

А.КОЗАННЕ Ж.ФЛЕРЕ Г.МЭТР М.РУССО
ОПТИКА И СВЯЗЬ

А.КОЗАННЕ
Ж.ФЛЕРЕ
Г.МЭТР
М.РУССО



**ОПТИКА
И
СВЯЗЬ**





OPTIQUE ET TÉLÉCOMMUNICATIONS

**TRANSMISSION ET TRAITEMENT OPTIQUES
DE L'INFORMATION**

par

Alain COZANNET

Ingénieur au C.N.E.T.

Jacques FLEURET

Professeur à l' E.N.S.T.

Henri MAITRE

Maître assistant à l' E.N.S.T.

Michel ROUSSEAU

*Ingénieur à la Compagnie Lyonnaise
de Transmissions Optiques*

PREFACE DE

M. FRANÇON

**Professeur à l'Université de Paris
Pierre et Marie CURIE**

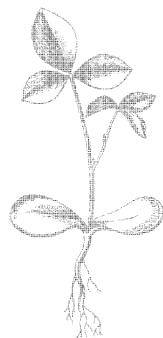
А. КОЗАННЕ
Ж. ФЛЕРЕ
Г. МЭТР
М. РУССО

**ОПТИКА
И
СВЯЗЬ**

ОПТИЧЕСКАЯ ПЕРЕДАЧА
И ОБРАБОТКА ИНФОРМАЦИИ

Перевод с французского
А. Г. КОЧЕТКОВА и Н. Г. СОКОЛОВОЙ
под редакцией канд. физ.-мат. наук
В. К. СОКОЛОВА

МОСКВА «МИР»
1984



Scan AAW

ББК 32.885

О627

УДК 595 + 654

О62 **Оптика и связь: Опт. передача и обраб. информ. Пер. с фр./**
А. Козанне, Ж. Флере, Г. Мэтр, М. Руссо — М.: Мир,
1984. — 504 с., ил.

Книга представляет собой первое в мировой литературе учебное пособие по оптической волоконной связи и оптическим методам обработки информации. Впервые в одной книге с единых позиций рассмотрен весь комплекс вопросов, связанных с этой областью знания: распространение света в направляющих системах, теория голографического метода регистрации и восстановления волновых полей, оптические методы обработки информации, измерение характеристик оптических волокон и методы их изготовления. Рассматривается также элементная база оптических систем передачи и обработки информации.

Для студентов, аспирантов и преподавателей, а также для инженеров и научных работников, работающих в области оптической волоконной связи и оптической обработки информации.

О $\frac{1704050000-279}{041(01)-84}$ **69—84, ч. 1**

ББК 32.885

6Ф0.1

Редакция литературы по физике

© Editions EYROLLES et CNET-ENST 1981

© Перевод на русский язык, «Мир», 1984

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Появление лазеров вновь пробудило интерес к оптическим методам передачи и обработки информации и стимулировало исследования в этом направлении

За прошедшее двадцатилетие были выполнены многочисленные исследования фундаментального и прикладного характера в различных областях физики и техники, которые привели к возникновению новых разделов оптики, таких, как голография, оптическая обработка информации, волоконная оптика, нелинейная оптика, интегральная оптика и др., составивших основу нового и весьма перспективного научно-технического направления, нацеленного на существенное повышение информационной пропускной способности и надежности систем связи и на увеличение быстродействия устройств обработки изображений. Эти исследования завершились разработкой теоретических основ оптических методов передачи и обработки информации и созданием в основном элементной базы, а также позволили более верно и объективно оценить место этих методов среди других, лучше понять их возможности и свойственные им ограничения. Это дает основание утверждать о переходе рассматриваемых методов из области фундаментальных исследований в сферу практических приложений.

Наибольшие практические успехи были достигнуты в области создания волоконно-оптических систем связи. Освоение промышленностью выпуска оптических волокон с предельно низкими потерями, разработка светодиодов и полупроводниковых лазеров, а также высокочувствительных фотодиодов — все это позволило создать первые волоконно-оптические системы передачи информации. В настоящее время во многих странах мира, как и в нашей стране, проходят опытную эксплуатацию волоконно-оптические линии связи, используемые для передачи телефонных разговоров и телевизионных программ. В ряде вузов началась подготовка специалистов по оптической связи и оптической обработке информации. Однако несмотря на большое число книг, посвященных различным аспектам волоконно-оптической связи и обработки информации, до сих пор нет учебника, в котором последовательно с единых позиций излагались бы все вопросы передачи и обработки информации оптическими методами.

Предлагаемая вниманию читателя книга А. Козанне, Ж. Флере, А. Мэтра и М. Руссо «Оптика и связь» в значительной мере заполняет имеющийся пробел. Книга написана известными французскими специалистами в области передачи и обработки информации оптическими методами, имеющими большой опыт преподавания в высшей школе. И хотя авторы в своем предисловии говорят, что их книга не предлагается в качестве стандартного учебника,

по нашему мнению, она представляет собой первый в мировой литературе учебник, в котором достаточно широко и глубоко, с единых позиций, рассмотрены все практически важные вопросы, связанные с передачей информации оптическими методами и с оптической обработкой изображений. Но это не совсем обычный учебник. Книга рассчитана не только, а может быть, и не столько на студентов вузов, сколько на специалистов в разных областях науки и техники, желающих серьезно ознакомиться с новыми методами передачи и обработки информации. Во-первых, в книге приводятся результаты самых последних исследований в рассматриваемых областях. Во-вторых, авторы не только излагают физические принципы оптических методов передачи и обработки информации, но и объективно характеризуют состояние рассматриваемых вопросов на сегодняшний день, повсюду наряду с преимуществами таких методов подчеркивая свойственные им недостатки, трудности и нерешенные проблемы. В-третьих, книга не перегружена математическими выкладками; везде, где это возможно, авторы отдают предпочтение физической ясности и наглядности изложения, а не математической строгости выводов, прибегая к различным упрощающим предположениям.

Содержание книги по главам подробно рассмотрено в предисловии профессора М. Франсона и в предисловии авторов. Поэтому мы отметим только, что, помимо физических принципов и элементной базы, в ней рассмотрены вопросы проектирования волоконно-оптических систем, основанных как на аналоговых, так и на цифровых методах передачи информации, а также вопросы экономической конкурентоспособности таких систем по сравнению с используемыми в настоящее время.

Что касается вопросов обработки информации, то в книге нашли отражение оптические методы пространственной фильтрации изображений, реализуемые средствами не только когерентной, но и, что особенно интересно для советского читателя, некогерентной оптики.

В заключение приведен обзор предложенных на сегодня комбинированных оптико-электронных систем обработки изображений и сделан прогноз развития чисто оптических систем передачи информации.

Книга снабжена обширной библиографией, содержащей практически все основополагающие работы по рассматриваемым вопросам.

Перевод книги выполнен А. Г. Кочетковым (гл. 1—4, 17, 18, приложения и предисловия) и Н. Г. Соколовой (гл. 5—16 и 19).

Б. СОКОЛОВ

ПРЕДИСЛОВИЕ К ФРАНЦУЗСКОМУ ИЗДАНИЮ

С тех пор, как появился лазер, оптика непрерывно развивается не только в плане фундаментальных исследований, но и в направлении практических приложений, главным образом для передачи и обработки информации.

Книга Козанне, Флере, Мэтра и Руссо заполнила пробел в данной области, выразившийся в отсутствии достаточно полной книги на эту тему на французском языке. Авторы начинают изложение с уравнений Максвелла, чтобы показать, как переменное электромагнитное поле не остается локализованным, а распространяется в соответствии с уравнением распространения. Они используют классический, а не квантовый подход при анализе эффектов поглощения и дисперсии света, которые играют важную роль в оптической связи. Эти вопросы являются предметом рассмотрения первых двух глав, и только в третьей главе изучаются явления отражения и распространения плоской волны в световоде. В четвертой главе на основе принципа Гюйгенса — Френеля и преобразования Фурье описаны дифракционные эффекты. Здесь дифракция изучается не с целью описать явления, наблюдаемые при дифракции на различных апертурах, а главным образом чтобы показать, как дифракция влияет на расходимость световых пучков. Авторы дают пример дифракции, наблюдаемой при последовательном отражении плоской волны от двух плоских зеркал, расположенных параллельно. Рассмотренные дифракционные эффекты играют важную роль в лазерных резонаторах. Конец четвертой главы посвящен основным понятиям когерентности, а также определению пространственной и временной когерентности. Голография становится предметом подробного рассмотрения в пятой главе. Здесь описаны различные типы голограмм, включая цифровые.

На основе материала, изложенного в предыдущих главах, авторы приступают к анализу проблем обработки информации методами когерентной и некогерентной оптики. Пространственная фильтрация применительно к задаче восстановления искаженных изображений, оптические корреляторы, согласованные фильтры и распознавание образов — таковы основные вопросы, рассматриваемые в шестой главе. В следующей главе изучаются методы обработки информации с использованием некогерентного освещения, а именно двухзрачковая пространственная фильтрация с использованием зрачковых функций, синтезируемых голографическими и цифровыми методами, а также методы пространственной фильтрации, не использующие свойств дифракции.

Начиная с 8-й главы рассматриваются исключительно проблемы оптической связи. Сначала дано подробное описание оптических волокон: распространение света в оптическом волокне, измерение их характеристик, методы изготовления оптических волокон и оптических кабелей на их основе. Этим

вопросам посвящены главы 8—11. В главе 12 рассматриваются лазеры с единственной целью дать представление о когерентном и некогерентном излучении, о характеристиках расходимости, монохроматичности или эффективности источников света, что необходимо для понимания излагаемого далее материала. После источников излучения в главе 13 рассматриваются фотоприемники. Однако необходимо уметь модулировать свет в соответствии с передаваемым сигналом, поскольку для передачи информации используется оптическая несущая частота. Этот вопрос является предметом главы 14, где авторы описывают электрооптические и акустооптические модуляторы света.

Три последующие главы посвящены вопросам использования конструктивных элементов, описанных в предыдущих главах, при проектировании волоконно-оптических линий связи. Глава 18 содержит данные не только о простой передаче сигналов, но также и об обработке оптических сигналов. И наконец, в последней главе авторы описывают интересные комбинированные методы обработки изображений, в которых оптические методы сочетаются с электронными.

Чтение этой очень хорошей книги не составляет труда благодаря тому, что изложение по большей мере обращено к интуиции читателя и сосредоточено на физическом смысле, а не на сложных математических формулах, которые здесь не нужны.

Приведенная очень обширная библиография свидетельствует о серьезности, с которой была написана эта книга. Книга будет очень полезна инженерам и студентам университетов и высшей школы.

М. ФРАНСОН

ПРЕДИСЛОВИЕ

Если попытаться установить, когда произошло слияние оптики и связи, то мы неизбежно потеряемся в седой древности, поскольку это случилось значительно раньше, чем появились сами эти слова и определились понятия, которые они обозначают.

Доисторический человек, несомненно, очень давно экспериментально исследовал пределы досягаемости человеческого голоса, что, возможно, толкнуло его на поиски акустических методов расширения этих пределов: так появились «там-тамы» и всякие сигнальные рога. Очень давно первобытный человек начал пользоваться и оптическими методами связи: ночью он использовал зажженные огни, днем — сигнальные дыммы.

Такие способы передачи информации были известны в Китае, у ассирийцев, в Египте, начиная с самых первых цивилизаций. Именно так, например, в 1084 г. до н. э. Агамемнон после взятия Трои передаст своей любимой жене в Аргос весть о победе и о своем возвращении. Известно, как оценит это Клитемнестра, но не в этом дело: 800 км с 9 эстафетами из огней на холмах — это самая длинная линия связи, которая будет превзойдена лишь через 3000 лет! А если Агамемнон существовал лишь в воображении вдохновенного поэта, то, стало быть, этот поэт, Эсхил, и явился отцом оптической связи в 450 г. до н. э. Такой метод сигнализации хорошо апробирован римлянами в начале нашей эры при завоевании Англии. Впоследствии пикты стали широко применять его против самих римлян. В будущем индейцы используют эту технику связи в своей борьбе с бледнолицыми.

Но не только сигналы исключительной важности, связанные с временами войн, передавались оптическими методами. Начиная с 9 в. до н. э. маяк на Сигейском мысе указывал мореплавателям вход в Геллеспонт¹⁾ и опасные скалы. Знаменитый маяк на острове Фарос в 3 в. до н. э. светил более чем на 150 км. Информация скудная, но кто будет отрицать, что она жизненно необходима?

В тот же исторический период Полибий описывает интересный метод, когда информация кодируется не глобально (война — мир, победа — поражение), а передается последовательно, буква за буквой, с кодированием на уровне букв. Для этого перед стеной перемещали 5 факелов, и по их взаимному расположению можно было угадывать символ. Таким принципом кодирования воспользуется Кардан в 1551 г., он его упростит и предложит использовать пять башен с огнями, которые можно закрывать и открывать.

¹⁾ Теперь пролив Дарданеллы. — *Прим. ред.*

Хотя методы в принципе уже существуют, с их разработкой и внедрением наблюдается отставание. Еще не сложилось понятие связи, необходимость в ней почти не ощущается. Первая Французская республика заполнит этот пробел и предложит нам первую настоящую линию связи: телеграф Шаппа (1791—1794). Отметим без всякого удивления, что это была оптическая линия связи. Поскольку она содержит в себе все элементы современной системы связи, мы их перечислим, пользуясь терминами, которые нам встретятся в последующих главах; передатчик — механический (управляемые рычаги); приемник — человек со вспомогательной оптикой (глаз + зрительная труба); канал — передача в воздушном пространстве; расстояние между ретрансляторами 5—10 км; кодирование — 196 знаков; информационная пропускная способность на линии Тулон — Париж (120 ретрансляторов) 50 знак/ч; отношение сигнала к шуму — сильно зависит от тумана.

Одновременно с этой системой англичанин Марри разработал телеграф с прерывателем (заслонками) (1795), который имел определенный успех, особенно в США.

Затем оптические методы пришли в забвение, они остались только в виде семафоров, вытесненные повсюду электрическими методами. В 1843 г. Морзе предложил свой телеграф¹⁾. С 1860 г. Казелли применил его для фототелеграфа. В 1876 г. Белл изобрел телефон, который похож на «веревочный телефон» Хука (1667), но в котором по проводам передаются электрические импульсы, а не акустические волны. Маркони дал способ связи без проводов (1896)²⁾. Белин, Рихти передают фотографии по телеавтографу (1902—1907), и с 1935 г. становится возможным прием движущихся изображений по телевидению³⁾.

Страница перевернута, оптические методы кажутся отжившими. «Фея электричества» побеждает, и рождение электроники может только упрочить этот триумф.

Взглянем более внимательно на условия столь стремительного развития и добавим к названным именам имена Максвелла, Планка, Бора, Эйнштейна, Дирака, Ферми... Перечень имен физиков, чьи работы отметили, как вехами, переход от электричества к электронике, от ключа Морзе к телевидению, был бы очень длинным. Менее чем за 50 лет утвердились самые смелые теории, вызывая за собой гигантские технические преобразования.

Но меняется не только техника. Общество тоже не стоит на месте. Спустя сто лет после своего рождения связь стала жизненно необходимой, и каждое десятилетие вносит новый вклад в развитие средств связи и в расширение их возможностей. Телевидение стало цветным, телефон занял место

¹⁾ Русский ученый П. Л. Шиллинг (1786—1837) создал в 1832 г. первый практически пригодный электромагнитный телеграф. — *Прим. ред.*

²⁾ Русский ученый, изобретатель радиосвязи А. С. Попов (1859—1905) 7 мая 1895 г. продемонстрировал изобретенный им первый в мире радиоприемник. — *Прим. ред.*

³⁾ Русский ученый Б. Л. Розинг (1869—1933) в 1907 г. предложил систему телевидения, использующую электронно-лучевую трубку, и в 1911 г. осуществил первую в мире передачу изображения по этой системе. — *Прим. ред.*

в каждом доме, как ванна и электричество. Газеты, подготавливаемые в Нью-Йорке, печатаются в Париже, и каждые 5 минут искусственный спутник посылает на землю столько информации, что она может составить книжку карманного формата. Каждый день возникают новые потребности, которые вызывают появление новых средств с замысловатыми названиями: видеоконференция, телетекст, телефотография, телекино, видеотелефон, телегазета. ЭВМ тоже хочет знать своих коллег и обмениваться с ними опытом.

Кажется, и электроника уже не удовлетворяет таким запросам; требуется увеличивать плотность передач и информации, повышать частоту, уплотнять каналы связи, если мы не хотим с утра до вечера набирать вызов. Таков первый мотив нашего обращения к оптическим длинам волн, обращения, на которое был получен ответ в 1966 г., когда Као и Хокхем увидели в оптическом волокне идеальную среду для передачи оптических сигналов. Их предсказание было подтверждено в 1970 г.: в лабораториях фирмы Corning Glass Works было получено первое оптическое волокно с потерями 20 дБ/км, действительно пригодное для практического применения. С того времени эти цифры непрерывно уменьшались, и в настоящее время потери в оптических волокнах составляют 0,5 и даже 0,2 дБ/км. Но, кроме того, необходимо обрабатывать, сжимать, фильтровать, распознавать передаваемую информацию. И поскольку в этой информации все большую роль играют изображения, то, может быть, имеет смысл не только передавать их, но и обрабатывать оптическими методами. Таков второй мотив обращения к оптическим методам.

Наша книга ставит своей целью дать какой-то ответ по этим двум пунктам, чтобы будущее не казалось совсем неясным.

Оптика и связь. Как вести изложение этой комплексной темы, на кого его ориентировать? Есть по крайней мере три группы людей, для которых эта область знаний весьма интересна. Первая группа — специалисты, активно работающие в этой области и развивающие ее дальше, другая группа — студенты, получившие основные знания физики и теории обработки сигналов и желающие специализироваться в данной области, и наконец, специалисты из иных областей науки и техники, желающие глубже изучить оптическую связь не для того, чтобы заниматься ею, а для того, чтобы почерпнуть здесь новые методы, дополняющие те, которыми они обычно пользуются как в области связи, так и в области обработки информации.

Наша книга адресуется не к представителям первой группы. Мы хотели, чтобы основы излагаемых здесь методов привлекли внимание читателя не менее, чем сами методы. Поэтому мы глубоко копнули корни соответствующих физических явлений и сделали это не для того, чтобы точными математическими выражениями описать тесную связь между, например, уравнениями Максвелла и оптической фильтрацией пространственных частот, а для того, чтобы подчеркнуть, какая последовательность предположений, абстракций и ограничений позволяет нам связать такие простые понятия, как режим гармонических колебаний, плоские волны или же принципы суперпозиции и пространственной инвариантности, чтобы прийти к более связному, более понятному, чем в начале, миру. Хотя в арсенале современной физики имеются точные решения рассматриваемых задач, мы часто будем обращаться к интуиции. Мы хотим обратить особое внимание на *внутреннюю* логику интуиции.

Так, предположить существование гармонических колебаний — значит ограничить сложность нашего мира. Это — упрощающее предположение, как и допущение существования точечных источников. Оба предположения соответствуют одному и тому же способу абстрагирования (как бы «предельному переходу»), но разница в объектах абстрагирования: в первом случае временно́е содержание волны, а во втором — ее пространственное содержание. Точно так же мы обращаем внимание на связь между основными предположениями, которые приходится принимать при решении уравнений Максвелла, и их отдаленными следствиями, ежедневно встречающимися на практике. Так принцип линейности приводит к понятиям мод в световоде и к понятию мультиплексирования фильтров, а пространственная инвариантность оптических систем естественным образом приводит нас к принципу параллельной обработки.

Таким образом, наша книга предназначена для обучения, но она не предлагается в качестве стандартного учебника. Области, которые мы рассматриваем, молоды. Ни одна из них не существовала до войны, и многие из них получили настоящее развитие только в последние пятнадцать лет. И нет еще какой-либо академической теории, авторитет которой был бы бесспорным. Мы выбрали подход, при котором читатель должен руководствоваться *интуицией* и *физическим смыслом*. Так мы подходим к общему изложению электромагнетизма, вопросов дифракции, голографии, распространения волн в оптических волокнах, исследуемого методами геометрической оптики, лазеров и т. д. Во многих случаях мы отсылаем читателя за более строгими доказательствами к другим работам.

Наша книга отличается от стандартного учебника еще и тем, что мы старались на каждом этапе излагать самые новые методы и самые последние результаты, даже если их преимущества и ценность еще не доказаны окончательно. Нам кажется, что обучение современной оптике в применении к связи должно состоять как в изложении ее успехов, так и в описании проб и ошибок, заблуждений, умозрений, сопровождающих ее развитие. Поэтому мы не упускаем случая иной раз заглянуть в «волшебный кристалл» и указать схемы из области будущего, которые, как нам кажется, пользуются благосклонностью разных лабораторий.

Теперь перейдем к структуре изложения материала, учитывая двойную задачу оптики в области связи: с одной стороны, обработка информации, а с другой — ее передача.

Первые четыре главы имеют целью развить из уравнений Максвелла теорию распространения света, которой мы будем пользоваться в следующих главах. Они показывают, как, исходя из общего решения уравнений электромагнетизма, можно прийти к более ограниченным, но и более подходящим для частной задачи о распространении световых лучей теориям. К ним относится геометрическая оптика, которая служит мощным средством анализа явлений в оптических волокнах и распространения света в световодах, и скалярная теория дифракции, используемая в голографии и пространственной фильтрации.

Указанные теории носят менее общий характер, нежели электромагнитная теория Максвелла. Поэтому нам придется постоянно указывать границы

их применимости. Но зато мы сильно упростим решение задач, перейдя от системы четырех векторных уравнений (Максвелла) к одному векторному (уравнению эйконала) или даже к одному скалярному уравнению (Гюйгенса — Френеля).

В этих первых главах мы также определим средства, которые позволяют возвести здание оптической связи: введем представления о пространственной и временной когерентности источников света, понятия, характеризующие область распространения (моды, потери), и понятия разных волн (однородных и неоднородных, плоских, сферических, гауссовых...).

Глава 5 дополняет теорию дифракции, изложенную в гл. 4, рассмотрением вопросов аналоговой и цифровой голографии и ее применений для хранения информации. Применения голографии для оптической обработки информации рассматриваются в гл. 6 и 7, которые посвящены описанию различных способов обработки информации, реализуемых методами когерентной и некогерентной оптики. В этих двух главах излагаются принципы оптической пространственной фильтрации, а также применения оптики для обработки изображений, автоматического чтения и распознавания образов.

На этом рассмотрение обработки информации в общем завершится. В последних главах мы вернемся к некоторым более общим вопросам обработки информации, вытекающим из нашего рассмотрения оптической передачи информации.

Глава 8 посвящена изучению распространения света в оптических волокнах. Исходя из упрощенной модели, основанной на геометрической оптике, выводятся многие важные характеристики. Далее они будут получены на основе более точной математической теории, позволяющей количественно исследовать различные явления: дисперсию, взаимодействие мод и т. д. В гл. 9 рассматриваются различные методы изготовления многомодовых оптических волокон.

Главы 10 и 11 посвящены, с одной стороны, изготовлению оптических волокон, а с другой — описанию оптических кабелей и соединительных устройств, которые позволят нам осуществлять оптическую связь. В следующих трех главах изучаются источники, приемники и модуляторы света. Изложенный в них материал предназначен главным образом для того, чтобы напомнить в плане оптической связи основные характеристики этих компонентов, принципиально необходимых для реализации любой системы связи.

Далее в гл. 15—17 мы приступаем к задаче проектирования оптических систем связи: выбор способа передачи информации (цифровые или аналоговые), типа кодирования, определение полосы пропускания, выбор основных компонентов системы, структуры сети. Мы будем широко обращаться к величинам, определенным в предшествующих главах, но при этом будем также учитывать стоимость, безопасность использования и надежность. Это позволит нам предложить логическую последовательность определения каждого из элементов — источник излучения, оптический кабель, ретранслятор, фотоприемник — в будущей разработке оптической системы связи.

Две последние главы более всего ориентированы в будущее. Прежде всего в гл. 18 предлагается некоторое расширение возможностей связи путем

временного, хроматического или пространственного уплотнения информации. Таким образом, оптическое волокно перестает играть роль простого канала передачи информации и в ряде случаев становится элементом ее обработки. К вопросам обработки мы обращаемся также в гл. 19. В этой главе показывается, как, используя оптическую обработку информации и вводя в оптическую систему электронные компоненты, можно получить новые системы, более приспособленные к современным методам передачи информации. В качестве примера предельных возможностей такого подхода здесь рассматриваются оптические цифровые вычислительные устройства и оптические коммутаторы для оптических систем связи.

ГЛАВА 1

ОСНОВЫ ЭЛЕКТРОМАГНЕТИЗМА

Данная книга посвящена основам новейших применений оптики в технике связи. Под этим понимаются в основном две области применения: передача информации и ее обработка оптическими методами.

Что касается передачи информации, то речь будет идти в основном о волоконных оптических волноводах (гл. 8 и далее). Что же касается ее обработки, то будут рассмотрены главным образом методы оптической обработки информации, основанные на голографии (гл. 5—7).

Теоретическую основу этих двух применений оптики составляют уравнения Максвелла.

Возьмем для примера волоконные оптические волноводы. Это стеклянные волокна, действующие как световоды. Принцип их действия можно объяснять двумя разными способами. При «электромагнитном», или «модовом», подходе изучают распространение в волокне электромагнитного поля, как и в классической теории волноводов. В приближении «геометрической», или «лучевой», оптики рассматривают зигзагообразное прохождение световых лучей в волокне с многократным отражением от его стенок. Второй подход, очевидно, не независим от первого, и, как увидим, он также основан на уравнениях Максвелла. Это позволит нам указать границы его применимости.

Что же касается голографии, то она базируется на теории дифракции, которая тоже основана на уравнениях Максвелла.

В данной главе мы хотим напомнить читателю фундаментальные уравнения электромагнетизма, которые, как предполагается, ранее были ему известны. Мы ограничимся тем, что кратко изложим классическую теорию и введем ряд понятий, которые необходимы для дальнейшего. Мы ни в коей мере не претендуем на оригинальность в данной области. Но при изложении обращаем особое внимание на те понятия, которые наиболее важны для нас.

§ 1. Уравнения Максвелла, соответствующие физические законы

Электромагнитное поле описывается следующими уравнениями [1]:

Основные уравнения

$$\operatorname{rot} \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t}, \quad (1.1)$$

$$\operatorname{rot} \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}. \quad (1.2)$$

Дополнительные уравнения

$$\operatorname{div} \mathbf{B} = 0, \quad (1.3)$$

$$\operatorname{div} \mathbf{D} = \rho. \quad (1.4)$$

Здесь \mathbf{E} и \mathbf{H} — векторы напряженности поля: \mathbf{E} — вектор напряженности электрического поля (В/м); \mathbf{H} — вектор напряженности магнитного поля (А/м²); \mathbf{D} и \mathbf{B} — векторы индукции, определяющие поведение волны в среде: \mathbf{D} — вектор электрической индукции, или электрического смещения (К/м²); \mathbf{B} — вектор магнитной индукции (Вб/м² или Т).

В заключение напомним закон сохранения заряда

$$\operatorname{div} \mathbf{j} + \frac{\partial \rho}{\partial t} = 0, \quad (1.5)$$

где ρ — объемная плотность электрического заряда (Кл/м³), \mathbf{j} — вектор электрической плотности тока (А/м²), и формулу для силы, действующей на заряд q в электромагнитном поле:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

Эти уравнения хорошо известны. Важно понимать их физический смысл, так как каждое уравнение соответствует вполне определенному физическому закону, который необходимо помнить. Чтобы вывести эти законы, нужно применить теорему Стокса — Остроградского.

Уравнения (1.1) — (1.5) выражают следующие физические законы.

А. Уравнение (1.1), называемое уравнением Максвелла — Фарадея (рис. 1.1)

$$\begin{aligned} \iint_{\Sigma} (\operatorname{rot} \mathbf{E}) \cdot \mathbf{n} \, ds &= \int_C \mathbf{E} \cdot d\mathbf{l} = - \frac{\partial}{\partial t} \iint_{\Sigma} \mathbf{B} \cdot \mathbf{n} \, ds, \\ \int_C \mathbf{E} \cdot d\mathbf{l} &= V = - \frac{\partial \Phi_{\mathbf{B}}}{\partial t}. \end{aligned} \quad (1.6)$$

где $\partial \Phi_{\mathbf{B}} = \iint_{\Sigma} \mathbf{B} \cdot \mathbf{n} \, ds$ — поток магнитной индукции.

Интеграл вектора электрического поля по замкнутому контуру Σ равен электродвижущей силе индукции.

Отсюда вытекает известный закон Ленца: при изменении магнитного потока в электрической цепи возникает «противо-электродвижущая» сила, пропорциональная скорости изменения потока и как бы противодействующая этому изменению.

Соотношение (1.6) позволяет, кстати, выявить одну важную аналогию. Как известно, V выражается в вольтах. Следовательно, напряженность электрического поля E выражается в единицах вольт/метр. В гл. 2, § 1 при изучении эффектов рас-

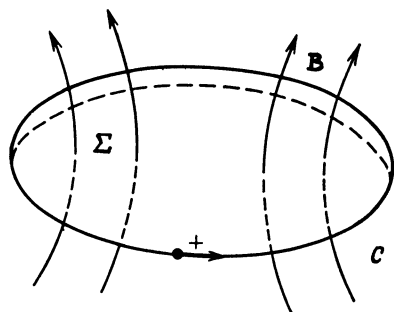


Рис. 1.1. Применение теоремы Стокса — Остроградского для вывода выражений (1.6) и (1.7).

C — замкнутый контур, ограничивающий элемент поверхности Σ , который пересекают линии индукции B .

пространения мы увидим, что, когда рассмотрение распространения волны можно свести к более классическому рассмотрению электрической линии передачи, аналогом напряженности электрического поля E будет потенциал V . Эта аналогия чисто формальная, но она позволяет легко получать уравнения распространения, а также энергетические соотношения из решения очень простой задачи о линии передачи. Эта аналогия такова:

$$\begin{array}{ccc} \text{Напряженность поля } E & \leftrightarrow & \text{Напряжение } V. \\ (\text{вольт/метр}) & & (\text{вольт}) \end{array}$$

Она вытекает из закона Максвелла — Фарадея (1.1).

Б. Уравнение (1.2), называемое уравнением Максвелла—Ампера

Величина $\partial D / \partial t$ имеет размерность плотности тока. Этот ток называют *током смещения*. Плотность *полного* тока дается выражением

$$J = j + \frac{\partial D}{\partial t},$$

где j — так называемый ток *проводимости*. Обозначим через I поток плотности *полного* тока:

$$I = \iint_{\Sigma} J \cdot n \, ds.$$

Его можно представлять себе как полный ток через поверхность Σ ; это то, что называют просто током и измеряют в амперах. Так же как и в § 1, п. А, можно получить следующее соотношение:

$$\int_C \mathbf{H} \cdot d\mathbf{l} = I. \quad (1.7)$$

Интеграл от напряженности магнитного поля по замкнутому контуру равен полному току через этот контур.

В данном случае можно провести следующую аналогию:

Напряженность поля \mathbf{H}	\leftrightarrow	Ток I .
(ампер/метр)		(ампер)

В. Уравнение (1.3), называемое уравнением непрерывности магнитного потока

В случае уравнений, содержащих дивергенцию, нужно рассматривать (рис. 1.2) объем, ограниченный замкнутой поверх-

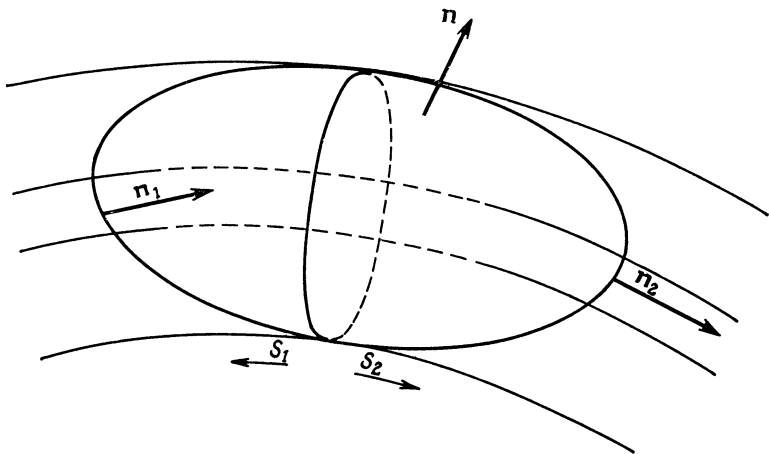


Рис. 1.2. Непрерывность магнитного потока.

ностью S . Получим

$$\iiint_V \operatorname{div} \mathbf{B} d\tau = \iint_S \mathbf{B} \cdot \mathbf{n} ds = 0,$$

или

$$\iint_{S_1} \mathbf{B} \cdot \mathbf{n}_1 ds = \iint_{S_2} \mathbf{B} \cdot \mathbf{n}_2 ds. \quad (1.8)$$

Поток магнитной индукции через замкнутую поверхность равен нулю. Поток, входящий через любую поверхность S_1 , опирающуюся на замкнутый контур C , равен выходящему потоку.

Таким образом, поток магнитной индукции не имеет источника (магнитного заряда).

Г. Уравнение (1.4), называемое теоремой Гаусса

$$\iiint_S \mathbf{D} \cdot \mathbf{n} \, ds = \iiint_V \rho \, d\tau. \quad (1.9)$$

Поток электрической индукции через замкнутую поверхность равен полному заряду в объеме, ограниченном этой поверхностью.

Соотношения (1.8) и (1.9) хорошо показывают единицы измерения величин \mathbf{B} (вебер/метр²) и \mathbf{D} (кулон/метр²).

Д. Уравнение (1.5), называемое уравнением сохранения заряда

$$\iiint_S \mathbf{j} \cdot \mathbf{n} \, ds + \frac{\partial}{\partial t} \iiint_V \rho \, d\tau = 0. \quad (1.10)$$

Заряд сохраняется: при наличии тока проводимости из объема V , ограниченного замкнутой поверхностью, полный заряд, находящийся в объеме V , уменьшается.

Необходимо учитывать, что в уравнение (1.10) входит только ток проводимости, т. е. ток, который соответствует перемещению действительных электрических зарядов.

§ 2. Материальные уравнения

Уравнения (1.1)–(1.5) не независимы: поскольку $\operatorname{div}(\operatorname{rot}) = 0$, из уравнений (1.1), (1.2) и (1.5) следует, что

$$\frac{\partial}{\partial t} (\operatorname{div} \mathbf{B}) = 0, \quad (1.11)$$

$$\frac{\partial}{\partial t} (\operatorname{div} \mathbf{D} - \rho) = 0 \quad (1.12)$$

В условиях, когда все величины зависят от времени и не содержат никаких статических составляющих, уравнения (1.3) и (1.4) вытекают из (1.1), (1.2) и (1.5). Этим объясняется то, что ранее уравнения (1.1) и (1.2) были названы «основными», а (1.3) и (1.4) — «дополнительными».

В трех уравнениях (1.1), (1.2) и (1.5) фигурируют шесть переменных. Среда распространения налагает три дополнительных соотношения, называемые *материальными* уравнениями,

которые дополняют нашу систему до шести уравнений с шестью неизвестными.

Далее будем предполагать, что речь идет об идеальной среде (смысл этого термина будет уточнен далее). Для такой среды материальные уравнения можно представить в виде

$$\mathbf{j} = \sigma \mathbf{E}, \quad (1.13)$$

где σ — удельная проводимость,

$$\mathbf{D} = \epsilon \mathbf{E}, \quad (1.14)$$

где ϵ — диэлектрическая проницаемость,

$$\mathbf{B} = \mu \mathbf{H}, \quad (1.15)$$

где μ — магнитная проницаемость; коэффициенты σ , ϵ и μ пока что считаем действительными величинами.

В случае *диэлектрика* (т. е. непроводящего материала), для которого $\sigma = 0$, вводят так называемый ток смещения

$$\mathbf{j}_D = \frac{\partial \mathbf{D}}{\partial t} = \epsilon \frac{\partial \mathbf{E}}{\partial t}.$$

Из уравнения (1.6) получаем

$$\frac{\partial V}{\partial t} = \int \frac{\partial \mathbf{E}}{\partial t} d\mathbf{l} = \frac{1}{\epsilon} \int \mathbf{j}_D d\mathbf{l} = \frac{1}{C} I,$$

где

$$C = \frac{\epsilon S}{L}.$$

В случае проводника ($\sigma = \infty$) соотношение (1.13) сводится к *закону Ома*:

$$\mathbf{j} = \sigma \mathbf{E}, \quad V = \int \mathbf{E} \cdot d\mathbf{l} = \frac{1}{\sigma} \int \mathbf{j} \cdot d\mathbf{l} = RI,$$

где $R = L/\sigma S$.

В общем случае в соответствии с (1.4) и (1.5) полный ток удовлетворяет соотношению $\operatorname{div} \mathbf{J} = 0$.

Таким образом, полный ток является замкнутым током.

Уточним теперь, что понимается под «идеальной» средой. Рассмотрим три материальных уравнения (1.13) — (1.15). Эти уравнения предполагают выполнение следующих условий.

1. *Локальность*: \mathbf{j} зависит только от напряженности \mathbf{E} в рассматриваемой точке. Следовательно, в пространстве имеется взаимно-однозначное соответствие между \mathbf{j} и \mathbf{E} . Кроме того, предполагается, что среда *гомогенная*, т. е. инвариантная по отношению к изменению начала координат.

2. *Отсутствие дисперсии*: имеется взаимно-однозначное соответствие во времени, например величина $\mathbf{D}(t)$ однозначно определяется величиной \mathbf{E} , взятой в момент времени t .

В § 5 и в гл. 2, § 5 будет рассмотрен более общий случай дисперсной среды, для которой материальные уравнения могут быть записаны в том же виде, что и ранее, но только для одной частоты, т. е. величины σ , ϵ и μ становятся зависящими от частоты. Например:

$$\mathbf{D}(\omega) = \epsilon(\omega) \mathbf{E}(\omega).$$

В этом случае дисперсия выражается через временную свертку.

Далее, уравнения (1.13) и (1.15) соответствуют *перманентной* среде, т. е. среде, инвариантной по отношению к временному сдвигу.

3. *Линейность среды*: любая линейная комбинация напряженностей полей может быть представлена в виде линейной комбинации плотностей тока \mathbf{j} и индукций \mathbf{D} и \mathbf{B} .

4. *Изотропность среды*: материальные уравнения инвариантны по отношению к повороту осей координат (в случае анизотропных сред коэффициенты σ , ϵ и μ будут тензорами).

Для вакуума имеем

$$\sigma = 0,$$

$$\epsilon = \epsilon_0 = 1/36 \pi \cdot 10^{-9} \text{ фарада/метр},$$

$$\mu = \mu_0 = 4 \pi \cdot 10^{-7} \text{ генри/метр}.$$

Таким образом, вакуум есть всего лишь некая диэлектрическая среда. Для идеального диэлектрика (непроводника) мы имеем $\sigma = 0$, а для идеального проводника $\sigma = \infty$. Известно, что в идеальном проводнике электромагнитное поле и плотность зарядов равны нулю.

В материальной среде отличие величин (1.14) и (1.15) от величин, соответствующих вакууму, можно объяснить появлением в данной среде поляризации за счет воздействия внешнего поля (\mathbf{E} , \mathbf{H}):

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \quad (1.16)$$

где \mathbf{P} — вектор электрической поляризации (плотность электрического момента или электрический момент единицы объема);

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}), \quad (1.17)$$

где \mathbf{M} — вектор намагниченности или намагниченность (плотность магнитного момента).

Аналогичным образом определяются и восприимчивости:

$$\mathbf{P} = \chi \epsilon_0 \mathbf{E}, \text{ где } \chi \text{ — абсолютная диэлектрическая восприимчивость,}$$

или коэффициент поляризуемости,

$$\mathbf{M} = \kappa \mathbf{H}, \text{ где } \kappa \text{ — абсолютная магнитная восприимчивость.}$$

Вводятся также относительная диэлектрическая проницаемость $\epsilon_r = \epsilon/\epsilon_0$ и относительная магнитная проницаемость $\mu_r = \mu/\mu_0$.

Поляризация материальных сред особенно интересна для нас, так как она позволяет частично объяснить поглощение света в оптических волноводах. В § 5 будет выведено выражение для мощности, рассеивающейся на диполях. В гл. 2, § 4, п. Б будет рассмотрена микроскопическая модель, объясняющая природу поляризации.

§ 3. Соотношения непрерывности [1, 2]

Из уравнений (1.1) — (1.4) следует, что при наличии поверхности раздела двух сред 1 и 2 [индексы 1 и 2 соответствуют

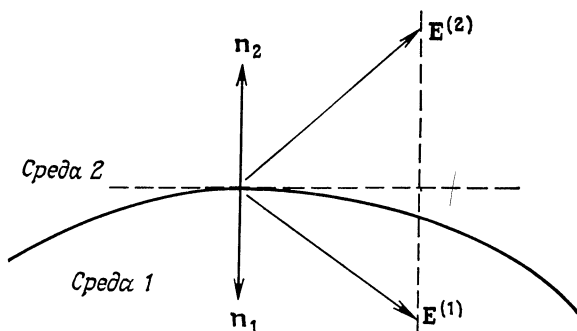


Рис. 1.3. Соотношение непрерывности.

Непрерывность тангенциальной составляющей электрического поля.

двум рассматриваемым средам; \mathbf{n}_1 и \mathbf{n}_2 — нормали к границе раздела соответственно в средах 1 и 2, рис. 1.3]:

1) тангенциальная составляющая вектора \mathbf{E} непрерывна:

$$\mathbf{n}_1 \times \mathbf{E}^{(1)} + \mathbf{n}_2 \times \mathbf{E}^{(2)} = 0; \quad (1.18)$$

2) плотность поверхностного тока выражается через тангенциальные составляющие вектора \mathbf{H} следующим образом:

$$\mathbf{n}_1 \times \mathbf{H}^{(1)} + \mathbf{n}_2 \times \mathbf{H}^{(2)} = \mathbf{j}_s; \quad (1.19)$$

3) нормальная составляющая вектора \mathbf{B} непрерывна:

$$\mathbf{n}_1 \cdot \mathbf{B}^{(1)} + \mathbf{n}_2 \cdot \mathbf{B}^{(2)} = 0; \quad (1.20)$$

4) плотность поверхностного заряда выражается через нормальные составляющие вектора \mathbf{D} следующим образом:

$$\mathbf{n}_1 \cdot \mathbf{D}^{(1)} + \mathbf{n}_2 \cdot \mathbf{D}^{(2)} = \rho_s. \quad (1.21)$$

Теперь мы имеем почти все формулы, необходимые для того, чтобы решать задачу электромагнетизма. Данные задачи таковы:

- 1) источники: распределение зарядов ρ и токов \mathbf{j} ;
- 2) пространственные границы и физические характеристики σ , ϵ и μ изучаемых сред.

Неизвестными являются напряженности полей и индукций во всех точках пространства в функции времени. Уравнения (1.1)—(1.5) и (1.13)—(1.15) вместе с соотношениями непрерывности (1.18)—(1.21) составляют систему уравнений в частных производных с граничными условиями (если исследуемая система имеет бесконечные границы, в нее необходимо добавить условие излучения в бесконечности). Такова постановка задачи. В большинстве практических случаев ее возможно точно решить. Вот почему, для того чтобы найти приближенные, но простые решения, мы должны будем далее принять ряд допущений, которые постепенно приведут нас к приближению геометрической оптики. Первое из таких допущений — гипотеза о гармонических колебаниях.

§ 4. Гармонические колебания, комплексные амплитуды

Каждый знает, что можно представить свет в виде волны. Часто предполагают, что все электрические поля и магнитные индукции изменяются *синусоидально с одинаковой частотой ω* . Такое допущение очень далеко от действительности. Ведь функция $\sin \omega t$ определена от $-\infty$ до $+\infty$ (в противном случае это уже не синусоида). Следовательно, процесс не имеет «начала», он как бы бесконечно долго длился до рассматриваемого момента (нет начала — значит нет причины). Иначе говоря, при такой постановке вопроса опускаются переходные процессы. В то же время процесс длится бесконечно долго, т. е. отсутствует затухание.

Таким образом, синусоидальную волну следует рассматривать только как способ расчета, не связанного с физической сущностью явления. Уравнения Максвелла линейны. Значит, допустимо применение преобразования Фурье. Решением системы уравнений будет линейная комбинация бесконечного числа синусоидальных составляющих. Но каждая синусоидальная волна не существует отдельно, сама по себе. Мы вернемся к этому вопросу далее, когда будут рассматриваться вопросы когерентности (гл. 4, § 4).

Важно четко представлять себе ограниченный характер приближения гармонических колебаний. Остальное — всего лишь вычисления. В произвольной точке пространства действительное

«физическое» поле можно записать в виде

$$\mathbf{E} = \mathbf{E}_0 \cos(\omega t + \varphi).$$

Обычно вводят *комплексную амплитуду* \mathbf{E}_c :

$$\mathbf{E}_c = \mathbf{E}_0 e^{i\varphi}.$$

Так как все операторы линейны, можно в расчетах оперировать только с комплексными амплитудами, а в конце вычислений определить физические поля в виде

$$\mathbf{E} = \operatorname{Re}(\mathbf{E}_c e^{i\omega t}) = \frac{1}{2}(\mathbf{E}_c e^{i\omega t} + \mathbf{E}_c^* e^{-i\omega t}).$$

Можно было бы также написать $\operatorname{Re}(\mathbf{E}_c e^{-i\omega t})$. Выбор знака в показателе экспоненты — простая условность. Мы везде далее будем брать знак плюс.

Уравнения Максвелла в комплексной форме непосредственно вытекают из рассмотренных ранее соотношений:

$$\left. \begin{aligned} \operatorname{rot} \mathbf{E}_c &= -i\omega \mu \mathbf{H}_c, \\ \operatorname{rot} \mathbf{H}_c &= (\sigma + i\omega \epsilon) \mathbf{E}_c, \\ \operatorname{div} \epsilon \mathbf{E}_c &= \rho, \\ \operatorname{div} \mu \mathbf{H}_c &= 0, \end{aligned} \right\} \quad (1.22)$$

$$\rho = \frac{i}{\omega} \operatorname{div}(\sigma \mathbf{E}_c). \quad (1.23)$$

§ 5. Энергетические соотношения, интенсивность света

Электромагнитная энергия определяется выражением

$$W = \iiint (\omega_e + \omega_m) d\tau,$$

где

$$\omega_e = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} = \frac{1}{2} \epsilon \mathbf{E}^2$$

есть плотность электрической энергии, а

$$\omega_m = \frac{1}{2} \mathbf{H} \cdot \mathbf{B} = \frac{1}{2} \mu \mathbf{H}^2 \quad (1.24)$$

есть плотность магнитной энергии.

В данном объеме V электромагнитная энергия W уменьшается:

- 1) за счет излучения (поток вектора Пойнтинга);
- 2) за счет тепловых (джоулевых) потерь (действие поля на свободные заряды). Таким образом,

$$\frac{\partial W}{\partial t} = - \iint_S \mathbf{R} \cdot \mathbf{n} ds - \iiint_V \sigma \mathbf{E}^2 d\tau, \quad (1.25)$$

где \mathbf{R} — вектор Пойнтинга, определяемый формулой

$$\mathbf{R} = \mathbf{E} \times \mathbf{H}. \quad (1.26)$$

В идеально непроводящей среде $\sigma = 0$, и, исходя из (1.25) и (1.26), получаем

$$\operatorname{div} \mathbf{R} + \frac{\partial \omega}{\partial t} = 0, \quad (1.27)$$

энергетическое соотношение, идентичное соотношению (1.5); в нем ω — полная плотность энергии, а вектор Пойнтинга дает полный поток энергии через единицу поверхности (интенсивность излучения):

$$|\mathbf{R}| = I = c\omega, \quad (1.28a)$$

\downarrow
 ватт/м²

\downarrow
 джоуль/м³

Это не что иное, как определение *интенсивности* света¹⁾ в оптике — плотности потока световой энергии. В соотношении

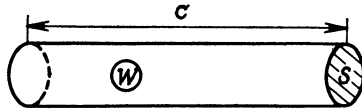


Рис. 1.4. Определение интенсивности света: $I = (1/S) (dW/dt)$.

Это энергия, проходящая через единицу поверхности в 1 с.

(1.28a) выражается тот факт, что это энергия, содержащаяся в цилиндре длиной c (рис. 1.4).

Полезно также записать соотношение (1.28a) в корпускулярном представлении. Для этого введем силу давления фотонов F на поверхность S , т. е. силу давления излучения $\bar{\omega}$:

$$d\omega = Fc dt;$$

тогда поток энергии (через единицу площади поверхности)

$$I = \frac{F}{S} c. \quad (1.286)$$

Это не что иное, как соотношение (1.28a), так как

$$(F/S) = \bar{\omega} = \omega.$$

В самом деле, рассмотрим импульс p потока фотонов:

$$p = \frac{W}{c} = \frac{\omega S c t}{c},$$

откуда

$$F = \frac{dp}{dt} = \omega S, \quad \frac{F}{S} = \omega.$$

¹⁾ Ее мы будем обозначать тем же символом I , что и в полный ток. Это не должно приводить к путанице, так как данные величины носят совершенно разный характер.

Таким образом, мы выразили [формула (1.28)] оптическую величину (интенсивность света) через *электромагнитные* величины (вектор Пойнтинга, плотность электромагнитной энергии).

Отметим, что интенсивность света, как она была нами определена, есть поток «лучистой» энергии и в нее не входят члены с σE^2 , соответствующие джоулевым тепловым потерям в реальной среде.

Но вернемся к энергетическому балансу [формула (1.25)]. Напомним, что в интересующей нас оптической области спектра частоты имеют порядок 10^{15} Гц. Имеющиеся фотоприемники могут воспринимать только среднее во времени значение измеряемых величин.

А. Средняя энергия, идеальный диэлектрик

В синусоидальном приближении выражения для средней энергии запишутся следующим образом (обратить внимание на множители $1/2!$):

$$\langle w_t \rangle = \frac{1}{2} \langle \mathbf{E} \cdot \mathbf{D} \rangle = \frac{1}{8} \langle (\mathbf{E}_c e^{i\omega t} + \mathbf{E}_c^* e^{-i\omega t}) (\mathbf{D}_c e^{i\omega t} + \mathbf{D}_c^* e^{-i\omega t}) \rangle,$$

или, поскольку среднее значение членов с 2ω равно нулю,

$$\begin{aligned} \langle w_e \rangle &= \frac{1}{8} (\mathbf{E}_c \cdot \mathbf{D}_c^* + \mathbf{D}_c \cdot \mathbf{E}_c^*) = \frac{1}{4} \operatorname{Re} (\mathbf{E}_c \cdot \mathbf{D}_c^*), \\ \langle w_e \rangle &= \frac{1}{4} \varepsilon |\mathbf{E}_c|^2, \\ \langle w_m \rangle &= \frac{1}{4} \mu |\mathbf{H}_c|^2. \end{aligned} \tag{1.29}$$

Здесь читателю полезно ознакомиться с аналогией $E \leftrightarrow V$ и $H \leftrightarrow I$ в плане анализа реактивных составляющих. Об этом можно прочитать несколько страниц в книге Фейнмана [1].

Аналогично получаем выражения для вектора Пойнтинга:

$$\langle \mathbf{R} \rangle = \frac{1}{2} \operatorname{Re} (\mathbf{E}_c \times \mathbf{H}_c^*),$$

и в этом случае закон сохранения энергии в идеальной среде записывается на основании равенства (1.27) в виде

$$\int_V \int \operatorname{div} \langle \mathbf{R} \rangle d\tau = \int_S \langle \mathbf{R} \rangle \cdot \mathbf{n} ds = 0,$$

поскольку в силу соотношения (1.29) $(\partial/\partial t) \langle w \rangle = 0$.

В случае идеальной среды в синусоидальном приближении поток Пойнтинга есть постоянная величина.

Б. Реальная среда

Очевидно, что в этом случае последнее утверждение неверно. Из закона сохранения энергии следует, что излучаемая энергия равна джоулевым тепловым потерям:

$$\operatorname{div} \langle \mathbf{R} \rangle + \sigma \langle \mathbf{E}^2 \rangle = 0.$$

В случае обычных материалов плотности энергии (1.24) определяются одинаково:

$$\begin{aligned} \frac{\partial w_e}{\partial t} &= \frac{\partial}{\partial t} \left(\frac{1}{2} \varepsilon_0 \mathbf{E}^2 \right) + \frac{1}{2} \mathbf{E} \cdot \frac{\partial \mathbf{P}}{\partial t}, \\ \frac{\partial w_m}{\partial t} &= \frac{\partial}{\partial t} \left(\frac{1}{2} \mu_0 \mathbf{H}^2 \right) + \frac{\mu_0}{2} \mathbf{H} \cdot \frac{\partial \mathbf{M}}{\partial t}. \end{aligned} \quad (1.30)$$

Энергия поля Энергия, пере-
в вакууме даваемая диполям

Обычные среды являются дисперсионными (рассеивающими). Поэтому следует учитывать частотные зависимости $\varepsilon(\omega)$ и $\mu(\omega)$. Если рассматривать волновой пакет малой ширины со средней частотой $\omega = \omega_0$, то полученные ранее выражения можно преобразовать в эквивалентные частотные соотношения. Для этого нужно использовать первые члены разложения в ряд частотных функций диэлектрической или магнитной проницаемости:

$$\begin{aligned} \varepsilon(\omega) &= \varepsilon(\omega_0) + (\omega - \omega_0) \left[\frac{\partial \varepsilon}{\partial \omega} \right]_{\omega=\omega_0} + \frac{1}{2} (\omega - \omega_0)^2 \left[\frac{\partial^2 \varepsilon}{\partial \omega^2} \right]_{\omega=\omega_0}, \\ \mu(\omega) &= \mu(\omega_0) + (\omega - \omega_0) \left[\frac{\partial \mu}{\partial \omega} \right]_{\omega=\omega_0} + \frac{1}{2} (\omega - \omega_0)^2 \left[\frac{\partial^2 \mu}{\partial \omega^2} \right]_{\omega=\omega_0}. \end{aligned}$$

Предполагается, что потери в среде отсутствуют, т. е. ε и μ — действительные величины в рассматриваемой области. После необходимых выкладок получаются следующие выражения, полный вывод которых можно найти в приложении 2:

$$\begin{aligned} w_e &= \frac{1}{2} \varepsilon(\omega_0) \mathbf{E}_0^2 + \frac{1}{4} \omega_0 \left[\frac{\partial \varepsilon}{\partial \omega} \right]_{\omega=\omega_0} \mathbf{E}_0^2, \\ \langle w_e \rangle &= \frac{1}{4} \mathbf{E}_0^2 \left[\varepsilon(\omega_0) + \omega_0 \left[\frac{\partial \varepsilon}{\partial \omega} \right]_{\omega=\omega_0} \right], \end{aligned}$$

или

$$\begin{aligned} \langle w_e \rangle &= \frac{1}{4} \mathbf{E}_0^2 \left[\frac{\partial (\omega \varepsilon)}{\partial \omega} \right]_{\omega=\omega_0}, \\ \langle w_m \rangle &= \frac{1}{4} \mathbf{H}_0^2 \left[\frac{\partial (\omega \mu)}{\partial \omega} \right]_{\omega=\omega_0}. \end{aligned}$$

ГЛАВА 2

РАСПРОСТРАНЕНИЕ ЭЛЕКТРОМАГНИТНЫХ ВОЛН

В данной главе мы продолжим анализ уравнений Максвелла, который позволит нам объяснить явления распространения электромагнитных волн, представляющие интерес с точки зрения оптической связи. (В частности, мы рассмотрим важные явления поглощения и дисперсии.) Будут введены также основные понятия геометрической, или лучевой, оптики, которые используются в гл. 8 при изучении оптических волноводов.

Отметим, что, хотя электромагнитная теория и позволяет вывести точные уравнения для разных задач, найти решения полученных уравнений не всегда легко. Эти громоздкие и сложные уравнения могут иногда скрывать некоторые физически очевидные соотношения. В то же время очень простое приближение геометрической оптики часто дает результаты, которые мало отличаются от точных решений, даваемых электромагнитной теорией. Электромагнитная же теория использует уж очень сложный математический аппарат сравнительно с оптикой. Поэтому инженеру во многих случаях лучше пользоваться теорией, хотя и приближенной, но позволяющей просто и быстро объяснить явления. Конечно, мы и не думаем противопоставлять *электромагнитной теории геометрическую оптику*. Инженер имеет возможность применять оба этих подхода. Для каждой задачи их можно использовать по своему усмотрению. Важно только четко представлять себе, что ты делаешь. Нужно знать, какие делаются допущения и какие ограничения накладываются при применении того или иного математического аппарата. Геометрическая оптика — это как бы бедовый ребенок электромагнитной теории. Она выводится из нее и позволяет решать многие задачи, но не следует ей полностью доверять. А если мы выходим за пределы применимости геометрической оптики, то необходимо уметь применить электромагнитную теорию.

Именно поэтому мы познакомим здесь читателя с некоторыми необходимыми основными понятиями оптики. Они будут введены на базе электромагнитной теории, что позволит каждый раз проводить связь между двумя теориями.

Данная глава посвящена главным образом свойствам плоских волн. Но мы увидим, что понятие плоской волны имеет не более физического смысла, чем понятие гармонических колебаний, введенное в гл. 1. Лишь в дальнейших главах мы по-

дойдем к типам волн, более точно описывающим действительность: модам диэлектрических волноводов (гл. 3, § 3), гауссовым пучкам (гл. 4, § 3), излучению лазеров (гл. 12).

§ 1. Волновые уравнения, плоские волны

В данном параграфе мы будем рассматривать идеальную диэлектрическую среду ($\sigma = 0$), предполагая, что источники отсутствуют ($\rho = 0, \mathbf{j} = 0$).

Из уравнений (1.1)—(1.4) легко получить следующие *уравнения распространения* волны:

$$\begin{aligned}\Delta \mathbf{E} - \epsilon \mu \frac{\partial^2 \mathbf{E}}{\partial t^2} &= 0, \\ \Delta \mathbf{H} - \epsilon \mu \frac{\partial^2 \mathbf{H}}{\partial t^2} &= 0.\end{aligned}\quad (2.1)$$

В приближении гармонических колебаний, исходя из уравнений (1.22), для случая идеального диэлектрика получаем

$$\begin{aligned}\Delta \mathbf{E}_c + \frac{\omega^2}{v_\phi^2} \mathbf{E}_c &= 0, \\ \Delta \mathbf{H}_c + \frac{\omega^2}{v_\phi^2} \mathbf{H}_c &= 0,\end{aligned}\quad (2.2)$$

где

$$v_\phi = \frac{1}{\sqrt{\epsilon \mu}}. \quad (2.3)$$

Уравнения (2.2) очень легко решить методом разложения в бесконечный ряд Фурье. Положим

$$\mathbf{E}_c(\mathbf{r}) = \int \int \int_{-\infty}^{+\infty} \mathbf{E}_0(\mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{k},$$

где $d\mathbf{k}$ — элемент объема в пространстве \mathbf{k} .

Подставив это выражение в (2.2), получим соотношение

$$\mathbf{k}^2 = \frac{\omega^2}{v_\phi^2}, \quad (2.4)$$

или

$$\mathbf{k}^2 = \omega^2 \epsilon \mu,$$

которое называется *дисперсионным уравнением*. Позднее мы остановимся на нем подробнее.

Таким образом, мы приходим к понятию *плоской гармонической волны*. По определению плоская гармоническая волна записывается в виде

$$\mathbf{E}_c = \mathbf{E}_0 e^{-i\mathbf{k} \cdot \mathbf{r}}, \quad \mathbf{H}_c = \mathbf{H}_0 e^{-i\mathbf{k} \cdot \mathbf{r}}, \quad (2.5)$$

где \mathbf{E}_0 , \mathbf{H}_0 и \mathbf{k} — постоянные векторы. Вектор \mathbf{k} называется *волновым вектором*. Он равен пространственной частоте волны, являющейся пространственным аналогом временной частоты ω . Если волна имеет пространственный период λ , то модуль волнового вектора \mathbf{k} равен

$$|\mathbf{k}| = \frac{2\pi}{\lambda};$$

его обычно называют *волновым числом*.

Чтобы точнее определить вектор \mathbf{k} , подставим выражения (2.5) в систему уравнений (1.22). Получим

$$\begin{aligned} \mathbf{k} \times \mathbf{E}_0 &= \omega \mu \mathbf{H}_0, \\ \mathbf{k} \times \mathbf{H}_0 &= -\omega \varepsilon \mathbf{E}_0, \end{aligned} \quad (2.6)$$

$$\begin{aligned} \mathbf{k} \cdot \mathbf{E}_0 &= 0, \\ \mathbf{k} \cdot \mathbf{H}_0 &= 0. \end{aligned} \quad (2.7)$$

Таким образом, три вектора \mathbf{E}_0 , \mathbf{H}_0 и \mathbf{k} образуют прямой трехгранный угол. Исключив вектор \mathbf{H}_0 из двух уравнений (2.6), мы определим модуль вектора \mathbf{k} другим способом. В результате снова приходим к уравнению дисперсии (2.4).

Фаза плоской волны (включая член с ωt) равна

$$\varphi(\mathbf{k}, \mathbf{r}) = \omega t - \mathbf{k} \cdot \mathbf{r},$$

а *фазовая скорость* волны (иначе говоря, скорость, с которой нужно перемещаться, чтобы видеть постоянную фазу) определяется выражением

$$v_\varphi = \frac{\omega}{k}.$$

Но, как нетрудно видеть, модули векторов \mathbf{E}_0 и \mathbf{H}_0 связаны соотношением

$$\frac{E_0}{H_0} = \frac{\omega \mu}{k} = \frac{k}{\omega \varepsilon} = \sqrt{\frac{\mu}{\varepsilon}} = Z. \quad (2.8a)$$

Воспользовавшись определением показателя преломления среды, получим

$$v_\varphi = \frac{c}{n},$$

где c — скорость (фазовая) плоской волны в вакууме.

Приведенные соотношения позволяют записать показатель преломления среды следующим образом:

$$n = \sqrt{\frac{\varepsilon}{\varepsilon_0}} = \frac{k}{k_0} = \frac{\lambda_0}{\lambda},$$

где $k_0 = 2\pi/\lambda_0$ — волновое число в вакууме. Тогда выражение (2.8а) можно представить в виде

$$Z = \frac{Z_0}{n}, \quad (2.86)$$

где $Z_0 = \sqrt{\mu_0/\epsilon_0}$ — волновое сопротивление вакуума.

А. Плоская волна и линия передачи

Вспомним аналогии $E \leftrightarrow V$ и $H \leftrightarrow I$ (гл. 1, § 1). Тогда очевидно, что величина Z аналогична сопротивлению; поэтому ее называют *волновым сопротивлением*. Эта очень важная величина далее будет широко использоваться. Она играет точно такую же роль, как и волновое сопротивление линии передачи.

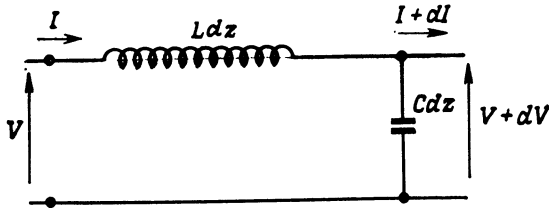


Рис. 2.1. Линия передачи.

Четырехполюсник, эквивалентный линии передачи без потерь.

Рассмотрим линию передачи (без потерь), схематически представленную на рис. 2.1, где L и C — погонные индуктивность и емкость линии передачи. Легко показать [3], что дифференциальные уравнения в частных производных для $V(t, z)$ и $I(t, z)$ имеют точно такой же вид, как и уравнения (2.1), а именно:

$$\begin{aligned} \frac{\partial^2 V}{\partial z^2} - LC \frac{\partial^2 V}{\partial t^2} &= 0, \\ \frac{\partial^2 I}{\partial z^2} - LC \frac{\partial^2 I}{\partial t^2} &= 0. \end{aligned} \quad (2.9)$$

В приближении гармонических колебаний их решение запишется в виде

$$\begin{aligned} V &= Ae^{i\omega(t - \frac{z}{v})} + Be^{i\omega(t + \frac{z}{v})}, \\ I &= \frac{1}{Z_c} \left(Ae^{i\omega(t - \frac{z}{v})} + Be^{i\omega(t + \frac{z}{v})} \right), \end{aligned}$$

где A и B — постоянные. Таким образом, это решение представляет собой линейную суперпозицию двух бегущих волн,

распространяющихся в противоположных направлениях с фазовой скоростью

$$v_{\varphi} = \frac{1}{\sqrt{LC}}, \quad (2.10)$$

Величина Z_c есть волновое сопротивление линии, т. е. отношение напряжения к току в режиме бегущей волны:

$$Z_c = \frac{V(t, z)}{I(t, z)} = \sqrt{\frac{L}{C}}. \quad (2.11)$$

Мощность, распространяющаяся вдоль линии (также в приближении гармонических колебаний, в режиме бегущей волны), будет равна

$$P = \frac{1}{2} Z_c I_0^2 = \frac{1}{2} \frac{V_0^2}{Z_c}. \quad (2.12)$$

Найдем еще средние значения электрической и магнитной энергии для случая плоской гармонической волны. Получаем

$$\langle w_e \rangle = \frac{1}{4} \varepsilon |E_0|^2 = \langle w_m \rangle = \frac{1}{4} \mu |H_0|^2.$$

Следовательно,

$$\begin{aligned} w_{\Sigma} &= \frac{1}{2} \varepsilon |E_0|^2, \\ w_{\Sigma} &= \frac{1}{2} \mu |H_0|^2. \end{aligned} \quad (2.13)$$

Найдем и среднее значение вектора Пойнтинга:

$$\begin{aligned} \langle \mathbf{R} \rangle &= \frac{1}{2} \operatorname{Re} (\mathbf{E}_0 \times \mathbf{H}_0^*) = \frac{1}{2} \operatorname{Re} \left(\frac{|\mathbf{E}_0|^2}{Z} \right), \\ \langle \mathbf{R} \rangle &= \frac{1}{2} \operatorname{Re} (Z |\mathbf{H}_0|^2), \end{aligned} \quad (2.14)$$

$$\langle \mathbf{R} \rangle = v_{\varphi} \langle w_{\Sigma} \rangle. \quad (2.15)$$

Соотношения (2.1), (2.3), (2.8) и (2.14) и соответствующие соотношения (2.9)—(2.12) ярко подчеркивают формальную аналогию между распространением плоской волны в диэлектрической среде и волн тока и напряжения в линии передачи.

Б. Характеристики плоской оптической волны, реальность этого понятия

Итак, мы имеем плоскую распространяющуюся волну. Подчеркнем, что в оптическом диапазоне частоты — порядка 10^{15} Гц, а длины волн — порядка 0,5 мкм.

Значения этих двух величин показывают большие возможности, которые можно реализовать как в области связи, так и в области хранения и обработки информации. Но, прежде чем перейти к соответствующим техническим приложениям, вернемся к вопросу о реальности понятия плоской волны [выражение (2.5)]. Здесь можно сказать то же, что было сказано в гл. 1, § 4 относительно приближения гармонических колебаний. Каким образом фаза может оставаться постоянной на бесконечной плоскости? Наблюдал ли кто-нибудь когда-либо бесконечную плоскость? Может ли фаза распространяться с постоянной скоростью в любой точке пространства? Анализ таких соображений, к которым мы вернемся при рассмотрении когерентности (гл. 4, § 4), приводит к следующему выводу.

Плоская волна в действительности не существует. Это лишь удобное математическое представление. В природе можно встретить лишь слабую имитацию этой теоретической модели.

Действительно, поскольку уравнения распространения линейны, общее решение уравнения (2.1) представляется в виде линейной комбинации плоских волн. Лишь разложение в бесконечный ряд Фурье даст всему некоторый физический смысл, но не плоской волне, а тому, что называется *спектром плоских волн*. В этом случае действительная волна $E(r, t)$ может быть представлена в виде

$$E(r, t) = \int_{-\infty}^{+\infty} e^{i\omega t} d\omega \iiint_{-\infty}^{+\infty} \underbrace{E_0(\omega, \mathbf{k})}_{\mathcal{E}(\omega, \mathbf{k})} e^{i\varphi(\omega, \mathbf{k})} e^{-i\mathbf{k} \cdot \mathbf{r}} d\tau.$$

Введем в рассмотрение неоднородную гармоническую волну, что позволит уточнить условие «существования» плоской волны.

§ 2. Гармоническая волна, приближение геометрической оптики

Рассмотрим случай неоднородной среды. Диэлектрическая проницаемость ϵ , или, что эквивалентно, показатель преломления n может изменяться в зависимости от радиуса r . (Напомним, например, что оптические волноводы — это стеклянные волокна, показатель преломления которых изменяется по радиусу.)

А. Гармоническая волна

Выше было показано, что решение уравнений распространения в однородной среде приводит к линейной зависимости фазы от пространственных координат (плоская волна).

В случае неоднородной среды мы будем искать несколько более общее решение системы (1.22) в таком виде:

$$\begin{aligned} \mathbf{E}_c &= \mathbf{E}_0(r) e^{-ik_0 S(r)}, \\ \mathbf{H}_c &= \mathbf{H}_0(r) e^{-ik_0 S(r)}, \end{aligned} \quad (2.16)$$

где $k_0 = 2\pi/\lambda_0 = \omega/c$ — величины, соответствующие вакууму.

Речь идет о *гармонической волне*. Величина $S(\mathbf{r})$ часто называется *эйконалом*. *Эквифазные*, или *волновые*, поверхности, называемые также *волновыми фронтами*, — это совокупности точек, удовлетворяющие условию

$$S(\mathbf{r}) = \text{const},$$

тогда как *поверхности равных амплитуд* будут определяться соотношениями

$$\begin{aligned} |\mathbf{E}_0(\mathbf{r})| &= \text{const}, \\ |\mathbf{H}_0(\mathbf{r})| &= \text{const}. \end{aligned}$$

Что касается *эквиэнергетических* поверхностей, то они определяются средним значением вектора Пойнтинга (напомним, что мы рассматриваем случай идеального диэлектрика):

$$\begin{aligned} \langle \mathbf{R} \rangle &= \frac{1}{2} \text{Re} (\mathbf{E}_c \times \mathbf{H}_c^*), \\ \langle \mathbf{R} \rangle &= \frac{1}{2} \text{Re} [\mathbf{E}_0(r) \times \mathbf{H}_0(r)] = \text{const}. \end{aligned}$$

Разумеется, *эквиэнергетические* поверхности не совпадают с *эквифазными* поверхностями, причем как те, так и другие априори не плоскости.

Б. Геометрическая оптика, уравнение эйконала

Подставим выражения (2.16) в систему (1.22). Мы не приводим подробностей расчета, которые можно найти в известных работах [4]. В предположении, что $\mathbf{k}(\mathbf{r}) = k_0 \text{grad}(S)$, получаем

$$\begin{aligned} \mathbf{k} \times \mathbf{E}_0 - \omega \mu \mathbf{H}_0 &= -i \text{rot } \mathbf{E}_0, \\ \mathbf{k} \times \mathbf{H}_0 + \omega \epsilon \mathbf{E}_0 &= -i \text{rot } \mathbf{H}_0. \end{aligned} \quad (2.17)$$

Допустим, что $\mathbf{E}_0(r)$ и $\mathbf{H}_0(r)$ не изменяются. В этом случае вновь получаем уравнения (2.6) и (2.7). Их решением будет плоская гармоническая волна, волновой вектор $\mathbf{k}(\mathbf{r})$ — постоянной величиной, волновые и *эквиэнергетические* поверхности будут плоскими и совпадающими друг с другом и т. д.

Но при каком условии можно допустить существование плоской волны в действительности? Поставим вопрос по-другому: при каком условии можно преобразовать (2.17) в (2.6)? Очевидно, что для этого необходимо отбросить правые члены

в (2.17). Поскольку ротор вектора содержит только пространственные производные, можно предположить, что векторы $\mathbf{H}_0(\mathbf{r})$ и $\mathbf{E}_0(\mathbf{r})$ изменяются сравнительно медленно. Исследуем подробнее одно из соотношений (2.17). Мы примем, что величина $\text{rot } \mathbf{E}_0$ пренебрежимо мала. Но по сравнению с чем? С $\mathbf{k} \times \mathbf{E}_0$ или с $\omega \mathbf{H}_0$, которые являются величинами одного порядка. Следовательно, предположение состоит в том, что

$$|\text{rot } \mathbf{E}_0| \ll |k_0 \mathbf{E}_0|. \quad (2.18a)$$

Это соотношение справедливо, если $\omega \rightarrow \infty$, или $\lambda_0 \rightarrow 0$, или $k_0 \rightarrow \infty$. Но всегда необходимо указывать, относительно чего определяются бесконечно малые или бесконечно большие величины. Обозначим через ΔE изменение напряженности поля на интервале Δx . Тогда условие (2.18a) приводит к допущению:

$$\begin{aligned} \frac{\Delta E}{E} &\ll \frac{\Delta x}{\lambda_0}, \\ \frac{\Delta H}{H} &\ll \frac{\Delta x}{\lambda_0}. \end{aligned} \quad (2.18б)$$

Иначе говоря, относительные изменения амплитуды напряженностей поля должны быть малы по сравнению с размерами рассматриваемой системы, измеряемыми в длинах волн.

Это и понятно, так как длина волны — хорошая мера для пространственного измерения.

Рассмотрим теперь этот же вопрос по-другому. Как правило, гармоническая волна имеет неплоские волновые поверхности. Нормаль к поверхности фронта волны $S(\mathbf{r}) = \text{const}$ задается вектором $\text{grad } S / |\text{grad } S|$. Предположение (2.18б) эквивалентно утверждению, что волна плоская в некоторой части пространства Δx .

Таким образом, на каждом малом участке волновую поверхность можно заменить частью плоскости, касательной к ней в рассматриваемой точке (рис. 2.2). В результате приходим к теории *световых лучей*, направленных по нормали к волновым поверхностям, т. е. к *геометрической оптике*. Величиной $\text{grad } S$ определяется световой луч в каждой точке волновой поверхности, и поэтому основным уравнением геометрической оптики является *уравнение эйконала*, которое получается из (2.17) и (2.7), когда условие (2.18) выполняется локально:

$$|\text{grad } S|^2 = c^2 \epsilon \mu = n^2, \quad (2.19)$$

где n — показатель преломления среды: $n = c/v_\phi = c \sqrt{\epsilon \mu}$.

Нужно хорошо усвоить, что физический смысл плоской волны и физический смысл световых лучей (иначе говоря, геометрической оптики) один и тот же.

Все это в сущности лишь математические понятия, некие модели физических процессов. Одна модель — плоская волна, другая — «точечный источник» света. Действительно, предположить существование плоской волны — это предположить существование пучка параллельных лучей и, следовательно, наличие точечного источника света в бесконечности. Таково же понятие «светового луча», математически определяемого направлением нормали к волновой поверхности.

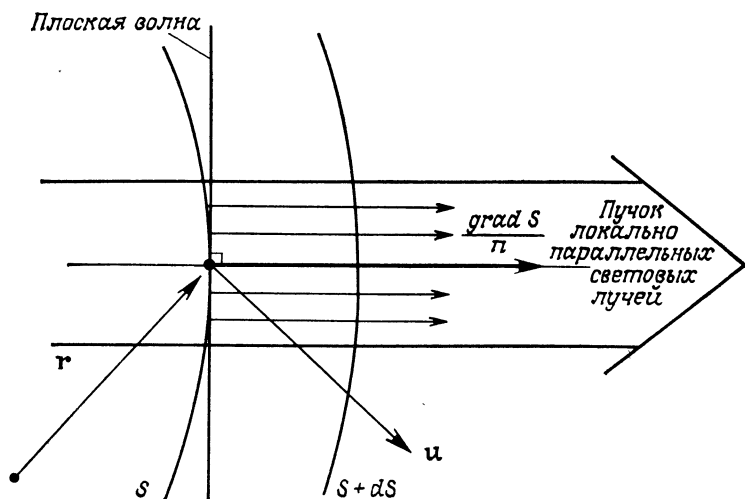


Рис. 2.2. Локальная аппроксимация волновой поверхности плоской волной.

Каждый малый элемент волновой поверхности заменяется касательной плоскостью. Локальный пучок, направленный по нормали к касательной плоскости, можно рассматривать как световой луч с макроскопической точки зрения

Но локально мы всегда имеем право заменить волновую поверхность касательной плоскостью. И тогда легко найти параметры плоской волны. Ее полная фаза (включая член с ωt) будет равна

$$\varphi(t, \mathbf{r}) = \omega t - k_0 S(\mathbf{r}).$$

Ее фазовая скорость в направлении единичного вектора \mathbf{u} определяется из соотношения

$$\omega dt = k_0 \text{grad } S \cdot \mathbf{u} dl = 0,$$

где l — криволинейная координата, измеряемая вдоль светового луча. В результате получаем

$$v_\varphi = \frac{dl}{dt} = \frac{c}{u \text{ grad } S}.$$

С другой стороны, в рамках сделанных допущений (среда не-дисперсная) вектор Пойнтинга будет равен

$$\langle \mathbf{R} \rangle = \frac{1}{2} \operatorname{Re} (\mathbf{E}_0 \times \mathbf{H}_0^*) = \frac{1}{2} Z_0 \mathbf{H}_0^2 \frac{\operatorname{grad} S}{n}$$

$$\left(Z_0 = \frac{\omega \mu}{k_0} \right).$$

И локальная аппроксимация плоской волной (рис. 2.2) позволяет вновь получить соотношение (2.19) и законы распространения лучей, которые из него вытекают. К тому же уравнение эйконала (2.19) — это не что иное, как дисперсионное уравнение (2.4) с тем отличием, что в уравнении эйконала рассматривают векторы распространения, являющиеся функциями радиус-вектора r . Поэтому очевидно, что уравнение эйконала применимо в неоднородном пространстве, и в частности в оптических волноводах.

В. Уравнение лучей

Величина $\operatorname{grad} S$ изменяется в зависимости от неоднородностей пространства, в котором распространяется волна, что приводит к «искривлению» световых лучей. Это вытекает из самого определения градиента

$$n \frac{\partial \mathbf{r}}{\partial l} = \operatorname{grad} S. \quad (2.20)$$

Читатель может сам (с помощью книги [4]) продифференцировать обе части равенства (2.20) по l . После применения уравнения эйконала (2.19) получим уравнение лучей

$$\frac{\partial}{\partial l} \left(n \frac{\partial \mathbf{r}}{\partial l} \right) = \operatorname{grad} n, \quad (2.21)$$

решениями которого, если пространство однородное, являются только прямые. Зато в неоднородном пространстве это соотношение позволит нам оценивать кривизну луча по формуле

$$\mathbf{C} = \frac{\partial^2}{\partial l^2} \mathbf{r},$$

откуда

$$n \mathbf{C} = \operatorname{grad} n - \frac{\partial n}{\partial l} \frac{\partial \mathbf{r}}{\partial l}. \quad (2.22)$$

Это уравнение нам пригодится, когда речь пойдет об оптических волноводах с градиентом показателя преломления (гл. 4, § 3).

Как видно из вышесказанного, касательная плоскость определяется через градиент показателя преломления среды (рис. 2.3). Этим, в частности, объясняется явление «миража»

(напомним, что в пустыне поверхность земли очень нагрета, а показатель преломления воздуха зависит от его плотности) и то, что лучи искривляются в сторону возрастания показателя преломления.

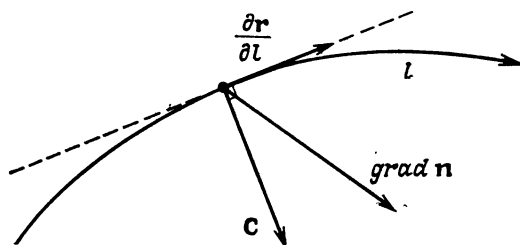


Рис. 2.3. Распространение «луча» в касательной плоскости.

Посмотрим теперь, как распространяется волна в дисперсной среде.

§ 3. Проникновение поля в частично проводящую среду, неоднородные волны

Выше мы рассмотрели гармонические волны, волновые поверхности которых не обязательно плоские и которые изменяются в пространстве из-за неоднородностей среды. Следовательно, это неоднородные волны. В данном параграфе мы рассмотрим другой пример неоднородных волн, распространяющихся *в однородной, но частично проводящей среде*. (Мы имеем в виду распространение волн в оптических волноводах, сделанных из стекла, которое, к сожалению, не является идеальным диэлектриком.)

Во всем этом параграфе мы будем рассматривать случай *однородной частично проводящей среды* (проводимость $\sigma \neq 0$) без источников ($\rho = 0$, $\mathbf{j} = 0$). Как будет распространяться световая волна в такой среде?

А. Введение комплексных констант

Введем эквивалентную диэлектрическую проницаемость

$$\epsilon_{\text{ЭКВ}} = \epsilon + \frac{\sigma}{i\omega}. \quad (2.23)$$

Тогда уравнения (1.22) можно записать в виде уравнений, соответствующих непроводящей среде:

$$\text{rot } \mathbf{E}_c = -i\omega \mu \mathbf{H}_c,$$

$$\text{rot } \mathbf{H}_c = i\omega \epsilon_{\text{ЭКВ}} \mathbf{E}_c,$$

$$\text{div } \epsilon_{\text{ЭКВ}} \mathbf{E}_c = 0,$$

$$\text{div } \mu \mathbf{H}_c = 0.$$

Заодно отметим, что, согласно (2.23), отнесение реального материала к «проводникам» или «диэлектрикам» зависит от частоты¹⁾. На бесконечно больших частотах материал всегда ведет себя как диэлектрик, а на очень низких — как проводник. Проводящая ли среда или диэлектрическая, это теоретически зависит от отношения

$$i\eta = \frac{\text{Ток смещения}}{\text{Ток проводимости}} = \frac{i\omega\epsilon}{\sigma}. \quad (2.24)$$

Для нас, оптиков, имеющих дело с очень высокими частотами, это хорошо в отношении потерь при распространении. Но дальше (§ 2, п. Г) увидим, что в действительности явления поглощения сложнее, чем описано здесь.

Решим уравнения Максвелла, введя в рассмотрение, как и в (2.16), волну вида

$$\begin{aligned} \mathbf{E}_c(r) &= \mathbf{E}_0 e^{-ik \cdot \mathbf{S}(r)}, \\ \mathbf{H}_c(r) &= \mathbf{H}_0 e^{-ik_0 \mathbf{S}(r)}, \end{aligned} \quad (2.25)$$

для которой априорно выбран простой случай, когда векторы \mathbf{E}_0 и \mathbf{H}_0 не изменяются, но принимают комплексные значения. В этом случае получим следующее выражение для вектора $\mathbf{S}(r)$:

$$\mathbf{S}(r) = \mathbf{S}'(r) - i\mathbf{S}''(r),$$

который также будет комплексным. Конечно, это видоизменит определение (гл. 2, § 2, п. А) эквифазных и эквиамплитудных поверхностей, в котором вектор $\mathbf{S}(r)$ предполагался действительной величиной. Поэтому рассмотрим эту задачу вновь.

Используемые нами операторы линейны. Положим

$$\begin{aligned} \mathbf{k}(r) &= k_0 \text{grad } S, \\ \mathbf{k}(r) &= k_0 \text{grad } S' - ik_0 \text{grad } S'', \\ \mathbf{k}(r) &= \mathbf{k}' - i\mathbf{k}'', \quad \text{где } \mathbf{k}' \text{ и } \mathbf{k}'' \text{ — мнимые величины.} \end{aligned}$$

И мы возвращаемся к уравнениям (2.17), но теперь \mathbf{k} — комплексная величина, а величина ϵ заменена величиной $\epsilon_{\text{экр}}$.

Как было сказано, ранее мы искали решения, для которых \mathbf{E}_0 и \mathbf{H}_0 — постоянные векторы, и сейчас фактически получили те же уравнения (2.6) и (2.7), но с комплексными \mathbf{k} , \mathbf{E}_0 и \mathbf{H}_0 .

Б. Неоднородные волны и уравнение дисперсии

Приведенные выше выражения описывают *неоднородную волну*, т. е. такую волну, в которой фазы и амплитуды распространяются неодинаково. Действительно, экспоненциальные

¹⁾ Для простоты мы часто будем называть частотой круговую частоту ω , как это нередко делается.

члены в выражениях (2.16) и (2.25) в первом приближении записываются в виде

$$e^{-ik \cdot r} = e^{-k'' \cdot r - ik' \cdot r}.$$

Это означает, что в данном случае распространение волны характеризуется двумя векторами: вектором $\mathbf{k}' = k_0 \text{grad } S'$, который локально определяет волновую поверхность и фазовую скорость, и вектором $\mathbf{k}'' = k_0 \text{grad } S''$, который локально определяет изменение амплитуды. Такой тип волны часто встречается в оптических волноводах (гл. 3, 8 и 14).

Что касается уравнения дисперсии, то оно легко выводится из уравнений (2.17).

В самом деле,

$$|\mathbf{k}|^2 = \omega^2 \epsilon_{\text{ЭКВ}}, \quad (2.26)$$

откуда

$$k = \pm \omega \sqrt{\mu \epsilon_{\text{ЭКВ}}}.$$

В. Энергия излучения и эффект Джоуля

Посмотрим, как распространяется энергия неоднородной волны. Определим вектор Пойнтинга. Для этого воспользуемся соотношением (2.6):

$$\langle \mathbf{R} \rangle = \frac{1}{2} \text{Re} \langle \mathbf{E}_c \times \mathbf{H}_c^* \rangle,$$

$$\langle \mathbf{R} \rangle = \frac{1}{2} e^{-2k'' \cdot r} \text{Re} (\mathbf{E}_0 \times \mathbf{H}_0^*).$$

Но прежде следует обобщить обозначения § 1, п. Г, поскольку теперь векторы \mathbf{E}_0 , \mathbf{H}_0 и \mathbf{k} априорно являются комплексными величинами.

Чтобы обойтись без скучных расчетов, рассмотрим простой частный случай: предположим, что \mathbf{E}_0 будет действительной величиной. Тогда (рис. 2.4), согласно (2.7), вектор \mathbf{E}_0 будет перпендикулярен векторам \mathbf{k}' и \mathbf{k}'' . И вектор \mathbf{H}_0 разложится на два вектора:

$$\mathbf{H}_0 = \frac{1}{\omega \mu} [\mathbf{H}'_0 - i \mathbf{H}''_0],$$

где $\mathbf{H}'_0 = \mathbf{k}' \times \mathbf{E}_0$, $\mathbf{H}''_0 = \mathbf{k}'' \times \mathbf{E}_0$ — действительные величины.

Вектор Пойнтинга также можно разложить на два вектора:

$$\langle \mathbf{R} \rangle = \frac{1}{2} e^{-2k'' \cdot r} [\text{Re} (\mathbf{E}_0 \times \mathbf{H}'_0) - \text{Re} (\mathbf{E}_0 \times i \mathbf{H}''_0)]. \quad (2.27)$$

Поскольку вектор \mathbf{E}_0 по предположению действительный, получаем $\mathbf{E}_0 \cdot \mathbf{k}' = \mathbf{E}_0 \cdot \mathbf{k}'' = 0$. Нетрудно убедиться, что только первый член в (2.27) дает ненулевой вклад. Итак, получаем

$$\langle \mathbf{R} \rangle = \frac{1}{2\omega \mu} e^{-2k'' \cdot r} |\mathbf{E}_0|^2 \mathbf{k}'. \quad (2.28)$$

Рассмотрим энергию, теряемую из-за эффекта Джоуля. Так как $\mathbf{j} = \sigma \mathbf{E}$, линии тока направлены вдоль \mathbf{E}_0 . Рассчитаем эту энергию в цилиндрическом объеме, ось которого перпендикулярна вектору \mathbf{k}'' (элементарный объем с размерами h , l , dr , рис. 2.5):

$$dW_I = \frac{1}{2} \left(\frac{1}{\sigma} \frac{h}{l dr} \right) \sigma^2 E_0^2 (l dr)^2 e^{-2\mathbf{k}'' \cdot \mathbf{r}},$$

$$dW_I = \frac{1}{2} \sigma E_0^2 e^{-2\mathbf{k}'' \cdot \mathbf{r}} hl dr.$$

Таким образом, частично проводящая среда является *диссипативной* средой, которая поглощает энергию в силу эффекта Джоуля.

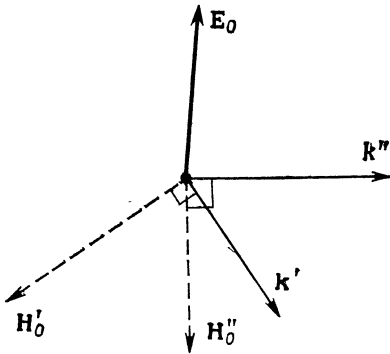


Рис. 2.4. Случай неоднородной волны с действительной величиной E_0 . Волна (E_0, H'_0) соответствует излучаемой энергии.

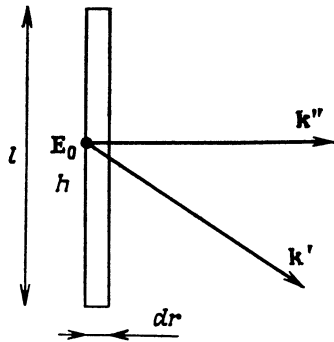


Рис. 2.5. Неоднородная волна. Расчет энергии в элементарном объеме. Излучаемая энергия распространяется вдоль вектора \mathbf{k}' .

Теперь исследуем в том же самом элементарном объеме уменьшение передаваемой энергии. Изменение потока вектора Пойнтинга, согласно (2.28), будет определяться выражением

$$dW_R = \frac{1}{2\omega\mu} E_0^2 e^{-2\mathbf{k}'' \cdot \mathbf{r}} hl \left(\mathbf{k}' \cdot \frac{\mathbf{k}''}{k''} \right) 2k'' dr.$$

Ничто не теряется, ничего не возникает. Сравнение двух предшествующих выражений приводит к уравнению дисперсии (2.26)

$$2\mathbf{k}' \cdot \mathbf{k}'' = \sigma\omega\mu.$$

Итак, мы показали, что неоднородная волна, рассмотренная ранее, может быть представлена в виде суммы двух волн:

1) плоской волны E_0, H'_0 , которая распространяется в направлении вектора \mathbf{k}' и переносит поток *лучистой энергии*, который определяется вектором Пойнтинга;

2) волны \mathbf{E}_0 , \mathbf{H}_0'' , которая не вносит никакого вклада в вектор Пойнтинга из-за сдвига фазы на $\pi/2$ между векторами электрической и магнитной напряженностей поля.

Вследствие диссипации энергии в среде энергия волны уменьшается в $e^{-2\mathbf{k}'' \cdot \mathbf{r}}$ раз в направлении \mathbf{k}'' . Уменьшение передаваемой энергии равно энергии, теряемой за счет эффекта Джоуля.

Мы привели пример распространения неоднородной волны в среде с потерями. Но далее мы увидим (§ 3, п. В), что такие волны могут также существовать в диэлектрических средах. В этом случае их существование следует из учета условий согласования на границах различных сред (явление полного внутреннего отражения, гл. 3, § 1, п. В; оптические волокна, гл. 8 и 9; интегральная оптика, гл. 14, § 3).

Г. Проникновение волны в рассеивающую среду

Вернемся снова к рассеивающей среде. Чтобы рассмотреть проникновение волны в такую среду, предположим для простоты, что векторы \mathbf{k}' и \mathbf{k}'' параллельны. В этом случае уравнение дисперсии (2.6) сведется к следующей системе двух уравнений с двумя неизвестными:

$$\begin{aligned} k'^2 - k''^2 &= \omega^2 \epsilon, \\ 2k'k'' &= \sigma \omega \mu. \end{aligned}$$

Эту систему уравнений можно решить в самом общем виде. Но, чтобы упростить расчет, рассмотрим только два частных случая в соответствии с определением η , данным формулой (2.24).

1. В случае слабо проводящей среды ($\eta \gg 1$), который будет нас особенно интересовать в дальнейшем, можно написать

$$\begin{aligned} k^2 &= \omega^2 \epsilon \mu \left(1 - i \frac{\sigma}{\epsilon \omega} \right), \\ k &\approx \pm \frac{\omega}{v_{\Phi}} \left(1 - \frac{i\sigma}{2\epsilon \omega} \right). \end{aligned}$$

Учитывая затем мнимую часть волнового вектора, легко найти экспоненциально затухающую волну вида

$$e^{-\mathbf{k}'' \cdot \mathbf{r}} = e^{-r/\delta}. \quad (2.29)$$

Эта волна не очень глубоко проникает в материал. Глубина проникновения определяется величиной δ , которая равна

$$\delta = \frac{2}{\sigma} \sqrt{\frac{\epsilon}{\mu}}. \quad (2.30)$$

Отсюда видно, что δ зависит от частоты только через σ или ϵ .

2. В случае материала со значительной проводимостью ($\eta \leq 1$), который нас интересует меньше, читатель легко найдет классический *поверхностный эффект* [2], когда глубина проникновения определяется выражением

$$\delta = \sqrt{\frac{2}{\sigma \omega \mu}}.$$

Величина δ обратно пропорциональна квадратному корню из частоты.

Замечание.

Выше мы не рассматривали экспоненциально *нарастающие волны*. Но волны вида $e^{+k'' \cdot r}$ также являются решениями рассматриваемых уравнений. И ничто не запрещает их использовать. К тому же они реально существуют: внутри лазерного резонатора происходит усиление света и, следовательно, существует волна такого типа. Правда, мы не имеем права рассматривать такие бесконечно нарастающие волны, если наша система имеет бесконечные размеры. Но если система ограничена по размерам, то общее решение представляет собой суперпозицию затухающих и нарастающих волн. Впрочем (если не учитывать фазовых явлений), затухает волна или нарастает, это зависит от направления наблюдения.

В данном параграфе мы рассмотрели распространение волн в частично проводящей среде. Это позволило привести пример неоднородной волны. В предыдущем параграфе мы имели дело с гармоническими волнами, которые также являются неоднородными волнами. Итак, с неоднородностью волны связано следующее: 1) пространственное изменение волновых поверхностей, 2) комплексный вектор распространения, чем, вообще говоря, обусловлено различие направлений распространения и поглощения энергии.

Получил объяснение и факт затухания волн из-за эффекта Джоуля. Теперь мы рассмотрим поглощение света с более физической точки зрения.

§ 4. Поглощение света

Что происходит, когда электромагнитная волна распространяется в материале? Атомы состоят из заряженных частиц, а потому на них действуют поля. Для простоты скажем лишь, что в присутствии поля изменяются их колебания. В результате энергия световой волны уменьшается, т. е. происходит поглощение света. Но колеблющиеся заряды снова излучают электромагнитную волну (явление значительно более сложное), которая... и т. д.

Сначала кратко остановимся на вопросе о поле, излучаемом ускоренным зарядом, что позволит нам сформулировать знаменитый *закон рассеяния Рэлея*. Далее мы рассмотрим классическую модель, описывающую действие поля на заряженную частицу.

Заметим, что мы не претендуем на точное *объяснение* рассматриваемых физических явлений. Тогда нам пришлось бы обратиться к квантовой механике. В то же время изучение этих

явлений затрагивает фундаментальные проблемы электромагнетизма, которые еще не решены. На эту тему читателю будет интересно прочитать или перечитать гл. 28 книги Фейнмана [1]. Формулы, которые будут приведены ниже, нужно рассматривать как некое эвристическое описание интересующих нас явлений. И наконец, отметим, что в данной главе мы останемся на уровне теоретической идеализации; реальные случаи поглощения будут рассмотрены далее, в гл. 8—10, посвященных оптическим волноводам.

А. Поле, излучаемое ускоренным зарядом, рассеяние света, закон Рэлея

Рассмотрим движущийся заряд. В тех случаях, которые нас интересуют, скорость перемещения зарядов в материале всегда очень мала по сравнению со скоростью света. Поэтому мы можем считать, что заряд, движущийся равномерно, создает статическое кулоновское поле: свет распространяется так быстро, что для него заряд как бы неподвижен.

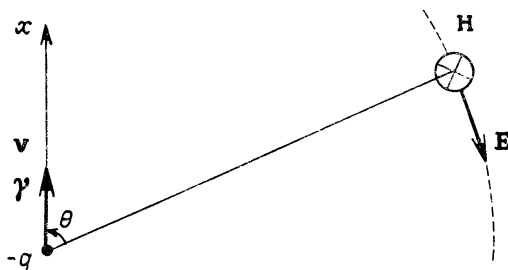


Рис. 2.6. Поле излучения ускоренно движущегося заряда.

Но нас не интересует статическое поле. Поэтому рассмотрим случай заряда, движущегося ускоренно. Мы будем следовать доказательству, приведенному в курсе физики Берклиевского университета [5], не воспроизводя его подробно.

Можно показать, что излучение обусловлено поперечной составляющей ускорения $\gamma_{\perp}(t)$. Напряженность электрического поля излучаемой волны дается выражением (рис. 2.6)

$$E(\mathbf{r}, t) = \frac{-q}{4\pi\epsilon_0 r c^2} \gamma_{\perp} \left(t - \frac{r}{c} \right)$$

(тогда как статическое поле просто пропорционально $1/r^2$). Допустим, что на очень большом расстоянии излучаемая волна становится плоской. Предположим теперь, что заряд колеблется по синусоидальному закону (гармонический осциллятор)

$$x = x_0 \cos \omega t.$$

Средняя излучаемая им мощность $\langle P \rangle$ легко рассчитывается путем интегрирования потока вектора Пойнтинга (2.14) через сферу. Предоставим это сделать читателю. Результат таков:

$$\langle P \rangle = \frac{\omega^4 q^2}{6\pi\epsilon c^3} \langle x^2 \rangle. \quad (2.31)$$

Электромагнитная мощность, излучаемая гармоническим осциллятором, пропорциональна величине ω^4 , т. е. величине λ^{-4} .

Это закон Рэлея. Позднее мы увидим, что при распространении в материале электромагнитная волна приводит в движение заряды и создает гармонические осцилляторы. Это вызывает рассеяние света в соответствии с законом (2.31). Естественно, на столько же уменьшается энергия падающей волны.

В соответствии с законом Рэлея рассеяние света обусловлено поглощением, которое уменьшается с увеличением длины волны (пропорционально λ^{-4}).

Теперь вернемся к исходному явлению, а именно действию световой волны на вещество.

Б. Микроскопическая модель поляризации

Напомним, что масса ядра атома значительно больше массы окружающих его электронов. Рассмотрим следующую задачу: плоская волна падает на покоящийся электрон с массой m и зарядом $-q$. Что при этом происходит?

Электрон связан с атомом. Поэтому будем считать, что на него действуют упругая возвращающая сила с постоянной $m\omega_0^2$ и сила «трения» $-a\mathbf{v}$. Получаем следующее уравнение движения электрона:

$$\mathbf{F} = -q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = m \frac{d^2 \mathbf{x}}{dt^2}.$$

Так как $v \ll c$ и $\mathbf{B}_0 = \mathbf{E}_0/c$, мы можем пренебречь членом с $[\mathbf{v} \times \mathbf{B}]$. Окончательно получаем

$$m \frac{d^2 x}{dt^2} + a \frac{dx}{dt} + m\omega_0^2 x = -qE_0 e^{i\omega t}.$$

Следовательно, комплексная амплитуда вынужденных колебаний электрона с частотой ω будет равна

$$x_c = \frac{-\frac{q}{m} E_0}{\omega_0^2 - \omega^2 + i \frac{a\omega}{m}}.$$

В дальнейшем мы будем опускать индекс «с» в обозначении комплексной амплитуды, что не отразится на ясности изложения, но позволит упростить запись.

Каков порядок величины ω_0 ? Если речь идет о заряде электрона в атоме, то известно, что соответствующие длины волн излучения расположены в ультрафиолетовой области спектра (эксперимент Франка и Герца) [6]. Следовательно, для рассматриваемого типа осциллятора можно принять, что

$$\omega_0 \gg \omega.$$

Найдем с учетом этого среднюю энергию осциллятора:

$$\langle x^2 \rangle \approx \frac{1}{2\omega_0^4} \left(\frac{q}{m} E_0 \right)^2.$$

Подставив это выражение в (2.31), находим

$$\langle P \rangle = \frac{\omega^4}{\omega_0^4} \frac{q^2}{12\pi\epsilon c^3} \left(\frac{q}{m} E_0 \right)^2.$$

Этот результат позволяет дополнить закон Рэлея: мощность, переизлучаемая диполями, создаваемыми электромагнитным полем, пропорциональна ω^4 .

Мы показали, что при взаимодействии поля с веществом возникают осциллирующие диполи. На микроскопическом уровне плотность дипольных моментов характеризуется вектором поляризации среды. Если в единице объема имеется η электронов, то вектор поляризации запишется в виде

$$\mathbf{P} = -\eta q \mathbf{x}.$$

Этот поляризационный член добавляется к $\epsilon \mathbf{E}$ при расчете электрической индукции [формула (1.16)]. Поэтому комплексная диэлектрическая проницаемость в данном случае будет равна

$$\frac{\eta q^2/m}{\omega_0^2 - \omega^2 + \frac{i a \omega}{m}}. \quad (2.32)$$

Напомним теперь, что среда, в которой распространяется волна, содержит не только электроны, но и атомы, молекулы, ионы и т. д. Взаимодействие любой заряженной частицы с полем описывается рассмотренной моделью. Более того, если взять стекло, то оно содержит в единице объема \mathcal{N} атомов стекла и различные *примеси* (в частности, ионы OH, а также ионы Fe, Cu, Sg и т. д.).

Для чистого стекла резонансный максимум действительно соответствует изложенной модели: условие $\omega_0 \gg \omega$ выполняется. Следовательно, в видимой и ближней инфракрасной областях спектра диэлектрическая проницаемость чистого стекла практически постоянна. Таким образом, для стекла без примесей имеем

$$\epsilon = \epsilon_0 + \frac{\mathcal{N} q^2}{m \omega_0^2}.$$

Это выражение можно переписать иначе, если ввести *элементарную статическую поляризуемость* стекла α_0 :

$$\epsilon = \epsilon_0 (1 + \mathcal{N} \alpha_0), \quad (2.33)$$

где

$$\alpha_0 = \frac{q^2}{m \epsilon_0 \omega_0^2}. \quad (2.34)$$

Для реального стекла в выражение (2.33) необходимо добавить столько членов вида (2.32), сколько имеется типов осцилляторов, обусловленных примесями.

Мы рассмотрим случай, когда стекло содержит лишь одну примесь с объемной плотностью η . Тогда выражение (2.33) примет вид

$$\epsilon = \epsilon_0 \left(1 + \mathcal{N} \alpha_0 + \frac{\eta \frac{q^2}{m}}{\omega_0^2 - \omega^2 + i \frac{a}{m} \omega} \right).$$

Чтобы найти частотную зависимость коэффициента поглощения k'' , достаточно решить уравнение дисперсии (2.26)¹⁾

$$k^2 = \omega^2 \mu \epsilon.$$

Для простоты допустим, что

1) *элементарная статическая поляризуемость одинакова для стекла и для примесей* [формула (2.34)];

2) *содержание примесей очень мало*.

В этом случае можно рассмотреть разложение в ряд, ограниченное двумя членами:

$$k' - ik'' \approx k_0 \sqrt{1 + \mathcal{N} \alpha_0} \left(1 + \frac{\eta \alpha}{2(1 + \mathcal{N} \alpha_0)} + \dots \right),$$

где $k_0^2 = \omega^2 \mu \epsilon_0$, а через α обозначена элементарная частотная поляризуемость:

$$\alpha = \frac{\frac{q^2}{m}}{\omega_0^2 - \omega^2 + i \frac{a}{m} \omega}.$$

Теперь остается только сделать выкладки. Введем следующие обозначения:

$$\Delta = a/m,$$

$$n_0 = \sqrt{1 + \mathcal{N} \alpha_0} \text{ — показатель преломления чистого стекла.} \quad (2.35)$$

¹⁾ Строго говоря, проведенное рассмотрение не вполне точно, так как мы не учитываем здесь действия поля, создаваемого другими соседними частицами (поправка на «локальное поле»). Кроме того, строгая теория взаимодействия поля со средой требует привлечения квантовой механики и приводит к появлению так называемых «сил осциллятора». Но в данной книге в этом нет необходимости.

Найдем для k'' форму резонансного максимума в окрестности частоты ω_0 (рис. 2.7). Поскольку обычно этот максимум достаточно узкий, запишем приближенное выражение, справедливое вблизи ω_0 . Обозначив через u разность $\omega - \omega_0$, получим

$$k'' = \frac{k_0''}{1 + \frac{4u^2}{\Delta^2}}, \quad (2.36)$$

где

$$k_0'' = \frac{k_0}{2} \frac{\eta a_0}{n_0} \left(\frac{\omega_0}{\Delta} \right). \quad (2.37)$$

Величина Δ — «полуширина» резонансного максимума (полная ширина на половине высоты). На практике формулой (2.37)

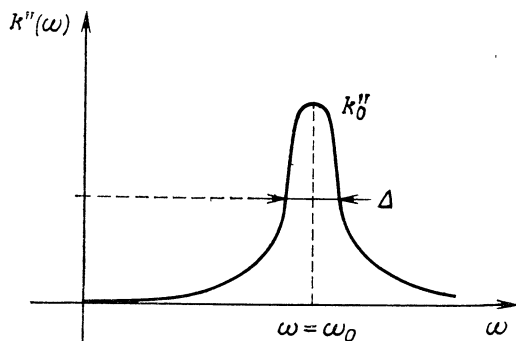


Рис. 2.7 Резонансная кривая (зависимость поглощения от частоты).

можно пользоваться для расчета содержания примесей. Для этого экспериментально определяют зависимость $k''(\omega)$ и величины ω_0 , k_0'' и Δ . Кроме того, измеряют показатель преломления n_0 . Затем по формулам (2.35) и (2.37) вычисляют отношение η/\mathcal{N} .

Итак, мы объяснили поглощение стекла: обычно коэффициент поглощения растет с частотой (закон Рэлея), но на эту кривую роста накладываются резонансные максимумы, соответствующие примесям (примеры таких кривых для реальных материалов читатель найдет в гл. 8—10, в частности на рис. 9.2). Эти резонансные максимумы, обусловленные примесями, вреднее в вопросах оптической связи, чем само рэлеевское рассеяние. В следующем параграфе мы рассмотрим другой вредный эффект, а именно дисперсию.

§ 5. Дисперсия

Известно, что в общем случае процесс распространения электромагнитных волн в среде является дисперсным, т. е. скорость распространения зависит от частоты. Поэтому, как было

сказано в гл. 1, предположение о чисто гармоническом режиме колебаний не соответствует физической реальности.

В действительности дисперсия складывается из двух различных явлений:

1) *дисперсия, обусловленная средой*, т. е. показатель преломления среды зависит от частоты; это — основной случай;

2) в оптическом волноводе имеется дисперсия, обусловленная неоднородностью среды (или наличием нескольких разных сред). Имеется в виду структура направленных мод. Как будет показано в гл. 3, каждой моде соответствует своя собственная дисперсия. Таким образом, существует *дисперсия мод* (или *модовая дисперсия*).

А. Дисперсия, обусловленная средой

Здесь мы рассмотрим только дисперсию, обусловленную средой распространения, а о дисперсии мод будет сказано в гл. 3, посвященной вопросу канализации волн.

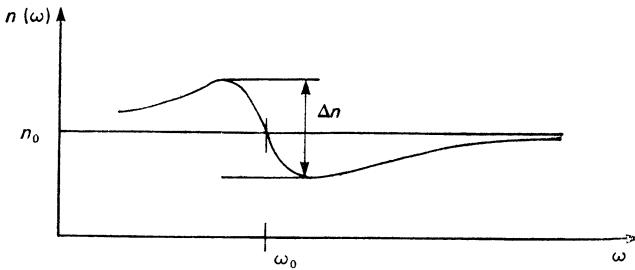


Рис. 2.8. Зависимость дисперсии, обусловленной средой распространения, от частоты в случае резонанса.

Впрочем, половина работы уже сделана. Чтобы найти k' , достаточно выделить действительную часть выражения (2.34). Эта операция полностью идентична расчету k'' . С теми же замечаниями, что и ранее [и используя главным образом соотношение (2.37)], находим зависимость показателя преломления среды от частоты:

$$\begin{aligned}
 n(\omega) &= c \sqrt{\epsilon \mu} = n_0 \left(1 + \frac{\alpha_0 \omega_0^2 (\omega_0^2 - \omega^2)}{2n_0^2 [(\omega_0^2 - \omega^2)^2 + \Delta^2 \omega^2]} \right) \approx \\
 &\approx n_0 - \frac{k_0'' 2u / \Delta}{k_0 \left(1 + \frac{4u^2}{\Delta^2} \right)},
 \end{aligned} \tag{2.38}$$

график которой приведен на рис. 2.8.

Как видно из рисунка, резонанс проявляется в резком изменении показателя преломления вблизи частоты ω_0 . Максимальное изменение показателя преломления равно

$$\Delta n = \frac{k_0''}{k_0} \quad (2.39)$$

Выражение (2.39) подчеркивает то обстоятельство, что дисперсия и поглощение тесно связаны друг с другом.

Можно показать, что в более общем случае действительная и мнимая составляющие магнитной восприимчивости χ' и χ'' в силу принципа причинности связаны между собой преобразованиями Гильберта или соотношениями Крамерса — Кронига [2].

Вот что дает нам физика. Посмотрим теперь, как распространяется пакет волн в дисперсной среде.

Б. Распространение пакета волн

В рассматриваемой нами дисперсной среде вектор k так же, как и показатель преломления n , зависит от частоты ω ; следовательно, и *фазовая скорость плоской волны также будет зависеть от частоты*.

Мы знаем, что плоских волн нет в природе. Значит, в средствах связи приходится передавать *пакет волн* с разными частотами. Но каждая частота распространяется со своей собственной скоростью, а в месте приема они суммируются без учета пожеланий передающего. Так, например, если передаются импульсы, то в общем случае они будут, к сожалению, расширяться и искажаться по форме.

1. Групповая скорость и соотношение неопределенностей

Предположим, что передается *узкий* волновой пакет, иначе говоря частоты близки к ω_0 ; распространение будет происходить вдоль оси z . При $z = 0$ поле может быть записано в форме интеграла Фурье

$$E(0, t) = \int_{\omega_1}^{\omega_2} \tilde{E}(\omega) e^{i\omega t} d\omega.$$

Поскольку функция $\tilde{E}(\omega)$ предполагается ограниченной по частоте, можно интегрировать по частоте в бесконечных пределах, от $-\infty$ до $+\infty$.

Так как каждая частота распространяется со своей собственной скоростью, то поле в точке z будет равно

$$E(z, t) = \int_{-\infty}^{+\infty} \tilde{E}(\omega) e^{i(\omega t - k(\omega)z)} d\omega. \quad (2.40)$$

Разложим показатель экспоненты в ряд и ограничимся первыми двумя членами:

$$i(\omega t - kz) \approx i(\omega_0 t - k_0 z) + i(\omega - \omega_0) \left[t - \left(\frac{dk}{d\omega} \right)_{\omega} z \right].$$

Внимание! Приведенное ограничение разложения допустимо, если член второго порядка

$$\frac{i}{2} (\omega - \omega_0)^2 \left(\frac{d^2 k}{d\omega^2} \right)_{\omega=\omega_0}$$

бесконечно мал.

Но, поскольку мы будем подставлять это ограниченное разложение в показатель экспоненты, необходимо, чтобы этот член второго порядка был мал не только по сравнению с членом первого порядка, но и по сравнению с 1, т. е. чтобы выполнялось условие

$$\exp \left[\frac{i}{2} (\omega - \omega_0)^2 \left(\frac{d^2 k}{d\omega^2} \right)_{\omega=\omega_0} \right] \approx 1.$$

Введем обозначение

$$v_g = \frac{1}{\left(\frac{dk}{d\omega} \right)_{\omega=\omega_0}}. \quad (2.41)$$

Эта величина называется *групповой скоростью*. Тогда соотношение (2.40) примет вид

$$E(z, t) \approx e^{i(\omega_0 t - k_0 z)} g \left(t - \frac{z}{v_g} \right),$$

где

$$g(t) = \int_{-\infty}^{+\infty} \tilde{E}(\omega) e^{i(\omega - \omega_0)t} d\omega.$$

В результате можно сделать следующие выводы:

1) групповая скорость v_g представляет собой скорость распространения огибающей g поля $E(z, t)$;

2) ширины функций g и \tilde{E} изменяются в противоположных направлениях, как это следует из преобразования Фурье (рис. 2.9). Но мы знаем, что можно рассматривать волну либо как функцию времени в данной точке пространства, либо как функцию точки пространства в данный момент времени. Таким образом,

а) в фиксированной точке z :

$$\Delta t \Delta \omega = \text{const}, \quad (2.42a)$$

где Δt — длительность волнового пакета, $\Delta \omega$ — ширина полосы частот;

б) в фиксированный момент времени t :

$$\Delta z \Delta k = \text{const}, \quad (2.42b)$$

где Δz — пространственная протяженность пакета, Δk — «монокromaticность», представляющая собой пространственный эквивалент ширины полосы частот, которая выражается функцией относительного изменения длины волны

$$\Delta k = 2\pi \frac{\Delta \lambda}{\lambda^2}$$

В гл. 4, § 1, п. А, 1 мы дадим другую интерпретацию пространственной протяженности волнового пакета Δz . В более общем случае полученные соотношения (которые представляют

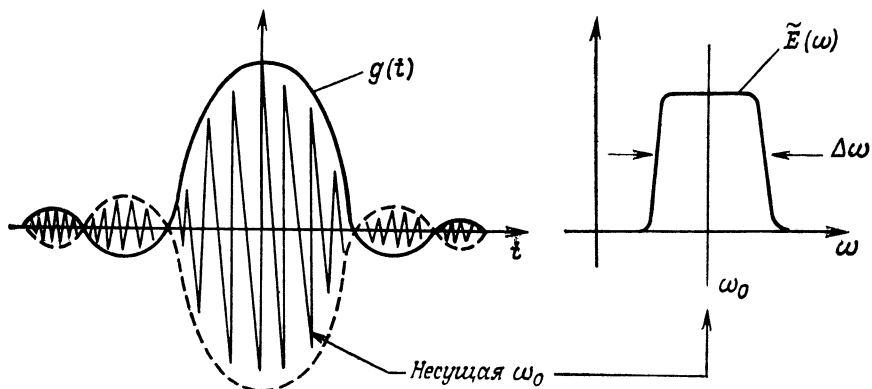


Рис. 2.9. Волновой пакет.

Функции $g(t)$ и $E(\omega)$ связаны между собой преобразованием Фурье.

собой не что иное, как соотношения неопределенностей) достаточно хорошо иллюстрируют некоторые характерные свойства лазерных лучей (гл. 4 и 12).

2. Расширение волнового пакета, обусловленное средой распространения

Волновой пакет, распространяющийся в дисперсной среде, постепенно расширяется. При перемещении на расстояние l временное уширение, связанное с вкладом полосы частот $\omega_0 \pm \Delta\omega$, приблизительно дается выражением

$$\frac{\Delta t}{\Delta \omega} = l \frac{|\partial v_g / \partial \omega|}{v_g^2}.$$

Коэффициент $\delta = \frac{\partial v_g}{\partial \omega}$ (2.43)

характеризует качество среды распространения в отношении дисперсии. Согласно формуле (2.42),

$$\Delta\omega \approx \frac{1}{\Delta t} = \frac{v_g}{\sqrt{l\delta}}. \quad (2.44)$$

Коэффициент (2.43) можно также выразить через функцию дисперсии $k(\omega)$. Действительно, из выражений (2.41) и (2.43) следует, что

$$\Delta\omega \approx \frac{1}{\sqrt{l \left| \frac{\partial^2 k}{\partial \omega^2} \right|}}. \quad (2.45)$$

Можно также ввести закон дисперсии в виде зависимостей $n(\omega)$ или $n(\lambda)$ ¹⁾, и читатель легко найдет соответствующие формулы. Например:

$$\Delta\omega \approx \sqrt{\frac{c}{l \left| 2 \frac{dn}{d\omega} + \omega \frac{d^2 n}{d\omega^2} \right|}}. \quad (2.46)$$

Однако на данном уровне было бы нецелесообразно применять для расчета производные из (2.38). При использовании формулы (2.45) достаточно иметь только экспериментально полученные зависимости $n(\omega)$. В заключение отметим:

Уширение волнового пакета, обусловленное средой распространения, пропорционально корню квадратному из пройденного расстояния. Полоса передаваемых частот $\Delta\omega$ выражается в единицах ГГц·км^{1/2}.

¹⁾ В этом случае вводят «групповой показатель преломления» $N = n - \lambda(dn/d\lambda)$. Это обозначение будет использовано при изучении оптических волноводов (гл. 8, § 5).

ГЛАВА 3

РАСПРОСТРАНЕНИЕ ВОЛН В СВЕТОВОДАХ

Если бы не было дифракции, то можно было бы говорить о плоской волне и световом луче. В этом случае световой пучок распространялся бы только прямолинейно и не расширялся. Чтобы передать сообщение с помощью светового пучка, было бы достаточно направить его на получателя, как бы далеко тот ни находился. Но, как будет показано в следующей главе, из-за дифракции световой пучок всегда постепенно расширяется. Поэтому приходится вмешиваться в процесс распространения пучка, чтобы компенсировать его расширение. Направленное распространение волн обеспечивается тем, что пучок света ограничивают некими направляющими поверхностями, и он распространяется, испытывая многократное отражение от них.

Хотя реальная световая волна не плоская, мы тем не менее воспользуемся удобным приближением плоских волн и представим поле в виде их суммы. Каждая плоская волна будет распространяться по своему закону.

Сначала мы напомним процесс отражения плоской волны от границы двух сред с разными показателями преломления, а затем рассмотрим металлические и диэлектрические волноводы. Мы познакомимся с некоторыми явлениями и понятиями, необходимыми для усвоения физики оптических волокон: с полным внутренним отражением, затухающей волной, световодными модами и т. д.

Основное внимание будет уделено диэлектрическим световодам, но для иллюстрации их основных характеристик мы будем сравнивать их с металлическими волноводами.

§ 1. Отражение плоской волны от границы раздела двух сред

Рассмотрим плоскую границу раздела двух сред с разными диэлектрическими проницаемостями ϵ_1 и ϵ_2 . Индексы i , r и t относятся к падающей, отраженной и прошедшей волнам.

А. Случай нормального падения

Напряженности поля будем рассматривать как скалярные величины, имея в виду направления, указанные на рис. 3.1 (в начальный момент напряженность $E_i > 0$ направлена в сторону $-y$, напряженность $H_i > 0$ — в сторону $+z$).

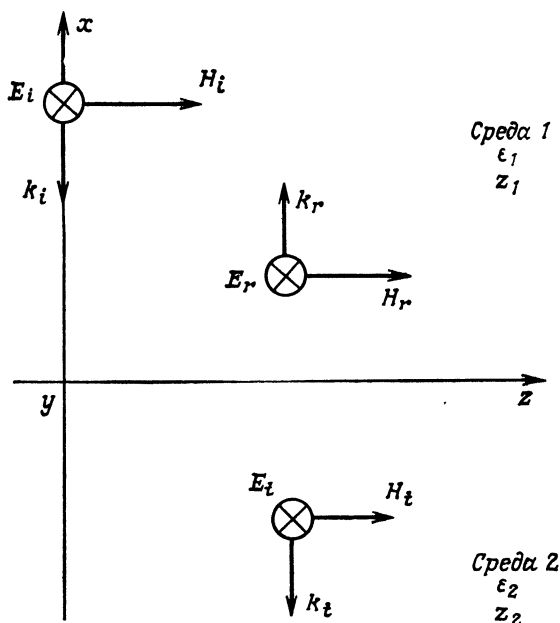


Рис. 3.1. Отражение плоской волны от границы раздела двух сред в случае нормального падения.

Выражения для волновых сопротивлений [формулы (2.8)] этих сред запишутся в виде

$$\frac{E_i}{H_i} = Z_1, \quad \frac{E_t}{H_t} = Z_2, \quad \frac{E_r}{H_r} = -Z_1.$$

(Знак «—» для отраженной волны получается из-за изменения направления распространения и нашего условия о скалярной форме записи напряженностей поля.)

Условия непрерывности (1.18) и (1.19) запишутся следующим образом: $E_i + E_r = E_t$, $H_i + H_r = H_t$.

Следует отметить также, что

$$\frac{E_{\text{полн}}}{H_{\text{полн}}} = \frac{E_i + E_r}{H_i + H_r} = \frac{E_t}{H_t} = Z_2.$$

При отражении волны в среде 1 от границы со средой 2 полное волновое сопротивление равно волновому сопротивлению среды 2.

Из пяти предшествующих уравнений легко получить амплитудные значения коэффициентов отражения и пропускания для напряженности электрического поля:

$$r = \frac{E_r}{E_i} = \frac{Z_2 - Z_1}{Z_1 + Z_2}, \quad t = \frac{E_t}{E_i} = \frac{2Z_2}{Z_1 + Z_2}. \quad (3.1)$$

Введем показатели преломления (гл. 2, § 1):

$$n_1 = \frac{c}{v_{\varphi_1}} = \sqrt{\frac{\epsilon_1}{\epsilon_0}}, \quad n_2 = \sqrt{\frac{\epsilon_2}{\epsilon_0}}, \quad \eta = \frac{n_2}{n_1} = \frac{Z_1}{Z_2}.$$

В результате получаем классические формулы:

$$r = \frac{1 - \eta}{1 + \eta}, \quad t = \frac{2}{1 + \eta}.$$

Чтобы вывести выражения для коэффициентов отражения R и пропускания T по мощности, нужно воспользоваться выражениями (2.14) для потока энергии. В итоге находим

$$R = \left| \frac{Z_2 - Z_1}{Z_2 + Z_1} \right|^2, \quad T = \frac{|E_t|^2/Z_2}{|E_i|^2/Z_1} = \frac{Z_1}{Z_2} |t|^2 = \frac{4Z_1 Z_2}{|Z_1 + Z_2|^2},$$

что согласуется с законом сохранения энергии:

$$R + T = 1.$$

Б. Случай произвольного падения

В этом случае плоскость xz (которую мы будем считать плоскостью падения) становится плоскостью симметрии.

1. Волны TE и TM , принцип дуальности

Покажем, что возможные электромагнитные поля представляют собой линейные комбинации двух типов волн (рис. 3.2):

1) волны TE (поперечной электрической волны), для которой вектор напряженности электрического поля \mathbf{E} перпендикулярен плоскости падения, и

2) волны TM (поперечной магнитной волны), для которой вектор напряженности магнитного поля \mathbf{H} перпендикулярен плоскости падения.

Их названия объясняются тем, что если рассматривать распространение волн в направлении, отличном от направления вектора \mathbf{k} , но в плоскости симметрии (например, в плоскости Ox или Oz), то только один из векторов напряженности электромагнитного поля не будет иметь продольной составляющей. Легко видеть, что в нашем случае, когда ищутся решения, для которых выполняется условие

$$\frac{\partial}{\partial y} = 0,$$

система уравнений Максвелла (1.1)–(1.4) разделяется на две независимые системы трех уравнений, каждая из которых содержит следующие величины:

1-я система $E_y, H_x, H_z \rightarrow$ волны TE ,

2-я система $H_y, E_x, E_z \rightarrow$ волны TM .

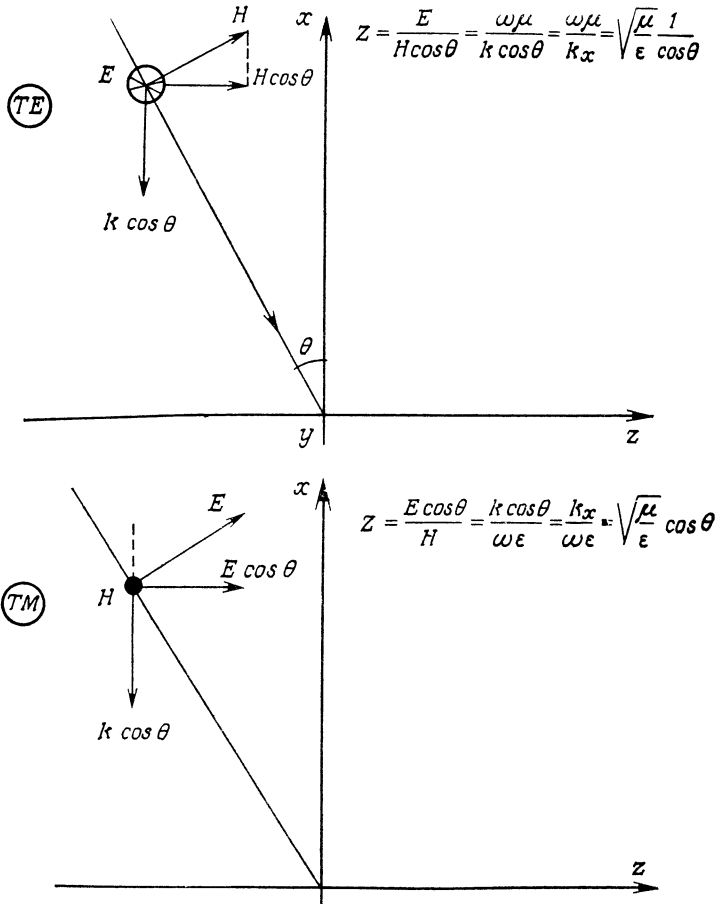


Рис. 3.2. Отражение плоской волны от границы раздела двух сред в случае наклонного падения.

При расчете полного волнового сопротивления учитываются только составляющие напряженностей полей, параллельные границе раздела сред.

Таким разделением уравнений Максвелла можно пользоваться при решении задач о распространении в волноводах. В общем случае показано [2], что электромагнитные поля можно разделять на разные типы волн в зависимости от симметрии системы. Такими соображениями мы будем руководствоваться далее (гл. 3, § 2 и § 3; гл. 8, § 4).

Говоря об упрощении задачи, следует отметить *принцип дуальности*. Речь идет о следующем свойстве уравнений Максвелла: уравнения Максвелла (1.1) и (1.2) переходят одно

в другое при замене

$$\mathbf{E} \rightarrow \mathbf{H}, \quad \mathbf{H} \rightarrow -\mathbf{E}, \quad \mu \rightarrow \epsilon, \quad \epsilon \rightarrow \mu.$$

Этот принцип имеет очень важное значение, так как он позволяет:

а) находить коэффициенты r и t для магнитных полей, зная коэффициенты для электрических полей;

б) находить выражения для моды TM , зная выражения для моды TE , и наоборот.

2. Формулы Френеля, угол Брюстера

Вернемся к своему вопросу. В случае произвольного падения (рис. 3.2) можно всегда разложить волну на две плоские волны: одну, распространяющуюся в направлении $x < 0$, другую — в направлении $z > 0$. Для этого достаточно разложить поле в плоскости падения на две составляющие, распространяющиеся вдоль осей Ox и Oz . Следовательно:

а) при распространении волны вдоль оси Oz не происходит никакого отражения;

б) распространение волны вдоль оси Ox приводит к появлению прошедшей и отраженной волн.

В результате в среде 1 мы получаем бегущую волну вдоль оси Oz и стоячую волну вдоль оси Ox .

Чтобы вычислить коэффициенты отражения и пропускания, достаточно рассмотреть распространение волны вдоль оси Ox . Задача сводится к случаю нормального падения с угловыми множителями, входящими в выражение для полного сопротивления (рис. 3.2). В результате с учетом (3.1) получаем следующие зависимости коэффициентов отражения r и пропускания t от углов падения i и отражения r :

для волны TE :

$$\begin{aligned} Z_1 &\sim \frac{1}{\cos i} \left\{ \begin{aligned} r_{TE} &= \frac{\cos i - \eta \cos r}{\cos i + \eta \cos r}, & R &= |r|^2, \\ Z_2 &\sim \frac{1}{\eta \cos r} \left\{ \begin{aligned} t_{TE} &= \frac{2 \cos i}{\cos i + \eta \cos r}, & T &= |t|^2, \end{aligned} \right. \end{aligned} \right. \end{aligned} \quad (3.2)$$

для волны TM :

$$\begin{aligned} Z_1 &\sim \cos i \left\{ \begin{aligned} r_{TM} &= \frac{\cos r - \eta \cos i}{\cos r + \eta \cos i}, & R &= |r|^2, \\ Z_2 &\sim \frac{\cos r}{\eta} \left\{ \begin{aligned} t_{TM} &= \frac{2 \cos r}{\cos r + \eta \cos i}, & T &= |t|^2. \end{aligned} \right. \end{aligned} \right. \end{aligned} \quad (3.3)$$

Таким образом, мы очень быстро вывели знаменитые формулы Френеля.

Вдумчивый читатель может заметить, что $R + T \neq 1$! Однако закон сохранения энергии выполняется, так как вектор Пойнтинга определяется одновременно двумя составляющими электромагнитного поля \mathbf{E} и \mathbf{H} .

Кривые зависимости коэффициентов отражения от угла падения приведены на рис. 3.3.

Здесь мы находим следующие известные выводы.

1. Для волны TM существует особый угол падения i_B , называемый *углом Брюстера*, при котором коэффициент отражения равен нулю. Это явление обычно используют для получения

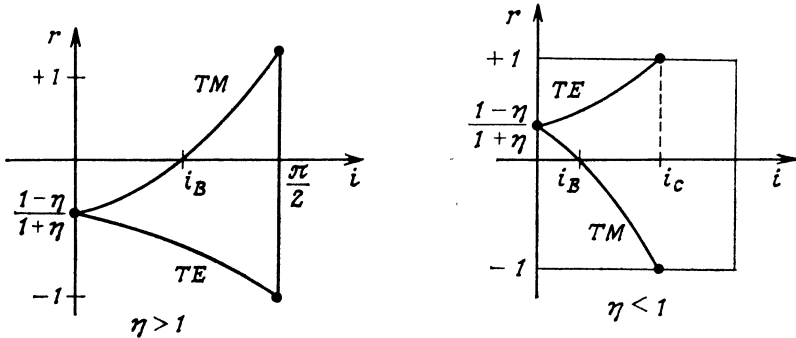


Рис. 3.3. Зависимость коэффициента отражения r от угла падения.

поляризованного света при отражении (в частности, в газовых лазерах с этой целью используют окно Брюстера; гл. 12, § 2, п. Б).

2. При нормальном падении волны TE на оптически более плотную среду ($\eta > 1$) она приобретает при отражении *фазовый сдвиг*, равный π .

3. При отражении волны TE от поверхности оптически менее плотной среды ($\eta < 1$) обнаруживается *предельный угол* падения i_c , при котором выполняется условие

$$\sin i_c = \eta \quad (3.4)$$

и который соответствует *полному внутреннему отражению*, поскольку в этом случае $|r| = 1$.

Что же происходит при углах падения, больших $i = i_c$? Это очень важный вопрос, особенно для анализа направленного распространения волн. Рассмотрим его подробнее.

В. Полное внутреннее отражение

Рассмотрим частный случай отражения волны TE при угле падения $i > i_c$ и при $\eta < 1$ [формула (3.2)]. В этом случае отношение волновых сопротивлений

$$\frac{Z_1}{Z_2} = \frac{\eta \cos r}{\cos i}$$

есть чисто мнимая величина, так как

$$\cos r = \sqrt{1 - \frac{\sin^2 i}{\eta^2}} = \pm i \sqrt{\frac{\sin^2 i}{\eta^2} - 1}. \quad (3.5)$$

Положим

$$\frac{Z_1}{Z_2} = -iz.$$

Тогда выражение для коэффициента отражения примет вид

$$r = \frac{1 + iz}{1 - iz}, \quad \text{где } |r| = 1. \quad (3.6)$$

Отсюда следуют два важных результата.

1. Происходит *полное внутреннее отражение*. Положим $r = e^{-i\Phi}$; тогда

$$\Phi = -\arg(r) = -2 \arctg(z) = \pm 2 \arctg \left[\frac{\sqrt{\sin^2 i - \eta^2}}{\cos i} \right]. \quad (3.7)$$

2. *Отраженная волна приобретает фазовый сдвиг, зависящий от угла падения.*

Далее мы вернемся к вопросу об этом фазовом сдвиге.

Поскольку происходит полное внутреннее отражение, полезно проанализировать энергетические соотношения. Рассмотрим подробнее волну в среде 2 (рис. 3.4). Если считать напряженность электрического поля \mathbf{E} , вектор которой параллелен оси Oy , действительной величиной, то две составляющие напряженности магнитного поля \mathbf{H} будут равны

$$H_x = H \sin r \quad \text{действительная,}$$

$$H_z = H \cos r \quad \text{мнимая,}$$

а соответствующие потоки P векторов Пойнтинга

$$P_z \sim E \times H_x^* \quad \text{действительный,}$$

$$P_x \sim E \times H_z^* \quad \text{мнимый.}$$

Таким образом, вдоль оси Oz мы имеем поток лучистой энергии, а вдоль оси Ox — поток реактивной энергии.

Следовательно, это эквивалентно наличию *неоднородной волны*, аналогичной той, которая рассматривалась в гл. 2, § 3, п. Б¹⁾). Исходя из этого, легко получить выражения для векторов \mathbf{k}' и \mathbf{k}'' :

$$k'_z = k_2 \frac{\sin i}{\eta} = k_1 \sin i, \quad (3.8)$$

$$k''_x = ik_2 \cos r, \quad (3.9)$$

¹⁾ Формально выражение для рассматриваемой волны идентично выражению из гл. 2, § 3, п. Б, но физические явления, лежащие в их основе, неодинаковы: в нашем случае среды — без потерь и без рассеяния энергии.

результатов, полученных для рассеивающей среды [формулы (2.29) и (2.30)].

Волны такого типа будут рассмотрены далее (гл. 8 и 9).

§ 2. Металлический световод

Прежде чем обратиться к изучению световодов, рассмотрим одномерные системы. Предположим, что волна распространяется в направлении Oz между двумя бесконечно длинными плоскостями, параллельными друг другу и оси Oz . Разумеется, при

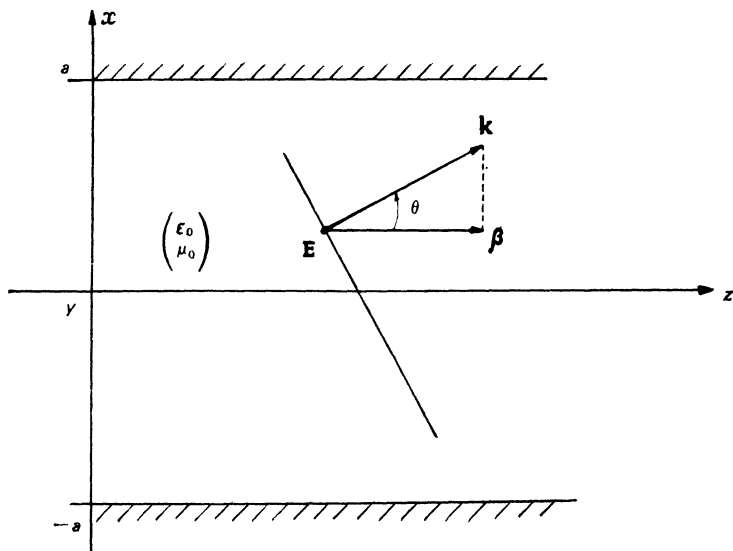


Рис. 3.5. Металлический одномерный световод.

таком выборе плоских световодов ограничивается общность наших выводов, поскольку реальные световоды имеют прямоугольное или круглое сечение. Но зато мы сможем легко понять основные явления. Далее в гл. 8 будут детально изучены цилиндрические оптические световоды. Другие виды световодов будут рассмотрены в гл. 14.

Подчеркнем, что мы останавливаемся на металлическом световоде с единственной целью проиллюстрировать те представления, на основе которых исследуется распространение света в диэлектрических световодах (оптическое приближение п. А; электромагнитное приближение п. Б; моды распространения п. В; затухающие волны п. Г). Читатель, уже знакомый с такими представлениями, может опустить этот материал.

Металлический световод представлен на рис. 3.5. Он образован двумя бесконечными *идеально проводящими* плоскостями,

уравнения которых таковы: $x = \pm a$. Заполняющая его среда (при $-a < x < +a$) — вакуум. Очевидно, что плоскость чертежа xz является плоскостью симметрии. Ранее отмечалось, что волны TE (E_y, H_x, H_z) и TM (H_y, E_x, E_z) можно рассматривать раздельно. Мы (произвольно) выберем волны TE . Для волн TM можно провести такой же анализ или воспользоваться принципом дуальности (§ 1, п. А).

А. Оптическое приближение

В световоде будет распространяться волна, форма которой соответствует его форме. Покажем, что эту волну можно рассматривать как линейную комбинацию плоских волн.

Исходя из этого, исследуем плоскую монохроматическую волну с длиной волны λ , распространяющуюся в вакууме¹⁾ между двумя проводящими плоскостями. Волновой вектор волны лежит в плоскости xOz и образует угол θ с осью Oz ($0 < \theta < \pi/2$). Такую волну называют «восходящей». Вектор напряженности электрического поля будем считать параллельным оси Oy :

$$E_y = E_0 e^{i(\omega t - kz \cos \theta - kx \sin \theta)}, \quad k = \frac{2\pi}{\lambda}.$$

В результате отражения от верхней плоскости появляется «нисходящая» волна ($-\pi/2 < \theta < 0$) с комплексной амплитудой rE_y . Коэффициент r определяется из граничного условия $E_y = 0$ (при всех z) на плоскости $x = a$:

$$e^{-ika \sin \theta} + r e^{ika \sin \theta} = 0,$$

откуда

$$r = -e^{-2ika \sin \theta}.$$

В результате интерференции этих двух волн во всех точках пространства между плоскостями образуется полное поле, определяемое выражением

$$E_{\text{полн}} = E_0 e^{i(\omega t - kz \cos \theta)} (e^{-ikx \sin \theta} - e^{ik(x-2a) \sin \theta}). \quad (3.10)$$

Очевидно, что граничные условия при $x = a$ выполняются. При $x = -a$ должно выполняться условие

$$e^{ika \sin \theta} - e^{-3ika \sin \theta} = 0,$$

откуда

$$2ka \sin \theta = n\pi, \quad (3.11)$$

¹⁾ Далее мы часто будем опускать индекс «0» у характеристик вакуума; таким образом, λ_0 и k_0 для вакуума будут записываться как λ и k .

где n — положительное целое число¹⁾ (тривиальный случай $\theta = 0$ не удовлетворяет условиям задачи).

Тогда выражение для полного поля запишется следующим образом:

$$E_{\text{полн}} = E_0 e^{i[\omega t - (k \cos \theta) z]} \left(e^{-i \frac{n\pi}{2a} x} - (-1)^n e^{i \frac{n\pi}{2a} x} \right).$$

Предположим, что n — четное число, т. е. $n = 2p$. Тогда поле (нечетное) примет вид

$$E_{\text{полн}} = -2iE_0 e^{i[\omega t - (k \cos \theta) z]} \sin\left(\frac{p\pi}{a} x\right); \quad (3.12)$$

если же n — нечетное число, т. е. $n = 2p + 1$, то поле (четное) будет

$$E_{\text{полн}} = 2E_0 e^{i[\omega t - (k \cos \theta) z]} \cos\left[\left(\frac{\pi}{2} + p\pi\right) \frac{x}{a}\right].$$

Что же означают соотношения (3.11) и (3.12)? Вдоль оси Oz мы имеем бегущую волну с *постоянной распространения в световоде*

$$\beta = k \cos \theta. \quad (3.13)$$

Вдоль оси Ox (рис. 3.6) мы имеем режим стоячих волн (струна, закрепленная в точках $x = \pm a$).

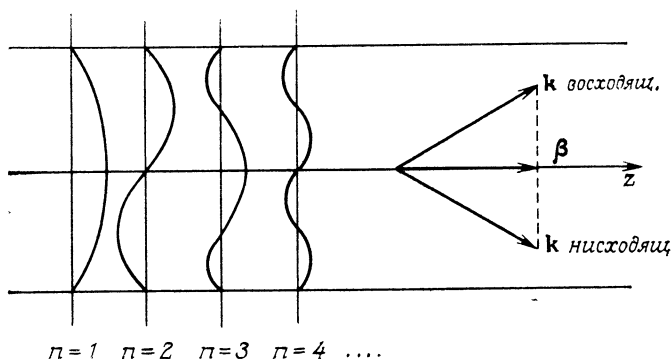


Рис. 3.6 Структура мод в одномерном металлическом световоде.

Условие (3.11) можно рассматривать по-разному.

Во-первых, как *условие согласования фаз*: если рассматривать однократное прохождение волны между плоскостями туда и обратно вдоль оси Ox , то полный фазовый сдвиг должен быть кратным 2π . В противном случае волна будет интерферировать со всеми последующими отраженными волнами и из-за гасящей интерференции быстро затухать. Под «полным» фазовым сдви-

¹⁾ Разумеется, его не следует путать с показателем преломления n .

гом следует понимать сумму фазового сдвига за счет распространения и сдвига фаз при отражениях. В металлическом световоде считать полный фазовый сдвиг легко, так как мы видели, что отражение от идеального проводника сопровождается фазовым сдвигом на π для напряженности электрического поля E_y (но не изменяет фазу напряженности магнитного поля H_z). В результате условие согласования фаз для металлического световода запишется в виде

$$(-1)^2 e^{4ika \sin \theta} = 1,$$

что и выражается соотношением (3.11).

Во-вторых, соотношение (3.11) можно рассматривать как *уравнение дисперсии*: оно позволяет определить постоянную распространения в световоде β в зависимости от частоты и геометрических параметров системы. Из (3.11) и (3.13) следует, что

$$\beta^2 = k^2 - \frac{n^2 \pi^2}{4a^2}. \quad (3.14)$$

Таким образом, в рассматриваемом световоде распространяются две плоские волны, удовлетворяющие соотношениям (3.11) и (3.12). Но встает вопрос: не образует ли вторая волна при отражении от плоскости $x = -a$ третью волну, которая при отражении от плоскости $x = +a$ в свою очередь порождает четвертую и т. д., и т. д.? Естественно, что это так. Здесь образуется как бы интерферометр Фабри—Перо [7]. Далее мы увидим (гл. 5, § 3 и гл. 12, § 1, п. А), что многократное отражение просто улучшает угловое разделение при углах θ , определяемых соотношением (3.11). Следовательно, наша модель двух волн вполне корректна. Нужно только учитывать, что каждая из этих волн есть сумма всех отраженных волн, четных или нечетных.

Б. Электромагнитное приближение

Вернемся к электромагнетизму. Можно решать нашу задачу и таким образом: искать напряженность электрического поля E_y в виде

$$E_y = E_0 f(x) e^{i(\omega t - \beta z)},$$

где β и $f(x)$ — неизвестные. Решив уравнение распространения (2.1), получим

$$\frac{f''(x)}{f(x)} = \beta^2 - \frac{\omega^2}{c^2}. \quad (3.15)$$

Поскольку оба члена в правой части здесь константы, запишем это равенство в виде

$$\frac{f''(x)}{f(x)} = -\gamma^2. \quad (3.16)$$

Решение данного уравнения таково:

$$f(x) = A \sin \gamma x + B \cos \gamma x.$$

Из граничных условий $f(\pm a) = 0$ следует равенство

$$A \sin \gamma a = B \cos \gamma a = 0.$$

Иначе говоря,

$$\text{если } \begin{cases} B = 0, \\ \sin \gamma a = 0, \quad \gamma = p \frac{\pi}{a}, \end{cases}$$

то $f(x) = A \sin \left(p \frac{\pi}{a} x \right)$ — нечетное решение;

$$\text{если } \begin{cases} A = 0 \\ \cos \gamma a = 0, \quad \gamma = \frac{1}{a} \left(\frac{\pi}{2} + p\pi \right), \end{cases}$$

то $f(x) = B \cos \left(\frac{\pi}{2} + p\pi \frac{x}{a} \right)$ — четное решение.

С другой стороны, из (3.15) и (3.16) следует соотношение

$$\beta^2 = \frac{\omega^2}{c^2} - \gamma^2,$$

т. е. уравнение дисперсии (3.14), в котором γ — величина, определенная выше.

Таким путем можно вновь получить оптическое приближение, о котором говорилось ранее. Достаточно учесть, что величина β должна быть действительной. Следовательно, должно выполняться условие

$$k^2 > \gamma^2;$$

положив

$$\sin \theta = \frac{\gamma}{k}, \quad (3.17)$$

получаем соотношение (3.11), в котором

$$\gamma = \frac{n\pi}{2a}, \quad (3.18)$$

откуда и вытекает интерпретация «плоских волн».

В. Моды распространения

Как же на практике формулируется задача о распространении волн в таком световоде? Имеется генератор, способный вводить в волновод волну заданной частоты $f = \omega/2\pi$. Нужно определить, под какими углами θ волны будут распространяться в волноводе.

Положим

$$\omega_c = \frac{\pi c}{2a}, \quad (3.19)$$

тогда (3.11) запишется в виде

$$\sin \theta = n \frac{\omega_c}{\omega}. \quad (3.20)$$

Отсюда следует, что при $\omega < \omega_c$ распространение волн невозможно. Частота ω_c называется *критической частотой*. При

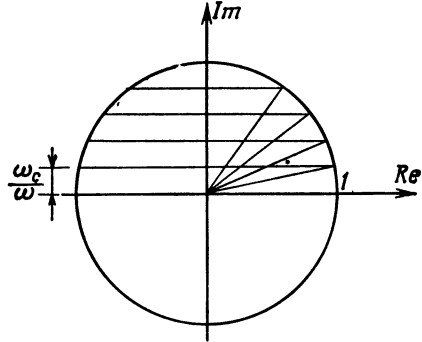


Рис. 3.7. Определение числа возможных мод N на частоте $f = \omega/2\pi$.

Построение состоит в том, чтобы разбить мнимую ось тригонометрической окружности на интервалы, равные ω_c/ω

$\omega > \omega_c$ решения получаются простым геометрическим построением, показанным на рис. 3.7. Из рисунка явствует, что при

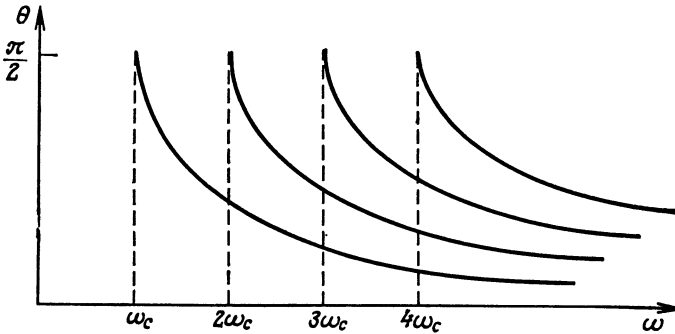


Рис. 3.8. Металлический световод.

Зависимость угла падения плоских волн от частоты для различных мод.

$N\omega_c \leq \omega \leq (N+1)\omega_c$ (где N — целое положительное число) мы имеем N решений:

$$\theta_n = \arcsin \left(n \frac{\omega_c}{\omega} \right), \quad \text{где } n = 1, 2, \dots, N. \quad (3.21)$$

Каждое такое решение называется *модой распространения*. Зависимости углов θ от частоты приведены на рис. 3.8.

При всех частотах, кратных ω_c , появляется дополнительная мода. На частотах, кратных $2\omega_c$, имеем $\theta = \pi/2$, т. е. волна поперечная и распространения нет. Но с увеличением частоты наклон волн уменьшается. В пределе, когда частота ω очень

велика (или когда длина волны λ очень мала), получается квазиосевая волна. Это, впрочем, и понятно, так как, поскольку отражающие стенки волновода находятся очень далеко друг от друга (расстояние между ними нужно измерять в длинах волн), их влиянием можно практически пренебречь. Следовательно, аксиальная квазиплоская волна может распространяться.

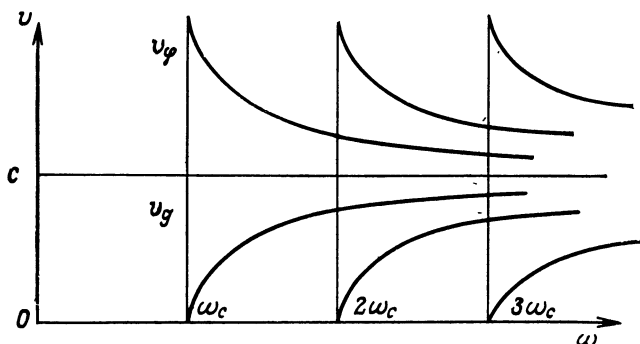


Рис. 3.9. Металлический световод.

Кривые дисперсии.

Фазовая скорость волн находится в прямой зависимости от их угла наклона, чем определяется величина $|\beta|$. Точнее (гл. 2, § 1),

$$v_\varphi = \frac{\omega}{\beta} = \frac{\omega}{k \cos \theta},$$

или с учетом (3.20)

$$v_\varphi = \frac{c}{\sqrt{1 - n^2 \frac{\omega_c^2}{\omega^2}}}.$$

Вычислим соответствующую групповую скорость:

$$v_g = \frac{d\omega}{d\beta} = c \sqrt{1 - n^2 \frac{\omega_c^2}{\omega^2}}. \quad (3.22)$$

Зависимость скоростей v_φ и v_g от частоты показана на рис. 3.9. Здесь мы видим, что *фазовая скорость больше скорости света c*! Это объясняется в приложении 1. Там же полученные ранее результаты объясняются на основе представления о длине волны. Выводится соотношение

$$\frac{1}{\lambda^2} = \frac{1}{\lambda_z^2} + \frac{n^2}{\lambda_c^2}, \quad (3.23)$$

где

$$\frac{\lambda_c}{n} = \frac{4a}{n} - \text{критическая длина волны}, \quad (3.24)$$

$$\lambda_z = \frac{2\pi}{\beta} - \text{кажущаяся длина волны вдоль оси } Oz. \quad (3.25)$$

Для моды порядка n величина λ_c/n равна расстоянию между двумя узлами вдоль оси Ox .

Вернемся к вопросу, поставленному раньше. Какая мода будет распространяться в световоде при заданной частоте? Мы показали, что на каждой частоте *может* распространяться определенное число мод. Общее решение будет линейной комбинацией этих мод с коэффициентами, зависящими в основном от условий на концах световода (излучатель и приемник).

При $\omega_c \leq \omega \leq 2\omega_c$ мы имеем *одномодовый* режим распространения. При более высоких частотах распространяется одновременно несколько мод. Каждая мода является дисперсной, и, что особенно важно, у каждой моды своя собственная дисперсия. На основании формулы (2.44) получаем для каждой моды

$$\Delta\omega = \frac{1}{n} \frac{1}{c\omega_c} \frac{1}{\sqrt{l}} (\omega v_g)^{3/2}. \quad (3.26)$$

Данной формулой выражается влияние *дисперсии моды*, и на этом мы завершаем изучение явления дисперсии, начатое в предыдущей главе.

Как нетрудно видеть, теоретически выгодно работать как можно ближе к одномодовому режиму на сравнительно низких частотах, поскольку произведение ωv_g есть возрастающая функция частоты.

Г. Затухающая волна

Что же происходит на частотах, меньших критической частоты? Поскольку при этом $\sin \theta$ получается больше 1, величина θ оказывается мнимой, так же как и λ_z . Если световод бесконечно длинный вдоль оси Oz , то мы имеем *затухающую волну*, так как

$$\cos \theta = \sqrt{1 - \frac{n^2 \lambda^2}{\lambda_c^2}} = \pm i \sqrt{\frac{n^2 \lambda^2}{\lambda_c^2} - 1}, \quad (3.27)$$

и выражение (3.12) принимает вид

$$E_{\text{полн}} = E_0 e^{i\omega t} e^{-kz} \sqrt{\left(\frac{n^2 \lambda^2}{\lambda_c^2} - 1\right)} \begin{cases} \sin\left(p \frac{\pi x}{a}\right), \\ \cos\left(\frac{\pi}{2} + p\pi\right) \frac{x}{a}, \end{cases}$$

откуда глубина проникновения волны в световод

$$\delta = \frac{\lambda}{2\pi \sqrt{\frac{n^2 \lambda^2}{\lambda_c^2} - 1}} = \frac{c}{\omega \sqrt{\frac{n^2 \omega_c^2}{\omega^2} - 1}}.$$

Таким образом, глубина проникновения δ тем меньше, чем дальше частота ω от критической частоты.

Затухающая волна подробнее рассматривается в приложении 2. Там показывается, что ее можно представить в виде суммы двух плоских неоднородных волн.

Теперь от металлических световодов перейдем к диэлектрическим.

§ 3. Диэлектрический световод

Рассмотрим (рис. 3.10) диэлектрический световод в виде диэлектрической полосы толщиной $2a$ с диэлектрической проницаемостью ϵ_1 (показатель преломления n_1) между двумя диэлектрическими полупространствами с диэлектрической проницаемостью ϵ_2 (показатель преломления n_2). Предположим, что

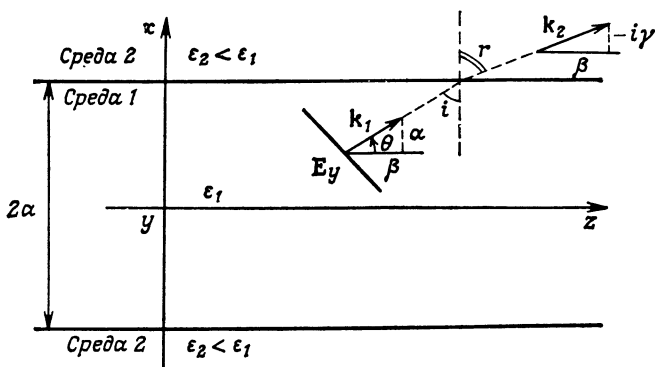


Рис. 3.10. Одномерный диэлектрический световод.

$\epsilon_2 < \epsilon_1$, т. е. $n_2 < n_1$. Это условие, очевидно, необходимо для того, чтобы в рассматриваемом волноводе распространялись волны. Известно, что свет отклоняется в сторону более высокого показателя преломления (гл. 2, § 2), а потому показатель преломления внутренней среды должен быть больше показателя преломления внешней среды. Мы будем рассматривать случай симметричного диэлектрического световода (обе внешние среды одинаковы), что облегчит нашу задачу.

Здесь можно повторить все сказанное относительно симметрии системы в случае металлического световода, и мы будем, так же как и там, рассматривать моды TE .

Попробуем, как и раньше, ввести плоскую волну внутрь световода. Ее волновой вектор будет определяться соотношением

$$\mathbf{k}_1 = \begin{cases} \beta & (\text{по оси } Oz), \\ \alpha & (\text{по оси } Ox), \end{cases} \quad |\mathbf{k}_1| = kn_1.$$

Волна преломится на границе сред, и в среде 2 ее волновой вектор будет таким:

$$\mathbf{k}_2 = \begin{cases} \beta & (\text{по оси } Oz), \\ -i\gamma & (\text{по оси } Ox), \end{cases} |\mathbf{k}_2| = kn_2.$$

Разумеется, проекции этих векторов на ось Oz одинаковы (равны β), поскольку вдоль всей плоскости раздела сред сохраняется равенство фаз.

Но нам необходимо полное внутреннее отражение, так как только в этом случае волна будет отражаться вверх и вниз на границах раздела и в результате оставаться внутри световода. Допустим, что угол падения волны удовлетворяет условию

$$i > i_c,$$

причем [формула (3.4)]

$$\sin i_c = \frac{n_2}{n_1} = \eta,$$

или, что эквивалентно, условию

$$\theta < \theta_c,$$

причем

$$\cos \theta_c = \eta.$$

В этом случае коэффициент отражения r становится мнимым и применима формула (3.5):

$$\cos r = \pm i \sqrt{\frac{\sin^2 i}{\eta^2} - 1} = \pm i \sqrt{\frac{\cos^2 \theta}{\eta^2} - 1}.$$

Выберем знак «—». Тогда составляющая вектора \mathbf{k}_2 , параллельная оси Ox , запишется в виде

$$kn_2 \cos r = -i\gamma,$$

где

$$\gamma = kn_2 \sqrt{\frac{\cos^2 \theta}{\eta^2} - 1}, \quad (3.28)$$

а напряженность электрического поля в среде 2 будет равна

$$E_y^{(2)} = E_0 e^{i(\omega t - \beta z)} e^{-\gamma x},$$

где

$$\beta = kn_2 \sin r = kn_1 \cos \theta. \quad (3.29)$$

Мы получили неоднородную волну, свойства которой нам хорошо известны. Ее амплитуда экспоненциально уменьшается при удалении в поперечном направлении от границ световода. Наш выбор знака в выражении (3.5) в действительности соответствует отражению от плоскости $x = +a$. Для отражения от

плоскости $x = -a$ нужно было бы взять другой знак, чтобы получить величину $e^{+\gamma x}$, удовлетворяющую условию

$$\lim_{x \rightarrow -\infty} e^{+\gamma x} = 0.$$

Итак, по обе стороны световода имеются две экспоненциально затухающие во внешних средах волны. А что же внутри? Внутри световода, по крайней мере формально, все происходит так же, как и в металлическом световоде. Суперпозиция падаю-

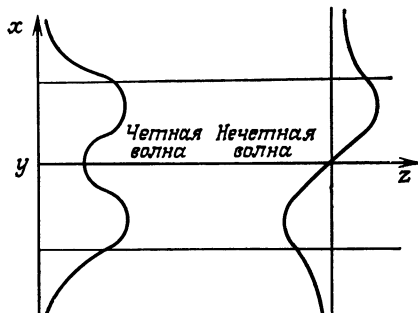


Рис. 3.11. Поперечные изменения составляющих $E(y)$.
Пример четной и нечетной волн [формулы (3.29) и (3.30)].

щей и отраженной волн даст нам *бегущую волну вдоль оси Oz* и *стоячую волну вдоль оси Ox* (рис. 3.11). При этом

$$E_y^{(1)} = E_0 e^{i(\omega t - \beta z)} \begin{cases} \cos \alpha x & \text{— «четная» волна,} \\ \sin \alpha x & \text{— «нечетная» волна.} \end{cases} \quad (3.30)$$

Можно сформулировать задачу следующим образом: известны характеристики световода a , n_1 , n_2 и частота волны ω (или, что эквивалентно, вектор \mathbf{k}). Нужно найти θ , α , β и γ .

А. Уравнения дисперсии и условие согласования фаз

На основании определения векторов \mathbf{k}_1 и \mathbf{k}_2 можно сразу написать

$$\begin{aligned} k_1^2 &= k^2 n_1^2 = \beta^2 + \alpha^2, \\ k_2^2 &= k^2 n_2^2 = \beta^2 - \gamma^2. \end{aligned} \quad (3.31)$$

Отсюда, исключив β , получаем

$$\alpha^2 + \gamma^2 = k^2 (n_1^2 - n_2^2). \quad (3.32)$$

Напишем теперь условие согласования фаз. Пусть Φ — фазовый сдвиг, возникающий при отражении от каждой границы раздела сред [формула (3.7)] как снизу, так и сверху. При-

равняем полный сдвиг фазы при прохождении волны от нижней стенки до верхней и обратно целому кратному числу 2π :

$$4kn_1a \sin \theta + 2\Phi = 2n\pi. \quad (3.33)$$

Чему же равен фазовый сдвиг Φ ? Вспомним выражение (3.7):

$$\Phi = -2 \operatorname{arctg} i \frac{Z_1}{Z_2} = -2 \operatorname{arctg} \left(i \frac{\eta \cos r}{\cos i} \right),$$

и с учетом выражений (3.5), (3.28) и определения

$$\alpha = kn_1 \sin \theta \quad (3.34)$$

найдем

$$\Phi = -2 \operatorname{arctg} \left(\frac{\gamma}{\alpha} \right). \quad (3.35)$$

Можно показать, что в действительности все происходит так, как если бы отражение происходило от металлической поверхности, расположенной на некотором расстоянии x_0 далее границы раздела двух сред (приложение 3). Таким образом, световые пучки смещаются при отражении от границы раздела двух сред (*эффект Гоосса — Хенхена* [8—13]).

Б. Решение задачи распространения, моды

Вернемся к уравнению дисперсии (3.33). Подставим в него выражения (3.34) и (3.35). Получаем два случая:

$$\begin{aligned} \text{если } n = 2p, \text{ то } \alpha \operatorname{tg} \alpha &= \gamma, \\ \text{если } n = 2p + 1, \text{ то } -\alpha \operatorname{ctg} \alpha &= \gamma. \end{aligned} \quad (3.36)$$

Итак, в общем случае нахождение зависимости α , β , γ и θ от частоты не простая задача, ибо мы приходим к трансцендентным уравнениям. Далее будем рассматривать случай $n = 2p$, учитывая, что при $n = 2p + 1$ рассуждения будут аналогичными.

Построим зависимость $\gamma\alpha$ от $\alpha\alpha$ (рис. 3.12). В результате приходим к следующему графическому решению: поскольку α и γ должны одновременно удовлетворять уравнениям (3.32) и (3.36), при каждом значении частоты они определяются точкой пересечения построенной кривой с дугой окружности, которая описывается уравнением (3.32) (естественно, мы ищем положительные решения для γ и α). При увеличении частоты радиус такой окружности увеличивается и, как нетрудно видеть, значения α и γ при этом также увеличиваются. Диаграмма, представленная на рис. 3.12, ясно показывает, как появляются разные моды распространения.

Так же как и в случае металлического световода, значения $n = 2p$ соответствуют нечетным модам, а значения $n = 2p +$

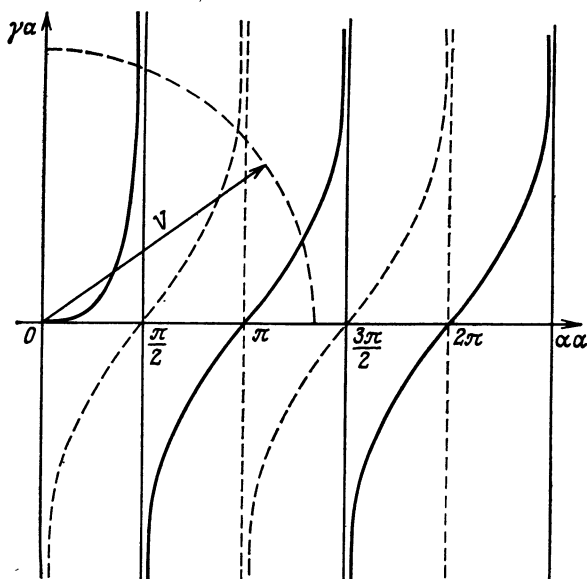


Рис. 3.12. Изменение мод в зависимости от приведенной частоты V .
Графическое построение зависимостей γ и α от частоты.

$+1$ — четным модам. Для удобства часто вводят *приведенную частоту*

$$V = ka \sqrt{n_1^2 - n_2^2} \quad (3.37)$$

(мы будем пользоваться ею в гл. 8). Окончательно характеристические уравнения мод записываются в виде [формула (3.36)]

$$aa \operatorname{tg} aa = \gamma a,$$

$$-aa \operatorname{ctg} aa = \gamma a,$$

или [формула (3.32)]

$$(\alpha^2 + \gamma^2) a^2 = V^2.$$

Схема вычисления полей в диэлектрическом световоде дается в приложении 4.

В. Поведение мод при изменении частоты, критические частоты

Поведение каждой моды при изменении частоты можно исследовать следующим образом. Критические частоты удовлетворяют условиям

$$aa = \frac{n\pi}{2},$$

$$\gamma = 0,$$

откуда получаем

$$\omega_c = \frac{\pi c}{2a \sqrt{n_1^2 - n_2^2}}, \quad (3.38)$$

$$\lambda_c = 4a \sqrt{n_1^2 - n_2^2}. \quad (3.39)$$

Новые моды появляются при всех частотах $n\omega_c$ ($n = 0, 1, 2, \dots$). Здесь наблюдается существенное отличие от металлического световода, а именно существование нулевой моды с $n = 0$. Для диэлектрического световода нет частотного порога. В самом деле, можно получить нулевой суммарный фазовый сдвиг в (3.33) за один проход от нижней стенки до верхней и обратно, так как, в отличие от металлического световода, в диэлектрическом световоде сдвиг фазы при распространении может быть скомпенсирован изменением фазы Φ при отражении.

Когда возникает некая мода распространения, мы имеем $\gamma = 0$, откуда в силу формулы (3.28) $\cos \theta_c = \eta$, т. е. угол — предельный. В этом случае распространение волны возможно, так как мы находимся на пределе преломления. Но условие $\gamma = 0$ означает, что волна полностью распространяется вне волновода, т. е. в среде 2¹⁾. Поэтому фазовая скорость равна величине $v_\phi^{(2)}$, т. е. фазовой скорости во внешней среде. При увеличении частоты величины α и γ возрастают, а угол θ убывает. Поле не так глубоко проникает во внешнюю среду, и волна концентрируется внутри световода. В пределе, когда частота ω стремится к бесконечности, величина γ также устремляется к бесконечности, $\theta = 0$ и волна полностью удерживается в среде 1. При этом ее фазовая скорость равна $v_\phi^{(1)}$. Такое изменение структуры моды распространения с увеличением частоты схематически показано на рис. 3.13.

В частном случае нулевой моды $n = 0$ показывается, что

$$aa \sim V \text{ при } \omega \rightarrow 0,$$

и мы снова получаем $\theta = \theta_c$ при $\omega = 0$.

Но, согласно (3.34), на критической частоте выполняются соотношения

$$\sin \theta_c = \frac{\alpha}{kn_1} = \frac{n\pi}{2a} \frac{\lambda}{2\pi} = \frac{n\lambda}{4a} = n \frac{\omega_c}{\omega},$$

где $\omega_c = \pi c / 2an_1$, т. е. ω_c — это критическая частота металлического световода, заполненного диэлектрической средой

¹⁾ Точнее говоря, электрическое поле во внешней среде равно нулю на критической частоте. Но вблизи критической частоты имеется слабое электрическое поле, однородное во внешней среде.

с показателем преломления n_1 . Предельный угол падения i_c такой же, как и в металлическом световоде для той же частоты.

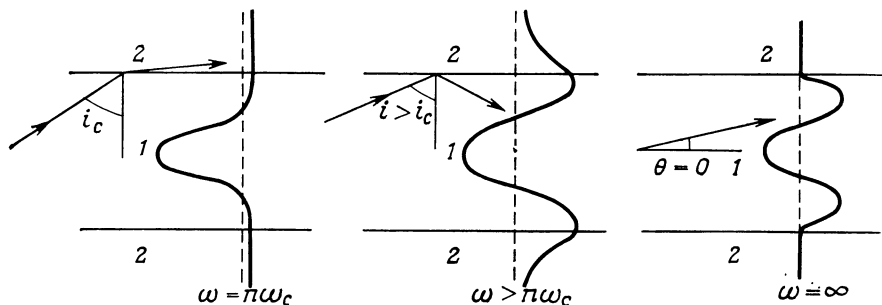


Рис. 3.13. Изменение структуры мод в зависимости от частоты ω .

На критической частоте поле внутри диэлектрического световода точно такое же, как в соответствующем металлическом световоде.

Г. Дисперсия

Как говорилось выше, при увеличении частоты ω от критической до бесконечности фазовая скорость волны изменяется от

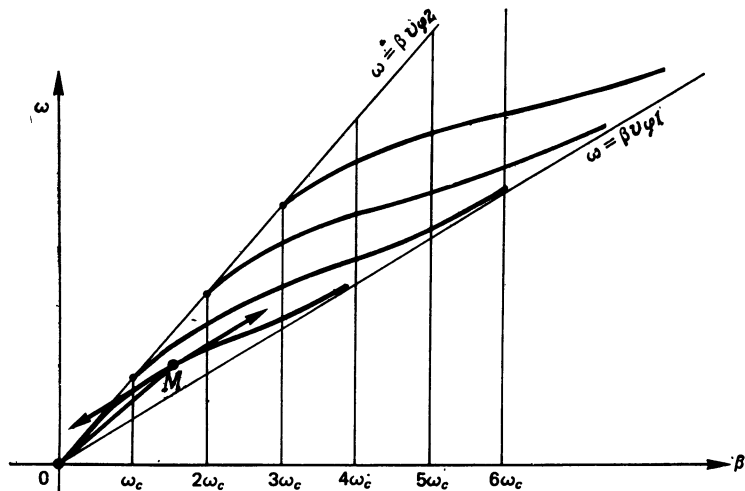


Рис. 3.14. Диэлектрический световод (кривые дисперсии).

Для данной моды на заданной частоте (точка M) фазовая скорость v_ϕ определяется наклоном прямой OM , а групповая скорость — наклоном касательной в точке M .

$v_\phi^{(2)}$ до $v_\phi^{(1)}$. Следовательно, диаграмма дисперсии $\omega(\beta)$ будет иметь такой вид, как на рис. 3.14.

Легко показать, что первый слева участок этой диаграммы соответствует всем углам $\theta > \theta_c$, т. е. континууму излучаемых мод. Поскольку такие моды распространяются в среде 2, они нас не интересуют. Всегда стараются либо их уменьшить, либо исключить совсем, чтобы можно было работать только с модами, *распространяющимися* в волноводе, которые образуют дискретный набор.

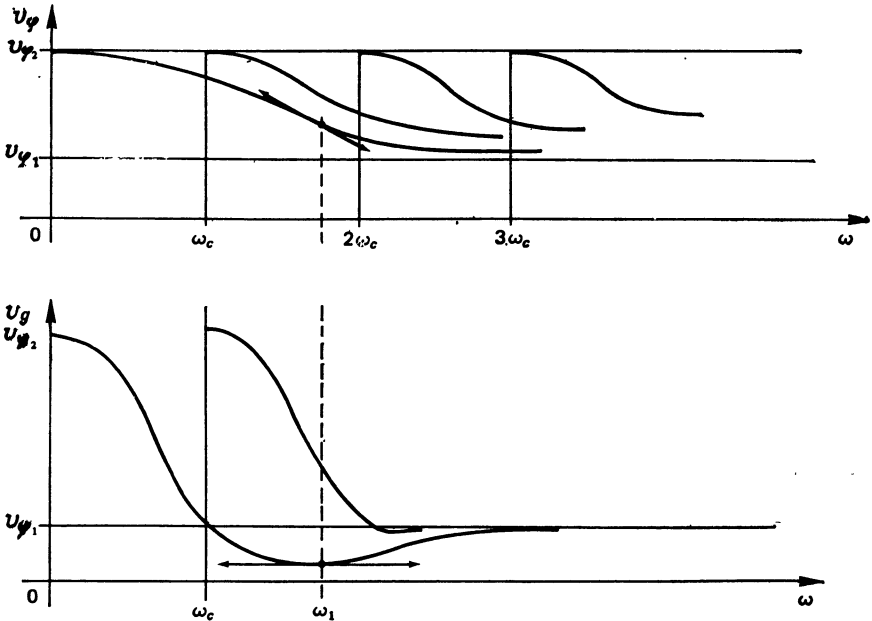


Рис. 3.15. Диэлектрический световод (зависимость фазовой и групповой скоростей от частоты).

Частота ω_1 соответствует точке перегиба кривой $v_\varphi(\omega)$ для основной моды.

Диаграмма рис. 3.14 позволяет легко определить зависимость фазовой и групповой скоростей от частоты ω . Действительно, $v_\varphi = \omega/\beta$ — коэффициент наклона прямой OM , а $v_g = d\omega/d\beta$ — коэффициент наклона касательной в точке M .

В результате получаем дисперсионные характеристики (рис. 3.15) диэлектрического световода.

Д. Одномодовый и многомодовый режимы распространения волн

Речь идет об одном важном различии. Казалось бы, что с точки зрения дисперсии желателен, как и в случае металлического волновода, одномодовый режим. Для этого, очевидно, необходимо обеспечить условие $\omega < \omega_c$. Максимальный размер

такого одномодового световода в соответствии с (3.39) определяется выражением

$$(4a)_{\text{одномод}} \leq \frac{\lambda}{\sqrt{n_1^2 - n_2^2}}.$$

Таким образом, можно получить одномодовый световод с поперечными размерами во много длин волн, если достаточно мала разность показателей преломления $\Delta n = n_1 - n_2$. То же самое будет показано в случае цилиндрических оптических волокон (гл. 8, § 4, п. Б, 1 и гл. 8, § 8).

Как нетрудно видеть (рис. 3.15), кривая зависимости групповой скорости каждой моды от частоты проходит через минимум в точке, соответствующий точке перегиба кривой $\omega(\beta)$. Поэтому не исключено, что окажется невозможным свести к минимуму групповую скорость для основной моды в одномодовом режиме. По этой причине или по какой-либо другой причине практического характера может оказаться необходимым работать в многомодовом режиме и учитывать одновременно дисперсию каждой моды и многомодовые эффекты.

Конечно, все это сложно. Тем более что в отличие от металлического волновода здесь нет точных формул. Очень часто приходится пользоваться приближенными выражениями. Так, например, можно рассмотреть случай $n = 0$ при сравнительно низких частотах. В этом случае доказывается, что

$$\gamma a \sim V^2 \text{ при } \omega \rightarrow 0.$$

Исходя из этого, непосредственно выводится приближенное выражение для $\beta(\omega)$. Далее решение задачи не встречает затруднений.

Е. Расширение волнового пакета

В гл. 2, § 5, п. Б мы рассматривали это явление с учетом только членов первого порядка по $d\beta/d\omega$. Позже мы ввели задним числом коэффициент второго порядка

$$\delta = \left| \frac{\partial v_g}{\partial \omega} \right|$$

и получили выражение (2.44). В действительности было бы интересно сохранить в разложении показателя экспоненты $i(\omega t - \beta z)$ в интеграле Фурье (2.40) члены до второго порядка по $d^2\beta/d\omega^2$, чтобы получить точную форму распространяющегося волнового пакета, так как деформация волнового пакета обусловлена членом второго порядка. Правда, в общем случае получающиеся выражения слишком сложны. Но вычисления возможны в частном случае *гауссова волнового пакета*. Предостав-

ляем читателю самому проверить (в качестве упражнения) следующий результат.

Гауссов волновой пакет длительностью T имеет вид

$$g(t) = e^{-\left(\frac{t}{T}\right)^2}.$$

Для моды TE_0 в пределе низких частот ($\omega \rightarrow 0$) ширина импульса T увеличивается в $\sqrt{2}$ раз на характеристическом расстоянии [14]

$$L = \frac{c^3 T^2}{6\omega a^2} \frac{n_2}{(n_1^2 - n_2^2)^2}.$$

Мы видим, что

а) длительность импульса входит в это выражение в квадрате;

б) частота входит в первой степени;

в) поскольку разность показателей преломления играет здесь очень важную роль, с точки зрения дисперсии выгодно использовать среды с очень близкими показателями преломления.

Не будем углубляться в этот вопрос. На практике нужно учитывать и другие факторы (световоды не обязательно симметричны, оптические волокна цилиндрические и т. д.), а потому изучение распространения даже малого числа мод — сложная задача. В гл. 8 мы рассмотрим модель многомодового оптического волокна, поперечные размеры которого велики по сравнению с длиной волны, так что к нему применимы методы геометрической оптики.

ГЛАВА 4

ТЕОРИЯ СКАЛЯРНОЙ ОПТИКИ

В предыдущих главах мы видели, как, исходя из уравнений Максвелла, постепенно приходят к оптическим понятиям. Так, из понятия напряженности электрического поля мы вывели понятия комплексной амплитуды и интенсивности света. При этом мы постоянно опирались на представление о плоской волне, которая, как нам известно, есть некая теоретическая идеализация реальной действительности. В данной главе мы продвинемся несколько далее в изучении оптической действительности. Мы увидим, что на деле параллельный пучок не остается долго параллельным.

В этой главе мы установим основные физические характеристики, связанные с дифракцией, которые будут использованы далее для исследования оптических методов обработки информации (гл. 5—7). Затем мы введем новое понятие гауссова пучка — основного теоретического инструмента анализа процесса формирования лазерных мод (гл. 12) и изучения распространения света в свободном пространстве и в световодах (гл. 8, § 1 и 8). В завершение дадим некоторые сведения о когерентности, необходимые для изучения голографии (гл. 5 и 7) и лазеров (гл. 12).

Строгое решение задач оптики основано на системе векторных уравнений с дополнительными условиями (рис. 4.1). В эту систему уравнений входят уравнения Максвелла (1.1)—(1.5) и определяющие соотношения (1.13)—(1.15). Условия бывают разных типов. Во-первых, задаются источники, т. е. распределение токов и зарядов или параметры волны, падающей на рассматриваемую область. Во-вторых, задаются условия вблизи поверхностей, ограничивающих изучаемую область [формулы (1.18)—(1.21)]. В-третьих, если эта область бесконечна, необходимо учитывать условие излучения на бесконечности, в котором, как отмечалось ранее, находит выражение закон сохранения энергии [2, 9]. Все это достаточно сложно. В лучшем случае можно надеяться найти векторные уравнения для каждого вектора напряженности или индукции поля, делая при этом хитрые исключения и учитывая специфические ограничения (например, симметрию) и т. п. Другой классический подход состоит в определении некой скалярной величины, приближенно удовлетворяющей нашей системе уравнений. Такой подход приводит к теориям Кирхгофа и Рэлея — Зоммерфельда, прекрасно изло-

женным в литературе [15]. Мы же выберем аксиоматический подход, который позволит достаточно быстро получить основные результаты.

Предположим, что *искомые световые характеристики полностью определяются некой (скалярной) комплексной величиной*. Такой величиной может быть, например, комплексная амплитуда. Далее в данной главе мы будем исходить из этого предположения, которое называют *предположением скалярной оптики*.

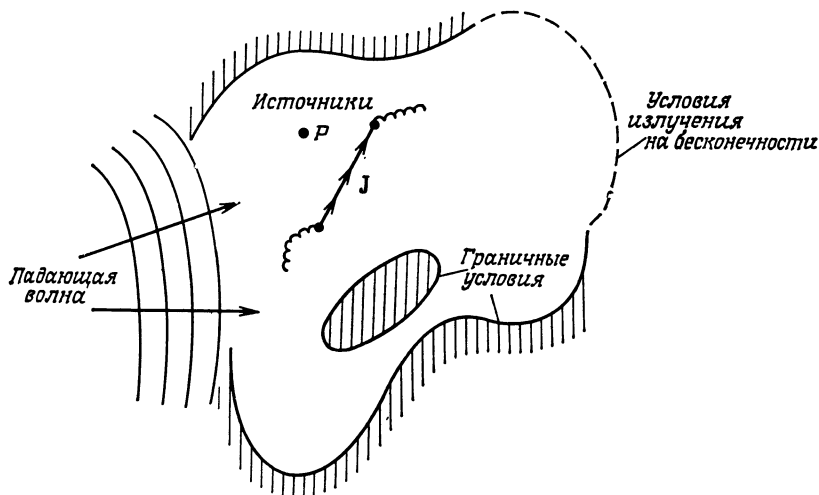


Рис. 4.1. Схема общего случая решения задачи распространения волны в пространстве, содержащем источники излучения и имеющем некоторые граничные условия.

Само собой разумеется, такое допущение справедливо не во всех случаях, но, как мы увидим, оно позволяет описать большое число явлений. Ограниченный объем книги не дает возможности подробно изложить теорию поляризации, а поэтому мы рекомендуем заинтересованному читателю обратиться к другим работам (например, [4, 7]). (Все, что касается поляризации, будет сказано по мере необходимости в гл. 8, § 4, 7 и 8, п. Д; гл. 14, § 1 и 3).

§ 1. Переход от сферической волны к плоской волне на основе дифракции

А. Переход от сферической волны к плоской волне

Рассмотрим сначала светящуюся монохроматическую «точку» (рис. 4.2). Этот «точечный» источник, испускающий лучи радиально, формирует идеальную сферическую волну

(предполагается, что пространство однородно и изотропно). Известно, что такого источника в природе не существует. Но всегда можно представить себе очень малое отверстие. Например, очень малое по сравнению с длиной волны отверстие в непрозрачном экране, который освещается плоской волной (рис. 4.3).

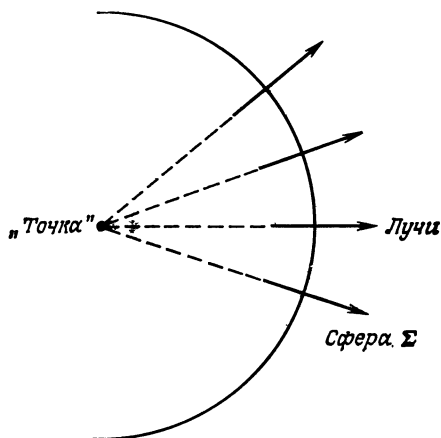


Рис. 4.2. «Точечный» источник. Излучаемая волна идеально сферическая, «лучи» расходящиеся.

Такую ситуацию легко смоделировать в бассейне с водой. Проведем следующий опыт: в бассейне с водой создадим волны, установим два барьера и посмотрим, как будут проходить

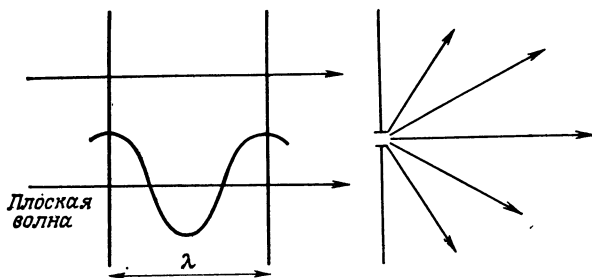


Рис. 4.3. Дифракция плоской волны на малом (по сравнению с длиной волны λ) отверстии.

волны между барьерами при уменьшении расстояния между ними. Очевидно, что постепенно определяющее значение приобретут краевые эффекты. Конечно, это не доказательство. Мы просто сформулируем следующий принцип: *точечный источник излучает свет равномерно во всех направлениях*.

Расширим отверстие, пока оно не превратится в «окно». Рассмотрим сразу случай очень большого (по сравнению с длиной волны λ) окна. Окно будет настолько большим, что влия-

нием его краев можно пренебречь (рис. 4.4). При прохождении через такое окно «лучи», несмотря на наличие его стенок, продолжают распространяться прямолинейно и остаются параллельными друг другу. Это значит, что применима *геометрическая оптика*.

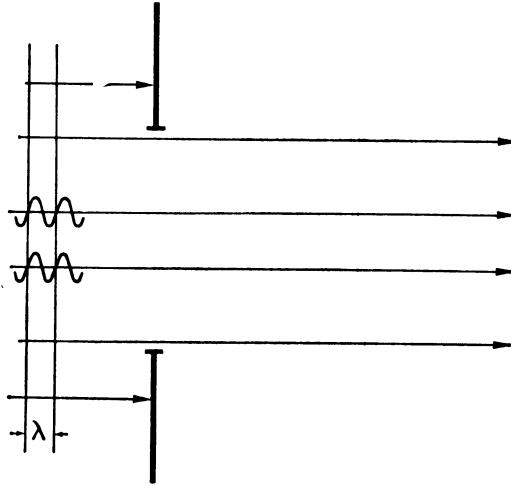


Рис. 4.4. Случай «большого» (по сравнению с длиной волны) окна.

Когда размер окна не очень велик по сравнению с длиной волны λ , справа от окна возникает «волнение»: происходит *дифракция*.

Б. Задача скалярной дифракции

Мы рассмотрим формальный подход, позволяющий быстро получать основные результаты скалярной теории дифракции.

Рассмотрим (рис. 4.5) плоское окно с функцией пропускания $f(x, y)$, освещенное плоской волной с длиной волны λ . Пространство предполагается *однородным и изотропным*. Считаем, что известна комплексная амплитуда слева от окна (при $z < 0$ волна плоская) и на самом окне [при $z = 0$ комплексная амплитуда будет равна $A_0 f(x, y)$, где A_0 — амплитуда плоской волны]. Рассмотрим распределение света на бесконечности в направлении единичного вектора \mathbf{n} с направляющими косинусами (α, β, γ) , удовлетворяющими условию

$$\alpha^2 + \beta^2 + \gamma^2 = 1. \quad (4.1)$$

Искомое распределение амплитуд будет функцией переменных α и β . Какова же эта функция $F(\alpha, \beta)$?

Подчеркнем, что далее мы будем наблюдать распределение комплексных амплитуд $F(\alpha, \beta)$ на *сфере бесконечного радиуса* S , а не на бесконечно удаленной плоскости P , и это приведет к заметным отличиям.

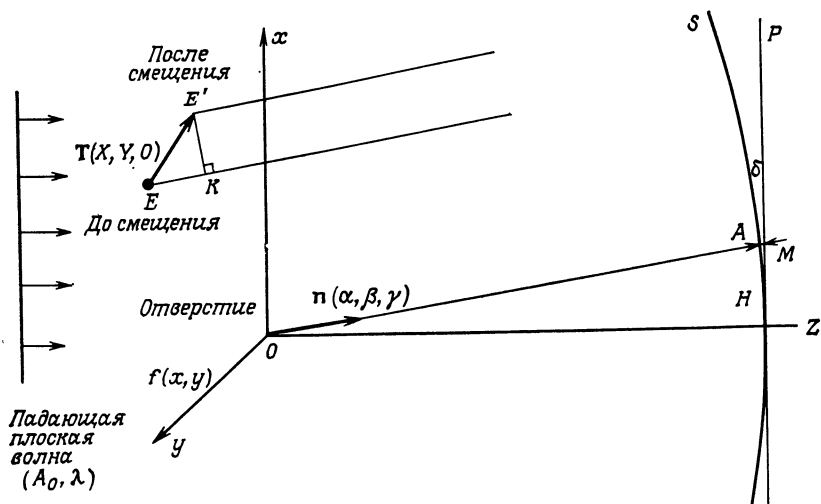


Рис. 4.5. Дифракция плоской монохроматической волны на окне $f(x, y)$, рассматриваемая на сфере бесконечного радиуса.

E — излучающая точка до смещения; E' — излучающая точка после смещения (на вектор T). Наблюдение ведется вдоль прямой с направляющими косинусами α, β, γ .

В силу однородности и изотропности пространства преобразование

$$f(x, y) \rightarrow F(\alpha, \beta)$$

будет, очевидно, *линейным*. Действительно, при любых a, b, f, g мы можем написать

$$af(x, y) + bg(x, y) \rightarrow aF(\alpha, \beta) + bG(\alpha, \beta).$$

Необходимо, однако, отметить, что это преобразование *не инвариантно по отношению к сдвигу*. Поместим в бесконечности на ось $\mathbf{n}(\alpha, \beta, \gamma)$ наблюдателя (рис. 4.5) и сместим окно в его плоскости в направлении вектора T (составляющие $X, Y, 0$). Можно было бы предположить, что наблюдатель, находящийся в бесконечности, не заметит конечного перемещения. Но это не так! После сдвига каждая излучающая точка E окна будет излучать свет в направлении \mathbf{n} по отличным от первоначальных траекториям (пропорционально отрезку EK на рис. 4.5). Очевидно, что смещение окна на величину T сопровождается появ-

лением линейно изменяющегося фазового сдвига

$$\varphi_1 = \frac{2\pi}{\lambda} \mathbf{T} \cdot \mathbf{n} = \frac{2\pi}{\lambda} (\alpha X + \beta Y)$$

(напомним, что в этом выражении величины α и β — два первых направляющих косинуса вектора \mathbf{n}).

На самом деле наблюдатель воспринимает только интенсивность света, и он не заметит сдвига окна.

В. Принцип Гюйгенса — Френеля и преобразование Фурье

Разобьем окно на очень малые, почти точечные площадки. Предположим, что каждая такая точечная площадка излучает в пространство сферическую волну синфазно с падающей волной. В этих двух предположениях и состоит принцип Гюйгенса — Френеля.

Теперь достаточно применить принцип *линейности* (суперпозиции), чтобы прийти к *преобразованию Фурье*. Каждый точечный источник излучает волну с амплитудой $f(x, y)$. При распространении в направлении \mathbf{n} световая волна, вышедшая из какой-либо точки (x, y) , получает соответствующий фазовый сдвиг относительно волны, вышедшей из начала координат. Таким образом, с учетом сказанного в предыдущем пункте параграфа мы можем написать

$$F(\alpha, \beta) = \iint_{-\infty}^{+\infty} f(x, y) e^{-\frac{2\pi}{\lambda} i(\alpha x + \beta y)} dx dy. \quad (4.2)$$

Написав (4.2), мы в неявной форме приняли, что выполняются следующие условия.

1. Гипотеза существования фурье-образа

Предполагается, что у функции $f(x, y)$ имеется фурье-образ и мы можем написать выражение (4.2).

2. Условие наблюдения

Напомним, что функция $F(\alpha, \beta)$ получена на сфере S бесконечного радиуса. Каково же распределение света на плоскости P ? Чтобы ответить на этот вопрос, достаточно вычислить расстояние δ от точки наблюдения M до точки на сфере S (рис. 4.6)

$$MH^2 = \rho^2 = MA \cdot MB,$$

$$\text{откуда } \delta = AM \approx \frac{\rho^2}{2z}, \quad (4.3)$$

$$\text{если предположить, что } \delta \ll z. \quad (4.4)$$

Отсюда следует, что единственное различие между наблюдением на сфере и на плоскости состоит в наличии *фазового сдвига, квадратично зависящего от ρ* :

$$\varphi_2 = \frac{2\pi}{\lambda} \delta = \frac{2\pi}{\lambda} \frac{\rho^2}{2z}. \quad (4.5)$$

Но это не имеет никакого значения, если фотоприемник чувствителен к интенсивности света.

3. Приближение малых углов

Условие (4.4) на деле означает, что угол между векторами Oz и \mathbf{n} должен быть мал. Если это не так, то, согласно нашей теории, выражение (4.2) остается справедливым при условии, что α , β и γ — направляющие косинусы вектора \mathbf{n} .

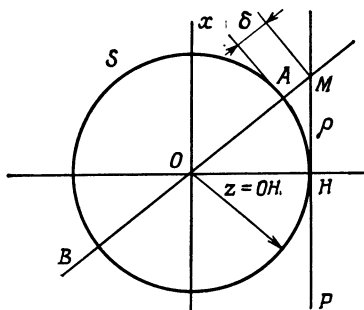


Рис. 4.6. Переход от сферы S бесконечного радиуса к плоскости P .

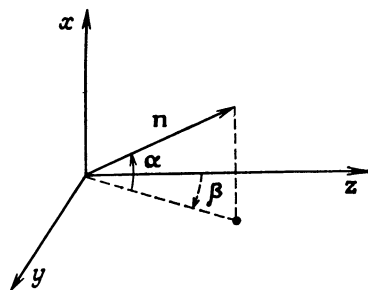


Рис. 4.7. Приближение малых углов.

\mathbf{n} — единичный вектор, т. е. косинусы углов α и β — направляющие.

Обозначим через $\bar{\alpha}$ и $\bar{\beta}$ углы, указанные на рис. 4.7. Без труда можно написать следующие соотношения для этих углов:

$$\begin{aligned} \alpha &= \sin \bar{\alpha}, \\ \beta &= \cos \bar{\alpha} \sin \bar{\beta}, \\ \gamma &= \cos \bar{\alpha} \cos \bar{\beta}. \end{aligned}$$

В приближении малых углов ($\bar{\alpha} \rightarrow 0$, $\bar{\beta} \rightarrow 0$) с точностью до членов разложения второго порядка они примут вид

$$\left. \begin{aligned} \alpha &\approx \bar{\alpha} \\ \beta &\approx \bar{\beta} \\ \gamma &\approx 1 \end{aligned} \right\} \text{ при } \bar{\alpha} \ll 1, \bar{\beta} \ll 1.$$

Поскольку это условие, как правило, выполняется, углы $\bar{\alpha}$ и $\bar{\beta}$ можно заменить их направляющими косинусами α и β .

4. Граница дальней зоны

Мы сказали: наблюдатель находится «очень далеко», в бесконечности. Проведем вновь расчет φ_1 , учитывая, что точка M находится не совсем в бесконечности (рис. 4.8).

Мы видим, что не был учтен малый фазовый сдвиг φ_2 , обусловленный разностью оптической длины пути между сферой Σ и плоскостью Π . При любом векторе сдвига T , лежащем в пределах окна, имеем

$$\varphi_2(T) \approx \frac{2\pi}{\lambda} \frac{|T|^2}{2z}. \quad (4.6)$$

Таким образом, слова «очень далеко» и «в бесконечности», которыми характеризовалось положение наблюдателя в точке M , в действительности

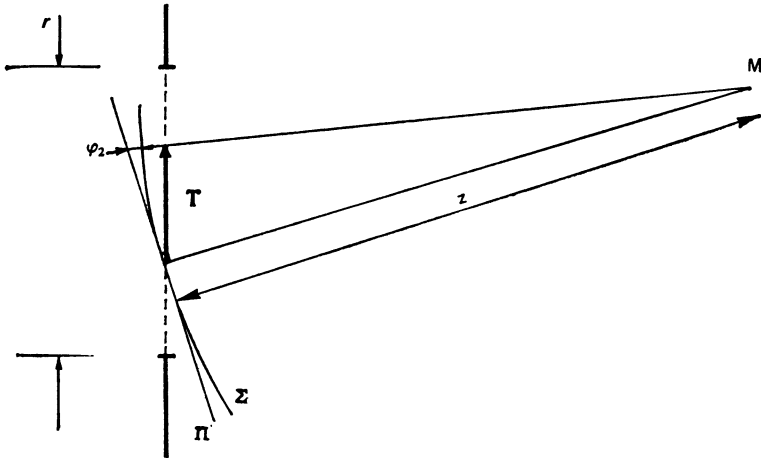


Рис. 4.8. Граница дальней зоны.

Расчет фазового множителя φ_2 , которым пренебрегают, когда точка наблюдения M «сильно удалена» ($z \gg z_0$).

означают следующее: любой квадратичный фазовый сдвиг $\varphi_2(T)$ не должен вносить заметного вклада в результат расчета по формуле (4.2) (каким бы ни был вектор T в пределах окна). Так, например, можно выбрать следующий критерий:

$$\varphi_2(T) \ll \pi \quad (4.7)$$

при любых T в пределах окна.

В результате получим следующее условие наблюдения на бесконечности: если r — порядок величины поперечных размеров окна, то, согласно формулам (4.6) и (4.7), необходимо помещать фотоприемник на расстоянии

$$z > z_0 \approx \frac{r^2}{\lambda}, \quad (4.8)$$

где z_0 — величина, называемая *границей дальней зоны*.

5. Соотношение между размерами окна и длиной волны

В нашей теории не было необходимости вводить условие $r \gg \lambda$, которое обязательно вводится при других подходах [15]. Из этого не следует, что наша теория лучше других. Мы лишь получаем те же самые результаты, не налагая ограничения $r \gg \lambda$. Но это возможно только в пределах наших исходных допущений, которые являются достаточно ограничивающими. От читателя требуется большая осторожность при использовании полученных результатов в более общем случае, например нескалярном, когда размеры окна сравнимы с длиной волны.

Г. Основные результаты

В данном параграфе мы ввели много фундаментальных оптических понятий. Мы использовали два основополагающих принципа.

1. *Принцип Гюйгенса — Френеля*: каждая точка окна рассматривается как точечный источник, излучающий сферическую волну, фаза которой равна фазе падающей волны.

2. *Принцип линейности* (линейной суперпозиции), который, по нашему предположению, справедлив и в общем случае бесконечной суммы.

Кроме того, мы сделали следующие допущения:

- а) для функции $f(x, y)$ существует фурье-образ;
- б) выполняется условие малых углов $\bar{\alpha} \ll 1$, $\bar{\beta} \ll 1$ (не необходимое);
- в) наблюдение ведется в дальней зоне $z \gg z_0 \approx r^2/\lambda$.

Получен следующий результат: *распределение комплексных амплитуд на сфере S есть двумерный фурье-образ функции пропускания окна*. Любое смещение окна в его плоскости приводит только к появлению линейного фазового сдвига.

Этот результат имеет фундаментальное значение для оптической обработки информации. Мы его получили на основе аксиоматического подхода. В нашу задачу не входило изложение общей теории дифракции, а потому мы отсылаем читателя к другим работам, где подробнее рассматривается математическое решение задачи [15, 16] и векторная теория дифракции [17]. Теперь разберем конкретнее явление дифракции.

Д. Расходимость и телесный угол пучка

Вернемся к двум примерам из § 1, п. А. Рассмотрим промежуточный случай окна $f(x, y)$, размеры которого не будут ни бесконечно малыми («точечное отверстие» и сферическая волна), ни бесконечно большими («широкое окно» и плоская волна).

Чтобы упростить запись, мы далее не будем упоминать о координате y .

Рассмотрим рис. 4.9. Пойдут ли «лучи» только прямо? Нет, так как, согласно формуле (4.2), дифракция на бесконечности

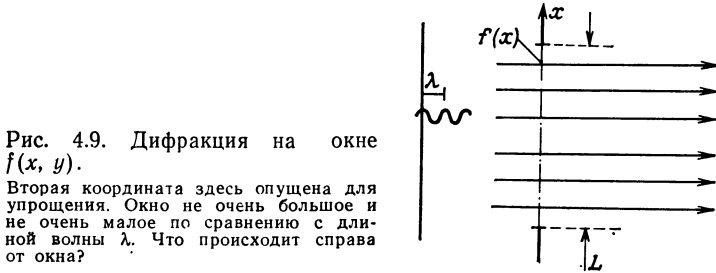


Рис. 4.9. Дифракция на окне $f(x, y)$.

Вторая координата здесь опущена для упрощения. Окно не очень большое и не очень малое по сравнению с длиной волны λ . Что происходит справа от окна?

определяется фурье-образом функции вида

$$A \operatorname{rect}\left(\frac{x}{L}\right) = \begin{cases} A & \text{при } |x| < L/2, \\ 0 & \text{при } |x| > L/2. \end{cases}$$

В этом случае

$$F(\alpha) = A \frac{\sin\left(\pi \frac{\alpha}{\lambda} L\right)}{\pi \frac{\alpha}{\lambda}}. \quad (4.9)$$

Введем обозначение sinc , которым мы часто будем пользоваться в дальнейшем:

$$F(\alpha) = AL \operatorname{sinc}\left(\frac{\alpha}{\lambda} L\right).$$

Таким образом, на расстоянии $z \gg z_0$ (рис. 4.10) мы получаем «дифракционное пятно», размеры которого можно характеризовать радиусом первого темного кольца:

$$\frac{\alpha}{\lambda} = \frac{1}{L}.$$

Если ввести пространственную координату u , соответствующую точке x в плоскости P , то получим

$$u_0 = \frac{\lambda z}{L}.$$

В то же время на бесконечности пучок расходится на угол

$$\alpha = \frac{\lambda}{L}. \quad (4.10)$$

Очевидно, что «лучи» не распространяются абсолютно прямо; это естественное явление: ограниченный параллельный световой

пучок не может долго оставаться параллельным. Он должен расходиться. После поглощения и дисперсии расходимость — третий враг в вопросах оптической связи.

Рассмотрим теперь более общий случай дифракции на двумерном окне. Пусть s — поверхность окна, а $\bar{\omega}$ — телесный угол,

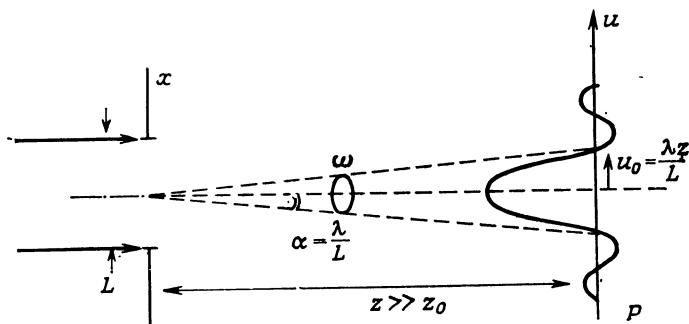


Рис. 4.10. Картина дифракции на окне шириной L , возникающая на бесконечности.

под которым видно дифракционное пятно. (Этот телесный угол должен быть определен в соответствии с некоторым критерием

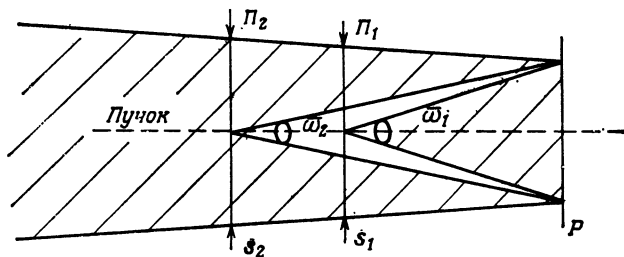


Рис. 4.11. Сохранение телесного угла пучка.

Рассматриваются три произвольных (параллельных) сечения пучка прямыми.

убывания амплитуды.) В таком случае выражение (4.10) преобразуется к виду

$$s\bar{\omega} = C\lambda^2 \quad (4.11)$$

(где C — константа, которая зависит от упомянутого выше критерия расходимости). Выражение (4.11) напоминает нам старый принцип сохранения телесного угла пучка (формула Клаузиуса [18]). Этот принцип, очевидный в приближении геометрической оптики, иллюстрируется на рис. 4.11, где рассматривается сечение пучка тремя произвольными параллельными плоско-

стями. Из выполненного на рисунке геометрического построения получаем

$$s_1 \bar{\omega}_1 = s_2 \bar{\omega}_2 \text{ для любых } \Pi_1 \text{ и } \Pi_2.$$

Эта постоянная величина является характеристикой рассматриваемого пучка. Она называется телесным углом пучка. Обозначим ее через Φ :

$$\Phi = s \bar{\omega} \text{ для любой плоскости } \Pi. \quad (4.12)$$

Сравнение выражений (4.11) и (4.12) показывает их полное сходство. Но выражение (4.12) получено из геометрического построения, а выражение (4.11) вытекает из физической оптики и относится к абсолютно когерентному световому пучку (иначе говоря, пучку, исходящему из плоской монохроматической волны). В § 4 мы уточним это понятие.

Учитывая сходство выражений (4.11) и (4.12), можно обобщить соотношение (4.11), написав

$$\Phi = C \lambda^2. \quad (4.13)$$

Поскольку любой отрезок когерентного светового пучка удовлетворяет соотношению (4.13), телесный угол пучка пропорционален квадрату длины волны. Коэффициент пропорциональности зависит от структуры пучка (и от принятых критериев определения s и $\bar{\omega}$).

Из рассмотренного принципа расходимости пучков можно сделать следующие выводы: 1) чем больше мы хотим сфокусировать пучок, тем сильнее он будет расходиться после точки фокусировки; 2) если мы хотим получить минимально расходящийся пучок, то необходимо формировать его возможно более широким.

Дальше нам встретится много примеров, иллюстрирующих этот принцип (например, расходимость луча полупроводниковых лазеров, гл. 12, § 4, п. Г, 2).

§ 2. Пространственная частота

Из выражения (4.2) следует, что в оптике легко осуществить комплексное двумерное преобразование Фурье. Для этого нужно удалиться в «бесконечность» или взять собирающую линзу с фокусным расстоянием f , которая перенесет изображение из бесконечно удаленной плоскости в свою фокальную плоскость (рис. 4.12). В действительности точное преобразование Фурье получается только тогда, когда транспарант $f(x, y)$ помещается в переднюю фокальную плоскость линзы [19]. В противном случае необходимо компенсировать фазовые сдвиги, обусловленные распространением (гл. 6, § 2).

Мы видим, что существует очень простая связь между каждой точкой фокальной плоскости (с координатами u, v) и направляющими косинусами α и β рассматриваемого направления распространения. В случае малых углов направляющие косинусы совпадают со своими углами α и β , определенными в § 1,

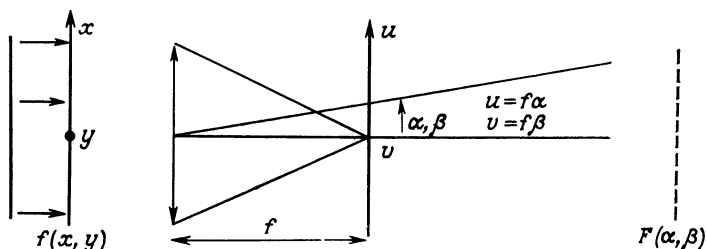


Рис. 4.12. Картина дифракции в фокальной плоскости линзы.

п. В, 3. В этом (и только в этом) случае распределение комплексной амплитуды в задней фокальной плоскости линзы дается выражением

$$G(u, v) = F\left(\frac{u}{f}, \frac{v}{f}\right).$$

На основании этого выражения в классической оптике определяют шкалу пространственных частот

$$\begin{aligned} f_x &= \frac{\alpha}{\lambda}, \\ f_y &= \frac{\beta}{\lambda}, \end{aligned} \quad (4.14)$$

или в случае малых углов

$$\begin{aligned} f_x &= \frac{u}{\lambda f}, \\ f_y &= \frac{v}{\lambda f}. \end{aligned} \quad (4.15)$$

В общем случае переход от (4.14) к (4.15) будет значительно более сложным.

Понятие пространственной частоты аналогично понятию временной частоты в электронике, но есть и важные различия.

1. Двумерный характер пространственных частот позволяет в дополнение к классическому частотному анализу выявлять зависимость от ориентации в плоскости $f(x, y)$.

2. Пространственно-частотный спектр функции $f(x, y)$ реален и физически доступен. Не нужно никаких спектроанализаторов! Возьмите линзу, осветите плоской волной транспарант с записью функции и наблюдайте распределение света в задней фокальной

плоскости линзы, которая теперь будет частотной плоскостью. Это легко и быстро. Притом можно выполнять над спектром определенные операции. Именно этим мы и займемся в следующих главах.

§ 3. Гауссовы пучки

Функция Гаусса — единственная функция, инвариантная по отношению к преобразованию Фурье. Поэтому случай окна с гауссовой функцией пропускания вида

$$f(x, y) = f_0 e^{-\frac{r^2}{a_0^2}}, \quad (4.16)$$

где $r^2 = x^2 + y^2$,

очень интересен с теоретической точки зрения. Но он интересен и с практической точки зрения, поскольку гауссовы пучки ближе к реальности, нежели плоская волна.

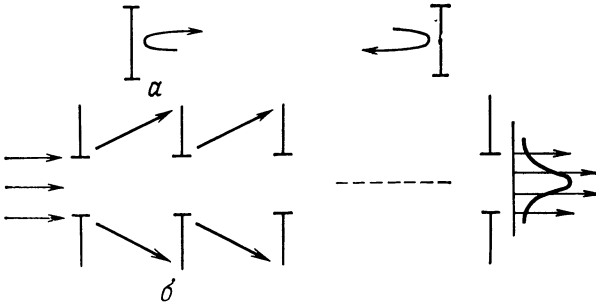


Рис. 4.13. Анализ работы лазерного резонатора по методу Фокса и Ли [20].

a — реальный случай; *b* — модель Фокса и Ли. При каждом отражении на зеркале происходит дифракция. Интерферометр Фабри — Перо можно заменить эквивалентной системой бесконечного числа диафрагм, на которых происходит дифракция. Если такую систему осветить плоской волной, то из-за большого числа диафрагм волна на выходе, «размытая» дифракцией, будет иметь распределение амплитуд, близкое к гауссову.

Для иллюстрации рассмотрим, следуя Фоксу и Ли [20], лазерный резонатор. Лазерный резонатор (гл. 12, § 1) образован двумя зеркалами, расположенными друг против друга (резонатор Фабри — Перо). В процессе последовательных отражений на этих зеркалах (конечной апертуры) световая волна многократно дифрагирует (рис. 4.13). В конце концов волна приобретает структуру, близкую гауссовой. Теперь мы имеем модель волны, на основе которой можно аппроксимировать многие естественные процессы.

Гауссова форма пучка дает еще одно преимущество. Но прежде чем объяснить его, вернемся к модели Фокса — Ли. На

каждой диафрагме происходит дифракция пучка, и большая часть дифрагированного света не проходит через следующую диафрагму. Таким образом, в результате многократной дифракции устраняется излишняя расходимость. Устраняются и влияния любых крупных неоднородностей (причин дифракции). В результате такой пространственной фильтрации формируется практически идеальный пучок, расходимость которого не может быть далее уменьшена (дифракционно-ограниченная расходимость). Иначе говоря, *гауссов пучок* — *наилучший с точки зрения расходимости*. Такому пучку соответствует наименьшая постоянная C в выражении для телесного угла пучка [формула (4.13)].

А. Распространение гауссова пучка

Возьмем окно с пропусканием (4.16), осветим его плоской волной и посмотрим, что будет происходить справа от него (рис. 4.14). Величина ω_0 есть поперечный размер пучка в плоскости $z = 0$ (на самом деле это минимальный диаметр пучка, или диаметр «минимального пятна»).

В бесконечности¹⁾ мы получим гауссово распределение света (фурье-образ гауссовой функции) в виде

$$F_0 e^{-\frac{\rho^2}{\omega^2}}, \quad (4.17)$$

где ρ — расстояние от оси в плоскости наблюдения — переменная, сопряженная с r (рис. 4.14).

Согласно формуле (4.10), угловая расходимость α будет пропорциональна λ/ω_0 (рис. 4.14). Следовательно, ширина гауссова пятна ω (определяется на уровне $1/e$ от значения поля на оси), полученная на (очень большом) расстоянии z , будет равна

$$\omega \approx \frac{\lambda}{\omega_0} z.$$

Нам известна теорема *гомотетии* для преобразования Фурье: когда функция «растягивается», ее фурье-образ «сжимается», и наоборот. Полученный нами результат является не чем иным, как выражением этой теоремы. Величина ω_0 есть мера пространственной протяженности «функции», а величина $\omega_0/\lambda z$ — соответствующая мера в плоскости фурье-образа [см. определение (4.15) пространственных частот, где следует заменить u на ω и f на z]. В итоге получаем

$$\omega_0 \frac{\omega}{\lambda z} = \text{const.}$$

¹⁾ Напомним, что «в бесконечности» для нас означает $z \gg z_0$, где $z_0 \approx \omega_0^2/\lambda$ [формула (4.8)].

В то же время бесконечно удаленный наблюдатель видит окно как некую точку. Следовательно, он воспринимает сферическую волну радиусом

$$R(z) = z.$$

Вот что мы получаем в бесконечности. Теперь посмотрим, как распространяется волна в пространстве от $z = 0$ до $z \gg z_0$.

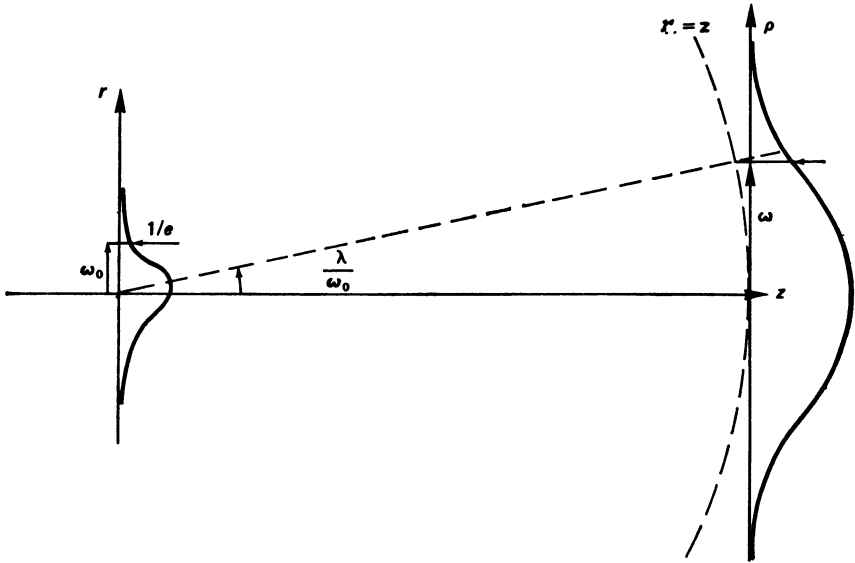


Рис. 4.14. Картина дифракции на окне с гауссовым законом изменения пропускания, возникающая на бесконечности.

Первыми на этот вопрос ответили Когельник и Ли [21]. Они показали, что гауссов пучок является решением скалярного уравнения распространения волны

$$\Delta u + k^2 u = 0$$

при выполнении следующих допущений, относящихся к распространению волны вдоль оси Oz . Прежде всего можно предположить, что фаза изменяется линейно вдоль оси z соответственно множителю $\exp(-ikz)$. Когельник и Ли [21] приняли, что нелинейные изменения фазы вдоль оси Oz сравнительно медленны. Мы не будем вдаваться в детали их расчета. Просто скажем, что это позволяет отбросить некоторые члены и получить в результате гауссов пучок как приближение второго порядка. В выражении для фазы имеются *квадратичные члены*, что может означать следующее: 1) наличие кривизны фронта волны, если показатель экспоненты — мнимая величина [формула

(4.5)]; 2) уменьшение амплитуды по гауссову закону, если показатель экспоненты — действительная величина [формула (4.16) или (4.17)].

Важно, что пучок остается гауссовым вдоль всей траектории распространения. Зависимость амплитуды волны от z и от расстояния до оси ρ описывается выражением

$$u(\rho, z) = \frac{\omega_0}{\omega} e^{-ikz} e^{i\varphi} e^{-\frac{\rho^2}{\omega^2}} e^{-ik\frac{\rho^2}{2R}}, \quad (4.18)$$

где $k = 2\pi/\lambda$.

Мы видим, что в сечении пучок является гауссовым и его ширина равна ω . Волна в пучке имеет сферический волновой

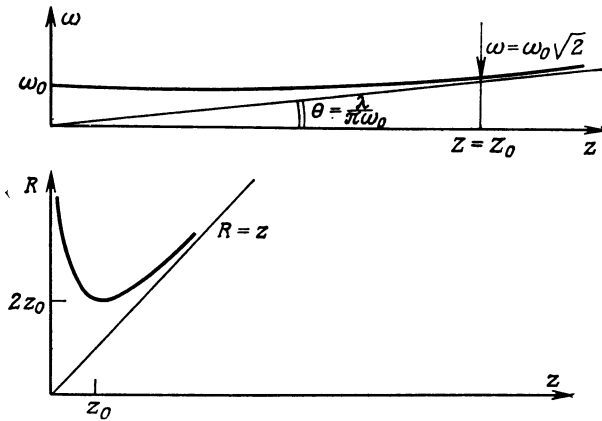


Рис. 4.15. Зависимость ширины ω и радиуса кривизны R гауссова пучка от расстояния z .

Изменив направление распространения пучка на обратное на верхнем рисунке, мы получим рис. 4.17.

фронт с радиусом кривизны R . Величины R и ω связаны с z следующими квадратичными зависимостями:

$$\left(\frac{\omega}{\omega_0}\right)^2 = 1 + \left(\frac{z}{z_0}\right)^2, \quad (4.19)$$

$$R = z \left[1 + \left(\frac{z_0}{z}\right)^2 \right], \quad (4.20)$$

$$\text{где } z = \frac{\pi\omega_0^2}{\lambda}. \quad (4.21)$$

Эти зависимости представлены на рис. 4.15.

Выражение для расходимости гауссова луча оказывается таким:

$$\alpha = \frac{\lambda}{\pi\omega_0}, \quad (4.22)$$

и им определяется постоянная в выражении для телесного угла соответствующего пучка.

В заключение отметим, что зависимость фазового сдвига от z не совсем линейна. На самом деле имеется дополнительный член

$$\varphi = \arctg\left(\frac{z}{z_0}\right).$$

Мы не будем более углубляться в детали теории гауссовых пучков. Отметим, однако, что гауссов пучок — самый простой. Можно рассматривать целое семейство «эрмитово-гауссовых» или «лагерово-гауссовых» пучков, характеристики распространения которых полностью идентичны гауссову пучку, но пространственная структура модулирована модальными членами (гл. 8, § 4, п. Б, 2). Подобные типы пучков наблюдаются в лазерном резонаторе. Поскольку пространственные структуры, упомянутые выше, имеют характеристики, достаточно сходные с модами, возникающими в световодах, мы не будем вдаваться в дальнейшие подробности. Рекомендуем читателю оригинальную работу Когельника [21], а также работу Ярива [22], в которых приведен ясный и точный анализ данного вопроса.

Итак, мы теперь вооружены гауссовым пучком — моделью, более близкой к реальности, нежели плоская или сферическая волна. Но те, кто работает с некоторыми типами лазеров (твердотельными и полупроводниковыми, гл. 12), с полным основанием могут сказать, что в гауссовом пучке есть еще, к сожалению, кое-что от идеальной теоретической модели.

Б. Фокусировка гауссова пучка

Что произойдет, если направить гауссов пучок на линзу?

Но сначала спросим, что такое линза. Согласно классическим законам оптики, теоретически идеальная линза — это оптическая система, преобразующая одну сферическую волну в другую сферическую волну или (частный случай) плоскую волну в сферическую (рис. 4.16). Следовательно, идеальная линза представляет собой «устройство квадратичной задержки», такое, что в каждой точке плоскости $z = 0$ формируемая комплексная амплитуда приобретает фазовый множитель вида

$$e^{\pm i \frac{2\pi}{\lambda} \frac{\rho^2}{2f}}.$$

Предоставляем читателю самому выполнить два небольших расчета, такие же, как и при выводе формулы (4.3). Величина f называется фокусным расстоянием линзы. Знак зависит от типа рассматриваемой линзы: собирающая (+) или рассеивающая (—).

Рассматриваемая линза идеальна в том плане, что: 1) она предполагается бесконечно тонкой; 2) она имеет бесконечно большую апертуру ($\rho \rightarrow \infty$); 3) вносимый ею фазовый сдвиг

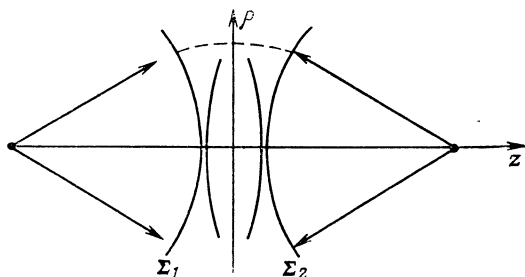


Рис. 4.16 Идеальная линза как квадратичная линза задержки.

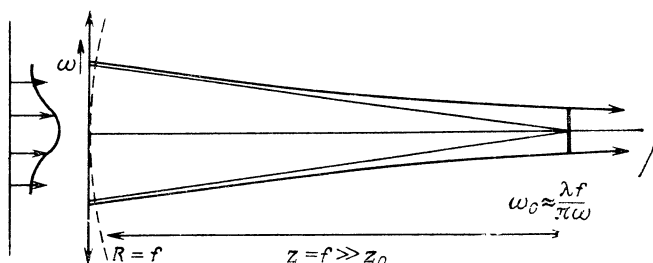


Рис. 4.17. Фокусировка гауссова пучка (см. рис. 4.15).

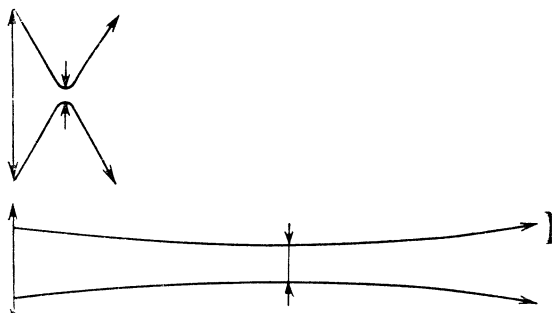


Рис. 4.18. Фокусировка гауссова пучка.

Вверху — короткофокусная линза: дифракционное пятно мало, но велика расходимость. Внизу — длиннофокусная линза: расходимость мала, но пучок плохо сфокусирован.

точно соответствует квадратичному закону, т. е. является параболическим (а не сферическим, что не одно и то же); 4) она однозначно характеризуется радиусом кривизны $\pm f$, который приобретает прошедшая через нее плоская волна.

Из этого очень легко получается схема фокусировки гауссова пучка. Направим плоскую волну с гауссовым распределением амплитуд на собирающую линзу (рис. 4.17). Достаточно представить себе, что на рис. 4.15 линза расположена справа, и изменить направление распространения волны на противоположное.

Радиус кривизны фронта волны, прошедшей сквозь линзу, будет равен $R = f$. Если предположить, что $f \gg z_0$ (а это условие практически всегда выполняется), то обращение выражения (4.19) с учетом (4.21) даст ширину дифракционного пятна в фокусе линзы:

$$\omega_0 \approx \frac{\lambda f}{\pi \omega}.$$

На рис. 4.18 схематически показано соотношение между фокусировкой и расходимостью.

В. Гауссов пучок и квадратичная среда

Предположим, что у нас есть гауссов пучок, изображенный на рис. 4.15. При $z < z_0 = \pi \omega_0^2 / \lambda$ он остается практически параллельным. Далее он расходится. Если мы хотим скорректировать

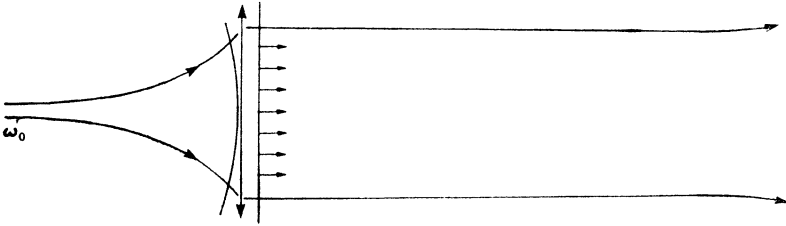


Рис. 4.19. Изменение расходимости пучка в зависимости от его ширины.

эту расходимость, то достаточно поместить на расстоянии $z \gg z_0$ линзу с фокусным расстоянием $f = R \approx z$. На выходе из этой линзы пучок станет параллельным, а поскольку ширина пучка стала больше, он дольше останется параллельным (рис. 4.19).

Чтобы не нужен был целый каскад линз, было предложено компенсировать естественную расходимость пучка непрерывно. С этой целью следует использовать среду распространения с изменяющимся показателем преломления $n(\rho, z)$. Разумеется, мы выбираем среду с квадратичным изменением показателя преломления вида

$$n(\rho, z) = n_0 - \frac{n_2}{2} \rho^2. \quad (4.23)$$

Направим на такую среду, начинающуюся в плоскости $z = 0$, гауссов пучок с рис. 4.15. В вакууме этот пучок вначале почти параллелен. Но потом он постепенно расходится. Действительно, в окрестности значения $z = 0$ можно написать

$$\omega = \omega_0 \sqrt{1 + \left(\frac{z}{z_0}\right)^2} \approx \omega_0 + \frac{\omega_0 z^2}{2z_0^2}.$$

Отсюда можно определить первоначальный радиус кривизны, который будет иметь гауссов пучок в вакууме:

$$R_g = \frac{z_0^2}{\omega_0} \quad (4.24)$$

(индекс g уточняет, что речь идет о естественной кривизне *гауссова* пучка).

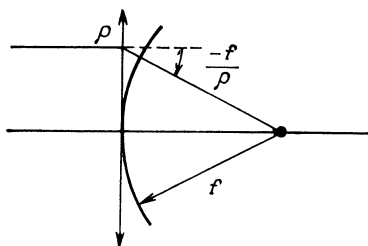


Рис. 4.20 Кривизна волнового фронта, создаваемая простой линзой.

Но у нас среда с неоднородным показателем преломления. Каждый слой dz такой среды действует почти как линза, которая будет компенсировать эту расходимость (4.24). Действительно, в силу формулы (2.22) кривизна фронта волны, вводимая этой средой, определяется соотношением

$$n\mathbf{C} = \text{grad } n = -n_2 \rho. \quad (4.25)$$

Приняв, что $n_2 \ll n_0$, получим радиус кривизны

$$R_n \approx -\frac{n_0}{n_2 \rho} \quad (4.26)$$

(индекс n уточняет, что речь идет о кривизне, обусловленной изменением показателя преломления). Величины R_n и R_g обратно пропорциональны смещению от оптической оси. В то же время R_n соответствует кривизне $-f/\rho$, приобретаемой световыми лучами после прохождения собирающей линзы с фокусным расстоянием f (рис. 4.20). Приравняв друг другу величины (4.24) и (4.26) при $\rho = \omega_0$, получаем

$$\frac{n_0}{n_2} = z_0^2. \quad (4.27)$$

Таким образом, если выполняется равенство (4.27), то естественная расходимость гауссова пучка будет полностью ком-

пенсироваться благодаря свойству квадратичной среды осуществлять самофокусировку. И если точно выполняются соотношения (4.23) и (4.27), то можно найти пучок с гауссовым распределением амплитуды по сечению, но с плоским фронтом (рис. 4.21), который будет распространяться не расходясь.

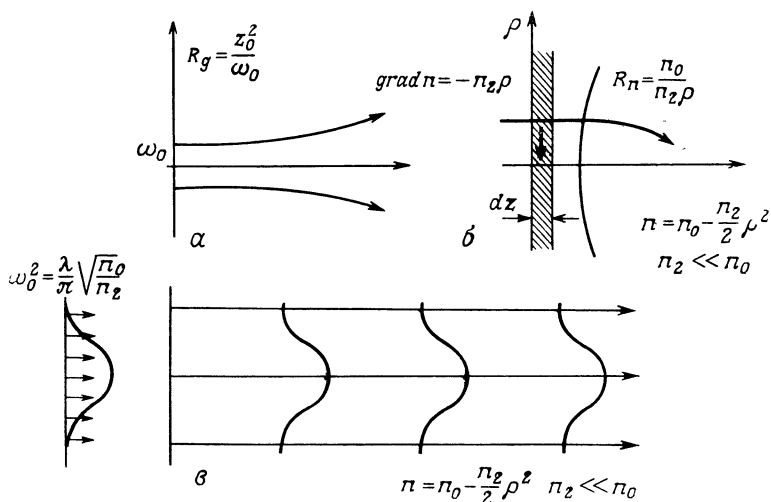


Рис. 4.21. Распространение гауссова пучка без расходимости в среде с идеально согласованным квадратичным изменением показателя преломления.

a — естественная кривизна гауссова пучка в вакууме, *б* — кривизна, создаваемая слоем dz «квадратичной» среды с $n(\rho) = n_0 - (n_2/2)\rho^2$; *в* — распространение пучка в согласованной среде. Распределение амплитуд в пучке остается гауссовым, но волна будет плоская.

В соответствии с (4.21) и (4.27) размеры такого пучка будут определяться соотношением

$$\omega_0^2 = \frac{\lambda}{\pi} \sqrt{\frac{n_0}{n_2}}.$$

В гл. 8 будут рассмотрены оптические волокна с градиентом (плавным изменением) показателя преломления, в которых используется этот принцип. Там мы начнем с более общего закона изменения показателя преломления, чем закон (4.23) (справедливость которого ограничена параксиальным приближением).

§ 4. Несколько слов о когерентности

Выше мы всюду неявно предполагали, что свет когерентный. Сначала мы рассмотрели *плоскую монохроматическую волну*. Это волна, *идеально когерентная* в том смысле, что *относительные изменения фазы постоянны во времени и в пространстве*.

Мы отметили, что такую плоскую волну можно рассматривать как волну, излучаемую точечным монохроматическим источником. Это идеальный источник. Затем мы рассмотрели гауссов пучок, минимальная ширина ω_0 которого есть конечная величина. Следовательно, гауссов пучок нельзя считать исходящим из точечного источника. Тем не менее такой пучок будет когерентным, так как он абсолютно монохроматичный, а относительные изменения фазы по фронту волны постоянны во времени.

Мы не собираемся подробно излагать здесь теорию когерентности и некогерентности. Отметим лишь отдельные физические моменты, которые пригодятся нам в дальнейшем, особенно когда речь пойдет об оптической обработке информации и характеристиках источников света.

Предположим, что у нас обычный источник света (протяженный, немонохроматичный) и мы хотим наблюдать интерференцию. Первое, что необходимо, это постараться сделать этот источник «когерентным».

А. Как получить когерентный источник?

Из сказанного ранее можно сделать вывод, что существуют две практически важные стороны когерентности: *протяженность источника (пространственная когерентность)* и его *монохроматичность (временная когерентность)*.

Чтобы, имея обычный источник света, сделать когерентный источник, достаточно, очевидно, поместить перед обычным источником точечную диафрагму, которая выделит очень малую излучающую площадку, и фильтр с очень высокой избирательностью, который выделит очень узкую спектральную полосу $\Delta\lambda$.

1. Временная когерентность, длина когерентности

Если бы источник был абсолютно монохроматическим, то он излучал бы чистую синусоиду в интервале времени от $t = -\infty$ до $t = +\infty$! Как уже отмечалось ранее (гл. 1, § 4), это невозможно: у любого излучения должно быть начало и должен быть конец. Обычно считается, что атомы излучают «волновые пакеты», или «цуги волн» (гл. 2, § 5, п. Б), которые не имеют **определенного фазового соотношения между собой** [23]. Предположим, что **длительность волнового пакета равна Δt** . Тогда **излучаемый спектр** (в частотах или длинах волн) не может быть **бесконечно узким**. Пусть $\Delta\lambda$ — ширина спектра излучения источника. **Каждый** знает, что бесполезно стараться получить **интерференционную картину** в интерферометре, в котором **разность** хода слишком велика: получим один пакет раньше другого.

Длину когерентности (и время когерентности) можно найти из соотношения (2.426):

$$l_c \approx \frac{\lambda^2}{\Delta\lambda}. \quad (4.28)$$

$$\tau_c = \frac{l_c}{c} \approx \frac{\lambda^2}{c\Delta\lambda}. \quad (4.29)$$

Длина когерентности — это практически важная характеристика излучения. Возьмем, например, очень хороший интерференционный фильтр, выделяющий полосу излучения $\Delta\lambda = 1$ нм. Тогда получим $l_c \approx 0,25$ мм, $\tau_c \approx 10^{-2}$ пс. Предположим, что таким фильтром мы фильтруем широкий спектр излучения. Волна будет «чистой» (синусоидальной) только в течение времени τ_c , которым и ограничивается среднее полезное время наблюдения. Получаемая при этом энергия излучения будет очень мала. Например, в случае лампы мощностью 1000 Вт она составит всего лишь

$$10^3 \cdot 10^{-14} = 10^{-11} \text{ Дж.}$$

Некоторые тепловые источники света испускают очень узкие спектральные линии, которые можно выделить при помощи интерференционных фильтров и получить таким образом излучение с длиной когерентности, достаточной для наблюдения интерференции (несколько миллиметров и даже сантиметры).

2. Пространственная когерентность

Рассмотрим теперь вопрос, связанный с размерами источника (рис. 4.22). Предположим, что у нас источник размером L , излучение которого проходит через интерференционный фильтр, т. е. источник, излучающий квазимонохроматический свет с длиной волны λ . Посмотрим, что происходит в фокусе линзы.

Если бы источник был эквивалентен диафрагме, освещаемой плоской волной, то в фокусе линзы мы наблюдали бы дифракционное пятно радиусом $\lambda f/L$ (§ 1, п. А.3). Но атомы источника испускают свет во всех направлениях и без определенных фазовых соотношений (рис. 4.22). Поэтому в фокальной плоскости линзы наблюдается рассеяние света в значительно более широкой области. Чтобы улучшить положение дел, поместим в эту плоскость диафрагму, диаметр которой d удовлетворяет условию

$$d \ll \frac{\lambda f}{L}. \quad (4.30)$$

(Например, при $L = 1$ см, $f = 1$ м и $\lambda = 0,5$ мкм получаем $d \ll 50$ мкм.)

Предположим, что источник излучает свет равномерно во всех направлениях (закон Ламберта). Дифрактама пропускает только очень малую долю энергии, равную

$$\left(\frac{d/f}{\pi}\right)^2 \ll \left(\frac{\lambda}{\pi L}\right)^2. \quad (4.31)$$

Грубо говоря, она выделяет из всех излучаемых источником волн одну плоскую волну. Но, согласно формуле (4.31), при $L = 1$ см коэффициент потерь составляет 10^{-10} !

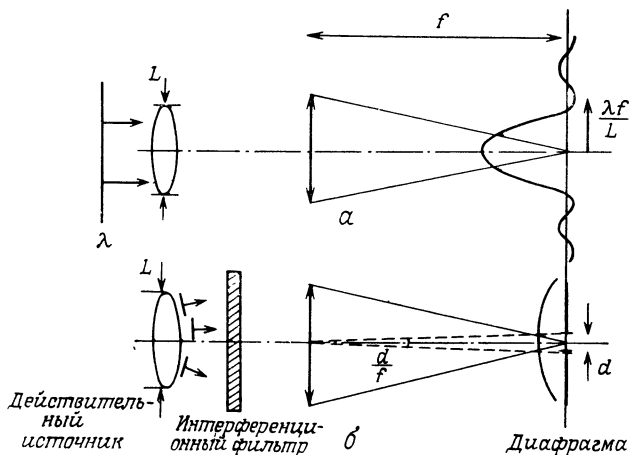


Рис. 4.22. Применение диафрагмы для улучшения пространственной когерентности.

a — картина дифракции плоской волны с ограниченной когерентностью; b — в фокальной плоскости линзы выделяется малая часть пространственного спектра, излучаемого реальным источником с фильтром

Если источник расположен очень далеко (например, Солнце), то нет необходимости в линзе. При угловом диаметре Солнца, равном ε (возьмем $\varepsilon = 10^{-2}$ рад), радиус дифракционного пятна по порядку величины равен

$$\frac{\lambda}{\varepsilon} \approx 50 \text{ мкм.}$$

Это означает, что можно (мысленно) повторить опыт Юнга с двумя освещенными Солнцем дырочками, но *расстояние между ними должно быть значительно меньше 50 мкм...* (напомним, что, кроме этого, требуется монохроматичность).

3. Энергетический баланс

Вернемся к нашей лампе мощностью 1000 Вт, излучение которой проходит через интерференционный фильтр и диафрагму (рис. 4.22, б). Энергия излучения на выходе такой системы со-

ставит всего лишь

$$10^{-11} \cdot 10^{-10} = 10^{-21} \text{ Дж}, \quad (4.32)$$

иначе говоря, менее одного фотона!

На основании такого расчета можно сделать следующий вывод: «теоретически» можно повысить когерентность какого-либо источника света при помощи классических устройств. Но энергетический баланс сводит все на нет.

Б. Как добиться интерференции?

Мы уже видели (п. А,1), что удовлетворительной временной когерентности можно добиться, выделив достаточно узкие спектральные линии. Самый простой способ получения интерференционной картины — закрыть источник экраном с двумя отверстиями.

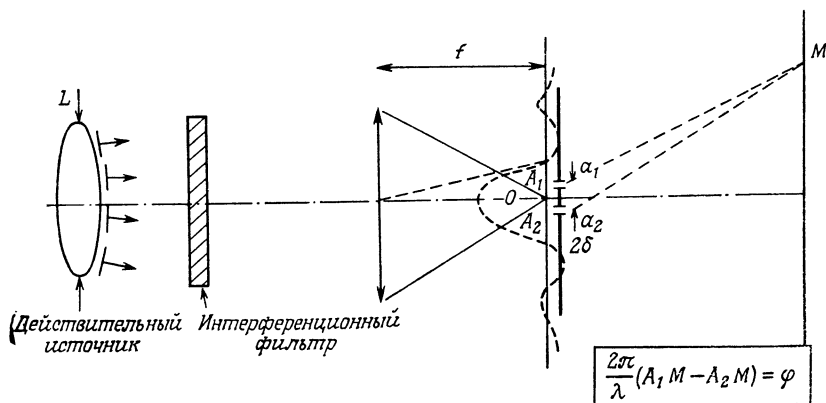


Рис. 4.23. Опыт Юнга.

Это и есть *опыт Юнга*. Однако с учетом сказанного в п. А,2 очень важное значение имеет расстояние между отверстиями. Рассмотрим схему, представленную на рис. 4.23.

Здесь возможны два крайних случая. В первом случае расстояние 2δ между отверстиями очень мало по сравнению с $\lambda f/L$. Можно принять, что они излучают синфазно. Все происходит так, как если бы у нас был точечный монохроматический источник. Интенсивность в точке M найдем путем сложения комплексных амплитуд, так как необходимо учитывать фазовые соотношения. Это случай *когерентной оптики*, и интенсивность света в точке M , полученная в результате суперпозиции двух оптических лучей с разностью фаз φ , определяется клас-

сической формулой

$$I = |a_1 + a_2 e^{i\varphi}|^2 = |a_1|^2 + |a_2|^2 + 2a_1 a_2 \cos \varphi. \quad (4.33)$$

Это и есть формула интерференции [23].

В другом крайнем случае, когда два отверстия разнесены очень далеко, нет определенного соотношения между фазами волн. Это случай *пространственной некогерентности*. Два пучка теперь не могут интерферировать из-за флуктуаций фазы между точками A_1 и A_2 . Поэтому в выражении (4.33) необходимо рас-

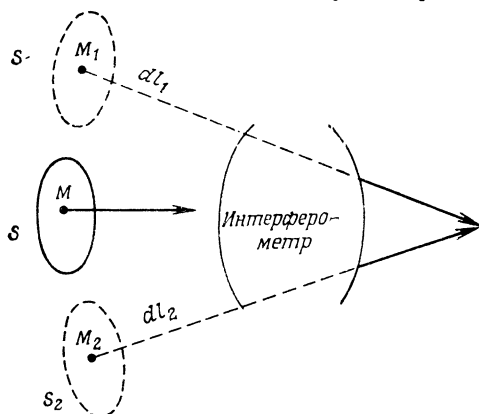


Рис. 4.24. Интерференция в случае протяженного монохроматического источника света

Интерферометр — это оптическая система, формирующая два изображения S_1 и S_2 источника S . Интерferируют попарно волны, излучаемые соответственными точками M_1 и M_2 в изображениях источника

считать усредненные во времени величины. В результате, считая излучения A_1 и A_2 некоррелированными, получаем

$$I = \langle |a_1|^2 \rangle + \langle |a_2|^2 \rangle = I_1 + I_2. \quad (4.34)$$

Общий вывод: в случае монохроматического протяженного источника нужно складывать интенсивности света. Только для интенсивности справедлив принцип линейной суперпозиции.

Отметим, что в случае очень близко расположенных отверстий (рис. 4.23) очень сильно уменьшается освещенность. Чтобы устранить такую трудность, применим другую схему, позволяющую осуществлять интерференцию при помощи протяженного источника. Для этого источник оптически «раздваивают» и наблюдают интерференцию между двумя соответственными точками полученных таким образом изображений (рис. 4.24). В этом случае формула интерференции для каждого точечного источника будет иметь вид

$$dI = dI_1 + dI_2 + 2\sqrt{dI_1 dI_2} \cos \varphi, \quad (4.35)$$

где dI_1 и dI_2 — элементарные интенсивности излучения двух соответственных точечных источников.

Таким методом мы воспользуемся в гл. 5, § 5, п. Е.

В двух следующих пунктах параграфа мы вернемся к теоретическим основаниям изложенного выше.

В. Теорема Цернике — Ван Ситтерта, пространственная когерентность и расходимость

Как мы видим, во всем, что касается пространственной когерентности, важную роль играет теория дифракции. В то же время переход от (4.33) к (4.34) получается в результате статистического временного усреднения, обозначаемого угловыми скобками $\langle \rangle$, излучения в точках A_1 и A_2 .

Действительно, колебательное состояние двух точек A_1 и A_2 , освещенных протяженным источником, можно характеризовать так называемой *степенью частичной когерентности*, которая определяется следующим образом:

$$\gamma(A_1 A_2) = \frac{\langle a_1(t) a_2^*(t) \rangle}{\sqrt{I_1 I_2}}. \quad (4.36)$$

В более общем случае с учетом (4.33) и (4.36) формула интерференции запишется в таком виде:

$$I = I_1 + I_2 + 2 \sqrt{I_1 I_2} \operatorname{Re}(\gamma e^{i\varphi}). \quad (4.37)$$

Как следует из формулы (4.37), коэффициентом γ (в общем случае комплексным) определяется контраст интерференционных полос и их смещение относительно начала координат. Выражение (4.34) было выведено в предположении об отсутствии корреляции между излучением в точках A_1 и A_2 .

Как коэффициент $\gamma(A_1, A_2)$ зависит от источника? На этот вопрос дает ответ теорема Цернике — Ван Ситтерта. Мы здесь не будем ее доказывать [23], а только сформулируем:

Степень частичной когерентности в двух точках A_1 и A_2 , освещенных протяженным монохроматическим источником, равна фурье-образу яркости источника.

Например, на рис. 4.22 степень частичной когерентности равна функции sinc (если яркость постоянна по всей площади источника):

$$\gamma(OA_1) = \operatorname{sinc}\left(\frac{\delta L}{\lambda f}\right),$$

где δ — расстояние от точки A_1 до точки на оси O .

В более общем случае частичная когерентность вычисляется по формулам теории дифракции [4].

При данной излучающей поверхности когерентность источника зависит от распределения яркости. От этого же зависит и его расходимость. На практике расходимость источника тесно связана с его когерентностью. Постоянная в выражении для телесного угла пучка S [формула (4.13)] тоже характеризует качество рассматриваемого источника. Так, у полупроводнико-

вых лазеров, излучающая поверхность которых всегда мала, по крайней мере в одном измерении, очень велика расходимость, в чем их большой недостаток (гл. 12, § 4).

Г. Теорема Винера — Хинчина, временная когерентность и спектр излучения

Выше мы установили, что длина когерентности связана со спектральной полосой излучения [формула (4.28)].

На самом деле длина когерентности — это лишь грубая характеристика временной когерентности. Рассмотрим точечный некогерентный источник. Он излучает временной сигнал вида

$$g(t) = \operatorname{Re} \{a(t) e^{i\omega_0 t}\}, \quad (4.38)$$

причем предполагается, что комплексная амплитуда $a(t)$ изменяется во времени медленно по сравнению с ω_0 .

Теорема Винера — Хинчина [23] состоит в следующем: спектр излучения источника

$$S(\omega) = \langle |G(\omega)|^2 \rangle, \\ \text{где } G(\omega) = \mathcal{F}(g(t)),$$

и автокорреляционная функция сигнала

$$\tilde{I}(\tau) = \langle g(t) g^*(t - \tau) \rangle$$

связаны между собой преобразованием Фурье.

Это общая теорема, относящаяся к случайным функциям, стационарным во втором порядке [24]. Но нас интересует возможность экспериментального приложения этой теоремы.

1. Измерение автокорреляции

Посмотрим сначала, что понимается в оптике под «автокорреляцией сигнала». С этой целью заставим интерферировать два пучка, идущие с разностью хода $l = c\tau$ (например, в интерферометре Майкельсона, рис. 4.25). В этом случае легко показать, что полная интенсивность будет меть вид

$$I(l = c\tau) = \langle |g(t) + g(t - \tau)|^2 \rangle, \\ I(l = c\tau) = 2I_0 + 2\operatorname{Re} \langle g(t) g^*(t - \tau) \rangle, \quad (4.39) \\ \text{где } I_0 = \langle |g(t)|^2 \rangle.$$

Введя степень частичной когерентности

$$\gamma(\tau) = \frac{\langle a(t) a^*(t - \tau) \rangle}{I_0},$$

получим выражение

$$I(\tau) = 2I_0[1 + \operatorname{Re}(\gamma(\tau)e^{i\omega_0\tau})],$$

которое полезно сравнить с формулой (4.37).

О полученных результатах можно сказать следующее.

1. Согласно формуле (4.39), интерферометр Майкельсона позволяет анализировать функцию автокорреляции сигнала.

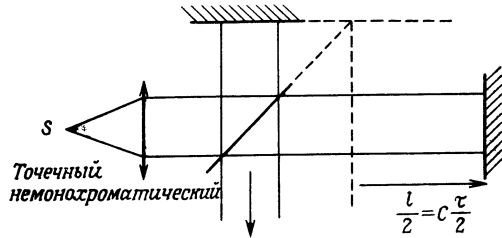
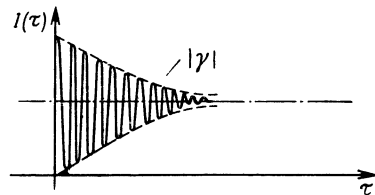


Рис. 4.25. Исследование автокорреляции при помощи интерферометра Майкельсона.

Осветителем служит точечный некогерентный источник света. Измеряется интенсивность света в зависимости от смещения зеркал.



2. Огибающая полученной кривой, которая в общем случае очень быстро флуктуирует из-за члена с ω_0 , есть степень частичной когерентности $\gamma(\tau)$. Длина когерентности — не что иное, как приближительная мера скорости убывания этой функции.

3. Это позволяет найти спектр излучения $S(\omega)$ как фурье-образ (метод «спектрального преобразования» [25]).

2. Исследование спектра

Что касается спектра, то его, как правило, исследуют классическими методами спектроскопии, т. е. пропуская через оптическую дисперсионную систему (призму, дифракционную решетку, интерферометр Фабри — Перо).

Предварительно отметим, что у лазеров (гл. 12) спектр состоит из равноудаленных друг от друга узких линий. Притом можно наблюдать некоторую периодичность функции $\gamma(\tau)$. Наконец, в рассматриваемом случае очень важны другие физические факторы, такие, как ширина и стабильность спектральных линий, которые существенны при получении большой длины когерентности.

Читатель встретится с практическим применением изложенного в гл. 12.

Д. Временной отклик оптического зрачка [26]

Теперь мы подходим к новой идее, которая представляется нам необычайно плодотворной в области «оптики и связи».

Рассмотрим одномерный зрачок (рис. 4.26), освещенный волной, исходящей из точечного *полихроматического* источника. Здесь нас будут интересовать временные изменения световых сигналов. Пусть $f(t)$ — сигнал, излучаемый источником, а $g(t)$ — сигнал, регистрируемый при наблюдении на бесконечности в

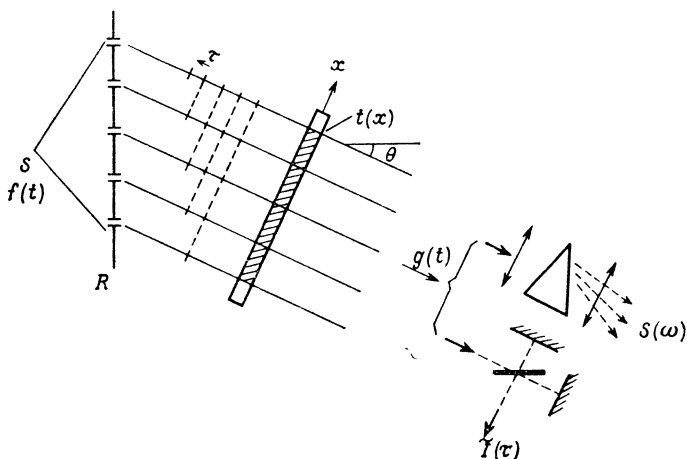


Рис. 4.26 Временной отклик на одномерной дифракционной решетке [26].

Точечный немонохроматический источник S наблюдается в бесконечности в направлении θ . Если излучается очень короткий импульс, то решетка сформирует серию очень коротких и равномерно распределенных в пространстве импульсов ($\tau < 1$ пс), аналогичным образом можно осуществлять временную модуляцию излучения одномерной пространственной функцией $f(x)$. В этом случае анализ принимаемого сигнала может осуществляться либо при помощи интерферометра Майкельсона, который сформирует функцию автокорреляции сигнала (рис. 4.25), либо при помощи классического спектрометра,

направлении θ . Очень легко убедиться, что функции $g(t)$ и $f(t)$ связаны между собой линейным и однородным преобразованием

$$af(t) \rightarrow ag(t) \quad \text{при} \quad a = \text{const},$$

$$f(t - \tau) \rightarrow g(t - \tau) \quad \text{при} \quad \tau \text{ действительном.}$$

Следовательно, функцию g можно найти путем линейной фильтрации функции f , т. е. выполнив свертку

$$g(t) = f(t) * h(t) \quad (4.40)$$

с некой функцией $h(t)$, называемой *временным «импульсным откликом»* и являющейся характеристической функцией зрачка. Мы найдем эту функцию как отклик оптической системы на функцию Дирака $f(t) = \delta(t)$. Выполним преобразование Фурье обеих частей равенства (4.40). При этом возникнет *временная передаточная функция* зрачка $H(\omega)$:

$$G(\omega) = F(\omega) H(\omega). \quad (4.41)$$

Мы не приводим здесь точные формулы, которыми $h(t)$ и $H(\omega)$ выражаются через функцию пропускания зрачка [26]. Скажем только, что длительность временного импульсного отклика очень мала — порядка пикосекунды!

На рис. 4.26 показано, как можно таким способом осуществить временную модуляцию полихроматического пучка и заставить его переносить пространственную информацию $t(x)$. Преимущество такого метода состоит в том, что он обещает огромную информационную пропускную способность. Но у нас нет столь быстродействующих источников и приемников. Как же тогда создать систему передачи, основанную на таком принципе?

Его авторы Фрели, Лакур и Вьено предлагают применять два типа регистрации, которые мы рассмотрели в п. Г, 1:

1) регистрировать функцию автокорреляции передаваемого сигнала C_g ,

$$C_g = C_f \otimes C_h, \quad (4.42)$$

при помощи интерферометра Майкельсона;

2) регистрировать энергетический спектр передаваемого сигнала

$$|G(\omega)|^2 = |F(\omega)|^2 |H(\omega)|^2 \quad (4.43)$$

при помощи классического спектрометра.

Очевидно, что при таких способах регистрации излучения можно установить оптическую связь с широкой полосой без применения электрооптических устройств. [Если, конечно, приемлема менее полная информация, соответствующая формулам (4.42) или (4.43), а не (4.40) и (4.41).] Чтобы устранить последний недостаток, названные авторы предлагают также метод *временной голографии* (гл. 18, § 3).

Изложенный метод пока что применяется в задачах метрологии [26, 27]. В области связи его возможности еще не реализованы.

Е. Отрицательные стороны когерентности

Из отрицательных сторон когерентности мы еще не касались проблемы *спеклов* (зернистости).

На рис. 4.27 представлен снимок дифракционного пятна. Это и есть спеклы.

Данное явление, крайне сложное, тесно связано с когерентностью света. Но оно также зависит от свойств поверхности освещенного объекта и равным образом от системы наблюдения. Спеклы создают высокий уровень шума при формировании изображений в лазерном свете (и могут служить помехой при передаче аналоговых сигналов по оптическим световодам [28, 29]).

Вот почему в настоящее время наметилась тенденция проводить оптическую обработку информации в некогерентном свете.

Таким образом, чаще всего спеклы рассматриваются как нежелательное явление. Но в некоторых случаях удалось использовать данное явление для обработки изображений [30]. Здесь мы не будем на этом останавливаться.

Среди других отрицательных сторон когерентности нужно также упомянуть общую сложность соответствующей теории. В данной книге мы не будем касаться теоретических вопросов, связанных с когерентностью, отсылая читателя к работам [4, 23, 30, 31], ставшим уже классическими.

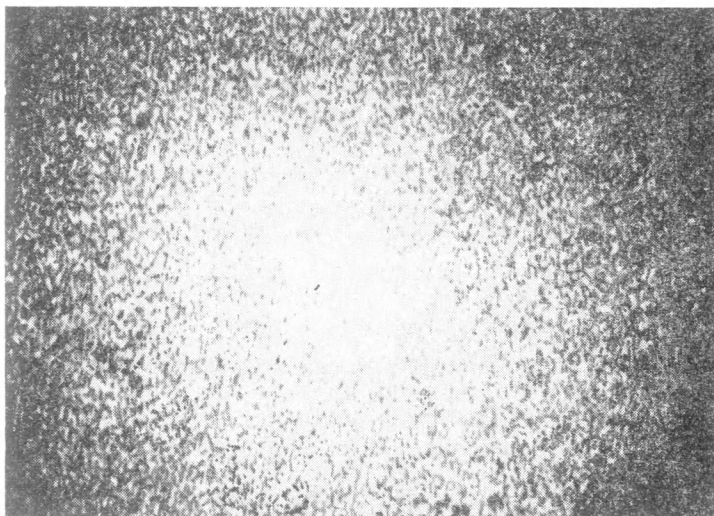


Рис. 4.27. Дифракционное пятно.

В заключение скажем, что открытие лазера представило собой гигантский шаг в этой области. Хотя лазеры еще далеки от генерации абсолютно когерентных волн (плоской волны по-прежнему не существует), в гл. 12 мы увидим, что газовые лазеры (гелий-неоновые) обеспечивают очень высокую когерентность. Такой лазер весьма распространенного промышленного типа дает излучение мощностью 10 мВт в пучке сечением 1 мм² при длине когерентности до 15 см. Для импульсных твердотельных лазеров характерна менее высокая когерентность, но они могут давать большую мощность (100 МВт и более) с излучающей поверхности площадью в несколько квадратных сантиметров за очень короткое время (~ 30 нс).

С открытием лазера когерентный свет стал в 10—20 раз более интенсивным (и более доступным).

Это колоссально! Но что можно делать с когерентным светом?

ГЛАВА 5

ГОЛОГРАФИЯ

В предыдущих главах мы видели, как теория электромагнетизма связана, с одной стороны, с геометрической оптикой, а с другой — с дифракцией. Прежде чем переходить к оптическим методам обработки информации, которыми мы займемся в следующих главах, мы познакомим читателя с принципами одного из таких методов, которым будем пользоваться в дальнейшем, а именно с принципами голографии. В этой новой области оптики, в которой по сути дела используются хорошо известные классические явления дифракции и интерференции, вместе с тем применяются оригинальные методы и вырабатывается особый способ мышления, чем и оправдано ее подробное рассмотрение.

В оптике наших дней голография снова возникает как некое неожиданное явление. Пять — десять лет она привлекала большое внимание исследователей ввиду того, что ее применения с первых же дней представлялись универсальными и удивительно многообразными. Очень благоприятная в то время для развития физических исследований международная обстановка тоже способствовала ее блестящей «популярности». Тем не менее пришлось все же признать, что развитие новой области оптики не дало столь быстро, как предвещалось, тех чудес, которых от нее ожидали. Суровая действительность технических ограничений сократила круг великолепных перспектив, а вместе с тем исчез энтузиазм и исчезли кредиты. Голография заняла свое место в ряду других областей оптики, богатая, как и они, плодотворными идеями и оригинальными приложениями, но в то же время и нерешенными проблемами.

Эту область оптики без ее прежней обманчивой привлекательности мы и будем изучать в данной главе.

Со времени Френеля (1816—1832 гг.) у оптиков были все теоретические сведения, необходимые для открытия голографии. Однако только в 1948 г. Габор [32] предложил идею голографии и получил первую голограмму¹⁾. Но возможности применения его открытия остаются весьма ограниченными до появления лазеров (Мейман, 1960) — источников света с большой

¹⁾ За работы по голографии Д. Габор (1900—1979 гг.) удостоен Нобелевской премии в 1971 г. — *Прим. ред.*

длиной когерентности¹⁾. Лейт и Упатниенс [33] (1962) предлагают затем наиболее распространенную сегодня схему записи голограмм, которая позволяет разделить объектный и опорный пучки²⁾. Одновременно с Вандер Люгтом [34] и Строуком [35] они исследуют основные области применения этого нового метода и выявляют наиболее перспективные из них [36].

Во Франции вслед за пионерскими работами Дюфье и затем Марешаля вклад в быстрое развитие голографии вносят различные лаборатории под руководством Франсона [37], Ловенталя [19] и Вьено [38, 39].

§ 1. „Голографическая“ запись

Как было показано в гл. 4, принцип Гюйгенса, видоизмененный Френелем, позволяет определить поле волны, дифрагировавшей на отверстии. Этот принцип мы рассмотрим здесь еще раз, чтобы объяснить основы голографии.

Известно, что принцип Гюйгенса — Френеля не дает точного решения задачи дифракции в общем случае. Но в случаях, когда размеры отверстия намного больше длины волны, он приводит к результатам, хорошо согласующимся с экспериментом.

Итак, рассмотрим освещенный объект (рис. 5.1). В каждой точке окна F амплитуда $A(x, y)$ и фаза $\varphi(x, y)$ отраженной объектом и дифрагировавшей на нем волны O_1 являются характеристиками объекта, так что, например, у наблюдателя сформируется на сетчатке глаза изображение, соответствующее объекту.

Теперь уберем объект и его освещение и возьмем волну O_2 , например плоскую. Предположим, что мы можем задать для этой волны O_2 в каждой точке окна F те же амплитуду $A(x, y)$ и фазу $\varphi(x, y)$, что и у волны O_1 .

Для наблюдателя, смотрящего через окно F , эти два случая абсолютно эквивалентны, поскольку в соответствии с принципом Гюйгенса — Френеля в обоих случаях будет одно и то же распределение вторичных источников, размещенных на окне F , с амплитудой $A(x, y)$ и фазой $\varphi(x, y)$ ³⁾. Во втором случае наблюдатель тоже видит объект, хотя его в действительности нет, и к тому же он видит его как реальный объект, т. е. с его

¹⁾ За создание первого квантового генератора — лазера — советским ученым Н. Г. Басову и А. М. Прохорову (совместно с американским ученым Ч. Таунсом) была присуждена Нобелевская премия по физике за 1964 г. Т. Мейман — создатель первого лазера на рубине. — *Прим. ред.*

²⁾ Большой вклад в развитие голографии внес советский ученый Ю. Н. Денисюк, предложивший схему записи голограмм во встречных пучках. За работы по голографии он удостоен Ленинской премии (1970 г.) и премии им. Д. Габора (1983 г.). — *Прим. ред.*

³⁾ Данное утверждение предполагает единственность решения обратной задачи дифракции; мы примем эту гипотезу.

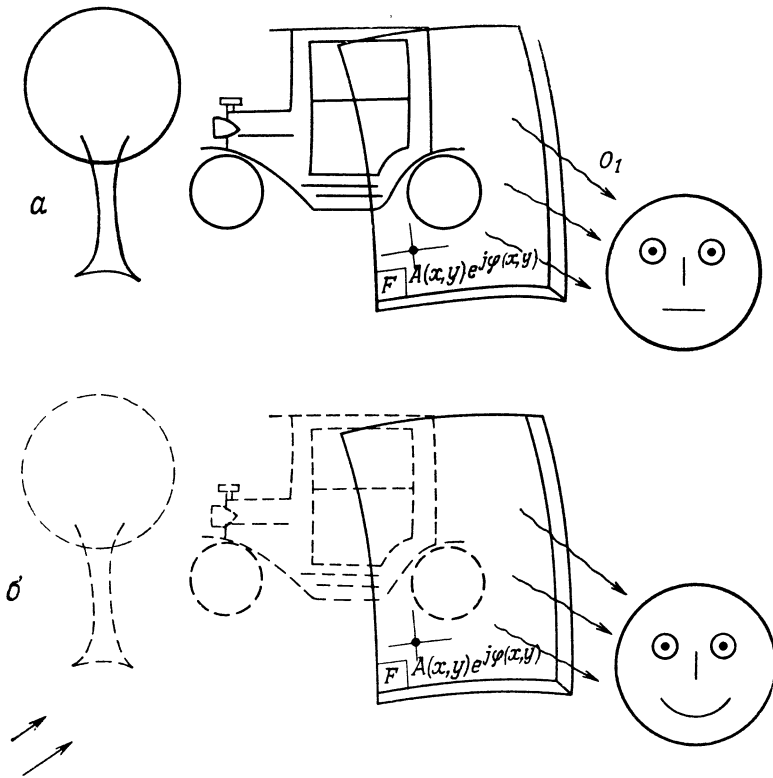


Рис. 5.1. Получение голограммы на основе принципа Гюйгенса — Френеля. *а* — объект (автомобиль + дерево) в результате отражения падающего на него света формирует в каждой точке (x, y) окна F волну O_1 , которая и позволяет наблюдателю «видеть» объект. *б* — в отсутствие объекта мы изменяем плоскую волну O_2 таким образом, чтобы в каждой точке окна F ее комплексная амплитуда была такой же, как и у волны O_1 ; в этом случае наблюдатель увидит объект таким, как если бы он действительно находился позади окна

объемностью (объемность создается благодаря способности наблюдателя воспринимать разницу в удаленности от него различных частей объекта) и его параллаксом (т. е. перемещением и деформацией объекта при перемещении наблюдателя) (рис. 5.1, *б*). Таким образом, если во втором случае мы сможем добиться того, чтобы волна O_2 приобрела фазу и амплитуду волны O_1 , то мы сможем реализовать «голографическую» запись (полную запись объекта).

Каким же способом придать волне O_2 комплексную амплитуду волны O_1 ? На первый взгляд, для этого подошли бы любые фоточувствительные материалы, преобразующие изменения комплексной амплитуды падающей волны в изменения комплексного пропускания за счет ослабления или фазового сдвига

(относительно таких материалов см. приложение 5). Но, к сожалению, большинство таких материалов являются квадратичными «детекторами», т. е. энергетическими детекторами, чувствительными к интенсивности I волны O_1 ,

$$I = |A \exp(i\varphi)|^2 = |A|^2.$$

При обычном фотографическом методе регистрации информация о фазе $\varphi(x, y)$ волны полностью теряется. Ниже мы увидим, каким образом Габор предложил устранить этот недостаток квадратичного детектирования.

§ 2. „Уравнение голографистов“

Идея голографической записи состоит в том, чтобы, кроме волны, которая нас интересует (называемой объектной волной), регистрировать дополнительно опорную волну, когерентную с объектной волной, таким образом, чтобы в результате их интерференции осуществлялось кодирование фазы φ . Введем обозначения

$O = o(M) \exp\{i\varphi(M)\}$ — объектная волна,

$R = r(M) \exp\{i\psi(M)\}$ — опорная волна.

Мы возьмем для записи среду, линейную по интенсивности (это может быть, например, идеальная фотоэмульсия, см. приложение 5). Она регистрирует распределение интенсивности, описываемое выражением

$$I = |O + R|^2 = |O|^2 + |R|^2 + OR^* + O^*R,$$

и ее пропускание после обычной фотохимической обработки (проявление, закрепление) будет пропорционально интенсивности I .

Осветим теперь эту пластинку опорной волной. Это будет волна O_2 , о которой говорилось в предыдущем параграфе. Волна, прошедшая сквозь пластинку, имеет вид

$$RI = (|O|^2 + |R|^2)R + O|R|^2 + O^*R^2. \quad (5.1)$$

Это выражение — основа голографии, и мы обязаны им Габору. В нем три слагаемых:

1. $(|O|^2 + |R|^2)R$ — опорная волна R , амплитуда которой модулирована коэффициентом $(|O|^2 + |R|^2)$.

2. $|R|^2O = r^2 o e^{i\varphi}$ — объектная волна O со своей фазой и со своей амплитудой. Она модулирована коэффициентом r^2 . Если в качестве опорной выбрать плоскую волну, то $r(x, y) = r_0$ и модуляция исчезает: объектная волна просто равномерно ослабляется в соответствии с коэффициентом r_0^2 .

3. $O^*R^2 = r^2 o e^{i(2\psi - \varphi)}$ — волна, которую называют комплексно-сопряженной с объектной волной. Это — паразитное слагаемое

(так же как и первое), которое появляется из-за квадратичного детектирования. Оно несет информацию, очень близкую к информации об объекте, но существенно отличается от объектной волны обратной фазой.

Итак, благодаря Габору мы решили половину нашей задачи: нам удалось определить фазу и амплитуду объектной волны, и мы способны их снова воссоздать на этапе, называемом «восстановлением» голограммы. Однако за этот успех пришлось заплатить введением двух паразитных слагаемых: опорной волны, которую мы ослабили, но не подавили совсем, и сопряженной волны. Как сделать, чтобы эти ненужные нам волны не испортили объектную волну O ? На этот вопрос мы ответим позднее, когда рассмотрим аналогию между голограммами и дифракционными решетками.

§ 3. Аналогия с дифракционными решетками

Рассмотрим самую простую голограмму, а именно голограмму точечного источника S_1 . В качестве опорного источника возьмем второй точечный источник S_2 , расположенный симметрично источнику S_1 относительно оптической оси. Как это вид-

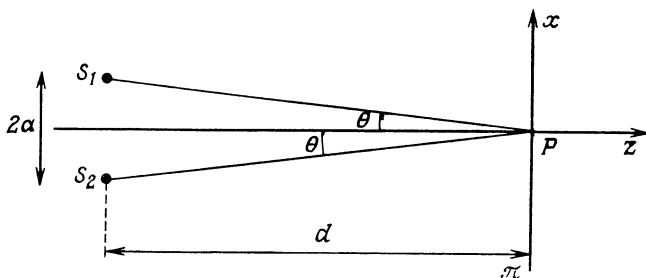


Рис. 5.2. Интерференция в случае двух точечных источников (опыт Юнга). Два когерентных точечных источника освещают фотоземлю, помещенную в плоскость π . Данный случай можно рассматривать как запись голограммы. Точка S_1 — объект, а точка S_2 — опорный источник

но из рис. 5.2, данная схема аналогична схеме опыта Юнга — интерференции в случае двух отверстий (гл. 4, § 4, п. Б). Для простоты будем считать, что обе волны O_1 и O_2 являются плоскими вблизи плоскости π . В первом приближении это условие выполняется достаточно точно, если выбрать $a/d \ll 1$ и рассматривать только небольшую окрестность вокруг точки P . Точный расчет с использованием сферических волн O_1 и O_2 приводит к тем же результатам, но путем более сложных вычислений.

Итак, объектная волна имеет комплексную амплитуду $O = \exp\{i(2\pi/\lambda)\theta\}$, если нормирована ее интенсивность и

$\sin \theta \approx \theta$. Комплексная амплитуда опорной волны при тех же условиях такова:

$$R = \exp \{ -i (2\pi/\lambda) \theta \}.$$

Интенсивность в картине интерференции этих двух волн равна

$$I = |O + R|^2 = 4 \cos^2 \left(\frac{2\pi}{\lambda} \theta \right) = 2 + 2 \cos \left(\frac{4\pi}{\lambda} \theta \right),$$

что, конечно, хорошо известно. Если мы зарегистрируем это распределение интенсивности на идеальной фотопластинке, помещенной в плоскость π , то получим дифракционную решетку с периодом, равным

$$\delta = \frac{\lambda}{2\theta} = \frac{\lambda d}{2a},$$

и амплитудным пропусканием вида

$$t(x) = t_0 \left[1 + \cos \left(\frac{4\pi}{\lambda} \theta \right) \right],$$

где t_0 — коэффициент пропорциональности.

Осветим теперь полученную дифракционную решетку опорной волной. Комплексная амплитуда волны, прошедшей сквозь решетку, будет равна

$$\begin{aligned} A = Rt &= t_0 R \left[1 + \cos \left(\frac{4\pi}{\lambda} \theta \right) \right] = \\ &= t_0 R + \frac{t_0}{2} \exp \left(i \frac{2\pi}{\lambda} \theta \right) + \frac{t_0}{2} \exp \left(-i \frac{6\pi}{\lambda} \theta \right). \end{aligned}$$

В этом выражении легко узнать три члена: R — опорный пучок, $\exp \{ i (2\pi/\lambda) \theta \}$ — объектный пучок O , $\exp \{ -i (6\pi/\lambda) \theta \}$ — пучок, комплексно-сопряженный с объектным.

Укажем направление всех этих пучков: 1) пучок R образует угол $+\theta$ с оптической осью, как и исходная падающая волна; 2) пучок O образует угол $-\theta$ с оптической осью и, следовательно, угол -2θ с опорным пучком; 3) комплексно-сопряженный пучок образует угол $+3\theta$ с оптической осью и, следовательно, угол $+2\theta$ с опорным пучком. Это общее свойство сопряженного пучка — он симметричен объектному пучку относительно опорного (рис. 5.3).

Здесь мы вновь обнаруживаем классическую дифракцию на синусоидальной дифракционной решетке. Действительно, как было показано ранее, распределение амплитуд на бесконечном удалении есть фурье-образ распределения амплитуд в падающей волне, умноженный на коэффициент пропускания дифракци-

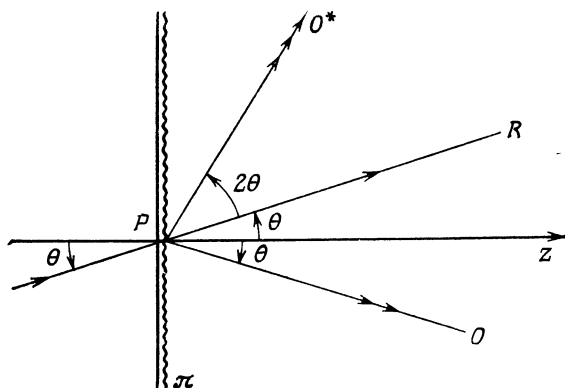


Рис. 5.3. Три пучка R , O и O^* , которые получаются, если осветить фотопластинку (экспонированную так, как показано на схеме рис. 5.2) точечным источником (т. е. восстанавливающей волной, идентичной опорной волне).

Пучки R , O и O^* можно интерпретировать двумя способами: а) в соответствии с теорией дифракционных решеток: R — недифрагированный пучок нулевого порядка, O и O^* — пучки порядков $+1$ и -1 , дифрагировавшие на синусоидальной решетке; б) в соответствии с теорией голографии: R — часть опорного пучка, не искаженная голограммой, O — волна, формирующая восстановленное изображение объекта, O^* — волна, формирующая комплексно-сопряженное изображение объекта.

рующей поверхности (гл. 4, § 1, п. В). Следовательно, можно написать

$$A(\alpha) = \mathcal{F}[Rt] = \mathcal{F}\left[\exp\left\{-i\frac{2\pi}{\lambda}\theta\right\} \times \right. \\ \left. \times \left(1 + \frac{1}{2}\exp\left\{i\frac{4\pi}{\lambda}\theta\right\} + \frac{1}{2}\exp\left\{-i\frac{4\pi}{\lambda}\theta\right\}\right)\right] = \\ = \delta(\alpha - \theta) + \frac{1}{2}\delta(\alpha + \theta) + \frac{1}{2}\delta(\alpha - 3\theta).$$

Итак, на бесконечности распределение амплитуд в поле дифракции имеет три максимума в направлениях $+\theta$, $-\theta$ и $+3\theta$. Это порядки дифракции 0 , -1 и $+1$. Период решетки был равен $\delta = \lambda/2\theta$, и поэтому порядки $+1$ и -1 отделены от нулевого порядка углом $\alpha = \lambda/\delta = 2\sin\theta$.

Для простоты мы предполагали, что угол θ очень мал. Это допущение удобно, но совсем не обязательно. С увеличением угла θ мы не можем больше приравнивать его величине $\sin\theta$ в формуле для амплитуды $A(\alpha)$ в дальней зоне. Тогда оба расчета приводят к одинаковым результатам, если угол дифракции α заменить величиной $\sin\alpha = \lambda/\delta = 2\sin\theta$.

Мы показали, каким образом голограмма двух точек просто сводится к интерференции и дифракционным решеткам. Прямое обобщение на основе принципа линейности, на котором базируется вся теория электромагнитного поля, позволило бы нам представить голограмму некоторого объекта в виде суммы

элементарных голограмм пар точек, связав виртуально каждую точку объекта с некой опорной точкой. Но мы не будем в это углубляться, так как введение многочисленных перекрестных членов быстро приводит к очень сложным формулам.

Тем не менее мы видим, что Юнг мог бы заниматься голографией, не подозревая об этом, если бы Ньепс¹⁾ родился на 30 лет раньше!

Нам постоянно нужно иметь в виду аналогию между голограммой и дифракционными решетками при исследовании голографического метода. Вот некоторые из его особенностей.

1. *Влияние длины волны.* Если голограмма записывается на длине волны λ_0 , а восстанавливается излучением с другой длиной волны λ , то каждая точка объекта формирует дифрагированную волну в направлении $\theta' = (\lambda/\lambda_0)\theta$, а не под углом θ . В результате формируется изображение с некоторым увеличением, отличным от единицы.

2. *Влияние нелинейности эмульсии.* Если в опыте Юнга при записи голограммы пропускание пластинки не является линейной функцией интенсивности, то мы все-таки запишем функцию с периодом δ , но ее профиль будет несинусоидальным. Поэтому поле дифракции опорной волны на такой дифракционной решетке будет состоять не из трех волн (0, $+1$ и -1), а из большего числа волн в направлениях θ_k , удовлетворяющих условию $\theta_k = k\lambda/\delta$ ($k = \pm 1, 2, 3, \dots$). Это гармоники, необходимые для восстановления профиля решетки при синтезе Фурье. На практике нелинейность эмульсии проявляется в уменьшении отношения сигнала к шуму и контраста изображения.

3. *Влияние толщины эмульсии.* Если мы возьмем толстую (по сравнению с длиной волны) эмульсию, то интерференционные полосы образуют слои в глубине самой эмульсии, эквивалентные трехмерной брэгговской дифракционной решетке. В отличие от тонкой решетки, которая дифрагирует всю падающую волну под некоторым углом, толстая решетка осуществляет селекцию падающих волн по углу. В самом деле (рис. 5.4), две волны, частично отраженные от разных слоев, окажутся в фазе только в том случае, если период решетки δ , угол падения β и длина волны λ удовлетворяют соотношению

$$\sin \beta = \frac{k\lambda}{2\delta}. \quad (5.2)$$

Данное соотношение, называемое «условием Брэгга», есть условие, при котором позади эмульсии существует дифрагировавший пучок. В силу этого условия осуществляется очень сильная угловая селекция пучков восстановления. Мы еще раз встретимся с данной особенностью толстой голограммы — ее свой-

¹⁾ Ж. Ньепс (1765—1833) — французский физик, один из создателей фототрафин. — Прим. ред.

ствами как трехмерной дифракционной решетки — в двух случаях: гл. 5, § 5, п. В, где мы качественно рассмотрим характеристики таких голограмм, а затем в гл. 5, § 7, п. Б, где речь пойдет об их применении для хранения информации (в голографических запоминающих устройствах).

Здесь же отметим, что очень тонкая эмульсия позволяет получать голограммы, которые могут быть восстановлены пучком с пространственной ориентацией, значительно отличающейся от ориентации опорного пучка. По мере того как увеличивают толщину, требования к углу падения становятся все более жесткими (гл. 14, § 2).

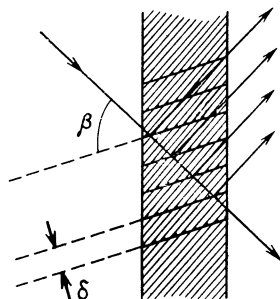


Рис. 5.4 Брэгговская дифракция на толстой голограмме.

4. *Влияние выбора метода кодирования интерференционной картины.* Мы знаем, что помимо периода дифракционной решетки важен ее профиль в пределах одного периода. В зависимости от того, воздействует ли этот профиль на фазу волны или амплитуду, а также от того, будет ли модуляция непрерывной или дискретной (бинарной), эффективность дифракционной решетки будет существенно изменяться. (Под эффективностью дифракционной решетки понимают отношение энергии, дифрагировавшей в первый порядок, к полной энергии, падающей на решетку.) Теми же свойствами обладают и голограммы, и, поскольку обычно желательно иметь максимальную дифракционную эффективность, мы укажем некоторые ее предельные значения, которых позволяют достичь разные методы кодирования (рис. 5.5) [40].

1. *Синусоидальная амплитудная модуляция*

$$t(x) = \frac{1}{2} + \frac{a}{2} \sin\left(\frac{2\pi x}{\delta}\right).$$

Это решетка с непрерывным изменением уровня серого. Максимальная дифракционная эффективность равна 6,2 % при $a = 1$.

2. *Бинарная амплитудная модуляция*

$$t(x) = \frac{1}{2} + \frac{a}{2} \text{rect}\left(\frac{x}{\delta}\right).$$

При $a = 1$ дифракционная эффективность достигает 10 %.

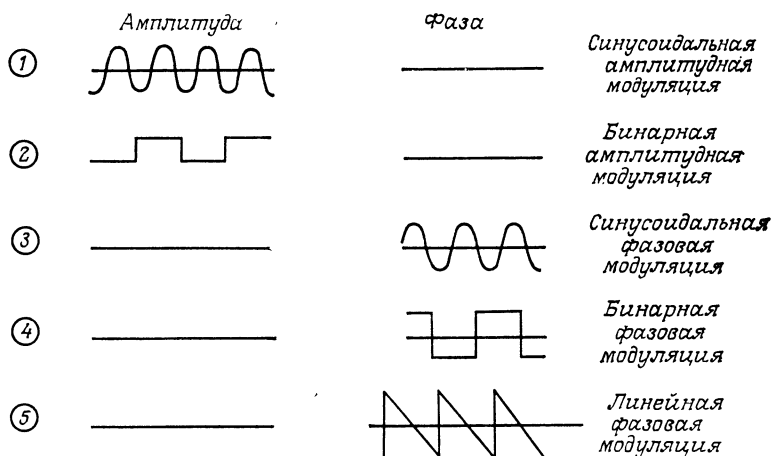


Рис. 5.5. Разные способы кодирования записи интерференционной картины. При экспонировании интерференционной картиной светочувствительный материал может изменять либо свое пропускание (в этом случае он будет модулировать восстанавливающую волну по амплитуде), либо показатель преломления или толщину (тогда он будет модулировать восстанавливающую волну по фазе)

3. Синусоидальная фазовая модуляция

$$t(x) = \exp \left[i \frac{\pi a}{2} \sin \left(\frac{2\pi x}{\delta} \right) \right].$$

Этот вид модуляции может быть получен, например, отбеливанием амплитудной решетки с синусоидальной модуляцией (приложение 5). Дифракционная эффективность такой фазовой решетки равна 34% при $a = 2$.

4. Бинарная фазовая модуляция

$$t(x) = \exp \left[i \frac{\pi a}{2} \operatorname{rect} \left(\frac{x}{\delta} \right) \right].$$

При $a = 2$ дифракционная эффективность достигает 41%.

5. *Линейная фазовая модуляция* (решетка с треугольным профилем)

$$t(x) = \exp i\pi \left(1 - \frac{2\bar{x}}{\delta} \right),$$

$$\text{где } \frac{\bar{x}}{\delta} = \frac{x}{\delta} \bmod (1).$$

Такая дифракционная решетка ведет себя подобно зеркалу, если амплитуда фазовой модуляции $2a$ равна 2π . В этом случае ее эффективность равна 100%.

Как видно из приведенных значений дифракционной эффективности, при записи голограмм весьма выгодно фазовое кодирование. Однако такое кодирование довольно сложно в практическом осуществлении и требует применения специальных носителей (приложение 5).

§ 4. Экспериментальная запись голограмм

Теперь нам нетрудно ответить на вопрос, который остался нерешенным: как сделать так, чтобы два паразитных пучка, записанных одновременно с объектной волной, не исказили ее? Выберем достаточно большой угол θ между опорным пучком

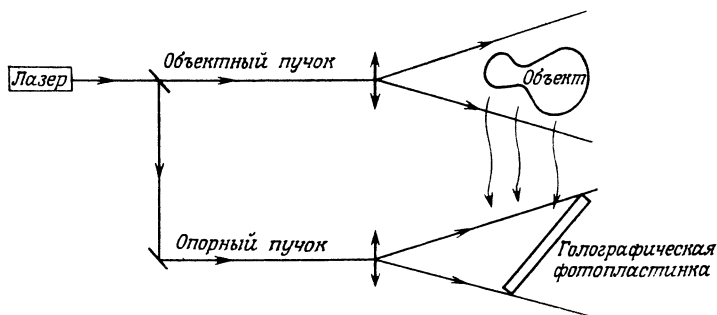


Рис. 5.6. Схема записи голограммы Лейта и Упатниекса, обеспечивающая разделение объектного и опорного пучков.

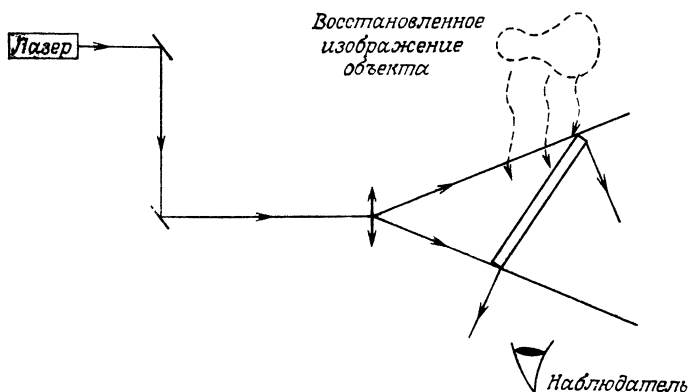


Рис. 5.7. Схема восстановления голограммы, полученной по схеме рис. 5.6

и объектной волной, чтобы эти два пучка перекрывались только в непосредственной близости от пластинки. Тем самым мы *разделим в пространстве* три совмещенных на голограмме слагаемых. Такое решение было предложено Лейтом и Упатниексом [33], и оно является основой большинства современных схем оптической голографии. На рис. 5.6 приведена упрощенная схема записи голограмм, использующая принцип пространственного разделения пучков, а на рис. 5.7 показана схема восстановления голограммы, записанной таким образом. Как видно

из рис. 5.7, в схеме восстановления имеется только пучок, использовавшийся при записи в качестве опорного. В этом случае наблюдатель видит объект, находящийся в том положении, в каком он был при регистрации.

Теперь мы знаем, в чем секрет голографии, и можем, казалось бы, сами записать голограмму. Но мы не учли всех следствий из допущений, сделанных в теории. Оказывается, что эти следствия, если ими пренебречь, могут привести к полной неудаче.

А. Когерентность

Все сказанное выше верно лишь при условии, что объектный и опорный пучки могут интерферировать. Поэтому временная когерентность источника (гл. 4, § 4) должна быть такой, чтобы разность хода в каждой точке голограммы между опорным пучком и всеми дифрагированными на объекте пучками была меньше длины когерентности. На практике этим условием существенно ограничиваются размеры объектов, при которых возможна голографическая запись.

Лазер, работающий в режиме непрерывной генерации (например, гелий-неоновый), обычно характеризуется длиной когерентности в несколько десятков сантиметров (путем точной селекции мод возможно получить длину когерентности в несколько метров и даже в несколько километров). Для обычных же спектральных ламп длина когерентности крайне редко бывает больше миллиметра, а в лампах накаливания она не превышает нескольких микрометров.

Пространственная когерентность также необходима для записи голограммы. Но если ее нет, то можно различными методами разделять волновой фронт и осуществлять интерференцию лучей, исходящих из одной и той же точки источника. Это те методы, которые использовались для записи голограмм с помощью некогерентных источников. Один такой пример мы встретим в § 5, п. Д.

Б. Фотоэмульсия

Мы уже отмечали, что голограммы можно записывать на разных светочувствительных материалах (приложение 5). Ниже мы опишем свойства и характеристики одного из этих материалов, чаще всего используемого на практике, и рассмотрим их влияние на запись простой голограммы, а именно голограммы точечного объекта, отделенного от точечного опорного источника углом θ .

1. Разрешение

Период дифракционной решетки, регистрируемой на фотоэмульсии, будет равен $\delta = \lambda / \sin \theta$. При длине волны $\lambda = 0,6$ мкм и $\theta = 60^\circ$ это величина порядка 0,7 мкм. Фотоэмульсия лишь тогда позволит зарегистрировать такую решетку, когда средний размер ее зерен значительно меньше этой величины. Требуется же записывать несколько тысяч «штрихов» на миллиметр. Следовательно, наши требования к разрешающей способности фотоэмульсии будут зависеть от того, какой угол мы хотим иметь между записываемыми пучками. Этим практически ограничивается разделение пучков при восстановлении.

2. Частотно-контрастная характеристика

Фотоэмульсия, однако, характеризуется не только размером зерен. Как всякую линейную и пространственно-инвариантную оптическую систему, ее свойства и поведение можно описать частотно-контрастной характеристикой (приложение 5,

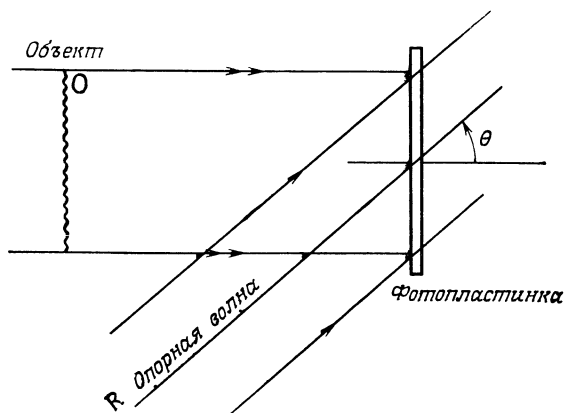


Рис. 5.8. К определению частотно-контрастной характеристики фотопленки. Объект O — синусоидальная решетка с пространственной частотой ν .

рис. П5.3). Посмотрим, как фотоэмульсия влияет на качество голограммы. Выберем в качестве объекта синусоидальную амплитудную решетку с частотой ν , а опорную волну направим на фотопластинку под углом θ (рис. 5.8), т. е.

$$O = o \exp(-i2\pi\nu x),$$

$$R = r \exp\{-i(2\pi/\lambda)x \sin \theta\} = r \exp(-i2\pi\nu_0 x).$$

Фотоэмульсия регистрирует распределение интенсивности вида

$$I = o^2 + r^2 + 2or \cos 2\pi x (\nu_0 + \nu).$$

На рис. 5.9 показано, как фотоэмульсия искажает однородный (в полосе пространственных частот $-\Delta\nu$, $\Delta\nu$) спектр объекта при голографической регистрации объекта. Таким образом, фотоэмульсия вносит контрастные искажения в восстановленное изображение, особенно заметные на высших пространственных частотах. Это проявляется также и в снижении разрешения (четкости) восстановленного изображения.

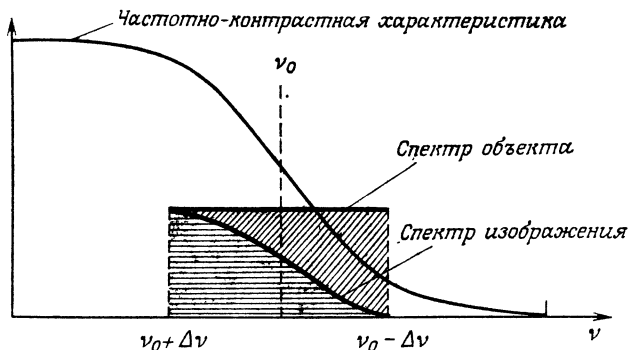


Рис. 5.9. Влияние ЧКХ фотопленки на пространственно-частотный спектр объекта, зарегистрированного голографическим методом.

По оси абсцисс отложена пространственная частота, по оси ординат — интенсивность восстановленной волны

Напомним, что разрешающая способность «мелкозернистых» эмульсий для любительских черно-белых пленок не превышает 200 линия/мм, а так называемых «быстрых» (высокочувствительных) эмульсий — от 30 до 100 линия/мм. Чувствительность же эмульсии находится в обратной зависимости от размеров ее зерна. Поэтому, если мы выбираем «голографическую» эмульсию (разрешающая способность от 1000 до 3000 линия/мм), то нам придется работать при длительных экспозициях или пользоваться очень яркими источниками света.

3. Линейность характеристики пропускания фотоэмульсии

К фотоэмульсии предъявляется еще и третье требование — линейности характеристики амплитудного пропускания. Выше мы предполагали, что пропускание фотопластинки линейно зависит от интенсивности света при ее экспонировании. На самом же деле характеристические кривые фотопленок имеют вид, приведенный на рис. 5.10 и П5.1. Из этих кривых видно, что фотоэмульсии обеспечат линейную в первом приближении запись голограмм только на участке AB . По этой причине интенсивность опорной волны R и объектной волны O в уравнении (5.1) выбирают таким образом, чтобы выполнялись условия $|O|^2 + |R|^2 = I_0$ и $2|OR| < B - A$.

Практически это заставляет выбирать интенсивность опорной волны, в 3—10 раз превышающую интенсивность объектной. Разумеется, такое соотношение интенсивностей приводит к заметному падению дифракционной эффективности голограммы, но при этом устраняются два главных неудобства, связанные с нелинейностью, а именно искажение распределения интенсивности на объекте и появление паразитных порядков дифракции. Отметим еще и такие свойства фотоэмульсий:

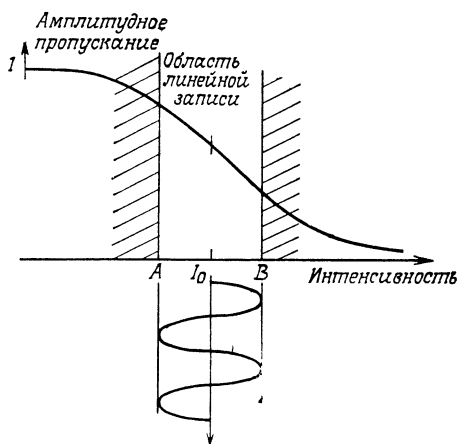


Рис 5.10. К выбору интенсивности опорной волны I_0 , обеспечивающей запись объекта на линейном участке (A, B) кривой зависимости амплитудного пропускания фотоэмульсии от экспозиции

1) *пассивность*, выражающаяся в том, что фотоэмульсии могут только поглощать свет или изменять его фазу, но не могут его усиливать;

2) ограниченный *динамический диапазон*, еще более сужающийся в связи с тем, что можно использовать лишь линейный участок характеристики амплитудного пропускания;

3) *шум* фотоэмульсий, который обусловлен, с одной стороны, зернистостью составляющих их химических веществ, а с другой — их неоднородностью. Шум фотоэмульсии обычно сильно зависит от уровня сигнала [41].

В. Стабильность

Представим себе, что во время записи голограммы, точно в середине экспозиции, некоторая точка A объекта сместится на такую величину, что все оптические пути дифрагированных точкой A лучей, попадающих на фотопластинку, увеличатся на половину длины волны. Что мы увидим при восстановлении полученной голограммы? Для всех точек (кроме точки A) записанная голограмма будет обычной, и потому эти точки объекта будут яркими, как если бы он был освещен опорным пучком, использованным при записи голограммы. Что же касается

точки A , то здесь на голограмме записаны два точечных источника A_1 и A_2 , причем световые волны приходят от них в противофазе. Следовательно, они погасят друг друга и точка A окажется темной. Чему же равна амплитуда смещения, при которой происходит такое гашение? Половина длины волны — это примерно 0,3 мкм!

Данный пример хорошо иллюстрирует те жесткие требования к стабильности схемы записи, которые необходимо выполнять при записи голограмм. *За время экспозиции ни один элемент оптической установки не должен сместиться хотя бы на долю длины волны.* В противном случае весь объект или его часть не будут записаны на голограмме, а следовательно, не будут и восстановлены. Это требование оказывается тем более жестким, что, как мы знаем, голографические эмульсии имеют низкую чувствительность и потому требуют длительных времен экспозиции. В связи с этим для получения голограмм приходится использовать специальные виброзащищенные установки (столы), причем все элементы оптической схемы (линзы, зеркала и т. п.) должны быть жестко закреплены. По той же причине нельзя надеяться, что удастся получить без особых затруднений голограммы людей и животных. Для этого потребовалось бы более совершенное оборудование: более мощные или импульсные лазеры.

В § 7, п. А мы увидим, что эта высокая чувствительность к перемещению голографируемого объекта оказывается выгодной для экспериментатора, поскольку она делает голографию одним из самых точных методов измерения микросмещений, вибраций и деформаций.

§ 5. Различные типы голограмм

Все сказанное в предыдущих параграфах и в том числе сделанные там выводы относятся в основном к голограммам Лейта — Упатниекса, которые обычно получают при регистрации трехмерных объектов. Существуют, однако, и другие типы голограмм, которые во многих случаях создавались для того, чтобы устранить тот или иной из отмеченных выше недостатков голограмм Лейта — Упатниекса.

А. Голограммы Габора [32]

Габор, в распоряжении которого были только источники света с очень малой длиной когерентности, смешал в одном пучке и объектную и опорную волны (рис. 5.11). Предложенная им схема, однако, предъявляет к объекту следующие три требования: 1) он должен быть почти всюду прозрачным, чтобы была возможна линейная регистрация отношения интенсивно-

стей опорной и объектной волн; 2) он должен быть плоским; 3) фотопластинка должна находиться в дальней зоне дифракции (гл. 4, § 1, п. В).

Для Габора эти требования имели второстепенное значение. Его цель состояла в создании микроскопа с большим увеличением и большой глубиной резкости, и почти все объекты, с ко-

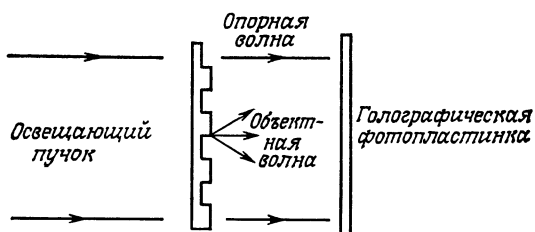


Рис. 5.11. Схема записи голограммы Габора.

Объектный и опорный пучки смешаны в одном освещающем объекте пучке.

торыми он работал, удовлетворяли трем перечисленным требованиям. В отношении же голографической регистрации они влекут за собой значительно более неприятные следствия: паразитные волны (опорная и сопряженная) восстанавливаются

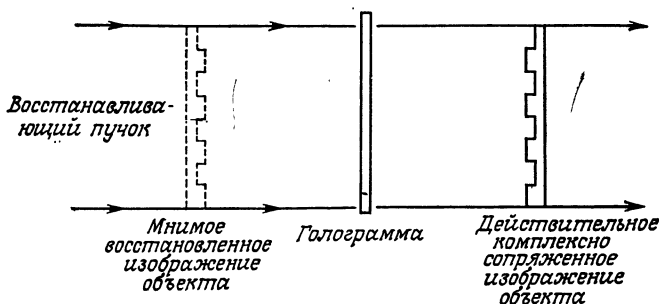


Рис. 5.12. При восстановлении голограммы Габора три дифрагировавших на ней пучка оказываются наложенными друг на друга.

в том же направлении, что и объектная волна. В результате очень яркий однородный фон и наличие сопряженного расфокусированного изображения объекта мешают наблюдателю видеть объект (рис. 5.12). Габор, а затем и другие исследователи предложили различные способы устранения этого недостатка. Но ни один из них не сравнится по эффективности с голограммами, использующими наклонный опорный пучок.

Б. Голограммы Фурье

Если голограммы Габора представляют главным образом исторический интерес, то голограммы Фурье — это один из самых распространенных видов голограмм и наиболее интересных для техники связи.

1. Принципы голографии Фурье

Голограмма Фурье получается, когда дифрагировавшая на объекте и опорная волны имеют одинаковую кривизну волнового фронта ρ в каждой точке x, y плоскости фотопластинки. При этом интерферирующие волны записываются в виде

$$R(x, y) = r \exp \left\{ i \frac{\pi \rho}{\lambda} [(x - x_r)^2 + (y - y_r)^2] \right\},$$

$$O(x, y) = o \exp \left\{ i \frac{\pi \rho}{\lambda} [(x - x_o)^2 + (y - y_o)^2] \right\},$$

где (x_o, y_o) , (x, y) и (x_r, y_r) — координаты точек на объекте, на голограмме и на опорном источнике (рис. 5.13). Квадратичные

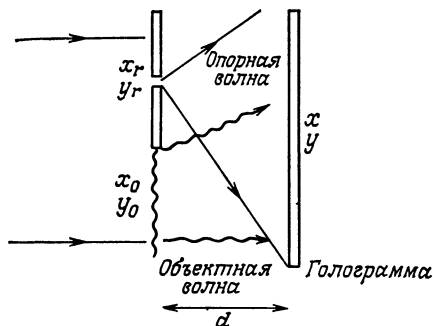


Рис. 5.13. Схема записи голограммы Фурье.

Точечный опорный источник расположен в одной плоскости с объектом. В данном случае нет необходимости в линзе.

члены описывают распространение сферической волны, имеющей радиус кривизны ρ [формулы (4.2) и (4.14)]. Следовательно, интенсивность, регистрируемая на голограмме, равна

$$I(x, y) = |o|^2 + r^2 + 2|o|r \cos \left\{ \frac{2\pi}{\lambda} \rho [x(x_o - x_r) + y(y_o - y_r) + (x_r^2 - x_o^2) + (y_r^2 - y_o^2)] \right\}. \quad (5.3)$$

Два члена $x_r^2 - x_o^2$ и $y_r^2 - y_o^2$ остаются неизменными на всей поверхности фотопластинки. Два других члена, напротив, указывают на то, что каждая точка объекта (x_o, y_o) записывается на голограмме в виде синусоидальной решетки, пространственные частоты которой таковы:

$$\frac{\rho}{\lambda} (x_o - x_r), \quad \frac{\rho}{\lambda} (y_o - y_r). \quad (5.4)$$

Таким образом, мы получили здесь *частотную модуляцию*. К этому важному результату мы еще вернемся в § 5, п. Б, 1.

Тем, что голографируемый объект представляется в виде суперпозиции элементарных синусоид, объясняется, почему этому способу записи голограмм присвоено имя Фурье. Голограммы Фурье особенно удобны при регистрации двумерных объектов. Далее мы увидим множество применений этого типа голограмм для оптической обработки информации, а также в голографической памяти. Пока что отметим, что здесь имеет место соответствие вида точка — плоскость между самим объектом, с одной стороны, и пространством пространственных частот — с другой. Это соответствие есть именно то самое преобразование Фурье, которое мы уже обнаружили, например, в выражении для дифракции света на бесконечности. Данное обстоятельство позволяет очень удобно синтезировать голограммы цифровыми методами с помощью ЭВМ.

2. Экспериментальное получение голограмм Фурье

Посмотрим теперь, как обычно записывают голограмму Фурье. На рис. 5.13 изображена схема записи, достаточно широко используемая на практике. Здесь точечный опорный источник расположен в той же плоскости, что и объект. В соответствии с принципом Гюйгенса из каждой точки объекта исходит

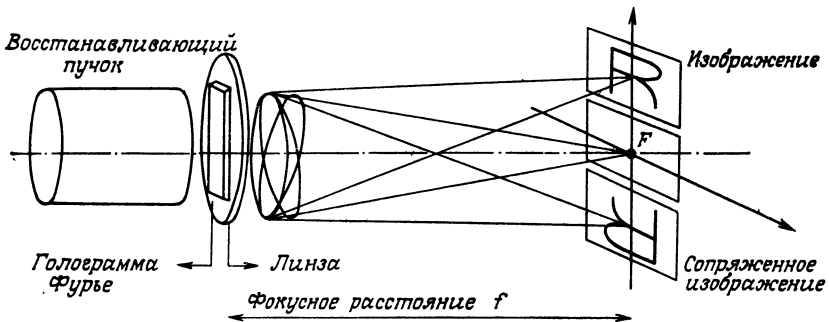


Рис. 5.14. Восстановление голограммы Фурье с помощью линзы.

Оба восстановленных изображения находятся в фокальной плоскости линзы и расположены симметрично относительно ее фокуса F .

сферическая волна, кривизна фронта которой равна кривизне фронта опорного пучка $\rho = 1/d$.

Если при восстановлении голограмму осветить параллельным пучком, то каждая элементарная решетка сформирует изображение соответствующей точки (x_0, y_0) в бесконечности.

Если теперь сразу за голограммой установить линзу, то все эти точки будут перенесены из бесконечности в фокальную плоскость линзы, как это показано на рис. 5.14. На оптической оси

мы снова найдем распределение, формируемое опорным пучком. Изображение объекта и сопряженное ему изображение теперь располагаются симметрично относительно оптической оси.

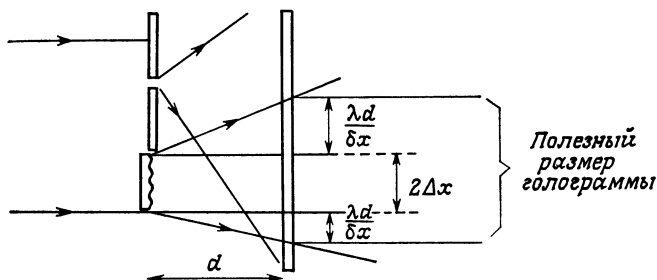


Рис. 5.15. К определению полезных размеров голограммы Фурье.

Выше мы предполагали, что у фотопластинки неограниченные размеры. Посмотрим теперь, каковы полезные размеры голограммы Фурье. Пусть δx — размер самой малой детали

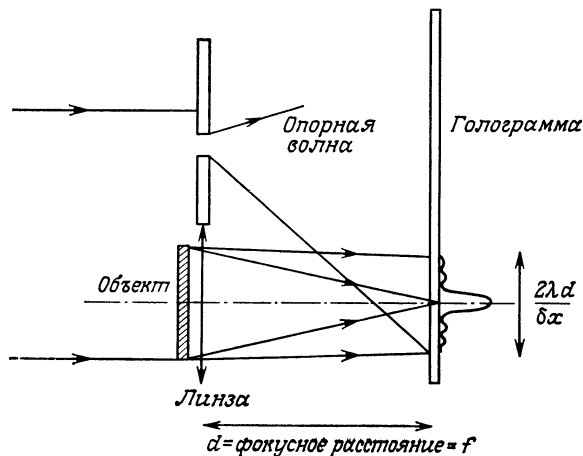


Рис. 5.16. Линза помещена в плоскость объекта, полезный размер голограммы минимален.

Под «голограммой Фурье» иногда подразумевают данный частный случай схемы записи, приведенной на рис. 5.13.

объекта. Тогда наивысшая пространственная частота в объекте будет равна $F = 1/\delta x$. Именно этой частотой и определяется разрешение объекта. Энергия, соответствующая этой пространственной частоте F , дифрагирует в плоскости голограммы в точку с абсциссой $x = \lambda d/\delta x + x_0$, где x_0 — абсцисса той точки, в которой находится мелкая деталь. Если размер объекта

равен $2\Delta x$, то голограмма Фурье будет иметь размер $2(\Delta x + \lambda d/\delta x)$ (рис. 5.15).

Чтобы уменьшить размер голограммы, можно воспользоваться схемой, приведенной на рис. 5.16. Здесь имеется линза, которая позволяет довести полезный размер голограммы до $2\lambda d/\delta x$. Этот выигрыш в размере достигается ценой значительного расширения динамического диапазона, регистрируемого голограммой распределения интенсивности, поскольку основная часть энергии содержится в области низких пространственных частот вблизи оптической оси. В поисках наилучшего компромисса между размером голограммы и ее динамическим диапазоном экспериментаторам часто приходится немного смещать фотопластинку из фокальной плоскости линзы и записывать так называемую голограмму Френеля.

Здесь мы прервем ненадолго наш обзор существующих типов голограмм и остановимся на некоторых особенно важных с точки зрения связи сторонах голографии, с которыми мы встретились при рассмотрении голограмм Фурье, но начнем, конечно, с классической голографии.

3. Голография и модуляция

Как было показано в § 2, после фотохимической обработки амплитудное пропускание голограммы пропорционально величине

$$O^2 + R^2 + OR^* + O^*R.$$

Если в качестве опорной выбрана плоская волна, наклоненная к объектной волне под небольшим углом θ , а дифрагированная на объекте объектная волна в плоскости фотопластинки имеет вид

$$O = o \exp \left(i \frac{2\pi\varphi}{\lambda} \right),$$

где o и φ — функции координат x и y , то пропускание полученной голограммы будет равно

$$t \sim o^2 + r^2 + or \exp \left\{ -i \frac{2\pi}{\lambda} (\theta - \varphi) \right\} + O^*R.$$

Таким образом, голографическая регистрация осуществляется за счет модуляции двух видов: 1) *амплитудной* — запись информации об амплитуде волны; 2) *фазовой* — запись информации об ее фазе.

В случае голографии Фурье имеются некоторые отличия. Мы уже видели, что вблизи фотопластинки дифрагированное объектом поле тесно связано с частотным содержанием объекта. В частности, в случае схемы, изображенной на рис. 5.16, это поле представляет собой фурье-образ распределения комплекс-

ных амплитуд на объекте (данное свойство линзы, с которым мы уже встречались в гл. 4, § 2, будет основой для изложения в гл. 6). Следовательно, полезный размер голограммы $2\lambda d/\delta x$ прямо связан с *шириной спектра пространственных частот* объекта. Но если в формуле (5.3) выделить частоты волны, то мы получим *частотную модуляцию* несущей — модуляцию, осуществляемую самими частотами объекта.

Эти аналогии с техникой связи часто будут нам полезны. Они естественным образом приведут нас к таким хорошо известным в электронике операциям, как пространственная фильтрация высоких или низких частот, полосовая фильтрация, демодуляция, мультиплексирование (уплотнение). Подобные параллели будут использованы в цифровой голографии (§ 6), а также в оптической пространственной фильтрации (гл. 6).

Вернемся теперь к обзору основных типов голограмм, а именно к объемным (толстым) голограммам — весьма важному виду голограмм, применяемых в устройствах оптической обработки информации и оптической памяти.

В. Объемные (толстые) голограммы [42]

В § 3 было показано, что свойства голограммы будут существенно различаться в зависимости от того, записана ли она на «тонкой» или «толстой» эмульсии. При этом мы подчеркнули, что тонкая голограмма восстанавливается приблизительно с одинаковой эффективностью независимо от ориентации

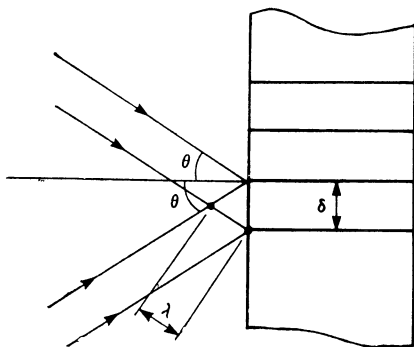


Рис. 5.17. Запись толстой голограммы.

восстанавливающего пучка, тогда как при восстановлении толстой голограммы для него должно выполняться условие Брэгга [формула (5.2)]. Это условие должно выполняться и для опорной волны. При записи голограммы объектный и опорный пучки образуют между собой угол 2θ (рис. 5.17). Слои почернения в эмульсии образуются по биссектрисе угла между пучками, а расстояние между слоями, которые соответствуют миниму-

мам в распределении интенсивности, будет равно $\delta = \lambda/2 \sin \theta$. Отсюда легко находятся углы, при которых выполняется условие Брэгга, в случае, когда восстанавливающая волна имеет ту же частоту, что и использованная при записи, а именно $\beta = +\theta$, а также $\beta = \pm(\pi - \theta)$. Решения вида $\sin \beta = k \sin \theta$ с $k \neq \pm 1$ существуют только при достаточно малых значениях β .

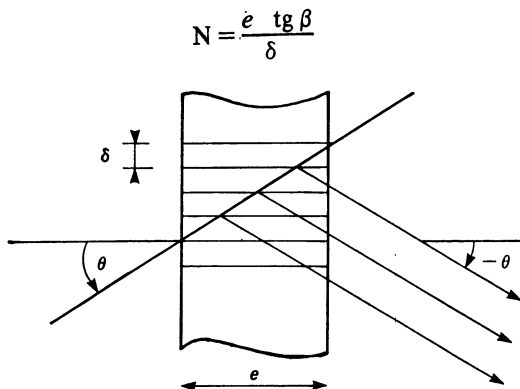


Рис. 5.18. Дифракция световой волны на толстой голограмме.

Интересно посмотреть подробнее, каким образом толщина эмульсии e влияет на пространственную селективность (избирательность) падающих лучей. Луч, падающий на голограмму под углом β , встретит на своем пути приблизительно N слоев (рис. 5.18):

$$N = \frac{e \operatorname{tg} \beta}{\delta}.$$

Предположим, что амплитуда волны не уменьшается при прохождении каждого слоя, поскольку отражение от него очень мало. Тогда полная амплитуда волны, отраженной в направлении $-\beta$, будет пропорциональна величине

$$A = 1 + e^{i\varphi} + e^{2i\varphi} + \dots + e^{i(N-1)\varphi},$$

где $\varphi = 4\pi\delta \sin \beta / \lambda$ — фазовый сдвиг между двумя соседними лучами, а распределение интенсивности в ее фронте будет равно

$$I = AA^* = \left(\frac{\sin \left(N \frac{\varphi}{2} \right)}{\sin (\varphi/2)} \right)^2 \quad (5.5)$$

На рис. 5.19 показано, как изменяется в этом случае интенсивность. Максимумы интенсивности тем больше, чем больше N . По существу формула (5.5) — это формула для интенсивности на выходе интерферометра Фабри — Перо [7], которой мы будем пользоваться при изучении лазерных резонаторов (гл. 12, § 1).

Максимумы интенсивности наблюдаются в тех точках, в которых выполняется условие

$$\sin \beta = k \sin \theta.$$

Ширина одного максимума $\Delta\theta$ равна удвоенной абсциссе первого нуля.

$$\sin \beta = \sin(\theta + \Delta\theta) = \left(k + \frac{1}{N}\right) \sin \theta,$$

$$\text{откуда } \Delta\theta = \frac{1}{N} \operatorname{tg} \theta = \frac{\lambda}{2e \sin \theta} = \frac{\delta}{e}.$$

Из полученной формулы видно, что угловая избирательность тем выше, чем толще эмульсия и чем больше угол между объ-

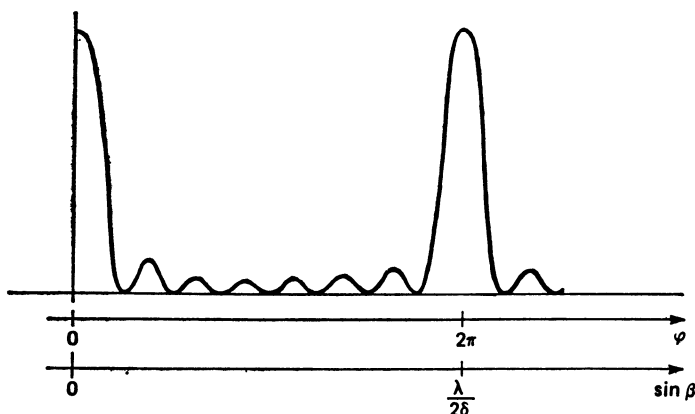


Рис. 5.19. Зависимость интенсивности волны, дифрагировавшей на толстой голограмме, от угла β восстанавливающего пучка.

ектной и опорной волнами. На рис. 5.20 представлена кривая зависимости угловой ширины дифракционного максимума при $\lambda/2e = 0,05$.

На практике толщина эмульсии может составлять 400 мкм, что соответствует допустимым отклонениям по углу порядка 1° . О применении толстых голограмм для хранения информации будет сказано в § 7, п. Б. Здесь же мы отметим одно оригинальное применение свойств толстых голограмм, тем более интересное, что оно на полстолетия опередило появление самого слова голография. Речь идет о цветной фотографии Липманна [43]. Ее принцип тот же, что и принцип объемной голографии, где объектная волна одновременно служит и опорной волной благодаря зеркалу, находящемуся в контакте с эмульсией (рис. 5.21).

В данном случае малой длины когерентности источника оказывается достаточно для записи интерференционных полос, рас-

полагающихся параллельно поверхности эмульсии (рис. 5.22). Предположим, что такого типа решетка была образована светом с длиной волны λ_0 . Если мы осветим ее белым светом, то сквозь нее пройдет только свет с длиной волны λ_0 , удовлетво-

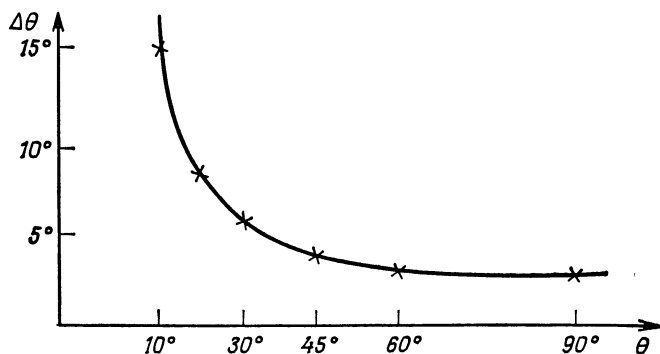


Рис. 5.20. Допустимое угловое отклонение $\Delta\theta$ направления восстанавливающего пучка в зависимости от угла θ между несущей и объектной волнами.

ряющей условию Брэгга. Фотопластинка в данном случае играет роль интерференционного фильтра.

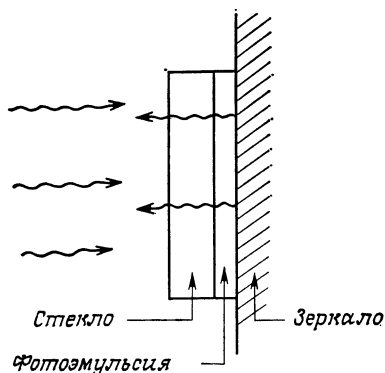


Рис. 5.21. Схема фоторегистрации по методу Липманна

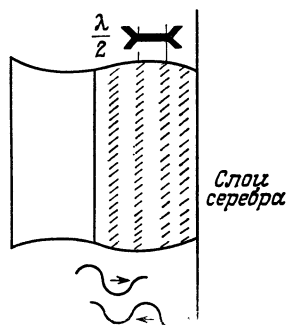


Рис. 5.22. Слой серебра, образующиеся в пучностях электрического поля при записи по методу Липманна.

Липманн использовал в своем методе три разные длины волны; свет каждой из них дает свою периодическую решетку и затем восстанавливает один из основных цветов в трехцветном синтезе. Данный метод, очень сложный в осуществлении, явился прообразом многих современных методов цветной голографии.

Г. Голограммы в белом свете [44]

С первых лет возникновения голографии представлялось привлекательным создать такие голограммы, которые, с одной стороны, не требовали бы лазера для восстановления, а с другой — позволяли бы восстанавливать не только форму объекта, но и его цвет.

Первое условие, по-видимому, выполняется в случае толстых голограмм. В самом деле, если мы запишем голограмму на длине волны λ_0 , а восстанавливать ее будем с помощью источника белого света, занимающего то же положение, которое занимал опорный источник при записи, то получим большое число источников с длиной волны λ , образующих одинаковый угол θ с фотопластинкой. И только те из них, которые будут удовлетворять условию $\sin \theta = k\lambda/2\delta$, дадут заметный вклад в восстановленное изображение. Это такие источники, для которых выполняется соотношение $\lambda = k\lambda_0$, а поскольку видимая область спектра не охватывает даже октавы (только от 0,4 до 0,7 мкм), то годится единственная длина волны λ_0 . Ширина спектра $\Delta\lambda$, пропускаемая голограммой, уменьшается с увеличением ее толщины e и определяется аналогично величине $\Delta\theta$ на основе формулы (5.5), а именно

$$\Delta\lambda = \frac{\lambda_0 \cos \theta}{2e \sin^2 \theta}.$$

Чтобы выполнить второе условие (восстановление цветного объекта), можно воспользоваться методом Липманна (рис. 5.21), приспособив его для нашей голографической задачи. Осветим объект излучением трех длин волн λ_1 , λ_2 и λ_3 , выбранных так, чтобы они соответствовали трем основным цветам трехкомпонентной теории цветового зрения (мы будем выбирать всегда красное излучение гелий-неонового лазера, а также зеленую и синюю линии излучения аргонового лазера, табл. 12.1). Далее, чтобы обеспечить стационарность интерференционной картины во всей толще эмульсионного слоя, направим три опорных пучка так, чтобы они проходили сквозь фотопластинку с обратной стороны (рис. 5.23). Тогда излучение каждой длины волны λ_i создаст решетку из слоев почернения, отстоящих друг от друга на расстояние $\delta_i = \lambda_i/2$, с амплитудой, пропорциональной амплитуде волны, дифрагированной объектом.

Восстановление полученной голограммы осуществляется белым светом. В каждой точке голограммы восстанавливаются только те длины волн, которые удовлетворяют условию

$$\lambda = 2\delta_i \pm \Delta\lambda \quad (i = 1, 2, 3).$$

Был предложен также другой метод записи цветных голограмм. В нем используется классическая схема Лейта — Упат-

ниекса, излучение трех длин волн и простая тонкая эмульсия (рис. 5.24).

На примере голограммы точечного источника мы покажем, как записывается и восстанавливается такая голограмма. При ее записи излучение каждой длины волны формирует решетку,

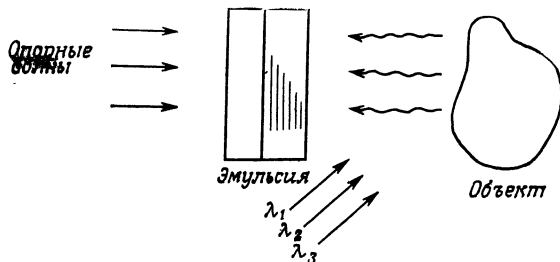


Рис. 5.23. Цветная голография, основанная на методе Липманна.

период которой δ_i связан с углом θ между объектным и опорным пучками и длиной волны λ_i соотношением

$$\delta_i = \frac{\lambda_i}{\sin \theta}.$$

Для восстановления голограммы мы возьмем один пучок, содержащий излучение трех длин волн λ_1 , λ_2 и λ_3 . Излучение

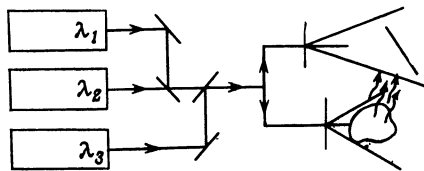


Рис. 5.24. Трехлазерная схема записи голограмм Лейта — Упатниекса.

с длиной волны λ_j будет дифрагировать на решетке δ_i под углом θ_{ji} , удовлетворяющим соотношению

$$\sin \theta_{ji} = \lambda_j / \delta_i = \lambda_j \sin \theta / \lambda_i.$$

Следовательно, каждая волна будет создавать три точечных изображения, причем только одно в направлении угла θ . Наблюдаемая картина будет, например, такой, как показано на рис. 5.25. Из семи восстановленных пучков шесть — паразитные. Слишком много! Тем не менее одно из восстановленных изображений будет содержать нужную комбинацию цветов. Если нам не нужна высокая дифракционная эффективность голограммы, то мы имеем здесь более легкий в реализации метод записи цветных голограмм, чем метод Липманна.

Правда, голограмма точечного объекта, даже цветная, не вызывает большого энтузиазма. Если мы хотим получить голограмму объекта конечных размеров, то встает вопрос: какие условия должны выполняться, чтобы изображение, восстановленное в направлении θ , не перекрывалось паразитными изображениями, восстановленными под углами θ_{12} , θ_{23} и т. д.? Возьмем красную линию излучения гелий-неонового лазера ($\lambda_1 = 0,6328$ мкм) и две линии аргонового лазера ($\lambda_2 = 0,5145$ мкм

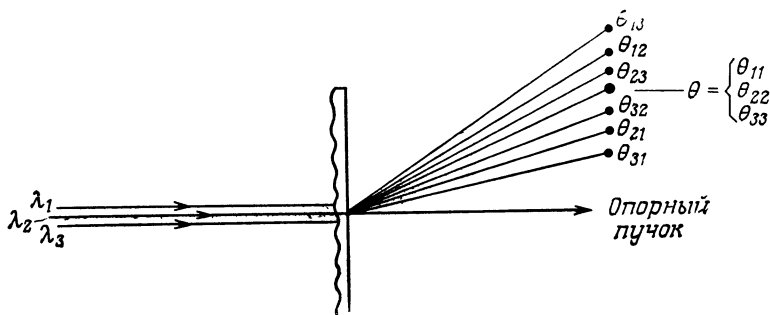


Рис. 5.25. Восстановление голограммы, полученной с помощью схемы рис. 5.24. Только пучок, дифрагированный под углом θ , содержит соответствующую информацию о 3 цветах, зарегистрированных на голограмме. Все другие пучки — паразитные.

и $\lambda_3 = 0,4881$ мкм). Свободный угол вокруг направления θ равен $\Delta\theta = (\theta - \theta_{32})/2 + (\theta_{23} - \theta)/2$. При малых углах θ получим $\Delta\theta/\theta = 0,05$, что практически совершенно недостаточно для записи цветной голограммы любого объекта.

Было предложено много способов устранения в цветных голограммах взаимосвязи между разными каналами. Один из наиболее интересных состоит в применении трех наложенных друг на друга эмульсионных слоев аналогично тому, как это делается в обычных цветных пленках. В этом случае при восстановлении каждая решетка взаимодействует только с одной длиной волны, а именно той, на которой она была записана. Другие способы основаны на применении линз и дифракционных решеток для преобразования в один пучок паразитных волн, возникающих в результате перекрестной связи между разными решетками.

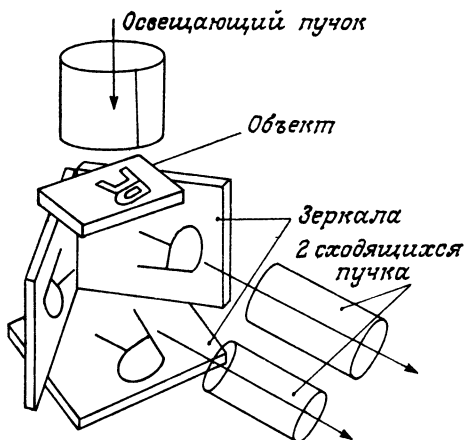
Д. Голограммы в некогерентном свете [45]

Еще одна интересная задача в голографии — как получать голограммы с помощью пространственно-некогерентных (например, протяженных) источников. Один из вариантов заключается в том, чтобы разделить объектный пучок на два одинаковых

пучка и затем заставить их интерферировать таким образом, чтобы каждая точка одного из пучков интерферировала с соответствующей ей точкой другого (рис. 4.24). Тогда при записи голограммы любой точки дифрагированную ею волну света можно одновременно рассматривать и как объектную, и как опорную. На рис. 5.26 представлена схема Строука и Рестрика

Рис. 5.26. Схема Строука и Рестрика [45] для записи голограмм в пространственно некогерентном свете

В результате последовательных отражений от 3 зеркал изображение объекта удваивается. Два его совершенно одинаковых изображения могут поточно интерферировать между собой



для записи таких голограмм. Позже было предложено множество разных усовершенствований схемы, которые, однако, не касались основного принципа — взаимно-однозначного соответствия между объектным пучком и его «зеркальным изображением».

Е. Радужные голограммы [46]

На этапе восстановления таких голограмм используется источник белого света, но восстанавливается не монохромное изображение на длине волны регистрации голограммы и не действительно цветное изображение, а непрерывная последовательность монохромных изображений, причем в каждом изображении преобладает одна определенная частота, которая воспроизводит весь спектр источника восстановления. Отсюда и название таких голограмм — радужные. Радужные голограммы отличаются превосходной дифракционной эффективностью, которая позволяет наблюдать их, например, днем при свете солнца. К сожалению, такие голограммы сохраняют параллакс только в горизонтальном направлении, и потому если в голографируемом сюжете имеется объемность в обоих направлениях, то нельзя будет видеть ни «верх» объектов, ни их «низ». Запись радужных голограмм осуществляется в два этапа. На первом этапе получают классическую голограмму H_1 объекта O , ис-

пользуя опорный пучок R (рис. 5.27, а). На втором этапе (рис. 5.27, б) полученную голограмму H_1 восстанавливают пучком R^* , сопряженным с опорным пучком R (это пучок, падающий на другую сторону голограммы H_1 и сходящийся в точечный источник, использованный на первом этапе). В результате мы восстанавливаем сопряженный пучок, который создает в месте расположения объекта O его действительное псевдоскопическое изображение O^* (т. е. изображение с обращенным рельефом, в котором выпуклостям соответствуют впадины и наоборот).

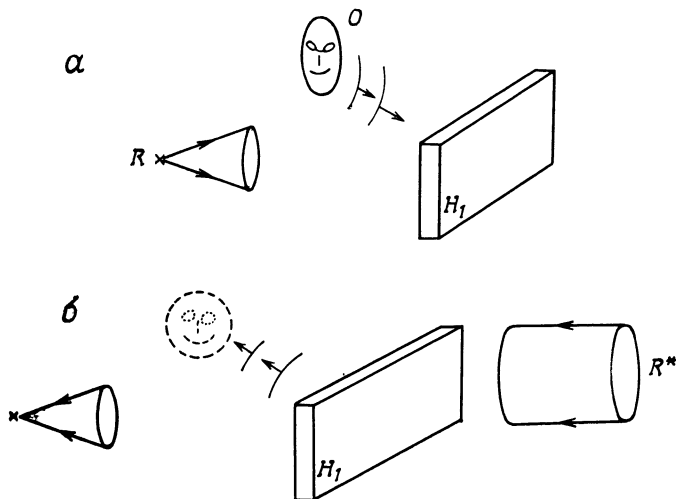


Рис. 5.27. «Радужная» голограмма [46].

a — запись голограммы H_1 ; b — восстановление голограммы H_1 с помощью пучка, «комплексно-сопряженного» с опорным. В этом случае восстановленное действительное изображение оказывается псевдоскопическим

Пока что мы описали двухэтапный метод голографии, часто применяемый для достижения особых эффектов, например для получения восстановленного изображения объекта перед фотопластинкой или на ней самой (а не за фотопластинкой, как обычно). Посмотрим теперь, как получают «радужную» голограмму.

Поместим перед голограммой H_1 маску в виде горизонтальной щели. Тогда восстановленное изображение O^* потеряет всю информацию о вертикальном параллаксе. Запишем в такой схеме вторую голограмму H_2 (рис. 5.28), используя в качестве объекта действительное изображение O , восстановленное голограммой H_1 , а в качестве опорной волны R'^* — волну, подобную R^* , но наклоненную в вертикальной плоскости. Полученную голограмму H_2 можно восстановить теперь с помощью волны R' двумя способами.

1. *С использованием монохроматической волны.* При этом мы восстановим псевдоскопическое изображение псевдоскопического изображения объекта O , т. е. сам объект, но он будет виден только тому наблюдателю, который смотрит непосредственно в щель в маске, использованной при записи голограммы H_2 .

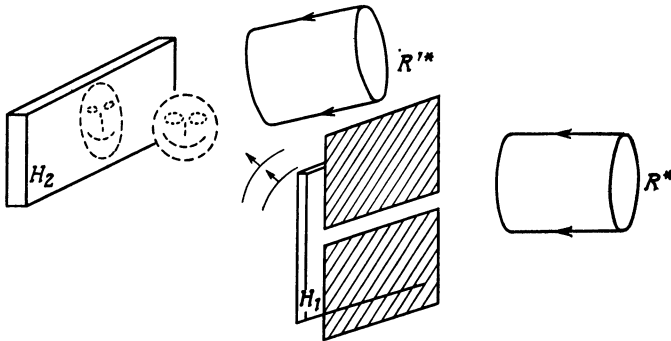


Рис. 5.28. Для получения «радужной» голограммы опорный пучок R^* наклоняют в вертикальной плоскости, а на голограмму H_1 накладывают непрозрачный экран с горизонтальной щелью.

2. *С использованием источника белого света.* При этом каждая длина волны в спектре источника сформирует свое восстановленное изображение в результате дифракции на решетке,

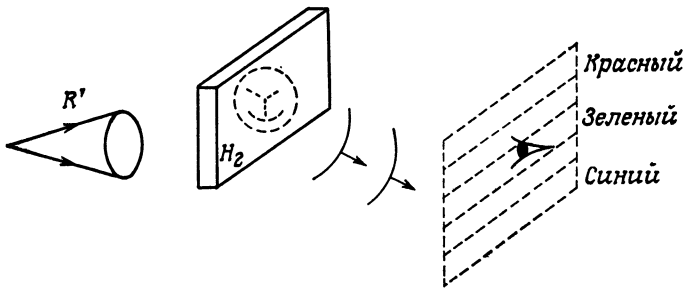


Рис. 5.29. Восстановление голограммы H_2 .

Наблюдаемое изображение окрашено в разные цвета в зависимости от положения глаз наблюдателя в вертикальной плоскости

образованной вследствие наклона пучка R' по отношению к R . Все эти изображения одного и того же объекта, накладываясь друг на друга, создают изображение объекта во всех цветах спектра, изменяющихся непрерывно от синего до красного. Дифракционная эффективность голограммы будет очень высокой, так как каждая длина волны источника дифрагирует только в малом телесном угле (рис. 5.29). Поэтому «радужные» го-

лограммы могут найти интересные применения в области визуализации информации для больших аудиторий: кино, реклама и т. п.

§ 6. Цифровые голограммы

Мы рассмотрели различные методы получения голограмм, которые можно назвать аналоговыми методами, поскольку в большинстве своем для кодирования фазы и амплитуды объектной волны они непосредственно используют оптику. Разумеется, для получения голограммы эти методы требуют наличия самого объекта.

Но во многих случаях желательно иметь цифровые голограммы, либо потому, что объект не существует (например, математическая функция для оптической пространственной фильтрации), либо потому, что он слишком дорого стоит (например, дифракционная решетка на искривленной поверхности или безабберационная линза), либо потому, что к нему нет легкого доступа (например, молекула ДНК, Эйфелева башня), либо, наконец, потому, что он очень часто изменяется (например, память для ЭВМ).

Синтез такой голограммы обычно осуществляется с помощью ЭВМ. Часто он проводится в три этапа:

1) *расчет голограммы* (вычисление пропускания фотопленки или другого носителя, который позволит ее реализовать);

2) *вывод результатов расчета на устройство*, формирующее изображение голограммы;

3) *перенос полученной голограммы на фотоэмульсию* с уменьшением, при котором ее можно было бы восстанавливать в лазерном свете.

Рассмотрим подробно все три этапа, оставив на конец самый важный из них — расчет и кодирование фазы и амплитуды.

А. Вывод голограммы

Для вывода голограмм из ЭВМ применяются графические или фотографические системы, управляемые ЭВМ. Это могут быть: графопостроитель (только для бинарных голограмм); телетайп [в случае полутонных голограмм (§ 6, п. В, 2) необходимую модуляцию по амплитуде можно обеспечить наложением друг на друга различных букв и цифр]; электронно-лучевая трубка с фотоприставкой; лазер с системой отклонения луча в двух направлениях, позволяющий исключить этап фотографического уменьшения.

От идеальной системы вывода голограммы из ЭВМ требуется:

а) возможность доступа к очень большому числу точек (часто порядка 1000×1000 в случае голограмм-фильтров и голографической памяти);

б) высокая точность адресации каждой точки;

в) быстроедействие и, разумеется, низкая стоимость!

У реальных систем вывода голограмм приемлемой стоимости имеются следующие основные недостатки: малая скорость записи (в случае графопостроителей); невысокая точность изображения (в случае электронно-лучевых трубок), а также ограниченное число синтезируемых точек, или, что эквивалентно, ограниченный размер выходного изображения (в случае теле-тайпов и графопостроителей, а также лазерных систем, использующих акусто-оптические и электро-оптические дефлекторы).

Что касается значительно более дорогих систем визуализации изображений, применяемых при цифровой обработке изображений (таких, как система с бегущим лучом или системы с вращающимся зеркальным барабаном), то у них очень хорошие характеристики. Это относится и к фотонаборным системам, управляемым ЭВМ, которые применяются при печатании газет. К сожалению, такое совершенное оборудование еще не часто можно встретить в вычислительных центрах, так что часто приходится довольствоваться более скромными средствами.

Б. Фотоуменьшение

Эта операция позволяет превратить полученную цифровую голограмму, площадь которой иногда достигает квадратного метра, в окончательную голограмму, пригодную для восстановления в свете лазера, размеры которой составляют несколько квадратных миллиметров. Следовательно, коэффициенты линейного уменьшения часто достигают 100. Это заставляет строго подходить к выбору фотографических объективов; к контролю за освещением, особенно при пересъемке полутонных голограмм; к выбору числовой апертуры используемой оптики (т. е. ее разрешения), чтобы обеспечить хорошую запись самых высоких пространственных частот, содержащихся в голограмме; к контролю за фокусировкой, особенно сложному при больших апертурах.

Отмеченными трудностями пересъемки голограммы с большим уменьшением, а также сопряженной с ней потерей качества голограммы объясняется то, что очень часто оптики разрабатывали системы микротелетайпов, обеспечивающие прямое получение окончательных размеров документа. Это заметно уменьшает время синтеза голограммы. Но в некоторых случаях [47] операция фотоуменьшения специально предусматривается для осуществления пространственной фильтрации в некогерент-

ном свете (для сглаживания), чтобы повысить дифракционную эффективность голограммы за счет подавления паразитных порядков дифракции, появляющихся вследствие дискретизации (рис. 5.30).

В. Расчет голограммы

Под расчетом голограммы обычно понимают не только определение требуемых распределений амплитуды и фазы волны, прошедшей через голограмму, но и кодирование, учитывающее те физические методы, которыми будет изготавливаться синтезированная голограмма. На самом деле это совершенно разные операции, и их ни в коем случае не следует путать. Приведем один пример. Сферической волне с определенным радиусом

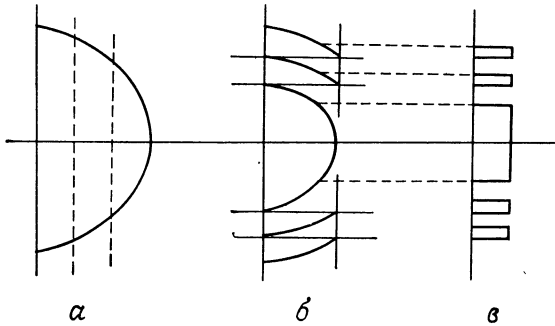


Рис. 5.31. Обычная линза (а), линза Френеля (б) и линза Сорэ (в).

Эти три вида линз позволяют сформировать один и тот же сферический волновой фронт на заданной длине волны.

соответствует только одно распределение амплитуд и фаз. Однако, по крайней мере частично, плоскую волну можно преобразовать в сферическую с помощью тонкой линзы, параболического зеркала, линзы Френеля, линзы Сорэ и еще многих других оптических элементов (рис. 5.31).

Аналогичным образом и при цифровом синтезе голограммы в нашем распоряжении будет много способов, когда нужно будет физически создать такую-то волну или такой-то пространственный фильтр или восстановить такой-то объект. Мы сначала покажем, каким образом выбором цифрового решения при синтезе голограммы будет ограничиваться набор волн, которые могут быть сформированы, а затем дадим обзор существующих методов синтеза голограмм, а также рассмотрим их принципы и достоинства.

1. Основные ограничения цифровых голограмм

Хотя можно рассчитывать голограммы всех типов, обычно синтезируют только голограммы Фурье. Расчет таких голограмм прост, а восстанавливаются они легко. Ниже мы остановимся на этом частном случае. Некоторые исследователи использовали для синтеза голограмм схему Френеля (§ 5, п. Б), преимущества которой мы увидим немного дальше.

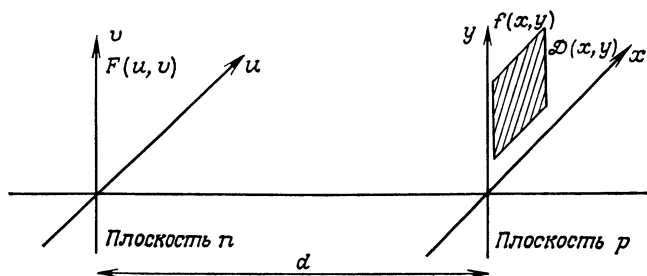


Рис. 5.32. К определению комплексной амплитуды волны в плоскости π по распределению $f(x, y)$, заданному в области \mathcal{D} плоскости P .

В самом общем виде задача синтеза голограммы представлена на рис. 5.32: нужно найти такое распределение комплексных амплитуд $F(u, v)$ непосредственно за голограммой, которое позволило бы получить заданное распределение комплексных амплитуд $f(x, y)$ в области $\mathcal{D}(x, y)$.

Такая задача математически неоднозначна. Плоскости π и P связаны между собой преобразованием Фурье. Поэтому любая функция, идентичная функции f в области \mathcal{D} и квадратично интегрируемая, имеет подходящий для нас фурье-образ. Следовательно, существует бесконечное множество функций F , и, например, для решения нашей задачи достаточно было бы выбрать такую, которая минимизирует потери энергии вне области \mathcal{D} .

Но математические решения задачи редко подчиняются физическим требованиям, тем более что последние оказываются весьма жесткими. Физические условия таковы:

1. *Область задания функции F ограничена* (по крайней мере носителем, на котором будет изготовлена голограмма, а также, вероятно, и используемыми оптическими системами). Шириной Δu голограммы непосредственно определяется разрешение в восстановленном изображении. Если обозначить минимальное расстояние между двумя независимыми точками в изображении (т. е. предел разрешения) через δx , то можно написать

$$\delta x = \frac{\lambda d}{\Delta u}.$$

2. Значения функции F также ограничены ($|F| \leq 1$). Действительно, голографическая запись пассивна, т. е. волна проходит сквозь голограмму без усиления. Это часто заставляет нас уменьшать яркость изображения $f(x, y)$, чтобы обеспечить нужный динамический диапазон. Разумеется, такое ослабление возможно лишь в ущерб общему энергетическому балансу системы.

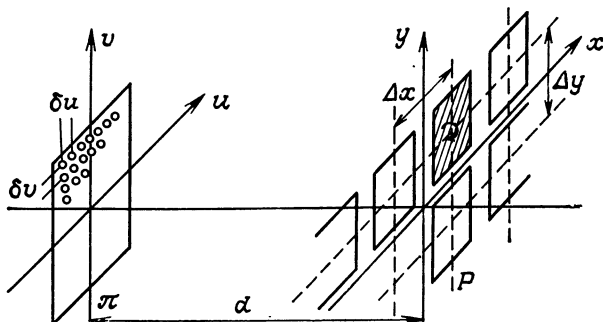


Рис. 5.33. Дискретизация голограммы (в плоскости π). Эта операция приводит к повторению участка \mathcal{D} в плоскости P .

3. Функция F должна быть дискретизованной по обоим пространственным переменным u и v . Если элементарные шаги дискретизации равны соответственно δu и δv , то непрерывную функцию F переводят в дискретизованную функцию \tilde{F} в соответствии с соотношением (рис. 5.33)

$$F(u, v) \leftrightarrow \tilde{F}(k \delta u, l \delta v) = \tilde{F}(u, v) \delta(u - k \delta u) \delta(v - l \delta v),$$

$$\left. \begin{aligned} -\frac{\Delta u}{2} &\leq u \leq \frac{\Delta u}{2} \\ -\frac{\Delta v}{2} &\leq v \leq \frac{\Delta v}{2} \end{aligned} \right\} \leftrightarrow \left\{ \begin{aligned} -K &\leq k \leq K, \\ -L &\leq l \leq L. \end{aligned} \right.$$

Эта операция, необходимая при обработке сигнала F цифровыми методами, приводит к периодическому повторению изображения $f(x, y)$, причем периоды повторения по соответствующим осям равны

$$\Delta x = \frac{\lambda d}{\Delta u}, \quad \Delta y = -\frac{\lambda d}{\Delta v}.$$

Таким образом, имеем следующее выражение для изображения:

$$f(x, y) \rightarrow \frac{\lambda^2}{\delta u} \frac{d^2}{\delta v} f(x, y) * \sum_{k=-\infty}^{+\infty} \delta(x - k \Delta x) * \sum_{l=-\infty}^{+\infty} \delta(y - l \Delta y).$$

Итак, дискретизация сопровождается потерей полезной энергии в области \mathcal{D} и появлением в плоскости наблюдения изображений паразитных порядков, что приводит к искажению изображения в области \mathcal{D} .

Напомним, что функция и ее фурье-образ не могут одновременно быть ограниченными в пространстве и что если мы ограничиваем размеры голограммы, то рискуем получить в плоскости P паразитные наложения. В этом случае невозможно на основании теоремы Шеннона ценой бесконечного числа выборок подавить паразитные повторы вне области задания функции $f(x, y)$.

Таким образом, остается сместить паразитные повторы $f(x, y)$ как можно дальше от области \mathcal{D} за счет возможно более плотной дискретизации функции F .

4. *Амплитуда функции F должна быть также квантованной.* Этим квантованием определяется точность, с которой предполагается воспроизводить значения функции F . Квантование должно учитывать источники шума, имеющиеся во всем процессе синтеза голограммы, а именно: при расчете, при физической реализации и при фотоумножении. Известно, что, как и в случае временных сигналов, грубое квантование менее чувствительно к воздействию шумов, однако оно требует более широкой полосы для передачи заданного объема информации. Предельный случай — бинарное (двухуровневое) квантование, преимущества которого будут рассмотрены чуть позднее.

Таковы в общих чертах основные ограничения синтеза цифровых голограмм. Нетрудно видеть, что здесь нет существенных отличий от классических ограничений, встречающихся при цифровой обработке временных сигналов, кроме следующих:

1. *Переход к двумерному случаю*, приводящий к очень быстрому росту машинного времени с увеличением размеров голограммы.

2. *Отсутствие принципа причинности*, который связан исключительно с односторонним течением времени и не имеет места в оптике. Пространственные частоты могут быть положительными или отрицательными в зависимости от выбора направлений осей координат. В действительном (не комплексном) изображении (при постоянном фазовом сдвиге) каждой положительной пространственной частоте соответствует равная ей отрицательная пространственная частота (это следует из эрмитовости преобразования Фурье).

Теперь мы можем приступить к рассмотрению различных типов цифровых голограмм, не претендуя на их исчерпывающий анализ (заинтересованный читатель может обратиться к работе [48]), но лишь с целью оценить основные тенденции и увязать их с аналогиями, имеющимися в области цифровой обработки временных сигналов.

2. Различные типы цифровых голограмм:

1. Голограммы с модуляцией пространственной частоты. Гаткие голограммы иногда неправильно называют «полутоновыми голограммами». Заметим все же, что за этими двумя названиями кроются разные свойства и они могут быть независимыми. В случае полутоновых голограмм требуется на основании «уравнения голографистов» (§ 2) записать на фотоэмульсии (или любом другом чисто амплитудном модуляторе) одновременно амплитуду и фазу волны F : к ней добавляют комплексно-сопряженную волну F^* . Чтобы разделить полученные таким образом пуски, пригоден метод Лейта и Упатниекса, который заключается в модуляции волн F и F^* одной пространственной частотой, так называемой несущей, т. е. в обеспечении наклона выходящих пучков к оптической оси.

Положим

$$G(u, v) = F(u, v) \sin(2\pi au).$$

Но распределение комплексных амплитуд $G(u, v) + G^*(u, v)$, если оно на самом деле действительное, непременно будет состоять из положительной и отрицательной частей и, следовательно, не может быть записано на фотоэмульсии, пропускание которой только положительно. Чтобы устранить это препятствие, нужно к $G + G^*$ добавить положительную функцию достаточной амплитуды. В противоположность аналоговым методам, при которых выбор такой функции ограничен, цифровые методы предоставляют широкие возможности. Хуанг [49] выделил три особенно интересных метода выбора такой функции:

а) классический голографический метод

$$\varphi_1 = 1 + |F|^2 + G + G^*;$$

б) добавление однородной постоянной

$$\varphi_2 = C + G + G^*,$$

где $C \geq 2|F|_{\max}$;

в) добавление усредненной функции

$$\varphi_3 = 2|F| + G + G^*.$$

На рис. 5.34, а показаны эти три типа цифровых голограмм, а на рис. 5.34, б приведено распределение интенсивности в плоскости изображения P .

Обозначим максимальную протяженность изображения, восстановленного с помощью F , через Δx . Эта величина характеризует ширину полосы пространственных частот, необходимую для получения F . Модуляция функции F с помощью несущей частоты a только сдвигает центр этой полосы частот в точку a . При добавлении комплексно-сопряженной волны G^* полоса

частот увеличивается до $2a + \Delta x$. Очевидно, что для надежного пространственного разделения изображений пространственную несущую следует выбирать из условия $a \geq \Delta x$.

1. В случае классической голографии пространственно-частотный спектр функции $|F|^2$ в 2 раза шире, чем функции F . При этом минимальная несущая частота, необходимая для разделения восстановленных изображений и пучка, распространяющегося вдоль оптической оси, будет определяться неравенством $a \geq 2\Delta x$, а ширина полосы — неравенством $\Delta x_1 \geq 4\Delta x$.

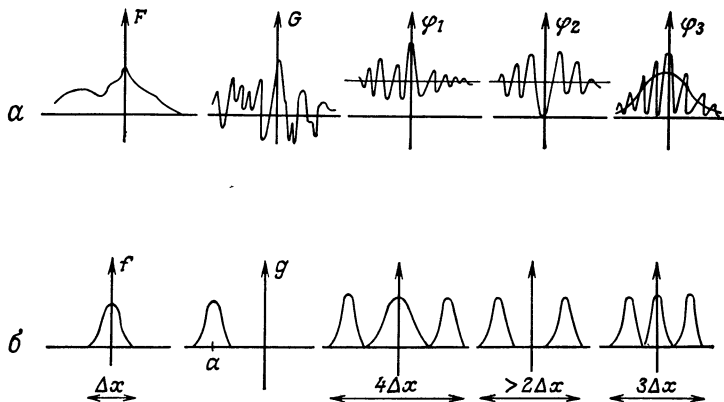


Рис. 5.34. Три типа цифровых голограмм.

a — функция F модулируется и получается функция G ; φ_1 , φ_2 и φ_3 — три способа обеспечить положительность голограммы при добавлении к G специальной функции; b — результат восстановления этих трех голограмм.

2. Метод добавления постоянного фона требует самой узкой полосы частот, поскольку на оптической оси формируется только функция Дирака. Следовательно, $\Delta x_2 \geq 2\Delta x$.

3. Метод добавления усредненной функции позволяет получить в каждой точке голограммы контраст, равный 1. Он обеспечивает наивысшую дифракционную эффективность голограммы. Требуемая в этом методе полоса частот $\Delta x_3 \geq 3\Delta x$ занимает промежуточное положение между двумя предыдущими.

Из изложенного следует, что при правильном выборе способа записи цифровой голограммы обеспечивается выигрыш по сравнению с традиционным голографическим методом либо в ширине полосы пропускания, либо в дифракционной эффективности. Но из табл. 5.1 видны и недостатки полутонного голографического процесса: число опорных точек, необходимых для получения голограммы изображения, состоящего из N точек, будет равно по крайней мере $2N$, а может быть и больше, так как при условии $u = \Delta x/2$, наложенном на функцию φ_2 , на самой границе изображения была бы светлая точка с очень боль-

шой энергией, которая, разумеется, исказила бы соседние точки изображения.

Таблица 5.1

Вид записи	F	G	$G + G^*$	$1 + F ^2 + G + G^*$	$C + G + G^*$	$2 F + G + G^*$
Ширина полос частот	Δx	Δx с центром в a	$2a + \Delta x$	$4\Delta x$	$2\Delta x$	$3\Delta x$

Далее мы увидим, что проблема чересчур плотной дискретизации, необходимой для получения голограммы, существует для всех типов цифровых голограмм и для других типов голограмм вопрос стоит даже более остро. Но самый трудный момент, встречающийся при практической реализации голограммы, — это динамический диапазон. Мы уже отмечали неоднородность распределения интенсивности в голограммах Фурье. Энергия сосредоточена главным образом на низких пространственных частотах, сгруппированных вокруг оптической оси. Уже в нескольких миллиметрах от оптической оси значения функции $F(u, v)$ составляют тысячные доли максимального значения $F(0, 0)$. Было предложено много способов для того, чтобы избежать такого плохого распределения энергии на голограмме и связанных с этим неприятных последствий: шумы, искажения, потеря разрешения и т. д.

1. Первый способ заключается в отказе от голографии Фурье и переходе в так называемую «зону Френеля» (§ 5, п. Б), для которой динамический диапазон значительно меньше. К сожалению, преобразование Френеля труднее осуществлять, и оно не позволяет использовать некоторые упрощающие расчет свойства, такие, как инвариантность по отношению к трансляции и симметрия.

2. Второй способ [47] основан на аналогии с классической голографией; в нем используется контролируемое насыщение наиболее интенсивных составляющих. Ценой небольших искажений можно обеспечить приемлемый динамический диапазон.

3. И наконец, третий способ основан на использовании рассеивателей, широко применяемых в аналоговой голографии. С помощью же ЭВМ определяют, какой рассеиватель $d(x, y)$, будучи помещен на объект $f(x, y)$, создаст наилучшее распределение $F(u, v)$ на голограмме. В работе [50] найдены наилучшие характеристики рассеивателя. Это чисто фазовый объект, удовлетворяющий условиям

$$|d(x, y)| = 1, \quad |D(u, v)| = 1.$$

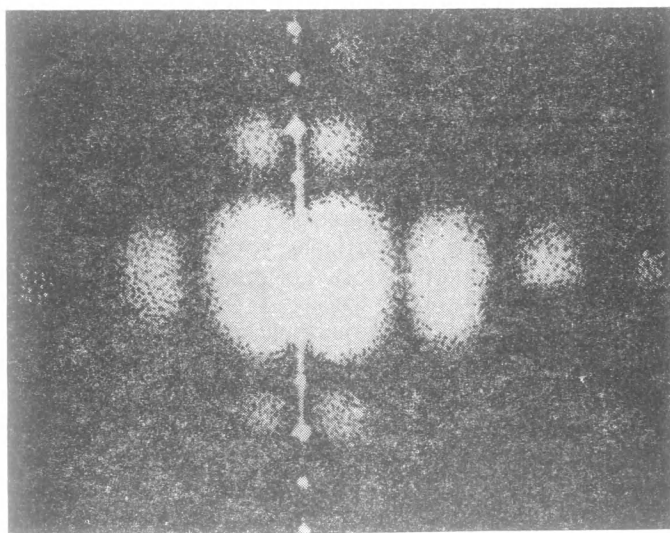
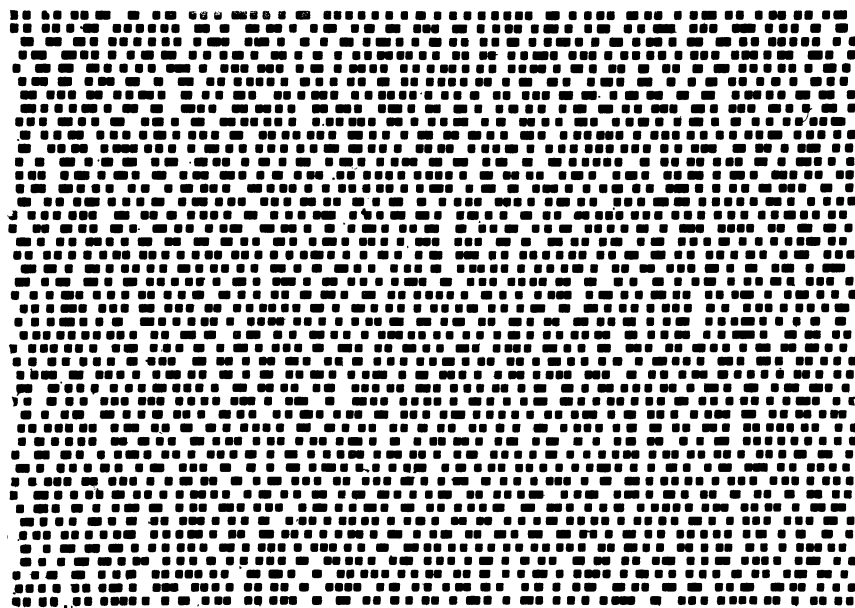


Рис. 5.35. Идеальный рассеиватель, синтезированный на ЭВМ.

Вверху — рассеиватель, который состоит из 64×64 элементов, распределенных (в одном направлении) по случайному закону. Внизу — картина дифракции на таком синтезированном по случайному закону рассеивателе. Состоит из многих дифракционных порядков в направлении осей u и v . Вокруг точки A (дифракционный порядок 1, 0) энергия дифрагировавшего света распределена практически равномерно.

Такой рассеиватель обеспечивает хорошую однородность амплитуды, большую избыточность голограмм (соответствие точка — плоскость между изображением и голограммой), а также отсутствие крупной зернистости (так называемого «макроспекла», гл. 4, § 4) (рис. 5.35).

Но среди самых важных факторов, затрудняющих синтез полутоновых голограмм, следует отметить необходимость в системе формирования изображения с изменяющейся амплитудой. Отсутствием таких систем частично объясняется, почему этот простой метод кодирования не применяется более широко.

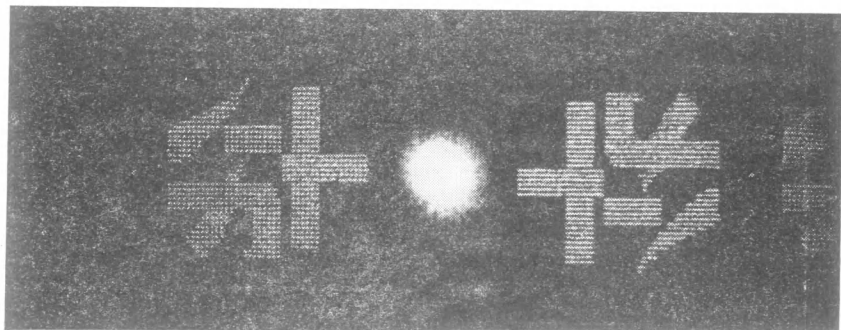


Рис. 5.36. Изображение, восстановленное с цифровой полутоновой голограммы, на котором видны два дифракционных порядка $+1$ и -1 , а также части паразитных порядков $+2$ и -2 .

На рис. 5.30 было показано увеличенное изображение цифровой голограммы, а на рис. 5.36 приведен результат восстановления другой голограммы, рассчитанной Р. Хауком.

2. Голограммы с фазовым смещением. Данный метод, предложенный Ломанном [51], открыл дорогу цифровой голографии. При таком методе амплитуда волны кодируется шириной импульса, а фаза — положением импульса. Таким образом, записанный сигнал оказывается бинарным. Всякая комплексная величина ζ представляется прозрачным окном в непрозрачном экране: площадь окна пропорциональна модулю величины ζ , а его смещение относительно некой правильной решетки пропорционально аргументу величины ζ .

Такой способ кодирования фазы называется фазовым смещением. Здесь фазовый сдвиг вводится аналогично сдвигу по фазе, обусловленному нарушением периодичности штрихов в дифракционной оптической решетке и вызывающему появление паразитных фигур дифракции, называемых «духами» (рис. 5.37).

На рис. 5.38 изображены две соседние ячейки, соответствующие двум комплексным величинам $F(k, l)$ и $F(k+1, l)$ в двух точках голограммы. Период правильной решетки ячеек равен

$\delta u = \delta v$, что соответствует увеличению на 1 индексов k и l . Ширины внутренних окон ячеек $\mu(k)$ и $\mu(k+1)$ пропорциональны величинам $|F(k, l)|$ и $|F(k+1, l)|$. Положением центров окон по отношению к центрам ячеек кодируются значения аргументов. Так, расстоянием $\delta'v$ между окнами определяется

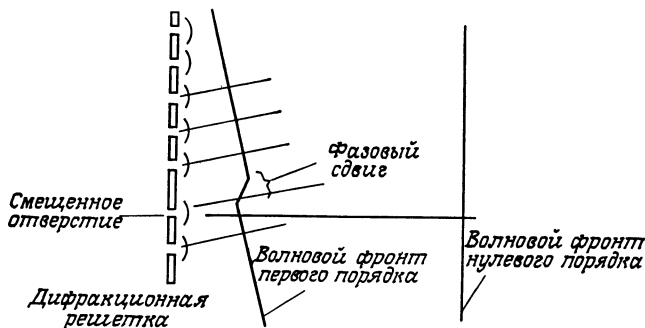


Рис. 5.37. Кодирование методом фазового смещения.

Смещение отверстия, нарушающее периодичность решетки, приводит к сдвигу фазы в волновом фронте, дифрагирующем в первый порядок. Для волны 2-го порядка дифракции сдвиг фазы будет вдвое больше

фазовый сдвиг между $F(k, l)$ и $F(k+1, l)$, равный $2\pi(\delta v - \delta'v)/\delta v$.

Как восстановить рассчитанную таким образом голограмму? Поскольку мы не прибегали ни к модуляции несущей частотой,

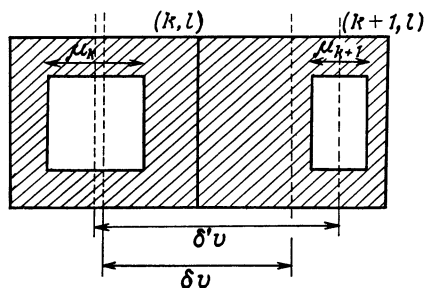


Рис. 5.38. Кодирование амплитуды и фазы по методу Ломанна.

ни к добавлению опорного или комплексно-сопряженного пучка, для кодирования изображения, состоящего из N точек, нам нужно только N ячеек. Это большой выигрыш по сравнению с полутоновой голографией, где требуется $2N$, $3N$ или $4N$ точек в зависимости от выбранного вида кодирования (табл. 5.1). Но этот выигрыш только кажущийся, так как разрешение бинарной голограммы определяется не расстояниями δu и δv между ячейками, а элементарным шагом ϵ , который должна будет обеспечить выходная система, рисующая каждое окно внутри ячейки. Что же это такое, элементарный шаг? Если желательно

получить точность по фазе, равную $2\pi/n_1$, при n_2 уровнях по амплитуде, то следует выбрать элементарный шаг равным $\varepsilon = \delta u/n$, где $n = \sup(n_1, n_2)$. Так, для обеспечения точности по фазе, равной $\pi/8$, при 20 уровнях амплитуды число элементарных точек на голограмме должно быть равно $nN = 20N$ а не $4N$, как в предельном случае для полутоновых голограмм. Следовательно, такой способ кодирования приводит к заметному расширению требуемой полосы частот. Этот факт хорошо известен и в теории связи: *кодирование с использованием широтно-импульсной модуляции (ШИМ) приводит к увеличению ширины полосы пропускания* (гл. 15, § 1). Тем не менее такой способ кодирования все больше применяется как в технике связи, так и в голографии, поскольку он малочувствителен к разного рода шумам: нерегулярностям системы формирования голограммы, неправильному освещению при фотоумножении, неоднородностям фотоподложек и т. д. Если добавить, что данный метод требует лишь самых обычных систем для формирования голограммы (это главным образом графопостроители), то станет ясно, что это действительно надежный и экономичный метод синтеза цифровых голограмм. У него в сущности два недостатка.

1. В теоретическом плане это прежде всего то, что, как показывает (в отличие от наших упрощенных рассуждений) строгий анализ, смещение δu на самом деле получается в результате итеративной процедуры, часто очень длительной [52].

2. В практическом отношении часто оказывается трудным обеспечить высокую точность кодирования (плотную дискретизацию) и большое число ячеек без высококачественного графопостроителя.

На рис. 5.39 приведено увеличенное изображение части голограммы рассмотренного типа, а на рис. 5.40 показан результат восстановления, но другой голограммы, предоставленный нам Х. Бартельтом.

3. Киноформы (фазовые голограммы). Мы рассмотрели два типа голограмм, в которых кодируется вся информация о комплексной функции F : ее амплитуда и ее фаза. Есть и другой подход к синтезу цифровых голограмм, ориентированный на кодирование одной фазовой информации. Он основывается на аналогии с дифракционными решетками и на очень интересных свойствах фазовых дифракционных решеток с пилообразным профилем (§ 3), обладающих оптимальной дифракционной эффективностью в первом порядке. Данное обстоятельство положено в основу нового типа цифровых голограмм, появившихся в 1969 г. [53]. В них совершенно не используется информация об амплитуде синтезируемой функции. Разумеется, для практического осуществления голограмм такого типа нужны фазово-модулирующие, а не амплитудные среды типа фотопленки. Правда, в

приложении 5 будет показано, что фазовую модуляцию можно осуществить и с помощью фотоэмульсий после их соответствующей фотохимической обработки. В то же время многие свето-

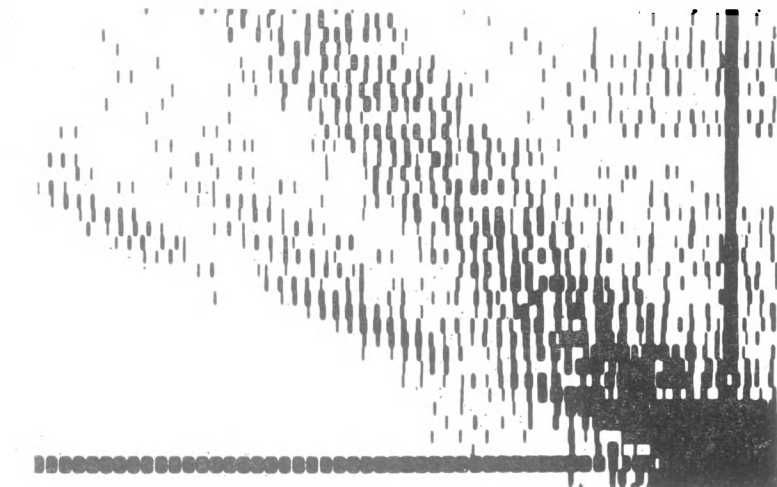


Рис. 5.39. Увеличенный участок голограммы, кодированной методом фазового смещения.



Рис. 5.40. Изображение, восстановленное с голограммы, кодированной методом фазового смещения.

чувствительные материалы относятся к фазовым модуляторам света, поскольку при экспонировании изменяется их толщина или показатель преломления.

Расчет киноформа особенно прост в связи с тем, что изменения фазы лежат в интервале от 0 до 2π . Подчеркнем, что такие голограммы сходны с линзами Френеля.

Чтобы еще больше упростить синтез киноформа, многие ис-

следователи занимались разработкой бинарных киноформов, в которых фаза принимает только два значения: 0 и π . Этот дополнительный шаг вполне оправдывается трудностью получения контролируемого фазового сдвига в пределах от 0 до 2π . Если говорить о линзах, то можно сказать, что он аналогичен переходу от линз Френеля к линзам Сорэ.

На рис. 5.31 было показано, каким образом с помощью этих двух киноформов осуществляется кодирование волны с параболическим законом изменения фазы.

Основные преимущества киноформа состоят в его высокой дифракционной эффективности и в простоте расчета. Отметим

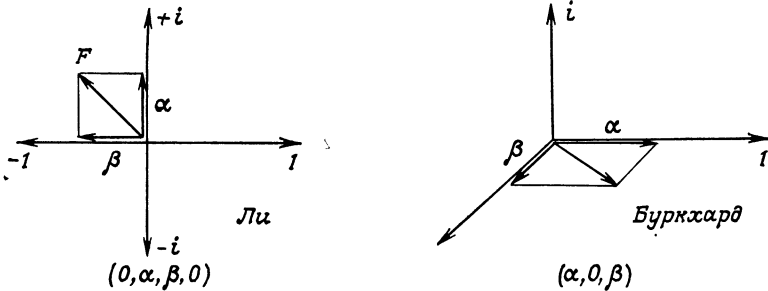


Рис. 5.41. Разложение комплексного числа F на 4 положительные составляющие (из которых 2 равны нулю) по методу Ли и на 3 положительные составляющие (из которых 1 равна нулю) по методу Буркхарда.

также, что здесь нет трудностей, связанных с динамическим диапазоном, которые очень значительны в случае амплитудных голограмм.

Недостатки же киноформа — это в основном искажения, которые вносятся из-за отказа от использования амплитудной информации, а также трудности осуществления непрерывной фазовой модуляции и, в частности, согласования отклонения фазы в пределах от 0 до 2π .

4. Голограммы Ли и Буркхарда. Подобно Ломанну, Ли [54] также предложил кодировать фазовую и амплитудную информацию на простом амплитудном модуляторе методом «фазового смещения». Но, вместо того чтобы представлять комплексную амплитуду $F(u, v)$ одной ячейкой, он раскладывает ее на четыре основных вектора $(1, 0)$, $(0, i)$, $(-1, 0)$, $(0, -i)$ и кодирует эти компоненты (рис. 5.41). Поскольку он считает все компоненты положительными, такое разложение оказывается единственным, а две компоненты обращаются в нуль. Буркхард [55] предложил сократить число основных векторов до 3, чтобы уменьшить число точек, требуемых для синтеза голограммы, с $4N$ до $3N$.

Преимущество такого метода синтеза цифровых голограмм заключается в основном в том, что теперь окна в непрозрачном экране центрируются относительно четырех (или трех) правильных решеток, так что положение каждого окна соответствует фазовым сдвигам основных векторов. Следовательно, теперь квантование фазы однозначно связано с квантованием амплитуды.

5. Осевая комплексная голограмма (ROACH)¹⁾. Этот метод синтеза цифровых голограмм предложен Гудменом и Чу [56] и основан на раздельном кодировании амплитуды и фазы волны с помощью трехслойной эмульсии, применяемой в цветных фотопленках. Один слой эмульсии служит для кодирования амплитуды, а два других — фазы.

Такая голограмма восстанавливает изображение объекта на оптической оси и потому не требует введения несущей частоты. Благодаря же раздельному кодированию амплитуды и фазы на разных эмульсиях при записи таких голограмм не требуется фотопленки с очень высоким разрешением.

6. Другие виды кодирования. В цифровой голографии были опробованы в большей или меньшей мере все способы кодирования, основанные на воздействии либо на фазу волны, либо на ее амплитуду. Отметим некоторые из них:

а) бинарные фазовые голограммы Сиверкена [57], использующие 8 базовых компонент для покрытия комплексной плоскости;

б) отбеленные голограммы Ломанна [51], в которых для кодирования амплитуды и фазы волны используется бинарная фазовая модуляция;

в) мультиплексные голограммы Далласа [58] или Ломанна, которые для одновременного кодирования двух волн используют две степени свободы, имеющиеся на плоскости;

г) «рассчитываемые в уме голограммы» Флере [59], используемые для кодирования голограмм бинарных объектов путем бинарной амплитудной модуляции;

д) голограммы «с последовательностью паритета» Чу и Гудмена [60], применяемые главным образом для дискретно распределенных в пространстве объектов, как, например, в оптической памяти.

§ 7. Применение голографии

Мы не ставим себе целью дать здесь полный обзор применений голографии, весьма многообразных и весьма разнородных. Но мы считаем целесообразным, прежде чем перейти к детальному изучению применения голографии для оптической обра-

¹⁾ Referenceless-On-Axis-Complex-Hologram.

ботки информации, подчеркнуть, что это не единственная область, где используется данный мощный метод, и что в других областях тоже много работают над ее развитием как в теоретическом, так и в техническом плане. Все области применения голографии можно объединить в следующие четыре основные группы: 1) обработка информации; 2) метрология и интерферометрия; 3) хранение информации (память); 4) отклонение световых пучков. Обработка информации объединяет самые плодотворные применения голографии в технике связи. Мы их подробно рассмотрим в гл. 6.

А. Метрология и интерферометрия [61, 62]

Применение голографии в метрологии и интерферометрии основывается на очень высокой чувствительности голографической схемы к микросмещениям ее элементов (§ 4, п. В). Как мы уже видели, если при записи голограммы некоторая точка смещается таким образом, что ее расстояние до фотопластинки изменяется на половину длины волны, то при восстановлении голограмма формирует два точечных источника, излучающие в противофазе, вследствие чего они скомпенсируют друг друга. В результате эта точка, хорошо освещенная при записи голограммы, окажется темной при восстановлении. Отсюда следуют три основных метода голографической интерферометрии.

1. Метод двух экспозиций

Во время первой экспозиции записывается голограмма объекта в исходном положении O_1 , однако фотопластинка не подвергается фотохимической обработке, а остается на месте. При второй экспозиции производится запись голограммы на ту же фотопластинку, с тем же опорным пучком, того же самого объекта, но в другом положении O_2 . После этого фотопластинку проявляют, фиксируют и затем восстанавливают обе голограммы одновременно. Все неподвижные точки объекта, а также и все те точки, расстояние которых до фотопластинки изменилось на целое число длин волн, дадут синфазный вклад в изображения O_1 и O_2 . Совершенно ясно, что эти точки будут яркими. Разумеется, этот метод применим для смещений любого вида, вращения, деформации и т. д.

2. Наблюдение в реальном времени

При таком методе записывается только голограмма объекта в положении O_1 . После фотохимической обработки она очень точно устанавливается на прежнее место, занимаемое ею при экспонировании. Если теперь объект смещается или деформи-

руется, занимая последовательные положения O_1, O_2, O_3 и т. д., то будет происходить интерференция между восстановленным с голограммы изображением объекта O_1 и его изображением O_i в момент времени t_i . В результате на поверхности объекта возникают интерференционные полосы, которые деформируются в реальном времени в зависимости от изменений объекта.

3. Метод усреднения во времени

Этот метод применяется главным образом для исследования периодических или случайных колебаний вокруг положения равновесия. Здесь тоже снимается одна голограмма, но время экспозиции намного больше периода колебаний объекта, так что в течение экспозиции объект свободно колеблется. Записывается очень сложный процесс, но в первом приближении можно считать, что все происходит так, как если бы за время экспозиции объект освещался только в крайних положениях. Следовательно, только эти крайние положения смещения объекта окажутся яркими при восстановлении, если они разделены разностью хода, кратной $\lambda/2$, и будут темными в противном случае. Обеспечиваемая голографической интерферометрией точность измерений обычно имеет порядок долей длины волны используемого излучения (0,1 мкм). Вместе с тем в особых случаях она может быть улучшена в 10—100 раз за счет использования различных схем. Такая точность соизмерима с точностью, обеспечиваемой традиционными методами интерферометрии, которые тоже не требуют контакта с объектом (в отличие от разного рода механических измерителей и маркерных устройств), но голографическая интерферометрия имеет ряд важных преимуществ. Отметим следующие из них: возможность наблюдения интерференции объектов в разные моменты времени или в разных местах; нечувствительность к разного рода дефектам оптики, используемой в голографическом интерферометре; возможность расчета совершенной опорной волны методами цифровой голографии; возможность применения любых диффузно отражающих поверхностей и т. д.

Любопытно, что актуальной задачей на сегодня является уменьшение чувствительности голографической интерферометрии, чтобы расширить область ее применения до более важного на практике диапазона деформаций. С подобной тенденцией мы встретимся и при рассмотрении оптических методов обработки информации (гл. 6, § 2, п. В), поскольку там высокая чувствительность голографических методов тоже иногда создает определенные трудности.

Основные применения голографической интерферометрии лежат в настоящее время в метрологии: измерение вибраций, микросмещений, состояний поверхности, определение режимов те-

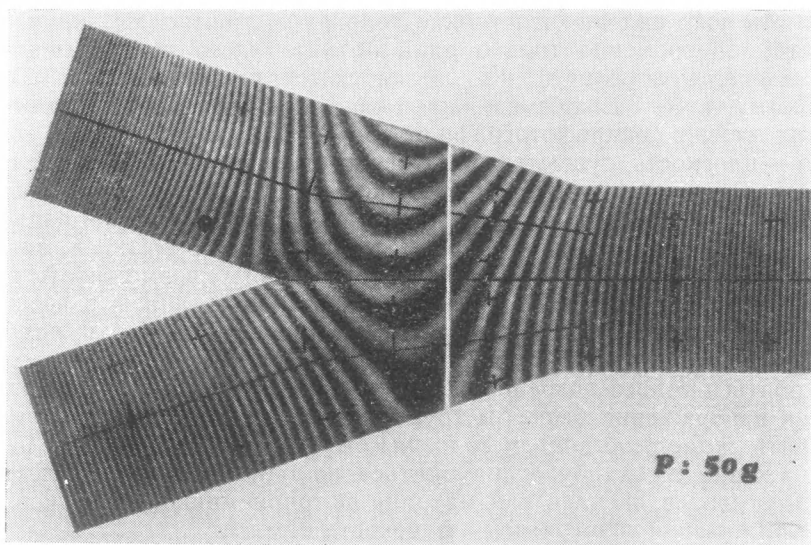
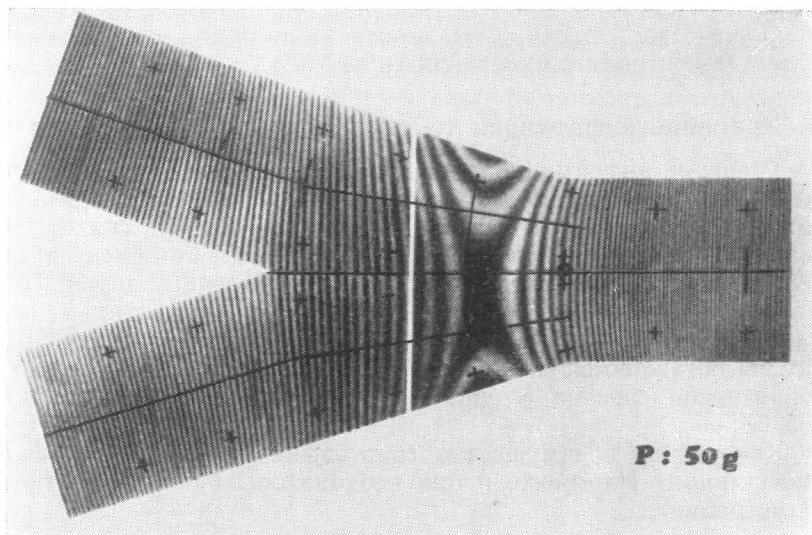


Рис. 5.42. Голографическая интерферометрия.
Интерферограмма модели участка шоссейной дороги под нагрузкой.

чений жидкостей и газов, загрязнений окружающей среды, контроль качества оптики. Она может найти применение во всех областях контроля и измерения (рис. 5.42).

Б. Хранение информации

Хранение информации принимает различные формы и находит самое разное применение. Мы не будем приводить примеры, которые не относятся непосредственно к технике связи, как, скажем, регистрация двумерных или трехмерных объектов для целей рекламы, сохранения или в художественных целях [63]. Не будем также касаться более отдаленных перспективных применений, таких, как голографические кино и телевидение [64]. Но мы остановимся на *голографической памяти* — области, которая тесно связана с обработкой и передачей информации [65, 66].

Каковы же преимущества голографии в области хранения информации? Их много, и они неодинаковы при разных типах регистрации.

1. Тонкие голограммы

Мы уже видели, что тонкие голограммы позволяют записывать одновременно только один объект (здесь нет объемного мультимплексирования). Какой же смысл регистрировать голограмму, а не сам объект, как в обычных архивах, использующих сейчас микрофотографию? Во-первых, соответствие точка — плоскость, существующее между точкой объекта и всей голограммой, гарантирует избыточность информации, благодаря чему при повреждении отдельных точек голографической записи не пропадает полностью часть содержимого памяти, а лишь слегка изменяется все содержимое. Во-вторых, возможность получать голограммы без линз или использовать линзы в весьма благоприятных для них условиях позволяет достигать *разрешения, близкого к теоретическому дифракционному пределу*. Добиться такого разрешения обычными средствами формирования изображения было бы трудно. И, в-третьих, можно так выбрать *геометрию записи голограммы*, что восстановленное изображение всегда будет находиться на одном и том же месте, например на дисплее или матрице фотоприемников, независимо от положения голограммы в плоскости памяти.

Какую же плотность записи информации можно получить таким методом? Расстояние между двумя соседними битами информации в плоскости объекта равно $\delta_x = 1,22\lambda\phi/d$, где ϕ — диаметр голограммы (будем считать ее круглой), а d — расстояние от нее до объекта (рис. 5.43). Следовательно, плотность записи информации равна $1/\delta_x^2$ и при обычных значениях параметров составляет 10^6 бит/мм². Это на несколько порядков

больше плотности записи, достигнутой в системах магнитной памяти с лентами, дисками или ферритами. Если добавить к этому и высокую скорость доступа, обеспечиваемую оптическими методами, то станет понятным тот интерес, который проявляется к голографической памяти. Правда, предельные значе-

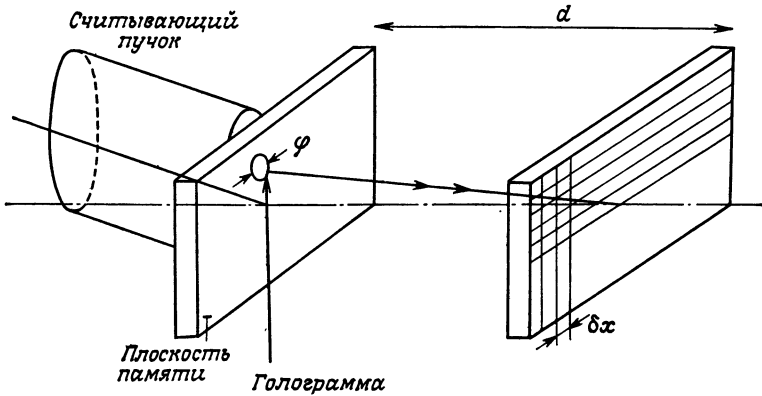


Рис. 5.43. Схема голографической памяти.

Плоская память представляет собой мозаику голограмм диаметром ϕ .

ния параметров еще не достигнуты. Но уже достигнута информационная пропускная способность порядка 100 Мбит/с [67].

2. Толстые голограммы

Толстые фотоэмульсии позволяют мультиплексировать голограммы за счет трехмерности среды. В § 5, п. В была рассчитана угловая избирательность голограммы, записанной на слое эмульсии толщиной e . Используя две степени свободы в ориентации считывающего пучка (рис. 5.44), можно на одном носителе записать от 20 до 200 голограмм, наложенных друг на друга. Такая высокая плотность записи голограмм сразу ставит много проблем: выравнивание дифракционных эффективностей голограмм; уменьшение их перекрестных искажений и повышение энергетической эффективности каждой голограммы. Сюда добавляются и технические трудности [68]: выбор оптики, обеспечение требуемых характеристик дефлекторов света (гл. 14, § 2) и фотоприемников (матриц или интегральных многодетекторных систем). Поэтому еще предстоит много работы, прежде чем будет достигнута плотность записи информации порядка 10^8 бит/мм² [69]. Разработка новых светочувствительных материалов (таких, например, как LiNbO_3 , $\text{Bi}_{12}\text{SiO}_{20}$, $\text{Bi}_{12}\text{GeO}_{20}$, PLZT и др.) взамен фотоэмульсии позволяет надеяться на создание оперативной голографической памяти. Используемые в такой памяти носители информации должны одновременно быть

чувствительными, прозрачными, обладать высоким разрешением и допускать многократную запись и стирание информации. Все эти свойства обычно не очень-то совместимы.

В настоящее время уже существуют архивные системы памяти большой емкости с быстрым доступом к записанной информации, работающие на принципе мультиплексирования голограмм (системы PHEDRE и SEMIRAMIS Национального

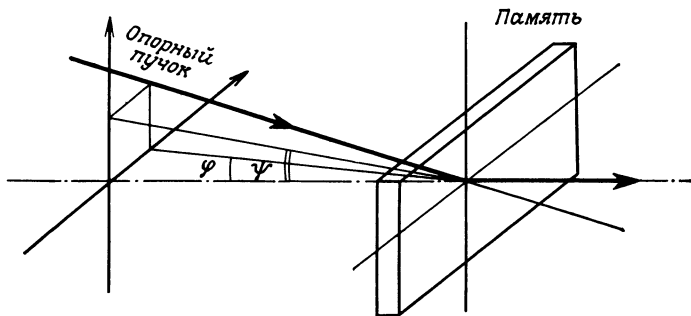


Рис. 5.44. Используя толстые регистрирующие среды и опорный пучок переменного наклона (углы ψ и ϕ), можно увеличить емкость голографической памяти.

центра исследований по технике связи [70]). Эти системы хранят и выдают потребителям главным образом аналоговую информацию, которая выводится на экран электронно-лучевой трубки. Что же касается цифровой информации, то здесь самые широкие потенциальные возможности применения — у систем голографической памяти большой емкости для ЭВМ.

В. Отклонение световых пучков

Отклоняемый лазерный луч может применяться для быстрого доступа к информации в оптической памяти, для записи на фотопленке или фоточувствительной бумаге фототерминала, для построения графиков и знаков на экране. Развертку можно осуществлять методами электрооптического отклонения световых пучков (гл. 14, § 2), но это очень дорого. Механические же методы сканирования (посредством автоматически управляемых зеркал) не обеспечивают нужного быстродействия. Но можно представить себе дефлекторы на основе рассчитанных голограмм, перемещаемых поперек пучка. Такие дефлекторы были бы не очень дороги и в то же время легко видоизменяемы. Суть метода состоит в наложении друг на друга голографических решеток, но, применяя для отклонения световых пучков решетки специальной геометрии, синтезируемые на ЭВМ, можно было бы также уменьшать aberrации и корректировать расстояние фокусировки в зависимости от точки адресации [71].

ГЛАВА 6

ОБРАБОТКА ИНФОРМАЦИИ МЕТОДАМИ КОГЕРЕНТНОЙ ОПТИКИ

В технике связи оптические методы предполагается использовать для решения задач передачи информации и ее обработки. Эта и следующие главы посвящены оптической обработке информации с использованием как когерентного, так и некогерентного света. Вопросы передачи информации, главным образом с помощью волоконных световодов, будут рассмотрены в последующих главах.

В настоящее время операции обработки сигналов все чаще и чаще осуществляются непосредственно в канале передачи. Большой объем передаваемой информации, необходимость ее передачи с большой скоростью и представления в заданном виде требуют доведения до уровня обычной, повседневной техники таких необходимых операций, как кодирование и декодирование, восстановление формы сигналов, уменьшение избыточности или их коррекция. В то же время из года в год все более важной в потоке передаваемых сообщений становится информация, переносимая в виде изображения: телевидение вчера, фототелеграфия и дистанционное зондирование сегодня, видеотелефон, телетекст и телекино завтра — все это действует в том же направлении и требует новых методов обработки информации, более быстрых и производительных.

Разумеется, электронные методы (аналоговые или цифровые) хорошо зарекомендовали себя при обработке сигналов всех видов. Это единственные методы, способные решать такие сложные задачи, как разного рода нелинейная фильтрация сигналов, анализ изображений и т. п. Однако электронные методы оказываются плохо приспособленными для обработки изображений, буквенно-цифровых текстов и графиков, которые по своей природе являются двумерными. При последовательном способе обработки изображений традиционными электронными методами, по-видимому, разрушается та квазиизотропная взаимозависимость между любой точкой изображения и окружающими ее соседними точками, которая составляет саму структуру изображения и которая учитывает его «костяк» и семантическое содержание. Известно, какие усилия затрачиваются на создание математического обеспечения для апостериорного восстановления этой взаимозависимости. Известно, какие новые системы обработки информации (специализированные процес-

соры, параллельные (матричные) вычислительные устройства, поточные процессоры) создаются для того, чтобы приспособить электронное оборудование для обработки изображений. И порой кажется, что эта техника зашла в область, которая ей не принадлежит, что эти усилия и что даже сами эти успехи — плоды лишь привычек общества, мысль которого на протяжении 40 лет работала в основном над развитием этой техники и которое уделяло ей столько внимания, как ни одной другой области научной деятельности.

Но если заново рассматривать проблему обработки изображений, то как не вернуться в первую очередь к оптике, которая всегда имела дело с информацией в виде изображений? Как не попытаться сохранить такое представление информации, столь удобное для использования человеком, который часто выступает в качестве последнего звена канала связи? Мы постараемся показать, что в ряде случаев оптика оказывается значительно лучше электронно-вычислительной техники при решении задач обработки изображений, что во многих случаях она дает другую, дополняющую информацию, а также что оптика во многих отношениях еще плохо освоенная область, в которой больше вопросов, нежели ответов. Именно эта сложность и придает сегодня волнующую привлекательность оптическим методам — сочетание больших возможностей, полной адекватности своему объекту, невразумительной техники и часто сомнительной теории.

Нужно не упускать из виду, что оптика — не волшебная палочка для обработки сигналов и никогда не будет ею. Оптика — это совокупность методов, очень подходящих для решения определенных задач, и такие методы естественным образом найдут свое применение наряду с другими методами. Для этого необходимо, чтобы не ослабевали усилия, которые за несколько десятилетий привели к созданию нового научного направления.

§ 1. Принципы обработки информации в когерентном свете

Пространственная фильтрация в когерентном свете основывается на том, что линза выполняет преобразование Фурье двумерного оптического сигнала. В гл. 4, § 1, п. Г, где речь шла о дифракции света, было показано, что распределение комплексных амплитуд $F(u, v)$ в задней фокальной плоскости P_2 линзы с точностью до фазового множителя равно двумерному фурье-образу распределения комплексных амплитуд $f(x, y)$ на объекте (рис. 6.1). Квадратичный фазовый множитель можно устранить, если транспарант с комплексным амплитудным профилем $f(x, y)$ поместить в переднюю фокальную плоскость

P_1 линзы (рис. 6.2); тогда

$$F(u, v) = \iint f(x, y) e^{-\frac{2\pi i}{\lambda f}(ux+vy)} dx dy.$$

По аналогии с терминологией, используемой в электронике, переменные $u/\lambda f$ и $v/\lambda f$ ассоциируемые с x и y , называются

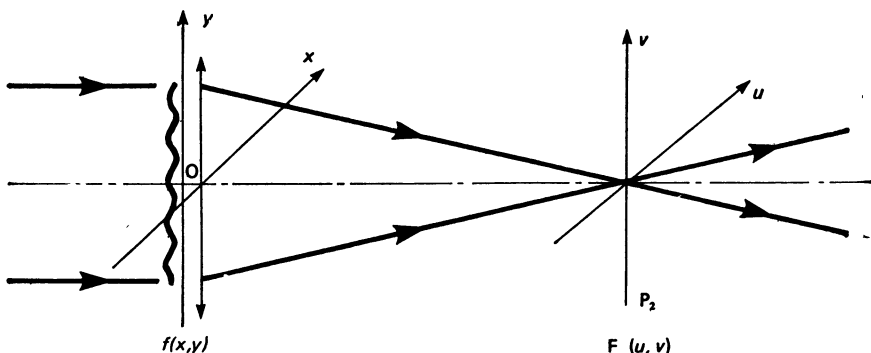


Рис. 6.1. Линза переносит картину дифракции, возникающую на бесконечности, в свою фокальную плоскость.

Распределения комплексных амплитуд $F(u, v)$ и $f(x, y)$ связаны между собой преобразованием Фурье (с точностью до квадратичного фазового множителя).

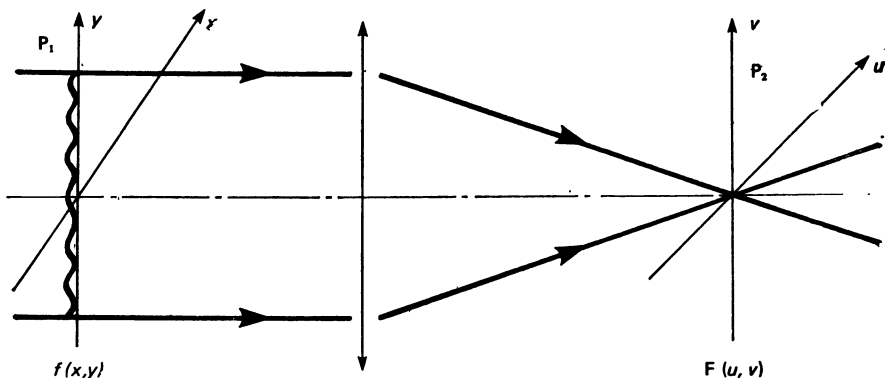


Рис. 6.2. Соотношение между $F(u, v)$ и $f(x, y)$ будет точным преобразованием Фурье, если транспарант с амплитудным пропусканием $f(x, y)$ поместить в переднюю фокальную плоскость линзы.

пространственными частотами (они выражаются в единицах м^{-1}). В гл. 4, § 2 мы уже имели возможность познакомиться с этими величинами.

Таким образом, поместив в плоскость P_2 транспарант с амплитудным пропусканием $H(u, v)$, можно осуществлять *филь-*

трацию пространственных частот с передаточной функцией $H(u, v)$:

$$G(u, v) = F(u, v) H(u, v).$$

Восстановление отфильтрованного изображения, наблюдаемого в плоскости P_3 (рис. 6.3), осуществляется очень просто с помощью одной дополнительной линзы (схема $4f$).

Рассмотренная система пространственной фильтрации называется также *системой двойной дифракции* [72]. В самом деле, первая дифракция на апертуре первой линзы эквивалентна преобразованию Фурье распределения комплексных амплитуд на

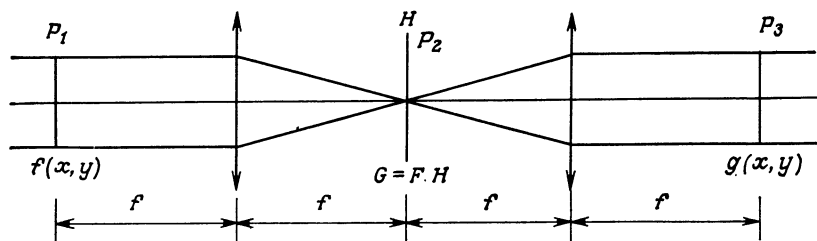


Рис. 6.3. Схема пространственной фильтрации вида $4f$. Это система с двойной дифракцией. Для нее нужны 2 линзы.

объекте. Затем свет проходит плоскость P_2 и, претерпевая вторую дифракцию, достигает плоскости изображения. При этом вторая дифракция может быть осуществлена с помощью второй линзы или просто в результате распространения света в свободном пространстве между задней фокальной плоскостью первой линзы и плоскостью изображения. В результате выполняется двойное преобразование Фурье

$$\mathcal{F}[\mathcal{F}[f(x)]] = f(-x).$$

Об условиях, при которых возможна такая операция, будет сказано несколько слов в приложении 6.

Очевидно, что в отсутствие пространственного фильтра H , в соответствии с геометрической оптикой в плоскости P_3 будет воспроизводиться функция $f(x, y)$ (если не считать, что у линз бывают аберрации). При наличии же в плоскости P_2 пространственного фильтра H на выходе системы пространственной фильтрации формируется функция

$$g(x, y) = f * h.$$

Комплексная функция $h(x, y)$, являющаяся фурье-образом передаточной функции $H(u, v)$, называется *импульсным откликом* пространственного фильтра. Она описывает изображение, фор-

мируемое в плоскости P_3 , когда в плоскости P_1 на оптической оси помещен точечный источник:

если $f(x, y) = \delta(x, y)$, то $g(x, y) = \delta * h = h(x, y)$.

Рассмотренная операция пространственной фильтрации оказывается линейной и инвариантной относительно смещения входного изображения в плоскости P_1 . Кратко остановимся на вопросе об инвариантности по отношению к смещению. Она выражается в том, что в какую бы точку X, Y ни сместился входной импульс, т. е. каков бы ни был импульс $\delta(x - X, y - Y)$, импульсный отклик системы пространственной фильтрации остается тем же самым, но смещенным, т. е. равным $h(x - X, y - Y)$.

Теперь мы видим основную особенность оптического метода обработки информации в когерентном свете, которая имеет две дуальные формы (в плоскости изображения P_3 и в плоскости фурье-образа P_2):

- а) можно мгновенно¹⁾ осуществлять двумерные свертки;
- б) можно визуализировать фурье-образ некоторого распределения комплексных амплитуд и мгновенно видоизменять его.

Таковы важнейшие следствия инвариантности по отношению к смещению (пространственной инвариантности) оптической системы с двойной дифракцией, которая позволяет проводить параллельную (одновременную) обработку большого числа данных, например точек изображения.

Прежде чем говорить о том, как использовать все многообразие операций пространственной фильтрации в когерентном свете, сделаем ряд общих замечаний.

1. Входные переменные системы пространственной фильтрации, так же как и переменные в плоскости фильтра и в выходной плоскости, являются комплексными величинами (амплитуда и фаза волны на входе и выходе, ослабление и фазовый сдвиг, вносимые пространственным фильтром). Теоретически для этих комплексных переменных нет никаких ограничений. Далее мы увидим, что в этом главное отличие пространственной фильтрации в когерентном свете от пространственной фильтрации в некогерентном свете. В частности, благодаря отмеченному отсутствию ограничений оказывается возможным:

а) обрабатывать всю падающую на вход системы волну, используя при этом всю информацию, содержащуюся в ней;

б) подвергать ее любой фильтрации с помощью пространственных фильтров с положительными, отрицательными или комплексными модуляционными характеристиками, которые можно практически осуществить;

в) получать на выходе любые функции, которые требуется.

¹⁾ Со скоростью света, что не так уж и плохо!

Такая свобода (теоретически) позволяет реализовать *любые линейные и гомогенные* операции пространственной фильтрации, которые предлагают математики. Практически же эта свобода будет ограничена только умением экспериментатора представлять математические величины с помощью имеющихся материалов (ограниченная ширина полосы пропускания, шумы, динамический диапазон и т. д.).

2. В схеме пространственной фильтрации вида $4f$ объект (транспарант) и фильтр симметричны относительно линзы. Для оптика комплексная амплитуда в передней фокальной плоскости линзы ничем существенно не отличается от комплексной амплитуды в ее задней фокальной плоскости. Поэтому для него фурье-образ функции такая же реальность, как и сама функция. С фурье-образом можно обращаться как с обычным изображением: увеличивать с помощью оптической системы, фотографировать и т. д. В частности, он может служить входным сигналом в оптической системе пространственной фильтрации.

3. Строго говоря, даже в отсутствие пространственного фильтра $H(u, v)$ изображение на выходе в плоскости P_3 не будет идентично объекту в плоскости P_1 , поскольку у линз ограниченная числовая апертура и они действуют как фильтры низких частот.

4. Необходимо контролировать фазовую и амплитудную характеристики пространственного фильтра H . Его амплитудная характеристика непосредственно определяется изменением прозрачности используемого материала (фотопленки, фоторезиста и т. п.). Фазовая характеристика фильтра зависит от многих параметров: показателя преломления, толщины, а также от оптической активности материала фильтра. Фазовые характеристики фильтра сложно контролировать, и обычно их труднее реализовать, чем амплитудные.

5. Пространственная переменная u в плоскости фурье-образа связана с *пространственной частотой* η соотношением $\eta = u/\lambda f$. Следовательно, пространственный фильтр только тогда определен полностью, когда известен коэффициент λf , характеризующий систему оптической фильтрации. На практике это приводит к оптическим фильтрам весьма малых размеров для оптических систем приемлемых габаритов. Так, например, при использовании линз с фокусными расстояниями 100 мм и транспарантов с размерами 24×36 мм на обычной фотопленке общая площадь фильтра не превышает 10 мм^2 .

Итак, мы познакомились с основными принципами оптики, на которых может быть основана обработка информации. Но, прежде чем переходить к применениям оптических методов обработки информации, мы напомним читателю единственную общую теорию, которая известна нам в этой области.

§ 2. Теория: метаксиальная пространственно-временная оптика [73, 74]

Сформулируем некоторые из сделанных ранее выводов в несколько более общем виде:

а) большинство интересующих нас оптических систем являются системами *линейными и пространственно-гомогенными* (§ 1);

б) эти же самые оптические системы можно рассматривать и как *временные фильтры* (гл. 4, § 4);

в) характерные параметры когерентности подчиняются определенным законам *теории дифракции* (гл. 4, § 4).

Все преобразования оптического сигнала от входа до выхода такой линейной пространственно-временной системы можно характеризовать *импульсным откликом* $h(x, y; \tau)$ или *передаточной функцией* $H(u, v; f)$. Боннэ [74] принадлежит заслуга обобщения всех этих положений в единой теории, основывающейся только на теории дифракции. Эта теория, названная «метаксиальной оптикой», состоит в следующем.

1. Она *обобщает параксиальную оптику*, которой мы пользовались ранее, на случай, когда вход и выход оптической системы рассматриваются на касательных сферах. Таким образом, в ней идет речь о приближении более высокого порядка, нежели в параксиальной теории (мы, например, уже подчеркивали в гл. 4, § 1, что фурье-образ получается на *сфере* бесконечного радиуса).

2. Она количественно описывает процесс передачи информации между входной и выходной сферами на основе теории дифракции; при этом рассматриваются *три* преобразования:

а) *преобразование Фурье*,

б) *пространственная фильтрация*,

в) *модуляция фазы*.

Мы имеем как бы обобщенное преобразование Френеля [15], которое в одномерном случае можно записать в следующем виде:

$$g(x') = \underbrace{e^{2\pi i C x'^2}}_{\text{Модуляция фазы}} \int e^{-2\pi i u x'} H(u) du \underbrace{\int f(x) e^{-2\pi i u x} dv}_{\text{Преобразование Фурье}}$$

Пространственная фильтрация

3. Она одинаково подходит к передаче комплексной амплитуды в когерентном свете и к передаче в условиях «*частичной когерентности*» при частично когерентном (пространственно-некогерентном) освещении.

4. Она рассматривает также передачу тех же величин и в случае *полихроматического* (некогерентного во времени) освещения.

В пунктах 3 и 4 принцип интерференции (гл. 4, § 4, п. Б) объединяется с теорией случайных процессов, лежащей в основе обработки информации при частично когерентном освещении (гл. 7).

5. Она приложима к общим вопросам пространственно-временной обработки информации оптическими методами.

За отсутствием места мы не можем более подробно остановиться на этой теории. Наша цель лишь в том, чтобы сообщить читателю о существовании общей теории, которая распространяется на всю оптическую обработку информации [73, 74].

Посмотрим теперь, каким образом можно использовать основные особенности когерентной оптической фильтрации.

§ 3. Применение оптической обработки информации в когерентном свете

На практике известны три вида оптической обработки информации (рассматриваемые ниже), которые различаются главным образом используемыми методами. Первый из них нашел весьма многочисленные применения, но лишь немногие из них имеют прямое отношение к технике связи. Поэтому они будут рассмотрены весьма кратко. Что же касается двух других, то на них мы остановимся подробнее.

А. Фильтрация пространственных частот

В рассматриваемом методе фильтрации пространственных частот чаще всего используется бинарный пространственный фильтр. Обычно это просто маска, помещаемая в частотной плоскости системы фильтрации. С помощью такого фильтра можно реализовать следующие виды пространственной фильтрации:

а) *низкочастотная* фильтрация, позволяющая ослабить шум, устранить растр на изображении, если оно растрено, или осуществить операцию сглаживания;

б) *высокочастотная* фильтрация, обеспечивающая повышение контраста изображения, улучшение его разрешения или выделение контуров на нем;

в) *полосовая* фильтрация, применяемая для выявления скрытых периодических структур на изображении или структур заданного размера; полосовой фильтр позволяет, например, сжать информацию, содержащуюся в странице печатного текста, сохраняя только характерные частоты, интересующие цензуру.

г) *направленная* (веерная) фильтрация, которая обеспечивает выделение линий с требуемой пространственной ориентацией;

д) мультипликация изображений в виде строки или матрицы с регулярным шагом при помощи одномерных и двумерных решеток.

Указанные виды пространственной фильтрации находят применение в самых различных областях: для анализа сейсмограмм при поиске нефти; для обработки фотографий, получаемых с помощью пузырьковых камер в ядерной физике; для определения участков, занимаемых интересующими нас сельскохозяйственными культурами, или зонами определенной растительности при исследовании природных ресурсов по аэрофотоснимкам и космическим снимкам; для идентификации тканей или особых волкнистых образований в биологии.

Во всех упомянутых случаях применения пространственной фильтрации очень часто она выполняет роль предварительной обработки изображений, которая позволяет производить преобразование и выбор информации, поступающей затем либо на другие процессоры для последующей обработки, либо непосредственно операторам для использования. Методы пространственной фильтрации просты в реализации, надежны, мощны и достаточно легко изменяемы в соответствии с разными задачами. Но результаты обработки оказываются не очень точными, и ими лучше пользоваться для приближенных оценок, нежели для измерений, поскольку они чувствительны к воздействию шумов, особенно когда необходимо сохранять высокие пространственные частоты. Их очень легко использовать в канале автоматической обработки информации, так как для любой заданной операции фильтрации параметры пространственного фильтра не зависят от характера обрабатываемого изображения.

Б. Восстановление искаженных изображений

Задачу восстановления искаженных изображений часто также называют задачей обращения свертки. В общем случае линейных систем она состоит в решении интегрального уравнения

$$g(x, y) = \iint f(x', y') h(x', y'; x, y) dx' dy',$$

где функция f описывает неискаженное изображение, g — искаженное изображение, а ядро h характеризует искажение изображения. В более простом случае, когда функция h однородна (пространственно-инвариантна), данное уравнение принимает вид

$$g(x, y) = \iint f(x', y') h(x - x', y - y') dx' dy' = f * h. \quad (6.1)$$

Именно эту упрощенную форму мы и будем рассматривать, поскольку первое уравнение еще не имеет общего решения в оптике.

Для полного восстановления изображения методом пространственной фильтрации нужно взять пространственный фильтр с модуляционной характеристикой вида

$$K(u, v) = \frac{1}{H(u, v)}. \quad (6.2)$$

Тогда распределение комплексных амплитуд в плоскости P_2 будет равно

$$GK = F, \quad (6.3)$$

т. е. фурье-образу восстановленного изображения.

Пространственный фильтр с модуляционной характеристикой вида (6.2) называется *инверсным фильтром*.

Для многих задач инверсный фильтр оказывается неосуществимым. В самом деле, передаточная функция H корректируемой системы обычно очень сильно убывает с ростом пространственной частоты (искажение снижает разрешение в изображении). Таким образом, существует полоса частот, вне которой величина H становится произвольно малой, а следовательно, величина K — столь угодно большой. Эта полоса простирается, в частности, далее максимальной пространственной частоты, содержащейся в восстанавливаемом изображении g (как всякое физическое изображение, g имеет ограниченные размеры). Но может случиться, что и в полосе частот изображения g функция H обращается в нуль в некоторых точках. Во всех таких точках (тех, которые лежат вне полосы, и тех внутри, где H равно нулю) мы теряем всю информацию, содержащуюся в F и соответствующую этим пространственным частотам. Поскольку же функция H линейная и однородная, эту информацию мы не найдем на других частотах. *Информация, содержащаяся в этих точках, полностью теряется.*

Невозможность восстановления такой информации совершенно очевидна и из выражения (6.2). На тех частотах, где H обращается в нуль, величина K должна быть бесконечно большой. Это невозможно осуществить практически, да и нежелательно, так как в тех точках, где сигнал равен нулю, наверняка имеется шум, который и будет тогда усилен.

Хотя инверсный пространственный фильтр и повышает разрешение в восстановленном изображении, отношение сигнала к шуму при этом оказывается очень низким. Это следует иметь в виду в случае изображений с высоким уровнем шумов. Правда, имеются другие фильтры без подобных отрицательных качеств. Но, прежде чем рассматривать их подробно, вернемся к вопросу усиления сигнала с помощью пространственного

фильтра. Известно, что оптические пространственные фильтры пассивны: их обычно изготавливают на фотопленке, фоточувствительных резистах, термопластиках и других материалах. Их пропускание не превышает 1, а значит, они не могут усиливать свет.

Поскольку, согласно формуле (6.3), величина K должна быть больше единицы, чтобы вернуть ее к интервалу $[0, 1]$, следует прибегнуть к нормализации. Это означает, что для сохранения динамического диапазона полутонного изображения приходится идти на значительное ослабление его яркости.

Среди фильтров, предложенных для улучшения характеристик инверсного фильтра, наиболее известен фильтр Винера [75]. Он учитывает наличие в обрабатываемом изображении случайного, аддитивного, не коррелированного с сигналом шума. В этом случае уравнение свертки (6.1) принимает вид

$$g = f * h + b. \quad (6.4)$$

Ищут такой фильтр, который сводил бы к минимуму среднеквадратичное отклонение ε отфильтрованного изображения \hat{g} от исходного изображения f ,

$$\varepsilon = \left\langle \iint_{-\infty}^{+\infty} |\hat{g} - f|^2 dx dy \right\rangle, \quad (6.5)$$

причем усреднение проводится по всем возможным реализациям шума.

Передаточная функция фильтра Винера следующим образом выражается через спектральные плотности мощности шума φ_b и сигнала φ_s :

$$K'(u, v) = \frac{H^*(u, v)}{|H(u, v)|^2 + \varphi_b(u, v)/\varphi_s(u, v)}. \quad (6.6)$$

Легко видеть, что при $\varphi_b \rightarrow 0$ фильтр Винера превращается в инверсный фильтр. Тем не менее величина K' всегда конечна, так как отношение φ_b/φ_s всегда строго положительно. На высоких пространственных частотах величина K' обращается в нуль так же, как и в полюсах инверсного фильтра. На рис. 6.4 показана передаточная функция инверсного фильтра, а также винеровского фильтра при двух значениях отношения φ_b/φ_s . (Когда нет никакой дополнительной информации о спектральных плотностях мощности сигнала и шума, для удобства можно считать их отношение постоянным и равным отношению сигнала к шуму в обрабатываемом изображении.) Выбранное здесь искажение изображения представляет собой смазанность, обусловленную равномерным движением фотокамеры во время экспозиции.

Инверсный фильтр повышает разрешение восстановленного изображения, а фильтр Винера — отношение сигнала к шуму.

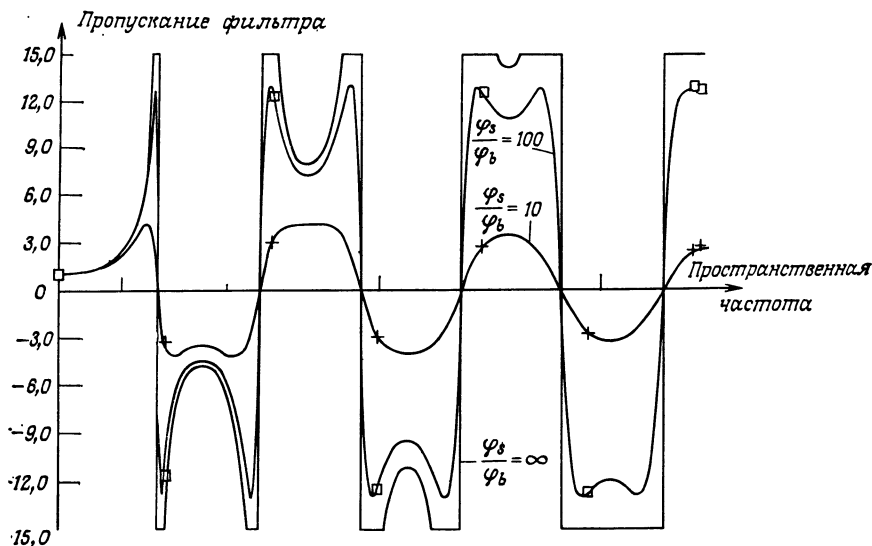


Рис. 6.4. Характеристики амплитудного пропускания инверсного и винеровского фильтров для коррекции изображений, искаженных смазыванием за счет линейного смещения с постоянной скоростью объекта или фотокамеры во время съемки (одномерный случай).

Импульсный отклик корректируемой системы формирования изображения равен $h(x) = \text{rect}(x/X)$, а следовательно, ее передаточная функция имеет вид $H(u) = \sin(2\pi u X/\lambda)/(2\pi u/\lambda)$. Передаточная функция инверсного фильтра (сплошная кривая) имеет регулярно повторяющиеся разрывы и полюсы. Фильтр Винера (здесь приведены его характеристики при двух значениях отношения сигнала к шуму $\varphi_s/\varphi_b = 10$ и $\varphi_s/\varphi_b = 100$) лучше как в отношении разрывов, так и в отношении асимптотических значений при очень больших u (пропускание фильтра, превышающее 15, приведено к значению 15).

Эти два вида коррекции изображений трудно совместимы. Дело в том, что разрешение изображения тесно связано с поведением изображающей системы на высоких частотах, но именно на этих частотах наибольший уровень шума.

Создание фильтра Винера явилось важным шагом в решении задачи восстановления сигналов. Однако лежащие в его основе статистические гипотезы не могут быть полностью применены к изображениям, поскольку:

1) шум фотоэмульсий мультипликативный, а не аддитивный;

2) он вовсе не независим от сигнала;

3) критерий минимума среднеквадратичного отклонения (6.4) не достаточен для оценки качества восстановления; его следовало бы дополнить критерием оценки субъективного качества изображения, имеющим отношение к наблюдателю. Однако такая оценка очень зависит от семантического содержания изображения и психофизиологии зрения.

В самом деле, глаз реагирует неодинаково на все детали изображения, он останавливается лишь на определенных, интересных для наблюдателя его фрагментах, для которых он и требует очень высокого качества воспроизведения. Причем эти фрагменты не отличаются от других участков изображения ни с точки зрения семантической, ни с точки зрения восприятия. Обычно это определенные прерывистые линии (контуры), определенные углы и т. д. Субъективный критерий качества изображения должен был бы выделить эти центры фиксации глаза и контуры вместо того, чтобы все усреднять. Однако задача отыскания такого критерия столь сложна, что в настоящее время к ней еще только робко приступают.

Указанные выше три момента очень сложно ввести в расчеты фильтра, обеспечивающего наилучшее восстановление изображения. Вот почему многие другие предложенные фильтры только частично удовлетворяют требованию высококачественного восстановления. К ним относятся:

1) *геометрический фильтр*, промежуточный между инверсным и винеровским фильтрами, обеспечивает некий компромисс между требованиями высокого разрешения и низкого шума [76];

2) *фильтр Филлипса* [77] минимизирует выражение (6.4), содержащее шум, а также и интеграл второй производной (описывает более гладкий профиль);

3) *фильтр Туми* [78] учитывает, кроме того, априорную информацию о существовании искомого решения в виде восстановленного изображения.

Известны также и другие восстанавливающие фильтры [76, 79]. Они обычно предназначаются для обработки изображений с низким уровнем шума, и их передаточная характеристика достаточно близка к характеристике винеровского фильтра. Ими можно пользоваться только тогда, когда заранее имеются определенные сведения об обрабатываемом изображении.

Какие же искажения могут быть скорректированы с помощью таких фильтров? Все линейные и гомогенные (пространственно-инвариантные) искажения. Приведем примеры таких искажений, которые чаще всего встречаются:

а) смазанность, обусловленная движением объекта во время съемки (поступательным перемещением, вращением или вибрацией);

б) нерезкость, обусловленная дефокусировкой (при условии, что объект плоский);

в) искажения, специфические для систем регистрации (например, электронных микроскопов, телевизионных передающих камер, ультразвуковых локаторов и т. д.);

г) нерезкость, обусловленная турбулентностью атмосферы (если времена экспозиции велики по сравнению с постоянной времени турбулентности);

д) искажения, специфические для отдельных приложений (например, рентгеновской томографии, радиолокационных станций бокового обзора и т. п.). Выше мы уже отмечали, что задача восстановления изображений является сложной и трудной. Это объясняется разными причинами.

1. В теоретическом плане это связано с тем, что обычно плохо известны величины, определяющие основные параметры восстанавливающего фильтра, а именно статистические характеристики исходного изображения, спектральная плотность и спектральное распределение шума, точные характеристики внешних в изображение искажений. Данное обстоятельство обычно заставляет экспериментаторов использовать предварительные оценки обрабатываемого изображения и определять наилучшие параметры восстанавливающего фильтра методом проб и ошибок.

2. В практическом плане оказывается достаточно сложной реализация комплексного пространственного фильтра с точными характеристиками (§ 4). Главные трудности связаны с реализацией резких изменений амплитудного пропускания фильтра и с обеспечением высокой точности его фазовой характеристики.

3. Кроме того, задача восстановления изображения основывается на понятии качества изображения, понятии, которое, как мы уже видели, трудно выразить математически.

Но есть еще проблема, связанная с самой природой системы пространственной фильтрации. Известно, что для света с высокой степенью когерентности характерна специфическая зернистость (так называемые спеклы, гл. 4, § 4, п. Е), обусловленная микроинтерференцией вдоль всего оптического пути; из-за этого когерентное освещение оказывается очень чувствительным к разного рода дефектам системы фильтрации: пылинкам, турбулентности, неоднородностям оптических элементов и т. д. В силу этих двух причин качество изображений, получаемых при когерентном освещении, уступает качеству изображений, формируемых при классическом некогерентном освещении, и это ухудшение качества особенно заметно, когда в фокальной плоскости линзы помещают пространственный фильтр. Таким образом, стараясь уменьшить искажения в восстанавливаемом изображении, оператор, сам того не желая, ухудшает общее качество обрабатываемого изображения. Следовательно, возможности восстановления изображений довольно ограничены и нужно согласиться, что на самом деле когерентная оптика еще не доказала практической эффективности своих методов в данной области.

В. Оптические корреляторы

Теперь мы познакомимся с другим способом использования способности оптической системы осуществлять преобразование

Фурье оптического сигнала. Известно, что произведение фурье-образов

$$G(u, v) = F(u, v) H(u, v)$$

соответствует свертке исходных функций в плоскости изображения P_3 (рис. 6.3) [формула (6.1)]:

$$g(x, y) = f * h = \iint f(x', y') h(x - x', y - y') dx' dy'.$$

На самом деле нам нужно найти функцию корреляции между f и s :

$$g(x, y) = f \otimes s = \iint f(x', y') s^*(x' - x, y' - y) dx' dy'. \quad (6.7)$$

Сравнив (6.6) и (6.7), мы видим, что для этого достаточно поместить в частотной плоскости P_2 пространственный фильтр с импульсным откликом $h(x, y)$ вида

$$h(x, y) = s^*(-x, -y). \quad (6.8)$$

Для определения передаточной функции фильтра воспользуемся свойствами эрмитовости фурье-образов

$$H(u, v) = S^*(-u, -v). \quad (6.9)$$

Переход $(u, v) \rightarrow (-u, -v)$ осуществляется простым поворотом пространственного фильтра в частотной плоскости на 180° ; что же касается комплексного сопряжения, то, хотя оно и кажется сложным для комплексных передаточных функций, позднее мы увидим, что эта операция реализуется очень легко.

По аналогии с переходом от инверсного фильтра к фильтру Винера мы можем повысить чувствительность обнаружения, если используем *согласованный фильтр* вида

$$H(u, v) = \frac{S^*(-u, -v)}{\Phi_b(u, v)}, \quad (6.10)$$

где Φ_b — спектральная плотность мощности шума. Позднее для простоты мы предположим, что шум имеет однородный спектр. Но такое допущение неверно в задачах распознавания образов, где все не интересующие нас образы можно рассматривать как шум.

Рассмотрим этот случай подробнее на примере. Пусть мы хотим идентифицировать надпись $s(x, y)$ среди других надписей $f_i(x, y)$ в атласе. Мы полагаем, что функция корреляции надписи s с каждой надписью f_i может служить мерой их сходства. Нам нужно последовательно помещать в плоскости объекта P_1 транспаранты со всеми надписями f_i , а в частотной плоскости поместить пространственный фильтр $H(u, v)$ с передаточной функцией, удовлетворяющей соотношению (6.9) или (6.10).

Тогда в плоскости P_3 мы тотчас получим нужную нам корреляцию $h_i = f_i \otimes s$. Вопрос в том, как осуществить фильтр с требуемой передаточной функцией. Обычно этот фильтр должен быть комплексным, и потому традиционные фотографические методы не позволяют его создать. Изящное решение задачи предлагает голография Фурье интересующей нас надписи s . На голограмме будут зарегистрированы следующие три слагаемых:

$$(|R|^2 + |S|^2) + R^*S(u, v) \exp\left(i \frac{2\pi u \theta}{\lambda}\right) + RS^*(u, v) \exp\left(-i \frac{2\pi u \theta}{\lambda}\right).$$

Третье слагаемое этого выражения является паразитным в голографии, но именно оно служит основой для нашей свертки. Если мы повернем установленную в частотной плоскости P_2

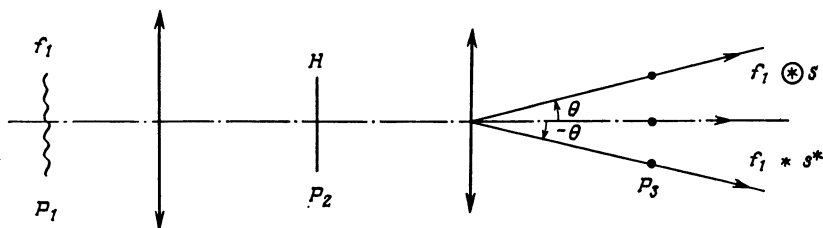


Рис. 6.5. Схема, используемая обычно в корреляторах, работающих в когерентном свете.

Если в частотную плоскость схемы пространственной фильтрации поместить фильтр-голограмму, то получим два дифракционных порядка $+1$ и -1 , причем в первом будет формироваться свертка, а во втором — функция корреляции входного изображения с импульсным откликом голографического фильтра.

пластинку с записью голограммы на 180° вокруг оптической оси, то получим нужный нам пространственный фильтр:

$$H'(u, v) = (|R|^2 + |S|^2) + R^*S(-u, -v) \exp\left(-i \frac{2\pi}{\lambda} u \theta\right) + RS^*(-u, -v) \exp\left(i \frac{2\pi}{\lambda} u \theta\right).$$

Такой фильтр даст три отфильтрованных изображения (рис. 6.5), разделенные углом θ , равным углу между сигнальным и опорным пучками при записи голограммы, а именно: первое — на оптической оси, описываемое не интересным для нас квадратичным членом; второе — под углом $-\theta$ к оптической оси, представляющее собой свертку f_i с s^* , т. е. $f_i * s$; третье — под углом θ к оптической оси, представляющее собой искомую корреляцию $f_i \otimes s$. Когда одна из надписей f_i на входном транспаранте будет s , в направлении угла θ будет максимальна энергия дифрагированного пучка.

Если входной сигнал f_i сместить в плоскости P_1 на некоторую величину, то на столько же сместится и отклик $f_i \otimes s$ в вы-

ходной плоскости P_3 , поскольку операция пространственной фильтрации является гомогенной, т. е. пространственно-инвариантной. Поэтому можно поместить в плоскость P_1 одновременно все надписи f_i , располагая их в точках (x_i, y_i) , и одновременно наблюдать в выходной плоскости сразу все корреляции. Тем самым обеспечивается параллельное распознавание сразу всех надписей, что очень повышает производительность обработки информации. На этом принципе созданы системы

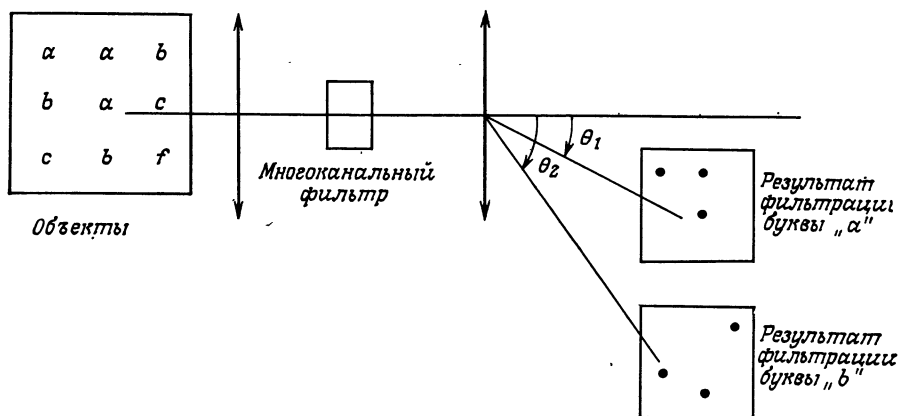


Рис. 6.6. Распознавание букв с помощью мультиплексного фильтра.

Фильтр, согласованный с буквой a , был изготовлен с использованием опорной волны, наклоненной под углом θ_1 , а при получении фильтра, согласованного с буквой b , был выбран угол наклона θ_2 . Операции распознавания букв a и b выполняются одновременно, но корреляционные пятна возникают в разных местах выходной плоскости. Для упрощения схемы мы здесь опустили симметричную часть выходной плоскости, где отображается поле сверток (гл. 6, § 1).

автоматического чтения. В частотной плоскости помещают голограмму-фильтр с записью буквы или слова, которое требуется распознать, а во входной плоскости устанавливают транспарант с записью текста, который нужно прочесть. Тогда в плоскости наблюдения в направлении под углом θ к оптической оси во всех местах расположения распознаваемой буквы появятся яркие точки.

Достаточно легко можно также реализовать параллельное распознавание большого числа образов s_j . Для этого на одной фотопластинке запишем столько наложенных голограмм при разных углах θ_j , сколько образов требуется распознать одновременно. (Разумеется, следует позаботиться о том, чтобы разные изображения не перекрывались.) Тогда в плоскости P_3 будет одновременно отображаться ансамбль объектов f_i , характеризующих корреляцию входного сигнала с каждым из образов s_j . При этом вокруг каждого направления θ_j будут отмечены положения s_j (рис. 6.6). Рассмотренный голографический метод

получения согласованного фильтра был предложен Вандер Люгтом [34] и им же распространен на мультиплексные (многоканальные) фильтры.

Рассмотрим подробнее основные свойства оптических корреляторов, использующих когерентное освещение. Такие корреляторы позволяют обнаруживать сигнал и определять его положение (координаты). Следовательно, коррелятор *не чувствителен к смещению сигнала* во входной плоскости. Но он очень чувствителен к повороту сигнала, его гомотетии, а также к любым изменениям формы сигнала.

Изящество решения задачи обнаружения сигнала и является главным достоинством когерентных корреляторов, которое позволяет распознавать сложные образы, различающиеся только деталями: отпечатки пальцев, рентгеновские снимки и т. д. Но у этого метода имеется недостаток — требуется совершенно точно знать распознаваемый образ. Так, например, при автоматическом чтении фильтры обычно согласованы только с одним шрифтом.

Есть еще один, более серьезный недостаток, связанный с самим принципом обнаружения, состоящим в отыскании максимумов функции корреляции. Вернемся к примеру обнаружения сигнала s среди образов f_i и предположим, что один из образов f_k , отличный от s , тем не менее содержит в себе сигнал s (например, буква «Е» содержит в себе букву «Г»). В этом случае становится ясным, что максимумы функции автокорреляции $s \otimes s$ и взаимной корреляции $s \otimes f_k$ будут идентичными. Мы оказываемся теперь в ситуации, когда шум и сигнал не являются больше взаимно независимыми. При этом отклик коррелятора на шум может даже превзойти отклик на сигнал.

Эlegantное решение этой проблемы предложили Ловенталь и Бельво [80]. Оно состоит в том, что на выходе коррелятора формируется функция автокорреляции не самого сигнала, а его производной. Дело в том, что специфическая информация, присутствующая данному образу, содержится в контурах образов, а не в однородных зонах. При этом передаточная функция требуемого фильтра теперь равна не $S(u, v)$, а $(u^2 + v^2)S(u, v)$. Это приводит к заметному повышению отношения сигнала к шуму на выходе¹⁾.

Другой подход к решению данной проблемы основан на применении фильтра FOUT (Filtre Optimisé Unique de Transcodage [81] — оптимальный фильтр с перекодированием). Принцип ра-

¹⁾ Здесь под отношением сигнала к шуму понимается отношение интенсивностей пика автокорреляции сигнала к пику взаимной корреляции с фоном. Если входное изображение искажено шумом, то операция предварительного «оконтуривания» (дифференцирования) может оказаться неэффективной в отношении повышения надежности обнаружения из-за возросшего влияния входного шума. — *Прим. ред.*

боты такого фильтра состоит в следующем. Найдем фильтр с таким импульсным откликом $p(x, y)$, что его отклик на любой сигнал $f_k(x, y)$ будет состоять из n бинарных точек. Известно, что если ансамбль подлежащих распознаванию образов состоит из K образов ($1 \leq k \leq K$), то для распознавания всех образов будет достаточно N бинарных откликов, где N — целая часть величины $(\log_2 K + 1)$, причем эти отклики образуют кодовое слово $b(k, n)$ ($n = 1, 2, \dots, N$), которое ассоциируется с каждым образом.

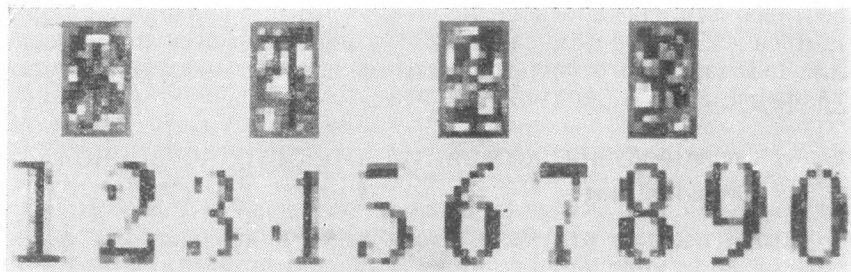


Рис. 6.7. Оптимальный фильтр с перекодированием, рассчитанный на распознавание 10 цифр.

Здесь 4 маски представлены отдельно. При оптическом выполнении операции распознавания они будут мультиплексированы.

Рассмотрим фильтр для линейной совокупности образов вида

$$p(x, y) = \sum_{k, n} a(k, n) f_k(-x + nx_0, -y), \quad (6.11)$$

где $a(k, n)$ — подлежащие определению неизвестные, а x_0 — смещение, соответствующее расстоянию, которое желательно ввести между двойными элементами b . Отклик такого фильтра на образ f_{k_0} записывается в виде

$$f_{k_0} * p = \sum_{k, n} a(k, n) f_k(x, y) \oplus f_{k_0}(x - nx_0, y). \quad (6.12)$$

Рассмотрим точку ($x = n_1 x_0, y = 0$) и введем корреляционную матрицу образов $C_{k_0 k}(0, 0)$. Тогда неизвестные $a(k, n)$ будут решениями системы NK уравнений с NK неизвестными

$$\begin{aligned} \sum_k a(k, n) C_{k_0 k}(0, 0) &= b(k_0, n), \\ k_0 &= 1, 2 \dots K, \\ n &= 1, 2 \dots N. \end{aligned} \quad (6.13)$$

Обычно такая система имеет решение и притом единственное даже в случае очень схожих друг с другом образов. При этом фильтр оказывается единственным, так как он мультиплексный.

Кроме того, он оптимальный в том смысле, что обеспечивает на выходе только бинарные отклики [т. е. $b(k, n) = 0$ или 1] и, следовательно, менее чувствителен к воздействию шумов. И наконец, он осуществляет перекодирование, поскольку любому непрерывному сигналу $f_k(x, y)$ на входе он ставит в соответствие дискретизованный и квантованный отклик на выходе $b(k, n)$.

Рассмотренный фильтр можно осуществить оптическими методами, поскольку он представляет собой линейную совокупность образов f_k [формула (6.11)]. Но для вычисления неизвестных $a(k, n)$ на основе уравнения (6.13) приходится обращаться к ЭВМ, и тем самым часто оправдывается применение для синтеза такого фильтра методов и средств цифровой голографии (рис. 6.7).

§ 4. Ограничения когерентной пространственной фильтрации

Выше мы уже отмечали, что когерентная оптическая фильтрация очень хорошо подходит для обработки изображений. Это объясняется следующими ее достоинствами:

- а) высоким быстродействием;
- б) двумерностью и параллельностью обработки информации;
- в) относительной универсальностью, поскольку (теоретически) могут быть реализованы любые линейные пространственно-инвариантные операции фильтрации над комплексными сигналами.

Однако достоинства метода пространственной фильтрации следует сопоставить с теми ограничениями, которыми в настоящее время тормозится его развитие. Эти ограничения часто связаны с развитием и состоянием современной технологии и потому могут казаться конъюнктурными, а не принципиальными. Тем не менее не следует их недооценивать, поскольку они весьма реальны. В предыдущих параграфах мы уже говорили о некоторых из слабых сторон когерентной оптической обработки. Подведем итог сказанному.

1. Требование когерентности источника света в пространстве и во времени *не позволяет обрабатывать самосветящиеся объекты*, такие, например, как изображения на телевизионном экране, или даже просто отражающие свет объекты, как, например, газеты и рисунки, или же трехмерные объекты. Мы можем обрабатывать только прозрачные изображения этих объектов, стало быть, главным образом их фотографии (точнее диапозитивы или негативы). Тем самым исключаются любые применения, требующие осуществлять обработку изображений в реальном времени и не допускающие включения этапа фотографирования, пока еще нет устройств для мгновенного пре-

образования, например, телевизионного изображения в транспарант, пригодный для работы в системе оптической фильтрации. Такие устройства постепенно создаются в оптических лабораториях. Их называют «оптическими реле»¹⁾. Для модуляции света в них используются либо жидкие кристаллы, либо материалы, обладающие электрооптическими свойствами (приложение 5).

2. Изготовление пространственных фильтров ставит ряд технологических проблем, к которым мы вернемся, когда речь пойдет о пространственной фильтрации в некогерентном свете (гл. 7). Эти проблемы связаны с *выбором материала для фильтра*. (Еще раз отсылаем читателя к приложению 5, где он найдет обзор основных материалов, используемых в качестве носителя информации при изготовлении фильтров.) Носители, которые модулируют только амплитуду светового луча, вынуждают прибегать к кодированию фазы либо голографическим методом Вандер Люгта (§ 3, п. В), либо методом фазового смещения (гл. 5, § 6, п. В). Пространственные фильтры, способные модулировать только фазу, сами по себе обычно не обеспечивают фильтрации, и поэтому их часто дополняют амплитудным фильтром.

3. Мы уже указывали на относительную универсальность метода пространственной фильтрации в когерентном свете, и у него действительно более широкие возможности, чем у метода пространственной фильтрации в некогерентном свете или у гибридных методов. Тем не менее пока что получила развитие только *линейная и пространственно-инвариантная фильтрация*. Хотя и было выполнено много работ по реализации как пространственно-неинвариантной, так и нелинейной фильтрации методами когерентной оптики, эти работы дают лишь частные решения, не поддающиеся обобщению на основе единого математического описания. Отметим среди них некоторые наиболее интересные:

- а) выполнение геометрических преобразований [82];
- б) осуществление некоторых видов нелинейной фильтрации [83];
- в) получение изофот [84];
- г) преобразование интенсивности по логарифмическому или любому другому закону [85];
- д) аналого-цифровое преобразование [86];
- е) пространственно-неинвариантное распознавание образов [87].

4. И наконец, остается проблема *качества изображения*, являющаяся одним из самых фундаментальных ограничений оп-

¹⁾ Речь идет о пространственно-временных модуляторах света, или управляемых транспарантах. — *Прим. ред.*

тических методов обработки информации в когерентном свете. Как известно из повседневного опыта экспериментаторам и как показано теоретически [88], когерентные оптические системы пространственной фильтрации более подвержены воздействию разного рода шумов, чем некогерентные. Шум вводится в оптическую систему либо на уровне объекта (например, шум фотоэмульсии, обусловленный зернистостью ее структуры), либо при прохождении когерентного оптического сигнала через линзы и фильтры. Снижение качества выходного изображения проявляется в снижении контраста, а также в уменьшении отношения сигнала к шуму. Поэтому для рассматриваемых методов более предпочтительны такие области применения, где качество изображения не является самым важным критерием. Это, например, распознавание образов, вычисления в цифровой форме и логические операции.

ГЛАВА 7

ОБРАБОТКА ИНФОРМАЦИИ МЕТОДАМИ НЕКОГЕРЕНТНОЙ ОПТИКИ

В предыдущей главе мы отметили недостатки и ограничения метода когерентной пространственной фильтрации и, в частности, подчеркнули, что эти ограничения в значительной мере связаны с требованием когерентности используемого света. Они состоят главным образом в высокой чувствительности когерентных оптических систем к разного рода шумам, а также в часто возникающей необходимости смены носителя, на котором была записана подлежащая обработке информация.

Поэтому теперь совершенно естественно вернуться к некогерентной оптике, чтобы попытаться осуществить более удовлетворительным образом двумерную параллельную пространственную фильтрацию двумерных массивов информации в реальном времени, которая показалась нам столь привлекательной стороной оптических методов. Но мы увидим, что на некогерентную оптику нельзя прямо переносить методы, используемые в когерентной оптике, и поэтому необходимо дать соответствующие теоретические основы.

§ 1. Теория некогерентной пространственной фильтрации

А. Теория

Прежде всего уточним, что мы понимаем под некогерентным источником света. Напомним, что для когерентной пространственной фильтрации требуется источник света, излучение которого когерентно в пространстве и во времени (гл. 4, § 4), т. е. источник должен иметь точечные размеры и быть абсолютно монохроматическим. Сначала мы сохраним монохроматичность источника, но увеличим его размеры, т. е. будем рассматривать протяженный источник, состоящий из множества излучателей со случайными начальными фазами. Затем рассмотрим полихроматический источник света как точечный, так и протяженный.

Излагаемая ниже теория некогерентной фильтрации была создана Дюффье [89] 30 лет назад.

Рассмотрим линзу, которая в плоскости P формирует изображение точечного источника. Распределение интенсивности в изображении точечного источника равно квадрату модуля

распределения амплитуды в картине дифракции на апертуре линзы $h(x, y)$. Оно называется картиной дифракции по интенсивности и равно

$$k(x, y) = |h(x, y)|^2. \quad (7.1)$$

Возьмем теперь пространственно-некогерентный объект, распределение интенсивности на котором обозначим через $o(x, y)$. (Разумеется, что в случае некогерентного объекта комплексная

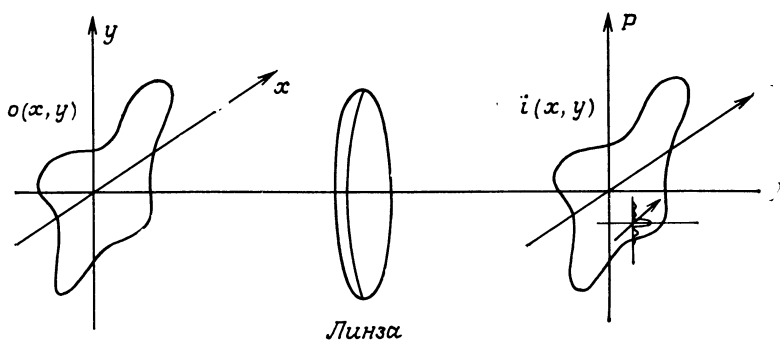


Рис. 7.1. Формирование изображения в некогерентном свете.

Каждая точка объекта дает вклад в изображение, соответствующий импульсному отклику изображающей системы для некогерентного света. Интенсивности всех отдельных таких вкладов складываются на выходе.

амплитуда не имеет смысла и его излучение можно характеризовать только интенсивностью.) Рассмотрим точечный источник $M(x_1, y_1)$ (рис. 7.1). Его картина дифракции по интенсивности имеет вид

$$k(x - x_1, y - y_1).$$

Эту величину следует умножить на интенсивность объекта в точке (x_1, y_1) :

$$o(x_1, y_1) k(x - x_1, y - y_1).$$

Поскольку излучение от всех точек объекта некогерентно, интенсивности всех картин дифракции складываются и полная интенсивность в плоскости P равна интегралу

$$i(x, y) = \iint o(x_1, y_1) k(x - x_1, y - y_1) dx_1 dy_1,$$

который представляет собой свертку распределения интенсивности на объекте с импульсным откликом оптической системы по интенсивности и символически записывается в виде

$$i = o * k. \quad (7.2)$$

На этом соотношении и основана обработка информации методами некогерентной оптики. Как нетрудно видеть, оно имеет тот же вид, что и при описании когерентной пространственной фильтрации (гл. 6, § 1). Но в него входят теперь другие величины:

а) при некогерентном освещении условиям линейной пространственно-инвариантной фильтрации удовлетворяет интенсивность освещения, тогда как при когерентном освещении эти условия выполняются для комплексных амплитуд;

б) импульсный отклик некогерентной системы пространственной фильтрации равен квадрату модуля импульсного отклика когерентной системы пространственной фильтрации [формула (7.1.)].

Выполнив преобразование Фурье левой и правой части равенства (7.1), можно переписать его в следующем виде:

$$K(u, v) = H(u, v) * H^*(-u, -v), \quad (7.3)$$

или, в символической форме,

$$K(u, v) = H(u, v) \oplus H(u, v).$$

Отсюда следует важный вывод: передаточная функция некогерентной системы пространственной фильтрации равна функции автокорреляции передаточной функции когерентной системы пространственной фильтрации.

Важнейшие характеристики оптических систем пространственной фильтрации представлены в табл. 7.1.

Таблица 7.1

	Когерентная пространственная фильтрация	Некогерентная пространственная фильтрация
Входная величина	f — комплексная амплитуда на объекте	$o = f ^2$ — интенсивность на объекте
Выходная величина	g — комплексная амплитуда в изображении	$i = g ^2$ — интенсивность в изображении
Передаточная функция	H — функция комплексного пропускания зрачка	$K = H \oplus H$ — функция автокорреляции зрачка
Импульсный отклик	h — картина дифракции на зрачке по амплитуде	$k = h ^2$ — картина дифракции на зрачке по интенсивности

В дальнейшем для краткости комплексное пропускание зрачка системы пространственной фильтрации $H(u, v)$ мы будем называть функцией зрачка или просто зрачком. В отсутствие

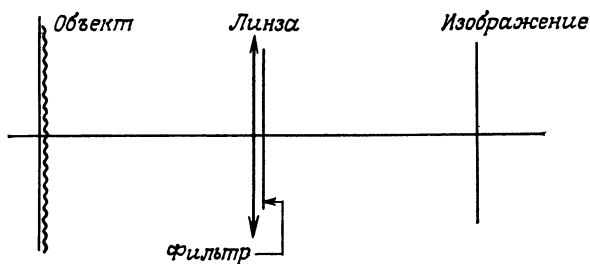


Рис. 7.2. Схема системы пространственной фильтрации в некогерентном свете, в которой имеется реальный фильтр в виде транспаранта, помещенного в плоскости зрачка линзы.

пространственного фильтра это просто зрачок самой оптической системы (рис. 7.2).

Б. Следствия

При первом взгляде на табл. 7.1 кажется, что при переходе от когерентной оптики к некогерентной возможности оптических методов обработки информации сильно уменьшаются в двух отношениях:

1) сужается круг допускающих обработку сигналов, так как теперь сигналы должны быть действительными и положительными (мы будем также употреблять термин «действительные однополярные сигналы» в противоположность двухполярным сигналам, которые могут принимать как положительные, так и отрицательные значения);

2) сужается круг практически осуществимых фильтров, так как они должны удовлетворять двум эквивалентным соотношениям (7.1) и (7.3).

Если внимательнее изучить области применений оптических методов обработки информации, то можно заметить, что только второе ограничение является действительно серьезным. На практике первое ограничение ослабляется в силу многих факторов. Во-первых, все подлежащие обработке изображения (кроме изображений, формируемых в микроскопе) обычно содержат всю семантическую информацию в виде изменения прозрачности по интенсивности. Во-вторых, фазовая модуляция, которая может иметь место, обычно является паразитной, вводимой без нашего ведома в результате плохого контроля качества параметров используемых оптических элементов: состояния поверхности фотоэмульсии, однородности стекла подложки и т. п.

И, в-третьих, даже при когерентной фильтрации мы не имеем простых средств для использования фазовой информации, содержащейся в выходной величине. Эта фазовая информация всегда пропадает, когда изображение на выходе наблюдают глазом или измеряют с помощью разного рода фотоэлектрических преобразователей (являющихся квадратичными по амплитуде детекторами).

Главным недостатком некогерентной фильтрации является *положительность импульсного отклика системы пространственной фильтрации*.

Исчезает также и такое преимущество когерентной фильтрации, как возможность доступа к частотной плоскости, где помещаются пространственные фильтры, и использования ее как плоскости изображения (гл. 6, § 1, п. Б). Ранее было показано, что при когерентном освещении задняя фокальная плоскость линзы является частотной плоскостью (гл. 6, § 1). Что же будет в этой плоскости при использовании абсолютно некогерентного источника освещения (образованного точечными источниками, фазы которых не коррелированы, т. е. имеют микроскопически малый радиус корреляции)? Если пренебречь дифракцией на участке объект — линза, то каждая точка источника S_1 создаст в фокальной плоскости линзы однородное распределение интенсивности. Очевидно, что тогда и суммарная интенсивность будет распределена однородно.

Таким образом, пространственная фильтрация при некогерентном освещении будет осуществляться совершенно иначе, нежели в случае когерентного освещения. Если при когерентном освещении нам было безразлично, как характеризовать пространственный фильтр, импульсным откликом или передаточной функцией, и затем мы могли физически реализовать требуемую передаточную характеристику фильтра на материале, то теперь мы не имеем больше возможности такого выбора. Мы должны *сначала задать импульсный отклик* некогерентной системы фильтрации, удовлетворяющий выражению (7.1), а затем вычислить соответствующую ему передаточную функцию. Но в данном случае это не та передаточная функция, которая будет реализована материально в виде некоторого транзистора. Теперь это зрачок оптической системы, которому, согласно выражению (7.3), соответствует требуемая передаточная функция. Следовательно, в случае некогерентного освещения *импульсный отклик системы пространственной фильтрации синтезируется косвенно*.

Итак, мы рассмотрели суть некогерентной обработки информации и ее требования к используемым методам. Само собой разумеется, что в рассматриваемом случае возможности оптических методов обработки информации оказываются меньше, чем методов, о которых говорилось в гл. 6. Тем не менее многие авторы (Ломанн, Ловенталь, Лукош, Мейлони, Роудс и др.)

занимались разработкой некогерентных методов и расширением области их применения до ее мыслимых пределов. Их работы привели к весьма удовлетворительному выводу: *«все операции по обработке информации, выполнимые в когерентном свете, могут быть также выполнены и при использовании некогерентного освещения»*. Правда, переход от когерентных оптических методов обработки информации к некогерентным часто сопровождается усложнением аппаратуры, которое заключается в следующем:

- а) в увеличении числа каналов обработки;
- б) в увеличении числа необходимых пространственных фильтров;
- в) в необходимости применения оптоэлектроники для реализации полной передаточной функции.

Поэтому такой переход оправдан лишь в том случае, если он приводит к повышению качества обработки или расширению области ее практического использования.

В следующих параграфах будет показано, каким образом можно расширить область применения пространственной фильтрации при некогерентном освещении. Мы рассмотрим некогерентные методы в порядке возрастания их сложности: сначала обработка в одном канале с однополярными зрачковыми функциями, затем двухканальная обработка (тоже с однополярными зрачками) и, наконец, некогерентная обработка, осуществляемая с помощью голографических фильтров.

§. 2. Области применения некогерентной пространственной фильтрации

А. Естественные (чистые) зрачки

Здесь мы рассмотрим зрачки, которые описываются действительными однополярными функциями и которые проще всех в реализации. К ним относятся, в частности, диафрагмы и различные маски, которые позволяли выполнять все операции пространственной фильтрации (гл. 6, § 2, п. А) в случае когерентного освещения.

Передаточная функция рассматриваемых зрачков при некогерентном освещении дается выражением [формула (7.3)]

$$K(u, v) = H * H^*(-) = H \otimes H.$$

Поскольку функция зрачка H однополярна, выполняется неравенство

$$K(u, v) \leq K(0, 0) \text{ при всех } u \text{ и } v. \quad (7.4)$$

Таким образом, некогерентная система пространственной фильтрации ведет себя как фильтр низких частот. Следовательно,

с ее помощью невозможно осуществить такие описанные в гл. 6, § 2, п. А операции, как повышение контраста изображения или его дифференцирование или же полосовая фильтрация пространственных частот, поскольку фильтры низких частот — это сглаживающие фильтры.

Как же найти зрачковую функцию, которая позволила бы выполнить требуемую операцию пространственной фильтрации? Это весьма сложная задача.

1. Если нам известна действительная и однополярная функция K , удовлетворяющая соотношению (7.3), то это еще не означает, что существует функция k , удовлетворяющая равенству (7.1). Нужно еще, чтобы фурье-образ k функции K был действительным и положительным.

2. Если даже k и удовлетворяет соотношению (7.1), то не существует общего метода для перехода от K к H . В самом деле:

а) уравнение (7.3) не имеет явного решения относительно H ;

б) уравнение (7.1) позволяет определить h , но без фазового множителя, и мы не знаем, как выбрать фазу, чтобы функция H была положительной.

Рассмотрим теперь одно важное и общее свойство некогерентной пространственной фильтрации. Оно вытекает из соотношения (7.3). Если сместить зрачковую функцию $H(u, v)$, т. е. произвести преобразование $H(u, v) \rightarrow H(u - u', v - v')$, то ее функция автокорреляции K останется прежней. Следовательно, *положение фильтра в плоскости зрачка не имеет значения*. В этом большое преимущество некогерентной фильтрации, так как мы знаем, что в фокальной плоскости когерентной оптической системы пространственной фильтрации пространственный фильтр необходимо устанавливать с очень высокой точностью. Этим свойством мы воспользуемся для синтеза мультиплексных многоканальных) пространственных фильтров (гл. 19, § 2).

Б. Двужрачковая некогерентная пространственная фильтрация

Пойдем дальше и предположим, что нам нужен пространственный фильтр с действительным, симметричным, но *двухполярным* импульсным откликом k . Из соотношения (7.1) следует, что такой фильтр нельзя осуществить, обойдясь одним зрачком. Представим k в виде суммы положительной и отрицательной частей (рис. 7.3):

$$k(x, y) = k^+ - k^-. \quad (7.5)$$

Здесь обе функции k^+ и k^- положительные и симметричные. Им соответствуют физически реализуемые зрачковые функции H^+ и H^- , если только для каждой из них можно обратить соот-

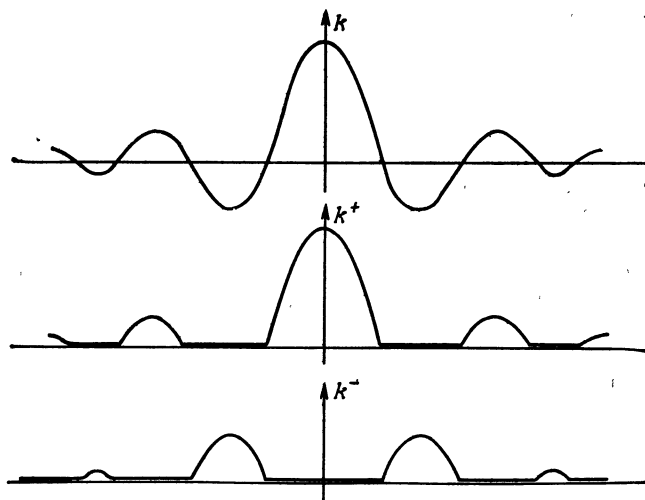


Рис. 7.3. Разложение биполярного импульсного отклика k на две однополярные компоненты k^+ и k^- .

ношение (7.1) или (7.3). Предположим, что этот этап осуществлен либо итерационными методами, либо за счет априорного знания задачи. Тогда мы увидим, что общее решение рассматриваемой задачи представляет собой разность двух частных решений:

$$i = o * k = o * (k^+ - k^-) = o * k^+ - o * k^- = i^+ - i^-.$$

Такое решение предполагает реализацию в виде двух идентичных оптических систем, одна из которых имеет зрачковую функцию H^+ , а другая H^- , и последующее вычитание изображения i^- из изображения i^+ . Такой эксперимент был осуществлен Ловенталем и Шавелем [90]. Сложным моментом данного метода некогерентной пространственной фильтрации, реализуемого с помощью двух зрачков, является операция вычитания промежуточных изображений. Некогерентные методы мало пригодны для реализации операции вычитания изображений, и поэтому для ее осуществления необходимо прибегать к некоторым хитростям. В обзоре Эберсоля [91] описаны различные методы вычитания изображений. Перечислим некоторые из них:

1) регистрация изображения посредством фотоприемника и последующее электронное вычитание [92];

2) модуляция изображений с помощью решеток, сдвинутых по фазе на 90° , с последующей высокочастотной фильтрацией [93];

3) модуляция изображений с помощью микроскопических рассеивателей [94];

4) голографическое вычитание изображений в когерентном свете [95];

5) применение двулучепреломляющих материалов [96] и поляризованного света;

6) применение пространственно-временных модуляторов света с оптической записью (приложение 5).

Изложенный подход к двузрачковой некогерентной фильтрации может быть обобщен на случай синтеза более сложных зрачковых функций. Так, например, Ломанн и Роудз [97] разработали очень общую теорию некогерентной фильтрации, охватывающую все типы пространственных фильтров.

В. Голографические зрочки

Пространственная фильтрация, осуществляемая с помощью одно- и даже двухполярных действительных передаточных функций, не позволяет выполнять все операции, выполняемые, например, в когерентной оптике. Напомним, что в гл. 6, § 3, п. В мы определили передаточную функцию согласованного с образом $s(x, y)$ фильтра в виде $S^*(-u, -v)$. Данное определение справедливо также и при некогерентном освещении. Из него с очевидностью следует, что если объект $s(x, y)$ не обладает осевой симметрией, то необходимо реализовывать комплексную передаточную функцию. Потребовать, чтобы функция K была равна передаточной функции согласованного фильтра, и надеяться затем найти зрачковую функцию H , удовлетворяющую соотношению (7.3), — это может показаться безнадежным делом. Но оптика делает все так, как нам нужно! Мы сейчас увидим, что подход, принятый в случае когерентного света, дает нам и сам фильтр!

Сравним оба подхода (когерентный и некогерентный, табл. 7.2). В левом столбце таблицы приведены основные характеристики когерентной согласованной фильтрации, используемой для распознавания образов и основанной на методе Вандер Люгта (гл. 6, § 3, п. В). Правый столбец оказывается аналогичным левому, но входной переменной здесь служит не комплексная амплитуда, а интенсивность.

Если теперь мы сравним столбцы по строкам, то к своему удивлению заметим, что в обоих случаях физически будет использоваться *только один фильтр*, пропускание которого описывается выражением

$$A + T f^{i\varphi} + T^* e^{-i\varphi}.$$

В самом деле, возьмем 4-ю строку. Имея в виду, что при некогерентной фильтрации зрачковая функция $H(u, v)$ выпол-

Таблица 7.2

	Когерентная пространственная фильтрация	Некогерентная пространственная фильтрация
Входной сигнал	Комплексная амплитуда f	Интенсивность $o = f ^2$
Нужная форма	s — комплексно	$\rho = s ^2$ — действительное положительное
Согласованный фильтр	$T = S^*(-u, -v)$ — комплексное	$Q = P^*(-u, -v)$ — комплексное
Голографический фильтр Вандер Люгта	$A + Te^{i\varphi} + T^*e^{-i\varphi}$	$B + Qe^{i\varphi} + Q^*e^{-i\varphi}$
Импульсный отклик	$A' + t*\delta(x - x_0) + t^*(-x)*\delta(x + x_0)$	$B' + p*\delta(x - x_0) + p^*(-x)*\delta(x + x_0)$
Выходной сигнал	$A'_1 + f*s^*(-x, -y)*\delta(x - x_0) + A'_2$	$B'_1 + o*p^*(-x, -y)*\delta(x - x_0) + B'_2$

няет ту же функцию, что и голографический пространственный фильтр $A + Te^{i\varphi} + T^*e^{-i\varphi}$ в когерентной фильтрации, можно написать

$$H = A + Te^{i\varphi} + T^*e^{-i\varphi}. \quad (7.6)$$

Воспользовавшись соотношением (7.3), вычислим передаточную функцию K для некогерентного случая

$$K = H \otimes H^* = (A + Te^{i\varphi} + T^*e^{-i\varphi})(A + T^*e^{-i\varphi} + Te^{i\varphi}) \quad (7.7)$$

и покажем, что она равна второму члену на 4-й строке:

$$K' = B + Qe^{i\varphi} + Q^*e^{-i\varphi}.$$

Вычислив фурье-образ обеих частей равенства (7.6), получим импульсный отклик когерентного пространственного фильтра:

$$h = A' + t*\delta(x - x_0) + t(-x, -y)*\delta(x + x_0). \quad (7.8)$$

Известно, что смещение x_0 (т. е. модуляцию фазы φ) следует выбирать таким образом, чтобы эти три слагаемых не перекрывались друг с другом. Найдем теперь импульсный отклик некогерентного пространственного фильтра, воспользовавшись формулой (7.8):

$$\begin{aligned} k = |h|^2 &= [A' + t*\delta(x - x_0) + t(-x, -y)*\delta(x + x_0)] \times \\ &\times [A'^* + t^*\delta(x - x_0) + t^*(-x, -y)*\delta(x + x_0)] = \\ &= |A'|^2 + |t|^2*\delta(x - x_0) + |t(-x, -y)|^2*\delta(x + x_0). \end{aligned}$$

Исключив четыре произведения пространственно разделенных функций и выполнив преобразование Фурье, мы легко находим

второй член 4-й строки и получаем соотношение

$$Q = T \oplus T.$$

Тем самым мы существенно расширили возможности применения некогерентной фильтрации. *Чтобы осуществить согласованную фильтрацию в некогерентном свете, достаточно в зрачок некогерентной оптической системы поместить голографический пространственный фильтр, который использовался для выполнения той же самой операции в когерентном свете.* При этом и результат фильтрации отображается аналогичным образом в первом порядке дифракции, соответствующем $x = x_0$.

На основе описанного метода Ловенталь и Вертс [98] разработали сложные пространственные фильтры для автоматического чтения.

Г. Зрачковые функции, синтезируемые методами цифровой голографии

Разумеется, можно реализовать зрачковые функции для некогерентной согласованной фильтрации и методами цифровой голографии. Как и в случае оптической голографии, ЭВМ позволяет синтезировать зрачковую функцию, которая описывается неким математическим выражением, не имея ее реального образца, аналогично методу Вандер Люгта. Однако в данном случае передаточная функция синтезируется косвенным образом, и это делает процесс формирования зрачка более сложным. На этом вопросе мы и остановимся здесь.

Итак, предположим, что нам нужно синтезировать зрачок с неизвестным пропусканием $H(u, v)$, зная его передаточную функцию $K(u, v)$. Поскольку уравнение (7.3) непосредственно не решается относительно H , перейдем к фурье-образам входящих в него величин.

Теперь мы можем написать $k = |h|^2$, где k известно, а h требуется определить. Положим $h_1 = \sqrt{k}$. Это уже будет решением нашей задачи, поскольку фурье-образ $H_1 = \mathcal{F}[h_1]$ удовлетворяет уравнению (7.3). Но оно, возможно, не будет лучшим решением. Все решения вида

$$h_i = h_1 \exp[i\varphi_i(x, y)]$$

также дадут зрачковые функции H_i , удовлетворяющие уравнению (7.3). В частности, можно так выбрать распределение фазы φ_i , чтобы, например, получить: а) самую простую функцию H_i ; б) или наилучшую эффективность фильтрации; в) или самый малый динамический диапазон изменения H_i и т. д.

Итак, мы имеем некоторую свободу выбора, позволяющую согласовать характеристики фильтра с экспериментальными условиями его применения. Это важное преимущество обусловлено неоднозначностью решения уравнения (7.3).

Д. Экспериментальные условия

Все рассмотренные нами методы фильтрации основывались на теории дифракции и уравнении фильтрации (7.2). Вот почему мы требовали, чтобы источник излучения был некогерентным (протяженным), но монохроматичным. Если размер источ-

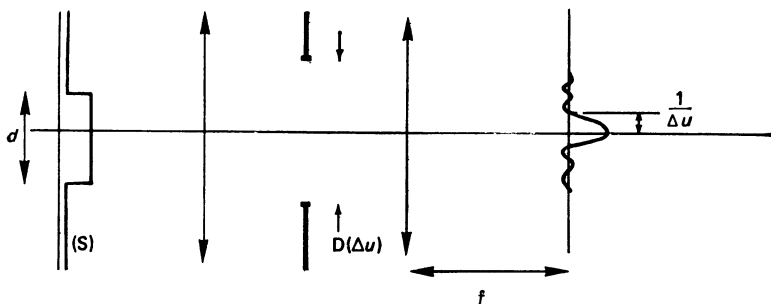


Рис. 7.4. Схема пространственной фильтрации, в которой использован монохроматический, но протяженный источник S.

Зрачок диаметром D дает дифракционное пятно шириной $1/\Delta u = \lambda f/D$. Действительное изображение источника в выходной плоскости представляет собой свертку его геометрического изображения с дифракционной картиной зрачка.

ника обозначить через d , а максимальную обрабатываемую оптической системой пространственную частоту через Δu , то очевидно, что должно выполняться неравенство (рис. 7.4)

$$d \ll \frac{1}{\Delta u} = \frac{\lambda f}{D},$$

где D — размер пространственного фильтра. В противном случае в плоскости изображения будет наблюдаться не только результат фильтрации, но также и свертка с источником.

Но реальные источники излучения не являются строго монохроматичными. Существует простое соотношение между разрешением оптической системы, которое требуется получить, и спектральной шириной источника $\Delta \lambda$ (рис. 7.5). Каждая синусоидальная составляющая фильтра будет давать в плоскости изображения пару светлых точек, расстояние между которыми пропорционально длине волны используемого освещения. В таком случае в первом приближении выходное изображение будет свернуто с функцией, характеризующей размеры источника и его спектральную ширину,

$$\frac{f}{\delta} \Delta \lambda. \quad (7.9)$$

Пусть размер изображения равен L , а требуемое разрешение равно l . По теореме Шенона число разрешаемых системой точек (или число ее степеней свободы) дается формулой

$$N = \frac{L}{l} = \frac{LD}{\lambda f} = \frac{D}{\delta}.$$

С учетом выражения (7.9) получаем

$$\frac{f}{\delta} \Delta\lambda \ll l.$$

Теперь легко вычислить допустимую спектральную ширину источника в зависимости от размера фильтруемого изображения L и требуемого разрешения l :

$$\frac{\Delta\lambda}{\lambda} \ll \frac{l}{L} = \frac{1}{N}, \quad (7.10)$$

где N — число независимых точек на ширине изображения.

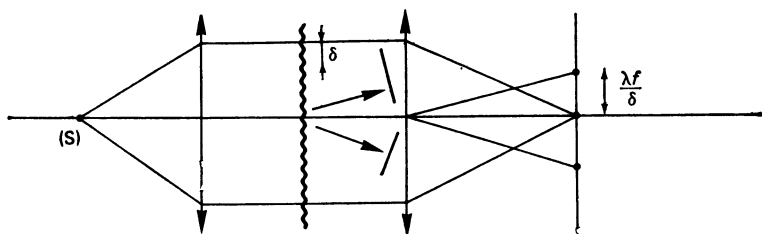


Рис. 7.5. Схема пространственной фильтрации, в которой использован точечный, но полихроматический источник S .

При заданной длине волны λ и данном пространственном периоде δ соответствующего фильтра два дифракционных порядка будут отстоять друг от друга на величину $2\lambda f/\delta$. Если же источник имеет ширину спектра излучения $\Delta\lambda$, то тот же пространственный период δ даст два дифракционных порядка, отстоящих друг от друга на $\Delta\lambda f/\delta$.

В случае обычных изображений этим условием заметно ограничивается выбор пригодных источников света, и поэтому очень часто для освещения объекта используется лазер, пространственная когерентность излучения которого нарушается с помощью движущихся рассеивателей, помещенных перед объектом. Пригодны также спектральные лампы (ртутные или натриевые) со светофильтрами. Если объект не прозрачный, а отражающий, его можно освещать непосредственно лазером. Дифракции лазерного излучения на шероховатой поверхности объекта обычно достаточно для обеспечения условия пространственной некогерентности.

Мейлони [99] создал систему распознавания знаков, взяв в качестве источника света электронно-лучевую трубку со специальным люминофором.

Некогерентные методы обработки информации по сравнению с когерентными предъявляют менее жесткие требования и к другим условиям эксперимента (стабильность, точность уста-

новки оптических элементов и др.). Единственный действительно трудный момент в некогерентной фильтрации, который мы уже отмечали, связан с операцией вычитания изображений, которая играет существенную роль в двух случаях:

1) при осуществлении двузрачковой фильтрации для получения разности двух изображений;

2) при однозрачковой фильтрации, когда для синтеза двухполярного импульса отклика добавляют равномерный фон, который поднимает отрицательные значения выше нуля.

Однако операция вычитания по самому своему принципу является источником шума. В самом деле, уменьшаемое и вычитаемое, как правило, одинаковы по порядку величины (по яркости), а их разность (а именно, она и интересует нас) намного меньше каждого из них. Поэтому отношение сигнала к шуму в разностном изображении будет значительно меньше, чем в двух исходных изображениях. Это справедливо независимо от способа осуществления вычитания (оптическое, электронное или цифровое), однако если процедура вычитания сама является источником шума, как это имеет место в большинстве аналоговых методов, то качество изображения ухудшается еще заметнее.

§ 3. Методы пространственной фильтрации, не использующие свойств дифракции

В подходе, который был использован в § 1 и 2 и который строго учитывал влияние зрачка оптической системы пространственной фильтрации на выходное изображение, мы явным образом принимали во внимание дифракцию света. С учетом дифракции выводятся уравнения фильтрации (7.2) и (7.3), а также требование монохроматичности источника (7.9).

Для некоторых задач пространственной фильтрации выполнение условия монохроматичности может оказаться слишком трудным. Так обстоит дело, например, в тех случаях, когда источником обрабатываемого изображения служит экран телевизора (черно-белого или цветного). В таких случаях можно применять методы пространственной фильтрации, основанные главным образом на геометрической оптике. Но применимость таких методов меньше, чем рассмотренных ранее. К тому же они ограничены в отношении разрешающей способности, поскольку не учитывают дифракцию.

А. Теневое маскирование (точечный источник белого света)

Итак, рассмотрим точечный источник, находящийся в бесконечности, и его изображение, $\delta(x, y)$, создаваемое линзой с фокусным расстоянием f (рис. 7.6). Поместим на линзу зрачок

с пропусканием по интенсивности $t(x, y)$ и исследуем изображение точечного источника в плоскости π , отстоящей на расстоянии z от фокальной плоскости.

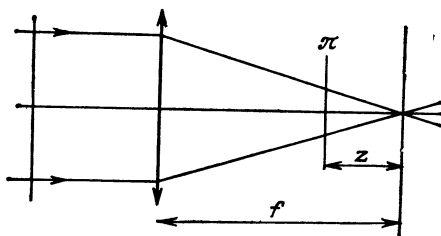


Рис. 7.6. Метод теневой маски.

Этот метод позволяет очень просто создавать пространственные фильтры путем синтеза импульсного отклика. Но его применимость ограничена приближениями геометрической оптики (не учитывается дифракция).

В пределах применимости геометрической оптики распределение интенсивности в изображении точки в плоскости π описывается сверткой

$$\delta(x, y) * t'(x, y),$$

где

$$t'(x, y) = t\left(\frac{z}{f}x, \frac{z}{f}y\right).$$

Следовательно, изображение объекта o в плоскости η будет таким:

$$i(x, y) = o(x, y) * t'(x, y).$$

Это выражение описывает отфильтрованное изображение, которое может быть получено с помощью рассматриваемой оптической системы. Но нужно установить границы применимости такого способа пространственной фильтрации.

1. Масштабный коэффициент z/f нельзя выбрать сколь угодно малым, поскольку плоскость π не может находиться слишком близко к фокальной плоскости. Хороший нижний предел для величины z — расстояние дальней зоны (гл. 4, § 3, п. В, 4).

2. Импульсный отклик $t(x, y)$ ограничен по разрешению. В самом деле, пусть δ — самая малая деталь зрачка. Влияние дифракции в промежутке между линзой и плоскостью π будет пренебрежимо малым, если расстояние между ними меньше расстояния дальней зоны для пучка, исходящего из этого элемента разрешения, т. е. если в соответствии с (4.8) выполняется условие

$$f - z \ll \frac{\delta^2}{\lambda}.$$

Таким образом, будучи чрезвычайно простым (импульсный отклик фильтра синтезируется прямо и просто), этот метод, од-

нако, оказывается существенно ограниченным по своим возможностям и главным образом по разрешению, или числу точек, фильтра. Поэтому его применение целесообразно только в тех задачах фильтрации, где импульсный отклик фильтра имеет простую форму.

Б. Другие методы пространственной фильтрации, основанные на геометрической оптике

Таких методов много, и в них часто используются особые свойства оптических систем или их компонентов (дифракционных решеток, призм, гофрированных экранов, растров и т. д.). Возможности их применения почти всегда ограничены, и они

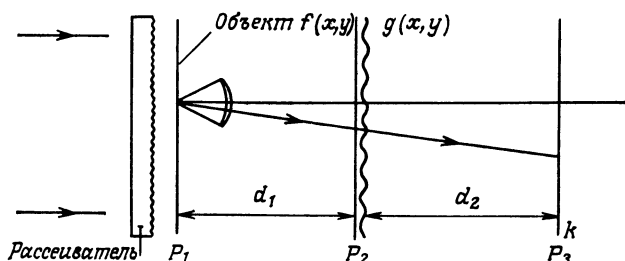


Рис. 7.7. Корреляция в белом свете с использованием рассеивателя.

Две величины, взаимную корреляцию которых требуется определить, задаются в плоскостях P_1 и P_2 . Каждая точка поверхности рассеивателя служит источником сферической волны.

пригодны только для реализации специфических видов пространственной фильтрации или для обработки особых изображений. Но зато они обычно очень просты в практическом осуществлении. Ряд интересных примеров их применения можно найти в книге Роджерса [100].

Мы рассмотрим только один метод, чтобы проиллюстрировать возможности весьма сложной оптической обработки в самом общем случае некогерентности — в случае протяженного источника белого света.

Итак, предположим (рис. 7.7), что объект f освещается с помощью рассеивателя. Тогда каждую точку объекта в первом приближении можно рассматривать как малый источник сферических волн. Поместим в плоскость P_2 пространственный фильтр g в виде некоторого транспаранта. Тогда интенсивность в точке x плоскости наблюдения P_3 будет равна

$$k(x) = \int_{-\infty}^{\infty} g(x - x_1) f\left(x - \frac{d_1 + d_2}{d_2} x_1\right) dx_1.$$

Полученный интеграл легко сводится к корреляционному путем замены переменных и изменения масштаба. Вот и все описание некогерентных методов обработки информации. Можно сделать три очевидных вывода:

1. Эти методы значительно видоизменяют области применения когерентной оптики, позволяя использовать в качестве носителей обрабатываемых изображений самые обычные носители информации (журналы, книги, фотографии на бумаге), а также использовать и самые быстродействующие источники информации (телевизионные системы, осциллографы и т. д.).

2. Их применение приводит к повышению качества обработанных изображений, поскольку устраняются два основных источника шума в когерентной оптике: зернистость и чувствительность к разного рода фазовым дефектам оптических элементов.

3. У них, к сожалению, имеются и недостатки, которые можно устранить лишь ценой усложнения обработки.

Теперь мы на некоторое время оставим оптическую обработку информации и перейдем к вопросам передачи информации оптическими методами. Мы познакомимся с техникой оптической связи и усвоим такие понятия, как полоса пропускания, пропускная способность, коммутация и т. п. Затем в гл. 19 мы вновь вернемся к вопросам обработки информации. Там мы на основе более «системного» подхода рассмотрим некие сочетания оптики и электроники, позволяющие осуществить некоторые структуры обработки информации, представляющиеся особенно богатыми по своим возможностям, и рискнем заглянуть в будущее... Но сначала обратимся к вопросам оптической связи.

ГЛАВА 8

РАСПРОСТРАНЕНИЕ СВЕТА В ОПТИЧЕСКИХ ВОЛОКНАХ

В данной главе мы изложим основы, необходимые для понимания явления распространения света в оптических волокнах с учетом элементов теории оптики и электромагнетизма, представленных в гл. 1, § 4. Глава 9 будет посвящена измерению основных характеристик, а в двух следующих главах (гл. 10 и 11) речь пойдет о технологии изготовления оптических волокон, кабелей и разъемов.

Поскольку тем самым канал передачи волоконно-оптической системы связи будет полностью охарактеризован, останется осветить вопрос об оконечных устройствах, осуществляющих преобразование электронов в фотоны и фотонов в электроны; это будет сделано в гл. 12—14. Охарактеризовав эти устройства и указав те специфические требования, которые должны выполняться при включении их в передающую систему, мы получим возможность исследовать всю такую цепь в гл. 15 и следующих за ней главах.

С появлением лазерных источников со слабой расходимостью возобновился интерес к передаче информации оптическими методами. Довольно скоро проблемы, связанные с передачей информации из одного пункта в другой в свободном пространстве, вызвали к жизни вопрос о распространении света в световодах, а затем — о замене свободного пространства замкнутым. Таким образом, возникла идея оптических световодов, основанных на том же принципе, что и светящиеся фонтаны: свет можно удерживать в тонком стеклянном волокне, которое представляет собой диэлектрический световод и называется обычно сердцевиной (рис. 8.1). Направленное распространение света обеспечивается посредством оболочки, показатель преломления которой ниже показателя сердцевины; в светящемся фонтане роль оболочки играет воздух.

До 1970 г. потери света в лучших сортах стекла составляли не менее нескольких сотен, а то и тысяч децибел на километр. Фирма Corning Glass Works первой объявила о получении стекла с потерями сначала 20, а потом и 5 дБ/км. Открывшиеся в результате этого многочисленные возможности различных применений дали необычайный толчок исследованиям как в направлении поиска новых материалов и методов производства волокна, так и в направлении разработки источников света и

фотоприемников. Полупроводниковые лазеры, электролюминесцентные диоды (светодиоды), фотодиоды с *PIN*-переходами и лавинные фотодиоды отныне фигурируют в каталогах многих фирм, и уже имеются оптические волокна с потерями всего 0,2 дБ/км на длине волны 1,55 мкм. Таким образом, после

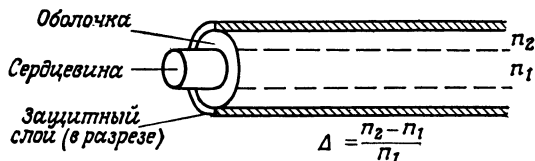


Рис. 8.1. Оптическое волокно.

Состоит из центральной части, называемой сердцевиной, и окружающей ее оболочки. Благодаря оболочке в сердцевине удерживается большая часть передаваемой по волокну световой энергии. Свет распространяется в оптическом волокне при условии $n_2 < n_1$, где n_1 — показатель преломления сердцевины, а n_2 — оболочки.

15-километрового пробега светового сигнала по волокну половина его исходной энергии оказывается еще не израсходованной!

§ 1. Некоторые сведения о распространении света в атмосфере

В гл. 4, § 3 мы установили, что с точки зрения расходимости лучше всего гауссов пучок. Таким образом, можно быстро оценить характеристики световодной системы, приняв, что в ней распространяется гауссов пучок.

Рассмотрим конфокальный резонатор, состоящий из двух одинаковых сферических зеркал радиусом R , расположенных на расстоянии R одно от другого. Пользуясь обозначениями гл. 4, § 3 (рис. 8.2), мы будем рассматривать луч минимальной апертуры ω_0 . Допустим, что ω_E — его ширина для передающей оптики, а ω_R — для принимающей; расстояния z_E и z_R , сумма которых равна R , есть решения уравнения (4.20), которое в данном случае примет вид

$$z^2 - Rz + z_0^2 = 0.$$

При помощи выражения (4.19) мы можем связать величину ω с расстоянием z в плоскости, где ширина пучка минимальна. Мы получаем, что произведение $\omega_E \omega_R$ есть следующая функция величин R и λ :

$$\omega_E \omega_R = \frac{R\lambda}{\pi}.$$

Если взять полезные размеры передающей и приемной антенн равными ω_E и ω_R , то будут некоторые потери (поскольку световой пучок расходится во всю полусферу), что эквива-

лентно ослаблению оптического сигнала. В табл. 8.1 представлены результаты вычислений [101] при $\omega_E = \omega_R = 1$ м для двух длин волн. Даже при значительных расстояниях потери оказываются не очень уж большими.

Таблица 8.1

Приближенная оценка ослабления оптического сигнала при распространении в вакууме между антеннами радиусом 1 м (дифракционные потери)

L	1000 км	10 000 км
λ		
0,63 мкм	0 дБ	20 дБ
10,6 мкм	20 дБ	40 дБ

Но атмосфера — это не вакуум. В гл. 2, § 4, п. Б мы уже рассматривали простые механизмы поглощения света. Реальная среда еще сложнее, и в атмосфере имеются три главных фактора, ухудшающие ее световодные характеристики.

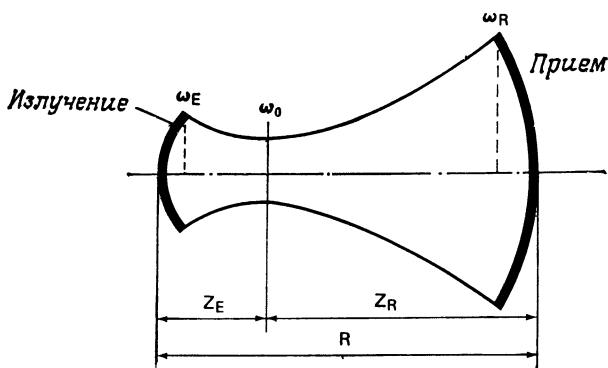


Рис. 8.2. Конфокальный несимметричный резонатор (в разрезе), образованный зеркалами-антеннами одинакового радиуса R , и гауссов пучок, ограниченный краями антенн.

1. *Молекулярное поглощение.* Оно сильно зависит от длины волны и определяется свойствами газов, входящих в состав атмосферы. Как видно из рис. 8.3, имеются так называемые окна прозрачности атмосферы, приходящиеся на пригодные для связи участки спектра.

2. *Поглощение, обусловленное аэрозолями* (к которым относятся, кроме тумана, взвешенные в воздухе мельчайшие твердые частицы). Единственный способ борьбы с этим видом поглощения — увеличить длину волны, чтобы она была больше размеров частиц (работать в инфракрасной области спектра).

3. *Атмосферная турбулентность.* Показатель преломления воздуха есть случайная функция координат и времени; таким образом, в параметрах принятой волны имеются статистические неопределенности. Основные последствия атмосферной турбулентности таковы: а) смещение пучка в точке приема; б) изменение амплитуды и фазы волны в плоскости приема.

Надежность передачи обычно оценивается временем, в течение которого атмосферные потери не превышают некоторую за-

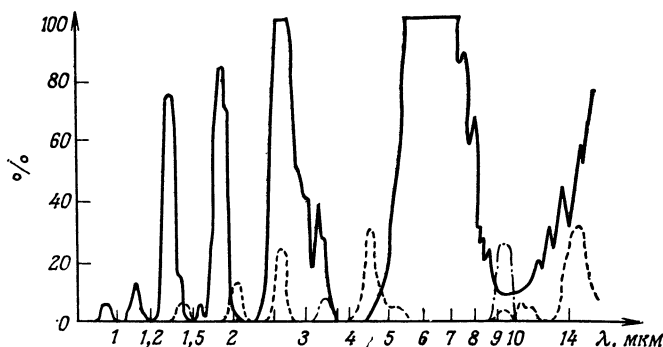


Рис. 8.3. Зависимость коэффициента поглощения от длины волны для разных компонент атмосферы [101].

Сплошная линия — для паров воды; штриховая — для углекислого газа; штрихпунктирная — для озона.

данную величину, и выражается в процентах от времени передачи. Она должна быть равна 90, даже 99,5 или 99,9%. Эксперимент показывает, что обеспечить такую высокую надежность связи можно только в исключительных климатических условиях. Отсюда вытекает необходимость проведения предварительных исследований при создании оптических систем связи в открытом пространстве. Такой вид оптической связи представляет интерес, например, для систем космической связи (правда, при этом возникают трудности вхождения в связь).

Для устранения вредного влияния атмосферы приходится использовать распространение света в замкнутой среде и, следовательно, различные способы направленного распространения света в световодах.

С этой целью были предложены и изучены многие методы. Мы лишь бегло коснемся их, прежде чем рассматривать оптические волокна.

1. *Диафрагменный световод.* После прохождения световым пучком ряда диафрагм устанавливается конфигурация поля с наименьшими потерями. При помощи диафрагм диаметром 4,5 см, расположенных на расстоянии 100 м друг от друга,

можно получить потери в 1 дБ/км. Главная трудность, разумеется, состоит в точном совмещении диафрагм [102]! Иллюстрация этого явления уже была дана в гл. 4, § 3.

2. *Линзовый световод*. Тонкая линза вводит такое же изменение оптического пути луча, что и вогнутое зеркало с тем же радиусом кривизны. Следовательно, гауссовы пучки представляют собой также моды, допускающие распространение в линзовом световоде (гл. 4, § 3, п. Б). Потери в линзовых световодах весьма незначительны (0,5 дБ/км [102]). Но такие световоды трудны в осуществлении: зависимость от положения линз, трудности направления света по кривой и т. д. К таким устройствам относятся системы газовых линз, действие которых основано на зависимости показателя преломления от температуры газа. При определенных условиях эти среды с градиентом показателя преломления становятся фокусирующими, но их практическое осуществление связано с большими трудностями. Укажем эксперименты, проведенные фирмой Bell Laboratories [103] и Национальным центром исследований по связи [104].

Самыми простыми в осуществлении световодами остаются оптические волокна.

§ 2. Оптические волокна

Оптическое волокно (рис. 8.1) представляет собой внутреннюю диэлектрическую среду (стекло, кварц и т. п.), в которой содержится основная часть световой энергии, передаваемой по волокну, и которая называется *сердцевиной*. Сердцевина может быть окружена слоем с более низким показателем преломления, который называется *оболочкой*. Для защиты от внешних воздействий и для повышения механической прочности сердцевина с ее оболочкой может быть покрыта слоем пластмассы.

Существуют разные типы волокон:

1. *Оптические волокна без оболочки*. Такие волокна представляют собой просто стеклянную нить. Они хрупки и неэффективны. У них большие потери, поскольку на границе двух сред электрическое поле не равно нулю, а граница весьма несовершенна. Притом, чтобы такое волокно было одномодовым, его диаметр должен быть 1 мкм [формула (3.39)].

2. *Оптические волокна с оптической оболочкой*. В 1954 г. было предложено [105, 106] покрывать световод очень тонкой оболочкой с показателем преломления несколько более низким, чем у сердцевины. Как мы увидим в дальнейшем (§ 4, п. Б, 1), распространение света в таком световоде зависит от приведенной частоты V [гл. 3, § 3, п. Б, формула (3.37)]. В нем возможна только одна мода распространения, если V меньше $V_c = 2,405$ (волокно со скачкообразным изменением показателя преломления). *Следует иметь в виду, что здесь речь идет о све-*

товодах круглого сечения. В лабораторных условиях удалось создать такие одномодовые волокна с потерями 0,2 дБ/км на длине волны 1,55 мкм [107—109]. Если увеличить диаметр сердцевины или разность показателей преломления сердцевины и оболочки, то оптический световод сможет пропускать очень много мод. В этом случае будет проще описывать свойства оптических волокон на основе геометрической оптики (гл. 2, § 2). С 1978 г. наилучший из опубликованных результатов с такого типа волокнами состоял в передаче цифрового сигнала со скоростью 32 Мбит/с на расстояние 53 км со средними потерями 0,7 дБ/км на длине волны 1,3 мкм [110].

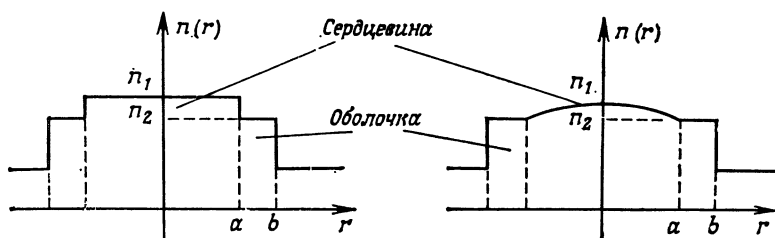


Рис. 8.4. Два основных типа оптического волокна.

а — волокно со скачкообразным изменением показателя преломления; *б* — волокно с градиентом показателя преломления (градиентное волокно).

В волокне с постоянным показателем преломления сердцевины, которое обычно называют *волокном со скачком показателя преломления*, распространение света обеспечивается за счет полного внутреннего отражения на границе между сердцевинной и оболочкой: свет распространяется зигзагообразно, и в зависимости от угла θ , под которым луч наклонен к оси (гл. 3, § 3), проходимое им расстояние и число отражений сильно меняются. Если сделать, чтобы показатель преломления изменялся в зависимости от расстояния r от оптической оси по параболическому закону вида

$$n(r) = n(0) - [n(0) - n(a)] \left(\frac{r}{a}\right)^2,$$

то оптические пути разных лучей будут практически одинаковыми. Как мы увидим позднее, это очень важно с точки зрения передачи информации. Оптические волокна, в которых показатель преломления изменяется по радиусу сердцевины (рис. 8.4, б), обычно называют волокнами с градиентом показателя преломления или градиентными волокнами.

Выбранный закон изменения показателя преломления может быть более или менее сложным (рис. 8.5, а). Направленное распространение света возможно также и в однородном материале, если ему придать определенную форму (рис. 8.5, б). Мы

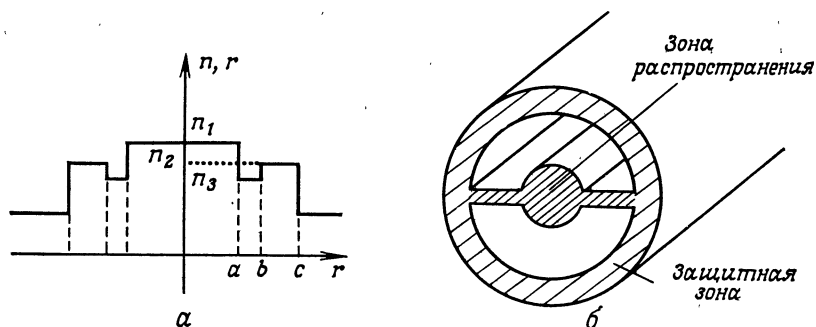


Рис. 8.5. Оптические волокна со сложным профилем показателя преломления.

a — W-образный профиль показателя преломления. Между точками a и b имеется провал, который облегчает выделение доминирующей моды; b — оптическое волокно из одного материала.

здесь не будем останавливаться на этом, а заинтересованный читатель найдет необходимую информацию в специальной литературе [111, 112].

§ 3. Исследование распространения света

Главы 3 и 4 подготовили нас к основной проблеме — проблеме изучения распространения света внутри одномерного световода. Число мод, возможных в световоде с приведенной частотой V , легко определить на основе результатов гл. 3, § 3, п. Б¹⁾: оно оказывается равным $V^2/2$, т. е. при разности показателей преломления 10^{-2} и радиусе сердцевины 50 мкм получаем около 1500 мод. Величина такого порядка, полученная недавно на практике, позволяет нам пользоваться геометрической оптикой, поскольку, как легко видеть, условия ее применимости здесь выполняются: относительные изменения комплексных амплитуд полей на расстояниях порядка длины волны малы.

Мы будем пользоваться методом геометрической оптики в § 3 и 5, рассматривая, в частности, условия ввода лучей в волокно и траектории лучей. Но этот метод не позволяет разрешить все поставленные проблемы, например проблему расчета потерь, в частности, в случае мод утечки, когда требуется знать остаточную амплитуду поля на некоторых каустиках. В этом случае нам придется обратиться к уравнениям Максвелла и разложить поле по модам. Расчеты оказываются слишком сложными, если не вводить упрощающих предположений (напри-

¹⁾ В данной главе рассматриваются световоды, имеющие два поперечных размера. Моды характеризуются двумя индексами, и их число пропорционально V^2 , а не V , как в гл. 3, § 3, п. Б.

мер, о малой разнице показателей преломления сердцевины и оболочки). Такой подход будет рассмотрен в § 4.

В § 5 и далее речь идет о распространении энергии и оптимизации волокон. В частности, рассматриваются улучшения, достигаемые в результате определения правильного градиента преломления, позволяющего обеспечить приблизительное равенство групповых скоростей разных мод (чаще говорят, впрочем, о времени распространения группы).

Исследование дисперсии, обусловленной материалом, продолжается в § 7, п. Д и может быть с легкостью обобщено на многомодовые волокна.

На практике проводится различие между одномодовыми и многомодовыми волокнами. Мы тоже рассматриваем одномодовые волокна отдельно в § 8 вместе с рядом специальных вопросов, которых мы не касаемся при рассмотрении многомодовых волокон.

А. Распространение света в многомодовых оптических волокнах

Многомодовые оптические волокна, диаметр которых составляет несколько десятков микрометров, а разница показателей преломления — порядка 10^{-2} , отвечают условиям, необходимым для использования геометрической оптики (гл. 2). Поэтому при изучении явления распространения света оказывается возможным локально заменить волновую поверхность ее касательной плоскостью и рассматривать траектории, ортогональные волновым поверхностям, т. е. *световые лучи*.

Мы будем исследовать уравнение лучей (2.21)

$$\frac{d}{dt} \left(n \frac{d\mathbf{r}}{dt} \right) = \text{grad} [n(\mathbf{r})]$$

в определенной среде, которой является оптическое волокно, и будем также привлекать уравнение (2.20), которое означает, что вектор, касательный по отношению к световому лучу, коллинеарен градиенту эйконала $S(\mathbf{r})$. Напомним, что равенство $S(\mathbf{r}) = \text{const}$ есть уравнение поверхности определенной волны.

В дальнейшем мы будем пренебрегать потерями, обусловленными поглощением в материале, которые, строго говоря, следовало бы учитывать (гл. 9, 10); пока нас занимают только явления, связанные с механизмом распространения света в волокне.

Мы покажем, что поведение луча зависит от двух параметров β и ν , роль которых будет пояснена анализом мод. Во многих случаях геометрическая интерпретация, связанная с параметрами β и ν , упрощает исследование траекторий лучей и граничных условий.

1. Обозначения

Мы будем рассматривать цилиндрические волокна с круговой симметрией относительно оси Oz . Показатель преломления зависит только от расстояния до оси. Он максимален на оси [$n(0) = n_1$], а в оболочке, т. е. при $r \geq a$ (где « a » — радиус сердцевинки волокна), принимает постоянное значение n_2 . Если особо не оговаривается, то будем считать, что $n_2 = n(a)$. Исходя из геометрии волокна, мы выбираем цилиндрическую систему координат r, ψ, z . Радиус-вектор \mathbf{r} есть сумма осевой составляющей $z\mathbf{u}_z$, где \mathbf{u}_z — единичный вектор по оси Oz , и поперечной составляющей ρ , которую мы можем записать в виде $r\mathbf{u}_r$, обозначив через \mathbf{u}_r единичный вектор радиального направления. Единичным вектором \mathbf{u}_ψ полярного угла ψ завершается ортонормальный трехгранник $\mathbf{u}_r, \mathbf{u}_\psi, \mathbf{u}_z$:

$$\mathbf{r} = \rho + z\mathbf{u}_z = r\mathbf{u}_r + z\mathbf{u}_z. \quad (8.1)$$

Подставим (8.1) в уравнение лучей (2.21). Поскольку показатель преломления $n(r)$ зависит только от радиального расстояния r , вектор $\text{grad}(n)$ имеет ненулевую составляющую только на \mathbf{u}_r . В проекции на трехгранник $\mathbf{u}_r, \mathbf{u}_\psi, \mathbf{u}_z$ получаем три скалярных дифференциальных уравнения, два из которых решаются весьма просто:

$$\text{на } \mathbf{u}_z \quad n(r) \frac{dz}{dl} = \frac{\beta}{k_0}, \quad (8.2)$$

$$\text{на } \mathbf{u}_\psi \quad n(r) r^2 \frac{d\psi}{dl} = \frac{\nu}{k_0}, \quad (8.3)$$

где через k_0 обозначена величина $2\pi/\lambda_0$. Мы выбрали постоянные интегрирования в виде β/k_0 и ν/k_0 , поскольку так проще получить произведение $k_0 n(r)$, которое представляет собой модуль волнового вектора, возникающего в текущей точке M . Константы β и ν определяются начальными условиями падения рассматриваемого луча.

Если обозначить через θ угол, образуемый вектором, касательным лучу, с осью Oz , то уравнение (8.2) примет вид

$$n(r) \cos \theta = n(r_0) \cos \theta_0 = \text{const}. \quad (8.4)$$

При заданном угле θ_0 вне волокна угол θ_0 может также зависеть от r_0 , т. е. от расстояния от оси до точки падения на входную плоскость. Равенство (8.4) попросту выражает первый закон преломления, а равенство (8.3) — второй закон: преломленный луч остается в той же плоскости, что и падающий. На основании решения (8.2) можно параметризовать луч в переменной z , и тогда уравнение (2.21) в проекции на плоскость

прямого сечения дает

$$\frac{d^2\rho}{dz^2} = \frac{k_0^2}{2\beta^2} \text{grad}[n^2(r)]. \quad (8.5)$$

Полученное уравнение (8.5) в механике описывает движение частицы в поле центральных сил. Равенство (8.3) после перехода к переменной z принимает вид

$$r^2 \frac{d\psi}{dz} = \frac{\nu}{\beta}.$$

Мы узнаем здесь *закон площадей*. Такая аналогия позволяет давать механическое толкование получаемых результатов.

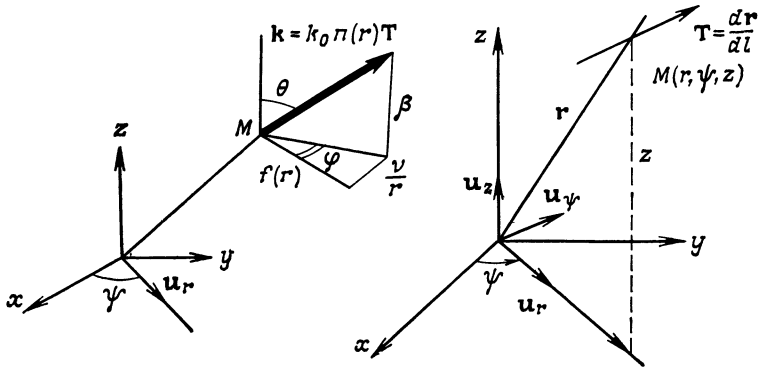


Рис. 8.6. Обозначения, используемые в тексте.

Согласно (2.20), волновой вектор \mathbf{k} , модуль которого равен $k_0 n(r)$, коллинеарен вектору $\mathbf{T} = d\mathbf{r}/dl$, т. е. единичному вектору, касательному к лучу. Используя (8.2) и (8.3), его можно записать следующим образом (рис. 8.6):

$$\mathbf{k} = k_0 n(r) \mathbf{T} = f(r) \mathbf{u}_r + \frac{\nu}{r} \mathbf{u}_\psi + \beta \mathbf{u}_z, \quad (8.6)$$

причем мы ввели обозначение

$$f^2(r) = k_0^2 n^2(r) - \beta^2 - \frac{\nu^2}{r^2}. \quad (8.7)$$

В дальнейшем мы всюду опускаем индекс «0», которым отмечаются значения величин k и λ в вакууме.

Распространение луча характеризуется главным образом функцией $f(r)$. Это означает, что, зная $n(r)$, с одной стороны, и β и ν — с другой, можно вычислить \mathbf{k} в любой точке волокна и, в частности, на его границах. Из (8.7) следует также, что величина $f^2(r)$ может быть равной нулю и отрицательной. Второе означает, что соответствующий луч становится комплексным, т. е. мы имеем *затухающую волну*.

Используя по-прежнему (2.20), находим проекцию вектора \mathbf{k} на направление \mathbf{u}_r : величина $f(r)$ пропорциональна производной dr/dl . С учетом формулы (8.2) окончательно получаем

$$f(r) = \beta \frac{dr}{dz}. \quad (8.8)$$

Для удобства выразим параметры β и ν через геометрические величины, которые позволят нам следить за распространением луча и описывать явления, связанные с распространением света. Обозначив через φ угол между проекцией вектора \mathbf{k} на плоскость прямого сечения и \mathbf{u}_r (рис. 8.6), получаем

$$\begin{aligned} \beta &= kn(r) \cos \theta, \\ \nu &= kn(r) \sin \theta \, r \sin \varphi, \\ f(r) &= kn(r) \sin \theta \cos \varphi. \end{aligned} \quad (8.9)$$

Заметим (это пригодится нам в § 4, п. А), что замена $\theta \rightarrow -\theta$ не меняет величины β , а замена $\nu \rightarrow -\nu$ приводит к точке, симметричной относительно плоскости \mathbf{u}_r , \mathbf{u}_z . Мы ограничимся случаями $\theta \geq 0$, $\nu \geq 0$, учитывая, что имеются четыре луча, которые ведут себя одинаково, если не считать изменения направления и различий в пределах симметрии.

2. Начальные условия

Ими определяются значения различных переменных на входном торце волокна. Предположим, что показатель преломления среды, в которой распространяется падающая волна, равен единице. Плоскую падающую волну будем характеризовать вектором ее нормали \mathbf{N} , указывающим направление лучей. Рассмотрим луч, падающий в точку I , расположенную на расстоянии r_0 от оси (рис. 8.7). Обозначим через Θ угол между векторами \mathbf{u}_z и \mathbf{N} , внешний по отношению к волокну. Плоскость падения в точке I проходит через векторы, эквивалентные векторам \mathbf{u}_z и \mathbf{N} . Угол преломления луча в точке I следующим образом выражается через локальный показатель преломления $n(r_0)$:

$$\sin \Theta = n(r_0) \sin \theta_0. \quad (8.10)$$

Тогда константы β и ν , соответствующие лучу, падающему в точке I , принимают вид

$$\beta = k \sqrt{n^2(r_0) - \sin^2 \Theta}, \quad (8.11)$$

$$\nu = k \sin \Theta \, r_0 \sin \varphi_0. \quad (8.12)$$

Выражения (8.11) и (8.12) позволяют сделать следующие выводы (рис. 8.7).

1. В волокне со скачкообразным изменением показателя преломления $n(r_0) = n_1$ плоская волна создает во всех точках входного торца лучи с одной и той же постоянной распространения β , определяемой формулой (8.11). Семейство прямых, касательных к одной и той же окружности радиусом r_1 , дает совокупность точек с одним и тем же параметром v , определяемым формулой

$$v = kr_1 \sin \Theta, \quad (8.13)$$

где $r_1 = r_0 \sin \varphi_0$. При заданном угле Θ можно из любой точки, удовлетворяющей условию $r_0 > r_1$, провести две такие прямые,

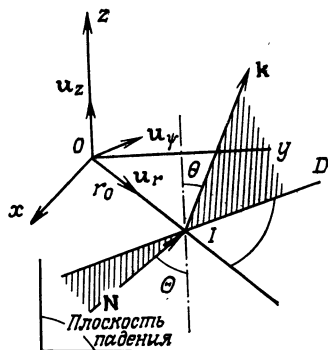


Рис. 8.7. Начальные условия.

Все точки прямой D удовлетворяют уравнению $r \sin \varphi = \text{const}$. Следовательно, при заданном значении внешнего угла Θ всем точкам этой прямой соответствует один и тот же параметр v , а точки, соответствующие одной и той же постоянной распространения β , лежат на окружности $r = r_0$.

которые представляют собой следы плоскостей падения на входном торце. Следовательно, одной точке на входе волокна соответствуют четыре луча с заданными параметрами β и v .

2. В волокне с градиентом показателя преломления точки, в которых одинаковый показатель преломления, лежат на одной окружности. Следовательно, плоская волна создает в волокне такого типа лучи с разными постоянными распространения β , соответствующими окружностям $r = r_0$, и интерпретация параметра v остается неизменной.

Мы еще вернемся к начальным условиям, когда речь пойдет об условиях на границах волокна, которым луч должен удовлетворять при распространении в волокне. Из всех вводимых в волокно лучей только небольшое их число будет удовлетворять условию распространения, поскольку большая часть лучей будет преломляться при первом падении на границе $r = a$.

3. Условия на границе двух сред

Согласно сказанному в гл. 3, волна, введенная в сердцевину волокна, будет удерживаться в ней за счет полного внутреннего отражения при некоторых условиях падения и определенной разности показателя преломления [113—118]. Полное внутреннее отражение выражается в следующем (гл. 3, § 1, п. В):

а) существует «предельный» угол падения θ_c на границе поверхностей сердцевина — оболочка;

б) имеется поток лучистой энергии вдоль оси Oz и поток реактивной энергии в радиальном направлении, причем амплитуда поля убывает по экспоненте.

Когда показатель преломления меняется непрерывно, полное внутреннее отражение приводит к явлению «миража»: световой луч искривляется в направлении увеличения показателя преломления с плавным изменением угла распространения. Однако волновые процессы, связанные с полным внутренним отражением, не меняются: появляется затухающая волна. Если она проходит сквозь диоптр, за которым снова становится активной, то некоторая часть энергии, первоначально распространяющейся в сердцевине, может передаваться в эту новую среду (как мы это увидим в гл. 14, где речь пойдет об устройствах ввода). Это будет наблюдаться в случае лучей утечки. Мы можем проследить за движением луча, рассматривая его радиальную компоненту $f(r)$, поскольку β и γ — действительные величины: 1) если радиальная составляющая $f(r)$ — величина действительная ($f^2(r) \geq 0$), то луч тоже действителен; 2) если же величина $f^2(r)$ отрицательна, то радиальная составляющая становится мнимой величиной и соответствующий луч будет комплексным (затухающая волна).

Значения r , при которых $f^2(r)$ обращается в нуль, представляют собой критические точки, в которых наблюдается явление полного внутреннего отражения. Если величина $f^2(r)$ остается положительной между двумя значениями r_1 и r_2 , то это значит, что луч колеблется между двумя цилиндрами, удовлетворяющими уравнениям $r = r_1$ и $r = r_2$. Здесь можно видеть аналогию с собственными колебаниями в объемном резонаторе и явлением стоячей волны: мы снова сталкиваемся с принципом квантования, который приводит к появлению мод.

Таким образом, нам нужно подробнее исследовать соотношение (8.7), которое описывает зависимость $f^2(r)$ от r и позволяет найти области, где луч, соответствующий константам β и γ , заданным начальными условиями, оказывается действительным.

Если единственной областью, где луч действителен, является сердцевина, то луч считается *распространяющимся* в волокне. Если же луч оказывается действительным в некоторой части оболочки, то он *распространяется с потерями* (с утечкой). И наконец, в том случае, когда луч действителен во всем объеме оболочки, мы имеем дело с *преломлением лучей* (а не распространением). Одно простое замечание поможет понять данное явление. Величина $f^2(r)$ равна разности двух функций, каждая из которых зависит только от одного параметра: $f_\beta(r) = k^2 n^2 - \beta^2$ и $f_\gamma(r) = v^2/r^2$. Когда постоянная β меняет значение, кри-

вые $f_\beta(r)$ смещаются по вертикали (рис. 8.8), а в оболочке функция $f_\beta(r)$ сохраняет постоянное значение, равное $f_\beta(a) = k^2 n_2^2 - \beta^2$. Функция $f_v(r)$ — монотонно убывающая и положительная (рис. 8.8). Относительное положение обеих функций при $r = a$ дает следующие характеристики явления в оболочке:

1) если $f_\beta(a) \leq 0$, то $f_\beta(r) < f_v(r)$, какой бы не была величина $r \geq a$, это случай распространения лучей, нет лучей потерь (утечки);

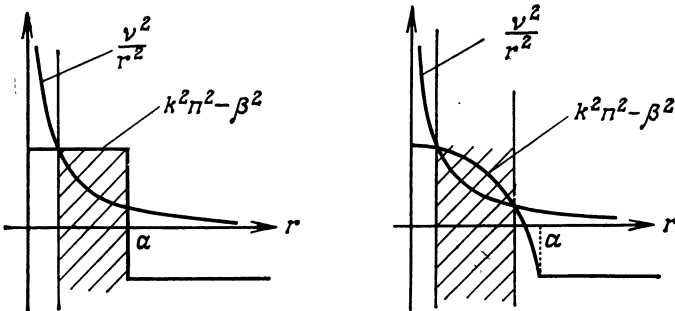


Рис. 8.8. Характер изменения функций $f_\beta(r)$ и $f_v(r)$ для волокна со скачкообразным изменением показателя преломления (слева) и градиентного волокна (справа).

В заштрихованной области разность $f_\beta - f_v$ положительна, и луч, который в ней распространяется, действителен.

2) если $f_\beta(a) > 0$, то следует различать два случая:

а) если $f_v(a) < f_\beta(a)$, то такая ситуация сохраняется при $r \geq a$, т. е. луч преломляется;

б) если $f_v(a) > f_\beta(a)$, то существует такое значение $r = r_3$, что $f_\beta > f_v$ при $r \geq r_3$.

Таким образом, характерным параметром является постоянная распространения β . Согласно формуле (8.9), величина β меньше произведения kn_1 (постоянной распространения плоской волны в среде с показателем преломления n_1). Для обеспечения режима распространения должно выполняться также условие $f_\beta(a) \leq 0$, так что постоянная распространения β должна лежать в пределах

$$\beta_c = kn_2 \leq \beta \leq kn_1. \quad (8.14)$$

Зона, в которой луч действителен, ограничивается цилиндрическими поверхностями $r = r_1$ и $r = r_2$ (заштрихована на рис. 8.9). Если β меньше величины kn_2 , называемой критической, то величина $f_\beta(a)$ будет положительной: в зависимости от параметра v величина $f_v(a)$ будет либо меньше $f_\beta(a)$ и соответствующий луч преломится при первом падении на границу

$r = a$, либо больше $f_\beta(a)$, и тогда существует такое значение r_3 , что луч снова становится действительным при $r > r_3$. В слое $r_2 \leq r \leq r_3$ луч будет комплексным (рис. 8.9).

С энергетической точки зрения здесь можно видеть аналогию с туннельным эффектом [119]: энергия из сердцевины может проникать в оболочку сквозь зону $r_2 < r < r_3$. Такие лучи называются *лучами утечки*, поскольку при распространении они постоянно создают потери энергии. На очень больших расстояниях вся энергия, имеющаяся в световоде, переносится лучами распространения, но на малых расстояниях, не превышающих

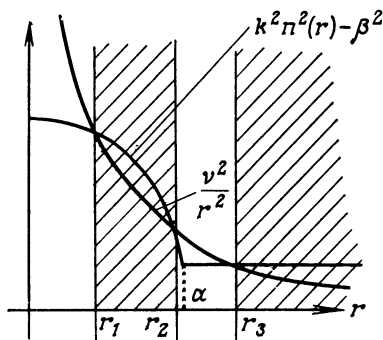


Рис. 8.9. Случай $\Theta > \Theta_c$, или $\beta < \beta_c$.

Если $\beta < \beta_c = kn_2$, то функция $f_\beta(r)$ при $r = a$ положительна. Следовательно, можно найти такие значения v , при которых соответствующий луч будет действительным при $r_1 < r < r_2$ и $r > r_3$ (заштрихованные области).

нескольких сотен метров, может быть существенной энергия, которую несут лучи утечки. На практике расстояние, на котором существенна роль лучей утечки, уменьшается из-за дефектов границы раздела и оптической оболочки (и из-за того, что ее толщина не бесконечна).

Рассмотрим отдельно волокна со скачком показателя преломления и градиентные волокна.

4. Оптические волокна со скачкообразным изменением показателя преломления

На основании формулы (8.14) можно найти максимальное значение угла θ_c из соотношения, за пределами которого постоянная распространения β меньше критического значения $\beta_c = kn_2$ и распространение света в световоде становится невозможным,

$$\beta_c = kn_1 \cos \theta_c = kn_2. \quad (8.15)$$

Этому углу соответствует угол падения, внешний по отношению к волокну Θ_c , определяемый равенством (8.10)

$$\sin \Theta_c = n_1 \sin \theta_c = \sqrt{n_1^2 - n_2^2}. \quad (8.16)$$

Величина $\sin \Theta_c$ называется *числовой апертурой волокна*. В случае оптического волокна со скачкообразным изменением пока-

зателя преломления числовой апертурой определяется максимальный угол ввода в волокно луча, распространяющегося без потерь. Ее часто обозначают буквами ЧА.

При значении β , удовлетворяющем условию (8.14), т. е. для распространяющегося луча, максимальное значение, которое может принять величина v при $r = a$, таково (рис. 8.10):

$$v_M = kn_1 a \sin \theta = ka \sin \theta. \quad (8.17)$$

Минимальное (по абсолютной величине) значение — это $v = 0$, что соответствует меридиональным лучам. При заданном значении v расстояние до оси луча проходит через минимум r_1 при $\varphi = \pi/2$, причем в точке $r = r_1$ производная dr/dz

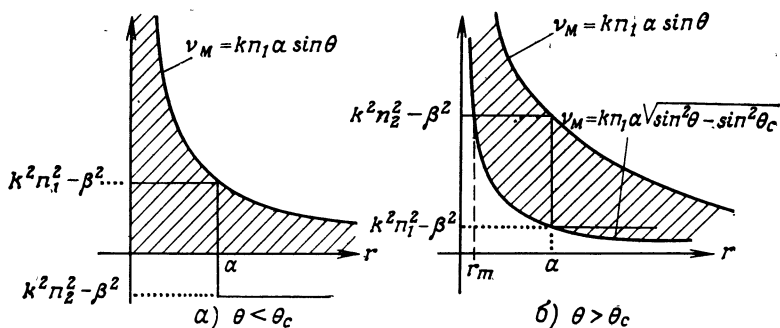


Рис. 8.10. Случай $\theta < \theta_c$ (или $\theta > \theta_c$) для волокна со скачкообразным изменением показателя преломления.

При $\theta < \theta_c$ всем значениям v , удовлетворяющим неравенству $0 \leq |v| \leq v_M$, соответствует действительный луч. При $\theta > \theta_c$ значениям v , удовлетворяющим неравенству $0 \leq |v| \leq v_m$, соответствует преломленный луч, а значениям v в интервале $v_m \leq |v| \leq v_M$ — луч потерь.

меняет знак (экстремум величины r). Таким образом, на поверхности $r = r_1$ происходит как бы полное внутреннее отражение, и эта поверхность общая для всех лучей с одинаковыми параметрами β и v . Можно говорить о *каустике* пучка или *конгруэнтности лучей* [118].

Если постоянная распространения β меньше β_c (случай $\theta > \theta_c$ на рис. 8.10), то меридиональные лучи перестают распространяться. Распространяются только лучи утечки, соответствующие значениям v , превышающим v_m (рис. 8.10), а при $0 < v < v_m$ лучи преломляются. В плоскости падения происходит разделение лучей по окружности радиусом r_m , который определяется двумя равенствами

$$v_m^2 = a^2 (k^2 n_2^2 - \beta^2) = r_m^2 k^2 n_1^2 \sin^2 \theta.$$

Отсюда находим

$$r_m = a \sqrt{1 - \frac{\sin^2 \Theta_c}{\sin^2 \Theta}}. \quad (8.18)$$

Мы можем ввести луч утечки в некоторой точке входного торца только в том случае, если произведение $r_0 \sin \varphi_0$ больше r_m (при заданном Θ). При заданном значении Θ этим условием

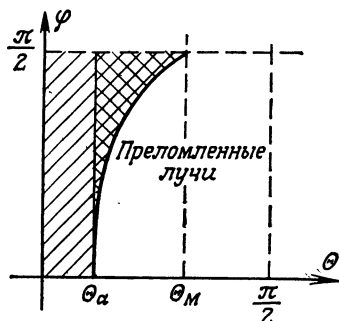


Рис. 8.11. Зависимость угла φ наклона луча в плоскости падения от угла Θ .

При $\Theta > \Theta_M$, где Θ_M — угол, определяющийся выражением

$$\sin \Theta_M = \frac{\sin \Theta_a(r_0)}{\sqrt{1 - r_0^2/a^2}},$$

все лучи будут преломленными. При $\Theta_a < \Theta < \Theta_M$ и $\varphi > \varphi_m$ будем иметь луч потерь (двойная штриховка). Простая штриховка — лучи распространения.

ограничивается полезный конус ввода лучей, поскольку угол φ должен превышать значение (рис. 8.11)

$$\varphi_m = \arcsin \left(\frac{r_m}{r_0} \right), \quad (8.19)$$

5. Волокна с градиентом показателя преломления

Мы знаем, что при одном и том же угле наклона Θ лучей на входе волокна постоянная распространения β будет зависеть от расстояния от точки падения до оси r_0 [формулы (8.9), (8.11), (8.12)]. Луч будет распространяться в волокне, если величина $f_\beta(r)$ отрицательна или равна нулю при $r = a$. С учетом формулы (8.11) мы можем представить это условие в виде

$$\sin^2 \Theta \leq n^2(r_0) \left(1 - \frac{n^2(a)}{n^2(r_0)} \right). \quad (8.20)$$

По аналогии со случаем волокон со скачкообразным изменением показателя преломления примем

$$\cos [\Theta_a(r_0)] = \frac{n(a)}{n(r_0)}. \quad (8.21)$$

Синус угла Θ_a будем называть *локальной числовой апертурой* волокна в точке с радиусом r_0 ; эта величина дается выражением

$$\sin [\Theta_a(r_0)] = n(r_0) \sin [\Theta_a(r_0)] = \sqrt{n^2(r_0) - n^2(a)}. \quad (8.22)$$

Такое определение вполне совместимо с определением (8.16), и любой луч, падающий в точку $r = r_0$ и попадающий внутрь

апертурного конуса $\Theta_a(r_0)$, распространяется после ввода в волокно. Величина $\Theta_a(r_0)$ называется *углом ввода* излучения в волокно в точке $r = r_0$, а *числовой апертурой* волокна при этом называется максимальное значение локальной числовой апертуры, т. е.

$$\text{ЧА} = \sin \Theta_a(0) = \sin \Theta_c.$$

Посмотрим теперь, что происходит при $r = r_0$, когда угол Θ больше угла ввода излучения в волокно $\Theta_a(r_0)$. Тогда при $r = a$ функцию $f_\beta(r)$ можно записать в виде $k^2(\sin^2 \Theta - \sin^2 \Theta_a)$, а чтобы луч, соответствующий значению v_0 параметра v , распространялся с утечкой, значение v_0 должно быть больше минимального критического значения v_m , задаваемого равенством $f_\beta(a) = f_v(a)$, т. е. (рис. 8.12)

$$v_m^2 = a^2 k^2 [\sin^2 \Theta - \sin^2 \Theta_a(r_0)]. \quad (8.23)$$

Сравним этот результат с формулой (8.18). Примем по аналогии

$$r_m = a \sqrt{1 - \frac{\sin^2 \Theta_a(r_0)}{\sin^2 \Theta}}, \quad (8.24)$$

что позволяет интерпретировать (8.23) геометрически. Если произведение $r_0 \sin \varphi_0$ меньше r_m , т. е. след плоскости падения пересекает окружность радиусом r_m , то соответствующий световой луч преломляется (рис. 8.13, D_1 — след плоскости падения), а в противном случае он становится лучом утечки. Угол ввода φ_0 должен быть больше угла $\varphi_m = \arcsin(r_m/r_0)$; таким образом, формула имеет тот же вид, что и для волокон со скачкообразным изменением показателя преломления.

Обобщим результаты, касающиеся светового луча, падающего в точку с $r = r_0$.

1. Если угол падения Θ меньше угла ввода излучения в волокно $\Theta_a(r_0)$, то луч распространяется в волокне при любом значении параметра v (или угла наклона φ).

2. Если угол падения больше угла ввода излучения в волокно, то при значениях v , лежащих между 0 и v_m , лучи преломляются; при значениях, больших v_m , они распространяются с потерями. Заметим, что этот последний случай будет невозможен, если кривая $f_v(r)$ не пересекает $f_\beta(r)$ иначе, как в точке с $r = a$.

При заданном Θ параметр φ_0 характеризует наклон плоскости падения относительно начального радиус-вектора. Потребовать, чтобы параметр v был больше v_m , равнозначно требованию $r_0 \sin \varphi_0 > r_m$. Ясно, что при $r_m > r_0$ это условие не выполняется: при этом значении v не существует луча утечки, исходящего из $r = r_0$.

Теперь мы сможем различать среди световых лучей, падающих на входной торец, лучи, *распространяющиеся без потерь*, лучи, *распространяющиеся с потерями*, и *преломляемые* лучи. Таким образом, мы можем просто рассчитать полезную мощность, вводимую в оптическое волокно неким источником света.

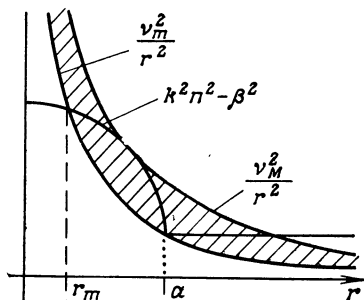


Рис. 8.12. Для лучей, распространяющихся с потерями, параметр v лежит между v_m и v_M .

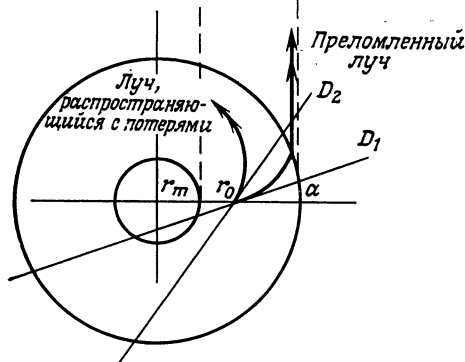


Рис. 8.13. Предельная окружность, разделяющая точки входного торца волокна.

При заданном θ [$\theta > \theta_a(r_0)$] любая точка внутри окружности радиусом r_m дает только преломленные лучи. Точка сердцевины, лежащая вне окружности, дает луч утечки, если прямая D , касательная к лучу в точке его возникновения, не пересекается с этой окружностью.

Возьмем простой случай ламбертова источника с яркостью в точке $r = r_0$, равной $B_0 \cos \theta$. В этой точке мощность, вводимая в волокно и распространяющаяся без потерь, равна

$$P_G(r_0) = 2\pi \int_0^{\theta_n(r_0)} B_0 \cos \theta \sin \theta d\theta = \pi B_0 \sin^2 \theta_a(r_0).$$

Она пропорциональна *квадрату локальной числовой апертуры*.

Вводимая мощность, которая приходится на лучи, распространяющиеся с утечкой, равна

$$P_F(r_0) = 4 \int_{\theta_a(r_0)}^{\theta_M} B_0 \cos \theta \sin \theta \int_{\varphi_m}^{\pi/2} d\varphi d\theta$$

(угол θ_M показан на рис. 8.11)

Мощность преломленных лучей равна $P_{\text{полн}} - P_G - P_F$.

Б. Траектория световых лучей

Выше мы говорили о лучах, вводимых в оптическое волокно. Теперь скажем несколько слов о самих траекториях [120].

1. Волокна со скачкообразным изменением показателя преломления

Показатель преломления сердцевинны есть постоянная величина, равная n_1 . Из (8.9) следует, что угол распространения θ сохраняет свое абсолютное значение и что произведение $r \sin \phi$ постоянно и, согласно (8.13), равно r_1 . Таким образом, между двумя последовательными полными внутренними отражениями траектория луча прямолинейная (рис. 8.14).

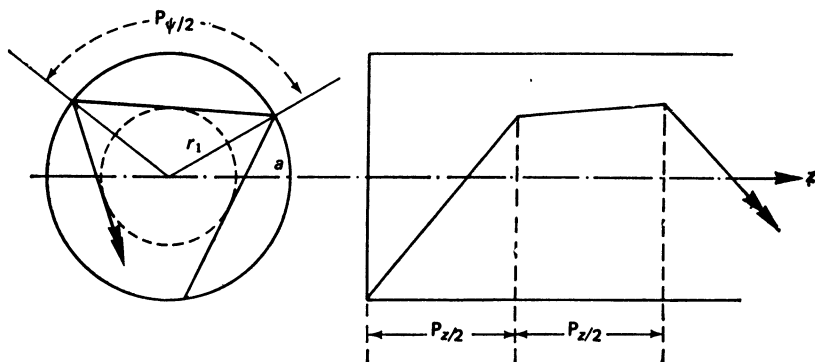


Рис. 8.14. Траектория распространения светового луча в волокне со скачкообразным изменением показателя преломления.

Траектория состоит из равных отрезков, получаемых один из другого путем смещения на $P_z/2$ и поворота на $P_\psi/2$ вокруг оси Oz . В проекции все они касаются одной и той же окружности радиусом r_1 . Величины P_z и P_ψ можно вычислить следующим образом [126]:

$$P_z = 4 \int_{r_1}^a dz = 2 \int_{r_1}^a \frac{\beta dr}{\sqrt{k^2 n_1^2 - \beta^2 - \frac{v^2}{r^2}}} = \frac{2a\beta}{k^2 n_1^2 - \beta^2} \sqrt{k^2 n_1^2 - \beta^2 - \frac{v^2}{a^2}}, \quad (8.25)$$

$$P_\psi = 4 \int_{r_1}^a d\psi = 2 \arccos \left[\frac{v}{a \sqrt{k^2 n_1^2 - \beta^2}} \right]. \quad (8.25a)$$

В случае меридиональных лучей ($\nu = 0$) окружность радиусом r_1 сжимается в точку $r = 0$, траектория становится периодической с периодом $P_z = 2a \operatorname{ctg} \theta$, а период вращения равен $P_\psi = 2\pi$.

2. Волокна с градиентом показателя преломления

Возьмем для примера случай, когда показатель преломления волокна изменяется по степенному закону [177]

$$n(r) = n_1 \sqrt{1 - 2\Delta \left(\frac{r}{r_1}\right)^\alpha}, \quad (8.26)$$

где Δ — относительная разница показателей преломления сердцевин и оболочки (в первом приближении), n_1 — показатель преломления сердцевины на оси волокна, а α — параметр, который может быть равен любому числу от 1 до бесконечности. При $\alpha = 2$ имеем закон псевдопараболического градиента¹⁾. Согласно формуле (8.22), числовая апертура волокна, описываемого формулой (8.26), максимальна на оси и равна

$$\sin \Theta_c = n_1 \sqrt{2\Delta}. \quad (8.27)$$

Подставив (8.26) в (8.5), получим

$$\frac{d^2 \rho}{dz^2} = \frac{2k^2 n_1^2 \Delta}{\beta^2 a^2} \rho. \quad (8.28)$$

Конец вектора ρ , удовлетворяющего уравнению (8.28), описывает эллипс, который в декартовых координатах, совпадающих с его осями, записывается с учетом начальных условий $(r_0, \theta_0, \varphi_0)$ [118] в виде

$$\begin{cases} X = A \cos \frac{\varphi_0}{2} \cos (\Omega z - \varphi), \\ Y = A \sin \frac{\varphi_0}{2} \sin (\Omega z - \varphi), \end{cases} \quad (8.29)$$

где $\Omega = \frac{kn_1 \sqrt{2\Delta}}{\beta a} = \frac{V}{a^2 \beta}$, V — приведенная частота [формула (3.37)], а

$$A = \sqrt{r_0^2 + \frac{\operatorname{tg}^2 \theta_0}{\Omega^2}},$$

$$\cos \varphi = \frac{r_0}{A}, \quad \sin \varphi = \frac{\operatorname{tg} \theta_0}{\Omega A}.$$

¹⁾ Псевдопараболического, так как в весьма распространенном случае малого Δ формула (8.26) при $\alpha = 2$ в первом приближении дает параболический закон изменения (гл. 4, § 3, п. В).

Чтобы соответствующий луч распространялся в волокне, большая ось эллипса не должна превышать размера сердцевины a . Это условие можно выразить через β и v :

$$k^2 n_1^2 - \beta^2 - \frac{v^2}{a^2} \leq \frac{V^2}{a^2}. \quad (8.30)$$

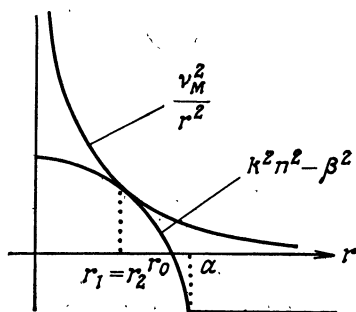
Неравенство (8.30) можно переписать в другом виде, допускающем простую геометрическую интерпретацию:

$$\sin \Theta_0 \sqrt{1 - \left(\frac{r_0 \sin \Phi_0}{a} \right)^2} \leq \sin \Theta_a(r_0). \quad (8.30a)$$

Поскольку корень квадратный здесь не превышает единицы, то имеется возможность вводить лучи с превышением предельного

Рис. 8.15. Определение максимального значения параметра v .

При $v=v_M$ кривые $f_\beta(r)$ и $f_v(r)$ касаются друг друга (в точке $r_1=r_2$); r_0 — то значение радиуса r , при котором $f_\beta(r)=0$.



угла $\Theta_a(r_0)$, которые будут распространяться с утечкой или преломляться. Найдем в данном частном случае изменения показателя преломления максимальный внешний угол ввода, при котором луч не будет преломляться. Он соответствует значению $\Phi_0 = \pi/2$, и тогда (8.30) принимает следующий вид (рис. 8.4):

$$\sin \Theta_{0 \text{ макс}} = n(r_0) \sin \theta_{0 \text{ макс}} = n_1 \sqrt{2\Delta} = \sin \Theta_c.$$

Это выражение показывает, что любой луч с углом падения Θ , превышающим Θ_c , преломляется, а лучи, падающие под углами от $\Theta_a(r_0)$ до Θ_c , будут либо лучами утечки, либо преломляться в зависимости от величины v (т. е. от положения следа плоскости падения относительно окружности радиусом r_m). Равенство предельного угла Θ_c и максимального угла ввода лучей утечки (рис. 8.15) объясняется тем, что в случае псевдопараболического закона изменения показателя преломления крутизна кривых v_M^2/r^2 и $k^2 n^2(r) - \beta^2$ одинакова при $r = a$. В случае когда закон изменения показателя преломления имеет вид (8.26) с показателем степени α , лежащим в пределах от единицы до бесконечности, максимальный угол ввода лучей утечки зависит от значения α ; при $\alpha = \infty$, как мы знаем (§ 3, п. А, 4), он равен $\pi/2$.

В частном случае $\alpha = 2$, рассмотренном выше, имеется ряд других интересных особенностей: а) траектория луча в проекции представляет собой замкнутую кривую (эллипс), а период вращения равен 2π ; б) период изменения вдоль оси Oz составляет $2\pi/\Omega = 2\pi a^2 \beta / V$. Заметим, что он зависит только от $\beta(r_0)$ (рис. 8.16) и не зависит от v . Отметим также, что оптическая длина пути луча в одном периоде не зависит от v .

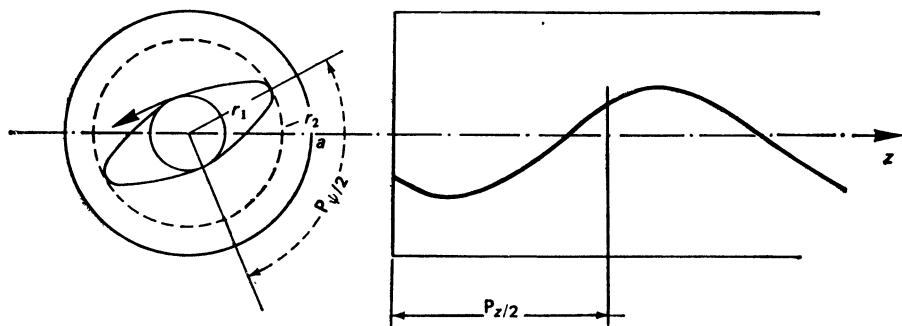


Рис. 8.16. Траектория распространения луча в градиентном волокне. Расстояние от луча до оси волокна периодически изменяется от r_1 до r_2 .

При $\alpha \neq 2$ и тем более при других законах изменения показателя преломления проекция траектории луча, вообще говоря, не дает замкнутую кривую, но все же можно определить псевдопериоды P_z и P_ψ . По соответствующим отрезкам кривой в этом случае можно построить всю траекторию луча, смещая их вдоль оси Oz и поворачивая вокруг нее (рис. 8.16).

§ 4. Моды распространения света

В случае световых лучей, распространяющихся в идеальном оптическом волокне, которое мы рассматривали, потери отсутствуют. И мы установили условия ввода, нашли постоянные распространения, траектории, т. е. геометрическая оптика вроде бы позволяет находить параметры, необходимые для исследования. Но если вернуться к гл. 3, то мы вспомним, что наряду с волной, распространяющейся в сердцевине, существует волна в оболочке, имеющая ту же фазу, т. е. оболочка участвует в распространении света, и если меняются условия распространения в оболочке, то также изменяется распространение света в сердцевине. Например, если у двух сред разный коэффициент поглощения (а так обычно и бывает), то в одной из них волна будет затухать быстрее и распространение будет нарушено!

Даже при одинаковом поглощении чисто геометрическая оптика не позволяет оценивать потери, обусловленные лучами

утечки; для этого нужно знать выражение для поля при $r \geq a$. Таким образом, в некоторых случаях необходимо рассчитывать электромагнитное поле. Обычно стараются найти некий базис, пригодный для разложения в ряд любого поля. Такой базис нам дает теория мод. Необходимые вычисления будут проводиться на основе результатов, полученных в гл. 3.

А. Уравнение дисперсии, расчет числа мод

Чтобы распространяющиеся в волокне лучи дали моду распространения, они должны так накладываться друг на друга, чтобы на прямом сечении волокна была сформирована стоячая волна как в радиальном направлении, так и по окружности.

Необходимая периодичность возникает в том случае, если для составляющей u_ψ волнового вектора \mathbf{k} на окружности радиусом r укладывается целое число периодов, т. е. если v — *целое число*. В самом деле, согласно (8.6), имеем

$$\int_0^{2\pi} u_\psi \cdot \mathbf{k} r d\psi = \int_0^{2\pi} \frac{v}{r} r d\psi = 2\pi v = 2\pi m. \quad (8.31)$$

Аналогичным образом и для радиальной составляющей на расстоянии между точками обращения r_1 и r_2 должно укладываться целое число полупериодов, к которым следует добавить изменение фазы при полном внутреннем отражении на каустиках:

$$p\pi = \int_{r_1}^{r_2} f(r) dr + \frac{\Phi_{r_1}}{2} + \frac{\Phi_{r_2}}{2}. \quad (8.32a)$$

Это уравнение представляет собой характеристическое уравнение мод распространения в волокне. Мы получаем в двумерном пространстве ситуацию, сходную с описанной в гл. 3: моды распространения получены путем квантования введенных констант, связанных с условиями на границах изучаемого световода. Для мод высокого порядка можно пренебречь фазовыми изменениями на каустике и несколько упростить уравнение. Иную, тоже интересную форму уравнения (8.32a) можно получить, если ввести приведенную частоту V и приведенную фазу B , определяемую выражением

$$B = \frac{(\beta/k)^2 - n_2^2}{n_1^2 - n_2^2}.$$

В частности, в случае профилей показателя преломления вида $n(r) = n_1 \sqrt{1 - 2\Delta g(r/a)}$, где $g(x)$ есть функция переменной

х, мы получаем

$$p\pi = \frac{1}{a} \int_{r_1}^{r_2} \left\{ V^2 \left[1 - g\left(\frac{r}{a}\right) \right] - B - \frac{v^2 a^2}{r^2} \right\}^{1/2} dr. \quad (8.32б)$$

При $p = p_c$ постоянная распространения β равна kn_2 , параметр B обращается в нуль и (8.32б) принимает упрощенный вид

$$p_c \pi = \frac{1}{a} \int_{r_1}^{r_2} \left\{ V^2 \left[1 - g\left(\frac{r}{a}\right) \right] - \frac{v^2 a^2}{r^2} \right\}^{1/2} dr. \quad (8.32в)$$

В случае волокон со скачком показателя преломления $[n(r) = n_1]$ и градиентных волокон с псевдопараболическим законом изменения показателя преломления $[g(r/a) = r^2/a^2]$ эти уравнения интегрируются без трудностей. В случае профиля $g(r/a) = (r/a)^\alpha$ с произвольным значением показателя степени α найдено приближенное решение уравнения (8.32), которое остается довольно точным, пока показатель степени α не слишком велик [121]; оно имеет вид

$$(k^2 n_1^2 - \beta^2)^{\frac{\alpha+2}{\alpha}} \sqrt{\frac{\alpha}{\alpha+2}} \left(\frac{2}{\alpha+2} \right)^{\frac{1}{\alpha}} = \left[v + 4 \left(\frac{2}{2+\alpha} \right)^{\frac{2}{\alpha}} p_c \right] V^{\frac{2}{\alpha}}.$$

Попытаемся теперь найти *число мод*, которые передает *многомодовое* волокно. На основании приближенной зависимости, вытекающей из (8.32) и (8.7), а именно:

$$p = \frac{1}{\pi} \int_{r_1}^{r_2} \left(k^2 n^2 - \beta^2 - \frac{v^2}{r^2} \right)^{1/2} dr, \quad (8.33)$$

мы можем сопоставить некоторым значениям v и β некоторое значение величины p ; в действительности существуют $4p$ моды, если учитывать поляризацию и круговую симметрию ($v| - v$) (§ 3, п. А, 1). Отметим, что при одном и том же значении величины v модам, для которых величина β больше, чем следует из уравнения (8.33), будет также соответствовать меньшее значение p . Таким образом, мы можем сопоставить величине β число мод $N(\beta)$, имеющих большую, чем β , постоянную распространения, сложив все числа p при изменении v от 0 до v_m . При этом максимальное значение v_m получается, когда оба корня r_1 и r_2 одинаковы (рис. 8.15). Поскольку величина v весьма быстро возрастает, мы рассматриваем ее как непрерывную переменную. Отсюда следует [117], что

$$N(\beta) = \frac{4}{\pi} \int_{v=0}^{v_m} dv \int_{r_1}^{r_2} \left(k^2 n^2 - \beta^2 - \frac{v^2}{r^2} \right)^{1/2} dr. \quad (8.34)$$

Изменяя порядок интегрирования, получаем

$$N(\beta) = \int_0^{r_0} [k^2 n^2(r) - \beta^2] r dr.$$

Как нетрудно видеть, величина $N(\beta)$ стремится к нулю, когда β приближается к своему максимальному значению. Общее число мод получается при $\beta = kn_2$:

$$N_t = \int_0^a (k^2 n^2 - k^2 n_2^2) r dr. \quad (8.35)$$

Итак, имеем:

а) для волокна со скачкообразным изменением показателя преломления

$$n(r) = n_1, \quad N_t = k^2 (n_1^2 - n_2^2) \frac{a^2}{2} = \frac{V^2}{2};$$

б) для градиентного волокна с псевдопараболическим профилем

$$N_t = V^2/4;$$

в) для градиентного волокна с произвольным показателем степени α

$$N_t = \frac{\alpha}{2(\alpha + 2)} V^2.$$

Эти результаты — лишь уточнение того, что уже было сказано в § 3.

Б. Формулы для полей

Если предположить, что среда изотропна и в ней нет ни токов, ни зарядов, то из уравнений Максвелла вытекает следующее уравнение распространения электромагнитной волны [формула (2.1)]:

$$\Delta \psi - \epsilon \mu \frac{\partial^2 \psi}{\partial t^2} = 0,$$

где вектор ψ может быть как электрической \mathbf{E} , так и магнитной \mathbf{H} составляющей поля. Учитывая геометрию волокон, мы будем пользоваться цилиндрической системой координат, и из уравнения (2.1) можно вывести систему шести скалярных уравнений. Поперечные составляющие можно выразить только через продольные составляющие E_z и H_z , которые в общем случае обе не нулевые (в таком случае мы имеем *гибридные моды*, ни ТЕ, ни ТМ). Будем искать решение в виде гармонических функций

переменных t и z :

$$\psi_z = \psi_1(r) \psi_2(\varphi) \exp i(\omega t - \beta z), \quad (8.36)$$

где β — составляющая вектора распространения по оси Oz .

Запишем поперечные составляющие, полученные путем проекции в цилиндрических координатах r , φ , z . Разделение переменных r и φ дает зависимость $\psi_2(\varphi)$ вида $\exp(i\nu\varphi)$, а радиальная зависимость $\psi_1(r)$ должна удовлетворять уравнению

$$\frac{d^2\psi_1}{dr^2} + \frac{1}{r} \frac{d\psi_1}{dr} + \left(k^2 n^2 - \beta^2 - \frac{\nu^2}{r^2}\right) \psi_1 = 0, \quad (8.37)$$

причем составляющие E_r , E_φ , H_r и H_φ записываются следующим образом [120]:

$$E_r = -\frac{i}{\kappa^2} \left(\beta \frac{\partial E_z}{\partial r} + \omega \frac{\mu}{r} \frac{\partial H_z}{\partial \varphi} \right), \quad E_\varphi = \frac{-i}{\kappa^2} \left(\frac{\beta}{r} \frac{\partial E_z}{\partial \varphi} - \omega \mu \frac{\partial H_z}{\partial r} \right), \quad (8.38)$$

$$H_r = -\frac{i}{\kappa^2} \left(\beta \frac{\partial H_z}{\partial r} - \omega \frac{\mu}{r} \frac{\partial E_z}{\partial \varphi} \right), \quad H_\varphi = \frac{-i}{\kappa^2} \left(\frac{\beta}{r} \frac{\partial H_z}{\partial \varphi} + \varepsilon \omega \frac{\partial E_z}{\partial r} \right),$$

где $k = 2\pi/\alpha$, $\kappa^2 = k^2 n^2 - \beta^2$.

Решение, найденное для E_z и H_z , должно отвечать *граничным условиям* рассматриваемого световода. Отметим, что коэффициент ψ_1 в формуле (8.37) есть не что иное, как квадрат радиальной составляющей волнового вектора, определяемого выражением (8.7).

Поясним наши расчеты в случае оптического волокна со скачкообразным изменением показателя преломления.

1. Решение волнового уравнения для волокна со скачком показателя преломления

Поскольку показатель преломления n представляет собой одну константу в сердцевине и другую — в оболочке, уравнение (8.7) оказывается дифференциальным уравнением Бесселя [122]. Его решения для разных областей волокна записываются следующим образом [115]:

$$\text{при } r < a \quad E_z = A J_\nu \left(u \frac{r}{a} \right) e^{i\nu\varphi}, \quad (8.39a)$$

$$H_z = B J_\nu \left(u \frac{r}{a} \right) e^{i\nu\varphi},$$

$$\text{где } u^2 = (k^2 n_1^2 - \beta^2) a^2;$$

$$\text{при } r > a \quad E_z = C K_\nu \left(w \frac{r}{a} \right) e^{i\nu\varphi}, \quad (8.39b)$$

$$H_z = D K_\nu \left(w \frac{r}{a} \right) e^{i\nu\varphi},$$

$$\text{где } w^2 = (\beta^2 - k^2 n_2^2) a^2.$$

Функции K_v — это модифицированные функции Бесселя [122]. Выражения (8.39) сходны с найденными в гл. 3, § 3 для случая плоского световода. Можно заметить, что величина $V^2 = u^2 + w^2$ есть характеристическая постоянная световода (поскольку в нее входят только радиус сердцевины a и показатели преломления сердцевины n_1 и оболочки n_2).

Постоянные A , B , C и D нельзя определить из уравнений Максвелла. Для того чтобы полученные решения представляли собой моды волокна, поля должны отвечать условиям непрерывности при $r = a$. Эти условия дают четыре однородных уравнения, которые имеют решения только тогда, когда главный определитель обращается в нуль, что приводит к *уравнению дисперсии*. [115]

$$\left(\frac{J'_v(u)}{uJ_v(u)} + \frac{K'_v(w)}{wK_v(w)} \right) \left(\frac{k^2 n_1^2 J'_v(u)}{uJ_v(u)} + \frac{k^2 n_2^2 K'_v(w)}{wK_v(w)} \right) = \beta^2 v^2 \left(\frac{1}{u^2} + \frac{1}{w^2} \right). \quad (8.40)$$

Решения уравнения (8.40) дают совокупность дискретных значений, и при $v = 0$ уравнение распадается на две части: происходит полное разделение поперечных электрических (TE) и магнитных (TM) мод, поскольку либо поле E_z , либо поле H_z оказывается равным нулю.

Если наблюдать за поведением полей в среде 2, можно убедиться, что проникновение волны в среду тем больше, чем меньше w . В пределе при $w \rightarrow 0$ распространение света в волокне не происходит. Следовательно, при $w = 0$ мы имеем $\beta_c = kn_2$ и $u_c = V$, т. е. в среде 2 решение имеет вид плоской волны (гл. 3, § 3, п. В).

Теперь можно найти предельную форму уравнения (8.39а) на граничной частоте при $w \rightarrow 0$. Получаются разные типы решений, которые можно, обозначив гибридные моды (когда ни E_z , ни H_z не являются нулевыми) через EH и HE , следующим образом классифицировать согласно уравнению, которому удовлетворяет граничная частота:

Обозначение моды		Граничная частота
HE_{11}		0
TE_{0m}	или TM_{0m}	m -й корень уравнения $J_0(u) = 0$
HE_{1m}	или EH_{vm}	m -й корень уравнения $J_v(u) = 0$
HE_{vm}	($v \neq 1$)	m -й корень уравнения $\left(\frac{n_1^2}{n_2^2} + 1 \right) J_{v-1}(u) = \frac{u}{v-1} J_v(u)$

Итак, существует мода HE_{11} , граничная частота которой равна нулю. Следовательно, можно так подобрать параметры световода, что будет распространяться только одна эта мода; это происходит при условии

$$V = \frac{2\pi a}{\lambda} \sqrt{n_1^2 - n_2^2} < 2,405. \quad (8.41)$$

Эти результаты сопоставимы с уже полученным в гл. 3 для одномерного диэлектрического световода.

В общем случае решения сложны и требуют применения ЭВМ.

Если принять, что разность показателей преломления n_1 и n_2 мала, то полученные результаты в значительной мере упрощаются.

Введем обозначения

$$\begin{aligned} J^- &= \omega J_{\nu-1}(u) K_{\nu}(\omega), \\ J^+ &= \omega J_{\nu+1}(u) K_{\nu}(\omega), \\ K^- &= u J_{\nu}(u) K_{\nu-1}(\omega), \\ K^+ &= u J_{\nu}(u) K_{\nu+1}(\omega). \end{aligned} \quad (8.42)$$

Тогда можно показать, что уравнение на собственные значения записывается следующим образом [123]:

$$\left(\frac{n_1^2}{n_2^2} J^- - K^-\right)(J^+ + K^+) + \left(\frac{n_1^2}{n_2^2} J^+ + K^+\right)(J^- - K^-) = 0. \quad (8.43)$$

При $n_1 \approx n_2 \approx n$ оно принимает вид

$$(J^- - K^-)(J^+ + K^+) = 0,$$

откуда получаем общее уравнение для мод $HE_{\nu m}$ и $EH_{\nu m}$ [124, 125]

$$\frac{J_{\nu \mp 1}(u)}{u J_{\nu}(u)} = \pm \frac{K_{\nu \mp 1}(\omega)}{\omega K_{\nu}(\omega)}, \quad (8.44)$$

где верхний знак относится к модам HE , а нижний — к модам EH . На основании рекуррентных соотношений для функций Бесселя можно показать, что имеется вырождение между модами $EH_{\nu-2, m}$ и $HE_{\nu, m}$, так как соответствующее уравнение дисперсии одно и то же; взяв их линейную комбинацию, мы получим «псевдомоды»¹⁾, поляризованные линейно, поля которых (на этот раз в декартовых координатах) могут быть записаны

¹⁾ «Псевдомоды», поскольку они не являются точными решениями. Они, однако, дают удобный базис для разложения, хотя постоянные распространения двух связанных мод слегка различаются.

в виде

$$\begin{aligned} \text{при } r < a \quad & \begin{cases} E_x = 0; E_y = A J_v \left(u \frac{r}{a} \right) \Big|_{\sin v\varphi}^{\cos v\varphi}, \\ H_y = 0; H_x = -A n \sqrt{\frac{e}{\mu}} J_v \left(u \frac{r}{a} \right) \Big|_{\sin v\varphi}^{\cos v\varphi}; \end{cases} \\ \text{при } r \geq a \quad & \begin{cases} E_x = 0; E_y = A \frac{J_v(u)}{K_v(w)} K_v \left(w \frac{r}{a} \right) \Big|_{\sin v\varphi}^{\cos v\varphi}, \\ H_y = 0; H_x = -A n \sqrt{\frac{e}{\mu}} \frac{J_v(u)}{K_v(u)} K_v \left(w \frac{r}{a} \right) \Big|_{\sin v\varphi}^{\cos v\varphi}. \end{cases} \end{aligned} \quad (8.45)$$

Если оценить порядок величин различных составляющих, то можно видеть, что эти поля почти поперечные, поскольку составляющие E_z и H_z , выраженные через u/ka , оказываются бесконечно малыми величинами первого порядка по сравнению с поперечными составляющими.

2. Решение для градиентного волокна

Векторное уравнение (2.1) разделяется на скалярные уравнения, и мы будем искать решение для электрического поля (например, для составляющей E_y) в форме, подсказываемой полученными ранее результатами:

$$E_y = \psi_1(r) \left(\frac{\cos v\varphi}{\sin v\varphi} \right) \exp(-i\beta z), \quad (8.46)$$

где, как и ранее, v — целое число. Мы снова приходим к уравнению (8.37) для функции $\psi_1(r)$, в котором теперь n^2 уже не константа, а функция переменной r .

В случае волокна с псевдопараболическим профилем показателя преломления, не усеченным при $r = a$, решение дифференциального уравнения имеет вид [122, 125]

$$\psi_1(r) = \left(\sqrt{2} \frac{r}{\omega} \right)^\omega L_q^\omega \left(\frac{2r^2}{\omega^2} \right) \exp \left\{ -\frac{r^2}{\omega^2} \right\}, \quad (8.47)$$

где L_q^ω — обобщенный полином Лагерра порядка v и степени q , а ω — параметр, связанный со световодом:

$$\omega^2 = \frac{2a}{kn_1 \sqrt{2\Delta}}. \quad (8.48)$$

В приближении малой разности показателей преломления геометрическое и электромагнитное исследование показало, что угол распространения остается небольшим. Это значит, что распространяющаяся волна почти поперечная, локально плоская: соответствующие моды поляризованы почти линейно и представляют собой моды LP_{vp} , где $p = q + 1$, причем q —

число узлов функции $\psi_1(r)$ на длине радиуса в соответствии с формулой (8.47). С другой стороны, соответствующая постоянная распространения дается выражением [115]

$$\beta = kn_1 \left[1 - \frac{\sqrt{2\Delta}}{kn_1 a} (4q + 2v + 2) \right]^{1/2},$$

которое, если учесть, что на практике величина Δ мала, с хорошим приближением записывается в виде

$$\beta = kn_1 - (2q + v + 1) \frac{\sqrt{2\Delta}}{a}. \quad (8.49)$$

Таким образом, постоянные распространения образуют арифметическую прогрессию с разностью $\sqrt{2\Delta}/a$, поскольку при переходе от одной моды к другой индекс v или q на единицу увеличивается.

§ 5. Оптимизация оптических волокон

Перейдем теперь к вопросу о том, как оптическое волокно передает световую энергию, и рассмотрим групповую скорость (гл. 2, § 5, п. Б). Мы уже упоминали, что оптическое волокно может передавать большое число мод, причем каждая мода имеет свою групповую скорость; в связи с этим наблюдается некоторый разброс во времени прибытия волны на выходе волокна, хотя на входе все волны были введены одновременно. Оптимизация волокна будет состоять в том, чтобы свести к минимуму эту временную дисперсию, добиваясь в качестве критерия широкой полосы пропускания или малого уширения импульса.

А. Оценка группового времени

Чаще измеряется не групповая скорость, а групповое время. На длине волны λ_0 групповое время на единицу длины определяется как

$$\tau_g = \frac{1}{c} \frac{\partial \beta}{\partial k} \Big|_{\lambda=\lambda_0} = \frac{1}{v_g}. \quad (8.50)$$

В данном параграфе мы ограничимся изучением волокон со степенным законом изменения профиля показателя преломления [формула (8.26)]

$$n(r) = n_1 \left[1 - 2\Delta \left(\frac{r}{a} \right)^\alpha \right]^{1/2}$$

(который, как мы помним, охватывает и случай псевдопараболического профиля с $\alpha = 2$, а также скачкообразное изменение

показателя преломления, когда $\alpha = \infty$). Задержку во времени τ_g можно найти, проследив за лучом: его траектория периодическая, и время прохождения расстояния L будет равно произведению временной задержки, равной периоду τ_z , на число периодов P_z , содержащихся в L . В свою очередь отрезок времени τ_z определяется интегралом

$$\tau_z = \int_{P_z} N dl, \quad (8.51)$$

где N — групповой показатель. Обозначим через N_1 значение величины N на оси (гл. 2, § 5, п. Б):

$$N_1 = \left[n(r) - \lambda \frac{dn(r)}{d\lambda} \right]_{r=0}. \quad (8.52)$$

В производную в выражении (8.50) нам придется ввести параметр дисперсии профиля P , равный

$$P = \frac{n_1}{N_1} \frac{\lambda}{\Delta} \frac{d\Delta}{d\lambda}. \quad (8.53)$$

Наконец, чтобы упростить запись, мы введем два полезных геометрических параметра: поскольку произведение $n(r) \cos \theta$ инвариантно, мы будем использовать угол θ_0 между осью и меридиональным лучом с той же постоянной распространения, определяемый соотношением

$$\beta = kn(r) \cos \theta = kn_1 \cos \theta_0, \quad (8.54)$$

а также максимальную удаленность от оси того же самого луча r_0 :

$$\beta = kn(r_0), \text{ т. е. } r_0 = a \left[\frac{a^2}{V^2} (k^2 n_1^2 - \beta^2) \right]^{1/\alpha}. \quad (8.55)$$

Параметром r_0 мы уже пользовались в формуле (8.30б) и на рис. 8.15. Рассмотрим теперь меридиональный луч ($v = 0$) с постоянной распространения $\beta = kn_1 \cos \theta_0 = kn(r_0)$. Мы можем записать период P_z в виде (см. формулу (8.25) и работу [126])

$$P_z = 2 \int_0^{r_0} \frac{\beta dr}{\sqrt{k^2 n^2 - \beta^2}}. \quad (8.56)$$

Произведя замену переменных $r = tr_0$ и $u = t^\alpha$, получим

$$P_z = \frac{2\beta r_0}{\sqrt{k^2 n_1^2 - \beta^2}} B\left(\frac{1}{2}, \frac{1}{\alpha}\right), \quad (8.56a)$$

где [122]

$$B\left(\frac{1}{2}, \frac{1}{\alpha}\right) = \int_0^1 u^{\frac{1-\alpha}{\alpha}} (1-u)^{-1/2} du.$$

Аналогичным образом можно вычислить время τ_z :

$$\tau_z = \frac{2}{c} \int_0^{r_0} N \frac{dl}{dr} dr = \frac{1}{c} \frac{N_1}{kn_1} \frac{r_0}{\alpha} \frac{1}{\sqrt{k^2 n_1^2 - \beta^2}} \times \\ \times \left(\beta^2 + (k^2 n_1^2 - \beta^2) \frac{\alpha + P}{\alpha + 2} \right) B\left(\frac{1}{2}, \frac{1}{\alpha}\right). \quad (8.57)$$

Отсюда можно найти временную задержку на длине L :

$$\tau = \tau_g(L) = \tau_z \frac{L}{P_z} = \frac{L N_1}{c} \frac{1}{kn_1} \left(\beta^2 + (k^2 n_1^2 - \beta^2) \frac{\alpha + P}{\alpha + 2} \right).$$

Величину τ можно также представить в виде функции параметра θ_0 :

$$\tau(\theta_0) = \frac{\tau_0}{\cos \theta_0} \left(1 - \frac{2-P}{\alpha+2} \sin^2 \theta_0 \right), \quad (8.58)$$

$$\text{где } \tau_0 = \tau(0) = \frac{N_1 L}{c}.$$

В этих вычислениях мы не учитывали возможной зависимости α от λ . Кроме того, поскольку L не равно целому числу периодов P_z , к выражению (8.58) добавляется поправочный член, которым, однако, можно пренебречь при больших L .

Полученный результат может быть обобщен на лучи с той же постоянной распространения β , но не являющиеся меридиональными [127]. Для волокна со скачкообразным изменением показателя преломления формула упрощается:

$$\tau(\theta_0) = \frac{\tau_0}{\cos \theta_0}. \quad (8.59)$$

Различие во времени между крайними модами в этом случае равно $\Delta\tau \approx \tau_0 \Delta$, что составляет примерно 1% — величину огромную. Более подробный анализ зависимости $\tau(\theta_0)$ от разных параметров позволит нам выявить возможности значительного улучшения характеристик оптических волокон.

Б. Зависимость группового времени распространения света от угла θ_0

В случае заданного монохроматического источника, т. е. при $P = \text{const}$, время распространения моды зависит от ее порядка. Чтобы выяснить характер этой зависимости, продифференцируем выражение (8.58) по θ_0 :

$$\frac{d\tau}{d\theta_0} = \frac{\tau_0}{\cos^2 \theta_0} \sin \theta_0 \left(\frac{\alpha + P}{2 - P} - \cos^2 \theta_0 \right) \frac{2 - P}{\alpha + 2}. \quad (8.60)$$

Далее будем считать, что $\alpha_1 = 2 - 2P$. Исследуем ход изменения производной $d\tau/d\theta_0$.

1. При $\alpha \geq \alpha_1 = 2 - 2P$ производная всегда положительная. Максимальный разрыв во времени между модами на концах волокон можно определить по формуле

$$\Delta\tau_{\max} = \tau(\theta_c) - \tau_0 = \tau_0 \left(\frac{1 - 2\Delta \frac{2-P}{\alpha+2}}{\sqrt{1-2\Delta}} - 1 \right). \quad (8.61)$$

2. При $\alpha < \alpha_1$ существует значение θ_1 величины θ_0 , при котором производная равна нулю. На практике (см., например,

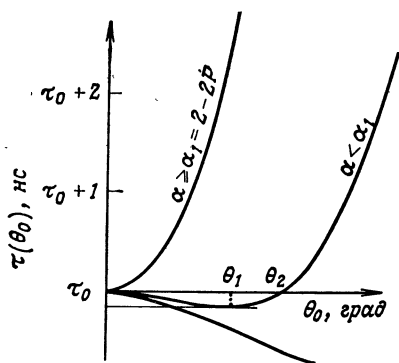


Рис. 8.17. Зависимость величины τ от угла θ_0 .

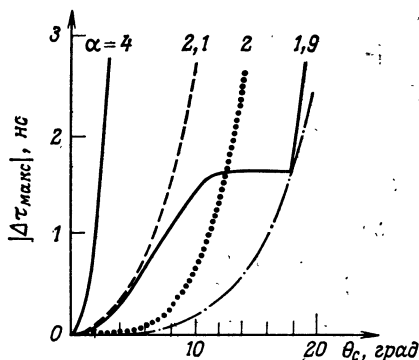


Рис. 8.18. Зависимость абсолютной величины максимального рассогласования по времени между лучами от угла θ_c .

[128]) оказывается, что величина $2 - 2P$ всегда остается положительной, что упрощает анализ хода изменения производной:

а) на участке от $\theta_0 = 0$ до $\theta_0 = \theta_1$ время τ медленно уменьшается от τ_0 до $\tau_1 = \tau(\theta_1)$;

б) на участке от $\theta_0 = \theta_1$ до $\theta_0 = \theta_2$ время τ увеличивается, но остается меньше τ_0 .

в) при $\theta_0 > \theta_2$ время τ быстро растет, превышая τ_0 .

Подобный ход изменения времени распространения мод показан на рис. 8.17. На рис. 8.18 представлена зависимость максимального разрыва во времени между связанными лучами от θ_c . Следует иметь в виду, что *при незначительных изменениях α изменения величины $\Delta\tau_{\max}$ весьма значительны*. Зависимость от угла θ_0 тоже остается весьма существенной всюду, кроме случаев, когда величина α близка к 2, как это будет показано ниже.

В. Зависимость дисперсии от α

Пусть на этот раз величины θ_c (или Δ) и P будут фиксированы, а величина α — переменной, т. е. будет изменяться профиль показателя преломления. На рис. 8.19 представлена

зависимость разности показателей преломления $n(\alpha)$ и $n(\alpha = 2)$, деленной на Δ , от расстояния до оси r/a . Как нетрудно видеть, при изменении α на 0,2 приведенная разность показателей преломления не превышает 3%.

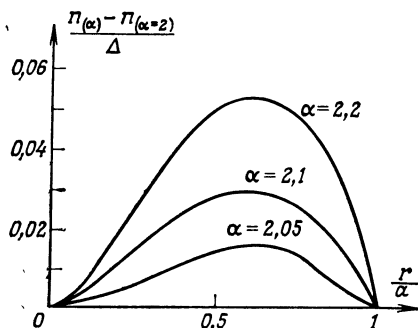


Рис. 8.19. Зависимость относительной разности показателей преломления от расстояния до оси волокна при нескольких значениях α , близких к 2. Разность показателей преломления поделена на Δ .

Учитывая сказанное ранее, можно обнаружить три зоны изменений в зависимости от значения α .

1. При $\alpha > \alpha_1$ мы имеем $\Delta\tau_{\max} = \tau(\theta_c) - \tau_0$.

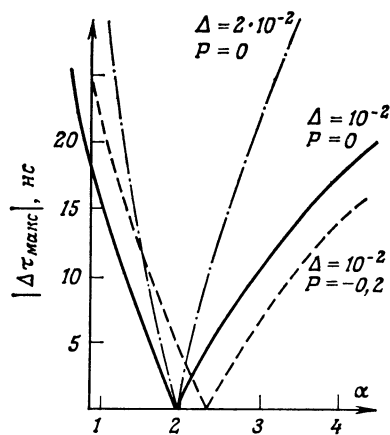


Рис. 8.20. Зависимость максимального рассогласования по времени $\Delta\tau_{\max}$ между лучами от α .

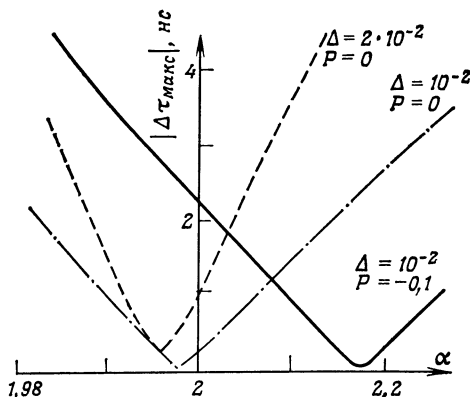


Рис. 8.21. Увеличенная часть рис. 8.20 вблизи значения $\alpha = 2$.

2. При $\alpha < \alpha_2 = (2 - 2P)(1 - 2\Delta) - P$ мы имеем $\Delta\tau_{\max} = \tau_0 - \tau(\theta_c)$.

3. При $\alpha_2 < \alpha < \alpha_1$ имеем следующее:

а) при $\theta_c < \theta_2$, где θ_2 зависит от α и P (рис. 8.17), справедливо выражение

$$\Delta\tau_{\max} = \tau_0 - \tau(\theta_1) = \tau_0 \left(1 - \frac{2}{\alpha + 2} \sqrt{(\alpha + 2)(2 - P)} \right); \quad (8.62)$$

б) при $\theta_c > \theta_2$ максимальное различие во времени равно

$$\Delta\tau_{\max} = \tau(\theta_c) - \tau(\theta_1).$$

Все сказанное представлено графически на рис. 8.20 и 8.21. Как нетрудно видеть, при заданных P и Δ существует оптимальное значение показателя степени α , при котором величина $\Delta\tau_{\max}$ минимальна¹⁾,

$$\alpha_{\text{опт}} = 2 - 2P - 2\Delta. \quad (8.63)$$

При переходе от $\alpha = \infty$ (волокно со скачкообразным изменением показателя преломления) к $\alpha = \alpha_{\text{опт}}$ достигается выигрыш в модовой дисперсии примерно в 800 раз. Таким образом, весьма высокая чувствительность $\Delta\tau_{\max}$ к изменениям α показывает, что разброс профилей показателей преломления играет очень важную роль.

Г. Зависимость максимального рассогласования по времени от P и λ

Проведенный выше анализ в приближении первого порядка (P — константа) показывает, что P входит в выражение для оптимального значения α : при изменении средней длины волны

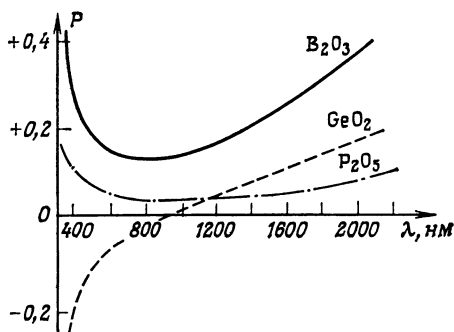


Рис. 8.22. Зависимость дисперсионного параметра P профиля показателя преломления от длины волны в кварце с разными добавками [128].

оптимальное значение α изменяется. На рис. 8.22 представлены результаты измерения величины P [128]: ее зависимость от длины волны в легированном кварце с разными примесями.

Эти экспериментальные данные показывают, что P изменяется незначительно (от $-0,4$ до $+0,2$). В волокнах с примесью фосфора наблюдаются крайне слабые изменения в диапазоне длин волн $0,6$ — $1,8$ мкм. Такие волокна остаются очень хорошими, поскольку величина α почти оптимальна во всем этом диапазоне волн. Во всех случаях, кроме особого случая легирования фосфором, изменение дисперсионного параметра профиля показателя преломления в зависимости от длины волны также дает существенный вклад в общую дисперсию.

¹⁾ Строго говоря, выражение должно быть несколько иным, так как мы не учли пространственного распределения энергии. Если учесть его, то $\alpha_{\text{опт}} = 2 - 2P - \Delta(4 - 2P)(3 - 2P)/(5 - 2P)$.

Д. Зависимость $\Delta\tau_{\text{макс}}$ от числовой апертуры

Мы не будем рассматривать общий случай. Отметим только, что оптимальное значение величины α зависит от Δ : при изменении числовой апертуры изменяется оптимальное значение α ; это существенно при изготовлении оптических волокон. Заметим, однако, что, выбирая значение, при котором величина α оптимальна, мы принимаем [формула (8.58)], что времена распространения света τ_0 и $\tau(\theta_c)$ одинаковы.

Тем не менее можно сделать одно замечание, основанное на анализе рис. 8.18. Поскольку максимальное временное несовпадение не зависит от θ_c в интервале $\theta_1 < \theta_c < \theta_2$, если мы зададим некоторое максимально допустимое временное несовпадение, то можно установить соответствующие границы изменения показателя Δ .

Например, если мы умеем осуществлять профиль с некоторым значением α , то все волокна одинакового состава с числовой апертурой, лежащей между θ_1 и θ_2 , будут иметь максимальное значение временной дисперсии на выходе, равное заданной величине.

Итак, мы видим, что групповая временная дисперсия в данном волокне определяется в основном следующими эффектами:

- а) *модовой дисперсией* (связанной с зависимостью τ от θ_0),
- б) *дисперсией, обусловленной материалом*, т. е. зависимостью τ от λ ; такая дисперсия называется *внутримодальной*, или *хроматической*.

§ 6. Импульсный отклик и передаточная функция

Найдя выражение для группового времени распространения $\tau(\theta, \lambda)$ [формула (8.58)], рассмотрим теперь источник излучения, который будем характеризовать светимостью $B(\theta, \lambda)$. Волокно осуществляет некую селекцию возбуждения разных лучей, и, строго говоря, мы тоже должны ввести в его отклик некий коэффициент ослабления $\mathcal{P}(\theta, \lambda)$, который может зависеть от длины волны и от угла θ . Тогда сигнал на выходе волокна будет иметь вид

$$s(t) = \int B(\theta, \lambda) \mathcal{P}(\theta, \lambda) e(t - \tau(\theta, \lambda)) dS d\Omega d\lambda, \quad (8.64)$$

где dS — элемент излучающей поверхности, расположенный на входном торце волокна, $d\Omega$ — элемент телесного угла, $e(t)$ — функция, учитывающая зависимость интенсивности источника от времени. Мы предполагаем, что оптическое волокно описывается оператором, линейным по отношению к оптической мощности. Это справедливо, если источник в достаточной мере некогерентен (т. е. величина, обратная времени когерентности источника,

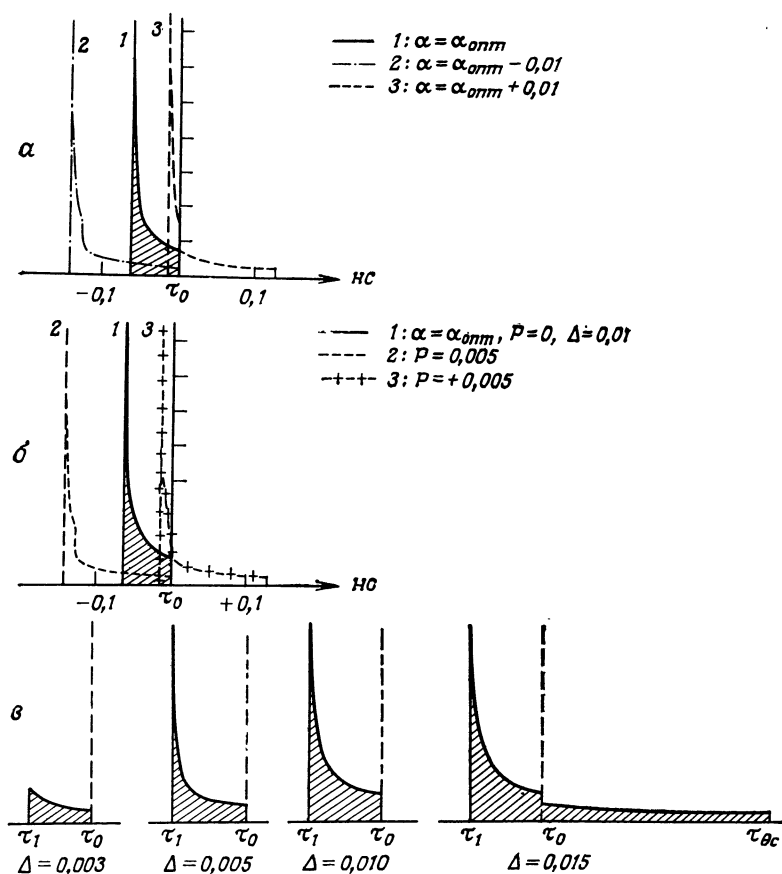


Рис. 8.23. Изменения импульсного отклика волокна при изменении параметров относительно своих оптимальных значений.

а) α изменяется, P и Δ постоянны; б) P изменяется, α и Δ постоянны; в) Δ изменяется, P постоянно, $\alpha = \alpha_{opt}$ при $\Delta = 0,01$.

намного больше полосы пропускания, в которой проводятся измерения [129]).

На рис. 8.23 представлены результаты исследования разных случаев при условии, что функция $e(t)$ временного возбуждения волокна на входе есть *импульс Дирака*, источник монохроматический и множитель $\mathcal{P}(\theta, \lambda)$, учитывающий потери, не зависит от угла. Когда источник не монохроматический, из-за слишком быстрых изменений $\Delta \tau_{max}$ в зависимости от α вблизи оптимума трудно предвидеть как сам результат, так и влияние на него различных факторов, поскольку изменение P при этом играет важную роль [128]; часто оказывается легче анализи-

ровать изменение, рассматривая полосу пропускания. Передаточную функцию $H(f)$ находят путем преобразования Фурье импульсного отклика или путем прямого измерения коэффициента усиления при изменении частоты *синусоидального возбуждения*:

$$s(f, t) = H(f) e^{i2\pi f t}.$$

На рис. 8.24 показано, как изменяется $H(f)$ в зависимости от α при постоянных значениях λ и Δ или в зависимости от Δ при фиксированных α и λ . На основании рис. 8.24 и 8.22 можно сделать следующие выводы:

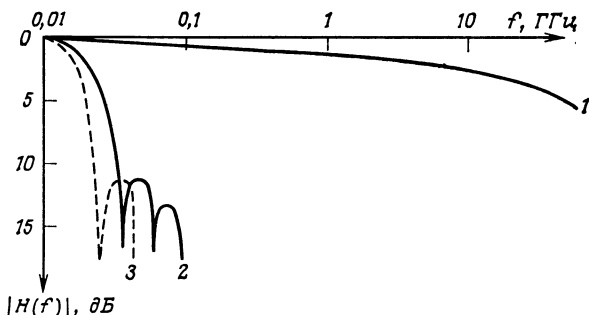


Рис. 8.24. Зависимость модуля передаточной функции волокна $H(f)$ от α и Δ . 1 — при $\alpha = \alpha_{\text{опт}}$, $\Delta = 0,01$; 2 — при $\alpha = \infty$, $\Delta = 0,01$; 3 — при $\alpha = \infty$, $\Delta = 0,015$.

1) полоса пропускания *кварцевых волокон, легированных германием*, очень сильно зависит от монохроматичности источника;

2) полоса пропускания *волокон, легированных фосфором*, весьма мало зависит от монохроматичности источника, поскольку P реагирует очень слабо на изменение длины волны.

§ 7. Влияние взаимодействия мод

Поскольку в волокнах имеются более или менее выраженные дефекты, между различными модами распространения света, вообще говоря, будет происходить энергетический обмен. С геометрической точки зрения это можно объяснить следующим образом: если граница волокна деформирована и составляет угол $d\varphi$ с его осью, то луч при отражении отклонится на угол $2d\varphi$. Поэтому его постоянная распространения изменится и изменится распределение полной мощности $P(z)$, передаваемой вдоль оси z , по различным модам распространения. Возможно также появление потерь, если угол распространения света превысит величину θ_c . Теоретически возможен энергетический обмен между каждой из мод и всеми прочими модами,

но мы для простоты примем [130], что каждая мода взаимодействует только с ближайшими к ней соседними модами. Если m — характерный параметр данной моды, то энергетический обмен для нее возможен только с модами $m - 1$ и $m + 1$.

Обозначим через $P_m(z)$ мощность, передаваемую вдоль оси z модой порядка m . Тогда полная мощность будет равна

$$P(z) = \sum_m P_m(z).$$

Сложность решения задачи зависит от типа световода (плоский или цилиндрический). В самом деле, если у всех мод разные и уменьшающиеся постоянные распространения ($\beta_m > \beta_{m+1} \dots$), то можно рассматривать каждую моду в отдельности. Но если существуют вырожденные моды, постоянные распространения которых одинаковы и соответствуют одному и тому же параметру m , то придется принять, что такие моды неразличимы и образуют группу мод. Тогда излагаемая здесь теория будет относиться к группам мод.

В одномерной модели плоского световода со скачкообразно изменяющимся показателем преломления [формула (3.11)] поперечное поле, связанное с модой m , оказывается квазисинусоидальным и угол распространения определяется выражением

$$\sin \theta = \frac{m\pi}{2ka},$$

а при малом θ , т. е. при малой разности показателей преломления, можно написать

$$\theta \approx \frac{m\pi}{2ka}.$$

Приращение $\Delta\theta$ от моды к моде постоянно и равно

$$\Delta\theta = \frac{\pi}{2ka}, \quad (8.65)$$

откуда $\theta = m\Delta\theta$.

В случае цилиндрического световода со скачкообразным изменением показателя преломления поперечная составляющая есть функция Бесселя J_v и при малой разности показателей преломления можно связать две моды $HE_{v,q}$ и $EH_{v-2,q}$ в псевдомоду LP (§ 4, п. Б, 1); индекс q есть число нулей функции J_v в пределах сердцевинны. Таким образом, можно показать [131, 132], что одно число $m = 2q + v$ может служить характеристикой группы из m мод с параметрами q и v .

А. Одномерная модель (плоский световод)

Допустим, что изменение мощности в моде m на отрезке dz линейно зависит от P_m и от разностей мощностей, передаваемых соседними модами. Тогда

$$\frac{dP_m}{dz} = -\alpha_m P_m + d_m (P_{m+1} - P_m) + d_{m-1} (P_{m-1} - P_m). \quad (8.66)$$

Если предположить, что в волокне очень много мод, то можно перейти от дискретного параметра m к непрерывному параметру θ :

$$\frac{dP_m}{d\theta} = \frac{P_{m+1} - P_m}{\theta_{m+1} - \theta_m}, \quad \text{т. е.} \quad P_{m+1} - P_m \approx \Delta\theta \frac{\partial P_m}{\partial \theta}.$$

Но разность

$$d_m \frac{\partial P_m}{\partial \theta} - d_{m-1} \frac{\partial P_{m-1}}{\partial \theta}$$

есть не что иное, как производная $(\partial/\partial\theta)(d \partial P/\partial\theta)$; таким образом, мы получаем уравнение

$$\frac{dP}{dz} = -\alpha(\theta) P + \Delta\theta \frac{\partial}{\partial\theta} \left[d \frac{\partial P}{\partial\theta} \right]. \quad (8.67)$$

Б. Цилиндрическая модель (волокно со скачкообразным изменением показателя преломления)

Параметр m относится к группе из m мод, и эти моды взаимодействуют с $m-1$ модами, параметр которых равен $m-1$ (§ 4, п. Б). В этом случае получаем

$$m \frac{dP_m}{dz} = -m\alpha_m P_m + m d_m (P_{m+1} - P_m) + (m-1)(P_{m-1} - P_m). \quad (8.68)$$

Приняв, что $\theta = m\Delta\theta$, мы можем снова перейти к непрерывности.

Разложим $\alpha(\theta)$ и $d(\theta)$ в ряд по возрастающим степеням θ . В силу симметрии примем внутренние потери равными

$$\alpha(\theta) = \alpha_0 + \alpha_1 \theta^2 + \dots$$

и ограничимся первым членом в выражении для коэффициента связи $d(\theta) = d_0$. Введем обозначение $D = d_0(\Delta\theta)^2$ в уравнении (8.68); тогда оно принимает вид уравнения диффузии. Зависимость от α_0 вводится в общее решение посредством множителя $e^{-\alpha_0 z}$, и мы можем принять следующую окончательную запись преобразованного уравнения [130, 131]:

а) в случае плоского световода

$$\frac{\partial P}{\partial z} = -\alpha_1 \theta^2 P + D \frac{\partial^2 P}{\partial \theta^2}, \quad (8.69)$$

б) в случае круглого световода

$$\frac{\partial P}{\partial z} = -\alpha_1 \theta^2 P - \frac{D}{\theta} \frac{\partial}{\partial \theta} \left(\theta \frac{\partial P}{\partial \theta} \right). \quad (8.70)$$

Решения этих уравнений ищут в сепарабельном виде $P = Q(\theta)e^{-\gamma z}$, причем функция $Q(\theta)$ должна удовлетворять одному из следующих дифференциальных уравнений:

а) в случае плоского световода

$$D \frac{d^2 Q}{d\theta^2} = [\alpha_1 \theta^2 - \gamma] Q; \quad (8.71)$$

б) в случае круглого световода

$$\frac{D}{\theta} \frac{d}{d\theta} \left(\theta \frac{dQ}{d\theta} \right) = [\alpha_1 \theta^2 - \gamma] Q.$$

Решениями этих уравнений будут функции Эрмита — Гаусса в первом случае и Лагерра — Гаусса — во втором. В обоих случаях решение, обеспечивающее минимальные потери, имеет вид $\exp(-\theta^2/\theta_0^2)$, причем $\theta_0^2 = 2\sqrt{D/\alpha_1}$ [130—133], а потери на связь характеризуются множителем $\exp(-\gamma_c z)$. Коэффициент γ_c равен $\sqrt{\alpha_1 D}$ для плоского волновода и $2\sqrt{\alpha_1 D}$ для круглого волновода.

При любой функции решение, зависящее от z , может быть построено из решений уравнения (8.71), а граничные условия задаются начальным распределением мощности в каждой из мод распространения. Не вдаваясь в подробности этой теории [132, 133], отметим некоторые важные результаты.

Взаимодействие мод приводит к дополнительным потерям при распространении света в световоде вследствие того, что энергия передается некоторым из нераспространяющихся мод. Предельное распределение энергии по длинам пробега z принимает вид

$$P(z) = P_0 \exp\{-(\alpha_0 + \gamma_c)z\}, \quad (8.72)$$

где γ_c характеризует потери, добавляющиеся вследствие преобразования мод. Пространственное распределение энергии зависит от потерь и от начальных условий ввода и *стремится к равновесному гауссовому распределению*, так что на практике можно говорить об эффективной числовой апертуре (гл. 9, § 4).

Если z мало по сравнению с $1/\gamma_c$, то взаимодействие мод не успевает оказать влияние на распространение света и модовая дисперсия линейно возрастает с расстоянием. Когда z велико по сравнению с $1/\gamma_c$, дисперсия растет по закону \sqrt{z} .

Передаточная функция световода принимает вид

$$H(f, z) \approx \exp \left\{ -\frac{sf^2 z}{\gamma_c} \right\}, \quad (8.73)$$

где s — параметр, зависящий от типа световода и от взаимодействия мод. Полоса пропускания при модуляции дается выражением

$$\text{ПП} \approx \sqrt{\frac{\gamma_c}{z}}. \quad (8.74)$$

Ценой дополнительных потерь, связанных с γ_c , можно увеличить ширину полосы пропускания волокна путем специального связывания мод (например, введением микроизгибов на определенных участках волокна), поскольку в зависимости от взаимодействия мод закон изменения полосы пропускания изменяется от вида $1/z$ до вида $1/\sqrt{z}$.

§ 8. Одномодовые волокна

А. Введение

Можно так подобрать параметры диэлектрического световода (§ 3), что будет распространяться только одна мода HE_{11} . Такое волокно, называемое одномодовым, имеет очень широкую полосу пропускания (несколько десятков гигагерц), но размеры его сердцевинны очень малы и соизмеримы с длиной волны. Соотношение (8.41), установленное для волокна со скачкообразно меняющимся показателем преломления, показывает, что произведение числовой апертуры на размер сердцевинны не должно превышать $V_c = 2,405$ для обеспечения распространения в волокне лишь одной моды. Рис. 8.25 иллюстрирует тот компромисс, которого нужно достичь: все точки, лежащие ниже кривой, удовлетворяют данному требованию. Для волокон на основе кварца окна передачи с очень малыми потерями расположены вблизи $\lambda = 1,3$ мкм и $\lambda = 1,55$ мкм. При $\lambda = 1,3$ мкм мы приближаемся к длине волны, на которой становится равной нулю дисперсия, связанная с материалом, $(\lambda/c)(d^2n/d\lambda^2)$ (длина волны, на которой исчезает эта дисперсия, зависит от используемых примесей, а в случае германия, в частности, и от числовой апертуры оптического волокна).

Подбирая параметр V , можно скомпенсировать влияние модовой дисперсии в световоде, а также хроматическую дисперсию для других значений длины волны. В частности, можно довести длину волны, на которой происходит компенсация диспер-

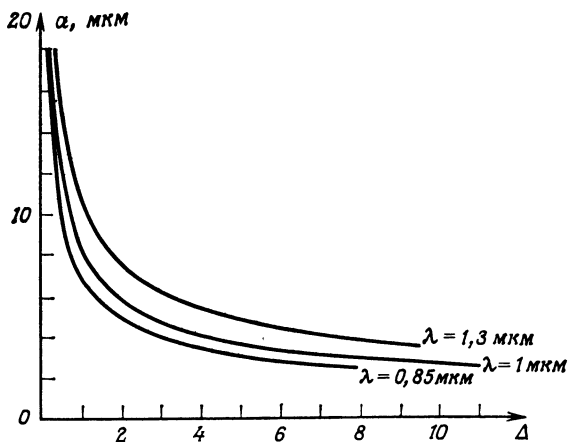


Рис. 8.25. Кривые максимальных значений α при заданных значениях Δ для одномодового волокна со скачкообразным изменением показателя преломления.

Показатель преломления оболочки равен $n_2 = n_1 \sqrt{1 - 2\Delta}$, волокно из кварца.

сий, до 1,55 мкм, и в этом случае качество световода значительно улучшается как в отношении потерь, так и в отношении полосы пропускания (§ 8, п. Д, 2).

Б. Характеристики электромагнитного поля

Используя принятые ранее обозначения, мы можем записать уравнение электромагнитного поля моды в общем виде

$$E(r, z) = E(r) \exp i(\omega t - \beta z). \quad (8.75)$$

Предположим сначала, что мы имеем волокно со скачкообразным изменением показателя преломления и с оболочкой бесконечной толщины. Амплитуда $E(r)$ была определена в § 4, п. Б, 1 и равна

$$E(r) = \begin{cases} J_0\left(u \frac{r}{a}\right) & \text{при } 0 < r \leq a, \\ \frac{J_0(u)}{K_0(w)} K\left(w \frac{r}{a}\right) & \text{при } r > a, \end{cases} \quad (8.76)$$

где u и w — параметры, которые удовлетворяют следующим уравнениям, вытекающим из граничных условий при $r = a$:

$$\begin{aligned} u^2 + w^2 &= V^2, \\ u \frac{J_1(u)}{J_0(u)} &= w \frac{K_1(w)}{K_0(w)}. \end{aligned} \quad (8.77)$$

Глож [133] дает приближенное решение этих уравнений в виде

$$u = \frac{(1+x)V}{1+(x^4+V^4)^{1/4}}, \quad (8.78)$$

причем относительная погрешность не превышает $4 \cdot 10^{-3}$ при $x = \sqrt{2}$. Форма поля, описываемая выражением (8.76), мало пригодна для анализа, но его можно выразить иначе (при $r \leq a$) — через гауссову функцию параметра ω_0 :

$$E(r) = A \exp \left\{ -\frac{r^2}{\omega_0^2} \right\}. \quad (8.79)$$

Параметр ω_0/a зависит только от приведенной частоты V , и его оптимальное значение можно найти путем максимизации эффективности (к.п.д.) ввода гауссова пучка в изучаемое волокно. На рис. 8.26 показано, как ω_0/a изменяется в зависимости от V . Заметим, что при $V > 2$ к.п.д. ввода превышает 99 %. В работе Маркузе [134] можно найти аналитические выражения, удобные для вычисления $\omega_0(V)$ с достаточно большой точностью. Правда, понятие размера (объема) моды, выражаемое формулой (8.79), становится спорным, когда в волокне имеется дефект на оси сердцевины (минимум показателя преломления).

Поле в дальней зоне, излучаемое волокном, можно тоже представить гауссовой функцией, и в воздухе расходимость, согласно формуле (4.22), будет равна

$$\theta_0 = \frac{1}{\pi \omega_0}.$$

Наконец, постоянная распространения одиночной моды дается выражением

$$\beta = \sqrt{k^2 n_1^2 - \frac{u^2}{a^2}}, \quad (8.80)$$

которое может быть также переписано в более общем виде с учетом типа оболочки, если использовать параметр B , введенный в § 4 [формула (8.33)]:

$$B = \frac{w^2}{V^2} = 1 - \frac{u^2}{V^2}, \quad (8.81)$$

$$\beta = k \sqrt{n_2^2 + B(n_1^2 - n_2^2)}. \quad (8.82)$$

Что будет, если показатель преломления сердцевины не постоянный, т. е. если волокно градиентное? Прежде всего это будет практический случай, поскольку при изготовлении волокон невозможно получить совершенно неизменяющийся показатель преломления.

Мы будем рассматривать в данном случае профиль показателя преломления, соответствующий формуле (8.26), для которого большая часть интересующих нас результатов была получена Гамблингом [135]. Они представлены на рис. 8.27, где по-

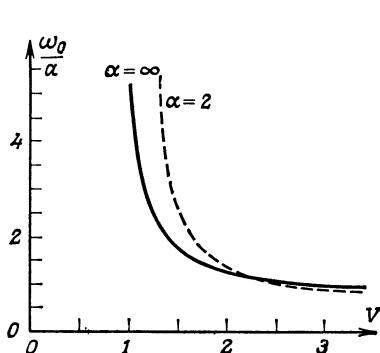


Рис. 8.26. Зависимость отношения ω_0/α от V .

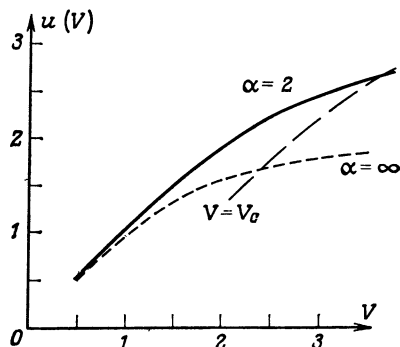


Рис. 8.27. Зависимость параметра u от V при $\alpha = 2$ [135].

Сплошная линия — для градиентного волокна; штриховая — для волокна со скачкообразным изменением показателя преломления.

казана зависимость параметра u от приведенной частоты V при разных значениях α . Мы нанесли также кривую критического значения V , выше которого волокно не является больше одно-модовым: в пределах от $\alpha = \infty$ (волокно со скачкообразным изменением показателя преломления) до $\alpha = 2$ это критическое значение изменяется от 2,405 до 3,518. Соотношение (8.78), описывающее $u(V)$, дает хорошее приближение, если при $\alpha = 2$ принять значение x , равное 2,52. Электромагнитное поле может быть также корректно описано гауссовой функцией с параметром ω_0 (кривая зависимости которого от V при $\alpha = 2$ проведена штриховой линией на рис. 8.26).

В. Потери в одномодовых волокнах

Потери связаны с многими неизбежными факторами. Кроме поглощения и рэлеевского рассеяния, можно назвать качество оболочки и ее толщину, кривизну волокна (рис. 8.28), его мик-

1. Влияние оболочки

Если использовать для поля выражение (8.79), то на внешней границе оболочки ($r = b$) поле будет равно $A \exp[-(b^2/\omega_0^2)]$. Таким образом, достаточно толщины оболочки, равной трем или четырем ω_0 , чтобы поле стало пренебрежимо малым; в дальнейшем мы будем без уменьшения общности считать, что толщина оболочки равна бесконечности.

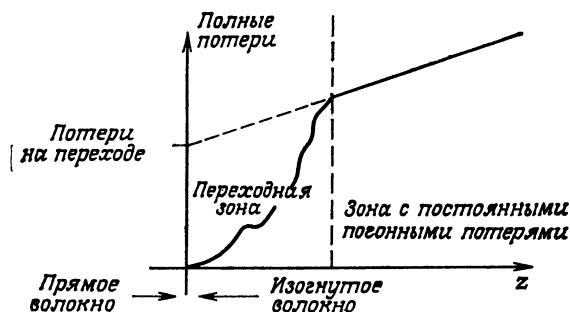


Рис. 8.28. Потери, обусловленные искривленностью волокна.

Перед зоной постоянных линейных потерь имеется переходная зона, потери в которой могут составлять 15 дБ.

Обозначив через γ^c и γ^g коэффициенты линейного ослабления в сердцевине и оболочке, можно написать следующее выражение для полного ослабления:

$$\gamma = \frac{\gamma^c P^c + \gamma^g P^g}{P_{\text{полн}}} = \gamma^c + (\gamma^g - \gamma^c) \exp\left\{-\frac{2a^2}{\omega_0^2}\right\}. \quad (8.83)$$

Если коэффициент ослабления в оболочке больше, чем в сердцевине, то следует, очевидно, учитывать дополнительные потери, обусловленные оболочкой, причем они тем значительнее, чем глубже в оболочку проникает поле.

2. Влияние кривизны волокна

Это весьма сложный вопрос, и мы приведем только некоторые хорошо известные результаты. Волокна с градиентом показателя преломления более чувствительны к искривленности, чем волокна со скачкообразным изменением преломления, причем с увеличением V потери уменьшаются. Как показал Гамблинг [135], анализ этого вида потерь затрудняется еще и тем, что до установления нового режима распространения света (с постоянными дополнительными линейными потерями, зависящими от радиуса кривизны) существует некоторый переходный режим большей или меньшей протяженности, зависящей от перехода от прямого к изогнутому волокну (рис. 8.28). Потери, обусловленные таким переходным режимом, могут достигать 15 дБ.

3. Влияние микроизгибов

Если после зоны перехода кривизна остается постоянной, то взаимодействие между модой распространения и модами излучения отсутствует, но, если кривизна меняется, возникают потери. Петерманн [136] показал, что важным параметром в этом случае является эффективный радиус моды ω_0 . В этом случае приходится вводить распределенные случайно микроискривления через функцию автокорреляции кривизны $1/R$, например, в таком виде:

$$\Gamma(l) = \frac{\lim_{Z \rightarrow \infty} \frac{1}{2Z} \int_{-Z}^{+Z} \frac{dz}{R(z)R(z+l)}}{\langle \frac{1}{R^2} \rangle}. \quad (8.84)$$

Можно также рассматривать спектр $\Phi(v)$, определяемый на соответствующих пространственных частотах: связь возможна только тогда, когда разность постоянных распространения между модами не является нулем спектра. Экспериментально показано, что приближение для спектра в виде

$$\Phi(v) = \frac{K}{v^2} \quad (8.85)$$

довольно близко к действительности: *линейные потери, связанные с микроискривлениями, пропорциональны ω_0^6 . Следовательно, совершенно необходимо минимизировать величину ω_0 .*

В отношении потерь из-за микроискривлений оба типа волокон (градиентные и со скачкообразным изменением показателя преломления), по-видимому, эквивалентны.

Г. Соединение одномодовых волокон

Способность одномодовых волокон «подключаться» к источникам света, а также жесткие допуски, которые должны выдерживаться в их соединении, — самые большие *недостатки одномодовых волокон*. В работе [134] приводятся данные о возможных порядках величин. На рис. 8.29 представлена зависимость относительных потерь в процентах от разных параметров.

1. Если волокна строго соосны, без осевого зазора, но характеризуются разными значениями ω_{01} и ω_{02} параметра ω_0 , то коэффициент передачи по мощности записывается в виде

$$T_w = \left(\frac{2\omega_{01}\omega_{02}}{\omega_{01}^2 + \omega_{02}^2} \right)^2. \quad (8.86)$$

Потери будут ниже 0,2 дБ при $\omega_{01}/\omega_{02} < 1,25$.

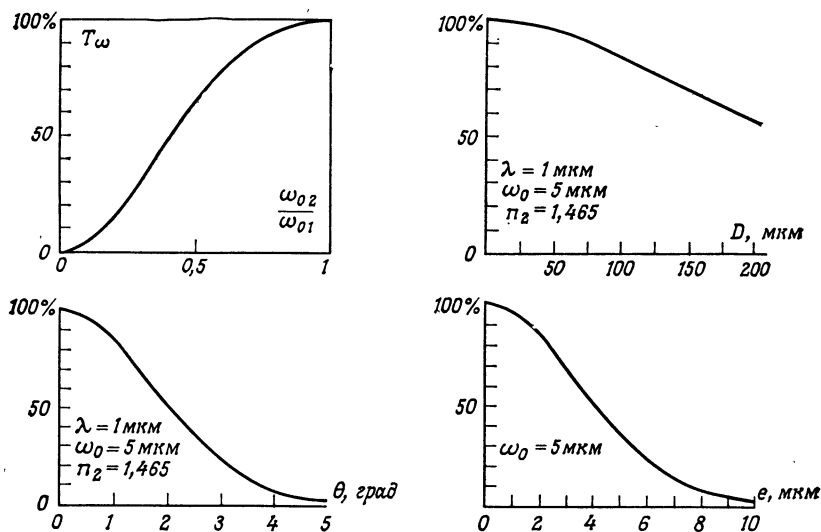


Рис. 8.29. Зависимость потерь (в процентах) в соединении двух одномодовых волокон от отношения ω_{01}/ω_{02} , ширины зазора D , углового несовпадения θ и несоосности e .

2. Если волокна одинаковы, строго соосны, но между ними есть зазор D , то коэффициент T будет равен

$$T_D = \frac{1 + 4Z^2}{Z^2 + (1 + 2Z^2)^2}, \quad (8.87)$$

$$\text{где } Z = \frac{D}{k n_2 \omega_0^2}.$$

3. Если волокна одинаковы, параллельны, без зазора, но не соосны и между их осями имеется расстояние e , то

$$T_e = \exp \left\{ - \frac{e^2}{\omega_0^2} \right\}. \quad (8.88)$$

4. Наконец, если волокна одинаковы, без зазора, центры их торцов совмещены, но оси составляют угол θ , то коэффициент T равен

$$T_\theta = \exp \left\{ - \left(\frac{\pi n_2 \omega_0 \theta}{\lambda} \right)^2 \right\}. \quad (8.89)$$

Д. Дисперсия

В случае многомодовых волокон двумя основными видами дисперсии были дисперсия, обусловленная различием в постоянных распространения различных мод, и дисперсия, обусловленная материалом, причем дисперсия внутри мод была незначи-

тельной. В одномодовых волокнах поляризация играет ту же роль, что и модовая дисперсия, и при наличии поляризации такое волокно можно рассматривать как двухмодовое из-за вырожденности моды HE_{11} .

1. Влияние поляризации

В совершенном волокне обе поляризации моды HE_{11} вырождены и имеют одни и те же характеристики. Но малейший дефект (который всегда присутствует на практике), такой, как эллиптичность сердцевины или двойное лучепреломление (естественное или связанное с микронискривлениями), снимает вырождение. Групповые скорости двух поляризаций в таком случае не совпадают, что приводит к взаимодействию между поляризациями, которое проявляется в виде биений, а также в значительной дисперсии группового времени распространения — порядка нескольких десятков пикосекунд [137]. Есть два способа бороться с такими трудностями:

1) повышать качество волокна, так чтобы разность двух постоянных распространения была незначительной — порядка $2\pi \text{ км}^{-1}$ или еще меньше, что позволяет предотвратить взаимодействие на расстоянии 1 км и более;

2) напротив, увеличить двойное лучепреломление, чтобы как можно больше разделить две поляризации и получить пространственную частоту биений порядка миллиметра, т. е. достаточно малую, при которой механическая связь затруднена; в результате поляризация сохранится, поскольку исчезнет возможность связи.

Однако при соединении разных волокон проблема остается нерешенной.

2. Групповое время

На основе выражения (8.50) с учетом формулы (8.82) получаем для группового времени формулу:

$$\tau_g = \frac{1}{c} \left[n_2 N_2 + B + \frac{1}{2} V \frac{dB}{dV} (n_1 N_1 - n_2 N_2) \right] \times \\ \times [n_2^2 + B(n_1^2 - n_2^2)]^{-1/2}, \quad (8.90)$$

обозначив через N_i ($i = 1$ или 2) групповые показатели сердцевины и оболочки. Поскольку относительные изменения $(n_1 - n_2)/n_2$ и $(N_1 - N_2)/N_2$ меньше единицы, это выражение упрощается и принимает вид

$$\tau_g = \frac{N_2}{c} + \frac{d(VB)}{dV} \frac{N_1 - N_2}{c}. \quad (8.91)$$

Продифференцировав (8.91), можно найти зависимость группового времени от длины волны. Обозначим коэффициент дисперсии материала через $M_i = (-\lambda/c) (d^2n/d\lambda^2)$. Тогда

$$\frac{d\tau_g}{d\lambda} = \underbrace{M_2 + (M_1 - M_2) \frac{d(VB)}{dV}}_{\text{Дисперсия материала}} - \underbrace{\frac{(N_1 - N_2)V}{c\lambda} \frac{d^2(VB)}{dV^2}}_{\text{Модовая дисперсия}}. \quad (8.92)$$

С учетом порядков величины разных членов это уравнение сводится к виду [138, 139]

$$\frac{d\tau_g}{d\lambda} = M_2 - \frac{N_1 - N_2}{c\lambda} V \frac{d^2(VB)}{dV^2}. \quad (8.93)$$

В табл. 8.2 сравниваются относительные величины двух оставшихся членов на разных длинах волны. Поскольку знаки двух членов могут быть противоположными, выбрав соответствующим образом параметры V и $N_1 - N_2$ [138—140], можно добиться, чтобы сумма этих двух членов была равна нулю на длине волны, большей 1,3 мкм.

Таблица 8.2

Дисперсия на длине волны 0,85 и 1,29 мкм и пример ее компенсации на длине волны 1,70 мкм

λ , мкм	0,85 ¹⁾	1,29 ¹⁾	1,70 ²⁾
M_2 , пс/нм · км	—84	2	31
Модовая дисперсия, пс/нм · км	3	2	30
Полная дисперсия, пс/нм · км	87	~ 0	~ 1

¹⁾ $V=1,4$, $N_1 - N_2 = 3 \cdot 10^{-3}$.

²⁾ $V=1,3$, $N_1 - N_2 = 1,4 \cdot 10^{-2}$, $\omega_0 = 4$ мкм

ИЗМЕРЕНИЕ ХАРАКТЕРИСТИК ОПТИЧЕСКИХ ВОЛОКОН

При разработке волоконно-оптических систем связи и усовершенствовании техники их изготовления и применения нужно владеть соответствующей методикой измерений рабочих характеристик оптических волокон. По большей части необходимые данные получают путем экспериментальных измерений, проводимых либо на основе теоретических исследований среды, в которой происходит распространение света, либо на основе упрощающих гипотез, требующих проверки.

Как мы увидим в гл. 17, одно из главных требований, предъявляемых к системе дальней связи,— это требование максимального расстояния между ретрансляторами. На практике это расстояние ограничивается разными взаимосвязанными факторами:

- 1) *ослаблением сигнала, обусловленным различными потерями;*
- 2) *искажением сигнала и уменьшением полосы пропускания с увеличением расстояния;*
- 3) *имеющейся мощностью излучения и чувствительностью фотоприемников.*

Если мы намереваемся стыковать несколько оптических волокон или отрезков кабеля, возникают другие проблемы: однородность диаметров волокон, их числовой апертуры, эксцентриситета сердцевины и т. д. Важно, чтобы разработчик оптической системы связи предвидел все это и предъявлял к изготовителю волокна требования, позволяющие реализовать нужные характеристики [141].

В этой главе мы не претендуем на создание строгой классификации измерений, необходимых физикам для проверки теоретических расчетов и инженеру по дальней связи для разработки системы с заданными характеристиками. Отметим, что международные организации¹⁾, ответственные за стандартизацию тех методов измерений характеристик оптических волокон, которые относятся ко второй категории, еще не составили окончательных рекомендаций. Поэтому мы ограничимся измерениями основных параметров, которые мы вводили, говоря о распространении света в волокне.

¹⁾ Например, Международный консультативный комитет по телеграфной и телефонной связи (МККТТС) и Международный комитет по электронике (МКЭ).

§ 1. Общие замечания

В гл. 8 мы рассматривали только идеальные и прямолинейные оптические волокна. В действительности геометрические размеры волокна и профиль показателя преломления могут изменяться вдоль оси z , причем дефекты могут быть локализованными или распределенными. Сам материал волокна поглощает свет, что приводит к уменьшению амплитуды полей, причем неодинаковому в сердцевине и в оболочке, в связи с чем возникают дополнительные искажения.

Различные неоднородности вызывают более или менее значительный обмен энергией между направлениями распространения света: это явление называется *взаимодействием мод* (гл. 8, § 7). Большая часть всех возможных мод диэлектрического волновода является модами распространения, но, к сожалению, имеются моды излучения, а также моды потерь. Дефекты волокна могут служить причиной возникновения взаимодействия между модами распространения и паразитными модами и приводить тем самым к дополнительным потерям.

Существуют и иные причины потерь: изогнутость волокна, рассеяние в волокне на микротрещинах или пузырьках, микроизгибы на поверхности волокна, образующиеся при его изготовлении или при изготовлении кабеля из волокон [142—144].

Такие неоднородности, как правило, приводят к возникновению взаимодействия между модами распространения и паразитными модами и тем самым увеличивают коэффициент линейного затухания, зависящий от z , и значительно усложняют интерпретацию различных измерений. Напомним (гл. 8, § 7), что при наличии взаимодействия мод *распределение энергии между модами* и характеристики распространения (линейные потери, затухание, диаграмма излучения) *стабилизируются на расстояниях*, превышающих так называемую *дистанцию равновесия* L_e . Оно может быть различным: от нескольких сантиметров для волокна со многими дефектами до нескольких километров для высококачественного волокна.

§ 2. Измерение потерь

В многокилометровой системе дальней связи ослабление, измеренное до достижения дистанции равновесия, обычно меньше ослабления, наблюдаемого на расстояниях, превышающих дистанцию равновесия. Кроме того, ослабление, измеренное до достижения равновесия, сильно зависит от условий ввода оптической волны в волокно и может изменяться от измерения к измерению. Чтобы исключить такие эффекты и получать результаты, которые позволяли бы предсказывать поведение волокна на участке передачи, принято измерять линейное затухание на расстояниях, *превышающих дистанцию равновесия* [144].

А. Прямой метод

Данный метод (рис. 9.1) состоит в простом сравнении мощностей P_1 и P_2 , переданных на расстояния L_1 и L_2 , что позволяет исключить влияние условий ввода излучения в волокно. Источником света может служить лампа с монохроматором или по-

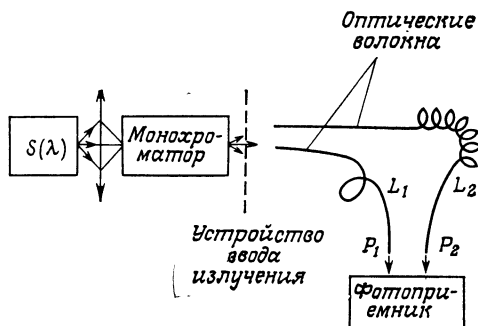


Рис. 9.1. Измерение ослабления.

Измерив мощности P_1 и P_2 , соответствующие длинам L_1 и L_2 , можно определить линейный коэффициент ослабления.

лупроводниковый лазер (гл. 12, § 4). Мощности P_1 и P_2 измеряют при помощи фотодиода (относительное измерение) или микроваттметра (абсолютное измерение), после того как убе-

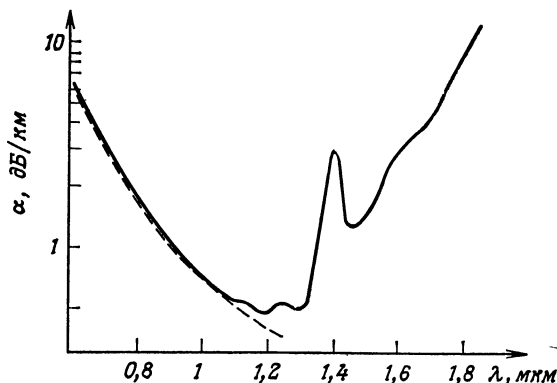


Рис. 9.2. Пример зависимости потерь в волокне от длины волны.

Сердцевина — фосфоросиликатная, оболочка — боросиликатная. Штриховая линия соответствует релеевскому закону $1/\lambda^4$. Краткое объяснение хода приведенной кривой можно найти в гл. 10, § 1.

дятся в достижении равновесия (например, установив, что диаграмма излучения на выходе волокон не зависит от условий ввода). Если же оно не достигнуто, то можно усилить взаимодействие мод, вставив волокно в устройство, называемое *смесителем мод* (имитатором равновесия) [145]. Измерив P_1 и P_2 , можно вычислить линейные потери α на длине волны λ , исходя из соотношения

$$P(z) = P_0 e^{-\alpha z},$$

откуда получаем

$$\alpha_{\text{дБ/км}} = \frac{10}{L_2 - L_1} \lg \frac{P_1}{P_2}. \quad (9.1)$$

Типичная зависимость величины α от длины волны λ представлена на рис. 9.2.

Б. Метод обратного рассеяния [146, 147]

При таком методе на одном конце волокна вводят короткий и мощный импульс излучения и затем *на этом же конце* наблюдают сигнал, отраженный разными рассеивающими цен-

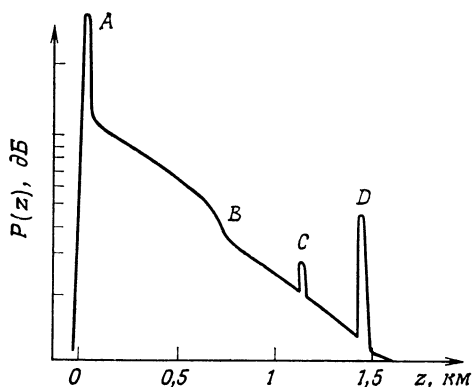


Рис. 9.3. Запись сигнала обратного рассеяния.

A — сигнал, отраженный от входного торца волокна; *B* — зона с повышенным ослаблением; *C* — сигнал, отраженный от точечного дефекта в волокне; *D* — сигнал, отраженный от выходного торца волокна.

трами. Если предположить, что дефекты распределены равномерно, то оптическая мощность отраженного оптического сигнала на входе как функция времени t ($t = 2nz/c$) выражается следующим образом:

$$P(t) = kP(0) e^{-2 \int_0^z \bar{\alpha} dz}. \quad (9.2)$$

Это соответствует сумме отражений от каждого из элементов волокна длиной dz , расположенного на расстоянии z от его конца, претерпевших ослабление при обратном пробеге. Измерив $P(t)$, можно вычислить средний коэффициент затухания $\bar{\alpha}$, который для некоторого элемента длины пробега $L_2 - L_1$ определяется в таком виде:

$$\bar{\alpha} = \frac{1}{2} \frac{\int_{L_1}^{L_2} (\alpha' + \alpha'') dz}{L_2 - L_1}, \quad (9.3)$$

где α' и α'' — линейные затухания при пробеге волны туда и обратно (они не обязательно одинаковы, поскольку обратное рассеяние может изменить модовый состав),

Данный метод хорош тем, что он позволяет получить не только полные потери на участке от L_1 до L_2 , но и *зависимость затухания $\bar{\alpha}(z)$ от z* (рис. 9.3). Главный же недостаток такого метода — низкий уровень отраженного сигнала (примерно на 40 дБ слабее входного сигнала).

Среди трудностей применения метода обратного рассеяния для измерения потерь в волокне можно назвать следующие.

1. Уменьшение отраженного сигнала на входе, обусловленное различием показателей преломления на входном диоптре. Эти потери можно уменьшить, если использовать на входе поляризованный источник излучения и регистрировать отраженный свет с применением скрещенного поляризатора или же если опустить конец волокна в сосуд с жидкостью, выравнивающей показатели преломления.

2. Трудности контроля за условиями ввода излучения в волокно.

3. Узкий динамический диапазон измерений и возможное насыщение детектора, если плохо скомпенсирован эффект, отмеченный в п. 1.

4. Высокий уровень шума в фотоприемнике в сравнении со слабым полезным сигналом.

§ 3. Измерение профиля показателя преломления

Существует целый ряд методов измерения профиля показателя преломления волокна.

А. Интерференционная микроскопия [148]

Оптическая разность хода, вносимая слоем толщиной e с показателем преломления n , равна $(n - 1)e$. Поэтому можно измерять изменение величины n , поддерживая толщину e постоян-

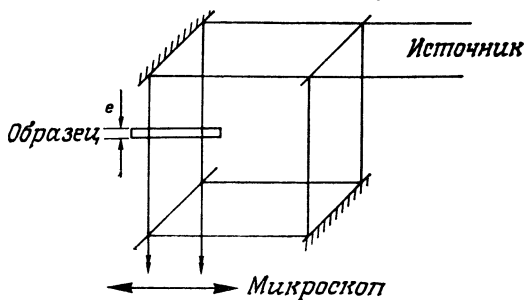


Рис. 9.4. Схема интерференционного микроскопа (интерферометра Маха — Цендера).

ной. С этой целью делают срез волокна, доводят полировкой его толщину до 100 мкм и помещают в одно из плеч интерферометра (рис. 9.4). По интерференционной картине рассчитывают профиль показателя преломления в срезе. Такой метод, довольно

сложный в применении (тонкая обработка образцов, трудность интерпретации полученной интерференционной картины), оказывается весьма точным (относительная погрешность измерений не более $10^{-4} - 10^{-3}$).

Б. Растровый электронный микроскоп

Такой микроскоп позволяет анализировать распределение примеси, и, если известно соотношение между концентрацией примеси и показателем преломления (что бывает не всегда), по распределению примеси можно рассчитать профиль показателя преломления. Здесь тоже приготовление образцов связано с трудностями, а, кроме того, метод неприменим при наличии нескольких примесей.

В. Измерение коэффициента отражения [149]

Соотношение между мощностью, отражаемой диоптром (торец волокна — воздух), и показателем преломления дается формулами Френеля. Поскольку размеры волокна малы (несколько десятков микрон), приходится применять сходящиеся пучки, чем ограничивается точность измерений. Прочие трудности связаны с получением хорошего качества поверхности торца волокна и достаточно высокого отношения сигнала к шуму.

Г. Метод ближнего поля [150]

Мы установили, что световая мощность, которую можно ввести в одной точке волокна, пропорциональна локальной числовой апертуре (гл. 8, § 3, п. А, 5). Если потери всех мод одинаковы (нет мод утечки и взаимодействия мод) и все моды одинаковым образом возбуждаются на входе, то мощность в любой точке выходного торца волокна можно вычислить по формуле

$$P(r) = P(0) \frac{n^2(r) - n^2(0)}{n_1^2 - n_2^2}. \quad (9.4)$$

К сожалению, этим очень простым выражением нельзя пользоваться без поправок, поскольку при малой длине волокна неизбежны моды утечки (гл. 8, § 3, п. А, 3). Следовало бы, очевидно, ввести поправочный множитель, который учитывал бы длину волокна, его диаметр и т. д. Но на практике можно сильно ослабить влияние мод утечки путем точного согласования числовой апертуры источника с апертурой волокна.

Схема измерения представлена на рис. 9.5. На входном торце отрезка волокна длиной от 0,5 до 2 м при помощи объектива микроскопа формируют изображение источника. Изображение выходного торца, сформированное объективом с большим относительным отверстием, сканируется при помощи малого фотодиода, перемещающегося по диаметру волокна. Для повышения точности измерений предусматриваются полосовые фильтры и ограничение пучка. Чтобы упростить процедуры измерения, можно заменить фотодиод передающей телевизионной

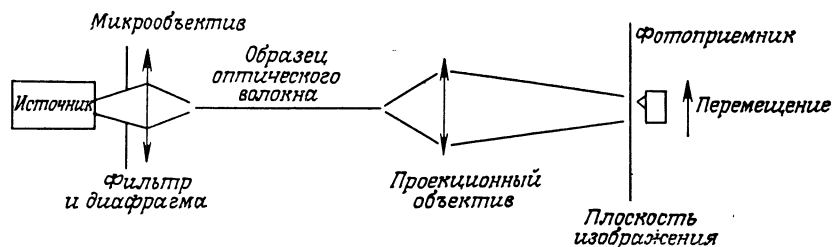


Рис. 9.5. Схема измерения показателя преломления $n(r)$ методом ближнего поля.

камерой и анализировать видеосигнал. Примеры данных, полученных таким методом, приведены на рис. 9.6.

Метод ближнего поля прост в осуществлении и поэтому широко применяется.

Производный метод (метод преломленного поля), разработанный Стюартом [151], основан на том, что на первых миллиметрах распространения исчезают все лучи, кроме лучей распространения и преломления. Кроме того, лучи распространения, но с потерями проникают из сердцевины в оболочку под весьма малым углом распространения и поэтому остаются вблизи границы между сердцевиной и оболочкой; их легко исключить при помощи небольшого кольца, центрванного на волокне. Принцип измерения весьма прост. Волокно помещают в пучок света, сходящийся в одной из точек его входного торца. Оно отбирает у этого пучка некоторую энергию, соответствующую лучам распространения. Затем измеряют световую мощность, оставшуюся после введения волокна, экранируя свет, передаваемый модами утечки, непрозрачным кольцом.

Применяется такой вариант метода [152]. Очень короткий отрезок волокна (2-3 см) погружают в иммерсионную кювету (что позволяет рассматривать оболочку как бесконечно толстую). Перемещая точку падения по диаметру волокна, можно найти дополнение к величине $P(r)$, определяемой выражением (9.4), без поправочного множителя. Такой метод значительно

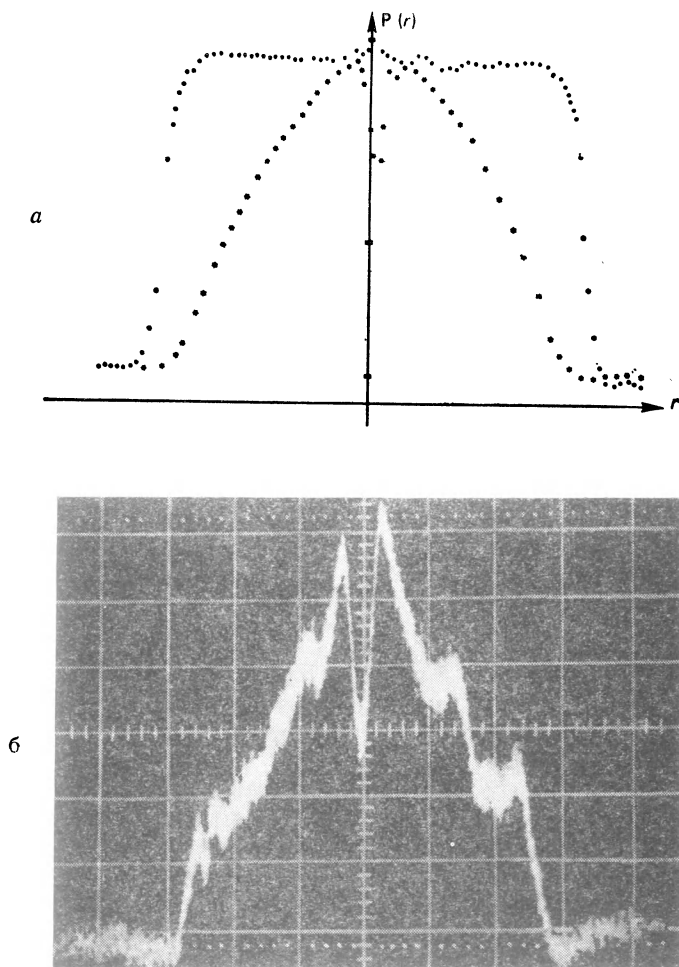


Рис. 9.6. Метод ближнего поля (типичные экспериментальные данные).

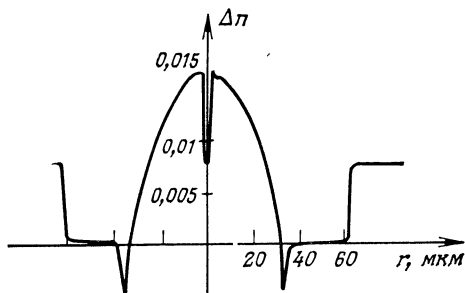
a — профиль показателя преломления, полученный при помощи фотодиода, в волокне со скачком показателя преломления (темные кружки) и в градиентном волокне (крестики, нижняя кривая). Провал в центре обусловлен дефектом изготовления волокна; *б* — осциллограмма видеосигнала строки, полученная при сканировании передающей телевизионной камерой.

точнее метода ближнего поля в пограничных зонах сердцевин и оболочки, где последний дает плохое разрешение (рис. 9.7).

Существуют также методы, основанные на анализе картины дифракции и требующие обработки данных на ЭВМ. Мы не будем останавливаться на этих методах, гораздо более сложных

Рис. 9.7. Профиль показателя преломления, полученный методом преломленного поля.

Здесь хорошо видны края волокна, незаметные на рис. 9.6.



[153, 154]. Отметим лишь то их преимущество, что они *не требуют разрушения исследуемого волокна*.

§ 4. Измерение числовой апертуры

Речь идет о том, чтобы определить *максимальный угол распространения* θ_c (или Θ_c), т. е. [формула (8.16)] максимальную разность показателей преломления

$$\text{ЧА} = \sqrt{n_1^2 - n_2^2}.$$

На практике из-за различия затуханий в сердцевине и оболочке и взаимодействия мод (гл. 8, § 7), кроме *теоретической максимальной числовой апертуры*, вводится еще *эффективная числовая апертура*

$$\text{ЧА}_{\text{эфф}} = \sin \Theta_m, \quad (9.5)$$

где Θ_m — половина угла при вершине конуса, в пределах которого передается 90% энергии. Этот параметр зависит от условий возбуждения и длины используемого волокна. Чтобы исключить эту зависимость, нужно измерять Θ_m после достижения условий равновесия. Измерение состоит в том, что снимают угловую диаграмму излучения в дальней зоне и затем по диаграмме вычисляют угол конуса, содержащего 90% излучаемой энергии.

§ 5. Измерение дисперсии, обусловленной материалом

В литературе имеются некоторые данные о дисперсии, но они не всегда согласуются между собой из-за различий в результатах измерений, проведенных на объемном образце

материала и на тонком волокне или на материале одного состава, но разного производства [155].

Эти различия обусловлены, с одной стороны, температурными условиями приготовления исследуемого образца, а с другой — нелинейной зависимостью дисперсии от состава [156, 157]. Важный параметр, характеризующий *дисперсию профиля показателя преломления*, был введен в гл. 8, § 5 [формула (8.53)], а именно:

$$P = \frac{n_1}{N_1} \frac{\lambda}{\Delta} \frac{d\Delta}{d\lambda},$$

где n_1 — показатель преломления на оси сердцевины, а $N_1 = n_1 - \lambda(dn_1/d\lambda)$ — групповой показатель [гл. 2, § 5 и формула (8.52)].

Один из методов измерения был разработан в Саутгемптонском университете [157]. Он состоит в том, что измеряют числовую апертуру, т. е. параметр Δ , исследуемого волокна в зависимости от длины волны, а затем находят соответствующие этой зависимости коэффициенты разложения Селлмейера

$$\text{ЧА}^2 = 2\Delta n_1^2 = A + B\lambda^2 + C\lambda^4 + \frac{D}{\lambda^2} + \frac{E}{\lambda^4}.$$

Аппроксимируя (по методу наименьших квадратов) этим выражением экспериментальную кривую, путем дифференцирования находят параметр P . О результатах такого исследования мы уже говорили (см. рис. 8.22). В изложенной выше теории неявно предполагается, что параметр P не зависит от состава материала. Чтобы проверить это, нужны дополнительные исследования.

§ 6. Измерение импульсного отклика и передаточной функции

Мы уже говорили (гл. 8, § 6; [129]) о том, что оптическое волокно эквивалентно фильтру, линейному по оптической мощности. Вследствие этого две функции — импульсный отклик и передаточная функция — оказываются связанными между собой преобразованием Фурье. Достаточно измерить одну из них, и тогда другую можно рассчитать. Напомним также о важном значении пространственного распределения энергии в волокне, которое должно быть максимально приближено к равновесному (§ 1).

Общая схема измерения представлена на рис. 9.8.

А. Измерение импульсного отклика [158 — 160]

Лазерным диодом можно управлять посредством очень коротких импульсов тока, превышающих пороговый ток (гл. 12, § 4). Это позволяет получать световые импульсы длительностью менее 200 пс (на половине высоты). Частота повторения импульсов выбирается так, чтобы их можно было наблюдать на экране осциллографа (например, 10 кГц). Корректное введение света на входе в волокно обеспечивается при помощи микрообъектива или цилиндрических линз, компенсирующих рас-

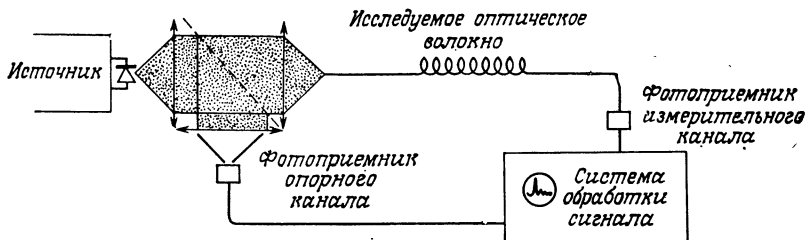


Рис. 9.8. Схема измерения импульсного отклика или частотной характеристики волокна (в зависимости от используемой системы обработки сигнала).

ходимость лазерного излучения. На выходе волокна свет регистрируется фотоприемником, за которым следует система визуализации и обработки сигнала: стробоскопический осциллограф, присоединенный через интерфейс к ЭВМ, которая по измеренному импульсному отклику вычисляет передаточную функцию в соответствии с программой быстрого преобразования Фурье (например, типа Кули — Тьюки).

Принцип возбуждения волокна при помощи лазерных диодов наряду с большим преимуществом возможности получения коротких и мощных импульсов на длине волны, выбранной для линии связи, имеет и недостатки. Вот некоторые из них: изменение числа возбуждаемых мод, дрейф центральной длины волны, изменение диаграммы направленности излучения. Чтобы хотя бы отчасти справиться с этими трудностями, можно использовать в качестве источника *лазерный диод, соединенный с возбуждающим волокном* длиной в несколько десятков сантиметров.

Другое возможное решение состоит в использовании лазеров с синхронизацией мод (например, ИАГ-лазера; гл. 12, § 3, п. В).

Б. Измерение передаточной функции

Если в системе обработки сигнала заменить осциллограф спектроанализатором, то можно сразу получить *модуль передаточной функции*. В самом деле, согласно теореме Планшереля,

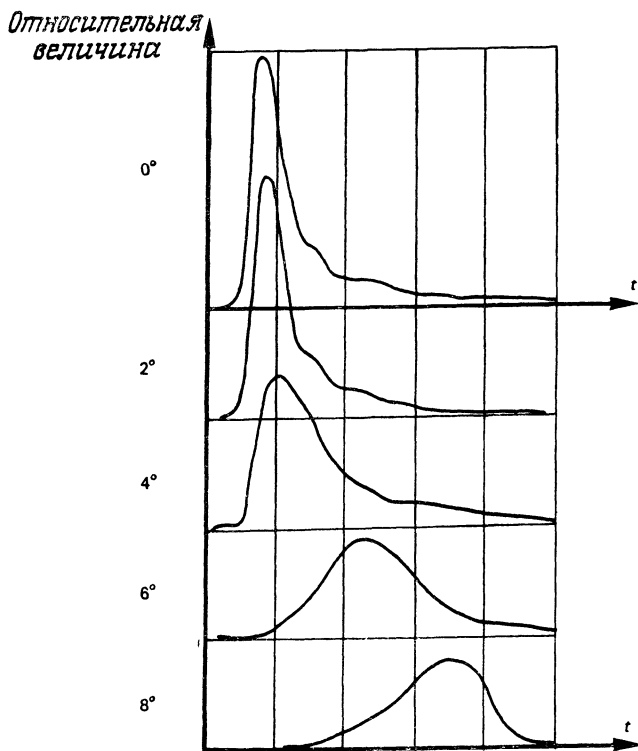


Рис. 9.9. Изменение импульсного отклика волокна.

Световой импульс $[0^\circ]$ вводится в исследуемое волокно в виде пучка света с числовой апертурой, вдвое меньшей, чем у волокна, которое поворачивают вокруг его оси. Среднее время пробега, а также форма и ослабление импульса изменяются в зависимости от угла поворота.

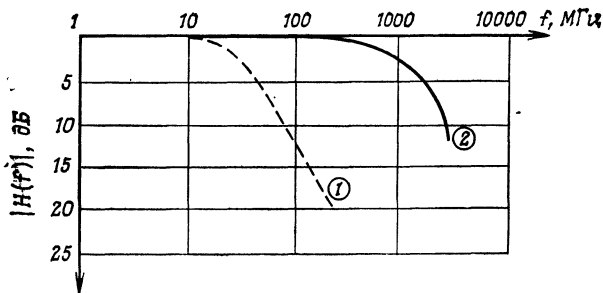


Рис. 9.10. Передаточные функции волокна, полученные экспериментально.

1 — волокно со скачкообразным изменением показателя преломления; 2 — градиентное волокно.

достаточно измерить модули спектров импульсов на входе $E(f)$ и выходе $S(f)$ волокна, выраженные в децибелах, чтобы получить модуль передаточной функции в виде разности

$$|H(f)|_{\text{дБ}} = |S(f)|_{\text{дБ}} - |E(f)|_{\text{дБ}}.$$

Вычислив затем на ЭВМ фурье-образ свертки

$$s(t) = h(t) * e(t),$$

можно найти полную передаточную функцию (модуль и фазу).

Можно измерить эти две величины более прямым способом, предусмотрев *синусоидальную модуляцию* лазерного источника и сравнивая излучение на входе и выходе исследуемого волокна. Для этого необходим фотоприемник с синхронным усилителем, который непосредственно сравнивал бы два сигнала [161, 162]. При таком методе приходится одновременно регистрировать измеряемый (выход волокна) и опорный (его вход) сигналы, а поэтому важное значение имеет линейность фотоприемников и требование использовать одни и те же части телесного угла излучения, чтобы обеспечить возбуждение волокна и фотоприемника в одних и тех же условиях. В идеальном случае частотная характеристика опорного канала идентична частотной характеристике канала, в котором находится испытываемое волокно.

Вообще говоря, необходима коррекция искажений, обусловленных дисперсией материала, но легче контролировать ширину спектра источника, если выбрать рабочую точку лазерного диода на линейном участке при малой глубине модуляции.

Таким методом можно проводить измерения на частотах до 1,5 ГГц, а на более высоких частотах затрудняется управление лазером (гл. 12 и 14).

Некоторые экспериментальные данные приведены на рис. 9.9 и 9.10. Они дополнительно иллюстрируют некоторые положения, высказанные в гл. 8. В частности, отметим следующее:

1) *в волокнах с градиентом показателя преломления меньше разброс времени пробега, что равнозначно большей ширине полосы пропускания;*

2) *при неоптимизованном градиенте показателя преломления время распространения зависит от среднего угла падающего пучка.*

§ 7. Измерение модовой дисперсии

Как мы установили, информационная пропускная способность многомодовых оптических волокон ограничивается в основном двумя факторами:

1) различием группового времени распространения у разных возбужденных мод (модовой дисперсией);

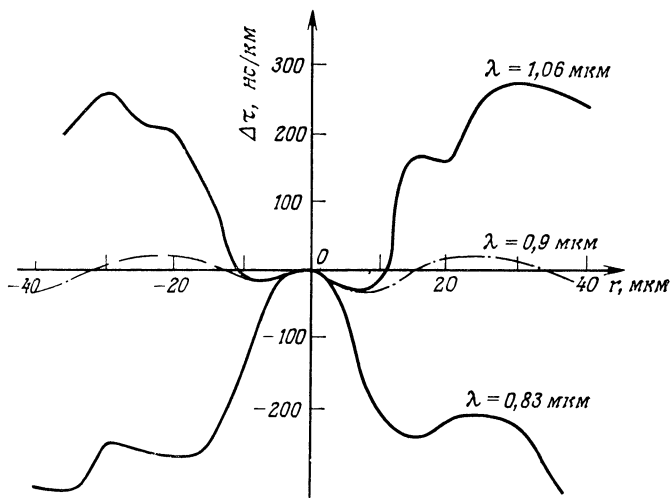


Рис. 9.11. Зависимость группового времени распространения $\Delta\tau$ от параметра моды.

2) зависимостью скорости распространения излучения от длины волны, обусловленной, как правило, материалом волокна (внутримодовой дисперсией).

Точность измерения показателя преломления такова, что не позволяет точно рассчитать модовую дисперсию в пределах оптимального профиля. Но, как говорилось в гл. 4, § 3, п. В, среда с квазипараболическим законом изменения показателя преломления компенсирует естественную расходимость гауссова пучка [22]. В таком случае можно показать, что гауссов пучок будет распространяться по волокну, сохраняя свой диаметр, и его средний луч будет вести себя как «геометрический луч» с параметрами r_0 или θ_0 , определяемыми выражением (8.54). Перемещая средний луч по входному торцу волокна, можно изменять такой параметр возбуждаемой моды и следить за изменением ее времени распространения.

С другой стороны, мы отмечали необычно сильную зависимость от параметра α (рис. 8.20 и 8.21). При изменении длины волны, на которой производится измерение, кривая величины $\Delta\tau_{\text{макс}}$ также изменяется, и мы можем найти оптимальную длину волны для исследуемого образца волокна и установить, в каком направлении нужно изменять α при изготовлении волокон, чтобы они были оптимальными на длине волны, выбранной для линии связи. Сказанное иллюстрируется данными, представленными на рис. 9.11 [163, 164].

ГЛАВА 10

ТЕХНОЛОГИЯ ИЗГОТОВЛЕНИЯ ОПТИЧЕСКИХ ВОЛОКОН

На основании сказанного ранее о распространении света можно предвидеть ряд трудно разрешимых проблем, которые возникают при изготовлении оптических волокон. Первая из них — проблема получения материалов, достаточно чистых для того, чтобы потери при передаче оптических сигналов не превышали нескольких децибел на километр. Прочие проблемы связаны с весьма малым диаметром волокон и высокими требованиями к его постоянству, с требуемой высокой точностью профиля показателя преломления. Наконец, совместимость с методами изготовления кабелей, защита от механических воздействий и механическая прочность волокон также имеют большое значение. Указав на разные причины потерь в оптических волокнах и их влияние на выбор способа получения волокон, мы весьма бегло коснемся промышленных методов получения заготовок, техники вытягивания из них оптических волокон и изготовления волоконно-оптических кабелей.

§ 1. Потери в оптических волокнах

Мы не будем здесь говорить о причинах потерь, связанных со структурой световода, таких, как потери мод утечки или потери из-за изгибов волокна, но рассмотрим только те из них, которые обусловлены качеством самого материала и способом изготовления волокна. Это потери двух видов: с одной стороны, потери, обусловленные *поглощением*, а с другой — потери, обусловленные *рассеянием* света при его распространении в волокне (гл. 2, § 4).

А. Потери за счет поглощения

В чистом стекле возможны два основных механизма поглощения световой энергии. Во-первых, из-за взаимодействия фотонов с колебательными уровнями молекул возникают полосы поглощения в инфракрасной области спектра, причем «фундаментальные» полосы такого поглощения лежат, как правило, выше 6 мкм. Во-вторых, существует резкий порог поглощения, когда энергия фотонов становится достаточной для того, чтобы вызвать переход электронов на более высокие уровни. Такой

вид поглощения наблюдается по большей части в ультрафиолетовой области ($\lambda < 0,4$ мкм).

Но в большинстве стекол имеются малые примеси, с чем связаны дополнительные потери, обусловленные переходами электронов между разными энергетическими уровнями, относящимися к незаполненным оболочкам металлов (Ti, Co, Fe, Cu, Ni, Mn, Cr, V), что дает широкий максимум поглощения и является причиной окрашенности стекол и кристаллов. В этом случае потери зависят главным образом от концентрации примеси и от состояния окисления металла. Состояние окисления ионов металла в значительной мере определяется технологией изготовления.

Центры окраски служат также центрами поглощения.

Гидроксильные ионы OH^- , основная частота колебаний которых соответствует примерно 2,73 мкм, также оказываются источником значительных потерь, в особенности 3-я гармоника (длина волны 0,91 мкм). На практике следует делать все возможное для того, чтобы получить очень чистое стекло, с уровнем примесей порядка 10^{-6} .

Б. Потери при рассеянии

Причин рассеяния света в оптических волокнах много. Ранее уже упоминалось *рэлеевское рассеяние*, неизбежное в любом материале и обусловленное флуктуациями показателя преломления волокна и наличием в нем микродефектов, размеры которых соизмеримы с длиной волны. Зависимость соответствующих потерь от длины волны следует закону λ^{-4} (гл. 2, § 4, п. А).

Кроме того, свет в волокне рассеивается *на пузырьках, микротрещинах, расстеклованных зонах*. Такие центры рассеяния особенно вредны на границе сердцевины с оболочкой в волокнах со скачком показателя преломления. Они приводят к ослаблению света в волокне, связывая моды распространения с модами потерь.

Можно значительно уменьшить концентрацию дефектов в волокне и ослабить их влияние, если при изготовлении строго соблюдать некоторые правила: никакой пыли, никаких неосторожных манипуляций с заготовками до изготовления из них волокон.

§ 2. Методы изготовления

Наряду с давно известными методами получения стекла, иногда весьма тонкими, разработаны новые методы, сближающиеся с технологией полупроводников вследствие необходимости иметь стекла с высокой степенью чистоты [165--167].

А. Стекольные методы

1. Метод двойного тигля

Самые классические методы получения стекла состоят в плавлении исходного сырья (стеклообразующее вещество SiO_2 , B_2O_3 и т. д. и модифицирующие прибавки окислов металлов Na_2O , CaO и т. д.), его проваривании (осветлении) и вытягивании волокна из расплава. На этом принципе основан метод изготовления оптического волокна с использованием двойного тигля (рис. 10.1), во внешнюю часть которого помещают стекло

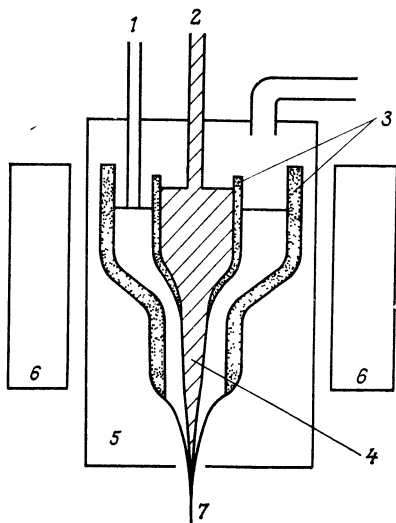


Рис. 10.1. Метод двойного тигля.

1 — стержень, подаваемый в тигель для оболочки; 2 — стержень, подаваемый в тигель для сердцевины; 3 — двойной тигель; 4 — зона ионного обмена; 5 — камера; 6 — печь; 7 — оптическое волокно.

для оболочки, а во внутреннюю — стекло для сердцевины. Можно получить градиент показателя преломления, если создать зону, в которой стекло оболочки и стекло сердцевины соприкасались бы в жидком и вязком состоянии между двумя фильерами. В таком случае может происходить катионный обмен между примесями, например между таллием и натрием. Данный метод был разработан фирмой Nippon Sheet Glass, изготавливающей оптические волокна типа Selfoc [168].

Поскольку конечное распределение примесей определяется диффузией, при постоянных времени, характерных для описанной технологии, трудно осуществить оптимальный профиль.

2. Метод раздвигания фаз [169]

Гораздо чаще, чем расплав, используется заготовка из стекла, которой придают в поперечном сечении нужную геометрию волокна. Для получения волокна ее достаточно подвергнуть вытяжке. Первый метод изготовления такой заготовки — это

метод разделения фаз и выщелачивания. Разработанный двадцать лет назад, этот метод приспособлен для получения оптических волокон: из боросиликатной ванны вытягивают цилиндрический стержень диаметром около 10 мм и длиной до 1 м. Этот стержень имеет такой состав, что при определенной температуре образуются две взаимосвязанные фазы, одна из которых богата двуокисью кремния, а другая — основанием. Эта доминирующая фаза, которая бедна двуокисью кремния, содержит большую часть окислов бора и натрия исходных стекол, а также и окислы переходных элементов: железа, хрома, никеля.

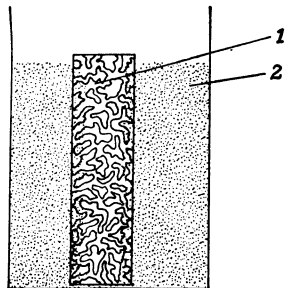


Рис. 10.2. Метод разделения фаз.

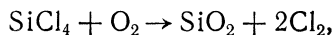
1 — пористый стержень; 2 — раствор стекла с показателем преломления более низким, чем у кварца. Исходный пористый стержень, практически не имеющий щелочной фазы, позволяет получить градиентный профиль показателя преломления путем пропитки его раствором, обеспечивающим более низкий показатель преломления.

Эту фазу можно удалить путем выщелачивания; тем самым обеспечивается очистка исходных материалов. В результате получается стержень с большим содержанием кремния, имеющий открытые поры, так что его можно пропитать раствором или подвергнуть специальной обработке, чтобы получить заготовку, имеющую градиентный состав, а следовательно, и градиентный показатель преломления. Данный метод тем более интересен, что при разделении фаз вследствие сегрегации примесей они концентрируются в щелочной фазе, которая затем удаляется. Таким образом, это и метод очистки, и метод изготовления (рис. 10.2). Первые результаты показали, что после удаления щелочной фазы потери уменьшились с 5000 до 100 дБ/км.

Далее мы изложим методы, позволяющие получать профиль с любым законом изменения показателя преломления путем последовательного наложения очень тонких слоев, показатель преломления которых можно изменять.

Б. Окисление в газовой фазе

Для получения оптических волокон обычно берут стекла с большим содержанием двуокиси кремния, но их высокая вязкость затрудняет применение обычных стекольных методов ввиду необходимости получения очень низкого содержания примесей (порядка 10^{-7}). Поэтому приходится получать стеклянные частицы SiO_2 с помощью химических реакций, таких, как



используя при этом два классических способа: окисление и гидролиз.

Добавляя примеси, можно изменять показатель преломления материала, а давление паров соединений кремния значительно больше давления паров гидридов и хлоридов переходных металлов, благодаря чему снижается возможность загрязнения.

Это «внутренний» метод, который был применен первым и который впервые позволил получить оптическое волокно с потерями меньше 20 дБ/км [170].

В горелку подают смесь примесей в виде хлоридов с SiCl_4 , чистым кислородом и горючим газом. Пары гидролизуются в пламени, продукты гидролиза переносятся во вращающуюся трубку и осаждаются на ее стенках. Трубку нагревают до температуры стеклования осажденного на ней пористого слоя, в результате чего образуется тонкая пленка силикатного стекла с примесью. Меняя концентрацию примесей, можно получить последовательность слоев с показателями преломления, изменяющимися в зависимости от расстояния до оси, т. е. градиент показателя преломления. Трубку вытягивают в волокно после предварительного смыкания ее полости.

Этот первый способ лег в основу других, более сложных.

1. Внешний способ (внешнее окисление в газовой фазе)

Этот способ представляет собой усовершенствованный вариант описанного выше, осуществленный фирмой Corning Glass

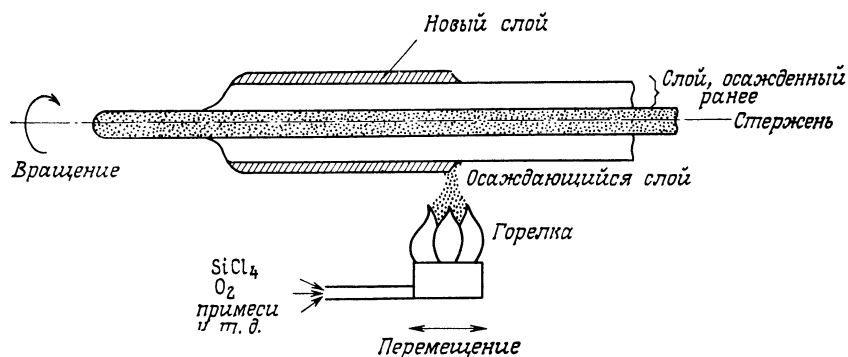


Рис. 10.3. «Внешний» способ.

Works [171]. Продукты гидролиза в пламени на этот раз осаждаются на поверхности стержня (рис. 10.3). Слой наращивают, вращая трубку над горелкой и перемещая ее вдоль оси. Градиент показателя преломления получают, изменяя концентрацию примесей, а слои стекла, которые образуют оболочку

волокна, осаждают в последнюю очередь. Одна из главных трудностей при таком способе — остекловывание полученного слоя. Затем удаляют стержень, смыкают просвет трубки и получают заготовку, т. е. стеклянный стержень с сечением, подобным сечению будущего оптического волокна.

2. Модифицированное химическое осаждение из газовой фазы

Это также усовершенствованный вариант первого способа, поэтому он называется еще «внутренним окислением в газовой фазе» [167, 172]. Газовая горелка, находящаяся снаружи (рис. 10.4), нагревает трубку до температуры, необходимой для окисления в газовой фазе. Трубка вращается и перемещается относительно горелки, как и в предыдущем случае.

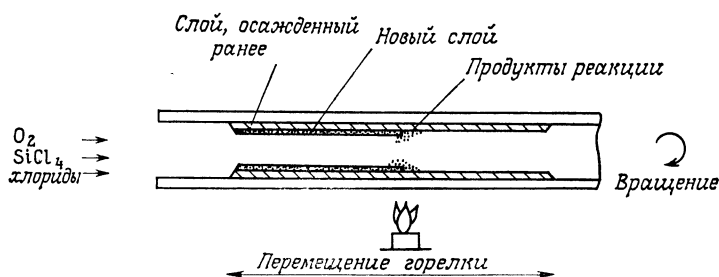


Рис. 10.4. Модифицированное химическое осаждение из газовой фазы.

В трубку вводится смесь необходимых для реакции газов (SiCl_4 , O_2 и примеси), и реакция происходит на стенках, нагретых до температуры $1300\text{--}1600^\circ\text{C}$, что вызывает осаждение стекла на внутренней поверхности. За один проход трубки наращивается слой толщиной в несколько микрон. Чтобы получить зону, которая после смыкания просвета трубки образует сердцевину, нужно ~ 50 оборотов. Оболочка же образуется из исходной трубки. В качестве основных добавок используются германий, бор, фосфор, фтор или их смесь. После смыкания просвета трубки получается заготовка.

3. Химическое осаждение из газовой фазы при помощи плазмы [173]

Так же как и в предыдущем способе, волокна с градиентом показателя преломления получают путем последовательного осаждения тонких слоев материалов с разными показателями преломления, образующихся в результате химической реакции в газе, проходящем по трубе. Отличие же от предыдущего спо-

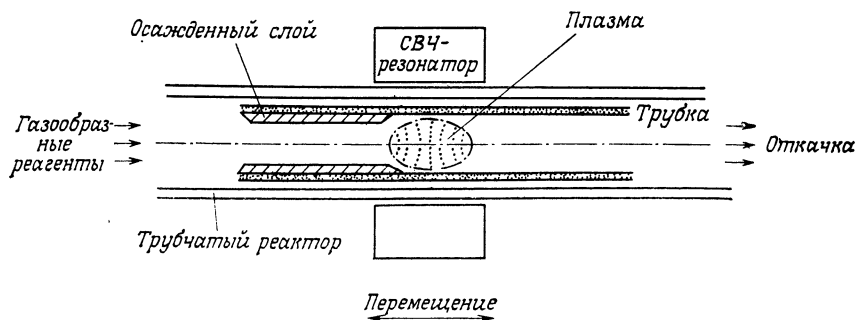


Рис. 10.5. Химическое осаждение с плазменным нагревом.

соба состоит в том, что химическая реакция вызывается неизо-термической плазмой низкого давления, создаваемой при по-мощи объемного СВЧ-резонатора, который находится вне трубки (рис. 10.5).

4. Осевое осаждение из газовой фазы

Этот метод, называемый также методом Вернейля [174], позволяет осуществить непрерывное изготовление заготовок, тогда как предыдущие методы дают заготовки длиной 30—40 см, самое большее 1 м. Заготовка растет в осевом направлении, а профиль показателя преломления задается путем изменения концентрации примесей по радиусу. Для получения частиц стекла используется такая же горелка, как и в описанном выше внешнем способе, что дает пористую заготовку, на которой при помощи еще одной горелки могут быть осаждены слои, образующие оболочку. Полученная пористая заготовка нагревается приблизительно до 1600°C , и в результате стеклования получается прозрачная заготовка (рис. 10.6).

§ 3. Вытягивание оптического волокна [165, 166]

В методе двойного тигля исходный материал в виде расплава непосредственно вытягивается в форме волокна. Другие методы изготовления волокон дают заготовки в промежуточном состоянии в виде заготовки, которая представляет собой стеклянный стержень, в сечении подобный волокну, но требующий дополнительного вытягивания для получения готового оптического волокна. На рис. 10.7 изображены различные элементы вытягивающего станка. Такой станок состоит, как правило, из следующих частей:

1) патрона для закрепления заготовки, допускающего ее перемещение;

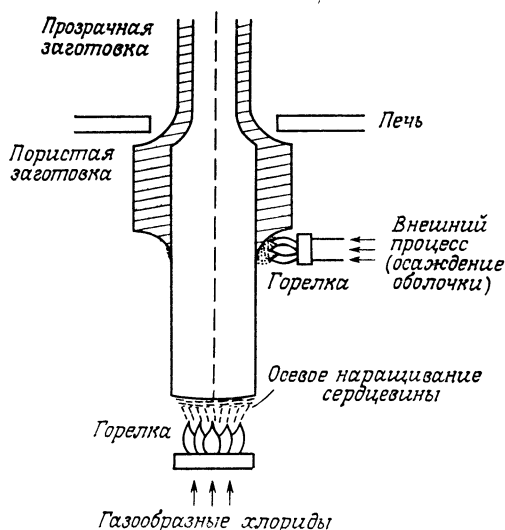


Рис. 10.6. Осевой метод



Рис. 10.7. Станок для вытягивания волокна.

2) печи (горелки, электронагревательной печи, ВЧ-печи) для размягчения заготовки: вещество становится текучим и дает нить;

3) принимающего барабана, на который наматывается волокно по мере вытягивания.

Между печью и барабаном обычно имеется система измерения диаметра нити и устройства для нанесения защитного покрытия волокна [например, слоя кайнара (поливинилиденфторида) толщиной в несколько микрон]. Скорость вытягивания — порядка 40 м/мин, при этом диаметр получаемых волокон составляет 50—150 мкм.

§ 4. Кварцево-силикатные волокна [175]

Оптическое волокно можно получить из простой кварцевой нити, если окружить ее оболочкой из силикатного стекла с показателем преломления, несколько меньшим, чем у кварца. Такой способ заслуживает внимания хотя бы из-за низкой стоимости волокна, но он пока не позволяет получать волокна с градиентом показателя преломления. Диаметр сердцевины, как правило, более значительный: 100—600 мкм. Кроме того, у таких волокон имеются другие существенные недостатки: низкое качество граничной поверхности между сердцевиной и оболочкой, неоднородность показателя преломления оболочки из силикатного стекла, повышенные потери при низких температурах.

ОПТИЧЕСКИЕ КАБЕЛИ И РАЗЪЕМЫ

Выше мы изложили технологию изготовления стеклянных нитей диаметром порядка 0,1 мм. Их механическая прочность весьма невелика, а поэтому приходится думать о способах их защиты. Начиная с нанесения первого слоя покрытия и кончая изготовлением кабеля, состоящего из десятков оптических волокон, технологу нужно искать все новые и новые способы, которые не повлияли бы на исходные характеристики волокна.

В зависимости от типа выбранного волокна и назначения кабеля — высокая или низкая информационная пропускная способность, передача сигналов на близкое или далекое расстояние, подземная, воздушная или подводная связь и т. д. — внешние воздействия будут различными и вызовут необходимость в создании различных типов кабелей.

Но волоконно-оптическая система связи (гл. 17) чаще всего состоит из отдельных световодов, соединенных концами; способы соединения волокон ставят нас перед необходимостью решать задачи точной механики. Качество соединения непосредственно связано с допусками при изготовлении, с отношением диаметров сердцевины и оболочки, с эксцентриситетом и некруглостью волокна и т. д. Решения, найденные для кабелей, отразятся и на оптических разъемах. Каждая структура кабеля имеет свой тип соединения, общепринятый или особый, который для него подходит.

§ 1. Характеристики волокон до изготовления кабеля

Стеклянные и кварцевые волокна характеризуются высокой прочностью на растяжение, но волокна, полученные сразу же после вытягивания, обнаруживают большое число трещин на внешней поверхности. При натяжении волокно ломается, так как эти трещины распространяются по прямому сечению, реже — из-за дефектов внутренней структуры, наличия микропузырьков. Кроме того, волокно чувствительно к пыли, влажности, к царапинам, поскольку все эти факторы, оказывая влияние на составляющие его материалы, ухудшают характеристики передачи оптических сигналов.

После вытягивания геометрические характеристики волокна (диаметр сердцевины и оболочки, эксцентриситет, профиль и

разность показателей преломления) уже заданы и улучшить их при нанесении защитных покрытий или при изготовлении кабеля невозможно. Поэтому очень важно соблюдать необходимые предосторожности в процессе вытягивания волокна.

А. Первичная защита волокна

Чтобы уменьшить влияние названных воздействий и предупредить разрушительное влияние внешней среды, сразу же после вывода из печи на волокно наносят очень тонкое (несколько микрометров) как можно более равное покрытие. Его назначение состоит в механическом заполнении микротрещин и предохранении волокна от влаги. При этом также на много порядков повышается прочность волокна на разрыв, что имеет первостепенную важность для дальнейших технологических операций (нанесение защитной оболочки, сборка кабеля)

Можно заранее сказать, что разность коэффициентов теплового расширения волокна и защитного слоя или же разных защитных слоев повысит нагрузку на волокно и, следовательно, приведет к росту потерь. Таким образом, выбор защитных материалов тесно связан с изменением температуры. Например, слой кайнара толщиной 4 мкм вносит дополнительные потери менее 10^{-2} дБ/°С км, тогда как потери, вносимые некоторыми другими пластмассовыми покрытиями, составляют $(7-8) \times 10^{-2}$ дБ/°С·км.

Б. Защитная оболочка (вторичное покрытие) волокна

В зависимости от типа кабеля (§ 2) волокна впоследствии покрываются защитной оболочкой или нет. Защитная оболочка делается по большей части на полиамидной, нейлоновой или полипропиленовой основе и чаще всего наносится путем экструзии на поточной линии. Различают два типа защитной изоляционной оболочки: тугая и свободная при внешнем диаметре в несколько десятых миллиметра.

Прочность на разрыв можно увеличить за счет распределения нагрузки между волокном и вторичным покрытием, толщину которого следует выбрать с учетом его коэффициента теплового расширения и его модуля Юнга.

§ 2. Кабели из оптических волокон

Были предложены и испытаны многие типы кабелей. Их можно разделить на три основные категории [176—178]: 1) *ленточная структура*, 2) *классическая структура*, 3) *цилиндрическая желобковая структура*.

Разумеется, в конкретных случаях возможны комбинации разных структур.

Упомянем также о пучках волокон (жгутах), которые исторически были использованы первыми. Но вследствие их низких качеств при передаче оптических сигналов, а также низкой механической прочности теперь ими пользуются только для освещения.

А. Кабели классической структуры [177, 179, 180]

Волокно в защитной оболочке обладает механической прочностью, позволяющей применять обычные способы изготовления кабеля: волокна располагаются вокруг центральной жилы при помощи специального сборочного механизма. Можно разместить на одной центральной жиле один или несколько слоев

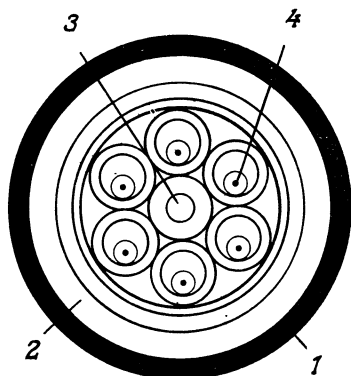


Рис. 11.1. «Классический» кабель из 6 оптических волокон.

1 — полиэтиленовая оболочка; 2 — алюминиевая трубка; 3 — медный несущий стержень или стеклянное волокно с покрытием; 4 — оптическое волокно с защитным покрытием.

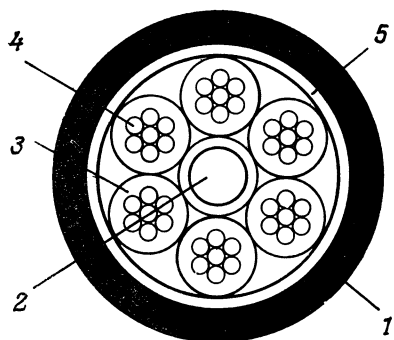


Рис. 11.2. «Классический» кабель, образованный 6 «многожильными» сборками из оптических волокон.

1 — полиэтиленовая оболочка; 2 — центральный металлический стержень; 3 — сборка из оптических волокон; 4 — оптическое волокно с защитным покрытием; 5 — алюминиевая трубка.

волокон (рис. 11.1, 11.6) или собрать кабель из нескольких многожильных элементов (рис. 11.2).

Наружная оболочка кабеля должна быть рассчитана на рабочие условия (проводка по трубе, прокладка в земле и т. п.), и ее материал следует выбирать, исходя из предполагаемых внешних воздействий.

Основные преимущества кабелей такого типа: волокно хорошо защищено, повышенная механическая прочность на разрыв, на сжатие и т. д., хорошо известны технология изготовления

и оборудование, довольно высокая температурная стабильность.

Из недостатков отметим следующие: громоздкость, поскольку упрочняется каждое отдельное волокно; высокая стоимость производства ввиду малой длины отрезков кабеля; необходимость предварительного удаления оболочки при индивидуальном соединении волокон.

Б. Кабели ленточной структуры [181, 182]

Этот тип оптического кабеля, разработанный фирмой Bell Laboratories, допускает весьма высокую плотность волокон в заданном поперечном сечении. Из волокон составляется матрица $m \times n$.

Матрица подразделяется на строки-ленты, которые легко накладываются друг на друга.

Лента состоит из уложенных параллельно n волокон, причем шаг между волокнами равен нескольким сотням микрометров. Волокна располагаются между двумя листами пластмассы, каждый из которых состоит из двух слоев: слоя полиэтилена с низким модулем упругости, который при нагревании прилипает к волокнам, и слоя полиэфира с высоким модулем упругости, который обеспечивает жесткость всей сборки (рис. 11.3).

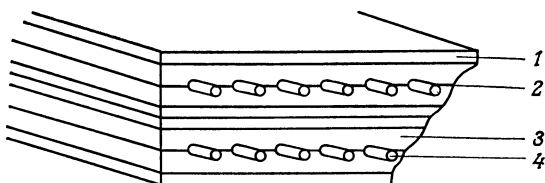


Рис. 11.3. Ленточный кабель в разрезе.

1 — слой полиэфирной пластмассы; 2 — граница двух слоев; 3 — полиэтиленовый слой; 4 — оптическое волокно.

Если наложить друг на друга m таких лент, то получим матрицу $m \times n$, которую затем скручивают и вводят как спираль в крепкую оболочку, армированную кордом с высоким сопротивлением на разрыв, который воспринимает внешние нагрузки, сохраняя матрицу неизменной.

Преимущества подобной структуры таковы:

1) возможность повышенной плотности упаковки волокон (пример: описанный кабель содержит $12 \times 12 = 144$ волокна при общем диаметре 12 мм [181]);

2) возможность одновременного присоединения всех $m \times n$ волокон;

3) меньшая стоимость, поскольку защита n волокон достигается одной операцией.

Из недостатков следует отметить:

- 1) наличие риска ломки волокон при изготовлении;
- 2) плохое распределение нагрузки на волокна при разных воздействиях на кабель: изменение затухания в волокне в этом случае будет сильно зависеть от положения волокна в матрице;
- 3) влияние температуры на столь компактную и неоднородную структуру;
- 4) сложность одновременного соединения $m \times n$ волокон.

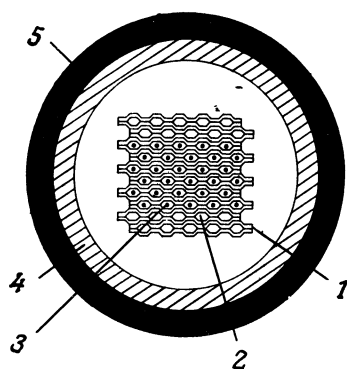


Рис. 11.4. Разновидность ленточного кабеля.

1 — лента из полиана; 2 — канал; 3 — оптическое волокно без второго защитного покрытия; 4 — алюминиевая трубка; 5 — полиэтиленовая оболочка.

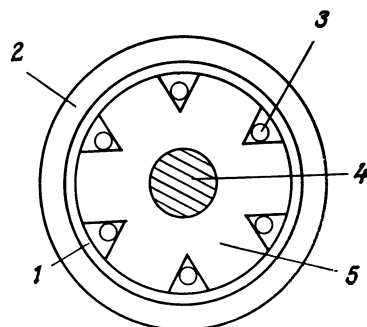


Рис. 11.5. Кабель цилиндрической желобковой структуры в разрезе.

1 — защитная лента; 2 — защитная оболочка; 3 — оптическое волокно без второго защитного покрытия; 4 — центральный несущий стержень; 5 — пластмассовый сердечник.

Иная разновидность ленточного кабеля (рис. 11.4), где лента образована ячейками из алюминия, позволяет избавиться от этих недостатков.

В. Кабели с желобковой цилиндрической структурой [183, 184]

В этом случае центральной частью кабеля является сердечник с желобами (пазами) в виде винтовой спирали, в которых в ненапрянутом состоянии располагаются волокна, предварительно покрытые защитной оболочкой (рис. 11.5).

Механическая прочность сердечника предохраняет волокна от всевозможных воздействий, которым подвергается кабель. Размеры и состояние поверхности желобов рассчитаны так, чтобы характеристики волокна оставались неизменными.

Защитный покров (ленточная обмотка) закрывает пазы, а функции механической защиты при эксплуатации выполняет внешняя оболочка, причем в кабеле может быть один или несколько сердечников описанного типа.

Желобковая цилиндрическая структура кабеля сочетает преимущества двух описанных выше: меньшая громоздкость, ограниченное число манипуляций с волокнами в процессе изготовления; возможность обеспечить одновременную защиту многим волокнам, что снижает стоимость. Преимуществом также является то обстоятельство, что волокно легко поддается фиксации в структуре, а также относительная легкость соединения сразу всех волокон каждого сердечника. Следует, однако, напомнить о необходимости производить операции с волокном без вторичного покрытия, о хрупкости волокон, ближайших к внешней оболочке кабеля, о необходимости располагать специальными сборочными механизмами.

§ 3. Оптические разъемы

Кабели из оптических волокон требуют надежных оптических разъемов. Для эффективного использования кабеля необходимо, во-первых, соединять между собой разные отрезки кабеля и, во-вторых, присоединять кабель на конечных пунктах к различным терминальным устройствам. По примеру кабелей с парами медных проводов были разработаны два описываемых далее метода: постоянное сращивание и разъемное соединение.

Если требования стойкости к внешним воздействиям (ударам, вибрациям, температурным условиям при хранении и использовании и т. п.) одинаковы как для классического соединения проводов, так и для оптического соединения волокон, то говорить об экранировании волокон друг от друга и перекрестной помехе между каналами в сущности не приходится. Однако весьма специфичной и трудной оказывается проблема ослабления сигнала соединением. Необходимо, чтобы такое ослабление было небольшим и не превышало ослабления в самом волокне, тем более что число соединений в системах очень велико. Это значит, что выгоднее использовать как можно более длинные кабели.

А. Основной принцип

В принципе соединение оптических волокон осуществить просто: достаточно совместить оси двух соединяемых волокон и торец одной сердцевины подвести к торцу другой. *На практике же возникает множество трудностей* [185]. В самом деле, очень малые размеры волокна в поперечнике (менее 500 мкм, рис. 11.6) вызывают необходимость в миниатюрных механических приспособлениях для удержания соединяемых волокон и в специальном наборе инструментов для монтажа.

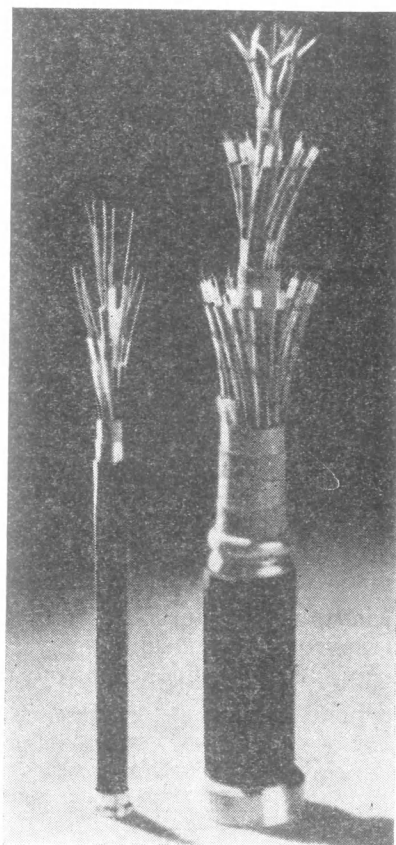


Рис. 11.6. Многожильный электрический кабель из коаксиальных парных проводов (справа) и соответствующий ему кабель из оптических волокон (слева).

1. Влияние нарушения взаимного положения волокон

Чтобы обеспечить минимальное ослабление сигнала в соединении, нужно выдерживать очень жесткие допуски на положение соединяемых волокон. Как это видно на рис. 11.7 и 11.8, зазор e , несовпадение осей a' и угловое отклонение θ должны быть крайне малы. (Кривые были получены для волокон с градиентом показателя преломления, но выводы остаются почти теми же для волокон со скачком показателя преломления).

Чтобы получить представление о порядке величины допустимых отклонений, возьмем волокно с числовой апертурой 0,2 и диаметром сердцевины 50 мкм. Тогда, чтобы потери в соединении не превышали 1 дБ, должны выполняться следующие условия: $e < 11$ мкм, $a' < 3$ мкм, $\theta < 2,2^\circ$.

2. Характеристики апертуры

Соединяемые волокна могут иметь различные характеристики как оптические (числовая апертура, профиль показателя преломления), так и геометрические (диаметр сердцевины).

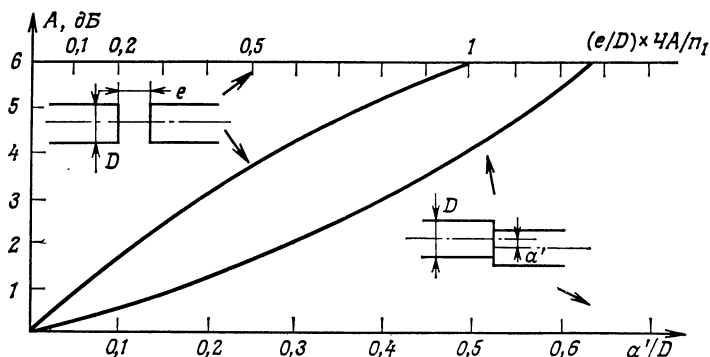


Рис. 11.7. Зависимость потерь в оптическом соединении от зазора между соединяемыми волокнами и их несоосности [186].

Очевидно, в частности, что переход света из одного волокна в другое, имеющее меньшую сердцевину или меньшую числовую апертуру, происходит с потерями, связанными с уменьше-

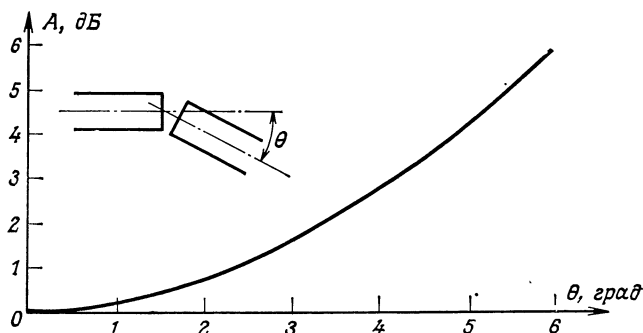


Рис. 11.8. Зависимость потерь в оптическом соединении от угла наклона между соединяемыми волокнами [187].

нием геометрических площадей (гл. 4, § 1, п. Д). В первом приближении потери пропорциональны отношению площадей сердцевин или отношению квадрата числовых апертур. Вернемся к рассмотренному выше числовому примеру. Если у «излучающего» волокна диаметр сердцевины составляет 50 мкм, то диаметр сердцевины «принимающего» волокна должен быть больше 45 мкм, чтобы потери не превышали 1 дБ (рис. 11.9). Кроме

того, как мы увидим дальше, большая часть соединений осуществляется путем совмещения наружных поверхностей волокна (оптической оболочки или первичного покрытия), а не сердцевин, которая в большинстве случаев оказывается недоступной. Из этого вытекает, что в случае недостаточной концентричности сердцевин относительно наружной поверхности

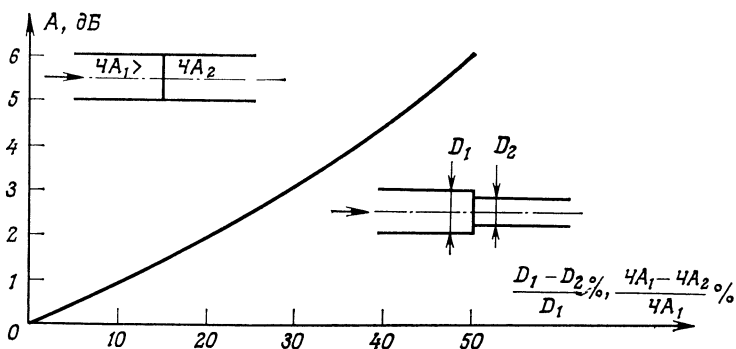


Рис. 11.9. Зависимость потерь в оптическом соединении от рассогласования числовых апертур (ЧА) и диаметров (D) соединяемых волокон.

(рис. 11.10) невозможно избежать несоосности сердцевин, а следовательно, и потерь. Поэтому от изготовителей требуется максимум усилий для того, чтобы свести к минимуму ошибки концентричности и эллиптичности волокон.

3. Влияние состояния поверхности

Все сказанное выше предполагает безупречное состояние торцевых поверхностей, и срез волокна должен быть строго перпендикулярен его оси. Для хорошего контакта между волокнами весьма важно, чтобы поверхности были гладкими. Для этого изготовители волокна разработали специальное режущее устройство, не такое простое, как широко используемые кусачки и инструменты для зачистки электрических кабелей, но тем не менее вполне пригодное для работы в полевых условиях, не требующее оптического контроля, полировки торцов волокон и дающее поверхности с удовлетворительными характеристиками.

Принцип работы этого устройства прост: волокно после удаления (химическим или механическим путем) защитных покрытий слегка натягивается на металлическом цилиндре (рис. 11.11). Алмазным острием делается надрез, и под действием натяжения волокно разрывается перпендикулярно своей оси [188, 189]. При таком распространении трещины состояние поверхности оказывается хорошим и не возникают осколки, неизбежные при обычных способах резки стекла.

Заметим, наконец, что из-за наличия тонкой воздушной прослойки между соединяемыми волокнами следует принимать в расчет и добавлять к предыдущим величинам внутренних потерь еще и френелевские потери на промежуточных поверхностях волокно — воздух и воздух — волокно. Для волокон, легированных кремнием, эти дополнительные потери составляют

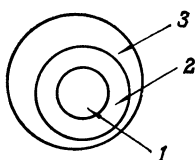


Рис. 11.10. Отсутствие концентричности между сердцевиной и наружной поверхностью волокна.

1 — сердцевина волокна; 2 — оптическая оболочка; 3 — наружная поверхность.

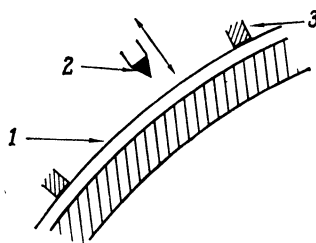


Рис. 11.11. Принцип действия приспособления для резки оптического волокна.

1 — волокно; 2 — алмазное острие; 3 — прижим.

примерно 0,5 дБ. Их можно устранить, смачивая концы волокон и заполняя воздушный промежуток жидкостью, согласовывающей показатели преломления, или, при стационарном соединении, сваривая волокна.

Б. Способы совмещения волокон

Экспериментально проверены многие способы совмещения волокон. При выборе того или иного способа следует руководствоваться следующими критериями:

а) *повторяемость результатов*: большая часть опробованных способов не дает хороших результатов при многократной разборке соединения;

б) *защита от влияния окружающей среды*: характеристики соединения не должны сильно зависеть от внешних условий;

в) *легкость осуществления*: все необходимое для выполнения операции совмещения должно изготавливаться легко и легко поддаваться монтажу;

г) *себестоимость*: должна исчисляться на многотысячных сериях, а не для одного опытного образца, который всегда стоит очень дорого.

Существующие методы можно, по-видимому, разделить на три большие категории [190] (рис. 11.12).

1. Волокно закрепляется в *наконечнике* очень точных размеров. Например, волокно можно фиксировать клеем внутри калиброванной трубки. Две такие трубки-наконечника совме-

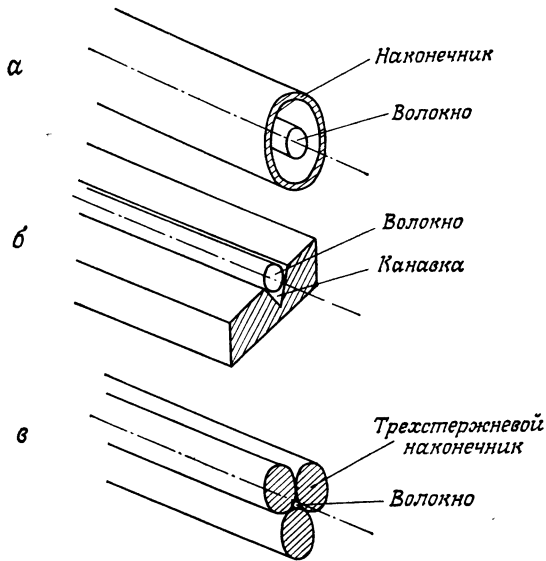


Рис. 11.12. Основные способы совмещения волокон при их соединении.

а — при помощи калиброванных трубчатых наконечников; *б* — при помощи наконечников с калиброванной V-образной канавкой; *в* — при помощи наконечников из 3 калиброванных цилиндрических стержней

щаются посредством внешнего, тоже калиброванного цилиндра. Преимущество такого способа в том, что здесь обеспечивается разъемное соединение, но он требует очень точного изготовления наконечников.

2. Волокно вкладывается в калиброванную V-образную канавку. Концы соединяемых волокон зажимаются съемными держателями, что позволяет разбирать соединение.

3. Волокно вкладывается в зазор между тремя калиброванными цилиндрами. Их размер выбран так, что диаметр волокна приходится точно на центр вписанной окружности. Все скреплено вместе и образует наконечник волокна. Соединение волокон осуществляется сжатием торцов двух наконечников, например внутри калиброванной трубки.

Три перечисленных основных принципа по-разному используются в *разъемных соединениях*. Разумеется, ими все не исчерпывается. Техника оптических волокон быстро развивается, и очевидно, что будут найдены и другие методы; в частности, отметим метод формовки волокон, который, по всей видимости, также дает хорошие результаты [190, 191]. Разборные разъемы, как правило, широко не применяются. Они довольно дороги (как при покупке, так и при установке), их установка в полевых условиях еще трудна (хотя некоторые конструкторы предлагают довольно простые в обращении монтажные инструмен-

ты), довольно значительны потери: в среднем 0,5 дБ при наличии иммерсионной жидкости и 1 дБ без нее. Вот почему оптические разъемы применяются лишь в определенных случаях:

1) для соединения кабеля с оптоэлектронной аппаратурой (стоечный разъем);

2) для разборных линий связи (распределитель, временные соединения);

3) в особых условиях проводки (сквозь перегородки в самолете, на корабле, в здании т. д.).

В. Сращивание волокон

Хотя изложенные выше методы соединения могут быть использованы для постоянного соединения, предпочтение отдается сварке. Это дает некоторые преимущества: во-первых, уменьшение потерь до 0,3 дБ (что связано главным образом с отсутствием воздушной прослойки между волокнами), а во-вторых, надежность соединения при длительной эксплуатации. Именно поэтому сварка используется при соединении отрезков кабеля в дальней оптической связи. Отметим, что малый диаметр оптического кабеля и его легкий вес позволяют протягивать его от одной опоры до другой на большие (1—2 км), чем для обычного кабеля (300—900 м), расстояния и уменьшать таким образом число сочленений.

Для сварки два аккуратно обрезанных конца волокон выравнивают и совмещают встык, например в калиброванной V-образной канавке. Плавление и сварка происходят за доли секунды при температуре 2000°C. Нагреватели могут быть трех типов: с дуговым разрядом между двумя электродами [188], возможно, в виде плазменной горелки [192], кислородно-водородная микрогорелка [192] или CO₂-лазер [193]. Если сварное соединение залить полиамидной массой, то образуется покрытие, предохраняющее его от механических повреждений, и прочность волокна на разрыв становится почти такой же, как и до сварки.

ГЛАВА 12

ЛАЗЕРЫ

В предыдущих главах особое внимание уделялось вопросу распространения света как в оптических системах обработки информации, так и в оптических волокнах при передаче сигналов. В данной же главе мы вернемся к вопросу о генерации света, ограничив себя, однако, только источниками света, пригодными для оптических систем связи. Мы не намерены подробно излагать теорию лазеров и электролюминесцентных светодиодов. Этому вопросу посвящено много книг (см., например, [194—196]). Наша цель в том, чтобы напомнить читателю основные принципы излучения когерентного и некогерентного света и рассмотреть характеристики расходимости, монохроматичности, эффективности и модуляции различных источников света, имеющих в нашем распоряжении, чтобы в последующих главах иметь возможность обосновать общий подход к оптическим системам обработки и передачи информации исходя из реальных характеристик. Поэтому здесь еще больше, чем в других главах, будет уделяться внимание упрощению, чтобы лучше выделить понятия, которые важны для последующих глав.

Открытие и разработка лазера — это, по-видимому, одно из самых важных событий в физике за последнюю четверть века. Если в технике и промышленности последствия этого открытия еще достаточно скромны, то в физике почти все ее разделы широко используют лазер, и он буквально перевернул методы физических исследований.

В технике связи лазер может найти применение двоякого рода:

а) в плане более близкой перспективы он оказывается самым подходящим источником света для волоконно-оптических систем связи с очень высокой информационной пропускной способностью;

б) в плане более далекой перспективы в рамках оптических систем обработки информации он может служить когерентным источником света для систем пространственной фильтрации и обеспечить мощность и когерентность, необходимые для использования очень интересных нелинейных эффектов.

Кроме того, излучение лазера уже широко используется в качестве переносчика информации в системах ближней связи в пределах прямой видимости (спутники, корабли, маяки и т. д.) в тех случаях, когда особенно желательна скрытность связи.

Мы рассмотрим прежде всего принцип действия лазера, затем опишем различные типы существующих лазеров и, наконец, покажем, как можно получить те или иные характеристики лазера за счет изменения режима его работы. Полупроводниковые лазеры, самые важные для техники связи, будут рассмотрены отдельно в § 4 после описания твердотельных, жидкостных и газовых лазеров.

§ 1. Принцип работы лазера

Лазер — это соединение двух основных элементов: 1) *усилителя световой волны* и 2) *цепи обратной связи в форме резонатора*.

Их сочетание и образует *генератор*, подобный многим другим известным в физике генераторам (механическим, электрическим). Тем не менее его использование в оптике столкнулось с техническими трудностями, обусловленными, с одной стороны, усилителем, а с другой — резонатором.

Усиление световой волны в усилителе основано на явлении *индуцированного излучения* фотона возбужденной частицей. Индуцированное излучение известно давно, еще начиная с работ Эйнштейна, но в оптическом диапазоне оно маскируется *спонтанным излучением*, т. е. обычным некогерентным излучением. Чтобы основную роль играло индуцированное излучение, нужно сильно изменить состояние усиливающей (активной) среды, а именно создать инверсию заселенности ее энергетических уровней.

Лазерный резонатор представляет собой открытый резонатор Фабри — Перо, т. е. систему двух плоских зеркал, расположенных одно против другого совершенно параллельно друг другу. Поскольку он должен содержать усиливающую среду, предъявляются исключительно жесткие требования к его настройке. Кроме того, требования высокой добротности резонатора (порядка 10^8 для газовых лазеров) часто приводят к необходимости в зеркалах исключительно высокого оптического качества.

Принцип работы лазера — хороший пример взаимодействия между веществом (усиливающей средой) и излучением (световой волной, которая возникает в резонаторе и затем многократно им отражается).

Строгий анализ проблем взаимодействия излучения с веществом требует привлечения аппарата квантовой механики или же по крайней мере полуквантового подхода (в этом случае квантуется только среда, а поля рассматриваются классическим образом). Тем не менее классическая механика, используемая с определенной осторожностью, позволяет весьма удовлетворительно объяснить большинство явлений.

Классический подход обладает тем достоинством, что обеспечивает очень наглядное понимание физики работы лазера. Именно по этой причине он выбран нами для дальнейшего рассмотрения. Сначала мы рассмотрим резонатор, а затем усилитель (§ 1, п. Б).

А. Резонатор Фабри — Перо

Резонатор Фабри — Перо представляет собой многолучевой интерферометр, образованный двумя плоскими зеркалами, расположенными параллельно друг другу [197].

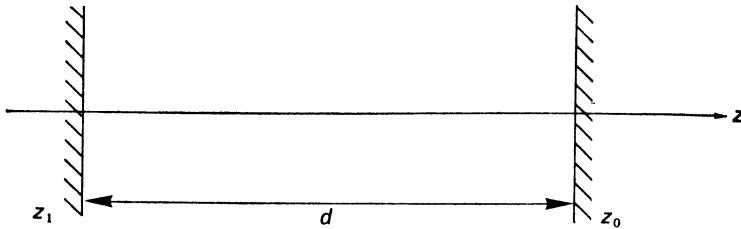


Рис. 12.1. Схема лазерного резонатора.

Мы будем рассматривать одномерный случай распространения излучения в среде, ограниченной двумя параллельными идеально проводящими плоскостями (рис. 12.1)

В этом случае решения уравнений Максвелла имеют вид

$$A \exp i\omega \left(t - \frac{z}{c} \right) + B \exp i\omega \left(t + \frac{z}{c} \right)$$

и представляют собой суперпозицию волн, распространяющихся вдоль оси Oz навстречу друг другу. Граничные условия в точках z_0 и z_1 :

$$A \exp i\omega \left(t - \frac{z_0}{c} \right) + B \exp i\omega \left(t + \frac{z_0}{c} \right) = 0,$$

$$A \exp i\omega \left(t - \frac{z_1}{c} \right) + B \exp i\omega \left(t + \frac{z_1}{c} \right) = 0,$$

должны выполняться для всех значений времени t . Таким образом, возможны только такие решения, которые удовлетворяют условию

$$\exp 2 \frac{i\omega}{c} (z_1 - z_0) = 1.$$

Введем обозначения $z_1 - z_0 = d$, $2\omega d/c = 2p\pi$, где p — целое число. Следовательно, в рассматриваемом резонаторе могут существовать только волны с частотой $\nu_p = pc/2d$. Послед-

нее соотношение можно записать также в виде

$$\lambda_p = \frac{2d}{p}. \quad (12.1)$$

Отсюда видно, каким образом зеркала резонатора осуществляют *квантование* частот, способных в нем распространяться.

Итак, амплитуда волны в резонаторе записывается в виде

$$2A \exp i\omega \left(t - \frac{z_0}{c} \right) \sin \frac{\omega}{c} (z - z_0);$$

такая форма характерна для стоячей волны. Все происходит так, как если бы внутри резонатора не было распространения волны.

Приведенными выше соотношениями определяются частоты, которые могут существовать в резонаторе. На самом деле они будут существовать только в том случае, если их возбудить. Имеются две возможности возбудить резонатор: первая — поместить источник излучения внутрь резонатора, вторая — осветить резонатор снаружи.

В первом случае волны будут складываться в резонаторе после каждого прохода, если частота источника равна ν_p . При этом интенсивность света увеличивается, пока источник излучает. Энергия не излучается наружу, если не сделать одну из плоскостей, образующих резонатор, слегка прозрачной. Тогда возникают *потери за счет вывода энергии* из резонатора. Эти потери немного уменьшают энергию в резонаторе, но появляется внешнее излучение. Лазер действует именно по такому принципу. Если же частота отличается от ν_p , то из-за повторяющейся интерференции волна ослабляется по мере ее возбуждения и энергия не накапливается в резонаторе.

Во втором случае в резонаторе постепенно накапливается энергия внешнего источника при $\nu = \nu_p$. Часть энергии выводится за счет связи во внешнюю среду, и, когда соответствующие потери становятся равными энергии, поступающей от источника, достигается равновесие. Если же частота ν отличается от ν_p , то, как и в предыдущем случае, энергия в резонаторе не накапливается.

Предположим, что в резонатор ввели энергию от внешнего источника, а затем убрали источник. Время τ , за которое при этом энергия колебаний в резонаторе уменьшится в e раз (e — основание натуральных логарифмов), называется *временем жизни* волны в резонаторе.

Добротность резонатора на частоте ν_p определяется как

$$Q = 2\pi\nu_p\tau.$$

Оценим величину τ для резонатора длиной $d = 1$ м с коэффициентом отражения зеркал $R = 0,99$ при освещении его зеленым светом ($\lambda_p = 0,5$ мкм, $\nu_p = 6 \cdot 10^{14}$ Гц).

После n проходов резонатора амплитуда волны станет равной

$$E(n) = E_0 R^{2n},$$

а прошедшее время будет равно $t = n\theta$, где $\theta = 2d/c$ — продолжительность двух проходов резонатора. Следовательно, изменение энергии в резонаторе описывается выражением

$$W(t) = W_0 R^{2 \frac{tc}{d}},$$

откуда получаем формулу для времени жизни волны в резонаторе:

$$\tau = -\frac{d}{2c} \frac{1}{\ln R} \approx 0,09 \text{ мкс.}$$

Тогда добротность резонатора составит $Q = 0,3 \cdot 10^9$. Это очень большая величина, она более чем в 1000 раз больше добротности СВЧ-резонаторов.

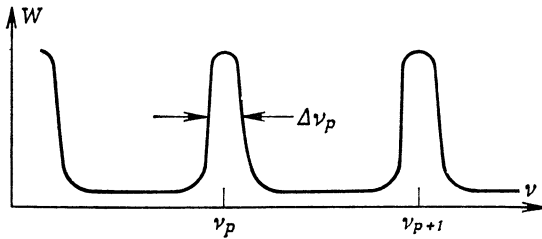


Рис. 12.2. Селекция частот в резонаторе Фабри — Перо.

Воспользовавшись соотношением (2.42), связывающим ширину полосы с временем жизни фотона, можно показать, что добротность резонатора является также хорошей мерой относительной ширины линии (рис. 12.2):

$$Q = \frac{\nu_p}{\Delta\nu_p}. \quad (12.2)$$

Таким образом, лазерные резонаторы характеризуются очень высокой избирательностью по длинам волн.

Расстояние между двумя соседними модами колебаний ν_p и ν_{p+1} также выражается очень просто [формула (12.1)]

$$\nu_{p+1} - \nu_p = \frac{c}{2d} = \frac{1}{\theta}.$$

Все сказанное применительно к одномерному случаю можно обобщить на резонатор любой формы, и при этом выводы те же: электромагнитные волны в объемном резонаторе суще-

ствуют в виде его мод, обозначаемых тремя целочисленными индексами (h, k, l), причем *длины волн этих мод и направления их распространения* оказываются *квантованными*. Картина очень похожа на распространение волн в оптическом волокне или в волноводе, хотя граничные условия иные. Более детальный физический анализ выявляет существование достаточно сложных мод (эрмитово-гауссовых волн).

В качестве полного руководства по резонаторам Фабри — Перо можно рекомендовать книгу [198].

При решении многих задач оказывается возможным уменьшить число индексов мод с трех до двух на основании симметрии системы (обычно ось Oz резонатора является осью симметрии).

Б. Усиливающая среда

Мы показали, что волна может оказаться в резонансе с модой резонатора Фабри — Перо. Но остается важный вопрос: как поддерживать эту волну в резонаторе, а может быть даже и усиливать ее? В самом деле, резонатор, о котором говорилось

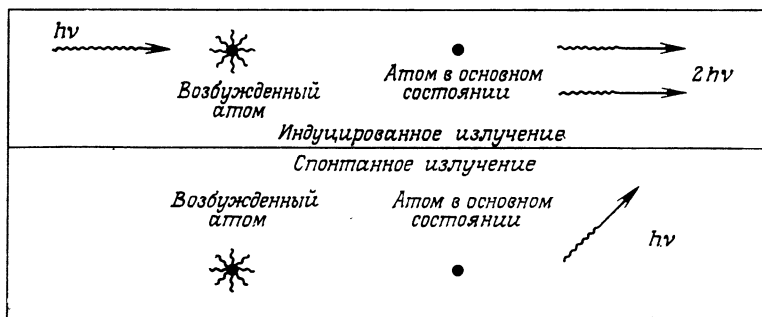


Рис. 12.3. Индуцированное и спонтанное излучение.

В процессе индуцированного излучения участвуют два фотона, процесс спонтанного излучения происходит независимо от других фотонов.

выше, рано или поздно отдает всю накопленную энергию и никогда не даст ее больше, чем получил.

Сообщать энергию волне для компенсации потерь должна усиливающая среда. Когда благодаря усилению в среде потери в резонаторе полностью компенсируются, любая падающая в резонатор волна с частотой ν_p будет сохранять свою интенсивность постоянной и отдавать потребителю за счет связи энергию, равную $(1 - R)^2 E_0^2$. Если же усиление превышает потери, то интенсивность волны в резонаторе так же, как и для потребителя, увеличивается, пока при очень высокой плотности энергии не произойдут изменения в среде, которые приведут к уменьшению коэффициента усиления.

Усиление в среде достигается за счет *индуцированного (вынужденного)* излучения (рис. 12.3). Оно соответствует одному из двух видов излучательных (с испусканием фотона) переходов, которые возможны для возвращения возбужденной частицы (атома или молекулы) на ее исходной (основной) уровень (существуют и другие виды передачи энергии, но без испускания света: столкновения между молекулами в газе, обмен энергией между фотоном и кристаллической решеткой в твердых телах). Эти два вида излучения, *индуцированное и спонтанное*, дают фотоны одинаковой частоты ν . В случае спонтанного излучения заранее не известны ни направление распространения этого фотона, ни момент его испускания (т. е. фаза волны).

Индукцированное излучение происходит только тогда, когда на возбужденный атом падает фотон с частотой ν . Тогда фаза и направление испускаемого фотона те же, что и у падающего. Таким образом, испускается второй фотон, неотличимый от первого.

1. Воздействие волны на возбужденный атом

Обратимся к уравнениям (предложенным Эйнштейном), которые описывают процесс взаимодействия вещества с падающим на него излучением [199, 200].

Рассмотрим некую совокупность атомов, обладающих двумя возможными энергетическими уровнями E_1 и E_2 , и предположим, что это невырожденные уровни. Пусть N_1 и N_2 — заселенности этих уровней. Пусть на эту совокупность атомов падает электромагнитная волна с частотой

$$\nu = \frac{E_2 - E_1}{h},$$

где h — постоянная Планка.

Вследствие поглощения dN_2 атомов переходят с уровня 1 на уровень 2 за время dt :

$$dN_2 = B\omega_\nu N_1 dt,$$

где ω_ν — плотность излучения в полосе частот ν , $\nu \pm d\nu$, а B — постоянная.

Вследствие спонтанного излучения переходит число dN'_2 :

$$dN'_2 = -AN_2 dt,$$

где A — постоянная.

Вследствие индуцированного излучения переходит число dN''_2 , пропорциональное плотности излучения:

$$dN''_2 = -B\omega_\nu N_2 dt.$$

Чтобы в стационарном режиме отношение N_2/N_1 приняло значение, определяемое законом Больцмана [199], нужно, кроме

того, чтобы A и B удовлетворяли соотношению

$$\frac{A}{B} = 8 \frac{h\nu^3}{c^3}.$$

В условиях термодинамического равновесия при наличии поля ω_ν выполняется равенство

$$B\omega_\nu N_1 - AN_2 - B\omega_\nu N_2 = 0,$$

откуда

$$\omega_\nu = \frac{A/B}{e^{h\nu/kT} - 1}.$$

Следовательно, отношение вероятности индуцированного излучения к вероятности спонтанного излучения будет равно

$$\frac{B\omega_\nu}{A} = \frac{1}{e^{h\nu/kT} - 1} = n,$$

где n — число фотонов с энергией $h\nu$.

Таким образом, при термодинамическом равновесии вероятность индуцированного излучения сильно зависит от отношения $h\nu/kT$. Оценим, например, вероятность индуцированного излучения при 300 К для двух диапазонов:

Видимый свет: $\lambda = 0,5$ мкм; $\nu = 6 \cdot 10^{14}$ Гц; $B\omega_\nu/A \approx 10^{-43}$.

СВЧ-излучение $\lambda = 3$ см; $\nu = 10$ ГГц; $B\omega_\nu/A \approx 1$.

При комнатной температуре граница между преимущественно спонтанным излучением и преимущественно индуцированным излучением проходит вблизи частоты $\nu = kT/h = 6 \cdot 10^{12}$, т. е. на длине волны порядка 50 мкм.

Заселенности обоих энергетических уровней удовлетворяют соотношению

$$\frac{N_2}{N_1} = \frac{B\omega_\nu}{A + B\omega_\nu} = \frac{n}{n + 1}. \quad (12.3)$$

Следовательно, в отсутствие внешнего излучения заселен только энергетический уровень 1. С увеличением числа фотонов заселенности обоих уровней постепенно выравниваются. В этом случае говорят, что переход с уровня 1 на уровень 2 насыщен.

2. Усиление

Рассмотрим пучок единичного сечения. Его интенсивность связана с плотностью излучения соотношением $SI_\nu = c\omega_\nu$. Вновь проведем баланс интенсивности, пренебрегая спонтанным излучением, и найдем изменение интенсивности при прохождении участка активной среды толщиной dx :

$$dI_\nu = h\nu I_\nu B(N_2 - N_1) dx.$$

Таким образом, коэффициент поглощения среды на частоте ν равен

$$\alpha_\nu = -\frac{1}{I_\nu} \frac{dI_\nu}{dx} = -\frac{h\nu}{c} (N_2 - N_1) B. \quad (12.4)$$

При термодинамическом равновесии $N_1 > N_2$ и, следовательно, $\alpha_\nu > 0$ и среда поглощает энергию. Если $N_2 > N_1$, то говорят, что имеется инверсия заселенности. В этом случае поглощение отрицательно, т. е. происходит усиление. После прохождения сквозь среду энергия волны больше, чем до прохождения.

Выше было показано [формула (12.3)], что если у частицы только два энергетических уровня, то невозможно реализовать условие $N_2 > N_1$, поскольку в пределе при очень сильном поле переход между уровнями насыщается и наступает равновесие $N_2 = N_1$. Каким же образом осуществить требуемую инверсию заселенности уровней? Прежде всего мы выбираем среду по крайней мере с тремя энергетическими уровнями с тем, чтобы, играя на временах рекомбинации между уровнями, мы смогли обеспечить инверсию заселенности для двух из них. А затем мы воздействуем на них мощной «накачкой». Этот этап является основным. Накачку можно осуществлять оптически, подвергая среду действию мощного излучения $\omega > \omega_\nu$, как мы это уже видели выше. Однако оптическая накачка возможна не для всех материалов. Иногда вместо нее прибегают к электронной, механической (за счет удара) или химической накачке.

3. Лазерная генерация

Поместим усиливающую среду с коэффициентом поглощения α_ν в резонатор, рассмотренный в § 1, п. А. Коэффициент усиления волны за один проход резонатора определяется из соотношения $I = GI_0$, где $G = \exp(-\alpha_\nu d)$, а d — активная длина усиливающей среды, которую мы примем равной длине резонатора.

Если частота колебаний ν , определяемая энергетическими уровнями усиливающей среды, отличается от частоты моды ν_p резонатора Фабри — Перо, то, как мы видели в § 1, п. А, такие частоты не будут существовать в резонаторе.

И наоборот, если частота колебаний соответствует моде резонатора, то амплитуда волны A_1 после двойного прохождения через резонатор (туда и обратно) будет равна

$$A_1 = A_0 R^2 \exp\left(-2\alpha_\nu - i\frac{4\pi}{\lambda}\right) d.$$

Здесь R^2 — произведение коэффициентов отражения двух зеркал.

После n двукратных проходов амплитуда волны будет равна

$$A_n = A_0 \left[R^2 \exp \left(-2\alpha_v - i \frac{4\pi}{\lambda} \right) d \right]^n.$$

Полная амплитуда волны в резонаторе будет равна сумме всех таких амплитуд:

$$A = \sum_{n=1}^{\infty} A_n = \frac{A_0}{1 - R^2 \exp \left(-2\alpha_v - i \frac{4\pi}{\lambda} \right) d}.$$

Полученная формула аналогична формуле Найквиста для резонаторов.

Действительно, если выполняется соотношение $R^2 \exp(-2\alpha_v - i4\pi/\lambda)d = 1$ (называемое *условием самовозбуждения*), то амплитуда волны возрастает беспредельно.

Кроме того, даже если величина A_0 практически равна нулю, то и в этом случае будет наблюдаться самовозбуждение лазера, начиная со сколь угодно слабой амплитуды волны с частотой ν_p . Тогда индуцированное излучение может начаться на основе спонтанного.

Из условия самовозбуждения вытекают два следствия:

$$1. \frac{4\pi}{\lambda} d = 2\pi, \text{ т. е. } \lambda = \frac{2d}{p};$$

лазер может самовозбудиться только на одной из мод резонатора [формула (12.1)].

$$2. R^2 \exp(-\alpha_v d) = 1,$$

т. е. минимальный требуемый коэффициент усиления соответствует компенсации потерь, обусловленных связью.

4. Рабочие режимы лазера

В начальный момент времени накачка создает в усиливающей среде инверсную заселенность ($N_2 > N_1$), характеризующую числом

$$\eta = \frac{N_2 - N_1}{N_2 + N_1}.$$

Коэффициент усиления среды α_v , пропорциональный η [формула (12.4)], растет при этом до порога самовозбуждения α_s , характеризуемого величиной η_s . Число η_s — это пороговое значение инверсной заселенности, при котором полностью компенсируются потери всех видов. При такой инверсной заселенности появляется *лазерное излучение*. Число индуцированных переходов увеличивается под влиянием возрастающего поля, что способствует также увеличению этого поля. Наблюдается лавин-

ный эффект, или «цепная реакция». Но эта цепная реакция не может быть неограниченной. С возрастанием числа актов индуцированного излучения уменьшается заселенность верхнего уровня, что приводит к уменьшению инверсной заселенности,

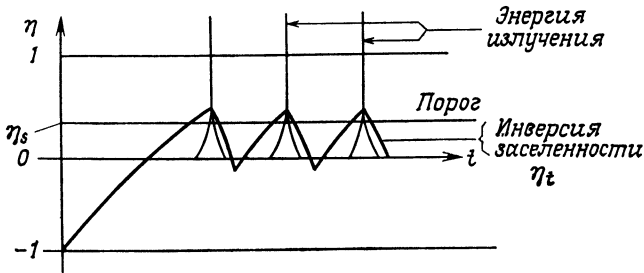


Рис. 12.4. Энергия излучения и инверсия заселенности энергетических уровней лазера с высоким коэффициентом усиления активной среды.

В этом случае лазер работает в режиме релаксационных колебаний (в импульсном режиме свободной генерации).

а значит, и коэффициента усиления α_v . Нарастание прекращается. При этом возможны два варианта.

1. Коэффициент усиления α_v очень большой, и накачка недостаточна для поддержания инверсной заселенности выше по-

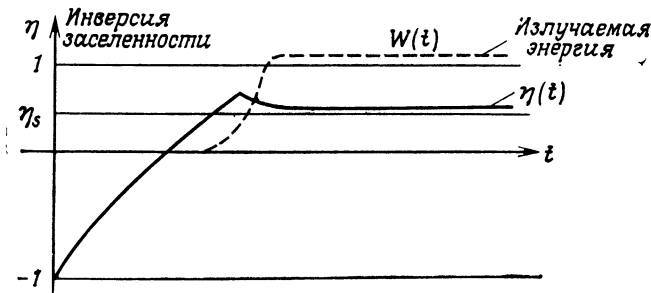


Рис. 12.5. Энергия излучения и инверсия заселенности лазера с низким коэффициентом усиления.

Инверсия заселенности остается выше порога, и лазер работает в режиме непрерывного излучения.

рогового значения η_s (рис. 12.4). В этом случае после излучения очень большой энергии лазер выключается на время, в течение которого накачка снова заставит величину η дойти до порогового значения η_s . Это так называемый режим *релаксационных колебаний* (импульсный режим свободной генерации). Такой режим работы характерен для лазеров большой мощности.

2. Коэффициент усиления α_v не очень велик, накачка позволяет поддерживать инверсную заселенность выше порога η_s . В этом случае лазер достигает некоего равновесного состояния и работает в режиме *непрерывного* излучения (рис. 12.5).

В. Ширина спектральных линий

Ширина спектральной линии излучения лазера зависит главным образом от двух факторов: от резонатора, с одной стороны, и усиливающей среды — с другой. Косвенно она часто определяется и другими, практическими соображениями, которые мы укажем ниже.

1. Естественная ширина лазерной линии

Усиливающая среда характеризуется шириной спектральной линии $\Delta\nu$, связанной с двумя ее энергетическими уровнями E_1 и E_2 . Неопределенностью ΔE измерения разности $E_2 - E_1$ определяется время жизни уровня 2, в соответствии с соотношением неопределенностей Гейзенберга [формула (2.42)]:

$$\Delta E \tau = h, \text{ т. е. } \Delta\nu = 1/\tau.$$

Выгоднее выбирать переходы с большим временем жизни. Правда, ширина $\Delta\nu$, называемая естественной шириной линии, обычно очень мала и ею можно пренебречь по сравнению с другими.

Пример: $\lambda = 0,5 \text{ мкм}$, $\tau \approx 10^{-8} \text{ с}$, $\Delta\lambda \approx 10^{-5} \text{ нм}$.

2. Ширина полосы пропускания резонатора Фабри — Перо

Ширина полосы пропускания резонатора Фабри — Перо зависит главным образом от его добротности Q . Теоретически она определяется соотношением (12.2):

$$\Delta\lambda = \frac{\lambda}{Q}.$$

Если $Q = 10^8$, $\lambda = 500 \text{ нм}$, то $\Delta\lambda = 5 \cdot 10^{-5} \text{ нм}$.

В действительности ширина полосы пропускания резонатора ограничивается допусками, с которыми изготавливается резонатор, и обычно составляет $\sim 10^{-4} \text{ нм}$.

3. Причины уширения лазерной линии

В зависимости от типа лазера имеется много различных причин уширения лазерной линии. Все они приводят к уширению естественной линии, а также к увеличению ширины полосы пропускания объемного резонатора.

1. *Доплеровское уширение.* Существенно в газовых лазерах. Оно состоит в том, что атомная линия частоты ν смещается под действием броуновского движения атомов со скоростью v (рис. 12.6), образующей угол θ с направлением излучения, в результате чего частота линии становится равной (в первом приближении)

$$\nu \left(1 - \frac{v}{c} \cos \theta \right).$$

По порядку величины доплеровская ширина линии обычно составляет 10^{-3} нм.

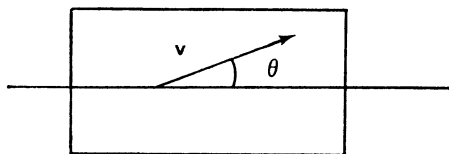


Рис. 12.6. Эффект Доплера.

Вектор скорости возбужденного атома и волновой вектор падающего фотона образуют угол θ

2. *Уширение давлением.* Движение атомов и молекул газа часто приводит к столкновениям между ними. Эти столкновения вызывают переход возбужденных атомов с уровня E_2 на уровень E_1 , снижая тем самым время жизни возбужденного уровня. В результате увеличивается ширина спектральной линии излучения (поскольку $\Delta\nu \tau = 1$).

3. *Штарковское уширение.* Под действием электрического поля уровни E_1 и E_2 могут расщепиться на несколько близких энергетических уровней (поле снимает вырождение энергетических уровней). Эффект Штарка наблюдается главным образом в твердотельных лазерах большой мощности. Так в случае лазера на рубине штарковское уширение линии излучения имеет порядок 0,1 нм.

4. *Уширение за счет дефектов в твердом теле,* создающих неоднородности в распределении потенциала.

Все эти виды уширения приводят к тому, что линия излучения лазера становится более широкой по сравнению с той, которая определяется добротностью резонатора. Таким образом, спектр излучения лазера представляет собой ряд равноотстоящих друг от друга спектральных линий, приходящихся на моды резонатора Фабри—Перо. Огибающей этих линий служит кривая, обусловленная главной причиной уширения для рассматриваемого типа лазера. Можно показать [200], что в действительности ширина каждой из элементарных линий меньше той ширины, которая обусловлена резонатором Фабри—Перо, и эта разница тем больше, чем больше отдаваемая мощ-

ность. Такое обострение спектральных линий обусловлено индуцированным излучением (рис. 12.7).

На практике важно различать два вида паразитного уширения линии излучения:

1) *однородное* уширение (штарковское или обусловленное давлением), при котором все атомы играют одинаковую роль в статистическом смысле;

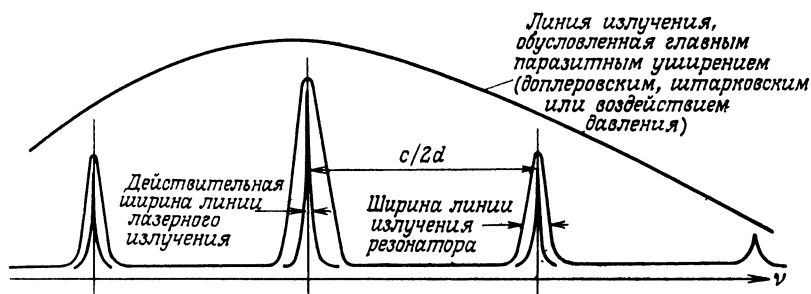


Рис. 12.7. Частотная избирательность различных явлений, используемых при лазерном излучении.

Огибающая различных спектральных линий излучения (основное уширение линии) зависит главным образом от типа усиливающей среды.

2) *неоднородное* уширение (доплеровское), когда волна частоты ν может взаимодействовать только с атомами, частота которых равна $\nu = \nu_0(1 - v \cos \theta/c)$.

Эти два вида уширения совершенно по-разному влияют на работу лазера.

§ 2. Различные типы лазеров

А. Общие сведения

Существование лазерного излучения в оптическом диапазоне волн требует наличия переходов $E_2 \rightarrow E_1$, для которых разность энергий равна 1—3 эВ (что соответствует $10\,000\text{ см}^{-1}$, или $5 \cdot 10^{14}\text{ Гц}$). Кроме того, эти переходы должны быть разрешенными (т. е. они должны удовлетворять законам квантовой механики о переходах) и быть излучательными (т. е. переход должен сопровождаться испусканием фотона).

При этом необходимо большое время жизни верхнего уровня перехода, чтобы можно было накопить на нем достаточное число атомов и обеспечить инверсию заселенности уровней.

Для лазерной генерации пригодны три типа структур.

1. *Атомы и ионы.* В этом случае используются переходы между их электронными оболочками (т. е. используется взаимодействие между ядром и электронной оболочкой атома).

2. *Молекулы.* Используя вращательно-колебательные спектры молекул, можно получить переходы с излучением в диапазоне от миллиметровых волн до ультрафиолетовой области спектра.

3. *Полупроводники.* Здесь используются переходы между зоной проводимости и валентной зоной. Длина излучаемых волн в данном случае приходится главным образом на ближнюю инфракрасную область спектра.

Б. Гелий-неоновый лазер

Этот тип лазера получил самое широкое применение. Газовая смесь состоит в основном из гелия (90%), но активную роль играет только неон. У этих газов близкие энергетические уровни, что и позволяет их использовать в смеси (рис. 12.8).

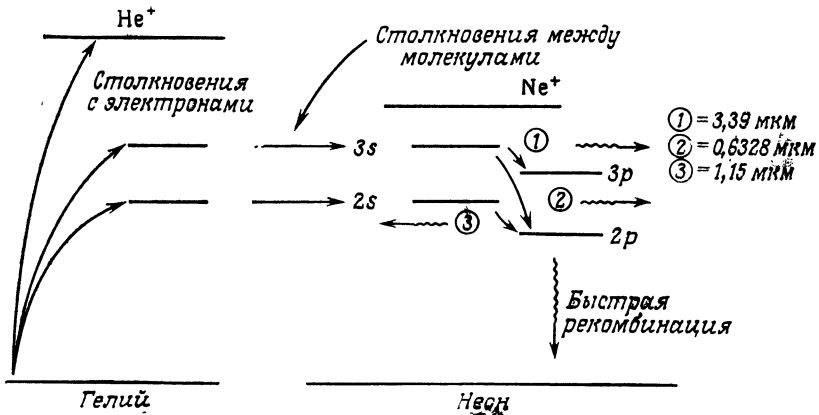


Рис. 12.8. Принцип действия гелий-неонового лазера.

Гелий не принимает прямого участия в излучении. Атомы гелия возбуждают атомы неона при столкновениях с ними. Обычно выделяется только красная спектральная линия излучения (0,633 мкм), но лазер может также генерировать излучение с длиной волны 3,39 и 1,15 мкм.

В электрическом разряде возбуждается гелий. За счет столкновений молекулы гелия передают свою энергию молекулам неона, избирательно заселяя при этом энергетические уровни 3s и 2s. Это быстро создает инверсию заселенности по отношению к уровням 3p и 2p, поскольку время жизни уровней 2p и 3p значительно меньше (четырёхуровневая схема лазера). В результате возникает возможность лазерной генерации на трех возможных длинах волн. Выбирая соответствующие зеркала резонатора, можно выделять ту или иную линию излучения (рис. 12.9). В видимой области спектра располагается единственная линия излучения с длиной волны 0,6328 мкм.

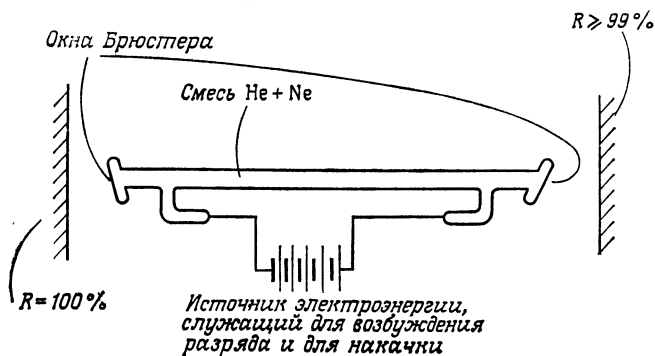


Рис. 12.9. Схема гелий-неонового лазера.

В настоящее время лазер такого типа наиболее распространен.

Гелий-неоновые лазеры характеризуются очень небольшим коэффициентом усиления активной среды, что дает им возможность работать в режиме непрерывного излучения. Их мощность излучения изменяется от 0,5 до 200 мВт.

В. Другие типы газовых лазеров

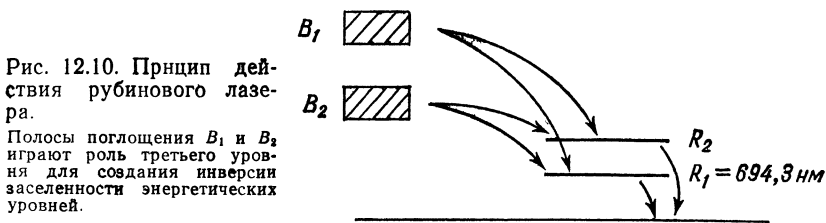
1. В ионном аргоновом лазере (Ar^+) имеются уровни $4s$ и $4p$, которые многократно вырождены. Благодаря этому возможны 9 возможных переходов в интервале от зеленого до фиолетового цветов. Спектральные линии излучения можно разделять в резонаторе с помощью дифракционной решетки или же использовать все линии одновременно. Самыми интенсивными оказываются линии с длинами волн 0,514 и 0,488 мкм. Аргоновый лазер на этих линиях излучения имеет мощность порядка 1 Вт. Суммарная мощность всех линий излучения может превышать 15 Вт. Ионные лазеры на криптоне и на ксеноне работают аналогичным образом.

2. В лазере на двуокиси углерода используется газовая смесь CO_2 и N_2 . Система накачки в этом лазере близка к используемой в гелий-неоновом лазере. Здесь азот играет роль гелия. Такой лазер характеризуется очень большим коэффициентом усиления активной среды и часто работает в импульсном режиме. Мощность такого лазера иногда достигает 1000 Вт. Спектральная линия излучения лежит в ИК-области при 10,6 мкм.

В отмеченных двух типах лазеров используется *четырёхуровневая схема* генерации: нижний уровень перехода не является основным. В такой схеме легче добиться инверсной заселенности.

Г. Твердотельные лазеры

1. В рубиновом лазере используются ионы хрома Cr^{3+} , распределенные в матрице Al_2O_3 . Полезные переходы создают только ионы хрома. Это типичный лазер с тремя уровнями. Ионы хрома образуют две широкие полосы поглощения B_1 и B_2 и две спектральные линии излучения, обозначаемые символами R_1 и R_2 . В результате накачки заселенность основного уровня уменьшается за счет переходов в полосы B_1 и B_2 . Последние характеризуются очень малым временем жизни, и их заселенность уменьшается из-за переходов на уровни R_1 и R_2 . В результате возникает инверсия заселенности и возникает лазерная генерация на линии R_1 . Накачка осуществляется с помощью



импульсных ламп-вспышек (оптическая накачка). Коэффициент усиления весьма высокий, и лазер может работать либо в импульсном, либо в непрерывном режиме. Линия излучения очень узкая (рис. 12.10).

2. Лазер на ионах неодима (Nd^{3+}): легированное стекло и иттриево-алюминиевый гранат (ИАГ). В этом лазере в качестве активного элемента используется редкоземельный элемент неодим. Особенностью лазера на неодиме является использование незаполненного уровня $4f$, который не является самым внешним в атоме. Это позволяет обеспечить с помощью внешней оболочки своеобразный экран против воздействия других атомов. Благодаря этому уширение линии излучения отсутствует даже при очень высоких концентрациях неодима.

Как и в рубиновом лазере, накачка осуществляется белым светом с помощью импульсной лампы благодаря наличию различных полос поглощения неодима в видимой области.

Это четырехуровневый лазер. Он характеризуется очень высоким к. п. д. Основная линия излучения — 1,06 мкм. Вначале в качестве матрицы использовалось силикатное стекло. В настоящее время переход Nd^{3+} чаще используют в матрице из иттриевого граната (окись алюминия $\text{Y}_3\text{Al}_5\text{O}_{12}$); это так называемый ИАГ-лазер.

Д. Лазеры на красителях

В таких лазерах активными центрами служат органические молекулы в растворе. Используются излучательные переходы между колебательными уровнями молекул. Обычно выбирают молекулы, состоящие из двух достаточно независимых подсистем A и B . Тем не менее A и B подвергаются относительным колебаниям, которые искажают уровни этих подсистем: в результате снимается их вырождение. Например, два уровня 1 и

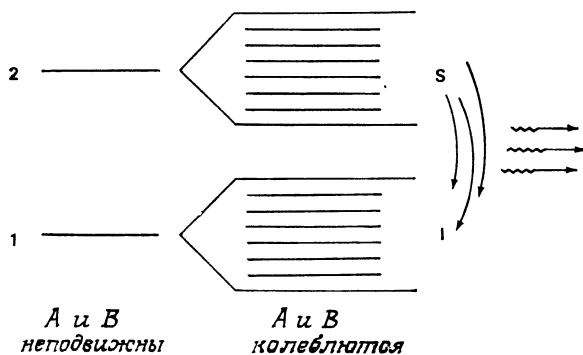


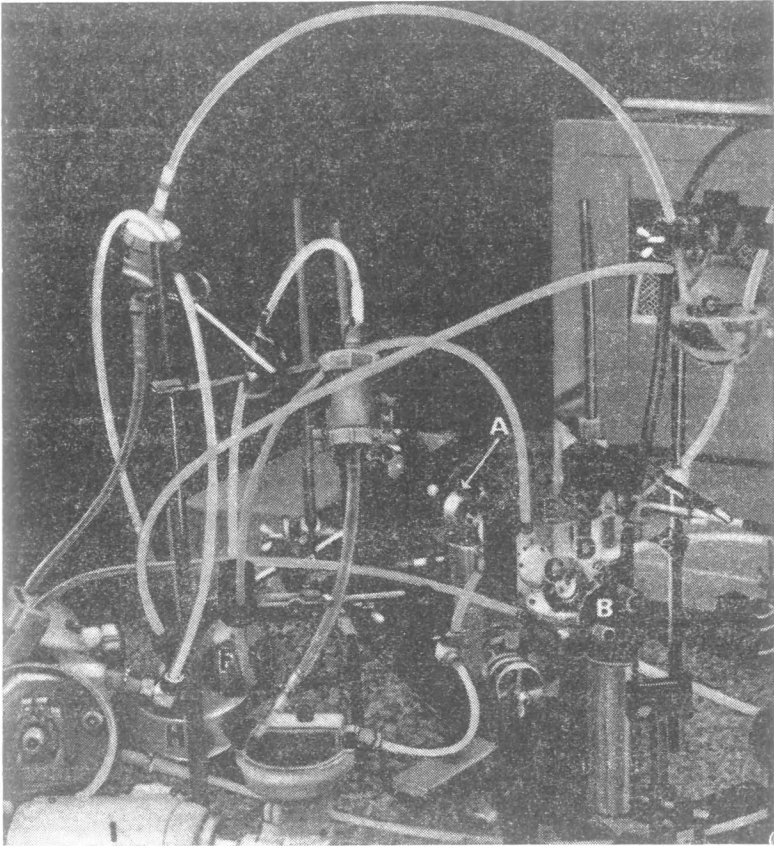
Рис. 12.11. Принцип действия лазера на красителе.

Между верхними и нижними энергетическими уровнями возможны переходы, соответствующие практически непрерывному спектру частот. Поместив на конце резонатора дифракционную решетку, можно выделить одну из этих частот.

2 преобразуются в две большие серии уровней S и I . Внутренняя рекомбинация в S и I , если она более быстрая, чем между S и I , может создать инверсию заселенности между нижними уровнями S и верхними уровнями I , хотя в целом серия I более заселена, чем S . Такая возможность используется для создания лазеров с *перестраиваемой длиной волны* в важной части видимого спектра (рис. 12.11 и фото на стр. 309).

Е. Полупроводниковые лазеры

Такие лазеры, будучи по своей физической природе сложными полупроводниковыми структурами, существенно отличаются от тех, которые мы рассмотрели раньше, и заслуживают того, чтобы описать их отдельно. Однако во многих моментах мы вновь найдем общие принципы, рассмотренные ранее, и на самом деле нет никакой принципиальной разницы. Существенное отличие состоит в том, что дискретные энергетические уровни атомов и молекул здесь преобразованы в широкие и непрерывные зоны, *валентную зону* и *зону проводимости*, а электроны подчиняются статистике Ферми—Дирака, а не стати-



Лазер на красителе.

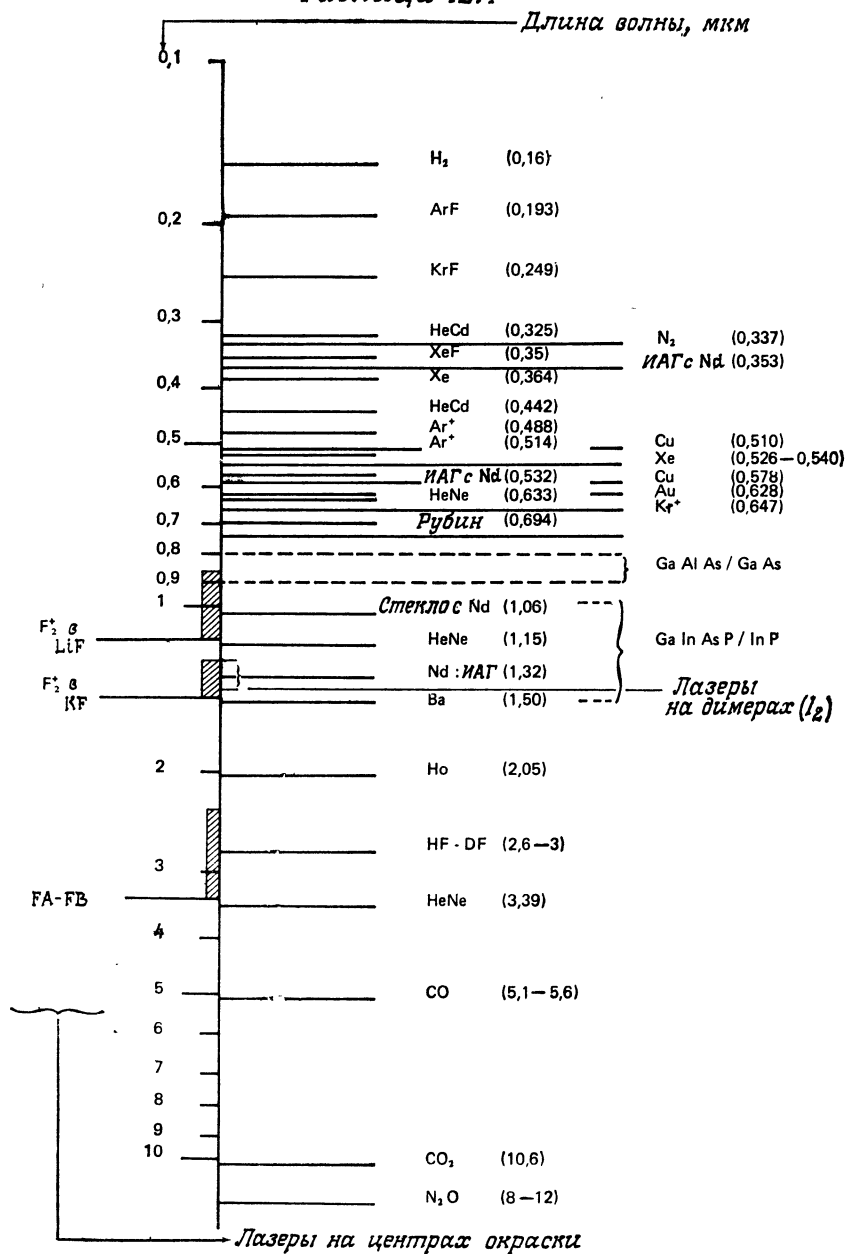
A — заднее зеркало; *B* — переднее зеркало; *C* — импульсная лампа; *D* — эталон Фабри — Перо для уменьшения ширины линии; *E* — усилительный резонатор; *F* — первый краситель (родамин 6G); *G* — второй краситель (кумарин 6); *H* — охлаждение импульсной лампы; *I* — насос, обеспечивающий циркуляцию красителя.

стике Больцмана, как в случае крупных частиц. Поскольку полупроводниковые лазеры очень подходят для волоконно-оптических систем связи, им будет целиком посвящен § 4.

Ж. Перспективы

Все типы лазеров, разработанных к настоящему времени, невозможно рассмотреть. Создается впечатление, что можно получить лазерную генерацию почти на всех веществах. В настоящее время работы ведутся в разных направлениях: создание лазеров на парах металлов, которые позволили бы расширить

Таблица 12.1



спектр лазерного излучения в сторону более коротких длин волн, создание *химических лазеров*, в которых используется испускание фотонов во время химической реакции и которые могли бы обеспечивать большие мощности в режиме непрерывного излучения, и, наконец, создание *лазеров на центрах окраски*, которые позволяют охватить очень большой спектральный диапазон. В табл. 12.1 приведены основные линии излучения лазеров, существующих в настоящее время.

§ 3. Некоторые общие вопросы лазерной техники

Из сказанного в § 1 явствует, что существуют многочисленные возможности улучшения характеристик лазера в том или ином отношении. Ниже мы покажем, каким образом можно видоизменить работу лазера, чтобы лучше приспособить его для решения некоторых частных задач. Представленные здесь некоторые примеры ни в коем случае нельзя считать исчерпывающими.

А. Поляризация

При общем изложении физики лазеров ничего не было сказано о поляризации излучаемого ими света, и нигде в уравнениях, используемых для описания объемного резонатора или усиливающей среды, поляризация не фигурировала. Действительно, лазер, вообще говоря, может излучать и неполяризованный свет, т. е. такой, который может быть разложен на две ортогональные поляризации (§ 4).

Эти две поляризации «конкурируют» между собой, т. е. свет с разной поляризацией генерируется за счет одной и той же инверсии заселенности. Часто это вредит стабильности работы лазера, поскольку если поле ω , соответствующее одной из поляризаций, оказывается более интенсивным, то она в большей мере использует инверсию заселенности и разница в интенсивностях еще более усиливается; однако ничто не мешает преобладанию другой поляризации спустя некоторый промежуток времени. В результате все время происходит переход от одной поляризации к другой. Во многих случаях применения лазеров (например, в интерферометрии или голографии) такие скачки поляризации излучения нежелательны.

Но очень просто обеспечить строго определенную поляризацию излучения на выходе лазера: достаточно ввести потери для одной из двух поляризаций. Это обычно осуществляется с помощью окошка (рис. 12.12), устанавливаемого в резонаторе под углом Брюстера (гл. 3, § 1, п. Б).

В твердотельных лазерах часто существует анизотропия в самом усилителе (например, анизотропия коэффициента усиления,

обусловленная естественным двулучепреломлением кристалла или эффектом Фарадея). В результате конкуренции мод в данном случае остается только одна поляризация. Такой эффект имеет место в лазере на рубине, который обладает двулучепре-

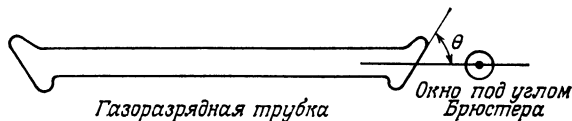


Рис. 12.12. Окно, установленное под углом Брюстера, для выделения определенной поляризации в газовом лазере.

ломлением и в котором оптическая ось никогда не совпадает с механической осью роста кристалла.

Б. Лазеры с модуляцией добротности

Такие лазеры уже очень давно применяются для получения очень больших энергий в форме очень коротких импульсов. Принцип их работы состоит в следующем: после самовозбуждения лазера с высоким усилением активной среды (когда инверсия заселенности превысит порог самовозбуждения η_s) заселенность более высоких энергетических уровней уменьшается из-за излучения вовне и коэффициент усиления падает.

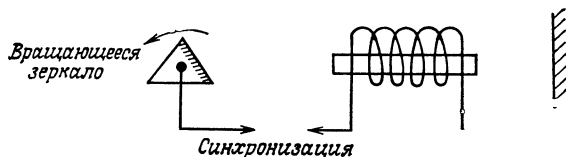


Рис. 12.13. Резонатор с вращающимся зеркалом.

Вращающееся зеркало, установленное вместо одного из двух неподвижных зеркал, позволяет выводить индуцированное излучение через определенные интервалы времени. В периоды, когда резонатор «открыт», инверсия заселенности непрерывно нарастает.

В лазерах с модуляцией добротности резонатор Фабри — Перо открыт в течение всего времени накачки. Его закрывают на очень короткое время только в тот момент, когда инверсная населенность приближается к 1 (это означает, что все атомы находятся в возбужденном состоянии) (см. рис. 12.4). В результате лазер излучает гигантский импульс очень малой длительности.

Пример: в непрерывном режиме излучается мощность 1 кВт за время 1 мкс, а в режиме с модуляцией добротности — 1 МВт за 10 нс.

В режиме генерации гигантских импульсов лазер не может обеспечить высокой частоты повторения импульсов, так как

после каждого импульса нужно время, чтобы рассеялось очень большое количество тепла.

Для запираания резонатора на соответствующее время может служить вращающееся зеркало (рис. 12.13). В этом случае накачка синхронизируется с вращением зеркала. Другие способы модуляции добротности основаны на использовании быстродействующих светозатворов: ячейки Керра или Поккельса, насыщающиеся фильтры, акустооптические ячейки (гл. 14, § 1).

В. Лазеры с синхронизованными модами

Как мы видели выше, только в одной линии излучения, например доплеровской, может существовать до десяти мод резонатора. Эти моды совершенно некогерентны, т. е. у них нет определенных фазовых соотношений. Но можно сфазировать эти моды, если ввести в резонатор высокочастотную модуляцию. Такую модуляцию можно осуществить с помощью акустооптического модулятора или насыщающегося поглотителя.

Г. Одномодовые лазеры, улучшение временной когерентности излучения

В резонаторе Фабри — Перо моды отстоят друг от друга по частоте на величину $\Delta\nu = c/2d$. Введем в него второй резонатор длиной $d' \ll d$ и настроим оба резонатора так, чтобы частота ν_0 одновременно удовлетворяла условиям резонанса в том и в другом. Первая мода, принадлежащая обоим системам, должна будет удовлетворять соотношению

$$\nu - \nu_0 = \frac{pc}{2d} = \frac{\sigma c}{2d'}.$$

При соответствующем выборе величины d' можно сделать так, что эта мода не попадет в полосу частот, усиливаемых за счет накачки.

Д. Улучшение пространственной когерентности излучения

Пространственная когерентность излучения газовых и твердотельных лазеров, как правило, весьма хорошая. Сечение пучка обычно — порядка 1 мм², а его угловая расходимость не превышает 1 мрад. Но можно еще больше улучшить пространственную когерентность, если подавить наименее «настроенные» моды резонатора, создав для них дополнительные потери, например, за счет небольшой диафрагмы, соответствующим образом расположенной в резонаторе.

В следующем параграфе мы, однако, увидим, что даже и без подобных ухищрений пространственная когерентность излучения газовых и твердотельных лазеров значительно выше, чем у полупроводниковых.

§ 4. Полупроводниковые лазеры

Излучение полупроводниковых лазеров обладает пространственной и временной когерентностью, и в этом смысле эти лазеры подобны описанным выше. Но у них имеются и некоторые специфические особенности:

1) излучательные переходы происходят между энергетическими зонами;

2) размеры лазера невелики, толщина активной зоны мала и вследствие этого сравнительно велика расходимость пучка излучения (гл. 4, § 1, п. Д);

3) накачка уровней производится путем пропускания тока через сам диод, этот ток и осуществляет требуемую инверсию заселенности уровней. При этом можно реализовать модуляцию излучения вплоть до весьма высоких частот (нескольких гигагерц), поскольку собственная частота излучателя очень высока.

В связи с тем что работа рассматриваемого типа лазера имеет специфические особенности, мы сначала напомним физику полупроводников. Затем мы рассмотрим основные характеристики полупроводниковых лазеров.

А. Основы физики полупроводников

1. Энергетические зоны

Будем сближать изолированные атомы, электроны которых обладают дискретными энергетическими уровнями. Волны, соответствующие электронам, взаимодействуют друг с другом, и каждый энергетический уровень расщепится на целый ряд различающихся, но очень близких друг к другу уровней, который и называется энергетической зоной. На межатомных расстояниях, соответствующих кристаллической решетке полупроводника, будут следующие энергетические зоны для электронов в направлении возрастания энергии: *валентная зона*, *запрещенная зона* (в которой нет ни одного разрешенного уровня) и *зона проводимости*.

В отсутствие энергетического возбуждения валентная зона будет полностью заполнена (все уровни заняты), а зона проводимости — пустая. Если, однако, на электрон воздействует энергия, достаточная для преодоления им запрещенной зоны, то он займет одно из состояний в зоне проводимости и оставит в валентной зоне свободное место (называемое дыркой).

Таблица 12.2

		GaAs	Si	Ge	In Sb
Ширина запрещенной зоны	ΔW , эВ	1,43	1,12	0,72	0,18
Критическая длина волны	λ_c , мкм	0,87	1,1	1,7	6,3

В табл. 12.2 приведены значения ширины запрещенной зоны ΔW и критической длины волны, при которой фотон передает электрону энергию, достаточную для преодоления этой зоны, для ряда интересных для нас полупроводниковых материалов [201].

Преодолев межатомную связь, валентный электрон становится свободным зарядом проводимости, способным участвовать в переносе электрического тока. Под влиянием теплового воздействия этот электрон способен перемещаться по закону броуновского движения, характеризуемого средней скоростью $v_{\text{тепл}}$, временем релаксации τ (временем между двумя столкновениями частиц) и средней длиной свободного пробега l . Для германия соответствующие значения этих величин таковы: $v_{\text{тепл}} = 10^5$ м/с; $\tau = 1$ пс; $l = 0,1$ мкм.

Аналогичным образом может перемещаться и дырка (недостающий валентный электрон связи будет получен от соседнего атома, который при этом сам ионизируется).

2. Легирование полупроводника

Легирование — это введение в кристаллическую решетку полупроводника инородных атомов. Можно показать, что созданные таким образом новые энергетические уровни располагаются в его запрещенной зоне. Например, в кремнии и германии, принадлежащих к IV группе элементов, атомы донорной примеси будут иметь 5 валентных электронов, а атомы акцепторной примеси — только 3. Энергия, необходимая для получения (высвобождения) одного носителя заряда $-q$ (заряда электрона), при наличии атома донора приблизительно в 50 раз меньше, чем у собственного атома решетки без примеси [202]. В соответствии с этим при комнатной температуре все атомы, осуществляющие легирование материала, оказываются ионизованными и, следовательно, соответствующие электрические заряды локализованы в кристаллической решетке полупроводника. В веществе, представляющем собой химическое соединение, понятия донорных и

акцепторных примесей сохраняют свой смысл, но в этом случае важное значение имеет не только валентность примесного атома, но и его положение в кристаллической решетке.

Произведение концентрации n носителей отрицательного заряда $-q$ (электронов) на концентрацию p носителей положительного заряда $+q$ (дырок) равно постоянной величине n_i^2 , которая зависит только от вида вещества и температуры [202],

$$np = n_i^2, \quad (12.5)$$

где n_i — концентрация носителей заряда в материале без примесей.

В целом полупроводник остается нейтральным. Можно показать, что если концентрация донорных примесей N_D намного больше концентрации акцепторных примесей N_A ($N_D \gg N_A$), то выполняются соотношения

$$n \gg p: \quad n \approx N_D, \quad p \approx \frac{n_i^2}{N_D},$$

и, наоборот, если $N_A \gg N_D$, то (12.6)

$$p \gg n: \quad p \approx N_A, \quad n \approx \frac{n_i^2}{N_A}.$$

В первом случае легированный полупроводник называют полупроводником n -типа (электронный полупроводник), в котором основными свободными носителями заряда являются электроны ($-q$). Во втором случае основными носителями заряда являются дырки ($+q$), и такой материал называется полупроводником p -типа (дырочный полупроводник).

3. Рекомбинация носителей

При возбуждении в полупроводнике возникает пара носителей (электрон и дырка), которая затем может рекомбинировать. Рекомбинация происходит тем чаще, чем выше концентрация носителей заряда.

В полупроводниковых приборах, как явствует из формулы (12.6), влияние новых носителей заряда, введенных в равном количестве, будет значительно сильнее на неосновные носители заряда (которых меньше). Процесс рекомбинации описывается экспоненциальным законом, и постоянная времени в показателе степени называется временем жизни носителей заряда. На время жизни носителей заряда влияют наличие в материале химических примесей, дислокаций решетки, загрязнений поверхности полупроводника, а также наличие так называемых «ловушек», которые создают промежуточные энергетические состояния.

Рекомбинация носителей заряда может быть излучательной с испусканием фотона, энергия которого приблизительно равна ширине запрещенной зоны ΔW , и *безызлучательной*, когда высвободившаяся энергия преобразуется в тепло. Протекание этих двух процессов в полупроводнике характеризуют внутренней эффективностью η_i и временем жизни носителей τ . Обозначив время жизни носителей по отношению к излучательной и безызлучательной рекомбинации через τ_r и τ_{nr} , получим следующее выражение для внутренней эффективности:

$$\eta_i = \frac{\tau_{nr}}{\tau_r + \tau_{nr}}. \quad (12.7)$$

Величиной η_i определяется частота излучательных актов рекомбинации носителей в общем числе рекомбинаций [203]

$$\tau = \frac{\tau_r \tau_{nr}}{\tau_r + \tau_{nr}}. \quad (12.8)$$

В так называемых полупроводниках с непрямыми переходами (Si, Ge, GaP, AlAs) рекомбинация требует участия фотона и протекает достаточно медленно. В других полупроводниках (GaAs, InAs, InP, $\text{Al}_x\text{Ga}_{1-x}\text{As}$, $\text{GaAs}_{1-y}\text{P}_y$, где $x < 0,37$ и $y < 0,45$) рекомбинация происходит прямо и время жизни по отношению к излучательной рекомбинации очень мало: от 10^{-10} до 10^{-8} с [203].

В заключение отметим, что хорошую внутреннюю эффективность значительно легче достичь в полупроводниках с прямыми переходами, нежели в полупроводниках с непрямыми переходами.

4. Инжекция неосновных носителей и токов

Если в полупроводниковом кристалле создать область с проводимостью p -типа, а рядом с ней зону с проводимостью n -типа, то мы получим $p-n$ (или $P-N$)-переход (рис. 12.14), который может иметь резкую или плавную границу в зависимости от скорости изменения концентрации электронов и дырок.

Носители заряда могут перемещаться в полупроводнике под действием электрического (или магнитного) поля и под действием разности концентраций носителей.

В первом случае под действием электрического поля видоизменяются траектории движения носителей заряда и появляется средняя скорость их *дрейфа*, которая по направлению совпадает с направлением электрического поля, а по величине пропорциональна напряженности поля. Коэффициент пропорциональности называется подвижностью соответствующих носителей заряда (обозначается через μ) и неодинаков для носителей разного типа.

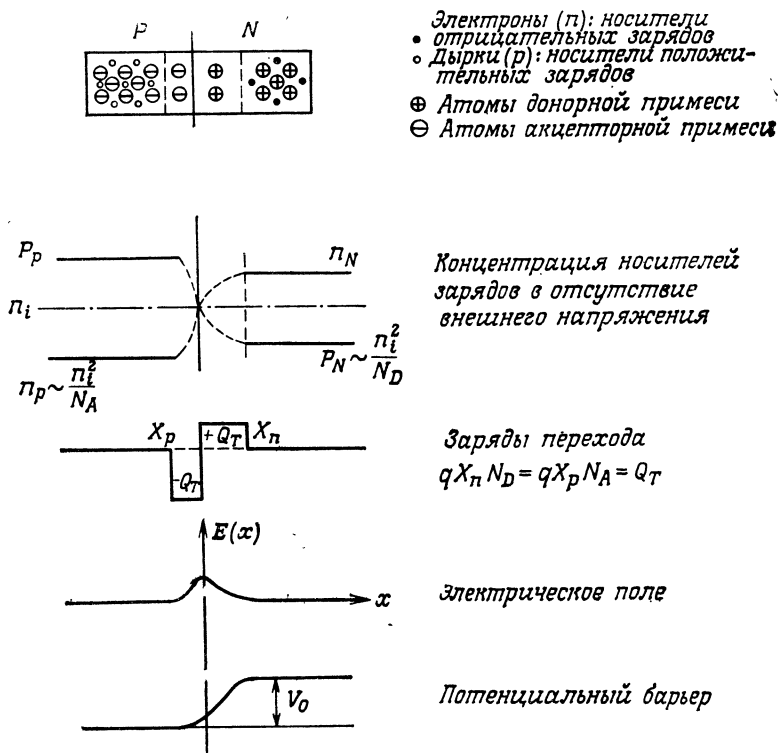


Рис. 12.14. Полупроводниковый $P-N$ -переход в условиях теплового равновесия.

Приравняв диффузионную составляющую тока составляющей тока проводимости, можно показать, что

$$n_p = n_n \exp \{-qV_0/kT\}, \quad p_n = p_p \exp \{-qV_0/kT\}.$$

Индекс указывает на тип материала (n или p) и тип носителя заряда.

Таким образом, можно написать

$$v_n = -\mu_n E; \quad v_p = \mu_p E. \quad (12.9)$$

В табл. 12.3 приведены значения подвижностей носителей заряда для некоторых типов полупроводниковых материалов.

Если в некоторой области полупроводника имеется избыток концентрации носителей заряда, то они диффундируют в соседние области с меньшей концентрацией, причем поток носителей пропорционален градиенту концентрации. Коэффициент пропорциональности обозначается через D_n или D_p в зависимости от типа носителя заряда. В отсутствие внешнего электрического

Таблица 12.3

		GaAs	Si	Ge	In Sb
Подвижность электронов	$\mu_n, \text{ м}^2/\text{В} \cdot \text{с}$	0,9	0,12	0,38	0,70
Подвижность дырок	$\mu_p, \text{ м}^2/\text{В} \cdot \text{с}$	0,05	0,05	0,19	0,10

поля число электронов и дырок, проходящих через единичную площадь, равно

$$D_n \frac{\partial (\Delta n)}{\partial x} - D_p \frac{\partial (\Delta p)}{\partial x},$$

где Δn — избыток электронов, а Δp — избыток дырок в направлении оси x .

Коэффициент диффузии D связан с подвижностью носителей μ соотношением Эйнштейна

$$D = \frac{kT}{q} \mu.$$

В процессе диффузии некоторое число неосновных носителей рекомбинирует. Среднее расстояние, которое пробегают носители до рекомбинации, называется длиной диффузии; последняя связана с коэффициентом диффузии и средним временем жизни носителя в данном материале формулой

$$L = \sqrt{D\tau}.$$

Миграция носителей заряда в направлении областей с меньшей концентрацией происходит с некоторой средней скоростью, которая называется скоростью диффузии, $v_{\text{дифф}}$ (гл. 13, § 3).

Таким образом, можно написать следующее выражение для полной плотности тока, обусловленного этими двумя видами смещения носителей заряда:

$$J_{\text{полн}} = J_d + J_{\text{дифф}}. \quad (12.10)$$

В $P-N$ -переходе основные носители P -области диффундируют в N -область, а основные носители N -области — в P -область. Поскольку заряды противоположного знака локализованы (на ионизированных атомах примеси) по разные стороны $P-N$ -перехода, нейтральность кристалла нарушается и возникает электрическое поле, которое противодействует диффузии носителей. В результате достигается состояние равновесия, когда влияние обоих факторов (диффузии и электрического поля) взаимно компенсируется. Электрическому полю в $P-N$ -

переходе соответствует разность потенциалов, которая в условиях равновесия равна (рис. 12.14)

$$V_0 = \frac{kT}{q} \ln \left(\frac{N_A N_D}{n_i^2} \right). \quad (12.11)$$

Если на концы монокристалла подать напряжение V , то оно будет складываться с потенциальным барьером V_0 и в зависимости от знака будет увеличивать его или уменьшать.

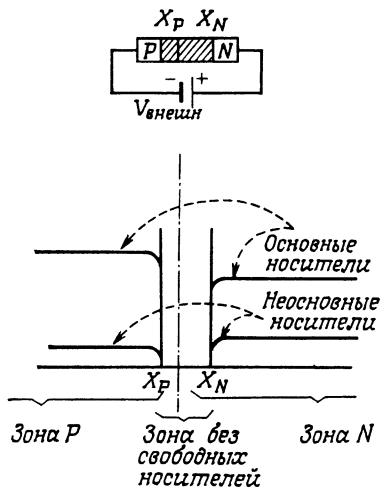


Рис. 12.15. Диаграмма концентрации носителей заряда при обратном смещении $P-N$ -перехода.

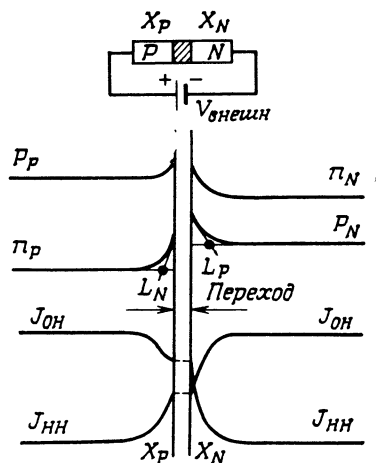


Рис. 12.16. Диаграмма концентрации носителей заряда и плотностей тока при прямом смещении $P-N$ -перехода.

Процессы рекомбинации в зоне перехода $X_P X_N$ не учитываются, поскольку они происходят с испусканием излучения и не способствуют лазерной генерации.

Рассмотрим сначала случай *обратного смещения* $P-N$ -перехода (рис. 12.15). В этом случае потенциальный барьер увеличивается и основные носители заряда уже не способны преодолеть его. Остается лишь ток, обусловленный прохождением через потенциальный барьер неосновных носителей заряда и называемый обратным током I_s .

В случае же *прямого смещения* (рис. 12.16) потенциальный барьер уменьшается и становится равным $-q(V_0 - V)$. Некоторая доля основных носителей заряда преодолевает этот барьер и вносит дополнительный вклад $I_s \exp(qV/kT)$ в общий ток $I_{\text{полн}}$, возникающий вследствие увеличения концентрации носителей заряда, которые диффундируют по ту сторону $p-n$ -пере-

хода и поэтому должны непрерывно пополняться. В результате окончательно получаем

$$I_{\text{полн}} = I_s \left(\exp \frac{qV}{kT} - 1 \right). \quad (12.12)$$

Б. Спонтанное и индуцированное излучение

Инжекция неосновных носителей создает в $p-n$ -переходе *инверсию заселенности* энергетических уровней и вызывает как спонтанное, так и индуцированное излучение света с центральной длиной волны, которая определяется шириной запрещенной зоны.

Поместив $p-n$ -переход в резонатор, можно выше определенного *порогового тока* получить усиление, превышающее оптические потери, и создать условия, при которых будет иметь место только индуцированное излучение. Как было показано в § 1, п. Б, длина волны излучения связана с модами, способными существовать в резонаторе. Обычно используют резонатор Фабри — Перо, зеркалами которого служат поверхности скола монокристалла. Поскольку на практике показатель преломления используемых материалов очень высок, это обеспечивает хорошее отражение световых волн от граней кристалла [204, 205].

Разумеется, могут быть использованы и другие способы создания инверсной заселенности, например оптическая накачка с использованием другого мощного лазера или катодолюминесценции (образование электронно-дырочных пар в результате столкновений с быстрыми электронами).

1. Условия эффективной работы

Пороговый ток первых лазерных диодов был очень большим (несколько десятков ампер), и поэтому приходилось работать в импульсном режиме, чтобы избежать разрушения прибора от перегрева. Чтобы перейти к непрерывному режиму работы, усилия были направлены на уменьшение порогового тока за счет увеличения усиления и на осуществление *удержания* носителей заряда и фотонов.

1. *Удержание носителей заряда.* Активная зона распространения неосновных носителей, инжектированных через $p-n$ -переход, ограничивается длиной их диффузии. Создав дополнительный потенциальный барьер на расстоянии, меньшем этой длины, например с помощью еще одного $p-n$ -перехода, можно ограничить зону распространения неосновных носителей, а заодно и уменьшить требуемую плотность тока (рис. 12.17).

Если по обе стороны $p-n$ -перехода один и тот же материал, то его называют *гомопереходом*, если же использованы

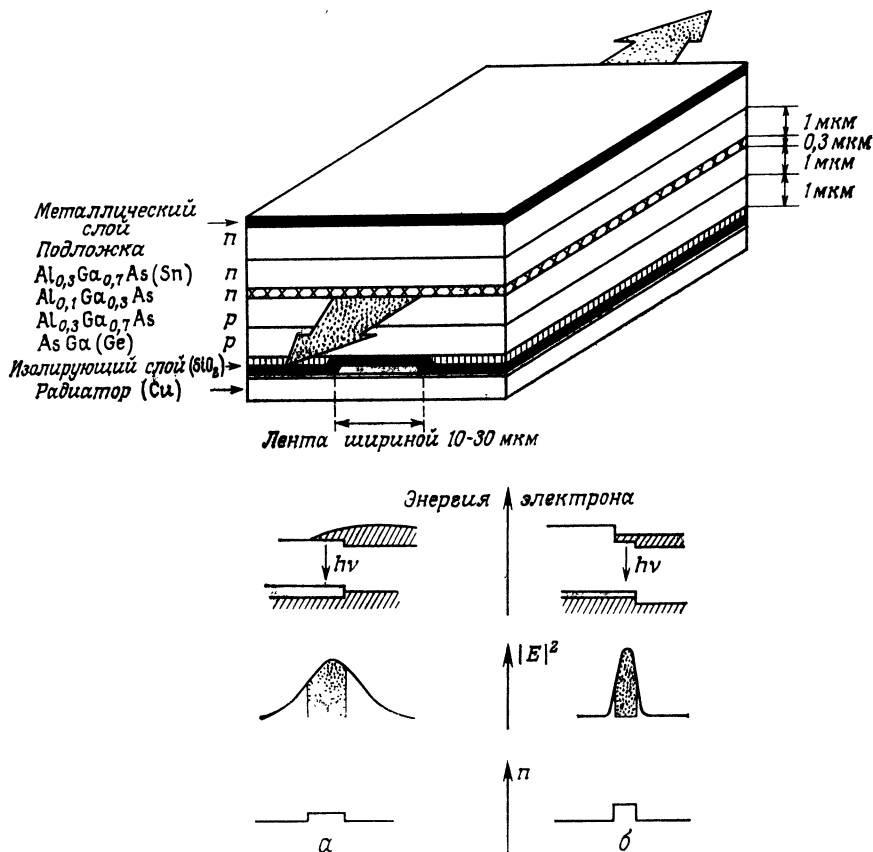


Рис. 12.17. Полупроводниковый лазер на двойном гетеропереходе с ленточной структурой, полученной методом избирательной изоляции.

Внизу для сравнения показаны энергетические зоны и ход изменения электрического поля и показателя преломления для лазера: α — на простом гетеропереходе и δ — на двойном гетеропереходе

разные материалы, то — *гетеропереходом* или *гетероструктурой*. Плотность порогового тока простой гетероструктуры — порядка 10^4 А/см², а в двойной гетероструктуре (рис. 12.17) она уменьшается до 10^3 А/см² [205].

Кроме того, при этом уменьшается полный ток, что приводит к уменьшению поперечных размеров p — n -перехода (ленточная структура, рис. 12.18).

2. *Удержание фотонов.* Изменение характера легирования, необходимого для обеспечения удержания носителей зарядов, в случае арсенида галлия приводит также и к уменьшению показателя преломления материала. Это полезный эффект, по-

скольку, как это было показано при рассмотрении оптических волокон, в этом случае имеет место явление направленного распространения света, чем и обеспечивается удержание фотонов, генерируемых в активной зоне.

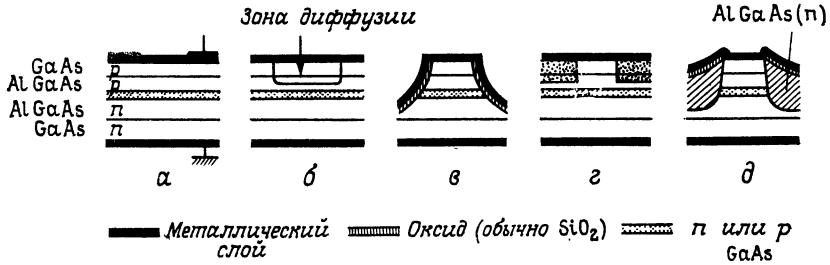


Рис. 12.18. Ленточные структуры (в разрезе), полученные разными способами. а — осаждением изолирующего оксида; б — избирательной диффузией цинка; в — технологией меза-структуры; г — формированием резистивных зон протонной имплантацией (бомбардировкой) на обеих сторонах ленточной структуры, д — методом «глубинной» ленты.

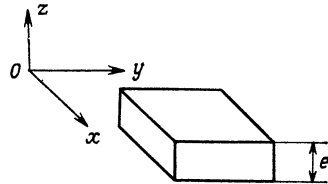
В заключение отметим, что к настоящему времени предложено очень много полупроводниковых структур для излучения света, но наиболее распространены двойные гетероструктуры с толщиной зоны рекомбинации порядка 0,3 мкм.

В. Оптические и электрические свойства полупроводниковых лазеров [206]

1. Моды излучения

Очевидно, что в объемном резонаторе существуют моды TE и TM . Анализируя трехмерный случай, можно выделить следующие типы мод (рис. 12.19).

Рис. 12.19. К определению типов мод в резонаторе полупроводникового лазера. Резонатор Фабри — Перо образован двумя гранями полупроводникового кристалла, параллельными плоскости zOy , причем толщина кристалла e всегда мала.



1. *Продольные моды* (распространяющиеся вдоль оси Oz). Это моды резонатора Фабри — Перо. Как будет показано в § 4, п. Г, 3, расстояние между такими модами составляет несколько ангстрем.

2. *Боковые моды* (распространяющиеся вдоль оси Oy). Они зависят от способа получения поверхностей кристалла. Обычно наблюдается несколько мод низкого порядка, отстоящих друг от друга на несколько десятков ангстрем.

3. *Поперечные моды* (вдоль оси Ox). Эти моды перпендикулярны плоскости $p-n$ -перехода и обусловлены разностью показателей преломления. Очевидно, что обычно стремятся к тому, чтобы возбуждалась только основная мода.

2. Пороговый ток

Пороговый ток сильно зависит от температуры, и эта зависимость имеет вид

$$I_s = I_0 \exp \frac{T}{T_0}, \quad (12.13)$$

где $T_0 \approx 120$ К.

В полупроводниковых лазерах с простой гетероструктурой пороговый ток уменьшается с увеличением расстояния между $p-n$ -переходом и гетероструктурой. В лазерах с двойной гетероструктурой он пропорционален объему резонатора [203—205].

3. Модуляция излучения

Самый простой метод модуляции излучения полупроводниковых лазеров состоит в модуляции тока инжекции I . Если инжекционный ток превышает пороговый ток, то наблюдается не-

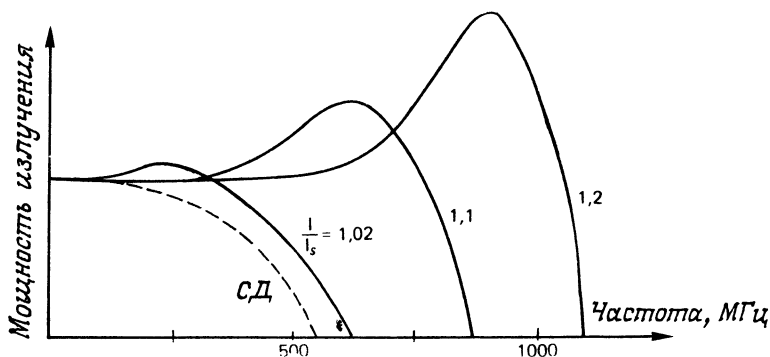


Рис. 12.20. Зависимость излучаемой мощности от частоты модуляции при разных значениях отношения тока модуляции к пороговому току. Ниже порога показана аналогичная зависимость для светодиода (штриховая линия).

которая задержка генерации излучения, описываемая приближенным выражением

$$t_e = t_0 \ln \left(\frac{I}{I - I_s} \right). \quad (12.14)$$

Поскольку электрические и оптические явления, наблюдающиеся в лазере, взаимосвязаны, при высоких частотах модуляции наблюдается резонанс между инжектируемыми носителями за-

ряда и электрическим полем. Резонансная частота тем выше, чем больше отношение I/I_S (рис. 12.20). Она дается формулой

$$\nu_r = \frac{m}{2\pi} \sqrt{\frac{1}{\tau_n \tau_p} \left(\frac{I}{I_S} - 1 \right)}, \quad (12.15)$$

где τ_n и τ_p — времена жизни электронов и дырок по отношению к спонтанной рекомбинации, а m — логарифм коэффициента усиления по току (типичные значения m лежат между 1 и 3). Рассматриваемые явления подобны тем, которые наблюдаются в газовых лазерах (рис. 12.5). На рис. 12.20 представлены типичные кривые зависимости мощности излучения от частоты модуляции тока накачки.

Г. Основные характеристики полупроводниковых лазеров

К числу основных характеристик полупроводниковых лазеров, определяющих возможность их использования в системах связи и передачи информации, относятся: мощность излучения и ее зависимость от тока модуляции, диаграмма направленности излучения, спектр излучения, срок службы.

1. Мощность излучения $P(I)$

По экспериментально измеренной зависимости мощности излучения полупроводникового лазера от тока накачки можно определить пороговый ток и энергоотдачу. На кривых такой за-

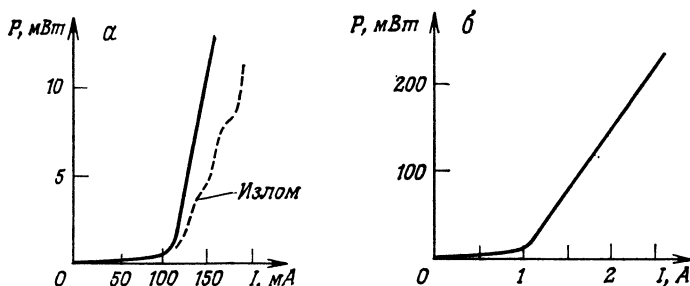


Рис. 12.21. Примерная зависимость излучаемой мощности от тока накачки $P(I)$ для полупроводникового лазера.

a — в непрерывном режиме; b — в импульсном режиме с коэффициентом заполнения менее 10%. КПД лазера измеряется крутизной $\Delta P/\Delta I$.

висимости (рис. 12.21) обнаруживаются нелинейности (перегибы и изломы), выраженные в той или иной степени. Кроме того, мощность, излучаемая лицевой гранью, ограничивается величиной, выше которой происходит разрушение граней кристалла, что делает несовместимыми большую мощность излучения и непрерывный режим работы.

2. Диаграмма направленности излучения

Направленность излучения полупроводниковых лазеров значительно меньше, чем газовых лазеров (рис. 12.22), по следующим причинам:

а) из-за большого показателя преломления полупроводникового кристалла оказываются большими углы выхода излучения из кристалла;

б) малая толщина активной зоны лазера оказывает сильное влияние на дифракционные эффекты (гл. 4, § 1, п. Д).

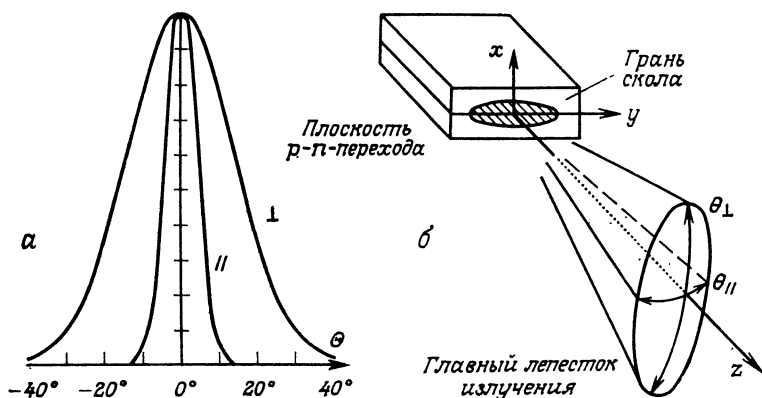


Рис. 12.22. Диаграмма направленности излучения полупроводникового лазера. а — диаграмма направленности; приведена только активная зона, ширина диаграммы направленности измеряется на уровне половинной мощности; б — зависимость излучаемой мощности от угла во взаимно перпендикулярных направлениях Ox и Oy

Диаграмма направленности излучения полупроводникового лазера *несимметрична*: ее ширина менее 20° в плоскости, параллельной переходу, и более 40° в перпендикулярной плоскости. В случае, когда существует несколько поперечных мод, наблюдается несколько боковых лепестков диаграммы направленности излучения в дальней зоне.

3. Спектр излучения

В спектральной области, где можно определить коэффициент усиления резонатора, существуют только такие продольные моды (гл. 12, § 4, п. В, 1), для которых длина пробега света в резонаторе туда и обратно равна целому числу длин волн. Если p — порядок моды, а λ — длина волны излучения, то для этой моды имеем

$$p\lambda = 2n(\lambda)L, \quad (12.16)$$

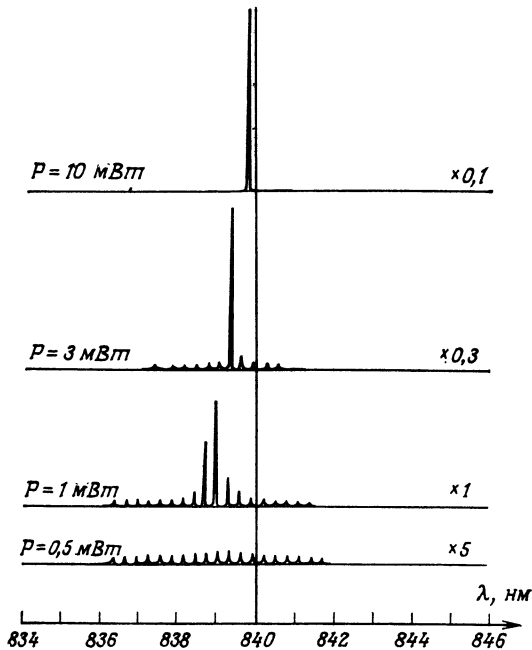


Рис. 12.23. Зависимость спектра излучения полупроводникового лазера от излучаемой оптической мощности.

Справа на горизонтальных осях указан масштабный множитель для вертикальной шкалы

а для следующей соседней моды порядка $p + 1$ с длиной волны λ' аналогично можно написать

$$(p + 1) \lambda' = 2n(\lambda') L.$$

Отсюда получаем следующую формулу для интервала между соседними модами:

$$\Delta\lambda = \frac{\lambda^2}{2nL \left(1 - \frac{\lambda}{n} \frac{dn}{d\lambda} \right)}. \quad (12.17)$$

Таким образом, кажущийся показатель преломления резонатора есть групповой показатель преломления (гл. 2, § 5 и гл. 8, § 5). Простой числовой пример ($\lambda = 0,9$ мкм, $n = 3,6$ и $L = 300$ мкм) показывает, что интервал между модами равен $\Delta\lambda \approx 0,4$ нм. Этот элементарный расчет объясняет и регулярную линейчатую структуру спектра излучения полупроводникового лазера (рис. 12.23).

4. Срок службы и надежность

Проблема долговечности и надежности полупроводниковых лазеров остается весьма сложной и еще нерешенной в настоящее время. Еще не создано полупроводниковых лазеров с высоким к. п. д. и сроком службы, превышающим значения, требуемые для использования в системах связи, а именно 10^5 ч для наземных и 10^6 ч для подводных ретрансляторов.

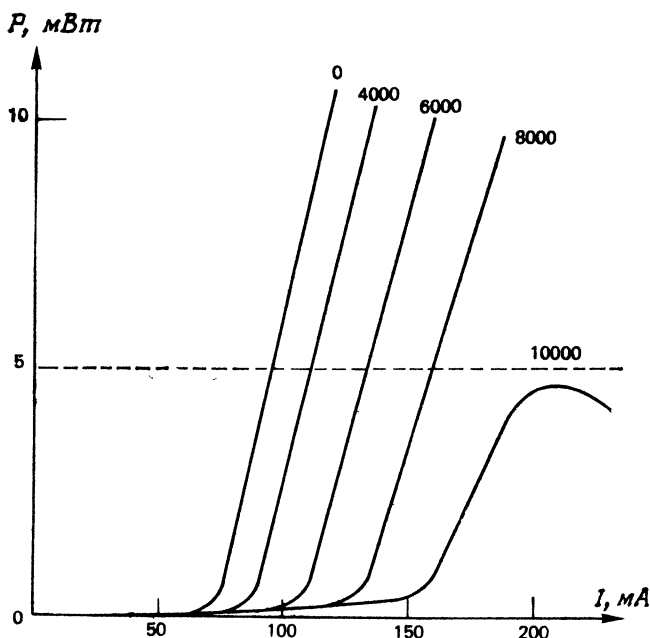


Рис. 12.24. Изменение характеристики $P(I)$ полупроводникового лазера при непрерывной работе с излучаемой мощностью 5 мВт.

Кривые были сняты после 0, 4000, 6000, 8000 и 10 000 ч работы.

Тем не менее развитие этой области идет быстро, и во многих лабораториях уже превышен срок службы полупроводниковых лазеров в 10^4 ч. Экстраполируя результаты испытаний при высоких температурах (от 60 до 80 °C), исследователи оценивают срок службы таких полупроводниковых лазеров величиной 10^5 ч и надеются достичь 10^6 ч. На рис. 12.24 показано, как изменяется рабочая характеристика полупроводникового лазера из-за деградации кристалла. В сущности нужно преодолеть два вида деградаций: катастрофическую деградацию, обусловленную световым потоком, и постепенную деградацию, обусловленную плотностью тока накачки, числом рабочих циклов, металлургическими процессами и т. д. Под металлургическими про-

цессами понимается образование в кристалле неизлучающих центров и неоднородных дефектов структуры кристалла, которые зависят от типа полупроводника, степени очистки материала и технологии изготовления лазера. Очевидно, что несоответствия кристаллических решеток материалов, образующих гетероструктуры, содействуют возникновению дислокаций и дефектов. Экспериментальные поиски ведутся в следующих направлениях:

- а) улучшение процесса эпитаксии (§ 4, п. Д);
- б) более тщательное изготовление омических контактов;
- в) добавка фосфора в слой GaAlAs для лучшего соответствия параметров решетки;
- г) пассивирование граней кристалла.

Д. Технология изготовления полупроводниковых лазеров

Не вдаваясь в подробности, можно сказать, что классический способ изготовления полупроводниковых лазеров — это *эпитаксиальный рост из жидкой фазы*, и задача состоит в том, чтобы вырастить на подложке гетерогенные пленки разной структуры и ввести в них различные примеси [207, 208].

Подложку поддерживают в тепловом равновесии с жидким материалом, образующим новую пленку. В результате постепенного охлаждения на поверхности подложки происходит кристаллизация и образуется пленка, толщина которой будет пропорциональна времени выдержки в соответствующей ванне. По порядку величины это время равно 1 мин при выращивании пленок зоны удержания и нескольким секундам для пленки активной зоны.

Если при переходе от одной пленки к другой возникают различия в кристаллической решетке из-за изменения примесей и их содержания, пленка отстает и растрескивается.

Существуют два способа получения ленточных структур: меза-технология и планарная технология. В последнем случае используются главным образом три метода.

1. Ленту размером от 10 до 20 мкм вырезают в слое изолирующей окиси кремния (SiO_2) и наносят металл. Омический контакт возникает только в зоне выреза.

2. Можно также проводить избирательную диффузию через зону *n*-типа к зоне *p*-типа.

3. Можно увеличить удельное сопротивление обеих сторон ленты путем бомбардировки протонами.

В заключение отметим, что в технологии изготовления полупроводниковых лазеров имеется много разных интересных возможностей. Поскольку генерируемые частоты определяются шириной запрещенной зоны, можно создать лазер с заданной

длиной волны излучения, изготовив соответствующий полупроводник. В этом отношении перспективны соединения $\text{Ga}_x\text{In}_{1-x}\text{As}$ и $\text{Ga}_x\text{Al}_{1-x}\text{As}$, в которых ширина запрещенной зоны является непрерывной функцией величины x . Изменять длину волны излучения можно, также изменяя рабочую температуру.

В то же время можно использовать более избирательную отрицательную обратную связь в резонаторе Фабри—Перо, осуществив, например, распределенную отрицательную обратную связь на основе дифракционных решеток. В этом случае длина волны излучения лазера оказывается особенно точной и меньше ширина линии [209, 210].

Технология полупроводниковых лазеров развивается быстро, и переход к большим длинам волн заставляет рассматривать, кроме соединений III—V групп, более сложные соединения, например четырехкомпонентные [211].

§ 5. Светодиоды [212]

Мы не будем очень подробно останавливаться на этих устройствах, в которых используется спонтанное излучение за счет инверсной заселенности энергетических уровней, создаваемой путем инжекции в $p-n$ -переход, смещенный в прямом направлении, неосновных носителей заряда.

Спонтанное излучение изотропно, и кривая усиления активной зоны дает спектральную ширину от 30 до 40 нм. Благодаря квазиламбертову распределению интенсивности в пучке излучения эффективность связи такого источника с оптическим волноводом не высока (1—10%). Однако *сравнительно простая конструкция светодиодов, их высокая надежность и достаточно слабая зависимость характеристик излучения от температуры делают их особенно подходящими для систем связи на короткие расстояния при невысокой информационной пропускной способности.*

В соответствии с характеристиками светодиодов, приведенными на рис. 12.25, излучаемая ими мощность на значительном участке почти пропорциональна току модуляции. Отметим, что обусловленное нагревом насыщение проявляется на высоком уровне мощности (приблизительно от 1 до 2 мВт) в непрерывном режиме работы.

Здесь тоже имеются *гомопереходы* и *гетеропереходы*.

Примером структуры с простым гомопереходом может служить диффузионный диод, получаемый диффузией атомов донорной примеси в пленку из GaAs, имеющую проводимость n -типа.

Что касается моделей гетеропереходов, то различают два вида структур этого типа.

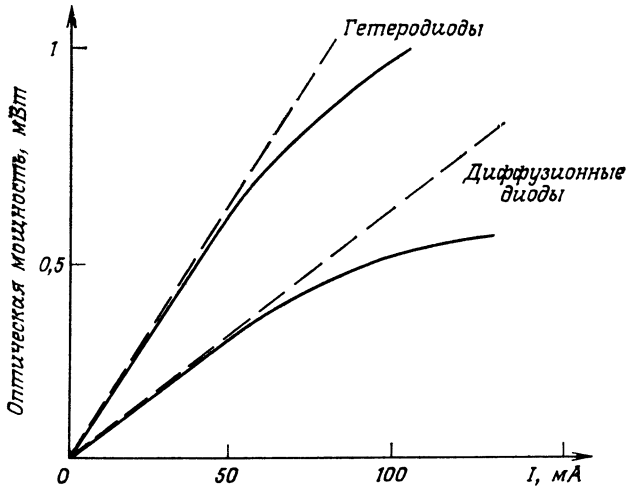


Рис. 12.25. Зависимость излучаемой светодиодом оптической мощности от тока возбуждения.

Сплошные линии — в непрерывном режиме, штриховые линии — в импульсном режиме.

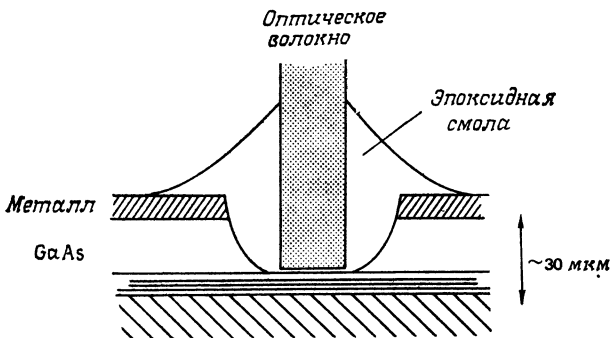


Рис. 12.26. Светодиод с двойной гетероструктурой (в разрезе), оптически связанный с волоконным световодом.

В верхнем слое протравлено углубление, чтобы можно было приблизить торец волокна к излучающей зоне СД.

1. Структура, в которой излучение происходит с поверхности [213] (рис. 12.26). Здесь зону рекомбинации приближают к теплоотводу (радиатору), а в подложке делают углубление для осуществления контакта с оптическим волокном.

2. Структура, в которой излучение происходит через ребро за счет направленного внутреннего распространения, как в

лазерных диодах. Здесь следует уменьшить длину активной зоны, чтобы исключить поглощение, поскольку в этом случае более длинный пробег.

В обоих случаях модуляция излучения светодиода осуществляется изменением тока возбуждения. Быстродействие светодиода ограничивается временем рекомбинации носителей заряда (1 нс) и его собственной емкостью. *Мощность излучения может быть больше 1 мВт, а полоса пропускания — больше 100 МГц.* У светодиодов на гетеропереходе с хорошим удержанием носителей заряда и фотонов достигнута полоса пропускания в 300 МГц [214].

ГЛАВА 13

ФОТОПРИЕМНИКИ

Известно, что световая энергия переносится фотонами и преобразование фотонов в электрический ток осуществляется фотоприемниками. Физические процессы, протекающие в полупроводниковых фотоприемниках, носят *обратный характер по отношению к процессам, рассмотренным в гл. 12*: под воздействием фотона с достаточно большой энергией электрон выбивается из валентной зоны и переходит в зону проводимости, в результате чего образуется пара свободных носителей заряда, которые, перемещаясь к противоположным зажимам фотоприемника, переносят фототок.

В данной главе мы ограничимся тем, что кратко рассмотрим полупроводниковые фотодиоды, применяемые в системах оптической связи. Напомним, что в $p-n$ -переходе, на который подано обратное смещение (гл. 12, § 4, рис. 12.5), существует зона, в которой нет свободных носителей заряда (обедненная зона) и в которой все определяется электрическим полем. Каждая возникающая в этой зоне пара электрон — дырка разделяется и дает вклад в ток, когда электрон и дырка выходят за границы обедненной зоны. Если пара носителей заряда возникает вне обедненной зоны, то она локально изменяет концентрацию свободных носителей заряда, что приводит к диффузии зарядов, направленной на выравнивание их концентрации. Если же пара электрон — дырка возникает на расстоянии от обедненной зоны, превышающем длину диффузии, то вероятность ее рекомбинации будет близка к единице и, следовательно, поглощенный фотон не даст вклада в ток. Если электрическое поле в обедненной зоне достаточно сильное, то ускоренные им носители заряда могут приобрести энергию, равную энергии ионизации, и выбить другие носители заряда, которые, тоже получив достаточное ускорение, в свою очередь ионизируют другие атомы и т. д., т. е. концентрация носителей заряда будет лавинно нарастать.

Характеристики фотоприемников зависят от двух основных факторов:

- 1) от *материала*, которым определяется коэффициент поглощения света, квантовый выход, а также подвижность носителей заряда;

- 2) от *конфигурации* фотоприемника и от подаваемого на него напряжения, чем определяется ширина запрещенной зоны,

время пролета, напряженность электрического поля и электрические шумы.

Важное значение имеет и предусилитель, который следует за фотоприемником в канале регистрации излучения [215—219]; подробнее данный вопрос будет рассмотрен в гл. 15.

§ 1. Основные характеристики фотодиодов

А. Поглощение излучения

Поглощение света в веществе описывается *экспоненциальным законом* (рис. 13.1)

$$\Phi = \Phi_0 \exp \{-\alpha x\}. \quad (13.1)$$

Параметр α называется *показателем поглощения* вещества. Обратная ему величина имеет размерность длины и равна тол-

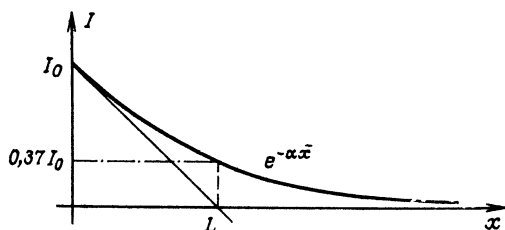


Рис. 13.1. Зависимость интенсивности света I от глубины x проникновения в среду

Ослабление света в среде обусловлено образованием электронно-дырочных пар.

щине вещества, при которой световой поток ослабляется в e раз, где e — основание натуральных логарифмов. Это соответствует поглощению около 63% излучения, падающего на вещество. Показатель поглощения α очень сильно зависит от длины волны λ , т. е. от энергии падающих фотонов и, значит, от их способности переводить электроны в валентную зону. Из приведенных на рис. 13.2 зависимостей показателя поглощения от длины волны следует, что свет создает заметный фототок только в определенной спектральной области, резко ограниченной с одной стороны критической длиной волны λ_c , которая соответствует минимальной энергии кванта, необходимой для преодоления запрещенной зоны ΔW вещества. Критическая длина волны дается формулой

$$\lambda_c = \frac{hc}{\Delta W}, \quad (13.2)$$

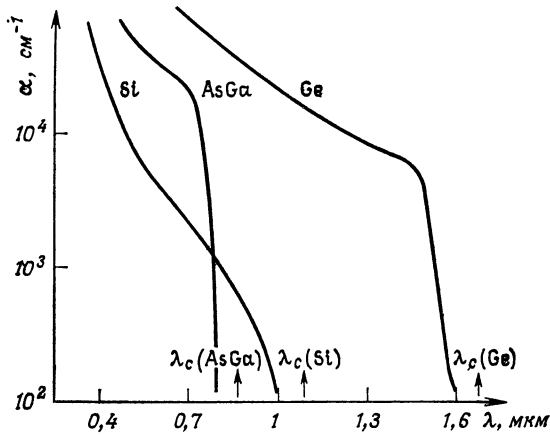


Рис. 13.2. Зависимость показателя поглощения α от длины волн. Стрелками указаны критические длины волн.

где h — постоянная Планка. Если энергию выразить в электрон-вольтах, а длину волны в микрометрах, то эта формула принимает вид

$$\lambda_c = \frac{1,24}{\Delta W}. \quad (13.2a)$$

Величины ΔW и λ_c для ряда используемых на практике полупроводниковых материалов приведены в табл. 13.1, которая служит дополнением к табл. 12.1.

Таблица 13.1

Материал	Si	Ge	GaAs	$\text{Ga}_x\text{In}_{1-x}\text{As}$	$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$
ΔW , эВ	1,1	0,72	1,43	1,43—0,36	1,35—0,36
λ_c , мкм	1,1	1,7	0,87	0,87—3,44	0,92—3,44

Другая граница фоточувствительности, со стороны коротких длин волн, обусловлена очень сильным поглощением излучения вблизи поверхности образца полупроводника, где очень мало время жизни носителей тока до их рекомбинации.

Б. Квантовый выход

Квантовый выход η — это число образовавшихся пар электрон — дырка, приходящееся на один упавший фотон. Следовательно, это как бы КПД данного процесса генерации

электрического тока. Фототок I_{ph} , создаваемый потоком фотонов φ_0 , будет равен

$$I_{ph} = q\eta\varphi_0, \quad (13.3)$$

где φ_0 — число фотонов, падающих в единицу времени.

Три фактора приводят к уменьшению квантового выхода η в фотодиодах: 1) потери, обусловленные отражением на входном диоптре; 2) объемная рекомбинация возникающих носителей заряда; поэтому стараются добиться того, чтобы генерация пар носителей заряда происходила в обедненной зоне, где электрическое поле разделяло бы образовавшуюся пару носителей, и уменьшить время пролета носителей, а следовательно, и вероятность их рекомбинации; 3) поверхностная рекомбинация, которая создает токи утечки; для уменьшения токов утечки применяют охранные кольца.

В. Постоянная времени [219]

Постоянная времени фотоприемника характеризует его быстродействие. Есть и другая характеристика быстродействия — время нарастания¹⁾. Постоянная времени фотоприемника зависит от многих параметров: от подвижности носителей заряда, ширины обедненной зоны, длины волны света. Она зависит и от того, движутся ли носители заряда под действием электрического поля или вследствие диффузии.

1. Влияние диффузии

Предположим, что толщина слоя объемного заряда мала по сравнению с длиной поглощения $L = 1/\alpha(\lambda)$. Тогда большая часть пар электрон — дырка будут перемещаться под действием диффузии и только те из них, которые достигнут обедненной зоны, дадут вклад в фототок. Следовательно, полезными будут те пары носителей заряда, которые генерируются на расстоянии, меньшем диффузионной длины, от обедненной зоны. Можно ввести понятие скорости диффузии носителей [202], которая пропорциональна логарифмической производной от локальной концентрации носителей заряда $C(x)$:

$$v_{\text{дифф}} = -D \frac{1}{C(x)} \frac{dC}{dx}. \quad (13.4)$$

¹⁾ Постоянная времени фотоприемника характеризует его реакцию на короткий импульс и равна времени от момента возникновения фототока до момента его возвращения к нулю. Время же нарастания характеризует отклик на воздействие в форме единичной ступенчатой функции (функции Хэвисайда) и равно времени, за которое фототок достигает установившегося значения. Таким образом, в постоянную времени входит время возврата в исходное состояние равновесия, и поэтому она лучше характеризует быстродействие, чем время нарастания.

Здесь D — коэффициент диффузии, который зависит от типа рассматриваемых носителей заряда (гл. 12, § 4). Если концентрация носителей распределяется по экспоненциальному закону, то скорость диффузии $v_{\text{дифф}}$ — постоянная величина, равная произведению $D\alpha$. Если предположить, что полезная толщина полупроводника e равна длине поглощения, то легко найти время, за которое носители ее проходят:

$$t_{\text{дифф}} = \frac{e}{v_{\text{дифф}}} = \frac{1}{\alpha^2 D}. \quad (13.5)$$

Влияние диффузии меньше, если p — n -переход расположен близко от поверхности и если велика толщина слоя объемного заряда.

2. Влияние времени перехода в обедненную зону

При достаточно высоких напряженностях электрического поля носители заряда будут перемещаться с предельной скоростью (гл. 12, § 4, п. А, 4), так что, согласно формуле (12.9), максимальное время перехода для электронов и дырок будет таким:

$$t_n = e/\mu_n E, \quad t_p = e/\mu_p E. \quad (13.6)$$

Следовательно, для уменьшения времени перехода следует уменьшать толщину e . Заметим, что формулы (13.6) относятся к случаю резкого p — n -перехода. В случае плавного p — n -перехода выражения для скоростей диффузии немного изменятся, но выводы останутся прежними.

3. Влияние емкости p — n -перехода

Поскольку на p — n -переходе возникают заряды $+Q$ и $-Q$ (рис. 12.14), у p — n -перехода имеется емкость, которая зависит от внешнего напряжения V и от ширины e слоя объемного заряда. Ей соответствует определенная постоянная времени, которая ограничивает полосу пропускания. Поэтому нужно увеличивать ширину e (например, для обеспечения полосы пропускания 3 ГГц величина e должна быть порядка 10 мкм).

Итак, приходится идти на компромисс между различными эффектами, снижающими быстродействие фотоприемника, потому что их учет приводит к противоположным требованиям.

Г. Чувствительность

Чувствительность фотоприемника — это полный КПД преобразования световой мощности в электрический ток:

$$S = \frac{I_{ph}}{q_0 h \nu},$$

где Φ_0 — поток световой энергии, т. е. произведение числа фотонов (в единицу времени) на энергию одного фотона. С учетом формулы (13.3) можно переписать выражение для чувствительности фотоприемника в виде

$$S = q\eta/h\nu. \quad (13.7)$$

Отсюда следует, что чувствительность фотоприемника тем выше, чем больше квантовый выход η , т. е. чем больше доля светового потока, поглощаемая в активной зоне (рис. 13.3).

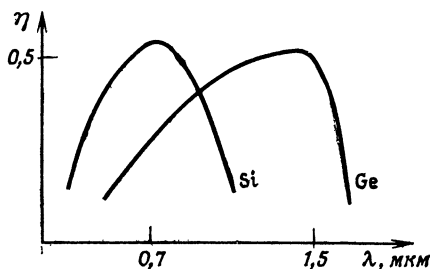


Рис. 13.3. Зависимость квантового выхода η от длины волны λ .

Если учесть противоречивые требования, предъявляемые к ширине активной зоны (п. В), то ее оптимальное с точки зрения полосы пропускания значение будет таким:

$$e = L = 1/\alpha(\lambda).$$

Такая ширина активной зоны обеспечивает квантовый выход $\sim 60\%$ и чувствительность на длине волны 1 мкм, равную $S = 0,5$ А/Вт.

Другой фактор, ограничивающий возможности фотодиодов, а следовательно, и систем оптической связи, — это *шум* (§ 6).

Д. Улучшение характеристик фотодиодов

Уменьшив степень *легирования* слоя N -типа, можно увеличить ширину слоя объемного заряда при том же рабочем напряжении. В пределе мы получаем беспримесный материал с собственной проводимостью (обозначаемый буквой I), к которому добавляется слой материала N -типа с малым удельным сопротивлением для обеспечения омического контакта. Такова структура PIN , обеспечивающая квантовый выход, близкий к единице, и очень высокую чувствительность.

Можно также повысить напряжение на фотодиоде до уровня, при котором возможно *лавинное усиление* (§ 4). Тогда ширина обедненного слоя будет определяться внешним напряжением и удельным сопротивлением материала (при постоянном напряжении величина e пропорциональна корню квадратному из удельного сопротивления материала). Кроме того, чем больше

ширина обедненного слоя, тем меньше емкость $p-n$ -перехода и тем меньше постоянная времени фотодиода.

В следующих параграфах мы подробно рассмотрим несколько отдельных структур фотодиодов.

§ 2. PIN -фотодиод

Фотодиод такого типа получается, если между слоем P -типа и слоем N -типа с малым удельным сопротивлением ввести слой I -типа с большим удельным сопротивлением, почти равным собственному удельному сопротивлению материала (рис. 13.4). На диод подается обратное напряжение, достаточное для обеднения свободными носителями заряда всего слоя с собственной

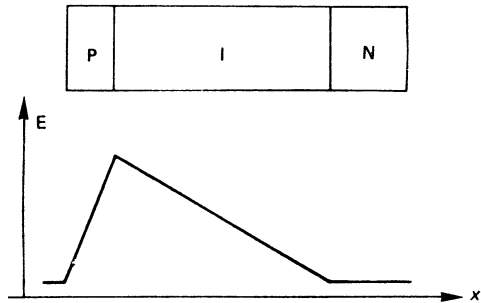


Рис. 13.4. Ход изменения электрического поля в PIN -фотодиоде.

проводимостью. Конструктивно PIN -фотодиод выполняется так, чтобы максимально уменьшить долю поглощения излучения вне слоя I -типа. С этой целью переход формируют у самой поверхности кристалла. Следовательно, постоянная времени такого диода определяется временем перехода носителей заряда через обедненный слой.

В установившемся режиме плотность полного тока, проходящего через $p-n$ -переход, будет равна [202]

$$J_{\text{полн}} = J_d + J_{\text{дифф}}, \quad (13.8)$$

где J_d — плотность тока носителей заряда, генерируемых в обедненном слое, а $J_{\text{дифф}}$ — плотность диффузионного тока носителей, генерируемых вне этого слоя. Если предположить, что тепловой ток пренебрежимо мал и что толщина слоя N -типа значительно меньше длины поглощения $L = 1/\alpha$, то скорость генерации пар носителей будет тогда равна (рис. 13.1)

$$G(x) = \varphi_0 \alpha e^{-\alpha x}. \quad (13.9)$$

Отсюда получаем следующее выражение для плотности тока:

$$J_d = q \int_0^e G(x) dx = q\varphi_0 (1 - e^{-\alpha e}). \quad (13.10)$$

Плотность неосновных носителей n_p при $x > e$ определяется из уравнения диффузии [203]

$$D_n \frac{\partial^2 n_p}{\partial x^2} - \frac{n_p - n_{p0}}{\tau_n} + G(x) = 0, \quad (13.11)$$

где D_n — коэффициент диффузии, τ_n — время жизни избыточных носителей, а n_{p0} — плотность электронов в равновесном состоянии.

Решение уравнения (13.11) имеет вид

$$n_p = n_{p0} + C_1 e^{-\alpha x} - (n_{p0} + C_1 e^{-\alpha e}) \exp \frac{e-x}{L_n}, \quad (13.12)$$

где введено обозначение $C_1 = \varphi_0 \alpha \tau_n / (1 - \alpha^2 L_n^2)$, причем $L_n = \sqrt{D_n \tau_n}$ — длина диффузии.

Из (13.12) находим плотность диффузионного тока:

$$J_{\text{дифф}} = -q D_n \frac{\partial n_p}{\partial x} \Big|_{x=e},$$

а затем на основе выражений (13.8) и (13.10) и полную плотность тока:

$$J_{\text{полн}} = q \varphi_0 \left(\frac{e^{-\alpha e}}{1 - \alpha L_n} \right) - q n_{p0} \frac{D_n}{L_n}. \quad (13.13)$$

Как правило, здесь можно пренебречь вторым слагаемым, содержащим n_{p0} , и тогда плотность полного тока оказывается пропорциональной падающему световому потоку. Она максимальна при двух условиях $\alpha e \gg 1$ и $\alpha L_n \gg 1$, которые противостоят требованию малой постоянной времени, так как увеличение e влечет за собой увеличение времени перехода. Для оценки влияния времени перехода на постоянную времени можно измерить фазовый сдвиг между фототоком и световым потоком, модулируемым высокой частотой. Для простоты предположим, что внешнее напряжение достаточно велико и поэтому в слое с собственной проводимостью нет свободных носителей заряда, а те носители, которые проходят через него, движутся с предельной скоростью при данном электрическом поле, т. е. $v = v_s$. Обозначив круговую частоту модуляции через ω , можно представить световой поток в виде $\varphi = \varphi_1 \exp \{i\omega t\}$. Вклад, вносимый слоем толщиной dx по оси x , будет равен

$$dJ = q \eta \varphi_1 \exp \left[i\omega \left(t - \frac{x}{v_s} \right) \right] dx,$$

если положить $\alpha = 0$. Следовательно,

$$J = q \eta \varphi_1 e^{i\omega t} \int_0^e \exp \left(-i\omega \frac{x}{v_s} \right) dx, \quad (13.14)$$

$$J \approx q \eta \varphi_1 e \frac{1 - \exp(i\omega t_r)}{i\omega t_r},$$

где $t_r = e/v_s$ — время, за которое носитель проходит через обедненный слой. Итак, ток проводимости амплитудно модулирован функцией вида $[1 - \exp(i\omega t_r)]/i\omega t_r$, график которой представлен на рис. 13.5. Если пренебречь влиянием тока смещения, который обусловлен внешним напряжением и не зависит от времени, то нетрудно видеть, что при $\omega t_r = 2,4$ эта функция умень-

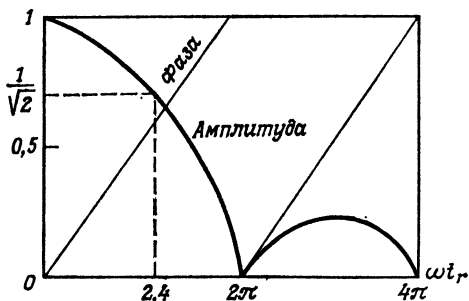


Рис. 13.5. Изменения амплитуды и фазы функции $[1 - \exp(i\omega t_r)]/i\omega t_r$.

шается на 3 дБ. Следовательно, полоса пропускания на уровне 3 дБ будет равна

$$f_{3\text{дБ}} = \frac{2,4}{2\pi t_r} \approx 0,4\nu_s. \quad (13.15)$$

Отсюда следует, что хороший компромисс между требованиями быстродействия и чувствительности достигается при $e = 1/\alpha$.

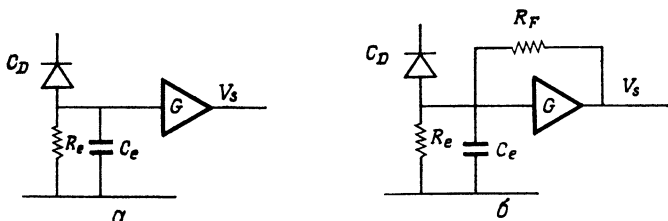


Рис. 13.6. Схемы включения фотодиода.

a — схема с усилителем напряжения: $V_S = R_e I_D [1 + (R_e/G)(C_e + C_D)p]^{-1}$; b — к фотодиоду подключается операционный усилитель с отрицательной обратной связью: $V_S = R_e I_D [1 + R_e(C_e + C_D)p]^{-1}$; C_D — емкость фотодиода.

В кремниевых фотодиодах при освещении с поверхности излучением полупроводникового лазера на GaAs ($0,85 < \lambda < 0,92$ мкм) или ИАГ-лазера ($\lambda = 1,06$ мкм) толщина обедненного слоя должна быть равна в первом случае 20—50 мкм, а во втором — 500 мкм. Это приводит к значительному времени перехода, которое, однако, можно уменьшить, если освещать фотодиод «с ребра», что увеличивает полезную длину поглощения без увеличения толщины e .

Освещаемый фотодиод эквивалентен генератору тока. Следовательно, возможны два варианта его включения (рис. 13.6).

1. Можно просто снимать напряжение, пропорциональное фототоку, с нагрузочного резистора R_e и подавать его на усилитель напряжения. При этом шумовой ток на входе будет тем меньше, чем больше сопротивление резистора. В сочетании с полевым транзистором такое решение позволяет регистрировать с низким уровнем шума очень малые световые мощности (десятки пиковатт).

2. Можно взять «операционный» усилитель с очень большим коэффициентом усиления и малым входным током, имеющий отрицательную обратную связь, благодаря чему его выходное напряжение пропорционально входному току.

Оба решения обеспечивают практически одинаковые характеристики с точки зрения шумов [215].

В заключение напомним самые важные полученные результаты: *PIN*-фотодиоды очень просты по своей структуре, обладают хорошей линейностью в широком динамическом диапазоне (от нескольких пиковатт до десятков милливатт), *просты в эксплуатации и дешевы*. В настоящее время они обеспечивают детектирование оптических сигналов, модулируемых частотами *гигагерцового* диапазона. Но непрерывно совершенствуемые фотодиоды с лавинным усилением могут вытеснить их в широкополосных системах.

§ 3. Плоскостной фотодиод

Эти фотодиоды, в которых поглощение происходит в основном вне обедненного слоя и, следовательно, постоянная времени определяется диффузией носителей. Если предположить, что толщина обедненного слоя мала и большая часть актов образования пар носителей заряда происходит вне этого слоя, то можно провести такой же расчет, как и в предыдущем параграфе. Но теперь временная задержка будет определяться диффузией [формула (13.5)] и мы получим следующее выражение для полосы пропускания:

$$f_{\text{злб}} = 2,4 \frac{1}{2\pi t_{\text{дифф}}} = 0,4\alpha^2 D, \quad (13.16)$$

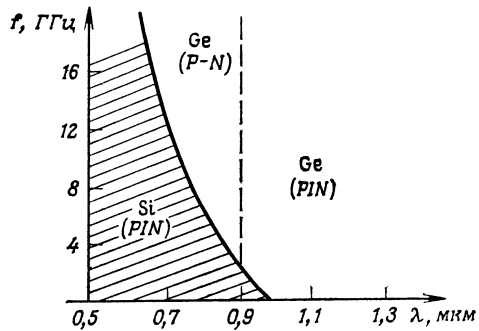
где по-прежнему $e = 1/\alpha(\lambda)$. В случае промежуточных структур, в которых имеются градиенты концентраций, или при плохом согласовании фотоприемника с рабочей длиной волны полюса пропускания будет лежать в пределах

$$0,4\alpha^2 D < f < 0,4\alpha v_s. \quad (13.17)$$

В тех случаях, когда рабочая длина волны и круговая частота модуляции известны заранее, следует одновременно опти-

Рис. 13.7. Зависимость граничной частоты f от длины волны λ для кремнивого и германиевого фотодиодов.

Кривая позволяет быстро определить материал или структуру фотодиода по известной рабочей частоте на заданной длине волны. (Пример: при частоте модуляции 12 ГГц на длине волны $\lambda = 0,8$ мкм следует взять германиевый фотодиод с $P-N$ -переходом)



мизировать и чувствительность и полосу пропускания. На рис. 13.7 приведена зависимость граничной частоты фотодиода от требуемой спектральной области.

§ 4. Лавинный фотодиод (ЛФД)

Когда световая мощность очень мала (~ 1 нВт), фототоки будут также весьма небольшими (~ 1 нА), и в этом случае для уменьшения влияния шума используют внутреннее усиление в фотоприемнике за счет эффекта лавинного умножения носителей заряда [220].

А. Лавинное умножение

В сильном электрическом поле энергия носителя заряда может превысить порог ионизации вещества и в результате его столкновения с атомом кристаллической решетки образуется

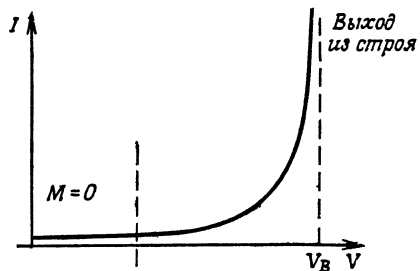


Рис. 13.8. Зависимость тока от напряжения в лавинном фотодиоде.

пара электрон — дырка. Если образовавшиеся носители тоже ускорятся до энергии, равной порогу ионизации, то они создадут другие носители и т. д., и в результате *проводимость нарастает за счет образования лавины носителей заряда* (рис. 13.8).

В идеальном процессе умножения каждый первичный электрон давал бы M вторичных электронов. Но процесс образова-

ния лавины — это вероятностный процесс, и величина M изменяется случайным образом. Флуктуации коэффициента умножения относительно среднего значения создают дополнительный шум, который весьма нежелателен.

Обозначим через α число пар электрон — дырка, создаваемых падающим носителем заряда на единице длины в направлении электрического поля, называемое скоростью ионизации α . Если толщина лавинного слоя равна e , то полное число носителей, создаваемых одним первичным носителем заряда, будет равно

$$N = \int_0^e \alpha dx. \quad (13.18)$$

Для простоты предположим, что скорость ионизации для электронов и дырок одинакова, что α не зависит от электрического поля и что рекомбинация носителей, которые достигли предельной скорости, пренебрежимо мала, так же как мало и влияние объемного заряда. Тогда можно показать [221], что коэффициент умножения M дается формулой

$$M = \frac{1}{1 - \int_0^e \alpha dx}. \quad (13.19)$$

При этом различают два режима проводимости.

1. *Режим проводимости с лавинным умножением.* В этом режиме число ионизаций ограничено и процесс не самоподдерживающийся. Если напряжение пробоя диода обозначить через V_B , то зависимость коэффициента умножения M от внешнего напряжения V может быть представлена в виде

$$1 - \frac{1}{M} = \left(\frac{V}{V_B} \right)^n,$$

где показатель степени n зависит от типа материала.

2. *Режим лавинного нарастания.* В этом случае коэффициент M растет до бесконечности и процесс самоподдерживающийся; нарастание ограничивается лишь объемным зарядом. На практике максимальный коэффициент умножения ограничивается преждевременным пробоем и выходом из строя определенных частей элемента. Следует также добавить, что коэффициент усиления зависит от лавинного тока. В самом деле, присутствующие в зоне лавины носители заряда уменьшают электрическое поле, оно уменьшается из-за падения напряжения на последовательном внутреннем сопротивлении и сопротивлении нагрузки. При сильном освещении ток пропорционален лишь

корню квадратному из первичного тока [222]. В действительности скорости ионизации для электронов α_n и дырок α_p неодинаковы и показано, что, если величина M меньше отношения α_n/α_p , полоса пропускания не уменьшается. Но при $M > \alpha_n/\alpha_p$ коэффициент умножения M изменяется при изменении частоты по закону

$$M = \frac{M_0}{\sqrt{1 + M_0^2 \omega^2 \tau^2}}.$$

Б. Фотодиод со структурой $P\pi PN$

Это частный случай лавинного фотодиода. В нем сочетаются преимущества лавинного и PIN -фотодиодов. К умножающему $P-N$ -переходу со стороны слоя P -типа добавляется слой с собственной проводимостью (рис. 13.9). При нормальных условиях

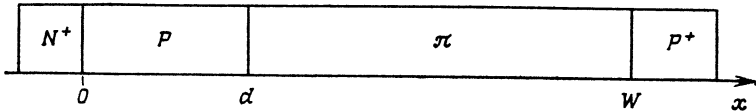


Рис. 13.9. Структура фотодиода типа $P\pi PN$.

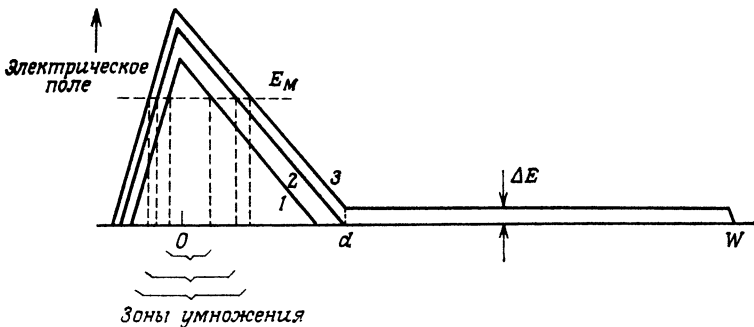


Рис. 13.10. Распределение электрического поля в структуре $P\pi PN$.

1 — при $V < V_p$; 2 — при $V = V_p$; 3 — при $V = V_p + \Delta E * (e - d)$. При напряжении $V > V_p$ увеличение напряжения V вызывает лишь слабое приращение электрического поля ΔE .

работы слой объемного заряда достигает слоя π раньше, чем наступает пробой $P-N$ -перехода. Поэтому большие изменения внешнего напряжения не приводят к большим изменениям поля в $P-N$ -переходе (рис. 13.10 и 13.11), а следовательно, и коэффициента умножения M . Таким образом, рабочий режим фотодиода становится более стабильным.

Но, чтобы не снизить полосы пропускания, нужно учитывать время прохождения слоя собственной проводимости. Если носители движутся со своей предельной скоростью, например в кремнии, то при толщине 100 мкм временная задержка имеет порядок 0,5 нс. Эта толщина должна быть сравнимой с глубиной проникновения света в полупроводник $L = 1/\alpha$. На длине волны 0,89 мкм величина L составляет 20 мкм. Следовательно,

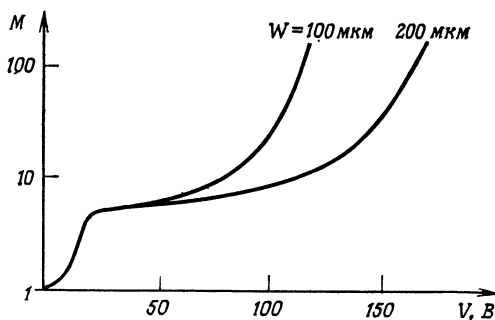


Рис. 13.11. Зависимость коэффициента умножения M от напряжения для структуры $P\pi PN$.

в фотодиодах со структурой $P\pi PN$ обеспечивается хороший компромисс между чувствительностью и быстродействием на этой длине волны. Зато вблизи 1,06 мкм величина L достигает ~ 100 мкм и такой компромисс невозможен.

§ 5. Фотоприемники из полупроводниковых соединений типа $AIII - BV$

Выше мы говорили о фотоприемниках, изготавливаемых на основе классических полупроводниковых материалов, таких, как кремний и германий [223]. Полупроводниковые соединения, образуемые элементами III и V группы периодической системы элементов (GaAs, InAs, InP, InSb, ...), имеют зонную структуру (гл. 12, § 4, п. А) и тоже пригодны для изготовления фотоприемников. На их основе в гетероструктурах с разной степенью легирования можно получить квантовый выход, близкий к 1, с заданной спектральной чувствительностью, поскольку ширина запрещенной зоны в этих соединениях зависит от концентрации примесей. Фотоприемники на этих соединениях работают при меньших напряжениях. Например, при одинаковом коэффициенте усиления, равном 50, для кремниевого лавинного фотодиода требуется напряжение питания, равное 150 В, а для лавинного фотодиода на основе соединений типа $AIII - BV$ всего 50 В. Правда, у таких фотоприемников значительно больше темновой ток (10 нА вместо 1 нА [224]).

§ 6. Шумы фотодиодов

А. Шумы в отсутствие лавинного усиления

В данном случае фототок пропорционален вызывающей его световой мощности. Такое соотношение выполняется в среднем, поскольку, как показывает более детальный анализ, фототок есть случайная величина, которая, как и всякая другая случайная величина, характеризуется различными моментами распределения вероятностей: средним значением, среднеквадратичным и т. д. Это и понятно, так как ток на выходе фотодиода равен сумме отдельных токов, соответствующих движению носителей заряда, возникающих в разные моменты времени.

Этот шум, который будет добавляться к шумам цепей усиления и обработки информации, искажает сигнал и так же, как потери в оптических волокнах, ограничивает дальность оптической связи. В гл. 15, § 3 мы подробно рассмотрим влияние шума на характеристики связи и покажем, какие меры нужно принимать в схемах, следующих за фотодиодом, чтобы уменьшить его влияние. Здесь же мы остановимся на микроскопических аспектах явления и рассмотрим процесс одновременной генерации сигнала и шума.

Обозначим через $p(t)$ мощность светового импульса, падающего на фотодиод, через $\{t_n\}$ — последовательность моментов времени, в которые рождается пара электрон — дырка, а через $u(t)$ — импульс напряжения на нагрузочном резисторе фотодиода, создаваемый парой носителей в момент $t = 0^1$). Полное напряжение на нагрузочном резисторе будет равно

$$v(t) = \sum_{n=1}^N u(t - t_n), \quad (13.20)$$

где N — полное число пар носителей заряда, генерируемых световым импульсом. Как показано в работах [217, 220], процесс генерации носителей, с которым связаны случайные переменные $\{t_n\}$ и N , описывается распределением Пуассона с параметром $\lambda(t)$, зависящим от времени. Таким образом, вероятность того, что в промежутке времени $(t, t + T)$ возникнет n пар носителей заряда, равна

$$\frac{\Lambda^n e^{-\Lambda}}{n!}, \quad (13.21)$$

$$\text{где } \Lambda = \frac{1}{T} \int_t^{t+T} \lambda(\tau) d\tau = \frac{1}{T} \int_t^{t+T} \frac{\eta p(\tau)}{h\nu} d\tau.$$

¹⁾ Величина $u(t)$ есть импульсный отклик канала фоторегистрации.

В таком случае среднее напряжение на выходе равно [217]

$$\langle v(t) \rangle = \frac{\eta}{h\nu} \int u(t - \tau) p(\tau) d\tau, \quad (13.22)$$

а его дисперсия такова:

$$\langle [v(t) - \langle v(t) \rangle]^2 \rangle = \frac{\eta}{h\nu} \int u^2(t - \tau) p(\tau) d\tau + \sigma_i^2. \quad (13.23)$$

Добавленное здесь слагаемое σ_i^2 учитывает шумы, главным образом тепловые, которые вносят электронные схемы, включенные на выходе фотодиода.

Положив $p(t) = \text{const} = p_0$, можно написать выражение для отношения сигнала к шуму:

$$\left(\frac{S}{B}\right)^2 = \frac{\langle v(t) \rangle^2}{\langle [v(t) - \langle v(t) \rangle]^2 \rangle} = \frac{\left(\frac{\eta p_0}{h\nu} \int u(\tau) d\tau\right)^2}{\frac{\eta p_0}{h\nu} \int u^2(\tau) d\tau + \sigma_i^2}. \quad (13.24)$$

Допустим также, что $u(t)$ — импульсный отклик идеального фильтра низких частот с полосой пропускания Δf ; тогда окончательно получим

$$\left(\frac{S}{B}\right)^2 = \frac{\left(\frac{\eta q p_0}{h\nu}\right)^2}{\frac{2\eta q^2 p_0}{h\nu} \Delta f + \sigma_i^2}. \quad (13.25)$$

Следовательно, существует такое значение световой мощности p_0 , при котором шум равен сигналу, т. е. $S/B = 1$ (или 0 дБ). Такая мощность называется *эквивалентной мощностью шума*. Чем меньше эквивалентная мощность шума, тем меньше оптическая мощность на входе приемника, необходимая для обеспечения заданного отношения сигнала к шуму. Обычно тепловой шум пропорционален полосе пропускания Δf , и поэтому эквивалентная мощность шума измеряется в единицах Вт·Гц^{-1/2} [220, 222, 225].

Б. Шумы при наличии лавинного усиления

Полученные выражения показывают ту важную роль, которую играют собственные шумы фотодиода в уменьшении полного отношения сигнала к шуму. Для уменьшения этого влияния можно использовать лавинный фотодиод с внутренним коэффициентом усиления M . Коэффициент усиления — случайная величина, распределение вероятности которой зависит от типа носителя заряда, вызывающего ионизацию. Обозначим через M_n значение коэффициента усиления в момент времени t_n , когда

рождается первая пара электрон — дырка. С учетом (13.20) полное напряжение на нагрузочном резисторе будет равно

$$v(t) = \sum_{n=1}^N M_n u(t - t_n). \quad (13.26)$$

Не вдаваясь в детали довольно сложного расчета [221, 222, 225], напомним сразу формулу для среднего квадрата:

$$\langle M_n^2 \rangle = F(M) M^2, \quad (13.27)$$

где M — среднее значение величины M_n , а $F(M)$ — коэффициент шума, характеризующий отклонения от постоянного коэффициента усиления, равного M . При данных условиях выражение (13.25) принимает вид

$$\left(\frac{S}{B}\right)^2 = \frac{nq \frac{M}{h\nu} p_0}{2\eta q^2 \frac{M^2 F(M) p_0 \Delta f}{h\nu} + \sigma_t^2}. \quad (13.28)$$

Аналогичным образом можно определить эквивалентную мощность шума для системы фотодиод — нагрузка. Отметим, что она зависит от коэффициента усиления M . Если предположить (гл. 15, § 3, п. В), что $F(M) = M^x$, то легко показать, что существует оптимальное значение M , при котором эквивалентная мощность шума минимальна [225].

МОДУЛЯЦИЯ СВЕТА, ВВЕДЕНИЕ В ИНТЕГРАЛЬНУЮ ОПТИКУ

В предыдущей главе мы рассмотрели работу фотоприемников. Однако для осуществления связи нужно еще уметь модулировать свет передаваемым сигналом.

Что же понимают под *модуляцией света*?

Модуляция — это изменение параметров света (главным образом его амплитуды или фазы, но можно также говорить об изменении поляризации, направления распространения, частоты, распределения мод и т. д.) в зависимости от управляющего (модулирующего) сигнала. Разумеется, что для нас модулирующий сигнал будет электрическим (ток, напряжение), но он может быть также акустическим, механическим и даже оптическим.

Существуют разные способы получить модулированный свет. Первый из них — использовать модулируемый источник света. Это так называемая *прямая модуляция*, примером которой может служить модуляция излучения полупроводникового лазера путем изменения тока накачки. Второй способ — модуляция излучения немодулируемого источника света. Это *внешняя модуляция*. И наконец, если есть соответствующий модулятор, то его можно ввести в лазерный резонатор и осуществить таким образом *внутреннюю модуляцию*. «Прямые» способы модуляции и «внутренняя» модуляция (являющаяся по существу разновидностью «прямой») уже рассматривались в гл. 12, § 3 и 4 (полупроводниковый лазер и твердотельный лазер в импульсном режиме). Поэтому в данной главе мы рассмотрим различные методы внешней модуляции.

Для модуляции необходимо, чтобы управляющий сигнал воздействовал на свет (рис. 14.1). Стало быть, модулятор — это некая система, в которой происходит взаимодействие света с веществом. Среди большого разнообразия различных по своей природе физических явлений — механических (или упругих), акустических, магнитных, электрических или оптических — используется ряд эффектов, которые можно сгруппировать так, как показано на рис. 14.2.

Процессы, происходящие при модуляции света, нередко весьма сложны. При выборе материала для модулятора часто предпочтение отдается кристаллам из-за их хороших механических, электрических и оптических свойств. Разумеется, мы не

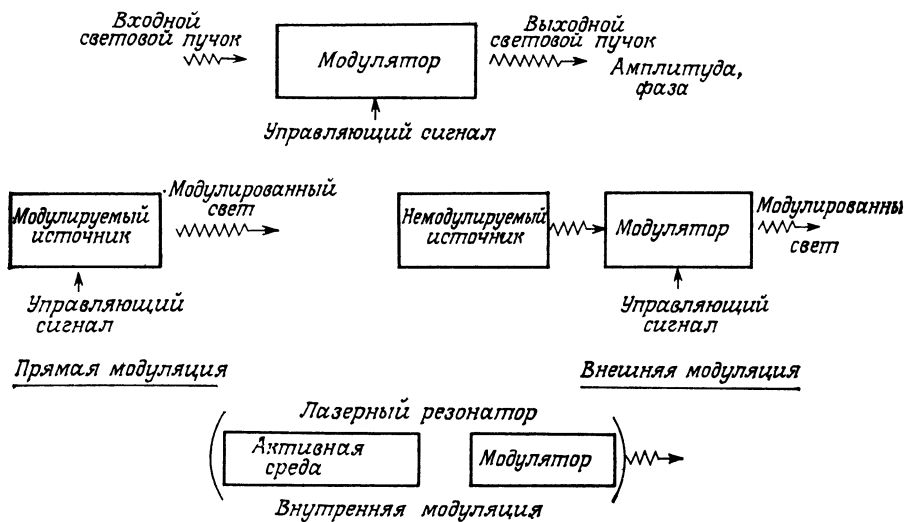


Рис. 14.1. Три вида модуляции света.

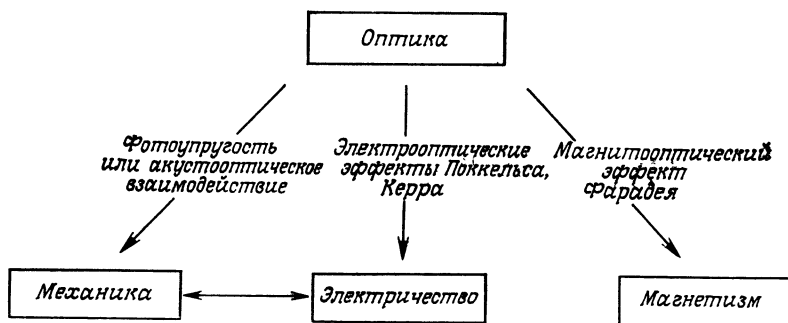


Рис. 14.2. Физические эффекты, используемые для модуляции света.

станем здесь рассматривать все свойства и особенности оптики кристаллов, которые читатель может изучить, обратившись к более фундаментальным трудам [226]. Наша цель состоит лишь в описании нескольких типов существующих модуляторов света, чтобы познакомить читателя с их характеристиками, особенно подчеркивая их недостатки. Вот почему мы очень кратко напомним физические свойства используемых в модуляторах материалов, предполагая, что читателю они известны. Мы ограничимся рассмотрением только электрооптических и акустооптических систем модуляции света, которые получили наибольшее распространение.

§ 1. Электрооптические модуляторы света

Модуляторы такого типа основаны на явлении двулучепреломления в анизотропных средах [227].

В таких средах напряженность электрического поля и электрическая индукция связаны между собой тензорным соотношением

$$\mathbf{D} = \hat{\epsilon} \mathbf{E},$$

где $\hat{\epsilon}$ — тензор диэлектрической проницаемости.

В дальнейшем нам удобнее будет иметь дело с тензором \hat{a} , который входит в обратное соотношение между \mathbf{E} и \mathbf{D} [227]:

$$\mathbf{E} = \hat{a} \mathbf{D}.$$

В большинстве кристаллов (называемых оптически неактивными) эти тензоры симметричны относительно главной диагонали, и поэтому достаточно оставить только 6 коэффициентов вместо 9. Кроме того, изменением осей координат легко диагонализировать соответствующую матрицу. Следовательно, в дальнейшем будем считать, что

$$(\mathbf{E}) = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix} (\mathbf{D}).$$

В кристаллах, обнаруживающих эффект двулучепреломления, в заданном направлении \mathbf{k} могут *распространяться две плоскополяризованные световые волны* [7]. Направления поляризации (совпадающие с векторами \mathbf{D} электрической индукции) и соответствующие показатели преломления, а следовательно, и фазовая скорость (гл. 2, § 1) определяются следующим образом:

а) строят эллипсоид показателя преломления, удовлетворяющий уравнению

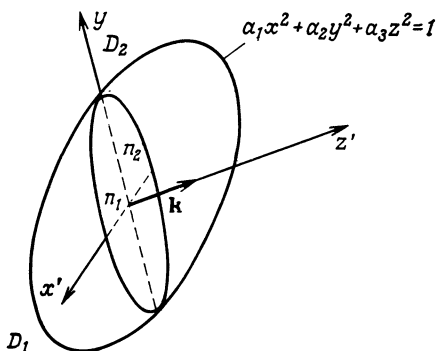
$$a_1 x^2 + a_2 y^2 + a_3 z^2 = 1;$$

б) рассматривают эллипс, который получается при пересечении этого эллипсоида с плоскостью, перпендикулярной вектору распространения световой волны \mathbf{k} .

Искомые направления поляризации (векторы \mathbf{D}_1 и \mathbf{D}_2) соответствуют главным осям полученного эллипса. Они называются *нейтральными осями* для направления распространения \mathbf{k} . Длины этих осей служат мерой показателей преломления для рассматриваемого направления распространения световой волны (рис. 14.3).

Вообще говоря, существуют два особых направления распространения \mathbf{k} , для которых эллипс превращается в окружность. Эти направления совпадают с *оптическими осями* кристалла. Следовательно, двулучепреломление не имеет места при распространении света вдоль оптической оси кристалла. Очевидно, что если эллипсоид представляет собой тело вращения, то кристалл одноосный, причем сфера характеризует изотропный материал.

Рис. 14.3. Эллипсоид показателя преломления и нейтральные оси, соответствующие заданному направлению распространения световой волны.



Рассмотрим линейно-поляризованную волну, направление поляризации которой \mathbf{D} образует угол α с \mathbf{D}_1 . Напомним, что направление поляризации плоской волны определяется направлением вектора электрической индукции. На выходе из среды две составляющие вектора поляризации в направлениях векторов \mathbf{D}_1 и \mathbf{D}_2 приобретут разный фазовый сдвиг. Если обозначить через l длину пробега волны внутри материала, то будем иметь

$$\left. \begin{aligned} \varphi_1 &= \frac{2\pi}{\lambda_0} n_1 l \\ \varphi_2 &= \frac{2\pi}{\lambda_0} n_2 l \end{aligned} \right\}, \text{ т. е. } \delta\varphi = \frac{2\pi}{\lambda_0} (n_1 - n_2) l. \quad (14.1)$$

Следовательно, на выходе обычно получают эллиптически поляризованную световую волну. При наличии анализатора, образующего угол θ с \mathbf{D}_1 , можно преобразовать такое изменение поляризации в изменение интенсивности за счет интерференции между составляющими поляризованной волны.

Интенсивность на выходе анализатора определяется следующим образом (рис. 14.4):

$$I = |OH' + OK'|^2 = |OH \cos \theta + OK \sin \theta|^2, \\ \text{где } OH = OM \cos \alpha e^{-i\varphi_1}, \quad OK = OM \sin \alpha e^{-i\varphi_2}.$$

После тригонометрических преобразований получим

$$I = I_0 \left[\cos^2(\theta - \alpha) - \sin 2\theta \sin 2\alpha \sin^2 \frac{\delta\varphi}{2} \right].$$

Представляет интерес простой частный случай, когда анализатор и поляризатор скрещены ($\theta - \alpha = \pi/2$). В этом случае интенсивность на выходе анализатора будет равна

$$I = I_0 \sin^2 2\alpha \sin^2 \frac{\delta\varphi}{2}.$$

Заметим, что пропускание кристалла будет максимальным при $\alpha = \pm\pi/4$. Это самая распространенная схема: кристалл

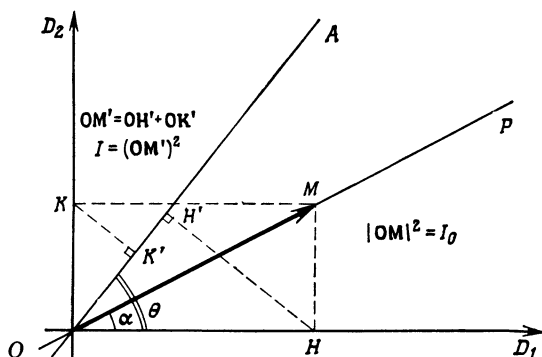


Рис. 14.4. Вычисление интенсивности на выходе системы поляризатор — анализатор, между которыми помещен электрооптический кристалл. P — поляризатор; A — анализатор; D_1 , D_2 — нейтральные оси кристалла, соответствующие рассматриваемому направлению распространения световой волны.

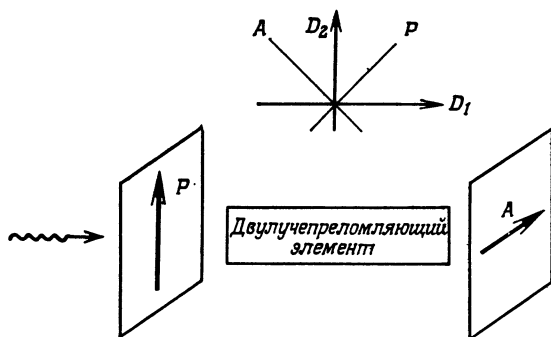


Рис. 14.5. Модуляция света по интенсивности при помещении кристалла между скрещенными поляризатором и анализатором.

помещают между скрещенными поляризатором и анализатором, расположенными по биссектрисам углов между нейтральными осями (рис. 14.5). Чтобы модулировать свет, изменяют разность фаз $\delta\varphi$ посредством внешнего электрического поля (в электри-

ческом поле некоторые материалы становятся двулучепреломляющими; к ним применимо все сказанное выше о двулучепреломлении). В электрооптическом материале характеристики среды (эллипсоид показателя преломления) зависят от напряженности внешнего поля в соответствии с тензорными соотношениями, описывающими изменения $a'_i - a_i$ в функции от поля. Известны два следующих классических электрооптических эффекта.

1. Линейный электрооптический эффект, или эффект Поккельса:

$$a'_i = a_i + r_{ij}E_j, \quad i = 1, \dots, 6, \quad j = 1, \dots, 3,$$

где r_{ij} — элементы электрооптического тензора \hat{r} .

2. Квадратичный электрооптический эффект, или эффект Керра:

$$a'_i = a_i + K_{ijk}E_jE_k,$$

где K_{ijk} — элементы тензора 3 ранга \hat{K} .

В общем случае ориентация нейтральных осей зависит от электрического поля. Приведенные выше тензорные соотношения позволяют определить *разность разностей фаз на выходе*

$$\Delta\varphi = \delta\varphi(\mathbf{E}) - \delta\varphi(0).$$

В дальнейшем будем полагать, что положение нейтральных осей остается неизменным. Тогда легко показать, что в случае эффекта Поккельса включение электрического поля приводит к следующим изменениям показателей преломления вдоль нейтральных осей:

$$\Delta n^{(1)} = K^{(1)}E, \quad \Delta n^{(2)} = K^{(2)}E,$$

где $K^{(1)}$ и $K^{(2)}$ — постоянные, зависящие от электрооптического тензора и от того, как поле ориентировано относительно осей материала. В этом случае относительный фазовый сдвиг будет следующим образом зависеть от внешнего электрического поля:

$$\Delta\varphi = \frac{2\pi}{\lambda_0} (K^{(1)} - K^{(2)})Eh,$$

где h — расстояние между электродами.

Важным параметром электрооптического модулятора является *полуволновое напряжение* V_0 , под которым понимают напряжение, обеспечивающее фазовый сдвиг, равный π :

$$V_0 = \frac{\lambda_0}{2|K^{(1)} - K^{(2)}|}. \quad (14.2)$$

На практике различают два вида электрооптических модуляторов света.

1. *Продольный модулятор*, в котором при помощи прозрачных электродов создается электрическое поле в направлении распространения света. В этом случае $h = l$, где l — протяженность материала в направлении распространения света. Зависимость вносимой разности фаз $\Delta\varphi$ от напряжения модуляции имеет вид

$$\Delta\varphi = \pi \frac{V}{V_0}.$$

2. *Поперечный модулятор*, в котором при помощи двух электродов, размещенных на разных сторонах кристалла на расстоянии h друг от друга, создается поле, поперечное относительно направления распространения света. Обычно $h < l$. На этот раз фазовый сдвиг определяется выражением

$$\Delta\varphi = \pi \frac{V}{V_0} \frac{l}{h}. \quad (14.3)$$

Может показаться, что в данном случае обеспечивается выигрыш в глубине модуляции по сравнению с продольным модулятором в l/h раз. Но следует учитывать, что хотя выражение (14.2) для V_0 внешне одинаково в обоих случаях, в нем неодинаковы коэффициенты $K^{(1)}$ и $K^{(2)}$, которые зависят от ориентации поля в кристалле.

Другой важный параметр модулятора — *требуемая управляющая мощность*. Представим модулятор в виде конденсатора емкостью C , включенного в колебательный контур, ширина Δf полосы пропускания которого удовлетворяет соотношению

$$\pi RC \Delta f = 1.$$

Тогда требуемая управляющая мощность будет равна

$$P = \frac{V^2}{2R} = \frac{\pi}{2} V^2 C \Delta f.$$

Если подставить сюда теоретическое выражение для емкости C и учесть найденное выше выражение для фазового сдвига $\Delta\varphi$, то получим следующее теоретическое выражение для требуемой управляющей мощности:

$$P = \frac{\pi}{2} \epsilon_0 \epsilon_r V_0^2 \Delta f \frac{h^2}{l} \left(\frac{\Delta\varphi}{\pi} \right)^2. \quad (14.4)$$

Таким образом, она пропорциональна ширине полосы пропускания и квадрату глубины модуляции $\Delta\varphi$. Это справедливо для всех типов модуляторов: электрооптических, акустооптических и магнитооптических. Поэтому можно ввести следующий критерий качества модулятора [228]:

$$\eta = \frac{P}{\Delta f (\Delta\varphi)^2}. \quad (14.5)$$

Из выражения (14.4) следует, что в модуляторах целесообразно использовать длинные и узкие кристаллы. Согласно формуле (14.3), выгодно также увеличивать длину поперечных модуляторов. Но нужно учитывать, что тогда увеличивается поглощение света в кристалле и к тому же технологически трудно получить кристаллы хорошего качества и большого размера. Не так проста и задача нанесения прозрачных электродов и пр. и пр. Поэтому порой подобного рода соображения и создают действительные ограничения вплоть до настоящего времени.

В случае *эффекта Керра* (квадратичного эффекта) на модулирующее поле необходимо накладывать постоянное поляризующее поле:

$$\Delta n = KE^2(t) \approx KE_0^2 + 2KE_0E(t).$$

При этом чем сильнее будет поляризующее поле, тем «легче» будет модуляция. Эффект Керра в некоторых жидкостях используется для модуляции добротности в твердотельных лазерах (гл. 12, § 3, п. Б). В этом случае имеет место не естественное двулучепреломление, а наведенное электрическим полем. Основной недостаток наведенного двулучепреломления состоит в том, что необходимо высокое напряжение (киловольты). Для снижения напряжения можно охладить кристалл до температуры, близкой к точке Кюри. В результате получают модуляторные ячейки, работающие при напряжении менее 100 В. Однако при этом необходима стабилизация температуры кристалла.

§ 2. Акустооптические модуляторы [229]

В фотоупругих материалах изменение эллипсоида показателя преломления вызывает деформация:

$$a'_i = a_i + p_{ik}S_k, \quad (14.6)$$

где p_{ik} — элементы тензора фотоупругости \hat{p} .

Требуемую деформацию можно получить, возбуждая в среде акустические волны при помощи пьезоэлектрического преобразователя. При этом в материале формируется дифракционная решетка, изменения показателя преломления в которой в каждой точке и в каждый момент времени, согласно формуле (14.6), будут пропорциональны акустической волне. В зависимости от условий согласования акустического импеданса на концах образца может быть создан либо режим бегущих волн, либо режим стоячих волн. Заметим, что поскольку скорость звука в среде $v \ll c$, то и в режиме бегущих волн свет «не будет замечать» перемещения акустических волн. Поэтому в первом

приближении можно рассматривать режим стоячих волн. В обоих режимах акустическая мощность изменяется с периодом $\Lambda/2$, где Λ — длина акустической волны в материале.

Итак, направим на образец, в котором возбуждена акустическая волна, плоскую световую волну с волновым вектором \mathbf{k}_1 (длина волны λ_1 , частота — ω_1), как это показано на рис. 14.6. Акустическая волна описывается соответствующим вектором $K(\Lambda/2, \Omega)$, а выходящая световая волна — вектором $\mathbf{k}_2(\lambda_2, \omega_2)$.

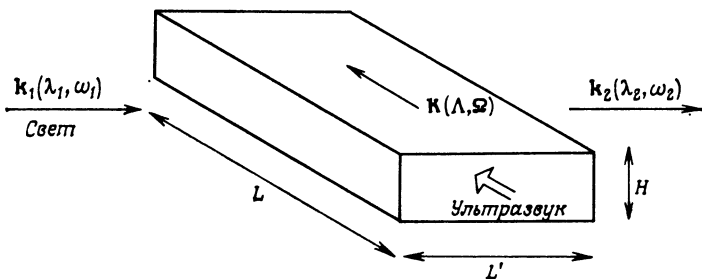


Рис. 14.6. Акустооптическая система.

На основании закона сохранения импульса $\hbar k$ и энергии $\hbar \omega$ можем написать

$$k_2 = k_1 + K, \quad (14.7)$$

$$\omega_2 = \omega_1 + \Omega. \quad (14.8)$$

Поскольку $\Omega \ll \omega_1$, согласно формуле (14.8), имеем

$$\omega_2 \approx \omega_1,$$

$$|k_2| \approx |k_1| \gg |K|.$$

Условия (14.7) и (14.8) говорят нам, как сделать акустический модулятор. Световая волна должна быть почти перпендикулярна возбуждаемой в образце акустической волне. Она падает под очень малым углом θ_i , и угол ее выхода из образца θ_e определяется диаграммой, представленной на рис. 14.7. Так как треугольник OA_1A_2 практически равнобедренный, из него мы находим уже известное нам условие Брэгга

$$\sin \theta_e = \frac{HA_2}{2|k_2|} \approx \frac{K}{2|k_1|} = \frac{\lambda_0}{n\Lambda}. \quad (14.9)$$

Угол отклонения выходящего пучка зависит от длины волны Λ или (что то же самое) от акустической частоты f :

$$\sin \theta_e = \frac{\lambda_0 f}{nv}. \quad (14.10)$$

Следовательно, нужно взять генератор с изменяющейся частотой и формировать внутри материала дифракционную ре-

шетку с изменяющимся периодом. Это позволит управлять углом отклонения пучка на выходе. В результате мы получим *акустооптический дефлектор*.

Существенным недостатком акустооптических дефлекторов является ограниченное число разрешимых направлений откло-

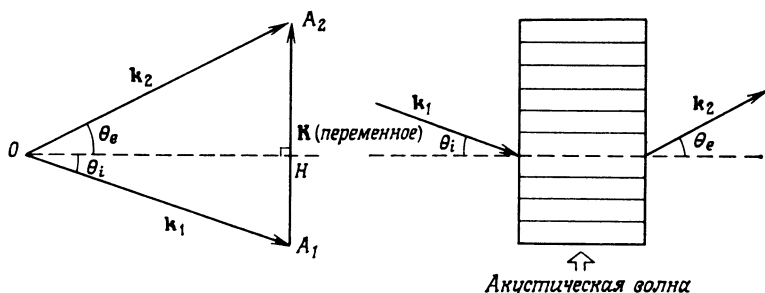


Рис. 14.7. Взаимодействие света с акустической волной, возбужденной в фотоупругом материале.

Показано положение волновых векторов k_1 и k_2 относительно акустической волны.

нения. Выше мы неявно предполагали, что взаимодействие световой и акустической волн происходит в среде неограниченных размеров. Однако пространственное ограничение среды взаимо-

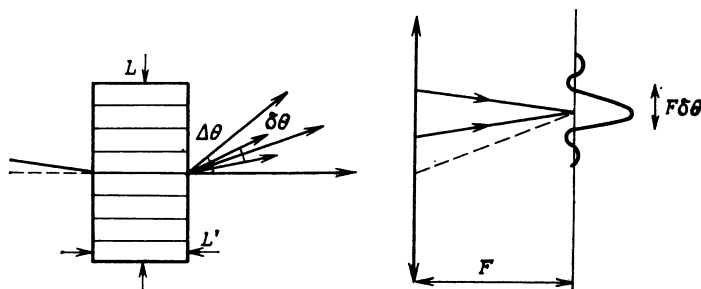


Рис. 14.8. Расчет максимального числа разрешаемых направлений отклонения на выходе акустооптического дефлектора.

действия вызывает дифракцию светового пучка. Поэтому, если через L обозначить ширину ультразвукового пучка (рис. 14.8), то угловая расходимость светового пучка на выходе дефлектора будет равна (гл. 4, § 1, п. В)

$$\delta\theta \sim \frac{\lambda}{L}.$$

Эту расходимость пучка можно обнаружить, например, наблюдая дифракционное пятно в фокусе линзы (рис. 14.8).

Но, как это следует из закона отклонения (14.10), полная угловая ширина $\Delta\theta$ связана с акустической шириной полосы пропускания соотношением

$$\Delta\theta = \frac{\lambda \Delta f}{v}.$$

Отсюда находим число разрешимых направлений отклонения:

$$N = \frac{\Delta\theta}{\delta\theta} = \frac{L}{v} \Delta f.$$

В это соотношение входит постоянная времени системы (время установления акустического режима во всем объеме материала)

$$\tau = \frac{L}{v}.$$

Окончательно получаем

$$N = \tau \Delta f. \quad (14.11)$$

Число разрешимых направлений отклонения в акустооптическом дефлекторе равно произведению постоянной времени на имеющуюся ширину полосы пропускания.

В действительности в определении величины $\delta\theta$ нужно указывать критерий различимости пучков на выходе. Чаще всего пользуются критерием Эри, который предусматривает, чтобы первый минимум каждого дифракционного пятна совпадал с максимумом интенсивности в соседнем дифракционном пятне. В зависимости от вида используемого критерия в выражении (14.11) появится тот или иной постоянный множитель, но это не меняет нашего вывода.

Выражение (14.11) хорошо известно в электронике. Оно дает максимальное число степеней свободы сигнала. В рассматриваемом нами случае это условие оказывается весьма ограничивающим. В самом деле, при увеличении τ снижается быстродействие дефлектора. Длина акустооптического взаимодействия L ограничивается поглощением акустических волн в материале. При увеличении Δf возникают другие трудности. Пьезоэлектрический преобразователь обычно обладает частотной характеристикой с ярко выраженным острым резонансным пиком. Это создает проблемы для электроники (в основном трудности связаны с согласованием импедансов на концах образца). В конечном счете частотная характеристика системы оказывается неравномерной, и иногда приходится вводить компенсацию затухания для крайних направлений. И наконец, можно было бы увеличить центральную частоту, но тогда возрастают акустические потери в материале.

Существенной характеристикой акустооптического дефлектора является также его *эффективность*, связанная с поглощением в материале, а также с толщиной используемого образца L' . В самом деле, при недостаточной толщине образца появляются многие порядки дифракции, а это, естественно, снижает эффективность дефлектора.

В зависимости от ширины L' ультразвукового пучка, на котором дифрагирует световая волна, обычно различают два предельных режима работы дефлектора. В первом режиме формируемая в нем дифракционная решетка может рассматриваться

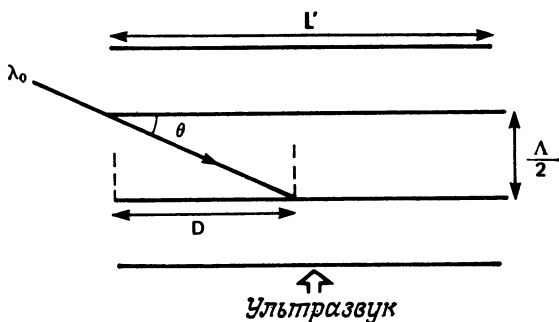


Рис. 14.9. Определение элементарного расстояния взаимодействия в случае толстой дифракционной решетки.

как тонкая, а во втором — как толстая (гл. 5). Можно ввести понятие элементарного расстояния акустооптического взаимодействия при малом угле θ (рис. 14.9):

$$D \approx \frac{\Lambda/2}{\theta} = \frac{n\Lambda^2}{\lambda_0}.$$

Соотношением величин L' и D можно характеризовать толщину формируемой решетки. При $L' \ll D$ мы имеем тонкую решетку. В противном случае мы имеем дело с толстой решеткой, где «лучи» могут «отражаться» на многих последовательных «слоях». Обычно для классификации режимов дифракции света на ультразвуке вводят параметр [230]

$$Q = 2\pi \frac{L'\lambda_0}{n\left(\frac{\Lambda}{2}\right)^2}.$$

В зависимости от его величины различают два режима дифракции:

1. *Режим дифракции Брэгга* ($Q > 10$).

В этом случае применимо условие Брэгга (14.9). Очень важное значение имеет угол падения пучка света. Можно показать,

что интенсивность модулированного света синусоидально зависит от глубины фазовой модуляции $\Delta\varphi$:

$$I = I_0 \sin^2 \left(\frac{\Delta\varphi}{2} \right).$$

В этом режиме теоретически достижима дифракционная эффективность, равная 100%.

2. Режим дифракции Рамана — Ната ($Q < 1$).

В этом случае имеют дело с тонкой решеткой и, следовательно, применима классическая формула для дифракционной решетки. В результате дифракции на решетке формируется много дифракционных порядков $0, \pm 1, \dots, \pm m$, причем условия освещения решетки не очень существенны:

$$\theta_m \approx \sin \theta_m = \frac{m\lambda_0}{n\Lambda},$$

где m — целое число.

Наконец, можно показать, что энергия, приходящаяся на m -й порядок дифракции, определяется выражением

$$I_m = I_0 J_m^2 \left(\frac{\Delta\theta}{2} \right),$$

где J_m — функция Бесселя первого рода порядка m . Как отмечалось выше, в данном случае 100%-ная модуляция невозможна.

При промежуточных значениях Q расчеты не так просты. В заключение подчеркнем, что во всех случаях паразитные дифракционные порядки снижают эффективность и создают затруднения при оптимизации величины N . Чтобы избежать перекрытия соседних порядков дифракции, приходится уменьшать интервал отклонения частоты Δf .

В настоящее время промышленностью выпускаются акустооптические дефлекторы со следующими характеристиками: число разрешаемых направлений — от 512 до 1024, эффективность (КПД) — от 50 до 80 %, постоянная времени — порядка 1 мкс.

Что касается модуляции света, осуществляемой методами акустооптики, то ее можно осуществлять двумя способами.

1. Путем блокирования одного из пучков на выходе дефлектора с $N \geq 2$. В зависимости от того, какая из двух частот возбуждается в дефлекторе, свет либо пройдет, либо нет. Следовательно, в данном случае мы получаем акустооптический прерыватель. В таком модуляторе поперечный размер L будет уменьшен. Благодаря этому можно получить весьма малые постоянные времени (доли наносекунды) при хорошем коэффициенте гашения.

2. Путем модуляции пучка 1-го порядка дифракции на выходе толстой решетки за счет изменения акустической мощности, подводимой к модулятору.

Что касается требуемой мощности управления модуляторами, то в настоящее время в этом отношении более выгодны, по-видимому, электрооптические, нежели акустооптические системы. Но техника развивается очень быстро и не позволяет очень уверенно делать долгосрочные прогнозы.

§ 3. Интегральная оптика, возможности и трудности

Вернемся к формулам (14.4) и (14.5). Параметр качества электрооптического модулятора имеет следующий вид:

$$\eta = \frac{1}{2\pi} \epsilon_0 \epsilon_r V_0^2 \frac{h^2}{l}.$$

Геометрические параметры модулятора входят только в виде отношения h^2/l . Выше мы уже отмечали, что в модуляторах выгодно использовать длинные образцы. Но величины h и l не независимы друг от друга. Из-за дифракции уменьшение h вызывает расширение пучка, что приводит к соответствующему уменьшению длины взаимодействия.

В лучшем случае, т. е. при использовании гауссова пучка (гл. 4, § 3), это отношение таково:

$$\frac{h^2}{l} \sim \lambda.$$

Отсюда вывод: если обеспечить направленное распространение света в среде взаимодействия, то вопрос о расходимости пучка отпадет; за счет направленного распространения света теоретически можно получить сколь угодно малый геометрический множитель.

Но можно заставить свет распространяться и можно связывать световые волны внутри устройства крайне малых размеров. Для этого применяется техника и технология тонких пленок, уже хорошо разработанная в электронике. При этом можно надеяться реализовать большинство операций по обработке оптической информации с помощью функциональных элементов в интегрально-оптическом исполнении.

Такая техника и технология получила название *интегральной оптики*. Мы не будем здесь рассматривать все разделы интегральной оптики, которая использует сложные физические явления и которая еще находится в состоянии быстрого развития. Мы ограничимся в основном интегрально-оптическими модуляторами.

А. Электрооптический фазовый модулятор

Этот вид модулятора был предложен Каминовом [228] на основе структуры полоскового световода (рис. 14.10) и реализован методом локальной диффузии титана в подложку из LiNbO_3 .

Необходимая мощность управления составляет 20 мкВт/МГц при напряжении 1,2 В. Полоса пропускания модулятора более 500 МГц. Однако используемые в модуляторе поляризатор и анализатор должны располагаться под углом 45° по отноше-

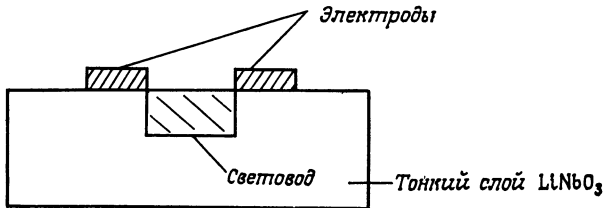


Рис. 14.10. Полосковый световод.

нию к плоскости структуры, что затрудняет их реализацию в интегрально-оптическом исполнении.

Б. Электрооптический амплитудный модулятор

Такой модулятор был создан совсем недавно [231], и в нем электрооптический эффект используется для прямого воздействия на профиль показателя преломления. Это прямо приводит к осуществлению *амплитудной модуляции* в неограниченном световоде. Получена глубина модуляции 13 дБ в полосе частот до 300 МГц (при напряжении питания $V = \pm 7,5$ В).

В. Устройства ввода-вывода излучения и устройства связи

Аналогичным образом можно осуществить связь между двумя полосковыми световодами, имплантированными на одной подложке. Если световоды расположены достаточно близко друг к другу (на расстоянии нескольких длин волн), то наблюдается взаимодействие между распространяющимися в них волнами (рис. 14.11) за счет затухающих волн, присутствующих в промежутке между световодами (гл. 3).

Такая система характеризуется коэффициентом связи κ , который зависит от профиля показателя преломления световодов, а также от степени перекрытия мод. Если в световод 1 ввести свет, то на определенном расстоянии от его торца часть световой волны может быть передана в световод 2. Очевидно, что эффективность связи может достигнуть 100 % только при условии согласования по фазе, т. е. при

$$\Delta\beta = \beta_2 - \beta_1 = 0 \quad (14.12)$$

(это все равно, что передавать что-нибудь из одного поезда в другой, когда они оба находятся в движении).

Если условие (14.12) выполняется, то наблюдается переход энергии из одного световода в другой и обратно с пространственным периодом, равным длине связи $1/\kappa$. В более общем

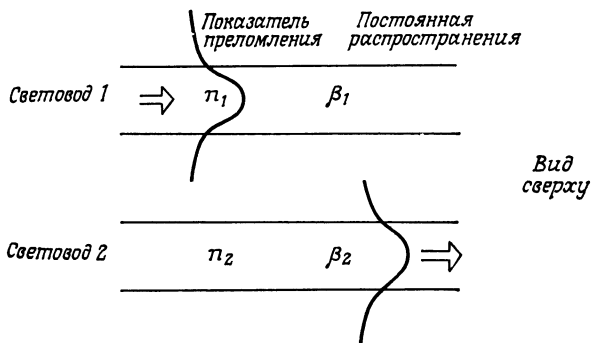


Рис. 14.11. Взаимодействие волн, распространяющихся в близко расположенных световодах.

случае длина связи \mathcal{L} зависит от коэффициента связи κ и от разности фаз $\Delta\beta$:

$$\mathcal{L} = \frac{2\pi}{\kappa \sqrt{1 + \left(\frac{\Delta\beta}{4\kappa}\right)^2}}.$$

Следовательно, в интегральной оптике можно управлять коэффициентом связи между световодами. Это явление может

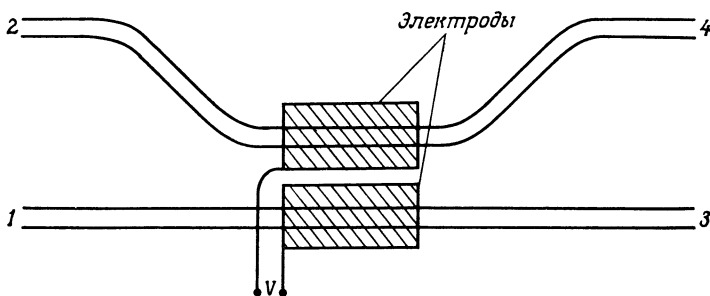


Рис. 14.12. Управляемый направленный ответвитель.

При включении напряжения V изменяется коэффициент связи, благодаря чему можно коммутировать выходы 3 и 4.

послужить основой для создания интересных схем: *направленных ответвителей, схем мультиплексирования* и т. д. Например, за счет электрооптического эффекта можно изменять и коэффициент связи κ и длину связи. Тем самым можно осуществлять изменяемую коммутацию (управляемый направленный ответвитель). Создан такой переключатель, который работает при напряжении 6 В [232] (рис. 14.12).

Наконец, можно реализовать схемы, состоящие из нескольких ответвителей-переключателей и модуляторов, выполненных на одной тонкой пленке. Так, например, выполнен *модулятор, основанный на принципе интерференции* [233]. Речь идет о настоящем перестраиваемом интерферометре (рис. 14.13), в ко-

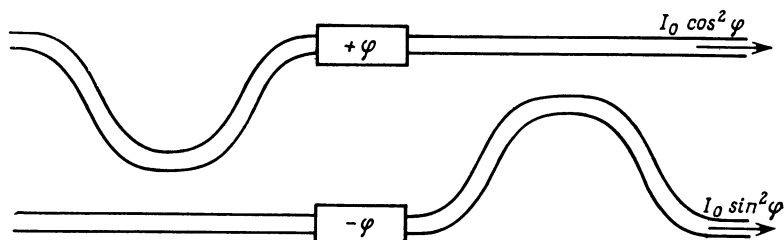


Рис. 14.13. Модуляция света за счет интерференции.

Модулятор состоит из двух устройств ввода с коэффициентами связи 50% и двух управляемых электрооптических фазовращателей.

тором два устройства ввода играют роль делителя пучка. Между двумя плечами интерферометра можно создавать управляемый фазовый сдвиг. В этой схеме была получена 70 %-ная модуляция при рабочем напряжении 20 В. В гл. 17, § 4 будут приведены устройства связи других типов.

Г. Акустооптические устройства

Что касается акустооптических устройств в интегральном исполнении, то в настоящее время реализованы только дефлекторы и модуляторы света [234]. Они представляют собой подложку из пьезоэлектрического материала (SiO_2 , LiTaO_3 , LiNbO_3 ...), на которой для возбуждения в ней поверхностных акустических волн нанесены преобразователи в виде гребенок, а сверху осаждены световодные слои из аморфного материала, обладающего большим коэффициентом фотоупругости (As_2S_3 , Ta_2O_3 , GeS).

Обычно для обеспечения больших коэффициентов модуляции эти устройства используют в режиме брэгговской дифракции. Требуемая акустическая мощность имеет порядок нескольких милливатт на мегагерц при напряжении в несколько вольт. Число разрешаемых направлений акустооптических дефлекторов в интегральном исполнении еще не велико.

Мы не будем больше рассматривать здесь интегральную оптику. В связи с вопросом о модуляции света было показано, что такая техника открывает очень интересные перспективы. У нее большие преимущества, связанные с миниатюризацией и интегральным исполнением. Мощность и управляющие напряжения существенно уменьшены, что приводит к меньшему

потреблению энергии. Кроме того, можно предвидеть реализацию очень широких полос пропускания благодаря возможности использования одномодового режима распространения света. Наконец, диапазон возможных операций по обработке сигналов весьма широк и иногда удивительно напоминает функции, разработанные в СВЧ-технике. Не будем более распространяться ни о различных технических возможностях [228, 232], ни о соответствующих теоретических аспектах, что является предметом рассмотрения целых книг [235—238].

Но если велики обещания, то значительно и число еще не решенных проблем и трудностей.

1. Жесткие допуски на размеры, требуемые при изготовлении некоторых схем интегральной оптики, требуют улучшения методов фотолитографии. Остаются еще очень большими потери в интегрально-оптических световодах (~ 1 дБ/см).

2. Еще не решены удовлетворительным образом проблемы соединения между отдельными схемами интегральной оптики, а также между оптическими волокнами и интегрально-оптическими схемами.

3. Высокая концентрация энергии в интегральных световодах увеличивает риск разрушения отдельных элементов.

4. И наконец, что касается устройств модуляции света, которые были предметом рассмотрения данной главы, то необходимо отметить следующее: *реализация требуемых характеристик интегрально-оптических устройств модуляции света часто зависит от возможностей получения кристаллов хорошего качества.*

Как и в обычных системах, такие параметры, как надежность, простота эксплуатации, срок службы, являются весьма важными и для устройств интегральной оптики.

ЦИФРОВЫЕ МЕТОДЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

В предыдущих главах были описаны основные элементы волоконно-оптической линии связи (ВОЛС): излучатель, световод и кабель, фотоприемник. В следующих трех главах мы рассмотрим возможности осуществления ВОЛС и покажем области их практического применения. Сначала мы остановимся на цифровых (данная глава) и аналоговых (гл. 16) методах передачи информации, а затем в гл. 17 подробно рассмотрим принципы проектирования полных систем связи.

Быстрое развитие вычислительной техники, значительное снижение стоимости логических схем и преимущества цифровых систем передачи информации (высокая помехозащищенность, нечувствительность к нелинейным искажениям) — все это привело к тому, что в телефонной сети аналоговые системы вытесняются цифровыми. Это касается как систем коммутации (временная электронная коммутация), так и систем передачи (использующих импульсно-кодovou модуляцию, ИКМ). Усилия в области исследований и разработок ВОЛС тесно связаны с этой эволюцией и вследствие этого ориентированы главным образом на использование цифровых методов передачи информации.

Чтобы убедиться в обоснованности такой ориентации, нам кажется целесообразным до начала изучения работы цифровых ВОЛС и до введения основных определений (кода передачи и построения приемника) рассмотреть принципы формирования сигналов, подлежащих передаче, и их назначение, а также напомнить некоторые величины и определения, касающиеся цифровых систем. Особенно важное по сравнению с другими областями значение имеет телефонная связь с применением импульсно-кодовой модуляции. Естественно, что именно на ней мы остановим свое внимание. Однако принципы ИКМ и соответствующие характеристики, рассматриваемые в § 1 и 2, достаточно общие, и потому читатель может распространить их и на другие виды сигналов (телеметрия, телеуправление и т. д.).

§ 1. Импульсно-коддовая модуляция (ИКМ)

Чтобы перевести аналоговый звуковой сигнал в цифровую форму, в которой его можно было бы подать на вход цифровой телефонной системы, необходимо выполнить ряд операций.

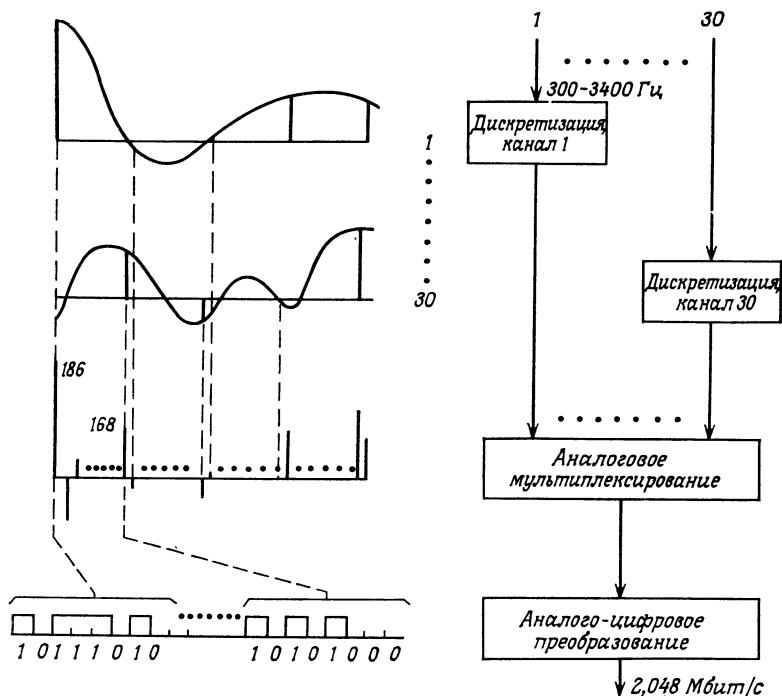


Рис. 15.1. Импульсно-кодовая модуляция (ИКМ).

Схема получения цифрового сигнала.

Они состоят в следующем (рис. 15.1): дискретизация, уплотнение дискретизированных импульсов, соответствующих разным телефонным каналам, и аналого-цифровое преобразование, часто называемое просто кодированием [239]. В установках ИКМ второго поколения (т. е. в тех, которые вводятся в настоящее время) такая последовательность операций единодушно одобрена французскими разработчиками [240]. В установках первого поколения исследовалась возможность аналого-цифрового преобразования до уплотнения [241]. Последнее решение дает некоторые преимущества, состоящие в том, что аналого-цифровой преобразователь работает с более низкой частотой тактовых импульсов, а также в возможности осуществления цифрового мультиплексирования. Однако при этом необходим преобразователь в телефонном канале.

Для восстановления аналогового сигнала из цифрового выполняется обратная последовательность операций: цифро-аналоговое преобразование, разделение каналов, низкочастотная фильтрация.

А. Дискретизация

Эта операция теоретически обосновывается теоремой Шеннона, использованной ранее для дискретизации цифровых голограмм (гл. 5, § 6, п. В). Однако в данном случае она применяется для одномерного сигнала.

Если временной сигнал дискретизируется при частоте следования стробирующих импульсов, в 2 раза превышающей верхнюю частоту спектра сигнала, то дискретный сигнал содержит всю информацию об исходном аналоговом сигнале.

На практике в случае телефонного сигнала, охватывающего спектр частот 300—3400 Гц, выбирают частоту дискретизации несколько выше $2 \cdot 3400$ Гц, т. е. 8000 Гц. Как показано на рис. 15.1, аналоговый сигнал преобразуется на выходе дискретизатора в последовательность импульсов с периодом 125 мкс, модулированных по амплитуде.

Б. Уплотнение каналов

В свободном интервале между двумя последовательными импульсами одного канала можно поместить дискретизованные импульсы 29 других независимых телефонных каналов, которые, следуя друг за другом, образуют новую единую последовательность из 30 импульсов. Такая операция называется уплотнением с разделением каналов во времени.

В. Аналого-цифровое преобразование

Амплитуда импульсов, полученных в результате дискретизации, измеряется и заменяется ближайшей величиной в двоичном коде (квантование). Каждый уровень квантования представляется восьмиразрядным двоичным числом (т. е. имеется $2^8 = 256$ возможных уровней квантования) и преобразуется в бинарные электрические импульсы.

Г. Иерархия цифровых систем связи

Вводя два дополнительных канала (один для синхронизации, а другой для сигнализации) в каждую группу из 30, получают информационную пропускную способность (измеряемую числом импульсов в секунду), оцениваемую величиной

$$(30 + 2) \cdot 8 \frac{1}{125 \cdot 10^{-6}} = 2,048 \text{ Мбит/с}$$

для передачи по 30 телефонным каналам, что равносильно $n (< 30)$ сигналам для передачи массива данных при пропускной способности 64 Кбит/с и $(30 - n)$ телефонным каналам.

В зависимости от важности информационного потока и от имеющегося канала связи можно либо передавать непосредственно последовательность цифр (рис. 15.2), либо объединять их, прибегая к цифровому мультиплексированию с тремя другими, и получать новую последовательность объемом 8,448 Мбит/с, которую можно передавать или объединять с другими, и т. д. Международный консультативный комитет по те-

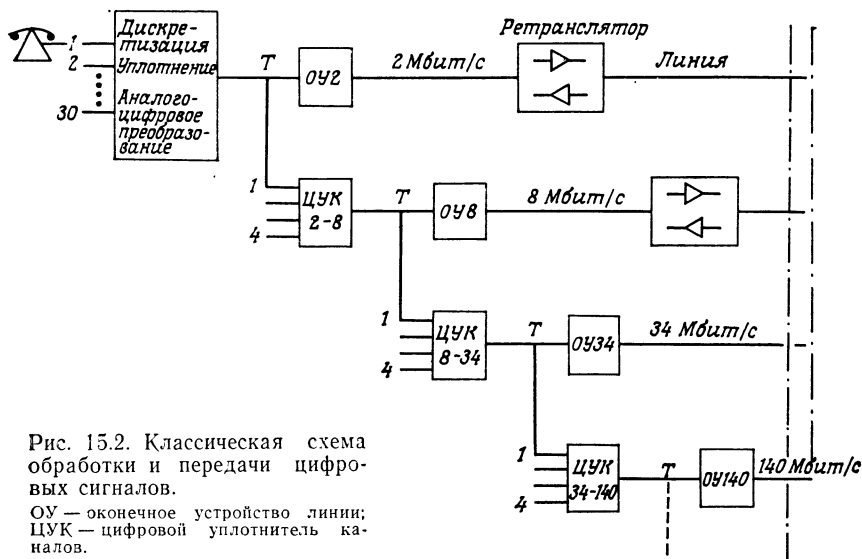


Рис. 15.2. Классическая схема обработки и передачи цифровых сигналов.

ОУ — оконечное устройство линии;
ЦУК — цифровой уплотнитель каналов.

леграфии и телефонии (МККТТ) установил правила уплотнения цифровых последовательностей и определил ступени их иерархии (табл. 15.1) [242].

Получается, что применение оптических волокон в телефонии ориентировано исключительно на нормализованную информационную пропускную способность (во Франции принята преимущественно пропускная способность 2,34 и 140 Мбит/с), но передача данных (между блоками ЭВМ или крупными вычислительными устройствами в промышленности и в военной области) не подчиняется этим нормам и занимает другие области спектра, например 100 Кбит/с, 1 Мбит/с и т. д.

Поскольку цифровое оборудование, предназначенное для дискретизации, уплотнения и аналого-цифрового преобразования передаваемых сигналов, строится в настоящее время на основе классических электронных схем, не зависящих от вида используемого в канале связи носителя информации, выбор носителя существен лишь в процессе передачи сигналов. Вот почему далее мы рассмотрим только специальное оборудование

Таблица 15.1

Иерархия цифровых систем передачи информации
(информационная пропускная способность выражена в единицах Мбит/с,
а информационная емкость есть число телефонных каналов)

	Порядок иерархии	1	2	3	4	5 ¹⁾
Европа	Пропускная способность	2,048	8,448	34,368	139,264	560—840
	Информационная емкость	30	120	480	1920	7680—11 520
США	Пропускная способность	1,544	6,312	44,736	274,176	
	Информационная емкость	24	96	672	4032	
Япония	Пропускная способность	1,544	6,312	32,064	97,728	396,200
	Информационная емкость	24	96	480	1440	5760

¹⁾ Находятся в стадии исследования.

для систем передачи информации на оптических волокнах, показанных на рис. 15.2 за точками *T*. Прежде чем приступить к их анализу, уточним важные параметры систем передачи информации.

§ 2. Основные параметры цифровых систем передачи информации

А. Код передачи

При цифровой передаче телефонных сигналов требуется восстанавливать в приемнике не только двоичные элементы, идентичные переданным, но и информацию о синхронизации по фазе и частоте. Последняя необходима для четкой работы приемника и задания точного времени выполнения операций по принятию решения, цифро-аналоговому преобразованию и разделению информационных каналов. Можно было бы предусмотреть отдельную передачу синхросигнала. Для этого потребовалась бы проводная линия или одно дополнительное оптическое волокно и пришлось бы обеспечить строго идентичные трассы, чтобы импульс сигнала прибывал одновременно с синхронизирующим импульсом. Но проще добавить синхросигнал непосредственно к двоичной информации.

Для этого, чтобы система отвечала требованиям эксплуатации, рассмотренным в § 3, сообщение следует передавать не в простой двоичной форме (0 или 1), а в виде сложных цифровых сигналов. Соответствие между передаваемой информацией в двоичной форме и цифровым сигналом, модулирующим излучатель, определяется *кодом передачи* (или *кодом в линии*). Операция по его осуществлению называется преобразованием кода.

Б. Регенерация импульсов

Как показано на рис. 15.6, в частном случае оптической модели (§ 4, п. А) приемник детектирует и усиливает принятые импульсы, чтобы компенсировать затухание сигнала в линии передачи. Затем производится регенерация сигналов, т. е. восстановление их формы, синхронно с частотой синхронизирующих импульсов. Восстановление сигналов осуществляется с помощью соответствующей схемы фазовой синхронизации. Она

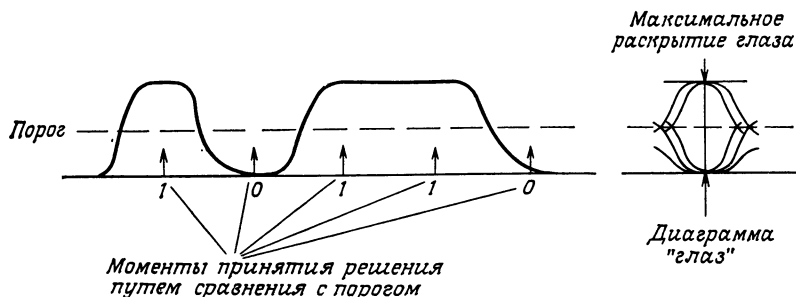


Рис. 15.3. Принцип синхронного детектирования.

позволяет фиксировать более удобные моменты времени для сравнения сигнала с порогом и для принятия решения о принимаемом импульсе (т. е. моменты стробирования). На примере, показанном на рис. 15.3, будет принято решение о приеме единицы (наличие импульса), если сигнал выше порога, и о приеме нуля (отсутствие импульса) в противоположном случае [243].

Заметим, что на практике допустимые отклонения момента принятия решения оцениваются непосредственно по *диаграмме «глаз»*, называемой так из-за ее сходства с человеческим глазом [244]. Чтобы наблюдать такую диаграмму на экране осциллографа, достаточно подать на вход вертикального отклонения сообщение, усиленное приемником, а на вход внешней синхронизации горизонтальной развертки — синхроимпульс. Вертикальным раскрытием «глаза» характеризуется устойчивость

к шуму (рис. 15.3): чем шире раскрыт «глаз», тем меньше влияние шума. Понятно, что оптимальный момент для принятия решения совпадает с центром «глаза». Контур «глаза» показывает чувствительность системы восстановления импульсов к положению момента стробирования. Чем круче контур, тем меньше момент стробирования должен быть удален от центра «глаза».

В. Коэффициент ошибок

Операция принятия решения (стробирования), в принципе очень простая, усложняется из-за искажения сигнала на выходе. Это происходит по двум причинам.

1. Из-за ограниченной полосы пропускания приемника, а в известных случаях и полосы пропускания линии связи форма принимаемого импульса отличается от прямоугольной формы передаваемого импульса. Наблюдаются хвосты импульса вне рабочего промежутка времени и напоявление на соседние импульсы — так называемая *межсимвольная помеха*. Это показано на рис. 15.3, где уровни, соответствующие символам «0», строго говоря, не являются нулевыми.

2. На полезный сигнал накладывается *паразитное напряжение случайного шума*, мгновенная амплитуда которого может быть сравнима с порогом.

Заметим, что на практике обычно нельзя уменьшить влияние обоих эффектов одновременно. В самом деле, как мы увидим из расчетов в § 4, п. Д, хотя при сужении полосы пропускания приемника шум снижается, при этом происходит растяжение фронтов импульса. Но на самом деле влияние последнего фактора достаточно мало по сравнению с влиянием шума, и в расчетах им можно пренебречь.

Впрочем, следует отметить, что в моменты стробирования межсимвольную помеху можно ликвидировать при определенной форме выходного напряжения (импульсы в форме «поднятого косинуса») [243].

Искажение сигнала, в особенности шумом, приводит к принятию ошибочных решений и поступлению к получателю некоторого числа ложных сигналов в двоичной форме. Качество восстановленного сообщения характеризуется *коэффициентом ошибок*, под которым понимают отношение числа ложных битов на выходе приемника к полному числу принятых битов.

Ниже мы найдем зависимость коэффициента ошибок от величины сигнала и шума до принятия решения. Мы предполагаем, что напряжение сигнала — это *случайная величина* с гауссовым законом распределения и стандартным отклонением σ_u . Она *центрирована* на u , когда излучается символ «1», и на нуле при передаче «0». Поэтому условная вероятность приема сим-

вола «0» при передаче символа «1» будет равна вероятности того, что напряжение на выходе приемника будет ниже порога u_s . Следовательно, выражение для этой вероятности имеет вид

$$P(0/1) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \int_{-\infty}^{u_s} \exp\left[-\left(\frac{x-u}{\sqrt{2\sigma_u^2}}\right)^2\right] dx.$$

Аналогично, условная вероятность приема символа «1», когда передается символ «0», будет равна

$$P(1/0) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \int_{u_s}^{\infty} \exp\left[-\left(\frac{x}{\sqrt{2\sigma_u^2}}\right)^2\right] dx.$$

Простая замена переменной приводит к следующим формулам:

$$P(0/1) = \frac{1}{2} \operatorname{erfc}\left(\frac{u-u_s}{\sqrt{2\sigma_u^2}}\right), \quad (15.1)$$

$$P(1/0) = \frac{1}{2} \operatorname{erfc}\left(\frac{u_s}{\sqrt{2\sigma_u^2}}\right), \quad (15.2)$$

где $\operatorname{erfc}(x)$ — дополнительная функция ошибок, равная

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-x^2) dx.$$

Символы «1» и «0» априори считаются равновероятными, а потому полная вероятность ошибки равна

$$P_e = \frac{1}{2} P(0/1) + \frac{1}{2} P(1/0),$$

и если учесть, что вероятность ошибки одинакова для каждого символа, то можно написать

$$P_e = P(0/1) = P(1/0).$$

С учетом (15.1) и (15.2) это выражение дает следующее значение порога:

$$u_s = \frac{u}{2}, \quad (15.3)$$

а выражение для вероятности ошибки (в случае длинной последовательности импульсов, равной наблюдаемому коэффициенту ошибок) примет вид

$$P_e = \frac{1}{2} \operatorname{erfc}\left(\frac{Q}{\sqrt{2}}\right), \quad (15.4)$$

$$\text{где } Q = \frac{u_s}{\sigma_u} = \frac{u}{2\sigma_u}. \quad (15.5)$$

Заметим, что, с одной стороны, как можно было бы интуитивно предвидеть, *порог равен половине амплитуды* (центр «глаза» на рис. 15.3), а с другой — задание коэффициента ошибок эквивалентно заданию величины Q . Введение последнего параметра оправдано, так как его величина пропорциональна отношению сигнала к шуму (S/B), определяемому обычно для цифровых систем связи как отношение максимального напряжения сигнала к действующему значению напряжения шума:

$$Q = \frac{1}{2} \frac{u}{\sigma_u} = \frac{1}{2} S/B. \quad (15.6)$$

Значения величины Q , соответствующие обычно принимаемым значениям коэффициента ошибок, указаны ниже:

P_e	10^{-6}	10^{-8}	10^{-10}
Q	4,75	5,61	6,36

§ 3. Код передачи

А. Критерии выбора

Основные факторы, которыми определяется выбор кода передачи в ВОЛС, таковы.

1. *Спектр сигнала в полосе пропускания.* Он должен быть узким и не должен иметь постоянной составляющей, т. е. должен быть ограничен сверху и снизу.

Первое условие позволяет, с одной стороны, передавать сообщения с использованием узкополосных линий связи (например, несколько километров волокна со скачкообразно изменяющимся показателем преломления), а с другой стороны, позволяет сузить полосу пропускания приемника и соответственно этому (§ 2, п. В; § 4, п. Д) уменьшить влияние шума.

Необходимая точность напряжения питания активных элементов приемника (фотодиода, усилителя) требует введения емкостной связи для развязки по переменному току. Поэтому, если не хотят прибегать к сложным методам (восстановление постоянной составляющей, управление порогом с помощью принимаемого слова), то в спектре сигнала не должны присутствовать низкочастотные составляющие. Следует заметить, что в спектрах сигналов, описываемых ниже, на нулевой частоте присутствует δ -функция Дирака (неправильно называемая спектральной линией), соответствующая среднему значению этих сигналов. *Второе условие не относится к этой линии*, которая не содержит полезной информации, а относится только к непрерывной части спектра.

2. *Восстановление частоты следования импульсов синхронизации.* Для принятия решения и восстановления формы импуль-

сов и их фазы (§ 2, п. Б) требуются импульсы синхронизации. Из принятого сообщения необходимо извлечь сигнал с частотой следования импульсов синхронизации (или кратной ей частотой), причем этот сигнал может быть синусоидальным или представлять собой последовательность прямоугольных импульсов.

Самый распространенный способ — это восстановление импульсов на основе переходов, имеющих в принимаемом сообщении. Добиться синхронизации и поддерживать ее тем легче, чем больше число переходов, которое в нем содержится. В особенно благоприятном случае, когда в спектре принимаемого сообщения имеется одна линия, содержащая гармонику частоты синхронизации, можно выделять ее фильтрацией (случай кода с «обращением», приводимого ниже).

3. Согласование с дискретным сообщением. Код передачи должен отображать любую двоичную последовательность.

4. Контроль за коэффициентом ошибок в окончном оборудовании. При эксплуатации телефонных систем необходимо контролировать качество связи, не прерывая ее. Поэтому требуется код передачи, обладающий избыточностью или характеристическими последовательностями. Тогда для контроля за работой линии связи достаточно на ее конце считать число нарушений правил образования кода. Можно также получить точные сведения о качестве связи, если предварительно теоретически или экспериментально определить соотношение между измеренным и истинным значениями коэффициента ошибок на символ.

5. Простота эксплуатации. Устройства кодирования, декодирования и контроля ошибок должны быть достаточно простыми.

6. Совместимость с передачей по оптическим волокнам. С одной стороны, форма импульсов должна соответствовать наилучшим характеристикам оптического волокна. С другой стороны, минимальная оптическая мощность при заданном коэффициенте ошибок с учетом ослабления при передаче и вводе оптического сигнала в волокно должна быть сравнима с мощностью источника излучения.















Поскольку импульсы излучаемой оптической мощности, разумеется, могут быть только положительными или нулевыми (нет «отрицательных» фотонов), мы не можем непосредственно использовать биполярные коды, которые применяются при передаче информации по проводным линиям.

Б. Применяемые коды

Были предложены и экспериментально проверены многие специальные, если не оригинальные, коды для систем связи [245—247]. Простейшие из них приведены в табл. 15.2.

Таблица 15.2

Определения используемых кодов. Правила кодирования

Код	Определение	Импульсы		Ширина полосы (при 90% энергии)
		«1»	«0»	
Без возвращения к нулю	«1»: уровень 1 на всем интервале «0»: уровень 0 на всем интервале			0,86/T
С возвращением к нулю	«1»: уровень 1 на первой половине интервала «0»: уровень 0 на всем интервале			1,72/T
Двух-фазный типа L	«1»: уровень 1 на первой половине интервала «0»: уровень 1 на второй половине интервала			2,93/T
Двух-фазный типа S	«1»: переход в начале интервала «0»: переход в начале и середине интервала			2,96/T
С обратением	«1»: попеременно уровни 1 и 0 на всем интервале «0»: уровень 1 на первой половине интервала			1,70/T
Электронно-фотонный 1	«1»: попеременно уровни 1 и 0 на всем интервале «0»: переход в середине интервала			1,52/T
Электронно-фотонный 2	«1»: попеременно уровни 1 и 0 на всем интервале «0»: попеременно уровень 1 на первой и второй половине интервала			1,70/T

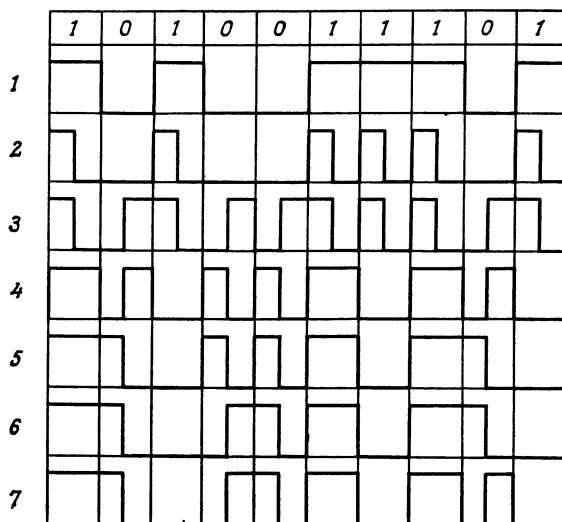


Рис. 15.4. Правила кодирования.

1 — код без возвращения к нулю; 2 — код с возвращением к нулю; 3 — двухфазный код типа *L*; 4 — двухфазный код типа *S*; 5 — код «с обращением» (coded mark inversion); 6 — электронно-фотонный код 1; 7 — электронно-фотонный код 2.

Правила кодирования объяснены на рис. 15.4, а соответствующие изменения спектральных энергетических плотностей — на рис. 15.5. Эти кривые рассчитаны в предположении случайных сообщений, в которых вероятность появления нулей и единиц равна $\frac{1}{2}$ (принцип «все или ничего») [248]. Они показывают ширину спектра сигнала. Полоса пропускания линии передачи (излучатель — волокно — приемник), слишком узкая по сравнению с шириной этих спектральных кривых, приведет к чрезмерной деформации импульсов, а слишком широкая — к нежелательному увеличению шума.

Коды без возвращения к нулю и с возвращением к нулю, хотя и указаны в таблице, не отвечают критерию равенства нулю постоянной составляющей; кроме того, длинные последовательности нулей не дают информации о частоте следования синхроимпульсов. Тем не менее благодаря своей простоте и узкой спектральной полосе кода без возвращения к нулю эти коды получили достаточное распространение (в США, Японии). Чтобы устранить длинные последовательности нулей и единиц, информацию пропускают через специальное устройство (регистр сдвига) на передающей стороне и восстанавливающее устройство в приемнике. Можно осуществлять автоматическое измерение коэффициента ошибок, добавляя к каждому m битам ($m \geq 15$) контрольный разряд четности.

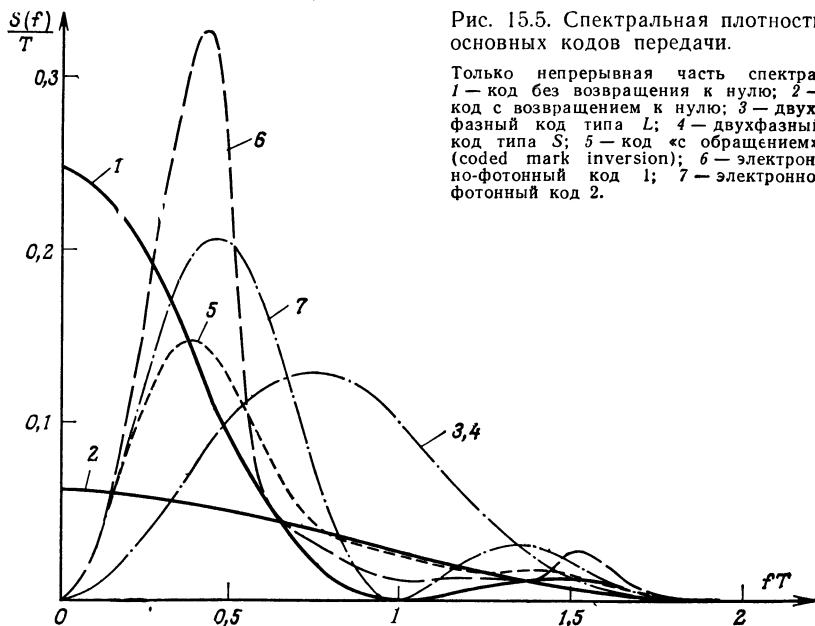


Рис. 15.5. Спектральная плотность основных кодов передачи.

Только непрерывная часть спектра. 1 — код без возвращения к нулю; 2 — код с возвращением к нулю; 3 — двухфазный код типа L ; 4 — двухфазный код типа S ; 5 — код «с обращением» (coded mark inversion); 6 — электронно-фотонный код 1; 7 — электронно-фотонный код 2.

Для высших уровней иерархии ИКМ (информационная пропускная способность 34 и 140 Мбит/с) исследуются коды блоков типа $mBnB$ [249]. Каждое слово передаваемого сообщения объемом m бит по определенному правилу преобразуется в слово, состоящее из n бит ($n > m$).

При низкой скорости передачи информации отношение сигнала к шуму обычно довольно велико, и поэтому можно использовать кодирование с широким спектром частот, очень простое в осуществлении. И наоборот, при высокой скорости передачи важно не расширять спектр частот, чтобы, с одной стороны, эффективно отфильтровать шум в приемнике, а с другой — упростить структуру усилительных схем и схем восстановления частоты синхрипульсов. При этом с учетом экспериментов, проводимых в настоящее время, можно ввести следующую классификацию.

1. Скорость передачи ниже 30 Мбит/с: код двухфазный типа L и S , ИКМ, электронно-фотонный код ЭФ1 и ЭФ2.

2. Скорость передачи от 30 до 200 Мбит/с: код с возвращением к нулю и с устранением длинных последовательностей, код типа 5B6B, ИКМ.

3. Скорость передачи выше 200 Мбит/с: код с возвращением к нулю и с устранением длинных последовательностей.

Заметим, что в связи с наличием нелинейных искажений, вносимых оптическими излучателями, здесь не рассматривались многоуровневые коды.

§ 4. Приемник

А. Общая схема

На рис. 15.6 представлена классическая (за исключением фотодиода) схема приемника.

Детектируемый фотодиодом ток усиливается и фильтруется, чтобы уменьшить влияние шума и получить достаточно высокий уровень сигнала на входе схемы принятия решений. Часто в схему приемника вводится схема автоматической регулировки

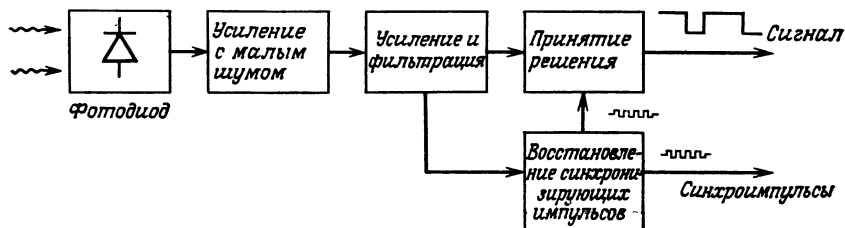


Рис. 15.6. Структурная схема цифрового приемника.

усиления (АРУ). Она компенсирует дрейф рабочей точки характеристик и изменения коэффициента усиления усилителей и коэффициента умножения (в случае использования фотоприемника лавинного типа) или входного уровня. Последний может колебаться из-за нестабильности излучателя или отличаться от теоретического значения, полученного при расчете.

Синхронное детектирование полностью аналогично описанному в § 2, п. 2 для систем с проводными линиями связи.

Следует, однако, отметить небольшое отличие на уровне восстановления синхросигнала. Выпрямление отрицательных импульсов, обычно практикуемое в классических системах, неприменимо непосредственно к системам на оптических волокнах, в которых принимаемые импульсы всегда положительны. Но его можно эффективно использовать в случаях, когда импульс синхронизации содержится в положительно-отрицательном сигнале, получаемом путем дифференцирования принятого сообщения.

Восстановление сигнала на выходе приемника позволяет путем подсчета нарушений правил кодирования проверять качество связи и, если оно недостаточно высокое, включать сигнальные устройства, установленные на станциях в конечном оборудовании линии связи. В случае, когда приемник расположен на промежуточном ретрансляторе, сигнал управляет

непосредственно оптическим излучателем и передается на следующий приемник. В окончательном же устройстве линии связи сигнал после возможного перекодирования, которое придает ему первоначальную форму, подается на разделители каналов (рис. 15.2).

Если, как мы видели ранее, восстановление импульсов синхронизации осуществляется в окончательном оборудовании линии связи, то это увеличивает стоимость и приводит к снижению информационной пропускной способности. Для устранения данного затруднения было предложено исключить эту операцию в промежуточных ретрансляторах. Ценой увеличения «дрожания» сигналов (т. е. неточности хронирования важных моментов в сообщении) и снижения отношения сигнала к шуму производится просто сравнение усиленного сигнала с порогом [250]. Установлено, что при идентичных характеристиках требуемые расстояния L между ретрансляторами должны соответствовать соотношению $L_{3R-3R} = 1,05L_{3R-2R} = 1,25L_{2R-2R}$.

Индексы соответствуют типу ретранслятора:

$3R$ = восстановление по форме — регенерация — восстановление по фазе (ретранслятор с полным восстановлением);

$2R$ = восстановление по форме — регенерация.

Как упоминалось в § 2, п. В, основной величиной, характеризующей качество связи, является коэффициент ошибок. Цифровой приемник принято оценивать *отношением, определяющим при заданной скорости передачи коэффициент ошибок, наблюдаемый на выходе, в функции от принимаемой оптической мощности*. Это отношение имеет то преимущество, что его можно использовать непосредственно для оценки системы, но оно неудобно тем, что зависит от кода передачи.

Например, если T — длительность элементарного интервала ($1/T$ — скорость передачи данных) в случае передачи сообщения, закодированного кодом без возврата к нулю, то для определения принимаемого символа достаточно сравнить сигнал с порогом в моменты T , $2T$, $3T$ и т. д. В случае сообщений, закодированных двухфазным кодом типа L , могут иметь место по крайней мере два возможных решения, что приводит к получению двух значений коэффициента ошибок при определенном значении оптической мощности (а значит, и отношения сигнала к шуму на выходе приемника). Кроме того, можно принимать решения в моменты $T/4$, $5T/4$ и т. д. и сравнивать с порогом сигналы в первой половине интервала (рис. 15.7, а) или принимать решение в моменты $T/4$, $3T/4$, $5T/4$, $7T/4$ и т. д. и сравнивать сигналы в первой и второй половинах интервала (рис. 15.7, б). В самом деле, по определению двухфазного кода типа L , представленного в табл. 15.2, сигнал в первой половине интервала идентичен символу, и поэтому, если разность между сигналом в первой половине и сигналом во второй половине отрицательна, то символ есть «0» (и «1» в противоположном случае) [251].

Ниже будет рассмотрена работа приемника и получено выражение для отношения сигнала к шуму и для коэффициента

Рис. 15.7. Принцип принятия решений для сообщений, кодированных двухфазным кодом типа L .

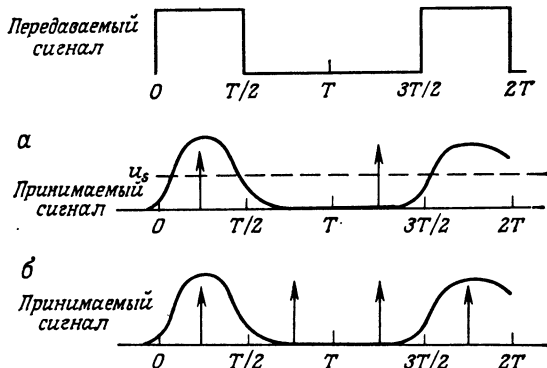
Стрелка показывает момент стробирования.

α — единственное решение:

$$u(T/4), u(5T/4), \dots \begin{cases} > u_s \rightarrow \langle 1 \rangle; \\ < u_s \rightarrow \langle 0 \rangle; \end{cases}$$

β — двойное решение:

$$u(T/4) - u(3T/4), \dots \begin{cases} > 0 \rightarrow \langle 1 \rangle, \\ u(5T/4) - u(7T/4), \dots < 0 \rightarrow \langle 0 \rangle. \end{cases}$$



ошибок в функции от принимаемой оптической мощности. Для упрощения расчетов и для того, чтобы исключить зависимость от кода передачи, допустим, что передаваемое сообщение кодируется кодом без возврата к нулю. Вид кодирования учитывается приближенно путем изменения оптических мощностей

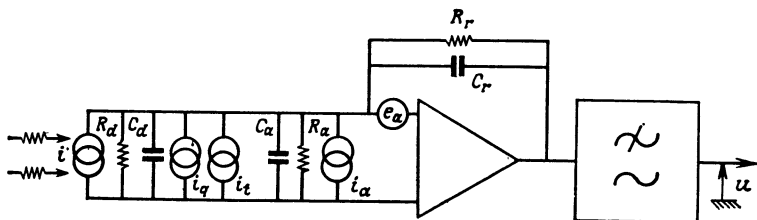


Рис. 15.8. Входные каскады приемника.

в соответствии с отношением полос пропускания (например, для сообщения, кодируемого двухфазным кодом типа L , она будет на 3 дБ больше). Кроме того, предположим, что синхронное детектирование осуществляется абсолютно точно, т. е. что решение принимается в центре диаграммы «глаз». Благодаря такому предположению можно ограничиться исследованием только той части схемы на рис. 15.8, в которой производится аналоговая обработка сигнала.

Б. Схемы приема

Нагрузочные элементы фотодиода R_d и C_d (емкость перехода) и входные элементы предусилителя R_a и C_a можно заменить элементами R_c и C_c :

$$R_c = \frac{R_a R_d}{R_a + R_d},$$

$$C_c = C_a + C_d.$$

Ниже (п. Г и Е) мы убедимся в том, что оптимизация отношения сигнала к шуму и соответственно коэффициента ошибок требует более высоких значений R_c и низких C_c . Очевидно, что это соответствует увеличению важной величины — сигнала на входе усилителя.

На практике для оптимизации отношения сигнала к шуму приходится обращаться к технологии гибридных интегральных микросхем при изготовлении предусилителя [252, 253]. Но входной каскад, действующий на сигнал как интегратор, слишком сильно деформирует импульсы, что требует применения дифференцирующей цепочки после усилителя. Назначение такой схемы, называемой схемой компенсации (выравнивающей схемой), состоит в том, чтобы придать импульсам такую форму, при которой точно срабатывает схема принятия решений. Кроме того, она снижает до минимума влияние входного каскада приемника, а если необходимо, то и волокна. Заметим, что компенсация узкой полосы пропускания канала связи, обычно практикуемая в системах с проводными линиями, недостаточно оправдана в случае оптических волокон, у которых широкая полоса пропускания.

Интегрирование приводит еще и к увеличению динамического диапазона сигнала, подлежащего усилению, между минимальным уровнем (длинный «0») и максимальным уровнем (длинная «1»), что создает трудности при изготовлении усилителя с линейной амплитудной характеристикой. К этому явлению добавляются уже упомянутые колебания принимаемой оптической мощности. Они объясняются флуктуациями излучателя и расхождением между номинальным уровнем сигнала, по которому регулируется коэффициент усиления усилителей, и реальных уровней, встречающихся в ходе эксплуатации. Такое расхождение объясняется дисперсией и ослаблением оптического сигнала при передаче по волокну. На практике подобные затруднения преодолевают с помощью системы комплексной отрицательной обратной связи в виде цепочки параллельно включенных R_r и C_r , соединяющей выход схемы предусилитель — усилитель с ее входом. Это схема, представленная на рис. 15.8, и она служит основой для теоретических расчетов. Заметим, что, как это часто бывает в электронике, выбор величины R_r должен быть компромиссным. С одной стороны, величина R_r должна быть большой, чтобы вносимый сопротивлением вклад в общий шум был пренебрежимо мал [254—256]. С другой стороны, для уменьшения флуктуаций уровня сигнала на выходе усилителя нужно, чтобы сопротивление R_r было мало.

В. Ток фотоприемника

Как было показано в гл. 13, § 2, п. Г, ток i на выходе фотоприемника пропорционален принимаемой оптической мощности p и дается выражением

$$i = \frac{\eta q}{h\nu} Mp + i_0, \quad (15.7)$$

где η — квантовый выход ($\eta < 1$); q — заряд электрона ($1,6 \cdot 10^{-19}$ К); $h\nu$ — энергия фотона (h — постоянная Планка, ν — оптическая частота); M — средний коэффициент умножения фотодиода (в случае PIN -фотодиода $M = 1$); i_0 — темновой ток.

Темновой ток обычно невелик (< 1 нА для кремниевых фотодиодов) и, кроме того, он не передается приемниками, в которых используется фильтрация низких частот. Следовательно, он не оказывает значительного влияния на характеристики приемников, работающих в диапазоне около 0,84 мкм, и его можно не принимать в расчет.

Г. Источники шума

На принимаемый сигнал накладываются разного рода шумы, имеющие различное происхождение и, несмотря на то что они носят случайный характер, их необходимо определять с большой точностью. Как достаточно хорошо подтверждено практикой, эти шумы можно считать белым шумом, т. е. они имеют однородный спектр и не коррелируют друг с другом. При вычислении отношения сигнала к шуму принимаются в расчет различные источники шума: дробовой шум, тепловой шум и шум предусилителя. Мы кратко остановимся на разных видах шумов и дадим выражения для их спектральной плотности, учитывая, что они подаются обратно на вход канала приема, как показано на рис. 15.8.

1. Дробовой шум

Такой шум обусловлен случайным пуассоновским распределением фотонов в принимаемой оптической волне и случайным характером процесса умножения в фотодиоде, если фотодиод лавинного типа или типа $P\pi PN$ (гл. 13, § 4). *Спектральная плотность этого шума в удвоенной полосе частот* дается следующим очень приближенным выражением [217, 257]:

$$N_{iq} = \frac{\eta q^2}{h\nu} M^2 F(M) p,$$

где $F(M)$ — коэффициент шума фотодиода, равный [222]

$$F(M) = k'M + \left(2 - \frac{1}{M}\right)(1 - k').$$

Здесь k' — постоянная, зависящая от типа полупроводника.

Часто используется следующее приближенное выражение для $F(M)$:

$$F(M) = M^x,$$

где x — дополнительная постоянная шума, равная приблизительно 0,5 для кремниевых и ~ 1 для германиевых фотодиодов.

Учитывая (15.7), получаем

$$N_{iq} = qiM^{1+x}. \quad (15.8)$$

Последнее выражение показывает существенное отличие приемников классических систем связи от приемников волоконно-оптических систем связи. *В оптических приемниках шум зависит от сигнала.*

Следовательно, в данном случае нельзя опираться на привычные представления теории оптимального приема [243], которая предполагает, что шум в приемнике постоянен и не зависит от сигнала.

Чтобы упростить нижеследующие расчеты, при оценке N_{iq} мы будем рассматривать только среднее значение тока [257]. Кроме того, если пренебречь током утечки и темновым током фотоприемника по сравнению с током сигнала, то мы также можем пренебречь и дробовым шумом, связанным с этими токами.

И наконец, заметим, что спектральная плотность шума связана со средним квадратом тока дробового шума следующим элементарным соотношением [243]:

$$\langle i_q^2 \rangle = 2N_{iq}B,$$

где B — полоса пропускания приемника.

2. Тепловой шум

Этот шум обусловлен резисторами R_c и R_r . Спектральная плотность теплового шума дается выражением

$$N_{it} = 2kT_k \left(\frac{1}{R_c} + \frac{1}{R_r} \right) = \frac{2kT_k}{R}, \quad (15.9)$$

где k — постоянная Больцмана ($1,38 \cdot 10^{-23}$ Дж/К), а T_k — абсолютная температура.

3. Собственные шумы предусилителя

Собственные шумы предусилителя оцениваются шумовым током i_a и шумовым напряжением e_a на его выходе и соответствующими спектральными плотностями N_{ia} и N_{ea} . Отметим, что N_{ia} измеряется в единицах $A^2/Гц$, а N_{ea} — в единицах $V^2/Гц$.

Д. Расчет отношения сигнала к шуму

На практике коэффициент усиления схемы предусилитель — усилитель высок и напряжение на выходе приемника u связано с входным током i соотношением [254]

$$u = \frac{R_r}{1 + 2\pi i f R_r C_r} i.$$

В случае, когда величина $1/2\pi R_r C_r$ незначительно превышает спектр частот сигнала Δf , это превышение должно компенсироваться схемой дифференцирования. В результате, какой бы ни была величина $\Delta f/2\pi R_r C_r$, выполняется элементарное соотношение

$$u = R_r i. \quad (15.10)$$

Расчет дисперсии σ_u^2 случайной величины u , называемой также мощностью шума, выполняется аналогичным образом. Для доли, вносимой шумовыми токами, обозначенной через $\sigma_u'^2$, получим

$$\sigma_u'^2 = \int_{-\Delta f}^{\Delta f} R_r^2 (N_{iq} + N_{it} + N_{ia}) df,$$

т.е. $\sigma_u'^2 = 2R_r^2 \Delta f (N_{iq} + N_{it} + N_{ia}). \quad (15.11)$

Точную оценку роли и вклада дробового шума дал Персоник [217], который сыграл важнейшую роль в области теоретического анализа приемника. Пусть $h(t)$ — импульсный отклик приемника; тогда напряжение на его выходе $u(t)$ связано с входным током $i(t)$ соотношением

$$u(t) = h(t) * i(t) = \int_{-\infty}^{\infty} i(t') h(t - t') dt'.$$

Вместо приближенного результата, даваемого формулой (15.11), а именно

$$\sigma_{q_1}^2 = 2R_r^2 \Delta f N_{iq} = 2qi R_r^2 \Delta f M^{1+x},$$

Персоник [217] получил точное выражение

$$\sigma_{q_2}^2 = \int_{-\infty}^{\infty} q M^{1+x} i(t') h^2(t - t') dt'.$$

Будем обозначать большими буквами фурье-образы временных функций. На основании соотношения Парсеваля преобразуем это выражение к виду

$$\sigma_{q_2}^2 = \int_{-\infty}^{\infty} q M^{1+x} I(f) (H(f) * H^*(f)) e^{-2\pi i f t} df.$$

Если изменения величины $H(f)$ малы по сравнению с изменениями величины $I(f)$ (случай, когда Δf больше спектра частот сигнала), то $\sigma_{q_2}^2$ сводится к $\sigma_{q_1}^2$.

Для оценки доли, вносимой напряжением шума $\sigma_u''^2$, нужно рассчитать передаточную функцию между e_a и выходным напряжением при нулевом токе сигнала. В соответствии с рис. 15.8 находим

$$\frac{u}{e_a} \Big|_{i=0} = \frac{R_r (1 + i2\pi f RC)}{(1 + i2\pi f R_r C_r)},$$

где с учетом (15.9) полагаем

$$R = \frac{R_r R_c}{R_r + R_c}, \\ C = C_r + C_c.$$

Приняв вышеупомянутое замечание в отношении компенсации знаменателя, получаем

$$\frac{u}{e_a} \Big|_{i=0} = \frac{R_r}{R} (1 + i2\pi f RC),$$

откуда следует выражение для дисперсии напряжения шума на выходе:

$$\sigma_u''^2 = \frac{R_r^2}{R^2} \int_{-\Delta f}^{\Delta f} (1 + 4\pi^2 R^2 C^2 f^2) N_{ea} df, \\ \text{т.е. } \sigma_u''^2 = 2 \frac{N_{ea} R_r^2}{R^2} \Delta f \frac{1 + 4\pi^2 R^2 C^2 (\Delta f)^2}{3}.$$

На практике выполняется соотношение (п. Ж)

$$\Delta f \gg \frac{1}{2\pi RC},$$

откуда следует

$$\sigma_u''^2 = \frac{8}{3} N_{ea} R_r^2 \pi^2 C^2 (\Delta f)^3. \quad (15.12)$$

Преобразуем это выражение в соответствии с (15.11). На основании *предположения о независимости источников шума* непосредственно получаем выражение для полной дисперсии шума:

$$\sigma_u^2 = 2\Delta f R_r^2 \left[N_{iq} + N_{it} + N_{ia} + \frac{4}{3} \pi^2 C^2 (\Delta f)^2 N_{ea} \right]. \quad (15.13)$$

В соответствии с определением (15.6) отношения сигнала к шуму, обозначаемого через S/B , можно написать

$$\left(\frac{S}{B} \right)^2 = \frac{u^2}{\sigma_u^2} = \frac{i^2}{2\Delta f} \frac{1}{N_{iq} + N_{it} + N_{ia} + \frac{4}{3} \pi^2 C^2 (\Delta f)^2 N_{ea}}.$$

Подставляя сюда выражение (15.9), окончательно получаем

$$\left(\frac{S}{B}\right)^2 = \frac{i^2}{2\Delta f} \frac{1}{N_{iq} + 2kT_k \left(\frac{1}{R_c} + \frac{1}{R_r}\right) + N_{ia} + \frac{4}{3} \pi^2 C^2 (\Delta f)^2 N_{ea}}. \quad (15.14)$$

Итак, мы убеждаемся что для оптимизации отношения сигнала к шуму нужно, чтобы источники шума были слабыми (это и так очевидно), входной ток (т. е. падающая на фотоприемник оптическая мощность) — значительным, а полные сопротивления нагрузки и цепи обратной связи — большими (C_c и C_r — малы, R_c и R_r — большие).

Е. Соотношение между оптической мощностью и коэффициентом ошибок

В § 2, п. В было установлено соотношение между коэффициентом ошибок и отношением сигнала к шуму. Однако этот расчет основан на предположении, что выходное напряжение является гауссовой случайной переменной. Допустив некоторую вольность в теореме о «центральном пределе» для набора случайных величин, примем, что напряжение на выходе оптического приемника, изображенного на рис. 15.8, тоже гауссова случайная величина. Более сложные, но и более точные расчеты (моделирование методом Монте-Карло, пределы Чернова) были выполнены Персоником [259, 260]. Они показывают, что рассматриваемая гипотеза приемлема и приводит к результатам, близким, если не считать значения порога u_s , к наблюдаемым на практике.

Предположим, что случайная переменная центрирована на

$$u = R_r i = R_r \frac{\eta q}{h\nu} p,$$

когда передается символ «1» (принимаемая оптическая мощность равна p), и на нуле, когда передается «0».

Тогда можно прямо применить результаты, полученные в § 3, п. В, и с помощью формул (15.4) и (15.6) рассчитать получаемый при этом коэффициент ошибок P_e в зависимости от принимаемой оптической мощности. Однако, как мы увидим в гл. 17, при проектировании оптической системы связи важно знать минимальную оптическую мощность, при которой будет получен заданный коэффициент ошибок.

Поэтому вместо P_e более предпочтительна переменная Q ; учитывая (15.6) и (15.14), получаем

$$4Q^2 = \frac{i^2}{2\Delta f} \frac{1}{q \frac{M^{1+\gamma_i}}{2} + N_{it} + N_{ia} + \frac{4}{3} \pi^2 C^2 (\Delta f)^2 N_{ea}}.$$

Вводя *средний ток* i_m (средний во времени), равный половине максимального значения тока, находим

$$2Q^2 = \frac{i_m^2}{\Delta f} \frac{1}{qM^{1+x}i_m + N_{it} + N_{ia} + \frac{4}{3}\pi^2 C^2 (\Delta f)^2 N_{ea}}. \quad (15.15)$$

Введем обозначения

$$A' = Q^2 \Delta f, \quad (15.16)$$

$$B' = N_{it} + N_{ia} + \frac{4}{3}\pi^2 C^2 (\Delta f)^2 N_{ea}. \quad (15.17)$$

Тогда из (15.15) получаем выражение для i_m и с учетом формулы (15.7) — выражение для *средней оптической мощности*:

$$p_m = \frac{h\nu}{\eta q} \{qA'M^x + \sqrt{q^2 A'^2 M^x + 2A'B'M^{-2}}\}, \quad (15.18)$$

в котором мы пренебрегли темновым током фотоприемника i_0 из-за его малости (п. В).

В случае PIN-фотодиода ($M=1$) формула (15.18) упрощается:

$$p_m = \frac{h\nu}{\eta q} \{qA' + \sqrt{q^2 A'^2 + 2A'B'}\}. \quad (15.19)$$

Заметим, что в случае лавинного фотодиода величина p_m в формуле (15.18) будет зависеть от M . Она сначала убывает благодаря члену с M^{-2} , а затем возрастает благодаря члену с M^x . Следовательно, *существует некое оптимальное значение M , равное M_0 , которое минимизирует оптическую мощность при заданном коэффициенте ошибок*. Приравняв нулю производную от выражения для p_m по M , получим

$$M_0 = \left(\frac{2B'}{q^2 A' x (x+2)} \right)^{\frac{1}{2+x}}, \quad (15.20)$$

откуда следует минимальное значение p_{m0} величины p_m :

$$p_{m0} = \frac{h\nu}{\eta q^2} \frac{2B'}{x} \left(\frac{q^2 A' x (x+2)}{2B'} \right)^{\frac{2+x}{2+x}}. \quad (15.21)$$

Ж. Числовой пример

Для количественной оценки зависимости p_m от скорости передачи информации и коэффициента ошибок зададимся типичными значениями величин параметров элементов, входящих в состав приемника:

1) Кремниевый фотодиод:

Чувствительность $(\eta q/h\nu) = 0,5$ А/Вт.

Дополнительная постоянная шума $x \approx 0,5$.

2) Полное сопротивление нагрузки: $R_c \approx 1 \text{ МОм}$, $C_c \approx 10 \text{ пФ}$.

3) Полное сопротивление цепи обратной связи: Наименьшее значение емкости, которое можно получить, равно $C_r = 0,5 \text{ пФ}$. Сопротивление резистора R_r выберем из условия обеспечения полосы пропускания усилителя, равной Δf , т. е.

$$R_r = \frac{1}{2\pi\Delta f C_r}.$$

И наконец, допустим, что для сигнала, кодированного кодом без возврата к нулю, необходима полоса пропускания, равная скорости передачи:

$$\Delta f = 1/T.$$

Тогда окончательно получим

$$R_r = \frac{T}{2\pi C_r}.$$

Заметим, что на практике обычно достаточна полоса пропускания, равная $0,7/T$ (рис. 15.5).

4) *Предусилитель*. При скоростях передачи данных, не превышающих 50 Мбит/с, предусилитель выполняется на полевых транзисторах, которые удовлетворяют требованиям высокого входного сопротивления и низкого шума. В таком случае величина N_{ia} пренебрежимо мала по сравнению с N_{it} , и мы получаем [261]

$$N_{ea} = \frac{2kT_k \cdot 0,7}{g},$$

где g — крутизна характеристики первого полевого транзистора. Принимаем $g = 5 \text{ мА/В}$.

При информационной пропускной способности больше 50 Мбит/с требование низкого уровня шума при большом коэффициенте усиления заставляет применять биполярные транзисторы. В этом случае имеем $N_{ia} = qI_B$, где $I_B \approx 30 \text{ мкА}$ — ток базы, и $N_{ea} = 2kT_k r_{bb'}$, где $r_{bb'} \approx 20 \text{ Ом}$ — сопротивление базы.

На рис. 15.9 приведена зависимость минимальной оптической мощности на входе приемника, необходимой для обеспечения коэффициента ошибок, равного 10^{-8} (принятое значение, обеспечивающее хорошее качество передачи), от информационной пропускной способности ВОЛС. Из графика видно, что введение биполярных транзисторов приводит к изменению крутизны кривых в точке, соответствующей 50 Мбит/с. В логарифмическом масштабе эти характеристики имеют вид прямых, уравнения которых таковы:

$$\begin{aligned} f < 50 \text{ Мбит/с} & \begin{cases} p_m = -55 + 11 \lg f, & (\text{PIN}) \\ p_{m0} = -70 + 10,5 \lg f, & (\text{ЛФД}) \end{cases} \\ f > 50 \text{ Мбит/с} & \begin{cases} p_m = -53 + 10 \lg f, & (\text{PIN}) \\ p_{m0} = -70 + 10 \lg f, & (\text{ЛФД}) \end{cases} \end{aligned}$$

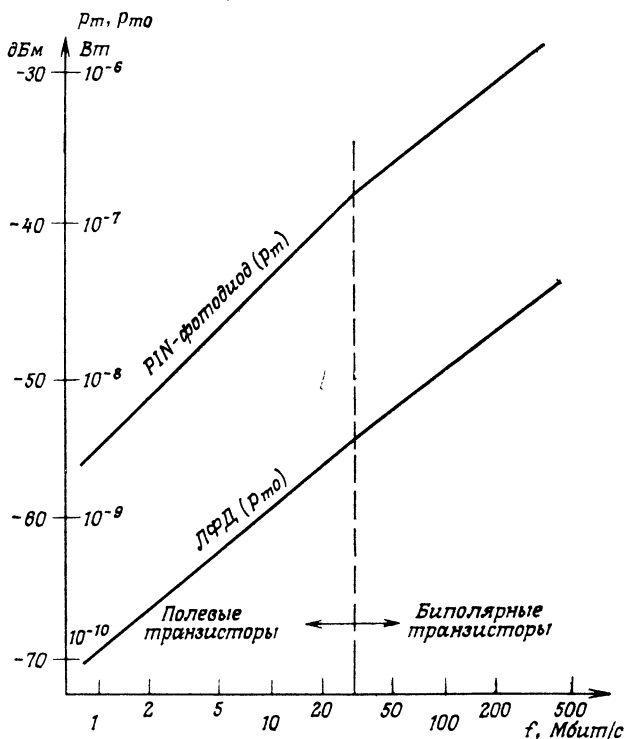


Рис. 15.9. Зависимость минимально необходимой оптической мощности на входе приемника от информационной пропускной способности линии при коэффициенте ошибок, равном 10^{-8} .

где величины p_m и p_{m0} измеряются в единицах дБм, а f — в единицах Мбит/с.

Напомним, что дБм — это единица, равная превышению в децибелах уровня мощности, равного 1 мВт, т. е.

$$p \text{ (дБм)} = 10 \lg p \text{ (мВт)}.$$

Полученные нами результаты сопоставимы с минимальным теоретическим значением для случая, когда можно полностью пренебречь шумами приемника, а учитывается только фотонный пуассоновский шум,

$$p_q = -85 + 10 \lg f.$$

Следовательно, влияние шума приемника оказывается значительным, поскольку приводит к увеличению требуемой мощности на входе приблизительно на 15 дБ при использовании лавинного фотодиода и на 30 дБ в случае PIN-фотодиода.

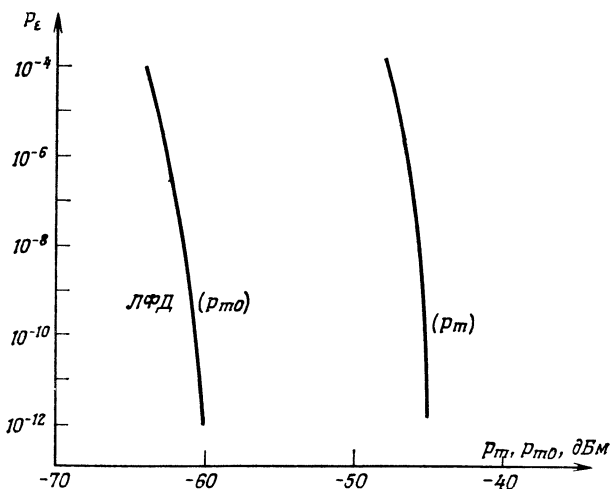


Рис. 15.10. Зависимость коэффициента ошибок от оптической мощности на входе приемника при информационной пропускной способности 8 Мбит/с. (Правая кривая — для PIN -фотодиода.)

Кроме того, отметим, что применение лавинного фотодиода тем более целесообразно, чем выше требуемая скорость передачи: выигрыш по чувствительности составляет около 14 дБ при 1 Мбит/с и 18 дБ при 500 Мбит/с. На практике трудно точно контролировать напряжение питания фотодиода, а следовательно, и точно поддерживать его коэффициент умножения равным оптимальному значению. Вследствие этого коэффициент усиления составляет 10—15 дБ.

Изложенные теоретические и практические выводы относятся к длинам волн около 0,84 мкм и к кремниевым фотоприемникам. Если же рассматриваются большие длины волн, вблизи 1,3 мкм (а именно такая тенденция наметилась в настоящее время), то необходимо использовать германиевый фотоприемник. В этом случае схема расчета остается прежней, но КПД и дополнительная постоянная шума фотоприемника x будут другими. Темновой ток будет больше ($i_0 > 0,5$ мкА), и, следовательно, нельзя больше пренебрегать связанным с ним дробовым шумом. Это приводит к более высоким значениям p_m и p_{m0} , а также к уменьшению того выигрыша, который обеспечивает лавинный фотодиод [262].

Несмотря на сделанные замечания, приведенные здесь числовые значения не только согласуются с результатами других исследований, но и приближаются к экспериментальным данным на длине волны 0,84 мкм [263, 264].

В дополнение к изложенному на рис. 15.10 приведена зависи-

мость коэффициента ошибок от принимаемой оптической мощности при скорости передачи 8 Мбит/с. Значительный наклон кривых объясняется просто резким изменением дополнительной функции ошибок.

Простым сравнением требуемой принимаемой оптической мощности и имеющейся излучаемой мощности (той и другой в единицах дБм) можно установить величину допустимых оптических потерь между излучателем и приемником. В гл. 17 будет показано, как использовать полученные результаты для оценки полных характеристик оптических систем связи при их проектировании.

АНАЛОГОВЫЕ МЕТОДЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

Как было показано в предыдущей главе, дальняя связь развивается в направлении цифровой передачи информации. Большинство исследований и разработок, связанных с волоконной оптикой, также ориентированы в этом направлении. Что же касается аналоговых методов связи на оптических волокнах, то они, если не считать первых экспериментов, носивших чисто демонстрационный рекламный характер (передача по волокну телевизионного сигнала), представляют сейчас интерес практически лишь в области *телеметрии и кабельного телевидения*.

В случае телеметрии можно осуществлять передачу результатов измерения между точками, находящимися под разным напряжением (например, при измерении тока в линии сверхвысокого напряжения). При этом необходимая электроизоляция обеспечивается самим оптическим волокном [265].

В области кабельного телевидения, достаточно развитого в Англии, США и Японии (проекты HI-OVI¹⁾), например, выполнены полностью на оптических волокнах) [266—288], для улучшения качества передачи и расширения предоставляемых услуг по сравнению с системами на коаксиальных кабелях используют волокна с низкими потерями и широкой полосой пропускания. Применение оптического волокна позволяет снизить до минимума число ретрансляторов между центральной станцией и абонентом, и даже полностью исключить их, а также увеличить число телевизионных каналов и передаваемых сигналов (факсимильных, телефонных и т. д.). Связь типа абонент — центральная станция позволяет вести диалог с центральной ЭВМ (получение разного рода справок, выбор и заказы товаров из дома) [244, 269, 270, 361].

Но пока что наша задача будет более скромной. Она состоит в том, чтобы описать и проанализировать работу аналоговых систем связи на оптических волокнах. Ниже рассматриваются три вида модуляции, оказывающие прямое влияние на качество передачи.

1. *Модуляция по интенсивности, или прямая модуляция*. Это самый простой в реализации способ модуляции. Передаваемый

¹⁾ High-Fidelity Optical Visual Information (англ.) — высоконадежная оптическая визуальная информация. — Прим. ред.

сигнал, по возможности усиленный, непосредственно управляет оптическим источником и модулирует интенсивность излучаемого им света.

2. *Частотная модуляция (ЧМ)*. Передаваемый сигнал предварительно модулирует по частоте электрическую поднесущую, а затем сформированный ЧМ-сигнал подается на оптический излучатель.

3. *Фазово-импульсная модуляция (ФИМ)*. В зависимости от изменения амплитуды сигнала изменяется положение излучаемых импульсов по отношению к положению опорных тактовых импульсов.

§ 1. Модуляция по интенсивности

А. Нелинейные искажения

Хотя модуляция по интенсивности и проста в осуществлении, для этого необходим оптический излучатель с линейной характеристикой, т. е. кривой зависимости излучаемой мощности от тока модуляции.

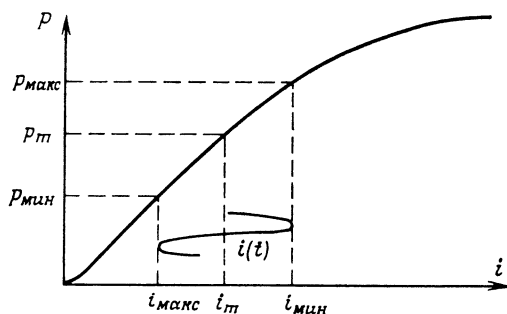


Рис. 16.1. Характеристика излучения светодиода.

На рис. 16.1 приведена типичная характеристика электролюминесцентного светодиода¹⁾. Рабочую точку, относительно которой осуществляется модуляция, обозначим через $M(p_m; i_m)$, а ток сигнала — через $i(t)$. Заметим, что модуляция осуществляется в классе А и выражение для коэффициента модуляции имеет вид

$$m = \frac{i_{\max} - i_{\min}}{i_{\max} + i_{\min}} = \frac{\sqrt{2} i_e}{i_m} = \frac{p_{\max} - p_{\min}}{p_{\max} + p_{\min}}, \quad (16.1)$$

где i_e — действующее значение тока сигнала модуляции.

¹⁾ В настоящее время модуляция по интенсивности применяется только в светодиодах [271]. Выпускаемые же промышленностью полупроводниковые лазеры имеют изломы на характеристиках (гл. 12, § 4, п. Г, 1), что исключает возможность применения в них прямой модуляции по интенсивности.

Если требуется хорошая линейность (малый уровень гармоник), то можно либо уменьшить глубину модуляции, что, однако, приводит, как это будет показано ниже, к снижению отношения сигнала к шуму при приеме, либо осуществить компенсацию вносимых нелинейных искажений. Предложено несколько методов компенсации нелинейности оптического излучателя [272].

1. *Введение предискажений* [273]. В передаваемый сигнал электронными методами предварительно вводят искажения, обратные тем, которые при модуляции внесет оптический излучатель.

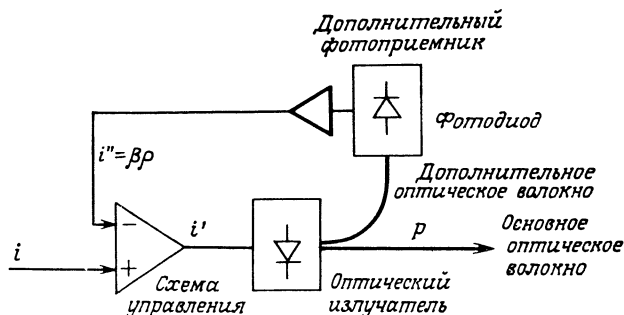


Рис. 16.2. Линеаризация модуляционной характеристики светодиода при помощи цепи обратной связи.

2. *Метод фазовой модуляции* [272]. Он состоит в одновременной модуляции двух светодиодов с идентичными характеристиками, равными по величине, но сдвинутыми по фазе на $\pi/2$ токами: $I \cos \omega t$ и $I \cos (\omega t + \pi/2)$. Последующее сложение излучаемых мощностей (например, с помощью смесительного волокна, описанного в гл. 17, § 4, п. Б) позволяет подавить вторые гармоники, т. е. слагаемые вида $\cos 2\omega t$ и $\cos (2\omega t + \pi)$.

3. *Метод полупредискажений* [274]. В этом методе используется опорный светодиод, который позволяет оценить уровень вносимых искажений и скорректировать их путем соответствующего изменения тока модуляции излучателя. Как и в предыдущем методе, основное неудобство состоит в том, что нужны два диода с идентичными характеристиками, которые можно изготовить только по специальному заказу.

4. *Метод отрицательной обратной связи* [275]. Такой способ линеаризации, широко используемый в усилителях, самый простой в реализации. Он состоит в том (рис. 16.2), что при помощи вспомогательного волокна отводят малую долю излучаемой мощности r , которая затем воздействует на ток, модулирующий светодиод.

Пусть i'' — ток обратной связи, т. е.

$$i'' = \beta r.$$

где β — коэффициент, учитывающий эффективность связи между излучателем и вспомогательным фотоприемником, а также усиление последнего. Следовательно, фактический модулирующий ток будет равен

$$i' = i - i''.$$

В хорошем приближении характеристику излучения светодиода можно считать параболой вида

$$p = a(i' - bi'^2),$$

где $a \approx 10^{-2}$ Вт/А, $b \approx 2 \text{ А}^{-1}$. Подставив сюда ток i , выраженный через ток сигнала i и излучаемую мощность p , получим

$$p = a[(i - i'') - b(i - i'')^2] = a[(i - \beta p) - b(i - \beta p)^2].$$

Решим полученное уравнение относительно $i - \beta p$:

$$i - \beta p = -\frac{1 + \beta a \pm \sqrt{(1 + \beta a)^2 - 4iab\beta}}{2ab\beta}.$$

Выбрав только одно решение, которое дает $p = 0$ при $i = 0$, и разложив его в ряд по степеням $(i + \beta a)$, получим, ограничившись двумя первыми членами,

$$i - \beta p = \frac{i}{1 + \beta a} + \frac{ab\beta i^2}{(1 + \beta a)^3},$$

$$\text{откуда } p = \frac{a}{1 + \beta a} \left(i - \frac{bi^2}{(1 + \beta a)} \right).$$

Если сравнить полученное выражение для излучаемой мощности с тем, которое было в отсутствие обратной связи,

$$p = a(i - bi^2),$$

то нетрудно видеть, что амплитуда искажений уменьшается в $(1 + \beta a)^2$ раз, но одновременно с этим и излучаемая мощность уменьшается в $(1 + \beta a)$ раз. Таким образом, коэффициент обратной связи β нужно выбирать на основе компромисса между требованием низкого уровня нелинейных искажений, с одной стороны, и требованием высокой стабильности схемы и высокой излучаемой мощности — с другой. Это обычное положение для систем с обратной связью.

Б. Отношение сигнала к шуму

Отношение сигнала к шуму при передаче информации аналоговыми методами, так же как и коэффициент ошибок при цифровой передаче, сильно зависит от характеристик приемника. Последний совершенно аналогичен усилительно-фильтрационной части цифрового приемника, изображенного на рис. 15.8. Различия имеются в схеме автоматической регулировки усиления

(АРУ) и в схеме восстановления постоянной составляющей, но они не существенны при оценке сигнала к шуму. Поэтому можно рассчитывать это отношение так, как это делалось в гл. 15, § 4, п. Д.

Если определить отношение сигнала к шуму как отношение действующих значений напряжений сигнала и шума (как это и принято в аналоговых системах), то можно написать

$$\left(\frac{S}{B}\right)^2 = \frac{u_e^2}{\sigma_u^2}. \quad (16.2)$$

Отсюда, учитывая (15.7) и вводя среднее значение тока i_m на выходе фотодиода, а также коэффициент модуляции m [формула (16.1)], получаем следующее выражение:

$$\left(\frac{S}{B}\right)^2 = \frac{m^2 i_m^2}{\Delta f} \frac{1}{N_{iq} + N_{it} + N_{ia} + \frac{4}{3} \pi^2 C^2 (\Delta f)^2 N_{ea}}, \quad (16.3)$$

где Δf — ширина частотного спектра сигнала.

Отсюда следует, что отношение сигнала к шуму растет с увеличением коэффициента модуляции.

В. Требуемая оптическая мощность

Как говорилось в предыдущей главе, при проектировании важно знать минимальную оптическую мощность, которая обеспечивает требуемое качество связи (характеризуемое обычно отношением сигнала к шуму). Чтобы определить эту мощность, можно воспользоваться (с некоторыми изменениями) расчетами, проведенными в гл. 15, § 4, п. Е.

Как и там, введем обозначения

$$A' = \left(\frac{S}{B}\right)^2 \frac{\Delta f}{2m^2}, \quad (16.4)$$

$$B' = N_{it} + N_{ia} + \frac{4}{3} \pi^2 C^2 (\Delta f)^2 N_{ea}. \quad (16.5)$$

Как и ранее, мы решаем уравнение

$$2A' = \frac{i_m^2}{q i_m + B'} \quad (16.6)$$

и находим ток и среднюю оптическую мощность.

Схема с PIN-фотодиодом. Выражение для средней оптической мощности таково:

$$p_m = \frac{h\nu}{\eta q} (qA' + \sqrt{q^2 A'^2 + 2A'B'}).$$

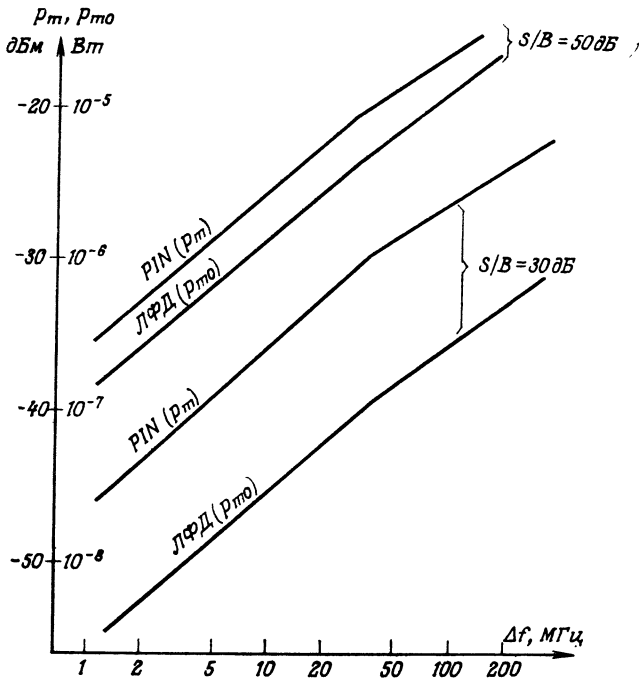


Рис. 16.3. Модуляция по интенсивности.

Зависимость минимальной оптической мощности на входе приемника от ширины спектра полезного сигнала при заданном отношении сигнала к шуму

Оно упрощается, если пренебречь дробовым шумом в сравнении с уровнем тепловых шумов:

$$p_m \approx \frac{h\nu}{\eta q} \sqrt{A'B'} = \frac{h\nu}{\eta q} \left(\frac{S}{B} \right) \frac{1}{m} \sqrt{\frac{B'\Delta f}{2}} \quad (16.7)$$

Схема с лавинным фотодиодом. С учетом того, что оптимальное значение величины M определяется формулой (15.20), а минимальное значение величины p_m — выражением (15.21), имеем

$$p_{m0} = \frac{h\nu}{\eta q^2} \frac{2B'}{x} \left(\frac{q^2 A' x (x+2)}{2B'} \right)^{\frac{2+x}{2+2x}}. \quad (16.8)$$

На рис. 16.3 представлены кривые зависимости величин p_m и p_{m0} от полосы частот сигнала при разных отношениях сигнала к шуму и при постоянном коэффициенте модуляции, равном 0,8, построенные с использованием типичных значений величин, приведенных в гл. 15, § 4, п. Ж. Расчеты были практически идентичны с выполненными в гл. 15, и полученные кривые подобны кривым рис. 15.9.

На рис. 16.4 представлена зависимость минимально допустимой мощности на входе приемника от отношения сигнала к шуму при передаче сигнала в полосе частот 6 МГц.

Мы видим, что *применение лавинного фотодиода тем более целесообразно, чем меньше отношение сигнала к шуму*. Объясняется это довольно просто. При больших отношениях сигнала к шуму требуемая оптическая мощность также значительна и

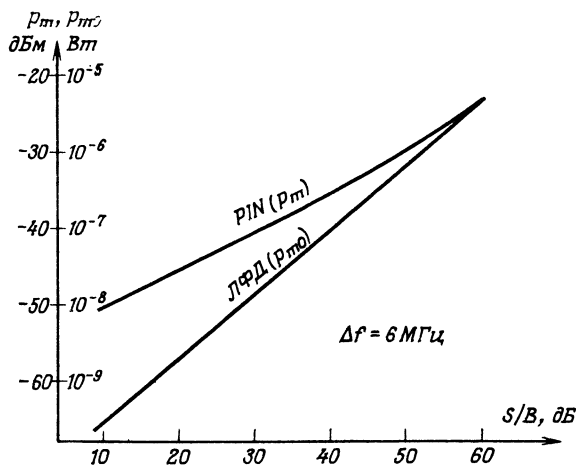


Рис. 16.4. Модуляция по интенсивности.

Зависимость минимально необходимой оптической мощности на входе приемника от отношения сигнала к шуму при полосе пропускания 6 МГц.

в соответствии с формулой (15.8) будет сравнительно большим дробовой шум. Для его снижения необходимо уменьшать коэффициент умножения, а тогда более целесообразно взять *PIN*-фотодиод. Отметим также, что в оптимальном случае дробовой шум, а также сумма теплового шума и шума усилителя — величины одного порядка. Поскольку уровень последних постоянен, при уменьшении принимаемой мощности должен уменьшаться и коэффициент M .

§ 2. Частотная модуляция¹⁾

Предварительная частотная модуляция поднесущей частоты передаваемым сигналом (на рис. 16.5) — это обычный метод *повышения отношения сигнала к шуму и снижения нелинейных искажений*, вносимых оптическим излучателем (разумеется, при

¹⁾ Под частотной модуляцией (ЧМ) здесь понимается совокупность двух операций: частотной модуляции поднесущей и модуляции по интенсивности оптического излучателя (рис. 16.5).

условии обеспечения хорошей линейности характеристик ЧМ-модуляторов и демодуляторов).

Прежде чем перейти к количественному анализу возможного увеличения отношения сигнала к шуму при использовании ча-

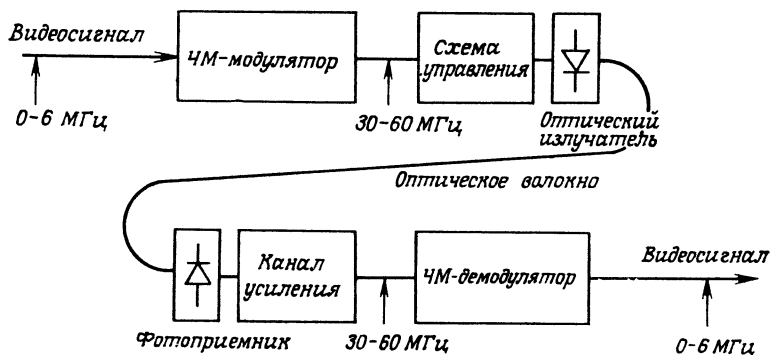


Рис. 16.5. Структурная схема аналоговой системы оптической связи, использующей частотную модуляцию.

Модулятор может представлять собой генератор, управляемый напряжением, а демодулятор — схему ограничитель — дискриминатор или схему фазовой синхронизации

стотной модуляции, напомним вкратце о ширине полосы пропускания, необходимой для передачи ЧМ-сигнала.

А. Полоса пропускания

Для обеспечения возможности количественного сравнения рассматриваемого способа со способом модуляции по интенсивности примем, как и в § 1, п. Б, ширину полезной полосы частот равной 6 МГц (передача телевизионного сигнала), а индекс модуляции — равным 1,5 (т. е. отношение девиации частоты к максимальной частоте модулирующего сигнала равно 1,5). По формуле Карсона [276] находим ширину спектра частотно-модулированного сигнала:

$$(\Delta f)_{\text{ЧМ}} = 2\Delta f(1 + \beta) = 30 \text{ МГц},$$

где β — индекс модуляции, а Δf — ширина спектра частот сигнала.

Чтобы исключить взаимодействие паразитных сигналов интермодуляции второго порядка с полезным сигналом, поднесущую частоту F выберем из условия, что спектр занимает ровно одну октаву, т. е.

$$2\left(F - \frac{(\Delta f)_{\text{ЧМ}}}{2}\right) \geq F + \frac{(\Delta f)_{\text{ЧМ}}}{2}.$$

Откуда находим $F = 45 \text{ МГц}$.

Б. Отношение сигнала к шуму

Итак, сигнал, который должен передаваться по волокну и которым модулируется оптический излучатель, занимает полосу частот от 30 до 60 МГц. Следовательно, приемник должен иметь такую же полосу пропускания с центральной частотой 45 МГц. Это мало изменяет схему приемника, представленную на рис. 15.8. Достаточно только заменить фильтр низких частот полосовым фильтром. Мало изменится и расчет отношения сигнала к шуму. Сначала определим отношение $(S/B)_{\text{чм}}$ сигнала к шуму для ЧМ-сигнала на поднесущей, а затем — отношение сигнала к шуму на выходе приемника, воспользовавшись простой формулой [277, с. 128]

$$\left(\frac{S}{B}\right)^2 = 3\beta^2 \left(\frac{S}{B}\right)_{\text{чм}}^2. \quad (16.9)$$

Для определения мощности шума на поднесущей достаточно выполнить расчет в рабочей полосе вокруг частоты F [277]. Так же как и в гл. 15, § 4, п. Д, получаем

$$\begin{aligned} \sigma_u'^2 &= R_r^2 \int_{F-\Delta f}^{F+\Delta f} (N_{iq} + N_{it} + N_{ia}) df, \\ \text{т. е. } \sigma_u'^2 &= 2\Delta f R_r^2 (N_{iq} + N_{it} + N_{ia}). \end{aligned} \quad (16.10)$$

Доля, вносимая напряжением, будет равна

$$\sigma_u''^2 = \frac{R_r^2}{R^2} \int_{F-\Delta f}^{F+\Delta f} (1 + 4\pi^2 R^2 C^2 f^2) N_{ea} df.$$

Как и в случае, рассмотренном в гл. 15, § 4, п. Д, преобладает второй член:

$$\sigma_u''^2 = \frac{8}{3} \pi^2 R_r^2 C^2 N_{ea} \Delta f [(\Delta f)^2 + 3F^2].$$

Пренебрегая величиной Δf по сравнению с F , получаем

$$\sigma_u''^2 = 8\pi^2 R_r^2 C^2 N_{ea} \Delta f F^2. \quad (16.11)$$

Из полученных соотношений выводим выражение для отношения сигнала к шуму на поднесущей:

$$\left(\frac{S}{B}\right)_{\text{чм}}^2 = \frac{i_e^2}{\sigma_u'^2 + \sigma_u''^2} = \frac{i_e^2}{2\Delta f} \frac{1}{N_{iq} + N_{it} + N_{ia} + 4\pi^2 C^2 F^2 N_{ea}}.$$

где индекс e указывает на то, что речь идет о действующих значениях.

Предположив, что оптический излучатель модулируется по интенсивности (рис. 16.5) с коэффициентом модуляции 100 %, находим следующее соотношение, связывающее среднее значение тока приемника с действующим значением тока поднесущей [рис. 16.1 и выражение (16.1)]:

$$i_m = \sqrt{2} i_e.$$

Окончательно получаем

$$\left(\frac{S}{B}\right)_{\text{чм}} = \frac{i_m^2}{4\Delta f} \frac{1}{N_{iq} + N_{it} + N_{ia} + 4\pi^2 C^2 F^2 N_{ea}}. \quad (16.12)$$

По формуле (16.9) можно вычислить отношение сигнала к шуму в основной полосе частот сигнала на выходе приемника.

В. Требуемая оптическая мощность

Чтобы рассчитать зависимость оптической мощности от отношения сигнала к шуму, введем, как и в гл. 15, § 4, п. Е, обозначения

$$A' = \frac{2}{3} \frac{\Delta f}{\beta^2} \left(\frac{S}{B}\right)^2, \quad (16.13)$$

$$B' = N_{it} + N_{ia} + 4\pi^2 C^2 F^2 N_{ea} \quad (16.14)$$

и затем найдем оптическую мощность, необходимую для получения заданного отношения сигнала к шуму; в случае *PIN*-фотодиода:

$$p_m \approx \frac{h\nu}{\eta q} \sqrt{2A'B'} = \frac{h\nu}{\eta q} \frac{2}{\beta} \left(\frac{S}{B}\right) \sqrt{\frac{B'\Delta f}{3}}, \quad (16.15)$$

и в случае лавинного фотодиода:

$$p_{m0} = \frac{h\nu}{\eta q^2} \frac{2B'}{x} \left(\frac{q^2 A' x (x+2)}{2B'} \right)^{\frac{2+x}{2+2x}}. \quad (16.16)$$

Поскольку полоса пропускания приемника занимает область от 30 до 60 МГц, имеет смысл выбрать биполярные транзисторы. На рис. 16.6 приведены зависимости величин p_m и p_{m0} от отношения сигнала к шуму при типичных значениях параметров схемы, которые были указаны в гл. 15, § 4, п. Ж. Сравнивая эти зависимости с аналогичными зависимостями для прямой модуляции по интенсивности, мы видим, что при отношении сигнала к шуму, равном 40 дБ, в случае лавинного фотодиода мы имеем выигрыш 2 дБ, а в случае *PIN*-фотодиода — проигрыш 10 дБ. В частном случае передачи телевизионного сигнала с использованием ЧМ, когда учитывается конкретный вид шума на выходе ЧМ-демодулятора и вводится схема коррекции предискажений

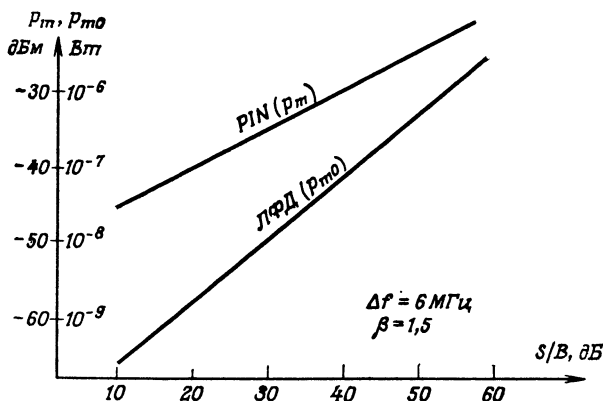


Рис. 16.6. Частотная модуляция.

Зависимость минимально необходимой оптической мощности на входе приемника от отношения сигнала к шуму при постоянной полосе пропускания.

[278], можно повысить отношение сигнала к шуму более чем на 15 дБ, в чем существенное преимущество рассматриваемого типа модуляции перед прямой модуляцией по интенсивности.

§ 3. Фазово-импульсная модуляция

Рассмотренные способы модуляции применяются главным образом при модуляции с помощью светодиодов, крутизна характеристики которых, хотя и не совсем постоянна, все же достаточно мала. В имеющихся же полупроводниковых лазерах дело обстоит иначе. С одной стороны, они имеют крутой наклон характеристики излучения при токах выше порога. С другой стороны, из-за наличия на характеристике нерегулярностей в виде изломов (гл. 12, § 4, п. Г) их трудно, а то и просто невозможно использовать при чисто аналоговых видах модуляции. Но в них возможна фазово-импульсная модуляция, которая позволяет использовать такое существенное преимущество ППЛ перед СД, как значительно бóльшая излучаемая и вводимая в волокно оптическая мощность.

А Принцип фазово-импульсной модуляции [277]

Принцип фазово-импульсной модуляции состоит в следующем: после стробирования (дискретизации) передаваемого сигнала, осуществляемого через строго одинаковые промежутки времени T_h , $2T_h$, $3T_h$, ..., каждый раз излучается один оптический импульс стандартной формы и постоянной амплитуды. При этом (рис. 16.7) задержка оптического импульса относительно

момента стробирования пропорциональна мгновенному значению сигнала в момент стробирования (причем коэффициент пропорциональности α также постоянный).

На приемной стороне по результатам измерения временных интервалов между тактовыми импульсами дискретизации и принимаемыми импульсами производится восстановление полезного сигнала (тактовые импульсы могут либо восстанавливаться на основе принятого сообщения, либо вырабатываться в приемном

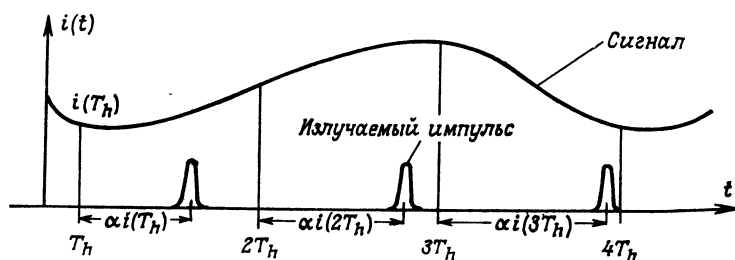


Рис. 16.7. Принцип фазово-импульсной модуляции.

оборудовании). Таким образом осуществляется преобразование *фазово-импульсной модуляции в широтно-импульсную модуляцию*, после чего простая низкочастотная фильтрация позволяет восстановить переданный сигнал.

Неточность измерения этих временных интервалов, связанная в большой степени с искажениями формы принимаемых импульсов, проявляется в виде шума, воздействующего на восстановленный сигнал.

Б. Отношение сигнала к шуму

Чтобы определить моменты поступления импульсов, нужно постоянно сравнивать принимаемый сигнал с неким порогом. Расчет отношения сигнала к шуму сложен, если импульсы не заданы в простом аналитическом виде. На основе наблюдений импульсов на выходе оптического волокна можно предположить, что регистрируемый фототок (т. е. с точностью до некоторого постоянного множителя — падающая на фотоприемник оптическая мощность) описывается выражением

$$i(t) = \begin{cases} I \cos^2 \frac{\pi t}{2\theta} & \text{при } |t| \leq \theta, \\ 0 & \text{при } |t| > \theta. \end{cases} \quad (16.17)$$

Это означает, что полоса пропускания оптического волокна должна быть больше $1/\theta$, а полный диапазон модуляции между двумя тактовыми импульсами равен

$$\tau = T_h - 2\theta. \quad (16.18)$$

Выбрав коэффициент модуляции равным 1, получаем, что величина τ будет пропорциональна удвоенной амплитуде полезного сигнала. В результате выражение для отношения сигнала к шуму, определяемого как отношение действующих значений сигнала и шума, примет вид

$$\left(\frac{S}{B}\right)_{\text{Фим}}^2 = \frac{1}{2} \left(\frac{\tau}{2}\right)^2 \frac{1}{\langle \delta t^2 \rangle} = \frac{\tau^2}{8 \langle \delta t^2 \rangle}, \quad (16.19)$$

где $\tau^2/8$ — мощность сигнала, выраженная через временной интервал, а $\langle \delta t^2 \rangle$ — дисперсия моментов поступления импульсов относительно среднего значения.

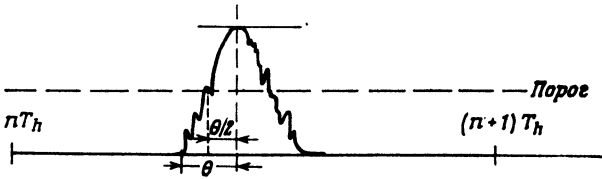


Рис. 16.8. Определение момента поступления импульса.

Чтобы определить неточность момента поступления импульса, нужно найти неточность момента прохождения сигналом порогового уровня. Как видно из рис. 16.8, пороговый уровень — это уровень, пересекаемый в момент $t = -\theta/2$ от вершины импульса. Эта неточность дается выражением [277]

$$\delta t = \tilde{u} \frac{1}{\left. \frac{du}{dt} \right|_{t = -\frac{\theta}{2}}}, \quad (16.20)$$

где \tilde{u} — мгновенное значение напряжения шума¹⁾.

Заметим, что, как можно было сказать заранее, неточность момента пересечения порога тем больше, чем меньше крутизна переднего фронта импульса.

С учетом соотношения (15.10) между выходным напряжением и входным током и выражения для тока (16.17) формула (16.20) принимает вид

$$\delta t = \frac{\tilde{u}}{R_r} \frac{2\theta}{\pi I}, \quad (16.21)$$

а после квадратичного усреднения

$$\langle \delta t^2 \rangle = \langle \tilde{u}^2 \rangle \left(\frac{2\theta}{R_r \pi I} \right)^2 = \sigma_u^2 \frac{4\theta^2}{R_r^2 \pi^2 I^2}. \quad (16.22)$$

¹⁾ Полный численный расчет показывает, что можно пренебрегать ложными отклонениями, которые наблюдаются вне непосредственной близости от импульса.

В результате получаем

$$\left(\frac{S}{B}\right)_{\text{Фим}}^2 = \frac{\tau^2 \pi^2 I^2 R_r^2}{32 \sigma_u^2 \theta^2}. \quad (16.23)$$

Для вычисления σ_u^2 предположим, что полоса пропускания приемника равна $1/\theta$. Это гарантирует отсутствие слишком больших искажений импульсов и хорошую фильтрацию шума. Теперь достаточно заменить Δf величиной $1/\theta$ в выражении (15.13) для мощности шума, чтобы получить окончательное выражение для отношения сигнала к шуму при фазово-импульсной модуляции:

$$\left(\frac{S}{B}\right)_{\text{Фим}}^2 = \frac{\tau^2 \pi^2 I^2}{64 \theta (N_{iq} + N_{it} + N_{ia} + \frac{4}{3} \frac{\pi^2 C^2}{\theta^2} N_{ea})}. \quad (16.24)$$

Вместо максимального тока I часто предпочитают использовать среднее значение тока i_m . Усредняя мгновенное значение тока за период, получаем

$$i_m = \frac{1}{T_h} \int_{-\frac{T_h}{2}}^{\frac{T_h}{2}} i(t) dt = \frac{\theta}{T_h} I. \quad (16.25)$$

В. Требуемая оптическая мощность

Как и ранее, нетрудно вывести выражение для оптической мощности, требуемой для обеспечения заданного отношения сигнала к шуму. Вводим обозначения

$$A' = \frac{32 \theta^3}{\pi^2 \tau^2 T_h^2} \left(\frac{S}{B}\right)_{\text{Фим}}^2 = \frac{32 \theta^2}{\pi^2 (T_h - 2\theta)^2 T_h^2} \left(\frac{S}{B}\right)_{\text{Фим}}^2, \quad (16.26)$$

$$B' = N_{it} + N_{ia} + \frac{4 \pi^2 C^2}{3 \theta^2} N_{ea}. \quad (16.27)$$

После необходимых выкладок получаем

$$\rho_m \approx \frac{h\nu}{\eta q} \sqrt{2A'B'}, \quad (16.28)$$

$$\rho_{m0} = \frac{h\nu}{\eta q^2} \frac{2B'}{x} \left(\frac{q^2 A' x (x+2)}{2B'} \right)^{\frac{2+x}{2+2x}}. \quad (16.29)$$

Для иллюстрации возьмем два значения полосы пропускания

$$\frac{1}{\theta} = 40 \text{ МГц}, \quad \frac{1}{\theta} = 80 \text{ МГц},$$

что соответствует длительности импульса на полувысоте

$$\theta = 25 \text{ нс}, \quad \theta = 12,5 \text{ нс}.$$

При ширине полосы, занимаемой полезным сигналом, равной 6 МГц, частота дискретизации будет равна (гл. 15, § 1, п. А)

$$1/T_h = 2 \cdot 6 = 12 \text{ МГц},$$

а период тактовых импульсов

$$T_h = 83,3 \text{ нс}.$$

При значениях параметров приемника на биполярных транзисторах, указанных в гл. 15, § 4, п. Ж, и при отношении сигнала к шуму, равном 40 дБ, находим следующие значения чувствительности приема:

При $\theta = 25 \text{ нс}$

$p_m = 1,9 \text{ мВт} (-27 \text{ дБм})$ для PIN-фотодиода,

$p_{m0} = 180 \text{ нВт} (-37 \text{ дБм})$ для лавинного фотодиода.

При $\theta = 12,5 \text{ нс}$

$p_m = 490 \text{ нВт} (-33 \text{ дБм})$ для PIN-фотодиода,

$p_{m0} = 13 \text{ нВт} (-49 \text{ дБм})$ для лавинного фотодиода.

Полученные результаты показывают, что чувствительность приема ухудшается с увеличением ширины излучаемых импульсов. При длительности импульса 25 нс (измеренной на полувысоте) найденные значения несколько хуже, чем в случае модуляции по интенсивности и частотной модуляции. И наоборот, разделив длительность импульса на 2, т. е. получив 12,5 нс (ширина полосы пропускания удваивается), и взяв лавинный фотодиод, можно выиграть от 8,5 до 10,5 дБ по отношению к этим видам модуляции.

§ 4. Сравнение разных видов модуляции

Если имеется возможность передавать очень короткие импульсы, то наименьшей мощности на входе оптического приемника при постоянном отношении сигнала к шуму на выходе требует фазово-импульсная модуляция. За ней идет частотная модуляция и лишь потом модуляция по интенсивности. Но на практике приходится принимать в расчет и другие технические факторы (простота реализации, вид передаваемого сигнала), а также экономические показатели (стоимость системы модулятор — демодулятор).

Модуляция по интенсивности проще всего в осуществлении (прямая модуляция, узкая требуемая полоса пропускания). Конечно, при такой модуляции существенны нелинейности характеристики излучателя, которые заставляют уменьшать глубину модуляции, что приводит к повышению уровня требуемой мощности, но появление в последнее время одномодовых полупро-

водниковых лазеров позволяет получить значительную оптическую мощность и компенсировать последний недостаток [29, 279, 280].

Частотная модуляция неудобна тем, что требует более широкой полосы пропускания, но при прочих равных условиях она обеспечивает более высокое отношение сигнала к шуму (выигрыш ~ 10 дБ) и малые искажения [265]. Добавим, что такая модуляция целесообразна, когда оптическое волокно должно служить наземной частью системы телевизионной связи через спутники или радиорелейные линии. В самом деле, сигналы, передаваемые по радиоканалу через атмосферу, уже модулированы по частоте и их можно передавать по оптическому волокну без существенных изменений [281, 282].

Фазово-импульсная модуляция наиболее сложна в осуществлении. С одной стороны, дискретизация приводит к обрезанию верхних частот сигнала, а с другой — малая длительность импульсов требует широкой полосы пропускания приемников, что усложняет их реализацию. По-видимому, эти недостатки позволяют устранить в какой-то мере (ценой некоторого ухудшения характеристик) такие менее сложные виды модуляции, как *широотно-импульсная модуляция* [283] и особенно *частотно-импульсная модуляция* [284, 285], соединяющая в себе преимущества импульсной и частотной модуляции.

ГЛАВА 17

ОСНОВЫ ПРОЕКТИРОВАНИЯ ВОЛОКОННО-ОПТИЧЕСКИХ ЛИНИЙ СВЯЗИ

Создание оптических волоконных кабелей с малым затуханием, разработка элементов оптоэлектроники и методов соединения оптических волокон, описанных в гл. 11—14, позволяет уже сейчас внедрять их в системы передачи информации. Специалисты надеются уже в ближайшее время получить эффект от использования всех или некоторых специфических свойств оптических волокон в зависимости от области применения (промышленная, массового обслуживания или военная). Кратко напомним эти свойства:

1) *малые потери*; это увеличивает расстояния, на которых возможна передача информации без промежуточного усиления (дальняя связь, кабельное вещание);

2) *большая полоса пропускания*; предварительное мультиплексирование сигналов позволяет уменьшить число волокон;

3) *нечувствительность к электромагнитным помехам*; не страшны помехи от посторонних источников (возможна связь вблизи линий электропередач или в районах с высокой магнитной активностью);

4) *отсутствие излучения*; обеспечивается скрытность обмена информацией и практически отсутствуют перекрестные помехи;

5) *малые размеры*; это очень важно в малых по объему помещениях (автомобили, самолеты, городской транспорт);

6) *малый вес*; снижение удельного веса является одним из основных направлений исследований в военной и авиакосмической промышленности;

7) *электрическая изоляция*; возможна передача информации между двумя точками с разными электрическими потенциалами без особых предосторожностей (корабли, электростанции);

8) *низкая предполагаемая стоимость*.

При проектировании волоконно-оптической линии связи нужно знать условия функционирования системы, вид передаваемой информации (аналоговая или цифровая, гл. 15 и 16) и учитывать нужды пользователя. Поэтому, прежде чем описывать различные структуры систем связи (§ 3), мы остановимся на вопросе о технических условиях, задаваемых заказчиком, и предварительно рассмотрим важнейшие свойства применяемых подсистем (излучатель — световод — приемник) на основе детального описания, приведенного в гл. 11 — 14.

Для простоты мы будем говорить в основном об элементарной двухабонентной линии. Но, учитывая возрастающий интерес к волоконно-оптическим линиям связи (ВОЛС), мы рассмотрим также и системы с более сложной структурой (кольцевые, многолучевые и т. д.).

§ 1. Технические условия

Проектирование ВОЛС начинается с тщательного изучения технических условий, поставленных пользователем. В частности, важно знать:

а) *вид передаваемого сигнала*: аналоговый или цифровой, полосу частот или требуемая информационная пропускная способность канала;

б) *качество, которое должна обеспечивать линия связи*: отношение сигнала к шуму или коэффициент ошибки, искажения, фазовую характеристику, дифференциальный коэффициент усиления и допустимое «дрожание» восстановленного синхросигнала;

в) *длину линии связи* и, возможно, требуемые длины отрезков оптического волоконного кабеля;

г) *уровни сигнала, входные и выходные импедансы линии связи*;

д) *структуру системы связи*: последовательная или параллельная, одно- или двунаправленная, двухабонентная, кольцевая или многолучевая;

е) *надежность оборудования*;

ж) *условия эксплуатации*: пределы изменения температуры, давления, влажности; воздушная или подземная линия;

з) *срок ввода в эксплуатацию*.

Все эти характеристики в разной мере нужны в зависимости от назначения линии связи и могут быть известны лишь частично. В общем случае три первых требования необходимы для определения основных технических решений: выбора вида модуляции, определения структурной схемы приемника; остальные требования служат ориентирами при выборе компонентов системы.

§ 2. Анализ линии связи

Прежде чем показать, как выбираются компоненты, наилучшим образом соответствующие техническим требованиям, напомним основные элементы волоконно-оптической линии связи и их важнейшие свойства [286—288].

А. Излучение

В соответствии с упрощенной схемой, показанной на рис. 17.1, первым элементом оптической системы связи является блок управления оптическим излучателем (в который может входить

фазовый модулятор или фазово-импульсный модулятор). Он приводит сигнал к уровню, совместимому с характеристиками излучателя. Кроме того, в случае *светодиода* (СД) он может компенсировать эффекты накопления, а в случае *полупроводникового лазера* (ППЛ) — наложить на сигнал ток смещения, чтобы обеспечить надпороговую модуляцию. Этот блок может также содержать цепи обратной связи, которые, как показано на рис. 16.2, снижают нелинейность характеристик светодиода и стабилизируют мощность излучения полупроводникового лазера

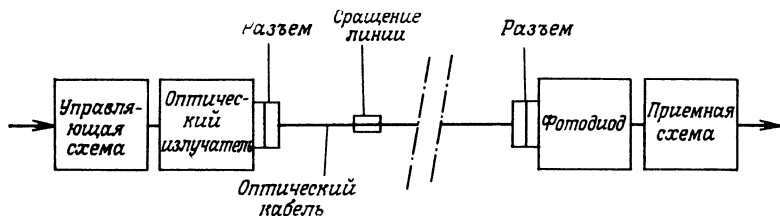


Рис 17.1. Структурная схема двухабонентной линии связи.

при помощи промежуточного фотодиода. Отметим, что в некоторых промышленных типах полупроводниковых лазеров применяется обратная связь, основанная на измерении фотодиодом выходной мощности лазера по его обратному излучению (излучению с задней поверхности резонатора). Учитывая, что изменения мощности излучения с передней и задней излучающих поверхностей лазера пропорциональны друг другу, такой вариант привлекает своей простотой технического решения по сравнению с использованием дополнительного оптического волокна (рис. 16.2) [289].

Как было отмечено ранее, оптическим излучателем может служить либо СД, либо ППЛ. Применение СД более предпочтительно благодаря меньшей стоимости, простоте осуществления модуляции (необходим ток в пределах от 50 до 100 мА) и большему сроку службы, что также является важным фактором. К недостаткам СД следует отнести малую мощность излучения (~ 1 мВт) и невысокий КПД соединения с оптическим волокном (пропорциональный корню квадратному из числовой апертуры). Так, например, максимальная мощность излучения, вводимая в оптическое волокно с числовой апертурой 0,2, равна ~ 40 мкВт. Полоса пропускания СД может достигать 100 МГц, что делает весьма широким круг его возможных применений.

У полупроводникового лазера много своих преимуществ: широкая полоса пропускания (свыше 500 МГц), при сравнительно небольшой средней излучаемой мощности (~ 10 мВт) высокий КПД согласования с оптическим волокном (потери согласования составляют несколько децибел). Однако необходимость темпера-

турной стабилизации, малый срок службы (10^6 ч для лабораторных и $\sim 5 \cdot 10^4$ ч для промышленных образцов) и высокая стоимость ставят серьезные препятствия на пути их широкого применения. Отметим, что нелинейность характеристик, затрудняющая использование аналоговой модуляции, обусловлена многомодовой структурой излучения. Применение одномодовых фильтров [29, 290] приводит к уменьшению нелинейных искажений.

Б. Передача

Оптические волокна бывают двух типов: со скачком показателя преломления (полоса пропускания до 35 МГц·км) и с градиентным (плавным) изменением показателя преломления (полоса пропускания свыше 300 МГц·км). Погонные потери на длине волны 0,84 мкм составляют от 3 до 10 дБ/км; для коротких линий связи (несколько десятков метров) могут применяться волокна с большими потерями (от 30 до 50 дБ/км). По мере уменьшения потерь в оптических волокнах и появления оптико-электронных элементов, позволяющих использовать окна с очень малыми потерями ($\sim 1,3$ мкм), *увеличиваются расстояния между ретрансляторами*. В гл. 8, § 2 приводится пример экспериментальной линии связи на 50 км, работающей на волне 1,25 мкм без ретранслятора [110]; отметим, что была также осуществлена связь на расстоянии 100 км с применением ИАГ-лазера, излучающего на волне 1,32 мкм [291]. Все говорит о том, что запросы потребителей растут: например, может оказаться, что в ближайшем будущем градиентное оптическое волокно с пропускной способностью 140 Мбит/с перестанет удовлетворять потребителя и потребуются серьезные разработки одномодовых волокон для обеспечения связи на большие расстояния [292, 293].

Что касается соединения оптических волокон, то наметилась тенденция (по крайней мере в системах массовой связи) применения сварных соединений (потери менее 0,3 дБ) при прокладке подземных линий и оптических разъемов (потери ~ 1 дБ) для подсоединения к внешнему оборудованию.

Необходимо отметить, что для передачи информации все меньше и меньше используются пучки оптических волокон, представляющие собой плотный жгут из десятков, а иногда и сотен волокон и передающих сигнал одновременно. Несмотря на их малую стоимость и легкость соединения, скромные технические характеристики (потери свыше 50 дБ/км и полоса пропускания менее 10 МГц·км) и хрупкость при одновременном снижении стоимости одиночных оптических волокон отодвигают их применение в область канализации света (освещение, волоконно-оптические лампы и т. д.).

В. Прием

В качестве фотоприемников применяются *PIN*-фотодиоды и лавинные фотодиоды. Они характеризуются увеличенной полосой пропускания (порядка гигагерц) и пренебрежимо малыми темновыми токами (менее 1 нА). Лавинный фотодиод труднее использовать, чем *PIN*-фотодиод (напряжение питания 200 В вместо 20 В, сильная зависимость коэффициента лавинного усиления от температуры), но, как показывают проведенные выше расчеты, у них более высокая чувствительность (приблизительно на 10 дБ при заданном значении отношения сигнала к шуму или коэффициента ошибки).

Сначала рассмотрим основы проектирования двухабонентной линии связи, а затем — проектирования линий связи, использующих более сложные структурные решения.

§ 3. Двухабонентная волоконно-оптическая линия связи

А. Основы проектирования

Процесс проектирования системы связи не есть жестко установленная логически последовательная цепочка операций определения ее элементов. Для полной оптимизации системы нужно проверить все возможные варианты решений — комбинации типов модуляторов, излучателя, световода и приемника — с учетом технических требований. Поэтому приходится не раз возвращаться назад и повторять расчеты, чтобы быть уверенным, что использовано все возможное [294].

В принципе проектирование можно начать с выбора любого из трех элементов — излучателя, световода или приемника — и проанализировать, какое влияние окажет выбор каждого из этих элементов на все другие элементы. Но обычно начинают с выбора приемника, причем здесь можно выделить следующие этапы.

1. *Берется самый простой вид модуляции или кодирования* (модуляция интенсивности аналоговым сигналом, код без возврата к нулю для цифрового сигнала).

2. *Выбирается самый простой приемник и рассчитывается минимальная мощность на его входе, обеспечивающая требуемое отношение сигнала к шуму или заданный коэффициент ошибки.* В зависимости от требований восстановления сигнала (предполагается или нет восстановление синхронизирующих импульсов) и контроля цифровой передачи (требуется ли контролировать качество передачи в линии) в случае необходимости следует изменить код передачи.

3. *Выбирается источник света с учетом полосы пропускания и требований надежности.* Если требуемый уровень нелинейных

искажений не может быть обеспечен при применении модуляции интенсивности, следует предусмотреть их коррекцию или сменить тип модуляции, но при этом необходимо повторить расчеты по п. 2.

4. *Рассчитываются допустимые потери в линии и полоса пропускания* и исходя из этого *выбирается тип оптического волокна*. Выбор оптического кабеля производится на основе компромисса между стоимостью, устойчивостью к внешним воздействиям при эксплуатации (температура, давление, вибрации и т. д.) и при прокладке (протягивание по трубам, прокладка в желобе). Кроме того, необходимо учитывать возможные дополнительные потери в соединениях отрезков кабеля между собой.

Если расчет дает слишком малые допустимые погонные потери, которые не могут быть обеспечены имеющимися световодами, то следует вернуться к выбору компонентов системы (применить лавинный фотодиод вместо *PIN*-фотодиода, ППЛ вместо СД) или изменить вид модуляции, затем повторить расчеты по п. 3 или по п. 2 и 3. Если, несмотря на это, потери в линии будут слишком большими, можно будет ввести один или несколько промежуточных ретрансляторов или, если речь идет об экспериментальной линии связи, сменить длину волны и вести передачу на длине волны, близкой к 1,3 мкм (в настоящее время из-за отсутствия необходимых элементов применять линии связи, работающие на длинах волн выше 1,3 мкм, можно только в лабораторных условиях [292]).

Если ширина полосы пропускания, приведенная к 1 км, не обеспечивается существующими в настоящее время оптическими волокнами, то следует либо рассмотреть возможность замены вида модуляции или компенсацию сужения полосы пропускания в приемнике [258] (при этом необходимо вновь вернуться к п. 2), либо, если линия связи экспериментальная, использовать одномодовые оптические волокна или применять компоненты, работающие на длине волны, близкой к 1,3 мкм (и снова вернуться к п. 2).

Отметим, что любой выбор, кажущийся в настоящее время оптимальным, через несколько лет может устареть из-за непрерывного совершенствования оптико-электронных элементов и повышения их надежности. Так, *переход от СД к ППЛ, от градиентного волокна к одномодовому, с длины волны 0,84 мкм на длину волны 1,3 мкм должен, по-видимому, изменить оптимальные структуры систем и дать новые возможности проектировщикам* [295].

Б. Баланс мощности в двухабонентной линии связи

Проведем для иллюстрации количественный расчет погонных потерь в линии, осуществляемый на этапе 4, в наиболее распространенном случае цифровой передачи данных.

На рис. 15.9 приведены минимальные мощности на входе приемника, рассчитанные в гл. 15, § 4 при заданном коэффициенте ошибки передачи 10^{-8} , для лавинных фотодиодов (p_m) и PIN-фотодиодов (p_{m0}). При использовании имеющихся в нашем распоряжении двух источников излучения (ППЛ и СД) ожидаемая мощность на входе приемника определяется из приведенного ниже баланса мощности:

	ППЛ		СД
Полная излучаемая мощность 10 мВт	10 дБм	1 мВт	0 дБм
Потери в соединениях	—4 дБ		—15 дБ
Потери в соединении с излучателем	—1 дБ		—1 дБ
Потери в соединении с приемником	—1 дБ		—1 дБ
Запас на компенсацию искажений	—6 дБ		—6 дБ
Потери в линии при передаче	— a дБ		— a дБ
Мощность на входе приемника	<u>—2 — a (дБм)</u>		<u>—23 — a (дБм)</u>

Приравнивая эти величины значениям p_m и p_{m0} (выраженным в дБм), получаем допустимые потери в линии связи:

ППЛ + лавинный фотодиод	$a = -2 - p_m$
ППЛ + PIN-фотодиод	$a = -2 - p_m$
СД + лавинный фотодиод	$a = -23 - p_{m0}$
СД + PIN-фотодиод	$a = -23 - p_m$

Соответствующие числовые данные приведены на рис. 17.2. С его помощью, зная длину линии связи, легко определить допустимое погонное затухание оптического кабеля и потери на соединения. Или же, наоборот, задавшись погонным затуханием, можно определить максимальную длину линии передачи без ретрансляторов.

Например, если взять оптический кабель с потерями 5 дБ/км на длине волны 0,84 мкм и со сварными соединениями (потери 0,3 дБ) через каждые 900 м, то можно определить максимальные расстояния между ретрансляторами для линий связи с различной информационной пропускной способностью из приводимой ниже таблицы. Для сравнения в ней также приводятся соответствующие максимальные расстояния для классических проводных линий связи, действующих в настоящее время.

Информационная пропускная способность, Мбит/с	2	8	34	140
Расстояние между ретрансляторами, км:				
Для волоконно-оптических линий связи	12	11	10	9
Для современных проводных линий связи	2	4	4	4

Возможность еще более увеличить длину линии связи без промежуточного усиления на длинах волн порядка 1,3 мкм

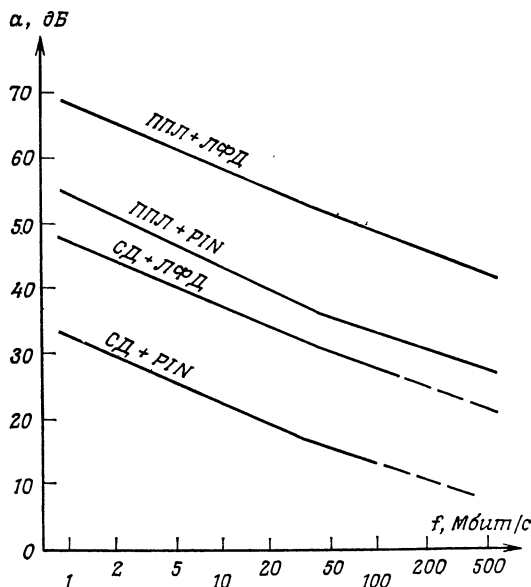


Рис. 17.2. Допустимое максимальное ослабление сигнала между излучателем и приемником цифровой линии связи (коэффициент ошибки $< 10^{-8}$).

и более [296] обеспечивает значительное преимущество волоконно-оптических линий связи. Чтобы проводимое сравнение было полным, остается провести экономический анализ различных систем и определить, в какой степени высокая стоимость оптических кабелей компенсируется уменьшением числа ретрансляторов.

В. Экономический анализ

Очевидно, что из всех возможных решений, удовлетворяющих техническим требованиям, предпочтение следует отдавать менее дорогостоящему. В настоящее время трудно определить типовые стоимости оптико-электронных элементов, поскольку речь идет не о промышленном, а экспериментальном производстве. Поэтому экономическое сравнение ВОЛС с проводными электрическими линиями связи, которое могло бы быть сделано на основании сегодняшних цен, не имело бы большого значения. Можно, однако, заглянуть в будущее и провести такое сравнение в рамках систем связи массового обслуживания (почта, телеграф, телефон). Для развития оптических систем связи подобного типа потребуется большое число одинаковых компонентов, и, следовательно, если учесть возможность их серийного производства, можно планировать снижение их стоимости.

Было проведено и опубликовано много экономических исследований, в которых предполагалось, что стоимость прокладки кабеля, а также ретрансляторов и внешнего оборудования оди-

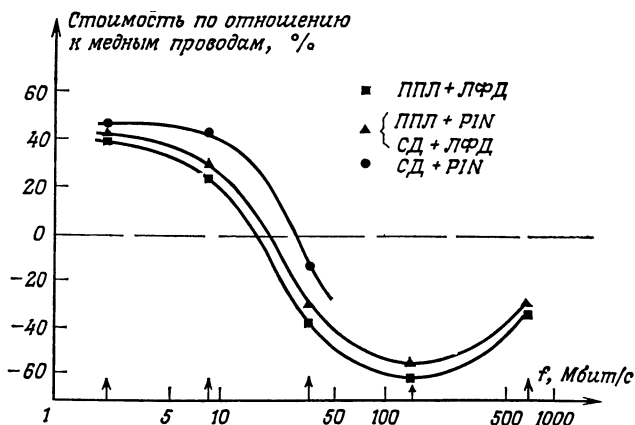


Рис. 17.3. Сравнение стоимости проводных и волоконно-оптических линий связи [298].

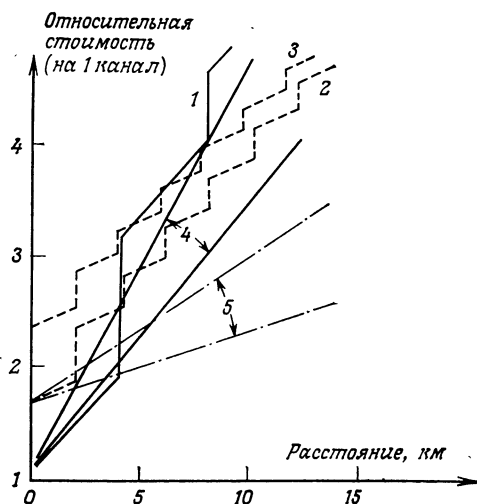


Рис. 17.4. Стоимость цифровых систем городской связи [299].

1 — 8 Мбит/с, коаксиальный кабель; 2 — 34 Мбит/с, коаксиальный кабель; 3 — 140 Мбит/с, коаксиальный кабель; 4 — 8 Мбит/с, волоконный кабель (2 кабеля по 24 волокна); 5 — 34 Мбит/с, волоконный кабель (1 кабель с 24 волокнами).

накова как для двухпроводных электрических линий, так и для ВОЛС. Самые важные результаты приведены на рис. 17.3 и 17.4. Эти зависимости были получены из расчета стоимости неизолированного оптического волокна, равной 0,5 франк/м (указанной

одной фирмой-изготовителем оптических волокон [297]). Расчетные данные рис. 17.3 [298] (кроме точек при 2 Мбит/с) относятся к системам междугородной связи, которые включают в себя следующие компоненты: 1 — ППЛ и лавинный фотодиод; 2 — ППЛ и *PIN*-фотодиод или СД с большой излучаемой мощностью и лавинный фотодиод; 3 — СД с большой излучаемой мощностью и *PIN*-фотодиод.

Как нетрудно видеть, для систем связи с пропускной способностью более 34 Мбит/с экономические преимущества ВОЛС несомненны. Имеющийся здесь запас экономической эффективности достаточно велик, так что этот вывод, вероятно, справедлив и в настоящее время, несмотря на повысившуюся с тех пор реальную стоимость оптического волокна.

Данные, представленные на рис. 17.4 [299], относятся исключительно к городским сетям связи. Они показывают, что ВОЛС с пропускной способностью 34 Мбит/с и длиной более 7 км экономически более выгодны. И наоборот, для расстояний менее 4 км более экономичными (с небольшим запасом) остаются системы на коаксиальных кабелях с пропускной способностью 8 Мбит/с. Отметим, что благодаря простоте прокладки оптических кабелей в городских условиях (меньшая глубина, малый вес) ВОЛС могут получить большой перевес над электрическими линиями связи гораздо раньше.

§ 4. Распределение данных¹⁾

Когда говорилось о технических условиях, задаваемых пользователем, мы отметили, что может возникнуть необходимость в системах связи с более сложной, нежели линейная, структурой (в виде кольца, многолучевой звезды и т. п.). Такие системы нужны в промышленности, где введение электрических магистральных линий и развитие информационной техники привели к большой гибкости в отношении обмена данными и их обработки. Поэтому важно, чтобы магистральные линии (шины) были выполнены на базе ВОЛС, не чувствительных к электромагнитным помехам.

Как показано на рис. 17.5, существуют две основные конфигурации подобных структур: *последовательная и многолучевая*; но возможны и другие варианты: кольцевая и комбинированная. Первоначально основу оптических линий связи представлял пучок оптических волокон [300]. Он имел преимущество в легкости выполнения оптических ответвлений и соединений (доста-

¹⁾ В многокомпьютерных системах сбора и обработки информации. — Прим. ред.

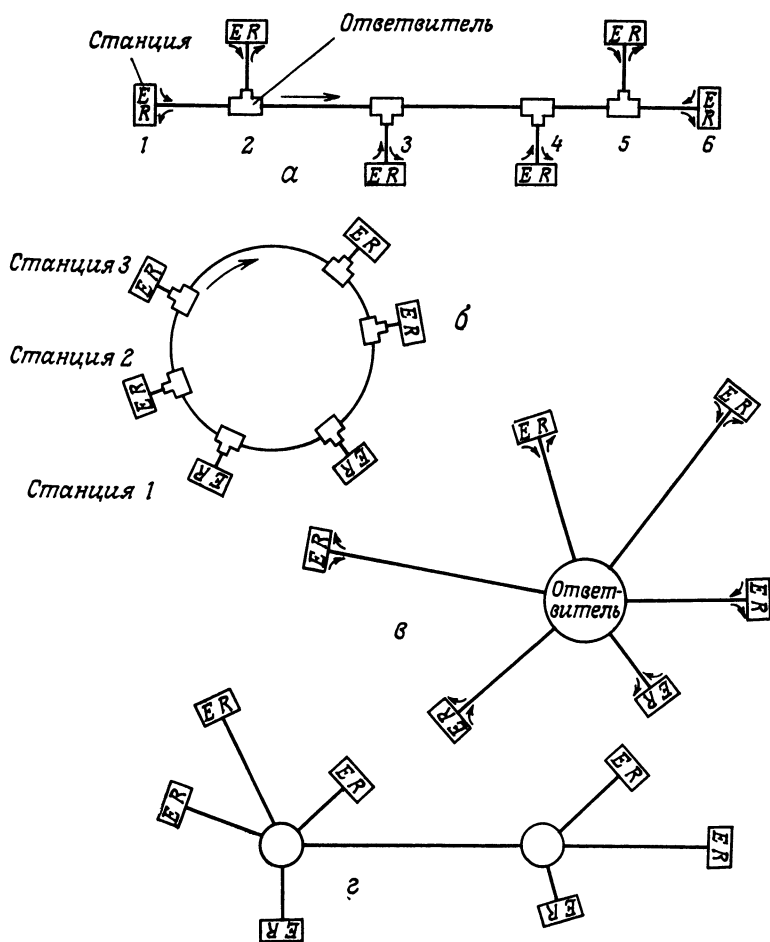


Рис. 17.5. Схемы распределения данных.

а — последовательная; *б* — кольцевая; *в* — многолучевая; *г* — комбинированная; *Е* — излучатель; *Р* — приемник.

точно извлечь оптическое волокно из общего пучка и направить его к пользователю), но обладал высокими потерями (гл. 17, § 2, п. Б). Чтобы устранить этот недостаток и расширить возможности оптических магистральных информационных линий (уменьшить число промежуточных ретрансляторов и увеличить расстояния передач), были исследованы одиночные оптические волокна. Эти исследования завершились разработкой ответвителей [301—303] и созданием экспериментальных систем связи [304, 305].

А. Последовательное распределение

Схема последовательного распределения данных наиболее близка к структуре электрических магистральных линий. Подключение станции к кабелю, по которому передается информация, осуществляется посредством Т-образного ответвителя или распределительной коробки. В зависимости от типа Т-образного ответвителя линия связи может быть одно- или двунаправленной. В первом случае не все станции могут обмениваться данными друг с другом (например, на рис. 17.5, *а* передача осуществляется от станции i к станции $i + k$). Только станции, расположенные после передающей станции, могут принять ее сообщение. Такая ситуация вполне приемлема, например, в случае кабельного вещания, когда станция 1 — передающий центр, а станции 2, 3, ... — приемные.

Вообще же говоря, требуется, чтобы каждая станция сети связи могла поддерживать связь с любой другой станцией. Это необходимо при автоматическом контроле за информацией, когда ЭВМ, находящаяся на станции 1, опрашивает одну за другой различные станции управления и контроля. Для этой цели можно организовать кольцевую структуру, соединив оба конца основного кабеля (при этом запрос передается от станции 1 к станции i , а ответ от станции i к станции 1 через станции $i + 1$, $i + 2$ и т. д.), или проложить второй оптический кабель (один для каждого направления передачи), или использовать двунаправленные ответвители.

1. Т-образные ответвители

Как показано на рис. 17.6, Т-образные ответвители могут быть активными. В этом случае они используются как ретрансляторы и позволяют обеспечивать большие расстояния между

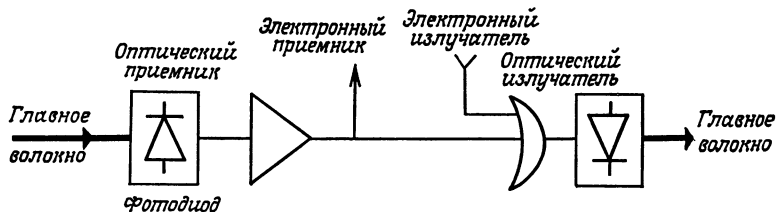


Рис. 17.6. Пример активного Т-образного ответвителя.

станциями. Отметим, что в принципе сложная система образуется путем объединения многих двухабонентных линий связи (от 1 к 2, от 2 к 3 и т. д.). Такая схема позволяет непосредственно расширять систему путем подключения новых станций. Однако *всякое нарушение в ответвителе* (отказ цепей питания или

усиления, проведение работ по техническому обслуживанию) прерывает передачу к следующим станциям. В основном по этой причине большинство заказчиков, а следовательно, и проектировщиков предпочитают применять пассивные Т-образные ответвители. Последние просто отводят к станции малую часть оптической мощности, распространяющейся по магистрали. Параллельное подключение станции с помощью оптических пассивных средств делает ее независимой от других.

Было исследовано много технических способов реализации оптических Т-образных ответвителей как на основе многомодовых оптических волокон [301—308], так и на основе одномодового волокна [309, 310]. Все они, кроме способа, описанного в работе [308], в котором применяются линзы с градиентом пока-

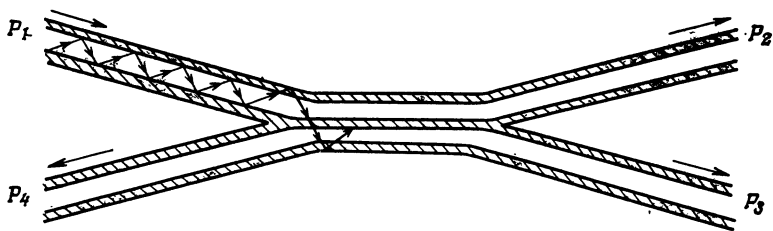


Рис. 17.7. Принцип действия пассивного Т-образного ответвителя.

зателя преломления, основаны на следующем принципе: создается локальная деформация основного оптического волокна с тем, чтобы возбудить в месте ответвления лучи утечки (гл. 8, § 3), которые улавливаются и канализируются к станции.

На рис. 17.7 представлен достаточно распространенный пример такого пассивного Т-образного ответвителя [301]. Основное оптическое волокно и волокно ответвителя прижимаются друг к другу и нагреваются кислородно-пропановой микрогорелкой. В результате оптическое волокно локально сплющивается (что обеспечит возбуждение лучей утечки), а оптическая оболочка расплавляется (что облегчает соединение). Расплавление оптических оболочек обеспечивает прочность соединения оптических волокон. Измеряя в процессе изготовления коэффициент ответвления в зависимости от приложенного давления и температуры нагрева, можно получить ответвители на 3, 10 и 20 дБ.

С учетом обозначений на рис. 17.7, где P_i — оптическая мощность, проходящая через сечение i , определим (в дБ):

$$\text{Коэффициент ответвления } C = 10 \lg \frac{P_1}{P_3}; \quad (17.1)$$

$$\text{Коэффициент направленного действия } D = 10 \lg \frac{P_1}{P_4};$$

во многих случаях применяется коэффициент направленного действия, равный 40 дБ, что позволяет пренебречь величиной P_4 по сравнению с P_1 ;

$$\text{Коэффициент потерь } P = 10 \lg \frac{P_1}{P_2 + P_3}, \quad (17.2)$$

который на практике получается достаточно малым — менее 1 дБ.

Такой Т-образный ответвитель может быть непосредственно применен в однонаправленной линии связи, а в сочетании с вторым Т-образным ответвителем на 3 дБ ($P_2 = P_3$) — в двунаправленной линии связи (рис. 17.8). На основании описанного основного элемента читатель сможет составить другие варианты подключения станций и другие структуры магистральных линий оптической связи.

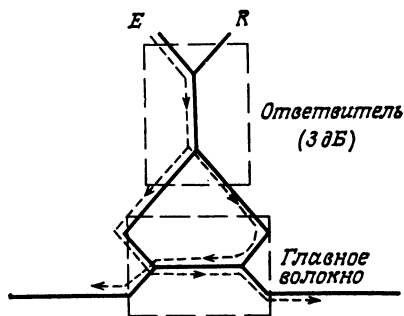


Рис. 17.8. Пример двунаправленного ответвителя.

Отметим, что ответвители, разработанные для интегральной оптики (гл. 14, § 3, п. В), смогут соперничать с описанными нами Т-образными ответвителями только тогда, когда будут снижены вносимые ими потери, и в тех случаях, когда требуется применение одномодовых волокон (например, в случае обмена данными при очень больших скоростях передачи информации).

2. Баланс мощности в линиях связи с последовательным распределением данных

Рассмотрим теперь характеристики систем, в которые входят Т-образные ответвители. Для этого проведем баланс мощностей в самом неблагоприятном случае, т. е. между самыми удаленными станциями: 1 и N . Поскольку нашей целью является выявление основных достоинств последовательной оптической магистрали связи, упростим расчет, ограничившись однонаправленной линией передачи (рис. 17.9). При этом переход к двунаправленным моделям (рис. 17.8) не представит трудностей (достаточно учесть ослабление за счет дополнительных ответвителей) и не изменит наших выводов.

Допустим, как это в общем и требуется, что для облегчения установки и обслуживания используются идентичные Т-образные ответвители, соединения которых с главным волокном, излучателями и приемниками характеризуются вносимыми потерями A_c . Допустим также, что оптические Т-образные ответвители

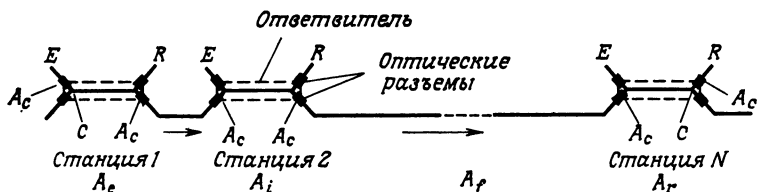


Рис. 17.9. Однонаправленная линейная оптическая система связи.

ли — это обратимые симметричные четырехполюсники. Учитывая обозначения на рис. 17.9, найдем потери (в децибелах):

а) между излучателем и главным волокном

$$A_e = A_c + C + A_c, \quad (17.3)$$

т. е. необходимо пройти входное и выходное контактные соединения и ответвитель;

б) между главным волокном и приемником

$$A_r = A_c + C + A_c; \quad (17.4)$$

в) при включении Т-образного ответвителя в главное волокно

$$A_i = A_c + 10 \lg \frac{P_1}{P_2} + A_c,$$

или с учетом (17.1) и (17.2)

$$A_i = 2A_c - 10 \lg \left(10^{-\frac{P}{10}} - 10^{-\frac{C}{10}} \right). \quad (17.5)$$

В результате получаем, что потери между излучателем станции 1 и приемником станции N равны

$$A_{1N} = A_e + (N - 2) A_i + A_r + A_f,$$

где A_f — полные потери в главном волокне. В первом приближении они пропорциональны расстоянию между двумя крайними точками магистральной линии. С учетом этого окончательно получаем

$$A_{1N} = 2NA_c + 2C - (N - 2) 10 \lg \left(10^{-\frac{P}{10}} - 10^{-\frac{C}{10}} \right) + A_f. \quad (17.6)$$

Величина A_{1N} является, с одной стороны, убывающей функцией переменной C за счет третьего члена в выражении (17.6),

а с другой — возрастающей функцией за счет члена $2C$. Следовательно, существует такое значение C , при котором максимальные потери минимальны. Оно определяется из условия

$$\frac{\partial A_{1N}}{\partial C} = - \frac{(N-2) 10^{-\frac{C}{10}}}{10^{-\frac{P}{10}} - 10^{-\frac{C}{10}}} + 2 = 0,$$

откуда

$$C = P + 10 \lg \frac{N}{2}. \quad (17.7)$$

Подставив этот результат в (17.6), получаем минимальные потери:

$$A_{\min} = N(2A_c + P) + 10 \lg \left(\frac{N^N}{(N-2)^{N-2}} \right) - 20 \lg 2 + A_f. \quad (17.8)$$

Отметим, что оптимальная величина C мало зависит от потерь в Т-образном ответвителе (поскольку P мало), а определяется главным образом общим числом станций. Минимальные же потери линейно зависят от P и A_c . Следовательно, эти величины необходимо уменьшать, особенно A_c . Однако основные потери обусловлены самим принципом ответвления [логарифмический член в выражении (17.8)], и эти потери не могут быть уменьшены. Отметим, что на практике число станций, заданное во время прокладки линии связи, не остается неизменным, так как в процессе эксплуатации часто бывает желательно расширить систему связи и подключить новые станции без модификации Т-образных ответвителей.

Для иллюстрации произведенных расчетов приведем графики зависимостей величины A_{\min} от числа станций для идеального случая $A_c = P = 0$ дБ, пренебрегая потерями в волокне A_f (что эквивалентно смещению кривой на рис. 17.10). Чтобы можно было судить количественно о важности уменьшения потерь в Т-образных ответвителях и в соединениях, мы приводим ту же зависимость при $A_c = P = 0,5$ дБ. Как и отмечалось, расстояние между кривыми тем больше, чем больше N . Чтобы лучше приблизиться к практическим условиям, мы приводим также кривую A_{\min} при тех же значениях $A_c = P = 0,5$ дБ, взяв коэффициент ответвления $C = 6$ дБ, не зависящий от N .

Эти результаты следует сравнить с максимально допустимыми потерями между излучателем и приемником, рассчитанными в § 3 и приведенными на рис. 17.2. Если ориентироваться на информационную пропускную способность 1 Мбит/с и на оптико-электронные компоненты, наиболее подходящие для промышленного применения в отношении надежности, простоты в работе и стоимости (§ 2, СД и PIN-фотодиоды), то потери

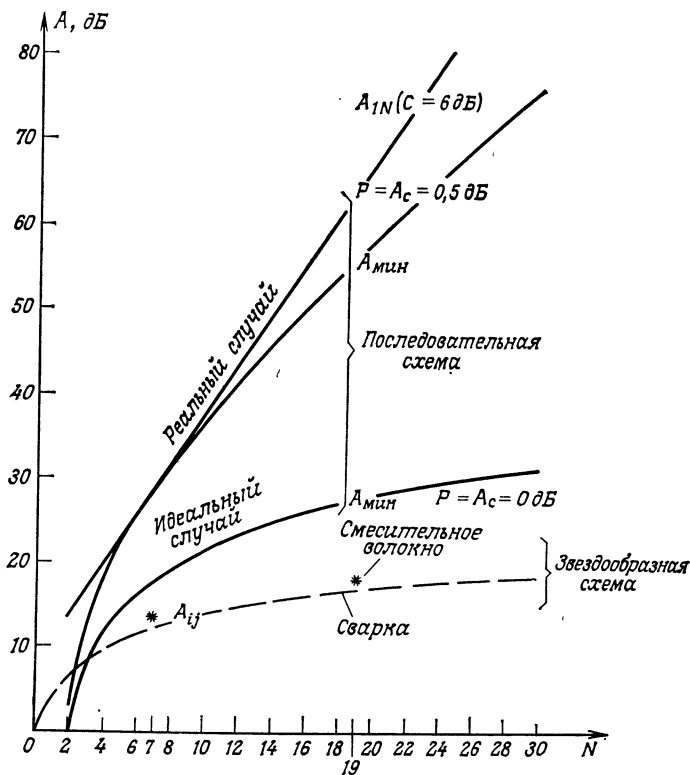


Рис. 17.10. Максимальные потери между двумя любыми станциями магистральной линии передачи данных в зависимости от общего числа станций. Потери в линии передачи не учитываются.

должны быть не более 34 дБ для обеспечения коэффициента ошибки не выше 10^{-8} . Это значит, что при $P=A_c=0,5$ дБ число станций не может превышать 8 (только 5 при $P=A_c=1$ дБ) и еще меньше, если потери в линии связи велики. Например, $N=6$, если $P=A_c=0,5$ дБ и $A_f=8$ дБ. При столь малом числе станций оптимизация C по N мало снижает потери и не оправдывает соответствующих трудностей реализации.

Чтобы значительно увеличить число станций при последовательной схеме распределения, можно в принципе заменить контактные соединения сварными (что, впрочем, дает снижение потерь лишь на 0,2 дБ; гл. 11, § 3, п. В) или использовать компоненты с лучшими характеристиками (ППЛ и лавинные фотодиоды).

Б. Многолучевое распределение

Другой способ увеличения числа абонентов при сохранении удовлетворительного качества передачи состоит в иной структуре системы связи, а именно в применении многолучевой схемы распределения данных. Как показано на рис. 17.5, б, такая система образуется при объединении абонентов в виде многолучевой звезды вокруг одного ответвителя или распределителя, который получает сигнал от любой станции и передает его на все другие станции [311].

1. Многолучевой ответвитель

Можно представить себе реализацию такого активного ответвителя с приемником и излучателем на основе активного Т-образного ответвителя, приведенного на рис. 17.6 [312]. Однако для снижения вероятности выхода его из строя обычно применяются пассивные устройства, одно из которых изображено на рис. 17.11 [313].

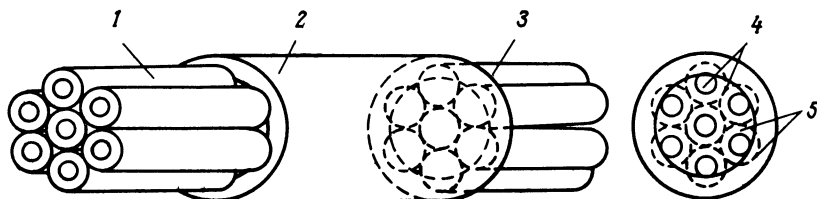


Рис. 17.11. Многолучевой ответвитель со смесительным световодом.

1 — входные волокна; 2 — смесительное волокно; 3 — выходные волокна; 4 — сердцевина; 5 — оптическая оболочка.

Световая энергия, излучаемая одной станцией, входит в смесительное волокно и после многократного отражения распределяется между разными приемными волокнами. Для уменьшения потерь распределения необходимо:

а) чтобы смесительное волокно имело ту же числовую апертуру, что и передающее волокно (этим обеспечивается полное внутреннее отражение лучей на поверхности раздела между сердцевиной и оболочкой);

б) чтобы была максимальной плотность упаковки передающих волокон (что возможно только при гексагональной упаковке — concentрическими кругами с 1, 6, 12, ... световодами [314]);

в) чтобы минимальная площадь, охватывающая сердцевинки передающих волокон, совпадала с сердцевиной смесительного волокна.

Предполагая равномерное распределение световой энергии на выходе смесителя (которое обеспечивается при длинах до

нескольких сантиметров), легко оценить потери между излучающим и любым приемным волокнами. Например, в случае ответвителя, состоящего из 7 волокон (рис. 17.11) с диаметром сердцевины 100 мкм и диаметром оболочки 140 мкм потери $A_{*}^{(7)}$ распределятся следующим образом:

Потери на распределение: $10 \lg 7 = 8,5$ дБ,

Потери на заполнение, определяемые как

$$10 \lg \left(\frac{\text{Площадь сечения сердцевины смесительного волокна}}{\sum_1^7 \text{Площадь сечения сердцевины передающего волокна}} \right),$$

с учетом указанных размеров составят

$$10 \lg \frac{\frac{\pi}{4} (380 \text{ мкм})^2}{7 \frac{\pi}{4} (100 \text{ мкм})^2} = 3,1 \text{ дБ}.$$

В результате получаем $A_{*}^{(7)} = 11,6$ дБ.

Нетрудно провести аналогичный расчет для 19 волокон: $A_{*}^{(19)} = 16,4$ дБ.

Отметим, что потери на заполнение пропорциональны отношению площадей сечения оболочек и сердцевин «смешиваемых» волокон. Следовательно, целесообразно использовать передающие волокна с малой толщиной оболочки. Так, например, волокна, применяемые для передачи оптических сигналов на большие расстояния и имеющие диаметры оболочки и сердцевин, равные 125 и 50 мкм, мало пригодны для этого и их использование приведет к увеличению потерь на 5 дБ по сравнению с полученными результатами.

Кроме больших потерь на заполнение, недостатком рассмотренного устройства является также *трудность его изготовления*. Необходимо, чтобы торцы волокон были плоскими, а положение передающих волокон строго определенным. Это приводит к необходимости в длительной подготовке концов волокон и в использовании очень точных приспособлений. Чтобы преодолеть эти затруднения и обеспечить промышленную реализацию многолучевого ответвителя, было предложено сваривать волокна методом, описанным в п. А. 1 и применяемым для изготовления Т-образных ответвителей [315—318]. В этом случае волокна скручиваются, вытягиваются и нагреваются. Происходит сплавление волокон в зоне сужения, которая и выполняет функции смесителя (рис. 7.12). Приведенные в работе [316] данные обнадещивают, так как полученные потери превышают расчетные ($10 \lg N$) только на 2—3 дБ.

Рассмотренные распределители со смесительным волокном и со сплавленными волокнами называются работающими на

пропускание: световая энергия входит с одной стороны и выходит с другой стороны. Это приводит к необходимости иметь на станции два волокна (одно для приема, другое для передачи) или один распределитель совместно с двумя Т-образными ответвителями на 3 дБ: один на станции, а другой соединен с многолучевым ответвителем. Этот Т-образный ответвитель можно убрать, если применить модель, работающую на отражение.

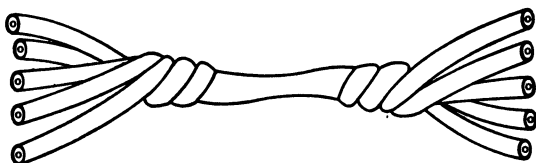


Рис. 17.12. Многолучевой ответвитель, полученный методом сплавления.

Для этого достаточно нанести зеркальный слой на выходной торец смесительного волокна, а в случае применения сплавленных волокон — разрезать зону скручивания посередине и нанести зеркальный слой на полученный торец [315]. Можно пойти дальше и исключить первый Т-образный ответвитель на станции, используя гибридный оптико-электронный элемент, разрабатываемый в настоящее время, который может работать либо как излучатель, либо как фотоприемник [319—321].

2. Баланс мощности многолучевого распределения

Расчет оптических потерь между двумя любыми станциями не сложен. При этом задача учета одно- или двунаправленности системы не возникает, так как все станции могут обмениваться информацией между собой. Важен лишь тип ответвителя. Чтобы сравнить количественно, рассмотрим распределители, работающие *на пропускание*, как наиболее легко осуществимые: со смесительным волокном и со сплавлением волокон. Получим следующее выражение для потерь:

$$A_{ij}(\text{дБ}) = 4A_c + A_*^{(N)} + A_f. \quad (17.9)$$

Оно показывает, что, если не считать потерь передачи A_f (от станции к распределителю и обратно), потери не зависят от номеров рассматриваемых станций i и j и незначительно возрастают с увеличением числа подключаемых станций. Используя равные отрезки оптических волокон, можно добиться, что потери A_f , а следовательно, и A_{ij} не будут зависеть от числа станций и можно будет унифицировать оборудование излучателей и приемников.

Чтобы проиллюстрировать наш расчет, положим $A_c = 0,5$ дБ, поскольку влияние этого члена предполагается малым. На

рис. 17.10 мы привели две точки ($N = 7$ и $N = 19$), соответствующие величинам, рассчитанным в случае применения смешительного волокна. Для модели распределителя со сплавлением волокон мы приняли, что

$$A_{\bullet}^{(N)} = 10 \lg N + 2,$$

$$\text{откуда } A_{ij} = 10 \lg N + 4,$$

что тоже (без учета A_f) представлено графически на рис. 17.10. Хотя полученные результаты носят теоретический характер и основываются на малом числе экспериментальных реализаций, они ясно показывают *преимущества многолучевого ответвителя* в отношении вносимых потерь. Даже с учетом потерь передачи (например, порядка 10 дБ) можно создать сеть связи с большим числом станций (например, 19), использующую только самые простые компоненты: светодиоды и PIN-фотодиоды.

В. Сравнение структур

Данные рис. 17.10 достаточно хорошо показывают решающее *преимущество многолучевой структуры* в плане потерь между станциями. Добавим к этому, что потери и время распространения между двумя станциями, соединенными через многолучевой распределитель, мало зависят от рассматриваемых станций. Это позволяет передающему знать, в какой момент времени сигнал принимается (что может быть важным при передаче распоряжений), и ограничить динамические характеристики приемников. В случае же последовательной схемы принимаемые оптические мощности, а следовательно, и уровни сигналов на входах приемников значительно отличаются от станции к станции. Например, (рис. 17.5, а), если станция 1 передает, то уровни мощности, принимаемые станциями 2 и 5 без учета потерь передачи, различаются по крайней мере на 20 дБ. Это вынуждает вводить блоки автоматической регулировки усиления в приемниках и усложняет изготовление оборудования.

Но у последовательной схемы имеются два важных преимущества.

1. *Модульность.* Можно добавлять Т-образные ответвители и станции без изменения существующей структуры системы в пределах, не нарушающих энергетического баланса.

2. *Требуемая длина оптического волокна.* Прокладка последовательной линии связи от одной станции к другой не встречает трудностей; с учетом иерархии станций это приводит к минимальному расходу оптического волокна. В виде упражнения читатель может сравнить необходимые длины оптических волокон

для последовательной и многолучевой схем системы связи, если абоненты расположены в вершинах квадрата, куба и т. д.

Очевидно, что решение использовать ту или иную структуру сети связи зависит от числа абонентов, их взаимного расположения и используемых компонентов. Применение полупроводникового лазера и лавинного фотодиода позволяет увеличить возможности последовательной структуры и предложить решения, которые хорошо согласуются с условиями прокладки кабеля (например, гибридная структура с многолучевыми распределителями там, где станции сконцентрированы, и с Т-образными ответвителями там, где они далеко разнесены друг от друга).

МУЛЬТИПЛЕКСИРОВАНИЕ И ПЕРЕДАЧА ИЗОБРАЖЕНИЙ

Описанные в предшествующих главах применения оптических волокон для целей связи незначительно отличаются от аналогичных применений обычных средств связи: симметричных и асимметричных кабелей, радиорелейных линий. Только носитель информации и концевые преобразователи (излучатель и фотоприемник) здесь не совсем обычны, но по своему характеру, а не по выполняемым функциям, состоящим в передаче электрического сигнала из точки А в точку Б.

Мы показали, что оптическое волокно вносит значительное упрощение в обеспечение связи и повышает ее качество, но при этом практически не оказывает влияния на средства обработки сигнала при передаче и приеме. Для кодирования и уплотнения речевых сигналов или для выделения из последовательности детектированных импульсных сигналов синхроимпульсов по-прежнему приходится обращаться к транзисторам, резисторам... Примером, свидетельствующим о том пути, который еще должна пройти оптика, может служить промежуточный ретранслятор. Он должен усиливать сигнал, в случае необходимости его восстанавливать, а также компенсировать потери в волокне. Для этого принятую оптическую волну необходимо зарегистрировать фотоприемником, усилить полученный электрический сигнал и вновь преобразовать его в оптическую волну при помощи излучателя. Конечно, при использовании оптических волокон уменьшается число ретрансляторов, но не упрощается их работа, поскольку прямого оптического усиления пока не существует. Первый шаг в этом направлении был сделан с появлением интегральной оптики, кратко рассмотренной в гл. 14

В данной главе, не претендуя на исчерпывающее изложение, мы попытаемся преодолеть узкие рамки использования оптики для передачи электрических сигналов и рассмотреть другие применения оптических систем в плане *прямой обработки оптических сигналов*. Мы покажем, как для временного уплотнения сигналов могут быть использованы линии задержки на оптических волокнах, как мультиплексирование по длинам волн позволит еще больше увеличить информационную пропускную способность ВОЛС. И наконец, мы рассмотрим возможности передачи изображений при помощи голограмм, а также возможности передачи изображений по одному волокну с использованием пространственного уплотнения и уплотнения по длинам волн.

§ 1. Линии задержки и временное мультиплексирование

Оптическое волокно обладает одним свойством, которое мы еще полностью не использовали, а именно свойством задерживать сигнал на время, пропорциональное своей длине. Действительно, в волокне со средним показателем преломления 1,5 время распространения оптического сигнала составляет 5 нс/м. Таким образом, соединяя последовательно многие километры оптического волокна, можно получить в небольших габаритах и при потерях в несколько децибел времени задержки, достигающие многих десятков микросекунд [322].

А. Оптическая линия задержки

Оптическая линия задержки имеет следующие преимущества перед акустической.

1. *Простота регулировки.* Достаточно обрезать волокно на расстоянии, соответствующем требуемой задержке, и при этом обеспечивается точность не ниже 1 нс.

2. *Постоянство характеристик в широком интервале температуры.*

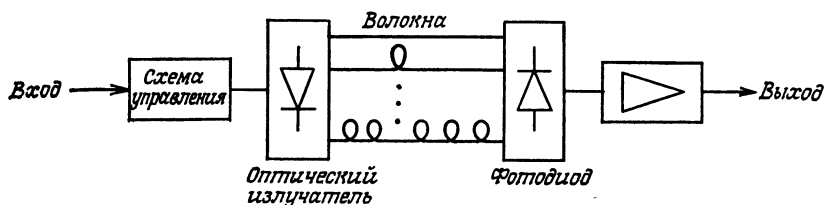


Рис. 18.1. Оптический гребенчатый фильтр.

3. *Независимость потерь от частоты сигнала.* Искажения отсутствуют, если ширина спектра сигнала меньше полосы пропускания оптического волокна. Это условие практически всегда выполняется для градиентных и одномодовых волокон, обеспечивающих полосу пропускания более 1 ГГц·км. В случае же акустических линий задержки потери пропорциональны корню квадратному из частоты, что приводит к искажению амплитуд и фаз задержки сигналов.

В области связи одно из применений линий задержки, а следовательно, и оптических волокон — изготовление гребенчатых фильтров [323]. Для синтеза требуемой передаточной функции гребенчатого фильтра применяется устройство, схематически представленное на рис. 18.1. В нем использованы только один оптический излучатель и один фотодиод. Оптические согласующие элементы позволяют равномерно освещать разные оптиче-

ские волокна и фокусировать полученные пучки света на фотоприемник.

Для иллюстрации найдем передаточную функцию фильтра, состоящего из N оптических волокон, длины которых нарастают в арифметической прогрессии: $l, 2l, \dots, Nl$, так что времена распространения сигнала в них равны $\tau, 2\tau, \dots, N\tau$. Пусть f — частота синусоидального сигнала, модулирующего излучатель. Тогда мощность, регистрируемая фотодиодом, будет равна

$$P = \sum_{n=1}^N \{p \cos [2\pi f(t + n\tau + \tau_0)] + p_m\},$$

где p_m — средняя мощность, p — амплитуда модуляции, связанная с p_m соотношением [формула (16.1)] $p = mp_m$ (m — коэффициент модуляции), а τ_0 — задержка, обусловленная оптическим излучателем и его схемой управления.

Путем несложных преобразований находим

$$P = p \frac{\sin N\pi f\tau}{\sin \pi f\tau} \cos \left[2\pi f \left(t + \tau_0 + \frac{N+1}{2} \tau \right) \right] + Np_m.$$

Поскольку в соответствии с (15.7) и (15.10) напряжение на выходе приемника пропорционально падающей на него мощности, то, пренебрегая задержкой $[\tau_0 + 1/2(N+1)\tau]$ и аддитивной константой, получаем для передаточной функции выражение

$$G(f) = \frac{1}{N} \frac{\sin N\pi f\tau}{\sin \pi f\tau},$$

приняв, что $G(0) = 1$.

Функция $G(f)$ максимальна и равна 1 при $f = k/\tau$ и минимальна при $f = (k + 1/2)/N\tau$. Таким образом, мы получили гребенчатый полосно-пропускающий фильтр, легко видоизменяемый путем изменения числа оптических волокон и их длины. Частный случай $N = 2$ соответствует дискриминатору.

Б. Временное мультиплексирование

Еще одно применение, может быть, более многообещающее, так как речь идет об оптическом способе осуществления того, что до сих пор осуществлялось только методами электроники, — *временное мультиплексирование* (уплотнение) сигналов. Если взглянуть на рис. 15.2, то мы увидим, что обработка и последующая передача цифровых последовательностей со скоростью 8, 34, 140 и т. д. Мбит/с производится после электронного мультиплексирования цифровых последовательностей, поступающих соответственно со скоростью 2, 8, 34 и т. д. Мбит/с. Первым шагом к созданию полностью оптической сети связи может стать мультиплексирование оптических сигналов по схеме, представленной на рис. 18.2.

В качестве примера рассмотрим четыре двоичные последовательности, формируемые отдельно со скоростью 2 Мбит/с импульсами длительностью 125 нс. Эти последовательности раздельно модулируют оптические излучатели и вводятся в оптические волокна длиной l , $l+25$, $l+50$ и $l+75$ м. В результате между этими последовательностями образуется временной сдвиг

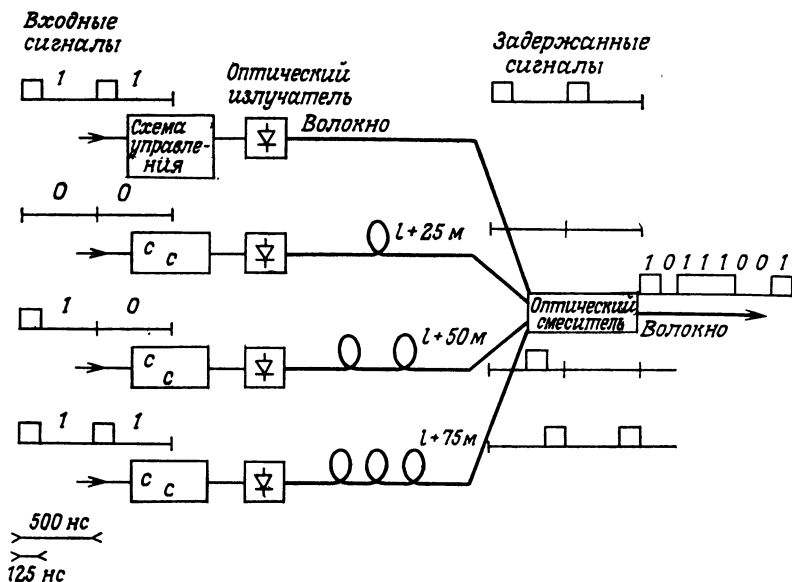


Рис. 18.2. Оптический временной мультиплексор. Справа: мультиплексированная последовательность 10111001.

в 125 нс. Если сложить оптические сигналы, например, при помощи смесительного оптического волокна, описанного в гл. 17, § 4, п. А, 1, то мы осуществим уплотнение четырех двоичных последовательностей и будем располагать оптическим сообщением с информационной емкостью 8 Мбит/с. Отметим, что теперь импульсы, составляющие последовательность 2 Мбит/с, имеют длительность до 500 нс (0,5 Мбит/с) и поэтому нужно будет уменьшать их длительность либо при их генерировании, либо позднее путем электронного или оптического стробирования.

Исходя из рассмотренного принципа мультиплексирования, читатель сможет сам составить другие схемы уплотнения сигналов, например уплотнения аналоговых выборок. Специалист по цифровым методам передачи информации сможет, добавив пятый канал (излучатель + волокно), ввести сигналы синхронизации.

§ 2. Мультиплексирование по длинам волн

В настоящее время для передачи нескольких сигналов по одному волокну мультиплексировуют электрические сигналы до передачи их в оптический излучатель. Мультиплексирование может быть временным (такое уплотнение, рассмотренное в предыдущем параграфе, хорошо подходит для цифровых сигналов) и частотным (главным образом для аналоговых сигналов). Такое уплотнение информации позволяет уменьшить число требуемых компонентов (источников излучения, светодиодов, фотоприемников), но может привести к неудобствам в других отношениях.

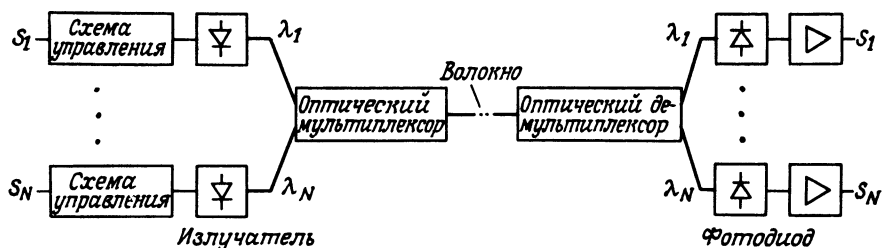


Рис. 18.3. Передача и прием с оптическим мультиплексированием по длинам волн.

1. С ростом числа мультиплексированных сигналов увеличивается оптическая мощность на входе приемника, необходимая для обеспечения нужного качества передачи (гл. 15, § 4; гл. 16, § 1), что может потребовать увеличения мощности излучателя, а это снижает надежность линии связи и в случае аналоговых сигналов приводит к увеличению нелинейных искажений, вносимых излучателем (гл. 16, § 1, п. А).

2. Мультиплексирование высокоинформативных сигналов, например при передаче цифровых последовательностей со скоростью 140 Мбит/с, приводит к очень высоким скоростям передачи данных, для чего требуются полосы пропускания оптических волокон, с трудом реализуемые в настоящее время.

3. Мультиплексирование различных по характеру сигналов, например аналоговых и цифровых, усложняет аппаратуру.

Чтобы уменьшить трудности, можно применить частичное электронное мультиплексирование и использовать отдельный излучатель для каждого из уплотненных сигналов. При этом оптические сигналы излучаются на разных длинах волн, которые затем смешиваются при помощи оптического мультиплексора и передаются по одному волокну. На приемной стороне демультиплексор длин волн осуществляет обратные операции и разделяет смешанные сигналы (рис. 18.3).

Чтобы надлежащим образом включаться в состав линии связи, система мультимплексор — демультимплексор должна удовлетворять определенным требованиям, связанным с условиями эксплуатации: а) малые потери (несколько децибел); б) низкие перекрестные помехи между каналами; в) малые размеры (геометрические размеры должны быть соизмеримы с размерами других оптоэлектронных компонентов); г) простота в эксплуатации (операции регулировки должны быть простыми и легко повторяемыми); д) и, естественно, низкая стоимость.

На основе этих требований были исследованы и экспериментально проверены различные принципы реализации такой системы [324—332]. Очевидно, что мультимплексор осуществить легче, чем демультимплексор, так как его задача сводится к простому смешению в одном световоде световых энергий, полученных от разных излучателей. В частности, для мультимплексирования двух волн были использованы оптические смесители, описанные в гл. 17, § 4, п. А, где не было речи о длине волны [327, 328]. Два излучающих оптических волокна малого диаметра были навиты вокруг волокна линии, как показано на рис. 17.12, и при этом вносимые потери составили ~ 1 дБ.

Проектирование демультимплексора основано на методах спектрального анализа (спектрографии) и состоит в использовании пассивных элементов с характеристиками, зависящими от длины волны. Как показано на рис. 18.4, такая характеристика может быть получена, например, за счет геометрического отклонения светового пучка с помощью призмы и дифракционной решетки [329, 330] или путем частотно-избирательного отражения от дихроичного фильтра [331]. В общем случае в качестве оптических элементов, необходимых для коллимирования светового пучка на входе и его фокусировки на выходе демультимплексора, применяются линзы с градиентом показателя преломления (гл. 4, § 3, п. Б), большим преимуществом которых являются малые аберрации [324]. Отметим, что описанные демультимплексоры могут быть использованы в качестве мультимплексоров, если изменить направление распространения света на противоположное. Это важное свойство, поскольку оно позволяет уменьшить число разрабатываемых элементов, что в конечном счете приводит к уменьшению стоимости их изготовления.

Существуют две основные задачи, решить которые позволяет мультимплексирование по длинам волн: *кабельное вещание* и *наращивание пропускной способности* уже проложенных линий связи.

Чтобы улучшить в будущем обслуживание абонента и скомпенсировать неединобразие передаваемых сообщений, можно передавать телефонный сигнал, один или несколько телевизионных сигналов и цифровые сигналы (видеотелефон, служебную

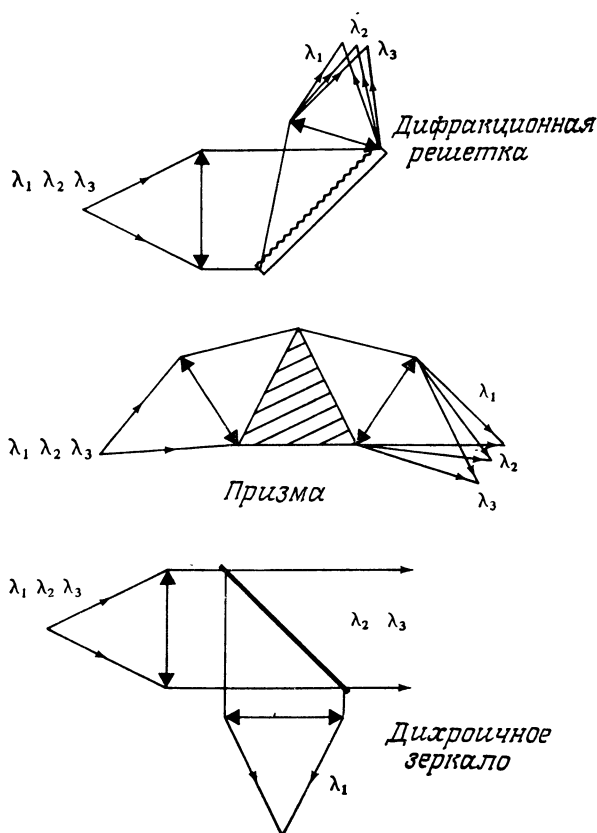


Рис. 18.4. Оптические демультиплексоры.

информацию, фототелеграфию...) по одному волокну (это уменьшит затраты), но на разных длинах волн (что облегчит эксплуатацию линии). Был проведен эксперимент по передаче информации по одному волокну одновременно на четырех длинах волн [332]. Три светодиода, излучающие на длинах волн 750, 830 и 900 нм, обеспечивали передачу в направлении центральный узел — абонент телефонных сигналов, первого и второго каналов телевидения. Четвертый светодиод, излучающий на длине волны 1060 нм, служил для передачи телефонных сигналов в направлении абонент — центральный узел.

Для удовлетворения возрастающих требований к средствам связи без замены уже проложенных оптических волоконных кабелей достаточно поместить на концах оптической линии связи комплект мультиплексор — демультиплексор и передавать дополнительные сигналы на другой длине волн. Это особенно

выгодно в тех случаях, когда прокладка нового кабеля связана с трудностями и требует больших затрат, например городские и подводные линии.

§ 3. Временная голография и передача изображений по оптическому волокну [26, 27, 38]

Идея использовать хроматический спектр белого света положена в основу оригинального метода обработки и передачи изображений, который существенно отличается от классических способов мультимплексирования по длинам волн.

В гл. 4, § 4, п. Д было показано, что дифракция полихроматического света на зрачке позволяет осуществлять очень интересную временную модуляцию света (рис. 4.26). Можно модулировать полихроматическую волну света изображением, а затем анализировать полученный сигнал одним из следующих способов:

- 1) при помощи интерферометра Майкельсона, который дает на выходе функцию автокорреляции передаваемого сигнала;
- 2) при помощи спектрометра, который дает на выходе энергетический спектр передаваемого сигнала

Сам сигнал получается новым методом — методом *временной голограммы*. Рассмотрим рис. 18.5 и сравним его с рис. 4.26. Мы видим, что добавлен «канал» немодулированного временного сигнала $r(t)$, который служит опорным сигналом. Зарегистрированный на выходе такой системы спектр представляет собой не что иное, как голограмму Фурье временного сигнала изображения $g(t)$. В действительности спектрометр регистрирует энергетический спектр суммарного сигнала

$$S(\omega) = |\mathcal{F}(g(t) + r(t))|^2.$$

Здесь мы имеем не что иное, как уравнение «голографиста» (гл. 5), но не в пространстве, как в случае фурье-голографии, а *во времени*. Другими словами, мы получили уравнение временной фурье-голограммы. Теперь предположим, что этот спектр $S(\omega)$ регистрируется на оптическом носителе. Остается восстановить информацию об «изображении» $g(t)$ так же, как мы восстанавливали бы обычную фурье-голограмму (в данном случае одномерную).

Такой метод обеспечивает высокую скорость временной дискретизации. Кроме того, он не требует специального канала связи.

Но в настоящее время многие вопросы еще находятся в стадии исследований, цель которых получить оптическое устройство, действительно способное передавать изображения. Речь идет о решении таких проблем, как а) двумерное кодирование;

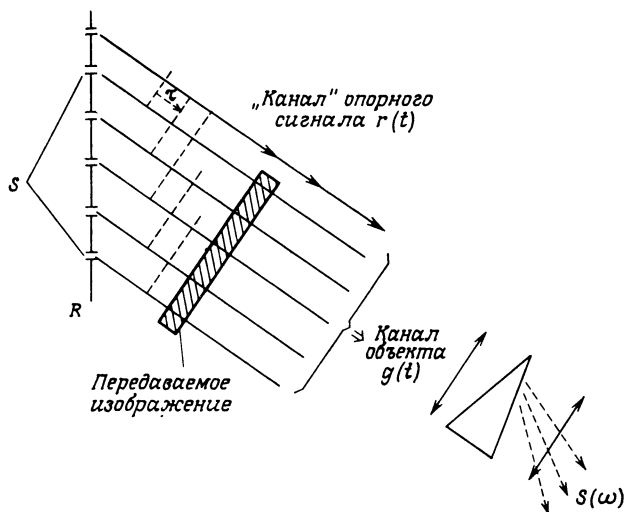


Рис. 18.5. Получение временной голограммы Фурье [26].

Спектрометр регистрирует квадрат модуля фурье-образа временного «объекта» $g(t)$, к которому добавляется «опорный» сигнал, приходящий по свободному «каналу» (не перекрывающему передаваемое изображение). Регистрация спектра $S(\omega)$ на оптическом носителе позволяет затем восстанавливать изображение $g(t)$ методом классической фурье-голографии.

б) регистрация сигнала $S(\omega)$; в) запись и восстановление временной голограммы в реальном времени.

И только будущее раскроет (как мы надеемся) все возможности этого действительно нового метода передачи изображений.

§ 4. Пространственное мультиплексирование [333]

Рассмотрим оптическое волокно со скачком показателя преломления (показатель преломления сердцевины n_1 , показатель преломления оболочки n_2), размеры которого мы полагаем такими, что к нему применима теория геометрической оптики. В гл. 8, § 3 было показано, что всей совокупности лучей, соответствующих плоской волне, падающей на входной торец волокна под углом Θ к его оси, отвечает одно значение параметра β . Но каждая точка входного торца волокна характеризуется своим значением параметра ν , и траектория луча представляет собой ломаную линию, «касающуюся» цилиндра внутри волокна (рис. 18.6). В результате многократного отражения в волокне луч может попасть в любую точку его выходного торца, но входящие лучи будут составлять с осью волокна угол Θ . Совокупность этих лучей сформирует в фокальной плоскости линзы L' с фокусным расстоянием F' окружность

радиусом $R = F' \operatorname{tg} \Theta$ (рис. 18.7). Таким образом, двум углам падения Θ_1 и Θ_2 соответствуют две окружности радиусами R_1 и R_2 , которые можно будет разделить, если они отстоят друг от друга на расстояние, большее, чем дифракционное пятно, диаметр которого определяется конечными размерами выходной апертуры волокна (каждая окружность-изображение имеет некую «ширину», обусловленную дифракцией).

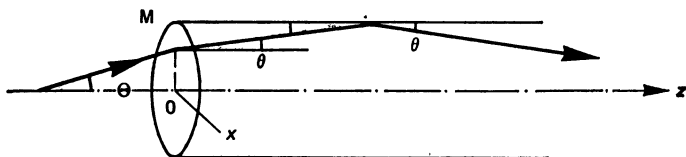


Рис. 18.6. Распространение луча в оптическом волокне со скачком показателя преломления.

В волокне со скачком показателя преломления все лучи, соответствующие плоской волне, падающей под углом Θ , распространяются с одинаковой постоянной распространения $\beta = k \sqrt{n_1^2 - \sin^2 \Theta}$ (внутренний угол θ определяется законом преломления $\sin \theta = n \sin \Theta$).

Из изложенного ясно, что по одному волокну можно передавать линейное световое распределение. На рис. 18.7 изображена оптическая схема, основанная на данном принципе. На оси Ox расположена полущель, линза L формирует соответ-



Рис. 18.7. Передача изображения полущели.

вующую плоскую волну, которая после прохождения через волокно с помощью линзы L' преобразуется в окружность радиусом

$$R = F' \operatorname{tg} \Theta = \frac{F'}{F} OA_0.$$

Если в фокальную плоскость линзы L' поместить щель, проходящую через центр окружности O' , то можно выделить две симметричные светящиеся точки окружности. В результате будет восстановлено первичное распределение (и симметричное ему) с масштабным коэффициентом F'/F .

Тем самым при помощи одного оптического волокна осуществлены мультиплексирование и передача сигналов, разнесенных в пространстве.

§ 5. Оптическое волокно и передача изображений

Изложенный принцип нетрудно применить для передачи двумерного изображения. Для этого достаточно перемещать изображение перед входной щелью и синхронно сканировать выходной щелью экран, на котором создается переданное по волокну изображение.

Возможны также другие методы, без механического сканирования изображения. Наиболее известен метод эндоскопии,

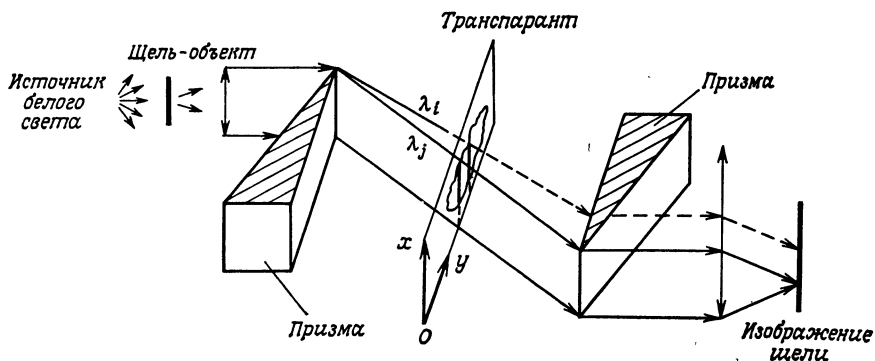


Рис. 18.8. Разложение изображения.

широко применяемый в медицине [334]. Принцип действия эндоскопа прост: по волокну передается изображение «точечного» объекта, но качество изображения и разрешающая способность невысокие. Другие, менее распространенные методы описаны в работах [335—337]. Читатель может сам судить об их оригинальности и о трудности их практического применения.

Недавно был предложен метод [338, 339], интересный как по своему принципу (совместно используются пространственное мультиплексирование и мультиплексирование по длинам волн), так и по простоте технической реализации. Его идея весьма проста (рис. 18.8). Тонкая узкая щель освещается белым светом, и через диспергирующий элемент (призму) проектируется объективом на передаваемое изображение. Такая оптическая система формирует на изображении цветные параллельные полосы, которые проходят через полупрозрачное изображение и модулируются распределением пропускания по строке этого изображения. Вторая проекционная система (призма и объектив) восстанавливает цветное изображение щели; остается

задача передачи изображения щели, которая была рассмотрена ранее (рис. 18.7).

Таким образом, два измерения изображения здесь передаются по-разному:

1) распределение пропускания (яркость) по оси Ox передается за счет углового разделения (пространственное мультиплексирование);

2) распределение пропускания по оси Oy передается за счет того, что каждая точка освещается своей длиной волны (λ -кодирование).

В этом методе принцип одномерного хроматического кодирования сочетается с принципом голографического временного кодирования (гл. 18, § 3) [340].

Все примеры этой и предшествующих глав иллюстрируют многочисленные возможности оптических методов и показывают их способность передавать и многие другие сигналы, отличные от уже существующих в электрической форме [341].

Диапазон применения оптических волокон широк, поскольку они обеспечивают возможность передавать на большие расстояния сигналы, мультиплексированные по времени, по частоте, по длинам волн и в пространстве. Развитие описанных в этой главе применений оптики, несомненно, будет следовать за развитием оптических линий дальней связи, являющихся подлинным локомотивом развития техники оптических волокон, и даже ускорять их развитие [361].

**ПУТИ РАЗВИТИЯ СИСТЕМ
ОБРАБОТКИ ИНФОРМАЦИИ,
КОМБИНИРОВАННЫЕ
ОПТИКО-ЭЛЕКТРОННЫЕ СИСТЕМЫ**

Волоконно-оптические системы *передачи* информации покинули область будущего, чтобы занять свое место в современной дальней связи. Для многих они символ самой передовой техники, которая служит одновременно и орудием, и движущей силой современного общества. Вот неопределенность понятия техники! Мы видели, что на самом деле это понятие охватывает не только овладение конечным продуктом, его изготовлением и его характеристиками, но также и познание физических законов, экономических условий и полное понимание нужд общества. Неважно, каковы будут пути к достижению успеха, возможно, они будут прослежены в других работах. Мы же просто подчеркнем два фактора, которые кажутся нам первостепенными и пример которых может способствовать развитию других новых методов:

1) *полная адекватность техники задаче, для которой она создавалась;*

2) *ее совместимость с существующими структурами и методами.*

И поскольку в этой книге мы пришли от проблем передачи информации к задачам обработки изображений, попробуем приложить эти два принципа эволюции методов и техники именно здесь, к новой области оптической *обработки* изображений. Следует признать, что если в первом пункте мы имеем, безусловно, основное преимущество оптических методов обработки изображений, то во втором остается много пробелов: несовместимость систем ввода с классическими носителями изображений (телевизионное изображение на экране ЭЛТ, фотография на бумаге и т. д.); несовместимость систем вывода с традиционными средствами хранения и обработки информации; несовместимость быстрого диалога между устройствами контроля (управления) информацией и устройствами принятия решения по результатам обработки.

Эта несовместимость объясняется теми существенными различиями, которые (как мы уже подчеркивали) имеются между оптическими методами обработки информации (параллельными и двумерными по своей природе), с одной стороны, и электронными цифровыми и аналоговыми методами (последовательными по своему принципу)—с другой. Оптику легко может показаться, что добиваться совместимости оптических

и электронных методов обработки изображений — значит стремиться к ограничению его возможностей, хотя на самом деле это открывает ему целый мир новых приложений.

Ниже мы рассмотрим несколько таких систем, которые в полной мере можно назвать *комбинированными*¹⁾. Речь здесь пойдет о том, чтобы разумно выделить одно или несколько электронных звеньев, которые, будучи включенными в канал оптической обработки, обеспечат либо лучшую интеграцию с электронными методами в глобальной системе передачи — обработка изображений, либо большую гибкость использования оптических методов.

Примеры, которые мы приведем, будут носить лишь предварительный характер, поскольку данное направление — самое новое в области обработки изображений. Бурное развитие некоторых типов электронных компонентов (запоминающих устройств, матриц, приборов с зарядовой связью и т. д.), так же как и прогресс оптоэлектроники (жидкие кристаллы, оптические пространственные модуляторы света) не позволяют определять тенденции по «состоянию в момент времени t ». И потому — у случая свои шансы! Многое еще предстоит найти и, конечно, в таких направлениях, которые мы еще себе даже не представляем.

Но сначала кратко остановимся на электронных методах обработки изображений, которые существуют сегодня.

§ 1. Электронные методы обработки изображений

Такие методы можно разделить на три большие группы: *цифровые, телевизионные и аналоговые*. Нам кажется логичным разделение двух последних методов, так как телевизионные методы основаны на применении очень специфических видов оборудования и техники, полностью совместимых друг с другом, что выделяет их в особую группу.

А. Цифровые методы

Нет необходимости еще раз подчеркивать большие возможности цифровых вычислительных машин при обработке изображений. Перечислим лишь их главные преимущества: 1) высокое качество обработки; 2) надежность и воспроизводимость результатов; 3) гибкость воздействия на параметры обработки; 4) широкий круг возможных видов обработки, а также их главные недостатки: 1) сравнительно высокая стоимость элементарной операции по обработке (которая, впрочем, постоянно снижается), 2) невозможность обработки изображений в реальном времени из-за огромного объема содержащейся в них инфор-

¹⁾ Или гибридными. — *Прим ред.*

мации; 3) сложность систем ввода и вывода изображений (поскольку при любой обработке изображение необходимо предварительно дискретизовать).

Ранее для обработки изображений использовали большие ЭВМ, в основном из-за их способности обрабатывать большие объемы информации за очень короткое время. Но потом их во многих случаях вытеснили специализированные мини-ЭВМ благодаря схеме с центральным процессором, в которой малый центральный процессор управляет огромными объемами памяти. Наконец, с появлением микропроцессоров устройства для обработки изображений будут ориентированы в новом направлении, к новым структурам, которые будут обеспечивать параллельную обработку на основе очень специальных алгоритмов благодаря применению системы малых центральных процессоров. Но пока что длительность их рабочего цикла часто слишком велика, чтобы они могли быть действительно конкурентоспособными.

Б. Телевизионные методы

Это методы обработки видеосигнала с выхода передающей телевизионной камеры в реальном масштабе времени при помощи разного рода аналоговых фильтров. Помимо частотной селекции, осуществляются главным образом операции по локальной нелинейной коррекции, по микшированию, коррекции временных задержек сигнала, а также некоторые эффекты, специфические для видеозвуковой техники, например инкрустацию в цветовом пространстве. Как и оптические методы, они кажутся прекрасно приспособленными для обработки изображений, но тем не менее не лишены значительных недостатков: 1) они только одномерные; 2) фильтры на видеочастоты (>1 МГц) оказываются трудноосуществимыми; 3) качество сигнала ограничено принятыми телевизионными стандартами.

Несмотря на свои недостатки, такие методы весьма ценны, поскольку: 1) вся техника апробирована и выпускается промышленностью (передающие камеры, видеоконтрольные устройства, линии задержки, видеоманитофоны, аппаратура управления и т. д.); 2) это методы обработки в реальном времени, хорошо приспособленные к работе в обычной обстановке, к каналам передачи и к наблюдению человеком. Наличие телевизионных стандартов очень облегчает и упрощает компоновку систем.

В. Аналоговые методы обработки информации

Аналоговая электронная техника предлагает нам в большей степени элементную базу (приборы и устройства), нежели методы обработки изображения. Говоря об элементной базе, нужно

иметь в виду всю оптоэлектронику: это интегральные линейки и матрицы фотоприемников и излучателей света, матрицы на жидких кристаллах, акустооптические и электрооптические модуляторы света и дефлекторы. Подчеркнем также ценность приборов с зарядовой связью (ПЗС), которые могут выполнять функции временной задержки, памяти или аналогового коррелятора на выходе оптического преобразователя.

Таблица 19.1

Методы обработки изображений	Преимущества	Недостатки
Оптические когерентные	<ol style="list-style-type: none"> 1. Возможность осуществления преобразования Фурье над комплексными функциями 2. Прямой доступ к частотной плоскости 	<ol style="list-style-type: none"> 1. Низкое качество изображения (спеклы) 2. Необходимость использования когерентного освещения (диапозитивов)
Оптические некогерентные	<ol style="list-style-type: none"> 1. Хорошее качество изображения 2. Простое освещение 	<ol style="list-style-type: none"> 1. Положительные передаточные функции 2. Сложность осуществления вычитания изображений
Телевизионные электронные	<ol style="list-style-type: none"> 1. Наличие хорошо отработанной техники 2. Отсутствие проблем, обусловленных носителем информации 3. Обработка в реальном времени 	<ol style="list-style-type: none"> 1. Ограниченные разрешение и качество изображений 2. Сложность обработки двумерных сигналов
Цифровые электронные	<ol style="list-style-type: none"> 1. Высокое качество изображений 2. Большая гибкость обработки 3. Универсальность 	<ol style="list-style-type: none"> 1. Высокая стоимость 2. Трудности ввода — вывода изображений 3. Трудности реализации обработки в реальном времени
Аналоговые электронные	<ol style="list-style-type: none"> 1. Быстрое развитие техники и элементной базы 	<ol style="list-style-type: none"> 1. Отсутствие готовых систем 2. Сложность обработки двумерных массивов

В табл. 19.1 мы сопоставляем основные преимущества и недостатки всех упомянутых выше методов обработки изображений. Как нетрудно видеть из таблицы, ни один из методов нельзя считать универсальным. Они скорее дополняют друг друга. В этой последней главе мы предполагаем привести примеры приложений, в которых недостатки одного метода ком-

пенсируются преимуществами другого. Мы будем идти в порядке убывающей сложности методов обработки: сначала рассмотрим когерентную пространственную фильтрацию, затем некогерентную, далее специализированные вычислительные устройства, которые предназначены для решения одной-единственной задачи, но задачи высокой сложности, наконец арифметические и логические блоки, которые совершенно естественно вернут нас к проблемам передачи информации, введя в привлекательный мир коммутации.

§ 2. Комбинированная система пространственной фильтрации в реальном времени

В гл. 6 мы подчеркнули те трудности, которые встречаются при необходимости быстрого синтеза пространственных фильтров в когерентной оптической системе пространственной фильтрации. Мы также подчеркивали важное значение человека-наблюдателя при оценке качества обработки, обусловленное тем, что невозможно определить объективный критерий качества изображения, который учитывал бы сложность физиологии зрения и процесса наблюдения человеком. В этом отношении интересен метод обработки, позволяющий *почти мгновенно синтезировать требуемый пространственный фильтр непосредственно в частотной плоскости оптической системы*. Он позволяет наблюдателю интерактивно подбирать оптимальные параметры обработки: характеристики искажения обрабатываемых изображений, отношение сигнала к шуму, используемую полосу пропускания и т. д.

Такая система должна использовать ЭВМ для расчета фильтра, оптоэлектронику для его синтеза и, разумеется, оптику для реализации самой процедуры обработки. Сложной частью этой системы является комбинированный оптико-электронный узел, содержащий реверсивные среды, допускающие многократную запись и стирание, причем управление процессом записи фильтра осуществляется посредством ЭВМ с достаточно высоким быстродействием. Сначала мы его назовем просто «фильтр»; немного дальше мы увидим, какие компоненты могут быть использованы для его реализации. Тогда структурная схема системы будет довольно проста (рис. 19.1). ЭВМ должна обеспечивать выполнение операций высокой сложности: например, двумерного преобразования Фурье. ЗУ согласует информационную пропускную способность вычислительного устройства и терминала. Его емкость в значительной степени зависит от характеристик фильтра, который надо реализовать: от нескольких сотен байтов в случае простой одномерной фильтрации до ста тысяч байтов в более сложных случаях.

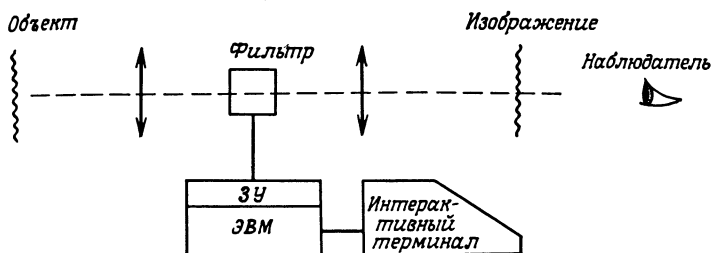


Рис. 19.1. Схема оптической системы пространственной фильтрации, работающей в реальном времени.

Теперь посмотрим, как можно осуществить «фильтр». Решения могут быть различными, и мы представим лишь два из них, которые кажутся нам наиболее типичными; другие представляют собой их варианты.

А. Комбинированные системы с оптическим управлением [342-344]

Различные регистрирующие материалы ведут себя по-разному в зависимости от длины волны излучения, а также от условий их использования (приложение 5). Например, они

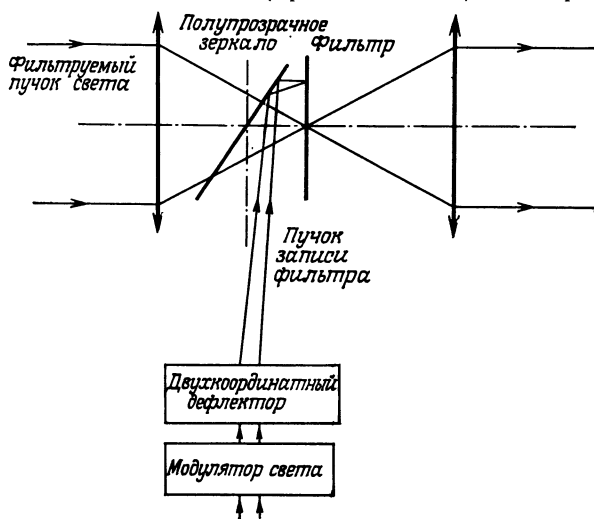


Рис. 19.2. Запись пространственного фильтра оптическим методом.

позволяют сначала записать информацию при помощи одного светового пучка, а затем, на втором этапе, при помощи другого светового пучка считать эту информацию. Эти два этапа представляют собой цикл запись — считывание. Можно стереть

записанную информацию и начать новый цикл. К таким реверсивным фоточувствительным материалам относятся фототермопластики; к ним также можно отнести пространственные модуляторы света типа Phototitus и Prom.

На рис. 19.2 показано, каким образом на таких материалах может быть реализована *оптическая система пространственной фильтрации, работающая в реальном времени*. Фильтрация осуществляется при помощи гелий-неонового лазера, а запись пространственного фильтра — при помощи аргонового лазера (гл. 12, § 2, п. В). Его луч сфокусирован, и производится точечная запись фильтра. Два дефлектора, управляемые ЭВМ, задают необходимые позиции сфокусированного пучка, а модулятор управляет энергией, подводимой к каждой точке. Управляющие логические схемы, свои в каждой подсистеме, обеспечивают синхронизацию этапов подготовки носителя к записи, самой записи, считывания и стирания.

Б. Комбинированные системы с электронным управлением

Мы только что видели, как можно осуществить запись фильтра оптическими средствами. Но в некоторых системах используется электронное управление. Таковы, в частности, системы с жидкими кристаллами, в которых пространственный фильтр формируется путем подачи напряжения на прозрачные электроды. Такое решение значительно упрощает оптическую схему пространственной фильтрации, но проблема усложняется на уровне коммутации сигналов при синтезе фильтра, поскольку необходимо $\sim 10^5$ ячеек с независимой адресацией, чтобы обеспечить необходимое для хорошей фильтрации число точек фильтра; две ортогональные гребенки прозрачных электродов часто позволяют решить проблему адресации.

§ 3. Комбинированная оптико-телевизионная система

Мы знаем, что когерентная пространственная фильтрация позволяет реализовать в одной операции самые сложные виды обработки; но мы также отметили, что эти возможности ограничиваются требованиями к качеству выполнения обработки (гл. 6, § 4). Поэтому здесь мы опишем две системы, в которых используется некогерентная пространственная фильтрация, связанная с использованием телевизионных средств. В первой из них задача электронных методов обработки сведена к нескольким существенным, но простым функциям. Во второй (§ 4) телевидение играет более важную роль, образуя цепь обратной связи.

Прежде всего посмотрим, каким образом можно наиболее рационально использовать телевизионную передающую камеру, чтобы обеспечить двухзрачковую фильтрацию при синтезе биполярных фильтров (гл. 7, § 2, п. Б) или трехзрачковую фильтрацию при синтезе комплексных фильтров. На рис. 19.3 показан принцип обработки изображений таким методом. Две телевизионные камеры, работающие синхронно, направлены на один и тот же объект; в плоскости зрачка каждой из них установлен фильтр. Один фильтр дает возможность реализовать по-

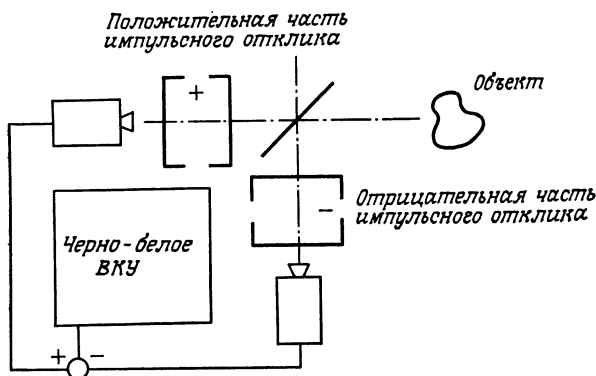


Рис. 19.3. Комбинированная оптико-телевизионная пространственная фильтрация изображений.

Вычитание изображений на выходе двух каналов обработки производится в форме видеосигналов.

ложительную часть импульсного отклика синтезируемой операции фильтрации, а другой — его отрицательную выпрямленную часть. В гл. 7, § 2, п. Б было показано, что для реализации требуемой пространственной фильтрации нужно из изображения, формируемого первой камерой, вычесть изображение с выхода второй камеры. Такая комбинированная система интересна простотой реализации операции вычитания изображений — путем вычитания видеосигналов. У нее следующие преимущества: а) пространственная фильтрация осуществляется в реальном времени без перезаписи информации на особый носитель; б) возможна фильтрация непосредственно трехмерных живых сцен; в) создается большое удобство использования благодаря стандартизации и наличию промышленного телевизионного оборудования; г) легко вводить локальную нелинейную фильтрацию благодаря имеющимся в цепи телевизионного канала регуляторам (управление усилением, линейностью, порогом) и благодаря простоте введения дополнительных требуемых для обработки компонентов (линий задержки, аналоговых фильтров

и т. д.); результаты обработки представляются в форме, очень удобной для наблюдения человеком.

На рис. 19.4 представлены два варианта реализации схемы, изображенной на рис. 19.3. Они дают возможность решить проблемы юстировки и синхронизации камер, но затрудняют для одной из них (а) обработку цветных объектов, а для другой (б) — обработку сцен в реальном времени.

Оптико-телевизионные методы нашли различные приложения при обработке изображений и распознавании образов [345].

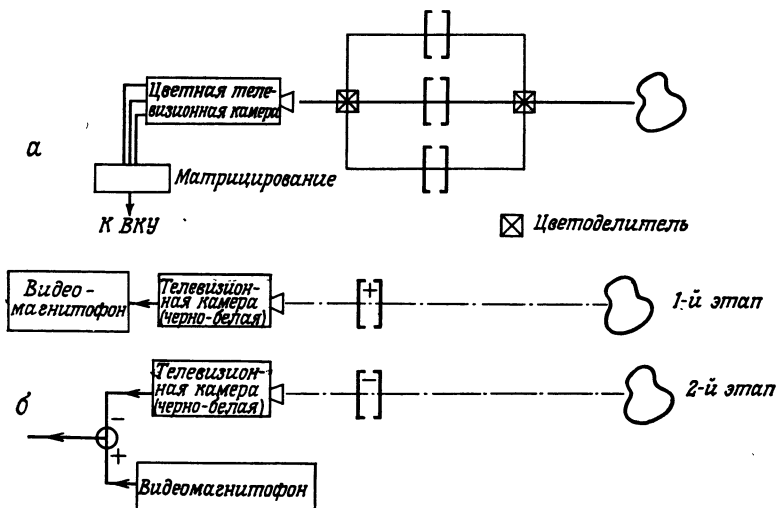


Рис. 19.4 Два варианта гибридной оптико-телевизионной пространственной фильтрации изображений.

а — с использованием трех каналов цветной телевизионной камеры; б — с использованием видеоманитофона для записи изображения на выходе одного из двух каналов обработки

Они могут быть расширены для обработки комплексных сигналов с использованием трех каналов; в этом случае цветная камера специально приспособляется для съемки. При этом этап вычитания видеосигналов должен быть заменен нелинейной обработкой в зависимости от того, интересуются ли модулем или квадратом модуля результата обработки. Можно значительно уменьшить сложность компоновки системы, используя свойство пространственной инвариантности фильтров в некогерентном свете по отношению к смещению (гл. 7, § 2, п. А). Тогда для совмещения трех расположенных рядом зрачков достаточно, чтобы апертура зрачка была достаточно большой. Благодаря наложению хроматических фильтров друг на друга каждый из них участвует лишь в фильтрации сигнала одного из трех каналов (рис. 19.5) [346, 347].

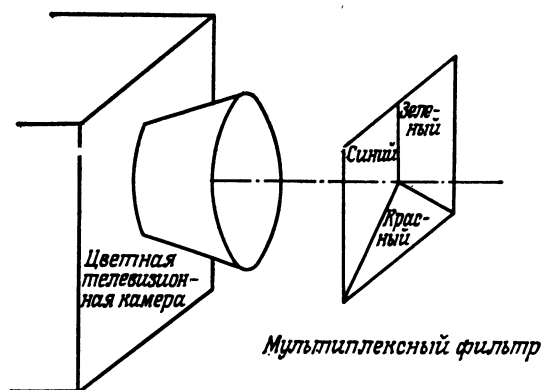


Рис. 19.5. Пространственное и хроматическое уплотнение трех каналов обработки изображения.

Три цветных фильтра расположены рядом в одной плоскости.

Такие характеристики рассмотренных здесь комбинированных методов обработки изображений, как легкость перестройки, обработка в реальном времени, наличие апробированной техники, а также и недостатки, не позволяющие получать высокое качество изображения, снижают их круг применения (робототехника, управление и контроль в промышленности).

§ 4. Комбинированная оптико-телевизионная обработка с обратной связью

Другой тип комбинированной пространственной фильтрации [348, 349], весьма перспективный, основан на том, что с использованием двух составляющих — «некогерентной фильтрации» и «видеофильтрации» — осуществляется функция, сложная при оптической обработке: обратная связь. Оптические системы с чисто оптической цепью обратной связи исследовались в работе [350]. Идея состояла в том, чтобы в слегка разъюстированный резонатор Фабри — Перо поместить когерентную оптическую систему пространственной фильтрации. Но осуществление обратной связи такого вида оказалось делом нелегким, а круг возможных видов обработки ограничивался трудностями динамического диапазона, связанными с пассивностью используемых пространственных фильтров. Эти два ограничения делают особенно привлекательным другое решение, в котором обратная связь обеспечена электронным каналом. Структурная схема оптического процессора с телевизионной обратной связью представлена на рис. 19.6. Она работает следующим образом.

1. Если исключить электронную часть (заштрихованный прямоугольник на рис. 19.6), то мы имеем обычную некогерентную оптическую систему пространственной фильтрации.

2. При помощи полупрозрачного зеркала L_1 часть обработанного сигнала вновь направляется в телевизионную камеру,

полученный видеосигнал затем фильтруется при помощи электронного фильтра, и изображение выводится на видеоконтрольное устройство (ВКУ).

3. Наконец, полупрозрачное зеркало L_2 обеспечивает сложение исходного изображения обрабатываемого объекта с его выведенным на ВКУ изображением, обработанным при помощи оптического, а затем электронного фильтров. Далее «круг» по петле обратной связи проделывает уже это составное изображение.

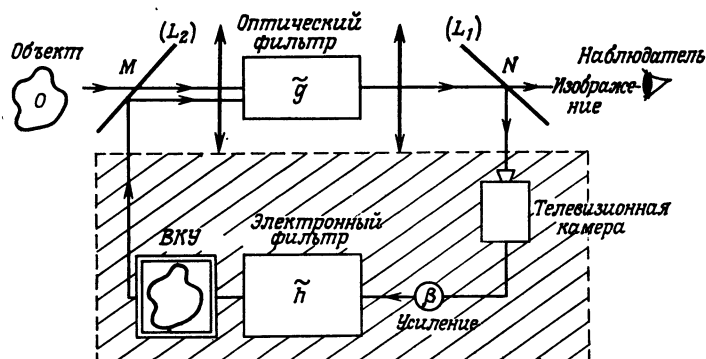


Рис. 19.6. Пространственная фильтрация изображений на основе опико-телевизионной системы с обратной связью

Следовательно, мы здесь имеем *аналогию с обратной связью в электронных системах*, но в данном случае сигналом, направляемым в петлю обратной связи при каждой итерации, является *двумерное изображение*, изменения которого во времени и будут управлять поведением систем в динамике.

Обратимся к обозначениям рис. 19.6. Сигналы объекта o и изображения i зависят от пространственных переменных x и y и от времени t . Прежде всего предположим, что в общем случае и оптическая и телевизионная системы описываются двумя операторами \tilde{g} и \tilde{h} , которые могут изменяться в пространстве и во времени, быть нелинейными, с памятью или без памяти. Возьмем сигнал в точке N изображения; в электронной цепи он преобразуется в $\tilde{h}[\beta i]$ и затем складывается в этой точке с сигналом объекта. Теперь два сигнала обрабатываются оператором \tilde{g} , что дает на выходе

$$i = \tilde{g}\{o + \tilde{h}[\beta i]\}. \quad (19.1)$$

Это основное уравнение остается справедливым, как мы уже отмечали, какими бы ни были операции обработки, осуществляемые над o и i . Однако было бы желательно записать это соотношение в более привычной форме $i = f[o]$. Обычно это

невозможно без некоторых допущений и ограничений, накладываемых на операторы \tilde{g} и \tilde{h} . Поэтому мы предположим, что это линейные и пространственно-инвариантные операторы. Тогда их воздействие может быть описано сверткой:

$$\begin{aligned}\tilde{g}[o] &= g * o, \\ h[i] &= h * i,\end{aligned}$$

где g и h — функции пространственных координат $g(x, y)$ и $h(x, y)$, являющиеся ядрами свертки. Выполнив над обеими частями равенства (19.1) преобразование Фурье относительно пространственных переменных, получаем выражение

$$I = G[O + \beta HI],$$

которое легко преобразуется к привычному виду

$$I = \frac{G}{1 - \beta GH} O \quad (19.2)$$

при условии, что выполняются некоторые условия по обеспечению устойчивости системы с обратной связью (мы их рассмотрим немного дальше).

Тут же отметим, что хотя выражение (19.2) теперь и имеет общий вид пространственной фильтрации $I = FO$, рассмотренной в гл. 6 и 7, оно тем не менее отличается от него в некоторых пунктах: 1) оно представляет собой только упрощенное выражение самого общего соотношения (19.1), допускающего осуществление *пространственно-зависимых* или *нелинейных* коррекций; 2) *временная* переменная входит в соотношение в явном виде; 3) мы располагаем теперь двумя функциями (G и H) для реализации одного фильтра F , что в большинстве случаев приводит к интересной возможности *оптимизации* фильтра в соответствии с источниками шума.

Мы имеем возможность осуществить электронное усиление сигнала в канале и получить так называемую «активную» обратную связь. Без усиления она «пассивная». Основной вопрос в любой системе с обратной связью — это вопрос ее устойчивости. При этом следует различать два случая.

1. В случае *положительной обратной связи* (когда в точке M рис. 19.6 обрабатываемые сигналы складываются) условие устойчивости ограничивает модуль коэффициента усиления β значениями, меньшими единицы.

2. В случае *отрицательной обратной связи* (когда сигналы вычитаются оптически или электронно) условие устойчивости учитывает физические параметры цепи: время послесвечения люминофоров, время накопления передающей трубки телевизионной камеры и т. д. Это обычно ограничивает модуль коэффициента усиления значениями порядка 4—6. Если желательно

реализовать, например, фильтр Винера [формула (6.5)], то нетрудно видеть, что наши возможности ограничиваются достаточно малыми значениями отношения сигнала к шуму, порядка β .

Но от систем фильтрации с обратной связью не всегда требуется устойчивость, и неустойчивые системы, находящиеся в режиме автоколебаний, могут представлять большой интерес для моделирования физических процессов. Именно в этом направлении и используется в настоящее время рассмотренный выше комбинированный метод [351, 352]; кроме того, он применяется для восстановления цветных или черно-белых изображений, содержащих изопланатические или пространственно-зависимые искажения [349, 350].

Правда, все возможности обработки такого типа еще не исследованы, как и возможности оптико-телевизионной обработки (§ 3). В частности, пока еще не рассматривается задача создания элементов с характеристиками, необходимыми специально для тех или иных применений: в этом случае нужно было бы рассмотреть возможности изменения разрешающей способности телевизионных камер и ВКУ или характеристик люминофоров и мишеней передающих трубок для их согласования с требованиями решаемой задачи.

Мы кратко описали три четко различающихся типа комбинированной фильтрации изображений. Все они основываются на принципе оптической фильтрации пространственных частот. Теперь мы рассмотрим новые приложения, в которых та часть обработки, которая приходится на оптику, все более и более специализируется.

§ 5. Быстродействующий преобразователь Фурье

В описываемом ниже решении задачи быстрого спектрального анализа изображений в основном сохранена исключительная способность когерентной оптики осуществлять двумерное преобразование Фурье (гл. 6, § 1). Обрабатываемое изображение обычно представляют в виде диапозитива или вводят в фурье-процессор непосредственно с помощью пространственного модулятора света. Линза, осуществляющая преобразование Фурье, отображает результат обработки на фотоприемник. Затем полученный фурье-образ обрабатывается на ЭВМ. Преимущества такой схемы обработки заключаются, в частности, в том, что он облегчает ЭВМ расчет преобразований Фурье, который занимает много времени (более 500 мс для одной строки из 1024 точек на машине IRIS80 при использовании алгоритма БПФ) и большой объем памяти (двумерное преобразование Фурье часто требует операций над матрицей $500 \times$

× 500). Основная трудность такой схемы связана с преобразованием оптического сигнала в электрический. Было предложено множество решений [353], использующих либо периодическое кадровое сканирование фокальной плоскости линзы (осуществляемое телевизионной камерой или матрицей фотоприемников), либо особое сканирование, более приспособленное к решаемой задаче (например, подвижная круглая или секторная маска в сочетании с интегральным фотоприемником).

Система подобного типа была предложена [354] для диагностики пневмоконоиоза в легочной рентгенографии. В этом случае входные данные имеют аналоговый вид, каждая рентгенограмма механически устанавливается во входной плоскости оптического преобразователя Фурье. Несколько другая схема [355] предложена для классификации термографических изображений.

Ограничения аналогового оптического преобразователя Фурье проявляются на двух уровнях: с одной стороны, они обусловлены большим *динамическим диапазоном получаемого сигнала*, что требует использования фотоприемников с изменяющейся чувствительностью; с другой стороны, они связаны с *отношением сигнала к шуму*, которое, как мы видели в гл. 6, должно быть высоким при обработке изображений методами когерентной оптики.

Зато информационная емкость в такой системе в настоящее время ограничена только *возможностями ввода-вывода информации в ЭВМ*. На практике система оптической преобразователь Фурье + фотоприемник позволяет обеспечить информационную пропускную способность значительно выше 10^6 байт/с при относительной точности вычислений не более $\sim 10^{-2}$.

§ 6. Быстродействующее специализированное вычислительное устройство

Разумеется, если есть возможность осуществлять двумерные преобразования Фурье, то можно реализовать и двумерные свертки вида

$$g(x, y) = f * h,$$

а также и операции взаимной корреляции

$$j(x, y) = f \otimes h = f * k,$$

$$\text{где } h(x, y) = k^*(-x, -y).$$

Эти операции целесообразно осуществлять оптическим методом, если ядра h и k всегда одни и те же для многих задач (математические операторы) или если они составляют небольшой набор, который можно было бы представить в виде на-

бора фильтров, установленных, например, на турели для обеспечения их быстрой смены.

Распространение такого метода вычислений на разнообразные математические операции продвинулось очень далеко. В качестве примера упомянем работу [356], в которой также оптическим методом осуществляется преобразование Меллина с помощью, правда, весьма специфической системы адресации. Отметим также работу [355], в которой оптическими методами осуществляется разложение Карунена — Лозва.

§ 7. Быстродействующие вычислительные устройства для выполнения логических операций

Мы подчеркнули недостатки вычислительных систем, использующих когерентную оптику для обработки изображений; чтобы избежать этих недостатков, многие авторы (см., например, [357]) сохранили в описанном выше специализированном вычислительном устройстве только возможность параллельной обработки, полностью пренебрегая возможностью обработки изображений. Кроме того, они предложили использовать в качестве входных сигналов лишь *дискретную* и *двоичную* информацию, представленную в виде точной копии машинного слова в ЭВМ. Так были решены проблемы динамического диапазона и шума. В этом случае источниками света служат светодиоды, причем как светодиоды, так и фотоприемники образуют интегральную структуру из $N \times N$ элементов. При испытании такой системы в настоящее время используются матрицы из 128×128 элементов; они позволяют осуществлять уже более 10^4 операций одновременно. Увеличение N до 1024 не заставит себя ждать.

Информация передается от излучателя к приемнику либо по воздуху, либо по оптическим волокнам, что более надежно. При прохождении оптического сигнала от излучателя к фотоприемнику осуществляется одна-единственная операция обработки либо оптическими методами (с помощью призм, зеркал), либо с использованием оптоэлектронных устройств (например, сэндвича из матрицы фотоприемников и матрицы излучателей). Такая схема обработки позволяет осуществлять одну логическую операцию параллельно во всех каналах. На основе подобных элементов можно реализовать не только *простые логические схемы* (схемы И, ИЛИ и т. д.), но и *регистры сдвига*, а также *опрокидывающие схемы* (рис. 19.7).

При времени коммутации, равном 5 мс, *полученная в настоящее время информационная пропускная способность превышает 10^6 бит/с*. С увеличением размерности используемых

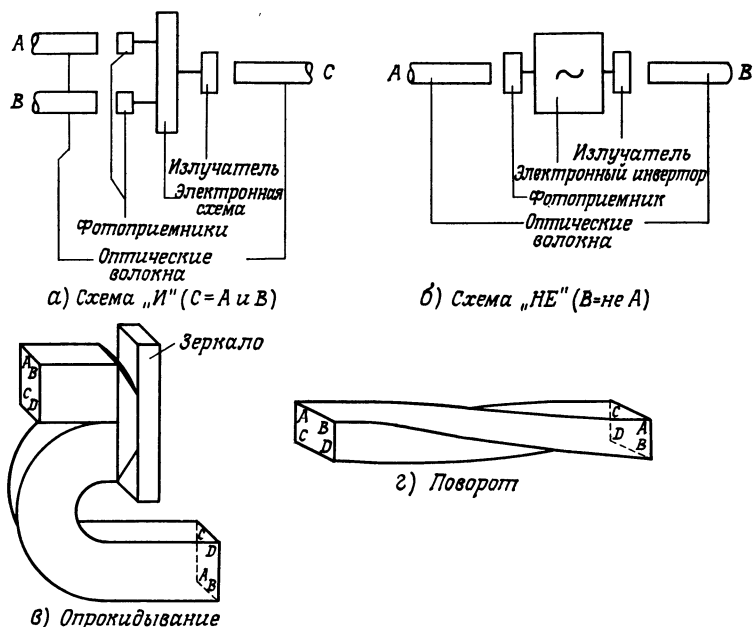


Рис. 19.7. Компоненты, используемые в оптических вычислительных устройствах [358].

а и б — оптико-электронные; в и г — оптические.

матриц и с уменьшением времени коммутации можно ожидать увеличения производительности вычислений в 10—1000 раз.

§ 8. К чисто оптическому цифровому вычислительному устройству

В этом направлении создания вычислительных устройств на основе оптических компонентов ряд авторов [358—360] предлагают сделать дополнительный шаг: использовать оптику для выполнения самых элементарных арифметических вычислений: сложения, умножения и т. д. Это позволяет сохранить присущую оптике высокую степень параллельности вычислений при высоком отношении сигнала к шуму за счет использования бинарных сигналов. Однако это требует разработки специфических компонентов, позволяющих выполнять самые употребительные арифметические операции. Прежде всего необходимо выбрать такой метод вычислений, который лучше подходил бы для реализации оптическими средствами. В этом отношении интересен *метод остаточных классов* (табл. 19.2). Кратко напомним, в чем состоит суть этого метода. Выбирают базу из N

Таблица 19.2

Сложение и вычитание методом остаточных классов

Сложение		Вычитание	
Модули	3457		3457
63 \longrightarrow	0330	29 \longrightarrow	2141
$+ 19 \longrightarrow$	$+ 1345$	$- 12 \longrightarrow$	$- 0025$
$= 82 \longleftarrow$	$= 1225$	$= 17 \longleftarrow$	$= 2123$

Два числа разлагаются по базовым модулям. Операции сложения и вычитания осуществляются над остатками. Преобразование «остаток» — десятичная система сложнее, чем разложение числа. Оно основывается на китайской теореме об остатке и дает десятичный результат вычислений в виде линейной комбинации его остатков.

взаимно простых чисел m_1, m_2, \dots, m_N , т. е. модулей. Разложение целого числа A по i -му модулю представляет собой остаток от деления A на m_i : $a_i = A \bmod (m_i)$. Таким образом, все положительные целые числа, меньшие $M = \prod_{i=1}^N m_i$, допускают

единственное разложение по M модулям. Каждая элементарная арифметическая операция имеет свой «перевод» в арифметике остаточных классов. Такая система вычислений особенно хорошо подходит для оптики в двух отношениях: с одной стороны, все вычисления проводятся параллельно, без передачи промежуточных результатов, как в обычной арифметике; с другой стороны, на каждом этапе необходима та точность, которая требуется при операциях с остатком a_i , а не со всем числом A .

Таким образом, «оптический процессор» будет представлять собой набор простых функций, или *таблиц*, очень быстро выполняющих наиболее часто встречающиеся операции: сложение, вычитание, умножение, разложение полиномов и в некоторых случаях даже корреляцию. Каждая таблица может быть осуществлена различными способами: либо с динамической адресацией, реализуемой при помощи дефлекторов, либо со статической адресацией, осуществляемой при помощи дифракционных решеток, зеркал, призм и т. д. (Некоторые функции уже осуществлены практически [358, 359].)

Одна таблица очень проста, если выбираются небольшие по величине модули, но общая сложность структуры быстро нарастает с увеличением числа взаимно связываемых таблиц. Структуры в интегрально-оптическом исполнении дадут в этом отношении большие преимущества компактности и совместимости с электронными схемами.

Из расчета элементарного времени переключения, равного 10 нс, в работе [358] мощность такого компьютера оценивается

величиной 10^{10} арифметических операций в секунду; это в 30 раз больше производительности ЭВМ на ПЗС и в 10 раз больше производительности цифровой ЭВМ.

Каково будущее вычислительных устройств такого типа, а также описанных в § 6? На этот вопрос очень трудно ответить. Самая осторожная оценка дает, безусловно, привлекательные цифры, не затрагивая последнего резерва оптики — мультиплексирования оптических сигналов по длинам волн. Но в истории современной оптики очень много такого «пилообразного» оптимизма: голографическая память, когерентная пространственная фильтрация изображений, оптическое автоматическое чтение, волоконная оптика. Если одни из перечисленных применений можно отнести к бесспорным успехам оптики, то другие, напротив, остаются еще как бы недоношенными. Есть опасения, что к ним относятся оптические калькуляторы, которые умерли еще до рождения, и не столько из-за отсутствия достоинств, сколько, быть может, из-за капризной Дамы Фортуны, которая решает, своевременно ли развитие в таком направлении или нет.

Но, возможно, суть этой главы не в содержании ее параграфов, а в их взаимосвязи. Поэтому напомним ход изложения.

С одной стороны, с учетом недостатков обработки изображений методами когерентной оптики нам потребовалось одновременно ускорить процесс синтеза фильтров и улучшить качество обработанных изображений. С другой стороны, с учетом трудностей некогерентной оптики нам потребовалось выполнять вычитание изображений более простым способом или расширить круг возможных видов обработки изображений. С этой точки зрения сочетание оптических методов с электронными (цифровыми или телевизионными) дает частные, но в большинстве случаев удовлетворительные решения проблемы, что мы уже неоднократно и подчеркивали. Тем не менее качество обработанных таким образом изображений остается невысоким и часто не соответствует принятым критериям оценки качества изображений.

Очень велико искушение забыть способность оптики давать изображения и сохранить лишь ее «суть», а именно способность быстрого выполнения преобразования Фурье и параллельность обработки информации. Последующие этапы обработки (измерение, классификация, распознавание и т. д.) обеспечиваются теперь цифровой ЭВМ. Но и тогда можно будет упрекнуть оптику в ее неточности как аналогового метода, в больших шумах и недостаточном динамическом диапазоне.

Если нужно сделать еще один шаг, то мы возьмем от нее лишь необычайные возможности параллельной обработки, представляя обрабатываемую информацию в форме бинарных изо-

бражений. Получим оптические арифметические и логические функциональные узлы, главные блоки любого вычислительного устройства, осуществляющие параллельную обработку целых матриц. Но больше уже не от чего отказываться! Нечего и думать о дальнейшем упрощении.

Правда, приложения, опирающиеся на последнюю основу, могут быть многочисленными; выше мы видели возможности приложений оптических методов к обработке информации с использованием логических функций и вычислений методом остаточных классов. Есть и еще одна возможность, относящаяся в большей степени к дальней связи. Это *оптическая коммутация* [361, 362].

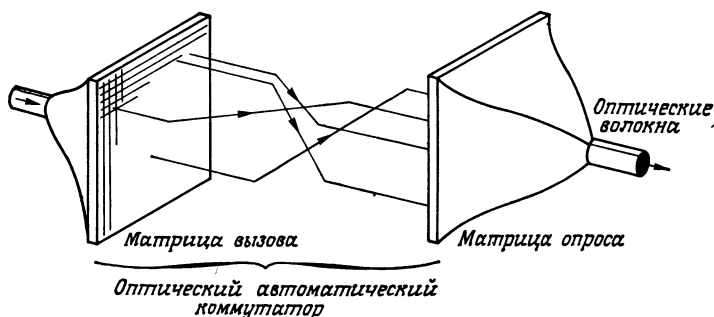


Рис. 19.8. Оптическая коммутация.

Задача в том, чтобы установить двухстороннюю взаимно-однозначную связь между двумя пучками оптических волокон, обеспечив при этом передачу некоторого числа вспомогательных служебных сигналов.

Учитывая нынешнее развитие сетей дальней связи, логично уже сейчас задуматься над вопросами коммутации в оптической связи. Известно, что чаще всего предлагаемое решение состоит в стандартизации связи при помощи оптических волокон, допускающей их подсоединение к любому центральному устройству. Это мудрое решение, которое тем более оправдано, что строительство автоматических узлов связи очень дорого и растягивается на долгие годы.

Но попробуем забежать вперед и поставим задачу соединить абонента А, который делает вызов, с абонентом Б, зная, что оба абонента связаны оптическими волокнами с одним и тем же центральным узлом. Преобразование двух световых сигналов в электрические сигналы, совместимые с современными коммутаторами, не всегда оправдано. Заметим сначала, что проблема коммутации идентична проблеме, решаемой арифметическим калькулятором: матрице с N входами (тех, кто вызывает) нужно поставить в однозначное соответствие матрицу с N выходами (тех, кого вызывают). (В действительности могут быть одновременно отобраны только n и N входов, рис. 19.8.)

Такое соответствие матриц осуществляется параллельно, и желательно иметь возможность очень быстро изменять его. Таким образом, нам нужно реализовать такую же «таблицу», как при вычислениях по методу остаточных классов, но с некоторыми отличиями. Во-первых, необходимо предусмотреть служебные функции, связанные с другими функциями таблицы и обусловленные процедурой телефонных соединений. Они будут обеспечивать, в частности, подачу звуковых сигналов, изменение тональности, а также и информацию о плате за разговор (не нужно забывать...). Во-вторых, связь должна быть двусторонней, идущей как от А к Б, так и от Б к А. Следовательно, нужно подумать о структурной схеме такой оптической коммутации. Как это часто бывает, у нас есть выбор между пространственной и временной коммутацией.

Пространственную коммутацию в принципе можно осуществлять двумя способами. 1. Либо приводя в соответствие при каждом вызове полной матрице излучателей некоторое изображение (но скорее всего голограмму из соображений компактности). Таким образом, каждый вызов постоянно находится в соответствии со всей совокупностью вызываемых абонентов, но как только связь устанавливается, все остальные переходят в состояние «занято». 2. Либо осуществляя динамическую адресацию при помощи голографических дифракционных решеток, записываемых на реверсивных носителях (например, сегнетоэлектриках, жидких кристаллах или термопластиках). В этом случае двусторонняя связь обеспечивается за счет возврата света, но передача служебных сигналов несколько проблематична. При временной коммутации каждому соединению отводится короткий момент коммутации. Это может быть осуществлено различными способами, с помощью акустооптических дефлекторов, матриц ПЗС или быстродействующих электронных коммутаторов. Здесь чисто оптические решения, по-видимому, менее привлекательны, поскольку исчезла параллельность соединений.

В этой области предстоит еще очень многое сделать, как в концептуальном, так и в экспериментальном плане. Но по крайней мере здесь мы пока не связаны соображениями рентабельности.

ПРИЛОЖЕНИЕ 1

МЕТАЛЛИЧЕСКИЙ СВЕТОВОД, МОДЫ В ЗАВИСИМОСТИ ОТ ДЛИНЫ ВОЛНЫ, ФАЗОВАЯ СКОРОСТЬ И СКОРОСТЬ СВЕТА

Почему фазовая скорость больше скорости света c ? В этом нет ничего удивительного.

Возьмем уравнение дисперсии (3.14), выраженное через длину волны. Имеем

$$k = \frac{2\pi}{\lambda}.$$

Критическая длина волны определяется как

$$\lambda_c = \frac{2\pi c}{\omega_c},$$

откуда в соответствии с формулой (3.19) получаем $\lambda_c = 4a$.

Введем также величину $\lambda_z = 2\pi/\beta$, которую назовем *кажущейся длиной* волны по оси Oz . Соотношение (3.14) примет вид [формула (3.23)]

$$\frac{1}{\lambda_z^2} = \frac{1}{\lambda^2} - \frac{n^2}{\lambda_c^2}.$$

Физический смысл критической длины волны λ_c/n явствует из аналогии с колеблющейся струной.

На обоих концах $x = \pm a$ должны быть узлы. Очевидно, что это возможно только в том случае, если кажущаяся длина волны λ_x по оси Ox удовлетворяет соотношению

$$2a = n \frac{\lambda_x}{2},$$

$$\text{откуда } \lambda_x = \frac{4a}{n} = \frac{\lambda_c}{n}.$$

Попутно снова находим условие согласования фаз, которое можно записать в виде

$$2\gamma 2a = 2n\pi,$$

откуда [формула (3.18)]

$$\gamma = \frac{n\pi}{2a}.$$

Таким образом, для моды порядка n величина λ_c/n равна расстоянию между двумя узлами вдоль оси Ox .

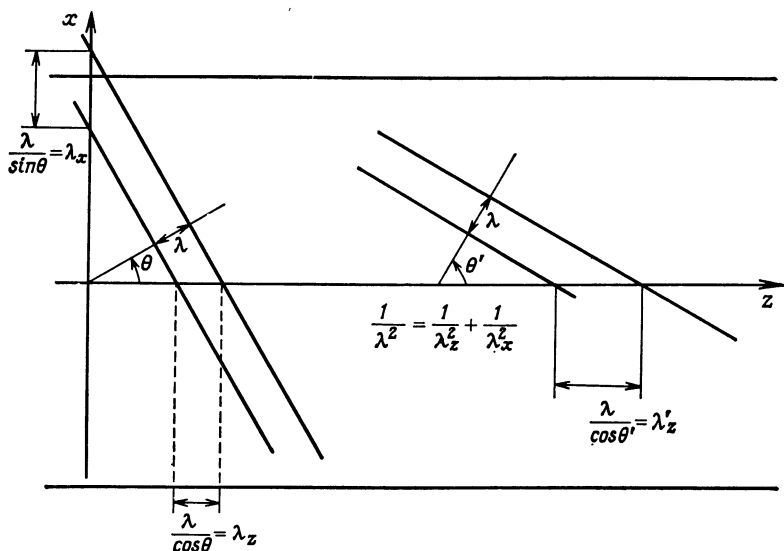


Рис. П1.1. Металлический световод.

Распространение плоской волны

Соотношение (3.23) графически иллюстрируется на рис. П1.1, где представлены две волновые поверхности, следующие одна за другой на расстоянии λ . Мы видим, что λ_z всегда больше

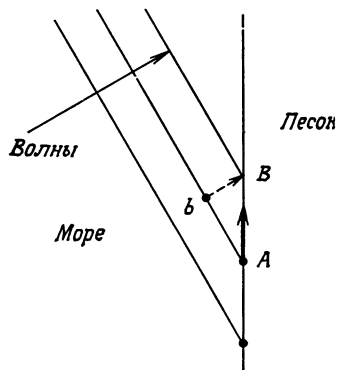


Рис. П1.2. Набегание морских волн на берег.

величины λ . Следовательно, фазовая скорость вдоль оси Oz всегда больше, чем скорость света c . При увеличении угла θ , когда он стремится к $\pi/2$, величина λ_z становится очень большой: фазовая скорость стремится к бесконечности.

Это понятно каждому, кто хотя бы раз побывал на морском пляже, когда волны набегают на берег под очень малым уг-

лом (рис. П1.2). При этом *кажется*, что морская пена очень быстро перемещается вдоль берега. Но это только кажется, так как на самом деле никакого перемещения материи между точками A и B не происходит. Пена, возникающая в точке B , приносится из точки b , но человеческий глаз воспринимает это как перемещение из точки A в точку B , и создается впечатление, что пена движется очень быстро. Ряд совершенно разных физических событий мысленно связывается в единую последовательность. Это аналогично следу светового зайчика, движущегося по стене.

В противоположность этому *групповая скорость v_g всегда меньше скорости света c* . При каждом появлении моды $v_g = 0$, затем величина v_g возрастает и стремится к c , когда ω стремится к бесконечности (плоская волна).

ПРИЛОЖЕНИЕ 2

МЕТАЛЛИЧЕСКИЙ СВЕТОВОД, ВЫРОЖДЕННАЯ ВОЛНА

Рассмотрим вырожденную волну, существующую в металлическом световоде, на частоте, меньшей критической частоты, с энергетической точки зрения.

Сначала необходимо рассчитать напряженность магнитного поля. Это нетрудно сделать [формула (1.1)]. На основании выражений (3.10) и (3.17) и результатов гл. 3, § 1 находим

$$H_x = \frac{k \cos \theta}{\omega \mu} E,$$

$$\text{где } E = E_0 e^{i[\omega t - (k \cos \theta) z]} [e^{-i y x} - (-1)^n e^{i y x}];$$

$$H_z = -\frac{\gamma}{\omega \mu} E_0 e^{i[\omega t - (k \cos \theta) z]} [e^{-i y x} + (-1)^n e^{i y x}].$$

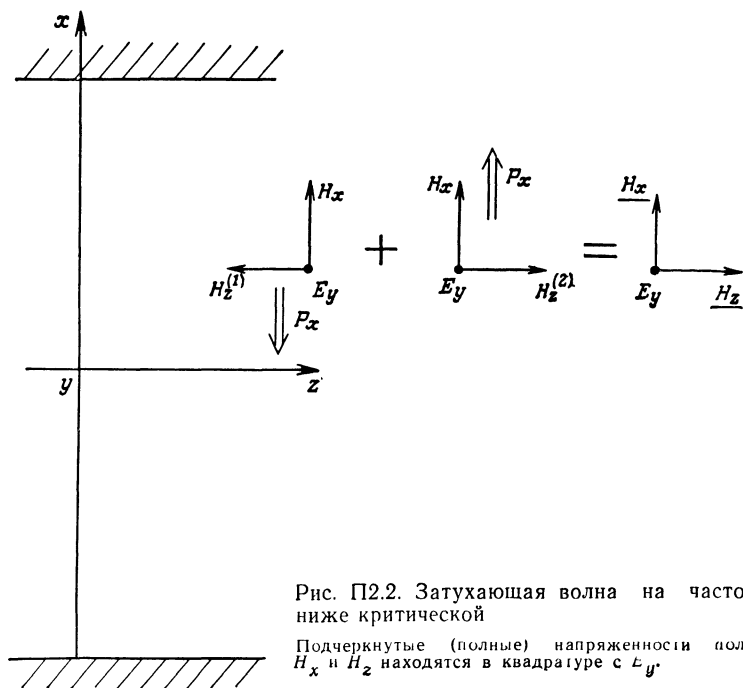
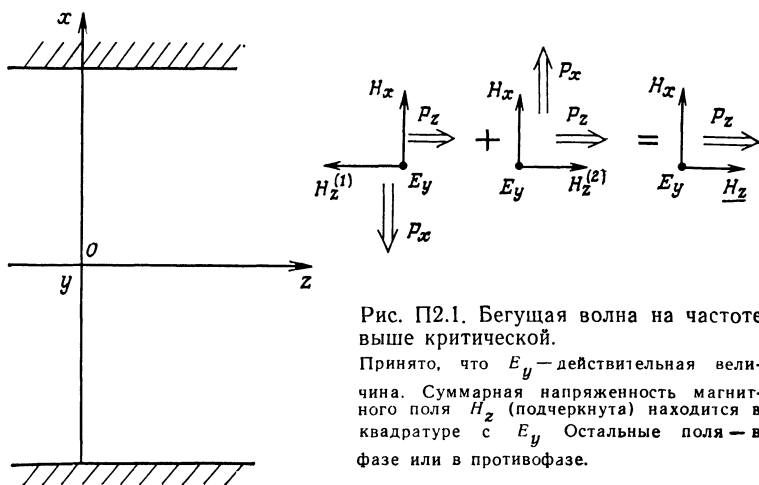
$\begin{array}{ccc} & \uparrow & \uparrow \\ & H_{(z)}^{(1)} & H_{(z)}^{(2)} \end{array}$

Полученные выражения справедливы на частотах как ниже, так и выше критической, разумеется, при условии, что величина $\cos \theta$ определяется формулой (3.27).

Таким образом, напряженность магнитного поля H_z равна сумме двух напряженностей, которые мы обозначим через $H_z^{(1)}$ и $H_z^{(2)}$.

1. Рассмотрим сначала случай бегущей волны, распространяющейся без затухания и имеющей частоту больше критической. В этом случае поле H_x синфазно с E_y .

Отсюда получаем вектор Пойнтинга, соответствующий распространению волны вдоль оси Oz (рис. П2.1). Чтобы исследовать распределение энергии в поперечном сечении световода, нужно рассмотреть пары векторов $(E_y, H_z = H_z^{(1)} + H_z^{(2)})$. Легко видеть, что поле $H_z^{(1)}$ находится в фазе, а поле $H_z^{(2)}$ — в противофазе с E_y . Но их сумма H_z сдвинута по фазе на $\pi/2$. Таким образом, имеются два вектора Пойнтинга, которые взаимно уничтожаются; в результате векторное произведение $E_y \times H_z^*$ оказывается чисто мнимым и соответствует *реактивной энергии*. Именно такой результат получается, когда стоячую волну представляют в виде суммы двух волн, распространяющихся в противоположных направлениях. *Результирующая волна является неоднородной волной.*



2. Рассмотрим тем же способом вырожденную волну, полученную на частоте ниже критической. В этом случае главное отличие состоит в том, что теперь вектор H_z находится в квадратуре с вектором E_y (рис. П2.2).

Отметим, что в данном случае волна представляется в виде суммы двух неоднородных волн. Как преобразуется энергия в этом случае? Если рассмотреть сначала суммарную волну, то видно, что никакой активной энергии нет. Имеется реактивная энергия, и мы наблюдаем только преобразование этой энергии из индуктивной в емкостную, и наоборот. Относительно энергетического баланса каждой взятой отдельно неоднородной волны можно заметить следующее: наблюдается поток излучаемой энергии вдоль оси Ox и затухающий поток вдоль оси Oz ; при суммировании этих двух волн остаются только реактивные составляющие энергии.

Таким образом, в случае вырожденной волны на частоте ниже критической восходящие и нисходящие плоские волны заменяются неоднородными волнами.

ПРИЛОЖЕНИЕ 3

СМЕЩЕНИЕ СВЕТОВОГО ПУЧКА ПРИ ОТРАЖЕНИИ ОТ ГРАНИЦЫ РАЗДЕЛА ДВУХ СРЕД, ЭФФЕКТ ГООСА — ХЕНХЕНА

Сначала покажем, что полное внутреннее отражение плоской волны на границе раздела двух сред сопровождается теми же эффектами, что и отражение от металлической плоскости, смещенной на некоторое расстояние x_0 от поверхности раздела рассматриваемых сред (рис. ПЗ.1).

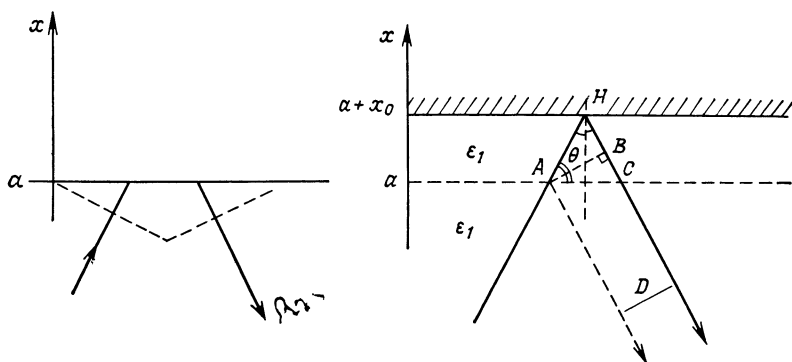


Рис. ПЗ.1. Полное внутреннее отражение плоской волны на границе раздела двух сред.

Эквивалентно отражению от металлической плоской поверхности (справа).

В гл. 3, § 1 было показано, что полное волновое сопротивление после отражения волны от плоскости, отделяющей среду 1 от среды 2, равно

$$Z = \frac{E_{\text{полн}}}{H_{\text{полн}}} = Z_2.$$

Исходя из этого, легко рассчитать это волновое сопротивление в точке $x = a$ для случая волны, отразившейся от металлической плоскости и остающейся в той же среде 1. Найдем [3]

$$Z = Z_1 \frac{e^{iax_0} - e^{-iax_0}}{e^{iax_0} + e^{-iax_0}} = iZ_1 \operatorname{tg}(ax_0).$$

Приравняв полученное выражение величине Z_2 , получаем соотношение

$$i \frac{Z_1}{Z_2} = -\operatorname{tg}\left(ax_0 + \frac{\pi}{2}\right),$$

которое полезно сравнить с (3.7). В результате находим зависимость φ от x_0 :

$$\varphi = 2\alpha x_0 + \pi. \quad (\text{ПЗ.1})$$

Само собой разумеется, что найденная фаза определяется с точностью до π . Таким образом, отражение может происходить от всех плоскостей, отстоящих друг от друга на величину, кратную $\pi/2\alpha$. Выбор, сделанный в формуле (ПЗ.1), мы обоснуем ниже.

Предшествующее выражение можно также получить, рассчитав оптический фазовый сдвиг при отражении от металлической плоскости. Как в классическом случае интерференции в тонких пластинках (рис. ПЗ.1), находим разность хода [7]

$$\delta = \overline{AH} + \overline{HB} = 2x_0 \sin \theta.$$

Внимание! Сказать, что $\delta = \overline{AH} + \overline{HC}$, было бы ошибкой, так как мы должны сравнивать фазу отраженной волны с той фазой, которую имела бы эта волна при отражении в точке А. С учетом этого получаем фазовый сдвиг

$$\varphi = kn_1\delta = 2x_0kn_1 \sin \theta = 2x_0\alpha,$$

к которому необходимо добавить скачок фазы, равный π , который появляется при отражении от металлической поверхности. В результате получаем соотношение (ПЗ.1).

Итак, можно сделать окончательный вывод: отражение светового пучка от границы раздела двух диэлектрических сред сопровождается его смещением на величину D , равную

$$D = 2x_0 \cos \theta.$$

Проделанные вычисления были проведены Ньютоном при исследовании им полного внутреннего отражения [8]. В действительности рассматриваемое явление значительно сложнее. В самом деле, такой опыт невозможно осуществить, поскольку не существует плоской волны. Поэтому придется использовать спектр плоских волн (гл. 2, § 1), для которого напряженности падающего $E(X, Z)$ и отраженного $E_r(X, Z)$ полей можно записать следующим образом (в координатах X, Z , связанных с координатами светового пучка соотношениями $x = X - a$, $z = Z$):

$$E(X, Z) = \int E_0(k_x) e^{i(k_x X + k_z Z)} dk_x,$$

$$k_x^2 + k_z^2 = k^2,$$

$$E_r(X, Z) = \int e^{-i\varphi(k_x)} E_0(k_x) e^{-i(k_x X + k_z Z)} dk_x.$$

Разложим фазу коэффициента отражения как функцию переменной k_x в ряд в окрестности среднего угла падения:

$$\varphi(k_x) = \varphi_0 + k_x \left[\frac{\partial \varphi}{\partial k_x} \right]_{k_x=0} + \frac{1}{2} k_x^2 \left[\frac{\partial^2 \varphi}{\partial k_x^2} \right]_{k_x=0} + \dots$$

Отсюда легко найдем смещение:

$$D = \left[\frac{\partial \varphi}{\partial k_x} \right]_{k_x=0}. \quad (\text{ПЗ.2})$$

Эта величина называется смещением Гооса — Хенхена [9], а выражение (ПЗ.2) — формулой Артманна [10].

Ограниченный объем книги не позволяет нам углубиться в данную область, в которой многие вопросы остаются еще спорными. Поэтому мы рекомендуем читателю обратиться к недавно опубликованным работам [11—13].

ПРИЛОЖЕНИЕ 4

ДИЭЛЕКТРИЧЕСКИЙ СВЕТОВОД, РАСЧЕТ НАПРЯЖЕННОСТЕЙ ПОЛЯ

Положив в формуле (3.30) $x = X - a$, можно найти полную напряженность электрического поля в диэлектрическом световоде в виде

$$E_{\text{полн}} \sim e^{i(\omega t - \beta z)}(e^{-i\alpha X} + r e^{+i\alpha X}),$$

где $r = \exp\{-i\varphi\}$ не что иное, как коэффициент отражения при $x = a$. При $X = x_0$ имеем $E_{\text{полн}} = 0$, т. е.

$$e^{-i\alpha x_0} + e^{-i\varphi} e^{i\alpha x_0} = 0.$$

Это, в сущности, другой вывод формулы (ПЗ.1). На критической частоте получим $r = 1$ и

$$E_{\text{полн}}(n\omega_c) \sim e^{i(\omega_c t - \beta z)} 2i \sin \alpha X \sim \sin \alpha(x + a).$$

Так как $\alpha a = n(\pi/2)$, имеем:

$$\text{Если } n = 2p, \text{ то } E(n\omega_c) \sim \sin\left(p \frac{\pi}{a} x\right).$$

$$\text{Если } n = 2p + 1, \text{ то } E(n\omega_c) \sim \cos\left(\frac{\pi}{2} + p\pi\right) \frac{x}{a}.$$

Эти выражения полностью идентичны формулам (3.12), относящимся к металлическому световоду.

Расчет напряженности магнитного поля аналогичен расчету напряженностей электрического поля, и мы его здесь проводить не будем.

ПРИЛОЖЕНИЕ 5

ОСНОВНЫЕ МАТЕРИАЛЫ ДЛЯ ОПТИЧЕСКОЙ ОБРАБОТКИ ИНФОРМАЦИИ

В данном приложении мы разделим носители оптической информации на простые материалы, такие, как фотоэмульсии, фоторезисты и т. п., и на более сложные структуры, основанные на этих простых базовых материалах, в сочетании с различными физико-химическими компонентами, позволяющими образовать фоточувствительную систему; это электрооптические преобразователи, жидкие кристаллы и термопластики. Читатель может найти полную информацию о носителях первой группы в работах [363, 364].

§ 1. Фотографическая эмульсия [365]

Этот носитель оптической информации используется наиболее широко, чему способствуют многие положительные факторы:

1) характеристики фотоэмульсии отличаются высокой стабильностью и могут быть прекрасно подобраны для каждого конкретного применения;

2) фотоэмульсии широко применяются во многих других областях науки и техники, так же как и химические вещества, используемые для их производства;

3) их применение отличается простотой и не требует специального оборудования.

Главный недостаток фотоэмульсий состоит в необходимости многоступенчатой химической обработки. Мы увидим, что другие недостатки, на которые иногда жалуются те, кто ими пользуется (нелинейность характеристики, низкая чувствительность, фотографический шум), в значительно большей степени присущи другим рассматриваемым нами материалам.

Фотографическая эмульсия — это материал с изменяемым пропусканием (прозрачностью). Как мы увидим, при необходимости ее можно преобразовать и в фазовый модулятор. Принцип записи информации на фотоэмульсиях основан на трех последовательных химических реакциях [363, с. 299], в результате которых кристаллы галлоида серебра преобразуются в металлосеребро.

Основной характеристикой фотографической эмульсии является кривая почернения, определяющая зависимость ампли-

тудного пропускания от падающей энергии света (рис. П5.1). (Эту кривую чаще представляют в логарифмических координатах, т. е. приводят не зависимость $t = f(E)$, а $d = \lg(t^2) = f(\lg E)$, где величина d называется оптической плотностью эмульсии.) Указанные кривые снимаются экспериментально.

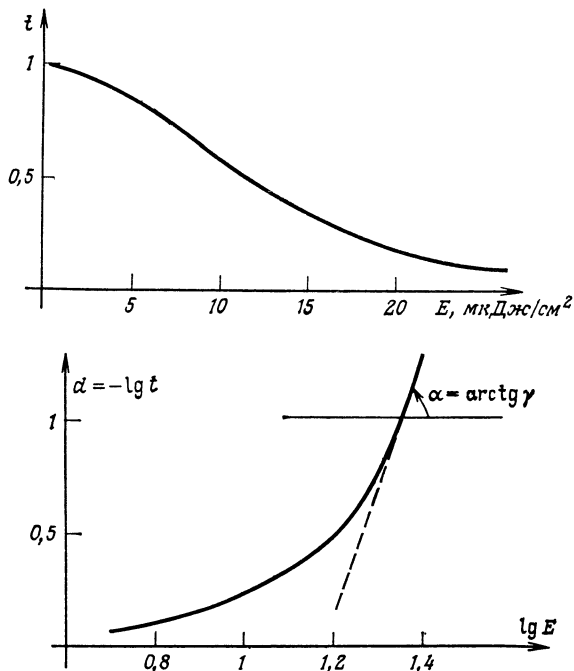


Рис. П5.1. Внизу — кривая почернения фотографической эмульсии; сверху — характеристическая кривая голографической эмульсии Agfa Gevaert Holotest 8E 75 (линейный масштаб); t — пропускание, E — энергетическая освещенность.

Для простоты эмульсию часто характеризуют просто крутизной кривой $d = f(\lg E)$ на ее линейном участке. Эта величина называется коэффициентом контрастности и обозначается через γ :

$$\gamma = \frac{df(\lg E)}{dE}.$$

Для идеальной эмульсии зависимость между амплитудным коэффициентом пропускания и получаемой ею энергией должна быть линейной. Это соответствует значению $\gamma = -2$. Отрицательный коэффициент контрастности означает, что речь идет о «позитивной» эмульсии, т. е. коэффициент пропускания неэкспонированной эмульсии равен нулю. Черно-белые эмульсии обычно относятся к «негативным» материалам.

Другой важной характеристикой фотографической эмульсии является частотно-контрастная характеристика (ЧКХ). На рис. П5.2 приведена ЧКХ одной из фотографических эмульсий.

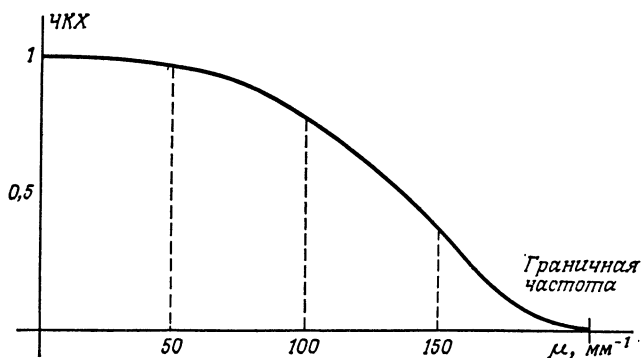


Рис. П5.2. Частотно-контрастная характеристика (ЧКХ) одной из эмульсий.

Основные параметры фотографических эмульсий, а также некоторых других материалов приведены в табл. П5.1.

Таблица П5.1

Характеристики наиболее распространенных фоточувствительных материалов

	Чувствительность, Дж/см ²	Разрешение, линия/мм	Эффект объемности	Вид модуляции	Возможность переписи
Фотографические эмульсии общего назначения (черно-белые)	10^{-9}	100	—	Амплитудная	Нет
Фотографические эмульсии голографические	10^{-5}	2000	Есть	То же	Нет
Бихромированная желатина	10^{-2}	2000	Есть	Фазовая	Нет
Фотохромные материалы	10^{-1}	3000	Есть	Амплитудная	Есть
Сегнетоэлектрики	$10^{-4} - 10^{-2}$	1000	Есть	Фазовая	Есть
Фоторезисты	10^{-2}	1000	Нет	То же	Нет
Фототермопластики	10^{-5}	1000	Нет	»	Есть
ПМС на эффекте Поккельса	10^{-5}	100	Нет	По поляризации	Есть
Жидкие кристаллы	—	100	Нет	Амплитудная и по поляризации	Есть

Примечание Указаны порядки величин для обычно применяемых материалов, а не для лучших образцов.

Фотографическая эмульсия позволяет получать чисто фазовую запись изображения. Для этого нужно удалить из слоя желатины металлическое серебро, в результате чего останутся пустые области, которые обеспечат изменение фазы. Существует много методов такого «отбеливания» эмульсии [363, с. 59]. Отметим лишь метод окисления серебра раствором соли (хлористого натрия), а также метод отбеливания парами галогенов, который позволяет избежать усадки эмульсии при сушке.

§ 2. Бихромированная желатина

Бихромированная желатина относится к фазово-модулирующим средам, которые можно использовать для толстых голограмм. Такие среды отличаются высоким разрешением, малым поглощением и слабым рассеянием света. Их основной недостаток в том, что в процессе обработки происходит изменение записанной информации.

Слои бихромированной желатины состоят из коллоидных растворов (конечно, желатины, а также альбумина, поливинилового спирта и т. д.), фотосенсибилизированных хроматами и бихроматами аммония, натрия и калия. Изменения, происходящие в среде после экспозиции, могут быть двух типов: изменение толщины слоя и изменение показателя преломления. Модуляция толщины слоя широко применялась в фотографии и типографском производстве. Модуляция показателя преломления представляет интерес для голографии и обработки информации [366]. Изменения показателя преломления могут составлять от 0,001 до нескольких десятых в зависимости от используемого способа, что обеспечивает изменение фазы, измеряемой в долях длины волны, $\Delta\varphi = 2\pi d\Delta n/\lambda$ (где d — толщина слоя желатины), от 1 до 2. Это позволяет получить оптимальные для фазовых голограмм значения дифракционной эффективности (гл. 5, § 3).

§ 3. Фотохромные материалы

Фотохромные материалы отличаются от рассмотренных ранее сред реверсивностью, т. е. способностью многократно записывать и стирать информацию. Поэтому они представляют большую ценность там, где необходима последовательная перепись (запись/стирание) информации, тем более что при этом не требуется никакой химической обработки. Фотохромные материалы характеризуются также полным отсутствием зернистости, чем обеспечивается высокое разрешение. Они являются модуляторами по интенсивности и используются для записи толстых голограмм.

К их главным недостаткам в настоящее время относятся малая дифракционная эффективность, низкая чувствительность, а также чувствительность записанного изображения к считывающему свету, что приводит к стиранию записанной информации.

В качестве основных фотохромных материалов используются фтористый кальций, титанат кальция и титанат стронция,

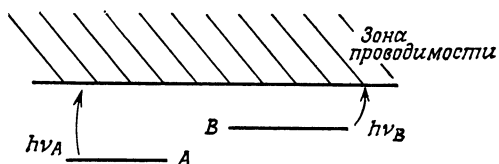


Рис. П5.3. Действие фотохромного материала.



легированные редкоземельными металлами. Их энергетические уровни имеют в запрещенной зоне ловушечные уровни *A* и *B* (рис. П5.3). Эти два уровня порождают две области поглощения, показанные на рис. П5.4. Когда материал возбуждается

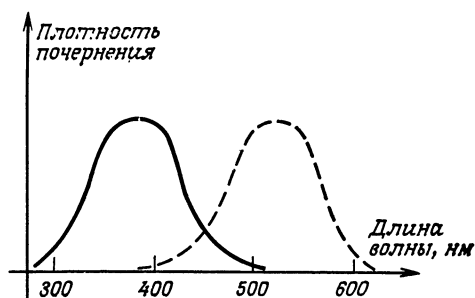


Рис. П5.4. Зависимость оптической плотности фотохромных материалов от длины волны в двух основных случаях.

Сплошная линия — заполнен уровень *A*, штриховая линия — занят уровень *B*.

экспонирующим излучением в области *A* (ультрафиолетовая область), его электроны переходят с уровня *A* в возбужденное состояние, затем спонтанно заполняют уровень *B*. Далее электроны очень медленно рекомбинируют на уровень *A*. Процесс рекомбинации можно ускорить с помощью зеленого излучения в полосе *B*. Таким образом, существуют два режима работы фотохромных материалов:

1) запись изображения ультрафиолетовым излучением и считывание зеленым светом;

2) «стирание» (т. е. запись) изображения зеленым светом на материале, равномерно засвеченном в ультрафиолетовом излучении, а затем его считывание снова в зеленом свете, но уже в «негативе».

Работа во втором режиме, несмотря на более низкую чувствительность, позволяет применять когерентный свет как при записи, так и при воспроизведении.

§ 4. Сегнетоэлектрические кристаллы

Это фазовые модуляторы света, реверсивные, имеющие хорошие оптические характеристики (разрешение, чувствительность, информационную емкость). Их главный недостаток в том, что в настоящее время сложно обеспечить одновременно все эти положительные свойства в одном экземпляре кристалла. В настоящее время к самым интересным и перспективным материалам для сегнетоэлектрических модуляторов относятся ниобат лития (LiNbO_3), титанат бария (BaTiO_3), а также керамика на основе полилантаната циркония и тантала (PLZT). В них используется полоса поглощения ионов железа Fe^{++} в видимой области для создания градиента концентрации свободных электронов, который создает электрическое поле и тем обеспечивает нужное изменение показателя преломления. Распределение этих свободных электронов во всем материале можно изменять, создавая в кристалле электрическое поле.

§ 5. Фоторезисты

Фоторезисты относятся к фазовым модуляторам, осуществляющим модуляцию света за счет изменения толщины слоя (таким образом, отсутствуют объемные эффекты). Фоторезисты характеризуются высоким разрешением, но весьма низкой чувствительностью. С другой стороны, если желательно обеспечить хорошую линейность записи, то не следует работать с большими коэффициентами модуляции.

Фоторезисты — это органические материалы, растворимость которых изменяется при облучении светом. Они могут быть как позитивными, так и негативными (П5, § 1). Как и фотографические пластинки, они выпускаются промышленностью в различных видах.

Фоторезисты применяются главным образом тогда, когда требуется много копий одного оригинала. Они позволяют легко получать рельефные слепки-копии.

Мы не будем описывать другие, менее разработанные материалы, такие, как магнитооптические пленки или пленки, получаемые осаждением металла. Далее мы рассмотрим более сложные структуры преобразователей изображения, для которых использование только одного материала недостаточно для обеспечения всего комплекса операций по записи и модуляции света.

§ 6. Термопластики

Термопластики — это аморфные вещества, размягчающиеся при нагревании. Они могут играть роль фазовых модуляторов света, сравнимых с тонкими дифракционными решетками, но допускают повторную запись. Пластические деформации создаются главным образом за счет электростатических сил, возникающих при образовании потенциального рельефа (распределения заряда) на противоположных поверхностях термопластика. Закрепление записанного изображения производится

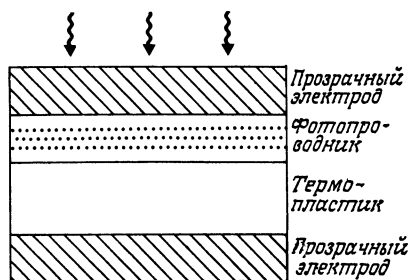


Рис. П5.5. Слоистая структура на фототермопластике.

путем быстрого охлаждения, а его стирание — путем общего нагрева с последующим равномерным распределением заряда по его поверхности.

Существуют два способа зарядки.

1. В первом способе используется электронная пушка, позволяющая избирательно наносить заряды путем сканирования электронным пучком всей поверхности термопластической мишени.

2. Второй способ состоит в том, что предварительно равномерно заряжают всю поверхность фотопроводника, а затем заставляют электроны мигрировать сквозь фотопроводник на поверхности термопластика под действием экспонирующего излучения и приложенного электрического поля. В этом случае термопластический модулятор представляет собой слоистую структуру, показанную на рис. П5.5.

Возможности применения термопластиков, обладающих хорошей чувствительностью и высокой разрешающей способностью, ограничиваются их очень узкой колоколообразной передаточной функцией: хорошо воспроизводятся лишь частоты, очень близкие к частоте резонанса, которая в свою очередь определяется толщиной слоя термопластика. Эти частоты соответствуют области от 500 до 1000 линия/мм. Невозможность записи постоянной составляющей позволяет использовать фототермопластические системы только для записи изображений типа голографической (на несущей частоте).

§ 7. Оптические пространственные модуляторы света на эффекте Поккельса [342, 367]

Пространственные модуляторы этого типа используют эффект Поккельса в кристаллах дейтерированного дигидрофосфата калия (KD_2PO_4), или просто KDP, охлажденных до точки Кюри. Адресация (т. е. запись изображения), как и в случае фототермопластика, осуществляется при помощи слоя фотопроводника и прозрачных электродов [342]. Здесь осуществляется модуляция световой волны по поляризации, которая при помощи скрещенных поляризаторов может быть преобразована в амплитудную модуляцию.

Запись изображений обычно производится на длине волны, отличной от длины волны, используемой при считывании. Эти две волны могут быть когерентными.

§ 8. Жидкие кристаллы [368]

Существует столько вариантов применения жидких кристаллов, что в нескольких строчках невозможно даже кратко все охватить. Отметим лишь два способа их использования, которые мы считаем основными:

- 1) слой жидкого кристалла помещают в кювету между двумя прозрачными электродами; пропускание материала изменяется под действием электрического поля, создаваемого напряжением на электродах;

- 2) создается слоистая структура, аналогичная приведенной на рис. П5.5; прозрачность жидкого кристалла изменяется под действием экспонирующего излучения.

В обоих случаях к основным недостаткам жидких кристаллов относятся невысокая разрешающая способность и ограниченное полное число элементов изображения.

ПРИЛОЖЕНИЕ 6

ДУВКРАТНОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ И ВЫРОЖДЕННЫЕ ВОЛНЫ

Попробуем обратить выражение (4.2), т. е. рассмотрим обратное преобразование Фурье

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(\alpha, \beta) e^{\frac{2\pi}{\lambda} i(\alpha x + \beta y)} d\alpha d\beta. \quad (\text{П6.1})$$

Здесь начинаются проблемы. Даже если фурье-образ функции $f(x, y)$ существует, нет гарантии, что интеграл (П6.1) берется. Мы получили соотношение (4.2) при условии, что [формула (4.1)]

$$\alpha^2 + \beta^2 + \gamma^2 = 1,$$

иначе говоря, при таких значениях α , β и γ , которые удовлетворяют неравенству

$$-1 \leq \frac{\alpha}{\gamma} \leq 1.$$

Но в уравнении (П6.1) бесконечные пределы интегрирования, и мы не знаем, как его проинтегрировать.

Отметим, однако, что выражение (П6.1) можно записать в ином виде. Мы можем его представить как спектр плоских волн (гл. 2, § 1, п. Б). Действительно, выражение вида

$$e^{\frac{2\pi}{\lambda} i(\alpha x + \beta y)} = \begin{cases} e^{\frac{2\pi}{\lambda} i(\alpha x + \beta y + \gamma z)} \\ \text{при } z = 0 \end{cases} \quad (\text{П6.2})$$

соответствует плоской волне с направляющими косинусами (α, β, γ) при условии, что выполняется соотношение (4.1). Рассмотрим два случая:

1) $\gamma^2 > 0$. В этом случае имеется плоская волна с амплитудой $F(\alpha, \beta)$, фаза которой определяется выражением (П6.2);

2) $\gamma^2 < 0$. В этом случае величина γ оказывается мнимой; следовательно, рассматриваемая волна есть неоднородная затухающая волна (гл. 2, § 3, п. Б) вида

$$e^{\frac{2\pi}{\lambda} [i(\alpha x + \beta y) - z \sqrt{\alpha^2 + \beta^2 - 1}]}$$

Разумеется, эти волны исчезают из-за затухания при распространении.

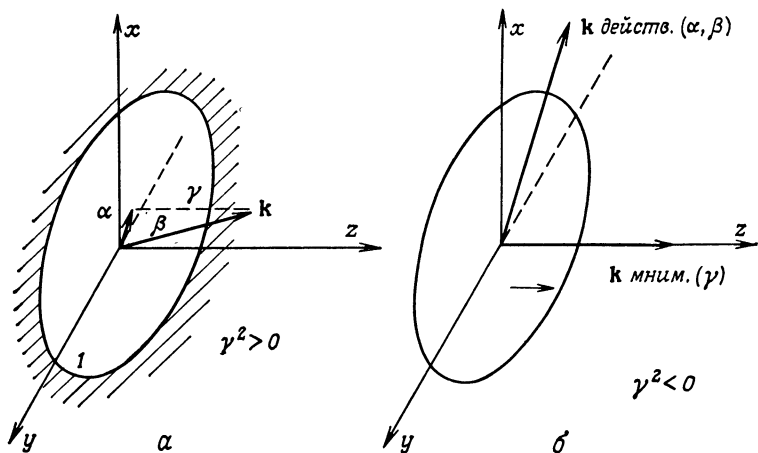


Рис. П6.1. Преобразование Фурье и спектр плоских волн.

α — случай $\gamma^2 > 0$, вектор распространения — действительная величина, $|\alpha|$ и $|\beta|$ меньше 1; δ — случай $\gamma^2 < 0$, действительная часть вектора распространения лежит в плоскости xOy . Волна неоднородная, вырожденная вдоль оси Oz .

На самом деле задачу о дифракции на окне $f(x, y)$ можно рассмотреть иначе следующим образом.

Прежде всего предположим, что для функции $f(x, y)$ существует фурье-образ $F(\alpha, \beta)$ ($-\infty < \alpha, \beta < +\infty$). Тогда можно, по крайней мере формально, написать соотношение (П6.1). Затем представим (П6.1) в виде спектра плоских волн. Если величина γ^2 положительна, то затруднений нет: в направлении $(\alpha, \beta, \sqrt{1 - \alpha^2 - \beta^2})$ получаем плоскую волну с амплитудой $F(\alpha, \beta)$; именно эту амплитуду мы обнаружим на сфере бесконечного радиуса. Отметим, что таким образом перекрываются все направления распространения в полупространстве справа (рис. П6.1, а). Если же величина γ^2 отрицательна, то тогда получаются неоднородные волны, затухающие вдоль оси z и распространяющиеся (фазовая скорость) параллельно плоскости окна (xOy) (рис. П6.1, б).

На практике двойная дифракция (прямое и обратное преобразование Фурье) не встречает трудностей в случаях, когда спектр плоских волн распространяется в параксиальной области. Система с двойной дифракцией описывается выражением

$$\mathcal{F}[\mathcal{F}(f(x, y))] = f(-x, -y),$$

чем и объясняется перевернутый характер оптических изображений.

ЛИТЕРАТУРА

1. *Feynmann R.*, Cours de physique, tome II Bilingua; Addison — Wesley, Londres, 1969.
2. *Vassallo C.*, Electromagnétisme classique dans la matière, Dunod, Paris, 1980.
3. *Gardiol F.*, Traité d'électricité, V. 3, Electromagnétisme, Ed. Georgi, 1977.
4. *Born M., Wolf E.*, Principles of Optics, Pergamon Press, 1970. [Имеется перевод 1-го издания *Борн М., Вольф Е.* Основы оптики. — М.: Наука, 1970.]
5. *Grawford F. S., Jr.*, Barkeley Physics Course, V. 3, McGraw-Hill, 1968.
6. *Young H. D.*, Fundamentals of Waves, Optics and Modern Physics, McGraw-Hill, 1976.
7. *Françon M.*, Vibrations lumineuses, Optique cohérente, Dunod, Paris, 1970.
8. *Newton I.*, Optiks, Dover, New York, 1952. [См. также: *Ньютон И.* Оптика или трактат об отражениях, преломлениях, изгибаниях и цветах света/ Пер. и примечания С. И. Вавилова, 2-е изд. — М., 1954.]
9. *Goos F., Hänchen G.*, Ann. Phys. (Leipz), **1**, 333 (1947).
10. *Artmann K.*, Ann. Phys., **6** (2), 87 (1948).
11. *McGuirk M., Carniglia C. K.*, An Angular Spectrum Representation Approach to the Goos — Hänchen Shift, Journ. Opt. Soc. Am., **67** (1), 103 (1977).
12. *Levy Y., Imbert C.*, Amplification des déplacements à la réflexion totale, Opt. Comm, **13**, 43 (1975).
13. *Hugonin J. P., Petit R.*, Etude générale des déplacements à la réflexion totale, Journ. Optics, **8** (2), 73 (1977).
14. École d'été d'électromagnétisme (Lannion, Juillet 1973), Tome III, Ondes électromagnétiques guidées par les diélectriques. Notes de cours de A. L. Cullen, rédigées par M. Chapet et M. Rousseau (Ed. CNET, Le Mezec).
15. *Goodman J. W.*, Introduction à l'optique de Fourier et à l'holographie, Masson, 1972. [Имеется перевод: *Гудмен Дж.* Введение в фурье-оптику. — М.: Мир, 1970.]
16. *Sommerfeld A.*, Optics. Lectures on Theoretical Physics. V. IV, Academic Press, 1954 [Имеется перевод: *Зоммерфельд А.* Оптика — М.: ИЛ, 1953.]
17. *Foissac Y.*, Formules nouvelles de rayonnement à l'infini des ouvertures planes. Ann. des télécoms., **30** (3—4), 97 (1975).
18. *Maréchal A.*, Traité d'optique instrumentale, l'ère section: formation des images, t. 1: Imagerie géométrique. Aberrations, Masson, 1952.
19. *Lowenthal S., Belvaux Y.*, Progrès récents en optique cohérente: filtrage des fréquences spatiales et holographie, Rev. d'Opt. th. et inst., **1** (1), 1 (1967). [Имеется перевод: *Ловенталь С., Бельво Н.* Пространственная фильтрация и голография — новое в когерентной оптике. — М.: Энергия, 1970.]
20. *Fox A. G., Li T.*, Resonant Modes in a Maser Interferometer. BSTJ., **40**, 489 (1961).
21. *Kogelnik H., Li T.*, Laser Beams and Resonators, Proc. IEEE, **54**, 1312 (1966). [Имеется перевод: ТИИЭР, 1966, т. 54, № 10.]
22. *Yariv A.*, Introduction to Optical Electronics (2nd ed.), Holt, Reinehart and Winston, 1976. [Имеется перевод: *Ярив А.* Квантовая электроника. — М.: Сов. радио, 1980.]

23. *Maréchal A., Françon M.*, Diffraction. Structure des images. Influence de la cohérence de la lumière, Masson, 1970. [Имеется перевод: *Марешаль А., Франсон М.* Структура оптического изображения. — М.: Мир, 1964.]
24. *Roubine E.*, Introduction a la théorie de la communication. t. 2, Signaux aléatoires, Masson, 1970.
25. *Bousquet P.*, Spectroscopie instrumentale, Dunod, Paris, 1969.
26. *Froehly C., Lacourt A., Viénot J. Ch.*, Notions de réponses impulsionnelles et de fonctions de transfert temporelles des pupilles optiques, justifications expérimentales et applications, *Nouv. Rev. Optique*, 4 (4), 183 (1973).
27. *Goedgebuer J. P., Lacourt A., Viénot J. Ch.*, Holographic Behaviour of Channelled Spectra in Space and Time Optics, *Opt. Comm.*, 16 (1), 99 (1976).
28. *Epworth R. E.*, The Phenomenon of Modal Noise in Analog and Digital Optical Systems, 4th European Conference on Optical Communication (Genova, Sept. 1978), p. 492.
29. *Ito K. et al.*, Optical Fiber Transmission of ITV Video Signal by Analog Baseband Modulation of Laser Diodes, 5th European Conference on Optical Communication (Amsterdam, Sept. 1979), Post dead-line paper.
30. *Françon M.*, Granularité laser. Speckle. Application en optique, Masson, 1978. [Имеется перевод: *Франсон М.* Оптика спеклов. — М.: Мир, 1980.]
31. *Beran M. J., Parent B. B., Jr.*, Theory of Partial Coherence, Prentice Hall, Englewood Cliffs, N. J., 1964.
32. *Gabor D.*, A New Microscope Principle, *Nature*, 161, 777 (1948).
33. *Leith E., Upatnieks J.*, Reconstructed Wavefronts and Communication Theory, *Journ. Opt. Soc. Am.*, 52, 1123 (1962).
34. *Vander Lugt A.*, Signal Detection by Complex Spatial Filtering, *IEEE Trans. Inf. Th.*, 139 (1964).
35. *Stroke G. W.*, Engineering Summer Conferences, Ann Arbor, Michigan, 1964.
36. *Leith E., Upatnieks J.*, Wavefront Reconstruction with Diffused Illumination and $3 \times D$ Objects Photography by Wave Front Reconstruction, *Journ. Opt. Soc. Am.*, 54 (11), 1295 (1964).
37. *Françon M.*, Holographie, Masson, Paris, 1969. [Имеется перевод: *Франсон М.* Голография. — М.: Мир, 1972.]
38. *Vienot J. C., Smigielski P.*, Holographie Optique: Développements et applications, Dunod, Paris, 1971. [Имеется перевод: *Вьено Ж.-Ш., Смигильский П., Руйе А.* Оптическая голография. — М.: Мир, 1973.]
39. *Optica Acta*, 24 (8) (Aug. 1977), Special Issue: Optics at Université de Franche Comté, Besançon, 1977.
40. *Brown B. R., Lohmann A. W.*, Computer Generated Dinary Holograms, *IBM Journ. Res. Developm.*, 13 (2), 160 (1969).
41. *Dainty J. C., Shaw R.*, Image Science, Academic Press, London, 1974, p. 276.
42. *Forshaw M. R. B.*, Thick Holograms: A Survey, *Optics and Laser Tech.*, 6, 1, 48, 28, 35 (1974).
43. *Lippmann G.*, La photo des couleurs, *Compt. Rend.*, 112, 274 (1891).
44. *Stroke G. W., Labeyrie A. E.*, White Light Reconstruction of Holographic Images Using the Lippmann — Bragg Diffraction Effect, *Phys. Letters*, 20 (4), 368 (1966).
45. *Stroke G. W., Restrick R. C.*, Holography With Spatially Non Coherent Light, *Appl. Phys. Letters*, 7/9, 229 (1965).
46. *Benton S. A.* — in: Applications of Holography and Optical Data Processing, *Proc. of the I. C. O.*, Jerusalem, 1976, p. 401.
47. *Fleuret J.*, Holographie calculée, *Nouv. Rev. d'Opt.*, 5 (4), 219 (1974).
48. *Lee W. H.*, Computer Generated Holograms Techniques and Applications, *Progress in optics* (Wolf ed.), Vol. XVI, North Holland, 1979.
49. *Huang T. S.*, Digital Holography, *Proc. IEEE*, 9, 1335 (1971). [Имеется перевод: — ТИИЭР, 1971, т. 52, № 9, с. 63]

50. Dallas W. J., Deterministic Diffusers for Holography, *Appl. Optics*, **12** (6), 1179 (1973).
51. Lohmann A. W., Paris A., Binary Fraunhofer Holograms Generated by Computer, *Appl. Optics*, **6** (10), 1739 (1967).
52. Chavel P., Hugonin J. P., High Quality Computer Holograms, *Journ. Opt. Soc. Am.*, **66**, 989 (1976).
53. Lesem L. B., Kirsch P. M., Jordan J. A., The Kinoform: a New Wavefront Reconstruction Device, *IBM Journ. Res. Dev.*, **2**, 150 (1969).
54. Lee W. H., Sampled Fourier Transform Hologram Generated by Computer, *Appl. Optics*, **9** (3), 639 (1970).
55. Burkhard C. B., A Simplification of Lee's Method of Generating Holograms by Computer, *Appl. Optics*, **9** (8), 1949 (1970).
56. Goodman J. W., Chu C., Spectrum Shaping With Parity Sequences, *Appl. Optics*, **11** (8), 1716 (1972).
57. Severcan M., Computed Holograms: Improving the Bred, Ph. D. Thesis, Technical Report No. 6415-6 NTIS Springfield.
58. Dallas W. J., Kinoform Fabrication — a New Method, *Opt. Comm.*, **8**, 341 (1973).
59. Fleuret J., Thèse d'Etat Université de Paris Sud, Mars 1977.
60. Chu R. C., Fienup J. R., Goodman J. W., Multi-Emulsion On-Axis Computer Generated Holograms, *Appl. Optics*, **12**, 1386 (1973).
61. 1er Congrès Européen de Métrologie optique, Proc. of the SPIEE, vol. 133, Strasbourg, 1977.
62. Schumann W., Dubas M., Holographic Interferometry, Topics in Applied Physics, XVI, Springer Verlag, Berlin, 1979.
63. Benton S. A., Holographic Displays. *Opt. Eng.*, **14** (5), 402 (1975).
64. Okoshi T., Three-Dimensional Imaging Techniques, Academic Press, 1976.
65. Knight G. R., Holographic Memories. *Opt. Eng.*, **14** (5), 453 (1975).
66. Chen D., Zook J. D., An Overview of Optical Data Storage, *Proc. IEEE*, **63** (8), 1207 (1975). [Имеется перевод: ТИИЭР, 1975, т. 63, № 8.]
67. Hacking K., Riley J. L., Digital Recording Using Holograms, BBC Report, August 1979.
68. Huignard J. P., Contribution à l'étude des mémoires holographiques, Thèse Université de Paris Sud, No. 207, Janv. 1974.
69. Hill B., Krumme J. P., Moch G., Polycube Optical Memory, *Appl. Optics*, **14** (11), 2607 (1975).
70. Bolusset S., Treneux M., Système de stockage d'informations analogiques par superposition de réseaux optiques modulés, *Nouv. Rev. d'Opt.*, **5** (6), 333 (1974).
71. Beiser L., Laser Scanning Systems. — In: *Laser Applications*, ed. M. Ross, Academic Press, New York, 1974.
72. Maréchal A., Croce P., Un filtre de fréquences spatiales pour l'amélioration du contraste des images optiques, *Compt. Rend.*, **237**, No. 12 (1953).
73. Bonnet G., Réponse percussive spatio-temporelle en diffraction polychromatique, *Ann. Télécomm.*, **30** (7, 8), 251 (1975).
74. Bonnet G., Introduction à l'optique métaxiale, *Ann. Télécomm.*, **33** (65), 1 (1978); **33** (66), 1 (1978).
75. Hellstrom C. W., Image Restoration by Method of Least Squares, *Journ. Opt. Soc. Am.*, **57**, (3), 297 (1967).
76. Andrews H. C., Hunt B. R., *Digital Image Restoration*, Prentice Hall, 1977.
77. Phillips D. L., A Technique for the Numerical Solutions of Certain Integral Equations of the First Kind, *Journ. Ass. Comp. Math.*, **9**, 84 (1962).
78. Twomey S., The Application of Numerical Filtering to the Solution of Integral Equations Encountered in Indirect Sensing Measurements, *Journ. Franklin Inst.*, **279**, 95 (1965).
79. Huang T. S., *Picture Processing and Digital Filtering*, Springer Verlag, New York, 1975. [Имеется перевод: Обработка изображений и цифровая фильтрация/Под ред. Т. Хуанга. — М.: Мир, 1979.]

80. Lowenthal S., Belvaux Y., Reconnaissance des formes en optique par traitement des signaux dérivés, *Compt. Rend.*, **262** (7), 413 (1966).
81. Fleuret J., Maître H., Optimization of a Binary Transcoding Single Filter Used for Character Recognition, *Opt. Comm.*, **17** (1), 64 (1976).
82. Bryngdahl O., Computer-Generated Holograms as Generalized Optical Components, *Opt. Eng.*, **14** (5), 426 (1975).
83. Lee S. H., Review of Coherent Optical Processing, *Appl. Phys.*, **10**, 203 (1976).
84. Marquet M., Tsujiuchi J., Interprétation des aspects particuliers des images obtenues dans une expérience de détramage, *Optica Acta*, **8**, 267 (1961).
85. Kato H., Goodman W. J., Nonlinear Transformations and Logarithmic Filtering in Coherent Optical Systems, *Opt. Comm.*, **8** (4), 378 (1973).
86. Strand T. C., Lohmann A. W., Analog to Digital Conversion of Pictures with Optical Means, *Proc. Elect. Optic. Conf. (Anaheim, 1975)*.
87. Casasent D., Psaltis D., Deformation-Invariant, Space-Variant Optical Pattern Recognition, *Progress in Optics* (ed. Wolf), **XVI**, 291 (1978).
88. Lowenthal S., Chavel P., Noise Problems in Optical Image Processing, *Proc. ICO (Jerusalem, 1976)*, Pergamon, New-York, 1977, p. 45.
89. Duffieux P. M., L'intégrale de Fourier et ses applications à l'optique, Université de Besançon, Réédition Masson, Paris, 1946.
90. Lowenthal S., Chavel P., A Method of Incoherent Optical Image Processing Using Synthetic Holograms, *Journ. Opt. Soc. Am.*, **66**, 14 (1976).
91. Ebersole J. F., Optical Image Subtraction, *Opt. Eng.*, **14** (5), 436 (1975).
92. Collins L. F., Difference Holography, *Appl. Optics*, **7**, 203 (1968).
93. Pennington W., Grid Coding: a Novel Technique for Image Processing, *Proc. IEEE*, **60**, 669 (1972).
94. Françon M., New Method of Optical Processing Using a Random Diffuser, *Optical Acta*, **20** (1), 1 (1973).
95. Belvaux Y., Lowenthal S., Subtraction (or Addition) of Illuminance, Space Optics, *Proc. of the 9th Congress of the ICO, 1972*, p. 613.
96. Debrus S., Kinany R., Detection of Differences Between Two Images Using The Weigert Effect, *Journ. Optics*, **10** (3), 119 (1979).
97. Lohmann A. W., Rhodes W. T., Two Pupils Synthesis of Optical Transfer Functions, *Appl. Optics*, **17** (7), 1141 (1978).
98. Lowenthal S., Werts A., Filtrage des fréquences spatiales en lumière incohérente à l'aide d'hologrammes, *Compt. Rend., Paris*, **B266**, 542 (1968).
99. Maloney W. T., Real Time Filtering of Oscilloscope Traces, *Appl. Optics*, **10** (11), 2554 (1971).
100. Rogers G. L., *Non Coherent Image Processing*, John Wiley, USA, 1977.
101. Aunis M., Communications optiques: les problèmes de propagation, Rapport interne CNET, Nov. 1971.
102. Goubau G., Schwering F., On the Guided Propagation of Electromagnetic Wave Beams, *IRE Trans. Anten. Propag.*, **3-AP-9**, 248 (Mai 1961).
103. Marcuse D., Miller S., Analysis of a Tubular Gas Lens, *BSTJ*, **43** (4), 1759 (1964).
104. Marie P., Guidage de la lumière cohérente par un guide hélicoïdal, *Ann. des Télécom.*, **25** (9—10), 320 (1970).
105. Van Heel A. C. S. A New Method of Transporting Optical Images Without Aberrations, *Nature*, **73**, 39 (1954).
106. Hopkins H. H., Kapany N. S., A Flexible Fibrescope Using Static Scanning, *Nature*, **73**, 39 (1954).
107. Miya T. et al., Ultimate Low Loss Single Mode Fibre at 1.55 M, *Electr. Letters*, **15** (4), 106 (1979).
108. Sentsui S. et al., Low Loss Monomode Fibers With P_2O_5 — SiO_2 Cladding In the Wavelength Region 1.2—1.6 M, Fifth European Conference on Optical Communication (Amsterdam, 17—19 September 1979), Paper 5.3.
109. Yokota H. et al., Long Length Single-Mode Fiber With Low Attenuation In the Dispersion-Free Region, Fifth European Conference on Optical Communication (Amsterdam, 17—19 September, 1979), Paper 5.4.

110. Ito T. et al., Non-Repeated 50 km Transmission Experiment Using Low-Loss Optical Fibres, *Electr. Letters*, **14** (16), 520 (1978).
111. Arnaud J. A., Theory of the Simple Material, Helicoidal Fiber, *BSTJ*, **53**(8), 1643 (1974).
112. Marcuse P., Theory of the Single Material Fiber, *BSTJ*, **53** (8), 1619 (1974).
113. Snyder A. W., Mitchell D. J., Leaky Rays Cause Failure of Geometric Optics on Optical Fibers, *Electr. Letters*, **9**, No. 19 (1973).
114. Snyder A. W., Leaky Ray Theory of Optical Waveguides of Circular Cross Section. *Appl. Physics*, **4**, 273 (1974).
115. Marcuse D., *Light Transmission Optics*, Van Nostrand Reinhold, 1972. [Имеется перевод: Маркузе Д. Оптические волноводы. — М.: Мир, 1974.]
116. Gloge D., Marcatili E. A. J., Impulse Response of Fiber with Ring Shaped Parabolic Index Distribution, *BSTJ*, **52** (7), 1161 (1973).
117. Gloge D., Marcatili E. A. J., Multimode Theory of Graded Core Fibers, *BSTJ*, **52** (9), 1563 (1973).
118. Cozannet A., Treneux M., Skew Rays in Optical Fibers, *Appl. Optics*, **14** (6), 1345 (1975).
119. Kittel C., *Introduction à la physique de l'état solide*, Dunod, 1970, p. 361. [Имеется перевод английского издания: Киттель Ч. Введение в физику твердого тела. — М.: Наука, 1978.]
120. Barnoski M. K., *Fundamentals of Optical Fiber Communications*, Academic Press, 1976.
121. Gloge D., Propagation Effects in Optical Fibers, *IEEE Trans. Microwave Theory Techn.*, **MTT-23** (1), 106 (1975).
122. Abramowitz M., Stegun I., *Handbook of Mathematical Functions*, Dover Publications, N. Y., 9 ed. [Имеется перевод: Справочник по специальным функциям с формулами, графиками и математическими таблицами/Под ред. М. Абрамовица, И. Стигана. — М.: Наука, 1979.]
123. Shesinger S. P., Diamant P., Vignants A., On Higher Hybrid Modes of Dielectric Cylinders, *IRE Trans. Microwave Theory Techn.*, (March 1960), p. 252.
124. Biernson G., Kinsley D. J., Mode Patterns in Dielectric Waveguide, *IEEE on Microwave Theory Tech.*: **MTT-13** (3), 345 (1965).
125. Snitzer E., Cylindrical Dielectric Waveguides Modes, *JOSA*, **51**, 491 (1961).
126. Bochove E. J., Cozannet A., Low Pass Filter Characteristics of Multimode Graded Index Fibers, *Opt. Quant. Electr.* **9**, 135 (1977).
127. Bochove E. J., On the Group Delay Time of Skew Rays in Graded Index Fibers, *Opt. Quant. Electr.*, **9**, 259 (1977).
128. Sladen F. et al., Profile Dispersion Measurements for Optical Fibers Over the Wavelength Range 350 NM to 1900 NM, 4th European Conference on Optical Communication (Genova, Sept. 1978), p. 48.
129. Vassallo C., Linear Power Responses of an Optical Fiber, *IEEE Trans. on MTT*, **MTT-25** (7), 572 (1977).
130. Gloge D., Optical Power Flow in Multimode Fibers, *BSTJ*, **51**, 1767 (1972).
131. Gloge D., Impulse Response of Clad Optical Multimode Fibers, *BSTJ*, **52** (6), 801 (1973).
132. Jeunhomme L., Thèse d'Etat: Rôle du couplage de modes dans la transmission de la lumière par fibres optiques multimodes, Paris, VI, 1978.
133. Gloge D., Weakly Guiding Fibers, *Appl. Optics*, **10**, 2252 (1971).
134. Marcuse D., Loss Analysis of Single Mode Fiber Splices, *BSTJ*, **56**, 703 (1977).
135. Gambling W. A., Matsumura H., Propagation in Radially Inhomogeneous Single Mode Fibers, *Opt. Quant. Electr.*, **10**, 31 (1978).
136. Petermann K., Theory of Microbending Loss in Monomode Fibers Arbitrary Refractive Index Profile, *AEÜ*, **30**, 337 (1976).
137. Rashleigh S. C., Ulrich R., Polarization Mode Dispersion in Single Mode Fibers, *Optic Lett.* (3 Aug. 1978), p. 60.
138. Gloge D., Dispersion in Weakly Guiding Fibers, *Appl. Optics*, **10**, 2442 (1971).

139. *White K. J., Nelson B. P.*, Zero Total Dispersion in Step Index Monomode Fibers at 1.3 and 1.55 M, *Electr. Letters*, **15** (13), 396 (1979).
140. *Gambling W. A. et al.*, Zero Mode Dispersion in Single Mode Fibers, *Electr. Letters*, **14**, 618 (1978).
141. *Keck D. B.*, Optical-Fiber Measurements, Topical Meeting on Optical Fiber Communication (Washington, March 1979), Paper WA5.
142. *Gloge D.*, Bending Loss in Multimode Fibers with Graded and Ungraded Core Index, *Appl. Optics*, **11**, 2506 (1972).
143. *Unger H. G.*, Planar Optical Waveguides and Fibers, Clarendon Press, Oxford, 1977, Ch. 6.
144. *Jeunhomme L. et al.*, Mesure des caractéristiques de transmission des conducteurs optiques, 2ème Colloque européen sur les transmissions par fibres optiques (Paris, 27—30 Sept. 1976), p. 123.
145. *Love W. F.*, Novel Mode Scrambler for Use in Optical-Fiber Bandwidth Measurements. — in: [141], Paper ThG2.
146. *Barnoski M. K. et al.*, OTDR Differential Modal Attenuation Measurements, 4th European Conference on Optical Communication (Genova, Sept. 1978), p. 133.
147. *Mathijsse P., de Blok C. M.*, Field Measurement of Splice Loss Applying the Backscattering Method, *Electr. Letters*, **15** (24), 795 (1979).
148. *Gloge D. et al.*, Profile Dispersion in Multimode Fibers: Measurement and Analysis, *Electr. Letters*, **11** (19), 469 (1975).
149. *Eickhoff W., Weidel E.*, Measuring Method for the Refractive Index Profile of Optical Glass Fibers, *Opt. Quant. Electr.*, **7**, 109 (1975).
150. *Sladen F. et al.*, Determination of Optical Fiber Refractive Index Profiles by Near Field Scanning Technique, *Appl. Phys. Letters*, **28** (5), 255 (1976).
151. *Stewart W.*, A New Technique for Measuring the Refractive Index Profiles of Graded Optical Fibers, IOOC, Tokyo, July 1977, C. 2.2.
152. *White K. I.*, The Measurement of the Refraction Index Profiles by the Refracted Near Field Technique. — in: [128], p. 146.
153. *Presby H. M.*, Refractive Index and Diameter Measurements of Unclad Optical Fibers, *JOSA*, **64** (3), 280 (1974).
154. *Okoshi T., Hotate K.*, Refractive Index Profile of an Optical Fiber: Its Measurement by the Scattering-Pattern Method, *Appl. Optics*, **15** (11), 2756 (1976).
155. *Sladen F., Payne D. N., Adams M. J.*, Measurement of Profile Dispersion in Optical Fibers: a Direct Technique, *Electr. Letters*, **13** (7), 212 (1977).
156. *Presby H. M. et al.*, Binary Silica Optical Fibers: Index and Dispersion Measurement, *Appl. Optics*, **15** (12), 3029 (1976).
157. *Payne D. M., Hartog A. H.*, Determination of the Wavelength of Zero Material Dispersion in Optical Fibers by Pulse-Delay Measurements, *Electr. Letters*, **13**, 627 (1977).
158. *Trenew M., Bouillie R.*, Bandwidth Properties of Optical Fibers, Topical Meeting on Optical Fibers (Williamsburg, Jan. 1975).
159. *Cozannet A., Jeunhomme L.*, Characterization of Optical Fibers and Components by Time Domain Measurements, Ecole d'été du Trégor, 1978, C. 14.
160. *Olshansky R., Keck D. R.*, Pulse Broadening in Graded-Index Optical Fibers, *Appl. Optics*, **15** (2), 483 (1976).
161. *Cohen L. G., Lin C.*, Pulse Delay Measurements in the Zero Dispersion Wavelength Region for Optical Fibers, *Appl. Optics*, **16** (12), 3136 (1977).
162. *Boisrobert C. et al.*, Sweep Frequency Transfer Function Measurement Applied to Optical Fiber, *IEEE Trans. Instrum Measur.* (Dec. 1976), p. 294.
163. *Jeunhomme L. et al.*, Wavelength Dependence of Modal Dispersion in Graded Index Optical Fibers, *Electr. Letters*, **14** (12.8), 364 (1978).
164. *Olshansky R.*, Differential Mode Attenuation in Graded-Index Optical Waveguides, IOOC, Tokyo, July 1977, p. 423.
165. *Passaret M., Regreny A.*, Technologie des fibres optiques, Technique de l'ingénieur, document E 2908

166. *Gloge D.*, Optical Fiber Technology, IEEE Press, New York, 1975.
167. *Schultz P. C.*, Progress in Optical Waveguide Process and Materials, Appl. Opt., **18** (21), 3684 (1979).
168. *Koizumi K. et al.*, New Light-Focusing Fibers Made by a Continuous Process, Appl. Optics, **13** (2), 255 (1974).
169. *Macedo P. et al.*, Molecular Stuffing of Phasil Glasses for Graded Index Optical Fibers, 2ème Colloque européen sur les communication optiques (Paris, Sept. 1976), p. 37.
170. *Keck D. B., Schultz P. C.*, 1973, US Patent 3711-262.
171. *Keck D. B. et al.*, 1973, US Patent 3737-292.
172. *French W. et al.*, Glass Fibers for Optical Communications, Ann. Rev. Material Science, **5**, 373 (1975).
173. *Kuppers D., Lydtin H.*, The Preparation of Optical Waveguides by Means of CVD Techniques, Atlanta CVD Conference (October 1977).
174. *Jzawa T. et al.*, Progress in Continuous Fabrication Process of High Silica Fiber Preforms. — in: [128], p. 30, Paper 1.5.
175. *Tanaka S. et al.*, Silicone Clad Fused Silica Core Fiber, Electr. Lett., **11** (7), 153 (1975).
176. *Bouillie R., Le Noane G.*, Optical Cable Technologies and Performances for Telecommunication Applications, Fourth European Conference on Optical Communication (Genova, September 1978), p. 190.
177. *Inada K. et al.*, Cabling of Ultra-Low Loss Optical Fibers, Fourth European Conference on Optical Communication (Genova, September 1978), p. 200.
178. *Yonechi S. et al.*, Characteristics of Optical Fiber Cables with Spacers, Topical Meeting on Optical Fiber Communication (Washington, March 6—8, 1979), Paper TUE2.
179. *Josteur R.*, Réalisation d'un câble à conducteurs optiques pour les transmissions numériques dans les systèmes de télécommunications, Deuxième Colloque européen sur les transmissions par fibres optiques (Paris, September 1976), p. 193.
180. *Ishihara K. et al.*, Determination of Optimum Structure in Coated Optical Fiber and Cable Unit, Fifth European Conference on Optical Communication (Amsterdam, September 1979), Paper 7.3.
181. *Schwartz M. I.*, Optical Cable Design Associated with Splicing Requirements. — in: [179], p. 325, 311.
182. *Inada K. et al.*, High Density Low-Loss Fiber Unit and the Cabling. — in: [180], Paper 7.5.
183. *Le Noane G.*, Optical Fiber Cable and Splicing Techniques, — in: [179], p. 247.
184. *King F. D., Kapron F. P.*, Rugged Optical Cables Having Attenuation Below 3 dB/km at 0.85 M. — in: [178], Paper TUE4.
185. *Fenton K. J., McCartney R. C.*, Loss Factors to Consider when Designing or Evaluating Single Fiber Separable Connectors for Fiber Optic Data Transmission Systems, Insulation Circuits, Aug. 1977, p. 27.
186. *Dalgleich J. F.*, Connections, Electronics (Aug. 1976), p. 96. [Имеется перевод: — Электроника, 1976, т. 49, № 16, с. 45.]
187. *Furuta H. et al.*, Optical Fiber Connector, Fujitsu Scientific and Technical Journal, **14** (1), 119 (1978).
188. *Sakamoto K. et al.*, The Automatic Splicing Machine Employing Electric Arc Fusion. — in: [176], p. 296.
189. *Ogal M. et al.*, Multifiber Splicing Using Slippery Fiber Holder. — in: [178], Paper WG7.
190. *Bowen T., Schumacher W.*, Fiber Optic Connector Developments: Moving to Annul Coupling Mismatches, Microwave Journ. (July 1979), p. 55.
191. *Le Noane G.*, Low-Loss Optical Fibre Connection Systems, Electr. Letters, **15** (1), 12 (1979).
192. *Jocteur R., Tardy A.*, Optical Fibers Splicing with Plasma Torch and Oxhydic Microburner. — in: [179], p. 261.

193. *Kinoshita K., Kobayashi M.*, End Preparation and Fusion Splicing of an Optical Fiber Array with a CO₂ Laser, *Appl. Optics*, **18** (19), 3256 (1979).
194. *Lengyel B. A.*, Introduction to Lasers Physics, Wiley, New York, 1966.
195. *Hartmann F.*, Les Lasers, PUF, Paris, 1974.
196. *Smith W. V., Sorokin P. P.*, The Laser, McGraw-Hill, New York, 1966.
197. *Bruhat G., Kastler A.*, Optique, Masson, Paris (6ème éd. 1975).
198. *Louisell W. H.*, Radiation and Noise in Quantum Electronics, McGraw-Hill, 1965.
199. *Landau L., Lifchitz E.*, Physique statistique, Éditions Mir, Moscou, 1967. [См. также: *Ландау Л. Д., Лифшиц Е. М.* Статистическая физика. — М. — Л.: Гостехиздат, 1951.]
200. *Bertein F.*, Electronique quantique, Eyrolles, Paris, tome 1, 1965; tome 2, 1969. [Имеется перевод: *Бертэн Ф.* Основы квантовой электроники. — М.: Мир, 1971.]
201. *Kressel H., Rutler J. K.*, Semiconductor Lasers and Heterojunction LEDs, Quantum Electronics, Academic Press, New York, 1977.
202. *Lyon Caen R.*, Diodes et transistors utilisés en commutation, Masson, Paris, 1965.
203. *Sze S. M.*, Physics of Semiconductor Devices, J. Wiley International, 1969.
204. *Kogelnik H., Li T.*, Laser Beams and Resonators, *Appl. Opt.*, **5**, 1550 (1966).
205. *Panish M. B.*, Heterostructure Injection Lasers, *IEEE Trans. on MTT*, **23** (Jan. 1975).
206. [201], Ch. 8.
207. [201], Ch. 12.
208. *Casey H. C., Jr., Panish M. R.*, Heterostructure Lasers, Quantum Electronics, Academic Press, New York, 1978.
209. *Scifres D. R. et al.*, Distributed Feedback GaAs—GaAlAs Diode Lasers, *SPIE*, Vol. 77, 1976.
210. *Sommers H. S.*, Performances of Injection Lasers with External Gratings, *RCA Review*, **38**, 33 (1977).
211. *Suematsu Y.*, Sources for Long Wavelength Optical Communication Systems. — in: [108], Paper 2.1.
212. *Bergh A. A., Dean P. J.*, Light Emitting Diodes, Clarendon Press, Oxford, 1976.
213. *Burrus C. A., Dawson R. W.*, Small-Area High-Current-Density GaAs Electroluminescent Diodes, *Appl. Phys. Letters*, **17**, 97 (1970).
214. *King F. D. et al.*, High Power Long-Lived Double Heterostructure LED's for Optical Communications, *Tech. Dig. Int. Electron Devices Meeting* (Washington D. C., Dec. 1975), p. 480.
215. *Smith D. R. et al.*, Receivers for Optical Communications: a Comparison of Avalanche Photodiodes with PIN-FET Hybrids, *Opt. Quant. Electr.*, **10**, 293 (1978).
216. *Boisrobert C. Y., Leboutet A.*, Utilization d'un photodétecteur type PIN dans un système de transmission à fibre optique, *Ann. Télécomm.*, **29**, (5—6), 227 (1974).
217. *Personick S. D.*, Receiver Design for Digital Fiber Optic Communications Systems, *BSTJ*, **52** (6), 843 (1973).
218. *Melchior H.*, Sensitive High Speed Photodetectors for the Demodulation of Visible and Near Infrared Light, *Journ. Luminescence*, **7**, 390 (1973).
219. *Melchior H. et al.*, Photodetectors for Optical Communication systems, *Proc. IEEE*, **58** (10), 1466 (1970). [Имеется перевод: — ТИИЭР, 1970, т. 58, № 10.]
220. *Webb P. P. et al.*, Properties of Avalanche Photodiodes, *RCA Review*, **35**, 235 (1974).
221. *Nishiba K.*, Japan Journ. Appl. Physics, **9**, 481 (1970).
222. *McIntyre R. J.*, Multiplication Noise in Uniform Avalanche Diodes, *IEEE Trans. Electron Devices*, **ED-13**, No 1 (1966).

223. Miller S. M., Avalanche Breakdown in Germanium, Phys. Rev., **99**, 1234 (1955).
224. Tomasetta L. R. et al., High Sensitivity Optical Receivers for 1.0—1.4 M Fiber-Optic Systems, IEEE Trans. Quant. Electr., **14** (11), 800 (1978).
225. Stillman G. E., Wolfe C. M., Avalanche Photodiodes, Semiconductor and Semimetals, vol. 12, Academic Press, New York, 1977, Ch. 5.
226. Nye J. F., Physical Properties of Crystals, Clarendon Press, 1976. [Имеется перевод: Най Дж. Физические свойства кристаллов и их описание при помощи тензоров и матриц. — М.: Мир, 1967.]
227. Le Mezeq J., Hanaff J., Chau N. N., L'onde électrique, t. XLVI (470) (1966).
228. IEEE Trans. on Microwave Theory and Techn., **MTT-23** (1) (1975), Special Issue: Integrated Optics and Optical Waveguides.
229. Sapriel J., L'acousto-optique, Collection de monographie de physique, Masson, 1976.
230. Klein W. R., Cook B. D., Unified Approach to Ultrasonic Light Diffraction, Trans. on Sonics and Ultrasonics, **SU-14** (3), 123 (1967).
231. Favre F., Rivoallan L., Modulateurs et commutateurs électro-optiques, Opt. Comm., **24** (1), 139 (1978).
232. Papuchon M., Puech C., L'optique intégrée: de nouvelles possibilités pour la modulation et la commutation de la lumière, Entropie, **77** (13), 13 (1977).
233. Omachi Y., Noda J., Appl. Phys. Lett., **27** (10), 544 (1975).
234. Hammer J. et al., IEEE Journ., **QE11**, No. 138 (1975).
235. Introduction to Integrated Optics, ed. M. K. Barnoski, Plenum Press, 1974. [Имеется перевод: Введение в интегральную оптику/Под ред. М. Барноски. — М.: Мир, 1977.]
236. Tamir T., Integrated Optics, Topics in Applied Physics, vol. 7, Springer Verlag, 1975. [Имеется перевод: Интегральная оптика/Под ред. Т. Тамира. — М.: Мир, 1978.]
237. Ostrowsky D. B., Fiber and Integrated Optics, Plenum Press, 1979.
238. Clair J. J., Télécommunications optiques, Introduction à l'optique intégrée, Masson, Paris, 1977 [Имеется перевод: Клар Ж.-Ж. Введение в интегральную оптику. — М.: Сов. радио, 1980.]
239. Cattermole K. W., Principles of Pulse Code Modulation, Illife, London, 1969.
240. Pressicaud J. M. et al., Equipements de multiplexage MIC 30 voies TNEI de la deuxième génération, Câbles et transmission, No. 2 (April 1978), p. 152.
241. Daguet J. L. et al., Système MIC à 30 voies à codage individuel, Câbles et Transmission, numéro particulier, transmissions numériques, Déc. 1975, p. 184.
242. Colavito C., La transmission numérique dans les systèmes radioélectriques, Journ. Télécomm., **45**, 374 (1978).
243. Lucky R. W. et al., Principles of Data Communication, McGraw Hill, New York, 1968.
244. Macchi C., Guilbert J. F., Téléinformatique, Dunod, Paris, 1979.
245. Rousseau M., Transmission Code and Receiver Selection for Optical Fibers PCM Communications, 1st European Conf. on Optical Fiber Communication, IEE Conf. Public., No. 132, p. 174.
246. Blackmore R. W., Fell P. H. 8.448 Mb/s Optical Fiber System. — in: [245], p. 182.
247. Takasaki Y. et al., Optical Pulse Formats for Fiber Optic Digital Communications, IEEE Trans. Comm., **COM-24**, 404 (1976).
248. Rousseau M., Application des chaines de Markov au calcul des densités spectrales de puissance de messages binaires, Ann. télécomm., **31** (1—2), 8 (1976).
249. Rousseau M., Block Codes for Optical Fiber Communication, Electr. Lett., **12** (18), 478 (1976).

250. *Osafune M., Higo Y.*, A Terminal Equipment for Optical Hybrid Digital Transmission Line with 2R-Repeater, 3rd European Conference on Optical Communication (Munich, Sept. 1977).
251. *Rousseau M.*, Réception des messages biphase dans les systèmes de télécommunications par fibres optiques, Ann. Télécom., **32** (7—8), 268 (1977).
252. *Le Men P. et al.*, Optimization of a 140 Mbit/s Optical Fiber Transmission System, 4th European Conference on Optical Communication (Genova, Sept. 1978), p. 530.
253. *Smith D. R. et al.*, Pin-Photodiode Hybrid Optical Receivers, Fifth European Conference on Optical Communications (Amsterdam, Sept. 1979), p. 13.4.
254. *Goranson R. W., Skipper J. D.*, Signal to Noise Ratio in Optical Photodiode Receivers, Proc. IEEE, (Oct. 1974), p. 1404. [Имеется перевод: — ТИИЭР, 1974, т. 62, № 10.]
255. *Ueno Y. et al.*, A 40 Mb/s and a 400 Mb/s Repeater for Fiber Optic Communication. — in: [245], p. 147.
256. *Hullett J. L., Muoi T. V.*, A Feedback Receive Amplifier for Optical Transmission Systems, IEEE Trans. Comm., **COM-24**, 1180 (1976).
257. *Hubbard W. M.*, Utilization of Optical-Frequency Carriers for Low and Moderate Bandwidth Channels, Bell Syst. Tech. Journ., **52** (5), 731 (1973).
258. *Dogliotti R. et al.*, Baseband Equalization in Fiber Optic Digital Transmission, Optical Quantum Electr., **8**, 343 (1976).
259. *Personick S. D. et al.*, A Detailed Comparison of Four Approaches to the Calculation of the Sensivity of Optical Fiber Receivers, IEEE Trans. Comm., **COM-25**, 541 (1977).
260. *Personick S. D.*, Receiver Design for Optical Fiber Systems, Proc. IEEE, **65** (12), 1670 (1977). [Имеется перевод: — ТИИЭР, 1977, т. 65, № 12.]
261. *Goell J. E.*, An Optical Repeater with High Impedance Input Amplifier, Bell Syst. Tech. Journ., **53** (4), 629 (1974).
262. *Hata S. et al.*, Performance of p—i—n Photodiode Compared with Avalanche Photodiode in the Longer-Wavelength Region of 1 to 2 M, Electr. Letters, **13** (22), 668 (1977).
263. *Kao C. K., Goell J. E.*, Design Process for Fiber-Optic Systems Follows Familiar Rules, Electronics (16 Sept. 1976), p. 113. [Имеется перевод: — Электроника, 1976, т. 49, № 19, с. 51.]
264. *Elion G., Elion H.*, Fiber Optics in Communications Systems, M. Dekker, 1978.
265. *Bassen H. I., Hoss R. J.*, An Optically Linked Telemetry System for Use with Electromagnetic-Field Measurement Probes, IEEE Trans. Electromagnetic Compatibility, **EMC-20** (4), 483 (1978).
266. *Hollis S., Ecker A.*, Fiber Optics for CATV in Perspective, IEEE Trans. Cable Telev., **CATV-2** (4), 154 (1977).
267. *Schrock C. B.*, Proposal for a Hub Controlled Cable Television System Using Optical Fiber, IEEE Trans. Cable Telev., **CATV-4** (2), 70 (1979).
268. *Kawahata M., Finley M. R.*, Optical Visual Information Systems — Some First Results. — in: [178], Paper TuB2.
269. *Glowinski A. et al.*, Prospective et évaluation des services future, 3rd World Telecommunication Forum, (Genève, 1979), Paper 3.1.6.
270. *Toms N.*, An Integrated Network Using Fiber Optics (INFO) for the Distribution of Video, Data and Telephony in Rural Areas, IEEE Transactions on Communications, **COM-26** (7), 1037 (1978).
271. *Kawamura K., Kohnushi K.*, Optical Picture Transmission System, 5th European Conference on Optical Communication (Amsterdam, Sept. 1979), Paper 20.
272. *Strauss J.*, Linear Transmitters for Fiber Optic Communications Using High Radiance LED's, 2ème Colloque Européen sur les transmissions par fibres optiques (Paris, Sept. 1976), p. 209.
273. *Asatani K., Kimura T.*, Linearization of LED Non-Linearity by Predistorsions, IEEE Trans. Electron Devices, **ED-25** (2), 207 (1978).

274. Carratt M., Tache J. P., Aerial Installation of an Optical Cable and Transmission of a 4 TV Channels on a Fiber. — in: [252], p. 664.
275. Ueno Y., Kajitani M., Color TV Transmission Using Light Emitting Diode, NEC Res. Developm. (35), 15 (1974).
276. Dupraz J., Théorie de la communication, Eyrolles, Paris, 1973.
277. Schwartz M. et al., Communication Systems and Techniques, McGraw-Hill, New York, 1966.
278. Libois L. J., Faisceaux hertziens et systèmes de modulation, Collec. scient. et techn. du CNET, Chiron, 1958.
279. Michaelis T. D., Laser Diode Evaluation for Optical Analog Link, IEEE Trans. on Cable Television, **CATV-4** (1), 30 (1979).
280. Sato K., Asatani K., Analogue Baseband TV Transmission Experiments Using Semiconductor Laser Diodes, Electr. Letters, **15** (24), 794 (1979).
281. Chen F. S. et al., Laser Transmitter for 70 MHz Entrance Links, BSTJ, **58** (7), 1617 (1979).
282. Albanese A., Video Transmission Tests Performed on Intermediate Frequency Lightwave Entrance Links, SMPTE Journ., **87**, 821 (1978).
283. Sato M. et al., A New Optical Communication System Using the Pulse Interval and Width Modulated Code, IEEE Trans. on Cable Television, **CATV-4** (1), 1 (1979).
284. Timmermann C., Signal to Noise Ratio of a Video Signal Transmitted by a Fiber-Optic System Using Pulse-Frequency Modulation, IEEE Trans. Broadcasting, **BC-23** (1), 12 (1977).
285. Takasaki Y., New Fiber Optic Analog Baseband Transmission Plan for Color TV Signals, IEEE Trans. on Communications, **COM-26** (6), 902 (1978).
286. Miller S. E., Chynoweth A. G., Optical Fiber Telecommunications, Academic Press, N. Y., 1979.
287. Midwinter J. E., Optical Fibers for Transmission, Wiley Interscience, N. Y., 1979.
288. Boirat M., Communications par fibres optiques, Entropie, No. 77 (1977).
289. Brosio A. et al., Level Control Circuit for Injection Laser Transmitters. — in: [252], p. 438.
290. Nagano K. et al., Optimizing Optical Transmitter and Receiver for Transmitting Multi-Channel Broadcasting TV Signals Using Laser Diodes. — in: [271], p. 13.1.
291. Shimada S. et al., 100 KM Optical Fibre Transmission System at 1.32 M Using Highpower C. W. Nd-YAG Laser, Electr. Letters, **15** (16), 484 (1979).
292. Machida S. et al., 1.5 M Optical Transmission Experiments Using Very Low-Loss Single Mode Fibres, Electr. Letters, **15**, No. 8 (1979).
293. Nakagawa K. et al., 800 Mb/s Fibre Transmission Test Using Low-Loss and Low-Dispersion Single-Mode Cable, Electr. Letters, **15**, 747 (1979).
294. McDewitt R., Systems Requirements Dictate Fiber-Optic Component Parameters, Electronics (14 Oct. 1976), p. 95. [Имеется перевод: — Электроника, 1976, т. 49, № 21, с. 37.]
295. Gloge D., Inside the New Region: Second Generation Fiber Systems, Optical Spectra (Nov. 1979), p. 50.
296. Nakagawa K., Ito T., Detailed Evaluation of an Attainable Repeater Spacing for Fiber Transmission at 1.3 M and 1.55 M Wavelengths, Electr. Letters, **15** (24), 776 (1979).
297. Lucy J. C., Fiberguide Projections Performance and Price. — in: [272], p. 47.
298. Colavito C. et al., Reliability and Cost Evaluation in View of the Introduction of Optical Fibers in Communications Networks. — in: [272], p. 389.
299. Bigi F., Colavito C., Optical Fibers in Local Digital Networks. — in: [250], p. 224.

300. *Hudson M., Thiel F.*, The Star Coupler: a Unique Interconnection Component for Multimode Optical Waveguide Communications, *Appl. Optics*, **13** (11), 2540 (1974).
301. *Tsujimoto Y. et al.*, Fabrication of Low-Loss 3 dB Couplers with Multimode Optical Fibers, *Electr. Letters*, **14** (5), 157 (1978).
302. *Kawasaki B. S., Hill K. O.*, Low-Loss Access Coupler for Multimode Optical Fiber Distribution Networks, *Appl. Opt.*, **16**, 1794 (1977).
303. *Opielka D., Rittich D.*, Low-Loss Optical Y-Branch, *Electr. Letters*, **15** (23), 757 (1979).
304. *Rawson E. G. et al.*, Fibernet: a Fiber Optic Computer Network Experiment. — in: [252], p. 636.
305. *Giertz H. W. et al.*, Experimental Fiber Optical Databus. — in: [252], p. 641.
306. *Jeunhomme L., Pocholle J. P.*, T-Coupler for Multimode Optical Fibers, Conference AGARD, No. 219, Optical Fibers, Integrated Optics and Their Military Applications, London (May 1977), p. 48.
307. *Farrington J. G., Chown M.*, An Optical Fiber Multiterminal Data System. — in: [306], p. 50.
308. *Uchida T., Sugimoto S.*, Micro-Optic Devices for Optical Communications. — in: [252], p. 374.
309. *Sheem S. K., Giallorenzi T. G.*, Single-Mode Fiber Multiterminal Star Directional Coupler, *Appl. Phys. Lett.*, **35** (2), 131 (1979).
310. *Weidel E., Gruchmann D.*, Tee Coupler for Single-Mode Fibers, *Electr. Letters*, **15** (23), 737 (1979).
311. *Campbell L. L.*, Multiterminal Fiber Systems, *Laser Focus* (June 1978), p. 42.
312. *McMahon D. H., Gravel R. L.*, Star Repeaters for Fiber Optic Links, *Appl. Optics*, **16** (2), 501 (1977).
313. *Meyer W.*, Star Couplers for Multimode Optical Waveguide Communication Systems. — in: [250], p. 166.
314. *Barnoski M. K.*, Data Distribution Using Fiber Optics, *Appl. Optics*, **14** (11), 2571 (1975).
315. *Ito T. et al.*, Bidirectional Tapered Fiber Star Couplers. — in: [252], p. 318.
316. *Rawson E. G.*, Optical Fibers for Local Computer Networks, Topical Meeting on Optical Fiber Communication (Washington, March 6—8 1979), Paper WE1.
317. *Villarruel C. A. et al.*, Couplers for Multiterminal Data Distribution Systems. — in: [316], Paper WG1.
318. *Ohshima S., Ozeki T.*, Rotation-Splice Tapered Fiber Star Coupler. — in: [271], Post dead-line paper 19.5.
319. *Carballes J. C. et al.*, EROS — an Opto-Electronic-Device for Bidirectional Links. — in: [252], p. 341.
320. *D'Auria L. et al.*, Libido Using Eros: Half-Duplex Optical Link, *Electr. Letters*, **15** (25), 820 (1979).
321. *McDonald R. I.*, Bidirectional Analogue Optical Transmission Using Semiconductor Junction Transceivers, *Electr. Letters*, **15** (4), 121 (1979).
322. *Wilner K., Van Den Heuvel A.*, Fiber-Optic Delay Lines for Microwave Signal Processing, *Proc. IEEE*, **64** (5), 805 (1976). [Имеется перевод: — ТИИЭР, 1976, т. 64, № 5.]
323. *Chang C. T. et al.*, Fibre-Optic Delay-Line Devices for R. F. Signal Processing, *Electr. Letters*, **13** (22), 678 (1977).
324. *Tomlinson W. J.*, Wavelength Multiplexing in Optical Fibers, *Appl. Optics*, **16** (8), 2180 (1977).
325. *Kobayashi K., Seki M.*, Micro-Optic Grating Multiplexers for Fiber Optic Communications. — in: [316], Paper WD1.
326. *Miki T., Ishio H.*, Viabilities of the Wavelength-Division-Multiplexing Transmission System over an Optical Fiber Cable, *IEEE Trans. on Communications*, **COM-26** (7), 1082 (1978).

327. Hara E. H. *et al.*, The Use of an Optical Power Combiner for Multiplexing Multiple Television Sources in Single-Fiber Optical Systems, IEEE Trans. on Cable Television, **CATV-4** (2), 49 (1979).
328. Hill K. O. *et al.*, Efficient Power Combiner for Multiplexing Multiple Sources to Single-Fiber Optical Systems. Appl. Physics Lett., **31** (11), 740 (1977).
329. Aoyama K. I., Minowa J. I., Optical Demultiplexer for a Wavelength Division Multiplexing System, Appl. Optics, **18** (8), 1253 (1979).
330. Aoyama K. I., Minowa J. I., Low-Loss Optical Demultiplexer for WDM Systems in the 0.8 μ m Wavelength Region, Appl. Optics, **18** (16), 2834 (1979).
331. Nosu K., Multireflection Optical Multi/Demultiplexer Using Interference Filters, Electr. Letters, **15** (14), 414 (1979).
332. Ishio H. *et al.*, A Two-Way Wavelength-Division-Multiplexing Transmission and its Application to a Switched TV-Distribution System. — in: [252], p. 646.
333. Levy U., Friesem A., Parallel Transmission of a One Dimensional Light Distribution by a Single Optical Fiber, Appl. Phys. Lett., **32**, No. 1 (1978).
334. Epstein M., Fiber Optics in Medicine, Proc. SPIE, vol. 77, Virginia, USA, 22 March 1976.
335. Ulrich R., Image Formation by Phase Coincidences in Optical Waveguides, Optics Comm., **13** (3), 259 (1975).
336. Gover A. *et al.*, Direct Transmission of Pictorial Information in Multimode Fibers, JOSA, **66** (4), 306 (1966).
337. Paxton K. B. *et al.*, Aberrations and Design of Graded Index (GRIN) Rods Used as Image Relays, Appl. Optics, **10** (1), 2090 (1971).
338. Friesem A., Levy U., Parallel Image Transmission by a Single Optical Fiber, Optics Letters, **2**, No. 5 (1978).
339. Bartelt H. O., Transmission of Two-Dimensional Images by Wavelength Multiplexing, Optics Comm., **28**, No. 1 (1979).
340. Ferriere R. *et al.*, Fourier Holograms Recorded and Decoded in Polychromatic Light, Optics Comm., **31** (3), 285 (1979).
341. Levy U., Friesem A. A., Direct Picture Transmission in a Single Optical Fiber with Holographic Filters, Optics Comm., **30** (2), 163 (1979).
342. Sprague R. A., Nisenson P., The PROM — A Status Report, Opt. Eng., **13**, 256 (1972).
343. Clainchard A. *et al.*, Filtrage optique en temps réel par inscription point par point sur des thermoplastiques, Congrès AFCET IRIA (Chatenay-Malabry, 21—23 Fév. 1978), p. 855.
344. De Bazelaere E., Lebreton G., Système de traitement optique du signal à tube Phototitius, Application au sonar actif, Colloque GRETSI (Nice, 26—30 Avril 1977), p. 79.
345. Fleuret J., Maitre H., Contour Detection: Numerical, Optical and Hybrid Methods, Symposium ICO-11 (Madrid, Sept. 1978), p. 383.
346. Wiersma P., A Three Colour Channel System to Synthesize Complex Point Spread Functions, Opt. Comm., **28** (3), 280 (1979).
347. Gorlitz D., Lanzl F., Colour Encoded Aperture Masks Used for Incoherent Filtering of Images, Opt. Comm., **28** (3), 283 (1979).
348. Lohmann A. W., Hausler G., Hybrid Image Processing with Feedback, Optics Comm., **21** (3), 365 (1977).
349. Sato T., Sasaki K., Yamamoto R., Image Processing System Using Incoherent Image Feedback, Appl. Optics, **17** (5), 717 (1978).
350. Jablonowski D. P., Lee S. H., A Coherent Optical Feedback System for Optical Information Processing, Appl. Physics, **8**, 51 (1975).
351. Hausler G., Simon M., Generation of Space and Time Picture Oscillations by Active Incoherent Feedback, Optica Acta, **25** (4), 327 (1978).
352. Gotz J., Hausler G., Sesselmann R., A TV Optical Operational Amplifier, Appl. Optics, **8**, 2754 (1979).

353. *Preston K., Jr.*, Coherent Optical Computers, McGraw-Hill, New York, 1972. [Имеется перевод: *Престон К.* Когерентные оптические вычислительные машины. — М.: Мир, 1974.]
354. *Stark H., Lee D.*, An Optical Digital Approach to the Pattern Recognition of Coal Workers Pneumoconiosis, IEEE Trans. on Syst. Man. Cyb., **SMC-6** (11), 788 (1976).
355. *Duvernoy J.*, Classification de thermographies par filtrage statistique des fréquences spatiales, Opt. Commun., **27** (3), 333 (1978).
356. *Casasent D., Psaltis D.*, Scale Invariant Optical Transform, Optical Engineering, **15** (3), 258 (1976).
357. *Schaefer D., Strong J.*, The computers, Proc. of the IEEE, **65** (1), 129 (1977). [Имеется перевод: — ТИИЭР, 1977, т. 65, № 1.]
358. *Huang A. et al.*, Optical Computation Using Residue Arithmetic, Appl. Optics, **18** (2), 149 (1979).
359. *Psaltis D., Casasent D.*, Optical Residue Arithmetic: a Correlation Approach, Appl. Optics, **18** (2), 133 (1979).
360. *Casasent D., Psaltis D.*, Decimal Residue Decimal Optical Converters, Opt. Comm., **27** (3), 317 (1978).
361. *Tréneux M.*, La révolution optique ou l'optique apprivoisée, L'écho des recherches, **100**, 21 (1980).
362. Commutation Optique: 4 technologies à l'avantgarde, Journ. Optics, **10**, 280 (1979).
363. *Smith M. M.*, Holographic Recording Materials, Topics in Applied Physics, vol. 20, Springer Verlag, Berlin, 1977.
364. *Kurtz R., Owen R.*, Holographic Recording Materials, A Review, Opt. Eng., **14** (5), 393 (1975).
365. *Kowaliski P.*, Théorie photographique appliquée, Masson, Paris, 1972.
366. *Shankoff T. A.*, Phase Holograms in Dichromated Gelatins, Appl. Optics, **7**, 2101 (1968).
367. *Donjon J. et al.*, A Pockels-Effect Light Valve: Phototitus, IEEE Trans. on Elec Dev., **ED-20** (11), 1037 (1973).
368. *Bleha W. P. et al.*, Application of the Liquid Crystal Light Valve to Real-Time Optical Data Processing, Opt. Eng., **17** (4), 371 (1978).

ОГЛАВЛЕНИЕ

Предисловие редактора перевода	5
Предисловие к французскому изданию	7
Предисловие	9
ГЛАВА 1. Основы электромагнетизма	15
§ 1. Уравнения Максвелла, соответствующие физические законы	15
§ 2. Материальные уравнения	19
§ 3. Соотношения непрерывности	22
§ 4. Гармонические колебания, комплексные амплитуды	23
§ 5. Энергетические соотношения, интенсивность света	24
ГЛАВА 2. Распространение электромагнитных волн	28
§ 1. Волновые уравнения, плоские волны	29
§ 2. Гармоническая волна, приближение геометрической оптики	33
§ 3. Проникновение поля в частично проводящую среду, неоднородные волны	38
§ 4. Поглощение света	43
§ 5. Дисперсия	48
ГЛАВА 3. Распространение волн в световодах	54
§ 1. Отражение плоской волны от границы раздела двух сред	54
§ 2. Металлический световод	62
§ 3. Диэлектрический световод	70
ГЛАВА 4. Теория скалярной оптики	80
§ 1. Переход от сферической волны к плоской волне на основе дифракции	81
§ 2. Пространственная частота	91
§ 3. Гауссовы пучки	93
§ 4. Несколько слов о когерентности	101
ГЛАВА 5. Голография	113
§ 1. «Голографическая» запись	114
§ 2. «Уравнение голографистов»	116
§ 3. Аналогия с дифракционными решетками	117
§ 4. Экспериментальная запись голограмм	123
§ 5. Различные типы голограмм	128
§ 6. Цифровые голограммы	144
§ 7. Применение голографии	160
ГЛАВА 6. Обработка информации методами когерентной оптики	167
§ 1. Принципы обработки информации в когерентном свете	168
§ 2. Теория: метаксиальная пространственно-временная оптика	173

§ 3. Применение оптической обработки информации в когерентном свете	174
§ 4. Ограничения когерентной пространственной фильтрации	186
ГЛАВА 7. Обработка информации методами некогерентной оптики	189
§ 1. Теория некогерентной пространственной фильтрации	189
§ 2. Области применения некогерентной пространственной фильтрации	194
§ 3. Методы пространственной фильтрации, не использующие свойств дифракции	202
ГЛАВА 8. Распространение света в оптических волокнах	206
§ 1. Некоторые сведения о распространении света в атмосфере	207
§ 2. Оптические волокна	210
§ 3. Исследование распространения света	212
§ 4. Моды распространения света	228
§ 5. Оптимизация оптических волокон	236
§ 6. Импульсный отклик и передаточная функция	242
§ 7. Влияние взаимодействия мод	244
§ 8. Одномодовые волокна	248
ГЛАВА 9. Измерение характеристик оптических волокон	257
§ 1. Общие замечания	258
§ 2. Измерение потерь	258
§ 3. Измерение профиля показателя преломления	261
§ 4. Измерение числовой апертуры	265
§ 5. Измерение дисперсии, обусловленной материалом	265
§ 6. Измерение импульсного отклика и передаточной функции	266
§ 7. Измерение модовой дисперсии	269
ГЛАВА 10. Технология изготовления оптических волокон	271
§ 1. Потери в оптических волокнах	271
§ 2. Методы изготовления	272
§ 3. Вытягивание оптического волокна	277
§ 4. Кварцево-силикатные волокна	278
ГЛАВА 11. Оптические кабели и разъемы	279
§ 1. Характеристики волокон до изготовления кабеля	279
§ 2. Кабели из оптических волокон	280
§ 3. Оптические разъемы	284
ГЛАВА 12. Лазеры	291
§ 1. Принцип работы лазера	292
§ 2. Различные типы лазеров	304
§ 3. Некоторые общие вопросы лазерной техники	311
§ 4. Полупроводниковые лазеры	314
§ 5. Светодиоды	330
ГЛАВА 13. Фотоприемники	333
§ 1. Основные характеристики фотодиодов	334
§ 2. PIN-фотодиод	339
§ 3. Плоскостной фотодиод	342
§ 4. Лавинный фотодиод (ЛФД)	343

§ 5. Фотоприемники из полупроводниковых соединений типа АІІІ — ВV	346
§ 6. Шумы фотодиодов	347
ГЛАВА 14. Модуляция света, введение в интегральную оптику	350
§ 1. Электрооптические модуляторы света	352
§ 2. Акустооптические модуляторы	357
§ 3. Интегральная оптика, возможности и трудности	363
ГЛАВА 15. Цифровые методы передачи информации	368
§ 1. Импульсно-кодовая модуляция (ИКМ)	368
§ 2. Основные параметры цифровых систем передачи информации	372
§ 3. Код передачи	376
§ 4. Приемник	381
ГЛАВА 16. Аналоговые методы передачи информации	395
§ 1. Модуляция по интенсивности	396
§ 2. Частотная модуляция	401
§ 3. Фазово-импульсная модуляция	405
§ 4. Сравнение разных видов модуляции	409
ГЛАВА 17. Основы проектирования волоконно-оптических линий связи	411
§ 1. Технические условия	412
§ 2. Анализ линии связи	412
§ 3. Двухабонентная волоконно-оптическая линия связи	415
§ 4. Распределение данных	420
ГЛАВА 18. Мультиплексирование и передача изображений	433
§ 1. Линии задержки и временное мультиплексирование	434
§ 2. Мультиплексирование по длинам волн	437
§ 3. Временная голография и передача изображений по оптическому волокну	440
§ 4. Пространственное мультиплексирование	441
§ 5. Оптическое волокно и передача изображений	443
ГЛАВА 19. Пути развития систем обработки информации, комбинированные оптико-электронные системы	445
§ 1. Электронные методы обработки изображений	446
§ 2. Комбинированная система пространственной фильтрации в реальном времени	449
§ 3. Комбинированная оптико-телевизионная система	451
§ 4. Комбинированная оптико-телевизионная обработка с обратной связью	454
§ 5. Быстродействующий преобразователь Фурье	457
§ 6. Быстродействующее специализированное вычислительное устройство	458
§ 7. Быстродействующие вычислительные устройства для выполнения логических операций	459
§ 8. К чисто оптическому цифровому вычислительному устройству	460
ПРИЛОЖЕНИЕ 1. Металлический световод, моды в зависимости от длины волны, фазовая скорость и скорость света	465

ПРИЛОЖЕНИЕ 2. Металлический световод, вырожденная волна	468
ПРИЛОЖЕНИЕ 3. Смещение светового пучка при отражении от границы раздела двух сред, эффект Гооса — Хенхена	471
ПРИЛОЖЕНИЕ 4. Диэлектрический световод, расчет напряженностей поля	474
ПРИЛОЖЕНИЕ 5. Основные материалы для оптической обработки информации	475
§ 1. Фотографическая эмульсия	475
§ 2. Бихромированная желатина	478
§ 3. Фотохромные материалы	478
§ 4. Сегнетоэлектрические кристаллы	480
§ 5. Фоторезисты	480
§ 6. Термопластики	481
§ 7. Оптические пространственные модуляторы света на эффекте Поккельса	482
§ 8. Жидкие кристаллы	482
ПРИЛОЖЕНИЕ 6. Двукратное преобразование Фурье и вырожденные волны	483
Литература	485

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, изд-во «Мир».

Ален Козанне, Жак Флере, Анри Мэтр, Мишель Руссо

Оптика и связь
Оптическая передача
и обработка информации

Ст. научный редактор Е. С. Куранский
Мл. научные редакторы Р. Х. Зацепина, Г. Г. Сорокина
Художник В. Г. Сорокин
Художественный редактор С. И. Кравцова
Технический редактор А. Г. Резоухова
Корректор С. А. Денисова

ИБ № 3743

Сдано в набор 13.10.83. Подписано к печати 27.03.84.
Формат 63×90¹/₁₆. Бумага типографская № 1. Гарнитура
литературная. Печать высокая Объем 15,75 бум. л Усл.
печ. л. 31,5. Усл. кр.-отт. 31,5. Уч.-изд. л. 29,01. Изд. № 2/2791.
Тираж 6700 экз. Зак. 669. Цена 3 руб.

ИЗДАТЕЛЬСТВО «МИР»
129820, Москва, И-110, ГСП, 1-й Рижский пер., 2.

Отпечатано с матриц Ленинградской типографии № 2 головного предприятия ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгений Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29 в Ленинградской типографии № 4 ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгений Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 191126, Ленинград, Социалистическая ул., 14.

Издательство «Мир» выпустит в 1985 г. книгу:

РОБЕРТСОН Б. Современная физика в прикладных науках: Пер. с англ. — 21 л. — 1 р. 80 к.

В книге достаточно полно и наглядно излагаются физические основы явлений, используемых при создании и функционировании многих современных технических устройств, в том числе в электронике (прежде всего твердотельной), лазерной оптике, ядерной энергетике и т. д. В книге 250 иллюстраций, каждая глава снабжена задачами, которые предназначены не для оперирования со знакомыми формулами, а для творческого углубления знаний. Материал представлен в 10 главах: «Специальная теория относительности»; «Электрон»; «Начала современной физики»; «Атомы в квантовой физике»; «Молекулы в квантовой физике»; «Прикладная атомная и молекулярная физика»; «Твердое состояние вещества»; «Прикладная физика твердого тела»; «Ядро»; «Практическое использование ядерной техники». По своему замыслу и значению книга близка к хорошо известной советскому читателю книге Р. Спроула «Современная физика» (М.: Наука, 1974).

Для инженерно-технических работников, а также преподавателей общей физики и студентов вузов (в качестве учебного пособия или развернутого справочника).

Уважаемый товарищ!

Если Вас интересует эта книга, то не забудьте оформить на нее предварительный заказ. Наилучшее время оформления заказов — май—июль 1984 г. Именно в этот период поступит большинство заказов, которые по существу и определяют окончательный тираж книги.

КНИГИ ПО КВАНТОВОЙ ЭЛЕКТРОНИКЕ И ОПТИКЕ

(выпуск 1979—84 гг.)

1. *Шуберт М., Вильгельми Б.* Введение в нелинейную оптику: Пер. с нем — 1979 — 25 л., ил. — 2 р. 70 к.
2. *Адаптивная оптика:* Сб. статей. Пер. с англ./Под ред Э. А. Витриченко. — 1980. — 23 л., ил — 3 р. 40 к.
3. *Кейси Х., Паниш М.* Лазеры на гетероструктурах: В 2-х частях. Пер. с англ. — 1981. — Часть 1: 17 л., ил., 2 р. 80 к. — Часть 2: 22 л., ил., 3 р. 60 к.
4. *Оптическая голография:* В 2-х томах. Пер. с англ./Под ред Х. Колфила — 1982. — 47 л., ил — 3 р. 90 к за комплект
5. *Адамс М Дж.* Введение в теорию оптических волноводов: Пер. с англ. — 1984. — 34 л., ил. — 3 р. 50 к.

