальнейшем, по мере расширения, роль вещества постепенно возрастает. Начиная с некоторого момента, оно начинает доминировать и вместе с этим постепенно ослабляется анизотропия.

Последнее выражается в том, что сжатие в одном из направлений замедляется и сменяется расширением, а затем скорости расширения по всем направлениям выравниваются. В конце концов, при очень больших временах от начала расширения динамика вещества и сама метрика становятся изотропными.

Релятивистское решение (4.1.21) — (4.1.25) демонстрирует возможность развития однородной и изотропной Вселенной из сильно анизотропного начального состояния. Это состояние в рассмотренной здесь модели является, правда, довольно специальным: оно однородно и пространственно евклидово. В более общем случае, когда это состояние неоднородно и не обладает пространственной евклидовостью, ситуация гораздо сложнее, и изотропизация просто за счет одного только расширения (адиабатическое охлаждение анизотропии, как это называют Мизнер, Торн и Уилер (1973)), вообще говоря, невозможна. Не исключено, однако, что физические процессы уже не просто кинематического характера (см. следующий параграф) могут обеспечить переход к изотропии от весьма произвольного анизотропного начального состояния.

§ 4.2. Некоторые особенности физических процессов в анизотропном мире

Анизотропная космология ставит целый ряд интересных и важных вопросов, касающихся влияния общей анизотропии мира на характер протекающих в нем физических процессов. В этом параграфе мы рассмотрим некоторые из них, обращая внимание в первую очередь на явления, существенные для понимания происхождения крупномасштабных космических структур.

Эволюция потенциальных движений. Обратимся к простейшему случаю метрики Казнера, для которой вещество можно считать пробным, и рассмотрим эволюцию потенциальных движений, наложенных на общее анизотропное расширение среды.

Регулярное расширение среды представляет собой движение в заданном внешнем поле, причем длины ребер некоторого однородного параллелепипеда с осями, параллельными трем осям казнеровской метрики, изменяются

пропорционально масштабным факторам:

$$l_i = a_i \chi_i, \quad (4.2.1)$$

где χ_i — разности лагранжевых координат соответствующих вершин. Если пользоваться ньютоновским языком, то можно сказать, что скорости изменения длин ребер суть

$$\dot{l}_i = \dot{a}_i \chi_i, \quad (4.2.2)$$

а соответствующие ускорения,

$$\ddot{l}_i = \ddot{a}_i \chi_i = \frac{\ddot{a}_i}{a_i} l_i, \quad (4.2.3)$$

пропорциональны расстояниям l_i .

Пусть область возмущения есть также однородный параллелепипед с осями, параллельными казнеровским осям, но длины его ребер связаны с разностями лагранжевых координат другими масштабными факторами $b_i(t)$, вообще говоря, не совпадающими с a_i :

$$\tilde{l}_i = b_i \chi_i. \quad (4.2.4)$$

Скорости изменения длин,

$$\dot{\tilde{l}}_i = \dot{b}_i \chi_i = \frac{\dot{b}_i}{b_i} \tilde{l}_i, \quad (4.2.5)$$

представляют собой потенциальные векторы, так как каждая компонента зависит только от «своей» координаты.

Ускорения

$$\ddot{\tilde{l}}_i = \ddot{b}_i \chi_i = \frac{\ddot{b}_i}{b_i} \tilde{l}_i \quad (4.2.6)$$

для тех же длин, т. е. в тех же точках, должны быть одинаковы и для регулярного и для возмущенного движения, так как эти ускорения определяются внешним полем. Поэтому

$$\ddot{b}_i = \frac{\ddot{a}_i}{a_i} b_i. \quad (4.2.7)$$

Такое уравнение движения (Дорошкевич, Зельдович и Новиков (1971)) справедливо для размеров, меньших

расстояния до горизонта,

$$\bar{l}_i \ll ct, \quad (4.2.8)$$

и при давлении среды, равном нулю.

Подставляя в (4.2.7)

$$a_i = t^{q_i},$$

где показатели q_i даются формулами (4.1.25), получим

$$\ddot{b}_i = q_i(q_i - 1) \frac{1}{t^2}. \quad (4.2.9)$$

Интегрируя дважды, находим:

$$b_i = \alpha_i t^{q_i} + \beta_i t^{1-q_i}, \quad (4.2.10)$$

где α_i, β_i — константы.

При $\beta_i = 0$ решение (4.2.10) переходит в невозмущенное: $b_i = \alpha t^{q_i} = a_i$. При $\alpha_i = 0$ движение имеет существенно иной характер. Объем области возмущения

$$\tilde{V} \propto b_1 b_2 b_3 \propto t^2 \quad (4.2.11)$$

возрастает быстрее, чем при регулярном расширении:

$$V \propto a_1 a_2 a_3 \propto t. \quad (4.2.12)$$

Если плотность в области возмущения была сначала выше средней, то при расширении объема по закону (4.2.11) это сгущение будет рассасываться. Обратная картина будет, например, в том случае, когда в метрике Казнера $q_1 = -1/3$, $q_2 = q_3 = 2/3$, а в решении (4.2.10) $\alpha_2 = \alpha_3 = \beta_1 = 0$:

$$\tilde{V} \propto t^{q_1 + (1-q_2) + (1-q_3)} = t^{1/3}. \quad (4.2.13)$$

Объем \tilde{V} расширяется медленнее, чем среда в целом, и контраст плотности в сгущении будет усиливаться со временем.

Возмущения, описываемые решением (4.2.10), не являются, вообще говоря, малыми; они во всяком случае не малы, если какое-либо из α_i полагается равным нулю в решении (4.2.10). Возмущения малы при

$$\delta_i = |\beta_i| t^{1-q_i} / \alpha_i t^{q_i} \ll 1. \quad (4.2.14)$$

При этом возмущение движения вдоль данного направления растет со временем, если

$$1 - 2q_i > 0, \quad (4.2.15)$$

и падает при выполнении обратного неравенства. В направлении, для которого показатель q_i отрицателен, возмущение является растущим. Так, при $q_1 = -1/3$, $q_2 = q_3 = 0$

$$\delta_1 \propto t^{5/3}. \quad (4.2.16)$$

Возмущение скорости есть

$$v_i = (\dot{b}_i - \dot{a}_i)\chi = \beta_i(1 - q_i)t^{-q_i}\chi. \quad (4.2.17)$$

Произведение собственной скорости v_i на соответствующий масштабный фактор есть, как следует отсюда, константа:

$$v_i a_i = \text{const.} \quad (4.2.18)$$

Относительная величина

$$\frac{v_i}{\dot{a}_i} \propto t^{1-2q_i} \propto \delta_i$$

возрастает или убывает вместе с δ . Относительное возмущение плотности (если, скажем, только один какой-либо из коэффициентов β_i отличен от нуля):

$$\frac{\delta\rho}{\rho} = \frac{\bar{\rho} - \rho}{\rho} = \frac{\frac{1}{\bar{V}} - \frac{1}{V}}{\frac{1}{V}} \propto \delta_i. \quad (4.2.19)$$

При $q_1 = -1/3$, $q_2 = q_3 = 0$ имеем

$$\frac{\delta\rho}{\rho} \propto t^{5/3}. \quad (4.2.20)$$

Из (4.2.20) видно, что развитие возмущений происходит гораздо быстрее, чем в изотропном мире с нулевым давлением, где $\delta\rho/\rho \propto t^{2/3}$ (см. § 3.1). Этот качественный результат сохраняется и в случае ультрарелятивистского уравнения состояния среды (Дорошкевич (1966)).

Рассмотренные возмущения могут, таким образом, оказаться растущими, как оказывались растущими потенциальные возмущения в изотропном мире. Однако причина этого роста в анизотропном мире существенно иная.

Если в изотропном случае нарастание плотности было обязано взаимному гравитационному притяжению частиц в области возмущения, и, следовательно, было проявлением гравитационной неустойчивости, то в анизотропном мире на вакуумной его фазе частицы являются пробными, их гравитационное взаимодействие не играет роли. Эволюция таких возмущений протекает под действием только внешнего гравитационного поля и, согласно (4.2.18), нарастание возмущений происходит только потому, что в одном из направлений происходит сжатие. Из-за этого случайное движение ускоряется фактически по той же причине, по которой увеличиваются скорости теплового движения частиц газа при адиабатическом сжатии занимаемого ими объема.

Воспользуемся теперь тем, что решение (4.2.10) позволяет описать и не малые собственные потенциальные движения вещества. Возмущения метрики при этом все же можно считать малыми и пренебрегать ими, так как само вещество считается пробным. В отсутствие собственных движений влиянием вещества можно пренебрегать, как показано в § 4.1, вплоть до сингулярности, по той причине, что плотность энергии вещества нарастает при приближении к особенности медленнее, чем «плотность энергии анизотропии», $\epsilon_a \propto t^{-2}$, которая служит мерой членов в левой, «геометрической», части уравнений (4.1.20).

При наличии немалых гидродинамических движений картина может оказаться принципиально иной. Действительно, выбирая частное решение, в котором при $q_1 = -1/3$, $q_2 = q_3 = 2/3$ коэффициенты $\alpha_1 = \beta_2 = \beta_3$ равны нулю, находим, что объем области возмущения изменяется как $t^{2(1-q_1)} = t^{5/3}$. Тогда плотность энергии в веществе без давления $\epsilon \propto t^{-3/2}$, и отношение $\epsilon/\epsilon_a \propto t^{-3/2}$ нарастает с уменьшением времени. Это означает, что условие вакуумности, $\epsilon/\epsilon_a \ll 1$, если оно выполняется с запасом в некоторый конечный момент времени в данной области среды, будет выполняться в этой области все хуже в более ранние времена. Можно сказать, что движения такого рода не вписываются в анизотропный мир, в котором материя «с самого начала» является пробной (Чернин (1972а), Мороз и Чернин (1974)). Возникает вопрос об анизотропных и, вообще говоря, неоднородных космологических моделях, в которых сингулярность имеет

не обязательно вакуумный характер или с течением времени может происходить, в частности, чередование вакуумных и невакуумных эпох. Один специальный пример анизотропной космологии с материальной, а не вакуумной особенностью, мы уже знаем — это модель, рассмотренная в предыдущем параграфе. В ней движение вещества вблизи сингулярности анизотропно, а метрика пространства-времени галилеева.

Нелинейные уравнения общей теории относительности допускают, по-видимому, несколько различных классов достаточно общих решений, в одном из которых содержатся в качестве частных решения, подобные решению Казнера, а в другом (или других) — решение Фридмана и родственные ему анизотропные и неоднородные решения с материальной особенностью. Эта возможность остается пока выясненной не до конца ввиду возникающих здесь значительных математических трудностей.

Усиление вихрей. Мы видели в § 3.2, что в космологии Фридмана существование изначальных вихревых движений невозможно. Если в некоторую эпоху в изотропном мире имеются слабые вихри, то либо это вихри не первичные, а порожденные в среде на достаточно поздней стадии расширения (см. гл. 5), либо вблизи сингулярности мир был нефридмановским. Как мы сейчас покажем, вихри могут быть «вписаны» с самого начала в анизотропную космологическую модель, которая в свою очередь должна переходить с течением времени в изотропную. При этом вихри не только сохраняются в ходе расширения, но сначала — при анизотропии расширения — могут даже усиливаться (Рубан и Чернин (1972)).

Это можно увидеть, обобщая на случай анизотропной деформации среды закон сохранения потока вихря $\text{rot } v$, или, в случае ультрарелятивистской среды, псевдовихря $M \text{ rot } v$, обсуждавшийся в § 3.2. Пусть линейные скорости вращения малы по сравнению со скоростями регулярного движения, так что «время оборота» вихря много меньше космологического времени, а размеры области, охваченной движением, малы по сравнению с расстоянием до горизонта. Эти условия обеспечивают малость возмущения и делают возможным применение упомянутых формул, полученных при рассуждениях на ньютоновском языке.

Если вихрь направлен вдоль одной из главных осей метрики Казнера, то деформация перпендикулярной ему площадки определяется масштабными факторами по поперечным направлениям. Условие сохранения потока псевдовихря в этом случае запишется в таком виде:

$$M(\text{rot } v)_i a_j a_k = \text{const}, \quad i \neq j \neq k. \quad (4.2.21)$$

Масса области, охваченной движением,

$$M = \rho V \propto \rho a_1 a_2 a_3,$$

причем при $p = 1/3\varepsilon$

$$\rho \propto V^{-1/3} = (a_1 a_2 a_3)^{-1/3}.$$

Тогда

$$(\text{rot } v)_i \propto \frac{(a_1 a_2 a_3)^{1/3}}{a_j a_k} \propto \frac{t^{1/3}}{a_j a_k}. \quad (4.2.22)$$

Если вихрь направлен вдоль оси наиболее быстрого расширения, то перпендикулярная ему площадка растягивается со временем медленно и вихрь возрастает. Например, если $q_i = 1$, $q_j = q_k = 0$, то $(\text{rot } v)_i \propto t^{1/3}$. Если воспользоваться параметрическим выражением для показателей (4.1.25), то легко увидеть, что для вихря, направленного вдоль оси «3», возрастание со временем имеет место при $-1 \leq U < -1/2$.

Усиление вихря не происходит в нерелятивистской среде. Действительно, при $p = 0$ вместо (4.2.21) имеем

$$(\text{rot } v)_i \propto \frac{1}{a_j a_k}. \quad (4.2.23)$$

Стоящая справа величина не возрастает со временем ни при каких допустимых значениях показателей в масштабных факторах.

Из анализа поведения вихрей в анизотропном мире можно сделать еще два существенных вывода. Во-первых существование вихрей возможно в анизотропном мире «с самого начала». При приближении к сингулярности вихри рассмотренного типа убывают, и поэтому, в отличие от изотропного случая, связанные с ними возмущения движения материи становятся сколь угодно малыми. Сама материя при этом, как мы видели ранее, является

пробной и не влияет на метрику вблизи сингулярности, так что вихри, свободно вписываясь в такую метрику, никак не искажают ее. Во-вторых, если изначальные вихри были сначала хаотически ориентированы по направлениям, то в ходе расширения должно происходить их выстраивание вдоль направления, которое отвечает их наиболее быстрому усилению.

Миксмастер. В изотропной космологии существует проблема причинной связи удаленных частей Вселенной. Возможность такой связи ограничена горизонтом событий (§ 1.4). Возникает вопрос: может ли горизонт событий отсутствовать в более общей космологической модели, например, в той, которая описывается метрикой (4.1.12)?

Наиболее быстрыми носителями причинной связи являются сигналы, распространяющиеся со скоростью света. Их мировые линии определяются равенством нулю интервала. Для такого сигнала, распространяющегося вдоль одного из направлений i в метрике (4.1.12), имеем

$$ds^2 = c^2 dt^2 - a_i^2(t) d\chi_i^2 = 0. \quad (4.2.24)$$

Формально наличие или отсутствие горизонта связано с характером зависимости масштабных факторов от времени: горизонт отсутствует, если $a(t)$ растет быстрее, чем t . В этом случае уравнение (4.2.24) имело бы решение

$$\chi_2 - \chi_1 = \int_0^t \frac{dt'}{a(t')}, \quad (4.2.25)$$

по которому разность лагранжевых координат частиц, связанных причинной связью за время t , считая от сингулярности $t = 0$, была бы сколь угодно большой из-за расходимости интеграла в сингулярности. В действительности, согласно решению (4.1.19) — (4.1.22) максимально быстрый рост масштабного фактора — линейный по времени, и он отвечает инерциальному разлету по одному из направлений. Для света, распространяющегося вдоль этого направления, интеграл (4.2.25) логарифмически расходится в сингулярности, и, значит, вдоль такого направления горизонт действительно отсутствует. Однако по двум другим направлениям, в которых масштабные факторы не меняются вблизи особенности, причинная связь по-прежнему ограничена горизонтом событий.

Это означает, что проблема все-таки не может быть решена в рамках столь простой метрики как (4.1.12), хотя последняя показывает, по-видимому, что надежда на ее решение не исключена.

Одной из возможностей является так называемая «миксмастер-модель» *) (Мизнер, Торн и Уилер (1973)). В этой модели сопутствующая веществу пространственная геометрия не является плоской, а похожа на геометрию закрытого мира Фридмана. Число частиц в ней конечно. Временное поведение модели на малых временных интервалах подобно казнеровскому; однако время от времени показатели степени в масштабных факторах, так сказать, циклически перебрасываются с одного направления на другое. Они остаются константами лишь на ограниченных интервалах, а, вообще говоря, являются функциями времени, так что сами масштабные факторы становятся осциллирующими функциями времени.

Подробный анализ осциллирующих космологических решений проделан Белинским, Лифшицем и Халатниковым (1970, 1971, 1972). Периоды осцилляции (рис. 11) сгущаются неограниченно по мере приближения к сингулярности, и полное их число бесконечно велико. Среди этих периодов имеются такие, что временно расширение по одному из направлений происходит линейно, а по двум другим отсутствует. Линейное (инерциальное) расширение перебрасывается в серии такого рода особых периодов с одного из направлений на другие. Так как, считая от сингулярности до любого конечного момента времени, сменилось бесконечно большое число периодов, то в каждом направлении, хотя и на короткие промежутки времени, могло происходить инерциальное расширение. Если один из этих периодов непосредственно примыкает к сингулярности, то горизонт вдоль соответствующего направления согласно (4.2.25), отсутствует. Если же он начинается через малое, но конечное время от сингулярности, то интеграл (4.2.25) конечен. Этих периодов очень много, и причинная связь может в принципе распространиться на все частицы в этой закрытой модели, так как в ней интервал значений χ конечен.

*) Миксмастер — сосуд для перемешивания, например, коктейлей. Эту модель называют иногда «перемешанным миром».

Анализ, проделанный Дорошкевичем, Лукашем и Новиковым (1971), показывает, что это достигается все же не при любом допустимом типе осциллирующего расширения, а только при некотором специальном. Не известно, почему в реальной Вселенной мог осуществляться именно такой, предпочтительный тип расширения, и проблема

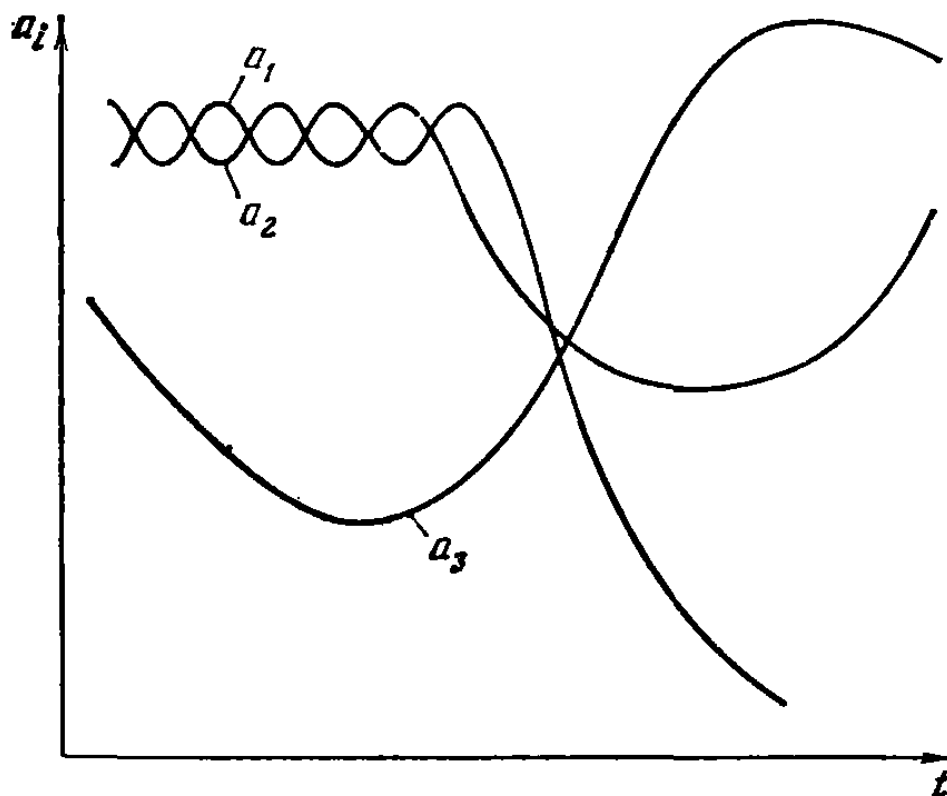


Рис. 11. Колебательный режим анизотропного расширения вблизи сингулярности.

причинной связи, сильно продвинутая благодаря привлечению анизотропной космологии, остается решенной не до конца.

Неадиабатические явления. Роль гравитации вещества, как мы видели, не слишком существенна в моделях с сильной анизотропией вблизи космологической сингулярности. Оказывается, однако, что одной только гравитацией картина не ограничивается. В последние годы выдвигаются интересные идеи о существенно неадиабатических явлениях в космологии, при которых «энергия анизотропии» диссипирует, увеличивая энтропию среды и даже число частиц в пространстве *).

*) См., например, книги Мизнера, Торна и Уилера (1973) и Зельдовича и Новикова (1975).

Во всякой среде, подвергающейся деформации, имеет место вязкость, причем, если деформация изотропна, то действует лишь вторая вязкость, а при анизотропной деформации действует также и первая вязкость, (см. § 3.3). Роль первой вязкости заключается тогда в том, что она препятствует анизотропии деформации, стремясь уменьшить эту анизотропию. Можно сказать, что та часть энергии движения среды, которая связана с анизотропией деформации («энергия анизотропии»), диссипирует благодаря первой вязкости, превращаясь в энергию теплового движения.

Это явление имеет место и в анизотропной космологии, в которой локальная деформация среды благодаря вязкости становится менее анизотропной, а вместе с тем уменьшается и анизотропия самой метрики. Микроскопический механизм этого явления можно представить себе следующим образом. Если бы частицы среды не взаимодействовали между собой, то компоненты импульса теплового движения частицы в направлениях, по которым происходит расширение, убывали бы со временем, а в направлении сжатия возрастали:

$$P_i a_i = \text{const.} \quad (4.2.26)$$

Это есть очевидное анизотропное обобщение формулы $Pa = \text{const}$, полученной в § 1.4 для изотропного расширения. Там она описывала «красное смещение» для квантов и частиц, совершающих случайные движения. Частным случаем этого общего соотношения является и формула (4.2.18).

На основании (4.2.26) можно сказать, что при анизотропном расширении импульсы в двух направлениях испытывали бы красное смещение, а в третьем — синее. Их средние квадратичные значения оказались бы в результате этого различными, а распределение стало бы анизотропным и неравновесным. Однако в действительности частицы взаимодействуют друг с другом при столкновениях и это ведет к изотропизации распределения импульсов. Энтропия при этом возрастает и происходит приближение к равновесию. Работа анизотропной деформации, которая выводит среду из термодинамического равновесия, термализуется, превращается в энергию теплового

движения. Сама же анизотропия деформации среды при этом уменьшается.

Убывание анизотропии деформации означает, что наступает некоторая несогласованность метрики и движения вещества: последнее становится менее анизотропным, чем первая. (Формально говоря, в синхронной системе отсчета, в которой описывается метрика, появляются гидродинамические движения.) Но роль вещества постепенно возрастает по мере удаления от особенности — это, вероятно, общее свойство анизотропных моделей, а не только решения, которое мы обсуждали выше. Поэтому изотропизация движения вещества должна вести со временем и к изотропизации метрики.

В вязком затухании анизотропии наибольшую роль играют в каждую эпоху частицы, у которых при той же плотности энергии больше длина свободного пробега: это могут быть фотоны, нейтрино, даже гравитоны, в зависимости от того, сколь ранний момент времени рассматривается. Процесс разыгрывается в областях среды, для которых во всяком случае выполнено условие причинной связности. Для Вселенной в целом он может быть, очевидно, в полной мере эффективным лишь тогда, когда горизонт событий отсутствует, хотя бы в том смысле, как в миксмастер-модели.

На самых ранних стадиях космологического расширения возможна, по-видимому, неадиабатичность иного рода, которая способна, вероятно, гораздо сильнее гасить анизотропию расширения, — это спонтанное рождение частиц.

Согласно одному из фундаментальных физических принципов — принципу неопределенности квантовой механики — в процессах, протекающих с характерным временем Δt , энергия сохраняется лишь с точностью до величины ΔE , определяемой соотношением

$$\Delta E \cdot \Delta t \sim \hbar.$$

В космологической задаче Δt есть не что иное, как возраст мира t , и потому квантовомеханическая неопределенность, или несохранение энергии в процессах, связанных с космологическим расширением, имеет порядок \hbar/t . В современном состоянии Вселенной эта величина пренебрежимо мала по сравнению, например, с энергией

поко: частиц. Однако при приближении к сингулярности время стремится к нулю, и неопределенность энергии неограниченно возрастает. Когда $t \sim \hbar/kT$ или \hbar/mc^2 , где T — температура, а m — масса покоя каких-то частиц, то неопределенность энергии (т. е. ее несохранение за этот промежуток времени) становится настолько большой, что фотоны и частицы могут рождаться или исчезать и притом только за счет того, что происходит космологическое расширение. Это дало основание предположить, что в начальную эпоху происходило интенсивное рождение частиц как с конечной массой покоя, например, барионов, так и с нулевой массой покоя, например, фотонов и нейтрино.

Рождение частиц может быть более интенсивным в анизотропном мире, подобно тому как вязкие процессы более эффективны при анизотропии. Частицы с нулевой массой покоя вообще могут рождаться лишь в анизотропном мире. Это связано с тем, что мировая линия частиц с нулевой массой покоя определяется уравнением $ds^2 = 0$. В изотропном мире последнее сводится к уравнению того же вида, что и в галилеевом мире:

$$ds^2 = a^2(t) (d\eta^2 - d\chi^2).$$

Но при таких условиях неадиабатичность не может играть роли. В отличие от этого, в анизотропном случае такое, как говорят, конформно-галилеево представление интервала невозможно. Если же частицы имеют конечную массу покоя, то интервал вдоль их мировой линии отличен от нуля, и масштабный множитель, зависящий от времени, не сокращается, так что ситуация к галилеевой не сводится.

Рождение частиц нарастает по мере приближения к сингулярности. С другой стороны, появление все большего числа частиц делает более существенным влияние материи на метрику. Это влияние, как мы видели, всегда ведет к изотропизации метрики. Можно сказать, что на рождение частиц тратится «энергия анизотропии», подобно тому как она диссипирует из-за вязкости. Рождение частиц в анизотропном мире — это в каком-то смысле проявление вязкости деформируемого вакуума. Возможно (Зельдович (1970б)), что рождение частиц уничтожает анизотропию расширения «с самого начала» расширения

за характерное планковское, или квантово-гравитационное, время

$$t_{\text{п}} \sim (G\hbar/c^5)^{1/2} \approx 10^{-43} \text{ сек.}$$

Эта величина получается из соображений размерности: если рассматривать скорость света в вакууме c , гравитационную постоянную G и постоянную Планка \hbar как независимые фундаментальные величины, то, по замечанию Планка, они составят полный набор, т. е. через их комбинации могут быть выражены величины любой размерности. Так, к $t_{\text{п}}$ добавляются элементарная квантово-гравитационная длина

$$l_{\text{п}} = ct_{\text{п}} \approx 10^{-33} \text{ см,}$$

плотность

$$\rho_{\text{п}} = (Gt_{\text{п}}^2)^{-1} \approx 10^{94} \text{ г/см}^3,$$

и масса

$$m_{\text{п}} = (\hbar c/G)^{1/2} \approx 10^{-5} \text{ г.}$$

Эти величины дают представление о временных и пространственных промежутках, на которых можно ожидать проявления квантовых эффектов в гравитации.

Нужно, однако, отметить, что затронутые здесь проблемы квантовых эффектов в космологии еще далеко не разработаны; эта интересная и важная область современной астрофизики находится в самом начале развития.

ГИДРОДИНАМИЧЕСКИЕ ПРОЦЕССЫ В МЕТАГАЛАКТИЧЕСКОЙ СРЕДЕ И ОБРАЗОВАНИЕ СКОПЛЕНИЙ ГАЛАКТИК

§ 5.1. Активная фаза космогонического процесса

«Мне кажется, что если бы вещество нашего Солнца и планет и вообще все вещество Вселенной было бы равномерно распределено по всему небесному пространству, а каждая частица испытывала бы врожденное тяготение ко всем остальным, и полный объем пространства, в котором рассеяно это вещество, был бы конечным, то вещество снаружи этого объема стремилось бы благодаря тяготению ко всему тому веществу, что внутри, и потому падало бы к центру этого пространства и составляло бы там единую большую сферическую массу. Но если бы это вещество было равномерно распределено по бесконечному пространству, оно никогда не слилось бы в единую массу; какая-то его часть сгущалась бы в одну массу, а другая — в другую, так что возникло бы бесконечное число больших масс, разбросанных по всему такому бесконечному пространству на большие расстояния друг от друга. И так могли бы образоваться Солнце и неподвижные звезды, если предположить еще, что по своей природе это вещество было светящееся. Но вот каким образом вещество могло бы разделиться на две части и та, что годится на это, слилась в светящееся тело, тогда как другая осталась темной или превратилась в темную, когда первая осталась неизменной,— это уже, на мой взгляд, нельзя объяснить одними только естественными причинами и я должен приписать это мысли и действию исполненного волей Творца».

Это отрывок из письма Ньютона (10 декабря 1692 года) Ричарду Бантли, ректору Тринити Колледжа в Кембридже. Будучи, как, по-видимому, и его адресат, человеком верующим, Ньютон разделил проблему на то, что подлежит и что не подлежит научному исследованию и объяснению. К последнему он отнес вопрос, почему звезды (светящееся вещество) светят, а планеты (темное вещество)

— нет. Это действительно трудный вопрос, и только через два с половиной века источник светимости звезд и Солнца был, наконец, понят и объяснен, но для этого понадобилась ядерная физика вместе со специальной теорией относительности и квантовой механикой.

Однако первая сторона проблемы — формирование небесных тел из однородного вещества — оказалась у Ньютона принципиально раскрытой. Значительна прежде всего сама идея эволюционности Вселенной. «Современная астрономия стала насквозь эволюционной... эволюция наблюдается на всех уровнях. Активные образования на поверхности Солнца, кометы, планетные атмосферы, газовые туманности, звезды, пульсары, ядра галактик, сами галактики, — наконец, — вся Вселенная как целое — эволюционируют с характерными для каждого случая масштабами времени» (Шкловский (1975)). Но во времена Ньютона не было никаких фактических указаний на что-либо подобное. Идея и механизм эволюции родились из закона всемирного тяготения и оказались определяющими для всего дальнейшего развития космогонии. Тяготение формировало космическую структуру, а ее усложнение и дифференциация дали начало эволюции совсем другого рода, в результате которой материя, существовавшая 15—18 миллиардов лет назад в состоянии однородно распределенной водородно-гелиевой плазмы, достигла высокой организации — во Вселенной возникла жизнь, а потом и разум, изучающий, среди прочего, и свою предысторию.

Механизм формирования небесных тел из однородно распределенного вещества, о котором говорит Ньютон, есть не что иное, как гравитационная неустойчивость. Фрагментация вещества происходит у Ньютона под действием одних только гравитационных сил в предположении, что никакие другие силы, если они вообще имеются, этому не препятствуют. Интересно при этом противопоставление двух случаев — конечного и бесконечного пространств (или, вернее, объемов), заполненных веществом. Бесконечность однородного распределения означает равноправность всех точек пространства и потому нет никакой выделенной точки, к которой как к центру могло бы обратиться «все вещество Вселенной». Единого центра для всей Вселенной нет, а однородное распределение неустойчиво и не может оставаться таковым сколь угодно

долго. Значит, должно быть бесконечно много центров, вокруг которых собираются отдельные массы вещества. С другой стороны, бесконечность объема означает фактически, что его пространственный размер очень велик по сравнению с каким-то другим размером, служащим в качестве характерного масштаба, а конечность не означает ничего, кроме соизмеримости или близости этих двух размеров. Если размер области распределения много больше характерного масштаба, то произойдет разбиение вещества на фрагменты; если же такого сильного неравенства нет, то масса останется единой и будет сжиматься как целое. Это, вероятно, и подразумевал Ньютон, но что скрывается за характерным масштабом длины, удалось выяснить только Джинсу в 1902 г. Пространственный масштаб — длина Джинса — задается силами, препятствующими гравитации, в первую очередь силой давления, обязанной упругости вещества (см. § 3.1).

Следующий шаг в теории гравитационной неустойчивости был сделан Лифшицем в 1946 г., когда он решил задачу устойчивости однородного и изотропного мира в линейном приближении (гл. 3). Основой для этого послужила космологическая модель Фридмана (гл. 1).

Открытие реликтового излучения (гл. 2) наполнило теорию гравитационной неустойчивости новым богатым содержанием, а измерения изотропии излучения позволили зафиксировать уровень отклонений от однородности и изотропии в одну конкретную эпоху эволюции Вселенной — эпоху рекомбинации, когда излучение «отключается» от вещества. В эту эпоху (ей отвечает красное смещение ~ 1000) относительная амплитуда возмущений плотности и скорости была заметно меньше единицы (§ 3.4). Такое слабо возмущенное состояние метagalактической среды может быть изучено, как мы видели, весьма полно, и, по-видимому, нет ни одного вопроса, касающегося поведения малых возмущений, на который нельзя было бы ответить, опираясь на теорию, развитую в гл. 3.

Из этой теории следует, в частности, вывод о существовании определенного рода догалактической структуры во Вселенной: малые по сравнению с единицей относительные амплитуды возмущений в эпоху рекомбинации (и даже раньше) должны были на множество порядков превосходить уровень статистических флуктуаций в среде

(§ 3.1). Мы знаем также, что сразу после «отключения» излучения от вещества случайные слабые потенциальные и вихревые движения, наложенные на общее расширение среды, могли сохраниться лишь в масштабах, отвечающих массам $\sim 10^{13} M_{\odot}$ и больше, тогда как в меньших масштабах они были подавлены диссипативными процессами (§ 3.3). В малых масштабах могли «выжить» лишь статические энтропийные возмущения, которые после рекомбинации способны превратиться в потенциальные возмущения (§ 3.1).

Гораздо меньше ясности в том, что касается природы возмущений, служащих «затравками» для гравитационной неустойчивости. Это вопрос совсем другого рода (Ньютон, наверное, отнес бы его к категории трансцендентных), для ответа на который требуется как минимум построение последовательной квантовой теории гравитационных явлений вблизи космологической сингулярности (см. § 4.2). Мы не знаем, были ли возмущения малыми «с самого начала», т. е. во все более ранние времена до эпохи рекомбинации, или, может быть, Вселенная в начале своего существования обладала сильной анизотропией, а возмущения в эпоху рекомбинации представляли собой остатки этой анизотропии, затухавшей на ранних фазах космологического расширения (гл. 4). Мы не можем «вывести» характеристики возмущений из первых принципов физики и вынуждены поэтому прибегать к тем или иным гипотезам относительно их типа, уровня их амплитуды, зависимости последней от масштаба (т. е. спектра возмущений) и т. д.

Характер предполагаемых возмущений в значительной степени определяет всю дальнейшую картину формирования космических структур. Если, например, возмущения в эпоху рекомбинации обязаны совместным движениям плазмы и излучения, существовавшим в более ранние эпохи, образование структур неизбежно начинается с самой крупной единицы — скопления галактик, масса которого превосходит границу диссипативного затухания. Если же допускаются и энтропийные возмущения, то этот процесс — в зависимости от спектра возмущений — мог бы начаться как со скоплений галактик, так и с шаровых скоплений звезд, масса которых на восемь порядков меньше (см. ниже).

Гравитационная неустойчивость и переход к активной фазе космогонического процесса. Каковы бы ни были предыстория или конкретный вид возмущений, служащих в эпоху рекомбинации элементами догалактической структуры, их дальнейшая эволюция должна так или иначе вести ко все более заметным локальным отклонениям от однородности и изотропии. В определенную эпоху эти отклонения становились сильными и создавали в среде сгущения вещества, способные дать начало галактикам и скоплениям галактик.

В этом смысле можно говорить об активной фазе космогонического процесса в расширяющемся мире, о переломном этапе превращения догалактической структуры в совокупность наблюдаемых космических тел. Это произошло не раньше, чем при красном смещении $z \sim 100$, которому отвечает средняя плотность Вселенной, совпадающая с типичной плотностью галактик, $\rho_G \approx 10^{-24} \text{ г/см}^3$. Мы увидим (§ 5.3), что эту нижнюю границу эпохи (или верхнюю границу соответствующего ей красного смещения) можно, по-видимому, сдвинуть до $z \approx 10$, если принять во внимание физические условия в метagalacticкой среде на активной фазе космогонического процесса.

В эту эпоху линейная теория гравитационной неустойчивости уже не применима, но момент перехода к состоянию сильных возмущений можно оценить по формулам этой теории на пределе их применимости, когда возмущения плотности $\delta\rho$ и скорости v достигают уровня

$$|\delta\rho| \approx \rho, v \approx V \approx \frac{\lambda}{t} \approx 5 \cdot 10^7 \Omega^{1/2} (1+z)^{1/2} \left(\frac{M}{10^{13} M_\odot} \right)^{1/2} \text{ см/сек.} \quad (5.1.1)$$

Здесь

$$t = \begin{cases} 5 \cdot 10^{17} \Omega^{-1/2} (1+z)^{-1/2} \text{ сек,} & z > 2/5 \Omega^{-1}, \\ 6 \cdot 10^{17} (1+z)^{-1} \text{ сек,} & z < 2/5 \Omega^{-1} \end{cases} \quad (5.1.2)$$

— текущий возраст мира, выраженный через красное смещение в соответствующую эпоху (см. §§ 1.2, 1.4 здесь принято $H = 50 \text{ км/сек Мпс}$); в формуле (5.1.1) и в дальнейших оценках (где обратное специально не оговаривается) считается, что $z \gtrsim 2/5 \Omega^{-1}$; хаббловская скорость V связана с массой M через масштаб λ :

$$\lambda(z) = \left(M \left| \frac{4\pi}{3} \rho \right. \right)^{1/2} \approx 10^{26} \Omega^{-1/2} (1+z)^{-1} \left(\frac{M}{10^{13} M_\odot} \right)^{1/2} \text{ см.}$$

Для указанной оценки нужно знать не только законы развития возмущений во времени, но и «начальные данные», значения величин $\delta\rho$ и ζ в эпоху рекомбинации. Верхние их пределы могут быть найдены, как говорилось, по наблюдательным ограничениям вариаций температуры реликтового фона (§ 3.4).

Случайные, хаотические движения, скорости которых определяются соотношением (5.1.1), могут быть достаточно сильными, чтобы преодолеть и скомпенсировать общее космологическое расширение в отдельных объемах метagalактической среды. Соответствующие массы обособляются таким образом от общего фона и начинают свою собственную индивидуальную эволюцию, мало зависящую от продолжающегося расширения Вселенной в целом. Существенно, что скорости движений $v \gtrsim V$, много больше скорости звука в метagalактической среде. В эпоху рекомбинации скорость звука составляла

$$u(z_{\text{rec}}) \approx \left(\frac{kT_{\text{rec}}}{m} \right)^{1/2} \approx 5 \cdot 10^5 \text{ см/сек}, \quad (5.1.3)$$

а после этого, примерно до $z \approx 100$, она изменялась по закону

$$u \propto a^{-1/2} \propto (1+z)^{1/2}, \quad (5.1.4)$$

так как из-за остаточной ионизации могло поддерживаться еще равенство температур вещества и излучения (см. § 2.1). При этом температура падала по закону

$$T \propto a^{-1} \propto (1+z),$$

характерному для излучения, ибо его теплоемкость много больше теплоемкости вещества. Заметим, что собственные тепловые скорости частиц, которые с точностью до численного коэффициента всегда совпадают со скоростью звука, изменяются здесь не так, как это было описано в § 1.4: они падают при расширении медленнее, ибо вещество в этот период подогревается фотонами. Позднее, при красных смещениях, меньших 100, подогрев становится недостаточным, и скорость звука в веществе падает по закону

$$u \propto T^{1/2} \propto a^{-1} \propto (1+z), \quad (5.1.5)$$

соответствующему показателю адиабаты вещества и общему соотношению для собственных движений пробных

частиц. Например, при красном смещении ≈ 10 скорость звука, вычисляемая в соответствии с (5.1.5), составляет

$$u(z = 10) \approx 10^4 \text{ см/сек}, \quad (5.1.6)$$

что много меньше скорости, даваемой соотношением (5.1.1) для масштаба галактики или скопления при том же красном смещении.

Масштаб галактики значительно превышает характерную джинсову длину во всю эпоху после рекомбинации. Как мы видели в § 3.1, джинсовой длине соответствует при рекомбинации масса

$$M_J(z_{\text{rec}}) \approx (10^{-5} - 10^{-6})M_{\odot}.$$

В период от z_{rec} до $z \approx 100$ величина этой массы,

$$M_J \propto \rho(ut)^3,$$

оставалась неизменной, как это следует из формулы (5.1.4). При $z < 100$ она падала в соответствии с формулой (5.1.5):

$$M_J \propto (1 + z)^{3/2}.$$

Значит, потенциальные движения масштаба галактик и скоплений (и даже масштаба звездных шаровых скоплений) усиливались после рекомбинации гравитационной неустойчивостью, так что их скорости возрастали, а с ними росло и отношение собственной скорости движений к хаббловской скорости. Это относится как к «исконно» потенциальным движениям, так и к потенциальным движениям, которые возникли после рекомбинации из первоначальных энтропийных возмущений.

С вихревыми движениями дело обстоит не так: их скорости падают после рекомбинации,

$$v_v \propto a^{-1} \propto (1 + z), \quad (5.1.7)$$

а в соответствии с этим падает и отношение

$$\frac{v_v}{V} \propto t^{-1} \propto (1 + z)^{3/2}. \quad (5.1.8)$$

Отсюда следует, что условие (5.1.1) не может быть выполнено для вихрей в более позднюю эпоху, если оно не выполнялось сразу при рекомбинации. Вероятнее всего, «изначальные» вихревые движения были слабее

при рекомбинации (если они вообще существовали; см. § 5.6).

Возмущения доджинсовых масштабов могут только затухать после рекомбинации; соответствующие им акустические скорости убывают по тому же закону, что и вихревые скорости.

Роль гидродинамических движений в космогоническом процессе. В последующих параграфах этой главы мы будем изучать общие свойства сильных гидродинамических движений в метagalacticкой среде, предполагая, что именно им принадлежит ведущая роль на активной фазе космогонического процесса. Характерной чертой таких движений является в первую очередь существование разрывов — ударных волн, тангенциальных разрывов и т. п. Мы будем привлекать в нашем рассмотрении выясненные в гидродинамике (Ландау и Лифшиц (1953), Зельдович и Райзер (1966), Седов (1961, 1967)) сведения об этих явлениях, применяя их к особым физическим условиям метagalacticкой среды. Мы будем считать, что движения имели безвихревой потенциальный характер и охватывали массы вещества, соответствующие группам и скоплениям галактик; они стали сильными в смысле, определяемом соотношениями (5.1.1), благодаря развитию гравитационной неустойчивости в послерекомбинационную эпоху. Мы будем считать также, что вместе с ними в метagalacticкой среде имелись в ту же эпоху и более слабые возмущения, соответствующие по массе галактикам и звездным скоплениям. Движения разных масштабов развивались независимо, пока нелинейные эффекты их взаимодействия были несущественны. С переходом к сильно возбужденному состоянию среды нелинейные гидродинамические эффекты превращаются в главный фактор космогонического процесса. В частности, именно им могли бы быть обязаны своим происхождением вихревые движения, возникающие в эту эпоху в метagalacticкой среде; такие вихри способны, по-видимому, обеспечить вращательные движения галактик (§§ 5.4, 5.6).

Общие выводы о характере физических процессов во Вселенной с сильными локальными возмущениями, к которым мы придем ниже, не зависят сами по себе от тех или иных конкретных гипотез о предыстории космической структуры и могут служить основой для построе-

ния различных по своим предпосылкам космогонических схем. Весьма вероятно, однако, что формирование крупномасштабной структуры Вселенной действительно началось с обособления самых больших газовых масс — протогрупп и протоскоплений галактик, — а затем произошла их фрагментация на газовые протогалактики, где в свою очередь возникли газовые комплексы, в которых рождались звезды. Представление о протоскоплении как о сгущении вещества, созданном крупномасштабными ударными волнами в метagalacticкой среде, выдвинуто Зельдовичем (1970а) и Сюняевым и Зельдовичем (1972) (см. § 5.5).

В пользу такой картины говорит целый ряд аргументов и в первую очередь, по-видимому, то очень важное обстоятельство, что подавляющее большинство галактик во Вселенной входит в группы или скопления. По данным Воронцова-Вельяминова (1972), доля одиночных галактик во всяком случае не больше 10—15% от числа объектов ярче некоторого предела; возможно, что «общее поле галактик состоит из сходящихся на нет по своей четкости скоплений и групп». В работе Фесенко (1976б) на основании большого статистического материала показано, что так называемые изолированные галактики являются скорее всего периферийными членами близких к ним рассеянных (неправильных) систем галактик.

Важно также, что согласно недавним исследованиям рентгеновского излучения различных внегалактических объектов (см. обзор Филда (1974)), многие скопления галактик содержат горячий межгалактический газ, причем иногда масса газа сравнима с суммарной массой галактик в богатых скоплениях.

Скопления галактик являются основной единицей крупномасштабной структуры Вселенной. По внешнему виду их разделяют на правильные и неправильные (Эйбл (1958), Рудницкий (1964), Воронцов-Вельяминов (1972)). Первые имеют более или менее правильную сферическую форму и регулярное, падающее наружу, распределение плотности с четко выраженным преобладанием наиболее массивных галактик в центральной области скопления. Таково, например, наиболее крупное из изученных скоплений — скопление в Волосах Вероники, или Кома. Суммарная масса галактик в богатом скоплении типа Комы составляет $\sim (1 \div 0,3) \cdot 10^{15} M_{\odot}$, радиус скопления

$\sim (1 \div 2) \cdot 10^{25}$ см $\approx 3-6$ Мпс. Средняя плотность, считая галактики и близкий к ним по общей массе горячий газ, $\sim 10^{-27} - 10^{-28}$ г/см³. Средняя скорость движения галактик (относительно центра скопления) $\sim 10^8$ см/сек.

Неправильные скопления не обнаруживают регулярного строения. Наиболее крупные из них, такие как скопление в Деве, содержат галактики с суммарной массой $\sim (1 \div 3) \cdot 10^{14} M_{\odot}$; их средняя плотность близка к плотности богатых правильных скоплений или ниже, а скорость движений галактик в несколько раз меньше.

Указанные количественные характеристики скоплений представляют собой не более как ориентировочные оценки, содержащие значительную неопределенность. Однако важным качественным и не вызывающим сомнений фактом является наличие двух типов скоплений. Особенно существенно, что имеется определенная корреляция между общим строением скопления и морфологическим типом преобладающих в нем галактик: в правильных скоплениях обнаруживаются прежде всего эллиптические галактики, а в неправильных — спиральные. Напомним, что спиральным галактикам присуще, как правило, весьма быстрое вращение (что возможно, и предопределило существование в них плоских подсистем со спиральными рукавами; см. § 6.1), — тогда как вращение эллиптических галактик чаще всего гораздо слабее. В этом смысле различие в типах скоплений можно рассматривать как указание на различие физических условий, и в первую очередь характера гидродинамических движений в соответствующих участках метагалактической среды на активной фазе космогонического процесса (§§ 5.5, 5.6).

Космогонические гипотезы. Космогоническую картину, в которой ведущая роль принадлежит крупномасштабным гидродинамическим движениям, а формирование структур начинается с обособления протоскоплений галактик, мы назвали вероятной, но ее нельзя считать единственно возможной на современном этапе исследований. Теоретически допустимы и другие космогонические схемы, включающие в себя, например, помимо гравитационной неустойчивости и гидродинамики также и другие физические процессы. Мы не будем обсуждать все известные по текущей литературе предложения и ограничимся тем, что в заключение этого параграфа кратко

изложим две наиболее интересные, по нашему мнению, гипотезы такого рода.

В одной из них, выдвинутой Дорошкевичем, Зельдовичем и Новиковым (1967), ключевым моментом являются взрывы сверхмассивных звезд, создающие в окружающей их среде сильные возмущения. Предполагается, что в начальном состоянии Вселенной имелись возмущения энтропийного типа, причем их относительная амплитуда была в среднем тем больше, чем меньше масштаб возмущений (падающий с масштабом спектр). После рекомбинации они растут благодаря гравитационной неустойчивости и быстрее всего становятся сильными возмущения, которым отвечает наименьшая возможная масса, т. е. джинсова масса $\sim (10^5 - 10^6)M_{\odot}$ (см. выше). Возникающие таким путем сгущения вещества превращаются, как полагают, в сверхмассивные звезды, или «празвезды».

Строение сверхмассивных звезд существенно отличается от строения обычных звезд типа Солнца. В сверхмассивной звезде давление равновесного излучения значительно превосходит давление газа, и именно излучение обеспечивает равновесие звезды, противодействуя силе тяжести. Разумеется, это равновесное излучение не имеет ничего общего с реликтовыми фотонами горячей Вселенной; оно возникает в сжимающемся газовом облаке, которое разогревается благодаря переходу в тепло части гравитационной потенциальной энергии его частиц. В равновесии давление излучения превосходит газовое давление, если масса звезды больше некоторой критической.

Последнее вытекает из простых оценок. Из баланса сил давления и гравитации

$$F_p \approx M \frac{p}{\rho R}, \quad F_G \approx \frac{GM^2}{R^2},$$

где M , R , ρ — масса, размер и средняя плотность звезды, а p — давление в ее недрах, находим

$$p \approx \rho \frac{GM}{R} \approx \frac{GM^2}{R^4}.$$

Если давление создается излучением, то

$$p \approx kT \left(\frac{kT}{\hbar c} \right)^3,$$

и характерная температура недр звезды

$$kT \approx (\hbar c)^{3/4} G^{1/4} M^{1/2} R^{-1}.$$

Газовое давление при этой температуре

$$\frac{\rho}{m} kT \approx G^{1/4} (\hbar c)^{3/4} m^{-1} M^{3/2} R^{-4}$$

меньше лучистого, если

$$M > m \left(\frac{\hbar c}{G m^2} \right)^{3/2} \approx M_{\odot}.$$

Радиус звезды не вошел в выражение для критической массы, которая представляется комбинацией одних только универсальных констант.

Более точный расчет, учитывающий распределение плотности и давления в звезде, дает, как показал Эддингтон (1926), примерно в сто раз больше значение критической массы.

Дорошкевич, Зельдович и Новиков (1967) предположили, что сверхмассивные звезды испытывают быструю эволюцию, в ходе которой выделяется, и, возможно, со взрывом, значительная часть их ядерной энергии. Как показал Бисноватый-Коган (1968), сверхмассивная звезда указанной массы ($\sim 10^5 \div 10^6 M_{\odot}$) действительно может испытать ядерный взрыв; звезды же бóльших масс (не вращающиеся) в конце своей эволюции неудержимо сжимаются под действием собственного тяготения, которому ядерное энерговыделение уже не способно препятствовать.

Примерно одна десятитысячная доля всего вещества должна быть заключена в быстро эволюционирующие сверхмассивные звезды, и тогда их ядерная энергия будет достаточна, чтобы разогреть остальной газ до температуры $10^5 - 10^6$ °К. В результате эволюции «празвезд» в метagalактической среде возникают возмущения масштаба галактик или скоплений галактик. При этом образование новых «празвезд» уже прекращается, а джинсовой длине соответствует при подходящих красных смещениях масса, близкая к типичной массе галактики. Вновь созданные возмущения большого масштаба и не слишком малой амплитуды усиливаются гравитационной неустойчивостью и, как предполагается, превращаются затем в галактики или квазары,

В другой картине, предложенной Пиблсом и Дикке (1968), нет ни взрывов, ни сильных гидродинамических движений. Исходные возмущения считаются энтропийными, но они порождают не сверхмассивные звезды, а шаровые скопления звезд. Джинсова длина в эпоху рекомбинации (и после нее, до красных смещений ~ 100 ; см. выше) отвечает массам вещества, которые действительно близки к массам реальных шаровых скоплений.

Как и в картине «празвезд», считается, что средняя амплитуда возмущений падает с масштабом, так что в послерекомбинационную эпоху наиболее существенны как раз возмущения с массой шаровых скоплений. Гравитационная неустойчивость превращает соответствующие сгущения вещества в сравнительно плотные облака. При этом не возникает ударных волн (из-за близости масштаба возмущений к джинсовой длине их формированию препятствует дисперсия — это объясняется в § 5.2). Облака не вращаются, так как любые вращательные движения, если они вообще имелись когда-либо, затухают в этом масштабе задолго до рекомбинации. Это согласуется с правильной сферической формой наблюдаемых шаровых скоплений; в противном случае вращение вызывало бы их уплощение.

Дальнейшая эволюция каждого облака сопровождается, согласно Пиблсу и Дикке, охлаждением (преимущественно благодаря возбуждению и высвечиванию уровней энергии молекул водорода), в результате чего становится возможной фрагментация газового облака на мелкие сгущения — протозвезды. Эта фрагментация также обязана гравитационной неустойчивости, развивающейся внутри облака. С другой стороны, гравитационная неустойчивость, развивающаяся в «газе», частицами которого являются сами эти облака как целое, ведет к их скучиванию в галактики, а галактики собираются затем в скопления.

Вся совокупность наблюдательных данных о звездном составе шаровых скоплений указывает на то, что они представляют собой наиболее старые образования во Вселенной. Это обстоятельство и послужило исходным мотивом гипотезы Пиблса и Дикке. Хотя галактики сильно различаются по массам, размерам, видимой форме, составляющие их звезды в основном различаются между собою гораздо слабее. Звездное население сферической

составляющей нашей Галактики весьма похоже на звезды эллиптических галактик, а население плоской составляющей галактики — на звезды неправильных галактик. Это, по-видимому, указывает на универсальность процесса звездообразования во всей Метагалактике, на его слабую чувствительность к вариации конкретных физических условий в различных галактиках. Сходство между галактиками проявляется также и в том, что во многих из них действительно обнаруживаются шаровые звездные скопления. Эти образования в нашей Галактике населяют сферическую составляющую и различаются между собой довольно слабо. Шаровые скопления в некоторых близких эллиптических и неправильных галактиках весьма похожи на шаровые скопления нашей Галактики.

В картине Пиблса и Дикке универсальность шаровых скоплений и их большой возраст находят прямое объяснение. Однако ряд других важных свойств звездных систем объяснить на этом пути не удастся. В первую очередь это относится к таким свойствам галактик, как их типичная масса и плотность. Кроме того, в последнее время появились возражения и против исходного пункта всей картины: по мнению Ван ден Берга (1975), шаровые скопления все же значительно различаются по своим свойствам. Содержание металлов в их звездах коррелирует с положением шаровых скоплений в галактике (эту трудность отмечали и сами авторы гипотезы), в других галактиках металличность шаровых скоплений коррелирует со светимостью (или массой) галактики. Трудно ожидать, наконец, что подавляющее число шаровых скоплений могло бы собраться в галактики; скорее, многие из них должны были бы остаться в межгалактическом пространстве, чего на самом деле не наблюдается.

Далеко не все ясно также в физических процессах на нелинейной фазе гравитационной неустойчивости облаков; остается не доказанным, например, что облака действительно распадаются на звезды, а не превращаются в единые тела — сверхмассивные звезды, фигурирующие в картине Дорошкевича, Зельдовича и Новикова.

Эти две космогонические гипотезы представляют собой построенные на основании астрономических наблюдений и физической теории эволюционные схемы, способные дать представление о возможной последовательности сс

бытий, приведших к наблюдаемой космической структуре. В § 5.7 мы говорим и о концепции Амбарцумяна, не связанной с идеей гравитационной неустойчивости; она не «вписывается», по-видимому, в теорию горячей Вселенной, но и не нуждается в ней как в основании или фоне. Каждая из этих гипотез выбирает в качестве исходного, ключевого, тот или иной астрономический факт или физический процесс, но такой выбор остается неоднозначным. Это относится и к картине сильных гидродинамических движений, которой мы отдаем здесь предпочтение (она является сейчас, по-видимому, наиболее разработанной).

В 1919 году Джинс писал: «При современном состоянии наших знаний любая попытка диктовать окончательные заключения по основным проблемам космогонии была бы ничем иным, как чистым догматизмом». Астрономические знания с тех пор невероятно возросли, огромные изменения произошли и в теоретической физике, но они не отменили это соображение.

§ 5.2. Сильные возмущения

Образования разрывов. Хорошо известный пример нелинейных эффектов в гидродинамике — «опрокидывание» звуковых волн.

Когда по среде распространяется звук, то в ней возникают и движутся состояния сгущения и разрежения. Говоря о скорости звука в среде, имеют в виду величину

$$u = \left(\frac{\partial p}{\partial \rho} \right)^{1/2}, \quad (5.2.1)$$

вычисляемую в пренебрежении слабыми неоднородностями при постоянной энтропии. Это соответствует линейному приближению в теории распространения звука. В действительности, однако, в областях уплотнения величина возрастает, а в областях разрежения убывает, как это следует из адиабатической связи между давлением и плотностью. В нерелятивистской среде с показателем адиабаты $\gamma = 5/3$

$$u \sim \rho^{1/2}. \quad (5.2.2)$$

и это учесть, что и будет соответствовать нелинейному

приближению в теории звука, то можно представить себе следующую картину. Пусть в среде распространяется звуковая волна, и в некоторый начальный момент состояния сгущения и разрежения распределены по синусоидальному закону. В линейном приближении это распределение состояний среды остается неизменным и происходит лишь его перемещение в пространстве. Однако такая неизменность только приближенна, и при более точном рассмотрении мы должны учесть изменения в распределении сгущений и разрежений. В той полуволне, где плотность падает в направлении распространения, профиль волны будет становиться все более крутым, т. е. уменьшение плотности будет происходить на все меньшем расстоянии. При этом состояние максимального уплотнения (гребень волны) распространяется в наиболее уплотненной им самой среде, и потому скорость его распространения наибольшая. Состояния перед гребнем распространяются по среде меньшей плотности и потому с меньшей скоростью. К этому добавляется еще различие самих скоростей в гребне и впадине. Гребень будет, так сказать, догонять эти состояния, и крутизна профиля рассматриваемой полуволны будет нарастать. В той полуволне, где плотность нарастает в направлении распространения, происходит обратное изменение профиля, и он будет становиться все более пологим. Можно сказать, что в результате такого процесса в волне появляются более высокие гармоники, обертона.

Изменение профиля волны будет происходить до тех пор, пока расстояние, на котором уменьшается плотность, не станет сравнимым по порядку величины с длиной свободного пробега частиц, так что гидродинамическое приближение окажется неприменимым. С макроскопической точки зрения это означает, что возникнет скачок, или разрыв, плотности, а с ней скорости и других макроскопических величин. Такой распространяющийся разрыв называют ударной волной *).

В ударной волне рождение новых гармоник из-за гидродинамической нелинейности компенсируется диссипативным затуханием, которое тем эффективнее, чем выше час-

*) Никакой периодичности движения в такой «волне», в отличие от звуковой, нет.

тота (или короче длина волны). Поэтому в среде может установиться стационарное разрывное течение.

Разрыв перемещается по находящемуся перед ним газу со скоростью, превышающей скорость звука в этом газе. Действительно, скорость распространения скачка близка к скорости звука на гребне породившей его звуковой волны, или, что то же, к скорости звука непосредственно за скачком. Эта скорость, очевидно, больше скорости звука в той полуволне, где происходило уменьшение плотности. В системе отсчета, связанной со скачком, газ, находящийся перед скачком, натекает на него со скоростью, которая для этого газа является сверхзвуковой.

Нелинейные гидродинамические процессы характеризуются временем

$$t_n \approx \lambda/v, \quad (5.2.3)$$

где v — скорость в данном масштабе λ . Это следует, например, из простой оценки членов гидродинамического уравнения движения (уравнения Эйлера):

$$\frac{\partial v}{\partial t} + (v\nabla)v = f.$$

Здесь f — полная сила, действующая на единицу массы среды. В линейном приближении по скорости левая часть уравнения содержит только производную по времени. Пусть скорость v изменяется на величину, сравнимую с ней самой, на расстоянии λ (в этом, собственно, и состоит смысл пространственного масштаба движения); тогда нелинейный член в уравнении Эйлера по порядку величины можно записать в виде

$$(v\nabla)v \sim v^2/\lambda.$$

Пусть, далее, гидродинамическая скорость в данном масштабе изменяется из-за нелинейности на величину, сравнимую с ней самой, за время t_n ; тогда

$$\frac{\partial v}{\partial t} \sim v/t_n.$$

Сравнение показывает, что нелинейный член имеет тот же порядок величины, что и линейный, т. е. нелинейность не мала, а существенна, когда t_n дается выражением (5.2.3).

В условиях расширяющейся метagalактической среды мерой эффективности нелинейного гидродинамического взаимодействия является отношение времени (5.2.3) к текущему возрасту мира, который задает темп изменения со временем общих условий во Вселенной. На активной фазе космогонического процесса скорость движений, наложенных на общее космологическое расширение, не меньше хаббловской скорости (5.1.1), так что гидродинамическое время не превышает текущий возраст мира.

«Опрокидывание» звуковых волн в расширяющейся метagalактической среде затрудняется из-за увеличения длины волны $\lambda \propto (1+z)^{-1}$, в результате чего гребню становится, так сказать, труднее догнать впадину. Тем не менее это может все же в принципе произойти. Например, на стадии преобладания излучения скорость акустических колебаний (в доджинсовых масштабах), как и скорость вихрей, остается постоянной; постоянна и скорость звука:

$$v = \text{const}, \quad u = c/\sqrt{3} \quad (5.2.4)$$

(при $\rho_r \gg \rho$ это есть следствие общего закона для собственных движений в среде; § 1.4). Кроме того, что скорость звука не меняется из-за космологического расширения, она не меняется еще и от сгущения к разрежению в звуковой волне. Гребень в этом случае движется быстрее впадины только из-за различия акустической скорости в этих состояниях волны. Возрастание нелинейности такого рода характеризуется увеличением и стремлением к единице отношения

$$\frac{v}{V} \propto V^{-1} \propto (1+z)^{-2}, \quad z > z_*. \quad (5.2.5)$$

Другими словами, для достижения гребнем впадины требуется конечное время или конечное число колебаний в волне:

$$\int_{t_1}^{t_2} \frac{v}{\lambda} dt \propto (t_2^{1/2} - t_1^{1/2}).$$

Нелинейность в этих условиях ведет реально к затуханию звуковых волн из-за диссипации их энергии, когда характерный масштаб перепада скорости в них (т. е. рас-

стояние между гребнем и впадиной) становится все меньше и рано или поздно оказывается меньше характерной длины диссипативного затухания (§ 3.3):

$$\lambda_v \approx \lambda_x \approx (ct/\sigma n)^{1/2} \propto (1+z)^{-3/2}. \quad (5.2.6)$$

В эпоху преобладания вещества, после рекомбинации, нелинейность этого типа не развивается, так как отношение собственной скорости к хаббловской падает для акустических волн в доджинсовых масштабах.

Эволюционная трансформация звуковых волн — не единственный механизм образования гидродинамических разрывов. Однако во всех случаях они возникают лишь при сверхзвуковом режиме движения газовых масс. Разрыв может возникнуть в результате эволюции слабого возмущения, когда ударные волны образуются в длительном процессе, либо при движении в среде со сверхзвуковой скоростью плотных тел, т. е. когда возмущение с самого начала не малó. К тому же результату может привести и движение облаков газа в разреженной среде со скоростью, которая для этой среды является сверхзвуковой. Ударные волны могут образоваться при столкновениях таких облаков, если их относительная скорость больше скорости звука в материале облаков. Разрывы такого происхождения называют неэволюционными.

Пусть, например, облака более или менее одинаковы и их столкновение — лобовое. В первый момент при соприкосновении внешних слоев облаков образуется тонкий слой уплотненного газа. Этот слой состоит из частиц, находившихся на поверхности облаков и затормозившихся при столкновениях на длине свободного пробега. На этот слой будет затем набегать слева и справа со сверхзвуковыми скоростями газ обоих облаков. Толщина уплотненной области будет при этом возрастать, а на ее границы с обеих сторон будет в возрастающем количестве набегать газ со сверхзвуковыми скоростями и испытывать торможение. Обе границы будут удаляться от первоначального слоя соприкосновения, распространяясь в обоих облаках и охватывая все большие их массы. Они будут двигаться навстречу набегающему газу и их скорость по отношению к газу будет обязательно сверхзвуковой. Картина становится, таким образом, подобной той, к которой мы пришли, рассматривая образование ударных волн при эволю-

ции слабого возмущения — звуковой волны. Конечным результатом столкновения двух облаков, движущихся друг относительно друга со сверхзвуковой скоростью, является образование возрастающей со временем области уплотнения, отделенной от окружающего ее газа фронтами ударных волн. Но этим картина не исчерпывается и мы вернемся к ней в § 5.6.

Парадокс Эрншоу и дисперсия «тяжелого звука». Отличие сильных возмущений от слабых ярко демонстрируется классическим результатом, известным как парадокс Эрншоу: при изоэнтропийных колебаниях газа плоские стационарные потенциальные волны конечной амплитуды невозможны. В линейном приближении, когда амплитуды считаются малыми, такие стационарные колебания существуют — это звуковые волны без затухания. Но в нелинейном приближении стационарности уже не может быть.

Доказательство проводится от противного. Предположим, что такие волны допустимы; тогда в системе отсчета, движущейся вместе с волной, уравнения движения (без гравитации) и неразрывности имеют вид

$$v \frac{dv}{dx} = - \frac{1}{\rho} \frac{dp}{dx}, \quad (5.2.7)$$

$$\frac{d}{dx}(\rho v) = 0. \quad (5.2.8)$$

Исключая из первого уравнения скорость

$$v = \frac{A}{\rho}, \quad A = \text{const}, \quad (5.2.9)$$

найденную из второго, имеем

$$dp = \frac{A^2}{\rho^2} d\rho. \quad (5.2.10)$$

Теперь интегрируем и находим давление как функцию плотности:

$$p = B - \frac{A^2}{\rho}, \quad B = \text{const} > 0. \quad (5.2.11)$$

Но соотношение (5.2.11) никак не похоже на адиабатическую связь давления с плотностью для нормальных га-

зов. Значит, стационарные изоэнтропийные волны конечной амплитуды действительно не существуют. Ударные волны конечной амплитуды могут быть стационарными, но в них происходит диссипация кинетической энергии, и энтропия возрастает (см. ниже).

Хотя при доказательстве не учитывалась гравитация среды, парадокс Эрншоу, несомненно, остается в силе для движений, масштаб которых значительно меньше джинсовой длины: в этом случае самогравитация заведомо не существенна. Ситуация, как оказывается, может измениться, когда масштаб движений сравним с джинсовой длиной. В этом случае самогравитация газа способна при определенных условиях не допустить «опрокидывания» звуковой волны.

Качественно роль гравитации в волновом процессе можно проследить на примере «тяжелого звука», распространяющегося в однородной среде. Если частота звука высока и его период мал по сравнению с характерным гравитационным временем $(G\rho)^{-1/2}$ (которое дает временную шкалу общего расширения однородной среды), то гравитацией (и общей нестационарностью) можно пренебречь; в нулевом приближении по малой величине отношения периода к гравитационному времени (или длины волны к джинсовой длине) мы имели бы обычный звук. В этом нулевом приближении частота звука связана с волновым числом $k = 2\pi/\lambda$ соотношением $\omega(k) = uk$, так что фазовая, ω/k , и групповая, $d\omega/dk$, скорости распространения волны равны между собой и совпадают со скоростью звука. Но в приближении, учитывающем малую величину $\lambda/R_J \approx (kR_J)^{-1}$, связь между частотой и волновым числом становится иной. Представление о ней дает соотношение

$$\omega^2 = u^2k^2 - 4\pi G\rho,$$

о котором мы упоминали в § 3.1 в связи с задачей Джинса о звуке в гравитирующей среде. Хотя это соотношение и не является вполне корректным по указанной в § 3.1 причине, оно приводит к правильной величине критической длины гравитационной неустойчивости и потому им можно воспользоваться также и для качественной оценки влияния гравитации на звук в приближении $\lambda/R_J < 1$

(ср. Каплан и Пикельнер (1974)). В этом случае

$$\frac{\omega}{k} \approx u \left(1 - \frac{2\pi G\rho}{u^2 k^3} \right),$$

$$\frac{d\omega}{dk} \approx \frac{u^2 k}{\omega}.$$

Отсюда видно, что имеет место, как говорят, дисперсия тяжелого звука, т. е. зависимость скорости его распространения (фазовой и групповой) от волнового числа. Но это означает, что гравитация вносит рассогласованность в распространение основной гармоники звука и более высоких гармоник, рождающихся благодаря нелинейности. Она нарушает их синфазность, ибо высокие гармоники отстают по фазе от основной. Тем самым затрудняется укрупнение фронта волны, и при определенных условиях ее опрокидывание может быть предотвращено. Для этого нужно, чтобы дисперсия была достаточно сильной и «отводила» рождающиеся высокие гармоники в достаточно быстром темпе. Если дисперсия и нелинейность одинаково эффективны и способны компенсировать одна другую, то волна может стать стационарной и распространяться с постоянной скоростью без изменения формы. Такая нелинейная волна отличается от ударной волны отсутствием необратимых процессов диссипации, сохранением энтропии.

Дисперсия, обязанная гравитации, существенна, если длина волны приближается к джинсовой *). Пример нелинейной стационарной волны в некоторой простой модели самогравитирующей среды построен и проанализирован в работе Антонова и Чернина (1977).

Ударные фронты. Ударным волнам масштаба скоплений или групп галактик принадлежит, по-видимому, важная роль в космогоническом процессе на активной его фазе. Поэтому мы кратко изложим сейчас выясненные в гидродинамике и необходимые для дальнейшего сведения об общих свойствах ударных фронтов.

Прежде всего на фронте всякой ударной волны должны выполняться три условия сохранения: потоков массы, импульса, энергии.

*) Это относится, в частности, к эволюции облаков в картине Пиблса и Дикке (§ 5.1).

Будем отмечать значком «1» величины, характеризующие состояние газа перед фронтом, и значком «2» — за фронтом. Удобно рассматривать скорость газа по отношению к перемещению скачка, т. е. к системе отсчета, связанной с фронтом ударной волны. Будем считать сначала, что скорость газа перпендикулярна к плоскости фронта (так называемые прямые скачки). Невозмущенный газ натекает на скачок со скоростью v_1 , а возмущенный газ уходит от скачка со скоростью v_2 . Значит, в единицу времени к единице площади фронта ударной волны приходит масса газа $\rho_1 v_1$, а за фронтом от него уходит масса $\rho_2 v_2$, где ρ_1, ρ_2 обозначают плотности газа перед фронтом и за ним. Масса, естественно, не исчезает и не рождается на скачке, поэтому

$$\rho_1 v_1 = \rho_2 v_2. \quad (5.2.12)$$

Это есть выражение первого из указанных трех условий.

Условие сохранения потока импульса имеет несколько более сложный вид. Импульс передается как макроскопическим, так и микроскопическим (тепловым) движением. Макроскопический поток импульса на единицу поверхности равен плотности потока частиц nv (где n — концентрация частиц), помноженной на импульс отдельной частицы mv :

$$mv \cdot nv = \rho v^2. \quad (5.2.13)$$

Микроскопический поток импульса, передаваемого тепловым движением, выражается через давление газа p . В некотором направлении, которое мы временно обозначим буквой « x », плотность потока частиц со случайной тепловой скоростью u_x равна $1/2 n(v) u_x$, где $n(v)$ есть число частиц со скоростью v ; множитель $1/2$ введен потому, что в среднем половина частиц движется в положительном, а половина в отрицательном направлении оси x . Так как речь идет о микроскопическом движении, то этот поток «затухает» на некотором расстоянии порядка длины свободного пробега; это означает, что частицы потока как бы отражаются от окружающей массы газа, изменив направление своей скорости на обратное, так как иначе в направлении x имел бы место средний макроскопический поток. При таком «отражении» каждая частица потока отдает окружающим частицам импульс $2u_x m$,

а следовательно, весь поток отдает в единицу времени единице площади окружающего газа импульс $n(v)tu_x$. Суммируя это по всем микроскопическим скоростям, найдем, что единица площади газа получает в единицу времени импульс

$$nt \langle u_x \rangle^2. \quad (5.2.14)$$

Импульс, передаваемый единице площади в единицу времени, есть действующая на нее сила. В данном случае это давление газа p , которое и равно, таким образом, микроскопической части плотности потока импульса. Поэтому полная плотность потока импульса есть

$$\rho v^2 + p. \quad (5.2.15)$$

Протекая через фронт волны, газ не изменяет своего импульса; даже если существуют какие-либо внешние силы, они не успеют оказать заметного действия на протяжении ширины скачка. Поэтому поток импульса одинаков по обе стороны скачка, что приводит ко второму условию непрерывности:

$$\rho_1 v_1^2 + p_1 = \rho_2 v_2^2 + p_2. \quad (5.2.16)$$

Наконец, существует третье условие непрерывности, следующее из сохранения энергии. Движущийся элемент газа переносит как энергию макроскопического движения, так и внутреннюю энергию. Плотность потока этих двух величин есть

$$\rho v \left(\frac{v^2}{2} + \frac{\epsilon_T}{\rho} \right),$$

где ϵ_T/ρ — тепловая энергия единицы массы газа. Однако, кроме этого, элемент газа оказывает давление на окружающую среду и при своем перемещении производит работу. Работа, произведенная элементом газа, есть

$$\int p dV = \int p dr ds = \int p v dt ds,$$

где dV — элемент объема газа, $dr = v dt$ — его перемещение за время dt , ds — элемент поверхности. Работа, совершаемая в единицу времени и рассчитанная на единицу площади, равна, следовательно, $p v$. Она производится за счет части энергии данного элемента газа, кото-

рая при этом передается лежащим впереди элементам газа. Так как работа давления тоже приводит к переносу энергии, полная плотность потока энергии есть:

$$\rho v \left(\frac{v^2}{2} + \frac{\epsilon_T}{\rho} \right) + pv = \rho v \left(\frac{v^2}{2} + w \right),$$

где $w = (\epsilon_T + p)/\rho$ — тепловая функция, или энтальпия, единицы массы. При этом расчете мы не учитывали возможность превращения энергии макроскопического движения в тепло, т. е. возрастание энтропии вещества, и считали течение по обе стороны скачка адиабатическим.

Поток энергии непрерывен при переходе газа через скачок по той же причине, по какой непрерывен и поток импульса. Следовательно, третье условие непрерывности имеет вид

$$\rho_1 v_1 \left(\frac{v_1^2}{2} + w_1 \right) = \rho_2 v_2 \left(\frac{v_2^2}{2} + w_2 \right). \quad (5.2.17)$$

С учетом непрерывности потока массы находим

$$\frac{v_1^2}{2} + w_1 = \frac{v_2^2}{2} + w_2. \quad (5.2.18)$$

Три условия непрерывности позволяют получить ряд важных следствий *). Обозначим через j постоянную величину плотности потока массы:

$$j = \rho_1 v_1 = \rho_2 v_2.$$

Тогда условия непрерывности потоков импульса и энергии можно записать следующим образом:

$$p_1 - p_2 = j^2 \left(\frac{1}{\rho_2} - \frac{1}{\rho_1} \right), \quad (5.2.19)$$

$$w_1 - w_2 = \frac{1}{2} j^2 \left(\frac{1}{\rho_2^2} - \frac{1}{\rho_1^2} \right). \quad (5.2.20)$$

Деля второе из этих уравнений на первое, получим:

$$\frac{w_1 - w_2}{p_1 - p_2} = \frac{1}{2} \left(\frac{1}{\rho_2} + \frac{1}{\rho_1} \right), \quad (5.2.21)$$

*) Они могут быть легко обобщены на случай косою падения газа на плоский фронт ударной волны или на случай искривленного фронта. При этом нормальная компонента скорости газа претерпевает разрыв, а тангенциальная остается непрерывной (см., например, Ландау и Лифшиц (1953)).

или

$$w_2 - w_1 = \frac{1}{2} (p_1 - p_2) \left(\frac{1}{\rho_1} + \frac{1}{\rho_2} \right). \quad (5.2.22)$$

Если известно уравнение состояния газа, т. е. зависимость p от ε_T , а следовательно, и от w , то последние соотношения дают возможность определить связь между давлением p_2 и плотностью ρ_2 газа за фронтом при заданных p_1 и ρ_1 .

Когда изменение величин ρ , p , w в скачке мало, то в первом приближении можно положить

$$\frac{1}{2} \left(\frac{1}{\rho_1} + \frac{1}{\rho_2} \right) \approx \frac{1}{\rho}.$$

Тогда уравнение (5.2.22) примет вид

$$dw = \frac{dp}{\rho}, \quad (5.2.23)$$

где dw и dp — малые изменения соответствующих величин. Но, как известно, общее выражение для дифференциала энтальпии имеет вид

$$dw = \frac{dp}{\rho} + T ds, \quad (5.2.24)$$

где T — температура, s — энтропия единицы массы. Сравнивая это с (5.2.23), мы видим, что если скачки величин p , w и ρ малы, то скачок энтропии есть величина более высокого порядка малости, которой поэтому можно пренебречь. Значит, в этом приближении все четыре потока — массы, импульса, энергии и энтропии — оказываются непрерывными при прохождении через достаточно слабый скачок. Этот вывод перестает быть справедливым, когда скачок не мал. В таком случае переход через скачок неадиабатичен уже в первом приближении, и скачок энтропии должен быть столь же большим, как и скачок p , w , ρ .

Наличие скачка энтропии имеет важное значение. Так как энтропия может только возрастать, то ее скачок при прохождении через разрыв может быть только в сторону увеличения. Можно показать, что в таком случае скачки давления и плотности должны быть также в направлении их увеличения. Именно с таким направлением изменений величин в ударной волне мы встретились в предыду-

щем разделе, когда рассматривали «опрокидывание» звуковой волны. Соотношение между величинами, характеризующими состояние вещества перед фронтом и за ним, определяется уже не адиабатой, а уравнением (5.2.21) или (5.2.22), которое теперь заменяет адиабату и которое называют ударной адиабатой, или адиабатой Гюгонио.

Скачки, определяемые этой адиабатой, обладают одной важной особенностью. В то время как скачок давления (т. е. отношение p_2/p_1) может быть в принципе сколь угодно велик, скачок плотности, а следовательно, и скачок скорости, всегда конечен и ограничен довольно узкими пределами. Чтобы показать это, воспользуемся тем обстоятельством, что движение по обе стороны разрыва происходит, как мы уже говорили, адиабатически. При этом мы ограничимся случаем классического идеального газа, теплоемкость которого постоянна. Тогда адиабатичность означает, что величины p_1 , ρ_1 , w_1 связаны соотношениями:

$$p_1 = \alpha_1 \rho_1^\gamma,$$

где $\gamma = c_p/c_v$ есть отношение теплоемкостей при постоянном давлении и постоянном объеме, α_1 — коэффициент пропорциональности, который выпадает из окончательных формул, и

$$w_1 = \frac{p_1 + \varepsilon_{T1}}{\rho_1} = \frac{p_1}{\rho_1} + \frac{c_v n_1 T_1}{\rho_1} = (c_v + 1) \frac{p_1}{\rho_1}.$$

Величина c_v есть теплоемкость на одну частицу, причем $c_p - c_v = 1$. Поэтому

$$w_1 = c_p \frac{p_1}{\rho_1} = \frac{c_p}{c_p - c_v} \frac{p_1}{\rho_1} = \frac{\gamma}{\gamma - 1} \frac{p_1}{\rho_1}.$$

Аналогичные соотношения (с заменой α_1 на α_2) имеют место для величин p_2 , ρ_2 , w_2 . Так как теплоемкость газа постоянна, показатель γ одинаков по обе стороны разрыва. Для одноатомного газа $\gamma = 5/3$.

Подставим теперь выражение энтальпии через давление и плотность в ударную адиабату (5.2.22). Тогда для скачка плотности получим выражение:

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma + 1) p_2 + (\gamma - 1) p_1}{(\gamma - 1) p_2 + (\gamma + 1) p_1}. \quad (5.2.25)$$

При сильном скачке давления, когда $p_2 \gg p_1$ и давлением перед разрывом можно пренебречь, мы получаем:

$$\frac{\rho_2}{\rho_1} \approx \frac{\gamma + 1}{\gamma - 1}. \quad (5.2.26)$$

Для одноатомного газа максимальное отношение ρ_2/ρ_1 равно 4. Вследствие непрерывности плотности потока j скачок скорости в разрыве

$$\frac{v_2}{v_1} = \frac{\rho_1}{\rho_2} < 1. \quad (5.2.27)$$

Значит, вещество оттекает от разрыва с меньшей скоростью, чем притекает к нему. Мы сейчас покажем, что скорость оттекания обязательно является дозвуковой. Для простоты ограничимся случаем сильного разрыва. Скорость звука (адиабатическая)

$$u = \left(\gamma \frac{p}{\rho} \right)^{1/2}. \quad (5.2.28)$$

Отсюда следует, что давление

$$p = \frac{1}{\gamma} \rho u^2.$$

Полагая в формуле (5.2.19) $j^2 = \rho_2^2 v_2^2$, учитывая формулу (5.2.25) и пренебрегая p_1 по сравнению p_2 , получим:

$$v_2^2 = \frac{\gamma - 1}{2} \frac{p_2}{\rho_2} = \frac{\gamma - 1}{2\gamma} u_2^2. \quad (5.2.29)$$

Отсюда видно, что при $\gamma > 1$ (когда эта формула вообще возможна)

$$\frac{v_2}{u_2} < 1. \quad (5.2.30)$$

Газ, проходя через фронт ударной волны, изменяет скачком свои характеристики; как мы видим, он сжимается, нагревается и теряет скорость. Его дальнейшая эволюция существенно зависит от характера тепловых процессов за фронтом. Для ударных волн на активной фазе космогонического процесса очень важно, что газ за их фронтами вновь оказывается ионизованным. При этом может (для не слишком малых красных смещений z) восстановиться

его взаимодействие с реликтовым излучением. Становится возможным также охлаждение газа благодаря излучению плазмы при тормозных процессах. Этот вопрос мы рассмотрим в следующем параграфе.

§ 5.3. Ударные волны в метagalacticкой среде

Характер гидродинамических движений в ударных волнах на активной фазе космогонического процесса определяется физическими условиями в метagalacticкой среде. Мы увидим, что благодаря этому появляются некоторые выделенные значения масштабов и масс, отвечающих типичным характеристикам галактик и их скоплений.

Ударные волны с сильным охлаждением. Когда на активной фазе космогонического процесса метagalacticкий газ разогревается и оказывается ионизованным в сильной ударной волне, между ним и уже довольно холодным к тому времени реликтовым излучением возникает значительный перепад температур. Поэтому появившиеся в газе свободные электроны начинают взаимодействовать с излучением посредством комптоновского рассеяния и передавать ему тепловую энергию; через посредство электронов и ионы также будут отдавать энергию излучению. Характерное время этого процесса найдено в § 2.2:

$$\tau_{er} = \frac{m_e c}{\sigma_T \epsilon_r} \approx 9 \cdot 10^{19} (1 + z)^{-4} \text{ сек.} \quad (5.3.1)$$

Охлаждение эффективно, если указанное время меньше характерного гидродинамического времени (5.2.3):

$$\tau_{er} < t_h \approx \lambda/v, \quad (5.3.2)$$

а так как последнее не должно в рассматриваемую эпоху превышать космологическое время (§ 5.2), то условие эффективности (5.3.2) может выполняться при

$$z > z_{er} \approx 10\Omega^{1/5}. \quad (5.3.3)$$

Кроме комптоновского охлаждения, обязанного реликтовому излучению, имеются и другие механизмы потери энергии газом: излучение газа при свободно-свободных и свободно-связанных переходах, а также

высвечивание в линиях (§ 2.2). Существенное отличие от комптоновского процесса состоит в том, что для последнего характерное время зависит лишь от красного смещения z , тогда как характерные времена процессов излучения в плазме определяются столкновениями ее частиц за фронтом волны, зависящими от ее температуры и плотности:

$$\tau = \frac{1}{n\sigma u}, \quad (5.3.4)$$

где n , u — концентрация и тепловая скорость сталкивающихся частиц, σ — сечение соответствующих процессов, которое зависит от температуры. В частности, для свободно-свободных переходов

$$\sigma \rightarrow \sigma_{ff} \approx \sigma_T \frac{(E_i m_e c^2)^{1/2}}{kT},$$

где E_i — энергия ионизации водорода. Этот процесс преобладает над другими тормозными процессами при температуре $T > T_{ff} \approx 5 \cdot 10^5$ °К. Характерное время для него:

$$\tau_{ff} \approx 3,6 \cdot 10^{11} (T \text{ °К})^{1/2} (\rho/m)^{-1}. \quad (5.3.5)$$

Если газ прозрачен для испускаемого им излучения, как в интересующих нас условиях, то потеря энергии будет происходить из всего объема газа, и τ_{ff} есть время, за которое он существенно охладится. (При непрозрачности излучение диффундирует в заполненной газом среде; охлаждение при этом происходит лишь с границ нагретой области и требует поэтому значительно большего времени.)

Времена охлаждения в любом случае гораздо больше времени, за которое происходит скачок гидродинамических величин в узком переходном слое с толщиной порядка длины свободного пробега частиц. Поэтому непосредственно за фронтом движение можно считать адиабатическим, и соотношения на скачке, выведенные в предыдущем параграфе, остаются в силе. Дальнейшее движение газа за фронтом существенно зависит от отношения времени охлаждения к гидродинамическому времени. Если это отношение велико, то охлаждением можно пренебречь, и состояние газа, прошедшего через скачок,

остаётся неизменным в течение всего характерного гидродинамического времени. За это время ударную волну можно считать стационарной. Ударную волну без теплопотерь называют адиабатической.

Дело обстоит иначе, когда указанные времена охлаждения малы по сравнению с гидродинамическим временем. В этом случае газ, нагретый на скачке, постепенно охлаждается по мере удаления от фронта. На некотором расстоянии от фронта его температура оказывается столь низкой ($\sim 10^4$ °К), что газ рекомбинирует, затем его атомы перейдут в основное состояние, и охлаждение, обязанное указанным выше процессам, прекратится. Дальнейшее течение газа будет происходить без потерь тепла, и его скорость, а также плотность и температура будут меняться затем в более медленном темпе, определяемом общим характерным гидродинамическим временем процесса. Ударные волны с сильным охлаждением принято называть изотермическими.

Значения установившихся величин плотности, скорости и давления в изотермической ударной волне можно получить, используя уравнения сохранения для плотности потока массы и импульса (для зоны, где $T = T_a \approx \approx 10^4$ °К, будем отмечать эти величины значком «а»):

$$\rho_1 v_1 = \rho_2 v_2 = \rho_a v_a, \quad (5.3.6)$$

$$p_1 + \rho_1 v_1^2 = p_2 + \rho_2 v_2^2 = p_a + \rho_a v_a^2. \quad (5.3.7)$$

Выражая v_a через v_2 с помощью (5.3.6) и подставляя во второе равенство в (5.3.7), получим

$$\left(\frac{\rho_a}{\rho_2}\right)^2 - \left(\frac{u_2}{u_1}\right)^2 \left[\left(1 + \gamma \frac{v_2^2}{u_2^2}\right) \frac{\rho_a}{\rho_2} + \gamma \frac{v_2^2}{u_2^2} \right] = 0. \quad (5.3.8)$$

Мы учли, что

$$p_i = \frac{1}{\gamma} \rho_i u_i^2, \quad i = 1, 2, a \quad (5.3.9)$$

(здесь, конечно, нет суммирования!)

В уравнении (5.3.8) в интересующем нас случае сильного охлаждения отношение ρ_a/ρ_2 между плоскостями «2» и «а» должно быть много больше единицы, что сейчас будет непосредственно показано. В то же время скорость

v_2 , как мы уже знаем, должна быть меньше u_2 . Поэтому

$$\frac{\rho_a}{\rho_2} \approx \left(\frac{u_2}{u_a} \right)^2 \approx \frac{T_2}{T_a}, \quad (5.3.10)$$

и это отношение на самом деле может значительно превышать единицу.

Во всей области за фронтом движение дозвуковое, хотя скорость звука и падает между плоскостями «2» и «а»:

$$\frac{v_a}{u_a} = \frac{\rho_2}{\rho_a} \frac{v_2}{u_a} = \frac{u_a}{u_2} \frac{v_2}{u_2} \ll 1. \quad (5.3.11)$$

Поэтому приближенно

$$\rho_a \approx \rho_2 \approx \rho_1 v_1^2. \quad (5.3.12)$$

Газ охлаждающийся по мере удаления от фронта, «поджимается» настолько, что, как мы видим, движение за фронтом происходит при почти постоянном давлении, почти изобарически.

Пользуясь формулами предыдущего параграфа, можно найти полный перепад плотности во всей области от «1» до «а»:

$$\frac{\rho_a}{\rho_1} = \frac{\gamma + 1}{\gamma - 1} \frac{\rho_a}{\rho_2} \approx \frac{\gamma + 1}{\gamma - 1} \frac{u_2^2}{u_a^2} \approx \frac{v_1^2}{u_a^2}. \quad (5.3.13)$$

Итоговое сжатие газа определяется, таким образом, квадратом отношения скорости набегающего на фронт потока к скорости звука в конечном состоянии.

Оценим теперь размер l_c области, в которой происходит охлаждение. Воспользуемся уравнением для плотности потока энергии. Вследствие теплопотери эта величина, в отличие от потока массы и импульса, не сохраняется. Если единица объема газа, находящаяся на расстоянии x от границы «1», теряет за единицу времени энергию $q(x)$, то полная потеря энергии за единицу времени, рассчитанная на единицу площади фронта, есть

$$Q = \int_1^a q(x) dx. \quad (5.3.14)$$

Поэтому уравнение для плотности потока энергии имеет

теперь вид

$$\left[\left(w_1 + \frac{v_1^2}{2} \right) - \left(w_a + \frac{v_a^2}{2} \right) \right] j = Q. \quad (5.3.15)$$

Как мы выяснили, при сильном охлаждении

$$v_a^2 \ll v_1^2, \quad w_a \approx \frac{p_a}{\rho_a} \approx u_a^2 \ll v_1^2.$$

Кроме того, для сильного скачка $v_1^2/2 \gg w_1$. Пренебрегая малыми величинами, получим из (5.3.15) следующее порядковое соотношение:

$$\rho_1 v_1^3 \approx Q. \quad (5.3.16)$$

С той же точностью потери энергии

$$q(x) \approx \frac{p_2}{\tau_c} \approx \rho_1 \frac{v_1^2}{\tau_c}.$$

Учитывая, что давление за фронтом практически не меняется за время охлаждения $\sim \tau_c$, вынесем p_2 за знак интеграла в (5.3.14). Тогда

$$Q \approx \rho_1 v_1^2 \int_0^{l_c} \frac{dx}{\tau_c(x)}. \quad (5.3.17)$$

Подставляя это в (5.3.16), получим

$$\int_0^{l_c} \frac{dx}{\tau_c(x)} \approx v_1. \quad (5.3.18)$$

Если газ охлаждается посредством комптоновского рассеяния электронов на реликтовом фоне, то время $\tau_c = \tau_{er}$ не зависит от его плотности и температуры. В этом случае

$$l_c \approx v_1 \tau_c. \quad (5.3.19)$$

По порядку величины это соотношение, как можно ожидать, приближенно справедливо и в общем случае. Так как $\tau_c < t$, длина l_c меньше масштаба движения λ . С другой стороны, эта длина сравнима с расстоянием, пройденным звуковой волной за время охлаждения:

$$l_c \approx u \tau_c. \quad (5.3.20)$$

Под u следует подразумевать некоторую среднюю скорость звука, сравнимую со скоростью u_2 непосредственно за фронтом; скорости же u_2 и v_1 не слишком отличаются друг от друга (для сильной ударной волны $u_2 \approx 1/2 v_1$).

Временные границы активной фазы космогонического процесса. Сделаем теперь некоторые численные оценки для скоростей и масштабов движений, характерных для активной фазы космогонического процесса. Прежде всего покажем, что ударные волны масштаба галактик и скоплений галактик, если они образуются при $z > z_{er}$, не могут быть адиабатическими.

Для быстрого комптоновского охлаждения в эту эпоху достаточно, чтобы газ оказался ионизованным. Ионизация газа имеет место при температурах на фронте $T_s \gtrsim 10^4$ °К. Если исходная скорость движения есть v , то в сильной ударной волне при показателе адиабаты $\gamma = 5/3$ она ослабляется, согласно (5.2.26), в четыре раза после прохождения через фронт; это значит, что на нагрев тратится $3/4 v$, что соответствует кинетической энергии на один атом $1/2 m (3/4 v_1)^2$. Эта энергия распределяется (после нагрева и ионизации) между электроном и ионом (или, с учетом гелия, между ионом и двумя электронами, но такое уточнение не слишком существенно), которые вместе приобретают тепловую энергию $3kT_s$. Отсюда следует связь между температурой газа за фронтом и гидродинамической скоростью набегающего потока:

$$T_s \approx 0,1 \frac{mv^2}{k} \text{ °К.} \quad (5.3.21)$$

Если $v \approx V \approx \lambda/t$, то, с учетом связи между масштабом и массой, $M = \frac{4\pi}{3} \rho \lambda^3$, находим

$$T_s \approx 5 \cdot 10^5 \Omega^{1/3} (1+z) \left(\frac{M}{10^{13} M_\odot} \right)^{2/3} \text{ °К.} \quad (5.3.22)$$

Эта температура превосходит 10^4 °К, если

$$M > M_a \approx 10^{10} \Omega (1+z)^{-3/2} M_\odot. \quad (5.3.23)$$

Массы $M_a \lesssim 3 \cdot 10^8 \Omega^{2/3} M_\odot$ меньше при $z > z_{er}$ типичных масс галактик и, значит, ударные волны масштаба скоплений, если они формировались при $z > z_{er}$, оказывались бы заведомо изотермическими благодаря быстрому ох-

лаждению на реликтовых фотонах. Согласно (5.3.13), они сжимали бы газ до плотностей

$$\rho_a \approx 4\rho \left(\frac{T_s}{T_a} \right) \approx 3 \cdot 10^{-28} \Omega^{1/2} (1+z)^4 \left(\frac{M}{10^{13} M_\odot} \right)^{2/3} \text{ г/см}^3. \quad (5.3.24)$$

Для ударных волн, охватывающих массы газа, сравнимые с массами скоплений галактик ($\sim 10^{13} - 10^{15} M_\odot$), эта плотность составила бы при $z \geq z_{er}$ величину $\approx 3 \cdot 10^{-24} - 8 \cdot 10^{-23} \text{ г/см}^3$, что заведомо превосходит типичные плотности скоплений галактик ($\leq 10^{-26} - 10^{-27} \text{ г/см}^3$) и даже типичную плотность галактик ($\sim 10^{-24} \text{ г/см}^3$). Это, очевидно, означает, что формирование газовых протогалактик в ударных волнах не могло начаться раньше, чем при красных смещениях $z < z_{er} \approx 10 \div 6$.

Для масс галактик $\sim 10^{11} M_\odot$ плотность (5.3.24) меньше характерной величины 10^{-24} г/см^3 при $z < 20 \div 15$, так что изолированные протогалактики в ударных волнах не могли возникать ранее этой эпохи. Указанные значения красного смещения определяют, таким образом, нижнюю временную границу активной фазы космогонического процесса в горячей Вселенной *).

Выделенные масштабы. В эпоху $z < 10$ комptonовское охлаждение уже не существенно для гидродинамических процессов, развивающихся быстрее общего космологического расширения; главную роль в охлаждении метагалактического газа играет излучение прозрачной плазмы, и скорость теплоотвода существенно зависит от температуры и плотности среды, а значит, и от масштаба движения. Мы покажем сейчас, что в масштабах, отвечающих скоплениям галактик, охлаждение газа за фронтом ударной волны происходит медленнее космологического расширения (и, следовательно, медленнее общего изменения гидродинамических характеристик движения), так что в эпоху $z < 10$ такие ударные волны являются адиабатическими.

Согласно (5.3.22) в ударной волне масштаба скопления ($M = 10^{13} - 10^{15} M_\odot$) температура превосходит значение

*) Это согласуется с результатами подсчетов далеких радиоисточников (радиогалактических квазаров), которые указывают на то, что эти объекты рождались преимущественно при $z = 2 \div 4$ (Дорошкевич, Лонгейр и Зельдович (1970)).

$T_{ff} \approx 5 \cdot 10^5$ °К, выше которого охлаждение плазмы, как говорилось, происходит в основном благодаря излучению при свободно-свободных переходах. Температуре $T_s = T_{ff}$ соответствует масса

$$M_{ff} \approx 10^{13} \Omega^{-1/2} (1+z)^{-3/2} M_{\odot}. \quad (5.3.25)$$

В интервале $10 > z > 2$ эта величина принимает значения $\sim 3 \cdot 10^{11} - 2 \cdot 10^{12} M_{\odot}$, сравнимые с массами наиболее крупных галактик. В ударной волне, охватывающей массу (5.3.25), время охлаждения

$$\tau_{ff}(M_{ff}) \approx 10^{19} \Omega^{-1} (1+z)^{-3} \text{ сек}, \quad (5.3.26)$$

заметно больше текущего космологического возраста t при красных смещениях $z < 7 \Omega^{-1/3}$. Время охлаждения тем больше, чем больше масштаб:

$$\tau_{ff} \approx 10^{19} \Omega^{-5/6} (1+z)^{-5/2} \left(\frac{M}{10^{13} M_{\odot}} \right)^{2/3} \text{ сек}, \quad (5.3.27)$$

и адиабатичность заведомо имеет место для ударных волн масштаба скоплений или групп галактик во всю эпоху после $z = z_{er}$. Поэтому средняя начальная плотность газового протоскопления равна учетверенной космологической плотности в эпоху формирования ударных волн:

$$4\rho(z) \approx 2 \cdot 10^{-29} \Omega (1+z)^3 \text{ г/см}^3. \quad (5.3.28)$$

Эта плотность не превосходит типичную плотность скоплений при $z < z_{er}$. Адиабатичность ударных волн масштаба скоплений не означает, конечно, что охлаждение в них вообще не происходит; газ, разогретый за их фронтами, постепенно остывает, но этот процесс медленнее общего изменения гидродинамических характеристик движения (в первую очередь скорости натекания свежего газа на фронт). Охлаждение может охватывать либо весь газ, либо только плотные области в нем (см. §§ 5.4, 5.5); во втором случае определенная доля газа могла бы оставаться в разреженном и горячем состоянии вплоть до современной эпохи. Эта доля не мала и составляет, скажем, половину полной массы M , если

$$M \gtrsim M_{\text{gas}} \approx 10^9 \Omega^{3/2} (1+z)^{15/2} M_{\odot}. \quad (5.3.29)$$

Величина M_{gas} соответствует равенству времени охлаждения (5.3.27) и современного возраста мира; при $10 > z > 2$ она может быть близка к массе богатых скоплений, которые, как мы знаем, действительно содержат горячий газ, излучающий в рентгеновской области спектра.

Длина свободного пробега частиц горячей плазмы за фронтом метагалактической ударной волны определяется сечением кулоновских столкновений, которое, как мы видели в § 2.2, падает с ростом температуры: $\sigma_0 = 2\pi(e^2/kT) \Lambda$, где $\Lambda = 30$ — кулоновский логарифм. По этой причине кулоновская длина пробега может оказаться весьма большой для ударных волн масштаба скоплений:

$$l_p = (4\sigma_0\rho/m)^{-1} \approx 2 \cdot 10^{21} \Omega^{-1/3} (1+z)^{-1} \left(\frac{M}{10^{13} M_\odot} \right)^{4/3} \text{ см.} \quad (5.3.30)$$

Она меньше пространственного масштаба движения $\lambda \approx \left(M / \frac{4\pi}{3} \rho \right)^{1/3}$ при *)

$$M < M_p \approx m\Lambda \left(\frac{e^2}{Gm^2} \right)^2 \approx 3 \cdot 10^{16} M_\odot. \quad (5.3.31)$$

По порядку величины масса M_p близка к массе сверхскопления или ячейки однородности во Вселенной; она не зависит ни от космологического параметра Ω , ни от момента формирования ударной волны, — еще один пример связи микроскопических констант с космологическими величинами (ср. § 2.1).

Рассмотрим теперь ударные волны масштаба отдельных галактик при $z < z_{er}$. Массы большинства галактик не превосходят величину M_{ff} ; температура газа на фронтах соответствующих им ударных волн ниже, чем $5 \cdot 10^5$ °К, и главную роль в охлаждении играют для них свободно-связанные переходы и высвечивание в атомных линиях. Характерное время охлаждения можно в этом случае представить в виде

$$\tau_c = \alpha(T)\tau_{ff}, \quad (5.3.32)$$

*) В больших масштабах гидродинамические процессы, и в частности, ударные волны, возможны при наличии коллективных возбуждений в плазме, взаимодействие с которыми и определяет в этом случае длину пробега частиц.

где $\alpha(T)$ — безразмерный множитель, меньший единицы при $T < T_{ff}$, причем в интервале между $2 \cdot 10^4$ и 10^5 °К он достигает минимума $\sim 10^{-3}$ (см. § 2.2). Благодаря сильной зависимости от температуры время (5.3.32) в указанном интервале температур меньше текущего космологического возраста. Ударные волны сопровождаются в этом случае сильным охлаждением практически во всех масштабах, меньших M_{ff} , вплоть до масштаба M_a , ниже которого ионизация газа на фронтах уже невозможна и охлаждение вовсе отсутствует. Это «окно» сильного охлаждения включает в себя, согласно сделанным выше оценкам, массы типичных галактик — от малых галактик до гигантов. Плотность газа, сжимаемого ударными волнами (при $z < z_{er}$) до значений (5.3.24), не превосходит при этом типичные плотности галактик:

$$\rho_a |_{M \ll M_{ff}} \lesssim 3 \cdot 10^{-28} \Omega (1+z)^3 \text{ г/см}^3. \quad (5.3.33)$$

Мы видим, что физические условия в метагалактической среде на активной фазе космогонического процесса определяют некоторые выделенные значения масштабов гидродинамического движения и соответствующих им масс, которые, как оказывается, близки к наиболее существенным характеристикам крупномасштабной структуры Вселенной (Чернин (1977)). Совпадение такого рода вряд ли является случайным; скорее оно указывает на реальную роль гидродинамических процессов в формировании галактик и их скоплений. Выделенные значения масштабов и масс никак не зависят от предположений о догалактической структуре в более раннюю эпоху. Единственное условие состоит в том, чтобы первоначальные возмущения оказались при $z_{er} > z > 2$ сильными в смысле, определенном в §§ 5.1 и 5.2.

Стационарное течение газа за фронтом. Процессы охлаждения в сжатом, разогретом и ионизованном газе за фронтами метагалактических ударных волн определяют не только их общие средние характеристики, которые мы оценили в предыдущем параграфе, но и всю картину распределения и движения вещества в соответствующих масштабах. Эта картина включает в себя, во-первых, ход скорости, плотности, температуры и давления газа в зависимости от расстояния до фронта ударной волны, а, во-вторых, — развивающуюся на этом фоне тепловую

неустойчивость, способную усиливать малые неоднородности в потоке газа, проходящего через фронт. Обсудим сначала первый из этих вопросов.

Если характерное время охлаждения значительно меньше характерного гидродинамического времени t_h и рассматриваются при этом промежутки времени, меньшие, чем t_h , но большие, чем время охлаждения, то можно считать, что гидродинамические характеристики движения в целом меняются мало. Пренебрегая такими малыми изменениями, можно считать гидродинамическое движение стационарным в этих промежутках времени и описывать его как некоторую распространяющуюся по газу волну, все величины в которой не зависят от времени в системе отсчета, движущейся вместе с волной.

Будем, как и выше, интересоваться прежде всего одномерными движениями и, кроме того, не будем учитывать гравитацию среды. Последнее возможно, когда характерное гравитационное время, $(G\rho)^{-1/2}$, больше времени охлаждения, что заведомо выполнено — во всяком случае, при не слишком большом сжатии газа, так как, согласно (5.1.1), время $(G\rho)^{-1/2} \approx t$ близко к гидродинамическому времени t_h . Роль гравитации мы примем во внимание, когда это условие перестанет выполняться и гравитационное время окажется сравнимым с тепловым (см. ниже).

В системе отсчета, движущейся вместе с волной, частные производные по времени от скорости, плотности и давления в уравнениях гидродинамики равны нулю, так что уравнения движения, неразрывности и теплового баланса принимают следующий вид:

$$\left. \begin{aligned} v \frac{dv}{dx} &= - \frac{dp}{dx} \frac{1}{\rho}, \\ \frac{d}{dx} (\rho v) &= 0, \\ nT v \frac{ds}{dt} &= \frac{1}{\gamma-1} v \frac{dp}{dx} - \frac{\gamma}{\gamma-1} v \frac{p}{\rho} \frac{d\rho}{dx} = -q. \end{aligned} \right\} (5.3.34)$$

Здесь γ , как всегда, показатель адиабаты, q — скорость тепловыделения из единицы объема, т. е. тепловая энергия единицы объема, деленная на характерное время охлаждения. При $q = 0$ энтропия s не меняется, и из

(5.3.34) следует адиабатическая связь давления с плотностью.

Для охлаждения на реликтовых фотонах, согласно (5.3.1):

$$q_{er} = \frac{3(\rho/m)kT}{\tau_{er}} \approx 3 \cdot 10^{-20} (1+z)^4 \rho \text{ эрг/см}^3 \cdot \text{сек.} \quad (5.3.35)$$

Для охлаждения посредством излучения при свободно-свободных переходах, в соответствии с (5.3.5),

$$q_{ff} = \frac{3(\rho/m)kT}{\tau_{ff}} \approx 10^{17} \rho^{1/2} \rho^{3/2} \text{ эрг/см}^3 \cdot \text{сек.} \quad (5.3.36)$$

(В числителях последних формул стоит коэффициент 3, а не 3/2, так как в ионизованном водороде концентрация частиц — протонов и электронов — равна удвоенному отношению ρ/m .)

Первые два уравнения в гидродинамической системе (5.3.34) совпадают с уравнениями (5.2.7) и (5.2.8), так что можно воспользоваться найденным в § 5.2 решением этих уравнений:

$$\left. \begin{aligned} v &= \frac{A}{\rho}, & A &= \text{const}, \\ p &= B - \frac{A^2}{\rho}, & B &= \text{const} > 0. \end{aligned} \right\} \quad (5.3.37)$$

При наличии охлаждения мы не должны ожидать адиабатической связи между давлением и плотностью, и потому, в отличие от изоэнтропийных волн (§ 5.2), в волне с охлаждением соотношения (5.3.37) допустимы.

Будем пользоваться решением (5.3.37) для описания движения за фронтом ударной волны и найдем константы A и B из условия, что на фронте должны выполняться условия непрерывности потоков массы и импульса (5.2.12), (5.2.16). Естественно считать, что в очень тонком слое, где происходит скачок гидродинамических величин, охлаждение не успевает сколько-нибудь заметно изменить тепловую энергию газа, так что эти соотношения, полученные для адиабатических (т. е. без охлаждения) ударных волн, здесь применимы. При сильных разрывах, которые и должны нас интересовать, скорость и плотность сразу за фронтом (отмечаем их, как и выше, значком «2») связаны со значениями этих величин перед фронтом (отмечаемыми значком «1») уравнениями (5.2.26), (5.2.27).

Из них следует, что

$$v_2 = \frac{A}{\rho_2} = \frac{\gamma - 1}{\gamma + 1} v_1,$$

$$\rho_2 = \frac{\gamma + 1}{\gamma - 1} \rho_1.$$

Отсюда

$$A = \rho_1 v_1. \quad (5.3.38)$$

Из сохранения потока импульса (5.2.16),

$$p_2 + \rho_2 v_2^2 = B - \frac{\rho_1^2 v_1^2}{\rho_2} + \rho_2 \left(\frac{\rho_1 v_1}{\rho_2} \right)^2 = p_1 + \rho_1 v_1^2 \approx \rho_1 v_1^2, *$$

находим (в пренебрежении давлением p_1):

$$B = \rho_1 v_1^2. \quad (5.3.39)$$

С учетом (5.3.38) и (5.3.39) решение (5.3.37) для зоны охлаждения газа за фронтом ударной волны принимает вид

$$p = \rho_1 \frac{1}{y}, \quad (5.3.40)$$

$$p = \rho_1 v_1^2 (1 - y), \quad (5.3.41)$$

где переменная y есть безразмерная скорость:

$$v = v_1 y. \quad (5.3.42)$$

Запишем еще следующее из (5.3.40) и (5.3.41) выражение для адиабатической скорости звука за фронтом:

$$u^2 = \gamma \frac{p}{\rho} = \gamma v_1^2 y (1 - y). \quad (5.3.43)$$

Величина y изменяется, уменьшаясь по мере удаления от фронта, от значения $y_2 = 1/4$ до значения y_a , которое определяет заднюю границу зоны охлаждения (рис. 12). Охлаждение прекращается, когда температура газа упадет настолько, что он снова окажется нейтральным; как мы уже говорили, рекомбинация в разогретом газе происходит при температуре $T_a \approx 10^4$ °К. Этой температуре отвечает скорость звука $u_a \approx 10^6$ см/сек. Поэтому согласно (5.3.43)

$$y_a = \frac{1}{\gamma} \frac{u_a^2}{v_1^2}. \quad (5.3.44)$$

Это соотношение определяет итоговое сжатие, достигаемое за фронтом ударной волны (в согласии с приближенной оценкой (5.3.13)), а также и характеристики

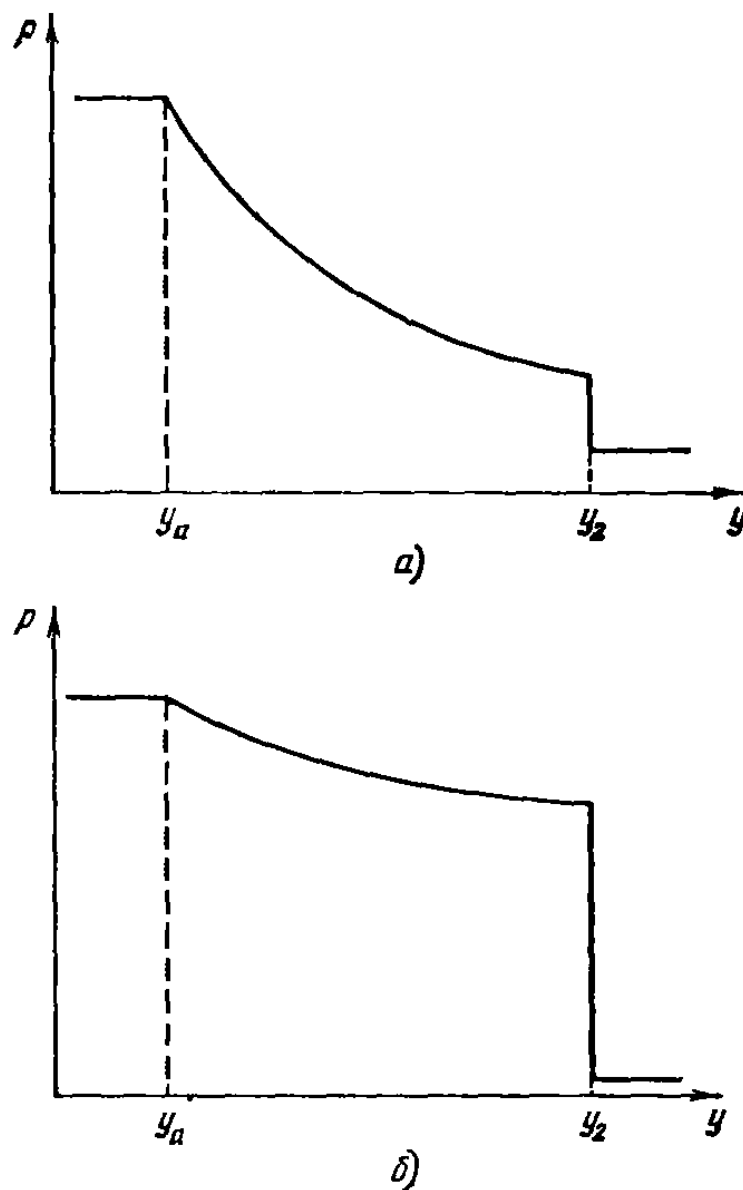


Рис. 12. Зависимость плотности (а) и давления (б) от скорости газа за фронтом ударной волны с быстрым охлаждением.

среды вне зоны охлаждения, где величины плотности и давления уже не меняются:

$$\left. \begin{aligned} v_a &= v_1 y_a, \\ \rho_a &= \rho_1 \frac{1}{y_a}, \\ p_a &= \rho_1 v_1^2 (1 - y_a). \end{aligned} \right\} \quad (5.3.45)$$

Решение (5.3.40) — (5.3.43) подтверждает и уточняет, в частности, сделанное выше заключение о приблизительно изобарическом характере движения за фронтом (см.

соотношение (5.3.12)). Действительно, отношение давлений на границе зоны охлаждения есть

$$\frac{p_2}{p_a} = \frac{1 - y_2}{1 - y_a} \approx \frac{3}{4}. \quad (5.3.46)$$

Градиент давления, обусловленный повышением давления по мере удаления от фронта, и служит причиной, по которой движение газа за фронтом тормозится, а его плотность возрастает. При этом возрастание плотности происходит быстрее, чем возрастание давления (формально из (5.3.37) следует, что при плотности, стремящейся к бесконечности, давление остается конечным).

Для полного решения задачи остается найти зависимость переменной y от координаты x , которую мы отсчитываем от фронта ударной волны. Воспользуемся третьим уравнением гидродинамической системы (5.3.34); при скорости, плотности и давлении, даваемыми решением (5.3.40) — (5.3.42), оно принимает вид

$$\rho_1 v_1^3 [\gamma - (\gamma + 1) y] \frac{dy}{dx} = -(1 - \gamma) q(y). \quad (5.3.47)$$

Для комptonовского охлаждения, согласно (5.3.35)

$$q_{er}(y) = \frac{\beta}{1 - \gamma} (1 - y) \rho_1 v_1^3 \text{ эрг/см}^3 \cdot \text{сек}, \quad (5.3.48)$$

где

$$\beta = 3 \cdot 10^{-20} (1 + z)^4 \frac{1 + \gamma}{v_1} \text{ см}^{-1}.$$

Тогда, интегрируя (5.3.47), находим при $\gamma = 5/3$:

$$\frac{8}{3} y - \ln(1 - y) = -\beta x + C. \quad (5.3.49)$$

Так как $x = 0$ при $y = y_2 = 1/4$, то

$$C = 11/12 \approx 1. \quad (5.3.50)$$

Учитывая, что $y \ll 1/4$, можно в первом приближении по этой малой величине написать так:

$$y \approx \frac{1}{4} (1 - \beta x). \quad (5.3.51)$$

Для области вблизи фронта, где $x \rightarrow 0$, можно, кроме того, разложить решение по степеням малого произведе-

ния βx ; ограничиваясь первыми степенями, находим:

$$\rho \propto (1 + \beta x), \quad p \propto \left(1 + \frac{3}{8} \beta x\right), \quad v \propto (1 - \beta x), \quad (\beta x \ll 1), \quad (5.3.52)$$

Вблизи задней границы зоны охлаждения, где $y \rightarrow y_a$,

$$\rho \propto y^{-1}, \quad p = \text{const}, \quad v \propto y, \quad (y \ll 1). \quad (5.3.53)$$

Следующая отсюда оценка размера зоны охлаждения ($x(y_a) = l_c$)

$$\frac{1}{4} (1 - \beta l_c) = y_a \ll 1, \quad l_c \approx \frac{1}{\beta} \approx v_1 \tau_{er}, \quad (5.3.54)$$

согласуется с приближенной оценкой (5.3.19), (5.3.20).

Если охлаждение происходит благодаря излучению при свободно-свободных переходах, то в соответствии с (5.3.36), (5.3.40) и (5.3.41)

$$q_{ff}(y) = \frac{b}{1 - \gamma} (1 - y)^{1/2} y^{-3/2} \rho_1 v_1^3 \text{ эрг/см}^3 \cdot \text{сек}, \quad (5.3.55)$$

где

$$b = 10^{17} \frac{1 + \gamma}{v_1^2} \rho_1 \text{ см}^{-1}.$$

Интегрируя снова уравнение (5.3.47), находим (при $\gamma = 5/3$) в первом приближении по малой величине y :

$$y = \left(\frac{3}{2}\right)^{2/5} (C - bx)^{2/5}. \quad (5.3.56)$$

Так как $x = 0$ при $y = y_2 = 1/4$, то

$$C = \frac{1}{48} \approx 0,02, \quad y = \frac{1}{4} (1 - 48bx)^{2/5}. \quad (5.3.57)$$

Для области вблизи фронта, где $x \rightarrow 0$, $48bx \ll 1$, решение принимает вид

$$\left. \begin{aligned} y &\propto (1 - 20bx), & \rho &\propto (1 + 20bx), \\ p &\propto (1 + 7bx), & v &\propto (1 - 20bx). \end{aligned} \right\} \quad (5.3.58)$$

Вблизи задней границы зоны охлаждения сохраняется асимптотика (5.3.53). Размер этой зоны можно оценить, положив в (5.3.56) $x = l_c$ при $y = y_a \ll 1$:

$$C - bl_c \ll 1, \quad l_c \approx \frac{0,02}{b}. \quad (5.3.59)$$

Учитывая в выражении для коэффициента b связь между скоростью v_1 и температурой (5.3.21), находим:

$$l_c \approx 0,01 v_1 \tau_{ff} (y_2). \quad (5.3.60)$$

В этом случае уточнение по сравнению с оценкой (5.3.19) оказывается более существенным, чем при комптоновском охлаждении. Нужно, правда, иметь в виду, что при снижении температуры газа до значений $T < 5 \cdot 10^5$ °К главный вклад в охлаждение дают, как говорилось, свободно-связанные переходы и излучение в линиях водорода и гелия. При таких температурах характерное время этих процессов, согласно сказанному в § 2.2, значительно меньше времени τ_{ff} . Но тогда размер зоны, где охлаждение происходит благодаря излучению при свободно-связанных переходах и высвечиванию в линиях, должен быть мал по сравнению с размером зоны, где преобладают свободно-свободные переходы; поэтому можно полагать, что соотношение (5.3.60) дает приближенную оценку для всей зоны охлаждения за фронтом ударной волны.

Тепловая неустойчивость. В реальных условиях метagalacticкой среды газ, натекающий на фронт ударной волны, никогда не может быть идеально однородным. В нем имеются сгущения и разрежения, представляющие собой малые возмущения той или иной природы. Оказываясь за фронтом, возмущения могут испытать значительное усиление, если охлаждение газа происходит достаточно быстро и обязано при этом тормозным процессам. Природу этого весьма общего явления, называемого тепловой неустойчивостью (Филд (1965)), легко понять качественно.

Если какой-либо элемент среды, охлаждающейся благодаря тормозным процессам, имеет плотность, несколько превышающую невозмущенную плотность газа, то частота столкновений частиц в нем больше, а потому больше и интенсивность тормозного излучения. Вследствие этого элемент охладится сильнее, чем окружающая его среда, и давление в нем упадет. Но тогда внешнее давление будет сжимать его, стремясь восстановить в нем прежнее давление. Плотность, а с ней интенсивность излучения, возрастут еще больше, и потому, раз начавшись, сжатие рассматриваемого элемента будет нарастать, пока тормозные процессы смогут осуществлять его охлаждение. Таким образом, явление действительно имеет характер неустой-

чивости. Она приводит к тому, что в горячем газе возникают более холодные и более плотные области. Их равновесие по давлению с окружающей средой означает, что в этих областях произведение плотности $\tilde{\rho}$ на температуру T имеет то же значение, что и для окружающей среды:

$$\tilde{\rho}T = \rho T. \quad (5.3.61)$$

Рассмотрим динамику тепловой неустойчивости за фронтом ударной волны в метagalacticкой среде (Гуревич и Чернин (1975), Зенцова и Чернин (1978)). Если охлаждение за фронтом существенно, но производится лишь комptonовским механизмом, неоднородности натекающего на фронт газа сохраняются, причем контраст плотности $\delta\rho/\rho$ остается за фронтом «замороженным», так как характерное время охлаждения и сжатия слоев газа $\tau_c = \tau_{er}$ не зависит в этом случае от их плотности. Если же преобладают тормозные процессы охлаждения, то время охлаждения и сжатия уже не будет универсальным, а станет функцией плотности. В тех местах, где плотность выше, время охлаждения меньше и сжатие происходит быстрее; в местах разрежения время охлаждения больше и сжатие идет медленнее, так что контраст плотности усиливается. При свободно-свободных переходах, например,

$$\tau_{ff} \propto \rho^{-3/2}, \quad p = \text{const}. \quad (5.3.62)$$

Чтобы получить представление о кинетике развития тепловой неустойчивости, воспользуемся снова уравнением теплового баланса, но будем применять его теперь к некоторому слою газа, с которым свяжем систему отсчета:

$$\frac{1}{\gamma-1} \frac{dp}{d\tau} - \frac{\gamma}{\gamma-1} \frac{p}{\rho} \frac{d\rho}{d\tau} = -q. \quad (5.3.63)$$

Время τ будем отсчитывать от момента пересечения фронта рассматриваемым элементом среды. Кроме того, будем полагать, что в пределах рассматриваемого слоя все величины слабо зависят от координаты x (для этого слой нужно считать достаточно тонким — см. ниже); в пренебрежении этой зависимостью полные производные по времени в (5.3.63) превращаются в частные. Учитывая сказанное выше, пренебрежем также изменением давления за фронтом ударной волны.

Для комptonовского охлаждения уравнение (5.3.63) после интегрирования дает невозмущенную плотность за фронтом как функцию времени τ :

$$\rho(\tau) = 4\rho(z) e^{\frac{6}{5} \frac{\tau}{\tau_{er}}}. \quad (5.3.64)$$

Здесь $\rho(z)$ — космологическая плотность, которая считается постоянной в течение времени τ .

Иначе происходит сжатие газа за фронтом, когда его охлаждение обусловлено излучением при свободно-свободных переходах. В этом случае, интегрируя (5.3.63), получим

$$\rho(\tau) = 4\rho(z) \left[1 - \frac{9}{5} \frac{\tau}{\tau_{ff}^0} \right], \quad (5.3.65)$$

где

$$\tau_{ff}^0 = \tau_{ff} \quad (\tau = 0).$$

Зависимость характерного времени τ_{ff} от плотности ρ приводит, во-первых, к степенному, а не экспоненциальному, как при комptonовском процессе, закону плотности; во-вторых, — к тепловой неустойчивости при наличии возмущений плотности в газе перед фронтом. При этом в формуле (5.3.65) под плотностью (и перед скобкой, и в выражении для τ_{ff}^0 — см. (5.3.5)) нужно понимать космологическую плотность $\rho(z)$ плюс возмущение $\delta\rho(z, r)$. Контраст плотности, как можно видеть из (5.3.65), нарастает по степенному закону. Это связано с тем, что неустойчивость развивается в нестационарной среде, которая испытывает общее сжатие.

Тепловая неустойчивость за фронтом ударной волны превращает возмущения плотности набегающего потока газа в энтропийные возмущения; для них давление считается невозмущенным при наличии неоднородностей плотности и температуры *). Выравнивание давления производится, как всегда в таких случаях, посредством звуковых волн; поэтому в процессе тепловой неустойчивости могут усиливаться сгущения, размер которых не пре-

*) Отличие таких возмущений от первичных энтропийных космологических возмущений на стадии преобладания излучения очевидно.

восходит расстояния, проходимого звуком за характерное время охлаждения, т. е.

$$\lambda < l_T \quad \text{и} \quad \tau_c.$$

Как скорость звука, так и время тормозных процессов уменьшаются со временем при падении температуры по мере удаления данного элемента газа от фронта; убывает и масса, содержащаяся в объеме с размером $\sim \tau_c$. Если, например, охлаждение осуществляется свободно-свободными переходами, то с учетом (5.3.40) — (5.3.42) найдем

$$M_T = \frac{4\pi}{3} \rho l_T^3 \propto y^5 (1 - y)^3. \quad (5.3.66)$$

По этой причине в масштабах, соответствующих меньшим массам, тепловая неустойчивость развивается дольше и в них достигается большее усиление контраста плотности. Если в пределах некоторого сгущения большой массы имеются какие-то сгущения меньших масс, то такое сгущение будет сначала выделяться из окружающей среды как целое с одновременным усилением контраста во всех внутренних масштабах, а затем тепловая неустойчивость «отключится» в масштабе всего сгущения и усиление будет продолжаться во все меньших внутренних его масштабах. В результате можно ожидать фрагментации сгущения, что, очевидно, должно быть существенно для всей картины космогонического процесса.

Развитию тепловой неустойчивости может препятствовать теплопроводность среды, стремящаяся сгладить температурные неоднородности. Характерное время выравнивания температуры определяется формулой

$$\tau_k \approx \frac{\lambda^2}{\kappa},$$

с которой мы уже встречались при рассмотрении диссипативных процессов в § 3.3. Здесь $\kappa \approx u_e l_e$, — температуропроводность, т. е. отношение теплопроводности к концентрации электронов, u_e — тепловая скорость и l_e — длина свободного пробега электронов. Перенос тепла в плазме осуществляется в этом случае именно электронами, так как для фотонов газ прозрачен и они только выносят энергию из системы. Что касается ионов, то их вклад в теплопроводность существенно меньше, так как скорость их

теплового движения меньше тепловой скорости электронов, а длина свободного пробега по отношению к столкновению друг с другом близка к l_e . Неравенство $\tau_x > \tau_c$, обеспечивающее сохранение и нарастание температурных неоднородностей, приводит к условию

$$\lambda > \lambda_x \approx (\kappa \tau_c)^{1/2}.$$

Отношение длины λ_x к характерной длине тепловой неустойчивости,

$$\frac{\lambda_x}{l_T} \approx \left(\frac{m_e}{m}\right)^{1/2} \left(\frac{l_e}{l_T}\right)^{1/2},$$

много меньше единицы, когда тепловая неустойчивость развивается в сгущении с массой галактики, а длина l_T соответствует по массе, например, шаровому скоплению.

Соотношения (5.3.61) и (5.3.65) не предполагают, что отклонение плотности в рассматриваемом элементе газа от невозмущенного значения мало; поэтому они относятся и к начальной линейной фазе развития тепловой неустойчивости, когда контраст плотности мал, и к более поздней, когда он велик. Формула (5.3.65) перестает быть справедливой, только когда температура в данном слое становится ниже, чем $5 \cdot 10^5$ °К. При меньших температурах охлаждение осуществляется излучением при свободно-связанных переходах и высвечиванием в линиях, и качественная картина остается прежней.

Развитие неустойчивости прекращается, когда температура в сжимающемся элементе среды приближается к 10^4 °К; скорость его охлаждения при этом снижается и сравнивается со скоростью охлаждения окружающей среды. Вскоре после этого охлаждение элемента становится несущественным, и его температура остается постоянной. Равновесие такого участка среды по давлению с окружающим его невозмущенным более горячим и разреженным газом требует, согласно (5.3.61), чтобы и плотность слоя оставалась неизменной. Но плотность невозмущенного газа при этом возрастает, так как его температура продолжает падать. В результате контраст плотности, достигнув некоторого максимального значения, станет после этого убывать. Неустойчивость сменяется устойчивостью,

и любое возмущение практически полностью рассосется, когда весь газ охладится до температуры порядка 10^4 °К.

Причиной, способной предотвратить сглаживание плотности в охлаждающемся газе, является собственная гравитация среды, благодаря которой тепловая неустойчивость за фронтом может смениться гравитационной неустойчивостью. Характерное время гравитационной неустойчивости $(G\rho)^{-1/2}$ для этого должно быть сравнимым со временем охлаждения или меньше его. В той зоне за фронтом, где преобладают свободно-свободные переходы, отношение времен,

$$(G\rho)^{1/2}\tau_{ff} \propto \rho^{-1} \propto y,$$

убывает, так что если гравитация среды не существенна вблизи фронта, то она тем более не играет роли во всей той зоне, где сгущение еще продолжает охлаждаться. Но прекращение охлаждения в сгущении означает неограниченное (формально) возрастание времени охлаждения, так что гравитация обязательно становится существенной, когда температура сгущения снижается до значения $\sim T_a$.

Кроме того, требуется, как всегда, чтобы размер сгущения превышал джинсову длину. Для $T = T_a$ и плотности $\rho \approx \rho_a$, которая является максимальной при заданном кинетическом давлении натекающего газа $\rho(z) v^2$, джинсова длина

$$R_J \approx \frac{u_a}{(G\rho_a)^{1/2}} \approx 2 \cdot 10^{23} \Omega^{-1/2} (1+z)^{-2} \left(\frac{M}{10^{13} M_\odot} \right)^{-1/2} \text{ см.} \quad (5.3.67)$$

В области с размером $\sim R_J$ содержится при этих условиях масса

$$M_J \approx 10^8 \Omega^{-1/2} (1+z)^{-3/2} M_\odot. \quad (5.3.68)$$

Ударной волне с сильным охлаждением соответствуют массы, сравнимые, как мы видели, с массами галактик. При $M \sim 10^{11} M_\odot$ и для интервала красных смещений $10 > z > 2$ величина (5.3.68) близка к массам звездных скоплений. Если галактики возникали благодаря сверхзвуковым движениям и ударным волнам в метagalacticкой среде или в газе протоскоплений*), то джинсову

*) В последнем случае их количественные характеристики несколько отличаются от указанных выше (см. §§ 5.6 и 5.7)

массу (5.3.68) нужно, вероятно, рассматривать как характерную массу тех газовых комплексов, в которых происходило образование звезд (см. §§ 6.1 и 7.1).

§ 5.4. Генерация вихрей в ударных волнах

Одной из отличительных особенностей ударных волн и других разрывных гидродинамических потоков является возможность генерации вихря изначально потенциальными движениями. Наличие разрывов делает неприменимой теорему Кельвина — Гельмгольца о сохранении циркуляции скорости (§ 3.2), запрещающую возникновение

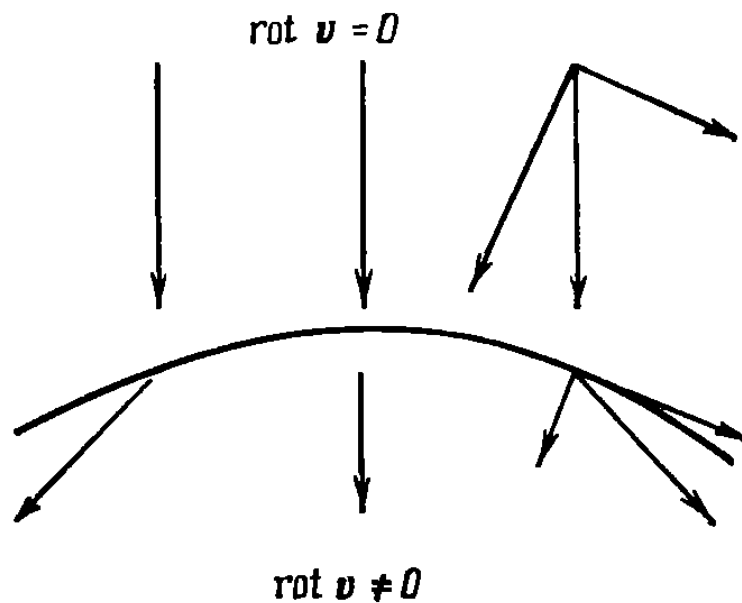


Рис. 13. Генерация вихря на искривленном фронте ударной волны.

(или уничтожение) вихрей в невязкой баротропной *) среде. Вот простой и наглядный пример появления вихря за фронтом ударной волны.

Пусть параллельный поток газа, в котором имеется только одна компонента скорости, натекает на искривленный фронт ударной волны (рис. 13). Из условий непрерывности трех потоков — массы, импульса и энергии — следует (§ 5.2), что нормальная к фронту компонента скорости потока претерпевает разрыв и уменьшается на фронте, а тангенциальная компонента остается неизменной. В результате за фронтом ударной волны поток уже не будет параллельным, и одномерность движения нарушится.

*) Напомним, что баротропия означает зависимость давления среды только от ее плотности.

При этом появится поперечная компонента скорости, а продольная ее компонента станет зависящей от поперечной координаты. Но это означает, что ротор скорости

$$\operatorname{rot} \mathbf{v} = \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z} \right) \mathbf{i} + \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x} \right) \mathbf{j} + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \right) \mathbf{k}$$

окажется, вообще говоря, отличным от нуля.

На активной фазе космогонического процесса в метagalактической среде, как мы говорили, должны существовать сверхзвуковые гидродинамические движения с разрывами, причем, вероятнее всего, они развиваются из слабых потенциальных движений, усиливаемых гравитационной неустойчивостью. Сильные сверхзвуковые движения создают значительные сгущения вещества, способные, по-видимому, дать начало протогалактикам и протоскоплениям галактик (см. § 5.5 и 5.6). В том же процессе они порождают и новые вихревые движения, так что в сильно возбужденном состоянии метagalактической среды бурные гидродинамические движения должны обладать существенной завихренностью.

Можно предполагать, что вращение галактик — это важное свойство крупномасштабной структуры Вселенной — обязано своим происхождением вторичным вихрям, генерируемым сверхзвуковыми движениями метagalактической среды (Чернин (1970а, б)).

Указанный в начале параграфа механизм генерации вихрей, очевидно, не является единственным. Мы рассмотрим сейчас и иные механизмы такого рода, а в § 5.6 обсудим некоторую общую космогоническую картину, в которой идея вторичных вихрей играет ключевую роль.

Рассеяние малых возмущений на фронте ударной волны. В среде, охваченной сверхзвуковыми движениями, неизбежно имеются как сильные, так и слабые возмущения различных масштабов и амплитуд. Нелинейные гидродинамические процессы взаимодействия таких возмущений и есть причина возникновения завихренности в среде. Взаимодействия слабых потенциальных движений друг с другом не могут породить вихри — для них выполняется теорема сохранения циркуляции; вихри рождаются при взаимодействиях слабых потенциальных возмущений с сильными и сильных друг с другом. Рассмотрим в ка-

честве примера взаимодействий первого рода рассеяние звуковой волны на плоском фронте ударной волны.

Звук может рассеиваться на фронте ударной волны, если он распространяется навстречу ей в набегающем на фронт газе или если он догоняет фронт, распространяясь вслед за ним в газе за фронтом. В первом случае (рис. 14, а) звуковая волна испытывает преломление, но не может отразиться от фронта, так как фронт движется по среде со

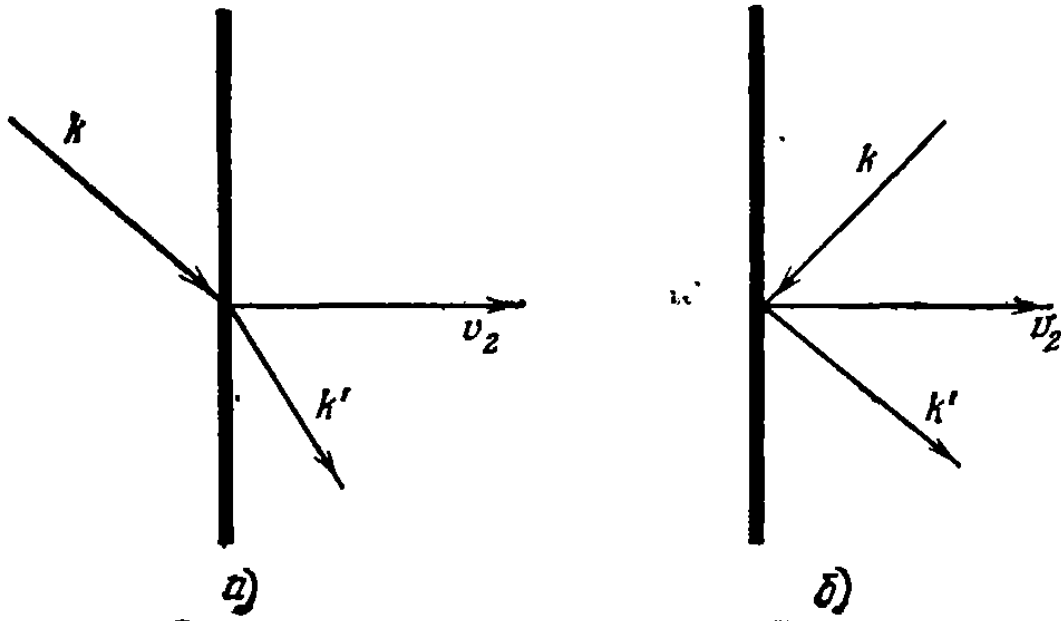


Рис. 14. Рассеяние звука на фронте ударной волны; а) — встречное падение; б) — падение вдогонку.

сверхзвуковой скоростью (или, что то же, газ, по которому распространяется звук, натекает на фронт со сверхзвуковой скоростью). Во втором случае (рис. 14, б) звук догоняет фронт, так как скорость оттока газа от фронта меньше скорости распространения звука. При этом он испытывает только отражение, а преломленная волна отсутствует, ибо в газе перед фронтом последний распространяется быстрее звука.

Но этими процессами не ограничивается взаимодействие звука с ударной волной. При преломлении (в случае падения навстречу фронту) или отражении (в случае падения вдогонку) звук возбуждает в среде малые возмущения двух других типов — вихревые и энтропийные. Только в одном случае — при прямом падении, когда плоскости фронтов звуковой и ударной волн параллельны, а волновой вектор звука нормален поверхности фронта, — генерация вторичного вихря невозможна. Действительно, в этом случае в задаче нет векторов, из которых можно

было бы образовать отличное от нуля векторное произведение, т. е. псевдовектор, необходимый для получения псевдовекторной величины — ротора скорости.

При любом косом падении звука отлично от нуля векторное произведение волнового вектора на нормаль к фронту. Это векторное произведение и определяет ориентацию ротора скорости. Принимая, что фронт лежит в плоскости (yz) прямоугольной системы координат так, что нормаль к нему имеет лишь x -проекцию, а волновой вектор звука — в плоскости (xy) , так что у него есть лишь x - и y -проекции, видим, что ротор скорости вторичного вихревого движения должен содержать только z -проекцию, а вектор вихревой скорости должен быть параллелен плоскости (xz) .

Необходимость возникновения вихревых движений (и энтропийных возмущений) при рассеянии звука на фронте ударной волны следует из сохранения трех потоков — массы, импульса и энергии. Мы вывели соответствующие условия в § 5.2, где движение предполагалось одномерным. Косое падение звука требует обобщить эти условия на двумерный случай, что достигается заменой скорости v в соотношениях (5.2.12), (5.2.16) и (5.2.18) ее нормальной к фронту компонентой; тангенциальная же компонента остается, как говорилось, без изменений. Эти условия нужно записать для скорости, давления, плотности и т. д., представляющих собой суммы соответствующих возмущенных и невозмущенных звуком величин. Нужно учитывать также, что под влиянием звукового возмущения сам фронт ударной волны начинает колебаться.

Рассмотрим сначала рассеяние звука вдогонку. В этом случае уравнения сохранения трех потоков сводятся (в первом порядке по амплитудам возмущений) к условию равенства нулю вариаций этих потоков в среде за фронтом; потоки перед фронтом не изменяются, так как туда звук не проникает.

Не будем выписывать, ввиду громоздкости, ни сами эти уравнения, ни их точные решения *), а приведем ре-

*) Процессы трансформации возмущений на фронтах ударных волн рассматривали Бюргерс, Бохинцев, Бриллюэн; наиболее полный анализ дан Конторовичем (1957) и Маккензи и Вестфалом (1970); роль этих процессов в космологии и космогонии анализировалась в работе Чернина, Зильберглейта и Троппа (1976).

зультат, справедливый для сильного разрыва при малых углах падения. Амплитуды ротора скорости \mathcal{R} и относительного возмущения энтропии \mathcal{S} , генерируемых за фронтом ударной волны, пропорциональны амплитуде D дивергенции скорости звука:

$$\mathcal{R} \sim D\vartheta, \quad \mathcal{S} \sim \frac{D}{uk}, \quad (5.4.1)$$

где u — скорость звука за фронтом, k — модуль волнового вектора, ϑ — угол между последним и нормалью к фронту. Амплитуда вторичного вихря обращается в нуль, как мы и ожидали, лишь при нулевом угле падения ϑ .

Вихрь рождается и при встречном косом падении звука на фронт ударной волны. При этом возможно явление, подобное полному внутреннему отражению: при углах, превосходящих некоторый критический, преломленный звук отсутствует (но отражение с самого начала запрещено!). Для сильных разрывов критический угол

$$\vartheta_c \approx \arctg \frac{v_1}{u} \approx 1, \quad (5.4.2)$$

где v_1 — скорость распространения ударной волны.

Если угол падения равен критическому, то направление скорости распространения звука за фронтом становится параллельным фронту (рис. 15). При больших углах падения звук экспоненциально затухает за фронтом, и возмущение складывается из вихревых и энтропийных волн.

Завихренность возникает за фронтом и в том случае, когда в натекающем потоке газа имеются замороженные энтропийные возмущения:

$$\mathcal{R} \sim uq\mathcal{S}\beta. \quad (5.4.3)$$

(Здесь q, β — волновое число и соответствующий угол

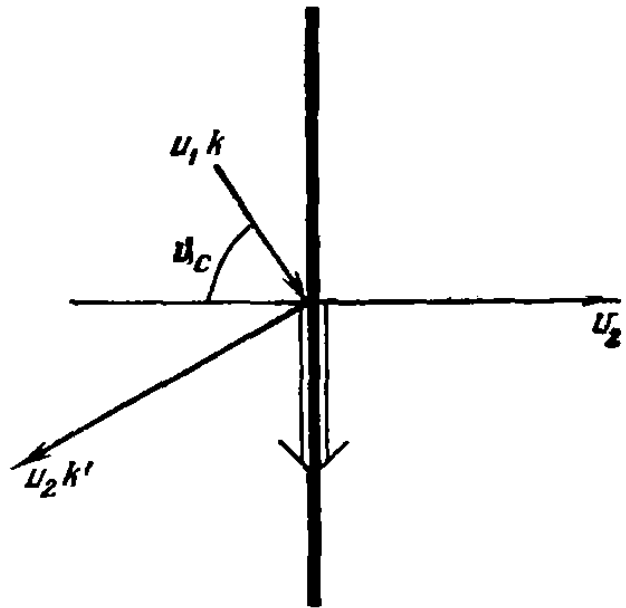


Рис. 15. Критический угол падения звука.

падения для энтропийного возмущения с относительной амплитудой ξ .) Эта приближенная формула справедлива, как и (5.4.1), для сильного разрыва при малых углах падения.

Можно сказать, что любые безвихревые возмущения в газе, по которому распространяется ударная волна, возбуждают вторичные вихри за ее фронтом, если только движение перед фронтом не является строго одномерным.

В реальных условиях метагалактической среды на активной фазе космогонического процесса существовал, по-видимому, весьма богатый набор случайных, хаотически ориентированных гидродинамических движений, содержащий как сильные, так и слабые возмущения, распространявшиеся в различных направлениях.

Хотя в теории, приводящей к соотношениям (5.4.1) — (5.4.3), не учитывается гравитация и космологическое расширение, сами эти оценочные соотношения как первое приближение остаются справедливыми для космологической ситуации, так как гидродинамические процессы в сильных возмущениях эффективнее, чем гравитация, и быстрее, чем космологическое расширение (см. § 5.1).

Если ударные волны охватывают массы газа, сравнимые с массами скоплений галактик, то взаимодействующие с ними слабые возмущения меньшего масштаба могли бы соответствовать по массе галактикам. За фронтом ударной волны эти возмущения можно описывать, пренебрегая (как и в сильных возмущениях) такими факторами, как охлаждение газа и собственная гравитация среды. В ударных волнах масштаба скоплений тепловые процессы происходят медленнее, чем изменение общих гидродинамических характеристик движения (§ 5.3). Это справедливо, очевидно, и для невозмущенного движения, и для возмущений. Пренебрежение собственной гравитацией возмущений возможно, если их пространственный масштаб меньше джинсовой длины. В ударной волне масштаба скопления это условие выполняется для масштаба, соответствующего по массе типичной галактике, ибо джинсова масса в этом случае близка, как легко видеть на основании соотношений § 5.3, к массе скопления в целом.

В газе, набегавшем на фронт, дело обстоит иначе. В эпоху $10 > z > 2$ безвихревые возмущения имеют, как

мы знаем, характер потенциальных движений, усиливаемых гравитационной неустойчивостью, их масштаб много больше джинсовой длины волны в метagalактической среде во всю эпоху после рекомбинации. Давление в этих масштабах не играет роли, а характерное гидродинамическое время, $t_h \approx \lambda/v$, больше t , так что такие потенциальные возмущения нельзя рассматривать как звуковые волны. Они развиваются со временем в том же темпе, в каком происходит космологическое расширение, т. е. изменение скорости благодаря гравитационной неустойчивости характеризуется временем t :

$$\left(\frac{\partial \ln v}{\partial t}\right)^{-1} \approx t.$$

Это время, очевидно, много больше времени прохождения области возмущения с размером λ через фронт ударной волны:

$$t \gg \lambda/v_1,$$

где v_1 — скорость натекания газа на фронт. Поэтому в первом приближении мы можем пренебречь изменением скорости малых возмущений при рассмотрении их взаимодействия с ударным фронтом. Но с тем же правом мы можем пренебречь и самой скоростью, учитывая, что в возмущениях $t_h \gg t$. В таком приближении для слабого потенциального движения существенной характеристикой является величина возмущения плотности, и в этом смысле оно формально подобно энтропийному возмущению, фигурирующему выше в стационарной (и без гравитации) задаче. Амплитуду вторичного вихря можно в этом случае оценить (для сильного разрыва и малых углов падения) по аналогии с (5.4.3):

$$\mathcal{R} \approx \frac{u}{\lambda} \frac{\delta\rho}{\rho} \vartheta, \quad (5.4.4)$$

где $\delta\rho/\rho$ — амплитуда относительного возмущения плотности в набегающем на фронт потоке газа, λ — размер возмущения.

Вихревое движение, рождающееся на фронте ударной волны, является, очевидно, дозвуковым, независимо от того, происходит ли рассеяние возмущений вдогонку или навстречу движению фронта. Скорости вихрей, существенные с космогонической точки зрения, должны

соответствовать удельному (на единицу массы) моменту вращения спиральных галактик:

$$v_G R_G \approx 10^{30} \text{ см/сек}, \quad (5.4.5)$$

где $v_G \approx 3 \cdot 10^7 \text{ см/сек}$, $R_G \approx 3 \cdot 10^{22} \text{ см}$ — характерные скорость и радиус галактики. В газе за фронтом ударной волны движение с такими скоростями действительно является дозвуковым, если масштаб ударной волны соответствует по массе скоплениям галактик. Согласно общим соотношениям для сильных разрывов (см. § 5.2) скорость звука за фронтом

$$u \approx \frac{1}{4} \sqrt{3\gamma} v_1. \quad (5.4.6)$$

Подставляя в качестве v_1 гидродинамическую скорость, отвечающую условию перехода к состоянию сильных возмущений (5.1.1), находим

$$u \approx 3 \cdot 10^7 \Omega^{1/2} (1+z)^{1/2} \left(\frac{M}{10^{13} M_\odot} \right)^{1/2} \text{ см/сек}. \quad (5.4.7)$$

В ударных волнах, для которых существенно охлаждение, скорость звука падает с удалением от фронта, и потому дозвуковые движения могут в этом случае стать со временем сверхзвуковыми. Скорость вихревого движения v_v , в данном слое газа, сжимающегося за фронтом ударной волны, возрастает с увеличением плотности:

$$v_v \propto \lambda^{-1} \propto \rho. \quad (5.4.8)$$

Мы учли здесь, что вектор ротора скорости для вихря, рожденного на фронте ударной волны, направлен параллельно оси z , а сжатие газа происходит в направлении оси x , так что сохранение момента импульса и сохранение потока вихря (см. § 3.2) означают в этом случае постоянство произведения скорости на размер возмущения λ вдоль направления сжатия. Если воспользоваться решением для ударной волны с охлаждением, найденным в предыдущем параграфе, то мы увидим, что отношение вихревой скорости и скорости звука,

$$\frac{v_v}{u} \propto y^{-3/2} (1-y)^{-1/2}, \quad (5.4.9)$$

возрастает с удалением данного элемента газа от фронта

(т. е. с уменьшением u) и может достигь единицы. Ясно, что переход к сверхзвуковому режиму вызывает существенную перестройку всей картины течения за фронтом: вихревые возмущения перестают быть малыми и в свою очередь становятся способными породить ударные волны и т. п.

Вихревая скорость за фронтом может, очевидно, остаться и дозвуковой, если при максимальном сжатии до плотности $\rho_a \approx \rho (v/u_a)^2$ она не превосходит величины $u_a = u (T_a) \approx 10^8$ см/сек. В этом случае момент вращения завихренного сгущения с массой галактики $\sim 10^{11} M_\odot$, которое могло бы возникнуть при фрагментации газа с плотностью ρ_a , ограничен величиной

$$u_a R_G \left(\frac{\rho_G}{\rho_a} \right)^{1/2} \approx 3 \cdot 10^{29} (1+z)^{-1/2} \left(\frac{M}{10^{13} M_\odot} \right)^{-1/2} \text{ см}^2/\text{сек.} \quad (5.4.10)$$

Даже при минимальном значении красного смещения $z = 2$ эта величина на порядок меньше типичного момента спиральных галактик (5.4.5).

Это означает, что вихри, интересные с космогонической точки зрения, обязательно являются сверхзвуковыми в таких слоях протоскопления, для которых существенно охлаждение. В протоскоплениях большой массы ($\sim 10^{15} M_\odot$) охлаждение происходит медленно и значительная часть газа долго остается горячей и более или менее однородной (§ 5.3); в протоскоплениях же умеренной массы ($\sim 10^{13} M_\odot$) некоторая часть газа довольно быстро успевает охладиться до температуры T_a (см. § 5.6).

Распространение малых возмущений в небаротропной среде. Если пренебречь в гидродинамическом уравнении движения (уравнении Эйлера) нелинейным по скорости членом и взять ротор от левой и правой его частей, то мы найдем:

$$\frac{d \text{rot}^2 v}{dt} \approx \text{rot} \frac{\nabla p}{\rho}. \quad (5.4.11)$$

При баротропии, когда давление есть функция только плотности, правая часть обращается в нуль, и вихрь, если он равен нулю в некоторый момент времени, остается равным нулю всегда. (Это заключение, как мы видели в § 3.2, справедливо и без пренебрежения нелинейным членом.)

Когда же условие баротропии не выполнено, вихрь может возникать и усиливаться в безвихревом потоке газа.

В интересующих нас условиях на активной фазе космогонического процесса небаротропные течения осуществляются за фронтами ударных волн, в которых существенно охлаждение, и плотность газа возрастает при удалении от фронта.

В слабом потенциальном возмущении, распространяющемся в такой неоднородной среде, давление и плотность изменяются на малые величины, которые мы обозначим через δp и $\delta \rho$, так что

$$p \rightarrow p(x) + \delta p(x, y, z, t), \quad \rho \rightarrow \rho(x) + \delta \rho(x, y, z, t).$$

Подставим эти выражения в правую часть (5.4.11):

$$\operatorname{rot} \frac{\nabla p}{\rho} = \operatorname{rot} \frac{\nabla(p + \delta p)}{\rho + \delta \rho} \approx \operatorname{rot} \frac{\nabla \delta p}{\rho} = - \left[\frac{\nabla \rho}{\rho^2} \times \nabla \delta p \right]. \quad (5.4.12)$$

Мы учли здесь, что давление в среде за фронтом слабо зависит от координаты и пренебрегли этой зависимостью; зависимость же плотности от координаты обусловила первый сомножитель векторного произведения в (5.4.12). Векторное произведение отлично от нуля, если градиент давления в малом потенциальном возмущении не параллелен градиенту плотности.

Для оценки вихря можно приближенно записать:

$$\nabla \rho \approx \frac{\rho}{l_c}, \quad \nabla \delta p \approx \frac{\delta p}{\lambda} \approx \frac{\delta \rho u^2}{\lambda}, \quad \frac{d \operatorname{rot} v}{dt} \approx \frac{\operatorname{rot} v}{\tau_v},$$

где $l_c = u \tau_c$ — размер зоны охлаждения, τ_c — время охлаждения, λ — характерный пространственный масштаб потенциального возмущения, τ_v — характерное время нарастания вихря. Тогда из (5.4.11), (5.4.12) следует:

$$\operatorname{rot} v \approx \frac{u}{l_c} \frac{\delta p}{\rho} \approx \frac{1}{\tau_c} \frac{\delta p}{\rho}. \quad (5.4.13)$$

Мы предположили при этом, что все величины в возмущенном потоке изменяются быстрее, чем в невозмущенном,

$$\lambda < l_c, \quad \tau_v < \tau_c$$

(это позволяет, в частности, не учитывать тепловую неустойчивость).

Потенциальное возмущение, распространяющееся в газе за фронтом ударной волны (т. е. звук, но в неоднородной среде), возбуждает, как мы видим, вихревую моду малых возмущений*). Тот же гидродинамический механизм способен, очевидно, усиливать вихрь, рожденный в данном слое газа на фронте ударной волны и сносимый общим потоком в область более высоких плотностей.

Заметим, что при наличии сильного охлаждения состояние газа за фронтом ударной волны можно, в известном смысле, рассматривать как сильное энтропийное возмущение, а генерацию вихря — как результат взаимодействия с этим возмущением слабого потенциального возмущения. В гидродинамике явление такого рода (но при энтропийном возмущении, считающемся слабым) известно как эффект Бьеркена (см. Монин и Яглом (1966)).

Неустойчивость фронта ударной волны в среде убывающей плотности. Ударная волна, распространяющаяся в однородной среде, устойчива относительно малых возмущений ее фронта. Это свойство, надежно установленное и теоретически, и экспериментально (Зельдович и Райзер (1966)), можно понять на основе того, что мы говорили выше о взаимодействии малых возмущений с фронтом ударной волны. Например, звуковые волны возмущают и саму ударную волну, но отклонение скорости распространения ударной волны от ее невозмущенного значения не нарастает со временем.

Можно, однако, предполагать, что при распространении в существенно неоднородной среде ситуация будет иной. Такая неоднородность означает сильное возмущение в газе перед фронтом, и потому картина не сводится к уже рассмотренной и представляет собой пример взаимодействия двух сильных возмущений друг с другом.

В гидродинамике ударные волны в неоднородной среде детально рассматривались, причем было показано, что в среде убывающей плотности скорость распространения ударной волны зависит от плотности (Чизнелл (1955), Уизем (1958), Седов (1962), Зельдович и Райзер (1966)):

$$v \propto \rho^{-\alpha}.$$

При положительных α ударная волна ускоряется в среде убывающей плотности.

*) Подробная количественная теория дана в работе Громова, Зильберглейта и Чернина (1978).

Вопрос об устойчивости ускоряющихся ударных волн поставлен Гуревичем и Румянцевым (1970) (см. также работу Румянцева и Чернина (1975)). Можно представить себе, что если какой-либо участок фронта случайно выдвинется вперед, то он окажется в том месте среды, в котором скорость распространения больше, чем скорость основного фронта. Поэтому «выдвинувшийся» участок фронта будет перемещаться дальше быстрее, чем основной фронт; его расстояние от основного фронта будет возрастать и соответственно будет возрастать относительная скорость этого участка и основного фронта. Значит, случайное начальное возмущение будет нарастать со временем. Наоборот, если какой-либо участок случайно отстанет от основного фронта, то он окажется в том месте среды, где скорость распространения меньше скорости основного фронта, и его отставание будет нарастать со временем. В этом качественном рассуждении предполагается, что размер участка возмущения вдоль фронта много больше расстояния, на которое он выдвигается (или отстает) от невозмущенного положения. Тогда участки возмущения, оказавшиеся впереди и позади невозмущенного фронта, движутся почти перпендикулярно к последнему, т. е. почти параллельно друг другу, не пересекаясь и не взаимодействуя между собой. Нарастание возмущений до заметной величины происходит за время $t_n \approx \lambda/v$, где λ — размер неоднородности плотности. При дальнейшем развитии такой неустойчивости рано или поздно станет существенным взаимодействие отдельных участков разрыва. Его результатом должно быть размывание фронта и дальнейшая хаотизация движений, в которых, вероятно, появляется при этом и вихревая составляющая.

§ 5.5. Формирование скоплений галактик

Ударные волны масштаба скоплений галактик являются, как мы видели в § 5.3, адиабатическими: охлаждение основной массы газа за их фронтами происходит медленнее изменения общих гидродинамических характеристик движения среды. По этой причине эволюция соответствующих сгущений вещества требует сложного анализа, учитывающего в первую очередь изменение со

временем скорости натекания газа на фронт ударной волны. Детальная модель формирования протоскопления в ударных волнах (при некоторых дополнительных предположениях о конкретном характере исходных гидродинамических движений) разработана Сюняевым и Зельдовичем (1972) (см. также книгу Зельдовича и Новикова (1975), в которой приводятся результаты подробных численных расчетов, выполненных Дорошкевичем и Шандариным (1973—1974), Дорошкевичем, Рябенкиным, Шандариным (1973)). Мы опишем эту модель чуть позже, а сейчас дадим общее качественное обсуждение всего процесса, включающего в себя как гидродинамические, так и гравитационные явления.

Качественная картина. Можно представить, что после образования гидродинамического разрыва скорость натекания газа на фронт возрастает в течение некоторого времени, пока не достигнет характерной величины, фигурирующей в общей качественной теории § 5.3 как скорость в данном масштабе движения. Связь последней с охваченной движением полной массой дается соотношением (см. (5.1.1))

$$v \approx V \approx 5 \cdot 10^7 \Omega^{1/3} (1+z)^{1/2} \left(\frac{M}{10^{13} M_{\odot}} \right)^{1/3} \text{ см/сек.} \quad (5.5.1)$$

Эта скорость, по-видимому, близка к максимальному значению скорости натекания и соответствует моменту, когда за фронтом оказалась масса газа, сравнимая с полной массой M (скажем, одна треть или половина ее). Затем скорость натекания должна уменьшаться, а движение в целом затухать. Весь этот процесс происходит за характерное гидродинамическое время, которое согласно общим соображениям (см. §§ 5.1, 5.2) порядка или меньше текущего возраста мира.

На первом этапе процесса скорость натекания может быть столь малой, что газ остается неионизованным и потому не остывает за фронтом, а только увеличивает свою плотность, сжимаясь адиабатически, раза в четыре (см. § 5.2). Это продолжается до тех пор, пока скорость не достигнет значения $v_a \approx 5 \cdot 10^8 \text{ см/сек.}$, соответствующего температуре за фронтом $T_s = T_a \approx 10^4 \text{ }^\circ\text{К}$, при которой начинается ионизация.

На следующем этапе скорость продолжает возрастать, температура на фронте повышается и газ все в большей степени ионизируется; при этом включаются процессы охлаждения, темп которых весьма велик для температур $2 \cdot 10^4 < T_s < T_{ff} \approx 5 \cdot 10^5$ °К (см. § 5.3). Газ быстро охлаждается и сжимается за фронтом до плотности

$$\rho_a = 4\rho(z) \frac{T_s(v)}{T_a}, \quad (5.5.2)$$

которая возрастает со временем по мере возрастания напора на фронт. При этом продолжает сжиматься и та первая порция газа, которая не испытала ионизации, так как во всей области за фронтом должно устанавливаться приблизительно постоянное давление, равное в каждый момент времени кинетическому давлению натекающего газа. На этом, втором, этапе процесса, продолжающемся до достижения скорости $v_{ff} \approx 4 \cdot 10^7$ см/сек, соответствующей температуре $T_{ff} \approx 5 \cdot 10^5$ °К, распределение плотности, давления и скорости за фронтом ударной волны устанавливается, вероятно, быстрее изменения скорости натекающего газа из-за быстрого его охлаждения благодаря высвечиванию в линиях и излучению при свободно-связанных переходах. Поэтому зависимость давления и плотности от скорости в каждый момент времени можно описывать с помощью стационарного решения (5.3.40) — (5.3.43).

Третий этап — адиабатический, как и первый; хотя газ ионизован, но скорость теплопотерь, обязанных теперь главным образом свободно-свободным переходам, мала, и тепловая энергия газа за фронтом практически не теряется за время возрастания скорости натекающего от значения v_{ff} до характерного значения (5.5.1). При этом за фронтом сохраняется значительный градиент температуры (теплопроводность плазмы недостаточно эффективна, чтобы его замыть). Температуру на фронте можно оценить по соотношению (5.3.22):

$$T_s \approx 5 \cdot 10^5 \Omega^{1/2} (1+z) \left(\frac{M}{10^{13} M_\odot} \right)^{2/3} \text{°К}. \quad (5.5.3)$$

Что же касается давления, то оно, по-прежнему, поддерживается приблизительно одинаковым во всей области за фронтом, так как в каждом из слоев — холодном (но

узком) и горячем (широком) — звуковые волны успевают распространиться на весь его размер за достаточно короткое время (ср. (5.3.20)). Отсюда следует, в частности, что плотность холодного ($T \sim 10^4$ °К) газа продолжает и на этом этапе увеличиваться в соответствии с формулой (5.5.2). Она достигает максимального значения, когда температура на фронте максимальна.

Последний, четвертый этап процесса характеризуется постепенным уменьшением скорости натекания газа на фронт при почти неизменной температуре газа за фронтом. При этом напор натекающего газа ослабляется, и вскоре его кинетическое давление становится ниже максимального давления, достигаемого за фронтом в момент, когда скорость натекания имеет наибольшее значение. Ясно, что равенство давлений (теплового и кинетического) за фронтом и перед ним должно в любом случае выполняться. Но это означает, что нагретый газ за фронтом начинает расширяться *) после достижения максимальной температуры и давления; его давление (в основном тепловое) уменьшается, стремясь к величине давления (в основном кинетического) набегающего газа:

$$p \approx \rho(z) v^2(t) \quad (5.5.4)$$

(напомним, что $\rho(z)$ есть средняя космологическая плотность при данном красном смещении). Это расширение, при медленности теплопотерь, является адиабатическим, так что плотность за фронтом падает с уменьшением скорости $v(t)$ по закону

$$\rho \propto v^{2/\gamma}, \quad (5.5.5)$$

следующему из (5.5.4). Вместе с плотностью падает и температура газа:

$$T \propto \rho^{\gamma-1} \propto v^{2 \frac{\gamma-1}{\gamma}}. \quad (5.5.6)$$

Характерное время охлаждения посредством излучения при свободно-свободных переходах при этом возрастает еще более:

$$\tau_{ff} \propto T^{1/2} \rho^{-1} \propto \rho^{-\frac{\gamma-3}{2}} \propto v^{\frac{\gamma-3}{\gamma}}, \quad (5.5.7)$$

*) В сжатом горячем газе возникает и распространяется (от фронта вглубь) так называемая волна разрежения.

Заметим, что связанная гравитацией система, находящаяся в квазистационарном состоянии, обладает отрицательной теплоемкостью: из (5.5.9) следует, что ее температура возрастает по мере отвода тепла (к этому свойству гравитационных систем мы вернемся еще в § 6.1). Так как согласно (5.5.9)

$$T \propto R^{-1}, \quad \tau_{ff} \propto R^{3/2}, \quad (5.5.10)$$

время теплоотвода уменьшается при квазистационарном сжатии и может оказаться меньше гравитационного времени, которое уменьшается при этом медленнее:

$$\tau_G \propto (G\rho)^{-1/2} \propto R^{3/2}. \quad (5.5.11)$$

Равенству времени τ_{ff} и τ_G отвечает размер (см. § 2.2, где дано точное выражение для τ_{ff})

$$\bar{R} \approx \frac{r_e^2 c \sqrt{E_i}}{G} \approx 0,8 \cdot 10^{23} \text{ см}, \quad (5.5.12)$$

выражающийся через одни лишь универсальные константы (Антонов и Чернин (1975), Антонов, Осипков и Чернин (1976)). Здесь r_e — классический радиус электрона, E_i — энергия ионизации водорода, фигурирующие в выражении для времени свободно-свободных переходов (§ 2.2). Размер (5.5.12) несколько меньше типичного размера больших скоплений галактик, так что нужно, вероятно, считать, что общее сжатие газового протоскопления прекращается еще до перехода в режим быстрого охлаждения.

Сжатие может остановиться лишь при условии, что прекращается отвод энергии из системы. Можно предполагать, что одновременно со сферизацией и начальным сжатием протоскопления в нем происходит фрагментация, в результате которой основная доля его вещества превращается в изолированные, достаточно плотные массы того или иного размера. Если фрагментация обязана тепловой неустойчивости, то плотность облаков-фрагментов, определяемая общим условием их равновесия с окружающим газом (5.3.61), может превосходить среднюю плотность протоскопления в отношении температур (ср. (5.3.61))

$$\frac{T}{T_a} \approx \frac{GmM}{RkT_a} \approx 10^4 \frac{\bar{R}}{R} \frac{M}{10^{13} M_\odot}. \quad (5.5.13)$$

Плотность облаков возрастает еще больше, когда каждое из них начинает сжиматься благодаря собственной гравитации; при большом уплотнении, превосходящем (5.5.13), динамическое равновесие облака со средой нарушается и последняя уже не препятствует его падению к центру системы.

Вероятно, большая часть возникших облаков падает не строго к центру, а отклоняется от него благодаря, например, притяжению других облаков, мимо которых данное облако пролетает, и которые способны добавить к его радиальному движению еще и поперечную составляющую. Но даже и не падая все вместе на центр, облака могут сталкиваться между собой при относительных движениях в формирующемся протоскоплении. Если суммарная масса облаков сравнима с полной массой системы, то характерное время контактных столкновений

$$\tau_{cc} \approx \frac{1}{n_c v \sigma} \approx \frac{R_c C}{v} \approx (G\rho)^{-1/2} \frac{R_c}{R} C. \quad (5.5.14)$$

Здесь n_c — концентрация облаков, $v = (GM/R)^{1/2}$ — их относительная скорость, R_c — типичный размер облака, C — скважность, т. е. отношение всего объема системы к суммарному объему облаков; $\sigma \approx R_c^2$.

Столкновения облаков неупруги, кинетическая энергия их относительного движения переходит в тепло (в ударных волнах) и выносится из системы излучением (время охлаждения отдельного облака по крайней мере в C раз меньше времени охлаждения для однородной газовой системы). Однако по мере уплотнения облаков скважность возрастает и столкновения постепенно становятся все более редкими, а затем и вовсе прекращаются. Но тогда прекращается вынос из системы энергии их относительного движения, и механическая энергия всей совокупности облаков-фрагментов больше не меняется. Так система в целом достигает стационарного состояния, ее общее сжатие прекращается, а каждое из облаков начинает свою собственную эволюцию, уже в гораздо меньшей степени зависящую от окружающих его условий. Газовое протоскопление превращается, таким образом, в систему фрагментов — протогалактик, погруженных в оставшийся после фрагментации газ. Очевидно,

же масштабах. Качественно оно основано на наглядной картине, в которой смена расширения сжатием в некотором объеме газа после гашения в нем хаббловских скоростей сопровождается постепенным уплотнением этого объема. Уплотнение вызвано тем, что начальные скорости сжатия не одинаковы в общем случае в различных направлениях и имеется направление, вдоль которого скорость сжатия больше, чем вдоль других. Это различие скоростей не сглаживается в ходе общего сжатия, по крайней мере до тех пор, пока давлением газа можно пренебречь. Такая кинематическая картина совместима и с динамикой, учитывающей гравитационное поле данной массы газа, так как поле также не устраняет различия начальных скоростей сжатия и может даже усилить это различие. В результате сжатие приводит к газовому образованию, размер которого вдоль одного из взаимно перпендикулярных направлений заметно меньше, чем вдоль двух других. Поэтому основные локальные характеристики такого образования можно описывать в одномерном приближении. Если выбрать ось x вдоль направления преимущественного сжатия, то в этом приближении, согласно Зельдовичу (1970),

$$x(t, \chi) = \alpha t^{2/3} \chi + \beta(\chi) t^{1/3}, \quad (5.5.15)$$

где $x(t, \chi)$ — эйлерова координата частицы с лагранжевой координатой χ ; α — константа; $\beta(x)$ — произвольная функция лагранжевой координаты. Первый член справа описывает невозмущенное движение — хаббловское расширение, которое для простоты считается параболическим (см. § 1.2), второй член — возмущение. Давление считается равным нулю. Движение по направлениям y и z рассматривается как невозмущенное: $y \sim z \sim \infty t^{2/3}$. В линейном приближении по возмущению решение совпадает с тем, которое получается в теории малых возмущений (§ 3.1).

Допущение состоит в том, что решение для сильных возмущений сохраняет тот же вид. Экстраполяция результатов теории малых возмущений ведет к тому, что эволюция, описываемая решением (5.5.15), не отличается качественно от инерциального движения без гравитации.

Возмущение нарастает в этом решении неограниченно, и если в некотором слое частиц производная $d\beta/d\chi$ отри-

дательна, то лагранжевы траектории этих частиц рано или поздно пересекутся; в этот момент ($t = t_s$) различным значениям лагранжевой координаты χ будет отвечать одно и то же значение эйлеровой координаты x :

$$\frac{\partial x}{\partial \chi} = \alpha t_s^{2/3} + \frac{d\beta}{d\chi} t_s^{1/3} = 0. \quad (5.5.16)$$

Сингулярность такого рода достигается при

$$t = t_s = \left(\frac{\alpha}{\left| \frac{d\beta}{d\chi} \right|} \right)^{3/2}, \quad (5.5.17)$$

Очевидно, что плотность вещества должна при этом обратиться в бесконечность. Формально это следует из того, что масса слоя данных частиц остается неизменной, а толщина слоя стремится к нулю при $t \rightarrow t_s$:

$$\rho \propto \left(\frac{\partial x}{\partial \chi} \right)^{-1} \quad (5.5.18)$$

(за сингулярность решение не может быть продолжено).

Если $d\beta/d\chi > 0$, то сингулярность не возникает, хотя возмущения могут быть сколь угодно велики: слои частиц движутся в этом случае строго друг за другом и расстояние между любыми двумя слоями возрастает со временем. Очевидно, что для всего движения это возможно лишь при особых начальных условиях, а в общем случае всегда найдутся слои, для которых $d\beta/d\chi < 0$, и сингулярность неизбежно возникнет.

Траектории (5.5.15) отвечает закон скорости

$$\dot{x} = \frac{2}{3} \alpha t^{-1/3} + \frac{4}{3} \beta(\chi) t^{1/3}. \quad (5.5.19)$$

Частный вид функции $\beta(\chi)$, принятый в работе Зельдовича (1970а),

$$\beta = \delta \cos k\chi, \quad k > 0, \quad 0 < \delta \leq 1, \quad (5.5.20)$$

позволяет проследить эволюцию профиля скорости при нарастании возмущения. В лагранжевых координатах он не меняется со временем, но в эйлеровых координатах искажается, так как «быстрые» слои догоняют «медленные». Слои, где $\beta(\chi) = 0$, движутся как невозмущенные, удаляясь друг от друга в соответствии с общим космологическим расширением. Другие слои движутся, приб-

лижаясь к переднему склону «волны» ($\beta = 0, \partial v / \partial \chi < 0$) и удаляясь от заднего склона ($\beta = 0, \partial v / \partial \chi > 0$). В результате передний склон укрупняется и именно на нем возникает в конце концов сингулярность.

Это похоже на нелинейную эволюцию звуковой волны (§ 5.2). Хотя укрупнение фронта происходит в решении (5.5.15) — (5.5.18) по чисто кинематической причине, как это было бы даже при инерциальном движении слоев с различными начальными скоростями, процесс и здесь имеет эволюционный характер, так как вблизи сингулярной плоскости первые сближающиеся слои газа обладают сколь угодно малой относительной скоростью. Формально в решении (5.5.15) — (5.5.20) давление равно нулю; фактически же при таких условиях оно должно быть существенно, так как малая, но конечная скорость звука в газе больше бесконечно малой скорости столкновения первых слоев. Поэтому первые порции газа сжимаются вблизи сингулярной плоскости адиабатически, а сверхзвуковой характер движения в целом проявляется позже, когда возрастающие относительные скорости новых порций газа становятся больше скорости звука в слое. Формирующиеся при этом разрывы представляют собой эволюционные ударные волны, расходящиеся в противоположные стороны от сингулярной плоскости.

Как показывает подробный расчет для массы $\sim 10^{13} M_{\odot}$, примерно половина всего вещества, охваченного движением, проходит через ударную волну за промежуток времени $\sim t_s$. Прошедший через фронт газ делится приблизительно поровну на плотную холодную ($\sim 10^4$ °К) и разреженную горячую ($\sim 10^6$ — 10^7 °К) фракции. Плотный газ составляет центральный слой, а разреженный окружает его и расположен позади фронтов ударных волн (рис. 16).

Предполагается, что плотный слой фрагментирует на протогалактики вследствие гравитационной неустойчивости *). Что же касается горячего разреженного газа, то он может частично сохраниться в объеме скопления и до настоящего времени; время его охлаждения при свободно-свободных переходах достаточно велико. Соответ-

*) В теории «блинов» сферизация скопления происходит, как предполагается, уже после распада слоя на фрагменты.

ствующее излучение может проявляться в рентгеновском диапазоне. Некоторые богатые скопления действительно обнаруживают, как мы говорили, рентгеновскую светимость.

О вращении протогалактик. Правильные скопления населены главным образом эллиптическими галактиками,

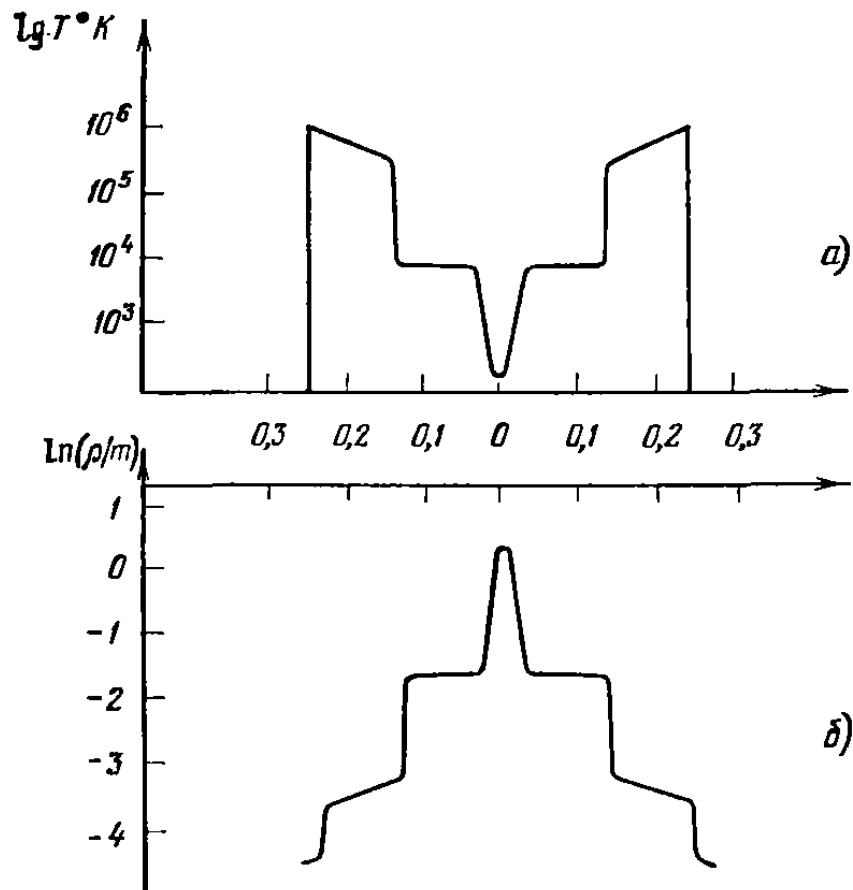


Рис. 16. Модель блина. Распределение температуры (а) и плотности (б) в момент, когда в слое между ударными волнами находится половина характерной массы; по горизонтальной оси доля характерной массы.

вращение которых, по-видимому, слабее, чем вращение спиралей. Попробуем выяснить, действительно ли в протоскоплениях, формирующихся рассмотренным выше путем, вращение фрагментов — протогалактик — должно быть слабым. Для этого оценим эффективность различных механизмов, способных в таких условиях придавать фрагментам вращение.

Одним из механизмов является приливное взаимодействие протогалактик. Как предполагал Ф. Хойл, если фрагменты несферичны, то, пролетая друг мимо друга, они могут получить вращение за счет взаимного гравита-

ционного воздействия (причем, разумеется, закон сохранения момента не нарушится). Сила, действующая со стороны одного тела на другое, различна для разных частей последнего, так как они находятся на разных расстояниях (рис. 17). Отсюда возникает вращательный

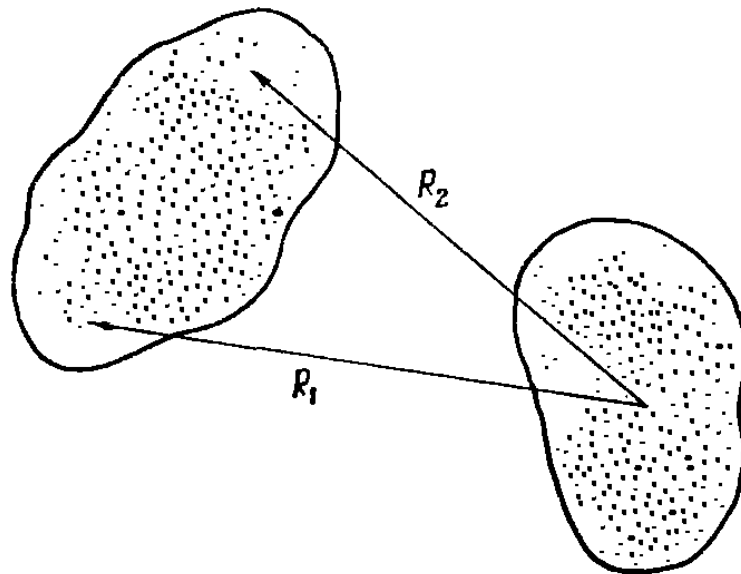


Рис. 17. Приливное взаимодействие протогалактик.

момент, который по порядку величины есть произведение соответствующей разности сил

$$GM_G \left(\frac{1}{R_1^2} - \frac{1}{R_2^2} \right) \approx \frac{2G\tilde{R}_G}{R_1^3} M_G$$

на массу тела M_G , его размер \tilde{R}_G и время пролета τ (R_1 и R_2 — расстояния, показанные на рисунке):

$$K \approx 2 \frac{GM_G^2}{R_1^3} \tilde{R}_G \tau.$$

Для эффективности приливного взаимодействия необходимы, очевидно, достаточно тесные пролеты, при которых расстояния между телами сравнимы с их собственными размерами. Время пролета должно оцениваться тогда как это расстояние, деленное на относительную скорость. Если скорость фрагментов определяется их движением в общем гравитационном поле скопления, $V \approx (GM/R)^{1/2}$ (где M и R — полная масса и размер скопления), то

$$K \approx 2 \sqrt{GM_G^3 R} \left(\frac{C}{N^2} \right)^{1/8} \approx K_G \left(\frac{C}{N^2} \right)^{1/8}. \quad (5.5.21)$$

Здесь N — число фрагментов (для простоты оценки они считаются приблизительно одинаковыми), C — скважность, которая уже фигурировала выше. Величина K_G представляет собой момент, соответствующий равенству центробежной и гравитационной сил, $GM_G/\bar{R}_G^2 \approx v_G^2/\bar{R}_G^2$, v_G — скорость вращения фрагмента; именно такое равновесие и осуществляется, по-видимому, в плоских подсистемах спиральных галактик.

Считая, что фрагменты образуются из холодного слоя газа с плотностью, близкой к плотности галактик, как в картине «блинов» (см. рис. 16), нужно положить, в (5.5.21) $R_1 \approx \bar{R}_G$, $C \approx 1$. Но тогда при $N \approx 100-1000$ момент окажется гораздо меньше момента спиральных галактик.

Набор момента, вероятно, более эффективен в том случае, когда галактики составляет пару (постоянную или временную), в которой относительная скорость галактик определяется их взаимным тяготением: $(GM_G/\bar{R}_G)^{1/2} \sim \sim (GM_G/R_1)^{1/2}$. В этом случае (при тех же прочих условиях) относительная скорость меньше, а время действия приливных сил больше, чем в (5.5.21), и тогда

$$K \approx K_G \approx (GM_G^3 \bar{R}_G)^{1/2}. \quad (5.5.22)$$

Последнее, впрочем, можно было заранее ожидать по соображениям размерности: (5.5.22) представляет собой единственную комбинацию нужной размерности, составленную из трех фигурирующих здесь величин.

Момент (5.5.22) имеет нужный для спиралей порядок величины, но оценка такого рода все же слишком груба и, как оказывается, завышена. Детальный расчет Пиблса (1969, 1974) показывает, что приобретаемый момент раз в пять меньше необходимого при самых благоприятных условиях.

Другим механизмом такого рода могла бы быть генерация вихрей на фронтах ударных волн, которую мы обсуждали в § 5.4; ее роль в космогонии галактик указана Черниным (1970а, б), а Дорошкевич (1973) исследовал возможность применения ее к условиям в «блинах»-протоскоплениях.

В плотном холодном слое газа, который, по предположению (см. предыдущий раздел), распадается на протогалактики, температура, как указывалось выше, со-

ставляет $\approx 10^4$ °К. Скорость звука при этом $\approx 10^3$ см/сек. Если вихри генерируются в дозвуковом режиме, что естественно предполагать, то соответствующие скорости слишком малы *), чтобы обеспечить вращение галактик, для которых типичными являются скорости 10^7 — $3 \cdot 10^7$ см/сек. (При этой оценке учитывается, что заметное усиление вращения при дальнейшем сжатии фрагментов с замороженными вихрями невозможно, так как сжатие не может быть сильным: плотность фрагментирующего слоя с самого начала близка к типичной плотности галактик).

Итак, в условиях газового сгущения, способного дать начало правильному скоплению галактик, вращение большей части фрагментов-протогалактик вряд ли может быть столь сильным, как то, которое наблюдается у спиральных галактик, подобных нашей Галактике. Вероятно, быстрое вращение фрагментов требует существенно иных условий в протоскоплении; но тогда нужно ожидать, что и само скопление в целом окажется в результате иным.

§ 5.6. Неправильные скопления и происхождение вращения галактик

Эмпирические данные об относительной распространенности скоплений двух типов, правильных и неправильных, остаются пока не вполне определенными; и все же правильные скопления, подобные Комае, встречаются реже, чем неправильные, подобные Деве (Воронцов-Вельяминов (1972)). Не слишком определены и данные об относительной распространенности галактик двух типов — эллиптических и спиральных; первых, однако, во всяком случае меньше раза в два или три (Воронцов-Вельяминов (1972)). Одно, в общем, кажется согласующимся с другим ввиду положительной корреляции эллиптических галактик с правильными скоплениями.

Возможную картину формирования правильных скоплений мы обсуждали в предыдущем параграфе, а здесь обратимся к неправильным скоплениям, в которых, как говорилось, всегда имеются в большом числе быстро вращающиеся спиральные галактики.

*) Если протогалактики возникают также и в горячем разреженном газе, их скорости могут быть больше.

Турбулентность метагалактической среды. Обсуждая в 1929 г. образование спиральных галактик при гравитационной конденсации разреженной среды, Дж. Джинс предполагал, что их угловой момент был первоначально рассеян в этой среде в форме беспорядочных потоков или движений соответствующих газовых масс.

Действительно, на фотографиях, сделанных с помощью больших телескопов, спиральные галактики выглядят яркими вихрями звездных облаков. Гидродинамические аналогии возникают сами собой и побуждают предположить, что эти звездные системы возникли благодаря вихревым движениям в дозвездном, догалактическом веществе Вселенной. Задолго до того идеи космических вихрей обсуждались в космогонических теориях Декарта, Канта, Лапласа применительно к Солнечной системе (см. § 7.2).

В 50-е годы Вейцеккер (1951) и Гамов (1952) рассматривали галактики или «замороженные вихри», выделившиеся из непрерывной метагалактической среды, охваченной турбулентным движением. Новые наблюдательные аргументы в пользу догалактической турбулентности приводит Оорт (1972).

Теоретический анализ особенностей турбулентных движений в расширяющейся среде дается в работах Нариан (1956) и Каплана (1959); после открытия реликтового излучения физические условия в турбулентной метагалактической среде исследовали Озерной и Чернин (1967, 1968), Харрисон (1971, 1973), Рис (1971), Джонс и Пиблс (1972), Силк и Эймс (1972), Силк и Ли (1973), Пиблс (1975), Зельдович и Новиков (1975); работы японской группы суммированы в обзоре Хайякавы и др. (1971); весьма полный обзор литературы дан Джонсом (1976).

Турбулентность представляет собой одно из самых интересных проявлений нелинейных эффектов в гидродинамике. Если в среде имеются гидродинамические движения различных типов, масштабов, интенсивностей, то их развитие происходит, вообще говоря, не независимо; между ними осуществляется взаимодействие, эффективность которого зависит в первую очередь от амплитуды скорости в данном масштабе движений. При очень малых скоростях взаимодействие движений является слабым и именно этот случай мы имели в виду в теории малых возмущений (гл. 3), когда рассматривали возможные типы

слабых возмущений, пренебрегая их взаимодействием между собой. При этом мы всегда ограничивались линейным приближением по скорости. Если же скорости движений не малы, то линейное приближение недостаточно, взаимодействие становится существенным, и оно может проявиться, например, в том, что движения одного типа будут порождать движения другого типа (см. § 5.4). При таких условиях становится возможным также взаимодействие различных масштабов движений, при котором движение данного масштаба вызывает возникновение движений других масштабов. Благодаря этим процессам гидродинамическая нелинейность способна привести к развитию турбулентности — возбужденного состояния среды с беспорядочными и запутанными хаотическими движениями различных типов, масштабов и интенсивностей. Нелинейность способна при определенных условиях привести такой набор движений к некоторой статистической упорядоченности, характеризуемой универсальным спектром скоростей.

Определим эти понятия несколько точнее. Относительная скорость двух элементов среды, находящихся на некотором расстоянии λ друг от друга, различна в разных местах среды и в разные моменты времени. Ее среднее значение по различным парам элементов, находящихся на заданном расстоянии λ , равно нулю, если движения полностью хаотичны, т. е. нет регулярного потока. Однако среднее значение квадрата скорости $\langle v^2 \rangle$ отлично от нуля, и под характерной средней скоростью турбулентности понимают корень квадратный из этой величины.

В соответствии с классификацией малых гидродинамических возмущений, указанной в гл. 3, будем различать турбулентность двух видов: вихревую, для которой $\text{div } v = 0$ и которая характеризуется распределением ротора скорости, и потенциальную, или акустическую, для которой, наоборот $\text{rot } v = 0$, так что она характеризуется распределением дивергенции скорости. Для обоих этих видов движений установление спектра обязано нелинейным гидродинамическим процессам, которые делают возможным переход кинетической энергии из одного масштаба движений в другой.

Нелинейные эффекты формально описываются членом $(v \nabla) v$ в уравнениях гидродинамики. При отсутствии

такого члена уравнения движения имеют решение вида

$$v(r, t) = v_k(t) e^{ikr}.$$

Здесь характерным масштабом является длина волны $\lambda = 2\pi/k$. При наличии нелинейного члена последнее выражение уже не является решением уравнений, но решение всегда может быть написано в виде ряда (или интеграла) Фурье:

$$v(r, t) = \sum_k v_k(t) e^{ikr}.$$

Взаимодействие масштабов становится возможным именно вследствие наличия нелинейного члена. Действительно, левая часть уравнения движения примет при таком выражении для скорости $v(r, t)$ вид

$$\sum_k e^{ikr} \left\{ \frac{\partial}{\partial t} v_k + i \sum_{k'} v_k (k' v e^{ik'r}) \right\}.$$

Различные масштабы $\sim k^{-1}$ и $\sim (k')^{-1}$ обязательно, как видно отсюда, связаны вторым, нелинейным членом в скобке. Это означает, что если в начальный момент возбуждено движение масштаба k^{-1} , то с течением времени кинетическая энергия будет, вообще говоря, переходить в другие масштабы.

Для стационарной вихревой турбулентности спектр движений был установлен Колмогоровым из простых соображений размерности. Энергия движений данного масштаба в расчете на единицу массы ($1/2 v^2$) может быть передана в другой, близкий к нему масштаб за характерное время $t_h \simeq \lambda/v$, которое служит единственной величиной размерности времени, могущей играть роль в этом процессе *). Значит, за единицу времени из одного масштаба в другой передается энергия $q \approx 1/2 v^3/\lambda$. Передача энергии должна происходить главным образом от больших масштабов к меньшим. Такое направление потока энергии объясняется тем, что в самых малых масштабах начинает играть роль вязкость, и энергия гидродинамических движений превращается в тепло. Мы видели (например, в § 3.3), что характерное время действия вязкости дейст-

*) Характерное гидродинамическое время определено в § 5.2 из условия эффективности нелинейных процессов.

вительно тем меньше, чем меньше масштаб. Для поддержания стационарного состояния необходим приток энергии именно в эти малые масштабы из основных энергосодержащих масштабов движения. В промежуточных масштабах, превышающих масштаб действия вязкости, диссипации нет, так что поток энергии из масштаба в масштаб должен быть одинаков для всех этих масштабов, $q = \text{const}$. Отсюда следует зависимость скорости от масштаба:

$$v \equiv (\langle v^2 \rangle)^{1/2} \propto \lambda^{1/3}. \quad (5.6.1)$$

Это и есть спектр стационарной вихревой турбулентности. Он устанавливается в тех масштабах, где гидродинамическое время меньше вязкого времени $\tau_v \simeq \lambda^2/\nu$:

$$\lambda > \lambda_{\min} \approx \frac{\nu}{v_{\min}} \approx \frac{\nu^{3/4}}{q^{1/4}}. \quad (5.6.2)$$

(Здесь ν — кинематический коэффициент вязкости.)

В случае акустической турбулентности спектр движений должен иметь по ряду причин иной вид. В этом случае простые соображения размерности недостаточны, так как, кроме величины $t_h = \lambda/\nu$, имеется еще одна величина размерности времени $t_u = \lambda/u$, где u — скорость звука. Тем не менее, согласно работе Захарова и Сагдеева (1970), спектр движений и в этом случае может иметь столь же универсальный характер, как и в вихревой турбулентности, но с другим показателем: $v \propto \lambda^{1/4}$. Если же в акустической турбулентности появляются разрывы и ударные волны (случай, особенно интересный для нас), то поток энергии уменьшается при переходе к меньшим масштабам и потому скорость v должна спадать несколько быстрее:

$$v \propto \lambda^{1/2}. \quad (5.6.3)$$

Вопрос о спектре такой турбулентности обсуждается в книгах Каплана (1958), Каплана и Пикельнера (1963), работе Кадомцева и Петвиашвили (1973). В обоих случаях показатели степени не слишком отличаются от колмогоровского.

В метагалактической среде турбулентные движения, наложенные на общее регулярное космологическое расширение, имеет различный характер в самых больших

масштабах, где скорость регулярного расширения должна быть, вероятно, больше скорости турбулентности, и в малых масштабах, в которых соотношение этих скоростей может быть обратным. Спектр в области больших масштабов целиком определяется в таком случае начальным состоянием движений, а в области малых масштабов — процессами взаимодействия движений и уже не зависит от начального состояния. Действительно, во втором случае за время, в течение которого гидродинамическое движение со скоростью v успеет охватить весь масштаб λ , не произойдет существенного регулярного расширения среды, так что турбулентность осуществляется практически как в стационарной среде и нелинейные процессы успевают привести движения к стационарному состоянию в каждую данную эпоху. Наоборот, в больших масштабах гидродинамическое движение не успевает, так сказать, охватить весь масштаб за время t , потому что расширение будет все время его «обгонять».

Масштаб, в котором гидродинамическая скорость сравнивается со скоростью регулярного расширения,

$$\lambda_{\max} \approx v_{\max} t \approx q^{1/2} t^{3/2}, \quad (5.6.4)$$

служит максимальным масштабом турбулентности. Он может возрасти со временем, так что в ходе космологического расширения в нелинейное взаимодействие вовлекаются движения все больших масштабов, энергия которых передается в меньшие масштабы, поддерживая в них состояние квазистационарной турбулентности.

Условие $t_h \lesssim t$ означает, что собственная скорость движений в данном масштабе больше скорости регулярного хаббловского расширения, но это сильное возмущение поля скорости не обязательно ведет к сильным возмущениям плотности. Действительно, если движения являются дозвуковыми (что, во всяком случае, следует считать выполненным на стадии преобладания излучения, когда скорость звука близка к скорости света), то возмущения плотности в потенциальных движениях масштаба, меньшего джинсовой длины, определяются, как и всегда в звуковой волне, не отношением v/V , а отношением v/u :

$$\frac{\delta\rho}{\rho} \approx \frac{v}{u} \approx \frac{v}{c}, \quad \rho_r \gg \rho. \quad (5.6.5)$$

Что же касается вихревых движений, то они вызывают возмущения плотности только во втором порядке по этому отношению:

$$\left(\frac{\delta\rho}{\rho}\right) \approx \left(\frac{v}{u}\right)^2, \quad v < u. \quad (5.6.6)$$

Это явление представляет собою нелинейную генерацию звуковых волн вихрями (Лайтхилл (1960), Кляцкин (1966а, б), Монин и Яглом (1967)); в космологии эффект рассматривали впервые Озерной и Чернин (1968). Считая условия $t_h < t$, $v < u$ выполненными, покажем, как приближенное соотношение (5.6.6) может быть получено из уравнений гидродинамики. Первое из указанных условий позволяет пренебречь расширением среды и в той же мере ее гравитацией, а второе гарантирует отсутствие ударных волн.

Пусть сначала в среде имелось одно лишь вихревое движение; по истечении времени $\sim t_h$ можно ожидать появления слабого потенциального движения, и чтобы оценить его скорость, запишем нерелятивистские уравнения движения и неразрывности для полной скорости движений, $v = v_v + v_p$, представляющей собою сумму вихревой и потенциальной компонент:

$$\frac{\partial v}{\partial t} + (v\nabla)v = -\frac{\nabla p}{\rho} = -u^2 \frac{\nabla\rho}{\rho},$$

$$\frac{\partial\rho}{\partial t} = -\operatorname{div}(\rho v) = -\rho \operatorname{div} v_p - v\nabla\rho.$$

Оценим члены этих уравнений по порядку величины:

$$\frac{\partial v}{\partial t} \sim \frac{v}{t_h}, \quad \nabla\rho \sim \frac{\delta\rho}{\lambda},$$

$$\frac{\partial\rho}{\partial t} \sim \frac{\delta\rho}{t_h}, \quad \operatorname{div} v_p \sim \frac{v_p}{\lambda}.$$

В нелинейном члене $(v\nabla)v$ нужно по смыслу рассматриваемой картины оставить только вихревую скорость:

$$(v\nabla)v \sim \frac{v_v^2}{\lambda}.$$

В уравнении неразрывности члены $\partial\rho/\partial t$ и $v\nabla\rho \sim v_v\delta\rho/\lambda$ имеют одинаковый порядок. Сравнивая эти члены с

членом $\rho \operatorname{div} v_p$, находим

$$\frac{\delta\rho}{\rho} \sim \frac{v_p}{\lambda} t_h.$$

В левой части уравнения движения оба члена на интервале времени $\sim t_h$ имеют одинаковый порядок v_v^2/λ , а правая часть

$$u^2 \frac{\nabla\rho}{\rho} \sim u^2 \frac{\delta\rho}{\lambda\rho} \sim u^2 \frac{v_p}{\lambda v_v}.$$

Отсюда следует соотношение

$$\frac{v_p}{v_v} \sim \left(\frac{v_v}{u}\right)^2, \quad (5.6.7)$$

а вслед за ним и (5.6.6).

Оценки такого рода справедливы и для релятивистских уравнений гидродинамики (Чернин и Эйдельман (1969, 1971)). Из них следует, например, что вихревые движения на стадии преобладания излучения порождают слабые потенциальные движения с относительными амплитудами

$$\frac{\delta\rho}{\rho} \sim \left(\frac{v_v}{c}\right)^2, \quad \frac{v_p}{v_v} \sim \left(\frac{v_v}{c}\right)^2. \quad (5.6.8)$$

Напомним, что нелинейный эффект другого рода — порождение вихря слабым потенциальным движением — запрещен теоремой сохранения циркуляции скорости (§ 3.2). Для релятивистской среды запрет налагается на псевдовихрь, играющий в релятивистской гидродинамике ту же роль.

В эпоху рекомбинации происходит сильное снижение скорости звука в среде. Если до рекомбинации, когда вещество и излучение взаимодействовали между собой и упругость среды определялась фотонами, скорость звука приближалась к скорости света, то после «отключения» излучения от вещества она упала, как мы знаем, более чем на четыре порядка. При таких условиях первоначальные движения, интересные с космогонической точки зрения *), оказываются сверхзвуковыми. Теперь выпол-

*) То есть, например, вихревые движения со скоростью, сравнимой с вращательной скоростью спиральных галактик $\sim 10^6$ — 10^7 см/сек.

нение условия $t_h < t$ неизбежно привело бы к сильным возмущениям. Но, как мы видели в §§ 5.1, 5.3, сильные возмущения, давшие начало наблюдаемым космическим структурам большого масштаба, развиваются позднее, в эпоху, характеризующуюся красным смещением < 100 или даже $\lesssim 10$. Поэтому нужно считать, что в выживающих к эпохе рекомбинации масштабах возмущений гидродинамическое время больше космологического возраста.

Для движений обоих рассмотренных здесь типов — потенциальных (в масштабах, меньших джинсова) и вихревых — отношение $t_h/t \approx v/V$ убывает на ранних стадиях изотропного расширения; поэтому условие эффективности нелинейных процессов не может выполняться для них в прошлом, до рекомбинации, раз оно не выполнено в данном масштабе в эпоху рекомбинации. Отсюда следует важный вывод: движения в области масштабов, выживших к послерекомбинационной эпохе, не могут находиться в статистическом упорядоченном состоянии, характеризуемом спектром универсального вида. Движения сохраняли к этой эпохе первоначальный спектр, определяемый физическими условиями вблизи сингулярного состояния, о которых можно только строить гипотезы. (Примеры гипотез такого рода см. в работах Зельдовича и Новикова (1969) и Зельдовича (1972)).

Результат мог бы быть иным только в случае немалых начальных возмущений при сильной неоднородности и изотропии ранней Вселенной (гл. 4). Не исключено, что общая изотропизация космологического расширения сопровождалась нелинейным гидродинамическим взаимодействием локальных возмущений, которые при этом диссипировали благодаря нелинейным эффектам и, становясь слабыми, одновременно приобретали спектр универсального вида благодаря тем же нелинейным эффектам. Эта возможность остается не изученной ввиду ее очевидной сложности.

Турбулентность с универсальным спектром может, однако, возникнуть позднее, на активной фазе космогонического процесса, когда гидродинамические движения, усиленные гравитационной неустойчивостью, ведут к появлению гидродинамических разрывов, ударных волн, сжимающих и разогревающих отдельные участки среды. В этих участках среды генерируются новые вихревые

движения, причем вторичная турбулентность такого происхождения развивается в дозвуковом режиме (см. ниже).

Первичные вихри. Вейдзеккер (1951) и Гамов (1952) представляли себе турбулентность догалактической среды как совокупность вихрей различных масштабов, способных погасить регулярное космологическое расширение в отдельных участках среды и тем самым привести к обособлению соответствующих газовых масс. «Вмороженные» в эти массы вихри обеспечивают собственные движения сгущений-протогалактик и в первую очередь их вращение.

Вихревые движения в эпоху обособления протогалактик должны иметь скорости, превосходящие скорость звука в среде, и в этом смысле догалактическая турбулентность должна быть сверхзвуковой. В гидродинамике такое состояние среды остается неизученным; насколько известно, в ней не было до того и самого понятия сверхзвуковой турбулентности (остающегося, впрочем, до сих пор определенным не слишком четко). Нет оснований допускать, что для сверхзвуковых вихревых движений могут выполняться универсальные статистические закономерности, установленные для локально изотропной стационарной вихревой турбулентности при дозвуковых скоростях. Эти закономерности предполагают передачу кинетической энергии от движений больших масштабов движениям меньших масштабов вплоть до самых малых, где эта энергия переходит в тепло. При сверхзвуковых движениях каскадная передача энергии от больших масштабов к малым невозможна без потерь, ее превращение в тепло происходит не только в самых мелкомасштабных движениях, где действует вязкость, но и в движениях всех масштабов, когда они вызывают появление ударных волн с сильной диссипацией энергии на фронтах. Для сверхзвуковых вихрей неприменимы, по-видимому, непосредственно и соображения, приводящие к универсальному спектру хаотических потенциальных движений с разрывами (см. выше). Можно лишь предполагать, что в состоянии хаотических сверхзвуковых движений должна существовать сложная картина движущихся, сталкивающихся, разрушающихся и вновь рождающихся сгущений вещества, «облаков», обладающих и спиновым и орбитальным вращением друг относительно друга.

Теоретический анализ наталкивается здесь на серьезные, до сих пор не преодоленные трудности. Намеченная в свое время Вейцеккером программа исследования отдельных ключевых процессов (таких, как распространение нестационарного разрыва, взаимодействие сильных разрывов и т. п.) осталась невыполненной (см. доклад Хорнера (1962) об итогах работы группы Вейцеккера). Проблема происхождения догалактических вихрей не входила в эту программу. Допускалось, что вихри существуют во Вселенной изначально и имеют ту же природу, что и общее регулярное расширение среды (Гамов (1952)).

Исследованию поведения вихревых движений в ранней Вселенной посвящены §§ 3.2 — 3.4, 4.2, 5.4; роль вихревых движений в процессе перехода к активной фазе космогонического процесса см. в § 5.1, а в предыдущем разделе настоящего параграфа рассматривались особенности вихревой турбулентности в метagalacticкой среде.

Мы знаем теперь, что вихревые движения нужной интенсивности могли быть дозвуковыми на ранней стадии расширения, когда вещество и излучение представляли собой ультрарелятивистскую жидкость, скорость звука в которой близка к скорости света. При этом начальные фазы космологического расширения должны быть анизотропными. На более поздней фазе величина скорости вихревого движения оставалась неизменной в ходе расширения, ставшего в больших масштабах изотропным. Переход движений в сверхзвуковой режим происходит при рекомбинации вещества, когда оно отделяется от излучения и скорость звука в нем, определяемая лишь температурой, сильно падает. Изначальные вихревые движения могут, таким образом, породить вскоре после рекомбинации сверхзвуковую турбулентность *).

Чтобы преодолеть космологическое расширение в определенных участках среды, вихри должны иметь скорость, сравнимую с хаббловской скоростью в соответствующем масштабе. Так как отношение первой скорости ко второй падает при космологическом расширении в послерекомбинационную эпоху (см. § 3.2), то это условие выполняется либо сразу после рекомбинации, либо никогда. Поэто-

*) Основные черты этой картины были выяснены Озерным и Черниным (1967, 1968).

му в картине первичных вихрей возникновение сверхзвуковой турбулентности, способной приводить к обособлению сгущений-протогалактик, нужно относить к эпохе рекомбинации. Но тогда плотность протогалактических облаков должна быть не меньше космологической плотности в эпоху рекомбинации ($\sim 5 \cdot 10^{-21} \Omega \text{ г/см}^3$). На самом деле в ударных волнах достигались бы во много раз большие плотности $\sim 10^{-18} \text{ г/см}^3$, как это следует из соотношения (5.3.24), которое дает плотность газа за фронтом ударной волны с сильным охлаждением; охлаждение действительно должно быть при этом сильным, так как время комптоновского охлаждения (5.3.1) много меньше возраста мира в эпоху рекомбинации.

Эти плотности значительно превышают типичную плотность галактик ($\sim 10^{-24} \text{ г/см}^3$), и потому приходится заключить, что первичные вихри не могли породить протогалактические сгущения ни в эпоху рекомбинации, ни позднее. По той же причине они не могли обеспечить достаточно быстрым вращением протогалактики, которые могли бы обособиться — по другим причинам — в нужную эпоху. Действительно, в спиральных галактиках (точнее, в их плоских подсистемах) вращение столь быстрое, что центробежная сила равняется гравитационной. Если такое равенство выполняется при плотности галактик, то при плотности в эпоху рекомбинации, которая порядка на три больше, в соответствующих масштабах преобладала бы центробежная сила. Ясно, что это снова отвечает скорости вращения, превосходящей хаббловскую скорость:

$$\frac{v^2}{\lambda} > \frac{GM_G}{\lambda^2}, \quad v > \left(\frac{GM_G}{\lambda} \right)^{1/2} \sim \lambda (G\rho)^{1/2} \sim \frac{\lambda}{t}.$$

Мы видим, что идея первичных вихрей приводит к некоторым противоречиям при попытке придать ей бóльшую определенность на основании новых космологических сведений, ставших известными после открытия реликтового излучения *). Но получаемые при этом конк-

*) Усложненные варианты космогонической схемы, связанные, например, с возможной задержкой рекомбинации из-за нагрева среды движениями, требуют слишком искусственного подбора параметров исходного движения (Чернин (1971, 1972а, б); см. также работы Озерного и Чибисова (1970), Озерного (1971), Курскова и Озерного (1974, 1975)).

ретные результаты способны послужить исходным материалом для дальнейших исследований в гидродинамике горячей Вселенной.

Протоскопление как турбулентный слой. Общие соображения Джинса, Вейцзеккера, Гамова, Оорта о турбулентном состоянии метagalактической среды, как и впечатляющие гидродинамические аналогии, о которых мы упоминали выше, не предполагают с необходимостью изначальное существование космических вихрей, и потому сделанное заключение о первичных вихрях не отменяет исходных предпосылок и мотивов турбулентной космогонии.

Наше заключение относится к предыстории догалактических вихрей и, вероятнее всего, его следует понимать как указание на «вторичную» природу этих движений. Можно предполагать, что вихри, давшие начало вращению галактик, действительно существовали в эпоху формирования звездных систем, но не были изначальными, а возникли незадолго до того в сверхзвуковых движениях метagalактической среды (Чернин (1970 а, б).

Некоторые механизмы генерации вторичных вихрей подробно проанализированы в § 5.4; они, по-видимому, не могли эффективно действовать в тех газовых сгущениях большого масштаба, которые превращались в правильные скопления галактик (§ 5.5). Образование таких сгущений происходило благодаря ударным волнам эволюционной природы, которые уплотняли большие массы газа; эти массы оказывались затем изолированными от окружающей их метagalактической среды и друг от друга и начинали индивидуальную эволюцию, слабо зависящую от внешних условий.

Обособление больших газовых масс благодаря эволюционным ударным волнам — только один из множества нелинейных процессов в среде, охваченной хаотическими сверхзвуковыми движениями. Сложная картина сильно возбужденного состояния среды обязательно включает в себя и процессы неэволюционного характера. Большие сгущения вещества, «облака», выделившиеся в той или иной степени из метagalактической среды, движутся друг относительно друга и при подходящих условиях могут испытывать контактные столкновения. Столкновение также, очевидно, представляет собой существенно нелиней-

ное явление — взаимодействие сильных возмущений друг с другом.

Для столкновений требуется, чтобы относительные собственные скорости облаков были не малы по сравнению с хаббловскими скоростями в соответствующих масштабах.

Это означает, что возмущения должны быть сильными не только в масштабе отдельного облака, но и в масштабе пар или групп облаков. В сильно возбужденном состоянии среды на активной фазе космогонического процесса имелась, вероятно, целая иерархия собственных гидродинамических движений, которая породила соответствующую иерархию плотных облаков газа. При этом одна и та же порция газа могла, например, претерпеть сначала значительное сжатие благодаря движениям какого-то одного масштаба, а затем была сжата дополнительно вместе с другими порциями газа благодаря движению некоторого большего масштаба, в котором она также с самого начала участвовала. При этом повторное сжатие могло осуществляться не эволюционно, а благодаря сверхзвуковым столкновениям облаков.

Характер столкновения существенно зависит от состояния вещества, вовлеченного в этот процесс. Физические явления при столкновении двух облаков должны значительно отличаться от тех более спокойных и постепенных изменений в состоянии движений, которые привели к обособлению и выделению самих этих облаков из метагалактической среды. Если выделение сгущений-облаков происходит эволюционным путем, подобным «опрокидыванию» звуковых волн (с тем, однако, отличием, что ведущим процессом является не гидродинамическая нелинейность, а гравитационная неустойчивость), то столкновение обособленных газовых масс является, как мы говорили в § 5.2, процессом неэволюционным в том смысле, что газ на поверхности одного из сталкивающихся облаков испытывает внезапное внешнее воздействие со стороны другого облака, налетающего на него.

Относительная скорость двух сталкивающихся облаков не меньше скоростей тех движений, которые породили эти облака. Отсюда следует, что скорость столкновения больше скорости звука в материале облаков, так как последняя не может превосходить скорость гидродинами-

ческих движений, которые сжали и нагрели материал облака. Сверхзвуковые столкновения газовых масс должны приводить к гидродинамическим разрывам, но они не дают непосредственно ударных волн, т. е. таких разрывов, на которых выполнены условия сохранения трех потоков (§ 5.2). Эти условия не могут быть сразу выполнены, так как физические состояния вещества по обе стороны поверхности разрыва в общем случае совершенно различны и полностью независимы. Разрывы такого происхождения не похожи на разрывы, образующиеся при эволюции звуковой волны, в которых с самого начала выполнены условия сохранения потоков массы, импульса и энергии.

Неэволюционные разрывы, очевидно, не могут быть стационарными; со временем они распадаются на некоторые «стандартные» разрывы, удовлетворяющие необходимым условиям сохранения. При этом в каждом случае формируется набор стандартных разрывов, обеспечивающий, так сказать, согласование состояний среды по обе стороны поверхности разрыва. Эти состояния характеризуются четырьмя величинами с каждой стороны разрыва: давлением, плотностью и двумя компонентами скорости: p_1, ρ_1, v_1, v_{1t} ; p_2, ρ_2, v_2, v_{2t} . Под скоростью v здесь нужно понимать (как и в § 5.2) нормальную, а под скоростью v_t — тангенциальную к поверхности разрыва компоненту скорости. Между этими величинами в общем случае не существует никаких связей, они вполне произвольны. так как при столкновении приходят в соприкосновение две массы газа, сжатые до различных плотностей и давлений и имеющие различные нормальные и тангенциальные к поверхности соприкосновения скорости. Поэтому для «стандартного» перехода от состояния «1» к состоянию «2» необходимо располагать четырьмя произвольными параметрами, что и определяет число стандартных разрывов, образующихся после распада неэволюционного разрыва.

Ударная волна характеризуется одним параметром — скоростью распространения (или, что то же, скоростью натекания на нее газа). Другой вид стандартного разрыва — контактный разрыв, при котором давление газа по обе стороны поверхности разрыва одинаково, скорости равны нулю, а плотности различны. В нерелятивистском газе

это соответствует скачку температуры и плотности, а с ними и энтропии газа. Такой разрыв представляет собой сильное возмущение энтропийного типа. Легко проверить с помощью формул § 5.2, что все необходимые условия сохранения на поверхности контактного разрыва, как мы его определили, выполняются. Как и ударная волна, такой разрыв характеризуется одним параметром, которым является, например, скачок плотности.

Третий вид стандартного разрыва — тангенциальный разрыв. Это контактный разрыв, но с той разницей, что помимо разрыва плотности в нем имеется еще разрыв тангенциальной компоненты скорости. Разрыв касательной, тангенциальной компоненты скорости невозможен в ударной волне; но если нормальная компонента скорости равна нулю и отсутствует скачок давления, то, как легко проверить, разрыв тангенциальной скорости допускается условиями сохранения трех потоков. Тангенциальный разрыв характеризуется двумя параметрами — скачками плотности и тангенциальной скорости. Таким образом, минимальный набор стандартных разрывов, на которые может распасться произвольный неэволюционный разрыв рассматриваемой природы, содержит две ударные волны и один тангенциальный разрыв (Ландау и Лифшиц (1953)): ударные волны движутся друг от друга по разные стороны первоначального разрыва, а на месте последнего остается тангенциальный разрыв.

Этот набор и должен на самом деле реализоваться при сверхзвуковом столкновении больших газовых масс. Тангенциальный разрыв во всяком случае необходим, если в начальном неэволюционном разрыве испытывали скачок тангенциальные скорости. Последний всегда возникает, если столкновение облаков не строго центральное (ср. § 5.2). Так как ни в ударной волне, ни в контактном разрыве тангенциальные скорости не меняются, то их скачок может происходить «стандартно» только на тангенциальном разрыве, формирующемся на месте первоначального разрыва. Тангенциальный разрыв неподвижен относительно газа; через него нет потока массы, что является следствием одинаковости давления по обе его стороны (рис. 18).

Весьма важным является то обстоятельство, что тангенциальный разрыв абсолютно неустойчив: любые малые

возмущения в скользящих один по другому слоях жидкости неограниченно возрастают. Природу этой неустойчивости легко понять, если представить себе, что в каком-то месте поверхность тангенциального разрыва искривилась

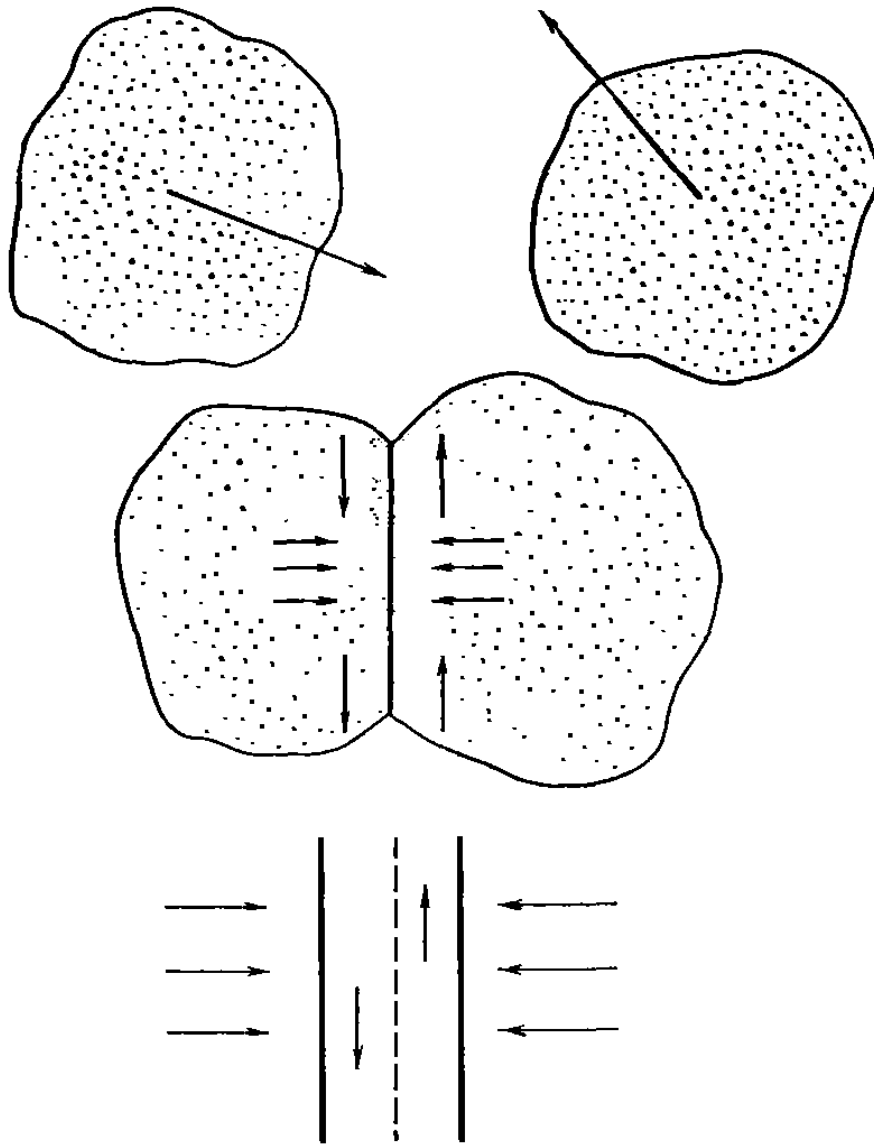


Рис. 18. Сверхзвуковое столкновение газовых масс. Возникновение неэволюционного разрыва и его распад на две ударные волны (сплошные линии) и тангенциальный разрыв (пунктир).

Можно сказать, что при этом изменились поперечные сечения потоков слева и справа от разрыва: одно сечение стало меньше, а другое больше. В таком случае, очевидно, должны в той же мере измениться и скорости потоков, если считать, что через изменившиеся сечения проходят те же порции газа. Такое допущение естественно, и из него следует, что в слое, где сечение стало меньше, скорость возрастет, а по другую сторону — уменьшится. Но стационарность потоков предполагает определенную

связь между давлением и скоростью, которая уже однажды была выяснена нами для одного простого случая. В § 5.2 мы видели, что в стационарном течении

$$p = B - Av, \quad v = \frac{A}{\rho}, \quad Av > 0,$$

т. е. давление тем меньше, чем больше скорость, и наоборот. (Это общее следствие теоремы Бернулли в гидродинамике.) Отсюда видно, что перед выпуклой частью поверхности тангенциального разрыва давление падает, а за ней возрастает. Но это должно, очевидно, вести к дальнейшему увеличению искривленности поверхности, откуда и следует, что тангенциальный разрыв неустойчив относительно малых возмущений такого рода.

Легко видеть, что характерное время усиления возмущений близко к отношению размера области возмущения вдоль поверхности к разности тангенциальных скоростей¹ слева и справа от нее (Ландау и Лифшиц (1953)). Граничная поверхность в высшей степени неустойчива, скорость нарастания возмущения с достаточно короткой длиной волны может быть сколь угодно велика.

Неустойчивость тангенциального разрыва была известна еще Гельмгольцу, который заметил, что границы струй, вытекающих из духовых труб, закручиваются в виде периодических спиралей. Граница струи как раз и представляет собой тангенциальный разрыв, отделяющий движущийся воздух от покоящегося. Хорошо известное явление — возникновение волн на воде при ветре — также обязано неустойчивости тангенциального разрыва. Разрыв находится на поверхности воды, вдоль которой движется воздух. В этом случае, правда, требуется учесть еще поверхностное натяжение воды и гравитационное поле. Это приводит к тому, что неустойчивость начинается при скорости, большей некоторой критической *).

Развитие неустойчивости в зоне тангенциального разрыва приводит к возбуждению хаотических движений

*) Эта скорость, вычислявшаяся теоретически еще Кельвином и Рэлеем, оказывается, однако, в несколько раз больше наблюдаемой, и убедительное объяснение расхождения до сих пор не найдено. Пример ветровых волн на воде лишней раз демонстрирует сложность явления гидродинамической неустойчивости в тангенциальном разрыве. Возможно, здесь существенны резонансные эффекты (Кадомцев и Конторович (1974)).

различных интенсивностей и масштабов, черпающих энергию из первоначального относительного движения слоев жидкости. В результате тангенциальный разрыв размывается в турбулентный слой (Ландау и Лифшиц (1953)). Характерной чертой турбулентности такого происхождения является наличие значительной вихревой компоненты в общем хаотическом движении. На это указывают в первую очередь многочисленные наблюдения и эксперименты. Однако полной и последовательной количественной теории даже для самых простых случаев до сих пор не построено. Здесь возникают серьезные трудности того же рода, что и при анализе других нелинейных гидродинамических явлений, о которых мы уже упоминали. Можно лишь предполагать, основываясь на общих соображениях и экспериментальных результатах, что при благоприятных условиях в области масштабов, удовлетворяющих условию (5.6.2), способно возникнуть локально изотропное турбулентное состояние с универсальным спектром (см. первый раздел этого параграфа). Под благоприятными следует, по-видимому, понимать условия, при которых в достаточно широкой зоне время развития неустойчивости и сравнимое с ним время нелинейного взаимодействия движений различных масштабов меньше гидродинамического времени всего процесса возникновения и поддержания первоначального разрыва.

Возвращаясь от общих гидродинамических соображений к картине распада неэволюционного разрыва, возникающего при сверхзвуковом столкновении газовых масс в метагалактической среде, мы можем теперь констатировать, что результатом столкновения является образование в зоне контакта облаков слоя уплотненного и разогретого газа, ограниченного расходящимися ударными волнами. В пределах этого слоя возникает вторичная турбулентность, содержащая значительную вихревую компоненту. Энергией для возбуждения и поддержания турбулентности служит тангенциальная компонента первоначального движения облаков. Энергия ударных волн обязана нормальной компоненте исходного относительного движения. Столкновения хаотически движущихся облаков чаще всего должны происходить так, что обе эти компоненты соизмеримы, а сильные неравенства можно ожидать только в вырожденных случаях. Поэтому для су-

щественной доли актов столкновений будет иметь место, вероятно, также и соизмеримость энергий ударных волн и турбулентности в зоне столкновения. При этом турбулентность должна развиваться в дозвуковом режиме.

Рассмотренные в § 5.4 механизмы генерации вторичных вихрей на ударных волнах могут действовать и в пределах слоя, возникающего при столкновении облаков. Например, звуковые возбуждения, отрывающиеся от тангенциального разрыва, могут достигать задних фронтов ударных волн и рассеиваться на них. При этом звук не выходит за пределы этого слоя, так как ударные волны движутся по внешней среде быстрее звука. Догоняющий ударную волну звук может только отражаться, возбуждая две другие—вихревую и энтропийную—моды малых возмущений. Вихри возникают также из-за небаротропии при распространении по неоднородному слою газа возмущений, порожденных распадом тангенциального разрыва; последние являются сначала потенциальными, так как раскачка неустойчивости в разрыве обязана потенциальной силе—силе давления (Зильбергейт и Чернин (1978)).

Турбулентный слой мало похож на сгущение эволюционного происхождения. Он обладает иным распределением температуры, плотности и скорости; но важнее наличие в нем значительных собственных вихревых движений, порожденных тангенциальным разрывом. Эти движения в силу общих гидродинамических закономерностей распространяются на все масштабы, вплоть до самых малых, где действует вязкость, стремясь образовать каскад вихрей того типа, который осуществляется в колмогоровской турбулентности. Характерная скорость этих движений меньше скорости распространения ударных фронтов по натекающему на них веществу. Различие, однако, не может быть очень большим (за исключением почти центрального столкновения, что должно считаться редким событием); это означает, что скорости вихревых движений в основном их масштабе по порядку величины сравнимы с начальной скоростью относительного движения облаков. Последняя же имеет порядок хаббловской скорости на разделяющем облака (точнее, их центры инерции) исходном расстоянии (5.5.1).

Турбулентность такого происхождения охватывает не всю метагалактическую среду (как это предполагалось

в картине Вейцеккера), а сосредоточена в отдельных слоях различной массы, плотности и температуры. Она зарождается в дозвуковом режиме, но движения определенных масштабов могут перейти затем в сверхзвуковой режим вследствие охлаждения газа и снижения скорости звука в нем. Возникающие благодаря этому ударные волны следующего поколения сжимают отдельные массы газа вместе с «вмороженными» в них вихревыми движениями. Поэтому турбулентные слои могут послужить гнездами фрагментации, в которых формируются протогалактические сгущения вещества, обладающие собственными движениями и в первую очередь вращением. Сами же турбулентные слои превращаются со временем в скопления галактик; при этом бурные движения препятствуют общей сферизации на газовой фазе и определяют хаотичность и иррегулярность распределения галактик в скоплениях.

Такова в общих чертах возможная картина турбулентной космогонии, сочетающая привлекательные идеи Вейцеккера (1951) и Гамова (1952), результаты Озерного и Чернина (1967, 1968) с общей теорией эволюции догалактической структуры в изотропной горячей Вселенной. Представление о вторичных вихрях позволяет объяснить происхождение сильно возбужденного, завихренного состояния метагалактической среды на активной фазе космогонического процесса, временные границы которой определяются комбинацией наблюдательных данных и анализа физических процессов взаимодействия вещества и излучения (§§ 5.1, 5.3). Дальнейшее развитие картины требует прежде всего выяснения количественных закономерностей вторичной турбулентности, ее характерных масштабов и скоростей. В качественном рассмотрении происхождения протогалактических сгущений никакие масштабы или массы не предполагались специально выделенными с самого начала. Не делалось и выбора между двумя типами исходных безвихревых возмущений. Требовалось лишь, чтобы в указанную эпоху соответствующие движения стали сильными. Предпочтительные масштабы оказываются, однако, выделенными физическими условиями в метагалактической среде.

В § 5.3 мы показали, что предпочтительные масштабы сильных движений определяются конкуренцией тепловых и

гидродинамических процессов. Эти результаты непосредственно применимы к картине формирования неправильных скоплений при столкновениях больших газовых масс. Единственное изменение, которое при этом нужно внести в соответствующие формулы и оценки, касается плотности вещества в турбулентном слое. Если до столкновения газовые массы были уплотнены ударными волнами (адиабатическими), то средняя начальная плотность слоя будет приблизительно равна учетверенной плотности облаков, которая в свою очередь приблизительно в четыре раза больше космологической плотности в соответствующую эпоху:

$$16 \rho(z) \approx 8 \cdot 10^{-29} \Omega (1+z)^3 \text{ г/см}^3. \quad (5.6.9)$$

Эта плотность меньше типичной плотности скоплений $\sim 10^{-26} - 10^{-27} \text{ г/см}^3$, если

$$1+z < (2,5 \div 5) \Omega^{-1/3}. \quad (5.6.10)$$

Очевидно, что соответствующие изменения в формулах § 5.3 сводятся формально к замене

$$\Omega \rightarrow 4\Omega.$$

При этом порядковые оценки меняются не слишком сильно. Принимая, что столкновения газовых масс, порождающие турбулентные слои, происходили, скажем, при $z = 8 \div 3$, найдем, с помощью соотношения (5.3.29), что наименьшая масса скопления, в котором горячий газ мог бы сохраниться до современной эпохи, составляет

$$M_{\text{gas}} \approx 5 \cdot 10^{12} \div 10^{15} M_{\odot}. \quad (5.6.11)$$

Эта масса отвечает равенству времени охлаждения газа современному возрасту Вселенной. Равенство же времени охлаждения возрасту Вселенной в эпоху образования ударных волн определяет, согласно соотношению (5.3.25), нижний предел массы в адиабатических ударных волнах, а с ним и нижний предел массы протоскоплений. Принимая снова, что скопления начали формироваться при $z = 8 \div 3$, найдем

$$M_{\text{ff}} \approx (1 \div 2) \cdot 10^{13} M_{\odot}. \quad (5.6.12)$$

Получающиеся таким образом характерные массы

скоплений галактик оказываются, как мы видим, близкими к реальным массам этих систем.

Охлаждение основной массы газа в турбулентном протоскоплении становится существенным не ранее, чем через несколько характерных времен исходного гидродинамического движения. Поэтому гидродинамическая неустойчивость тангенциального разрыва успевает развиться еще в горячем газе и, значит, в дозвуковом режиме. Дальнейшее постепенное снижение температуры переводит вторичные движения, обязанные распаду тангенциального разрыва, в сверхзвуковой режим во всех масштабах, где их скорость выше, чем $\approx 10^6$ см/сек (последняя величина соответствует скорости звука в газе, охлажденном до предельной температуры $\approx 10^4$ °К). Этим определяется нижняя граница масс фрагментов-протогалактик, которую, с учетом необходимых изменений, можно оценить по соотношению (5.3.23), принимая также во внимание (5.6.10):

$$M_G \geq (3 \cdot 10^9 \div 10^{10}) \Omega^{3/2} M_{\odot}. \quad (5.6.13)$$

Нужно, по-видимому, относить этот нижний предел прежде всего к спиральным галактикам, так как именно они и должны, скорее всего, образовываться в завихренной турбулентной среде. Интересно, что согласно Холмбергу (1969), эмпирическая функция распределения спиральных галактик по массам действительно резко падает при массах $\approx 10^9 M_{\odot}$ и обрывается при $M_G \approx 2 \cdot 10^8 M_{\odot}$ (рис. 19). Качественные соображения о фрагментации турбулентного слоя и количественная оценка (5.6.13) находятся в хорошем согласии с этим фактом и могут рассматриваться как возможное его объяснение.

С функцией масс Холмберга согласуется и оценка (5.6.12), которую можно понимать не только как нижний предел масс для групп и скоплений, но и как верхний предел масс для галактик.

Положение максимума на функции масс для спиральных галактик сдвинуто в сторону меньших масс, так что основная доля галактик имеет массы $\sim 10^9 - 10^{10} M_{\odot}$, не слишком отличающиеся от нижнего предела. Это обстоятельство могло бы быть следствием спектра движений в турбулентном слое. Действительно, если вихревые скорости падают с масштабом не быстрее, чем по линейному

закону,

$$v \propto \lambda^n, \quad n < 1,$$

(как в случае универсальных спектров; см. первый раздел настоящего параграфа), то необходимое условие формирования ударных волн, $t_h(\lambda) \approx \lambda/v(\lambda) \leq t$, легче всего

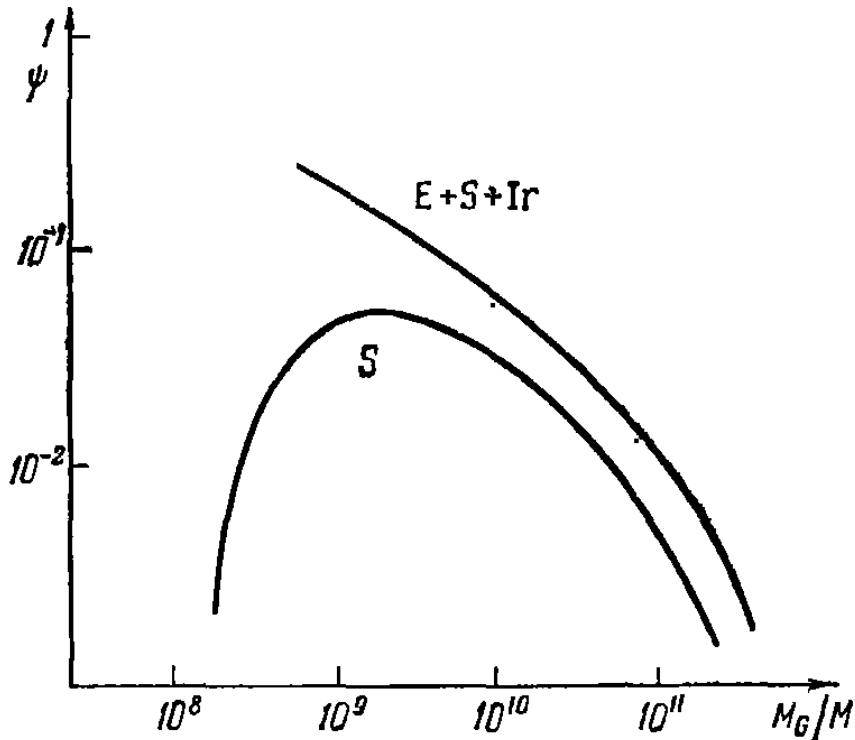


Рис. 19. Функция масс по Холмбергу (1969). ψ — отношение числа галактик с массой между M и $M + dM$ к полному числу галактик в данном объеме пространства.

выполнялось в самых малых масштабах движения, и нижний предел масс галактик в (5.6.13) должен поэтому находиться не очень далеко от масс наиболее распространенных галактик.

Наконец важной величиной является скорость вращения фрагментов-протогалактик. Вихревые движения турбулентного слоя оказываются замороженными в обособляющиеся массы газа, и соответствующий им вращательный момент сохраняется в каждой протогалактике. Далее протогалактики сжимаются под действием собственной гравитации, пока не достигнут типичной плотности галактик, $\rho_g \approx 10^{-24} \text{ г/см}^3$. По мере сжатия протогалактика вращается все быстрее; в конечном состоянии, если момент достаточно велик, устанавливается равновесие центробежной и гравитационной сил, что и наблюдается в плоских подсистемах гигантских спиральных галактик, подобных нашей Галактике.

Оценку итоговой скорости вращения v_G можно сделать, приняв, что при образовании сгущения-протогалактики вихревая скорость v в данном масштабе λ была порядка λ/t (и больше), как это требуется для образования ударной волны. Сохранение момента

$$M_G v \lambda = \text{const}$$

обеспечивает тогда при плотности ρ_G скорость (как минимум)

$$v_G \approx 10^8 \Omega^{-1/6} (1+z)^{-1/2} \left(\frac{M_G}{10^{11} M_\odot} \right)^{1/2} \text{ см/сек}, \quad (5.6.14)$$

которая в интервале красных смещений $10 > z > 2$ действительно близка к типичной скорости вращения спиральных галактик.

Итак, существенные характеристики галактик и скоплений появляются в картине протогалактических вихрей как следствие физических условий в метagalacticкой среде на активной фазе космогонического процесса. Соотношения для этих величин (5.6.11)—(5.6.14) не содержат произвольных параметров, связанных с «затравочными» возмущениями, а типичные массы и скорости вращения галактик в скоплении мало зависят от параметров скопления.

Полная картина космогонического процесса включает в себя образования, подобные «блинам», и турбулентные слои как предельные случаи в сложном многообразии нелинейных гидродинамических явлений в сильно возбужденной метagalacticкой среде на активной фазе космогонического процесса. Относительную частоту событий того и другого рода *) в реальных физических условиях довольно трудно оценить теоретически. Теория процессов такого рода в общей гидродинамике еще очень далека от полноты и количественной определенности. По этой причине остаются пока не объясненными такие важные характеристики крупномасштабной космической структуры, как доля вещества, вошедшая в галактики, распределение галактик по морфологическим типам и т. п.

*) Стоит заметить, что столкновения больших масс, сравнимых с массами скоплений, автоматически появлялись как пересечения «блинов» при численном моделировании формирования «блинов» (Дорошкевич и др. (1976)).

«...Теория вихрей чревата множеством трудностей». Эти слова, сказанные Ньютоном по поводу космогонии Декарта, можно вполне отнести и к обсуждаемой здесь современной теории космических вихрей. Трудности, по-видимому, не имеют все же характера внутренних противоречий и потому можно надеяться, что они не окажутся непреодолимыми.

§ 5.7. Скопления галактик: распад или скрытые массы?

Скопления галактик обнаруживают одну замечательную особенность, о которой мы упоминали в первом параграфе книги; для многих из них масса, определенная по скоростям собственного движения галактик в скоплении, оказывается заметно больше массы, определенной по общей светимости галактик. Первый способ нахождения массы основан на предположении, что скопления представляют собою стационарные гравитационно-связанные системы. Полная механическая энергия каждой такой системы должна быть отрицательной, причем кинетическая энергия должна составлять (в среднем по времени) половину модуля гравитационной потенциальной энергии:

$$E_k = \frac{1}{2} |E_G|. \quad (5.7.1)$$

Это соотношение представляет собой так называемую вириальную теорему классической механики. Она получается как прямое следствие уравнений движения при общей стационарности системы. Например, для частицы массы M_1 , обращающейся по стационарной круговой орбите вокруг центральной массы $M \gg M_1$, скорость движения v удовлетворяет условию равенства (по модулю) центробежной и гравитационной сил:

$$\frac{v^2}{R} = \frac{GM}{R^2},$$

где R — радиус орбиты. Отсюда видно, что кинетическая энергия частицы $\frac{1}{2}M_1v^2$ составляет половину модуля ее гравитационной потенциальной энергии GM_1M/R , что и доказывает соотношение (5.7.1) для такой системы из двух

частиц. Легко провести доказательство и для любой системы произвольного числа частиц (см. § 6.1). Соотношение

$$v^2 \approx \frac{GM}{R} \quad (5.7.2)$$

для гравитационно-связанной системы следует, по существу, просто из соображений размерности, если понимать под v среднюю квадратичную скорость частиц в системе, называемую дисперсией скоростей, а под M и R — ее полную массу и характерный размер.

Если размер скопления и дисперсия скоростей галактик в нем известны, то из (5.7.2) следует оценка массы скопления:

$$M_{VT} \approx \frac{v^2 R}{G}. \quad (5.7.3)$$

Определенную таким способом массу называют вириальной.

Другой способ определения массы состоит в том, что полную наблюдаемую светимость скопления умножают на принимаемое за стандартное отношение масса/светимость, найденное независимо для отдельных галактик. Это последнее отношение различно для галактик различных типов, но, если известно, что в данном скоплении преобладают галактики какого-то определенного типа, то суммарную массу этих галактик M_L можно таким способом действительно оценить. И вот оказывается, что суммарная масса галактик почти всегда меньше вириальной массы скопления:

$$M_{VT} > M_L. \quad (5.7.4)$$

Это обстоятельство (его называют вириальным парадоксом) было установлено в 30-е годы Цвикки, который впервые применил вириальную теорему к скоплениям галактик. Результаты новых детальных исследований наблюдательных данных по этой проблеме имеются в книге Воронцова-Вельяминова (1972), в работах Эйбла (1961), Рудницкого (1964), Караченцева (1965, 1966, 1968), Караченцевой (1967), Караченцева и Терехина (1968), Руда (1974), Руда и др. (1972), Фесенко (1976а, б).

Вириальный парадокс исчезает, если скопления не являются стационарными системами, так что к ним неприменима вириальная теорема. Однако если они все же стационарны, то из него следует наличие в скоплениях значительных масс темного, не светящегося вещества помимо вещества самих галактик, причем эта, как говорят, «скрытая» масса должна быть в 3—10 раз больше общей массы галактик в скоплениях. Имеется ряд аргументов в пользу этой второй возможности (мы рассмотрим их ниже), но и первая внимательно изучается в последние годы.

Если скопления нестационарны и никакой другой массы, сравнимой с массой галактик, в них нет, то неравенство (5.7.4) должно означать, что механическая энергия скоплений положительна. В этом случае движения галактик в скоплениях характеризуют не его полную массу, а степень нестационарности системы. Такая точка зрения была высказана Амбарцумяном (1960, 1972); она послужила одним из центральных пунктов космогонической гипотезы, развиваемой в Бюраканской астрофизической обсерватории.

Амбарцумян предположил, что скопления галактик возникают в результате «взрывов» плотных тел; каждое такое тело способно делиться на разлетающиеся фрагменты, которые превращаются затем в галактики. Энергия «взрыва» сообщает галактикам большие гиперболические скорости, так что механическая энергия системы положительна. Образование звезд происходит в результате дальнейшего дробления фрагментов; но в центре галактики может остаться какая-то часть вещества в первоначальной форме. Дозвездное тело в ядре галактики может испытывать гигантские взрывы, сопровождающиеся выбросами значительных масс вещества, интенсивным радиоизлучением и т. д.

Природа плотных дозвездных тел не конкретизируется в этой космогонической концепции. Делается лишь предположение, что эти тела не обязательно должны удовлетворять известным сейчас законам физики; в частности, при их взрывах не обязательно выполнение закона сохранения энергии в той форме, в которой этот закон обычно формулируется. Закон сохранения момента также не обязателен для них. Допускается и нарушение закона сохранения числа барионов.

Гипотезы, основанные на столь радикальных допущениях *), трудно развивать теоретически. Однако с наблюдательной точки зрения представление об активности ядер галактик оказалось важным и плодотворным. Разнообразные явления активности ядер были действительно недавно обнаружены. Наиболее важный вклад в эти исследования внесли бюраканские астрономы (Амбарцумян (1972)).

Роль явлений такого рода в общей эволюции галактик весьма существенна. Последовательной их теории еще нет, возникающие здесь задачи очень сложны; но все же с нашей точки зрения это не может служить основанием для выхода за рамки «обычной» физики. Богатство, мощь и универсальность идей современной физики применительно к фундаментальным проблемам космологии и космогонии убедительно демонстрируются классическими результатами Эйнштейна, Фридмана, Лифшица, Гамова, о которых мы говорим в нашей книге.

Если скопления стационарны и не распадаются, то вириальный парадокс нужно рассматривать как указание на присутствие в группах и скоплениях галактик значительных невидимых, темных масс.

Такая гипотеза подкрепляется в первую очередь следующим соображением. Если бы скопления распадались, то их возраст (или время распада), оцениваемый как размер, деленный на скорость движения галактик, оказался бы меньше возраста галактик. Согласно средним характеристикам систем галактик, указываемым Караченцевым (1968), возраст распадающихся групп и скоплений составил бы от одного до двух миллиардов лет, тогда как обычно определяемый по звездному составу (и на основе теории звездной эволюции) возраст галактик примерно в десять раз больше (о возрасте звезд и галактик см. книги Шварцшильда (1961), Каплана (1970), Шкловского (1975)). Даже геологический возраст Земли, составляющий приблизительно 5 миллиардов лет, больше возраста распадающихся систем. Правильная структура скоплений, например, скопления Кома, служит еще одним аргументом:

*) Джинс (1929), обсуждая проблему спиральных рукавов, заметил, что неудача в попытке объяснить их происхождение наводит на подозрение, что центры галактик — это сингулярные точки, через которые вещество, может быть, проникает в нашу Вселенную из каких-то других миров.

сфероидальная форма и регулярное распределение плотности свойственны обычно стационарным гравитационно-связанным системам.

Детальные наблюдательные исследования действительно установили в последние годы наличие в скоплениях галактик ранее не учитываемого (при оценке по светимости) вещества. Де Вокулер (1969), Арп и Бертола (1969), Корменди и Бакалл (1974) обнаружили вокруг ряда массивных галактик слабые ореолы, простирающиеся на расстояния до половины мегапарсека и более. При столь большой протяженности даже слабая поверхностная яркость обеспечивает в ряде случаев удвоение общей светимости галактики. Можно предполагать, что по крайней мере в два раза должна увеличиться в связи с этим и оценка общей массы таких галактик. Массивные галактики с ореолами обнаружены в центральных областях некоторых скоплений галактик, в частности, в центре Комы.

Мы уже упоминали о том, что исследование излучения скоплений галактик в рентгеновской области показало, что многие из них (и в первую очередь снова Кома) являются протяженными рентгеновскими источниками в энергетической области 2—8 кэв; это излучение интерпретируется как результат свободно-свободных переходов в ионизованном горячем (до 10^8 °К) газе *). Судя по интенсивности потока, масса газа сравнима с массой галактик в скоплении, определяемой по их светимости (Филд (1974)).

Открытие межгалактического газа в скоплениях вместе с увеличением оценки масс гигантских галактик по наблюдениям ореолов сокращает разрыв между вириальной массой скоплений и суммой всех непосредственно регистрируемых масс в скоплениях. Возьмем, например, наиболее хорошо изученное скопление Кома. Его параметры (Руд и др. (1972)), приведенные к значению фактора Хаббла $H = 55$ км/сек·Мпс, таковы: вириальная масса $M_{VT} = 5 \cdot 10^{15} M_{\odot}$, средняя скорость галактик относительно центра скопления $v \approx 1000$ км/сек, полная светимость $L = 3 \cdot 10^{13} L_{\odot}$. Если принять отношение масса/светимость $\sim 30 M_{\odot}/L_{\odot}$, то масса галактик составит $M_L = 9 \cdot 10^{14} M_{\odot}$,

*) Ранее вопрос о природе межгалактического газа и его тепловом режиме рассматривался в работах Гинзбурга (1964), Гинзбурга и Озерного (1965), Вайнштейна и Сюняева (1966), Курта и Сюняева (1967, 1970).

т. е. примерно двадцать процентов от вириальной массы. С учетом горячего газа доля обнаруживаемой массы возрастает до тридцати-сорока процентов. Отношение $M/L = 30M_{\odot}/L_{\odot}$ считается типичным для эллиптических галактик, которые преобладают в Комае *). Для массивных галактик с ореолами это отношение доходит, возможно, до $70 M_{\odot}/L_{\odot}$ (Брандт и Рузен (1969)).

Недавние исследования динамики галактик в парах и группах, а также внутренней динамики галактик (Эйнасто и др. (1974а, б, в; 1975, 1976), Острайкер и др. (1974)) указывают на то, что отдельные массивные галактики или группы массивных галактик обладают протяженными коронами, создающими дополнительное поле тяготения. Судя по величине этого поля (в предположении, что группы и скопления стационарны), масса корон может быть в 3—10 раз больше масс самих галактик. Связь корон, обнаруживаемых по динамике, со слабыми протяженными ореолами, обнаруживаемыми по оптическому излучению, остается пока не вполне ясной, но размеры тех и других, по-видимому, сравнимы. Не исключено, что масса, приписываемая коронам, дополняет массы скоплений до вириальных.

Можно предполагать, что «скрытая» масса представляет собой совокупность большого числа звезд низкой светимости. Другой вероятный кандидат на роль скрытых масс — ионизованный газ умеренной температуры — создавал бы, если бы он имелся в скоплениях в нужных количествах, излучение в ультрафиолетовом, оптическом и радиодиапазонах. Из наблюдений известны верхние пределы на интенсивность такого излучения. Основанные на них оценки ограничивают массу газа примерно десятью-двадцатью процентами вириальной массы (Филд (1972, 1974), Тартер и Силк (1974), Чернин и др. (1976), Комберг и Новиков (1975)). Другие возможности (черные дыры, пыль, плотные облака нейтрального водорода), как можно предполагать, менее вероятны (Тартер и Силк (1974), Комберг (1974)).

Если короны действительно существуют и не сводятся к слабым ореолам вокруг галактик, то в чем, помимо ди-

*) Детальный сравнительный анализ различных методов определения масс галактик и скоплений дан Эйнасто (1972), Эйнасто и др. (1976).

намики, это могло бы еще проявляться? Такой вопрос исследовался в работе Чернина и др. (1976) в рамках газодинамической модели, предполагающей, что как видимая, так и скрытая структура в системах галактик образовались из большой конденсации догалактического вещества. Уплотнения различного масштаба внутри облака могли бы образовывать галактики — от гигантов до карликов, тогда как основная часть материала оставалась еще диффузной, создавая газовую протокорону системы галактик. Массивные галактики должны, вероятно, занимать центральную область такой системы, а карлики могли бы населять весь ее объем.

Карлики, движущиеся сквозь газ протокороны, могут сохранить свой собственный межзвездный газ, если гравитационная связь газа со звездами сильнее, чем встречный «ветер» коронарного газа. «Ветер» преобладает над гравитационной связью вблизи центра системы, где плотность коронарного газа больше. В этой области карлики не могут сохранить свой межзвездный газ и быстро теряют его. Содержание газа — важная характеристика галактик, коррелирующая с морфологическим типом: спиральные и неправильные галактики богаты газом, а эллиптические бедны. После выдувания газа в карликовых спиральных и неправильных галактиках остается только сферический остов старых звезд, подобный сферической составляющей нашей Галактики, что и делает их морфологически неотличимыми от эллиптических галактик. В этом смысле газодинамическое действие протокороны может создавать морфологический эффект в системах галактик: карлики эллиптические должны преобладать во внутренней области систем, а неэллиптические, спирали и неправильные, — во внешней.

Такой эффект действительно был обнаружен в результате обработки большого наблюдательного материала по карликовым галактикам в близких системах галактик (Эйнасто и др. (1974б)). Ранее сходный эффект отмечался Караченцевой (1973).

Эта закономерность демонстрирует повышенную степень организации и регулярности строения групп галактик, где она обнаруживается; такие группы, представляющие собой особый тип метагалактических популяций, были названы гипергалактиками.

С космогонической точки зрения «скрытые» массы и короны ставят целый ряд вопросов, еще не нашедших обоснованного ответа. Применительно к газодинамической модели, о которой шла речь выше, это в первую очередь вопрос о дальнейшей судьбе газа протокороны. Не ясно, как большая его часть, да еще при малой средней плотности, могла бы превратиться в звезды низкой светимости. Этот вопрос возникает, впрочем, и вне связи с конкретными моделями, если только такие звезды действительно составляют основную долю «скрытой» массы.

В 1959 г. Цвикки писал, что скопление Кома, а с ним и все другие скопления галактик, для которых известна дисперсия скоростей, оказываются в несколько сотен раз массивнее, чем об этом можно было бы судить по оценкам масс самых ярких галактик в скоплениях, выводимых из их светимостей. За прошедшие с тех пор годы представление о гравитационной связанности и стационарности скоплений получило новые наблюдательные и теоретические подтверждения, на основании которых с большей уверенностью можно полагать, что скопления действительно имеют те массы, которые следуют из теоремы вириала. Но разрыв между вириальной и полной обнаруженной массой в Кома, как мы видим, оценивается сейчас уже не в два порядка *). Это достигнуто уточнением отношения масса/светимость, для которого принято Цвикки значение (~ 1) оказалось сильно заниженным. Возможно и снижение оценки вириальной массы Кома. По мнению Фесенко (1976б), это может быть связано с исключением галактик фона, которые случайно проектируются на скопление и создают видимость большей, чем на самом деле, дисперсии скоростей галактик. Другая возможность предполагает учет группирования галактик внутри скопления в отдельные подскопления, что способно увеличить при той же массе общий гравитационный потенциал скопления. Связанная со всем этим неопределенность оценивается множителем 2 или 3; с учетом сказанного выше это может означать, что вириальный парадокс для Кома теперь,

*) В оценках Цвикки фактор Хаббла принимался равным $550 \text{ км/сек} \cdot \text{Мпс}$, что раз в десять больше действительного значения. Но, как легко видеть (см., например, Тартер и Силк (1974)), отношение вириальной массы к светящей не зависит от фактора Хаббла.

после открытия горячего газа в скоплениях и ореолов, или корон вокруг массивных галактик, более не существует. Не может быть исключена, разумеется, и противоположная точка зрения, особенно когда речь идет о других скоплениях галактик, для которых ситуация гораздо менее ясна.

Уточнение массы галактик, оцениваемой по светимости, и открытие горячего межгалактического газа в скоплениях; обнаружение ореолов вокруг массивных галактик и пространственного разделения эллиптических и неэллиптических карликовых галактик в гипергалактиках вместе с новыми динамическими указаниями на присутствие дополнительных масс в системах галактик — все это приближает нас к решению проблемы, поставленной Цвикки в тридцатые годы, когда он впервые применил теорему вириала к скоплениям галактик.

{

ДИНАМИЧЕСКАЯ ЭВОЛЮЦИЯ ЗВЕЗДНЫХ СИСТЕМ



§ 6.1. От протогалактики к звездной системе

В истории каждой галактики была короткая, но полная событиями эпоха, когда ее вещество в виде облака газа, только что выделившегося из газового протоскопления, сжималось под действием собственной гравитации. В этом процессе рождались первые звезды, формировались подсистемы (такие, как плоская и сферическая подсистемы нашей Галактики), устанавливалось распределение звезд по скоростям. Облако газа превращалось в звездную систему, а звездная система достигала стационарного состояния. После этого изменения в звездной системе происходили гораздо медленнее и имели совсем иной характер.

Бурные события могли развиваться лишь в центральных областях галактик благодаря активным процессам в их ядрах. В спиральных галактиках продолжалось, кроме того, формирование молодых звезд из газа и пыли их плоских подсистем, возникал очерчиваемый этими звездами спиральный узор.

В настоящей главе мы рассмотрим некоторые из явлений, существенных для общей динамической эволюции звездных систем, а затем, в § 7.1, обсудим процесс образования звезд и формирование спирального узора. Природа активности галактических ядер и родственных им квазаров представляет собой одну из самых важных проблем современной астрофизики; она, однако, выходит за рамки нашей книги, так же как и ряд других интересных проблем физики галактик.

Сжатие газовой протогалактики. Обособление протогалактических сгущений происходит, по-видимому, благодаря гидродинамическим и тепловым процессам в газовом протоскоплении (§§ 5.5, 5.6). Дальнейшая эволюция

сгущений определяется прежде всего их собственной гравитацией.

Гравитация «включается», когда размер сгущения сравним с джинсовой длиной или превосходит ее. Для этого требуется, чтобы давление в облаке было не слишком велико (ср. § 3.1):

$$p \lesssim \rho \left(\frac{GM}{R} \right). \quad (6.1.1)$$

Здесь ρ — плотность сгущения, M , R — его масса и характерный размер. Этому условию отвечает ограничение на температуру облака:

$$T \lesssim \frac{GM}{k} \rho^{1/3} M^{2/3} \approx 5 \cdot 10^5 \left(\frac{\rho}{\rho_G} \right)^{1/3} \left(\frac{M}{10^{10} M_{\odot}} \right)^{2/3} \text{°К.} \quad (6.1.2)$$

(Здесь, как и ранее, k — постоянная Больцмана, m — масса атома водорода, $\rho_G \approx 10^{-24} \text{ г/см}^3$ — типичная плотность галактик.)

Гравитационные процессы, и прежде всего общее сжатие сгущения под действием собственной тяжести происходят быстрее тепловых процессов, если характерное гравитационное время,

$$t_G \approx (G\rho)^{-1/2} \approx 3 \cdot 10^{15} \left(\frac{\rho}{\rho_G} \right)^{-1/2} \text{ сек,} \quad (6.1.3)$$

меньше времени охлаждения. В состоянии, когда силы давления и гравитации в сгущении сравниваются (этому отвечает знак приближенного равенства в формулах (6.1.1) и (6.1.2)), время охлаждения посредством излучения при свободно-свободных переходах

$$\tau_{ff} \approx 2 \cdot 10^{14} \left(\frac{\rho}{\rho_G} \right)^{-5/6} \left(\frac{M}{10^{10} M_{\odot}} \right)^{1/3} \text{ сек.} \quad (6.1.4)$$

Время (6.1.4) больше гравитационного времени (6.1.3), если масса протогалактического сгущения не очень велика (ср. соотношения (5.5.10)—(5.5.12)):

$$M < M_1 \approx 3 \cdot 10^{13} \left(\frac{\rho}{\rho_G} \right) M_{\odot}. \quad (6.1.5)$$

Протогалактика с массой $10^{11} M_{\odot}$ удовлетворяет этому

условию *), если ее начальная плотность не меньше, чем $3 \cdot 10^{-27} \text{ г/см}^3$. Начиная с этой (или несколько большей) плотности, протогалактика такой массы сжимается свободно под действием собственной гравитации. Свободное сжатие претерпевают и протогалактики меньшей массы, в которых охлаждение происходит либо посредством свободно-свободных переходов, когда выполняется неравенство (6.1.6), либо посредством свободно-связанных переходов и высвечивания в линиях.

Охлаждение возможно, если температура не ниже $T_a \approx \approx 10^4 \text{ }^\circ\text{К}$. Последней отвечает джинсова масса

$$M_J(T_a) \approx 3 \cdot 10^7 \left(\frac{\rho}{\rho_G} \right)^{-1/2} M_\odot, \quad (6.1.7)$$

которая при не слишком малых плотностях (они во всяком случае не должны быть меньше средней плотности групп и скоплений $\approx 10^{-26} - 10^{-27} \text{ г/см}^3$) меньше типичных галактических масс.

Сгущения, массы которых велики и не удовлетворяют ограничению (6.1.5) (но удовлетворяют критерию Джинса), сначала медленно, — по мере охлаждения, — сжимаются в квазистационарном режиме при равновесных в каждый момент значениях температуры и плотности, а затем сжатие все равно переходит в свободное падение с быстрым охлаждением, так как массы всех галактик меньше коэффициента перед отношением плотностей в соотношении (6.1.5).

Итак, сжатие протогалактики до типичной галактической плотности происходит (с самого начала или начиная с некоторой плотности) в режиме свободного падения; гравитационная потенциальная энергия сгущения переходит при этом в кинетическую энергию его сжатия как целого и тепловую энергию газа. Тепло, однако, быстро отводится излучением, так что сжатие происходит почти изотермически с температурой около $10^4 \text{ }^\circ\text{К}$.

*) В этой оценке масса все же должна быть достаточно велика, чтобы равновесная температура (соотношение (6.1.2) со знаком равенства) превышала характерное для свободно-свободных переходов значение $5 \cdot 10^5 \text{ }^\circ\text{К}$:

$$M > M_2 \approx 10^{10} \left(\frac{\rho}{\rho_G} \right)^{-1/2} M_\odot. \quad (6.1.6)$$

Такое неравенство совместимо с (6.1.5) при плотности $\geq 5 \cdot 10^{-27} \text{ г/см}^3$.

Охлаждение снижает критическую джинсову массу *) до значения (6.1.7); по этой причине в сжимающемся газе возможно развитие гравитационной неустойчивости, способной привести к его дроблению на сравнительно малые массы. «Затравками» для гравитационной неустойчивости, приводящей к такой фрагментации, могли бы послужить уплотнения, созданные в газовой протогалактике тепловой неустойчивостью (той же природы, что и в § 5.3), которая развивалась при охлаждении газа благодаря тормозным процессам.

Среди фрагментов должны, по-видимому, преобладать сначала такие, масса которых $3 \cdot 10^7 - 10^8 M_{\odot}$ (ср. (6.1.7)), но затем и сами эти фрагменты, сжимаясь изотермически до плотностей, превосходящих среднюю плотность галактик, дробились на еще более мелкие фрагменты с массами звездных скоплений **) или отдельных звезд (см. § 7.1).

Прекращение общего сжатия и переход системы как целого к стационарному состоянию становится возможным, когда основная доля ее вещества распалась на малые фрагменты различной массы, которые уже не сталкиваются между собой контактно, а взаимодействуют только благодаря взаимному притяжению (ср. дробление протоскопления в § 5.5). Каждый из таких фрагментов продолжает «свободно падать» в общем гравитационном поле, создаваемом всеми остальными массами, но орбиты большинства фрагментов финитны, и для системы в целом выполняется вириальная теорема (ее доказательство см. ниже):

$$E_K = |E_G|/2,$$

где E_K , E_G — средние по времени значения кинетической и потенциальной энергии всех масс.

На фазе общего сжатия и фрагментации полная механическая энергия системы $E = E_K + E_G$ перераспреде-

*) Если (как в теории «блинов» — см. § 5.5) фрагментация и образование протогалактик происходят в холодном слое, то джинсова масса с самого начала меньше массы протогалактики.

**) Большая часть скоплений такого рода должна была затем разрушаться в результате их внутренней эволюции (§ 6.3), их взаимного приливного влияния или приливного действия ядра галактик, мимо которого они пролетали; могут быть существенны также вспышки сверхновых звезд, приводящие к выбросу из скопления остатков газа; это уменьшает компактность скопления, делает его менее прочным (Пикельнер (1976)).

лялась в пользу кинетической энергии. Если свободное сжатие началось из состояния покоя ($E_K = 0$, $E_G = E \approx \approx -GM/R_1$), то вириальное соотношение между E_K и E_G установилось бы, когда размер системы уменьшился вдвое ($E_K = \frac{1}{2} |E_G|$, $E_G = -\frac{GM}{R^2}$, $R_2 = \frac{1}{2} R_1$), при условии, что полная энергия E не менялась.

В действительности, однако, сжатие протогалактики было, по-видимому, более сильным. Согласно Эггену, Линден-Беллу и Сэндиджу (1962), кинематические характеристики самых старых звезд Галактики указывают на то, что в ходе свободного сжатия размеры протогалактики могли уменьшиться раз в 5—10. В таком случае должно было произойти значительное уменьшение механической энергии системы благодаря излучению газа. Из-за этого на завершающем этапе сжатия протогалактика могла бы иметь довольно значительную светимость.

Свободное падение — первый этап эволюции галактики. На этом нестационарном фоне происходило образование звездных комплексов, которые взаимодействовали гравитационно друг с другом и с протогалактикой в целом; взаимодействие обеспечивало, как полагает Линден-Белл (1967) (см. также работу Эггена, Линден-Белла и Сэндиджа (1962)), быструю бесстолкновительную релаксацию системы (см. ниже).

Наиболее сложной проблемой является, по-видимому, формирование в этом процессе подсистем галактик, которые, как, например, в нашей Галактике, сильно различаются по кинематике. Возможно, первые звезды возникали еще в начале сжатия протогалактики, когда ее вращение было не слишком сильным. Такие звезды могли приобрести сглаженное, более или менее сферическое распределение благодаря быстрой релаксации. Оставшийся газ должен был продолжать падение, причем по мере его уплотнения вращение убыстрялось, пока центробежная сила не остановила сжатие поперек оси вращения; сжатие же вдоль оси продолжалось и привело в конце концов к образованию вращающегося диска — плоской подсистемы (см. Каплан и Пикельнер (1963), Реддиш (1967)). В нашей Галактике массы обеих подсистем, по-видимому, сравнимы, а скорости вращения в районе Солнца различаются в 4—5 раз.

Вириальная теорема. Каждая звезда галактики движется в суммарном гравитационном поле остальных звезд и в то же время участвует в создании общего поля системы. При этом орбиты звезд могут быть и круговыми и более или менее сильно вытянутыми. Для подавляющего большинства звезд орбиты финитны, и при своих движениях звезды остаются в пределах некоторого конечного объема.

Объем, заполняемый орбитами звезд, может быть сферическим, или эллипсоидальным, или подобным диску. Это зависит от того, обладает ли звездная система в целом вращением, причем в случае сферы это вращение слабо или отсутствует, а в случае диска оно столь сильно, что орбиты звезд являются почти круговыми. Мы знаем, что даже в пределах одной звездной системы возможны отдельные большие группы звезд, вращающиеся как целое вокруг общего центра системы с различными угловыми скоростями и обладающие аксиально-симметричным распределением. Такие группы звезд образуют подсистемы различной степени сплюснутости, в наиболее плоской из которых находятся обычно спиральные рукава.

Звездные системы — это системы многих «частиц», которыми служат отдельные звезды или группы звезд. Для их описания необходимы методы статистики. Однако, в отличие от систем, обычно изучаемых в статистической физике, гравитационные системы управляются не молекулярными силами, быстро спадающими с расстоянием, и не электрическими силами, которые экранируются зарядами противоположного знака, а силами тяготения, которые медленно спадают с расстоянием и не экранируются. Вследствие этого потенциальная яма, в которой движется отдельная звезда, определяется главным образом не полем ее ближайших соседей, а полем основной массы звезд системы, которые могут находиться на большом расстоянии от рассматриваемой звезды. Поля всех этих удаленных звезд складываются друг с другом, а их случайные изменения, связанные с перемещениями звезд, почти компенсируются в установившемся состоянии системы. В большинстве реальных звездных систем результирующее поле обладает высокой степенью регулярности в пространстве и стационарности во времени.

Гравитационные системы, как всякие механические системы, можно характеризовать величиной механической

энергии, складывающейся из потенциальной гравитационной энергии и кинетической энергии движения звезд. В связанных системах полная механическая энергия отрицательна, а при наличии общей стационарности имеет место еще и определенная связь между средними по времени значениями потенциальной и кинетической энергий системы, о которой мы уже не раз упоминали. Это так называемая вириальная теорема, которой мы дадим сейчас доказательство.

Обозначим через i номера частиц системы, а через m_i , v_i и p_i — их массы, скорости и импульсы. Тогда кинетическая энергия системы

$$\begin{aligned} E_k &= \frac{1}{2} \sum_i m_i v_i^2 = \frac{1}{2} \sum_i p_i \frac{dr_i}{dt} = \\ &= \frac{1}{2} \frac{d}{dt} \sum_i p_i r_i - \frac{1}{2} \sum_i r_i \frac{dp_i}{dt}. \end{aligned} \quad (6.1.8)$$

Усредним это выражение по какому-нибудь большому промежутку времени Δt , т. е. вычислим выражение вида

$$\bar{f}(t) = \frac{1}{\Delta t} \int_{t - \frac{1}{2} \Delta t}^{t + \frac{1}{2} \Delta t} f(t') dt'.$$

Среднее от первого слагаемого справа в (6.1.8) равно

$$\frac{1}{\Delta t} \sum_i \left[(r_i p_i)_{t + \frac{1}{2} \Delta t} - (r_i p_i)_{t - \frac{1}{2} \Delta t} \right]. \quad (6.1.9)$$

Если рассматриваемая гравитационная система связана, так что ее частицы не удаляются на бесконечность и движутся с конечными скоростями, то при $\Delta t \rightarrow \infty$ это выражение стремится к нулю.

Фактически промежуток усреднения Δt должен быть больше характерного времени изменения суммарной кинетической энергии всех звезд системы. Эти изменения, или флуктуации, связаны с тем, что кинетическая энергия отдельной звезды изменяется при ее собственном движении в объеме системы (например, для звезды на вытянутой орбите — при перемещении от апоцентра к перигею),

а также с не полной компенсацией изменений полей других звезд — близких и далеких, создающих для данной звезды потенциальную яму. Но так как у разных звезд эти изменения никак не связаны и случайны, то характерное время флуктуаций суммарной кинетической энергии всех частиц системы не превосходит типичного времени оборота звезды. Последнее по порядку величины есть $(G\rho)^{-1/2}$, где ρ — средняя плотность системы.

Для усреднения второго слагаемого в (6.1.8) используем уравнение движения

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial E_G}{\partial \mathbf{r}_i}, \quad (6.1.10)$$

где

$$E_G(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = -\frac{1}{2} \sum_{i \neq j} \frac{Gm_i m_j}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (6.1.11)$$

— потенциальная энергия системы при каком-то мгновенном расположении частиц, а $\partial E_G / \partial \mathbf{r}_i$ — вектор с компонентами $\partial E_G / \partial x_i$, $\partial E_G / \partial y_i$, $\partial E_G / \partial z_i$. Тогда мы получим среднее значение кинетической энергии за указанный большой промежуток времени:

$$\bar{E}_K = \frac{1}{2} \sum_i \mathbf{r}_i \frac{\partial E_G}{\partial \mathbf{r}_i}. \quad (6.1.12)$$

Как видно из выражения для $E_G(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$, потенциальная энергия представляет собой однородную функцию степени -1 переменных r_i и потому, по теореме Эйлера об однородных функциях *), мы получаем окончательный результат: |

$$\bar{E}_K = -\frac{1}{2} \bar{E}_G. \quad (6.1.13)$$

Таким образом, среднее значение кинетической энергии связанной и стационарной системы равно половине сред-

*) Для однородной функции $y(x_1, x_2, \dots, x_N)$ степени n , т. е. при $y(\alpha x_1, \alpha x_2, \dots, \alpha x_N) = \alpha^n y(x_1, x_2, \dots, x_N)$, имеет место соотношение:

$$\sum_i x_i \frac{dy}{dx_i} = ny.$$

него значения ее потенциальной энергии с обратным знаком. Полная энергия системы

$$E = E_K + E_G = -\frac{1}{2} E_G. \quad (6.1.14)$$

Статистическое равновесие гравитирующей системы. Реальные звездные системы в подавляющем большинстве далеки от состояния статистического равновесия. О близости к равновесию, да и то с определенными оговорками (см. ниже), можно говорить лишь применительно к таким системам, как рассеянные и шаровые звездные скопления. Тем не менее исследование статистического равновесия как предельного состояния, к которому обязаны стремиться — в силу общих статистических закономерностей — все гравитационные системы, полезно и необходимо для понимания тенденций их динамической эволюции. Поэтому мы рассмотрим сейчас основные черты статистического равновесия в гравитационных системах.

Равновесная система характеризуется так называемым модулем канонического распределения θ , связанным со средней кинетической энергией частиц соотношением

$$\frac{E_K}{N} = \frac{3}{2} \theta. \quad (6.1.15)$$

(Для молекулярных систем $\theta = kT$, где T — температура, k — постоянная Больцмана.) Распределение частиц по скоростям имеет в равновесии вид максвелловского распределения:

$$n_M = A e^{-\frac{mv^2}{2\theta}}. \quad (6.1.16)$$

Здесь n_M есть число частиц, приходящееся на единичный интервал импульсов $(dp) = dp_x dp_y dp_z$ и на единицу объема, так что полное число частиц

$$N = \int n_M (dp) dV, \quad (6.1.17)$$

где интегрирование производится в интервале импульсов от $-\infty$ до $+\infty$ и по всему объему системы. Из этого равенства определяется постоянная распределения A . Если в системе имеется регулярное вращение, то под скоростью v нужно понимать пекулярную, т. е. остаточную после вычитания регулярного движения, скорость частиц.

Распределение частиц в пространстве определяется в равновесии законом Больцмана:

$$n(r) \propto e^{-\varphi/\theta}, \quad (6.1.18)$$

где φ — потенциальная энергия отдельной звезды. Для этой величины имеет место уравнение Пуассона

$$\Delta\varphi = 4\pi G m_0^2 n, \quad (6.1.19)$$

которое определяет общее сглаженное гравитационное поле системы, называемое в этом случае самосогласованным *).

Используя соотношение (6.1.15), можно выразить полную энергию системы (6.1.14) через модуль распределения:

$$E = -\frac{3}{2} N\theta. \quad (6.1.20)$$

Но отсюда следует, что теплоемкость системы, $dE/d\theta$, отрицательна. Из термодинамики известно, что система с отрицательной теплоемкостью неустойчива (Ландау и Лифшиц (1964)). Эта неустойчивость имеет место и для системы, управляемой гравитационными силами, и там она проявляется в виде следующего парадокса (Антонов (1962)). Представим себе систему, в которой звезды можно разделить на две подсистемы с различными температурами. Пусть в некоторый начальный момент эти две системы почти одинаковы по структуре, но подсистема 1 несколько «горячее» системы 2, т. е. кинетическая энергия звезд в первой подсистеме несколько больше, чем во второй. Тогда при случайных сближениях звезд этих подсистем звезды подсистемы 1 вследствие стремления к равномерному распределению кинетической энергии будут в среднем отдавать свою энергию звездам подсистемы 2. Однако при этом температура подсистемы 1 отнюдь не будет понижаться, а температура подсистемы 2 не будет повышаться. Вместо этого в подсистеме 1 уменьшение полной энергии будет в силу ее отрицательности означать увеличение ее абсолютного значения, а вместе с тем, по вириальной теореме, и увеличение кинетической энергии, а следовательно, и

*) Для простоты считаем здесь, что массы звезд m_0 одинаковы.

температуры; наоборот, в подсистеме 2 полная энергия будет увеличиваться, а ее абсолютное значение будет уменьшаться и вместе с ним будет уменьшаться средняя кинетическая энергия и температура. Таким образом, температуры этих двух подсистем будут не сближаться друг с другом, как этого требует основная теорема статистической физики (*H*-теорема), а наоборот, их разность будет возрастать. Мы увидим, однако (§ 6.3), что такое аномальное поведение гравитационных систем длится лишь в течение некоторой конечной эпохи их эволюции и заканчивается «победой» *H*-теоремы.

Воспользуемся еще раз вириальной теоремой, чтобы найти энергию испарения отдельной частицы из системы, E_e , т. е. абсолютное значение потенциальной энергии, характеризующей ее связь с системой. Потенциальная энергия частицы

$$\varphi_i = -Gm_i \sum_j \frac{m_j}{|r_i - r_j|}, \quad (6.1.21)$$

а среднее значение этой величины, взятое с обратным знаком, и есть энергия испарения E_e . Если считать, что все звезды в системе имеют одинаковые массы, то

$$E_e N = \sum_{i \neq j} \frac{Gm_i m_j}{|r_i - r_j|} = 2|E_G|, \quad (6.1.22)$$

откуда

$$E_e = \frac{2|E_G|}{N}, \quad (6.1.23)$$

или, с учетом (6.1.15),

$$E_e = 4 \frac{E_K}{N}. \quad (6.1.24)$$

Таким образом, в гравитационных системах существует универсальное соотношение между средними значениями кинетической энергии и энергии связи отдельных частиц. Для систем, близких к равновесию,

$$E_e = 6\theta. \quad (6.1.25)$$

Из этого обстоятельства вытекает важное следствие: для гравитационной системы максвелловское распределение по скоростям никогда не может осуществляться

полностью. В отличие, например, от молекулярных систем, где энергия испарения может превышать среднюю энергию частиц во сколько угодно раз, в гравитационной системе конечная доля звезд (порядка процента) при максвелловском распределении имеет кинетическую энергию, превышающую энергию связи, или энергию испарения. Эти звезды не могут удерживаться системой и покидают ее. Система оказывается, таким образом, лишенной высокоэнергичных своих членов, а стремление к установлению максвелловского распределения приводит все время к появлению новых «быстрых» звезд, так что испарение звезд из системы происходит непрерывно. Кинетика этого процесса будет специально рассмотрена в § 6.3.

Явление испарения звезд из гравитационных систем указывает, в частности, на принципиальное ограничение возможной длительности их стационарного состояния. Мы увидим, что для галактик или скоплений галактик, такая стационарность (или квазистационарность) существует в течение многих десятков миллиардов лет, тогда как для шаровых или рассеянных скоплений промежутки времени, в течение которого состояние системы можно считать стационарным, меньше их возраста.

Нестационарность и бурная релаксация. Полное статистическое равновесие гравитационной системы недостижимо, но происходящие в ней процессы так или иначе стремятся приблизить его. Эти процессы, называемые в статистике процессами релаксации, действуют и в установившемся квазистационарном состоянии и на самой первой нестационарной фазе эволюции галактики. Мы увидим в § 6.2, что релаксация в установившемся состоянии крайне медленна для галактик и может быть существенна только для звездных скоплений. Что же касается релаксации при сильной нестационарности системы, то согласно Линден-Беллу (1967), она развивается очень быстро — в темпе общего сжатия системы.

Если основная масса звезд галактики образовалась на фазе ее сжатия, то можно предположить, что изменение гравитационного поля при сильной общей нестационарности было случайным во времени и приводило к картине звездных движений, напоминающей турбулентность. При таких условиях естественно ожидать появления распределения звезд по скоростям максвелловского типа. Что-

механизм, который стремится привести систему к состоянию статистического равновесия.

Тесные звездно-звездные сближения. Кинетическая энергия $1/2 m_0 v^2$ одной из сближающихся звезд (будем считать для простоты их массы m_0 одинаковыми) изменится в результате сближения существенно, если звезды сближаются на такое расстояние r_0 , на котором потенциальная энергия их взаимодействия, — Gm_0^2/r_0 , по абсолютной величине сравнима с кинетической:

$$\frac{Gm_0^2}{r_0} \approx m_0 v^2, \quad r_0 \approx \frac{Gm_0}{v^2}. \quad (6.2.1)$$

В этой формуле v есть относительная скорость движения звезд, имеющая разные значения при различных сближениях. Однако для большинства сближений, существенных для всего процесса приближения к равновесию, характерная скорость, близкая по порядку величины к средней скорости движения звезд. Так мы и будем понимать эту величину в формуле (6.2.1). Для невращающейся системы, находящейся в стационарном (или квазистационарном) состоянии, из вириальной теоремы находим по порядку величины:

$$\frac{GN^2 m_0^2}{R} \approx Nm_0 v^2.$$

Здесь слева стоит абсолютная величина гравитационной энергии системы с числом звезд N и размером R . Отсюда

$$v^2 \approx \frac{GNm_0}{R}. \quad (6.2.2)$$

Можно выразить эту величину и через среднюю плотность системы:

$$v \approx R (G\rho)^{1/2}. \quad (6.2.3)$$

Для галактики со средней плотностью $\rho = 10^{-24}$ г/см³ и размером $R = 3 \cdot 10^{22}$ см величина $v \approx 10^7$ см/сек.

Расстояние r_0 , играющее роль прицельного параметра для гравитационных столкновений *), можно связать

*) Обратим внимание на аналогию с кулоновским радиусом в плазме (который мы тоже обозначали через r_0); см. § 2.2.

теперь со средними параметрами системы:

$$r_0 \approx \frac{m_0}{R^2 \rho} \approx R \frac{m_0}{M} \approx \frac{R}{N}, \quad (6.2.4)$$

где $M = m_0 N$ — полная масса системы. Сравним это со средним расстоянием между звездами

$$\bar{r} \approx \frac{R}{N^{1/3}}.$$

Отношение

$$\frac{r_0}{\bar{r}} \approx N^{-2/3} \quad (6.2.5)$$

много меньше единицы, и это сразу указывает на то, что интересующие нас сближения звезд должны быть редкими событиями.

Число сближений, которые звезда испытывает в единицу времени, равно потоку других звезд через сечение

$$\sigma_0 \approx r_0^2. \quad (6.2.6)$$

Плотность потока звезд, движущихся со скоростью v , есть $n_0 v$, где $n_0 \approx N/R^3$ — концентрация звезд в системе. Поэтому через площадку с размером r_0 проходит в единицу времени число звезд, равное $n_0 v \sigma_0$. Это есть частота сближений данной звезды, а обратная величина,

$$\tau \approx \frac{1}{n_0 v \sigma_0}, \quad (6.2.7)$$

есть время, за которое данная звезда испытает сближение, существенно изменяющее ее энергию. Так как эти соображения можно отнести к любой звезде, скорость которой не слишком отличается от средней, т. е. к основной массе звезд, то через время τ значительное число звезд изменит свою энергию существенно, и значит, вся система заметно приблизится к состоянию равновесия. Время τ , являющееся, таким образом, временем релаксации, выражается с помощью предыдущих формул через среднюю плотность системы и число звезд в ней:

$$\tau \approx N (G\rho)^{-1/2}. \quad (6.2.8)$$

Время, определяемое этой формулой, характеризует скорость приближения гравитационной системы к равновесию посредством таких тесных сближений звезд, при

которых их энергии меняются на величину, сравнимую с самими этими энергиями. Сделаем численную оценку этого времени для параметров типичной галактики. При числе звезд $N = 10^{11}$ и средней плотности $\rho = 10^{-24}$ г/см³ находим

$$\tau \approx 3 \cdot 10^{26} \text{ сек} \approx 10^{19} \text{ лет.}$$

Это время на множество порядков превышает и возраст галактики, и возраст Вселенной $\sim (1 \div 2) \cdot 10^{10}$ лет. Следовательно, в галактиках тесные звездно-звездные сближения практически не происходят и не могут играть роли в их релаксации.

Соображения о тесных сближениях можно применить не только к звездам в галактиках, но и к целым галактикам, движущимся друг относительно друга в скоплениях. Это законно, так как расстояния между галактиками в скоплениях много больше размеров галактик. Если в скоплении содержится, скажем, 10^3 галактик, то при средней плотности скопления $\rho \approx 10^{-27}$ г/см³ найдем

$$\tau \approx 10^{20} \text{ сек} \approx 3 \cdot 10^{12} \text{ лет,}$$

что также больше возраста галактик, хотя разрыв не столь значителен.

Иной результат получается на другом конце шкалы звездных систем, именно для рассеянных и шаровых звездных скоплений. Средняя плотность рассеянных скоплений раз в 10 больше, чем плотность галактики в целом, но зато число членов в них $N \approx 10^2 \div 10^3$. Поэтому для рассеянных скоплений

$$\tau \approx 10^{17} - 10^{18} \text{ сек} \approx 3 \cdot 10^9 - 3 \cdot 10^{10} \text{ лет.}$$

Плотность в их центральных областях существенно больше и потому для этих областей время τ может оказаться еще в 10—30 раз меньше, и будет сравнимо с возрастом рассеянных скоплений или меньше его. В этом случае можно сказать, что такие системы должны успеть приблизиться к статистическому равновесию в той мере, в какой это вообще возможно для гравитационных систем. То же относится, по-видимому, и к центральным областям шаровых скоплений. Эти системы действительно обнаруживают близость к равновесию, проявляющуюся в правильности их структуры, сфероидальности формы.

Диффузия звезд по энергиям. Процесс релаксации посредством тесных звездно-звездных сближений предполагает существенное изменение энергий звезд после одного или нескольких близких прохождений для каждой звезды. Очевидно, однако, что изменение энергии звезды происходит и при более далеких взаимных прохождениях звезд. В каждом акте такого взаимодействия энергия звезды меняется гораздо меньше, но зато более далекие пролеты должны происходить гораздо чаще, чем тесные сближения. Поэтому можно ожидать, что релаксация гравитационных систем посредством далеких прохождений также возможна; оказывается, что она даже более эффективна, чем релаксация посредством тесных сближений.

Роль далеких прохождений может быть существенна потому, что гравитационные силы гораздо медленнее убывают с расстоянием, чем, скажем, силы межмолекулярного взаимодействия в газах. Звезда взаимодействует гравитационно с самыми отдаленными частями системы, причем особенно важны сближения на наибольших расстояниях, на которых наложение взаимодействий с различными партнерами еще не компенсирует их среднего результата. Эти расстояния должны быть, очевидно, близки к среднему расстоянию между звездами \bar{r} .

При $r = \bar{r}$ энергия взаимодействия двух звезд с массами m_0 ,

$$\frac{Gm_0^2}{\bar{r}} \approx \frac{Gm_0^2}{R} N^{1/2},$$

меньше энергии отдельной звезды,

$$E_i \approx \frac{Gm_0^2}{R} N,$$

в отношении $N^{-2/3}$. Чтобы энергия звезды изменилась на величину, сравнимую с ней самой, требуется много таких сближений, так что это изменение энергии $(\Delta E)_i$ складывается из многих отдельных малых приращений:

$$(\Delta E)_i = \sum_i (\delta E)_i.$$

При этих многочисленных сближениях величины $(\delta E)_i$ одинаково или почти одинаково часто могут быть и положительными, и отрицательными. Поэтому представление о

результатирующем изменении энергии может дать лишь квадрат этой величины. А так как этот квадрат столь же случаен, как и отдельные изменения, то необходимо определить его среднее значение. При этом усреднении средние значения удвоенных произведений $2(\delta E)_i(\delta E)_k$ можно считать равными нулю, так как эти изменения независимо друг от друга имеют положительные или отрицательные значения. Поэтому средние значения квадратов изменений оказываются аддитивными:

$$\overline{(\Delta E)_i^2} = \sum_i \overline{(\delta E)_i^2}.$$

Изменение какой-либо величины посредством многочисленных мелких шажков, в каждом из которых эта величина меняется случайным образом, представляет собой диффузию. Можно сказать, что изменение энергии звезды в процессе многократных случайных сближений происходит посредством диффузии в пространстве энергий.

Гравитационные силы зависят от расстояния по тому же закону, что и кулоновские силы, действующие между заряженными частицами в плазме. Поэтому естественно ожидать, что учет далеких сближений звезд должен приводить к увеличению эффективного сечения взаимодействия на множитель, подобный кулоновскому логарифму:

$$\sigma_0 = r_0^2 \Lambda. \quad (6.2.9)$$

В плазме $\Lambda = 2\pi \ln(r_D/r_0)$, где $r_D = (kT/4\pi n e^2)^{1/2}$, $r_0 = e^2/kT$ — дебаевский и кулоновский радиусы. В системе частиц, взаимодействующих благодаря гравитационным силам, роль энергии kT играет величина Gm_0M/R , а роль квадрата заряда — величина Gm^2 . Подставляя еще $n = N/R^3$, получим вместо кулоновского логарифма выражение, которое по порядку величины есть $10 \ln N$. Соответствующее время релаксации *)

$$\tau_E = \frac{N}{(G\rho)^{1/2}} \frac{1}{10 \ln N}. \quad (6.2.10)$$

Для звездных скоплений, рассеянных и шаровых, время такой диффузионной релаксации оказывается меньше в

*) Детальный анализ различных аспектов столкновительной релаксации звездных систем имеется в книгах Чандрасекара (1948) и Огородникова (1958).

30—50 раз. Именно оно и определяет в них основное направление эволюции, одной из существенных черт которой является испарение звезд (см. следующий параграф).

Что же касается галактик, то для них даже и время τ_E , очевидно, слишком велико по сравнению с возрастом. В них ни тесные, ни далекие сближения звезд неэффективны. В нынешнем состоянии этих систем никакие эффекты такого рода не могут играть существенной роли.

Разнообразие типов галактик, значительные неоднородности распределения звезд в них, наличие подсистем и т. п. не отменяет этого вывода, сделанного на основании самых общих соображений и порядковых оценок, так как разрыв между временами τ и τ_E и возрастом галактик составляет несколько порядков. При всех условиях галактики являются бесстолкновительными системами.

Это обстоятельство приводит к важной и до сих пор не решенной проблеме. Звезды нашей Галактики в окрестностях Солнца обнаруживают некоторое регулярное распределение по скоростям. Оно представляет собой анизотропное обобщение распределения Максвелла. Удобнее в этом случае выражать функцию распределения как функцию декартовых компонент скоростей и выбрать систему координат так, чтобы одна из осей шла в направлении к центру Галактики, вторая — в плоскости галактического диска и перпендикулярно к первой, а третья — перпендикулярно к плоскости диска. В этой системе координат функция распределения имеет вид (распределение Шварцшильда):

$$n(v_1, v_2, v_3) dv_1 dv_2 dv_3 = A \exp\left(-\frac{v_1^2}{2\sigma_1^2} - \frac{v_2^2}{2\sigma_2^2} - \frac{v_3^2}{2\sigma_3^2}\right) dv_1 dv_2 dv_3. \quad (6.2.11)$$

Здесь v_1, v_2, v_3 — компоненты peculiarной скорости, $\sigma_1, \sigma_2, \sigma_3$ — дисперсии, равные среднеквадратичным значениям соответствующих скоростей: $\sigma_i^2 = \overline{v_i^2}$. Наибольшее значение имеет дисперсия σ_1 , соответствующая движению к центру Галактики и от него, а наименьшее — σ_3 , соответствующая движению перпендикулярно к плоскости галактического диска. Эти величины $\lesssim 20$ км/сек, причем дисперсии оказываются различными для звезд разных спектральных классов,