

ИЗДАТЕЛЬСТВО
«МИР»

INTRODUCTION TO FOURIER OPTICS

by

Joseph W. GOODMAN

Department of Electrical Engineering
Stanford University

MCGRAW-HILL BOOK COMPANY
SAN FRANCISCO • NEW YORK • ST. LOUIS
• TORONTO • LONDON • SYDNEY
1968

Дж. Гудмен

ВВЕДЕНИЕ В ФУРЬЕ-ОПТИКУ

Перевод с английского
В. Ю. ГАЛИЦКОГО и М. П. ГОЛОВЕЯ

Под редакцией

Г. И. КОСУРОВА

А. И. Косуров
1971.

ИЗДАТЕЛЬСТВО «МИР»

МОСКВА • 1970

Книга содержит систематическое и вместе с тем достаточно элементарное изложение теории оптического изображения с точки зрения интегральных преобразований. В ней рассматриваются вопросы оптической обработки информации, фильтрация, основы метода голографии и ее возможные применения.

Книга предназначена для физиков и инженеров, занимающихся теорией, а также разработкой и применением оптических приборов. Она также может быть использована как дополнительное пособие студентами старших курсов, специализирующимися по оптике и радиофизике. Разделы, посвященные возможным применениям новых методов в оптике, представляют большой интерес и для читателей-неспециалистов.

Редакция литературы по физике

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Для современной теории оптического изображения и применения оптики для обработки информации характерно широкое использование операционных методов решения задач на основе преобразования Фурье. Эти методы, разработанные главным образом в радиофизике и теории связи, не только вооружили теорию оптического изображения гибким и мощным математическим аппаратом, но и открыли широкие возможности для плодотворных аналогий между оптикой и теорией связи. Общность математического аппарата и физических идей потребовала, естественно, и унификации терминологии, которая произошла на базе теории связи. При этом в теорию изображения вошли такие непривычные для слуха оптика термины, как импульсное воздействие (вместо точечного источника света), импульсный отклик (вместо изображения точечного источника), передаточная функция, частотная характеристика и т. д.

Отдавая должное существенному влиянию теории связи на оптику, необходимо все же четко представлять, что основы современной теории изображения были заложены работами Релея, Майкельсона, Аббе и других выдающихся оптиков прошлого на основе идей, возникших в самой оптике. Так, самые элементарные соображения теории дифракции Фраунгофера показывают, что распределение амплитуд света в фокальной плоскости определяется преобразованием Фурье распределения амплитуд в плоскости отверстия объектива. Вычисление разрешающей силы оптического прибора применительно к двум близким точечным источникам тоже сводится к преобразованию Фурье и применению принципа суперпозиции. Наконец, теория изображения в микроскопе, созданная Аббе еще в прошлом веке, звучит совсем по современному. Возможности влиять на изображение

путем манипуляций в частотной плоскости, блестящей демонстрацией чего могут служить метод темного поля Теплера и метод фазового контраста Цернике, были открыты в оптике без привлечения идей из радиофизики и даже раньше аналогичных идей относительно преобразования фазовой модуляции в амплитудную. Таким образом, классические работы по оптике уже содержали основные элементы современной теории изображения.

До тех пор, пока качество изображения оценивалось только глазом наблюдателя или фотографической пластинкой, в дальнейшей разработке теории и в уточнении критериев не было необходимости. Развитие телевидения, радиолокации и радиоастрономии, включивших оптическое изображение в общую систему каналов связи, потребовало единого подхода к оценке качества изображения с общих позиций теории информации и вызвало приток новых идей и методов в теорию формирования изображения. Решающую роль сыграло также появление когерентных источников света, открывших широкие возможности для оптической обработки информации и вызвавших к жизни новую область оптики — голографию.

Аналогия между теорией связи и теорией изображения — не чисто механическая и не всегда прямая. Дело не только в том, что при переходе от теории связи к теории изображения функции времени заменяются функциями координат, но и в том, что изображение всегда по меньшей мере двумерно и преобразование изображения соответствует многоканальной системе. В этом смысле возможности оптики при обработке информации шире возможностей обычных каналов связи. Кроме того, некоторые операции, естественные для оптики, в системах связи оказываются искусственными и второстепенными. Так, одной из простейших и основных операций в оптике — преобразованию спектра пространственных частот изменением масштаба изображения — в системах связи соответствует двухступенчатый процесс записи сигнала и воспроизведения с другой скоростью. Обычному в системах связи смещению спектра частот путем гетеродинирования в оптике соответствует двухступенчатый метод голографии.

Большая часть материала по современной теории оптического изображения, по когерентной оптике и голографии разбросана в периодической литературе и в специальных монографиях, которые не могут быть рекомендованы для первого чтения. Вместе с тем как радиоинженеры, так и оптики ощущают необходимость в пособии, в котором содержалось бы изложение оптики с точки зрения интегральных преобразований и которое могло бы служить если не учебником, то путеводителем по специальной литературе.

Этим задачам в значительной степени удовлетворяет книга Гудмена, русский перевод которой предлагается читателям. Книга возникла на основе курса лекций, которые автор читал аспирантам Стенфордского университета, что, безусловно, нашло отражение в характере изложения и подборе иллюстративного материала.

Книга содержит восемь глав. В первой, вводной главе, излагаются общие соображения относительно линейности и аналогии между системами связи и оптическими системами. Кроме того, дается краткий обзор содержания книги. Гл. 2 содержит математические основы теории преобразования Фурье в области двух независимых переменных. В ней приводятся основные теоремы, а также ряд прямых и обратных преобразований наиболее употребительных функций. В конце главы формулируется теорема выборки или теорема о дискретном представлении функции с ограниченным спектром. Гл. 3 содержит вывод принципа Гюйгенса — Френеля из волнового уравнения в разных формах, зависящих от выбора функций Грина. Эта глава написана несколько более академично, чем другие, однако последние параграфы, посвященные разложению светового поля в спектр плоских волн, представляют интерес и с практической стороны.

Рассмотрение вопросов физической оптики начинается с гл. 4, где излагается теория дифракции в обычных приближениях Фраунгофера и Френеля. Дифракция Фраунгофера иллюстрируется примерами прямоугольного и круглого отверстий, а также амплитудной и фазовой синусоидальных решеток. Дифракция Френеля, т. е. дифракционная теория теней, рассматривается на

примере прямоугольного отверстия, однако до конца вычисления проводятся для не очень удачного случая геометрической тени, для которого принятые приближения, строго говоря, не выполняются. В гл. 5 математический аппарат, развитый в предыдущих главах, применяется к теории изображения, создаваемого линзами. Гл. 6 посвящена рассмотрению теории изображения с точки зрения пространственно-частотных характеристик. Рассматриваются также вопросы когерентности в применении к теории изображения. Вводятся понятия когерентной передаточной функции и оптической передаточной функции. Сравняются свойства когерентной и некогерентной оптических систем (в частности, их разрешающая сила). В последнем параграфе в общих чертах рассматривается интересный вопрос о разрешении выше дифракционного предела, однако следует иметь в виду, что возможности этого метода сильно ограничены шумами.

Последние две главы, 7 и 8, занимающие половину объема книги, посвящены различным применениям, главным образом когерентной оптики. В гл. 7 рассматриваются фильтрация и оптическая обработка информации. Много внимания уделено фильтрации в плоскости пространственных частот и комплексным сопряженным фильтрам. В конце главы очень кратко рассматривается метод оптической обработки радиолокационных данных. Гл. 8 является довольно полным, хотя и элементарным, обзором голографии, начиная с основных принципов и кончая самыми различными ее применениями.

Таким образом, в книге с общих позиций интегральных преобразований рассмотрен широкий круг вопросов, связанных с теорией оптического изображения. К сожалению, в нее не вошли такие важные вопросы современной прикладной оптики, как фурье-спектрометрия, в основе которой тоже лежат интегральные преобразования. Вне поля зрения остались также вопросы статистической оптики; в частности, не проводится глубокий анализ когерентности, что было бы очень полезно хотя бы в применении к голографии.

В книге сравнительно мало конкретных примеров. Однако это в значительной мере компенсируется хорошо

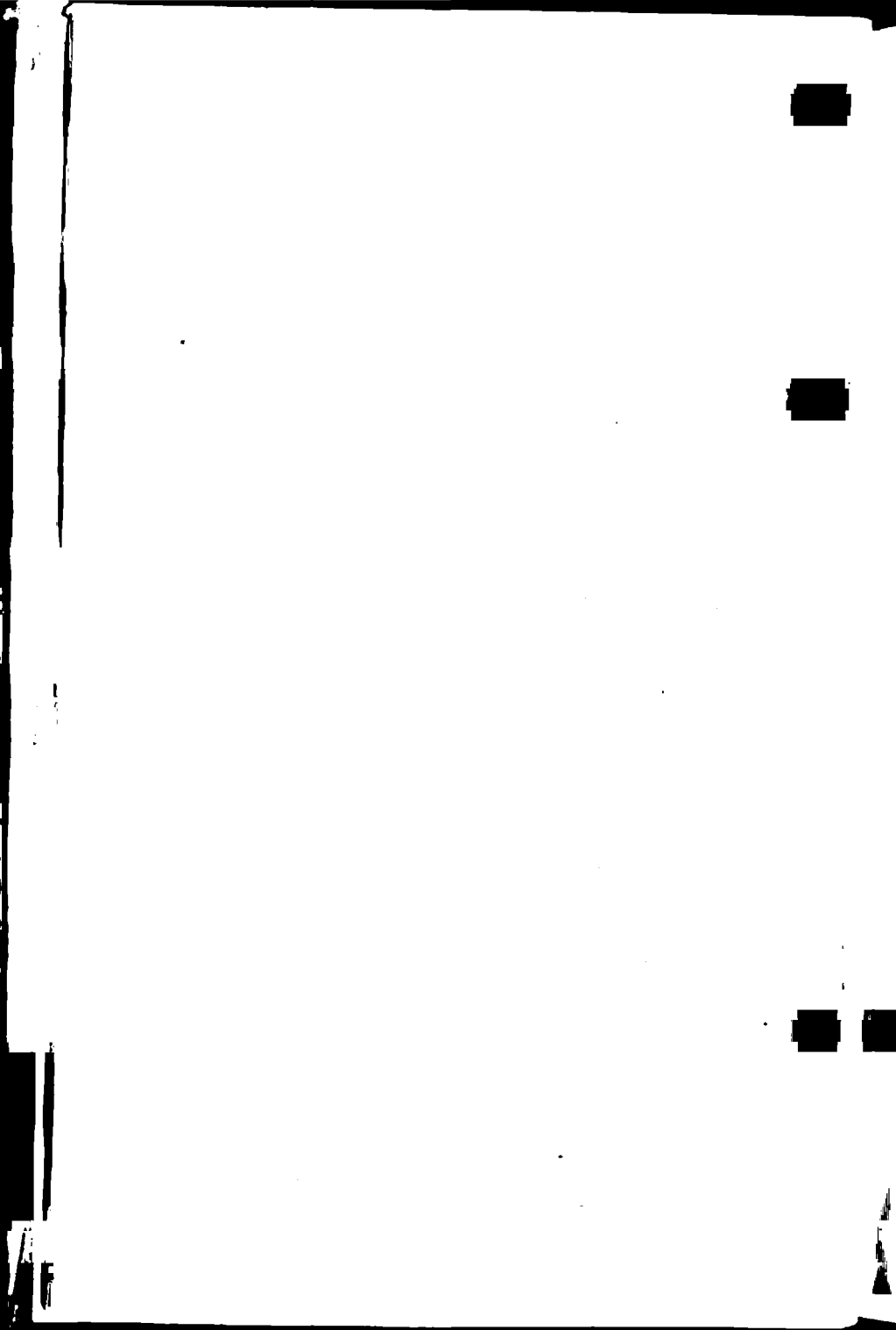
подобранными к каждой главе задачами, содержание которых часто выходит за рамки непосредственно книги и предполагает использование дополнительной литературы, как учебной, так и оригинальной, рекомендованной в конце каждой главы. Серьезный читатель не должен пренебрегать ни дополнительной литературой, ни решением задач. При рассмотрении теоретических вопросов автор не стремится к наглядности, а целиком основывается на математическом анализе. Имея в виду, что для многих читателей наглядность представлений облегчит освоение математического аппарата, при переводе мы дополнили список литературы не только книгами, близкими по характеру к литературе, рекомендованной автором, но и такими работами, в которых основное внимание уделяется наглядности физических представлений.

При подготовке русского издания возник ряд терминологических вопросов. Как правило, мы старались выбирать используемые в теории связи термины, наиболее близкие к буквальному переводу. Перевод выполнили В. Ю. Галицкий (гл. 1—6) и М. П. Головей (гл. 7 и 8 и приложения).

Мы надеемся, что книга Гудмена окажется полезным дополнительным пособием для студентов радиотехнических и оптических специальностей, а также для радиоинженеров и физиков, работающих в области оптической обработки информации, когерентной оптики и голографии.

Г. И. Косоуров

От редакции. В связи с безвременной кончиной Г. И. Косоурова работа над книгой была завершена Б. Н. Гречушниковым.



ПРЕДИСЛОВИЕ

При написании этой книги я ставил себе цель создать учебник оптики, предназначенный для инженеров-электриков. Как, вероятно, известно читателю, сейчас имеется несколько прекрасных книг, посвященных применению анализа Фурье и понятий теории линейных систем в оптике. Однако по собственному опыту я знаю, что эти книги годятся скорее для физиков, которые уже хорошо знакомы с принципами классической оптики, но, вероятно, менее знакомы с математическими методами, так плодотворно применяемыми в современной теории формирования изображения. Инженеры-электрики, напротив, очень хорошо знакомы с математическими методами (поскольку такие методы широко используются в анализе цепей), но относительно слабо разбираются в принципах классической оптики. Поэтому мне представлялось целесообразным изложить принципы оптики для инженеров-электриков таким образом, чтобы максимально использовать уже известные им математические методы. Анализ Фурье и теория линейных систем образуют фундамент, на котором построены теория формирования изображения, оптической обработки информации и голографии.

В основу книги лег конспект тридцати лекций по фурье-оптике, прочитанных на факультете электротехники Стенфордского университета в течение четверти семестра. Слушателями лекций были преимущественно аспиранты первого или второго года обучения. По мере того, как объем материала при пересмотре сделанных ранее записей увеличивался, становилось все труднее полностью изложить его в 30-ти лекциях. Поэтому в последующих вариантах курса я нашел необходимым опустить (или, вернее, изложить только кратко) материал, содержащийся в гл. 2, § 3; гл. 3, § 5 и 6; гл. 4, § 3; гл. 6,

§ 6; гл. 7, § 7 и частично материал гл. 8, § 8 и 9. Если курс лекций читается в течение полного семестра, то, вероятно, можно включить весь материал.

Я благодарен многим людям за их помощь и поддержку. Вероятно, больше всего я обязан сотрудникам Радиолокационной и Оптической лабораторий Мичиганского университета, которые стимулировали мой интерес к оптике и оказали существенное влияние на мою точку зрения. В частности, я обязан А. Зейгману и Н. Абрамсону, которые оказали мне поддержку в начальный период работы, когда это было особенно необходимо, Э. О'Нейлу, Е. Лейту, Л. Катрона и Х. Лотшу, которые тщательно просмотрели рукопись и сделали ряд ценных замечаний, М. Лейману, который затратил много времени на подготовку фотографий, совершенно необходимых в этой книге, Д. Джексону и В. Хантли за ряд полезных обсуждений, миссис Алисе Лескеллит, которая переписала на машинке большую часть рукописи, и многим студентам, которые способствовали повышению квалификации своего учителя. Наконец, я хочу поблагодарить мою жену Хон Мей, без участия и поддержки которой эта книга не смогла бы стать реальностью.

Джозеф У. Гудмен

ГЛАВА I

ВВЕДЕНИЕ

§ 1. ОПТИКА И ТЕОРИЯ СВЯЗИ

Начиная с конца 30-х годов нашего столетия постепенно расширялись и крепили связи между такой почетной отраслью физики, как оптика, и электротехническими дисциплинами: теорией информации и теорией связи. Эта тенденция понятна, ибо системы связи, так же как и системы, создающие изображение, предназначены для накопления или передачи информации. В первом случае информация обычно носит временную природу (т. е. форма волны тока или напряжения модулирована), в то время как во втором случае она имеет пространственную природу (т. е. амплитуде или интенсивности света свойственно пространственное распределение), но это различие, если подходить к вопросу с абстрактной точки зрения, довольно несущественно.

Вероятно, наиболее глубокая связь между двумя названными дисциплинами заключена в сходстве используемого в них математического аппарата, а именно анализа Фурье и теории «систем». Основной причиной такого сходства служит не только общий интерес к «информации», но и существование вполне определенных фундаментальных свойств, присущих как системам связи, так и системам, создающим изображение. Например, многим электронным схемам, так же как и приборам, создающим изображение, присущи такие свойства, как *линейность* и *инвариантность* (см. гл. 2). Любую схему или прибор (электронный, оптический или иного типа), обладающий этими двумя свойствами, можно довольно просто описать математически, используя методы *частотного анализа*. Таким образом, если усилитель низкой частоты удобно описывать, пользуясь его временной частотной характеристикой, то систему, создающую изображение, столь же удобно описывать,

пользуясь ее пространственной частотной характеристикой.

Сходство между системами не исчезает, даже если они не обладают свойствами линейности и инвариантности. У некоторых нелинейных оптических элементов (в частности, фотопленки) соотношения между входным и выходным сигналами совершенно аналогичны соответствующим характеристикам нелинейных элементов электронных схем (диодов, электронных ламп и т. п.), так что в обоих случаях могут быть использованы близкие математические методы.

Особенно важно знать, что сходство математических аппаратов можно использовать не только с целью анализа, но и с целью синтеза. Подобно тому как мы сознательно в соответствии с заданной программой управляем спектром временной функции, можно любым наперед заданным образом изменять спектр пространственной функции. На современном этапе развития оптики ряд выдающихся результатов получен благодаря применению методов синтеза Фурье. В качестве примера можно назвать создание микроскопа фазового контраста Цернике, оптических согласованных фильтров и разработку различных методов усиления контраста. В будущем, несомненно, выявится еще больше преимуществ такого тесного контакта между двумя названными дисциплинами, который обогатит как оптику, так и теорию связи.

§ 2. ОБ ЭТОЙ КНИГЕ

Предполагается, что читатели этой книги уже имеют довольно прочные знания в области анализа Фурье и в теории линейных систем. В гл. 2 дается обзор необходимых основных положений. Чтобы не наскучить читателям, хорошо знакомым с анализом временных сигналов и с системами связи, обзор проводится для функций двух независимых переменных. Такие функции имеют, разумеется, первостепенное значение в оптике. Переход от одной независимой переменной к двум открывает новые возможности применения математической теории, позволяя изучать явления, не имеющие прямых аналогов в теории электронных сигналов и систем.

Физическое явление, называемое *дифракцией*, играет предельно важную роль в теории оптических систем, создающих изображение. В гл. 3 излагаются основы скалярной теории дифракции, включая приближения Кирхгофа, Релея — Зоммерфельда и метод «углового спектра». В гл. 4 общие результаты дополняются приближениями Френеля и Фраунгофера, а также примерами расчетов дифракционной картины.

В оптических системах, образующих изображение, и в оптических системах обработки информации линзы, несомненно, представляют собой главные элементы. В гл. 5 обсуждаются линзы и их свойства. Тонкие линзы рассматриваются как элементы, выполняющие фазовое преобразование; это позволяет вывести обычную формулу линзы и некоторые свойства положительных линз как элементов, выполняющих преобразование Фурье.

Гл. 6 посвящена применению методов частотного анализа к системам, создающим изображение как при когерентном, так и при некогерентном освещении. Определяются соответствующие передаточные функции, и обсуждаются их свойства для систем с аберрациями и без них. С различных точек зрения сравниваются системы, использующие когерентное и некогерентное освещение. В конце главы рассматриваются возможности получения разрешения, превышающего «классический» предел, определяемый дифракцией.

В гл. 7, темой которой является пространственное фильтрование и оптическая обработка информации, особое внимание уделяется проблемам синтеза. Сначала обсуждается историческое развитие идей синтеза Фурье, а затем кратко излагаются основные свойства фотопленки. Исследуются системы оптической обработки информации, основанные на использовании законов геометрической оптики, после чего обсуждаются более мощные методы пространственно-частотной обработки. Значительная часть главы посвящена так называемому «фильтру Вандер Лугта», который позволяет проводить синтез комплексных передаточных функций посредством масок в пространственно-частотной плоскости. Оптическое распознавание образов и оптическая обработка

данных, полученных методом синтетических апертур, — вот основные вопросы, рассматриваемые в этой главе.

Последняя глава посвящается получению изображения методом восстановления волнового фронта, или голографии. Сначала восстановление волнового фронта обсуждается с весьма общей точки зрения, после чего несколько более детально рассматриваются частные методы, разработанные Габором и Лейтом и Упатником. Обсуждается также развитие трехмерной безлинзовой фотографии. Особое внимание уделяется свойствам фотопленки и их влиянию на изображение, полученное путем восстановления волнового фронта; при этом затрагиваются эффекты, связанные с ограниченностью пространственно-частотного отклика пленки, нелинейностью и толщиной эмульсии. Рассматривается также обобщение разработанных методов на случай голографии движущихся предметов, голографии при некогерентном освещении и голографии в отраженном свете. В конце главы в общих чертах описаны возможные применения голографии, начиная с микроскопии и кончая получением изображения через искажающую среду.

АНАЛИЗ ДВУМЕРНЫХ ЛИНЕЙНЫХ СИСТЕМ

Из опыта известно, что многим физическим явлениям присуще общее фундаментальное свойство: реакция на одновременное воздействие нескольких источников в точности равна сумме реакций, вызываемых каждым источником, действующим отдельно. Такие явления называются *линейными*, а это их общее свойство — *линейностью*. Электрические цепи, составленные из сопротивлений, емкостей и индуктивностей, обычно линейны по отношению к входным сигналам самой разнообразной величины и формы. Как мы скоро увидим, линейность волнового уравнения, описывающего распространение света в большинстве сред, позволяет рассматривать образование оптического изображения как линейное преобразование распределения света в «предмете» в распределение света в «изображении».

Свойство линейности, позволяющее значительно упростить математическое описание многих явлений, лежит в основе математического аппарата теории, которую мы в дальнейшем будем называть *теорией линейных систем*. Большое преимущество линейных систем заключается в том, что их реакцию (напряжение, ток, амплитуду света или его интенсивность) на какое-либо сложное воздействие можно выражать через отклики на некоторые «элементарные» воздействия. Таким образом, если воздействие можно представить в виде линейной комбинации элементарных воздействий, каждое из которых вызывает отклик известного простого вида, то благодаря линейности полная ответная реакция может быть найдена в виде соответствующей линейной комбинации таких откликов.

В этой главе мы кратко рассмотрим ряд математических методов, полезных при описании линейных явлений, и остановимся на некоторых математических разложениях, часто используемых при анализе. В после-

дующих главах мы познакомимся с воздействиями (входными сигналами системы) и откликами (выходными сигналами системы), которые будут представлять собой разные физические величины. Если освещение, используемое в оптической системе, обладает свойством, называемым *пространственной когерентностью*, то свет целесообразно описывать в виде пространственного распределения *поля комплексных амплитуд*. Если же освещение не обладает пространственной когерентностью, то свет следует описывать в виде пространственного распределения *действительных значений интенсивности*. Наше внимание будет сосредоточено на анализе линейных систем, входные сигналы которых выражены в комплексной форме; результаты для входных сигналов, представленных в действительной форме, рассматриваются как частные случаи теории.

§ 1. ДВУМЕРНЫЙ АНАЛИЗ ФУРЬЕ

Анализ Фурье — чрезвычайно мощный метод математического анализа как линейных, так и нелинейных явлений. Этот метод широко используется при изучении электрических цепей и систем связи; предполагается, что читатель сталкивался с такими приложениями теории Фурье и поэтому знаком с анализом функций одной независимой переменной (например, времени). Обзор основных математических понятий можно найти в книгах Папулиша [1] и Брэйсуэлла [2]. Мы же займемся изложением анализа функций *двух* независимых переменных и, не стремясь к излишней математической строгости, воспользуемся приближением, которое чаще всего применяется при решении инженерных задач.

Определение и условия существования

Преобразование Фурье, или *фурье-образ* (иначе *спектр Фурье*, или *частотный спектр*) комплексной функции¹⁾ g двух независимых переменных x и y , которое

¹⁾ Комплексные функции будут выделены жирным или готическим шрифтом.

мы будем обозначать символом $\mathcal{F}\{g\}$, определяется выражением¹⁾

$$\mathcal{F}\{g\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \exp[-j2\pi(f_x x + f_y y)] dx dy. \quad (2.1)$$

Преобразование такого вида представляет собой функцию двух независимых переменных f_x и f_y , которые мы обычно называем *частотами*. Точно так же *обратное преобразование Фурье* функции $G(f_x, f_y)$, которое будет обозначаться символом $\mathcal{F}^{-1}\{G\}$, определяется выражением

$$\mathcal{F}^{-1}\{G\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f_x, f_y) \exp[j2\pi(f_x x + f_y y)] df_x df_y. \quad (2.2)$$

Заметим, что математические операции прямого и обратного преобразований отличаются только знаком экспоненты в подинтегральном выражении.

Прежде чем обсуждать свойства прямого и обратного преобразований Фурье, мы должны сначала определить условия, при которых выражения (2.1) и (2.2) имеют смысл. Для некоторых функций эти интегралы могут не существовать в обычном математическом смысле, и поэтому наше изложение было бы не полным без краткого рассмотрения «условий существования». Из всех возможных наборов условий, *достаточных* для существования интеграла (2.1), как правило, используется следующий:

1. Функция g должна быть абсолютно интегрируема по бесконечной плоскости xy .

2. Функция g должна иметь только конечное число разрывов и конечное число максимумов и минимумов в пределах любого прямоугольника конечных размеров.

3. Функция g не должна иметь разрывов второго рода.

Обычно любое из этих условий можно смягчить за счет усиления одного или двух других, но

¹⁾ Когда выше или ниже символа двойного интеграла указывается только один предел интегрирования, то этот предел относится к интегрированию по обоим переменным,

рассмотрение этого вопроса увело бы нас довольно далеко в сторону от наших основных задач.

Как отметил Брэйсуэлл [2], «физическая возможность осуществления — это достаточно обоснованное условие существования преобразования». Однако при анализе систем часто удобно выражать истинную физическую форму волн при помощи идеализированных математических функций, для которых отдельные приведенные выше условия существования могут и не выполняться. Например, сильный короткий импульс принято изображать при помощи так называемой δ -функции Дирака ¹⁾

$$\delta(t) = \lim_{N \rightarrow \infty} N \exp(-N^2 \pi t^2).$$

Точно так же идеализированный точечный источник света часто описывают при помощи двумерной δ -функции:

$$\delta(x, y) = \lim_{N \rightarrow \infty} N^2 \exp[-N^2 \pi (x^2 + y^2)]. \quad (2.3)$$

Такие функции равны нулю везде, кроме начала координат, где они стремятся к бесконечности, имея разрыв второго рода, и поэтому не удовлетворяют третьему условию существования. Нетрудно привести другие важные примеры; например, функции

$$f(x, y) = 1 \text{ и } f(x, y) = \cos(2\pi f_x x)$$

не удовлетворяют первому условию существования.

Очевидно, для того чтобы анализ Фурье можно было применить к большинству интересующих нас функций, следует несколько обобщить соотношение (2.1). К счастью, для функций, которые, строго говоря, не удовлетворяют условиям существования, часто можно найти имеющее смысл преобразование, если эти функции удастся определить как предел последовательности функций, поддающихся преобразованию Фурье. Преобразуя каждый член определяющей последовательности, мы получаем соответствующую последовательность фурье-образов, предел которой называют *обобщенным фурье-образом* исходной функции. С обобщенными

¹⁾ Более детальное обсуждение δ -функции, включающее другие определения, приведено в приложении I.

фурье-образами можно обращаться точно так же, как и с обычными, и различие между ними, как правило, можно не учитывать; в тех случаях, когда говорят о фурье-образе функции, не удовлетворяющей условиям существования, фактически имеют в виду обобщенный фурье-образ. Более подробное рассмотрение такого обобщения анализа Фурье можно найти в книге Лайтхилла [3].

В качестве примера расчета обобщенного фурье-образа рассмотрим δ -функцию Дирака, которая, как мы видели, не удовлетворяет третьему условию существования. Заметим, что каждый член определяющей последовательности (2.3) удовлетворяет условиям существования и имеет следующий фурье-образ (см. табл. 1):

$$\mathcal{F} \{N^2 \exp[-N^2\pi(x^2 + y^2)]\} = \exp\left[-\frac{\pi(f_x^2 + f_y^2)}{N^2}\right].$$

Соответственно обобщенный фурье-образ функции $\delta(x, y)$ имеет вид

$$\mathcal{F} \{\delta(x, y)\} = \lim_{N \rightarrow \infty} \left\{ \exp\left[-\frac{\pi(f_x^2 + f_y^2)}{N^2}\right] \right\} = 1. \quad (2.4)$$

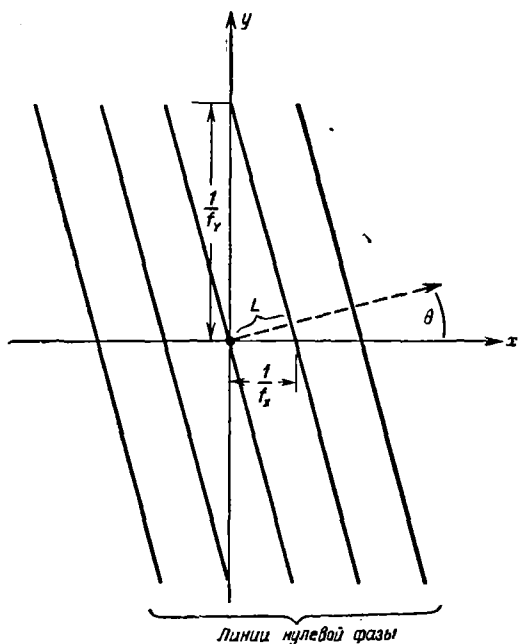
Отсюда видно, что спектр δ -функции одинаков во всем диапазоне изменения частот.

В табл. 1 (см. стр. 32) приведены другие примеры обобщенных фурье-образов.

Преобразование Фурье как разложение

Как уже упоминалось, сложный сигнал на входе линейных систем часто бывает удобно представлять в виде ряда более простых сигналов. Тогда, вычислив отклики системы на каждый из этих «элементарных» сигналов и сложив их, можно найти полную реакцию системы. В основе такого разложения лежит анализ Фурье. Возьмем известное выражение обратного преобразования

$$g(t) = \int_{-\infty}^{\infty} G(f) \exp(j2\pi ft) df,$$



Фиг. 1. Линии нулевой фазы для функции $\exp[j2\pi(f_x x + f_y y)]$.

выражающее функцию времени g через ее частотный спектр. Это выражение можно рассматривать как представление функции $g(t)$ в виде линейной комбинации (т. е. интеграла) элементарных функций характерного вида $\exp(j2\pi f t)$. Очевидно, комплексное число $G(f)$ представляет собой просто весовой коэффициент, на который следует умножить элементарную функцию частоты f при синтезе искомой функции $g(t)$.

Точно так же *двумерное* преобразование Фурье можно рассматривать как представление функции $g(x, y)$ в виде линейной комбинации элементарных функций типа $\exp[j2\pi(f_x x + f_y y)]$. Такие функции обладают рядом интересных свойств. Заметим, что для любой определенной пары частот (f_x, f_y) фаза соответствующей элементарной функции равна нулю вдоль линии, кото-

рая описывается уравнением

$$y = -\frac{f_x}{f_y} x + \frac{n}{f_y} \quad (n - \text{целое число}).$$

Таким образом, можно считать, что элементарная функция в плоскости xy «направлена» (по отношению к оси x) под углом θ (фиг. 1). Величина этого угла равна

$$\theta = \arctg \frac{f_y}{f_x}. \quad (2.5)$$

Пространственный период (т. е. расстояние между линиями нулевой фазы), очевидно, можно записать в виде

$$L = \frac{1}{\sqrt{f_x^2 + f_y^2}}. \quad (2.6)$$

Таким образом, и в этом случае обратное преобразование Фурье можно рассматривать как метод разложения математических функций. Спектр Фурье G функции g представляет собой просто набор весовых множителей, на которые следует умножать каждую из элементарных функций, чтобы получить функцию g . Несомненное достоинство использования этого разложения станет вполне очевидным только после обсуждения инвариантных линейных систем.

Теоремы относительно преобразования Фурье

Из основного определения (2.1) преобразования Фурье вытекает ряд математических следствий, облегчающих нахождение фурье-образа. Рассмотрим несколько основных свойств преобразования, которые будут широко использоваться в дальнейшем. Эти свойства формулируются в виде математических теорем, за которыми следует краткое изложение их физического смысла. Так как эти теоремы представляют собой непосредственное обобщение аналогичных утверждений для одномерного случая, доказательство их приводится в приложении.

1. Теорема линейности. $\mathcal{F}\{\alpha g + \beta h\} = \alpha \mathcal{F}\{g\} + \beta \mathcal{F}\{h\}$, т. е. фурье-образ суммы двух функций равен сумме их фурье-образов.

2. Теорема подобия. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\mathcal{F}\{g(ax, by)\} = \frac{1}{|ab|} G\left(\frac{f_x}{a}, \frac{f_y}{b}\right),$$

т. е. «растяжение» координат в пространственной области (x, y) приводит к сокращению координат в области частот (f_x, f_y) и к изменению общей амплитуды спектра.

3. Теорема смещения. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\mathcal{F}\{g(x-a, y-b)\} = G(f_x, f_y) \exp[-j2\pi(f_x a + f_y b)],$$

т. е. смещение функции в пространственной области вызывает линейный фазовый сдвиг в области частот.

4. Теорема Парсеваля. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)|^2 dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |G(f_x, f_y)|^2 df_x df_y.$$

Эта теорема обычно интерпретируется как закон сохранения энергии.

5. Теорема свертки. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$ и $\mathcal{F}\{h(x, y)\} = H(f_x, f_y)$, то

$$\mathcal{F}\left\{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) h(x-\xi, y-\eta) d\xi d\eta\right\} = G(f_x, f_y) H(f_x, f_y).$$

Свертка двух функций в пространственной области (операция, как будет показано, часто встречающаяся в теории линейных систем) полностью эквивалентна более простой операции умножения их образов.

6. Теорема автокорреляции. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\mathcal{F}\left\{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) g^*(\xi-x, \eta-y) d\xi d\eta\right\} = |G(f_x, f_y)|^2.$$

Аналогично

$$\mathcal{F}\{|g(\xi, \eta)|^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(\xi, \eta) G^*(\xi + f_x, \eta + f_y) d\xi d\eta.$$

Эту теорему можно рассматривать как частный случай теоремы свертки.

7. Интегральная теорема Фурье. В каждой точке непрерывности функции g

$$\mathcal{F}\mathcal{F}^{-1}\{g(x, y)\} = \mathcal{F}^{-1}\mathcal{F}\{g(x, y)\} = g(x, y).$$

В каждой точке разрыва функции g два последовательных преобразования дают среднее значение величины g в окрестности этой точки. Иначе говоря, производя последовательно прямое и обратное преобразования функции, мы снова получаем эту функцию, за исключением точек разрыва.

Рассмотренные выше теоремы преобразования интересны не только с теоретической точки зрения. Эти теоремы будут часто использоваться, так как они дают нам основные правила проведения преобразований Фурье и могут существенно упростить решение задач, связанных с использованием анализа Фурье.

Функции с разделяющимися переменными

Функция двух независимых переменных называется *функцией с разделяющимися переменными* в определенной системе координат, если ее можно записать в виде произведения двух функций, каждая из которых зависит только от одной независимой переменной. Иными словами, g есть функция с разделяющимися переменными в прямоугольных координатах (x, y) , если

$$g(x, y) = g_x(x) g_y(y), \quad (2.7)$$

и в полярных координатах (r, θ) , если

$$g(r, \theta) = g_r(r) g_\theta(\theta). \quad (2.8)$$

Функциями с разделяющимися переменными часто удобнее пользоваться, чем более общими функциями,

так как их свойства позволяют свести сложные двумерные действия к более простым одномерным. Например, функция с разделяющимися переменными в прямоугольной системе координат имеет особенно простое свойство: ее двумерный фурье-образ можно представить в виде произведения одномерных фурье-образов, о чем свидетельствует следующее соотношение:

$$\begin{aligned} \mathcal{F}\{g(x, y)\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \exp[-j2\pi(f_x x + f_y y)] dx dy = \\ &= \int_{-\infty}^{\infty} g_x(x) \exp[-j2\pi f_x x] dx \int_{-\infty}^{\infty} g_y(y) \exp[-j2\pi f_y y] dy = \\ &= \mathcal{F}_x\{g_x\} \mathcal{F}_y\{g_y\}. \quad (2.9) \end{aligned}$$

Таким образом, сам фурье-образ функции g может быть представлен в виде произведения двух сомножителей, один из которых зависит только от f_x , а другой — только от f_y ; процесс двумерного преобразования упрощается и принимает вид последовательности более привычных одномерных действий.

Если мы имеем функцию, у которой разделяются переменные в полярной системе координат, то в этом случае дело обстоит несколько сложнее, чем при разделении переменных в прямоугольной системе координат, однако и в этом случае обычно двумерные операции можно свести к ряду одномерных. Например, преобразование Фурье функции с разделяющимися переменными в полярной системе координат рассматривается в задаче 2.7, где читателю предлагается убедиться, что двумерный спектр может быть найден путем выполнения ряда одномерных операций, называемых *преобразованиями Ханкеля*.

Функции, обладающие осевой симметрией. Преобразование Фурье — Бесселя

Вероятно, наиболее простой класс функций с разделяющимися переменными в полярной системе координат образуют функции, обладающие *осевой симметрией*. Го-

ворят, что функция g обладает осевой симметрией, если ее можно записать как функцию только радиуса r , т. е.

$$g(r, \theta) = g_R(r). \quad (2.10)$$

Такие функции играют особенно важную роль в интересующих нас задачах, поскольку в большинстве оптических систем мы встречаемся именно с этим видом симметрии. Соответственно особое внимание будет уделено фурье-преобразованиям функций, обладающих осевой симметрией.

Фурье-образ функции g в прямоугольной системе координат имеет вид

$$G(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \exp[-j2\pi(f_x x + f_y y)] dx dy. \quad (2.11)$$

Чтобы полностью использовать свойство осевой симметрии функции g , перейдем к полярным координатам как в плоскости xy , так и в плоскости $f_x f_y$, воспользовавшись следующими формулами:

$$\begin{aligned} r &= \sqrt{x^2 + y^2}, & x &= r \cos \theta, \\ \theta &= \arctg\left(\frac{y}{x}\right), & y &= r \sin \theta, \\ \rho &= \sqrt{f_x^2 + f_y^2}, & f_x &= \rho \cos \varphi, \\ \varphi &= \text{arctg}\left(\frac{f_y}{f_x}\right), & f_y &= \rho \sin \varphi. \end{aligned} \quad (2.12)$$

На этот раз запишем фурье-образ как функцию радиуса и угла

$$\mathcal{F}\{g\} = G_0(\rho, \varphi).$$

Переходя в соотношении (2.11) к другим переменным с помощью формул (2.12), фурье-образ функции g можно записать в виде

$$\begin{aligned} G_0(\rho, \varphi) &= \\ &= \int_0^{2\pi} d\theta \int_0^{\infty} dr \cdot r g_R(r) \exp[-j2\pi\rho(\cos\theta \cos\varphi + \sin\theta \sin\varphi)], \end{aligned}$$

или, иначе,

$$G_0(\rho, \varphi) = \int_0^{\infty} dr \cdot r g_R(r) \int_0^{2\pi} d\theta \exp[-j2\pi r \rho \cos(\theta - \varphi)]. \quad (2.13)$$

Наконец, чтобы упростить выражение (2.13), используем следующее тождество:

$$J_0(a) = \frac{1}{2\pi} \int_0^{2\pi} \exp[-ja \cos(\theta - \varphi)] d\theta, \quad (2.14)$$

где J_0 — функция Бесселя первого рода нулевого порядка. В результате подстановки (2.14) в (2.13) фурье-образ G_0 становится функцией радиуса ρ , не зависящей явно от угла φ :

$$G_0(\rho) = 2\pi \int_0^{\infty} r g_R(r) J_0(2\pi r \rho) dr. \quad (2.15)$$

Итак, фурье-образ функций, имеющих осевую симметрию, сам обладает осевой симметрией и может быть найден путем выполнения одномерного действия (2.15). Этот особый вид преобразования Фурье встречается настолько часто, что заслуживает специального названия: выражение (2.15) называется *преобразованием Фурье — Бесселя*, или, иначе, *преобразованием Ханкеля нулевого порядка*. Для краткости мы будем пользоваться первым названием.

На основании аналогичных рассуждений *обратное* фурье-преобразование функции $G_0(\rho)$, обладающей осевой симметрией, можно представить как

$$g_R(r) = 2\pi \int_0^{\infty} \rho G_0(\rho) J_0(2\pi r \rho) d\rho.$$

Следовательно, для функций, обладающих осевой симметрией, различие между действиями прямого и обратного преобразований отсутствует.

Непосредственно из интегральной теоремы Фурье следует, что в каждой точке r , в которой функция $g_R(r)$ непрерывна, можно записать, обозначая действие пре-

образования Фурье — Бесселя через $\mathcal{B}\{\}$:

$$\mathcal{B}\mathcal{B}^{-1}\{g_R(r)\} = \mathcal{B}^{-1}\mathcal{B}\{g_R(r)\} = g_R(r).$$

Кроме того, непосредственное применение теоремы подобия показывает, что (см. задачу 2.4)

$$\mathcal{B}\{g_R(ar)\} = \frac{1}{a^2} G_0\left(\frac{\rho}{a}\right).$$

При использовании выражения (2.15) для нахождения преобразования Фурье — Бесселя читатель должен помнить, что оно представляет собой не более чем частный случай двумерного преобразования Фурье и поэтому любое известное свойство преобразования Фурье имеет полный аналог в преобразованиях Фурье — Бесселя.

Некоторые часто используемые функции и полезные пары преобразований Фурье

Ряд математических функций настолько широко используется в последующем изложении, что для удобства следует особым образом обозначить сами эти функции. Соответственно мы примем следующие обозначения некоторых часто встречающихся функций:

Прямоугольная функция

$$\text{rect}(x) = \begin{cases} 1 & |x| \leq \frac{1}{2}, \\ 0 & \text{в остальных случаях.} \end{cases}$$

sinc функция

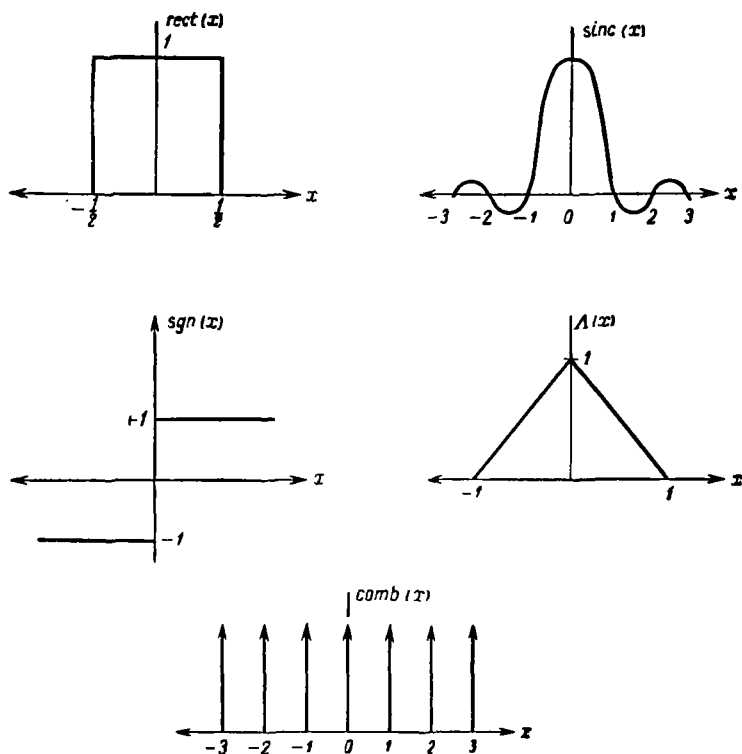
$$\text{sinc}(x) = \frac{\sin \pi x}{\pi x}.$$

Функция знака

$$\text{sgn}(x) = \begin{cases} 1 & x > 0, \\ 0 & x = 0, \\ -1 & x < 0. \end{cases}$$

Треугольная функция

$$\Lambda(x) = \begin{cases} 1 - |x| & |x| \leq 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$



Фиг. 2. Специальные функции.

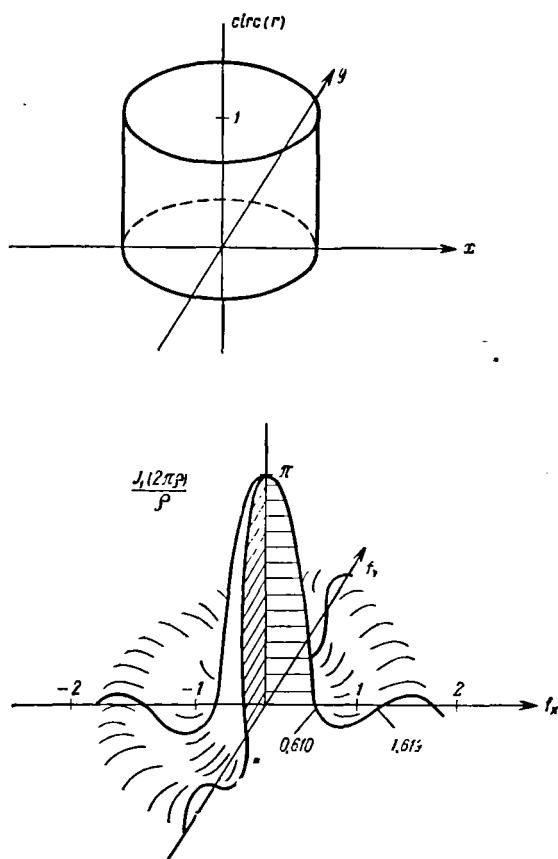
Гребенчатая функция

$$\text{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x-n).$$

Круговая функция

$$\text{circ}(\sqrt{x^2 + y^2}) = \begin{cases} 1 & \sqrt{x^2 + y^2} \leq 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Первые пять из этих функций, изображенные на фиг. 2, зависят только от одной переменной, однако, образуя произведение этих функций, мы можем получить в двумерном пространстве множество функций с разделяющимися переменными. Круговая функция — единствен-



Фиг. 3. Круговая функция и ее образ.

ная из них функция двух независимых переменных; на фиг. 3 показан вид этой функции.

В заключение кратко рассмотрим анализа Фурье мы представим несколько конкретных примеров двумерного преобразования. В табл. 1 приведен ряд функций с разделяющимися переменными в прямоугольной системе координат и их фурье-образы. Все эти образы можно найти непосредственно как произведения известных

Таблица 1

Пары преобразований некоторых функций с разделяющимися переменными в прямоугольной системе координат

Функция	Образ
$\exp[-\pi(x^2 + y^2)]$	$\exp[-\pi(f_X^2 + f_Y^2)]$
$\text{rect}(x) \text{rect}(y)$	$\text{sinc}(f_X) \text{sinc}(f_Y)$
$\Lambda(x) \Lambda(y)$	$\text{sinc}^2(f_X) \text{sinc}^2(f_Y)$
$\delta(x, y)$	1
$\exp[j\pi(x + y)]$	$\delta\left(f_X - \frac{1}{2}, f_Y - \frac{1}{2}\right)$
$\text{sgn}(x) \text{sgn}(y)$	$\frac{1}{j\pi f_X} \cdot \frac{1}{j\pi f_Y}$
$\text{comb}(x) \text{comb}(y)$	$\text{comb}(f_X) \text{comb}(f_Y)$

одномерных образов. Поэтому доказательство этих соотношений предоставляем читателю (см. задачу 2.2).

Наоборот, преобразования большинства функций, обладающих осевой симметрией, нельзя получить просто с помощью известных одномерных преобразований. Из числа функций, обладающих осевой симметрией, чаще всего встречается функция

$$\text{circ}(r) = \begin{cases} 1 & r \leq 1, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Попытаемся теперь найти преобразование этой функции. Используя выражение (2.15) для преобразования Фурье — Бесселя можно записать

$$\mathcal{B}\{\text{circ}(r)\} = 2\pi \int_0^1 r J_0(2\pi r \rho) dr.$$

Производя замену переменных $r' = 2\pi r \rho$ и используя тождество

$$\int_0^x \xi J_0(\xi) d\xi = x J_1(x),$$

это преобразование можно переписать в виде

$$\mathcal{F}\{\text{circ}(r)\} = \frac{1}{2\pi\rho^2} \int_0^{2\pi\rho} r' J_0(r') dr' = \frac{J_1(2\rho)}{\rho}, \quad (2.16)$$

где J_1 — функция Бесселя первого рода первого порядка. На фиг. 3 показан вид круговой функции и ее образа. Отметим, что, как и следовало ожидать, образ обладает осевой симметрией и состоит из центрального пика и ряда концентрических колец уменьшающейся амплитуды. Отличительная особенность этого образа заключается в том, что его нули расположены вдоль радиуса на разных расстояниях друг от друга. С рядом других пар функций, связанных преобразованием Фурье — Бесселя, читатель может познакомиться при решении задач (см. задачу 2.4).

§ 2. ЛИНЕЙНЫЕ СИСТЕМЫ

Для удобства дальнейшего обсуждения постараемся дать понятию *система* настолько общее определение, чтобы его можно было применять как в достаточно известном случае электрических схем, так и в менее известном случае оптических устройств, создающих изображение. В соответствии с этим под системой следует понимать все то, что осуществляет преобразование ряда входных функций в ряд выходных функций. В электрических схемах входные и выходные сигналы представляют собой действительные функции (напряжения или токи) одной независимой переменной (времени), тогда как в системах, создающих изображение, входные и выходные сигналы могут быть действительными (интенсивность) или комплексными (амплитуда поля) функциями двух независимых переменных (пространственных координат). Вопрос о том, какую величину в данной системе следует считать переменной — интенсивность или амплитуду поля, будет обсуждаться несколько позднее.

Если ограничиться рассмотрением детерминированных (неслучайных) систем, то можно считать, что определенный входной сигнал должен преобразовываться в единственный выходной сигнал. Однако вовсе не

обязательно, чтобы каждый выходной сигнал соответствовал единственному входному, ибо, как мы увидим, целый ряд входных функций может совсем не создавать сигналов на выходе. Поэтому сначала мы обратимся к рассмотрению систем, отвечающих одинаковым образом на множество сигналов.

Для представления системы удобно использовать математический оператор $\mathcal{P}\{\}$, который показывает, как нужно подействовать на входные функции, чтобы получить выходные. Таким образом, если функция $g_1(x_1, y_1)$ описывает входной сигнал системы, а $g_2(x_2, y_2)$ — соответствующий выходной сигнал, то при таком определении оператора $\mathcal{P}\{\}$ эти две функции оказываются связанными соотношением

$$g_2(x_2, y_2) = \mathcal{P}\{g_1(x_1, y_1)\}. \quad (2.17)$$

Если не задать более подробно свойства оператора $\mathcal{P}\{\}$, то соотношение (2.17) фактически исчерпывает все, что мы можем сказать о системе. В дальнейшем, за редким исключением, мы будем главным образом (хотя не всегда) иметь дело с ограниченным классом систем, которые называются *линейными*. Предположение о линейности правильно отражает поведение многих простых и имеющих физический смысл систем; оно также позволяет установить полезные соотношения между входными и выходными сигналами.

Линейность и интеграл суперпозиции

Систему называют *линейной*, если для всех входных функций t и s и для всех комплексных постоянных a и b выполняется следующее свойство суперпозиции:

$$\mathcal{P}\{as(x_1, y_1) + bt(x_1, y_1)\} = a\mathcal{P}\{s(x_1, y_1)\} + b\mathcal{P}\{t(x_1, y_1)\}. \quad (2.18)$$

Как упоминалось ранее, основное преимущество линейности заключается в том, что для системы, обладающей этим свойством, ее отклик на произвольный входной сигнал можно выразить через отклики на определенные «элементарные» функции, на которые следует разложить входной сигнал. Поэтому самое важное теперь — найти простой и удобный способ разложения входного сигнала. Такое разложение можно получить,

используя так называемое *фильтрующее свойство* δ -функции (ср. приложение I), которое позволяет записать

$$g_1(x_1, y_1) = \int \int_{-\infty}^{\infty} g_1(\xi, \eta) \delta(x_1 - \xi, y_1 - \eta) d\xi d\eta. \quad (2.19)$$

Можно считать, что это соотношение является представлением g_1 в виде линейной комбинации взвешенных и смещенных δ -функций; элементарными функциями разложения служат именно эти δ -функции.

Чтобы найти отклик системы на входной сигнал g_1 , подставим (2.19) в (2.17):

$$g_2(x_2, y_2) = \mathcal{P} \left\{ \int \int_{-\infty}^{\infty} g_1(\xi, \eta) \delta(x_1 - \xi, y_1 - \eta) d\xi d\eta \right\}.$$

Будем теперь рассматривать число $g_1(\xi, \eta)$ просто как весовой множитель, с которым следует брать элементарную функцию $\delta(x_1 - \xi, y_1 - \eta)$, и, привлекая свойство линейности (2.18), подействуем оператором $\mathcal{P}\{ \}$ на отдельные элементарные функции. Тогда оператор $\mathcal{P}\{ \}$ можно ввести под знак интеграла, что дает

$$g_2(x_2, y_2) = \int \int_{-\infty}^{\infty} g_1(\xi, \eta) \mathcal{P} \{ \delta(x_1 - \xi, y_1 - \eta) \} d\xi d\eta.$$

Наконец, символом $h(x_2, y_2; \xi, \eta)$ обозначим отклик на выходе в точке с координатами (x_2, y_2) на сигнал вида δ -функции в точке с координатами (ξ, η) на входе, т. е.

$$h(x_2, y_2; \xi, \eta) = \mathcal{P} \{ \delta(x_1 - \xi, y_1 - \eta) \}. \quad (2.20)$$

Функция h называется *импульсным откликом* системы. Входной и выходной сигналы системы теперь можно связать простым соотношением

$$g_2(x_2, y_2) = \int \int_{-\infty}^{\infty} g_1(\xi, \eta) h(x_2, y_2; \xi, \eta) d\xi d\eta. \quad (2.21)$$

Это фундаментальное соотношение, известное как *интеграл суперпозиции*, выражает тот очень важный

факт, что линейная система полностью характеризуется ее откликом на входные импульсы. Чтобы полностью определить выходной сигнал, необходимо, вообще говоря, знать эти отклики для любых координат и импульсов в плоскости входного сигнала. В случае если *оптическая*¹⁾ система линейна, этот вывод имеет интересную физическую интерпретацию, заключающуюся в том, что действие создающих изображение элементов (линз, диафрагм и т. д.) может быть полностью описано, если определены (возможно, комплексные) изображения *точечных источников*, расположенных по всему полю предмета.

Инвариантные линейные системы. Передаточные функции

Определив для общей линейной системы соотношения между входными и выходными сигналами, обратимся теперь к важному подклассу линейных систем, которые называются *инвариантными* линейными системами. Говорят, что электрическая схема обладает *временной инвариантностью*, если ее импульсный отклик $h(t; \tau)$ (т. е. отклик в момент t на воздействие единичного импульса, приложенного в момент τ) зависит только от промежутка времени $(t - \tau)$. Электрические схемы, составленные из сопротивлений, емкостей и индуктивностей определенных значений, обладают временной инвариантностью, так как их характеристики не изменяются со временем.

Точно так же говорят, что линейная оптическая система *пространственно инвариантна* (или, иначе, *изопланарна*), если ее импульсный отклик $h(x_2, y_2; \xi, \eta)$ зависит только от расстояний $(x_2 - \xi)$ и $(y_2 - \eta)$. Для таких систем можно, очевидно, написать²⁾

$$h(x_2, y_2; \xi, \eta) = h(x_2 - \xi, y_2 - \eta). \quad (2.22)$$

¹⁾ Автор пользуется английским термином „imaging system”, употребляя слова «оптическая система» в более широком смысле. — *Прим. ред.*

²⁾ Так как переменные x_2, y_2 и ξ, η относятся к разным пространствам, то такое утверждение имеет смысл только при соответствующем согласовании осей координат и масштабов, что автор и делает, например, в гл. 5, § 3. — *Прим. ред.*

Таким образом, система, создающая изображение, считается пространственно инвариантной, если изображение точечного источника меняет только положение, но не свою функциональную форму по мере того, как этот источник пробегает поле предмета.

Практически системы, создающие изображение, редко бывают изопланарными по всему их предметному полю, однако обычно удается поле предмета разбить на небольшие области (*изопланарные участки*), внутри которых система приблизительно инвариантна. Чтобы полностью описать систему, создающую изображение, необходимо определить импульсный отклик, соответствующий каждому изопланарному участку, но если интересующий нас участок предметного поля имеет малые размеры, то часто оказывается достаточно рассмотреть лишь изопланарный участок, лежащий на оси системы. Заметим, что для инвариантной системы интеграл суперпозиции (2.21) принимает особенно простой вид

$$g_2(x_2, y_2) = \int \int_{-\infty}^{\infty} g_1(\xi, \eta) h(x_2 - \xi, y_2 - \eta) d\xi d\eta. \quad (2.23)$$

В правой части соотношения (2.23) нетрудно узнать двумерную *свертку* функции предмета и импульсного отклика системы. В дальнейшем будет удобно использовать более короткое обозначение для свертки типа (2.23), записывая это соотношение в виде

$$g_2 = g_1 * h;$$

звездочка между двумя функциями обозначает операцию свертки этих функций.

Для класса инвариантных линейных систем математический аппарат разработан гораздо полнее, чем для более общего класса всех линейных систем, и именно благодаря этому с инвариантными системами значительно легче иметь дело. Простота инвариантных систем является особенно наглядно, если применить преобразование Фурье к свертке (2.23).

Именно, производя преобразование правой и левой частей соотношения (2.23) и применяя теорему свертки, мы видим, что спектры $G_2(f_x, f_y)$ и $G_1(f_x, f_y)$ выходного

и входного сигналов системы связаны простым соотношением

$$G_2(f_x, f_y) = H(f_x, f_y) G_1(f_x, f_y), \quad (2.24)$$

где H — фурье-образ импульсного отклика

$$H(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\xi, \eta) \exp[-j2\pi(f_x \xi + f_y \eta)] d\xi d\eta. \quad (2.25)$$

Функция H , называемая *передаточной функцией* системы, показывает, как действует система в «частотной области». Заметим, что сравнительно трудоемкая операция свертки (2.23), необходимая для нахождения выходного сигнала системы заменяется часто в соответствии с (2.24) более простой последовательностью, состоящей из преобразования Фурье, умножения образов и обратного преобразования Фурье.

Вместе с тем соотношения (2.24) и (2.25) показывают, что для линейных инвариантных систем входной сигнал можно разлагать на элементарные функции, более удобные, чем δ -функции, входящие в (2.19). Такими элементарными функциями являются, очевидно, комплексные экспоненциальные функции. Преобразуя g_1 , мы просто разлагаем входной сигнал на комплексные экспоненциальные функции разных пространственных частот (f_x, f_y) . Умножая затем спектр входного сигнала G_1 на передаточную функцию H , мы учитываем воздействие системы на каждую элементарную функцию. Заметим, что эти воздействия приводят только к изменению амплитуды и сдвигу фазы. Об этом свидетельствует тот факт, что для каждой пары частот (f_x, f_y) мы просто умножаем спектр входного сигнала на комплексное число $H(f_x, f_y)$. Обратное преобразование спектра выходного сигнала G_2 восстанавливает выходной сигнал g_2 путем сложения всех видоизмененных элементарных функций.

В заключение следует особо отметить, что использование теории передаточной функции позволяет упростить рассмотрение только в случае *инвариантных* линейных систем. С применением теории Фурье в анализе неста-

ционарных электрических цепей и пространственно-неоднородных оптических систем читатель может познакомиться по работам [4, 5].

§ 3. ДВУМЕРНАЯ ТЕОРИЯ ВЫБОРКИ

Для обработки информации и для целей математического анализа часто оказывается удобным задавать функцию $g(x, y)$ набором ее выборочных значений, взятых в дискретной совокупности точек в плоскости x, y . Интуитивно ясно, что если эти значения взяты достаточно близко друг к другу, то выборочные данные являются хорошим представлением исходной функции; иначе говоря, функцию g можно восстановить с большой точностью путем простого интерполирования. Менее очевидным представляется тот факт, что для определенного класса функций (известных под названием функций с *ограниченной шириной спектра*) можно осуществить *точное* восстановление, если потребовать, чтобы интервал между выборочными значениями не превышал определенного предельного значения. Это обстоятельство было впервые отмечено Уиттекером [6] и позднее изложено в общедоступной форме Шенноном [7] в его работах по теории информации.

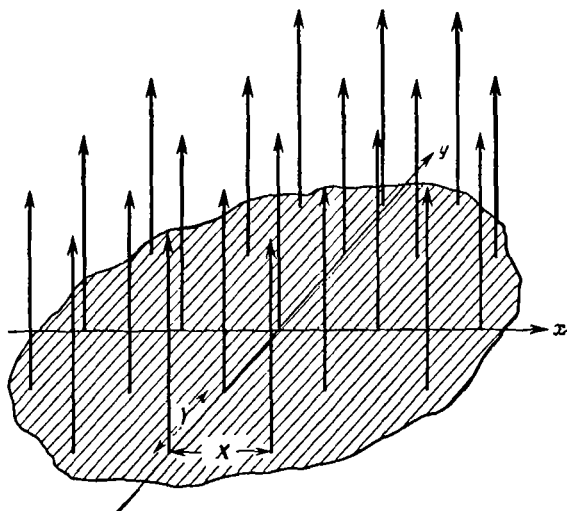
Теорема выборки справедлива для класса функций с ограниченной шириной спектра. К этому классу мы относим функции, фурье-образы которых не равны нулю только в пределах конечной области \mathcal{R} пространства частот.

Сначала рассмотрим формулировку этой теоремы, совершенно аналогичную одномерной теореме Шеннона. Позднее мы кратко отметим те изменения, которые могут быть внесены в отдельных двумерных случаях.

Теорема Уиттекера — Шеннона

Чтобы получить самый простой вариант теоремы выборки, рассмотрим прямоугольную сетку выборок функции g , которая определяется следующим образом:

$$g_s(x, y) = \text{comb}\left(\frac{x}{X}\right) \text{comb}\left(\frac{y}{Y}\right) g(x, y). \quad (2.26)$$



Фиг. 4. Выборочная функция.

Таким образом, выборочная функция g_s состоит из ряда δ -функций, расположенных на расстоянии X в направлении x и на расстоянии Y в направлении y , как показано на фиг. 4. Площадь под каждой δ -функцией пропорциональна значению функции g в данной точке прямоугольной выборочной сетки. Из теоремы свертки следует, что спектр G_s функции g_s можно найти путем свертки образов функций $\text{comb}(x/X) \text{comb}(y/Y)$ и g , или

$$G_s(f_x, f_y) = \mathcal{F} \left\{ \text{comb} \left(\frac{x}{X} \right) \text{comb} \left(\frac{y}{Y} \right) \right\} * G(f_x, f_y),$$

где звездочка опять обозначает операцию свертки. Теперь, используя табл. 1 и теорему подобия, имеем

$$\mathcal{F} \left\{ \text{comb} \left(\frac{x}{X} \right) \text{comb} \left(\frac{y}{Y} \right) \right\} = XY \text{comb}(Xf_x) \text{comb}(Yf_y),$$

в то время как из решения задачи 2.1, п. «б», следует

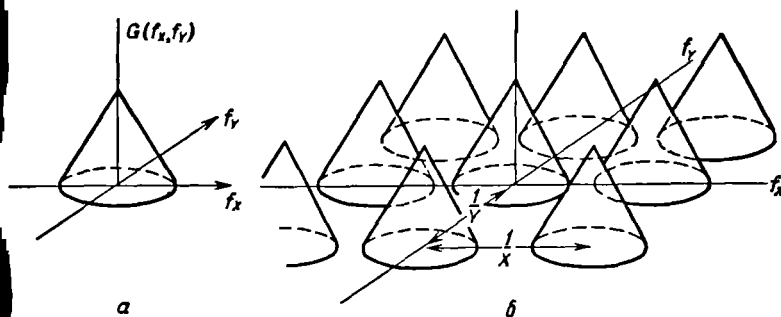
$$XY \text{comb}(Xf_x) \text{comb}(Yf_y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \delta \left(f_x - \frac{n}{X}, f_y - \frac{m}{Y} \right).$$

Отсюда вытекает, что спектр выборочной функции имеет следующий вид:

$$G_s(f_x, f_y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} G\left(f_x - \frac{n}{X}, f_y - \frac{m}{Y}\right). \quad (2.27)$$

Таким образом, спектр функции g_s можно найти просто путем построения спектра функции g возле каждой точки $(n/X, m/Y)$ в плоскости $f_x f_y$, как это показано на фиг. 5.

Предполагается, что функция g имеет ограниченный спектр; следовательно, функция G отлична от нуля только в конечной области \mathcal{R} пространства частот. Как видно из соотношения (2.27), область, в которой спектр выборочной функции отличен от нуля, можно найти путем построения области \mathcal{R} около каждой точки $(n/X, m/Y)$ в плоскости частот. Теперь ясно, что если X и Y достаточно малы (т. е. выборочные значения расположены достаточно близко друг к другу), то расстояния $1/X$ и $1/Y$ между различными спектральными областями будут достаточно большими, чтобы смежные области не перекрывались (см. фиг. 5). Таким образом, чтобы осуществить точное восстановление исходного спектра G по спектру G_s , нужно пропустить выборочную функцию g_s через линейный фильтр, который пропускает без искажения члены соотношения (2.27) с индексами $n = 0, m = 0$, тогда как остальные члены полностью исклю-



Фиг. 5. Спектры исходной функции (а) и выборочных значений (б).

чаются. На выходе этого фильтра мы, очевидно, получим точное воспроизведение первоначальной функции $g(x, y)$.

Как было показано выше, для успешного восстановления исходных данных выборочные значения должны быть расположены достаточно близко друг к другу; это необходимо для того, чтобы различные области спектра G_s не перекрывались. Чтобы определить максимально допустимое расстояние между выборочными значениями, рассмотрим *наименьший* прямоугольник¹⁾, который полностью заключает в себе область \mathcal{R} , и обозначим через $2B_x$ и $2B_y$ его стороны в направлениях f_x и f_y соответственно. Так как различные составляющие в спектре выборочных данных (2.27) находятся друг от друга на расстояниях, равных $1/X$ и $1/Y$ соответственно в направлениях f_x и f_y , то можно считать, что разделение спектральных областей заведомо имеет место, если

$$X \leq \frac{1}{2B_x} \quad \text{и} \quad Y \leq \frac{1}{2B_y}. \quad (2.28)$$

Таким образом, для точного восстановления исходной функции *максимальные* интервалы выборочной сетки должны быть равны $(2B_x)^{-1}$ и $(2B_y)^{-1}$.

После того как мы нашли максимально допустимые интервалы между выборочными значениями, остается определить точный вид передаточной функции фильтра, через который должны быть пропущены выборочные данные. В большинстве случаев при этом имеется широкий выбор, так как для многих возможных форм области \mathcal{R} существует множество передаточных функций, которые будут пропускать составляющую спектра G_s с $n = 0$, $m = 0$, исключая при этом все остальные члены. Нам, однако, достаточно знать, что в том случае, когда соотношение (2.28) удовлетворяется, имеется такая передаточная функция, которая всегда будет давать желаемый

¹⁾ Для простоты мы предполагаем, что центр этого прямоугольника расположен в начале координат. Если это не так, то нетрудно так изменить аргументы, чтобы распространить теорему выборки и на этот случай.

результат независимо от формы области \mathcal{A} , а именно

$$\mathbf{H}(f_X, f_Y) = \text{rect}\left(\frac{f_X}{2B_X}\right) \text{rect}\left(\frac{f_Y}{2B_Y}\right). \quad (2.29)$$

Возможность точного восстановления функции \mathbf{G} по функции \mathbf{G}_s очевидна, если учесть, что спектр выходного сигнала после такого фильтра имеет вид

$$\mathbf{G}_s(f_X, f_Y) \text{rect}\left(\frac{f_X}{2B_X}\right) \text{rect}\left(\frac{f_Y}{2B_Y}\right) \equiv \mathbf{G}(f_X, f_Y).$$

Эквивалентное тождество в пространстве координат можно записать следующим образом:

$$\left[\text{comb}\left(\frac{x}{X}\right) \text{comb}\left(\frac{y}{Y}\right) g(x, y) \right] * \mathbf{h}(x, y) = g(x, y), \quad (2.30)$$

где \mathbf{h} — импульсный отклик фильтра:

$$\mathbf{h}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left(\frac{f_X}{2B_X}\right) \text{rect}\left(\frac{f_Y}{2B_Y}\right) \exp[j2\pi(f_X x + f_Y y)] df_X df_Y = 4B_X B_Y \text{sinc}(2B_X x) \text{sinc}(2B_Y y).$$

С учетом того, что

$$\begin{aligned} \text{comb}\left(\frac{x}{X}\right) \text{comb}\left(\frac{y}{Y}\right) g(x, y) &= \\ &= XY \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g(nX, mY) \delta(x - nX, y - mY), \end{aligned}$$

соотношение (2.30) принимает вид

$$\begin{aligned} g(x, y) &= 4B_X B_Y XY \times \\ &\times \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g(nX, mY) \text{sinc}[2B_X(x - nX)] \text{sinc}[2B_Y(y - mY)]. \end{aligned}$$

Наконец, когда интервалы X и Y между выборочными значениями имеют максимально допустимые значения, это тождество можно записать в виде

$$\begin{aligned} g(x, y) &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g\left(\frac{n}{2B_X}, \frac{m}{2B_Y}\right) \times \\ &\times \text{sinc}\left[2B_X\left(x - \frac{n}{2B_X}\right)\right] \text{sinc}\left[2B_Y\left(y - \frac{m}{2B_Y}\right)\right]. \quad (2.31) \end{aligned}$$

Соотношение (2.31) представляет собой фундаментальный результат, который мы будем называть *теоремой выборки Уиттекера — Шеннона*. Она означает, что, имея расположенный соответствующим образом прямоугольный ряд выборочных значений функции с ограниченной шириной спектра, можно осуществить точное ее восстановление.

Для восстановления необходимо ввести в каждой точке выборки интерполяционную функцию, равную произведению sinc функций.

Полученный выше результат не единственно возможный вид теоремы выборки. В наших рассуждениях мы два раза сделали довольно произвольный выбор. При других предположениях получатся иные варианты теоремы выборки. Первый элемент произвольности — это использование *прямоугольной* выборочной сетки, второй — выбор особого вида передаточной функции (2.29).

Другие варианты теоремы, которые получаются при использовании иных, отличных от наших предположений относительно выборочной сетки и передаточной функции, конечно, не менее справедливы, чем теорема (2.31). В некоторых случаях альтернативные теоремы даже могут быть более «эффективными» в том смысле, что для обеспечения полного восстановления достаточно иметь меньше выборок на единицу площади. Читателю, интересующемуся многомерной теорией выборки, мы рекомендуем обратиться к работам Брэйсуэлла [8] и Петерсона и Миддлтона [9]. Кроме того, в работе Линдена [10] обсуждаются теоремы выборки, включающие не только значения функции, но и значения ее производных.

ЗАДАЧИ

2.1. Доказать следующие свойства δ -функций:

$$a) \delta(ax, by) = \frac{1}{|ab|} \delta(x, y),$$

$$b) \text{comb}(ax) \text{comb}(by) = \frac{1}{|ab|} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \delta\left(x - \frac{n}{a}, y - \frac{m}{b}\right).$$

2.2. Доказать следующие соотношения для фурье-образов:

а) $\mathcal{F}\{\text{rect}(x)\text{rect}(y)\} = \text{sinc}(f_X)\text{sinc}(f_Y)$;

б) $\mathcal{F}\{\Lambda(x)\Lambda(y)\} = \text{sinc}^2(f_X)\text{sinc}^2(f_Y)$.

Доказать следующие соотношения для обобщенных фурье-образов:

в) $\mathcal{F}\{1\} = \delta(x, y)$;

г) $\mathcal{F}\{\text{sgn}(x)\text{sgn}(y)\} = \left(\frac{1}{j\pi f_X}\right)\left(\frac{1}{j\pi f_Y}\right)$.

2.3. Доказать следующие теоремы для фурье-образов:

а) $\mathcal{F}\mathcal{F}\{g(x, y)\} = \mathcal{F}^{-1}\mathcal{F}^{-1}\{g(x, y)\} = g(-x, -y)$

во всех точках непрерывности функции g ;

б) $\mathcal{F}\{g(x, y)h(x, y)\} = \mathcal{F}\{g(x, y)\} * \mathcal{F}\{h(x, y)\}$;

в) $\mathcal{F}\{\nabla^2 g(x, y)\} = -4\pi^2(f_X^2 + f_Y^2)\mathcal{F}\{g(x, y)\}$,

где ∇^2 — оператор Лапласа

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

2.4. Доказать следующие соотношения для преобразований Фурье — Бесселя:

а) если $g_R(r) = \delta(r - r_0)$, то

$$\mathcal{B}\{g_R(r)\} = 2\pi r_0 J_0(2\pi r_0 \rho);$$

б) если $g_R(r) = 1$ для $a \leq r \leq 1$ и $g_R(r) = 0$ в остальных случаях, то

$$\mathcal{B}\{g_R(r)\} = \frac{J_1(2\pi\rho) - aJ_1(2\pi a\rho)}{\rho};$$

в) если $\mathcal{B}\{g_R(r)\} = G(\rho)$, то

$$\mathcal{B}\{g_R(ar)\} = \frac{1}{a^2} G\left(\frac{\rho}{a}\right);$$

г) $\mathcal{B}\{\exp(-\pi r^2)\} = \exp(-\pi \rho^2)$.

2.5. Выражение

$$p(x, y) = g(x, y) * \left[\text{comb}\left(\frac{x}{X}\right)\text{comb}\left(\frac{y}{Y}\right)\right]$$

определяет периодическую функцию с периодом X в направлении x и периодом Y в направлении y .

а) Показать, что фурье-образ функции p может быть записан в виде

$$P(f_X, f_Y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} G\left(\frac{n}{X}, \frac{m}{Y}\right) \delta\left(f_X - \frac{n}{X}, f_Y - \frac{m}{Y}\right),$$

где G — фурье-образ функции g .

б) Схематически изобразить функцию $p(x, y)$, если

$$g(x, y) = \text{rect}\left(2\frac{x}{X}\right) \text{rect}\left(2\frac{y}{Y}\right),$$

и найти соответствующий фурье-образ $P(f_X, f_Y)$.

2.6. Пусть операторы преобразования $\mathcal{F}_A\{\}$ и $\mathcal{F}_B\{\}$ имеют вид

$$\mathcal{F}_A\{g\} = \frac{1}{a} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) \exp\left[-j\frac{2\pi}{a}(f_X\xi + f_Y\eta)\right] d\xi d\eta,$$

$$\mathcal{F}_B\{g\} = \frac{1}{b} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) \exp\left[-j\frac{2\pi}{b}(x\xi + y\eta)\right] d\xi d\eta.$$

а) Найти простое выражение для

$$\mathcal{F}_B\{\mathcal{F}_A\{g(x, y)\}\}.$$

б) Объяснить результаты для $a > b$ и для $a < b$.

2.7. Пусть $g(r, \theta)$ — функция с разделяющимися переменными в полярных координатах.

а) Показать, что если $g(r, \theta) = g_R(r) e^{im\theta}$, то

$$\mathcal{F}\{g(r, \theta)\} = (-j)^m e^{im\varphi} \mathcal{H}_m\{g_R(r)\},$$

где $\mathcal{H}_m\{\}$ — преобразование Ханкеля порядка m ;

$$\mathcal{H}_m\{g_R(r)\} = 2\pi \int_0^{\infty} r g_R(r) J_m(2\pi r \rho) dr,$$

а (ρ, φ) — полярные координаты в пространстве частот.

Указание:

$$\exp(ja \sin x) = \sum_{k=-\infty}^{\infty} J_k(a) \exp(jkx).$$

б) Показать, что для более общего случая произвольной угловой зависимости $g_R(\theta)$ преобразование Фурье может быть выражено при помощи следующей бесконечной последовательности преобразо-

ваний Ханкеля:

$$\mathcal{F}\{g(r, \theta)\} = \sum_{k=-\infty}^{\infty} c_k (-j)^k e^{jk\theta} \mathcal{H}_k\{g_R(r)\},$$

где

$$c_k = \frac{1}{2\pi} \int_0^{2\pi} g_\theta(\theta) e^{-jk\theta} d\theta.$$

2.8. Предположим, что на входе линейной системы действует синусоидальный сигнал

$$g(x, y) = \cos[2\pi(f_x x + f_y y)].$$

При каком (достаточном) условии выходной сигнал будет действительной синусоидальной функцией той же самой пространственной частоты, что и входной сигнал? Выразить амплитуду и фазу этого выходного сигнала через соответствующие параметры системы.

2.9. Показать, что функция, спектральные компоненты которой отличны от нуля вне круга радиусом B в плоскости частот, удовлетворяет следующей теореме выборки:

$$g(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g\left(\frac{n}{2B}, \frac{m}{2B}\right) \left\{ \frac{2\pi B^2 J_1[2\pi B \sqrt{(x-n/2B)^2 + (y-m/2B)^2}]}{2\pi B \sqrt{(x-n/2B)^2 + (y-m/2B)^2}} \right\}.$$

2.10. Действие оператора преобразования Фурье можно рассматривать как отображение функций в их образы; следовательно, оно удовлетворяет определению системы, принятому в этой главе.

а) *Линейна* ли эта система?

б) Можно ли определить *передаточную функцию*, которая характеризует эту систему? Если да, то какова она? Если нет, то почему?

2.11. «Эквивалентную площадь» Δ_{xy} функции $g(x, y)$ можно определить так:

$$\Delta_{xy} = \left| \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) dx dy}{g(0, 0)} \right|,$$

в то время как «эквивалентная ширина спектра» функции g связана с ее образом G следующим соотношением:

$$\Delta_{f_x f_y} = \left| \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f_x, f_y) df_x df_y}{G(0, 0)} \right|.$$

Показать, что $\Delta_{xy} \Delta_{f_x f_y} = 1$.

2.12. Некоторая комплексная функция двух независимых переменных (x, y) имеет пространственный фурье-образ, который тождественно равен нулю при всех частотах, кроме области $|f_x| \leq B_x$, $|f_y| \leq B_y$. Показать, что часть этой функции, простирающаяся за пределы области $|x| \leq X$, $|y| \leq Y$ в конфигурационном пространстве, может быть (приблизительно) определена при помощи $32B_x B_y X Y$ действительных чисел. Почему это только приближение и когда такое приближение допустимо? (Число $16B_x B_y X Y$ обычно называется пространственно-полосовым произведением рассматриваемой части функции.)

2.13. Входной сигнал некоторой оптической системы представляет собой распределение комплексного поля предмета $U_0(x, y)$, спектр пространственных частот которого неограничен. Выходной сигнал этой системы представляет собой распределение поля изображения $U_i(x, y)$. Предполагается, что система, создающая изображение, действует как линейный пространственно-инвариантный низкочастотный фильтр, передаточная функция которого тождественно равна нулю во всей частотной области, кроме участка, где $|f_x| \leq B_x$, $|f_y| \leq B_y$. Показать, что существует «эквивалентный» предмет $U'_0(x, y)$, состоящий из прямоугольной сетки точечных источников и дающий точно такое же изображение U_i , как и истинный предмет U_0 , и что распределение поля по эквивалентному предмету можно записать в виде

$$U'_0(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_0(\xi, \eta) \operatorname{sinc}(n - 2B_x \xi) \times \right. \\ \left. \times \operatorname{sinc}(m - 2B_y \eta) d\xi d\eta \right] \delta \left(x - \frac{n}{2B_x}, y - \frac{m}{2B_y} \right).$$

ЛИТЕРАТУРА¹⁾

1. Papoulis A., The Fourier Integral and Its Applications, New York, 1962.
2. Bracewell R. N., The Fourier Transform and Its Applications, New York, 1965.
3. Lighthill M. J., Introduction to Fourier Analysis and Generalized Functions, New York, 1960.
4. Kailath T., в книге Lectures on Communication System Theory, ed. E. J. Baghdady, New York, 1960.
Характеристики канала. Дисперсные каналы, меняющиеся во времени.
5. Lohmann A. W., Paris D. P., Journ. Opt. Soc. Am., 55, 1007 (1965).
Формирование пространственно-переменного изображения.

¹⁾ Здесь и далее литература, отмеченная звездочкой, добавлена редактором перевода. — Прим. ред.

6. Whittaker E. T., Proc. Roy. Soc. Edinburgh, Sect. A, 35, 181 (1915).
О функциях, которые могут быть представлены разложениями теории интерполяции.
7. Shannon C. E., Proc. IRE, 37, 10 (1949).
Связь в присутствии шумов.
8. Bracewell R. N., Australia Journ. Phys., 9, 297 (1956).
Двумерное сглаживание в радиоастрономии.
9. Peterson D. P., Middleton D., Information and Control, 5, 279 (1962).
Выборка и восстановление функций с ограниченным волновым спектром в N -мерном евклидовом пространстве.
10. Linden D. A., Proc. IRE, 47, 1219 (1959).
Обсуждение теорем выборки.
- 11*. Снеддон И., Преобразование Фурье, ИЛ, 1955.
- 12*. Гельфанд И. М., Шиллов Г. Е., Обобщенные функции и действия над ними, М., 1959.
- 13*. Конторович М. И., Операционное исчисление и процессы в электрических цепях, М., 1964.
- 14*. Харкевич А. А., Спектры и анализ, М., 1962.
- 15*. Харкевич А. А., Теоретические основы радиосвязи, М., 1957.
- 16*. Боде Г., Шеннон К., Теория информации и ее приложения, М., 1959.

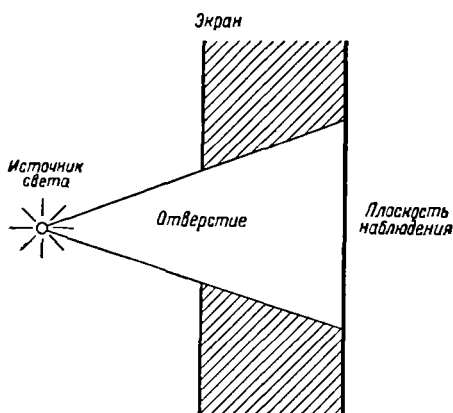
ГЛАВА 3

ОСНОВЫ СКАЛЯРНОЙ ТЕОРИИ ДИФРАКЦИИ

Явление *дифракции* играет чрезвычайно важную роль в разделах физики и техники, так или иначе связанных с вопросами распространения волн. В этой главе мы рассмотрим некоторые основные положения скалярной теории дифракции. Хотя обсуждаемая здесь теория имеет достаточно общий характер и применима в разных областях, в частности к рассмотрению распространения акустических волн и радиоволн, основное внимание будет сосредоточено на ее использовании в физической оптике. Чтобы полностью понять свойства оптических систем, образующих изображение, и оптических систем обработки информации, важно уметь оценивать влияние дифракции и обусловленных ею ограничений характеристик системы. В последующих параграфах мы дадим ряд ссылок на работы, содержащие более подробное изложение теории дифракции.

§ 1. ИСТОРИЧЕСКИЙ ОБЗОР

Зоммерфельд [1] удачно определил *дифракцию* как «любое отклонение световых лучей от прямой линии, которое нельзя объяснить отражением или преломлением». Первое сообщение с точным описанием такого явления было сделано Гримальди; оно было опубликовано в 1665 г., вскоре после его смерти. Описанные им измерения проводились при помощи устройства, схема которого изображена на фиг. 6: отверстие в непрозрачном экране освещается источником света, а в плоскости, расположенной на некотором расстоянии позади экрана, измеряется освещенность. Из корпускулярной теории света, которая в то время была общепринятой для объяснения оптических явлений, следовало, что тень позади

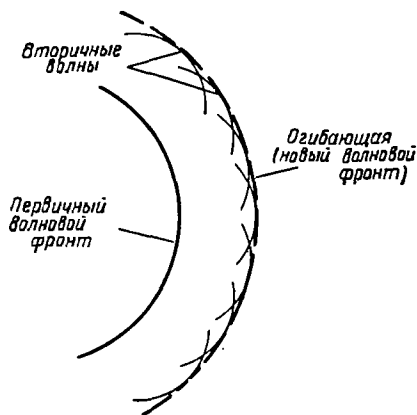


Фиг. 6. Схема наблюдения дифракции света.

экрана должна иметь четкие границы. Тем не менее наблюдения Гримальди указывали на то, что переход от света к тени происходит постепенно, а не резко. Если бы Гримальди располагал более совершенным источником света, то он смог бы наблюдать такое поразительное явление, как чередование светлых и темных полос в области геометрической тени экрана¹⁾. Такие эффекты не могли найти удовлетворительного объяснения на основе корпускулярной теории света, согласно которой в отсутствие отражения и преломления лучи света должны распространяться прямолинейно.

Первый шаг в развитии теории, которая позволяла объяснить такие эффекты, был сделан в 1678 г. основоположником волновой теории света Христианом Гюйгенсом. Гюйгенс выдвинул интуитивное утверждение, которое можно сформулировать следующим образом: если каждую точку волнового фронта светового возмущения рассматривать как новый источник «вторичного»

¹⁾ Для экрана с отверстием полосы, заходящие в область тени, появляются только на больших расстояниях, соответствующих дифракции Фраунгофера (см. гл. 4, § 2). На малых расстояниях осцилляции интенсивности наблюдаются в освещенной области (см. фиг. 52). — *Прим. ред.*



Фиг. 7. Построение огибающей Гюйгенса.

сферического возмущения, то в любой последующий момент времени волновой фронт можно найти путем построения огибающей вторичных слабых волн, как показано на фиг. 7. Интуитивные идеи Гюйгенса были существенно преобразованы в 1818 г. в знаменитом мемуаре Огюстена Жана Френеля, который дополнил гюйгенсово построение огибающей принципом интерференции Юнга. Сделав некоторые довольно произвольные допущения относительно эффективных амплитуд и фаз вторичных источников Гюйгенса и предположив, что вторичные волны интерферируют друг с другом, Френель смог с большой точностью рассчитать распределение света в дифракционных картинах.

Идеи Гюйгенса и Френеля в 1882 г. были поставлены на более прочную математическую основу Густавом Кирхгофом, которому удалось показать, что особенности амплитуд и фаз, приписываемые Френелем вторичным источникам, логически следуют из волновой природы света. Кирхгоф основывал свои математические построения на двух предположениях относительно граничных значений светового возмущения, падающего на поверхность препятствия, расположенного на пути света. Позднее Пуанкаре (в 1892 г.) и Зоммерфельд

(в 1894 г.) доказали, что эти предположения несовместимы друг с другом¹⁾).

В свете этих критических замечаний предложенную Кирхгофом формулировку так называемого *принципа Гюйгенса — Френеля* следует рассматривать как первое приближение, хотя в большинстве случаев она дает поразительное согласие с экспериментальными данными. Котлер [2] пытался разрешить противоречия путем сведения задачи Кирхгофа о *граничных значениях* к задаче с разрывными решениями. Теория Кирхгофа была видоизменена также Зоммерфельдом, который исключил одно из вышеупомянутых граничных условий путем использования теории функций Грина. Эта так называемая *теория дифракции Релея — Зоммерфельда* будет изложена в § 4.

С самого начала следует подчеркнуть, что в теориях Кирхгофа и Релея — Зоммерфельда используются некоторые существенные упрощения и приближения. Наиболее важное из них состоит в том, что эти теории являются *скалярными*, т. е. рассматривается только скалярная амплитуда одной поперечной компоненты электрического или магнитного поля. При этом предполагается, что любые другие представляющие интерес компоненты можно рассматривать независимо таким же образом. При таком подходе по существу полностью игнорируется тот факт, что различные компоненты векторов электрического и магнитного полей *связаны* уравнениями Максвелла и поэтому их нельзя рассматривать независимо. К счастью, эксперименты в микроволновой области спектра [3] показали, что скалярная теория дает очень точные результаты, если выполняются два условия: 1) отверстия в экранах велики по сравнению с длиной волны и 2) дифрагированные волны наблюдаются не слишком близко от экранов. Эти условия хорошо удовлетворяются в рассматриваемых нами задачах. Более полное обсуждение применимости скалярной теории в оптике приборов читатель может найти в работе [4]. Тем не менее в ряде важных задач

¹⁾ Более подробное рассмотрение этих противоречий приведено в § 4.

необходимые условия не удовлетворяются; в качестве примера можно назвать теорию дифракционных решеток высокого разрешения [5]. Подобные задачи мы не будем рассматривать, так как для получения достаточно точных результатов *необходимо* учитывать векторную природу полей. Векторные обобщения теории дифракции действительно существуют, причем первое удовлетворительное рассмотрение выполнено Котлером [6].

Первое по-настоящему строгое решение дифракционной задачи дал в 1896 г. Зоммерфельд [7], который рассмотрел двумерный случай плоской волны, падающей на бесконечно тонкую идеально проводящую полуплоскость. Котлер [8] позднее сопоставил решение Зоммерфельда с соответствующими результатами скалярной теории Кирхгофа.

Само собой разумеется, что наше историческое введение в предмет, так широко представленный в литературе, не может быть полным. Поэтому читателю мы рекомендуем обратиться к всестороннему изложению теории дифракции, которое содержится, например, в работах [9—11].

§ 2. ПРЕДВАРИТЕЛЬНЫЕ МАТЕМАТИЧЕСКИЕ СВЕДЕНИЯ

Прежде чем начать изложение теории дифракции, дадим сначала ряд математических сведений, которые образуют основу последующего рассмотрения теории дифракции. Попутно мы будем вводить обозначения, которые используются в дальнейшем.

Уравнение Гельмгольца

Пусть световое возмущение в точке P в момент времени t описывается скалярной функцией $u(P, t)$; в случае линейно поляризованных волн можно считать, что эта функция представляет напряженность электрического или магнитного поля. Сначала ограничимся случаем чисто монохроматических волн; обобщение результатов на случай полихроматических волн рассматривается в § 5.

Развернутое выражение для поля монохроматической волны имеет вид

$$u(P, t) = U(P) \cos [2\pi\nu t + \varphi(P)], \quad (3.1)$$

где $U(P)$ и $\varphi(P)$ — соответственно амплитуда и фаза волны в точке P , а ν — оптическая частота. Используя комплексные обозначения, выражение (3.1) можно переписать в более компактном виде

$$u(P, t) = \operatorname{Re} [U(P) e^{-i2\pi\nu t}], \quad (3.2)$$

где $U(P)$ — комплексная функция координат (иногда называемая *фазором*):

$$U(P) = U(P) e^{-i\varphi(P)}, \quad (3.3)$$

а Re — сокращенное обозначение *действительной части* функции, заключенной в квадратных скобках.

Если действительная функция возмущения $u(P, t)$ описывает оптическую волну, то в каждой точке, где нет источников, она должна удовлетворять скалярному волновому уравнению

$$\nabla^2 u - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = 0; \quad (3.4)$$

здесь ∇^2 — оператор Лапласа:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Задание комплексной функции $U(P)$ достаточно для описания возмущения, так как зависимость U от времени известна заранее. Подставляя выражение (3.2) в уравнение (3.4), получаем, что комплексная функция возмущения U должна удовлетворять не зависящему от времени уравнению

$$(\nabla^2 + k^2) U = 0, \quad (3.5)$$

где величина

$$k = 2\pi \frac{\nu}{c} = \frac{2\pi}{\lambda}$$

называется волновым числом.

Уравнение (3.5) носит название *уравнения Гельмгольца*; мы можем принять на будущее, что комплексная амплитуда любого монохроматического оптического возмущения, распространяющегося в свободном пространстве, должна подчиняться такому уравнению.

Теорема Грина

Расчет комплексной функции возмущения U в рассматриваемой точке пространства может быть выполнен при помощи соотношения, называемого *теоремой Грина*. Эта теорема, которая приводится в большинстве учебников (см., например, [12]), может быть сформулирована следующим образом.

Теорема Грина. Пусть $U(P)$ и $G(P)$ — две произвольные комплексные функции координат, а S — замкнутая поверхность, ограничивающая объем V . Если функции U , G и их первые и вторые частные производные однозначны и непрерывны внутри объема, ограниченного поверхностью S , и на самой поверхности S , то

$$\iiint_V (G \nabla^2 U - U \nabla^2 G) dv = \iint_S \left(G \frac{\partial U}{\partial n} - U \frac{\partial G}{\partial n} \right) ds, \quad (3.6)$$

где $\partial/\partial n$ обозначает частную производную в каждой точке поверхности S , взятую по направлению *внешней* нормали к этой поверхности.

Эта теорема во многих отношениях является основным звеном скалярной теории дифракции. Однако только при осторожном выборе так называемой *функции Грина* G и замкнутой поверхности S теореме можно непосредственно применять к дифракционным задачам. Обратимся теперь к первой из этих задач, а именно рассмотрим выбор функции Грина, сделанный Кирхгофом, и интегральную теорему, следующую из этого выбора.

Интегральная теорема Гельмгольца — Кирхгофа

Формулировка дифракционной задачи, предложенная Кирхгофом, основана на интегральной теореме, которая выражает решение однородного волнового уравнения в произвольной точке через значения этого решения и его первой производной на произвольной замкнутой поверхности, окружающей рассматриваемую точку. Эта теорема была ранее выведена в акустике Гельмгольцем.

Обозначим выбранную точку через P_0 , а через S — окружающую ее произвольную замкнутую поверхность, как показано на фиг. 8. Задача состоит в том, чтобы выразить оптическое возмущение в точке P_0 через его значения на поверхности S . Для решения этой задачи, следуя Кирхгофу, используем теорему Грина и выберем в качестве функции Грина G сферическую волну единичной амплитуды, распространяющуюся из точки P_0 (т. е. так называемую функцию Грина *свободного пространства*). Таким образом, для функции G в произвольной точке P_1 имеем

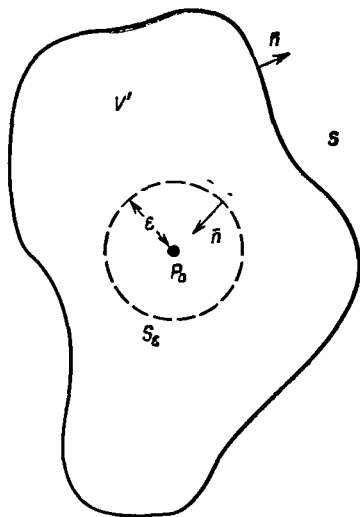
$$G(P_1) = \frac{\exp(jkr_{01})}{r_{01}}, \quad (3.7)$$

где через r_{01} мы обозначили длину вектора \vec{r}_{01} , направленного из точки P_0 в точку P_1 .

Чтобы мы имели право пользоваться теоремой Грина, функция G и ее первая и вторая производные должны быть непрерывны в объеме V , ограниченном поверхностью S . Поэтому для того, чтобы исключить точку разрыва, окружим P_0 небольшой сферической поверхностью S_ϵ радиусом ϵ . Затем применим теорему Грина, причем интегрирование будем вести по объему V' , заключенному между поверхностями S и S_ϵ , так что поверхностью интегрирования будет составная поверхность

$$S' = S + S_\epsilon,$$

как показано на фиг. 8. Заметим, что «внешняя» нормаль к составной поверхности направлена наружу (в общепринятом смысле) на поверхности S и внутрь на поверхности S_ϵ (по направлению к P_0).



Фиг. 8. Поверхность интегрирования.

Внутри объема V' возмущение \mathbf{G} , представляющее собой расходящуюся сферическую волну, удовлетворяет уравнению Гельмгольца

$$(\nabla^2 + k^2) \mathbf{G} = 0. \quad (3.8)$$

Используя оба уравнения Гельмгольца (3.5) и (3.8) для преобразования левой части формулы Грина, находим

$$\iiint_{V'} (\mathbf{G} \nabla^2 \mathbf{U} - \mathbf{U} \nabla^2 \mathbf{G}) dv = - \iiint_{V'} (\mathbf{G} \mathbf{U} k^2 - \mathbf{U} \mathbf{G} k^2) dv \equiv 0.$$

Тогда теорема (3.6) запишется следующим образом:

$$\iint_{S'} \left(\mathbf{G} \frac{\partial \mathbf{U}}{\partial n} - \mathbf{U} \frac{\partial \mathbf{G}}{\partial n} \right) ds = 0,$$

или

$$- \iint_{S_\varepsilon} \left(\mathbf{G} \frac{\partial \mathbf{U}}{\partial n} - \mathbf{U} \frac{\partial \mathbf{G}}{\partial n} \right) ds = \iint_S \left(\mathbf{G} \frac{\partial \mathbf{U}}{\partial n} - \mathbf{U} \frac{\partial \mathbf{G}}{\partial n} \right) ds. \quad (3.9)$$

Теперь заметим, что для произвольной точки P_1 на поверхности S'

$$\mathbf{G}(P_1) = \frac{\exp(jkr_{01})}{r_{01}} \quad (3.10)$$

и

$$\frac{\partial \mathbf{G}(P_1)}{\partial n} = \cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}) \left(jk - \frac{1}{r_{01}} \right) \exp \frac{jkr_{01}}{r_{01}}, \quad (3.11)$$

где $\cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01})$ представляет собой косинус угла между направлением внешней нормали $\bar{\mathbf{n}}$ и вектором $\bar{\mathbf{r}}_{01}$, соединяющим точки P_0 и P_1 . Для частного случая, когда точка P_1 лежит на поверхности S_ε , имеем $\cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}) = -1$, и эти выражения принимают вид

$$\mathbf{G}(P_1) = \frac{e^{jke}}{\varepsilon} \quad \text{и} \quad \frac{\partial \mathbf{G}(P_1)}{\partial n} = \frac{e^{jke}}{\varepsilon} \left(\frac{1}{\varepsilon} - jk \right). \quad (3.12)$$

Если ε стремится к нулю, то в силу непрерывности функции \mathbf{U} (и ее производных) в точке P_0 мы можем за-

писать

$$\iint_{S_e} \left(\mathbf{G} \frac{\partial \mathbf{U}}{\partial n} - \mathbf{U} \frac{\partial \mathbf{G}}{\partial n} \right) ds = 4\pi e^2 \left[\frac{\partial \mathbf{U}(P_0)}{\partial n} \frac{e^{jke}}{e} - \mathbf{U}(P_0) \frac{e^{jke}}{e} \left(\frac{1}{e} - jk \right) \right]_{e \rightarrow 0} = -4\pi \mathbf{U}(P_0).$$

Подстановка этого результата в (3.9) дает

$$\mathbf{U}(P_0) = \frac{1}{4\pi} \iint_S \left\{ \frac{\partial \mathbf{U}}{\partial n} \left[\frac{\exp(jkr_{01})}{r_{01}} \right] - \mathbf{U} \frac{\partial}{\partial n} \left[\frac{\exp(jkr_{01})}{r_{01}} \right] \right\} ds. \quad (3.13)$$

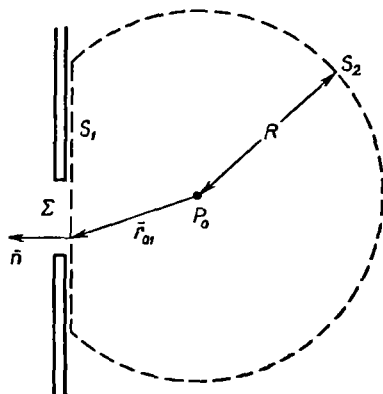
Это соотношение, которое носит название *интегральной теоремы Гельмгольца — Кирхгофа*, играет существенную роль в скалярной теории дифракции, так как позволяет выразить поле в любой точке P_0 через «граничные значения» волны на любой замкнутой поверхности, окружающей эту точку. Как мы увидим ниже, при помощи этого соотношения можно вывести скалярные уравнения дифракции.

§ 3. ФОРМУЛИРОВКА ЗАДАЧИ О ДИФРАКЦИИ НА ПЛОСКОМ ЭКРАНЕ, ПРЕДЛОЖЕННАЯ КИРХГОФОМ

Рассмотрим задачу о дифракции на отверстии в бесконечном непрозрачном экране. Предполагается, что волновое возмущение падает слева на экран с отверстием, как показано на фиг. 9; необходимо рассчитать поле в точке P_0 позади отверстия.

Применение интегральной теоремы

Чтобы найти поле в точке P_0 , воспользуемся интегральной теоремой Гельмгольца и Кирхгофа, выбрав соответствующим образом поверхности интегрирования. Следуя Кирхгофу, возьмем замкнутую поверхность S таким образом, чтобы она состояла из двух частей, как показано на фиг. 9. Пусть плоская поверхность S_1 , лежащая сразу за дифрагирующим экраном, замыкается большим сферическим колпаком S_2 радиусом R с центром в рассматриваемой точке P_0 . Полная замкнутая



Фиг. 9. К формулировке задачи о дифракции на плоском экране, предложенной Кирхгофом.

поверхность S представляет собой просто сумму S_1 и S_2 . Применяя (3.13), получаем

$$U(P_0) = \frac{1}{4\pi} \iint_{S_1+S_2} \left(\frac{\partial U}{\partial n} \mathbf{G} - U \frac{\partial \mathbf{G}}{\partial n} \right) ds,$$

где, как прежде,

$$\mathbf{G} = \frac{\exp(jkr_{01})}{r_{01}}.$$

По мере увеличения R поверхность S_2 принимает форму большой полусферической оболочки.

Функции U и \mathbf{G} уменьшаются пропорционально $1/R$, вследствие чего подынтегральное выражение будет в конечном итоге стремиться к нулю. При этом невольно напрашивается вывод, что интегрирование по поверхности S_2 должно давать нулевой вклад. Однако такой вывод будет ошибочным, поскольку площадь интегрирования возрастает пропорционально R^2 . Заманчивым представляется также предположение, что благодаря конечной скорости распространения возмущения, равной c , величину R в конце концов можно взять настолько большой, что волны не достигнут S_2 и подынтегральное выражение на этой поверхности будет равно нулю. Но эти

соображения несовместимы с нашим предположением о монохроматичности возмущения, которое, по определению, должно существовать постоянно. Очевидно, прежде чем отбросить вклад от S_2 , необходимо провести более тщательное исследование.

Исследуя эту задачу более детально, мы увидим, что на поверхности S_2

$$\mathbf{G} = \frac{\exp(jkR)}{R},$$

а из (3.11) следует, что

$$\frac{\partial \mathbf{G}}{\partial n} = \left(jk - \frac{1}{R} \right) \frac{\exp(jkR)}{R} \approx jk\mathbf{G},$$

где последнее приближение справедливо для больших R . Рассматриваемый интеграл, таким образом, можно свести к следующему виду:

$$\iint_{S_2} \left[\mathbf{G} \frac{\partial \mathbf{U}}{\partial n} - \mathbf{U} (jk\mathbf{G}) \right] ds = \int_{\Omega} \mathbf{G} \left(\frac{\partial \mathbf{U}}{\partial n} - jk\mathbf{U} \right) R^2 d\omega,$$

где Ω — телесный угол с вершиной в точке P_0 , стягиваемый поверхностью S_2 . Величина $|\mathbf{R}\mathbf{G}|$ равномерно ограничена на поверхности S_2 . Поэтому полный интеграл по S_2 будет стремиться к нулю по мере стремления R к бесконечности при условии, что

$$\lim_{R \rightarrow \infty} R \left(\frac{\partial \mathbf{U}}{\partial n} - jk\mathbf{U} \right) = 0 \quad (3.14)$$

равномерно во всем телесном угле. Это требование называется *условием Зоммерфельда для излучения* [13] и удовлетворяется, если возмущение \mathbf{U} стремится к нулю со скоростью, по крайней мере равной той скорости, с которой расходитя сферическая волна (см. задачу 3.1). Так как возмущение, падающее на отверстие, всегда представляет собой сферическую волну или линейную комбинацию сферических волн, можно быть уверенным, что это требование будет практически удовлетворяться и, следовательно, интеграл по S_2 не будет давать вклада в общий интеграл.

Граничные условия Кирхгофа

Устранив интегрирование по поверхности S_2 , теперь можно выразить возмущение в точке P_0 через возмущение и его производную по нормали, взятые на бесконечной плоскости S_1 , расположенной непосредственно за экраном, т. е.

$$U(P_0) = \frac{1}{4\pi} \iint_{S_1} \left(\frac{\partial U}{\partial n} \mathbf{G} - U \frac{\partial \mathbf{G}}{\partial n} \right) ds. \quad (3.15)$$

Экран непрозрачен везде, кроме открытого отверстия, которое мы обозначим через Σ . Поэтому интуитивно понятно, что основной вклад в интеграл (3.15) должны давать точки поверхности S_1 , расположенные в пределах отверстия Σ , где подынтегральная функция, очевидно, будет иметь максимальное значение.

Кирхгоф [14] соответственно принял следующие предположения:

1. На отверстии Σ распределение поля U и его производная $\partial U / \partial n$ имеют точно такие же значения, какие они имели бы в отсутствие экрана.
2. На той части поверхности S_1 , которая лежит в области геометрической тени экрана, распределение поля U и его производная $\partial U / \partial n$ тождественно равны нулю.

Эти условия известны под названием *граничных условий Кирхгофа*. Первое условие позволяет нам определить возмущение, падающее на отверстие, пренебрегая наличием экрана, а второе дает возможность пренебречь интегрированием по всей поверхности, за исключением той ее части, которая лежит непосредственно в пределах самого отверстия. Таким образом, выражение (3.15) принимает вид

$$U(P_0) = \frac{1}{4\pi} \iint_{\Sigma} \left(\frac{\partial U}{\partial n} \mathbf{G} - U \frac{\partial \mathbf{G}}{\partial n} \right) ds. \quad (3.16)$$

Хотя граничные условия Кирхгофа значительно упрощают результат, важно понимать, что ни одно из них не может быть абсолютно справедливым. Присутствие экрана будет неизбежно вызывать некоторое возмуще-

ние поля на отверстии Σ , так как вдоль края отверстия должны выполняться определенные граничные условия, что не требовалось в случае отсутствия экрана. Кроме того, тень за экраном никогда не бывает резкой, так как поле проникает за экран на расстояние нескольких длин волн. Однако если размеры отверстия велики по сравнению с длиной волны, то этими краевыми эффектами заведомо можно пренебречь¹⁾ и использовать оба граничных условия. При этом получаются результаты, которые хорошо согласуются с экспериментом.

Формула дифракции Френеля — Кирхгофа и принцип Гюйгенса — Френеля

Дальнейшее упрощение выражения для $U(P_0)$ можно произвести, если учесть, что расстояние r_{01} от отверстия до точки наблюдения обычно во много раз больше длины волны λ , следовательно, $k \gg 1/r_{01}$; при этом выражение (3.11) запишется следующим образом:

$$\begin{aligned} \frac{\partial G(P_1)}{\partial n} &= \cos(\bar{n}, \bar{r}_{01}) \left(jk - \frac{1}{r_{01}} \right) \frac{\exp(jkr_{01})}{r_{01}} \approx \\ &\approx jk \cos(\bar{n}, \bar{r}_{01}) \frac{\exp(jkr_{01})}{r_{01}}. \end{aligned}$$

Подставляя это приближенное выражение и выражение (3.7) для G в (3.16), находим, что

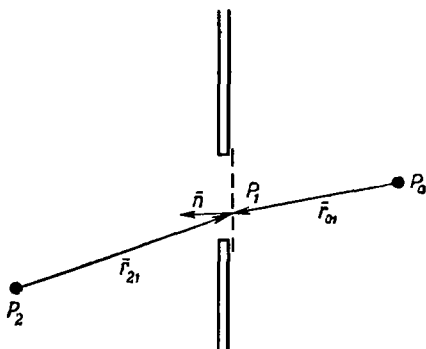
$$U(P_0) = \frac{1}{4\pi} \iint_{\Sigma} \frac{\exp(jkr_{01})}{r_{01}} \left[\frac{\partial U}{\partial n} - jkU \cos(\bar{n}, \bar{r}_{01}) \right] ds. \quad (3.17)$$

Теперь предположим, что на отверстие падает сферическая волна

$$U(P_1) = \frac{A \exp(jkr_{21})}{r_{21}},$$

исходящая из одиночного точечного источника, расположенного в точке P_2 на расстоянии r_{21} от P_1 (фиг. 10). Если расстояние r_{21} во много раз больше длины волны,

¹⁾ Как мы увидим ниже, возражения против применения граничных условий Кирхгофа возникают не вследствие наличия краевых эффектов, а вследствие некоторых внутренних противоречий.



Фиг. 10. Освещение плоского экрана точечным источником.

то (3.17) сразу можно упростить (см. задачу 3.2):

$$U(P_0) = \frac{A}{j\lambda} \iint_{\Sigma} \frac{\exp[jk(r_{21} + r_{01})]}{r_{21}r_{01}} \left[\frac{\cos(\bar{n}, \bar{r}_{01}) - \cos(\bar{n}, \bar{r}_{21})}{2} \right] ds. \quad (3.18)$$

Этот результат, справедливый только для случая, когда отверстие освещается одиночным точечным источником, известен под названием *формулы дифракции Френеля — Кирхгофа*.

Заметим, что выражение (3.18) симметрично по отношению к точечному источнику и точке наблюдения. Таким образом, точечный источник, помещенный в точку P_0 , будет давать в точке P_2 такой же эффект, какой дает в точке P_0 точечный источник равной интенсивности, помещенный в точку P_2 . Этот результат иногда называют *теоремой взаимности Гельмгольца*.

Наконец, рассмотрим одну интересную и полезную интерпретацию формулы дифракции (3.18). Перепишем ее следующим образом:

$$U(P_0) = \iint_{\Sigma} U'(P_1) \frac{\exp(jkr_{01})}{r_{01}} ds, \quad (3.19)$$

где

$$U'(P_1) = \frac{1}{j\lambda} \left[\frac{A \exp(jkr_{21})}{r_{21}} \right] \left[\frac{\cos(\bar{n}, \bar{r}_{01}) - \cos(\bar{n}, \bar{r}_{21})}{2} \right]. \quad (3.20)$$

Исходя из выражения (3.19), мы можем считать, что поле в точке P_0 создается бесконечным множеством вообразимых «вторичных» точечных источников, расположенных в пределах самого отверстия. Амплитуда $U'(P_1)$ вторичного источника, расположенного в точке P_1 , очевидно, пропорциональна амплитуде $[A \exp(jkr_{21})]/r_{21}$ волны, падающей в точку P_1 , но отличается от этой амплитуды в трех отношениях. Во-первых, амплитуда отличается от амплитуды падающей волны множителем λ^{-1} . Во-вторых, эта амплитуда уменьшается также за счет так называемого коэффициента наклона $\frac{1}{2}[\cos(\bar{n}, \bar{r}_{01}) - \cos(\bar{n}, \bar{r}_{21})]$, который никогда не превышает единицы и всегда положителен; в результате каждому вторичному источнику отвечает анизотропная «картина направленности». Наконец, фаза излучения вторичного источника в точке P_1 отличается от фазы падающей волны на 90° .

Эти любопытные свойства вторичных источников представляют интерес в историческом смысле, так как в более ранней работе Френеля было найдено, что путем сочетания построения огибающей Гюйгенса и принципа интерференции Юнга можно точно предсказать дифракционную картину только в том случае, если вторичным источникам приписать свойства, подобные перечисленным. Для получения точных результатов Френель предположил, что вторичные источники обладают именно такими свойствами. Математический вывод Кирхгофа показал, что эти свойства естественно вытекают из волновой природы света.

Заметим, что в приведенном выше выводе так называемого принципа Гюйгенса — Френеля мы ограничились тем случаем, когда отверстие освещается только одной расходящейся сферической волной. Однако вскоре мы вернемся к принципу Гюйгенса — Френеля и покажем, что он носит более общий характер, чем следует из этого первого рассмотрения.

§ 4. ФОРМУЛИРОВКА РЕЛЕЯ — ЗОММЕРФЕЛЬДА ЗАДАЧИ О ДИФРАКЦИИ НА ПЛОСКОМ ЭКРАНЕ

Теория Кирхгофа, как это следует из эксперимента, дает удивительно точные результаты и широко применяется на практике. Однако эта теория содержит

некоторые внутренние противоречия, что заставляет искать более удовлетворительное математическое решение. Затруднения теории Кирхгофа возникают из-за того, что граничные условия налагаются как на напряженность поля, так и на ее производную по нормали. В частности, известная теорема теории потенциала гласит, что если двумерная потенциальная функция и ее производная по нормали обращаются в нуль на любом конечном отрезке кривой, то потенциальная функция *должна обращаться в нуль во всей плоскости*. Точно так же, если решение трехмерного волнового уравнения обращается в нуль на любом конечном элементе поверхности, то оно должно обращаться в нуль во всем пространстве. Таким образом, два граничных условия Кирхгофа, взятые вместе, означают, что повсюду за отверстием поле тождественно равно нулю, в противоречии с известной физической ситуацией. Другое указание на противоречивость теории состоит в том, что, как можно показать, формула дифракции Френеля — Кирхгофа не позволяет получить принятые граничные условия при стремлении точки наблюдения к краю экрана или отверстия. Если принять во внимание эти противоречия, то поистине удивительно, что теория Кирхгофа дает такие точные результаты.

Противоречия теории Кирхгофа были устранены Зоммерфельдом, который исключил необходимость одновременного наложения граничных условий на поле и его производную по нормали. Эта так называемая теория Релея — Зоммерфельда составляет тему настоящего параграфа. Сравнение теорий Кирхгофа и Релея — Зоммерфельда можно найти в работе [15]¹⁾.

Выбор альтернативных функций Грина

Рассмотрим снова выражение (3.15), которое связывает напряженность поля в точке наблюдения с полем падающей волны и его производной по нормали по всему

¹⁾ Тот факт, что одна теория противоречива, а другая нет, не обязательно означает, что вторая *более точна*, чем первая. Сопоставление точности формулировок Кирхгофа и Релея — Зоммерфельда все еще составляет предмет активного исследования.

экрану,

$$U(P_0) = \frac{1}{4\pi} \iint_{S_1} \left(\frac{\partial U}{\partial n} G - U \frac{\partial G}{\partial n} \right) ds. \quad (3.15)$$

Предположим, что функция Грина G теории Кирхгофа видоизменена таким образом, что либо сама функция G , либо ее производная $\partial G/\partial n$ обращается в нуль на всей поверхности S_1 , но вывод вышеприведенного выражения все же остается справедливым. В любом случае отпала бы необходимость наложения граничных условий одновременно на U и $\partial U/\partial n$ и противоречия теории Кирхгофа были бы устранены.

Зоммерфельд показал, что функция Грина с требуемыми свойствами действительно существует. Предположим, что функция G создается не только точечным источником, помещенным в точку P_0 , но также и вторым точечным источником, расположенным в точке \bar{P}_0 , которая представляет собой зеркальное изображение точки P_0 и лежит по другую сторону экрана (фиг. 11). Пусть источник в точке \bar{P}_0 имеет ту же длину волны λ , что и источник в точке P_0 , и предположим, что излучения этих источников сдвинуты по фазе на 180° . Функция Грина в этом случае описывается выражением

$$G_-(P_1) = \frac{\exp(jkr_{01})}{r_{01}} - \frac{\exp(jk\bar{r}_{01})}{\bar{r}_{01}}, \quad (3.21)$$

где \bar{r}_{01} — расстояние между точками \bar{P}_0 и P_1 . Соответствующая производная по нормали от функции G_- равна

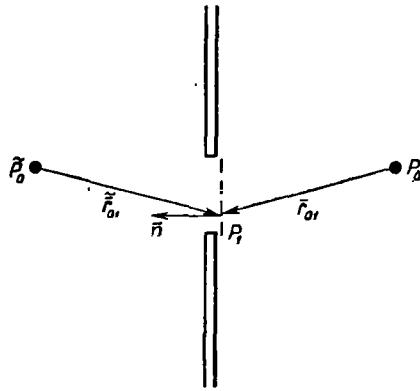
$$\begin{aligned} \frac{\partial G_-}{\partial n} = & \cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}) \left(jk - \frac{1}{r_{01}} \right) \frac{\exp(jkr_{01})}{r_{01}} - \\ & - \cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}) \left(jk - \frac{1}{\bar{r}_{01}} \right) \frac{\exp(jk\bar{r}_{01})}{\bar{r}_{01}}. \end{aligned} \quad (3.22)$$

Теперь для точки P_1 на поверхности S_1 имеем

$$\begin{aligned} r_{01} &= \bar{r}_{01}, \\ \cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}) &= -\cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}), \end{aligned}$$

и, следовательно, на этой поверхности

$$\begin{aligned} G_-(P_1) &= 0, \\ \frac{\partial G_-(P_1)}{\partial n} &= 2 \cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}) \left(jk - \frac{1}{r_{01}} \right) \frac{\exp(jkr_{01})}{r_{01}}. \end{aligned} \quad (3.23)$$



Фиг. 11. К формулировке задачи о дифракции на плоском экране, предложенной Релеем и Зоммерфельдом.

Таким образом, функция G_- обращается в нуль на всей поверхности S_1 .

При условии, что оба точечных источника излучают в фазе, альтернативная (и в той же мере пригодная) функция Грина G_+ определяется выражением

$$G_+(P_1) = \frac{\exp(jkr_{01})}{r_{01}} + \frac{\exp(jk\bar{r}_{01})}{\bar{r}_{01}}. \quad (3.24)$$

Легко показать (см. задачу 3.3), что *производная по нормали* от этой функции обращается в нуль на экране и на отверстии.

Формула дифракции Релея — Зоммерфельда

Подставим функцию Грина G_- вместо G в выражение (3.15). С учетом (3.23) непосредственно получим

$$U(P_0) = \frac{1}{j\lambda} \int \int_{S_1} U(P_1) \frac{\exp(jkr_{01})}{r_{01}} \cos(\bar{n}, \bar{r}_{01}) ds, \quad (3.25)$$

где предполагается, что $r_{01} \gg \lambda$.

Теперь граничные условия Кирхгофа налагаются только на функцию U , что дает весьма общий результат:

$$U(P_0) = \frac{1}{j\lambda} \int \int_{\Sigma} U(P_1) \frac{\exp(jkr_{01})}{r_{01}} \cos(\bar{n}, \bar{r}_{01}) ds, \quad (3.26)$$

Так как для $\partial U/\partial n$ граничные условия учитывать не нужно, противоречия теории Кирхгофа, очевидно, снимаются.

Сравним эти результаты с результатами теории Кирхгофа. Пусть отверстие освещается сферической волной, расходящейся из точечного источника, расположенного в точке R_2 (см. фиг. 10). Тогда

$$U(P_1) = A \frac{\exp(jkr_{21})}{r_{21}}$$

и

$$U(P_0) = \frac{A}{j\lambda} \int_{\Sigma} \int \frac{\exp[jk(r_{21} + r_{01})]}{r_{21}r_{01}} \cos(\bar{n}, \bar{r}_{01}) ds. \quad (3.27)$$

Это последнее выражение известно под названием *формулы дифракции Релея — Зоммерфельда*, и его можно сравнить с соответствующим результатом теории Кирхгофа — выражением (3.18). Заметим, что эти выражения отличаются только значениями коэффициентов наклона.

Рассмотрение соответствующих результатов в случае, когда в качестве функции Грина выбирается функция G_+ , вынесено в задачи. Выражение для $U(P_0)$, полученное с этой альтернативной функцией Грина, разумеется, столь же справедливо, как и найденное выше.

Дальнейшее обсуждение принципа Гюйгенса — Френеля

В предшествующем выводе принципа Гюйгенса — Френеля предполагалось, что отверстие освещено одной расходящейся сферической волной. Из выражения (3.26) следует, что в действительности принцип справедлив для освещения произвольного вида. Такое обобщение этого частного случая действительно возможно, так как произвольный источник можно всегда представить в виде ряда (возможно, бесконечного) точечных источников, а в силу линейности волнового уравнения принцип Гюйгенса — Френеля можно применить к каждому точечному источнику отдельно. Выражение (3.26) для принципа Гюйгенса — Френеля представляет собой не что иное, как *интеграл суперпозиции*, который рассматривался в

гл. 2. Исходя из этого, можно переписать (3.26) в виде

$$U(P_0) = \int_{\Sigma} h(P_0, P_1) U(P_1) ds, \quad (3.28)$$

где весовая функция $h(P_0, P_1)$ определяется выражением

$$h(P_0, P_1) = \frac{1}{j\lambda} \frac{\exp(jkr_{01})}{r_{01}} \cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}). \quad (3.29)$$

Появление интеграла суперпозиции в результате проведенного анализа явления дифракции не может вызвать удивления. Как мы уже видели, основным требованием при получении такого результата является *линейность* — свойство, которое предполагалось с самого начала нашего анализа.

§ 5. ОБОБЩЕНИЕ ТЕОРИИ РЕЛЕЯ — ЗОММЕРФЕЛЬДА НА СЛУЧАЙ НЕМОНОХРОМАТИЧЕСКИХ ВОЛН

Во всех предыдущих случаях волновое возмущение предполагалось идеально монохроматическим. Волны, достаточно близкие к монохроматическим, нетрудно получать практически, и их особенно легко анализировать. Однако теперь мы кратко обсудим более общий случай немонохроматического возмущения. При этом ограничимся рассмотрением выводов теории Релея — Зоммерфельда, хотя подобные результаты могут быть получены и из анализа Кирхгофа.

Рассмотрим скалярное возмущение $u(P_0, t)$, наблюдаемое позади отверстия Σ в непрозрачном экране, когда на отверстие падает возмущение $u(P_1, t)$. Функции времени $u(P_0, t)$ и $u(P_1, t)$ могут быть представлены с помощью обратных преобразований Фурье:

$$u(P_1, t) = \int_{-\infty}^{+\infty} U(P_1, \nu) \exp(j2\pi\nu t) d\nu, \quad (3.30a)$$

$$u(P_0, t) = \int_{-\infty}^{+\infty} U(P_0, \nu) \exp(j2\pi\nu t) d\nu, \quad (3.30б)$$

где $U(P_0, \nu)$ и $U(P_1, \nu)$ — просто фурье-спектры функций $u(P_0, t)$ и $u(P_1, t)$ соответственно, а ν — оптическая частота.

Преобразуем выражения (3.30а) и (3.30б), произведя замену переменных $\nu' = -\nu$. В результате получим

$$u(P_1, t) = \int_{-\infty}^{+\infty} U(P_1, -\nu') \exp(-j2\pi\nu't) d\nu', \quad (3.31а)$$

$$u(P_0, t) = \int_{-\infty}^{+\infty} U(P_0, -\nu') \exp(-j2\pi\nu't) d\nu'. \quad (3.31б)$$

Теперь эти выражения можно рассматривать как представление немонахроматических функций времени $u(P_1, t)$ и $u(P_0, t)$ в виде линейной комбинации монотонахроматических функций времени вида (3.2). Монахроматические элементарные функции зависят от частоты ν' , причем комплексные амплитуды возмущения на частоте ν' равны $U(P_1, -\nu')$ и $U(P_0, -\nu')$ ¹⁾.

Используя линейность явления распространения волн и применяя результаты § 4, мы можем найти комплексную амплитуду каждой монотонахроматической составляющей возмущения в точке P_0 и, суммируя их, получить в результате общую функцию времени $u(P_0, t)$.

Итак, с помощью (3.26) можно сразу написать

$$U(P_0, -\nu') = -j \frac{\nu'}{c} \iint_{\Sigma} U(P_1, -\nu') \frac{\exp\left(j2\pi\nu' \frac{r_{01}}{c}\right)}{r_{01}} \cos(\bar{n}, \bar{r}_{01}) ds, \quad (3.32)$$

где $\lambda'\nu' = c$. Подставляя (3.32) в (3.31б) и меняя порядок интегрирования, получаем

$$u(P_0, t) = \iint_{\Sigma} \frac{\cos(\bar{n}, \bar{r}_{01})}{2\pi c r_{01}} \int_{-\infty}^{+\infty} -j2\pi\nu' U(P_1, -\nu') \times \\ \times \exp\left[-j2\pi\nu' \left(t - \frac{r_{01}}{c}\right)\right] d\nu' ds.$$

¹⁾ Точнее говоря, $U(P_1, -\nu')$ не амплитуда возмущения на частоте ν' , а плотность распределения амплитуды. — Прим. ред.

Наконец, используя равенство

$$\begin{aligned} \frac{d}{dt} u(P_1, t) &= \frac{d}{dt} \int_{-\infty}^{+\infty} U(P_1, -v') \exp(-j2\pi v' t) dv' = \\ &= \int_{-\infty}^{+\infty} -j2\pi v' U(P_1, -v') \exp(-j2\pi v' t) dv', \end{aligned}$$

можно записать

$$u(P_0, t) = \iint_{\Sigma} \frac{\cos(\bar{n}, \bar{r}_{01})}{2\pi c r_{01}} \frac{d}{dt} u\left(P_1, t - \frac{r_{01}}{c}\right) ds. \quad (3.33)$$

Таким образом, мы видим, что волновое возмущение в точке P_0 пропорционально производной по времени от возмущения в каждой точке P_1 отверстия. Так как для распространения возмущения от точки P_1 к точке P_0 требуется время r_{01}/c , наблюдаемая волна зависит от производной падающей волны в «запаздывающие» моменты времени $t - (r_{01}/c)$.

Это более общее рассмотрение показывает, что изучение дифракции идеально монохроматических волн имеет отнюдь не чисто академический интерес, так как результаты для более общего случая временных возмущений легко можно синтезировать из результатов, полученных для монохроматических волн. К этому следует добавить, что оптические источники, которые для многих целей можно считать идеально монохроматическими, легко доступны (например, гелий-неоновый газовый лазер). Дополнительно немонахроматические волны рассматриваются в задаче 3.6.

§ 6. ДИФРАКЦИЯ НА ГРАНИЦАХ

При установлении принципа Гюйгенса — Френеля нам было удобно считать каждую точку отверстия новым источником сферических волн. Было отмечено, что такие источники введены только для удобства математического описания и не имеют реального физического смысла. Более физический подход, впервые качественно

сформулированный Томасом Юнгом в 1802 г., заключается в следующем: наблюдаемое поле рассматривается как суперпозиция падающей волны, прошедшей через отверстие без искажения, и дифрагированной волны, источником которой служит *край* отверстия. В силу того обстоятельства, что источником новой волны служит материальная среда — край отверстия, этот подход является более физичным.

Качественные аргументы Юнга подкрепляются строгим решением задачи о дифракции на полубесконечном идеальном проводящем отражающем экране, которое получил Зоммерфельд [7]. Это строгое решение показывает, что поле в области геометрической тени экрана имеет форму цилиндрической волны, источником которой служит край экрана, тогда как в освещенной области за экраном поле представляет суперпозицию этой цилиндрической волны и прошедшей без искажения падающей волны.

Применимость такой точки зрения к решению более общих задач дифракции была исследована Магги [16] и Рубиновичем [17], которые показали, что формулу дифракции Кирхгофа действительно можно преобразовать к виду, отвечающему идеям Юнга. Позднее Миямото и Вольф [18] расширили теорию дифракции на границе. Подробное обсуждение этих идей читатель найдет в вышеупомянутых работах.

Геометрическая теория дифракции, развитая Келлером [19], представляет собой иной подход, который также тесно связан с идеями Юнга. В этом методе для нахождения поля позади препятствия используются принципы геометрической оптики, видоизмененные путем введения представления о «дифрагированных лучах», исходящих из некоторых точек самого препятствия. Предполагается, что новые лучи образуются на краях, углах, остриях и поверхностях препятствия. Эта теория часто применяется для расчета дифракции волн на препятствиях, которые имеют слишком сложную форму, чтобы дифракцию на них можно было рассчитать другими методами. В силу сказанного такая теория приобретает все большую популярность.

§ 7. УГЛОВОЙ СПЕКТР ПЛОСКИХ ВОЛН

Скалярную теорию дифракции можно сформулировать таким образом, чтобы она была похожа на теорию линейных инвариантных во времени фильтров. В физической оптике эта формулировка применяется не так широко, как теория Кирхгофа, тем не менее для читателей, знакомых с теорией линейных систем, она должна быть интуитивно ясной, поэтому имеет смысл изложить ее в этой книге. Как мы увидим далее, при фурье-анализе распределения комплексного поля в любой плоскости различные пространственные составляющие фурье-образа можно отождествить с плоскими волнами, распространяющимися в различных направлениях. Амплитуду поля в любой точке можно вычислить путем сложения амплитуд этих плоских волн, учитывая фазовые сдвиги, которые они приобретают, приходя в рассматриваемую точку. Подробное рассмотрение этого приближения теории дифракции, а также его применение в теории распространения радиоволн читатель может найти в работе Рэтклиффа [20].

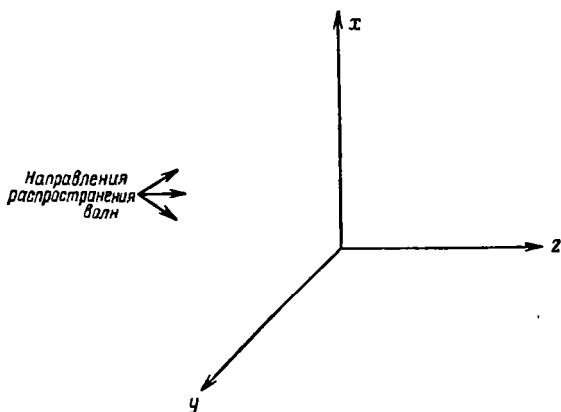
Угловой спектр и его физический смысл

Предположим, что волна, созданная некоторой произвольной системой монохроматических источников и распространяющаяся в положительном направлении оси z , достигает плоскости xy , как показано на фиг. 12. Пусть комплексное поле в этой плоскости описывается функцией $U(x, y, 0)$; наша конечная цель — рассчитать результирующее поле $U(x, y, z)$ в точке P_0 с координатами (x, y, z) .

В плоскости xy двумерный фурье-образ функции U имеет вид

$$A_0(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x, y, 0) \times \\ \times \exp[-j2\pi(f_x x + f_y y)] dx dy. \quad (3.34)$$

Как указывалось в гл. 2, операцию преобразования Фурье можно рассматривать как представление сложной



Фиг. 12. Система координат для расчета углового спектра.

функции в виде совокупности более простых комплексных экспоненциальных функций. Чтобы подчеркнуть это обстоятельство, запишем функцию U в виде обратного фурье-образа ее спектра

$$U(x, y, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_0(f_x, f_y) \exp[j2\pi(f_x x + f_y y)] df_x df_y. \quad (3.35)$$

Теперь напомним, что уравнение для плоской волны единичной амплитуды с направляющими косинусами нормали к фронту волны (α , β , γ) имеет вид

$$V(x, y, z) = \exp\left[j \frac{2\pi}{\lambda} (\alpha x + \beta y + \gamma z)\right],$$

где

$$\gamma = \sqrt{1 - \alpha^2 - \beta^2}.$$

Таким образом, в плоскости $z = 0$ комплексную экспоненциальную функцию $\exp[j2\pi(f_x x + f_y y)]$ можно рассматривать как плоскую волну с направляющими косинусами

$$\alpha = \lambda f_x, \quad \beta = \lambda f_y, \quad \gamma = \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}. \quad (3.36)$$

Комплексная амплитуда этой плоской волны равна $A_0(f_x, f_y) df_x df_y$, где $(f_x = \alpha/\lambda, f_y = \beta/\lambda)$. По этой причине функцию

$$A_0\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x, y, 0) \exp\left[-j2\pi\left(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y\right)\right] dx dy. \quad (3.37)$$

называют *угловым спектром* возмущения $U(x, y, 0)$.

Распространение углового спектра

Теперь рассмотрим угловой спектр возмущения U в плоскости, параллельной плоскости xy , но расположенной на расстоянии z от нее. Пусть функция $A(\alpha/\lambda, \beta/\lambda; z)$ представляет угловой спектр функции $U(x, y, z)$, т. е.

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x, y, z) \times \\ \times \exp\left[-j2\pi\left(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y\right)\right] dx dy. \quad (3.38)$$

Если связь между $A_0(\alpha/\lambda, \beta/\lambda)$ и $A(\alpha/\lambda, \beta/\lambda; z)$ известна, то можно определить влияние распространения волны на ее угловой спектр.

Чтобы найти нужное соотношение, отметим, что возмущение U можно записать в виде

$$U(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z\right) \times \\ \times \exp\left[j2\pi\left(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y\right)\right] d\frac{\alpha}{\lambda} d\frac{\beta}{\lambda}. \quad (3.39)$$

Кроме того, во всех точках, где нет источника, функция U должна удовлетворять уравнению Гельмгольца

$$\nabla^2 U + k^2 U = 0.$$

Непосредственно применяя это условие к выражению (3.39), находим, что функция A должна удовлетворять

дифференциальному уравнению

$$\frac{d^2}{dz^2} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z\right) + \left(\frac{2\pi}{\lambda}\right)^2 [1 - \alpha^2 - \beta^2] A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z\right) = 0.$$

Частное решение этого уравнения можно записать в виде

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z\right) = A_0\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) \exp\left(j \frac{2\pi}{\lambda} \sqrt{1 - \alpha^2 - \beta^2} z\right), \quad (3.40)$$

откуда следует, что в том случае, когда направляющие косинусы (α, β) удовлетворяют условию

$$\alpha^2 + \beta^2 < 1,$$

распространение волны на расстояние z проявляется только в изменении относительных фаз различных составляющих углового спектра. Фазовые сдвиги возникают вследствие того, что плоские волны, распространяясь под разными углами, проходят разные расстояния, пока достигнут рассматриваемой точки.

Однако если направляющие косинусы (α, β) удовлетворяют условию

$$\alpha^2 + \beta^2 > 1,$$

то требуется другое истолкование. В этом случае квадратный корень в выражении (3.40) будет мнимым, и это выражение можно записать в виде

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}; z\right) = A_0\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) \exp(-\mu z), \quad (3.41)$$

где

$$\mu = \frac{2\pi}{\lambda} \sqrt{\alpha^2 + \beta^2 - 1}.$$

Так как μ — положительное действительное число, то эти волновые составляющие сильно ослабляются при распространении. Такие составляющие углового спектра называются затухающими волнами; они совершенно аналогичны волнам, возникающим в СВЧ волноводе ниже его граничной частоты. Заметим, что предельный случай

$$\alpha^2 + \beta^2 = 1$$

соответствует плоским волнам, которые распространяются в направлениях, *перпендикулярных* к оси z , и не переносят энергию в направлении z ¹⁾.

Наконец, следует упомянуть, что возмущение в точке (x, y, z) можно выразить через угловой спектр путем обратного преобразования выражения (3.40)

$$U(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_0\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) \exp\left(j \frac{2\pi}{\lambda} \sqrt{1 - \alpha^2 - \beta^2} z\right) \times \\ \times \exp\left[j2\pi\left(\frac{\alpha}{\lambda} x + \frac{\beta}{\lambda} y\right)\right] d\frac{\alpha}{\lambda} d\frac{\beta}{\lambda}. \quad (3.42)$$

Влияние ограничивающего отверстия на угловой спектр возмущения

Предположим, что бесконечный непрозрачный экран с отверстием Σ помещен в плоскость $z = 0$ (см. фиг. 12). Рассмотрим влияние этого экрана на угловой спектр возмущения. Пусть угловой спектр возмущения, которое достигает экрана, есть $A_i(\alpha/\lambda, \beta/\lambda)$, а *амплитудный коэффициент пропускания* экрана определим соотношением

$$t(x, y) = \begin{cases} 1 & (x, y) \text{ в } \Sigma, \\ 0 & \text{в остальных точках.} \end{cases}$$

Применим граничные условия Кирхгофа к возмущению U , предполагая, что экран не возмущает волны, падающие на отверстие Σ , и что поле в геометрической тени экрана тождественно равно нулю. Таким образом, поле в плоскости, расположенной непосредственно за экраном, можно записать в виде

$$U_t(x, y, 0) = U_i(x, y, 0)t(x, y),$$

где $U_i(x, y, 0)$ представляет собой поле падающей на экран волны. Применяя теорему свертки анализа Фурье, угловой спектр $A_t(\alpha/\lambda, \beta/\lambda)$ пропущенных волн можно

¹⁾ Заметим, что затухающие волны возникают только в тех случаях, когда применение скалярной теории не вполне законно. Поэтому более правильно рассматривать их в векторном приближении.

записать в виде

$$A_t\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = A_t\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) * T\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right), \quad (3.43)$$

где

$$T\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t(x, y) \exp\left[-j2\pi\left(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y\right)\right] dx dy. \quad (3.44)$$

Мы видим, что угловой спектр прошедшего возмущения определяется сверткой спектра падающего возмущения и углового спектра, характеризующего само отверстие.

Для частного случая одной плоской волны единичной амплитуды, падающей нормально на отверстие, результат имеет особенно простую форму. В этом случае

$$A_t\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = \delta\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right)$$

и

$$A_t\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = \delta\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) * T\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right) = T\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}\right).$$

Таким образом, пропущенный угловой спектр можно найти просто преобразованием Фурье функции пропускания отверстия.

Подводя итог, можно сказать, что введение отверстия, которое пространственно ограничивает падающую волну, приводит к уширению углового спектра возмущения. Если отверстие освещается плоской волной, то угловой спектр пропущенного света можно найти непосредственно путем преобразования Фурье функции пропускания отверстия. Чем меньше отверстие, тем шире угловой спектр за отверстием. Аналогичный эффект имеет место, конечно, и в случае импульсных электрических сигналов; ограничение длительности функции во времени приводит к уширению ее частотного спектра.

Линейный пространственный фильтр, эквивалентный распространению волн

Рассмотрим снова распространение света от плоскости $z = 0$ (см. фиг. 12) к параллельной плоскости, расположенной на расстоянии z . Можно считать, что

возмущение $U(x, y, 0)$, падающее на первую плоскость, отображается благодаря процессу распространения в новое распределение поля $U(x, y, z)$. Такое отображение соответствует нашему предыдущему определению системы. Действительно, мы покажем, что явление распространения обуславливает такой же эффект, как линейная пространственно-инвариантная система, и что оно характеризуется относительно простой передаточной функцией.

Линейность явления распространения уже обсуждалась выше; она непосредственно следует из линейности волнового уравнения, или, иначе, из интеграла суперпозиции (3.28). Свойство пространственной инвариантности легче всего продемонстрировать, вычислив выражение для передаточной функции, которая описывает эффекты распространения. Если отображение имеет передаточную функцию, то оно должно быть пространственно-инвариантно.

Чтобы найти передаточную функцию, вернемся к понятию углового спектра. Однако вместо того, чтобы представлять угловые спектры в виде функций направляющих косинусов (α, β), сначала возьмем спектры в виде функций пространственных частот (f_x, f_y). Пространственные частоты и направляющие косинусы непосредственно связаны соотношением (3.36).

Пусть функция $A(f_x, f_y, z)$ снова представляет пространственный спектр функции $U(x, y, z)$, а функция $A_0(f_x, f_y)$ — спектр функции $U(x, y, 0)$. Таким образом, $U(x, y, z)$ можно выразить как

$$U(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(f_x, f_y, z) \exp [j2\pi (f_x x + f_y y)] df_x df_y.$$

Но, кроме того, из (3.42) имеем

$$U(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_0(f_x, f_y) \times \\ \times \exp \left[j2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \right] \times \\ \times \exp [j2\pi (f_x x + f_y y)] df_x df_y.$$

Из сравнения этих двух соотношений находим

$$\begin{aligned} A(f_X, f_Y; z) &= \\ &= A_0(f_X, f_Y) \exp\left[j2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_X)^2 - (\lambda f_Y)^2}\right]. \end{aligned} \quad (3.45)$$

Таким образом, явление распространения характеризуется передаточной функцией H , имеющей вид

$$\begin{aligned} H(f_X, f_Y) &= \frac{A(f_X, f_Y; z)}{A_0(f_X, f_Y)} = \\ &= \exp\left[j2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_X)^2 - (\lambda f_Y)^2}\right]. \end{aligned} \quad (3.46)$$

Если расстояние z равно по меньшей мере нескольким длинам волн, то затухающими волнами можно пренебречь. В результате получим передаточную функцию

$$\begin{aligned} H(f_X, f_Y) &= \\ &= \begin{cases} \exp\left[j2\pi \frac{z}{\lambda} \sqrt{1 - (\lambda f_X)^2 - (\lambda f_Y)^2}\right] & f_X^2 + f_Y^2 < \frac{1}{\lambda^2}, \\ 0 & \text{в остальных} \\ & \text{случаях.} \end{cases} \end{aligned} \quad (3.47)$$

В заключение отметим, что явление распространения можно рассматривать как действие линейного диспергирующего пространственного фильтра с конечной полосой пропускания пространственных частот. Пропускание фильтра равно нулю вне круга радиусом λ^{-1} в плоскости частот. Внутри этой круглой области пропускания модуль передаточной функции равен единице, но при этом появляются зависящие от частот фазовые сдвиги. Фазовая дисперсия системы наиболее значительна на высоких пространственных частотах и обращается в нуль при стремлении f_X и f_Y к нулю.

ЗАДАЧИ

3.1. Показать, что расходящаяся сферическая волна удовлетворяет условию Зоммерфельда для излучения.

3.2. Показать, что соотношение (3.17) преобразуется в соотношение (3.18), если отверстие освещается расходящейся сферической волной.

3.3. Рассмотреть применение функции Грина

$$G_+(P_1) = \frac{\exp(jkr_{01})}{r_{01}} + \frac{\exp(jk\bar{r}_{01})}{\bar{r}_{01}}$$

в теории Релея — Зоммерфельда.

а) Показать, что производная по нормали от функции G_+ обращается в нуль в плоскости отверстия.

б) Используя эту функцию Грина, выразить $U(P_0)$ через произвольное возмущение в плоскости отверстия. Какие граничные условия нужно при этом учесть?

в) Используя результат, полученный в п. «б», найти выражение для $U(P_0)$ в том случае, когда отверстие освещается сферической волной, расходящейся из точки P_2 .

3.4. Предполагая, что освещение представляет собой нормально падающую плоскую волну единичной амплитуды, найти угловой спектр в случае

а) круглого отверстия диаметром d ,

б) круглого непрозрачного диска диаметром d .

3.5. Рассмотрим немонахроматическое возмущение $u(P, t)$ с центральной частотой $\bar{\nu}$ и шириной спектра $\Delta\nu$.

Пусть комплексное возмущение $u_-(P, t)$ содержит только составляющие возмущения $u(P, t)$ с отрицательными частотами. Таким образом,

$$u_-(P, t) = \int_{-\infty}^0 U(P, \nu) \exp(j2\pi\nu t) d\nu,$$

где $U(P, \nu)$ — спектр Фурье функции $u(P, t)$.

Используя фиг. 9, показать, что если

$$\frac{\Delta\nu}{\bar{\nu}} \ll 1 \quad \text{и} \quad \frac{1}{\Delta\nu} \gg \frac{r_{01}}{c},$$

то

$$u_-(P_0, t) = \frac{1}{j\lambda} \iint_{\Sigma} u_-(P_1, t) \frac{\exp(jkr_{01})}{r_{01}} \cos(\bar{n}, \bar{r}_{01}) ds,$$

где $\lambda = c/\bar{\nu}$, а $k = 2\pi/\lambda$.

ЛИТЕРАТУРА

1. Sommerfeld A., Vorlesungen über theoretische Physik, Bd. IV, Optik, Wiesbaden, 1950. (Имеется перевод: А. Зоммерфельд, Оптика, ИЛ, 1953.)
2. Kottler F., в книге Progress in Optics, ed. E. Wolf, vol. 4, Amsterdam, 1965.
Дифракция на черном экране.

3. Silver S., Journ. Opt. Soc. Am., 52, 131 (1962).
Микроволновые щелевые антенны и теория дифракции.
4. Born M., Wolf E., Principles of Optics, 2nd rev. ed., London — New York, 1964.
5. Stroke G. W., Proc. Third Intern. Conf. Quantum Electron., vol. 2, 1964.
Теория, получение и использование оптических решеток в спектроскопии высокого разрешения.
6. Kottler F., Ann. Phys. (4), 71, 457 (1923).
Электромагнитная теория дифракции на черных экранах.
7. Sommerfeld A., Math. Ann., 47, 317 (1896).
Математическая теория дифракции.
8. Kottler F., Ann. Phys. (4), 70, 405 (1923).
К теории дифракции на черных экранах.
9. Baker H. B., Copson E. T., The Mathematical Theory of Huygens' Principle, 2nd. ed., Oxford, 1949.
10. Вонкманн С. J., в книге Rep. Progr. Phys., ed. A. C. Strickland, vol. 17, London, 1954.
11. Hoenl H., Maue A. W., Westpfahl K., в книге Handbuch der Physik, Hsg. S. Fluegge, Bd. 25, Berlin, 1961. (Имеется перевод: X. Хёнл, А. Мауэ, К. Вестпфаль, Теория дифракции, изд-во «Мир», 1964.)
12. Hildebrand F. B., Advanced Calculus for Engineers, New York, 1948.
13. Sommerfeld A., Jahresber. Deut. Math. Ver., 21, 309 (1912).
Функция Грина волнового уравнения.
14. Kirchhoff G., Wiedemann Ann. (2), 18, 663 (1883).
К теории световых лучей.
15. Wolf E., Marchand E. W., Journ. Opt. Soc. Am., 54, 587 (1964).
Сравнение теорий дифракции на отверстиях, предложенных Кирхгофом и Релеем — Зоммерфельдом.
16. Maggi G. A., Ann. Matematica, 16, 21 (1888).
О распространении света через изотропную среду при наличии и в отсутствие препятствий.
17. Rubinowicz A., Ann. Phys. (14), 53, 257 (1917).
Дифракционные волны в теории дифракционных явлений Кирхгофа.
18. Rubinowicz A., в книге Progress in Optics, ed. E. Wolf, vol. 4, Amsterdam, 1965.
Дифракционная волна Миямото — Вольфа.
19. Keller J. B., Journ. Opt. Soc. Am., 52, 116 (1962).
Геометрическая теория дифракции.
20. Ratcliffe J. A., в книге Rep. Progr. Phys., ed. C. Strickland, vol. 19, London, 1956.
Некоторые аспекты теории дифракции и их применение к ионосфере.
- 21*. Дитчберн Р., Физическая оптика, М., 1965, гл. 6.
- 22*. Горелик Г. С., Колебания и волны, М., 1950.
- 23*. Ландау Л. Д., Лифшиц Е. М., Теория поля, М., 1962, гл. 7.
- 24*. Ландау Л. Д., Лифшиц Е. М., Электродинамика сплошных сред, М., 1957, гл. 9.

ДИФРАКЦИЯ ФРЕНЕЛЯ И ФРАУНГОФЕРА

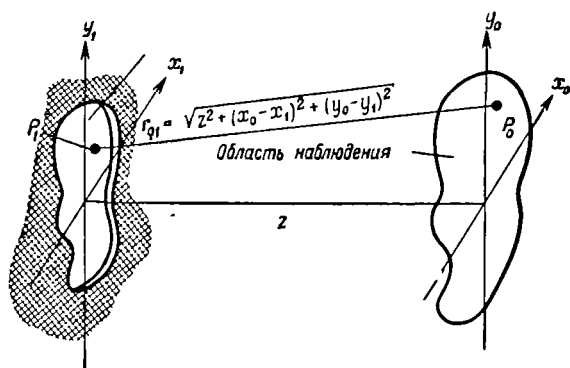
В предыдущей главе выводы скалярной теории дифракции были представлены в наиболее общем виде. Обратимся теперь к рассмотрению ряда приближений к общей теории, которые позволяют свести расчет дифракционной картины к более простым математическим действиям. Эти приближения, применяемые обычно в большинстве задач, связанных с распространением волн, мы будем в дальнейшем называть *приближениями Френеля и Фраунгофера*. Наше рассмотрение будет отличаться от обычных трактовок дифракции Френеля и Фраунгофера тем, что мы в соответствии с принятым нами представлением явления распространения как некоторой «системы» попытаемся найти приближения, справедливые для широкого класса «входных» распределений поля.

§ 1. ПРИБЛИЖЕНИЕ, ДАВАЕМОЕ ПРИНЦИПОМ ГЮЙГЕНСА — ФРЕНЕЛЯ

Снова обсудим дифракцию монохроматического света на бесконечном непрозрачном экране с отверстием Σ конечных размеров. Рассматривается плоский экран, с которым жестко связана прямоугольная система координат (x_1, y_1) (фиг. 13). Кроме того, предполагается, что область наблюдения представляет собой плоскость, параллельную плоскости экрана и расположенную на расстоянии z по нормали от него. С плоскостью наблюдения связана система координат (x_0, y_0) , оси которой параллельны осям плоскости (x_1, y_1) .

Исходные приближения

Используя математическое выражение принципа Гюйгенса — Френеля, в частности (3.28) и (3.29), не-



Фиг. 13. Геометрическая схема дифракционной задачи.

трудно найти амплитуду поля в точке (x_0, y_0)

$$U(x_0, y_0) = \int_{\Sigma} \int h(x_0, y_0; x_1, y_1) U(x_1, y_1) dx_1 dy_1, \quad (4.1)$$

где

$$h(x_0, y_0; x_1, y_1) = \frac{1}{j\lambda} \frac{\exp(jkr_{01})}{r_{01}} \cos(\bar{n}, \bar{r}_{01}). \quad (4.2)$$

В дальнейшем такой интеграл суперпозиции мы будем записывать с бесконечными пределами; при этом предполагается, что в соответствии с граничными условиями Кирхгофа функция $U(x_1, y_1)$ за пределами отверстия Σ тождественно равна нулю. Таким образом,

$$U(x_0, y_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x_0, y_0; x_1, y_1) U(x_1, y_1) dx_1 dy_1. \quad (4.3)$$

Наши приближения будут основываться на предположении, что расстояние z между отверстием и плоскостью наблюдения значительно превышает максимальный линейный размер отверстия Σ . Кроме того, будем предполагать, что в плоскости наблюдения рассматривается только конечная область вблизи оси z и что расстояние z много больше максимального размера этой области. С учетом этих допущений можно считать, что

с точностью не хуже 5%

$$\cos(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01}) \approx 1,$$

если угол $(\bar{\mathbf{n}}, \bar{\mathbf{r}}_{01})$ не превышает 18° . При этом величина r_{01} в знаменателе выражения (4.2) будет лишь незначительно отличаться от z . Тогда для функции \mathbf{h} справедливо следующее выражение:

$$\mathbf{h}(x_0, y_0; x_1, y_1) \approx \frac{1}{j\lambda z} \exp(jkr_{01}). \quad (4.4)$$

Заметим, что величину r_{01} в экспоненте нельзя просто заменить на z , так как возникающая при этом погрешность умножается на очень большое число k и фазовые погрешности становятся много больше 2π рад.

Приближения Френеля

Дальнейшее упрощение удается получить, принимая некоторые приближения для величины r_{01} в (4.4). Из фиг. 13 видно, что точная формула для расстояния имеет вид

$$\begin{aligned} r_{01} &= \sqrt{z^2 + (x_0 - x_1)^2 + (y_0 - y_1)^2} = \\ &= z \sqrt{1 + \left(\frac{x_0 - x_1}{z}\right)^2 + \left(\frac{y_0 - y_1}{z}\right)^2}. \end{aligned} \quad (4.5)$$

Для аппроксимации удобно воспользоваться разложением квадратного корня

$$\sqrt{1+b} = 1 + \frac{1}{2}b - \frac{1}{8}b^2 + \dots, \quad |b| < 1. \quad (4.6)$$

Предположим, что квадратный корень достаточно хорошо аппроксимируется первыми двумя членами этого разложения. Тогда

$$\begin{aligned} z \sqrt{1 + \left(\frac{x_0 - x_1}{z}\right)^2 + \left(\frac{y_0 - y_1}{z}\right)^2} &\approx \\ &\approx z \left[1 + \frac{1}{2} \left(\frac{x_0 - x_1}{z}\right)^2 + \frac{1}{2} \left(\frac{y_0 - y_1}{z}\right)^2 \right]. \end{aligned}$$

С учетом этого приближения, которое мы будем называть приближением Френеля, можно переписать функ-

цию h в виде

$$h(x_0, y_0; x_1, y_1) = \frac{\exp(jkz)}{j\lambda z} \exp\left\{j \frac{k}{2z} [(x_0 - x_1)^2 + (y_0 - y_1)^2]\right\}. \quad (4.7)$$

Когда расстояние z достаточно велико для того, чтобы это выражение можно было считать точным, говорят, что наблюдатель находится в области *дифракции Френеля*.

Заметим, что слабым местом нашего приближения была замена сферических вторичных волн Гюйгенса поверхностями *второго порядка*. Ясно, что такое приближение налагает определенные ограничения на относительные размеры отверстия, размеры области наблюдения и расстояние z . В качестве достаточного условия, обеспечивающего точность, можно потребовать, чтобы максимальное изменение фазы, вносимое членом более высокого порядка в разложении корня, было много меньше 1 рад. Это условие будет выполнено, если z удовлетворяет соотношению

$$z^3 \gg \frac{\pi}{4\lambda} [(x_0 - x_1)^2 + (y_0 - y_1)^2]_{\max}^2. \quad (4.8)$$

Однако это требование обычно не является необходимым. Для справедливости приближения Френеля требуется только, чтобы члены более высокого порядка в разложении не изменяли значения интеграла суперпозиции (4.3). Для этого необязательно, чтобы максимальные значения добавляемых фазовых множителей были много меньше 1 рад. В том случае, когда расстояние z мало и условие (4.8) не удовлетворяется, величина $k/2z$, как правило, будет такой большой, а колебания квадратичного фазового множителя в (4.7) такими быстрыми, что основной вклад в интеграл будут давать только точки, которые расположены рядом с точкой $(x_1 = x_0, y_1 = y_0)$, где скорость изменения фазы минимальна. В окрестности таких точек «стационарной фазы» величиной фазового члена следующего, более высокого порядка часто можно пренебречь. Подробное рассмот-

рение так называемого *принципа стационарной фазы* приведено в работе [3.4]¹⁾.

Считая приближение Френеля справедливым, интеграл суперпозиции можно представить в любой из двух эквивалентных форм. Во-первых, $U(x_0, y_0)$ можно рассматривать как *свертку* функций $U(x_1, y_1)$ и h , т. е.

$$U(x_0, y_0) = \frac{\exp(jkz)}{j\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x_1, y_1) \times \\ \times \exp\left\{j \frac{k}{2z} [(x_0 - x_1)^2 + (y_0 - y_1)^2]\right\} dx_1 dy_1. \quad (4.9)$$

Во-вторых, после разложения квадратичных членов в экспоненте получим

$$U(x_0, y_0) = \frac{\exp(jkz)}{j\lambda z} \exp\left[j \frac{k}{2z} (x_0^2 + y_0^2)\right] \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ U(x_1, y_1) \exp\left[j \frac{k}{2z} (x_1^2 + y_1^2)\right] \right\} \times \\ \times \exp\left[-j \frac{2\pi}{\lambda z} (x_0 x_1 + y_0 y_1)\right] dx_1 dy_1. \quad (4.10)$$

Таким образом, с точностью до амплитудного и фазового множителей, которые не зависят от (x_1, y_1) , функцию $U(x_0, y_0)$ можно найти как *фурье-образ* функции $U(x_1, y_1) \exp\left[j \frac{k}{2z} (x_1^2 + y_1^2)\right]$ для частот $f_x = x_0/\lambda z$, $f_y = y_0/\lambda z$, что обеспечивает правильный масштаб в плоскости наблюдения.

Возвратимся к первому виду записи интеграла суперпозиции (4.9). Представление его в виде свертки сразу наводит на мысль, что исследование дифракции Френеля в плоскости пространственных частот позволит нам лучше понять суть дела. Соответственно мы можем подвергнуть прямому преобразованию Фурье пространственно-инвариантную весовую функцию (4.7).

¹⁾ Мы используем две системы ссылок на литературу. Работы, входящие в список литературы к данной главе, обозначаются, как обычно, цифрами в квадратных скобках. Для работ, приведенных в списках литературы к «чужим» главам, применяется двойная нумерация, где первая цифра указывает номер главы. — *Прим. ред.*

При этом получаем передаточную функцию

$$H(f_x, f_y) = \exp(jkz) \exp[-j\pi\lambda z (f_x^2 + f_y^2)], \quad (4.11)$$

которая описывает эффекты распространения в пространстве при дифракции Френеля. Первый экспоненциальный множитель в этом выражении представляет собой общую фазовую задержку, которую приобретает каждая составляющая углового спектра при своем распространении между двумя плоскостями, находящимися на расстоянии z . Второй экспоненциальный множитель описывает фазовую дисперсию, которая зависит от частоты по квадратичному закону. Из сравнения выражений (4.11) и (3.46) следует, что найденная выше передаточная функция представляет собой просто приближение к более общей передаточной функции, полученной в гл. 3, § 6.

Приближение Фраунгофера

Расчет дифракционной картины становится еще проще, если принять более жесткие ограничения, чем в приближении Френеля. В частности, мы уже видели, что в области дифракции Френеля напряженность поля $U(x_0, y_0)$ можно найти путем фурье-преобразования произведения распределения $U(x_1, y_1)$ в отверстии на квадратичную фазовую функцию $\exp[j(k/2z)(x_1^2 + y_1^2)]$. Если к тому же принять более строгое предположение Фраунгофера

$$z \gg \frac{k(x_1^2 + y_1^2)_{\max}}{2}, \quad (4.12)$$

то квадратичный фазовый множитель становится равным приблизительно единице по всему отверстию; тогда наблюдаемое распределение поля можно найти непосредственно, вычисляя фурье-образ распределения в самом отверстии. Таким образом, в области дифракции Фраунгофера

$$U(x_0, y_0) = \frac{\exp(jkz) \exp\left[j \frac{k}{2z} (x_0^2 + y_0^2)\right]}{j\lambda z} \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x_1, y_1) \exp\left[-j \frac{2\pi}{\lambda z} (x_0 x_1 + y_0 y_1)\right] dx_1 dy_1. \quad (4.13)$$

Без учета множителей, стоящих перед интегралом, это выражение представляет собой фурье-образ распределения поля в отверстии, вычисленный для частот $f_x = x_0/\lambda z$, $f_y = y_0/\lambda z$.

В области оптических частот условия, необходимые для того, чтобы уравнения дифракции Фраунгофера были справедливы, в действительности могут быть довольно жесткими. Например, при длине волны $6 \cdot 10^{-7}$ м (красный свет) и ширине отверстия 2,5 см расстояние должно удовлетворять условию

$$z \gg 1600 \text{ м.}$$

Тем не менее в ряде важных случаев требуемые условия удовлетворяются. Кроме того, картины дифракции Фраунгофера можно наблюдать и на расстояниях, не удовлетворяющих условию (4.12), если отверстие освещается сферической волной, сходящейся к наблюдателю (см. задачу 4.3), или если между наблюдателем и отверстием (см. гл. 5) соответствующим образом расположена собирающая линза.

Наконец нужно заметить, что не существует передаточной функции, которую можно связать прямо с дифракцией Фраунгофера, так как приближение (4.12) не удовлетворяет условию пространственной инвариантности уравнения дифракции (ср. задачу 2.10). Однако, так как дифракция Фраунгофера представляет собой просто предельный случай дифракции Френеля, выражение (4.11) для передаточной функции остается справедливым всегда как для условий Френеля, так и для условий Фраунгофера.

§ 2. ПРИМЕРЫ ДИФРАКЦИОННЫХ КАРТИН ФРАУНГОФЕРА

Рассмотрим ряд примеров расчета дифракции Фраунгофера. Дополнительные примеры приведены в задачах 4.1—4.3. Результаты предыдущего параграфа можно непосредственно использовать для нахождения распределения комплексного поля фраунгоферовской дифракционной картины на любом отверстии. Однако следует иметь в виду, что реальные детекторы излучения, в том числе и глаз, реагируют на оптическую ин-

тенсивность¹⁾, а не на амплитуду поля. Следовательно, в конечном итоге дифракционные картины должны описываться распределением интенсивности.

Прямоугольное отверстие

Сначала рассмотрим случай прямоугольного отверстия с амплитудным коэффициентом пропускания

$$t(x_1, y_1) = \text{rect}\left(\frac{x_1}{l_X}\right) \text{rect}\left(\frac{y_1}{l_Y}\right).$$

Постоянные l_X и l_Y — размеры отверстия соответственно в направлении x_1 и y_1 . Если отверстие освещается нормально падающей монохроматической плоской волной единичной амплитуды, то распределение поля в отверстии описывается коэффициентом пропускания t . Таким образом, используя (4.13), дифракционную картину Фраунгофера можно описать выражением

$$U(x_0, y_0) = \frac{\exp(jkz) \exp\left[j \frac{k}{2z} (x_0^2 + y_0^2)\right]}{j\lambda z} \times \mathcal{F}\{U(x_1, y_1)\} \Big|_{\substack{f_X = x_0/\lambda z \\ f_Y = y_0/\lambda z}}$$

Учитывая, что

$$\mathcal{F}\{U(x_1, y_1)\} = l_X l_Y \text{sinc}(l_X f_X) \text{sinc}(l_Y f_Y),$$

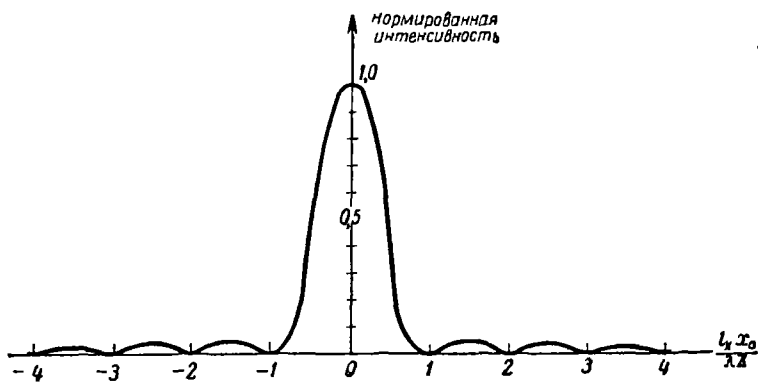
находим

$$U(x_0, y_0) = \frac{\exp(jkz) \exp\left[j \frac{k}{2z} (x_0^2 + y_0^2)\right]}{j\lambda z} \times l_X l_Y \text{sinc}\left(\frac{l_X x_0}{\lambda z}\right) \text{sinc}\left(\frac{l_Y y_0}{\lambda z}\right)$$

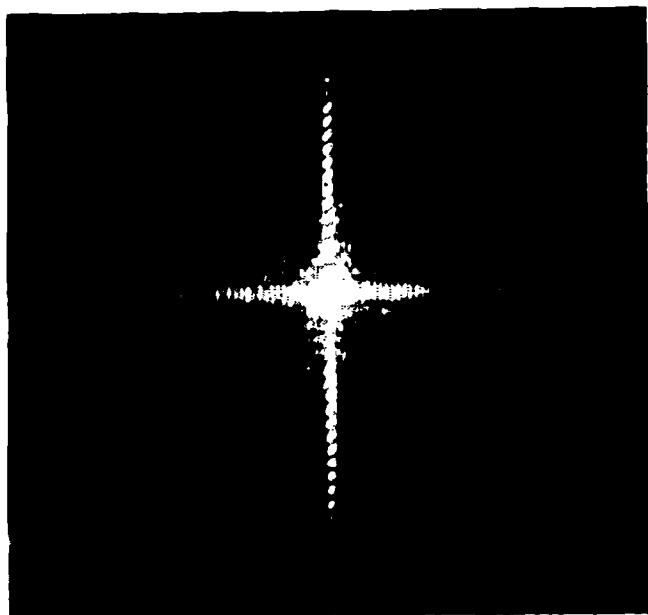
и

$$I(x_0, y_0) = \frac{l_X^2 l_Y^2}{\lambda^2 z^2} \text{sinc}^2\left(\frac{l_X x_0}{\lambda z}\right) \text{sinc}^2\left(\frac{l_Y y_0}{\lambda z}\right). \quad (4.14)$$

¹⁾ Интенсивность оптического поля определяется здесь соотношением $I(P) = 2 \langle u^2(P, t) \rangle$, где угловые скобки означают усреднение по бесконечному промежутку времени. В случае монохроматического поля с комплексной амплитудой $U(P)$ это выражение имеет вид $I(P) = |U(P)|^2$.



Фиг. 14. Распределение интенсивности в картине дифракции Фраунгофера на прямоугольном отверстии.



Фиг. 15. Картина дифракции Фраунгофера на прямоугольном отверстии ($l_x/l_y = 2$).

На фиг. 14 изображена дифракционная картина Фраунгофера в плоскости $y_0 = 0$. Заметим, что расстояние между первыми двумя нулями (ширина главного лепестка) равно

$$\Delta x_0 = 2 \frac{\lambda z}{l_x}. \quad (4.15)$$

На фиг. 15 представлена фотография картины дифракции на прямоугольном отверстии с отношением сторон $l_x/l_y = 2$.

Круглое отверстие

Рассмотрим дифракцию на отверстии, которое имеет форму круга, а не прямоугольника. Пусть диаметр этого отверстия равен l . Тогда если r_1 — величина радиуса-вектора в плоскости отверстия, то

$$t(r_1) = \text{circ} \left(\frac{r_1}{l/2} \right).$$

Осевая симметрия этой задачи позволяет заменить преобразование Фурье в (4.13) преобразованием Фурье — Бесселя. Таким образом, если r_0 — величина радиуса-вектора в плоскости наблюдения, то

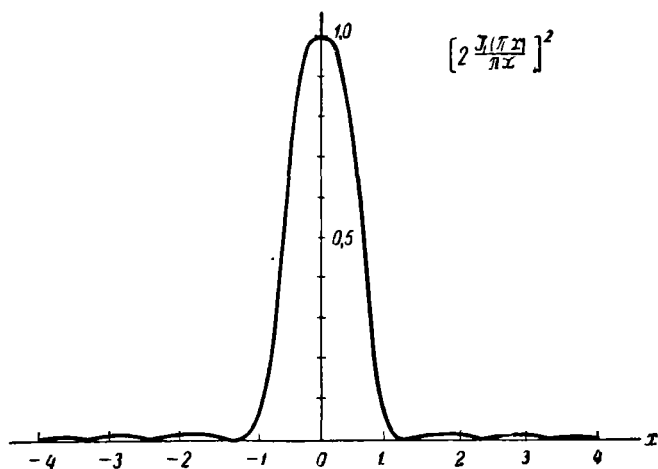
$$U(r_0) = \frac{\exp(jkz)}{j\lambda z} \exp \left(j \frac{kr_0^2}{2z} \right) \mathcal{B} \{ U(r_1) \} \Big|_{\rho=r_0/\lambda z}. \quad (4.16)$$

В случае освещения нормально падающей волной единичной амплитуды $U(r_1) = t(r_1)$; кроме того,

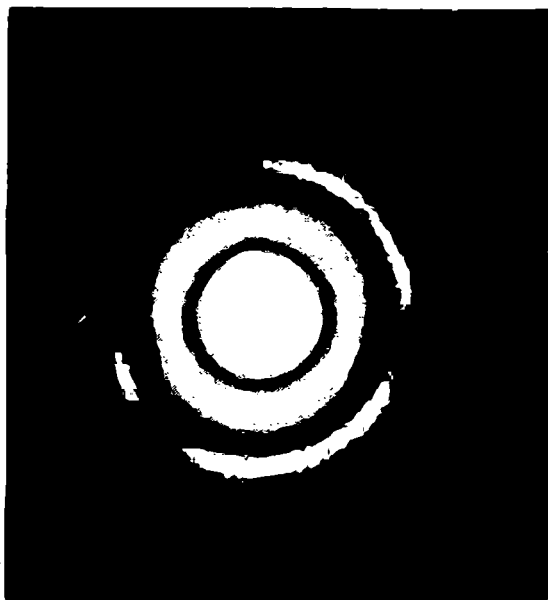
$$\mathcal{B} \left\{ \text{circ} \left(\frac{r_1}{l/2} \right) \right\} = \left(\frac{l}{2} \right)^2 \frac{J_1(\pi l \rho)}{l \rho / 2}.$$

Таким образом, видно, что распределение амплитуды в дифракционной картине Фраунгофера имеет вид

$$\begin{aligned} U(r_0) &= \frac{\exp(jkz)}{j\lambda z} \exp \left(j \frac{kr_0^2}{2z} \right) \left[\left(\frac{l}{2} \right)^2 \frac{J_1(\pi l r_0 / \lambda z)}{l r_0 / 2 \lambda z} \right] = \\ &= \exp(jkz) \exp \left(j \frac{kr_0^2}{2z} \right) \frac{kl^2}{j8z} \left[2 \frac{J_1(kl r_0 / 2z)}{kl r_0 / 2z} \right], \end{aligned}$$



Ф и г. 16. Распределение интенсивности, называемое картиной Эри:



Ф и г. 17. Картина дифракции Фраунгофера на круглом отверстии.

тогда как распределение интенсивности описывается выражением

$$I(r_0) = \left(\frac{kl^2}{8z}\right)^2 \left[2 \frac{J_1(klr_0/2z)}{klr_0/2z}\right]^2. \quad (4.17)$$

Это распределение обычно называют *картиной Эри* по имени Г. Б. Эри, который первый исследовал ее. В табл. 2 приведены значения интенсивности в последовательных максимумах и минимумах картины Эри.

Таблица 2

x	$\left[2 \frac{J_1(\pi x)}{\pi x}\right]^2$	Максимум или минимум
0	1	макс.
1,220	0	мин.
1,635	0,0175	макс.
2,233	0	мин.
2,679	0,0042	макс.
3,238	0	мин.
3,699	0,0016	макс.

Из таблицы видно, что радиус первого темного круга определяется выражением

$$\Delta r_0 = 1,22 \frac{\lambda z}{l}. \quad (4.18)$$

На фиг. 16 изображена картина Эри в плоскости $y_0=0$, а на фиг. 17 представлена фотография картины дифракции Фраунгофера на круглом отверстии.

Синусоидальная амплитудная решетка

В предыдущих примерах предполагалось, что дифракция происходит на отверстиях в непрозрачных бесконечных экранах. Обобщим теперь наше представление о предмете, на котором происходит дифракция.

Определим комплексный амплитудный коэффициент пропускания экрана $t(x, y)$ в соответствующих точках как отношение амплитуды поля непосредственно за экраном к амплитуде поля волны, падающей на экран. До сих пор рассматривался только следующий

коэффициент пропускания:

$$t(x, y) = \begin{cases} 1 & \text{в пределах отверстия,} \\ 0 & \text{за пределами отверстия.} \end{cases}$$

Однако в пределах данного отверстия амплитудный коэффициент пропускания можно сделать наперед заданной функцией координат. Зависимость поглощения от координат можно вводить, например, при помощи фотографического диапозитива, который позволяет получить все действительные значения t между нулем и единицей. Координатную зависимость фазового сдвига можно вводить при помощи прозрачных пластинок переменной толщины, что позволяет реализовать все значения t , лежащие в комплексной плоскости внутри окружности единичного радиуса.

В качестве примера этого более общего типа экрана, на котором происходит дифракция, рассмотрим *синусоидальную амплитудную решетку*, определяемую коэффициентом пропускания

$$t(x, y) = \left[\frac{1}{2} + \frac{m}{2} \cos(2\pi f_0 x) \right] \text{rect} \left(\frac{x}{l} \right) \text{rect} \left(\frac{y}{l} \right), \quad (4.19)$$

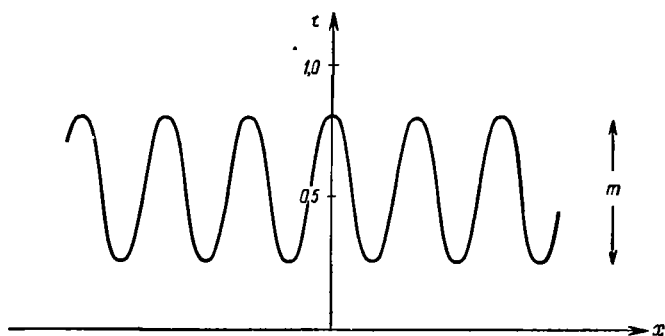
где для простоты мы предположили, что структура решетки ограничена квадратом со стороной l . Параметр m определяет разность между максимальным и минимальным коэффициентами пропускания, а f_0 — частота решетки. На фиг. 18 изображен график коэффициента пропускания.

Если экран освещен нормально падающей монохроматической плоской волной единичной амплитуды, то распределение поля в отверстии описывается просто функцией t . Для нахождения дифракционной картины Фраунгофера сначала найдем преобразование Фурье этого распределения. Учитывая, что

$$\mathcal{F} \left\{ \frac{1}{2} + \frac{m}{2} \cos(2\pi f_0 x) \right\} = \frac{1}{2} \delta(f_x, f_y) + \\ + \frac{m}{4} \delta(f_x + f_0, f_y) + \frac{m}{4} \delta(f_x - f_0, f_y)$$

и

$$\mathcal{F} \left\{ \text{rect} \left(\frac{x_1}{l} \right) \text{rect} \left(\frac{y_1}{l} \right) \right\} = l^2 \text{sinc}(lf_x) \text{sinc}(lf_y),$$



Фиг. 18. Функция пропускания синусоидальной амплитудной решетки.

можно, пользуясь теоремой свертки, записать

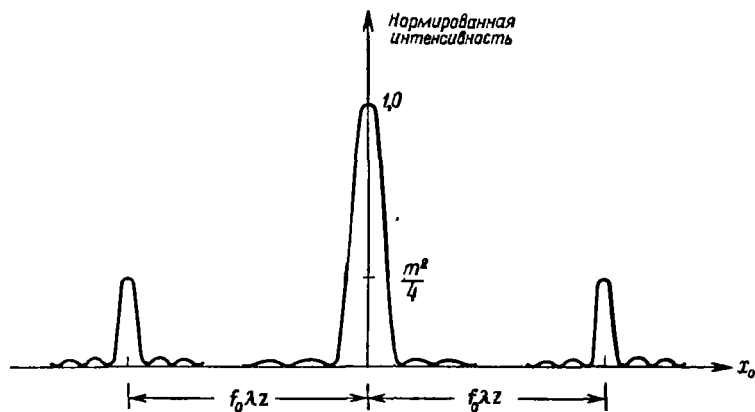
$$\mathcal{F}\{U(x_1, y_1)\} = \frac{l^2}{2} \operatorname{sinc}(lf_y) \left\{ \operatorname{sinc}(lf_x) + \frac{m}{2} \operatorname{sinc}[l(f_x + f_0)] + \frac{m}{2} \operatorname{sinc}[l(f_x - f_0)] \right\}.$$

Дифракционную картину Фраунгофера теперь можно представить в виде

$$U(x_0, y_0) = \frac{l^2}{j2\lambda z} \exp(jkz) \exp\left[j \frac{k}{2z} (x_0^2 + y_0^2)\right] \times \\ \times \operatorname{sinc}\left(\frac{ly_0}{\lambda z}\right) \left\{ \operatorname{sinc}\left(\frac{lx_0}{\lambda z}\right) + \frac{m}{2} \operatorname{sinc}\left[\frac{l}{\lambda z} (x_0 + f_0\lambda z)\right] + \frac{m}{2} \operatorname{sinc}\left[\frac{l}{\lambda z} (x_0 - f_0\lambda z)\right] \right\}. \quad (4.20)$$

Наконец, соответствующую интенсивность можно найти, возводя в квадрат выражение (4.20). Заметим, что если частота решетки f_0 много больше $2/l$, то перекрытием трех sinc функций, зависящих от x_0 , можно пренебречь. В результате получим

$$I(x_0, y_0) = \left(\frac{l^2}{2\lambda z}\right)^2 \operatorname{sinc}^2\left(\frac{ly_0}{\lambda z}\right) \left\{ \operatorname{sinc}^2\left(\frac{lx_0}{\lambda z}\right) + \frac{m^2}{4} \operatorname{sinc}^2\left[\frac{l}{\lambda z} (x_0 + f_0\lambda z)\right] + \frac{m^2}{4} \operatorname{sinc}^2\left[\frac{l}{\lambda z} (x_0 - f_0\lambda z)\right] \right\}. \quad (4.21)$$



Фиг. 19. Распределение интенсивности в картине дифракции Фраунгофера на синусоидальной амплитудной решетке.

На фиг. 19 представлена картина распределения интенсивности. Заметим, что синусоидальное изменение пропускания по отверстию обуславливает перераспределение энергии: некоторая доля ее переходит от центрального дифракционного максимума к двум дополнительным боковым максимумам. Центральный дифракционный максимум называется составляющей *нулевого порядка* дифракционной картины Фраунгофера, тогда как два боковых максимума называются составляющими *первого порядка*. Составляющие первого порядка отделены в пространстве от составляющей нулевого порядка расстоянием $f_0 \lambda z$, а ширина каждой из них пропорциональна $\lambda z/l$. Отсюда следует, что разрешающая способность решетки пропорциональна величине $f_0 \lambda z / (\lambda z/l) = f_0 l$, т. е. числу периодов решетки, и не зависит от расстояния z до плоскости наблюдения. Наконец, нужно подчеркнуть, что наши рассуждения основаны на скалярной теории и, следовательно, справедливы только в том случае, когда частота решетки мала по сравнению с λ^{-1} . Если это условие не удовлетворяется, то необходимо использовать векторное представление, основанное на уравнениях Максвелла.

Синусоидальная фазовая решетка

В качестве последнего примера расчета дифракции Фраунгофера рассмотрим *синусоидальную фазовую решетку*, определяемую коэффициентом пропускания

$$t(x_1, y_1) = \exp\left[j \frac{m}{2} \sin(2\pi f_0 x_1)\right] \operatorname{rect}\left(\frac{x_1}{l}\right) \operatorname{rect}\left(\frac{y_1}{l}\right), \quad (4.22)$$

где соответствующим выбором начальной фазы мы исключили множитель, представляющий среднюю по решетке фазовую задержку. Параметр m определяет разность между максимальным и минимальным значением фазового сдвига.

Если экран освещается нормально падающей плоской волной единичной амплитуды, то распределение поля непосредственно за экраном точно описывается выражением (4.22). Анализ можно упростить, используя тождество

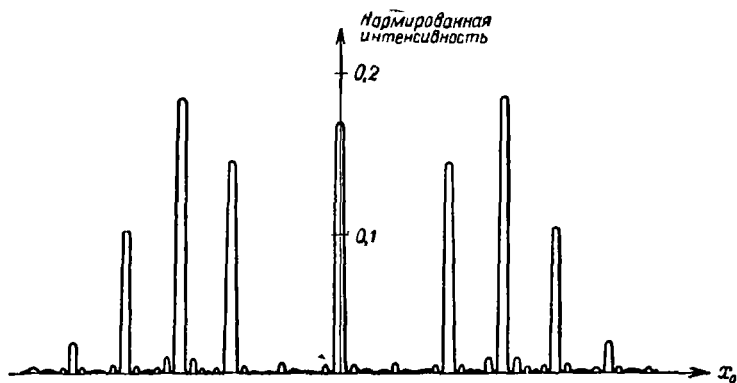
$$\exp\left[j \frac{m}{2} \sin(2\pi f_0 x_1)\right] = \sum_{q=-\infty}^{\infty} J_q\left(\frac{m}{2}\right) \exp(j2\pi q f_0 x_1),$$

где J_q — функция Бесселя первого рода, порядка q . Таким образом,

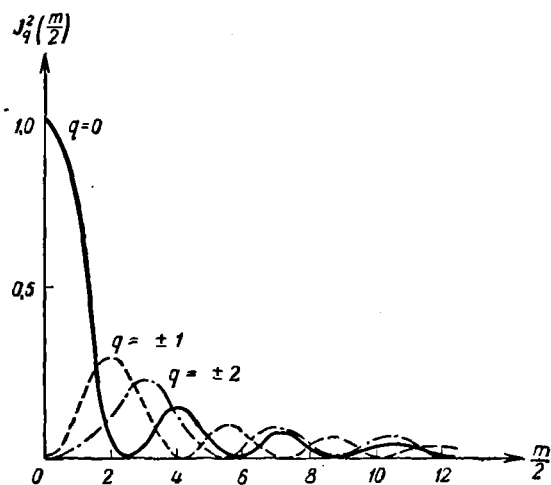
$$\begin{aligned} \mathcal{F}\{U(x_1, y_1)\} &= \mathcal{F}\{t(x_1, y_1)\} = \\ &= [l^2 \operatorname{sinc}(lf_x) \operatorname{sinc}(lf_y)] * \left[\sum_{q=-\infty}^{\infty} J_q\left(\frac{m}{2}\right) \delta(f_x - qf_0, f_y) \right] = \\ &= \sum_{q=-\infty}^{\infty} l^2 J_q\left(\frac{m}{2}\right) \operatorname{sinc}[l(f_x - qf_0)] \operatorname{sinc}(lf_y), \end{aligned}$$

и напряженность поля в дифракционной картине Фраунгофера можно записать в виде

$$\begin{aligned} U(x_0, y_0) &= \frac{l^2}{j\lambda z} \exp(jkz) \exp\left[j \frac{k}{2z}(x_0^2 + y_0^2)\right] \times \\ &\times \sum_{q=-\infty}^{\infty} J_q\left(\frac{m}{2}\right) \operatorname{sinc}\left[\frac{l}{\lambda z}(x_0 - qf_0 \lambda z)\right] \operatorname{sinc}\left(\frac{ly_0}{\lambda z}\right). \quad (4.23) \end{aligned}$$



Ф и г. 20. Распределение интенсивности в картине дифракции Фраунгофера на синусоидальной фазовой решетке ($m=8$).



Ф и г. 21. График функции $J_q^2\left(\frac{m}{2}\right)$ для трех значений $\pm q$.

Если мы опять будем считать $f_0 \gg 2/l$, то перекрытием различных дифракционных порядков можно пренебречь, и соответствующее распределение интенсивности принимает вид

$$I(x_0, y_0) = \left(\frac{l^2}{\lambda z}\right)^2 \sum_{q=-\infty}^{\infty} J_q^2\left(\frac{m}{2}\right) \operatorname{sinc}^2\left[\frac{l}{\lambda z}(x_0 - qf_0\lambda z)\right] \operatorname{sinc}^2\left(\frac{ly_0}{\lambda z}\right). \quad (4.24)$$

Введение синусоидальной фазовой решетки, таким образом, приводит к тому, что энергия перераспределяется между составляющей нулевого порядка и множественной составляющих более высокого порядка. Максимальная интенсивность составляющей q -го порядка равна просто $[l^2 J_q^2(m/2)/\lambda z]^2$, а расстояние этой составляющей от центра дифракционной картины равно $qf_0\lambda z$. На фиг. 20 изображено распределение интенсивности для случая, когда разность между максимальным и минимальным значением фазы (m) равна 8 рад . На фиг. 21 представлен график зависимости J_q^2 от $m/2$ для различных значений q . Отметим, что если $m/2$ является корнем функции J_0 , то максимум нулевого порядка исчезает.

§ 3. ПРИМЕР ДИФРАКЦИОННОЙ КАРТИНЫ ФРЕНЕЛЯ

Рассчитать дифракционную картину Френеля обычно бывает сложнее, чем дифракционную картину Фраунгофера. В качестве иллюстрации случая дифракции Френеля рассмотрим дифракцию только на квадратном отверстии; другие примеры расчета дифракции Френеля даны в задачах 4.3—4.5.

Предположим, что квадратное отверстие со стороной l освещено нормально падающей монохроматической плоской волной единичной амплитуды. Тогда распределение поля непосредственно за отверстием можно записать в виде

$$U(x_1, y_1) = \operatorname{rect}\left(\frac{x_1}{l}\right) \operatorname{rect}\left(\frac{y_1}{l}\right). \quad (4.25)$$

При рассмотрении дифракции Френеля удобнее всего использовать представление возмущения в виде свертки. При этом получается следующее выражение для дифрагированного поля:

$$U(x_0, y_0) = \frac{\exp(jkz)}{j\lambda z} \int_{-l/2}^{l/2} \int_{-l/2}^{l/2} \exp \left\{ j \frac{k}{2z} [(x_1 - x_0)^2 + (y_1 - y_0)^2] \right\} dx_1 dy_1.$$

Это выражение можно представить в виде произведения двух интегралов

$$U(x_0, y_0) = \frac{\exp(jkz)}{j\lambda z} \mathfrak{S}(x_0) \mathfrak{S}(y_0), \quad (4.26)$$

где

$$\mathfrak{S}(x_0) = \int_{-l/2}^{l/2} \exp \left[j \frac{k}{2z} (x_1 - x_0)^2 \right] dx_1,$$

$$\mathfrak{S}(y_0) = \int_{-l/2}^{l/2} \exp \left[j \frac{k}{2z} (y_1 - y_0)^2 \right] dy_1.$$

Интегралы существенно упрощаются после замены переменных

$$\xi = \sqrt{\frac{k}{\pi z}} (x_1 - x_0), \quad \eta = \sqrt{\frac{k}{\pi z}} (y_1 - y_0),$$

и переходят в следующие интегралы:

$$\mathfrak{S}(x_0) = \sqrt{\frac{\pi z}{k}} \int_{\xi_1}^{\xi_2} \exp \left(j \frac{\pi}{2} \xi^2 \right) d\xi,$$

$$\mathfrak{S}(y_0) = \sqrt{\frac{\pi z}{k}} \int_{\eta_1}^{\eta_2} \exp \left(j \frac{\pi}{2} \eta^2 \right) d\eta,$$

где пределы интегрирования определяются соотношениями

$$\begin{aligned} \xi_1 &= -\sqrt{\frac{k}{\pi z}} \left(\frac{l}{2} + x_0 \right), & \xi_2 &= \sqrt{\frac{k}{\pi z}} \left(\frac{l}{2} - x_0 \right), \\ \eta_1 &= -\sqrt{\frac{k}{\pi z}} \left(\frac{l}{2} + y_0 \right), & \eta_2 &= \sqrt{\frac{k}{\pi z}} \left(\frac{l}{2} - y_0 \right). \end{aligned} \quad (4.27)$$

Интегралы $\mathfrak{Z}(x_0)$ и $\mathfrak{Z}(y_0)$ можно выразить через табулированные функции, известные под названием *интегралов Френеля*. Последние определяются выражениями

$$C(\alpha) = \int_0^{\alpha} \cos \frac{\pi t^2}{2} dt, \quad S(\alpha) = \int_0^{\alpha} \sin \frac{\pi t^2}{2} dt. \quad (4.28)$$

Учитывая, что

$$\int_{\xi_1}^{\xi_2} \exp\left(j \frac{\pi}{2} \xi^2\right) d\xi = \int_0^{\xi_2} \exp\left(j \frac{\pi}{2} \xi^2\right) d\xi - \int_0^{\xi_1} \exp\left(j \frac{\pi}{2} \xi^2\right) d\xi,$$

можно записать

$$\mathfrak{Z}(x_0) = \sqrt{\frac{\pi z}{k}} \{ [C(\xi_2) - C(\xi_1)] + j[S(\xi_2) - S(\xi_1)] \}, \quad (4.29)$$

$$\mathfrak{Z}(y_0) = \sqrt{\frac{\pi z}{k}} \{ [C(\eta_2) - C(\eta_1)] + j[S(\eta_2) - S(\eta_1)] \}.$$

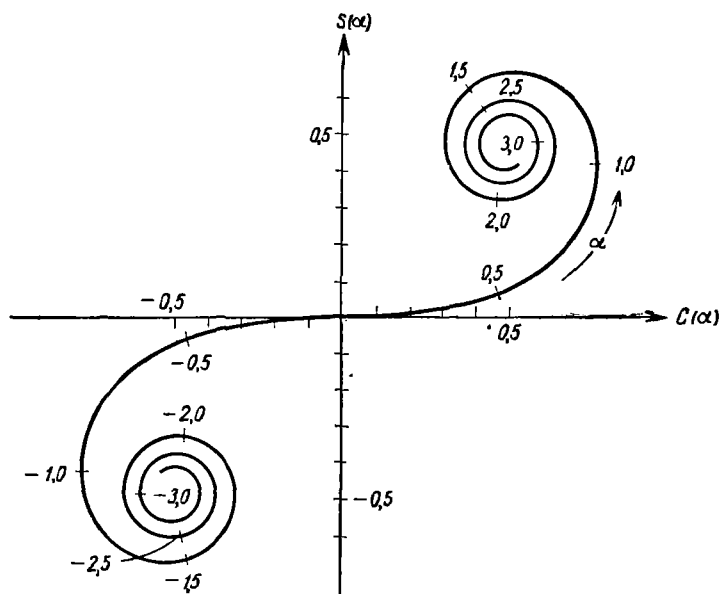
Наконец, подставляя (4.29) в (4.26), получаем распределение комплексного поля

$$U(x_0, y_0) = \frac{\exp(jkz)}{2j} \{ [C(\xi_2) - C(\xi_1)] + j[S(\xi_2) - S(\xi_1)] \} \times \\ \times \{ [C(\eta_2) - C(\eta_1)] + j[S(\eta_2) - S(\eta_1)] \} \quad (4.30)$$

и соответствующее распределение интенсивности

$$I(x_0, y_0) = \frac{1}{4} \{ [C(\xi_2) - C(\xi_1)]^2 + \\ + [S(\xi_2) - S(\xi_1)]^2 \} \{ [C(\eta_2) - C(\eta_1)]^2 + [S(\eta_2) - S(\eta_1)]^2 \}. \quad (4.31)$$

Для интерпретации этих выражений удобно воспользоваться графическим построением, которое называется *спиралью Корню*; спираль Корню (фиг. 22) представляет собой совместный график зависимости $C(\alpha)$ и $S(\alpha)$



Фиг. 22. Спираль Корню.

от параметра α . Заметим, что величину $C(\alpha) + jS(\alpha)$ можно считать комплексным фазором, соединяющим начало координат с точкой α на спирали. Следовательно, величина $\{[C(\xi_2) - C(\xi_1)] + j[S(\xi_2) - S(\xi_1)]\}$ представляет собой фазор, определяемый участком спирали между точками ξ_1 и ξ_2 . Используя подобный графический метод, можно вычислить значения выражений (4.30) и (4.31) в каждой точке дифракционной картины.

Для иллюстрации расчета такого типа рассмотрим распределение поля $U(x_0, y_0)$ в предельном случае дифракции Френеля на очень малом расстоянии z от отверстия, но таком, чтобы приближение Френеля все же выполнялось. Для достаточно малых z величина $k/\pi z$ — очень большое число, так что можно принять

следующие приближенные значения пределов (4.27) ¹⁾:

$$\xi_1 \approx \begin{cases} -\infty & x_0 > -\frac{l}{2}, \\ +\infty & x_0 < -\frac{l}{2}, \end{cases} \quad \xi_2 \approx \begin{cases} -\infty & x_0 > \frac{l}{2}, \\ +\infty & x_0 < \frac{l}{2}, \end{cases}$$

$$\eta_1 \approx \begin{cases} -\infty & y_0 > -\frac{l}{2}, \\ +\infty & y_0 < -\frac{l}{2}, \end{cases} \quad \eta_2 \approx \begin{cases} -\infty & y_0 > \frac{l}{2}, \\ +\infty & y_0 < \frac{l}{2}, \end{cases}$$

Используя спираль Корню, получаем соответствующие значения $C(\alpha)$ и $S(\alpha)$:

$$C(\xi_1) = S(\xi_1) \approx \begin{cases} -\frac{1}{2} & x_0 > -\frac{l}{2}, \\ +\frac{1}{2} & x_0 < -\frac{l}{2}, \end{cases}$$

$$C(\xi_2) = S(\xi_2) \approx \begin{cases} -\frac{1}{2} & x_0 > \frac{l}{2}, \\ +\frac{1}{2} & x_0 < \frac{l}{2}, \end{cases}$$

$$C(\eta_1) = S(\eta_1) \approx \begin{cases} -\frac{1}{2} & y_0 > -\frac{l}{2}, \\ +\frac{1}{2} & y_0 < -\frac{l}{2}, \end{cases}$$

$$C(\eta_2) = S(\eta_2) \approx \begin{cases} -\frac{1}{2} & y_0 > \frac{l}{2}, \\ +\frac{1}{2} & y_0 < \frac{l}{2}. \end{cases}$$

Подставляя эти результаты в (4.30), имеем

$$U(x_0, y_0) = \exp(jkz) \operatorname{rect}\left(\frac{x_0}{l}\right) \operatorname{rect}\left(\frac{y_0}{l}\right). \quad (4.32)$$

¹⁾ Принимая такие пределы интегрирования, автор фактически отказывается от сделанной выше оговорки о применимости приближения Френеля в том виде, который принят в этой книге. Строго говоря, переход к геометрической оптике выходит за рамки принятого приближения. Переход к геометрической оптике означает, что в рассматриваемой плоскости мы пренебрегли шириной области, в которой по крайней мере один из пределов интегрирования соответствует малым значениям аргумента α по сравнению с размерами области, где оба предела имеют большие α . — *Прим. ред.*

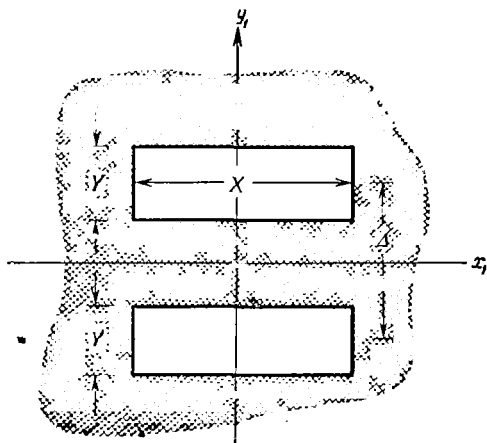
Таким образом, распределение поля $U(x_0, y_0)$ в предельном случае дифракции Френеля соответствует геометрической проекции распределения поля в отверстии $U(x_1, y_1)$, и, следовательно, мы приходим к геометрической оптике. Этот результат, разумеется, справедлив не только для прямоугольного отверстия, но и для отверстий произвольной формы, как это можно показать при помощи принципа стационарной фазы (см. работу [3.4]).

ЗАДАЧИ

4.1. Экран освещается нормально падающей монохроматической волной единичной амплитуды.

а) Найти распределение интенсивности поля в дифракционной картине Фраунгофера на двух щелях (фиг. 23).

б) Изобразить распределение интенсивности в дифракционной картине вдоль осей x_0 и y_0 плоскости наблюдения, считая $X/\lambda z = 10 \text{ м}^{-1}$, $Y/\lambda z = 1 \text{ м}^{-1}$, $\Delta/\lambda z = 3/2 \text{ м}^{-1}$, где z — расстояние до плоскости наблюдения, а λ — длина волны.



Фиг. 23.

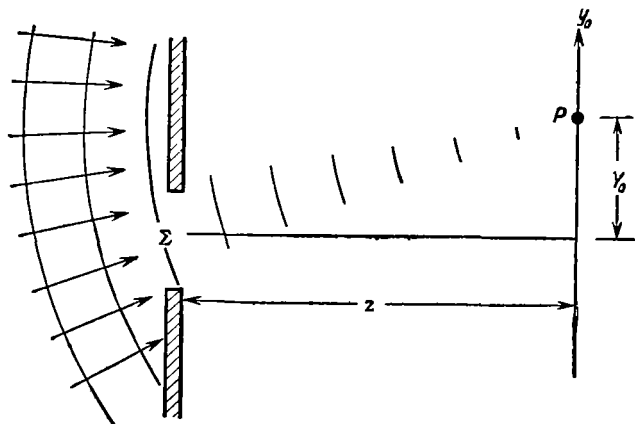
4.2. а) Изобразить препятствие, описываемое коэффициентом пропускания

$$t(x_1, y_1) = \left\{ \left[\text{rect} \left(\frac{x_1}{X} \right) \text{rect} \left(\frac{y_1}{Y} \right) \right] \cdot \left[\frac{1}{\Delta} \text{comb} \left(\frac{y_1}{\Delta} \right) \delta(x_1) \right] \right\} \text{rect} \left(\frac{y_1}{N\Delta} \right),$$

где N — нечетное целое число, а $\Delta > Y$.

б) Найти выражение для распределения интенсивности в картине дифракции Фраунгофера на этом отверстии, считая, что оно освещается нормально падающей плоской волной, а $N \gg 1$.

в) Каким должно быть соотношение между Y и Δ для того, чтобы максимумы четного порядка имели наименьшую величину, тогда как максимум нулевого порядка оставался приблизительно неизменным?



Фиг. 24.

4.3. Отверстие Σ в непрозрачном экране освещается сферической волной, сходящейся в точке P , расположенной в плоскости, параллельной экрану на расстоянии z за отверстием (фиг. 24).

а) Найти квадратичное приближение для волнового фронта падающей волны в плоскости отверстия, считая, что 1) точка P лежит на оси z ; 2) имеет координаты $(0, Y_0)$.

б) Считая, что дифракция Френеля на отверстии наблюдается в плоскости, содержащей точку P , показать, что в обоих указанных выше случаях распределение интенсивности представляет собой дифракционную картину Фраунгофера на отверстии с центром в точке P .

4.4. Найти интенсивность поля на оси картины дифракции Френеля на отверстиях со следующими коэффициентами пропускания (предполагая, что освещение осуществляется нормально падающей плоской волной единичной амплитуды):

а) $t(x_1, y_1) = \text{circ } \sqrt{x_1^2 + y_1^2}$;

$$б) t(x_1, y_1) = \begin{cases} 1 & a \leq \sqrt{x_1^2 + y_1^2} \leq 1, \\ 0 & \text{в остальных случаях,} \end{cases}$$

где $a < 1$.

4.5. Принято считать, что две дискретные спектральные линии можно «навверняка разрешить» при помощи дифракционной решетки, если дифракционный максимум q -го порядка для излучения с длиной волны λ_1 точно совпадает с первым дифракционным минимумом q -го порядка для излучения с длиной волны λ_2 . *Разрешающая способность* решетки равна отношению средней длины волны λ к минимальной разрешаемой разности длин волн $\Delta\lambda$. Показать, что разрешающая способность синусоидальной фазовой решетки, рассмотренной в данной главе, равна

$$\frac{\lambda}{\Delta\lambda} = qM = qM,$$

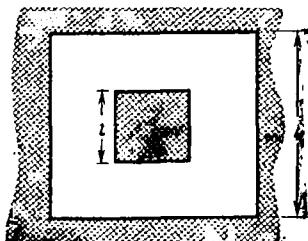
где q — порядок дифракции, используемый при измерении, а M — число штрихов решетки. Какое явление ограничивает использование произвольно высоких порядков дифракции?

4.6. а) Считая, что освещение осуществляется нормально падающей плоской волной единичной амплитуды, и пренебрегая конечностью размеров отверстия, найти картину дифракции Френеля на отверстии, которое имеет следующий коэффициент пропускания:

$$t(x, y) = \frac{1}{2} (1 + m \cos 2\pi f_0 x).$$

б) Учитывая, что $m \ll 1$, определить, на каких расстояниях z от отверстия распределение поля в плоскости, параллельной экрану, 1) модулировано в пространстве только по амплитуде, 2) модулировано в пространстве только по фазе?

4.7. Найти выражение для распределения интенсивности в картине дифракции Фраунгофера на отверстии, изображенном на фиг. 25. Предполагается, что освещение осуществляется нормально падающей плоской волной единичной амплитуды.



Фиг. 25.

ГЛАВА 5

ЛИНЗЫ КАК ЭЛЕМЕНТЫ, ВЫПОЛНЯЮЩИЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ И СОЗДАЮЩИЕ ИЗОБРАЖЕНИЕ

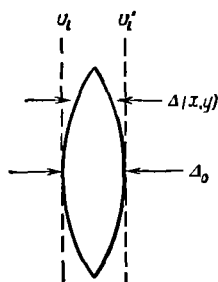
Наиболее важными компонентами оптических систем, образующих изображения, и оптических систем обработки информации являются, конечно, линзы. Хотя подробное обсуждение свойств линз было бы полезно, это потребовало бы довольно длинного экскурса в основные принципы геометрической оптики. Так как в нашем случае полное обсуждение не обязательно, мы воспользуемся другой точкой зрения, хотя и не основанной непосредственно на принципах геометрической оптики, однако согласующейся с ними в конечных пределах. Мы ограничимся рассмотрением монохроматического освещения, откладывая обобщения на случай немонохроматического освещения до гл. 6.

§ 1. ТОНКАЯ ЛИНЗА КАК ЭЛЕМЕНТ, ВЫПОЛНЯЮЩИЙ ФАЗОВОЕ ПРЕОБРАЗОВАНИЕ

Линзы изготавливаются из оптически плотного материала, как правило, из стекла, в котором скорость распространения оптического возмущения меньше, чем в воздухе. Линза называется *тонкой*, если луч, входящий в точке с координатами (x, y) на одной поверхности линзы, выходит в точке приблизительно с такими же координатами на противоположной поверхности, т. е. если смещением луча внутри линзы можно пренебречь. Таким образом, тонкая линза просто задерживает фронт падающей волны на величину, пропорциональную толщине линзы в каждой точке.

Пусть максимальная толщина линзы (фиг. 26) равна Δ_0 , а толщина ее в точке с координатами (x, y) равна $\Delta(x, y)$. Тогда в точке с координатами (x, y) полное изменение фазы после прохождения волны через линзу можно записать в виде

$$\varphi(x, y) = kn\Delta(x, y) + k[\Delta_0 - \Delta(x, y)],$$



Ф и г. 26. Функция толщины.

где n — показатель преломления материала линзы, $kn\Delta(x, y)$ — изменение фазы, введенное линзой, а $k[\Delta_0 - \Delta(x, y)]$ — изменение фазы за счет оставшейся области свободного пространства между двумя плоскостями. То же изменение фазы можно представить в виде произведения фазовых преобразований

$$t_l(x, y) = \exp[jk\Delta_0] \exp[jk(n-1)\Delta(x, y)]. \quad (5.1)$$

Комплексное поле $U_l'(x, y)$ в плоскости, расположенной непосредственно позади линзы, связано с комплексным полем $U_l(x, y)$ волны, падающей на плоскость, расположенную непосредственно перед линзой, соотношением

$$U_l'(x, y) = t_l(x, y) U_l(x, y). \quad (5.2)$$

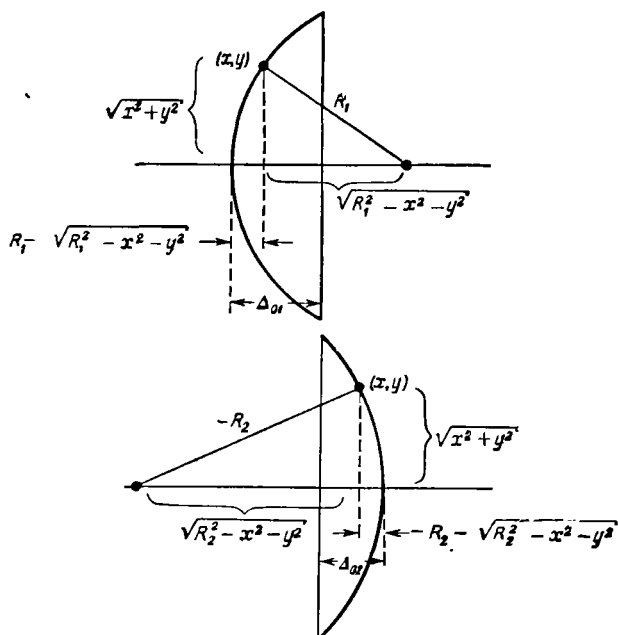
Чтобы понять действие линзы, остается установить математический вид функции толщины $\Delta(x, y)$.

Функция толщины

Чтобы установить вид фазовых преобразований для различных типов линз, прежде всего введем следующее правило знаков. Условимся считать радиус кривизны каждой *выпуклой* поверхности, встречающейся на пути луча, распространяющегося слева направо, *положительным*, а радиус кривизны *вогнутой* поверхности — *отрицательным*. Таким образом, на фиг. 26 радиус кривизны левой поверхности выражается положительным числом R_1 , а радиус кривизны правой поверхности — отрицательным числом R_2 .

Чтобы найти толщину $\Delta(x, y)$, разделим линзу на две части, как показано на фиг. 27, и запишем полную функцию толщины в виде суммы двух отдельных функций толщины

$$\Delta(x, y) = \Delta_1(x, y) + \Delta_2(x, y). \quad (5.3)$$



Фиг. 27. Расчет функции толщины.

Для линзы, изображенной в верхней части фиг. 27, имеем

$$\begin{aligned} \Delta_1(x, y) &= \Delta_{01} - (R_1 - \sqrt{R_1^2 - x^2 - y^2}) = \\ &= \Delta_{01} - R_1 \left(1 - \sqrt{1 - \frac{x^2 + y^2}{R_1^2}} \right). \end{aligned} \quad (5.4)$$

Точно так же

$$\begin{aligned} \Delta_2(x, y) &= \Delta_{02} - (-R_2 - \sqrt{R_2^2 - x^2 - y^2}) = \\ &= \Delta_{02} + R_2 \left(1 - \sqrt{1 - \frac{x^2 + y^2}{R_2^2}} \right), \end{aligned} \quad (5.5)$$

где положительное число $-R_2$ вынесено из-под знака радикала. Комбинируя соотношения (5.3) — (5.5),

получаем выражение для полной функции толщины

$$\Delta(x, y) = \Delta_0 - R_1 \left(1 - \sqrt{1 - \frac{x^2 + y^2}{R_1^2}} \right) + \\ + R_2 \left(1 - \sqrt{1 - \frac{x^2 + y^2}{R_2^2}} \right), \quad (5.6)$$

где $\Delta_0 = \Delta_{01} + \Delta_{02}$.

Параксиальное приближение

Выражение для функции толщины можно существенно упростить, если ограничиться рассмотрением части волнового фронта вблизи оси линзы, или, что то же самое, если рассматривать только *параксиальные* лучи. Таким образом, мы будем рассматривать только те достаточно малые значения x и y , для которых справедливы следующие приближения:

$$\sqrt{1 - \frac{x^2 + y^2}{R_1^2}} \approx 1 - \frac{x^2 + y^2}{2R_1^2}, \\ \sqrt{1 - \frac{x^2 + y^2}{R_2^2}} \approx 1 - \frac{x^2 + y^2}{2R_2^2}. \quad (5.7)$$

Разумеется, результирующее фазовое преобразование будет правильно описывать линзу только в ограниченной области, однако это не более жесткое ограничение, чем обычное параксиальное приближение геометрической оптики. Заметим, что соотношения (5.7) означают замену сферических поверхностей линзы параболическими. С учетом этих приближений функция толщины принимает вид

$$\Delta(x, y) = \Delta_0 - \frac{x^2 + y^2}{2} \left(\frac{1}{R_1} - \frac{1}{R_2} \right). \quad (5.8)$$

Фазовое преобразование и его физический смысл

Подставляя выражение (5.8) в (5.1), получаем следующее параксиальное приближение для преобразования, произведенного линзой:

$$t_l(x, y) = \exp[jkn\Delta_0] \exp \left[-jk(n-1) \frac{x^2 + y^2}{2} \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \right].$$

Физические характеристики линзы (т. е. n , R_1 и R_2) можно объединить в одном числе f , которое называется *фокусным расстоянием* и определяется выражением

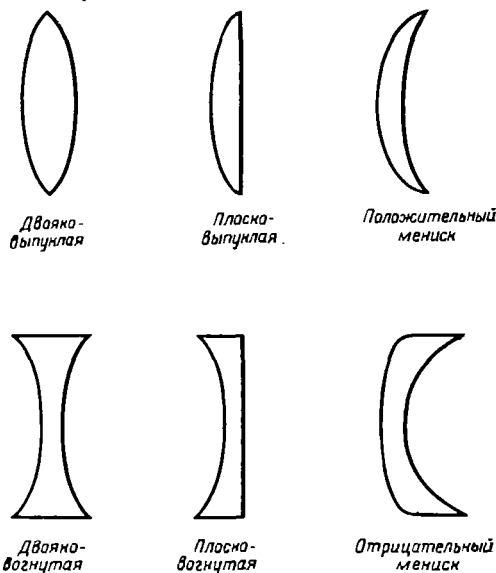
$$\frac{1}{f} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right). \quad (5.9)$$

Таким образом, фазовое преобразование можно записать в виде

$$t_l(x, y) = \exp[jkn\Delta_0] \exp\left[-j\frac{k}{2f}(x^2 + y^2)\right]. \quad (5.10)$$

Именно этим основным выражением мы будем пользоваться, рассматривая действие линзы на падающее возмущение.

Заметим, что, хотя вывод этого выражения сделан для линзы определенной формы, изображенной на фиг. 26, принятое правило знаков позволяет применять



Фиг. 28. Различные типы линз.

полученные результаты к линзам других типов. На фиг. 28 представлено несколько различных типов линз с разными сочетаниями выпуклых и вогнутых поверхностей. Принятое здесь правило знаков означает, что фокусное расстояние f двояковыпуклой, плосковыпуклой линз и положительного мениска *положительно*, а двояковогнутой, плосковогнутой линз и отрицательного мениска *отрицательно* (мы предлагаем читателю убедиться в этом самостоятельно; см. задачу 5.1). Таким образом, если знак фокусного расстояния выбран правильно, то для описания любой из линз можно использовать выражение (5.10).

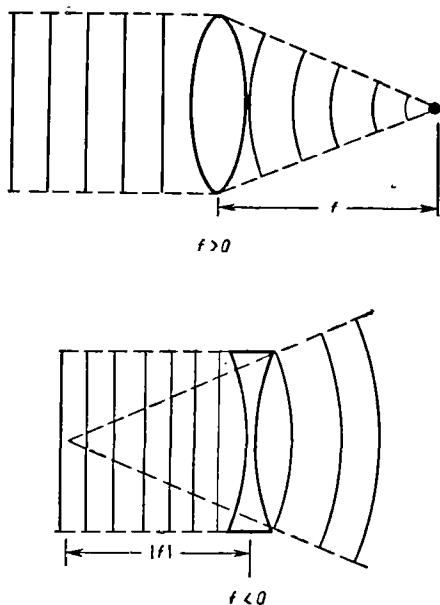
Чтобы лучше понять физический смысл преобразования, осуществляемого линзой, рассмотрим действие линзы на нормально падающую плоскую волну единичной амплитуды. В этом случае распределение поля U_i в плоскости перед линзой однородно; из соотношения (5.2) вытекает следующее выражение для возмущения U'_i за линзой;

$$U'_i(x, y) = \exp[jkn\Delta_0] \exp\left[-j\frac{k}{2f}(x^2 + y^2)\right].$$

Первый член представляет собой постоянный фазовый сдвиг, тогда как второй член аппроксимирует в квадратичном приближении сферическую волну.

Если фокусное расстояние f положительно, то сферическая волна сходится на оси линзы в точке, расположенной на расстоянии f за линзой. Если же фокусное расстояние отрицательно, то сферическая волна расходится из точки на оси линзы, расположенной на расстоянии f перед линзой. Оба эти случая показаны на фиг. 29. В связи с этим линза с положительным фокусным расстоянием называется *собирающей* (или *положительной*), а линза с отрицательным фокусным расстоянием — *рассеивающей* (или *отрицательной*).

Наше заключение относительно того, что линза со сферическими поверхностями преобразует падающую плоскую волну в сферическую, справедливо лишь в параксиальном приближении. Если условие параксиальности не выполняется, то даже при идеально сфериче-



Фиг. 29. Действие собирающей и рассеивающей линз на нормально падающую плоскую волну.

ских поверхностях линз фронт волны на выходе будет иметь отклонения от идеальной сферичности, называемые *абберациями* (см. гл. 6, § 4). На практике для улучшения сферичности волнового фронта на выходе линзы часто «исправляют», придавая путем шлифовки их поверхностям несферическую форму.

Необходимо, однако, подчеркнуть, что результаты, которые могут быть получены при использовании фазового преобразования (5.19), в действительности имеют более общий характер, чем можно ожидать на основании анализа, с помощью которого это преобразование было получено. Тщательный анализ большинства систем из хорошо исправленных линз, выполненный на основе геометрической оптики, показывает, что практически их свойства совпадают с предсказаниями нашей ограниченной теории.

§ 2. ЛИНЗЫ КАК ЭЛЕМЕНТЫ, ВЫПОЛНЯЮЩИЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ

Способность собирающей линзы выполнять двумерное преобразование Фурье — одно из наиболее замечательных и полезных ее свойств. С операцией преобразования Фурье обычно ассоциируются громоздкие, сложные и дорогие электронные спектральные анализаторы, однако эту трудную аналоговую операцию может с предельной простотой выполнять когерентная оптическая система.

Далее мы рассмотрим три различные конфигурации, при помощи которых выполняется операция преобразования. Во всех случаях освещение считается монохроматическим и рассматривается распределение амплитуды света в задней фокальной плоскости линзы¹⁾. Эти конфигурации представлены на фиг. 30, *a—в*. В случае *a* преобразуемый предмет помещается непосредственно перед линзой. В случае *б* предмет помещается на расстоянии d_0 от линзы, а в случае *в* он размещается за линзой на расстоянии d от фокальной плоскости. Дополнительные сведения относительно линз как элементов, выполняющих преобразование Фурье, можно найти в работах [1—3].

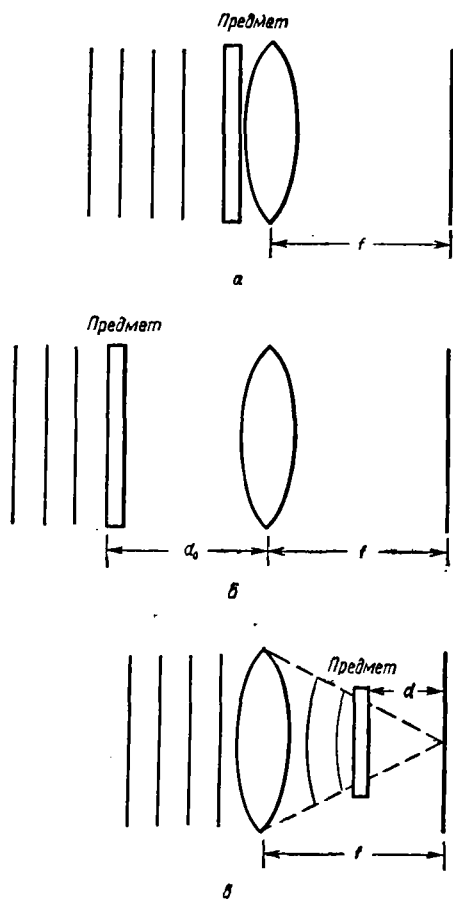
Предмет, расположенный вплотную к линзе

Пусть плоский предмет с амплитудным коэффициентом пропускания $t_0(x, y)$ расположен непосредственно перед собирающей линзой с фокусным расстоянием f , как показано на фиг. 30, *a*. Предполагается, что предмет равномерно освещен нормально падающей монохроматической плоской волной с амплитудой A ; при этом амплитуда волны, падающей на линзу, равна

$$U_i(x, y) = At_0(x, y). \quad (5.11)$$

Конечный размер апертуры линзы можно учесть, связывая с линзой *функцию зрачка* $P(x, y)$, которая

¹⁾ Передней и задней фокальными плоскостями линзы называются плоскости, перпендикулярные к оси линзы и расположенные соответственно на расстоянии $|f|$ перед линзой и позади нее.



Фиг. 30. Оптические схемы, осуществляющие преобразование Фурье.

а — предмет расположен непосредственно перед линзой; б — предмет расположен на некотором расстоянии перед линзой; в — предмет расположен позади линзы.

определяется следующим образом:

$$P(x, y) = \begin{cases} 1 & \text{внутри апертуры линзы,} \\ 0 & \text{вне апертуры линзы.} \end{cases}$$

Таким образом, с учетом (5.10) распределение амплитуды позади линзы принимает вид

$$U'_l(x, y) = U_l(x, y) P(x, y) \exp\left[-j \frac{k}{2f}(x^2 + y^2)\right]. \quad (5.12)$$

Постоянная фазовая задержка, связанная с преобразованием, осуществляемым линзой, опущена, так как она не оказывает существенного влияния на конечный результат.

Для нахождения распределения $U_f(x_f, y_f)$ амплитуды поля в задней фокальной плоскости линзы используется формула дифракции Френеля (4.10). Таким образом, полагая $z = f$, находим

$$U_f(x_f, y_f) = \frac{\exp\left[j \frac{k}{2f}(x_f^2 + y_f^2)\right]}{j\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U'_l(x, y) \times \\ \times \exp\left[j \frac{k}{2f}(x^2 + y^2)\right] \exp\left[-j \frac{2\pi}{\lambda f}(xx_f + yy_f)\right] dx dy, \quad (5.13)$$

где постоянный фазовый множитель опущен. Подставляя (5.12) в (5.13), мы видим, что квадратичные фазовые множители в подынтегральном выражении сокращаются. Тогда имеем

$$U_f(x_f, y_f) = \frac{\exp\left[j \frac{k}{2f}(x_f^2 + y_f^2)\right]}{j\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U_l(x, y) \times \\ \times P(x, y) \exp\left[-j \frac{2\pi}{\lambda f}(xx_f + yy_f)\right] dx dy. \quad (5.14)$$

Следовательно, распределение поля U_f пропорционально двумерному фурье-образу части исходного поля, вырезанной отверстием линзы. Если физический размер предмета меньше апертуры линзы, то множитель $P(x, y)$

можно не учитывать. В этом случае

$$U_f(x_f, y_f) = \frac{A \exp \left[j \frac{k}{2f} (x_f^2 + y_f^2) \right]}{j\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_0(x, y) \times \\ \times \exp \left[-j \frac{2\pi}{\lambda f} (xx_f + yy_f) \right] dx dy. \quad (5.15)$$

Очевидно, что амплитуда и фаза волны в точке с координатами (x_f, y_f) определяются амплитудой и фазой фурье-компонент предмета, имеющих частоты $(f_x = x_f/\lambda f, f_y = y_f/\lambda f)$.

Заметим, что соотношение между распределением амплитуды в плоскости предмета и в фокальной плоскости отличается от точного выражения для преобразования Фурье квадратичным фазовым множителем, который стоит перед интегралом. Хотя фазовое распределение в фокальной плоскости не совпадает с фазовым распределением в спектре предмета, разница между ними представляет простое фазовое искривление. В большинстве случаев нас интересует распределение *интенсивности* в фокальной плоскости. Измерение распределения интенсивности дает информацию о *спектре мощности* объекта; фазовое распределение при этом не играет роли. Таким образом,

$$I_f(x_f, y_f) = \\ = \frac{A^2}{\lambda^2 f^2} \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_0(x, y) \exp \left[-j \frac{2\pi}{\lambda f} (xx_f + yy_f) \right] dx dy \right|^2. \quad (5.16)$$

Предмет, расположенный перед линзой

Рассмотрим другое, более общее расположение предмета и линзы, изображенное на фиг. 30, б. Предмет, помещенный на расстоянии d_0 перед линзой, освещается нормально падающей плоской волной амплитуды A . Амплитудный коэффициент пропускания предмета снова обозначим через t_0 . Кроме того, пусть $F_0(f_x, f_y)$ представляет спектр Фурье света, пропущенного предметом, а $F_l(f_x, f_y)$ — спектр Фурье света, падающего на линзу,

т. е.

$$\mathbf{F}_0(f_x, f_y) = \mathcal{F}\{A t_0\}, \quad \mathbf{F}_l(f_x, f_y) = \mathcal{F}\{U_l\}.$$

Если считать, что в том случае, когда свет распространяется на расстояние d_0 , приближения Френеля справедливы, то \mathbf{F}_0 и \mathbf{F}_l можно связать при помощи (4.11) следующим образом:

$$\mathbf{F}_l(f_x, f_y) = \mathbf{F}_0(f_x, f_y) \exp[-j\pi\lambda d_0 (f_x^2 + f_y^2)], \quad (5.17)$$

где постоянный фазовый сдвиг не учитывается.

Не будем пока учитывать конечность апертуры линзы. Таким образом, полагая $P = 1$, перепишем выражение (5.14) в виде

$$U_f(x_f, y_f) = \frac{\exp\left[j\frac{k}{2f}(x_f^2 + y_f^2)\right]}{j\lambda f} \mathbf{F}_l\left(\frac{x_f}{\lambda f}, \frac{y_f}{\lambda f}\right). \quad (5.18)$$

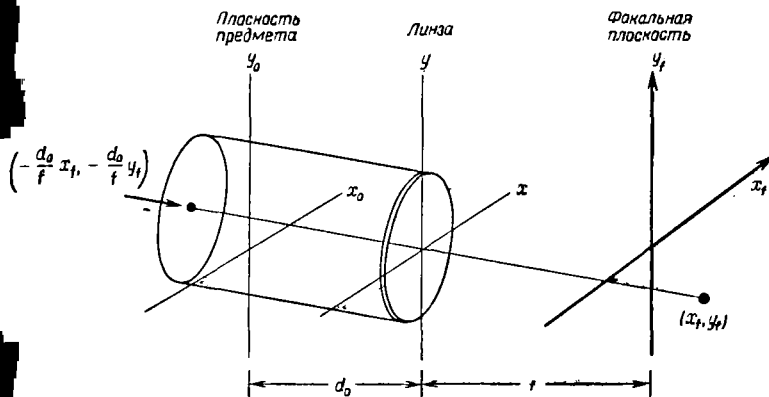
Подставляя (5.17) в (5.18), получаем

$$U_f(x_f, y_f) = \frac{\exp\left[j\frac{k}{2f}\left(1 - \frac{d_0}{f}\right)(x_f^2 + y_f^2)\right]}{j\lambda f} \mathbf{F}_0\left(\frac{x_f}{\lambda f}, \frac{y_f}{\lambda f}\right),$$

$$U_f(x_f, y_f) = \frac{A \exp\left[j\frac{k}{2f}\left(1 - \frac{d_0}{f}\right)(x_f^2 + y_f^2)\right]}{j\lambda f} \times$$

$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_0(x_0, y_0) \exp\left[-j\frac{2\pi}{\lambda f}(x_0 x_f + y_0 y_f)\right] dx_0 dy_0. \quad (5.19)$$

Следовательно, амплитуда и фаза волны в точке с координатами (x_f, y_f) опять связаны с амплитудой и фазой спектра предмета на частотах $(x_f/\lambda f, y_f/\lambda f)$. Заметим, что перед интегралом преобразования опять стоит фазовый множитель, но в частном случае $d_0 = f$ показатель экспоненты обращается в нуль. *Очевидно, когда предмет расположен в передней фокальной плоскости линзы, фазовое искривление исчезает. При этом мы получаем соотношение, точно совпадающее с преобразованием Фурье!*



Фиг. 31. Вильетирирование предмета.

В предыдущих вычислениях мы совершенно не учитывали конечность размеров апертуры линзы. Чтобы учесть влияние апертуры линзы, воспользуемся приближением геометрической оптики. Такое приближение справедливо, если расстояние d_0 достаточно мало, чтобы для данной апертуры линзы расположение предмета соответствовало предельному случаю дифракции Френеля. В большинстве задач, представляющих интерес, это условие удовлетворяется. Амплитуда в точке с координатами (x_f, y_f) , показанной на фиг. 31, равна сумме амплитуд всех лучей с направляющими косинусами $\alpha \approx x_f/f$, $\beta \approx y_f/f$. Однако только часть таких лучей собирается апертурой линзы. Таким образом, конечный размер апертуры линзы можно учесть, проектируя эту апертуру опять в плоскость предмета; при этом центр проекции располагается на линии, соединяющей точку (x_f, y_f) с центром линзы (см. фиг. 31)¹⁾. Как и прежде, апертура линзы ограничивает эффективный размер предмета, но на этот раз та часть t_0 , которая определяет U_f , зависит от координат (x_f, y_f) . Как следует из фиг. 31,

¹⁾ Другими словами, при достаточно малом d_0 основной вклад в амплитуду вырезаемых апертурой линзы частей плоских волн, по которым складывается объект, дают точки объекта, лежащие в области, определяемой геометрической оптикой. — Прим. ред.

величину U_f в точке (x_f, y_f) можно определить путем преобразования Фурье части предмета с центром в точке $[x_0 = -(d_0/f)x_f, y_0 = -(d_0/f)y_f]$, которая вырезается функцией зрачка P . Таким образом,

$$U_f(x_f, y_f) = \frac{A \exp \left[j \frac{k}{2f} \left(1 - \frac{d_0}{f} \right) (x_f^2 + y_f^2) \right]}{j\lambda f} \times \\ \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_0(x_0, y_0) P \left(x_0 + \frac{d_0}{f} x_f, y_0 + \frac{d_0}{f} y_f \right) \times \\ \times \exp \left[-j \frac{2\pi}{\lambda f} (x_0 x_f + y_0 y_f) \right] dx_0 dy_0. \quad (5.20)$$

Ограничение эффективного размера предмета, обусловленное конечностью апертуры линзы, называется *виньетированием*. Заметим, что виньетирование предмета минимально в том случае, когда предмет расположен вблизи линзы и когда апертура линзы значительно больше предмета. На практике, чтобы уменьшить виньетирование, часто помещают предмет непосредственно перед линзой, хотя при анализе обычно удобно помещать предмет в передней фокальной плоскости, где преобразование Фурье выполняется точно.

Предмет, расположенный за линзой

Наконец, рассмотрим случай, когда предмет находится за линзой, как показано на фиг. 30, в. Предмет снова имеет коэффициент пропускания, равный t_0 , но теперь он расположен на расстоянии d от фокальной плоскости линзы. Пусть линза освещена нормально падающей плоской волной с амплитудой A . Тогда на предмет падает сферическая волна, сходящаяся в задней фокальной плоскости линзы.

В приближении геометрической оптики амплитуда сферической волны, падающей на предмет, равна Af/d ; освещенная область предмета ограничена линией пересечения сходящегося конуса лучей с плоскостью предмета. Если диаметр линзы равен L , то в месте расположения предмета освещается круговая область диаметром

Idf. Конечный размер освещенной области можно представить математически, проектируя функцию зрачка линзы на предмет конусом лучей; при этом мы получаем эффективную функцию зрачка $P[x_0(f/d), y_0(f/d)]$ в плоскости предмета. Заметим, что объект с коэффициентом пропускания t_0 сам имеет конечную апертуру; эффективная апертура в плоскости предмета определится, таким образом, пересечением истинной апертуры предмета с проекцией функции зрачка. Если апертура предмета полностью освещена, то проекция функции зрачка выпадает из рассмотрения.

Используя параксиальное приближение для сферической волны, освещающей предмет, амплитуду поля, прошедшего предметом, можно записать в виде

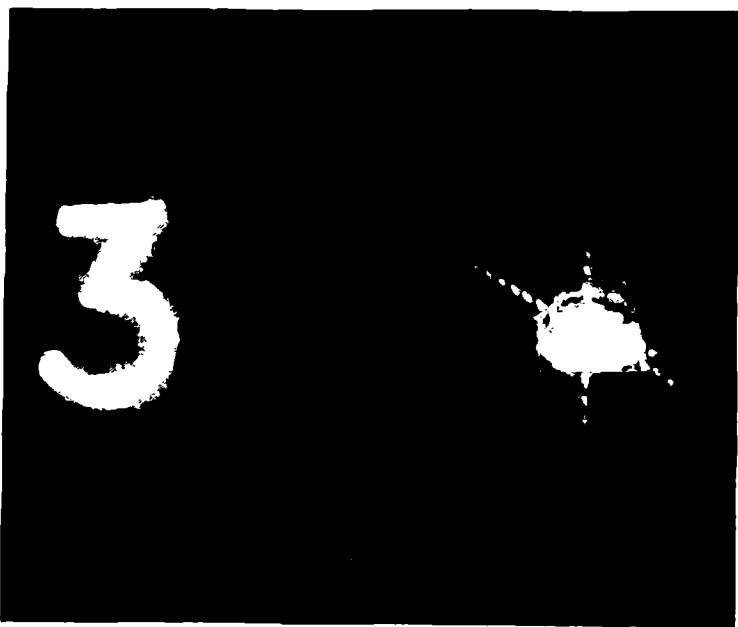
$$U_0(x_0, y_0) = \left\{ \frac{Af}{d} P\left(x_0 \frac{f}{d}, y_0 \frac{f}{d}\right) \times \right. \\ \left. \times \exp\left[-j \frac{k}{2d}(x_0^2 + y_0^2)\right] \right\} t_0(x_0, y_0). \quad (5.21)$$

Предполагая, что дифракция от предмета, наблюдаемая в фокальной плоскости, есть дифракция Френеля, и используя (4.10), можно найти

$$U_f(x_f, y_f) = \frac{A \exp\left[j \frac{k}{2d}(x_f^2 + y_f^2)\right]}{j\lambda d} \frac{f}{d} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t_0(x_0, y_0) \times \\ \times P\left(x_0 \frac{f}{d}, y_0 \frac{f}{d}\right) \exp\left[-j \frac{2\pi}{\lambda d}(x_0 x_f + y_0 y_f)\right] dx_0 dy_0. \quad (5.22)$$

Таким образом, с точностью до квадратичного фазового множителя распределение амплитуды в фокальной плоскости представляет собой фурье-образ части предмета, ограниченный проекцией апертуры линзы.

Результат (5.22) фактически совпадает с результатом, полученным для случая расположения предмета непосредственно перед линзой. Однако при таком расположении достигается чрезвычайная гибкость эксперимента, а именно масштаб фурье-образа находится под контролем экспериментатора. При увеличении d пространственный масштаб образа увеличивается по крайней мере до тех пор, пока предмет не окажется



Фиг. 32. Фурье-спектр образа цифры «3», полученный оптическим способом.

непосредственно вблизи линзы ($d=f$). При уменьшении d масштаб образа уменьшается. Такую гибкость можно широко использовать при осуществлении пространственной фильтрации (см. гл. 7), где некоторая возможность регулирования масштаба преобразования часто оказывается очень полезной.

Наконец, проиллюстрируем на типичном примере вид двумерного анализа Фурье, который можно легко выполнить оптическим способом. На фиг. 32 слева изображена прозрачная цифра «3», которая, будучи помещена перед положительной линзой, при когерентном освещении дает в задней фокальной плоскости линзы картину (фурье-спектр), изображенную справа. На фотографии хорошо видны высокочастотные составляющие, соответствующие прямолинейным участкам цифры.

§ 3. ФОРМИРОВАНИЕ ИЗОБРАЖЕНИЯ. МОНОХРОМАТИЧЕСКОЕ ОСВЕЩЕНИЕ

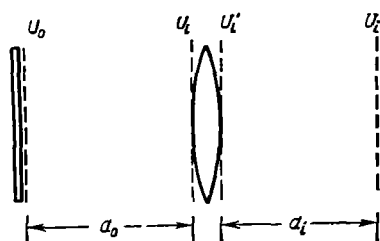
Конечно, наиболее известным свойством линз является их способность формировать изображение. Если предмет помещен перед линзой и освещен, то при определенных условиях в другой плоскости возникает распределение интенсивности, которое очень напоминает предмет. Это распределение называется *изображением* предмета. Изображение может быть *действительным* в том смысле, что в плоскости за линзой возникает действительное распределение, и *мнимым* в том смысле, что свет за линзой кажется исходящим из новой плоскости, расположенной перед линзой.

Сначала обсудим формирование изображения при некоторых ограничивающих предположениях. Во-первых, будем рассматривать свободную от аберраций положительную тонкую линзу, которая формирует действительное изображение. Во-вторых, нас будет интересовать только *монохроматическое* освещение. Это ограничение означает, что система, создающая изображение, линейна относительно комплексной амплитуды поля. Оба ограничения будут сняты в гл. 6, где задача о формировании изображения будет рассматриваться при помощи гораздо более общего метода.

Импульсный отклик положительной линзы

Предположим, что плоский предмет, находящийся на расстоянии d_0 перед положительной линзой, как показано на фиг. 33, освещен монохроматическим светом. Обозначим комплексное поле непосредственно за предметом через $U_0(x_0, y_0)$. Распределение поля, которое возникает на расстоянии d_i за линзой, обозначим через $U_i(x_i, y_i)$. Наша задача — определить условия, при которых распределение поля U_i можно с уверенностью назвать «изображением» распределения поля в плоскости предмета U_0 .

Ввиду линейности явления распространения волн поле U_i всегда можно представить в виде интеграла



Фиг. 33. Схема получения изображения.

суперпозиции

$$U_i(x_i, y_i) = \iint_{-\infty}^{\infty} h(x_i, y_i; x_0, y_0) U_0(x_0, y_0) dx_0 dy_0, \quad (5.23)$$

где $h(x_i, y_i; x_0, y_0)$ — амплитуда в точке (x_i, y_i) поля, созданного точечным источником единичной амплитуды, расположенным в точке (x_0, y_0) предмета. Таким образом, свойства системы, создающей изображение, будут полностью описаны, если можно определить импульсный отклик h .

Чтобы оптическая система давала высококачественное изображение, поле U_i должно как можно меньше отличаться от U_0 . Это означает, что импульсный отклик должен приближенно походить на δ -функцию, т. е.

$$h(x_i, y_i; x_0, y_0) \approx K \delta(x_i \pm Mx_0, y_i \pm My_0), \quad (5.24)$$

где K — комплексная постоянная, M — увеличение системы, а знак плюс или минус учитывает возможность как прямого, так и обратного изображения. Поэтому «плоскостью изображения» мы будем называть ту плоскость, где (5.24) выполняется лучше всего.

Для определения h допустим, что предмет представлен δ -функцией (точечный источник) в точке с координатами (x_0, y_0) . В этом случае на линзу будет попадать сферическая волна, расходящаяся из точки (x_0, y_0) . В параксиальном приближении эту волну можно записать в виде

$$U_i(x, y) = \frac{1}{j\lambda d_0} \exp \left\{ j \frac{k}{2d_0} [(x - x_0)^2 + (y - y_0)^2] \right\}. \quad (5.25)$$

После прохождения через линзу распределение поля принимает вид

$$U'_i(x, y) = U_i(x, y) P(x, y) \exp\left[-j \frac{k}{2f}(x^2 + y^2)\right]. \quad (5.26)$$

Наконец, предполагая, что при распространении света на расстояние d_i справедливо соотношение (4.9), полученное для случая дифракции Френеля, можно написать

$$h(x_i, y_i; x_0, y_0) = \frac{1}{j\lambda d_i} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U'_i(x, y) \times \\ \times \exp\left\{j \frac{k}{2d_i} [(x_i - x)^2 + (y_i - y)^2]\right\} dx dy, \quad (5.27)$$

где опущен постоянный фазовый множитель. Комбинируя (5.25), (5.26) и (5.27), получаем несколько громоздкое окончательное выражение

$$h(x_i, y_i; x_0, y_0) = \frac{1}{\lambda^2 d_0 d_i} \exp\left[j \frac{k}{2d_i} (x_i^2 + y_i^2)\right] \times \\ \times \exp\left[j \frac{k}{2d_0} (x_0^2 + y_0^2)\right] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(x, y) \exp\left[j \frac{k}{2} \left(\frac{1}{d_0} + \frac{1}{d_i} - \frac{1}{f}\right) \times \right. \\ \left. \times (x^2 + y^2)\right] \exp\left[-jk \left[\left(\frac{x_0}{d_0} + \frac{x_i}{d_i}\right) x + \left(\frac{y_0}{d_0} + \frac{y_i}{d_i}\right) y\right]\right] dx dy. \quad (5.28)$$

Соотношения (5.23) и (5.28) являются формальным решением, определяющим зависимость между U_0 и U_i . Однако без дальнейших упрощений трудно определить условия, при которых распределение U_i можно с уверенностью назвать изображением распределения U_0 .

Устранение квадратичных фазовых множителей. Формула линзы

Самые неприятные члены в приведенном выше выражении для импульсного отклика — это члены, содержащие квадратичные фазовые множители. Заметим, что два из них, а именно

$$\exp\left[j \frac{k}{2d_i} (x_i^2 + y_i^2)\right] \quad \text{и} \quad \exp\left[j \frac{k}{2d_0} (x_0^2 + y_0^2)\right],$$

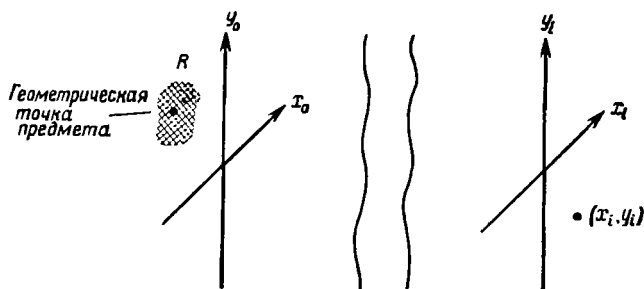
не зависят от координат линзы (x, y) . Эти члены определяют фазовое искривление в плоскостях $x_i y_i$ и $x_0 y_0$. Если бы мы решили рассматривать формирование изображения между двумя сферическими поверхностями, а не между двумя плоскостями, эти члены можно было бы исключить. Однако можно показать, что и в случае формирования изображения между двумя плоскостями оба эти члена несущественны.

Опуская множитель $\exp [j (k/2d_i)(x_i^2 + y_i^2)]$, заметим, что в подавляющем большинстве представляющих интерес случаев конечной целью задачи формирования изображения является получение некоторого распределения света, которое будет воспринято детектором, реагирующим только на интенсивность (например, фотопленкой). Так как рассматриваемый член изменяет только распределение фазы, он никак не будет влиять на результаты измерения интенсивности и, следовательно, может быть опущен.

К сожалению, от фазового множителя $\exp [j (k/2d_0)(x_0^2 + y_0^2)]$ не удается освободиться столь же просто, поскольку он зависит от переменных интегрирования (x_0, y_0) интеграла суперпозиции¹⁾. Однако в большинстве случаев, представляющих интерес, от него тоже можно избавиться. Если система, создающая изображение, ведет себя приблизительно так же, как идеальная система, для которой справедливо соотношение (5.24), то амплитуда волны в точке с координатами (x_i, y_i) будет определяться вкладом только очень малой области в пространстве предмета с центром в точке, соответствующей идеальному геометрическому изображению (фиг. 34). Если внутри этой малой области аргумент $\exp [j (k/2d_0)(x_0^2 + y_0^2)]$ изменяется не более чем на долю радиана, то можно использовать приближение

$$\exp \left[j \frac{k}{2d_0} (x_0^2 + y_0^2) \right] \approx \exp \left[j \frac{k}{2d_0} \left(\frac{x_i^2 + y_i^2}{M^2} \right) \right]. \quad (5.29)$$

¹⁾ Для случая некогерентного освещения можно показать, что оптическая система с импульсным откликом $|\mathfrak{h}|^2$ линейна в отношении интенсивности. В этом случае оба множителя можно опустить. Подробности см. в гл. 6.



Фиг. 34. Область R , в которой функция h для точки с координатами (x_i, y_i) имеет значительную величину.

Теперь экспоненциальный член можно опустить, так как он не зависит от (x_0, y_0) и, следовательно, не влияет на результат измерения интенсивности в плоскости $x_i y_i$. Это приближение и условия его применимости более подробно рассмотрены в задаче 5.7.

Воспользовавшись приведенными соображениями, перепишем выражение для импульсного отклика в виде

$$\begin{aligned}
 h(x_i, y_i; x_0, y_0) &\approx \\
 &\approx \frac{1}{\lambda^2 d_0 d_i} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(x, y) \exp \left[j \frac{k}{2} \left(\frac{1}{d_0} + \frac{1}{d_i} - \frac{1}{f} \right) (x^2 + y^2) \right] \times \\
 &\times \exp \left\{ -jk \left[\left(\frac{x_0}{d_0} + \frac{x_i}{d_i} \right) x + \left(\frac{y_0}{d_0} + \frac{y_i}{d_i} \right) y \right] \right\} dx dy. \quad (5.30)
 \end{aligned}$$

Чтобы получить совсем простой результат, рассмотрим случай, когда плоскость наблюдения расположена на таком расстоянии d_i от линзы, что удовлетворяется соотношение

$$\frac{1}{d_0} + \frac{1}{d_i} - \frac{1}{f} = 0. \quad (5.31)$$

Это соотношение известно из геометрической оптики, где оно называется *формулой линзы*. Соотношение (5.31) определяет расположенную за линзой точку, в которой пересекаются лучи, исходящие из одной точки предмета (точка изображения). В приближении геометрической

оптики выполнение формулы линзы означает, что импульсный отклик системы достаточно близок к идеальному. Предположение о выполнении формулы линзы позволяет свести импульсный отклик к виду

$$h(x_i, y_i; x_0, y_0) \approx \frac{1}{\lambda^2 d_0 d_i} \iint_{-\infty}^{\infty} P(x, y) \times \\ \times \exp \left\{ -jk \left[\left(\frac{x_0}{d_0} + \frac{x_i}{d_i} \right) x + \left(\frac{y_0}{d_0} + \frac{y_i}{d_i} \right) y \right] \right\} dx dy. \quad (5.32)$$

Определяя увеличение системы как

$$M = \frac{d_i}{d_0}, \quad (5.33)$$

находим последний упрощенный вид импульсного отклика

$$h(x_i, y_i; x_0, y_0) \approx \frac{1}{\lambda^2 d_0 d_i} \iint_{-\infty}^{\infty} P(x, y) \times \\ \times \exp \left\{ -j \frac{2\pi}{\lambda d_i} [(x_i + Mx_0)x + (y_i + My_0)y] \right\} dx dy. \quad (5.34)$$

Таким образом, если формула линзы справедлива, то импульсный отклик соответствует картине дифракции Фраунгофера на апертуре линзы, причем центр картины находится в точке изображения ($x_i = -Mx_0$, $y_i = -My_0$). Появление формулы дифракции Фраунгофера не должно вызывать удивления. Выбирая расстояние d_i в соответствии с формулой линзы, мы исследуем плоскость, в которой сходится сферическая волна, прошедшая через линзу. Из результатов задачи 4.3 следует, что распределение света около точки, в которой сходится волна, будет точно представлять картину дифракции Фраунгофера на апертуре линзы, ограничивающей протяженность волнового фронта.

Соотношение между предметом и изображением

Рассмотрим сначала свойства изображения, предсказываемые геометрической оптикой. Чтобы найти это идеальное изображение, допустим, что длина волны λ стре-

мится к нулю. В этом случае дифракционные эффекты становятся несущественными. Производя замену переменных

$$\tilde{x} = \frac{x}{\lambda d_i}, \quad \tilde{y} = \frac{y}{\lambda d_i}, \quad (5.35)$$

выражение для импульсного отклика (5.34) можно переписать в виде

$$\begin{aligned} h(x_i, y_i; x_0, y_0) \approx M \int_{-\infty}^{\infty} \int P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y}) \times \\ \times \exp\{-j2\pi[(x_i + Mx_0)\tilde{x} + (y_i + My_0)\tilde{y}]\} d\tilde{x} d\tilde{y}. \end{aligned} \quad (5.36)$$

Так как λ стремится к нулю, область значений (\tilde{x}, \tilde{y}) , где функция зрачка P равна единице, будет безгранично увеличиваться, что дает возможность заменить P единицей, оставив те же пределы интегрирования. Таким образом,

$$\begin{aligned} h(x_i, y_i; x_0, y_0) \rightarrow \\ \rightarrow M \int_{-\infty}^{\infty} \int \exp\{-j2\pi[(x_i + Mx_0)\tilde{x} + (y_i + My_0)\tilde{y}]\} d\tilde{x} d\tilde{y} = \\ = M\delta(x_i + Mx_0, y_i + My_0) = \frac{1}{M} \delta\left(\frac{x_i}{M} + x_0, \frac{y_i}{M} + y_0\right). \end{aligned} \quad (5.37)$$

Подставляя этот результат в интеграл суперпозиции (5.23), получаем соотношение, связывающее распределения амплитуды в точках предмета и в точках изображения

$$U_i(x_i, y_i) = \frac{1}{M} U_0\left(-\frac{x_i}{M}, -\frac{y_i}{M}\right). \quad (5.38)$$

Отсюда следует, что изображение, получаемое в приближении геометрической оптики, представляет собой точную копию предмета, увеличенную и перевернутую в плоскости изображения.

Выводы геометрической оптики, конечно, приближены. Более точное представление о соотношении между предметом и изображением можно получить только при учете дифракционных эффектов. Чтобы найти такое соотношение, вернемся к выражению (5.36) для

импульсного отклика и произведем следующую дополнительную замену переменных:

$$\tilde{x}_0 = -Mx_0, \quad \tilde{y}_0 = -My_0. \quad (5.39)$$

Импульсный отклик в этом случае будет равен

$$\begin{aligned} h(x_i, y_i; \tilde{x}_0, \tilde{y}_0) = M \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y}) \times \\ \times \exp\{-j2\pi[(x_i - \tilde{x}_0)\tilde{x} + (y_i - \tilde{y}_0)\tilde{y}]\} d\tilde{x} d\tilde{y}. \end{aligned} \quad (5.40)$$

Заметим, что h теперь *пространственно-инвариантная* величина, зависящая только от разности координат $(x_i - \tilde{x}_0, y_i - \tilde{y}_0)$. С введением еще одного определения

$$\tilde{h} = \frac{1}{M} h \quad (5.41)$$

интеграл суперпозиции (5.23) принимает вид

$$\begin{aligned} U_i(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0) \times \\ \times \left[\frac{1}{M} U_0\left(-\frac{\tilde{x}_0}{M}, -\frac{\tilde{y}_0}{M}\right) \right] d\tilde{x}_0 d\tilde{y}_0. \end{aligned} \quad (5.42)$$

В этом выражении мы узнаем *свертку* импульсного отклика \tilde{h} и идеального изображения. Для удобства определим новую функцию

$$U_g(x_i, y_i) = \frac{1}{M} U_0\left(-\frac{x_i}{M}, -\frac{y_i}{M}\right). \quad (5.43)$$

Свертку (5.42) тогда можно переписать в упрощенном виде:

$$U_i(x_i, y_i) = \tilde{h}(x_i, y_i) * U_g(x_i, y_i), \quad (5.44)$$

где

$$\begin{aligned} \tilde{h}(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y}) \times \\ \times \exp[-j2\pi(x_i \tilde{x} + y_i \tilde{y})] d\tilde{x} d\tilde{y}. \end{aligned} \quad (5.45)$$

Соотношения (5.44) и (5.45) представляют собой конечный результат настоящего анализа. Они показывают,

что при учете дифракционных эффектов изображение нельзя больше считать точной копией предмета. Полученное изображение дает несколько сглаженный облик предмета, что является следствием неравенства нулю ширины импульсного отклика \tilde{h} . Это сглаживание может привести к значительному ослаблению мелких деталей предмета и соответственно к потере точности воспроизведения изображения. Точно такое же явление можно наблюдать в случае, когда электрический сигнал проходит через линейную электрическую схему. Если длительность импульсного отклика схемы велика по сравнению с «временем пульсаций» входного сигнала, то схема будет сглаживать входной сигнал. Таким образом, быстрые изменения входного сигнала не будут воспроизводиться на выходе.

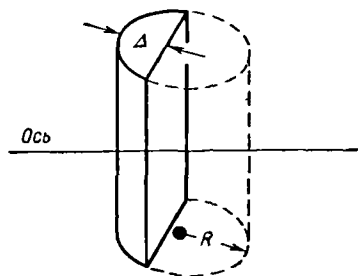
В случае инвариантных во времени электрических схем потерю точности воспроизведения сигнала часто гораздо удобнее описывать в частотной области. Если входной сигнал содержит компоненты со слишком высокой частотой, которые не могут быть пропущены данной схемой, то сигнал на выходе будет воспроизводиться неточно. Большое удобство частотного подхода в случае электрических схем наводит на мысль о том, что подобные представления полезно было бы применить при изучении оптических систем. Применение понятия фильтрации к системам, создающим изображение, действительно имеет огромную важность. Этому будет посвящена следующая глава.

ЗАДАЧИ

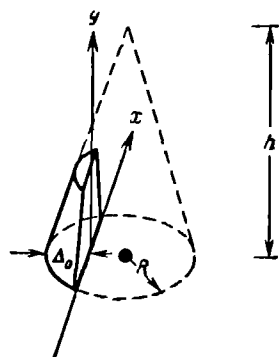
5.1. Показать, что фокусные расстояния двояковыпуклой и плосковыпуклой линз и положительного мениска всегда положительны, а фокусные расстояния двояковогнутой и плоскогнутой линз и отрицательного мениска всегда отрицательны.

5.2. а) Найти параксиальное приближение для фазового преобразования, осуществляемого линзой, которая представляет собой часть цилиндра, как изображено на фиг. 35.

б) Какое влияние оказывает такая линза на плоскую волну, распространяющуюся вдоль оси?



Фиг. 35.



Фиг. 36.

5.3. Рассмотрим линзу, которая представляет собой часть конуса (фиг. 36).

а) Показать, что в параксиальном приближении фазовое преобразование, выполняемое такой линзой, имеет вид (линза предполагается тонкой)

$$t_l(x, y) = \exp \left\{ jk \left[n\Delta_0 - \frac{(n-1)Ry}{h} - \frac{x^2}{2f(y)} \right] \right\},$$

где

$$f(y) = \frac{R(1-y/h)}{n-1}.$$

б) Как влияет такая линза на плоскую волну, распространяющуюся в направлении, перпендикулярном плоскости xy ?

5.4. Предметная функция U_0 , ограниченная круглой апертурой диаметром l , задана в передней фокальной плоскости собирающей линзы диаметром L . Распределение интенсивности измеряется в задней фокальной плоскости линзы. Предполагается, что $L > l$.

а) Найти выражение для максимальной пространственной частоты, для которой измеренная интенсивность точно равна квадрату модуля спектра Фурье предмета.

б) Чему равна эта частота (период/мм), если $L = 4$ см, $l = 2$ см, фокусное расстояние $f = 50$ см, а $\lambda = 6 \cdot 10^{-7}$ м?

в) Каково значение частоты, выше которой измеренный спектр исчезает, несмотря на то что спектр предмета может иметь отличные от нуля составляющие и на более высоких частотах?

5.5. Амплитудный коэффициент пропускания экрана описывается функцией, обладающей осевой симметрией

$$t(r) = \left(\frac{1}{2} + \frac{1}{2} \cos \alpha r^2 \right) \text{circ} \left(\frac{r}{l} \right).$$

- а) Почему этот экран действует подобно линзе?
 б) Найти выражение для фокусного расстояния экрана.
 в) Какие особенности могут серьезно ограничить применение этого экрана в качестве системы, создающей изображение, особенно для полихроматических предметов?

5.6. Последовательность одномерных предметных функций можно записать как $U_0(x, y_k)$, где $y_1, y_2, \dots, y_k, \dots, y_N$ представляют собой N фиксированных y -координат. Желательно выполнить преобразование Фурье для всех N функций в направлении x ; при этом мы должны получить ряд преобразований

$$G_0(f_x, y_k) = \int_{-\infty}^{\infty} U_0(x, y_k) \exp(-j2\pi f_x x) dx.$$

Пренебрегая конечными размерами линзы и предмета и используя рассмотренные в этой главе свойства линз как элементов, выполняющих преобразование Фурье, показать, как можно осуществить такое преобразование при помощи:

- а) двух цилиндрических линз с разными фокусными расстояниями,
 б) цилиндрической и сферической линз с одинаковыми фокусными расстояниями.

Указание. Необходимо найти только $|G_0|^2$, поэтому фазовые множители можно опустить.

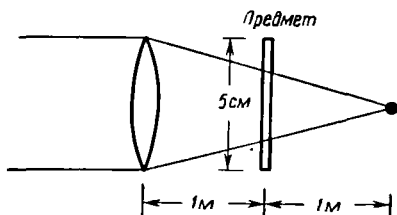
5.7. Учитывая приближение (5.29):

- а) определить, при каком радиусе r в плоскости предмета множитель $\exp[j(k/2d_0)(x_0^2 + y_0^2)]$ отличается от его значения в начале координат точно на π радиан;
 б) найти радиус первого нуля функции h (в плоскости предмета), считая, что круглый зрачок имеет радиус α , а начало координат находится в точке наблюдения в плоскости изображения;
 в) на основании этих результатов определить, при каком соотношении между α , λ и d_0 фазовый множитель $\exp[j(k/2d_0)(x_0^2 + y_0^2)]$ можно не учитывать, считая, что точка наблюдения находится вблизи оси линзы?

5.8. Монохроматическая плоская волна единичной амплитуды нормально падает на собирающую линзу, диаметр которой равен 5 см, а фокусное расстояние 2 м (фиг. 37). На расстоянии 1 м за линзой на ее оси помещен предмет, амплитудный коэффициент пропускания которого

$$t(x_0, y_0) = \frac{1}{2} (1 + \cos 2\pi f_0 x_0) \operatorname{rect}\left(\frac{x_0}{L}\right) \operatorname{rect}\left(\frac{y_0}{L}\right).$$

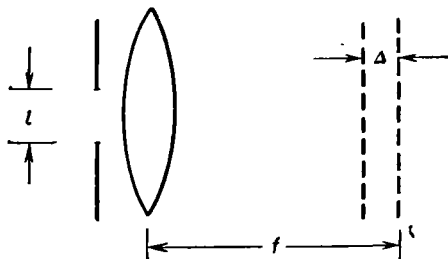
Считая $L = 1$ см, а $f_0 = 100$ период/см, изобразить распределение ин-



Фиг. 37.

тенсивности по оси x , фокальной плоскости, отмечая численные значения расстояний между дифракционными максимумами, а также ширину (между первыми нулями) отдельных максимумов.

5.9. Нормально падающая монохроматическая волна единичной амплитуды освещает предмет, максимальный линейный размер которого равен l . Предмет помещен непосредственно перед собирающей линзой с фокусным расстоянием f ; диаметр линзы превышает l . Из-за погрешности установки распределение интенсивности измеряется в плоскости, находящейся на расстоянии $f - \Delta$ за линзой (фиг. 38). Насколько малой должна быть ошибка Δ , если измеряемое распределение интенсивности должно точно представлять картину дифракции Фраунгофера на предмете?

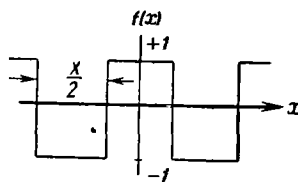


Фиг. 38.

5.10. Некоторый экран с амплитудным коэффициентом пропускания

$$t(r) = \left[\frac{1}{2} + \frac{1}{2} \operatorname{sgn}(\cos ar^2) \right] \operatorname{circ}\left(\frac{r}{l}\right)$$

нормально освещен монохроматической плоской волной единичной амплитуды. Показать, что такой экран действует как линза с множеством фокусных расстояний. Определить значения этих фокусных расстояний и относительные значения энергии, переносимой через фокус в соответствующие фокальные плоскости. Экран, взятый в этой задаче, известен под названием *зонной пластины Френеля*.



Ф и г. 39.

Указание. Прямоугольная волна, изображенная на фиг. 39, может быть представлена рядом Фурье

$$f(x) = \sum_{n=-\infty}^{\infty} \left[\frac{\sin(\pi n/2)}{\pi n} \right] \exp\left(j \frac{2\pi n x}{X}\right).$$

ЛИТЕРАТУРА

1. Rhodes J., Am. Journ. Phys., 21, 337 (1953).
Анализ и синтез оптических изображений.
2. Cutrona L. J., et al., IRE Trans. Inform. Theory, IT-6 386 (1960).
Оптические системы фильтрации и обработки информации.
3. Preston K., Jr., в книге Optical and Electrooptical Information Processing, Cambridge, 1965.
Использование способности линз осуществлять преобразование Фурье для анализа спектра сигналов.

ГЛАВА 6

ЧАСТОТНЫЙ АНАЛИЗ ОПТИЧЕСКИХ СИСТЕМ

В длинной и яркой истории оптики методы частотного анализа и теория линейных систем стали играть важную роль только сравнительно недавно. Однако за этот короткий период вышеупомянутые методы нашли столь широкое применение и оказались настолько плодотворными, что сейчас они занимают ведущее место в теории систем, образующих изображение.

Методы Фурье довольно стихийно проникли в анализ оптических систем в конце тридцатых годов нашего столетия, когда многие исследователи предложили применять синусоидальные тесты при расчете систем. Начал этому направлению положил французский ученый Дюоффе, самые значительные результаты работы которого изложены в книге, посвященной применению методов Фурье в оптике и опубликованной в 1946 г. [1]. К сожалению, эта книга до сих пор не переведена на английский язык и доступна не многим. В Соединенных Штатах большой интерес к этим работам возник благодаря деятельности инженера-электрика Отто Шейда, который очень успешно применил методы теории линейных систем и теории связи для анализа и усовершенствования линз телевизионных камер [2]. Однако в действительности основы оптики Фурье были заложены значительно раньше 1940 г., в частности в работах Эрнста Аббе (1840—1905) и лорда Релея (1842—1919).

В этой главе мы рассмотрим роль анализа Фурье в теории образования изображения при когерентном и некогерентном освещении. Случай некогерентного освещения является более общим; с ним мы сталкиваемся на всех этапах развития оптики. С другой стороны, когерентное освещение используется для формирования изображения в микроскопе, а в последнее время роль его еще больше возросла в связи с созданием лазеров.

Дальнейшее рассмотрение различных аспектов этого предмета можно найти в книгах О'Нейла [3], Франсона [4], Линфута [5] и Строука [6].

§ 1. ОБОБЩЕНИЕ ПРЕДСТАВЛЕНИЯ О СИСТЕМАХ. ОБРАЗУЮЩИХ ИЗОБРАЖЕНИЕ

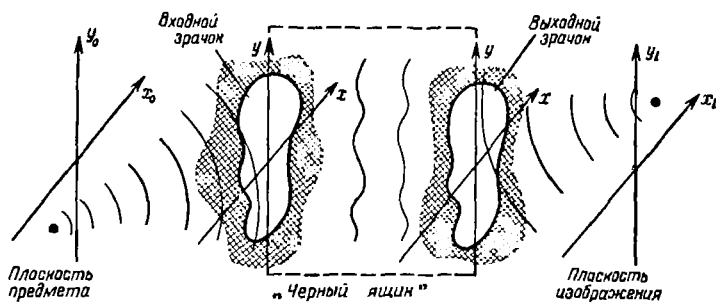
В предыдущей главе, рассматривая свойства линз как элементов, создающих изображение, мы ограничились частным случаем одной тонкой собирающей линзы и монохроматического освещения. Постараемся теперь дополнить свои рассуждения и получить результаты, применимые к более общим системам линз. Затем снимем ограничение относительно монохроматичности освещения, что позволит найти результаты для случая «квази-монохроматического» света, как пространственно когерентного, так и некогерентного.

Обобщенная модель

Предположим, что рассматриваемая оптическая система состоит не из одной линзы, а из нескольких линз, среди которых могут быть как положительные, так и отрицательные. Линзы могут и не быть «тонкими» в определенном ранее смысле. Будем предполагать, однако, что система в конечном счете дает *действительное* изображение, но фактически это не ограничение, так как если система дает мнимое изображение, то оно может быть преобразовано в итоге в действительное, например глазом. Значит, в подобном случае нам следует включить глаз в качестве конечного элемента в нашу систему.

При рассмотрении свойств системы линз будем считать, что все элементы, создающие изображение, помещены в один «черный ящик» и что основные свойства системы можно полностью описать, определяя только *конечные свойства* этого устройства.

Согласно фиг. 40, входом этого черного ящика служит *входной зрачок*, представляющий собой отверстие конечных размеров (эффективное или действительное), через которое должен проходить свет прежде, чем он достигнет элементов, создающих изображение, а



Фиг. 40. Обобщенная модель оптической системы.

выходом — *выходной зрачок* (также эффективный или действительный), представляющий собой отверстие конечных размеров, через которое свет проходит после создающих изображение элементов на пути к плоскости изображения. Обычно считают, что путь света между входной и выходной плоскостями может быть достаточно точно описан в приближении геометрической оптики. Таким образом, конечный размер обоих зрачков можно найти, строя по законам геометрической оптики проекцию наименьшей апертуры системы соответственно в пространстве предметов и пространстве изображений. Поскольку размеры получающихся зрачков определяются размерами изображения эффективного отверстия, существующего где-то внутри системы, они могут быть меньше действительных физических размеров отверстий в входной и выходной плоскостях. Заметим, что при таком определении входной зрачок всегда является изображением выходного зрачка и наоборот.

Оптическая система называется *дифракционно ограниченной*, если она преобразует ¹⁾ расходящуюся сферическую волну, исходящую из любого точечного источника, в новую идеальную сферическую волну, которая сходится в точке, лежащей в плоскости изображения. Таким образом, конечное свойство дифракционно ограниченной системы линз заключается в том, что она преобразует расходящуюся сферическую волну, падающую

¹⁾ В приближении геометрической оптики. — *Прим. ред.*

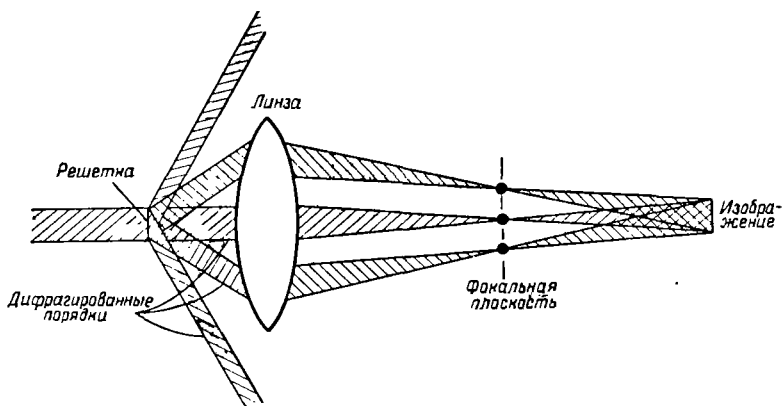
на входной зрачок, в сходящуюся сферическую волну, выходящую через выходной зрачок. Для любой реальной оптической системы это свойство выполняется в лучшем случае только для конечной области в плоскости предмета. Если рассматриваемый предмет не выходит за пределы этой области, систему можно отнести к дифракционно ограниченной.

Если в действительности фронт волны от точечного источника после выходного зрачка значительно отличается от идеальной сферической формы, то говорят, что оптическая система имеет *абберации*. Абберации будут подробно рассмотрены в § 4, где будет показано, что они приводят к образованию искажений в пространственно-частотном отклике системы, создающей изображение.

Дифракционные эффекты

Геометрическая оптика с достаточной точностью описывает прохождение света от входного зрачка к выходному, поэтому дифракционные эффекты играют заметную роль только на пути света от предмета к входному зрачку и от выходного зрачка к изображению. Действительно, *все* ограничения, налагаемые дифракцией, можно связать с *любым* из этих двух участков пути распространения света. Утверждения о том, что разрешение изображения ограничивается входным зрачком конечных размеров или выходным зрачком конечных размеров, полностью эквивалентны. Основная причина эквивалентности заключается в том, что один зрачок представляет собой просто изображение другого.

Представление о том, что обсуждаемые дифракционные эффекты обусловлены *входным* зрачком конечных размеров, было впервые введено Эрнстом Аббе [7] в 1873 г. Согласно теории Аббе, только определенная часть дифракционных максимумов, созданных сложным предметом, пропускается входным зрачком конечных размеров. Не пропускаются зрачком те максимумы, которые соответствуют высокочастотным составляющим предмета. Это положение иллюстрируется фиг. 41, где предметом служит простая решетка, а оптическая система состоит из одной положительной линзы.



Фиг. 41. Формирование изображения по Аббе.

В 1896 г. Релей [8] высказал другую, фактически эквивалентную точку зрения, в соответствии с которой дифракционные эффекты обусловлены *выходным* зрачком конечных размеров. Такой подход был использован в гл. 5, § 3, и мы будем придерживаться его и здесь.

Снова амплитуда в точках изображения представляется интегралом суперпозиции ¹⁾

$$U_i(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x_i, y_i; x_0, y_0) U_0(x_0, y_0) dx_0 dy_0, \quad (6.1)$$

где h — амплитуда в точке плоскости изображения с координатами (x_i, y_i) при действии точечного источника, расположенного в точке (x_0, y_0) . В отсутствие aberrаций отклик h создается сходящейся сферической волной (ограниченной протяженности), идущей от выходного зрачка к точке идеального изображения ($x_i = Mx_0, y_i = My_0$). В этом более общем случае увеличение M может быть отрицательным или положительным в зависимости от того, какое изображение наблюдается — прямое или перевернутое.

¹⁾ Здесь мы сохраняем предположение о монохроматичности освещения, но в следующем параграфе снимем это ограничение.

Из задачи 4.3 и гл. 5, § 3 следует, что распределение амплитуды света около точки идеального изображения представляет собой просто картину дифракции Фраунгофера на выходном зрачке, центр которой находится в точке с координатами (Mx_0, My_0) . Таким образом,

$$h(x_i, y_i; x_0, y_0) = K \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(x, y) \times \\ \times \exp\left\{-j \frac{2\pi}{\lambda d_i} [(x_i - Mx_0)x + (y_i - My_0)y]\right\} dx dy, \quad (6.2)$$

где K — комплексная постоянная, а функция зрачка P равна единице внутри зрачка и нулю за его пределами. При записи этого уравнения мы опять пренебрегаем квадратичными фазовыми множителями для плоскостей предмета и изображения, что было обосновано в гл. 5, § 3. После замены переменных

$$\tilde{x} = \frac{x}{\lambda d_i}, \quad \tilde{y} = \frac{y}{\lambda d_i}, \quad \tilde{x}_0 = Mx_0, \quad \tilde{y}_0 = My_0$$

получим

$$h(x_i - \tilde{x}_0, y_i - \tilde{y}_0) = K \lambda^2 d_i^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y}) \times \\ \times \exp\{-j 2\pi [(x_i - \tilde{x}_0) \tilde{x} + (y_i - \tilde{y}_0) \tilde{y}]\} d\tilde{x} d\tilde{y}. \quad (6.3)$$

Вводя определения

$$\tilde{h} \approx \frac{1}{K \lambda^2 d_i^2} h,$$

$$U_g(\tilde{x}_0, \tilde{y}_0) = K \frac{\lambda^2 d_i^2}{M^2} U_0\left(\frac{\tilde{x}_0}{M}, \frac{\tilde{y}_0}{M}\right),$$

интеграл суперпозиции (6.1) можно в конце концов записать в виде

$$U_I(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0) U_g(\tilde{x}_0, \tilde{y}_0) d\tilde{x}_0 d\tilde{y}_0, \quad (6.4)$$

где U_g — идеальное изображение в приближении геометрической оптики, а

$$\hat{h}(x_i, y_i) = \iint_{-\infty}^{\infty} P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y}) \exp[-j2\pi(x_i \tilde{x} + y_i \tilde{y})] d\tilde{x} d\tilde{y}. \quad (6.5)$$

Таким образом, в этом более общем случае изображение снова можно представить как свертку идеального изображения и импульсного отклика, который определяется выходным зрачком системы.

Немонохроматическое освещение. Когерентный и некогерентный случаи

При рассмотрении оптических систем мы до сих пор всегда вводили предположение о строгой монохроматичности освещения. Это очень жесткое ограничение, так как свет реальных источников (включая и лазеры) никогда не бывает идеально монохроматическим. Статистический характер изменений во времени амплитуды и фазы света может в действительности весьма радикально влиять на свойства системы, создающей изображение. Поэтому мы временно отклонимся от нашей основной темы и обсудим наиболее существенные эффекты, обусловленные немонхроматичностью.

Чтобы полностью разобраться в этом вопросе, следовало бы проделать длинный окольный путь и рассмотреть *теорию частичной когерентности*. Однако в нашем случае такой подход нецелесообразен. Поэтому мы подойдем к вопросу с двух точек зрения, одна из которых является чисто эвристической, а другая — более строгой, но недостаточно общей. Более полное обсуждение вопроса читатель найдет в книге Берана и Паррента [9].

В случае монохроматического освещения амплитуду поля было удобно изображать при помощи комплексного фазора U , который представлял собой функцию пространственных координат. В случае немонхроматического освещения это приближение можно обобщить и представлять поля при помощи *меняющихся во времени фазоров*, которые зависят как от времени, так и от пространственных координат. Для квазимонохроматичес-

кого света амплитуду и фазу меняющегося во времени фазора легко отождествить с огибающей и фазой оптической волны. Однако понятие меняющегося во времени фазора можно обобщить и на случай неквазимонохроматических волн (см., например, [9], гл. 2).

Рассмотрим свойства света, пропущенного или отраженного предметом, который освещается немонохроматической волной. Так как изменения амплитуды фазора во времени носят статистический характер, поле можно удовлетворительно описать только при помощи статистических понятий. Как мы уже видели, каждая точка предмета обуславливает импульсный отклик в плоскости изображения. Если изменение во времени амплитуды и фазы света в точке предмета носит случайный характер, то полная амплитуда и фаза импульсного отклика будут меняться соответствующим образом. Следовательно, статистические зависимости между амплитудами в разных точках предмета будут обуславливать статистические зависимости между соответствующими импульсными откликами в плоскости изображения. Наличие таких статистических зависимостей влияет главным образом на результат усреднения по времени, в итоге которого и получается конечное распределение интенсивности в изображении.

Здесь мы обсудим только два типа освещения предмета. Сначала рассмотрим такое освещение, при котором изменения амплитуды фазора во всех точках предмета происходят *одновременно*. Таким образом, абсолютные фазы поля в двух произвольных точках будут меняться одинаково, а их относительная разность фаз будет сохраняться. Такое освещение называется *пространственно когерентным*. Затем мы рассмотрим освещение предмета, обладающее противоположным свойством, которое состоит в том, что амплитуды фазора во всех точках в плоскости предмета изменяются статистически независимо или по законам, не связанным друг с другом. Такое освещение называется *пространственно некогерентным*. В дальнейшем мы будем называть эти типы освещения просто *когерентным* и *некогерентным*. В тех случаях, когда свет исходит из одной точки,

освещение всегда будет когерентным¹⁾. Наиболее известным примером источника такого света является лазер, хотя и более распространенные источники, например циркониевые дуговые лампы, могут давать когерентный свет, однако меньшей интенсивности, чем лазер, если их излучение пропустить сначала через малое отверстие. Некогерентный свет дают рассеивающие или протяженные источники, например газоразрядные трубки²⁾.

Если освещение предмета когерентно, различные импульсные отклики в плоскости изображения меняются одновременно и, следовательно, их комплексные амплитуды должны складываться. Таким образом, когерентная система образования изображения оказывается линейной в отношении комплексной амплитуды. Поэтому можно непосредственно использовать результаты анализа, проведенного выше для монохроматического освещения; необходимо лишь учитывать, что комплексная амплитуда U теперь представляет собой амплитуду инвариантного во времени фазора, который описывает *относительные* амплитуды и фазы света.

Если предмет освещается некогерентным светом, импульсные отклики в плоскости изображения изменяются по статистически независимым законам. Следовательно, должны складываться их мощности, или *интенсивности*. Так как интенсивность любого данного импульсного отклика пропорциональна интенсивности света в точке предмета, ответственной за его возникновение, система при некогерентном освещении линейно отображает *интенсивность*; при этом импульсный отклик пропорционален квадрату модуля когерентного импульсного отклика.

Приведенные выше соображения носят чисто эвристический характер и, кроме того, содержат определен-

¹⁾ Это условие является достаточным, но не необходимым. Например, в том случае, когда свет от точечного источника проходит через неподвижный рассеиватель, относительная фаза света в любых двух точках за рассеивателем остается независимой от времени. Следовательно, прошедший свет все еще пространственно когерентен, несмотря на то что он не исходит из единичного точечного источника.

²⁾ Возможны также другие определения когерентности в терминах квантовой теории поля (см., например, [10, 11]), но нам будет достаточно приведенных выше упрощенных понятий.

ные предположения и приближения. Поэтому мы обратимся к более строгому исследованию задачи. Прежде всего напомним, что в случае монохроматического освещения мы пришли к фазорному представлению, опускающую составляющую положительной частоты в спектре косинусоидального поля и приписывая комплексную амплитуду оставшейся составляющей отрицательной частоты. Чтобы обобщить это представление на случай немонохроматической волны $u(P, t)$, опустим все составляющие спектра Фурье с положительной частотой; при этом мы получим новую (комплексную) функцию $u_-(P, t)$. Если, далее, записать

$$u_-(P, t) = U(P, t) \exp(-j2\pi\bar{\nu}t),$$

где $\bar{\nu}$ — средняя частота возмущения, то комплексную функцию $U(P, t)$ можно рассматривать как меняющееся во времени фазорное представление возмущения $u(P, t)$.

Из решения задачи 3.5 следует, что если удовлетворяются два специальных условия, то фазор $U(P, t)$ подчиняется тем же самым законам распространения, что и не зависящий от времени фазор монохроматического возмущения частоты $\bar{\nu}$. Эти условия имеют вид

$$\frac{\Delta\nu}{\bar{\nu}} \ll 1 \quad (\text{предположение квазимонохроматичности})$$

и

$$\frac{1}{\Delta\nu} \gg \frac{d}{c},$$

где $\Delta\nu$ — ширина полосы спектра возмущения; d — наибольшая длина оптического пути, а c — скорость света. Из этих двух условий непосредственно следует, что с помощью соотношения (6.4) амплитуду фазора изображения можно представить в виде свертки

$$U_i(x_i, y_i; t) = \int \int_{-\infty}^{\infty} \tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0) U_g(\tilde{x}_0, \tilde{y}_0; t) d\tilde{x}_0 d\tilde{y}_0, \quad (6.6)$$

где U_g — амплитуда фазора идеального изображения, а \tilde{h} определяется соотношением (6.5), если λ заменить средней длиной волны $\bar{\lambda}$.

В конечном счете нас интересует именно *интенсивность* в плоскости изображения¹⁾

$$I_i(x_i, y_i) = \langle U_i(x_i, y_i; t) U_i^*(x_i, y_i; t) \rangle. \quad (6.7)$$

Чтобы определить величину интенсивности, подставим (6.6) в (6.7); изменяя порядок интегрирования и усредняя, получаем

$$I_i(x_i, y_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\tilde{x}_0 d\tilde{y}_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\tilde{x}'_0 d\tilde{y}'_0 \tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0) \times \\ \times \tilde{h}^*(x_i - \tilde{x}'_0, y_i - \tilde{y}'_0) \langle U_g(\tilde{x}_0, \tilde{y}_0; t) U_g^*(\tilde{x}'_0, \tilde{y}'_0; t) \rangle. \quad (6.8)$$

Так как статистические свойства U_g и U_0 одинаковы, то окончательная величина в угловых скобках зависит, как можно видеть, от *пространственной когерентности* освещения предмета.

Если освещение предмета абсолютно *когерентно*, амплитуды фазора в плоскости предмета отличаются только комплексными постоянными. Иначе говоря, можно записать

$$U_g(\tilde{x}_0, \tilde{y}_0; t) = U_g(\tilde{x}_0, \tilde{y}_0) \frac{U_g(0, 0; t)}{\langle |U_g(0, 0; t)|^2 \rangle^{1/2}}, \quad (6.9) \\ U_g(\tilde{x}'_0, \tilde{y}'_0; t) = U_g(\tilde{x}'_0, \tilde{y}'_0) \frac{U_g(0, 0; t)}{\langle |U_g(0, 0; t)|^2 \rangle^{1/2}},$$

где не зависящие от времени величины U_g представляют собой амплитуды фазора, *отнесенные* к меняющейся во времени амплитуде фазора в начале координат. Подставляя эти соотношения в (6.8) и проводя интегрирование по $\tilde{x}'_0, \tilde{y}'_0$, получаем

$$I_i(x_i, y_i) = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0) U_g(\tilde{x}_0, \tilde{y}_0) d\tilde{x}_0 d\tilde{y}_0 \right|^2. \quad (6.10)$$

Наконец, определяя $U_i(x_i, y_i)$ как амплитуду фазора $U_i(x_i, y_i; t)$, отнесенную к амплитуде U_g в начале ко-

¹⁾ Угловые скобки опять означают усреднение по времени.

ординат, находим, что когерентная оптическая система описывается уравнением

$$U_i(x_i, y_i) = \iint_{-\infty}^{\infty} \tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0) U_g(\tilde{x}_0, \tilde{y}_0) d\tilde{x}_0 d\tilde{y}_0. \quad (6.11)$$

Таким образом, мы приходим к выводу, что в случае когерентного освещения предмета система образования изображения линейна относительно *комплексной амплитуды*.

Если освещение предмета абсолютно *некогерентно*, амплитуды фазора в разных точках предмета меняются по статистически независимым законам. Это идеализированное представление можно описать соотношением

$$\begin{aligned} \langle U_g(\tilde{x}_0, \tilde{y}_0; t) U_g^*(\tilde{x}_0', \tilde{y}_0'; t) \rangle = \\ = \kappa I_g(\tilde{x}_0, \tilde{y}_0) \delta(\tilde{x}_0 - \tilde{x}_0', \tilde{y}_0 - \tilde{y}_0'), \end{aligned} \quad (6.12)$$

где κ — действительная постоянная. Однако, как можно показать, такое представление не вполне точно, ибо оно означает, что через предмет пройдут только затухающие волны. В действительности когерентность существует в пределах минимального расстояния порядка длины волны. (Подробности см. в работе [9], разд. 4.4.) Тем не менее в большинстве практических случаев применимо соотношение (6.12). При подстановке его в (6.8) получим

$$I_i(x_i, y_i) = \kappa \iint_{-\infty}^{\infty} |\tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0)|^2 I_g(\tilde{x}_0, \tilde{y}_0) d\tilde{x}_0 d\tilde{y}_0. \quad (6.13)$$

Таким образом, интенсивность изображения определяется как свертка идеальной интенсивности I_g и импульсного отклика $|\tilde{h}|^2$. Мы приходим к выводу, что в случае некогерентного освещения предмета оптическую систему следует рассматривать как линейное преобразование *интенсивности*. Кроме того, импульсный отклик пропорционален квадрату модуля импульсного отклика, получаемого при когерентном освещении.

§ 2. ЧАСТОТНЫЙ ОТКЛИК ДИФРАКЦИОННО ОГРАНИЧЕННОЙ КОГЕРЕНТНОЙ ОПТИЧЕСКОЙ СИСТЕМЫ

Обратимся теперь к основной теме этой главы — частотному анализу оптических систем. Сначала рассмотрим только оптические системы с когерентным освещением. Системы, в которых используется некогерентное освещение, анализируются в § 3.

Как подчеркивалось ранее, при когерентном освещении оптическая система линейна относительно комплексной амплитуды поля. Это значит, что в такой системе интенсивность отображается совсем не линейно. Если использовать обычные представления частотного анализа, то следует говорить о линейном отображении амплитуды.

Передаточная функция при когерентном освещении

При анализе когерентных систем мы получили пространственно-инвариантную форму отображения амплитуды, которую можно представить с помощью свертки

$$U_i(x_i, y_i) = \iint_{-\infty}^{\infty} \tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0) U_g(\tilde{x}_0, \tilde{y}_0) d\tilde{x}_0 d\tilde{y}_0. \quad (6.14)$$

В таком случае следует ожидать, что к этой системе непосредственно применимо понятие передаточной функции. Чтобы это было действительно так, определим частотные спектры на входе и выходе системы следующим образом:

$$G_g(f_x, f_y) = \iint_{-\infty}^{\infty} U_g(\tilde{x}_0, \tilde{y}_0) \exp[-j2\pi(f_x \tilde{x}_0 + f_y \tilde{y}_0)] d\tilde{x}_0 d\tilde{y}_0, \quad (6.15)$$

$$G_i(f_x, f_y) = \iint_{-\infty}^{\infty} U_i(x_i, y_i) \exp[-j2\pi(f_x x_i + f_y y_i)] dx_i dy_i.$$

Кроме того, определим передаточную функцию как фурье-образ пространственно-инвариантного импульс-

ного отклика

$$\mathbf{H}(f_X, f_Y) = \iint_{-\infty}^{\infty} \tilde{\mathbf{h}}(x_i, y_i) \exp[-j2\pi(f_X x_i + f_Y y_i)] dx_i dy_i. \quad (6.16)$$

Теперь, применяя теорему свертки к (6.14), находим

$$\mathbf{G}_i(f_X, f_Y) = \mathbf{H}(f_X, f_Y) \mathbf{G}_g(f_X, f_Y). \quad (6.17)$$

Таким образом, мы выразили действие дифракционно ограниченной системы (по крайней мере формально) в частотной области. В дальнейшем функцию $\mathbf{H}(f_X, f_Y)$ будем называть *когерентной передаточной функцией*. Остается теперь связать \mathbf{H} непосредственно с физическими свойствами самой оптической системы.

С этой целью заметим, что, хотя соотношение (6.16) определяет \mathbf{H} как фурье-образ функции $\tilde{\mathbf{h}}$, эта последняя функция сама определяется преобразованием Фурье [ср. (6.5)]

$$\tilde{\mathbf{h}}(x_i, y_i) = \iint_{-\infty}^{\infty} P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y}) \exp[-j2\pi(x_i \tilde{x} + y_i \tilde{y})] d\tilde{x} d\tilde{y}.$$

Таким образом, мы приходим к выводу, что для дифракционно ограниченной системы

$$\mathbf{H}(f_X, f_Y) = \mathcal{F}\{\mathcal{F}\{P(\lambda d_i \tilde{x}, \lambda d_i \tilde{y})\}\} = P(-\lambda d_i f_X, -\lambda d_i f_Y). \quad (6.18)$$

Это крайне важное соотношение, так как оно дает информацию относительно поведения дифракционно ограниченных когерентных систем в частотной области. Так как функция зрачка P всегда равна или единице или нулю, то же самое справедливо и для когерентной передаточной функции. Это, естественно, означает, что в частотной области дифракционно ограниченная система имеет конечную полосу пропускания, внутри которой все частотные составляющие пропускаются без искажения амплитуды и фазы¹⁾. На границе этой полосы

¹⁾ Заметим, что этот вывод справедлив только для системы без аберраций. Как будет видно из § 4, система с аберрациями вносит искажения внутри полосы пропускания.

пропускания частотный отклик сразу падает до нуля, в силу чего частотные составляющие вне полосы пропускания полностью подавляются.

В заключение заметим, что когерентная передаточная функция пропорциональна *отраженной* функции зрачка; это обстоятельство выражается знаками минус в аргументе функции P . В дальнейшем можно устранить это неудобство обозначений, определив P в отраженной системе координат, а не в обычной системе, изображенной на фиг. 40. Таким образом, запишем

$$H(f_x^*, f_y) = P(\lambda d_i f_x, \lambda d_i f_y). \quad (6.19)$$

Различие между функциями (6.18) и (6.19) для нас несущественно, так как все рассматриваемые нами задачи симметричны относительно координат f_x и f_y .

Примеры передаточных функций при когерентном освещении для дифракционно ограниченных систем

Чтобы составить представление о частотном отклике дифракционно ограниченных систем при когерентном освещении, рассмотрим передаточные функции систем с квадратным (ширина l) и круглым (диаметр l) входными зрачками. Для этих двух случаев имеем соответственно

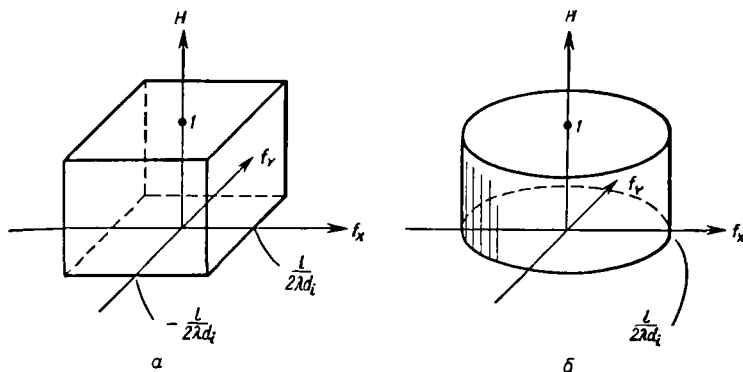
$$P(x, y) = \text{rect}\left(\frac{x}{l}\right) \text{rect}\left(\frac{y}{l}\right),$$

$$P(x, y) = \text{circ}\left(\frac{\sqrt{x^2 + y^2}}{l/2}\right).$$

Таким образом, в соответствии с (6.19) передаточные функции при когерентном освещении для этих двух случаев имеют вид

$$H(f_x, f_y) = \text{rect}\left(\frac{\lambda d_i f_x}{l}\right) \text{rect}\left(\frac{\lambda d_i f_y}{l}\right),$$

$$H(f_x, f_y) = \text{circ}\left(\frac{\sqrt{f_x^2 + f_y^2}}{l/2\lambda d_i}\right).$$



Фиг. 42. Передаточные функции для дифракционно ограниченных когерентных систем в случае квадратного (а) и круглого (б) выходного зрачка.

Эти функции изображены на фиг. 42. Заметим, что в обоих случаях предельная частота f_0 определяется выражением

$$f_0 = \frac{l}{2\lambda d_i}, \quad (6.20)$$

причем для круглого зрачка она одинакова во всех направлениях частотной плоскости, тогда как в случае квадратного зрачка такое ограничение относится только к осям f_x и f_y . Чтобы оценить порядок величины f_0 , положим $l = 2$ см, $d_i = 10$ см, $\lambda = 10^{-4}$ см. Тогда предельная частота будет равна 100 период/мм.

§ 3. ЧАСТОТНЫЙ ОТКЛИК ДИФРАКЦИОННО ОГРАНИЧЕННОЙ НЕКОГЕРЕНТНОЙ ОПТИЧЕСКОЙ СИСТЕМЫ

Как мы видели, между функцией выходного зрачка и когерентной передаточной функцией существует очень простая прямая связь. Если освещение предмета некогерентно, передаточная функция оптической системы, как мы увидим, снова определяется выходным зрачком. Правда, теперь связь является менее непосредственной, но иногда она оказывается более интересной. Поэтому

теория образования изображения при некогерентном освещении имеет некоторые яркие особенности, отсутствующие в случае когерентного освещения. Теперь обратимся к рассмотрению этой теории, снова концентрируя внимание на *дифракционно ограниченных* системах, хотя последующие рассуждения применимы ко всем некогерентным системам независимо от их аберраций.

Оптическая передаточная функция и ее связь с когерентной передаточной функцией

Было показано, что для оптических систем, в которых используется некогерентное освещение, *интенсивность* определяется сверткой

$$I_i(x_i, y_i) = \kappa \iint_{-\infty}^{\infty} |\tilde{h}(x_i - \tilde{x}_0, y_i - \tilde{y}_0)|^2 I_g(\tilde{x}_0, \tilde{y}_0) d\tilde{x}_0 d\tilde{y}_0. \quad (6.21)$$

Поэтому при частотном анализе таких систем следовало бы считать, что они осуществляют линейное преобразование распределения интенсивности. Исходя из этого, введем *нормированные* частотные спектры распределений I_g и I_i , которые определяются выражениями

$$\mathcal{G}_g(f_x, f_y) = \frac{\iint_{-\infty}^{\infty} I_g(\tilde{x}_0, \tilde{y}_0) \exp[-j2\pi(f_x \tilde{x}_0 + f_y \tilde{y}_0)] d\tilde{x}_0 d\tilde{y}_0}{\iint_{-\infty}^{\infty} I_g(\tilde{x}_0, \tilde{y}_0) d\tilde{x}_0 d\tilde{y}_0}, \quad (6.22)$$

$$\mathcal{G}_i(f_x, f_y) = \frac{\iint_{-\infty}^{\infty} I_i(x_i, y_i) \exp[-j2\pi(f_x x_i + f_y y_i)] dx_i dy_i}{\iint_{-\infty}^{\infty} I_i(x_i, y_i) dx_i dy_i}. \quad (6.23)$$

Нормировка спектров при помощи их значений для «нулевой частоты» вводится отчасти для математического удобства, отчасти по более важным причинам. Так как интенсивность неотрицательная величина, то распределение интенсивности всегда содержит не равную нулю постоянную составляющую, или постоянный фон. Визуальное качество изображения в значительной степени зависит от «контраста», т. е. от относительной интенсивности несущих информацию участков изображения и всегда присутствующего фона. По указанной причине спектры нормируются на этот фон.

Нормированную передаточную функцию системы можно определить таким же образом:

$$\mathfrak{H}(f_x, f_y) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\tilde{h}(x_i, y_i)|^2 \exp[-j2\pi(f_x x_i + f_y y_i)] dx_i dy_i}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\tilde{h}(x_i, y_i)|^2 dx_i dy_i}. \quad (6.24)$$

Применение теоремы свертки к интегралу (6.21) приводит тогда к следующему соотношению в частотной области:

$$\mathfrak{G}_i(f_x, f_y) = \mathfrak{H}(f_x, f_y) \mathfrak{G}_g(f_x, f_y). \quad (6.25)$$

Функция \mathfrak{H} известна под названием *оптической передаточной функции* (сокращенно ОПФ) системы. Ее модуль $|\mathfrak{H}|$ называют *модуляционной передаточной функцией* (МПФ). Заметим, что $\mathfrak{H}(f_x, f_y)$ просто определяет вводимый системой комплексный весовой множитель частотной составляющей в точке (f_x, f_y) , *отнесенный* к весовому множителю составляющей нулевой частоты.

Так как когерентная передаточная функция и оптическая передаточная функция содержат функцию \tilde{h} , можно ожидать, что между ними имеется некоторая связь. Такая связь действительно существует; ее можно найти при помощи теоремы автокорреляции (см. гл. 2). Так как

$$\mathfrak{H}(f_x, f_y) = \mathcal{F}\{\tilde{h}\}$$

и

$$\mathfrak{F}(f_X, f_Y) = \frac{\mathcal{F}\{|\tilde{h}|^2\}}{\mathcal{F}\{|\tilde{h}|^2\}_{f_X=0, f_Y=0}},$$

то непосредственно получаем

$$\mathfrak{F}(f_X, f_Y) = \frac{\int \int_{-\infty}^{\infty} H(\xi', \eta') H^*(\xi' + f_X, \eta' + f_Y) d\xi' d\eta'}{\int \int_{-\infty}^{\infty} |H(\xi', \eta')|^2 d\xi' d\eta'}. \quad (6.26)$$

Простой заменой переменных

$$\xi = \xi' + \frac{f_X}{2}, \quad \eta = \eta' + \frac{f_Y}{2}$$

приходим к симметричному выражению

$$\begin{aligned} \mathfrak{F}(f_X, f_Y) = & \\ & \frac{\int \int_{-\infty}^{\infty} H\left(\xi - \frac{f_X}{2}, \eta - \frac{f_Y}{2}\right) H^*\left(\xi + \frac{f_X}{2}, \eta + \frac{f_Y}{2}\right) d\xi d\eta}{\int \int_{-\infty}^{\infty} |H(\xi, \eta)|^2 d\xi d\eta}. \end{aligned} \quad (6.27)$$

Выражение (6.27) определяет основную связь между свойствами когерентных и некогерентных систем. Заметим, что оно полностью справедливо для систем как с абберациями, так и без аббераций.

Основные свойства ОПФ

Ряд самых простых и общих свойств ОПФ можно получить непосредственно из выражения (6.27). Наиболее важные свойства перечислены ниже:

1. $\mathfrak{F}(0, 0) = 1$,
2. $\mathfrak{F}(-f_X, -f_Y) = \mathfrak{F}^*(f_X, f_Y)$,
3. $|\mathfrak{F}(f_X, f_Y)| \leq |\mathfrak{F}(0, 0)|$.

Первое свойство получаем непосредственно подстановкой $f_x = 0$, $f_y = 0$ в выражение (6.27). Доказательство второго свойства предлагаем провести читателям в качестве упражнения (это свойство представляет собой не более чем утверждение, что преобразование Фурье действительной функции эрмитово).

Доказательство того, что МПФ на любой частоте всегда меньше, чем ее величина на нулевой частоте, равная единице, значительно сложнее. Для доказательства третьего свойства используем *неравенство Шварца* [12], которое можно сформулировать следующим образом: если $X(\xi, \eta)$ и $Y(\xi, \eta)$ — любые две комплексные функции переменных ξ, η , то

$$\left| \int \int XY d\xi d\eta \right|^2 \leq \int \int |X|^2 d\xi d\eta \int \int |Y|^2 d\xi d\eta. \quad (6.28)$$

Знак равенства ставится тогда и только тогда, когда $Y = KX^*$, где K — комплексная постоянная. Полагая

$$X(\xi, \eta) = H\left(\xi - \frac{f_x}{2}, \eta - \frac{f_y}{2}\right)$$

и

$$Y(\xi, \eta) = H^*\left(\xi + \frac{f_x}{2}, \eta + \frac{f_y}{2}\right),$$

находим

$$\begin{aligned} & \left| \int \int_{-\infty}^{\infty} H\left(\xi - \frac{f_x}{2}, \eta - \frac{f_y}{2}\right) H^*\left(\xi + \frac{f_x}{2}, \eta + \frac{f_y}{2}\right) d\xi d\eta \right|^2 \leq \\ & \leq \int \int_{-\infty}^{\infty} \left| H\left(\xi - \frac{f_x}{2}, \eta - \frac{f_y}{2}\right) \right|^2 d\xi d\eta \times \\ & \times \int \int_{-\infty}^{\infty} \left| H^*\left(\xi + \frac{f_x}{2}, \eta + \frac{f_y}{2}\right) \right|^2 d\xi d\eta = \\ & = \left[\int \int_{-\infty}^{\infty} |H(\xi, \eta)|^2 d\xi d\eta \right]^2. \end{aligned}$$

Нормируя правую часть неравенства, находим, что величина $|\mathcal{G}(f_x, f_y)|$ всегда меньше или равна единице.

Наконец, нужно отметить, что, хотя ОПФ всегда равен единице на нулевой частоте, это не означает, что абсолютный уровень интенсивности фона изображения совпадает с абсолютным уровнем интенсивности фона предмета. В действительности абсолютный уровень интенсивности всегда уменьшается из-за конечных размеров входной апертуры системы, но это уменьшение не очевидно вследствие нашей нормировки ОПФ.

ОПФ системы без аберраций

До сих пор наши рассуждения были одинаково применимы к системам с аберрациями и без них. Рассмотрим теперь частный случай дифракционно ограниченной системы. Напомним, что для когерентной системы

$$H(f_x, f_y) = P(\lambda d_i f_x, \lambda d_i f_y).$$

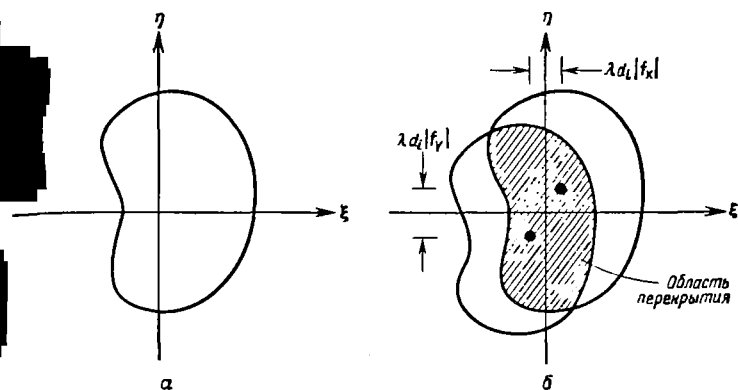
Для некогерентной системы, как следует из выражения (6.27) (с простой заменой переменных), имеем

$$\begin{aligned} \mathfrak{H}(f_x, f_y) &= \\ &= \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P\left(\xi - \frac{\lambda d_i f_x}{2}, \eta - \frac{\lambda d_i f_y}{2}\right) P\left(\xi + \frac{\lambda d_i f_x}{2}, \eta + \frac{\lambda d_i f_y}{2}\right) d\xi d\eta}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\xi, \eta) d\xi d\eta}, \end{aligned} \quad (6.29)$$

где в знаменателе мы заменили P^2 на P в силу того, что функция P равна или единице или нулю.

Выражение (6.29) для \mathfrak{H} допускает чрезвычайно важную геометрическую интерпретацию. Числитель представляет область перекрытия двух смещенных функций зрачка, центр одной из которых лежит в точке с координатами $(\lambda d_i f_x/2, \lambda d_i f_y/2)$, а центр другой — диаметрально противоположной точке $(-\lambda d_i f_x/2, -\lambda d_i f_y/2)$. Знаменатель просто нормирует эту область перекрытия к полной площади зрачка. Таким образом,

$$\mathfrak{H}(f_x, f_y) = \frac{\text{Область перекрытия}}{\text{Полная площадь}}.$$



Фиг. 43. Геометрическая интерпретация ОПФ дифракционно ограниченной системы.

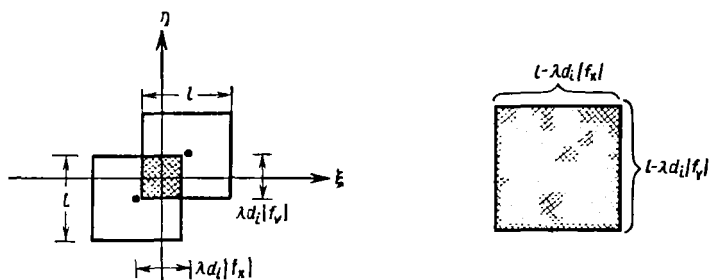
а — функция зрачка; полная площадь равна знаменателю ОПФ; б — две смещенные функции зрачка; заштрихованная площадь равна числителю ОПФ.

Расчет ОПФ дифракционно ограниченной системы можно сразу выполнить с помощью последовательности действий, основанных на этой интерпретации, иллюстрируемой фиг. 43. В случае простых геометрических форм удастся найти замкнутые выражения для нормированной области перекрытия (см. приведенные ниже примеры). Для зрачков очень сложной формы значения ОПФ для ряда дискретных частот можно рассчитать при помощи планиметра или цифровой вычислительной машины.

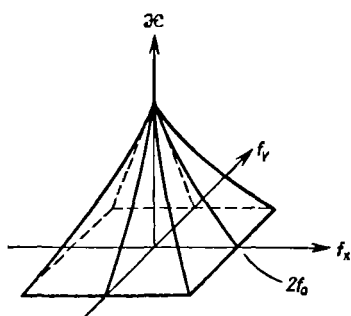
Наконец, заметим, что такая геометрическая интерпретация показывает, что ОПФ дифракционно ограниченной системы всегда действительна и неотрицательна. Однако она не обязательно должна быть монотонно убывающей функцией частот (см., например, задачу 6.1).

Примеры ОПФ дифракционно ограниченных систем

Рассмотрим теперь ОПФ, соответствующие дифракционно ограниченным системам с квадратным (ширина l) и круглым (диаметр l) выходными зрачками. Фиг. 44, а иллюстрирует расчет для случая квадратного



а



б

Фиг. 44. ОПФ дифракционно ограниченной системы с квадратным выходным зрачком.

а — расчет области перекрытия; б — вид полученной ОПФ.

зрачка. Площадь области перекрытия, очевидно, равна

$$\mathcal{A}(f_x, f_y) = \begin{cases} (l - \lambda d_i |f_x|)(l - \lambda d_i |f_y|) & |f_x| \leq l/\lambda d_i, \\ & |f_y| \leq l/\lambda d_i, \\ 0 & \text{в остальных случаях.} \end{cases}$$

После нормировки этой области на полную площадь l^2 получим

$$\mathcal{G}(f_x, f_y) = \Lambda\left(\frac{f_x}{2f_0}\right) \Lambda\left(\frac{f_y}{2f_0}\right), \quad (6.30)$$

где Λ — треугольная функция, введенная в гл. 2, а f_0 — предельная частота той же системы при когерентном освещении

$$f_0 = \frac{l}{2\lambda d_i}.$$

Заметим, что предельная частота при некогерентном освещении системы равна $2f_0$ вдоль осей f_x и f_y ¹⁾. ОПФ, описываемая выражением (6.30), изображена на фиг. 44, б.

В случае круглого выходного зрачка расчет выполняется несколько сложнее. Так как ОПФ будет, очевидно, обладать осевой симметрией, достаточно вычислить \mathfrak{F} вдоль положительной оси f_x . Как видно из фиг. 45, можно считать, что область перекрытия равна учетверенной площади заштрихованной части B кругового сектора $A + B$. Но площадь сектора равна

Площадь $(A + B) =$

$$= \left[\frac{\theta}{2\pi} \right] \left[\pi \left(\frac{l}{2} \right)^2 \right] = \left[\frac{\arccos(\lambda d_i f_x / l)}{2\pi} \right] \left[\pi \left(\frac{l}{2} \right)^2 \right],$$

тогда как для площади треугольника A имеем

$$\text{Площадь } (A) = \frac{1}{2} \left(\frac{\lambda d_i f_x}{2} \right) \sqrt{\left(\frac{l}{2} \right)^2 - \left(\frac{\lambda d_i f_x}{2} \right)^2}.$$

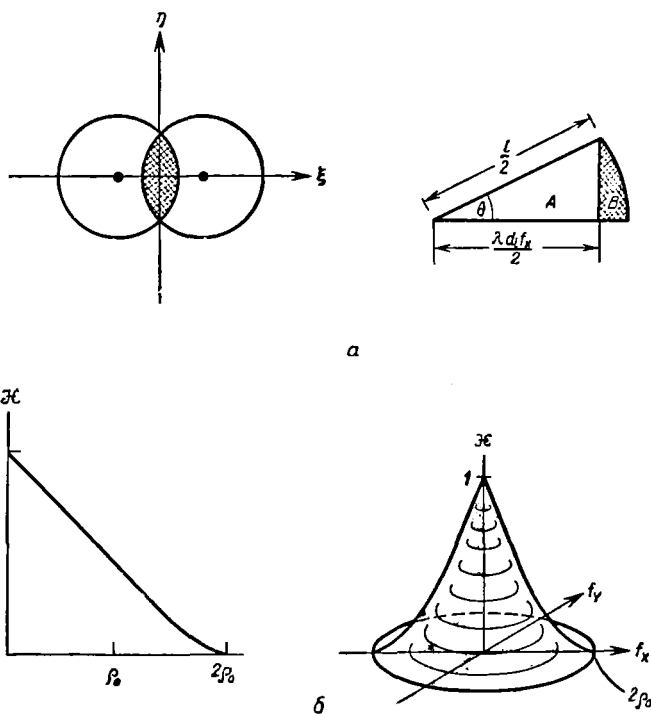
Наконец, можно написать

$$\mathfrak{F}(f_x, 0) = \frac{4 [\text{Площадь } (A + B) - \text{Площадь } (A)]}{\pi (l/2)^2}$$

или для произвольного радиального смещения ρ в частотной плоскости

$$\mathfrak{F}(\rho) = \begin{cases} \frac{2}{\pi} \left[\arccos\left(\frac{\rho}{2\rho_0}\right) - \frac{\rho}{2\rho_0} \sqrt{1 - \left(\frac{\rho}{2\rho_0}\right)^2} \right], & \rho \leq 2\rho_0, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (6.31)$$

¹⁾ Это не означает, что некогерентная система имеет вдвое большую разрешающую способность, чем когерентная система; см. § 3.



Фиг. 45. ОПФ дифракционно ограниченной системы с круглым выходным зрачком.

a — расчет области перекрытия; *b* — вид полученной ОПФ.

Величина ρ_0 представляет собой предельную частоту когерентной системы

$$\rho_0 = \frac{l}{2\lambda d_l}.$$

Как следует из фиг. 45, *b*, ОПФ опять отлична от нуля вплоть до частоты, в два раза превышающей предельную частоту когерентной системы.

§ 4. АБЕРРАЦИИ И ИХ ВЛИЯНИЕ НА ЧАСТОТНЫЙ ОТКЛИК

При разработке обобщенной модели оптической системы предполагалось, в частности, что наличие точеч-

ного источника обуславливает появление в выходном зрачке идеальной сферической волны, которая сходится в точке идеального изображения. Такая система называлась *дифракционно ограниченной*. Теперь рассмотрим влияние *аббераций*, или отклонений волнового фронта в выходном зрачке от сферической формы. Аберрации могут быть обусловлены различными причинами, начиная от самых простых, как, например, ошибка фокусировки, и кончая довольно сложными, связанными со свойствами самих идеально сферических линз, как, например, сферическая аберрация. Подробное обсуждение аберраций и их влияния на частотный отклик не входит в нашу задачу. Мы остановимся на самых общих эффектах, иллюстрацией которых послужит один относительно простой пример. Более полное рассмотрение различных типов аберраций и их влияния на частотный отклик содержится, например, в работе [3.4], гл. IX.

Обобщенная функция зрачка

Если оптическая система является дифракционно ограниченной, то импульсный отклик (при когерентном освещении), как мы видели, представляет собой картину дифракции Фраунгофера на выходном отверстии с центром в точке идеального изображения. Это обстоятельство подсказывает удобный прием, который позволит непосредственно учесть аберрации в наших предыдущих результатах. В частности, в случае искажения волнового фронта можно представить, что выходной зрачок освещается идеальной сферической волной, но в пределах отверстия находится фазовая пластинка, деформирующая выходящий из зрачка фронт волны. Если фазовая ошибка в точке (x, y) выходного зрачка изображается как $kW(x, y)$, где $k = 2\pi/\lambda$, а W — эффективная погрешность длины пути, то комплексный коэффициент пропускания P воображаемой фазовой пластинки можно представить в виде

$$P(x, y) = P(x, y) \exp [jkW(x, y)]. \quad (6.32)$$

Комплексную функцию P можно считать *обобщенной функцией зрачка*. Импульсный отклик когерентной

системы с aberrациями представляет собой просто картину дифракции Фраунгофера на отверстии с коэффициентом пропускания P . Импульсный отклик некогерентной системы с aberrациями, разумеется, равен квадрату модуля импульсного отклика когерентной системы.

Влияние aberrаций на когерентную передаточную функцию

При рассмотрении дифракционно ограниченной когерентной системы передаточная функция была определена с учетом того, что 1) импульсный отклик соответствует фурье-образу зрачка и 2) когерентная передаточная функция является фурье-образом импульсного отклика. Как следствие соотношений между двумя преобразованиями Фурье было найдено, что передаточная функция пропорциональна функции зрачка P . Подобными рассуждениями можно воспользоваться и при наличии aberrаций, если заменить функцию P обобщенной функцией зрачка P . Таким образом, передаточная функция при когерентном освещении запишется следующим образом:

$$\begin{aligned} H(f_x, f_y) &= P(\lambda d_{ifx}, \lambda d_{ify}) = \\ &= P(\lambda d_{ifx}, \lambda d_{ify}) \exp[jkW(\lambda d_{ifx}, \lambda d_{ify})]. \end{aligned} \quad (6.33)$$

Очевидно, что в случае когерентного освещения ограничение полосы пропускания передаточной функции, которое обусловлено конечным размером выходного зрачка, не зависит от наличия aberrаций. Aberrации вводят только *фазовые искажения* в пределах полосы пропускания. Разумеется, фазовые искажения могут оказывать некоторое влияние на точность воспроизведения.

Вышеприведенными соображениями почти исчерпывается все, что можно сказать в общем относительно влияния aberrаций волнового фронта на когерентную систему. В этом случае результат весьма прост; результат же для некогерентной системы, как мы сейчас увидим, опять более сложен и во многих отношениях более интересен.

Влияние аберраций на ОПФ

После того как выяснено влияние аберраций на когерентную передаточную функцию, при помощи выражения (6.27) можно определить их влияние на оптическую передаточную функцию. Для упрощения обозначений введем функцию $\mathcal{A}(f_X, f_Y)$, определяемую как *область перекрытия функций*

$$P\left(\xi - \frac{\lambda d_i f_X}{2}, \eta - \frac{\lambda d_i f_Y}{2}\right) \quad \text{и} \quad P\left(\xi + \frac{\lambda d_i f_X}{2}, \eta + \frac{\lambda d_i f_Y}{2}\right).$$

Таким образом, ОПФ дифракционно ограниченной системы в этих новых обозначениях имеет вид

$$\Phi(f_X, f_Y) = \frac{\iint_{\mathcal{A}(f_X, f_Y)} d\xi d\eta}{\iint_{\mathcal{A}(0, 0)} d\xi d\eta}.$$

При наличии аберраций подстановка (6.33) в (6.27) дает

$$\begin{aligned} \Phi(f_X, f_Y) &= \\ &= \frac{\iint_{\mathcal{A}(f_X, f_Y)} \exp\left\{jk\left[W\left(\xi - \frac{\lambda d_i f_X}{2}, \eta - \frac{\lambda d_i f_Y}{2}\right) - W\left(\xi + \frac{\lambda d_i f_X}{2}, \eta + \frac{\lambda d_i f_Y}{2}\right)\right]\right\} d\xi d\eta}{\iint_{\mathcal{A}(0, 0)} d\xi d\eta}. \end{aligned} \quad (6.34)$$

Это выражение позволяет нам непосредственно связать погрешности волнового фронта и ОПФ.

Как важное общее свойство, отметим тот факт, что аберрации *никогда не увеличивают* МПФ (модуль ОПФ). Для доказательства этого свойства используем неравенство Шварца (6.28). Пусть функции X и Y в этом неравенстве имеют следующий вид:

$$\begin{aligned} X(\xi, \eta) &= \exp\left[jkW\left(\xi - \frac{\lambda d_i f_X}{2}, \eta - \frac{\lambda d_i f_Y}{2}\right)\right], \\ Y(\xi, \eta) &= \exp\left[-jkW\left(\xi + \frac{\lambda d_i f_X}{2}, \eta + \frac{\lambda d_i f_Y}{2}\right)\right]. \end{aligned}$$

Замечая, что $|X|^2 = |Y|^2 = 1$, получаем

$$\begin{aligned}
 & |\mathfrak{H}(f_x, f_y)|_{\text{с aberrациями}}^2 = \\
 & \left| \frac{\iint_{\mathcal{A}(f_x, f_y)} \exp \left\{ ik \left[W \left(\xi - \frac{\lambda d_i f_x}{2}, \eta - \frac{\lambda d_i f_y}{2} \right) - \right. \right. \right. \\
 & \qquad \qquad \qquad \left. \left. \left. - W \left(\xi + \frac{\lambda d_i f_x}{2}, \eta + \frac{\lambda d_i f_y}{2} \right) \right] \right\} d\xi d\eta \right|^2 \\
 & = \frac{\iint_{\mathcal{A}(0,0)} d\xi d\eta}{\iint_{\mathcal{A}(0,0)} d\xi d\eta} \ll \\
 & \ll \left[\frac{\iint_{\mathcal{A}(f_x, f_y)} d\xi d\eta}{\iint_{\mathcal{A}(0,0)} d\xi d\eta} \right]^2 = |\mathfrak{H}(f_x, f_y)|_{\text{без aberrаций}}^2
 \end{aligned}$$

Таким образом, aberrации будут в основном уменьшать контраст каждой пространственно-частотной составляющей интенсивности изображения. Абсолютное значение предельной частоты остается неизменным, но сильные aberrации будут настолько уменьшать высокочастотные части ОПФ, что эффективная предельная частота окажется много меньше соответствующей величины в дифракционно ограниченной системе. Кроме того, aberrации могут привести к отрицательным значениям ОПФ в некоторых диапазонах частот, что совершенно невозможно в системах, свободных от aberrаций. Если ОПФ отрицательна, то составляющие изображения на этой частоте испытывают реверсирование контраста, т. е. пики интенсивности превращаются в нули интенсивности и наоборот. Пример такого эффекта приведен в следующем параграфе.

Пример простой aberrации — ошибка фокусировки¹⁾

Одной из самых простых aberrаций, легко описываемых математически, является ошибка фокусировки. Но

¹⁾ Ошибку фокусировки можно отнести к aberrациям системы только формально. Aberrации определяют отклонение выходящей из системы волны от наиболее близкой к ней сферической волны. В дан-

даже в этом простом случае для того, чтобы не усложнять математические выкладки, проще рассмотреть квадратное отверстие (а не круглое, которое представляет больший практический интерес).

Как было показано в гл. 5, § 3, условия, при которых предмет находится в фокусе, определяются формулой линзы

$$\frac{1}{d_i} + \frac{1}{d_o} - \frac{1}{f} = 0.$$

Если плоскость изображения не совпадает с фокальной плоскостью, то справедливо более общее соотношение

$$\frac{1}{d_i} + \frac{1}{d_o} - \frac{1}{f} = \varepsilon.$$

Сравнивая (5.30) и (6.33), находим функцию аберрации

$$W(x, y) = \frac{\varepsilon(x^2 + y^2)}{2}. \quad (6.35)$$

Для квадратного отверстия со стороной l максимальная фазовая ошибка на краю отверстия (вдоль оси x или y) равна $\varepsilon kl^2/8$. Таким образом, максимальная ошибка в оптической длине, которую мы обозначим ω , равна

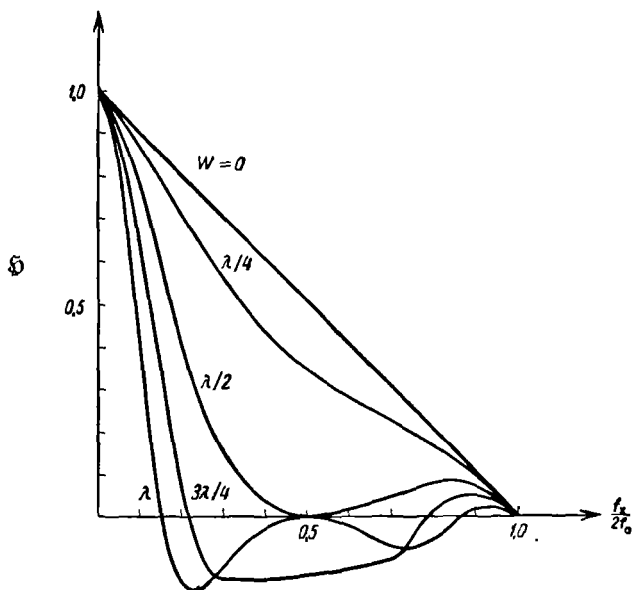
$$\omega = \frac{\varepsilon l^2}{8}. \quad (6.36)$$

Число ω — удобный показатель величины ошибки фокусировки.

Если выражение (6.35) для функции аберрации W подставить в выражение (6.34) для ОПФ, то после ряда простых преобразований получим

$$\begin{aligned} \Phi(f_x, f_y) &= \Lambda\left(\frac{f_x}{2f_o}\right) \Lambda\left(\frac{f_y}{2f_o}\right) \times \\ &\times \operatorname{sinc}\left[\frac{8\omega}{\lambda}\left(\frac{f_x}{2f_o}\right)\left(1 - \frac{|f_x|}{2f_o}\right)\right] \operatorname{sinc}\left[\frac{8\omega}{\lambda}\left(\frac{f_y}{2f_o}\right)\left(1 - \frac{|f_y|}{2f_o}\right)\right]. \end{aligned} \quad (6.37)$$

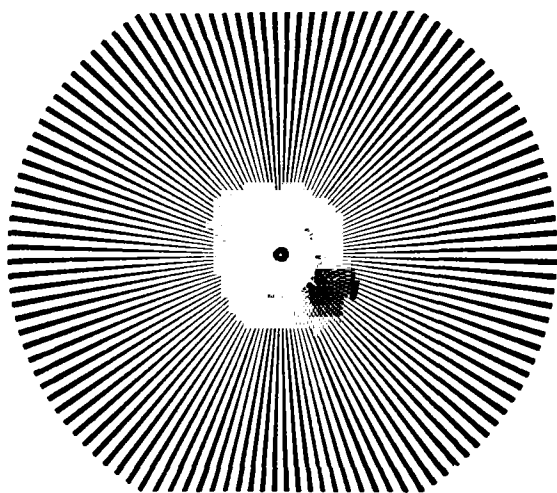
ном случае речь идет об отклонении сферической волны, сходящейся в точке изображения, от сферы с центром в точке наблюдения. — Прим. ред.



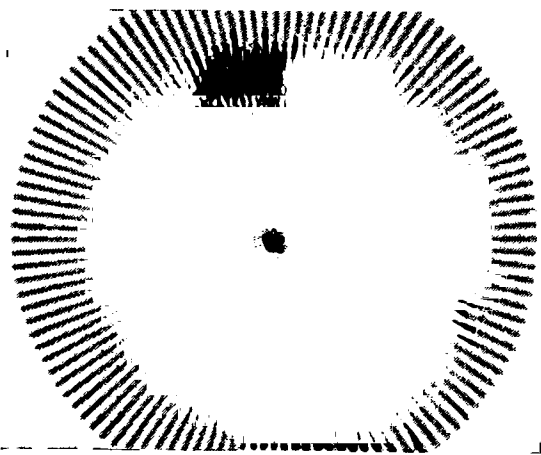
Фиг. 46. График ОПФ при неточной фокусировке в случае квадратного отверстия.

График этой ОПФ показан на фиг. 46. Заметим, что при $w = 0$ мы действительно получаем ОПФ для дифракционно ограниченной системы. Отметим также, что для значений $w > 1/2$ знак ОПФ меняется на обратный. Эти реверсирования контраста можно легко наблюдать, если в качестве предмета использовать изображенную на фиг. 47, а миру, составленную из «спиц». Если система расфокусирована, то при увеличении пространственной частоты (т. е. уменьшении радиуса) происходит постепенное ослабление контраста и ряд реверсирований контраста, как видно на фиг. 47, б.

Наконец, рассмотрим вид ОПФ в случае очень большой ошибки фокусировки (т. е. при $w \gg 1$). В этом случае частотный отклик падает до нуля для относительно небольших значений $f_x/2f_0$ и $f_y/2f_0$. Следовательно,



a



б

Ф и г. 47. Сфокусированное (*a*) и расфокусированное (*б*) изображение миры, составленной из «спиц».

можно записать

$$1 - \frac{|f_x|}{2f_0} \approx 1, \quad 1 - \frac{|f_y|}{2f_0} \approx 1$$

и выражение для ОПФ принимает вид

$$\mathfrak{F}(f_x, f_y) \approx \text{sinc} \left[\frac{8\omega}{\lambda} \left(\frac{f_x}{2f_0} \right) \right] \text{sinc} \left[\frac{8\omega}{\lambda} \left(\frac{f_y}{2f_0} \right) \right]. \quad (6.38)$$

Читатель может убедиться (см. задачу 6.6), что это есть точное выражение для ОПФ, предсказываемое геометрической оптикой! Вообще в случае значительных аберраций любого вида для получения хорошего приближения для ОПФ системы достаточно подвергнуть преобразованию Фурье импульсный отклик, предсказываемый геометрической оптикой.

§ 5. СРАВНЕНИЕ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЯ ПРИ КОГЕРЕНТНОМ И НЕКОГЕРЕНТНОМ ОСВЕЩЕНИИ

Как следует из § 2 и 3, ОПФ дифракционно ограниченной системы отлична от нуля вплоть до частоты, вдвое превышающей предельную частоту когерентной передаточной функции. Естественно напрашивается вывод, что изображения при некогерентном освещении неизменно будут «лучше», чем при когерентном, при условии, что в обоих случаях используется одна и та же оптическая система. Как мы сейчас увидим, этот вывод, вообще говоря, *неверен*: сравнение двух типов освещения значительно сложнее, чем может показаться при таком поверхностном рассмотрении.

Слабым местом вышеприведенных соображений является непосредственное сравнение предельных частот в этих двух случаях. В действительности их нельзя сравнивать непосредственно. Предельная частота при когерентном освещении определяет составляющую максимальной частоты для *амплитуды* изображения, тогда как предельная частота при некогерентном освещении относится к частотным составляющим *интенсивности* изображения; однако в обоих случаях в конечном итоге наблю-

дается интенсивность. Несомненно, любое прямое сравнение этих двух систем должно проводиться по одной и той же величине — интенсивности изображения.

Даже если величина, по которой сравниваются системы, выбрана правильно, сравнение по-прежнему будет затруднительным в силу еще одного важного обстоятельства: неопределенным остается понятие «лучше». Мы не имеем универсального качественного критерия, на котором могли бы основывать свои суждения. Можно рассмотреть ряд возможных критериев этого типа (например, наименьшая среднеквадратичная разность между интенсивностями предмета и изображения), но, к сожалению, конечная реакция наблюдателя так сложна и малопонятна, что трудно установить правильный и имеющий смысл критерий.

Из-за отсутствия подходящего качественного критерия сравнение двух типов изображения неизбежно касается лишь каких-то его параметров и мало говорит о полном качестве изображения. Тем не менее такие сравнения весьма полезны, поскольку выявляют некоторые различия между двумя типами освещения.

Частотный спектр интенсивности изображения

Одной из простых характеристик интенсивности изображения, по которой можно производить его сравнение в двух случаях, служит *частотный спектр*. В отличие от некогерентной системы, которая линейна в отношении интенсивности, когерентная система сильно нелинейна по этой переменной. Таким образом, нахождение спектра интенсивности в случае когерентного освещения несколько затруднено.

В случае некогерентного освещения интенсивность изображения представляется сверткой

$$I_i = |\tilde{h}|^2 * I_g = |\tilde{h}|^2 * |U_g|^2.$$

С другой стороны, в случае когерентного освещения имеем

$$I_i = |\tilde{h} * U_g|^2.$$

Будем обозначать символом \star операцию автокорреляции

$$\begin{aligned} X(f_x, f_y) \star X(f_x, f_y) &= \\ &= \int_{-\infty}^{\infty} \int X(\xi, \eta) X^*(\xi + f_x, \eta + f_y) d\xi d\eta. \end{aligned} \quad (6.39)$$

Тогда частотный спектр интенсивности изображения для двух случаев можно записать сразу следующим образом:

$$\begin{aligned} \text{Некогерентное освещение } \mathcal{F}\{I_i\} &= [H \star H][G_g \star G_g], \\ \text{Когерентное освещение } \mathcal{F}\{I_i\} &= H G_g \star H G_g, \end{aligned} \quad (6.40)$$

где G_g — спектр функции U_g , а H — передаточная функция при когерентном освещении.

Из общих соотношений (6.40) не следует, что в смысле частотного состава изображения один вид освещения лучше другого. Однако они показывают, что частотные спектры в этих двух случаях могут быть различными и, кроме того, результаты любого подобного сравнения будут в значительной степени зависеть от распределений как интенсивности, так и фазы в плоскости предмета.

Для подтверждения последнего вывода рассмотрим два предмета с одинаковыми коэффициентами пропускания по интенсивности, но с различными распределениями фазы. При этом можно будет сказать, что изображение одного из них лучше в когерентном свете, а другого — в некогерентном. Пусть коэффициент пропускания по интенсивности для предметов в обоих случаях имеет вид

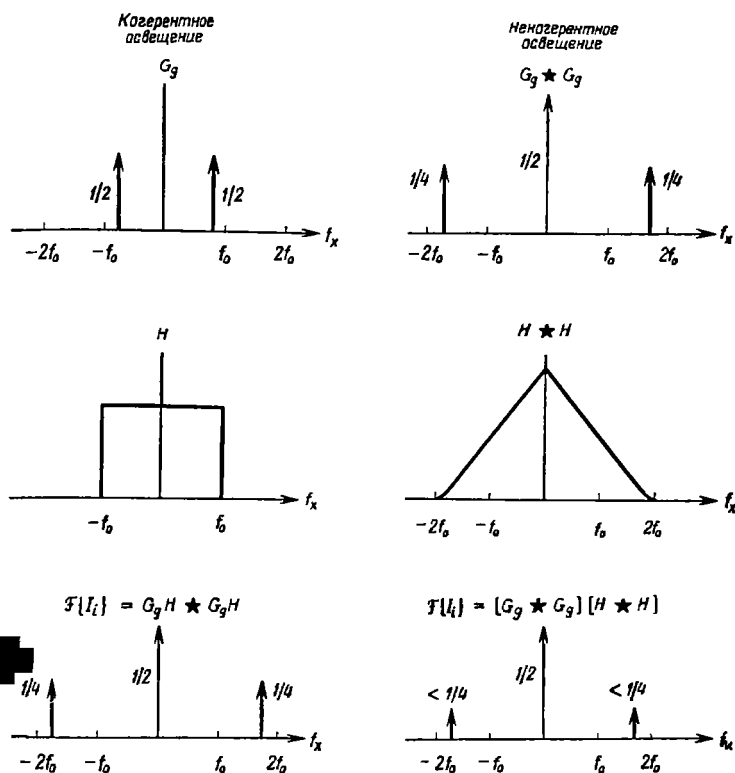
$$\tau(x_0, y_0) = \cos^2 2\pi \bar{f} x_0,$$

где мы считаем, что

$$\frac{f_0}{2} < \bar{f}_0 < f_0,$$

причем f_0 — предельная частота при когерентном освещении. Амплитудные коэффициенты пропускания для двух предметов выбраны следующим образом:

$$\begin{aligned} A: \quad & t_a(x_0, y_0) = \cos 2\pi \bar{f} x_0, \\ B: \quad & t_b(x_0, y_0) = |\cos 2\pi \bar{f} x_0|, \end{aligned} \quad (6.41)$$



Фиг. 48. Расчет спектра интенсивности изображения для предмета А.

т. е. предметы различаются только периодическим распределением фазы. Фиг. 48 иллюстрирует последовательные действия в частотной области, которые в итоге дают спектр интенсивности изображения предмета А. Во всех случаях предполагается, что оптическая система дифракционно ограничена. Заметим, что контраст распределения интенсивности изображения для некогерентного освещения хуже, чем для когерентного. Таким образом, можно сказать, что для данного предмета *лучше* когерентное освещение.

Соответствующее сравнение для предмета B осуществляется проще. Амплитуда предмета теперь представляет собой периодическую функцию с частотой основной гармоники $2f$. Но так как $2f > f_0$, то в случае когерентного освещения интенсивность изображения меняться *не будет*, тогда как некогерентная система формирует то же самое изображение, как и для предмета A . Таким образом, во втором случае *лучше* некогерентное освещение.

В итоге ответ на вопрос, какой тип освещения лучше с точки зрения спектрального состава изображения, зависит в значительной степени от структуры предмета и, в частности, от распределения фазы в точках предмета. *Нельзя* сказать, что один вид освещения предпочтителен во всех случаях. Сравнение обычно сложно, хотя существуют и простые случаи, подобные только что рассмотренному. Другой простой пример рассмотрен в задаче 6.3.

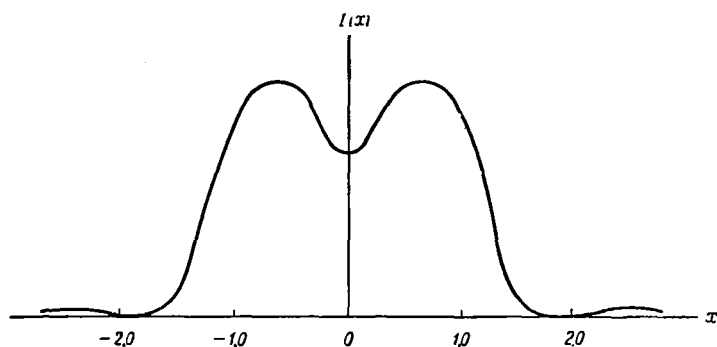
Разрешение двух точек

Второй возможный критерий сравнения основан на способности соответствующей системы различать два близко расположенных точечных источника. Такой критерий давно использовался как характеристика качества оптических систем, в частности в астрономии, где он играет наиболее важную практическую роль.

Согласно так называемому *критерию Рэлея*, два некогерентных точечных источника «разрешаются» (дифракционно ограниченной) системой, если центр диска Эри, созданного одним источником, совпадает с нулем первого порядка дифракционной картины, созданной вторым источником. Минимальное разрешаемое расстояние между идеальными изображениями, следовательно, равно

$$\delta = 1,22 \frac{\lambda d_i}{l}, \quad (6.42)$$

где l — диаметр выходного зрачка [ср. соотношение (4.18)]. На фиг. 49 представлено распределение интенсивности изображения для этого минимального расстояния. Заметим, что провал в центре составляет около 19% от максимальной интенсивности.



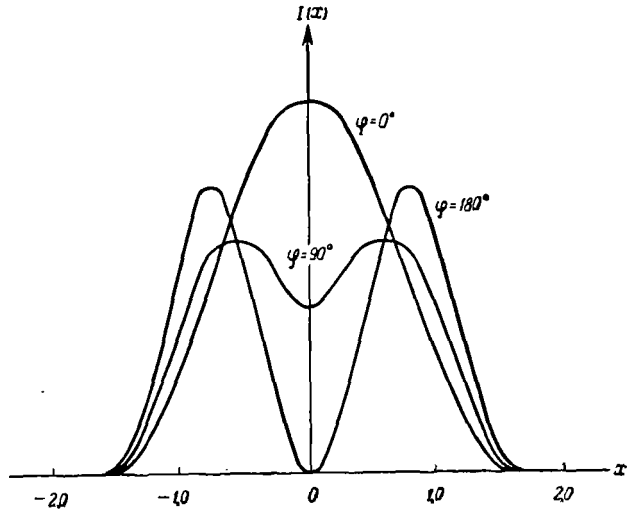
Фиг. 49. Распределение интенсивности в изображении двух некогерентных точечных источников, находящихся на минимальном разрешаемом расстоянии.

Теперь мы можем спросить: при каком освещении, когерентном или некогерентном, легче различать два точечных источника, разделенных расстоянием Релея? Для астрономических объектов, где мы всегда имеем некогерентное освещение, этот вопрос, разумеется, носит чисто академический характер. Однако в области микрокопии, где по желанию можно выбрать когерентное или некогерентное освещение, этот вопрос представляет определенный интерес.

Как и в предыдущих случаях, ответ зависит от *распределения фазы*, связанного с предметом. Распределение интенсивности в изображении можно сразу записать в нормированных координатах изображения следующим образом:

$$I(x) = \left| 2 \frac{J_1[\pi(x-0,61)]}{\pi(x-0,61)} + e^{i\varphi} 2 \frac{J_1[\pi(x+0,61)]}{\pi(x+0,61)} \right|^2,$$

где φ — относительная разность фаз двух точечных источников. На фиг. 50 приведены распределения интенсивности изображения для точечных источников, излучающих в фазе ($\varphi = 0^\circ$), в противофазе ($\varphi = 180^\circ$) и с разностью фаз в четверть периода ($\varphi = 90^\circ$). Распределение интенсивности в изображении для источников, сдвинутых по фазе на 90° , тождественно распределению для некогерентных точечных источников. Когда разность

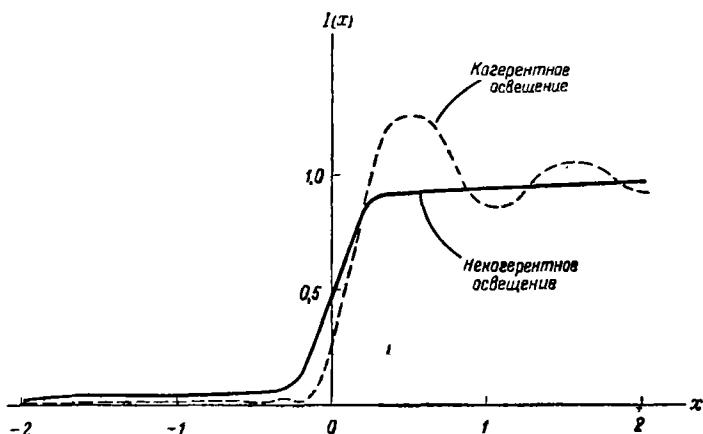


Фиг. 50. Распределение интенсивности в изображении двух взаимно когерентных точечных источников, находящихся на минимальном разрешаемом расстоянии.

фаз $\varphi = 0^\circ$, провал на кривой интенсивности изображения отсутствует и, следовательно, две точки труднее различать, чем в случае некогерентного освещения. Наконец, если точечные источники излучают в противофазе, то провал составляет больше 19% и при когерентном освещении две точки различимы лучше, чем при некогерентном. Таким образом, снова нельзя сделать вывод относительно того, какой вид излучения предпочтительнее с точки зрения разрешающей способности.

Другие эффекты

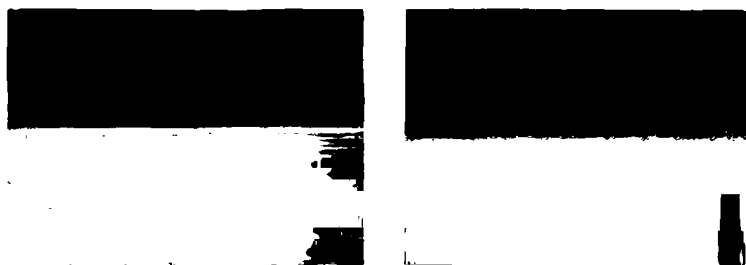
Существует ряд свойств изображений, о которых следует упомянуть при сравнении когерентных и некогерентных изображений [13]. Во-первых, отклики некогерентной и когерентной систем на возмущение, распространяющееся от резко очерченного края, существенно отличаются друг от друга. На фиг. 51 представлены



Фиг. 51. Теоретические кривые интенсивности откликов системы на функцию скачка в пространстве предмета при когерентном и некогерентном освещении.

вычисленные теоретически отклики системы с квадратным выходным зрачком на функцию скачка в пространстве предмета. На фиг. 52 приведены фотографии изображения края в двух случаях. Видно, что в случае когерентного освещения резко выражены осцилляции. Это свойство аналогично «дребезжанию», которое имеет место в видеоусилительных схемах, если передаточные функции падают слишком резко. Передаточная функция когерентной системы имеет резкие переходы, тогда как изменения ОПФ присходят более плавно. Из графиков, приведенных на фиг. 51, также видно, что кажущиеся положения края (т. е. абсциссы точек, в которых интенсивности отклика равны половине максимальной) различны для двух видов освещения, причем когерентное освещение вводит незначительную погрешность.

Кроме того, мы должны упомянуть об *эффекте зернистости*, который легко наблюдается при высокой степени когерентности освещения. На фиг. 53 приведены фотографии предмета в когерентном и некогерентном свете. Зернистость изображения, полученного в когерентном свете, — прямое следствие оптической грубости



Фиг. 52. Фотографии функции скачка в пространстве предмета при когерентном (слева) и некогерентном (справа) освещении.

поверхности предмета и высокой когерентности света лазера, который служил источником (см., например, [14, 15]). Можно показать, что размер отдельных *зерен* приблизительно равен размеру изображения разрешаемого элемента. Таким образом, если размер рассматриваемого предмета приблизительно равен *пределу разрешения*



Фиг. 53. Изображения, иллюстрирующие эффект зернистости при когерентном (а) и некогерентном (б) освещении.

оптической системы, эффекты зернистости могут быть особенно неприятны. Часто можно избавиться от зернистости, помещая кусочек матового стекла в поток света и сдвигая его в процессе наблюдения. Степень когерентности света частично уменьшается, и зернистость в значительной степени «размывается».

Наконец, освещение с высокой степенью когерентности особенно чувствительно к оптическим дефектам, которые могут встретиться на пути света. Например, мельчайшие частички пыли на линзе могут приводить к появлению ярко выраженных дифракционных колец, которые будут накладываться на изображение. Эти эффекты можно свести к минимуму, помещая подвижный рассеиватель, подобный матовому стеклу, в поток света.

§ 6. РАЗРЕШАЮЩАЯ СПОСОБНОСТЬ, ПРЕВЫШАЮЩАЯ КЛАССИЧЕСКИЙ ДИФРАКЦИОННЫЙ ПРЕДЕЛ

До недавнего времени было широко распространено мнение, что дифракционные эффекты накладывают принципиальные ограничения на характеристики оптической системы. Считалось, что невозможно получить разрешение, превышающее классический дифракционный предел¹). Даже для бесконечно большой линзы появление затухающих волн ограничивает разрешение величиной порядка оптической длины волны. Последние работы [17—19] показали, однако, что для некоторых типов предметов можно получить разрешение, превышающее классический предел²). Действительно, как мы покажем в этом параграфе, для пространственно *ограниченных* предметов в *принципе* можно получить сколь угодно высокое разрешение.

1) Критерий предела разрешения зависит от соглашения относительно контрастной чувствительности приемника, например глаза. — *Прим. ред.*

2) Особый интерес представляет работа Торальдо ди Франсиса [17], который с помощью теории сверхнаправленных антенн показал, что привычный «предел Релея» для разрешения носит практический, а не теоретический характер.

Основные математические положения

Существуют весьма основательные математические соображения, показывающие, что разрешение может превышать классический предел. Эти соображения основаны на двух главных математических принципах, которые мы приведем здесь в виде теорем. Их доказательства можно найти, например, в работе [20].

Теорема 1. Двумерный фурье-образ пространственно ограниченной функции есть аналитическая функция в плоскости $f_x f_y$.

Теорема 2. Если любая аналитическая функция в плоскости $f_x f_y$ точно известна в сколь угодно малой (но конечной) области этой плоскости, то полная функция (единственная) может быть найдена методом аналитического продолжения.

Далее, для любой оптической системы как при когерентном, так и при некогерентном освещении информация, содержащаяся в изображении, определяется конечной частью спектра предмета¹⁾, а именно частью, пропущенной передаточной функцией системы. Если эту конечную часть спектра предмета можно точно определить по изображению, то для ограниченного предмета *полный* спектр предмета можно найти аналитическим продолжением. Если найден полный спектр предмета, то предмет может быть восстановлен с бесконечной точностью.

Пример схемы восстановления

Хотя основные математические принципы легче всего формулировать в терминах теории аналитического продолжения, восстановление можно осуществить многими способами; результат при этом не изменится. Особенно простой метод рассмотрен в работе Гарриса [18]. Предположим, что предмет ограничен прямоугольником со сторонами L_x и L_y . Тогда, используя теорему Уиттекера — Шеннона, можно записать спектр предмета $G_0(f_x, f_y)$ через его выборочные значения в точках

¹⁾ Под *спектром предмета* мы подразумеваем, конечно, спектр амплитуды предмета при когерентном освещении и спектр интенсивности предмета при некогерентном освещении.

$(n/L_X, m/L_Y)$ ¹⁾:

$$G_0(f_X, f_Y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} G_0\left(\frac{n}{L_X}, \frac{m}{L_Y}\right) \times \\ \times \operatorname{sinc}\left[2L_X\left(f_X - \frac{n}{2L_X}\right)\right] \operatorname{sinc}\left[2L_Y\left(f_Y - \left(\frac{m}{2L_Y}\right)\right)\right]. \quad (6.43)$$

Вследствие ограниченности полосы пропускания оптической системы значения $G_0(n/L_X, m/L_Y)$ можно найти только для нескольких малых чисел (n, m) . Предположим, что мы сумели каким-то образом определить G_0 для больших целых чисел, скажем $(n = \pm N, m = \pm M)$, так что с помощью приближенной формулы

$$G_0(f_X, f_Y) \approx \sum_{n=-N}^N \sum_{m=-M}^M G_0\left(\frac{n}{L_X}, \frac{m}{L_Y}\right) \times \\ \times \operatorname{sinc}\left[2L_X\left(f_X - \frac{n}{2L_X}\right)\right] \operatorname{sinc}\left[2L_Y\left(f_Y - \frac{m}{2L_Y}\right)\right] \quad (6.44)$$

можем определить точные значения $G_0(f_X, f_Y)$ по крайней мере в некоторой конечной области *вне* полосы пропускания системы. Другими словами, формула (6.44) позволяет установить спектр предмета по крайней мере в некоторой конечной области *вне* полосы пропускания, если известно достаточное число *выборочных значений* вне полосы пропускания.

Чтобы определить выборочные значения вне полосы пропускания, можно использовать формулу (6.44) иным образом: мы *измеряем* значения G_0 на любых $(2N+1)(2M+1)$ частотах *внутри* полосы пропускания. Если любое из этих измеренных значений подставить вместо G_0 в левую часть формулы (6.44), получим одно уравнение с $(2N+1)(2M+1)$ неизвестными ²⁾:

$$G_0\left(\frac{n}{L_X}, \frac{m}{L_Y}\right): n = 0, \pm 1, \dots, \pm N, m = 0, \pm 1, \dots, \pm M.$$

¹⁾ Аналитичность функции G_0 можно доказать непосредственно из (6.43) путем разложения sinc функции в степенной ряд, откуда следует, что G_0 тоже раскладывается в степенной ряд.

²⁾ Фактически некоторые из этих неизвестных могут лежать в пределах полосы пропускания и, следовательно, быть известными, но это упрощение только облегчает восстановление.

Таким образом, $(2N+1)(2M+1)$ известных значений G_0 дают $(2N+1)(2M+1)$ линейных уравнений с $(2N+1)(2M+1)$ неизвестными. Эту систему уравнений можно в принципе решить и получить ряд выборочных значений, которые лежат также и вне полосы пропускания. Теорема выборки затем позволяет найти значения G_0 между выборками и получить по крайней мере часть спектра предмета, лежащую вне полосы пропускания. Преобразование Фурье дает затем изображение с более высоким разрешением, чем допускает классический предел.

Практические ограничения

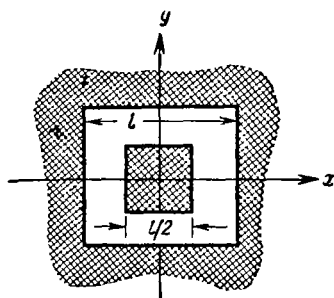
Математические принципы, лежащие в основе теории восстановления предмета, теперь ясны. Остается выяснить, какой результат можно получить, применяя эти принципы к реальным оптическим системам. Основное затруднение представляет собой *шум*, который неизменно сопровождает любое измерение спектра предмета в пределах полосы пропускания. Неточность измерений в пределах полосы пропускания может существенно увеличиться, если система линейных уравнений решается для выборочных значений вне полосы пропускания, поскольку такие выборочные значения выражаются линейными комбинациями *всех* измеренных значений, каждое из которых вносит некоторый шумовой вклад. Проблема восстановления остается предметом активных исследований; только со временем выяснится, насколько успешно эти методы могут применяться на практике.

ЗАДАЧИ

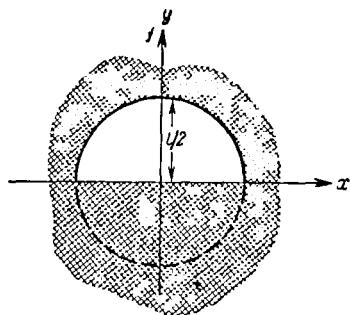
6.1. Некогерентная оптическая система имеет зрачок в виде квадрата со стороной l . Квадратный экран со стороной $l/2$ расположен в центре зрачка, как показано на фиг. 54.

а) Построить график оптической передаточной функции при наличии и в отсутствие экрана.

б) Изобразить вид оптической передаточной функции в предельном случае, когда размер экрана приблизительно равен размеру выходного зрачка.



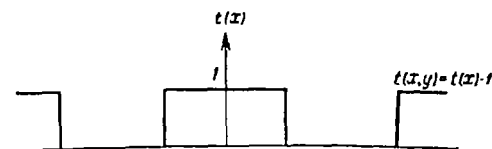
Фиг. 54.



Фиг. 55.

6.2. Некогерентная оптическая система имеет круглый выходной зрачок диаметром l . Диафрагма в виде полуплоскости помещена в выходной зрачок, в результате чего он принимает форму, показанную на фиг. 55. Найти выражения для оптической передаточной функции вдоль осей f_x и f_y .

6.3. Изображение предмета с амплитудным коэффициентом пропускания в виде прямоугольной волны (фиг. 56) образуется при помощи линзы, имеющей круговой зрачок. Фокусное расстояние линзы равно 10 см , частота первой гармоники прямоугольной волны равна 1000 период/см , расстояние до предмета равно 20 см , длина волны 10^{-4} см . Найти минимальный диаметр линзы, при котором будет еще заметно некоторое изменение интенсивности в плоскости изображения для случаев а) когерентного освещения предмета, б) некогерентного освещения предмета.



Фиг. 56.

6.4. Синусоидальный тест-объект с амплитудным коэффициентом пропускания

$$t(x, y) = \frac{1}{2} + \frac{1}{2} \cos 2\pi \tilde{f}x$$

помещен перед когерентной оптической системой. Предполагая, что частота \tilde{f} достаточно мала, чтобы быть пропущенной системой, и пренебрегая увеличением и общим ослаблением системы

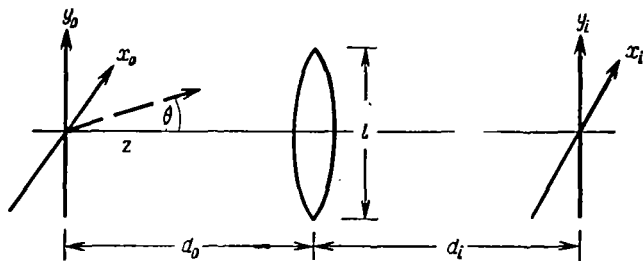
а) найти распределение интенсивности в плоскости изображения для системы, свободной от aberrаций;

б) показать, что то же самое распределение интенсивности получается в бесконечном множестве плоскостей пространства изображения, не соответствующих плоскости фокусировки.

6.5. Синусоидальная амплитудная решетка с амплитудным пропусканием

$$t(x, y) = \frac{1}{2} + \frac{1}{2} \cos 2\pi\tilde{f}x$$

расположена перед собирающей круглой линзой (диаметр l , фокусное расстояние f) и освещается монохроматической плоской волной, распространяющейся под углом θ в плоскости x_0z , как показано на фиг. 57.



Фиг. 57.

а) Каков фурье-спектр распределения амплитуды света, пропущенного решеткой?

б) Полагая $d_0 = d_i = 2f$, определить максимальный угол θ , для которого в плоскости изображения будет заметно некоторое изменение интенсивности.

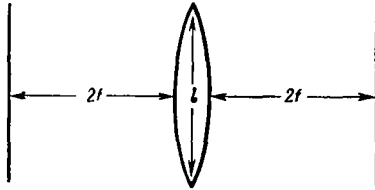
в) Предполагая, что этот максимальный угол известен, определить распределение интенсивности в плоскости изображения и сравнить его с соответствующим распределением интенсивности для $\theta = 0$.

г) Предполагая, что θ принимает максимальное значение, определить максимальную частоту \tilde{f} , при которой изменения интенсивности еще будут наблюдаться в плоскости изображения. Сравнить эту частоту с граничной частотой при $\theta = 0$.

6.6. Используя соотношение (6.38), показать, что полученная в приближении геометрической оптики ОПФ для нефокусированной системы с квадратным зрачком равна

$$\Phi(f_x, f_y) = \text{sinc} \left[\frac{8\omega}{\lambda} \left(\frac{f_x}{2f_0} \right) \right] \text{sinc} \left[\frac{8\omega}{\lambda} \left(\frac{f_y}{2f_0} \right) \right],$$

где w — максимальная погрешность в оптической длине пути в направлении осей x или y , а f_0 — предельная частота при когерентном освещении.



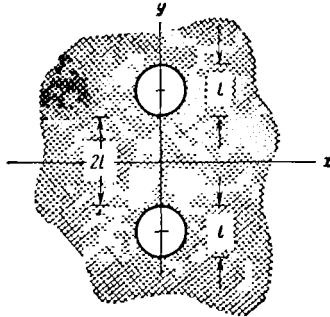
Фиг. 58.

6.7. Предмет имеет коэффициент пропускания по интенсивности

$$\tau(x, y) = \frac{1}{2} (1 + \cos 2\pi \tilde{f} x)$$

и создает постоянный однородный фазовый сдвиг в плоскости предмета. Изображение этого предмета формируется при помощи положительной линзы диаметром l и фокусным расстоянием f , как показано на фиг. 58. Сравнить максимальные частоты \tilde{f} , пропущенные системой, для случаев когерентного и некогерентного освещения.

6.8. Для некогерентной системы, зрачок которой имеет форму, показанную на фиг. 59, изобразить сечения оптической передаточной функции плоскостями $\tilde{f}_x = \text{const}$ и $\tilde{f}_y = \text{const}$ и отметить значения предельных частот.



Фиг. 59.

6.9. Все частотные составляющие амплитудного коэффициента пропускания некоторого предмета лежат в плоскости частот на

окружностях, радиусы которых определяются соотношением

$$\rho_m = \sqrt{2ma}, \quad m = 0, 1, 2, 3, \dots$$

Предполагая, что предмет освещается плоской однородной волной, и пренебрегая дифракционными эффектами, обусловленными конечными размерами предмета, показать, что такой предмет является «самонзображающимся» в том смысле, что за предметом без линз образуются не периодически повторяющиеся идеальные изображения.

6.10. Величину aberrаций некогерентной системы очень удобно характеризовать фактором резкости Стрела \mathcal{D} , который определяется как отношение интенсивности света в максимуме изображения точечного источника (импульсный отклик) к максимуму интенсивности в изображении точечного источника тем же прибором в отсутствие aberrаций. Доказать, что фактор \mathcal{D} равен нормированной площади под оптической передаточной функцией прибора с aberrациями, т. е.

$$\mathcal{D} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathfrak{D}(f_x, f_y) |_{\text{с aberrациями}} df_x df_y}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathfrak{D}(f_x, f_y) |_{\text{без aberrаций}} df_x df_y}.$$

ЛИТЕРАТУРА

1. Duffieux P. M., L'Intégrale de Fourier et ses Applications à L'Optique, Besançon, 1946.
2. Schade O. H., RCA Rev., 9, 5 (Part I), 245 (Part II), 490 (Part III), 653 (Part IV) (1948).
Электрооптические характеристики телевизионных систем.
3. O'Neill E. L., Introduction to Statistical Optics, London, 1963. (Имеется перевод: Э. О'Нейл, Введение в статистическую оптику, изд-во «Мир», 1966.)
4. Françon M., Modern Applications of Physical Optics, New York, 1963.
5. Linfoot E. H., Fourier Methods in Optical Image Evaluation, London, 1964.
6. Stroke G. W., An Introduction to Coherent Optics and Holography, New York, 1966. (Имеется перевод: Дж. Строк, Введение в когерентную оптику и голографию, изд-во «Мир», 1967.)
7. Abbe E., Archiv. Mikroskopische Anat., 9, 413 (1873).
8. Rayleigh, Phil. Mag. (5), 42, 167 (1896).
О теории оптических изображений, в особенности в микроскопе.
9. Bergan M. J., Parrent G. B., Jr., Theory of Partial Coherence, New York, 1964.
10. Mandel L., Wolf E., Rev. Mod. Phys., 37, 231 (1965). [Имеется перевод: УФН, 87, вып. 3 (1965); 88, вып. 2, 4 (1966).] Когерентные свойства оптических полей.

11. Glauber R. J., Phys. Rev., 131, 2766 (1963).
Когерентные и некогерентные свойства полей излучения.
12. Schwartz M., Information, Transmission, Modulation and Noise, New York, 1959, p. 285.
13. Considine P. S., Journ. Opt. Soc. Am., 56, 1001 (1966).
Влияние когерентности на оптические системы.
14. Oliver B. M., Proc. IEEE, 51, 220 (1963).
Зернистость и случайная дифракция.
15. Goodman J. W., Proc. IEEE, 53, 1688 (1965).
Эффект мерцания мишени при оптической локации.
16. Skipner T. J., Journ. Opt. Soc. Am., 53, 1350A (1963).
Эффект шероховатости поверхности при изображении в когерентном свете.
17. di Francia, Toraldo G., Nuovo Cimento, Suppl. (9), 9, 426 (1952).
Сверхнаправленные антенны и оптическая разрешающая способность.
18. Harris J. L., Journ. Opt. Soc. Am., 54, 931 (1964).
Дифракция и разрешающая способность.
19. Barnes C. W., Journ. Opt. Soc. Am., 56, 575 (1966).
Восстановление предмета в дифракционно ограниченной оптической системе.
20. Guillemin E. A., The Mathematics of Circuit Analysis, New York, 1949, p. 288.
- 21*. Марешаль А., Франсон М., Структура оптического изображения, изд-во «Мир», 1964.
- 22*. Королев А. Н., УФН, 96, вып. 2 (1968).
Повышение разрешающей способности оптических систем.

ГЛАВА 7

ПРОСТРАНСТВЕННАЯ ФИЛЬТРАЦИЯ И ОПТИЧЕСКАЯ ОБРАБОТКА ИНФОРМАЦИИ

В предыдущей главе мы убедились в несомненной плодотворности концепции линейных систем при анализе оптических устройств. Однако если бы эта концепция была полезна *только* для целей анализа, то она играла бы в современной оптике значительно меньшую роль, чем это оказывается в действительности. Фактическое ее значение можно понять только после рассмотрения удивительных возможностей *синтеза* с помощью оптических систем.

Существует много примеров, когда применение концепции линейных систем к задачам проектирования оптических устройств принесло большую пользу. Одно из направлений применения теории линейных систем связано с использованием пространственно-частотного рассмотрения для усовершенствования различных типов оптических приборов. Примеры проблем такого рода в их историческом аспекте будут обсуждаться в § 1.

Существуют и другие в такой же мере важные применения, которые непосредственно не относятся к системам, формирующим изображение как таковое, и их лучше отнести к общей области *обработки информации*. Эти применения основаны на способности оптических систем выполнять общие линейные преобразования поступающих на вход данных. Объем данных в некоторых случаях бывает настолько большим, что существенно превышает возможности наблюдателя. Тогда для *уменьшения* количества данных решающую роль будет играть линейное преобразование, выделяющее нужные группы таким образом, чтобы они могли восприниматься наблюдателем. Пример подобных приложений приведен в § 6 при обсуждении вопроса о *распознавании образов*. Возможны также случаи, когда большая часть данных имеет такую форму, что не может непосредственно восприниматься наблюдателем; тогда линейное преобразо-

вание может придать им нужный вид. В качестве примера можно назвать обработку данных антенн синтетической апертуры, что рассматривается в § 7.

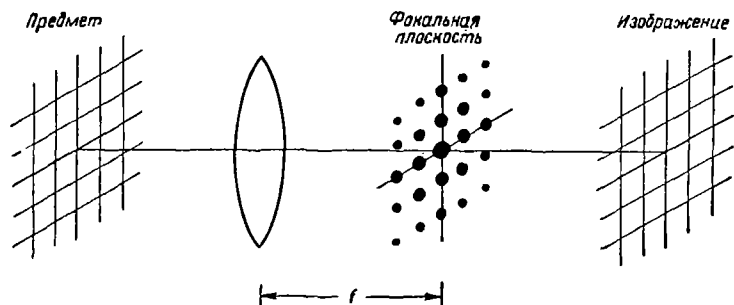
Круг вопросов, связанных с оптической обработкой информации, слишком широк, чтобы его можно было всесторонне рассмотреть всего в одной главе. Действительно, в настоящее время существует уже несколько книг, посвященных исключительно этой теме (см., например, [1, 2]). В настоящей главе мы разберем только несколько общих принципов и приведем примеры, позволяющие проиллюстрировать эти принципы. При этом, несомненно, из нашего поля зрения выпадет множество интересных и полезных методов, но читатель, интересующийся затронутым вопросом, может составить более полное представление о предмете, обратившись к цитированной литературе.

§ 1. ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ

Можно сказать, что история метода синтеза Фурье начинается с того момента, когда стало применяться сознательное воздействие на спектр изображения. Первое сообщение об экспериментах такого рода было опубликовано Аббе [6.7] в 1873 г., а затем Портером [3] в 1906 г. В обоих случаях конечной целью экспериментов была проверка созданной Аббе теории формирования изображений в микроскопе и исследование пределов ее применимости. Принимая во внимание изящество и простоту этих экспериментов, мы рассмотрим их более подробно.

Эксперимент Аббе — Портера

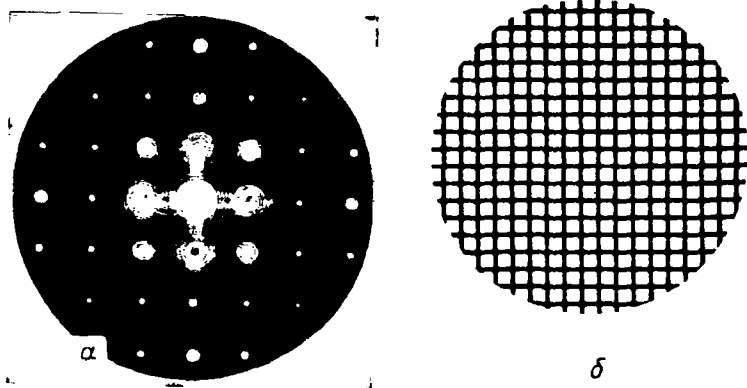
Эксперименты, осуществленные Аббе и Портером, позволяют наглядно представить не только механизм формирования изображения в когерентной системе, но и основные принципы анализа Фурье. Идея этих экспериментов иллюстрируется схемой на фиг. 60. Предмет, которым служит сетка из тонкой проволоки, освещается когерентным светом. В задней фокальной плоскости линзы получается фурье-спектр периодической сетки.



Фиг. 60. Схема эксперимента Аббе — Портера.

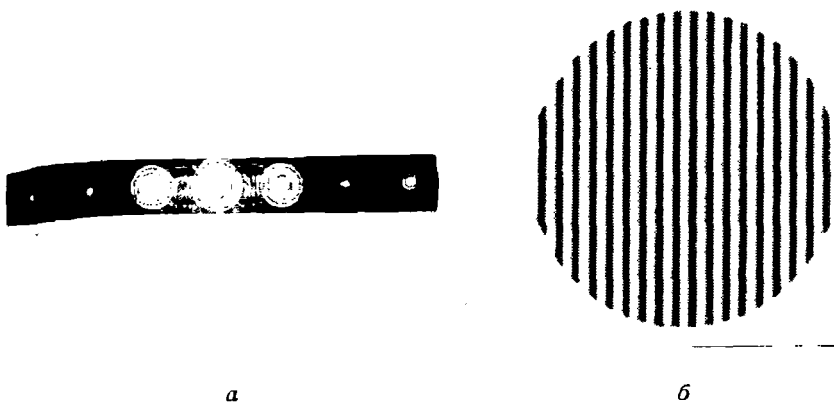
Различные фурье-компоненты, прошедшие через линзу, складываясь, дают в плоскости изображения точную копию решетки. Помещая в фокальную плоскость различные препятствия (например, ирисовую диафрагму, щель, маленький экран), можно непосредственно различным образом воздействовать на спектр изображения.

На фиг. 61, *а* приведена фотография исходного никак не измененного спектра сетки; на фиг. 61, *б* — соответ



Фиг. 61. Фотографии исходной сетки и ее спектра.

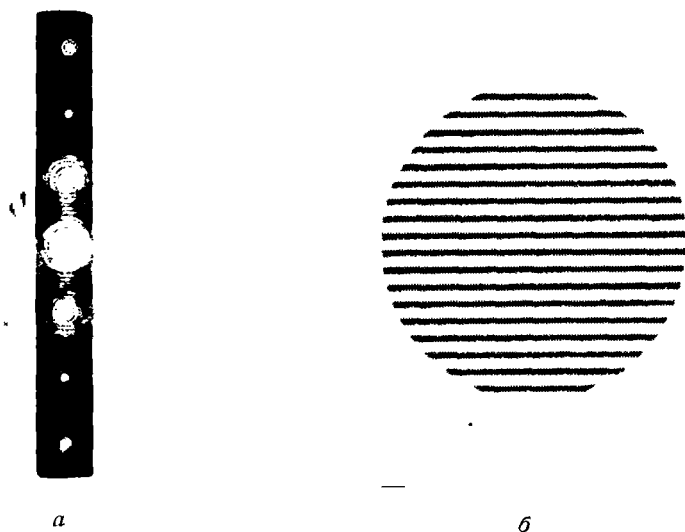
а — спектр; *б* — изображение.



Фиг. 62. Спектр сетки, отфильтрованный горизонтальной щелью (а), и соответствующее изображение (б).

ствующее изображение. Из-за периодичности предмета в фокальной плоскости появляется набор отдельных спектральных компонент; ширина каждой из них определяется размерами круглой оправы, в которую вставлена сетка. Яркие пятна вдоль горизонтальной оси в фокальной плоскости соответствуют комплексным экспоненциальным компонентам, направленным горизонтально (ср. фиг. 1); яркие пятна вдоль вертикальной оси соответствуют вертикально направленным комплексным экспоненциальным компонентам. Внеосевые пятна соответствуют компонентам, направленным под соответствующим углом в плоскости предмета.

Возможности метода *пространственной фильтрации* хорошо видны на примере эксперимента, при котором в фокальную плоскость помещают узкую щель так, чтобы через нее проходил только один ряд спектральных компонент. На фиг. 62, а показан спектр, получающийся при использовании горизонтальной щели. Соответствующее изображение, приведенное на фиг. 62, б, содержит только *вертикальную* структуру сетки; отсюда с очевидностью следует, что именно горизонтально направленные комплексные экспоненциальные компоненты дают вклад в



Фиг. 63. Спектр сетки, отфильтрованный вертикальной щелью (а) и соответствующее изображение (б).

вертикальную структуру изображения. Горизонтальная структура совершенно исчезает.

Если повернуть щель на 90° так, чтобы через нее проходил лишь вертикальный ряд (фиг. 63, а), то получающееся изображение содержит только горизонтальную структуру (фиг. 63, б). Легко обнаружить и другие интересные эффекты. Например, если в фокальную плоскость поместить ирисовую диафрагму и установить ее так, чтобы через нее проходила только осевая фурье-компонента, то при последовательном расширении ирисовой диафрагмы можно шаг за шагом проследить фурье-синтез сетки. Если же убрать ирисовую диафрагму, а вместо нее на оси линзы в фокальной плоскости поместить маленький экран, который закрывал бы только центральный порядок, или компоненту «нулевой частоты», то мы получим изображение сетки с обращенным контрастом (см. задачу 7.1).

Фазово-контрастный микроскоп Цернике

Многие интересные для микроскопии объекты обладают высокой прозрачностью и поэтому совсем или почти совсем не поглощают свет (например, бесцветные бактерии). При прохождении света через такой объект основным эффектом будет появление сдвига фазы, величина которого различна в разных точках. Подобный эффект, конечно, нельзя наблюдать с помощью обычного микроскопа и приемника, реагирующего на интенсивность света. Уже давно разработано много методов наблюдения таких объектов; к их числу относятся интерферометрический способ, *метод темного поля*, в котором используется небольшой экран, закрывающий только центральную спектральную компоненту нулевого порядка (см. задачу 7.2), и *метод свилей*, в котором устраняются все спектральные компоненты по одну сторону от нулевого порядка (см. задачу 7.3). Все эти методы обладают одним и тем же недостатком: наблюдаемые изменения интенсивности *нелинейно* связаны со сдвигом фазы и поэтому не могут служить непосредственной характеристикой вариаций толщины объекта.

В 1935 г. Фриц Цернике [4] предложил новый *фазово-контрастный* метод, основанный на принципах пространственной фильтрации. Достоинство этого метода заключается в том, что наблюдаемая интенсивность света (при определенных условиях, которые будут рассмотрены в дальнейшем) *линейно* связана со сдвигом фазы, вносимым объектом. Это один из первых результатов, полученных путем использования идеи о синтезе, поэтому мы остановимся на нем подробнее.

Предположим, что прозрачный объект с амплитудным коэффициентом пропускания

$$t(x, y) = e^{i\varphi(x, y)} \quad (7.1)$$

освещен когерентно в оптической системе. Для упрощения математических расчетов допустим, что увеличение равно единице и что можно пренебречь ограниченностью размеров входного и выходного зрачков системы. Кроме того, необходимое условие получения линейной связи между сдвигом фазы и интенсивностью состоит в том,

что сдвиг фазы φ должен быть много меньше 1 рад . В таком случае амплитудный коэффициент пропускания можно аппроксимировать выражением

$$t(x, y) \approx 1 + j\varphi(x, y). \quad (7.2)$$

В этом приближении мы пренебрегаем членами порядка φ^2 и выше. Заметим, что первый член в (7.2) соответствует интенсивной волне, проходящей через образец без изменения, в то время как второй член отвечает более слабому дифрагированному свету, отклоняющемуся от оси системы.

Интенсивность изображения, получающегося в обычном микроскопе, можно записать в виде

$$I_i \approx |1 + j\varphi|^2 \approx 1,$$

где, следуя нашему приближению, мы пренебрегли членом φ^2 . Цернике понял, что дифрагированный свет нельзя наблюдать из-за того, что он *сдвинут по фазе на 90* по отношению к интенсивному фону, и что если бы удалось как-то видоизменить указанное фазовое соотношение, то интерференция этих двух членов могла бы обуславливать видимые изменения интенсивности изображения. Принимая во внимание, что фон фокусируется в точку, лежащую на оси в фокальной плоскости, а дифрагированный свет (содержащий более высокие пространственные частоты) отклоняется в сторону от фокуса, он предложил поместить в фокальной плоскости фазовую пластинку, изменяющую соотношение фаз между сфокусированным и дифрагированным светом.

Фазовая пластинка может представлять собой стеклянную подложку, на которую нанесено маленькое прозрачное диэлектрическое пятнышко. Пятнышко помещается в центр фокальной плоскости и имеет такую толщину и показатель преломления, что изменение фазы сфокусированного света по отношению к изменению фазы дифрагированного света составляет $\pi/2$ или $3\pi/2 \text{ рад}$. В первом случае интенсивность в плоскости изображения равна

$$I_i = \left| \exp\left(j \frac{\pi}{2}\right) + j\varphi \right|^2 = |j(1 + \varphi)|^2 \approx 1 + 2\varphi, \quad (7.3)$$

а во втором случае имеем

$$I_1 = \left| \exp\left(j \frac{3\pi}{2}\right) + j\varphi \right|^2 = | -j(1 - \varphi) |^2 \approx 1 - 2\varphi. \quad (7.4)$$

Таким образом, интенсивность изображения становится линейной функцией сдвига фазы φ . Когда фаза фона изменяется на $\pi/2$, чему соответствует соотношение (7.3), мы имеем *положительный фазовый контраст*, а при изменении фазы на $3\pi/2$ говорят об *отрицательном фазовом контрасте* [соотношение (7.4)]. Кроме того, контраст изображения можно улучшить, если пятнышко сделать частично поглощающим (см. задачу 7.4).

Метод фазового контраста — один из методов преобразования пространственной фазовой модуляции в пространственную модуляцию интенсивности. Для читателя, знакомого с электроникой, вероятно, будет неинтересно узнать, что через год после открытия Цернике Армстронг [5] предложил совершенно аналогичный метод преобразования амплитудно-модулированных электромагнитных волн в фазово-модулированные. Как мы уже убедились в гл. 6 и увидим в настоящей главе, за годы, прошедшие после этих открытий, между радиоэлектроникой и оптикой установились более тесные связи.

Улучшение качества фотографий. Метод Марешаля

В начале 50-х годов сотрудники Оптического института Парижского университета начали активно исследовать возможности использования метода когерентной пространственной фильтрации для улучшения качества фотографий. Наиболее существенными были исследования Марешаля, которые привели к дальнейшему возрастанию интереса к вопросам оптической обработки информации.

Марешаль понимал, что недостатки фотографий являются следствием соответствующих недостатков передаточной функции оптической системы, с помощью которой получают эти фотографии. В дальнейшем он установил, что если фотографические диапозитивы установить в предметной плоскости другой (когерентной) системы, то,

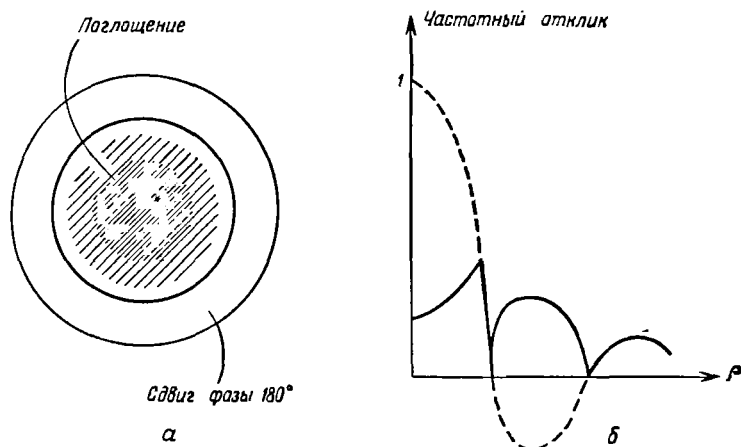
помещая соответствующие поглощающие и фазосдвигающие пластинки в фокальной плоскости, можно создать *компенсирующий фильтр*, который по крайней мере частично устраняет нежелательные дефекты. Хотя передаточная функция исходной оптической системы может быть неудовлетворительной, при использовании произведения этой передаточной функции на передаточную функцию компенсирующей системы должен получаться более правильный частотный отклик.

Маршалль и его сотрудники нашли множество различных способов улучшения качества фотографий. Было показано, например, что контраст малых деталей изображений можно улучшить, если просто ослабить низкочастотные компоненты в спектре объекта. Значительные успехи были также достигнуты и в отношении устранения размытости изображений. В этом случае исходная оптическая система расфокусировалась так, что получался импульсный отклик, состоящий в приближении геометрической оптики из одинаковых кружков. Соответствующая передаточная функция имеет вид

$$\mathfrak{H}(\rho) \approx 2 \frac{J_1(a\rho)}{a\rho},$$

где a — постоянная величина, а $\rho = \sqrt{f_x^2 + f_y^2}$. Чтобы получить компенсирующий фильтр, в фокальную плоскость фильтрующей системы помещались поглощающая и фазовая пластинки. Поглощающая пластинка ослабляла низкочастотный интенсивный пик функции \mathfrak{H} , а фазовая пластинка изменяла фазу первого отрицательного максимума этой функции на 180° . Исходная и компенсированная передаточные функции приведены на фиг. 64.

В качестве другого примера укажем, что с помощью пространственного фильтра можно устранить периодическую структуру, связанную с процессом передачи полутоновых изображений (фотографий) при типографской печати (например, в газетах). Процесс передачи полутонов во многом аналогичен процедуре периодической выборки, рассмотренной в гл. 2, § 3. Спектр типографски отпечатанного таким способом рисунка содержит периодическую структуру, во многих отношениях сходную с



Фиг. 64. Компенсация нерезкости изображения.

a — фильтр, помещаемый в фокальной плоскости; *b* — первоначальная передаточная функция (пунктирная кривая) и та же функция после компенсации (сплошная кривая).

показанной на фиг. 5. Если в фокальной плоскости фильтрующей системы поместить ирисовую диафрагму, то можно добиться того, чтобы через нее проходили только гармоники, близкие к нулевой частоте. Таким образом удается устранить периодическую структуру на рисунке, тогда как вся желательная структура изображения сохраняется.

Возникновение подхода, основанного на теории связи¹⁾

В начале 50-х годов текущего столетия стало очевидно, что взаимный обмен между теорией связи и оптикой сулит большие выгоды. Многие проблемы, с которыми сталкиваются исследователи, занимающиеся оптикой, имеют очень большое сходство с проблемами

¹⁾ Мы не будем пытаться перечислить здесь все ранние работы, способствовавшие сближению этих гочек зрения, а в основном сосредоточим внимание на публикациях, которые представляют наибольший интерес для читателя, имеющего подготовку в области электроники.

оптимальной фильтрации, обнаружения и оценки в теории связи. Установлению такого обмена в большой мере способствовали работы Петера Элиса, занимающегося теорией связи, и его сотрудников Грея и Робинсона, в особенности опубликованная в 1952 г. статья «Анализ Фурье в оптике» [6], а также статья Элиса «Оптика и теория связи» [7] (1953). Настоящее сближение двух точек зрения началось после того, как в 1956 г. физик О'Нейл опубликовал свою статью «Пространственная фильтрация в оптике» [8]. Точнее говоря, сближению оптики и теории связи вообще в значительной мере способствовало то большое впечатление, которое произвели на ученых-оптиков идеи О'Нейла, высказанные им в его научных работах и лекциях. Такая взаимосвязь была и продолжает оставаться плодотворной, о чем свидетельствует большой интерес к работам, представленным на состоявшемся в 1960 г. симпозиуме «Вопросы теории связи и информации в современной оптике» [9], и высокий уровень этих работ. С тех пор слияние этих двух точек зрения стало настолько полным, что иногда бывает уже трудно сказать, следует ли ту или иную часть работы печатать в оптическом или электротехническом журнале.

Применение когерентной оптики к более общим задачам обработки информации

Если в начале 50-х годов часть физиков стала все больше и больше понимать, что некоторые аспекты радиоэлектроники можно с успехом использовать в оптике, то в конце 50-х и начале 60-х годов многие радиоинженеры стали представлять, что системы пространственной фильтрации можно успешно применять при решении более общих задач обработки информации. Возможности когерентной фильтрации особенно очевидны в области обработки радиолокационных сигналов. Практически эти возможности впервые использовал Катрона с сотрудниками (Радиолокационная лаборатория Мичиганского университета). В 1960 г. Мичиганская группа опубликовала статью «Оптическая обработка информации и системы фильтрации» [10], которая вызвала большой интерес к этой методике как среди радиоинженеров, так и

среди физиков. Одним из наиболее успешных приложений метода когерентной фильтрации в области радиолокации является обработка результатов, полученных от антенн «синтетической апертуры» [11]. На этом вопросе мы кратко остановимся в § 7. Обзор литературы начиная с середины 60-х годов показывает, что методика когерентной обработки результатов нашла широкое применение в самых различных областях, начиная, например, от фурье-спектроскопии [12] до анализа сейсмических волн [13].

§ 2. ФОТОГРАФИЧЕСКАЯ ПЛЕНКА

Фотографическая пленка играла и продолжает играть важнейшую роль в оптических и фильтрующих системах. При фильтрации она может выполнять три очень важные функции. Во-первых, она может служить материалом, на котором записываются вводимые в оптическую систему подлежащие обработке исходные данные. Во-вторых, она может служить материалом для фильтра, определяющего характеристику системы. Например, она может быть ослабителем, который необходимо вводить в частотную плоскость, чтобы получить нужную передаточную функцию. В-третьих, она может служить материалом для записи выходных сигналов в системе обработки данных. Часто она выполняет все три функции.

Существует много других материалов, применяемых с некоторыми ограничениями для замены пленки в одной или нескольких ее функциях. К числу таких материалов относятся: жидкие акустические элементы, термопластические ленты, фотохромные пленки и стекла и др. Совершенно очевидно, что роль каждого из этих материалов будет расти, но представляется также весьма маловероятным, что пленка будет когда-нибудь совершенно вытеснена как основной элемент систем оптической обработки информации.

По этой причине мы немного отклонимся от нашей основной темы и рассмотрим важнейшие свойства фотопленки. Более исчерпывающую информацию читатель может найти в работах [14, 15].

Физические основы экспонирования

Неэкспонированная фотографическая пленка или пластинка в основном состоит из множества крошечных зерен галоидного серебра, взвешенных в желатине. Последний в свою очередь нанесен на твердую «основу», представляющую собой обычно ацетат у пленок или стекло у пластинок. Когда светочувствительный материал экспонируется, т. е. подвергается воздействию света, зерна галоидного серебра поглощают энергию оптических волн и в них происходят сложные физические изменения. Обнаружено, что зерна, поглотившие достаточное количество энергии, содержат крошечные частички металлического серебра; эти частички называются *центрами проявления*. Затем экспонированная пленка подвергается химической обработке — *проявлению*. При этом в каком-либо зерне одного крошечного центра проявления может явиться причиной превращения всего галоидного серебра в этом зерне в металлическое серебро. Зерна, не содержащие центров проявления, не подвергаются таким изменениям. Затем пленка «фиксируется», т. е. подвергается химической обработке, при которой зерна галоидного серебра удаляются, а металлическое серебро остается. Таким образом в дальнейшем предотвращается возможность превращения не подвергавшихся действию света зерен в металлическое серебро. Естественно, зерна серебра почти непрозрачны на оптических частотах; следовательно, коэффициент непрозрачности проявленной пленки или пластинки будет зависеть от плотности зерен серебра на каждом участке пленки.

Определение основных величин

Экспозиция E , которой подвергается пленка или пластинка, определяется как *энергия, приходящаяся на единицу площади* в каждой точке светочувствительной поверхности. Если интенсивность падающего света равна \mathcal{I} ¹⁾, а продолжительность экспозиции составляет

¹⁾ В дальнейшем мы будем использовать букву \mathcal{I} для обозначения интересующей нас интенсивности до или во время экспозиции и букву I для обозначения интенсивности после проявления.

T (сек), то мы имеем

$$E = \mathcal{U}T.$$

Коэффициент пропускания по интенсивности $\tau(x, y)$ проявленного диапозитива определяется следующим образом:

$$\tau(x, y) = \frac{\left\{ \begin{array}{l} \text{Прошедшая интенсивность в точке } (x, y) \\ \text{Падающая интенсивность в точке } (x, y) \end{array} \right\}_{\text{лок. средн.}}}{}$$

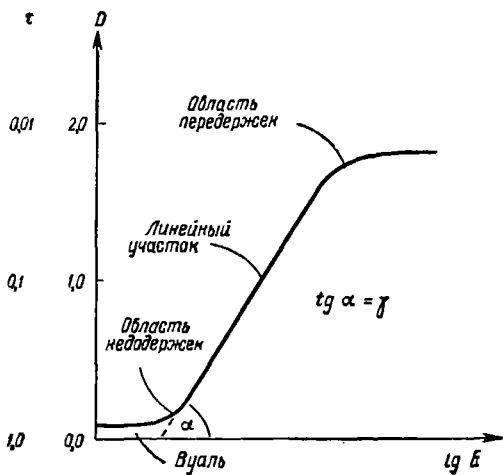
где локальное усреднение производится по области, размеры которой велики по сравнению с размерами зерен, но малы по сравнению с величиной участка, на котором интенсивность прошедшего света заметно меняется.

В 1890 г. Хертер и Дрифилд опубликовали классическую работу, в которой показали, что величина $\lg(1/\tau)$ пропорциональна массе серебра на единице площади проявленного диапозитива. Соответственно они определили *фотографическую плотность почернения* D как

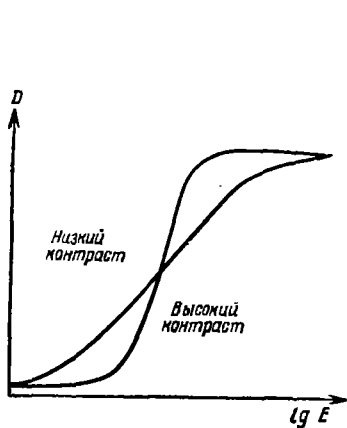
$$D = \lg\left(\frac{1}{\tau}\right). \quad (7.5)$$

Для описания светочувствительных свойств фотопленок, как правило, используется график зависимости плотности почернения от логарифма экспозиции, обычно называемой кривой *Хертера и Дрифилда* (или для краткости кривой X и D) ¹⁾. На фиг. 65 приведена типичная кривая почернения негативных фотоматериалов. Если экспозиция меньше некоторой определенной величины, то плотность почернения не зависит от экспозиции и имеет минимальное значение, называемое *вуалью*. В «области недодержек» при увеличении экспозиции начинает расти и плотность почернения. Затем следует участок кривой, на котором плотность почернения прямо пропорциональна логарифму экспозиции; наклон этого линейного участка кривой характеризует *коэффициент контрастности* γ пленки. Наконец, в области, называемой

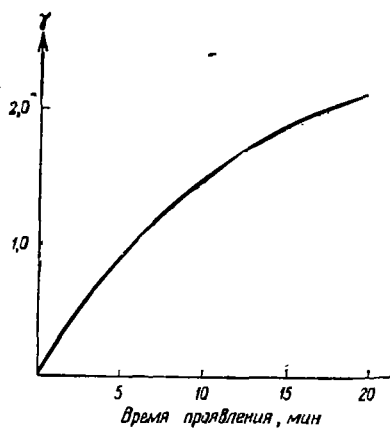
¹⁾ В русской литературе обычно употребляется термин «характеристическая кривая», или «кривая почернения». Мы будем пользоваться последним термином. — *Прим. ред.*



Фиг. 65. Кривая Хертера и Дрифилда (кривая почернения).



Фиг. 66. Кривые почернения высоко- и низкоконтрастных пленок,



Фиг. 67. Зависимость контрастности пленки γ от времени проявления.

областью *передержек*, наступает насыщение и опять при увеличении экспозиции плотность почернения не изменяется.

В обычной фотографии, как правило, используется линейный участок кривой почернения. Пленки с большим значением γ называются *высококонтрастными*, а пленки с малыми γ — *низкоконтрастными*. На фиг. 66 приведены кривые почернения низко- и высококонтрастных пленок. Величина γ зависит от трех основных факторов: во-первых, от марки эмульсии¹⁾ (например, у низкоконтрастных пленок Plus-X и Tri-X $\gamma = 1$ или меньше, а у пленки High Contrast Copy $\gamma = 2$ или 3); во-вторых, от типа применяемого проявителя; в-третьих, от времени проявления. На фиг. 67 показана типичная зависимость γ от времени проявления. При разумном выборе пленки, проявителя и времени проявления можно добиться требуемой величины γ с нужной степенью точности.

Пленка в некогерентной оптической системе

Во многих случаях пленку можно рассматривать как элемент, который осуществляет преобразование распределения интенсивности света \mathcal{I} , падающего на пленку при экспозиции, в распределение интенсивности света I , проходящего через нее после проявления. Такая точка зрения особенно удобна, когда пленка используется в качестве одного из элементов в некогерентной оптической системе. Мы детально рассмотрим природу такого рода преобразования.

Если используется линейный участок кривой почернения, то плотность D можно записать в виде

$$D = \gamma_n \lg E - D_0 = \gamma_n \lg (\mathcal{I}T) - D_0, \quad (7.6)$$

где γ_n — наклон линейного участка кривой, а $-D_0$ — значение D в точке, где прямолинейное продолжение кривой (если пренебречь областью недодержек) пересекает ось D . Индекс « n » означает, что мы имеем дело с негативом.

¹⁾ *Эмульсией* называют только светочувствительную часть пленки (без основы).

Интенсивность света, действовавшего при экспонировании, можно связать с коэффициентом пропускания после проявления, принимая во внимание определение фотোগрафической плотности почернения

$$D = \lg \left(\frac{1}{\tau_n} \right).$$

После подстановки этого определения в (7.6) получим

$$\lg \tau_n = -\gamma_n \lg (\mathcal{I}T) + D_0.$$

или, в другой записи,

$$\tau_n = 10^{D_0} (\mathcal{I}T)^{-\gamma_n}.$$

Окончательно

$$\tau_n = K_n \mathcal{I}^{-\gamma_n}, \quad (7.7)$$

где K_n — положительная константа. Отметим, что преобразование \mathcal{I} в τ_n , описываемое этим соотношением, имеет нелинейный характер для любых положительных значений γ пленки.

Можно добиться, чтобы показатель степени в соотношении, связывающем коэффициент пропускания с интенсивностью света, падающего на пленку при экспонировании, был положительным, хотя для этого, как правило, требуется двухступенчатый процесс. На первом этапе получают обычным образом негатив, на втором этапе свет, прошедший через негатив, используется для экспонирования другой пленки; в результате получается диапозитив. Для количественного рассмотрения этого процесса положим, что пропускание негатива в соответствии с (7.7) можно записать в виде

$$\tau_{n1} = K_{n1} \mathcal{I}^{-\gamma_{n1}}.$$

Если этот негатив находится в непосредственном контакте с другим куском неэкспонированной пленки и освещается светом с интенсивностью I_0 , то интенсивность света, действующего на вторую пленку, равна просто $\tau_{n1} I_0$ и результирующее пропускание будет

$$\tau_p = K_{n2} (I_0 \tau_{n1})^{-\gamma_{n2}} = K_{n2} I_0^{-\gamma_{n2}} K_{n1}^{-\gamma_{n2}} \mathcal{I}^{\gamma_{n1} \gamma_{n2}},$$

или

$$\tau_p = K_p \mathcal{I}^{\gamma_p}, \quad (7.8)$$

где K_p — положительная константа, а $\gamma_p = \gamma_{n1}\gamma_{n2}$ — полная контрастность двухступенчатого процесса. Очевидно, диапозитив дает линейное отображение интенсивностей тогда и только тогда, когда полная контрастность равна единице.

Таким образом, линейное преобразование интенсивности света, действовавшего при экспозиции, в интенсивность света, прошедшего через диапозитив после проявления, как мы видели, имеет место только при весьма специфических условиях. Тем не менее можно показать, что пленка обеспечивает линейное отображение *малых* изменений интенсивности для гораздо более широкого класса условий (см. задачу 7.5).

Пленка в когерентной оптической системе

Когда пленка используется в качестве одного из элементов *когерентной* оптической системы, правильнее рассматривать ее как элемент, который либо обеспечивает преобразование интенсивности света, действовавшего при экспозиции, в комплексную амплитуду света, проходящего через пленку после проявления, либо осуществляет преобразование комплексной амплитуды света, действовавшего при экспозиции, в комплексную амплитуду света, проходящего через пленку после проявления. Вторая точка зрения возможна, конечно, только в том случае, когда проявленная пленка освещается когерентным светом.

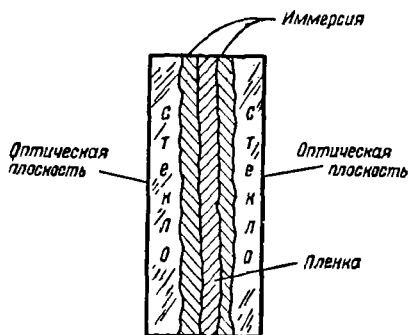
Так как, согласно обеим точкам зрения, комплексная амплитуда света, прошедшего через пленку, служит параметром когерентной системы, пропускание диапозитива необходимо описывать с помощью комплексного *амплитудного* коэффициента пропускания t . Сразу же представляется возможным определить t просто как положительный квадратный корень из коэффициента пропускания по интенсивности τ . Однако такое определение не учитывает фазовых сдвигов, появляющихся при прохождении света через пленку. Такие сдвиги фазы возникают из-за колебаний толщины пленки, которые могут быть обусловлены двумя различными причинами. Во-первых, толщина основы пленки меняется случайным

образом, т. е. основа не является оптически плоской. Во-вторых, установлено, что толщина эмульсии проявленного диапозитива меняется в зависимости от плотности серебра. Изменения толщины в последнем случае играют основную роль и сильно зависят от колебаний экспозиции, которой подвергалась пленка. Таким образом, окончательное выражение для амплитудного коэффициента пропускания пленки следует записать в виде

$$t(x, y) = + \sqrt{\tau(x, y)} \exp [j\varphi(x, y)], \quad (7.9)$$

где $\varphi(x, y)$ описывает распределение сдвигов фаз, получающиеся после прохождения светом диапозитива.

Во многих случаях желательно совершенно исключить всякие колебания толщины пленки, так как их чрезвычайно трудно контролировать. Влияние этих колебаний можно устранить с помощью устройства, называемого *жидкостным компенсатором*. Такое устройство имеет оптически плоские стенки, между которыми в виде сэндвича, как показано на фиг. 68, находятся пленка и иммерсия. Показатель преломления иммерсии должен быть каким-то компромиссным, так как он не может быть равен одновременно показателю преломления и основы, и эмульсии, и стекла (которое, как правило, плоское только с одной стороны). Однако при соответствующем выборе иммерсии оптическую длину пути через



Фиг. 68. Жидкостный компенсатор колебаний толщины пленки.

жидкостный компенсатор можно сделать приблизительно постоянной. Тогда амплитудный коэффициент пропускания пленки и компенсатора запишется в виде

$$t(x, y) = + \sqrt{\tau(x, y)}. \quad (7.10)$$

Если сдвиги фаз устранены, то, комбинируя (7.10) с (7.7) и (7.8), можно найти амплитудный коэффициент пропускания:

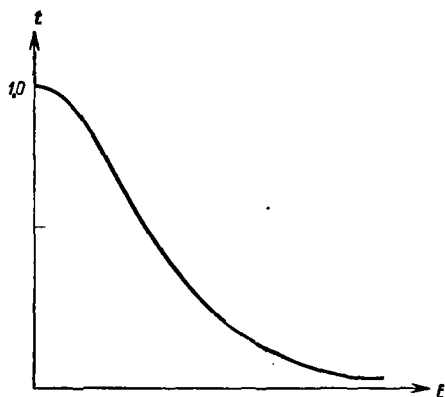
$$\begin{aligned} \text{Негатив:} \quad t_n &= k_n \mathcal{G}^{-\gamma_n/2} = k_n |U|^{-\gamma_n}, \\ \text{Диапозитив:} \quad t_p &= k_p \mathcal{G}^{\gamma_p/2} = k_p |U|^{\gamma_p}, \end{aligned} \quad (7.11)$$

где U — комплексная амплитуда поля при экспозиции, а

$$k_n = K_n^{1/2}, \quad k_p = K_p^{1/2}.$$

Как мы увидим из большого количества примеров, которые будут рассмотрены в последующих параграфах, часто бывает желательно, чтобы реакция пленки на комплексную амплитуду соответствовала *квадратичному* закону. Этого можно достигнуть различными способами. В частности, как следует из соотношения (7.11), можно изготовить диапозитив с полной контрастностью (γ_p), равной 2. Чтобы получить максимальный динамический диапазон экспозиций, в котором выполняется это соотношение, контрастность на первой ступени двухступенчатого процесса, как правило, выбирается меньше единицы (например, $\gamma_{n1} = 1/2$), а на второй больше 2 (например, $\gamma_{n2} = 4$), так, чтобы их произведение равнялось 2.

Однако для ограниченного динамического диапазона можно получить квадратичную зависимость как для позитива, так и для негатива с любой контрастностью. В этом легко убедиться, если отказаться от традиционного описания свойств пленки с помощью кривой почернения, а вместо этого построить непосредственно график зависимости амплитудного коэффициента пропускания от экспозиции (в линейном масштабе). Такое описание применял Марешаль (см. [9], стр. 48); им с успехом пользовался также Косма [16] при анализе влияния нелинейности пленки. На фиг. 69 показан график зависимости амплитудного коэффициента пропускания от



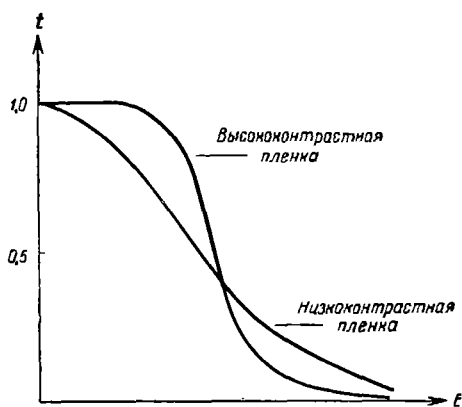
Фиг. 69. Типичная кривая зависимости пропускания от экспозиции для негатива.

экспозиции (t — E -кривая) для типичного негатива. Если рабочая точка пленки «смещена» в область линейного участка t — E -кривой, то для небольших изменений амплитуды квадратичный закон преобразования будет выполняться в некотором диапазоне изменения амплитуды. Таким образом, если обозначить через E_b экспозицию «смещения», а через t_b — соответствующее смещение коэффициента пропускания, то t — E -кривую на линейном участке можно аппроксимировать выражением

$$t_n \approx t_b + \beta(E - E_b) = t_b + \beta' |\Delta U|^2, \quad (7.12)$$

где β — наклон кривой в рабочей точке, ΔU — малое изменение амплитуды, а β' — произведение β на время экспозиции. Заметим, что для негатива β и β' — отрицательные числа.

На фиг. 70 сравниваются t — E -кривые высоко- и низкоконтрастных пленок. Мы видим, что кривые для высококонтрастных пленок имеют больший наклон, поэтому такие пленки более эффективно осуществляют превращение малых изменений экспозиции в изменения коэффициента пропускания. Однако увеличение эффективности сопровождается уменьшением динамического диапазона экспозиций, в котором t — E -кривая остается



Ф и г. 70. Кривые зависимости пропускания от экспозиции высоко- и низкоконтрастных пленок.

линейной. Интересно, что экспозиция смещения, для которой получается максимальный динамический диапазон, как правило, очень сильно сдвинута в область недодержек соответствующей кривой почернения.

Читателю, желающему получить дополнительную информацию о поведении фотопленки в когерентных оптических системах, мы рекомендуем обратиться к работам [17, 18].

Модуляционная передаточная функция

До настоящего момента мы молчаливо допускали, что любые изменения экспозиции, как бы ни был мал их пространственный масштаб, будут преобразовываться в изменения фотографической плотности почернения пленки в соответствии с кривой почернения эмульсии. На практике установлено, однако, что если пространственный масштаб изменений экспозиции слишком мал, то на диапозитиве может и *не получиться* соответствующих изменений плотности. В этом случае мы можем сделать весьма общее утверждение: данная пленка имеет ограниченный частотный отклик.

Как было отмечено ранее, в общем случае поведение пленки является существенно нелинейным, поэтому, конечно, некорректно связывать с отображением, осуществляемым пленкой, передаточную функцию. Однако физические явления, ограничивающие частотный отклик, по большей части линейны (например, рассеяние света эмульсии при экспонировании и химическая диффузия при проявлении). Следовательно, есть основания полагать, что линейные явления, ограничивающие частотный отклик, могут быть отделены от нелинейных свойств. Действительно, в большинстве случаев такой подход дает хорошие результаты. Поэтому, вообще говоря, можно рассматривать фотографическое отображение осуществляемое пленкой, как каскад линейного пространственно-инвариантного фильтра, имеющего некоторую передаточную функцию, и элемента с так называемой *нелинейностью без памяти*¹⁾, обладающего всеми рассмотренными выше свойствами²⁾.

Для измерения характеристик линейного фильтра, определяющего свойства пленки, подвергнем ее косинусоидальной по координате экспозиции:

$$E = E_0 + E_1 \cos 2\pi f x. \quad (7.13)$$

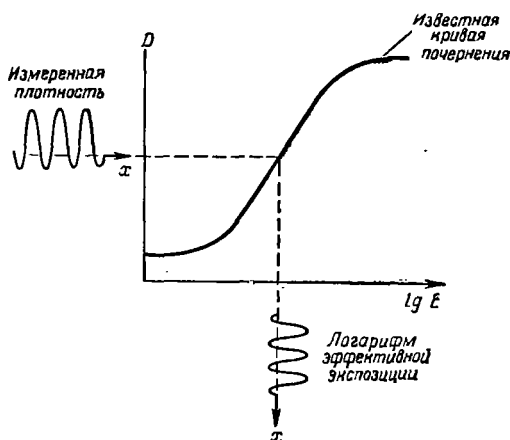
«Модуляция», связанная с такой экспозицией, определяется как отношение пикового значения переменной составляющей к уровню фона, т. е.

$$M_t = \frac{E_1}{E_0}. \quad (7.14)$$

Если в полученном диапозитиве измерить изменения плотности, то с помощью кривой почернения можно обратным преобразованием найти вычисленное, или «эффективное», косинусоидальное распределение экспозиции, как показано на фиг. 71. Модуляция $M_{\text{эфф}}$ распределения эффективной экспозиции всегда оказывается

¹⁾ То есть элемент, не вводящий никаких новых переходных процессов. — *Прим. ред.*

²⁾ Более точным представлением является каскад: линейный фильтр, элемент с нелинейностью без памяти и второй линейный фильтр. Однако для наших целей достаточно описанного выше простого представления.



Фиг. 71. Нахождение эффективного распределения экспозиции.

меньше, чем модуляция M_i действительного распределения экспозиции. В соответствии с этим модуляционная передаточная функция пленки определяется следующим образом:

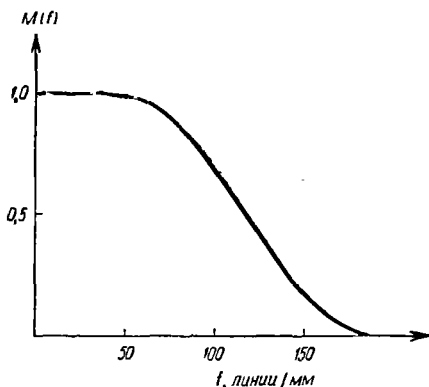
$$M(f) = \frac{M_{\text{эфф}}(f)}{M_i(f)},$$

где мы использовали запись, подчеркивающую зависимость от пространственной частоты экспозиции. Кроме того, распределение эффективной экспозиции зависит от детального вида функции рассеяния точки или импульсного отклика линейного фильтра, соответствующего пленке, поэтому оно может быть сдвинуто по фазе на $\Omega(f)$ по отношению к действительному распределению экспозиции. Таким образом, частотный отклик линейного каскада можно описать комплексной передаточной функцией

$$M(f) = M(f) e^{-j\Omega(f)}.$$

Следовательно, распределение эффективной экспозиции на входе нелинейного каскада можно записать в виде

$$E = E_0 + M(f) E_1 \cos [2\pi f x - \Omega(f)]. \quad (7.15)$$



Фиг. 72. Типичная модуляционная передаточная функция пленки.

На фиг. 72 изображена типичная частотная зависимость МПФ пленки. Диапазон частот, для которого получается заметный отклик, изменяется в широких пределах от пленки к пленке и зависит от размеров зерна, толщины эмульсии и других факторов. Например, пленка Plus-X дает разрешение приблизительно до 50 *линий/мм*, а спектроскопические пластинки Кодак 649F — по крайней мере до 2500 *линий/мм*.

§ 3. НЕКОГЕРЕНТНЫЕ СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ, ОСНОВАННЫЕ НА ГЕОМЕТРИЧЕСКОЙ ОПТИКЕ

Принцип действия значительной группы систем оптической обработки результатов характеризуется тем, что он полностью основан на законах геометрической оптики. Такие системы по большей части предназначены для работы с некогерентным освещением. Они сравнительно просты по конструкции и в обращении, но, как мы увидим, эта простота иногда достигается за счет ограничения эффективности обработки данных. Системы, которые будут обсуждаться ниже, следует рассматривать только как гипотетические конструкции.

Передача изображения

Одну из самых простых систем оптической обработки можно осуществить, если просто воспользоваться способностью линзы передавать изображения. Такую систему предложил еще в 1927 г. Э. Голдберг (Германия, Дрезден) и в Германии же получил на ее патент. Голдберг, получивший в 1931 г. патент и в США [19], полностью создавал возможность применения своего изобретения в области, которая известна теперь под названием *распознавание образов*.

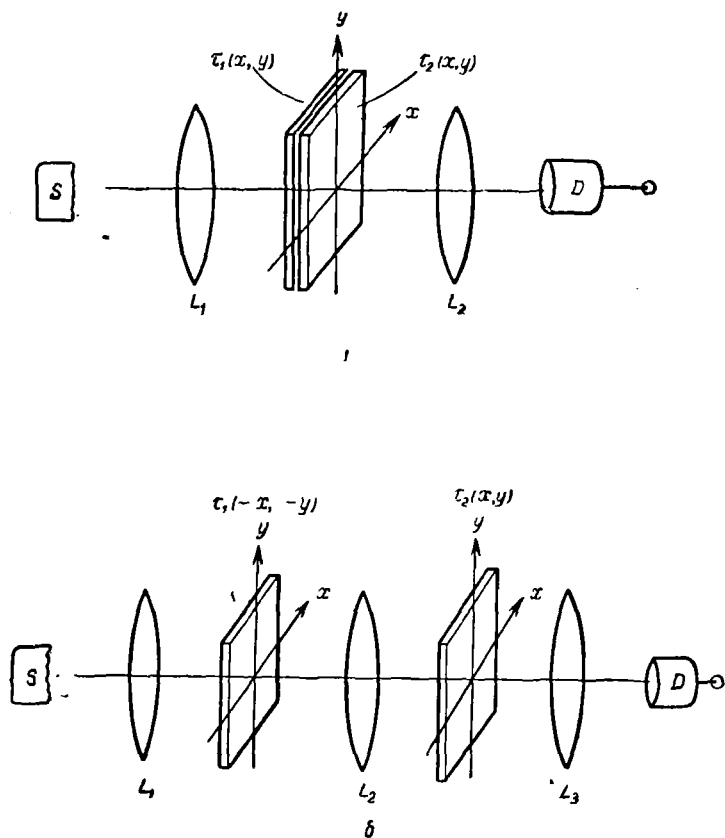
Система передачи изображения основана на весьма простых принципах. Если диапозитив с коэффициентом пропускания по интенсивности τ_1 отображается на другой диапозитив с коэффициентом пропускания τ_2 , то интенсивность света в каждой точке за вторым диапозитивом пропорциональна произведению $\tau_1\tau_2$. Для измерения *полной* интенсивности света, прошедшего через два диапозитива, можно воспользоваться фотодетектором, ток которого I будет определяться соотношением ¹⁾

$$I = k \int \int_{-\infty}^{\infty} \tau_1(x, y) \tau_2(x, y) dx dy. \quad (7.16)$$

На фиг. 73 показаны два способа осуществления этой операции. В методе, представленном на фиг. 73, а, линза L_1 отображает с увеличением равномерно светящийся некогерентный источник S на два диапозитива, помещенные вплотную друг к другу. Затем линза L_2 собирает свет, прошедший через диапозитив τ_2 , на маленький фотодетектор D . Фототок в этом случае описывается соотношением (7.16).

Тесный контакт диапозитивов приводит к трудностям механического характера в тех случаях, когда необходима быстрая замена одного или обоих диапозитивов. Поэтому целесообразно пространственно разделить диапозитивы, как показано на фиг. 73, б. Линза L_1 опять

¹⁾ Записывая бесконечные пределы интегрирования, мы полагаем, что ограниченность размеров диапозитивов уже учтена в определении τ_1 и τ_2 .



Фиг. 73. Системы для получения интегрального произведения.

дает увеличенное изображение источника на τ_1 . Линза L_2 отображает τ_1 на τ_2 , и свет, прошедший через τ_2 , собирается линзой L_3 на фотодетектор. Заметим, что диапозитив τ_1 должен быть перевернут соответствующим образом, чтобы компенсировать инверсию, вводимую линзой L_2 . В этом случае фототок также определяется соотношением (7.16).

Несмотря на то что указанная выше операция оказывается весьма полезной во многих случаях, включая

распознавание образов, все же часто бывает желательно осуществить родственную ей, но более общую операцию свертки. Одномерную свертку функций τ_1 и τ_2 можно получить с помощью любой из вышеприведенных систем, если перемещать один из диапозитивов с постоянной скоростью и измерять сигнал фотодетектора как функцию времени. Более точно, обращаясь к фиг. 73, б, будем считать, что диапозитив τ_2 установлен в перевернутом положении, так что операция (7.16) имеет вид

$$I = k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tau_1(x, y) \tau_2(-x, -y) dx dy.$$

Если перемещать диапозитив τ_2 в сторону отрицательных значений x со скоростью v , то сигнал детектора будет равен

$$I(t) = k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tau_1(x, y) \tau_2(vt - x, -y) dx dy.$$

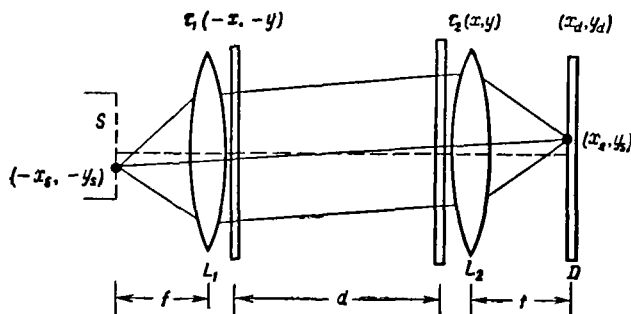
Если сканирование по x последовательно повторять с различными смещениями $-y_m$ вдоль y , то для сигнала детектора получим

$$I_m(t) = k \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tau_1(x, y) \tau_2(vt - x, y_m - y) dx dy, \quad m = 1, 2, 3, \dots \quad (7.17)$$

Множество функций $I_m(t)$ представляет собой полную двумерную свертку, дискретную по y .

Свертка без перемещений

Описанный выше метод осуществления свертки чрезвычайно неудобен и трудоемок из-за необходимости механического сканирования. Если видоизменить оптическую схему, то эту же операцию можно выполнить и без относительного перемещения диапозитивов [20]. Обращаясь к фиг. 74, предположим, что протяженный источник S находится в передней фокальной плоскости линзы L_1 . Сразу же за L_1 поместим диапозитив



Фиг. 74. Система, позволяющая осуществлять свертку без перемещения.

с коэффициентом пропускания $\tau_1(-x, -y)$. На расстоянии d от диапозитива τ_1 и непосредственно перед линзой L_2 установим диапозитив $\tau_2(+x, +y)$. Распределение интенсивности в задней фокальной плоскости линзы L_2 будем измерять с помощью фотопленки, хотя можно воспользоваться и электронным приемником (например, видиконом).

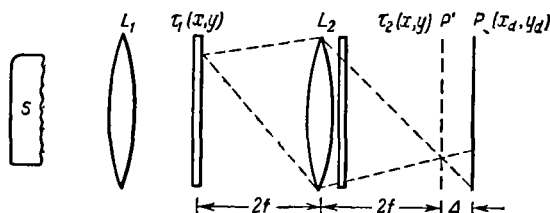
Чтобы понять работу этой системы, рассмотрим вначале распространение света от какой-то точки источника, имеющей координаты $(-x_s, -y_s)$. Лучи, выходящие из этой точки, после линзы L_1 (и τ_1) идут параллельным пучком и дают на τ_2 распределение интенсивности, пропорциональное $\tau_1[-x + (d/f)x_s, -y + (d/f)y_s]$. После прохождения через τ_2 лучи фокусируются на приемнике в точке (x_d, y_d) ; при этом мы полагаем, что обе линзы имеют одинаковые фокусные расстояния. Таким образом, распределение интенсивности на приемнике может быть записано в виде

$$I(x_d = x_s, y_d = y_s) = k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tau_1\left(\frac{d}{f}x_s - x, \frac{d}{f}y_s - y\right) \tau_2(x, y) dx dy, \quad (7.18)$$

что и представляет собой требуемую свертку.

Синтез импульсных откликов в расфокусированных системах

Непосредственный синтез нужного импульсного отклика можно осуществить с помощью «расфокусированной» системы, показанной на фиг. 75¹⁾ [21]. Линза L_1 опять отображает протяженный источник S на диапозитив с коэффициентом пропускания τ_1 , стоящий на «входе» системы. Линза L_2 дает изображение диапозитива τ_1 в плоскости P' . Для простоты положим, что τ_1 и P' находятся на расстояниях $2f$ от линзы L_2 , поэтому увеличение равно единице. Диапозитив с пропусканием τ_2 , соответствующим требуемому импульсному отклику, помещается непосредственно за L_2 ; сигнал на выходе системы регистрируется в плоскости P , находящейся на расстоянии Δ от плоскости изображения P' .



Фиг. 75. Синтез импульсного отклика с помощью расфокусированной системы.

Работу этой системы легче всего понять, если взять на τ_1 точечный источник единичной интенсивности с координатами (x, y) и найти распределение интенсивности в плоскости P . В приближении геометрической оптики лучи, проходящие через τ_2 , собираются в точку в плоскости P' , затем расходятся и дают уменьшенную проекцию τ_2 на плоскость P . Центр проекции имеет координаты $\{x_d = -[1 + (\Delta/2f)]x, y_d = -[1 + (\Delta/2f)]y\}$, а масштаб уменьшения диапозитива τ_2 равен $\Delta/2f$. Если учесть инверсию при проектировании τ_2 , то сигнал от точечного

¹⁾ Фактически эта система почти идентична системе на фиг. 74, но здесь используется иной подход, что оправдывает отдельное обсуждение схемы.

источника можно записать в виде

$$|h(x_d, y_d; x, y)|^2 = \\ = k\tau_2 \left\{ -\frac{2f}{\Delta} \left[x_d + \left(1 + \frac{\Delta}{2f} \right) x \right], -\frac{2f}{\Delta} \left[y_d + \left(1 + \frac{\Delta}{2f} \right) y \right] \right\}.$$

Следовательно, интенсивность на выходе в точке $(-x_d, -y_d)$ можно записать в виде интеграла свертки

$$I(-x_d, -y_d) = k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tau_1(x, y) \times \\ \times \tau_2 \left\{ \frac{2f}{\Delta} \left[x_d - \left(1 + \frac{\Delta}{2f} \right) x \right], \frac{2f}{\Delta} \left[y_d - \left(1 + \frac{\Delta}{2f} \right) y \right] \right\} dx dy. \quad (7.19)$$

Ограничения

Некогерентные системы обработки информации, основанные на законах геометрической оптики, обладают следующим общим недостатком: из-за некогерентной природы освещения данные на входе и импульсный отклик на выходе *не могут иметь отрицательных* распределений (интенсивности). В настоящее время известны методы получения эквивалентов отрицательных распределений (см., например, [21]), но они, как правило, громоздки и неудобны. Пока что не существует простого и прямого метода, позволяющего превращать биполярные входные сигналы в биполярные импульсные отклики, используя некогерентное освещение.

Кроме того, все системы, сконструированные на основе геометрической оптики, должны удовлетворять общему условию: геометрические размеры системы должны быть выбраны так, чтобы дифракционные эффекты были совершенно незначительны. Это требование хорошо удовлетворяется в системе на фиг. 73, а и не представляет большой трудности в системе на фиг. 73, б. Однако системы, изображенные на фиг. 74 и 75, как будет видно из последующего рассмотрения, в этом смысле весьма критичны. Из задачи 2.12 следует, что удобным критерием сложности данной функции служит ее пространственно-полосовое произведение. Эта величина представляет с достаточной хорошей точностью число независимых значений, которые задают данную функцию. Для полу-

чения максимальной эффективности системы обработки данных необходимо сконструировать ее так, чтобы на вход можно было подавать функции с наибольшим возможным пространственно-полосовым произведением.

Рассмотрим систему, изображенную на фиг. 74. Чтобы добиться максимального пространственно-полосового произведения системы, мы должны попытаться разместить на диапозитиве максимально возможное количество независимых данных. Однако чем мельче становится пространственная структура диапозитивов, тем больше света дифрагирует на первом диапозитиве, и тем меньше лучей подчиняется законам геометрической оптики. Следовательно, способ рассмотрения работы системы становится все менее точным и данные на выходе все более отличаются от ожидаемых результатов.

Аналогичные рассуждения применимы к системе фиг. 75. По мере того как структура второго диапозитива становится все мельче и мельче, изображение в плоскости P (которое, как мы говорили выше, образуется светом из точечного источника, находящегося на τ_1) будет все меньше и меньше напоминать геометрическую проекцию диапозитива τ_2 . Следовательно, метод анализа, с помощью которого были получены характеристики системы, будет тем менее точным, чем больше величина пространственно-полосового произведения.

Несмотря на то что мы рассмотрели только отдельные примеры, основной вывод совершенно очевиден: если в апертуру определенного размера необходимо поместить большое число данных, то в конечном счете надо принять во внимание дифракционные эффекты. Поэтому при использовании системы, созданной на основе геометрической оптики, весьма существенно иметь уверенность в том, что она работает при таких условиях, при которых обеспечивается точное соблюдение законов геометрической оптики.

§ 4. СИНТЕЗ В ЧАСТОТНОМ ПРОСТРАНСТВЕ

Операцию линейной фильтрации можно также осуществить в *частотном пространстве*, а не в пространстве координат. Если используется когерентное освещение, то

желаемую операцию фильтрации можно осуществить непосредственно, подбирая амплитудный коэффициент пропускания диапозитива в задней фокальной плоскости отображающей линзы. Мы уже познакомились с примерами такого рода систем при рассмотрении фильтрации фотографических изображений (Маршалль) и в фазово-контрастном микроскопе (Цернике).

Если для освещения используется некогерентный свет, то синтез в пространстве частот тоже возможен: так как между амплитудным пропусканием зрачка и оптической передаточной функцией системы существует простое автокорреляционное соотношение. Однако некогерентные системы имеют два очевидных недостатка. Во-первых, класс передаточных функций, который можно непосредственно синтезировать, ограничен классом автокорреляционных функций. Другими словами, могут быть осуществлены только действительные неотрицательные импульсные отклики. Во многих технических приложениях необходимы передаточные функции более общего вида. Во-вторых, даже если необходимая передаточная функция и соответствует неотрицательному действительному импульсному отклику, то требуемая функция зрачка системы может оказаться не единственной. В настоящее время неизвестно какой-либо последовательности операций, позволяющей найти *простейшую* функцию зрачка, которая давала бы желаемую передаточную функцию.

По этой причине мы ограничимся рассмотрением вопросов обработки в пространстве частот только для случая когерентного света. Однако это не означает, что обработка в пространстве частот для случая некогерентного света вообще не представляет интереса. Примеры некогерентных¹⁾ систем, в которых используется синтез в пространстве частот, можно найти в работе [22].

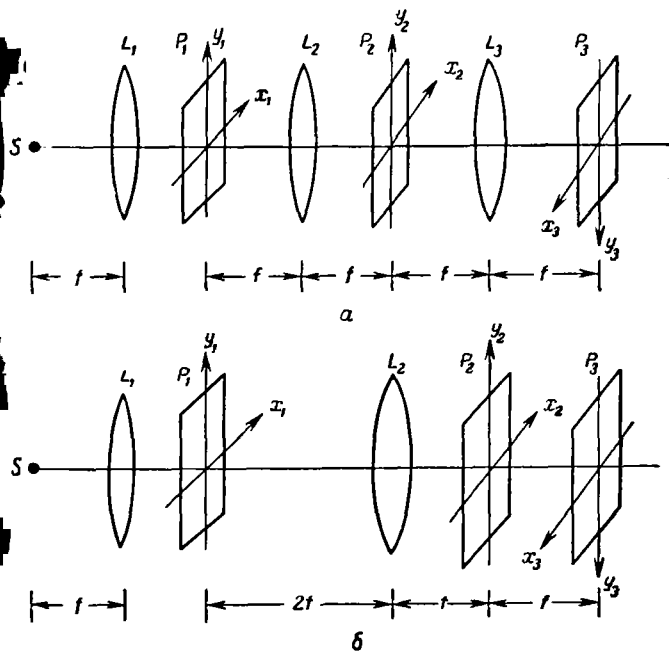
¹⁾ На первой ступени в этих системах также используется когерентный свет, но только для того, чтобы преобразовать сигналы на входе в особую форму, которая позволяет выполнять желаемую операцию в некогерентном свете,

Некоторые виды когерентных систем

Так как когерентные системы линейны по отношению к комплексной амплитуде, они могут осуществлять операцию вида

$$I(x, y) = K \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta \right|^2. \quad (7.20)$$

На фиг. 76 показаны две возможные схемы, позволяющие осуществлять операцию такого типа. В системе фиг. 76, а свет точечного источника S коллимируется линзой L_1 . Предназначенный для фильтрации сигнал подается на вход системы в плоскость P_1 , куда поме-



Фиг. 76. Две возможные конфигурации систем когерентной обработки.

щается диапозитив с изменяющимся амплитудным коэффициентом пропускания $g(x_1, y_1)$. Линза L_2 выполняет фурье-преобразование¹⁾ функции g , давая в плоскости P_2 распределение амплитуды $k_1 G(x_2/\lambda f, y_2/\lambda f)$, где G — фурье-образ функции g , а k_1 — комплексная константа. Чтобы воздействовать на амплитуду и фазу спектра G , в эту плоскость помещают фильтр. Если H — фурье-образ функции h , то амплитудный коэффициент пропускания частотного фильтра должен иметь вид

$$t(x_2, y_2) = k_2 H\left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f}\right). \quad (7.21)$$

Следовательно, распределение амплитуды за фильтром пропорционально GH . Наконец, линза L_3 преобразует это распределение амплитуды и дает в плоскости P_3 распределение интенсивности

$$I(x_3, y_3) = K \left| \iint_{-\infty}^{\infty} g(\xi, \eta) h(-x_3 - \xi, -y_3 - \eta) d\xi d\eta \right|^2.$$

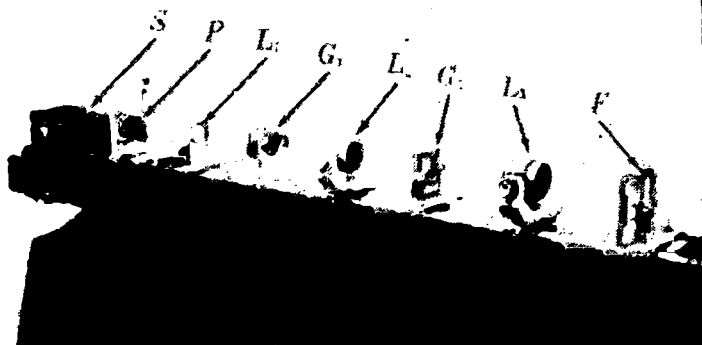
Знак минус перед x_3 и y_3 появляется из-за инверсии, которая получается в результате двух последовательных преобразований Фурье, а не из-за того, что после фурье-преобразования осуществляется *инверсное* фурье-преобразование (ср. задачу 2.3а). Это неудобство можно устранить, если изменить направления осей координатной системы (x_3, y_3) , как показано на фиг. 76, а. Такая операция (в дальнейшем она будет осуществляться повсюду) позволяет записать распределение интенсивности на выходе в виде

$$I(x_3, y_3) = K \left| \iint_{-\infty}^{\infty} g(\xi, \eta) h(x_3 - \xi, y_3 - \eta) d\xi d\eta \right|^2. \quad (7.22)$$

На фиг. 77 приведена фотография системы оптической обработки информации вышеописанного типа.

Другое устройство для выполнения той же операции показано на фиг. 76, б. Как и раньше, линза L_1 коллимирует свет от точечного источника S . В данном

¹⁾ Конечно, необходимо чтобы диаметр линз был достаточно большим, чтобы уменьшить эффект виньетирования (см. задачу 5.4).



Фиг. 77. Фотография когерентной системы оптической обработки информации.

S — лазерный источник; P — точечная диафрагма; L_1 — коллимирующая линза; G_1 — жидкостный компенсатор (входная плоскость); L_2 — преобразующая линза; G_2 — жидкостный компенсатор (частотная плоскость); L_2 — преобразующая линза; F — держатель пленки (выходная плоскость).

случае линза L_2 выполняет две функции, т. е. осуществляет фурье-преобразование и формирует изображение. Входным сигналом служит диапозитив с коэффициентом пропускания g , помещенный в плоскость P_1 , которая в данном случае находится на расстоянии $2f$ перед линзой (L_2)¹⁾. В задней фокальной плоскости L_2 получается спектр G , но на этот раз с квадратичным фазовым множителем $\exp[-j(k/2f)(x_2^2 + y_2^2)]$, так как входной сигнал теперь уже не находится в передней фокальной плоскости [см. (5.19)]. Частотный фильтр с амплитудным коэффициентом пропускания, пропорциональным H , как и раньше, помещается в плоскость P_2 . Плоскость P_3 находится на расстоянии $2f$ за линзой;

¹⁾ Это расстояние может и не быть равно $2f$. Необходимо только, чтобы формула линзы связывала плоскости P_1 и P_3 .

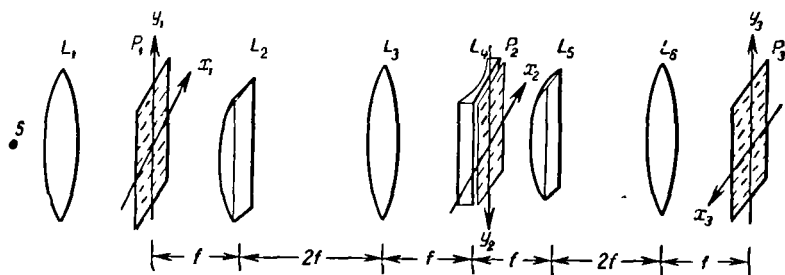
следовательно, формула линзы связывает плоскости P_1 и P_3 . Так как спектр объекта умножается на \mathbf{H} , распределение интенсивности в плоскости P_3 опять будет иметь вид

$$I(x_3, y_3) = K \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) h(x_3 - \xi, y_3 - \eta) d\xi d\eta \right|^2.$$

У второй системы есть два существенных для практики недостатка. Прежде всего, расстояние в системе b от входа до преобразующей линзы L_2 вдвое больше, чем в системе a . Это означает, что для устранения влияния виньетирования при данном фокусном расстоянии и данном размере входного диапозитива линза L_2 в системе b должна иметь большие размеры, чем в системе a . Вторым недостатком связан с соображениями, с помощью которых было получено приближенное соотношение (5.29) при анализе формирования когерентного изображения линзой. В этом выводе мы опирались на то, что в амплитуду в любой данной точке изображения дает вклад только небольшая область точек вблизи геометрической точки объекта¹⁾. Если операция фильтрации, которой соответствует функция \mathbf{H} , характеризуется большой величиной пространственно-полосового произведения, то импульсный отклик h будет обладать конечной площадью; тогда сигнал на выходе системы b надо рассматривать как результат фильтрации функции $g(x_1, y_1) \exp[j(k/4f)(x_1^2 + y_1^2)]$, а не просто функции $g(x_1, y_1)$. С этой проблемой мы не сталкиваемся в системе a , в которой проектируется не сфера на сферу, а изображение плоскости P_1 на плоскость P_3 .

Наконец, следует заметить, что часто когерентные системы конструируются так, чтобы они обрабатывали не одну функцию двух независимых переменных, а множество функций одной независимой переменной. Примером может служить так называемый *астигматический* оператор, показанный на фиг. 78. Данные на вход системы поступают в плоскость P_1 , расположенную за

¹⁾ То есть вблизи точки объекта, изображающейся в данную точку в приближении геометрической оптики. — Прим. ред.



Фиг. 78. Пример астигматического оператора.

коллимирующей линзой L_1 . Данные на входе задаются в виде вертикального набора одномерных функций пропускания, каждая из которых изменяется в горизонтальном направлении. Далее на расстоянии фокуса f от плоскости P_1 расположена цилиндрическая линза L_2 , действующая в вертикальной плоскости. На расстоянии $2f$ за линзой L_2 помещается сферическая линза L_3 , которая также имеет фокусное расстояние f . «Частотной плоскостью» теперь служит P_2 . Комбинация линз L_2 и L_3 осуществляет двойное преобразование Фурье в направлении y и, следовательно, резко изображает вертикальное измерение. Так как линза L_2 не действует в направлении x , сферическая линза L_3 осуществляет фурье-преобразование горизонтального измерения с фазовым множителем $\exp[-j(k/f)x_2^2]$ в плоскости P_2 . Этот фазовый множитель можно устранить, если непосредственно перед P_2 поместить отрицательную цилиндрическую линзу с фокусным расстоянием $f/2$ и таким образом получить еще одно преобразование фазы

$$t_1(x_2) = \exp\left(j\frac{k}{f}x_2^2\right), \quad (7.23)$$

которое устраняет изменение фазы. Если совокупность входных сигналов есть множество коэффициентов пропускания $g_k(x_1)$, $k = 1, 2, \dots, K$, то в плоскости P_3 мы получим соответствующее множество фурье-образов $G_k(x_2)$, $k = 1, 2, \dots, K$ с обратным порядком нумера-

ции по вертикали, получающимся в результате операции отображения.

В плоскости P_2 можно поместить набор линейных одномерных фильтров. Линзы L_4 , L_5 , как и раньше, формируют изображение в y -направлении и осуществляют фурье-преобразование в x -направлении, сохраняя, таким образом, структуру набора, но возвращая множество функций в «пространство координат». Фазовый множитель, связанный с последним преобразованием Фурье, в общем случае не существует.

Ограничения

Несмотря на то что когерентные системы обладают в общем случае большей гибкостью и имеют большую емкость, чем некогерентные, тем не менее и в этом случае на возможные типы операций налагаются некоторые ограничения. Рассмотрим ограничения, присущие системам, в которых требуемая передаточная функция синтезируется путем независимого изменения амплитудного и фазового коэффициентов пропускания в плоскости частот. Как мы увидим в § 5, при более сложных методах получения частотных масок, основанных на интерферометрической записи, некоторые из этих ограничений снимаются.

Обычно для осуществления данной передаточной функции в частотной плоскости помещают независимые амплитудную и фазовую маски. Амплитуду пропускания, как правило, меняют с помощью фотопластинки, помещенной в жидкостный компенсатор. Фазовое пропускание можно регулировать, устанавливая прозрачную пластинку, толщина которой меняется нужным образом. С помощью такого метода, вообще говоря, довольно трудно получить больше двух значений возможной фазы (например, относительная задержка может составлять 0 или 180°). Фазовые пластинки можно наносить на основу почти так же, как наносят дифракционные решетки; кроме того, их можно получать методами испарения или другими методами. Все эти методы неудобны и могут быть успешно осуществлены только в том случае, когда требуется относительно простая дву-

значная регулировка фазы. Таким образом, из-за определенных практических ограничений можно осуществлять передаточные функции только довольно простого типа.

Следует заметить, что даже в том случае, когда необходим импульсный отклик простого вида (например, образ «Р»), соответствующую передаточную функцию, во-первых, трудно заранее вычислить аналитически, и, во-вторых, она слишком сложна для того, чтобы ее можно было синтезировать этим методом.

Наконец, если функции на входе, имеющие большое пространственно-полосовое произведение, должны быть отфильтрованы так, чтобы импульсный отклик на выходе имел тоже большое пространственно-полосовое произведение, то в ряде случаев точность установки фильтрующей маски в частотной плоскости должна быть очень высокой (см. задачу 7.6). Однако, как правило, эту трудность можно преодолеть, если воспользоваться микроманипулятором.

Следовательно, подводя итог, можно сказать, что возможности обычной когерентной системы серьезнее всего ограничиваются трудностью одновременного контроля в частотной плоскости распределений амплитуды и фазы; исключения составляют лишь самые простые случаи. Только в 1963 г., когда появились интерферометрически записанные фильтры, это серьезное ограничение удалось в основном преодолеть. Последние достижения в области оптической обработки информации будут обсуждаться в следующем параграфе.

§ 5. ФИЛЬТРЫ ВАНДЕР ЛЮГТА

В 1963 г. Вандер Люгт, сотрудник Радиолокационной лаборатории Мичиганского университета, предложил и продемонстрировал новый метод синтеза масок в плоскости частот когерентных систем [23, 24]. Маски, полученные этим методом, обладают замечательным свойством: они позволяют успешно регулировать как амплитуду, так и фазу передаточной функции, несмотря на то что состоят исключительно из *поглощающих* эле-

ментов¹⁾. Посредством этого метода можно устранить два наиболее серьезных ограничения обычных когерентных систем обработки.

Синтез маски в частотной плоскости

Частотная маска для фильтра Вандер Люгта синтезируется с помощью интерферометрической системы типа показанной на фиг. 79. Линза L_1 коллимирует свет от точечного источника S . Часть этого света падает на маску P_1 , амплитудный коэффициент пропускания которой соответствует распределению амплитуды требуемого импульсного отклика h . Линза L_2 осуществляет фурие преобразование распределения амплитуды h , давая на пленке распределение амплитуды $(1/\lambda f) H(x_2/\lambda f, y_2/\lambda f)$. Другая часть коллимированного света проходит над маской P_1 , отклоняется призмой P и попадает на пленку под углом θ , как показано на фиг. 79.

Суммарная интенсивность в каждой точке пленки определяется интерференцией двух пучков света, каждый из которых характеризуется своим распределением амплитуд. Наклонная плоская волна, выходящая из призмы, дает на пленке распределение поля

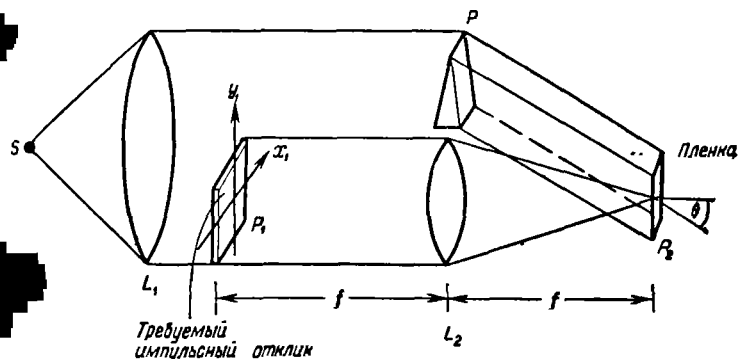
$$U_r(x_2, y_2) = r_0 e^{-i2\pi\alpha y_2}, \quad (7.24)$$

где пространственная частота определяется выражением

$$\alpha = \frac{\sin \theta}{\lambda}. \quad (7.25)$$

Суммарное распределение интенсивности можно пред

¹⁾ Исторически применению фильтра Вандер Люгта предшествовал очень похожий на него метод, известный в настоящее время под названием резко отсекающего (hard clipped) фильтра. Для таких фильтров частотная маска, как правило, рассчитывается численно, а не записывается оптически. Несмотря на то что такой фильтр использовался при обработке радиолокационных данных еще в 1961 г., по вполне понятным причинам его описание не появлялось в открытой печати вплоть до 1965 г. [25]. (Резко отсекающими такие фильтры называются потому, что они состоят из прозрачных и непрозрачных полос и производят в связи с этим только фазовую фильтрацию, — *Прим. ред.*)



Фиг. 79. Запись частотной маски для фильтра Вандер Люгта.

ставить следующим образом:

$$\begin{aligned} \mathcal{U}(x_2, y_2) &= \left| r_0 \exp(-j2\pi a y_2) + \frac{1}{\lambda f} \mathbf{H} \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right|^2 = \\ &= r_0^2 + \frac{1}{\lambda^2 f^2} \left| \mathbf{H} \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right|^2 + \frac{r_0}{\lambda f} \mathbf{H} \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \times \\ &\times \exp(j2\pi a y_2) + \frac{r_0}{\lambda f} \mathbf{H}^* \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \exp(-j2\pi a y_2). \end{aligned} \quad (7.26)$$

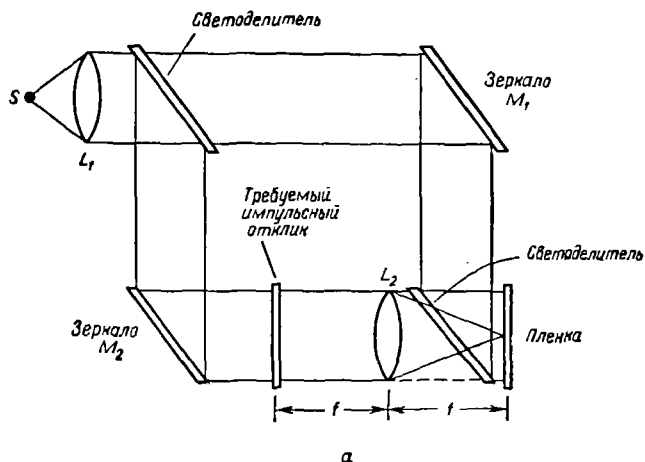
Заметим, что если комплексная функция \mathbf{H} характеризуется распределением амплитуды A и распределением фазы ψ , т. е. если

$$\mathbf{H} \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) = A \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \exp \left[-j\psi \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right],$$

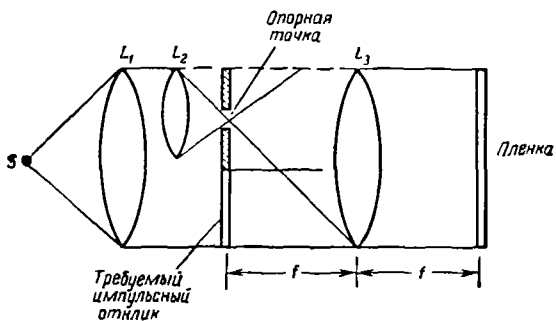
то выражение для \mathcal{U} можно записать в виде

$$\begin{aligned} \mathcal{U}(x_2, y_2) &= r_0^2 + \frac{1}{\lambda^2 f^2} A^2 \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) + \\ &+ \frac{2r_0}{\lambda f} A \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \cos \left[2\pi a y_2 - \psi \left(\frac{x_2}{\lambda f}, \frac{y_2}{\lambda f} \right) \right]. \end{aligned} \quad (7.27)$$

При такой записи сразу видно, каким образом с помощью интерферометрического процесса комплексная функция \mathbf{H} записывается на детектор, чувствительный к интенсивности: амплитудная и фазовая информация записывается соответственно как амплитудная и фазовая модуляция *высокочастотной несущей*, которой служит наклонная «опорная» волна, выходящая из призмы,



а



б

Фиг. 80. Две другие схемы записи частотных масок.
 а — видоизмененный интерферометр Маха — Цендера; б — видоизмененный интерферометр Релея.

Существуют, конечно, и другие оптические системы, которые могут дать распределение интенсивности, описываемое соотношением (7.27). На фиг. 80 показаны еще два возможных варианта. Система а есть не что иное, как видоизмененный интерферометр Маха — Цендера. Если наклонять зеркало M_1 , то в плоскость пленки будет приходиться наклонная плоская волна. В нижнем плече

интерферометра линза L_2 , как и раньше, осуществляет фурье-преобразование требуемого импульсного отклика. Наконец, светоделитель позволяет получать в плоскости пленки сумму этих двух волн.

Система б, представляющая собой видоизмененный интерферометр Релея, позволяет осуществить третий способ получения того же самого распределения (см. работу [2], стр. 128). За коллимирующей линзой L_1 стоит меньшая по размерам линза L_2 , фокусирующая часть коллимированного света в яркое пятнышко в передней фокальной плоскости линзы L_3 . Сферическая волна, исходящая из этой «опорной точки», коллимируется линзой L_3 и дает в плоскости пленки наклонную плоскую волну. Маска, соответствующая импульсному отклику, осуществляет, как обычно, фурье-преобразование амплитуды прошедшей через нее волны. Следовательно, в плоскости пленки опять получается распределение интенсивности, аналогичное распределению (7.27).

На последней ступени синтеза частотной маски экспонированная пленка проявляется таким образом, чтобы получился диапозитив с амплитудным коэффициентом пропускания, пропорциональным распределению интенсивности света при экспозиции. Следовательно,

$$t(x_2, y_2) \sim r_0^2 + \frac{1}{\lambda^2 f^2} |H|^2 + \frac{r_0}{\lambda f} H \exp(j2\pi a y_2) + \frac{r_0}{\lambda f} H^* \exp(-j2\pi a y_2). \quad (7.28)$$

Отметим, что если не учитывать комплексный экспоненциальный множитель, то третий член в выражении для коэффициента пропускания пропорционален H и в точности удовлетворяет требованиям, предъявляемым к коэффициенту пропускания, необходимому для синтеза фильтра с импульсным откликом h . Теперь остается только показать, как можно использовать этот член и исключить другие члены.

Обработка данных, поступающих на вход

После того как частотная маска синтезирована, ее можно установить в обычную систему обработки информации, изображенную выше на фиг. 76, а. Если на вход

поступает предназначенная для фильтрации функция $g(x_1, y_1)$, то в частотной плоскости на маску падает свет с распределением амплитуды $(1/\lambda f) G(x_2/\lambda f, y_2/\lambda f)$. Для напряженности поля, прошедшего через маску, имеет следующее соотношение:

$$U_2 \sim \frac{r_0^2 G}{\lambda f} + \frac{1}{\lambda^2 f^2} |H|^2 G + \frac{r_0}{\lambda^2 f^2} H G \exp(j2\pi\alpha y_2) + \\ + \frac{r_0}{\lambda^2 f^2} H^* G \exp(-j2\pi\alpha y_2)$$

Последняя линза L_3 на фиг. 76, а осуществляет фурье-преобразование поля U_2 (с множителем $1/\lambda f$). Учитывая, что система координат в P_3 инверсно преобразована, находим, что в этой плоскости напряженность поля пропорциональна следующей величине:

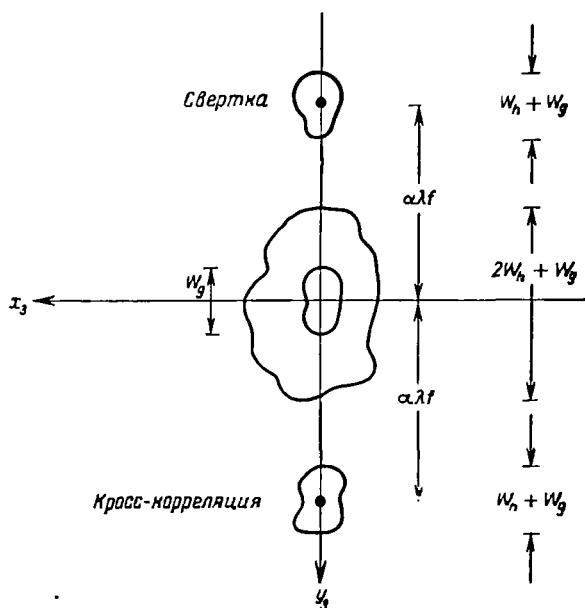
$$U_3(x_3, y_3) \sim r_0^2 g(x_3, y_3) + \\ + \frac{1}{\lambda^2 f^2} [h(x_3, y_3) * h^*(-x_3, -y_3) * g(x_3, y_3)] + \\ + \frac{r_0}{\lambda f} [h(x_3, y_3) * g(x_3, y_3) * \delta(x_3, y_3 + \alpha\lambda f)] + \\ + \frac{r_0}{\lambda f} [h^*(-x_3, -y_3) * g(x_3, y_3) * \delta(x_3, y_3 - \alpha\lambda f)]. \quad (7.29)$$

Третий член в этом выражении представляет для нас особый интерес. Поскольку

$$h(x_3, y_3) * g(x_3, y_3) * \delta(x_3, y_3 + \alpha\lambda f) = \\ = \int \int_{-\infty}^{\infty} h(x_3 - \xi, y_3 + \alpha\lambda f - \eta) g(\xi, \eta) d\xi d\eta, \quad (7.30)$$

находим, что эта составляющая сигнала на выходе есть свертка h и g , смещенная в плоскости $x_3 y_3$ по оси y_3 на $-\alpha\lambda f$. Для дальнейшего четвертый член соотношения (7.29) можно переписать в виде

$$h^*(-x_3, -y_3) * g(x_3, y_3) * \delta(x_3, y_3 - \alpha\lambda f) = \\ = \int \int_{-\infty}^{\infty} g(\xi, \eta) h^*(\xi - x_3, \eta - y_3 + \alpha\lambda f) d\xi d\eta, \quad (7.31)$$



Фиг. 81. Положение различных компонент на выходе оператора

что представляет собой *кросс-корреляцию* функций g и h , смещенную по оси y_3 на $\alpha\lambda f$.

Заметим, что первый и второй члены соотношения (7.29), не представляющие интереса в обычных операциях фильтрации, не смещены относительно начала координат в плоскости x_3y_3 . Таким образом, если «несущая частота» α выбрана достаточно большой или, другими словами, если опорная волна падает под достаточно большим углом, то компоненты, соответствующие свертке и кросс-корреляции, будут, очевидно, отклоняться от оси (в противоположных направлениях) так, что каждую из них можно будет наблюдать отдельно. Чтобы найти свертку h и g , наблюдатель просто должен исследовать распределение света вблизи точки $(0, -\alpha\lambda f)$. Чтобы найти кросс-корреляцию h и g , необходимо вести наблюдение вблизи точки $(0, \alpha\lambda f)$.

Чтобы более точно определить требования, налагаемые на α , рассмотрим ширину различных компонент на выходе системы, показанных на фиг. 81. Если W_h — максимальная ширина функции h в направлении y , а W_g — соответствующая величина для g , то ширины различных составляющих на выходе будут

1. $r_0^2 g(x_3, y_3) \rightarrow W_g$,
2. $\frac{1}{\lambda^2 f^2} [h(x_3, y_3) * h^*(-x_3, -y_3) * g(x_3, y_3)] \rightarrow 2W_h + W_g$
3. $\frac{r_0}{\lambda f} [h(x_3, y_3) * g(x_3, y_3) * \delta(x_3, y_3 + \alpha \lambda f)] \rightarrow W_h + W_g$,
4. $\frac{r_0}{\lambda f} [h^*(-x_3, -y_3) * g(x_3, y_3) * \delta(x_3, y_3 - \alpha \lambda f)] \rightarrow W_h + W_g$.

Из рисунка ясно, что полное разделение будет достигнуто, если

$$\alpha > \frac{1}{\lambda f} \left(\frac{3W_h}{2} + W_g \right),$$

или, эквивалентно, если

$$\theta > \frac{3W_h}{2f} + \frac{W_g}{f}, \quad (7.32)$$

где ввиду малости углов мы воспользовались приближением $\sin \theta \approx \theta$.

Преимущества фильтра Вандер Люгта

Использование фильтра Вандер Люгта устраняет два наиболее серьезных недостатка обычных схем когерентной обработки. Во-первых, если нужно получить импульсный отклик заданного вида, то отпадает необходимость в математически сложной и трудоемкой операции по нахождению требуемой передаточной функции; система, в которой синтезируется маска для частотной плоскости, осуществляет *оптическим* путем фурье-преобразование импульсного отклика. Во-вторых, комплексная передаточная функция общего вида синтезируется с помощью одной *поглощающей* маски; в этом случае уже

нет необходимости каким-то сложным образом влиять на изменение фазы при прохождении волны через частотную плоскость. Чтобы устранить все фазовые сдвиги, поглощающая маска просто помещается в жидкостный компенсатор.

Фильтр Вандер Люгта также очень чувствителен к точности установки маски в частотной плоскости, но не более чувствителен, чем обычные системы когерентной обработки. Для записи модулированной высокочастотной несущей необходима эмульсия с более высоким разрешением, чем для обычного синтеза маски, но такие материалы существуют (например, спектроскопические пластинки Кодак), так что выполнение этого требования не вызывает особых трудностей.

Отметим, что метод Вандер Люгта открывает новые возможности для когерентной обработки. Если раньше основной практической проблемой был синтез маски в плоскости частот, теперь трудности опять сосредоточены в *пространстве координат*. Но в пространстве координат трудности гораздо менее серьезны, так как требуемые импульсные отклики часто имеют простую форму, так что необходимые маски можно изготовить обычными фотографическими методами. Таким образом, фильтр Вандер Люгта распространил сферу использования когерентных операторов на новую, ранее недоступную область, в которую попадают многие наиболее перспективные приложения метода когерентной обработки.

§ 6. ПРИЛОЖЕНИЕ К РАСПОЗНАВАНИЮ ОБРАЗОВ

Практическое применение оптической обработки, которое находилось в центре внимания в течение ряда лет, относится к области *распознавания образов*. Как мы увидим в дальнейшем, эта операция служит превосходным примером того, как осуществляется обработка информации в случае простого импульсного отклика (тогда как передаточные функции могут и не быть простыми). Поэтому метод синтеза Вандер Люгта особенно удобен при решении таких задач.

Согласованный фильтр

Понятие *согласованного фильтра* играет важную роль в проблеме распознавания образов. По определению, мы называем линейный пространственно-инвариантный фильтр *согласованным* с данным сигналом $s(x, y)$, если его импульсный отклик $h(x, y)$ имеет вид

$$h(x, y) = s^*(-x, -y). \quad (7.33)$$

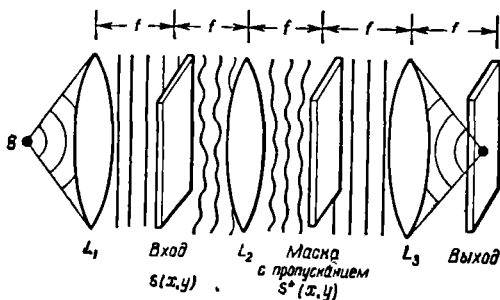
Если входной сигнал $g(x, y)$ приложен к фильтру, согласованному с сигналом $s(x, y)$, то на выходе получится сигнал

$$\begin{aligned} v(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x - \xi, y - \eta) g(\xi, \eta) d\xi d\eta = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta) s^*(\xi - x, \eta - y) d\xi d\eta, \end{aligned} \quad (7.34)$$

который, как можно заметить, представляет собой кросс-корреляцию функций g и s .

Исторически понятие согласованного фильтра впервые возникло в связи с вопросами обнаружения сигналов; если необходимо выделить сигнал известной формы из «белого» шума, то согласованный фильтр позволяет осуществить линейную операцию, в результате которой получается максимальное значение отношения мгновенной мощности сигнала (в определенный момент) к средней мощности шума [26]. Однако ниже мы будем полагать, что образы не имеют шумов, и приведем другие причины необходимости использования операции фильтрации.

Сущность операции согласованной фильтрации можно понять, обращаясь к оптической интерпретации, которая иллюстрируется схемой на фиг. 82. Положим, что фильтр, согласованный с входным сигналом $s(x, y)$, осуществляется в виде частотной маски в обычной схеме когерентной обработки. Путем фурье-преобразования импульсного отклика (7.33) находим, что требуемая пере-



Фиг. 82. Оптическая схема операции согласованной фильтрации.

даточная функция имеет вид

$$H(f_x, f_y) = S^*(f_x, f_y), \quad (7.35)$$

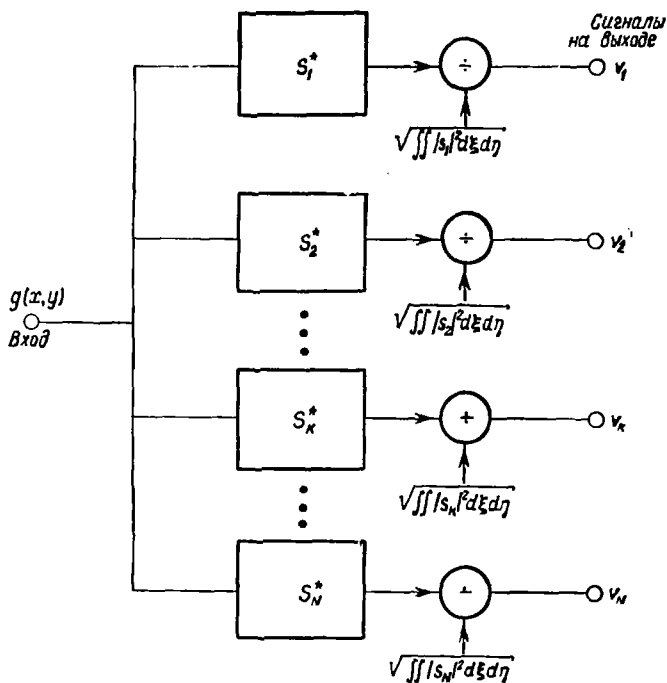
где $H = \mathcal{F}\{h\}$, а $S = \mathcal{F}\{s\}$. Следовательно, частотная маска должна иметь амплитудный коэффициент пропускания, пропорциональный S^* .

Рассмотрим теперь вид распределения поля, прошедшего через маску, когда на вход подается сигнал s (с которым фильтр согласован). Распределение поля, падающего на маску, пропорционально S , а распределение поля, прошедшего через маску, пропорционально SS^* . Эта последняя величина *действительна*; иными словами, маска точно компенсирует кривизну падающего на нее волнового фронта S . Таким образом, поле, прошедшее через маску, представляет собой *плоскую волну*, которая фокусируется в яркую точку в фокусе последней преобразующей линзы. Если на вход поступает сигнал, отличный от $s(x, y)$, то кривизна волнового фронта в общем случае *не компенсируется* маской и прошедший через систему свет уже *не будет собираться* в яркую точку в фокусе последней линзы. Таким образом, предположение о присутствии на входе сигнала s можно проверить, измеряя интенсивность света в фокусе последней преобразующей линзы. (Если сигнал s на входе смещен относительно начала координат, то яркая точка в выходной плоскости просто смещается на соответствующее расстояние; ср. задачу 7.7.)

Проблема распознавания образов

Рассмотрим следующую специфическую проблему распознавания образов: сигнал g на входе системы обработки информации может состоять из N каких-либо отдельных образов s_1, s_2, \dots, s_N ; с помощью оптического оператора необходимо определить заданный образ. Как мы покажем, процесс идентификации можно осуществить, если приложить входной сигнал к совокупности N фильтров, каждый из которых согласован с одним из возможных образов на входе.

Блок-схема распознающего устройства показана на фиг. 83. Входной сигнал одновременно (или последовательно) подается на N согласованных фильтров с пере-



Фиг. 83. Блок-схема системы распознавания образов.

даточными функциями S_1^* , S_2^* , ..., S_N^* . Отклик каждого фильтра нормируется путем деления на корень квадратный из полной энергии, приходящей от образа, с которым фильтр согласован. Эта нормировка, которая может быть выполнена с помощью электронной схемы после получения сигналов на выходе фильтров, учитывает то, что различным образам на входе в общем случае могут соответствовать разные энергии. На выходе системы сравниваются квадраты модулей сигналов $|v_1|^2$, $|v_2|^2$, ..., $|v_N|^2$. Как мы сейчас покажем, если данный образ

$$g(x, y) = s_k(x, y)$$

действительно присутствует в сигнале на входе, то на выходе сигнал $|v_k|^2$ будет наибольшим из N откликов.

Чтобы доказать это утверждение, заметим, во-первых, что, согласно соотношению (7.34), максимум сигнала $|v_k|^2$ на выходе правильно согласованного фильтра равен

$$|v_k|^2 = \frac{\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s_k|^2 d\xi d\eta \right]^2}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s_k|^2 d\xi d\eta} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s_k|^2 d\xi d\eta. \quad (7.36)$$

С другой стороны, отклик $|v_n|^2$ ($n \neq k$) неправильно согласованного фильтра равен

$$|v_n|^2 = \frac{\left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_k s_n^* d\xi d\eta \right|^2}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s_n|^2 d\xi d\eta}. \quad (7.37)$$

Воспользовавшись неравенством Шварца, можно написать

$$\left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_k s_n^* d\xi d\eta \right|^2 \leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s_k|^2 d\xi d\eta \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s_n|^2 d\xi d\eta.$$

Отсюда непосредственно следует

$$|v_n|^2 \leq \int \int_{-\infty}^{\infty} |s_k|^2 d\xi d\eta = |v_k|^2; \quad (7.38)$$

равенство имеет место только при условии, что

$$s_n(x, y) = \kappa s_k(x, y).$$

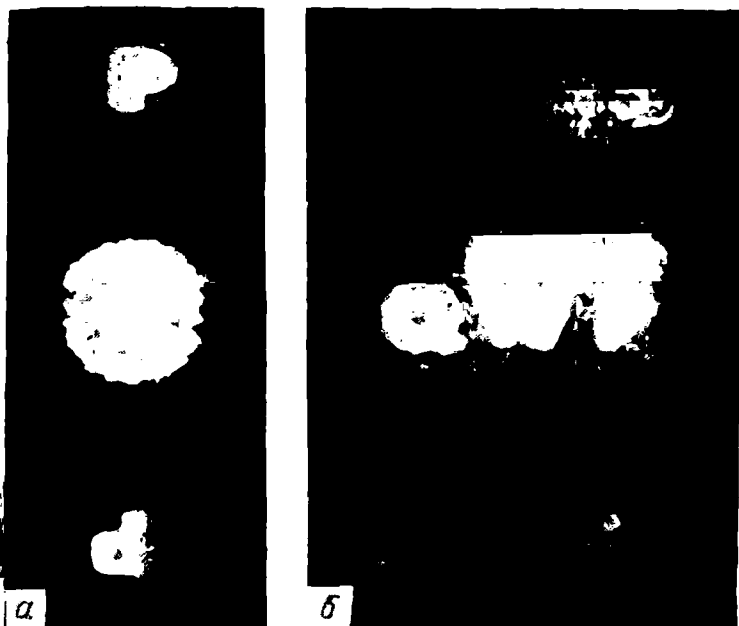
Из полученного результата с очевидностью вытекает, что согласованный фильтр дает возможность установить, какой образ из возможного набора образов действительно присутствует в системе. Следует подчеркнуть, что такая способность не присуща исключительно согласованным фильтрам. Действительно, в некоторых случаях можно улучшить дискриминацию образов путем модификации (т. е. рассогласования) всех фильтров (см., например, работу [2], стр. 130—133).

Оптический синтез устройства для распознавания образов

Операцию согласованной фильтрации легко осуществить с помощью рассмотренного выше метода Вандер Люгта. Это выполняется особенно просто, поскольку, как мы уже говорили, один из сигналов на выходе фильтра Вандер Люгта есть кросс-корреляция сигнала на входе и исходного образа, с помощью которого был синтезирован фильтр. Если сосредоточить внимание на соответствующей области выходного пространства, то легко обнаружить сигналы на выходе согласованного фильтра.

На фиг. 84, а приведена фотография импульсного отклика фильтра Вандер Люгта, который был синтезирован для образа «Р». Верхняя часть отклика дает свертку сигнала на входе с символом Р, а нижняя часть — кросс-корреляцию сигнала на входе с буквой Р. Центральная часть отклика является помехой и не представляет никакого интереса.

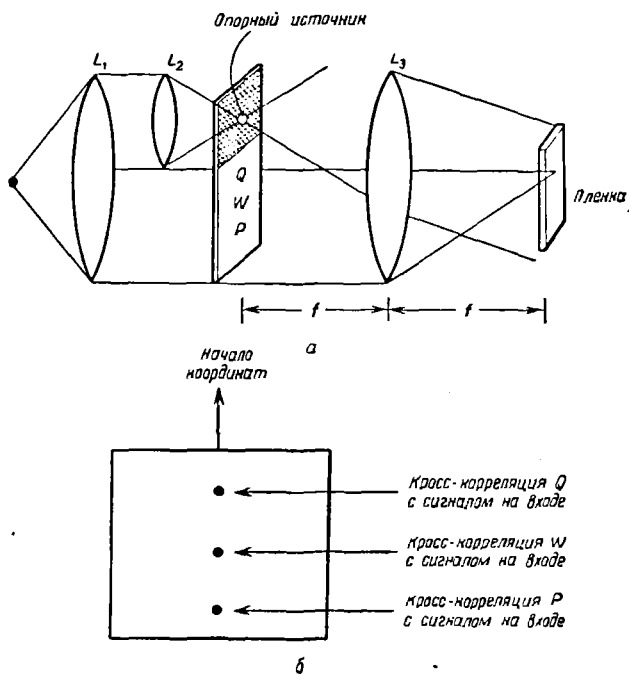
На фиг. 84, б показан отклик на выходе согласованного фильтра на буквы Q, W и P. Отметим, что наличие



Фиг. 84. Фотографии импульсного отклика фильтра Вандер Люгта (а) и отклика на выходе фильтра, согласованного для буквы «Р», на входной сигнал — буквы Q, W и P (б).

яркой точки в отклике на букву Р говорит о высокой корреляции между сигналом на входе и буквой, с которой согласован фильтр.

Чтобы получить блок согласованных фильтров, показанный на фиг. 83, можно синтезировать N отдельных фильтров Вандер Люгта и сигнал на входе подавать последовательно на каждый фильтр. Однако если число N не слишком велико, то можно синтезировать весь блок фильтров на одной частотной маске. Это можно сделать, или модулируя одну несущую частоту различными сигналами, или записывая на один диапозитив различные частотные маски с разными несущими частотами. На фиг. 85, а показан один из способов записи фильтра с



Фиг. 85. Синтез совокупности согласованных фильтров на одной частотной маске.

а — синтез частотной маски; б — поле зрения согласованного фильтра.

одной несущей, модулированной несколькими сигналами. Пучки света, идущие от букв Q, W и P, составляют различные углы с опорной волной; в результате кросс-корреляции между Q, W и P и образом на входе оказываются на разных расстояниях от начала координат (фиг. 85, б).

Число различных фильтров, которые можно осуществить этим методом, ограничено величиной динамического диапазона, который может быть получен на маске. Вандер Люгту удалось осуществить синтез девяти отдельных импульсных откликов от одной маски (см. работу [2], стр. 133—139).

Чувствительность к изменению масштаба и вращению

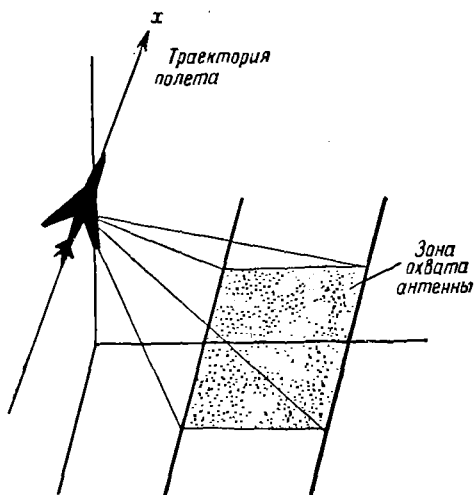
Метод когерентно-оптического распознавания образов обладает недостатками, которые присущи всем способам решения задачи распознавания образов с помощью согласованных фильтров. Говоря более конкретно, фильтры обычно довольно чувствительны к изменению масштаба и поворотам образов на входе системы. Когда образ на входе системы имеет неверную угловую ориентацию или неправильное увеличение, величина отклика правильно согласованного фильтра уменьшается и в процессе идентификации образа возможны ошибки. Чувствительность к вращению и изменению масштаба сильно зависит от структуры данного образа (например, образ «L» более чувствителен к вращению, чем «O»), однако эту чувствительность можно до некоторой степени уменьшать при достаточно умелом осуществлении фильтрации. Часто разумным решением такой задачи является механический поиск нужного положения среди различных масштабов и ориентаций образа. Однако это довольно трудоемкая операция. Более подробно затронутый вопрос разобран в работе [2].

§ 7. ОБРАБОТКА ДАННЫХ АНТЕНН СИНТЕТИЧЕСКОЙ АПЕРТУРЫ

Одним из самых замечательных приложений методики оптической обработки информации является построение карт земной поверхности с высоким разрешением по данным, полученным от так называемых радиолокаторов синтетической апертуры. Все нижеприведенное рассмотрение основано на статье Катроны и др. [11], к которой читатель может обратиться за дополнительными сведениями.

Создание синтетической апертуры

Обращаясь к фиг. 86, рассмотрим радиолокационную систему, находящуюся на самолете, летящем с постоянной скоростью v_a по прямолинейной траектории в направлении оси x . С помощью радиолокатора необходимо



Фиг. 86. Геометрическая схема радиолокации бокового обзора.

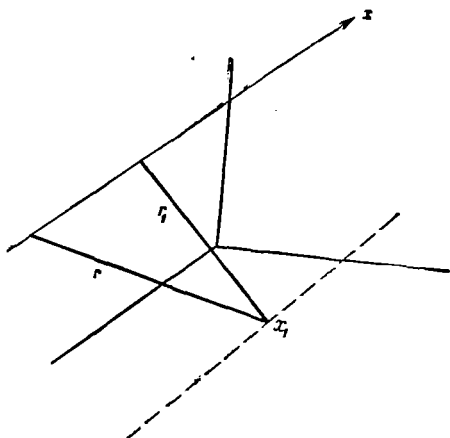
получить подробную карту отражающей способности поверхности вдоль траектории полета. Разрешение в направлении, перпендикулярном траектории полета, получают, посылая импульсные сигналы и записывая отраженные сигналы в функции времени. Разрешение по азимуту можно в принципе получить, если ширина луча радиолокатора в этом направлении будет очень малой. Однако азимутальное разрешение, достижимое на расстоянии R с помощью антенны, имеющей линейный размер D , грубо говоря, равно $\lambda_r R/D$. Так как в микроволновой области длина волны λ_r на несколько порядков больше длины оптических волн, для получения разрешения, сравнимого с разрешением систем аэрофотосъемки, необходимы антенны практически неосуществимых размеров.

Эта проблема решается с помощью антенн синтетической апертуры. Предположим, что на самолете находится маленькая антенна с *широким лучом*, направление бокового обзора которой постоянно относительно самолета. Радиолокатор посылает импульсы из точек, расположенных последовательно вдоль траектории полета; во всех

этих точках производится запись (амплитуды и фазы) отраженного сигнала в зависимости от времени. Каждую такую запись можно рассматривать как сигнал, полученный от одного элемента ряда антенн. Чтобы синтезировать эффективную апертуру, длина которой может достигать сотен и тысяч метров, требуется только соответствующим образом скомбинировать записи от различных элементов. Отметим, что для получения синтетического ряда наибольшей длины антенна радиолокатора облучает данную точку поверхности в течение максимальной возможной части траектории полета. Таким образом, чем больше ширина луча антенны радиолокатора, тем выше разрешение, которое может быть достигнуто при обработке полученных данных. Однако для реализации полного разрешения необходимо, чтобы на самолете имелся опорный сигнал со стабильной фазой, что позволяло бы записывать как амплитуду, так и фазу отраженного сигнала на протяжении полета.

Прием и развертка сигналов

Для более подробного изучения процесса получения данных рассмотрим вначале схему, изображенную на фиг. 87. Расстояние вдоль траектории полета обозначим через x . Для простоты положим, что в точке с координатой x_1 имеется точечный рассеиватель, находящийся на расстоянии r_1 от траектории по перпендикуляру к ней. Для еще большего упрощения примем, что радиолокатор излучает волну в виде непрерывной синусоиды частоты f_r . Импульсный характер действительно излучаемых радиолокатором сигналов проявляется просто в периодической выборке из сигналов, которые получались бы в случае непрерывного синусоидального сигнала. Согласно теории выборки (см. гл. 2, § 3), при рассмотрении азимутальных сигналов дискретностью выборки можно пренебречь, если расстояние, которое пролетает самолет за промежуток времени между импульсами, меньше $(2B_x)^{-1}$, где B_x — ширина спектра сигнала, отраженного от поверхности. Рассмотрение импульсного характера сигнала существенно для разрешения в другом измерении — расстоянии от траектории полета (дальности).



Фиг. 87. Схема траектории полета.

Сигнал, возвращающийся на самолет от исследуемой рассеивающей точки, может быть записан в виде фазора, зависящего от времени

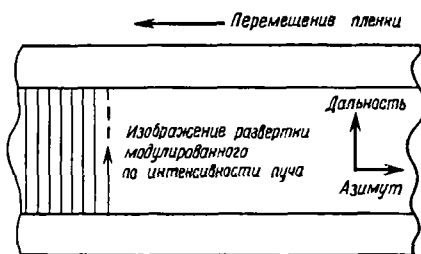
$$s_1(t) = \sigma_1 \exp \left[j2\pi f_r \left(t - \frac{2r}{c} \right) \right], \quad (7.39)$$

где f_r — частота радиолокатора; r — расстояние от самолета до цели; c — скорость света, а σ_1 — комплексный амплитудный множитель, зависящий от таких параметров, как мощность излучения, отражательная способность цели, сдвиг фазы и ослабление сигнала с расстоянием, которое пропорционально четвертой степени расстояния от рассеивающей точки до самолета. Расстояние r можно выразить через r_1 , x_1 и x (координату вдоль траектории полета):

$$r = \sqrt{r_1^2 + (x - x_1)^2} \approx r_1 + \frac{(x - x_1)^2}{2r_1}. \quad (7.40)$$

Следовательно, принимаемый сигнал будет

$$s_1(t) = \sigma_1(x_1, r_1) \exp \left\{ j \left[2\pi f_r t - \frac{4\pi r_1}{\lambda_r} - \frac{2\pi (x - x_1)^2}{\lambda_r r_1} \right] \right\}. \quad (7.41)$$



Фиг. 88. Лента с записью.

При движении самолета пространственная переменная x связана с переменной времени t соотношением

$$x = v_a t.$$

Если поверхность земли, находящуюся на расстоянии r_1 , рассматривать как совокупность большого числа точечных рассеивателей, то полный принимаемый сигнал можно записать следующим образом:

$$\begin{aligned} s(t) &= \sum_n s_n(t) = \\ &= \sum_n \sigma_n(x_n, r_1) \exp \left\{ j \left[2\pi f_r t - \frac{4\pi r_1}{\lambda_r} - \frac{2\pi}{\lambda_r r_1} (v_a t - x_n)^2 \right] \right\}. \end{aligned} \quad (7.42)$$

Принимаемый сигнал синхронно детектируется, что эквивалентно просто уменьшению средней частоты принятого сигнала, которая теперь равна не f_r , а меньшему значению f'_r . В результате получим

$$s'(t) = \sum_n |\sigma_n(x_n, r_1)| \cos \left[2\pi f'_r t - \frac{4\pi r_1}{\lambda_r} - \frac{2\pi}{\lambda_r r_1} (v_a t - x_n)^2 + \varphi_n \right]. \quad (7.43)$$

При этом мы отказались от фазорной записи.

Детектированный сигнал используется для модуляции интенсивности электронного луча катодной трубки. После каждого переданного импульса электронный луч развертывается по вертикали в течение интересующего

нас интервала времени. На фиг. 88 показана запись, которая получается, если перед экраном осциллографа движется пленка с горизонтальной скоростью v_f . Вдоль вертикальных линий меняется дальность, а азимут данной рассеивающей точки по отношению к радиолокатору меняется в горизонтальном направлении.

Фокусирующие свойства пленки

Как и раньше, будем рассматривать только расстояние r_1 и, следовательно, ограничимся данными, записанными на пленке вдоль линии $y = y_1$, и опять будем пренебрегать импульсной природой посылаемого сигнала. Если соответствующим образом подобрать экспозицию, то сведения об азимуте принятого сигнала будут записаны в виде амплитудного коэффициента пропускания

$$I(x, y_1) = t_b + \chi \sum_n |\sigma_n(x_n, r_1)| \times \\ \times \cos \left[2\pi f_x x - \frac{4\pi r_1}{\lambda_r} - \frac{2\pi}{\lambda_r r_1} \left(\frac{v_a}{v_f} x - x_n \right)^2 + \varphi_n \right], \quad (7.44)$$

где t_b — постоянное пропускание (смещение), вводимое для того, чтобы можно было записать биполярные видеосигналы, а χ — просто константа. При написании (7.44) мы воспользовались соотношением¹⁾

$$x = v_f t,$$

из которого следует, что

$$f_x = \frac{f_r'}{v_f}.$$

Заменяя косинус в соотношении (7.44) двумя комплексными экспонентами, можно записать пропускание в виде суммы постоянного пропускания t_b и еще двух

¹⁾ Заметим, что переменная x используется теперь для обозначения координат пленки, в то время как раньше ею обозначались координаты радиолокатора.

слагаемых:

$$t_a(x, y) = \frac{\chi}{2} \sum_n \sigma'_n(x_n, r_1) \exp \left\{ j \left[2\pi f_x x - \frac{2\pi v_a}{\lambda_r r_1 v_f} \left(x - \frac{v_f}{v_a} x_n \right)^2 \right] \right\} \quad (7.45a)$$

и

$$t_b(x, y) = \frac{\chi}{2} \sum_n \sigma''_n(x_n, r_1) \exp \left\{ -j \left[2\pi f_x x - \frac{2\pi}{\lambda_r r_1} \frac{v_a}{v_f} \left(x - \frac{v_f}{v_a} x_n \right)^2 \right] \right\}, \quad (7.45b)$$

где постоянная фаза $4\pi r_1/\lambda_r$, так же как и фазы φ_n , включена в σ'_n .

Если ограничиться рассмотрением только одной рассеивающей точки, скажем, с номером $n = N$, то соответствующая ей компонента t_a будет

$$t_a^{(N)}(x, y) = \frac{\chi}{2} \sigma'_N(x_N, y_1) \exp(j2\pi f_x x) \times \exp \left[-j \frac{2\pi}{\lambda_r r_1} \frac{v_a}{v_f} \left(x - \frac{v_f}{v_a} x_N \right)^2 \right]. \quad (7.46)$$

В этой компоненте прошедшего через пленку света первый экспоненциальный множитель, имеющий линейную фазовую зависимость, вводит просто наклон волнового фронта. Угол наклона θ фронта волны к плоскости диапозитива может быть найден из соотношения

$$\sin \theta = \lambda_0 f_x, \quad (7.47)$$

где λ_0 — длина волны света.

Обращаясь ко второму экспоненциальному множителю, замечаем его сходство с функцией пропускания положительной цилиндрической линзы с центром в точке $x = x_0$

$$t_l(x) = \exp \left[-j \frac{\pi}{\lambda_0 f_1} (x - x_0)^2 \right], \quad (7.48)$$

где f_1 — фокусное расстояние. Приравнявая (7.46) к (7.48), находим, что этот множитель в t_a соответствует свойствам положительной цилиндрической линзы

с фокусным расстоянием

$$f_1 = \frac{1}{2} \frac{\lambda_r}{\lambda_0} \frac{v_f}{v_a} r_1, \quad (7.49)$$

ось которой имеет координату

$$x = \frac{v_f}{v_a} x_N. \quad (7.50)$$

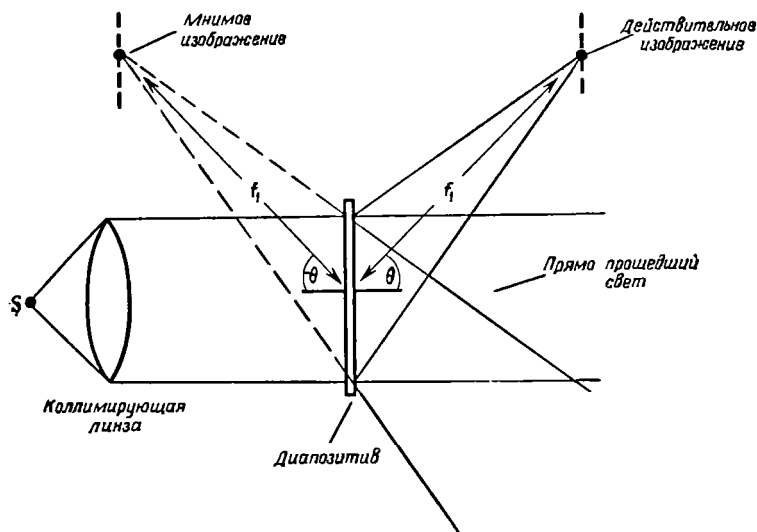
Аналогично N -я компонента t_β

$$t_\beta^{(N)}(x, y_1) = \frac{\chi}{2} \sigma_N^*(x_N, y_1) \exp(-j2\pi f_x x) \times \\ \times \exp \left[j \frac{2\pi}{\lambda_r r_1} \frac{v_a}{v_f} \left(x - \frac{v_f}{v_a} x_N \right)^2 \right] \quad (7.51)$$

также состоит из двух экспоненциальных сомножителей. Первая экспонента определяет наклон волнового фронта в противоположном направлении, т. е. под углом $-\theta$, а вторая соответствует пропусканию отрицательной линзы, имеющей центр в точке $x = (v_f/v_a)x_N$ и фокусное расстояние f_1 , определяемое соотношением (7.49).

На фиг. 89 показаны три компонента, на которые разлагается свет, прошедший через пленку, в случае когда рассеивателем является одна точка. Наличие постоянной компоненты пропускания t_b приводит к тому, что проходящая через диапозитив световая волна просто равномерно ослабляется, не испытывая никаких других изменений. Относительно компонент $t_\alpha^{(N)}$ и $t_\beta^{(N)}$ пропускания можно сказать, что они ответственны за возникновение пары «изображений», позволяющей определить азимут рассеивающей точки. Действительно, за счет компоненты $t_\alpha^{(N)}$ свет фокусируется в яркую линию за диапозитивом, а за счет компонент $t_\beta^{(N)}$ мы получаем волну, которая кажется исходящей из линейного протяженного источника, расположенного перед диапозитивом (см. фиг. 89).

Если на расстоянии r_1 находится множество рассеивающих точек, то при оптической обработке записи каждая из них будет соответствовать своя собственная пара действительных и мнимых линий фокусировки. Относительное азимутальное расположение точек рассеяния определяет относительное положение центров структур на



Ф и г. 89. Свет, прошедший через область пленки, соответствующую линии $y = y_1$.

пленке, действующих подобно линзам, а следовательно, и относительное положение соответствующих линий фокусировки. Таким образом, в определенных плоскостях перед и за диапозитивом воссоздается полная картина азимутального распределения точечных рассеивателей, находящихся на расстоянии r_1 . Мы еще раз подчеркиваем, что эта картина растянута в направлении y , так как в этом направлении пленка не обладает фокусирующими свойствами.

Получение конечного изображения

В конечном счете нам хотелось бы получить изображение не только азимутального распределения точечных рассеивателей, но и их распределения по дальности. Дальность (расстояние по перпендикулярно) рассеивающей точки передается на пленку y -координатой

записи азимутов рассеивающих точек. Следовательно, необходимо отобразить изменение пропускания пленки в y -направлении непосредственно в плоскость фокусировки азимутальных сигналов. Эта задача усложняется тем, что места фокусировки для различных значений азимута зависят от расстояния r_1 [см. (7.49)], а следовательно, от рассматриваемой координаты y . Чтобы получить с помощью радиолокатора карту местности, очевидно, необходимо отобразить изменения пропускания в направлении y на *наклонную* плоскость, в которой имеет место азимутальная фокусировка.

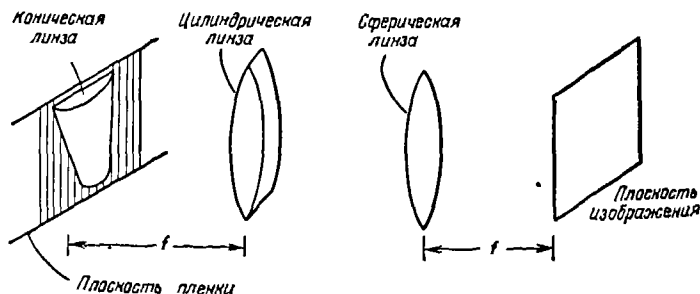
Эту задачу можно решить с помощью оптической системы, показанной на фиг. 90. Положительная *коническая* линза (см. задачу 5.3) помещается непосредственно за диапозитивом. Функция пропускания этой линзы имеет вид

$$t_l(x, y) = \exp\left(-j \frac{\pi}{\lambda_0 f_1} x^2\right), \quad (7.52)$$

а ее фокусное расстояние линейно зависит от координаты дальности (или, эквивалентно, от координаты y)

$$f_1 = \frac{1}{2} \frac{\lambda_r}{\lambda_0} \frac{v_f}{v_a} r_1. \quad (7.53)$$

Сравнение соотношений (7.51) — (7.53) показывает, что эта линза переносит изображение всей наклонной плоскости, содержащей *мнимые* линейные источники, в бесконечность, но в то же время она никак не влияет



Фиг. 90. Оптическая система для воспроизведения изображения.

на изменения пропускания вдоль оси y . Информация об азимутах остается записанной в виде направлений на расположенные в бесконечности мнимые линейные источники. Если пленку поместить в фокусе цилиндрической линзы, то получим в бесконечности мнимое изображение структуры в y -направлении. Изображения в x - и y -направлениях теперь совпадают, и их следует просто перенести обратно из бесконечности, чтобы получить действительное изображение. Эту последнюю операцию можно осуществить, если после цилиндрической линзы на фокусном расстоянии от плоскости наблюдения поместить сферическую линзу. Такая линза дает действительное изображение обеих координат (дальности и азимута) исходного рельефа, причем обе они резко фокусируются в плоскости выхода.

Заключительные замечания

Плодотворность метода оптической обработки информации для построения карт поверхности по данным, полученным с помощью антенн синтетической апертуры, объясняется чрезвычайной простотой, с которой оптическая система осуществляет довольно сложное и запутанное линейное преобразование, необходимое для получения изображения. Если подходить к вопросу с точки зрения основных принципов, то можно сказать, что такая простота определяется в основном одинаковым характером законов (с некоторым масштабным множителем), описывающих распространение волн, отраженных объектом при радиолокации, с одной стороны, и оптических волн в системе обработки — с другой. Что касается информации об азимуте, то систему обработки можно рассматривать как устройство, в котором воспроизводится точный оптический аналог волн, падающих на набор синтетических антенн. Если оптический аналог получен, то формирование изображения может быть выполнено стандартными оптическими методами.

Следует подчеркнуть, что создание оптического аналога радиолокационного метода основано на способности фотографического диапозитива сохранять как амплитудную, так и фазовую информацию. При этом очень

важную роль играет отличная от нуля промежуточная частота f'_r , остающаяся после синхронного детектирования, и соответственно пространственная частота f_x на пленке. Как следует из соотношения (7.44), изменения во времени амплитуды и фазы азимутального сигнала записываются как амплитудная и фазовая модуляции пространственной несущей частоты f_x . Именно из-за присутствия этой несущей действительная и мнимая линии фокусировки на фиг. 89 наклонены к оси диапозитива под углами соответственно θ и $-\theta$. Если бы не было этого смещения, то, например, мнимую линию фокусировки невозможно было бы наблюдать без искажений, вызванных ее интерференцией с прямо прошедшим светом и светом, фокусирующимся в действительную линию.

Таким образом, мы еще раз наглядно убедились в важности использования пространственной несущей частоты, что подчеркивалось и выше при обсуждении фильтра Вандер Люгта. Как мы увидим в последней главе, процесс получения изображения путем восстановления волнового фронта, или голография, приобрел в настоящее время такое важное значение в основном также благодаря использованию этой методики.

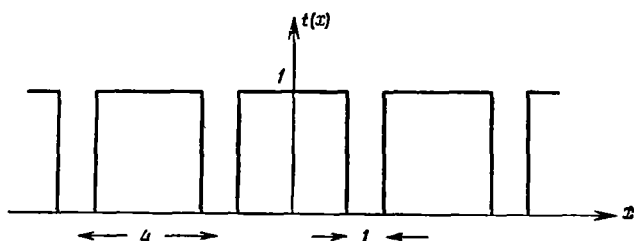
ЗАДАЧИ

7.1. Объект имеет периодический амплитудный коэффициент пропускания, который можно записать в виде

$$t(x, y) = t(x) \cdot 1,$$

где функция $t(x)$ показана на фиг. 91. Этот объект установлен в предметной плоскости системы, изображенной на фиг. 60, а на оси линзы в ее фокальной плоскости помещен маленький непрозрачный экран. Нарисовать приблизительно распределение *интенсивности* в плоскости изображения.

7.2. Фазовые объекты можно наблюдать с помощью так называемого *метода темного поля*, помещая маленький непрозрачный экран в заднюю фокальную плоскость отображающей линзы так, чтобы он закрывал недифрагированный свет. Полагая, что при прохождении через объект фаза изменяется меньше, чем на 1 рад , найти интенсивность изображения в зависимости от изменений фазы, вносимых объектом.



Ф и г. 91.

7.3. Наблюдение фазовых объектов методом *свилей* осуществляют, помещая в фокальной плоскости лезвие бритвы так, чтобы оно закрывало одну половину дифрагированного света. Следовательно, амплитудный коэффициент пропускания для фокальной плоскости можно записать следующим образом:

$$t_f(x, y) = \frac{1}{2} (1 + \operatorname{sgn} x).$$

а) Считая увеличение равным единице и пренебрегая инверсией изображения, показать, что амплитуда изображения связана с амплитудой объекта U_i соотношением

$$U_i(x_i, y_i) = \frac{1}{2} \left[U_0(x_i, y_i) + \frac{j}{\pi} \int_{-\infty}^{\infty} \frac{U_0(x', y_i)}{x_i - x'} dx' \right].$$

б) Предположим, что поле объекта имеет вид

$$U_0(x_0, y_0) = \exp [j\varphi(x_0, y_0)],$$

где $\varphi \ll 1$. Показать, что интенсивность изображения приближенно описывается соотношением

$$I_i(x_i, y_i) \approx \frac{1}{4} \left[1 - \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\varphi(x', y_i)}{x_i - x'} dx' \right].$$

в) Найти и схематически изобразить распределение интенсивности изображения, когда

$$\varphi(x_0, y_0) = \varphi_0 \operatorname{rect} \left(\frac{x_0}{X} \right)$$

при $\varphi_0 \ll 1$.

7.4. Найти выражение для интенсивности изображения в том случае, когда фазовая пластинка в фазово-контрастном микроскопе Цернике является также частично поглощающей с коэффициентом пропускания по интенсивности, равным α ($0 < \alpha < 1$).

7.5. Фотопленка экспонируется малоcontrastным распределением интенсивности:

$$\mathcal{I}(x, y) = \mathcal{I}_0 + \Delta\mathcal{I}(x, y) \quad |\Delta\mathcal{I}| \ll \mathcal{I}_0,$$

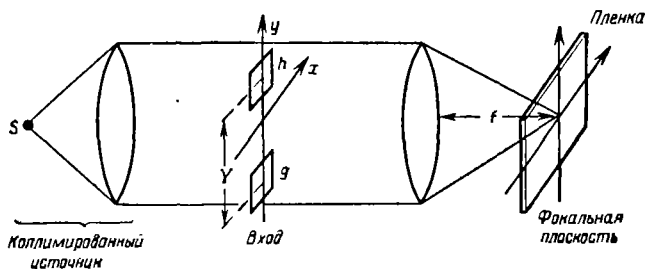
затем из нее изготавливается негатив. Полагая, что \mathcal{I}_0 смещает рабочую точку пленки в линейный участок кривой почернения, показать, что в том случае, когда контраст $\Delta\mathcal{I}/\mathcal{I}_0$ достаточно мал, распределение контраста в свете, прошедшем через диапозитив, линейно связано с распределением контраста при экспозиции.

7.6. Ширина входной апертуры некоторой когерентной системы обработки равна 3 см. Фокусное расстояние первой преобразующей линзы 10 см, длина волны света 0,6328 мк. С какой точностью необходимо установить частотную маску в фокальной плоскости, если считать, что масштаб структуры маски сравним с масштабом структуры спектра на входе?

7.7. Показать, что положение входа относительно согласованного фильтра влияет на сигнал на выходе только через изменения положения максимума ожидаемого отклика.

7.8. Объект в виде диапозитива с комплексным амплитудным коэффициентом пропускания $t(x, y)$ помещен непосредственно перед собирающей линзой. Объект освещается нормально падающей монохроматической плоской волной; распределение интенсивности в задней фокальной плоскости записывается на фотопластику. В результате получают позитивный транспарант с $\gamma = 2$. Затем диапозитив освещают по нормали плоской волной, причем непосредственно за диапозитивом помещают точно такую же собирающую линзу. Какое соотношение существует между амплитудным коэффициентом пропускания исходного объекта и распределением интенсивности в задней фокальной плоскости линзы на второй ступени процесса?

7.9. Два диапозитива с амплитудными коэффициентами пропускания $h(x, y)$ и $g(x, y)$ устанавливаются на вход системы перед собирающей линзой, как показано на фиг. 92, так, что их центры находятся в точках $(x = 0, y = Y/2)$ и $(x = 0, y = -Y/2)$. Записыва-



Фиг. 92.

вается распределение интенсивности в задней фокальной плоскости линзы и изготавливается позитивный транспарант с $\gamma = 2$.

Полученный диапозитив помещается перед той же самой линзой и опять преобразуется. Показать, что амплитуда света в задней фокальной плоскости линзы содержит кросс-корреляцию h и g . При каких условиях можно отделить на выходе кросс-корреляцию от других компонент?

7.10. Фазовый объект с амплитудным коэффициентом пропускания $t(x, y) = \exp[j\varphi(x, y)]$ находится в предметной плоскости когерентной оптической системы. В задней фокальной плоскости системы устанавливается ослабляющая пластинка (постоянной толщины) с пропусканием по интенсивности

$$\tau(x, y) = \alpha(x^4 + 2x^2y^2 + y^4).$$

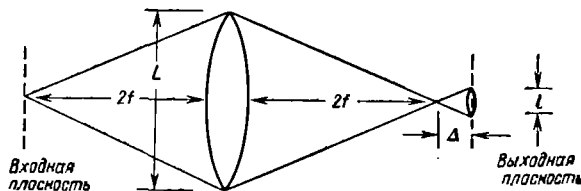
Как связана интенсивность изображения с фазой объекта?

7.11. На фотопластинку с модуляционной передаточной функцией $M(f)$ падает свет с распределением амплитуды

$$U(x, y) = \exp(j2\pi f_1 x) + \exp(j2\pi f_2 x).$$

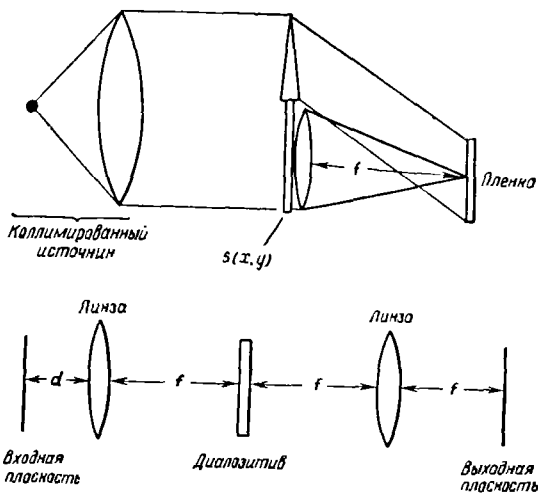
После проявления получен позитив с $\gamma = 4$. Полагая, что рабочая точка находится на линейном участке кривой почернения, построить частотный спектр распределения комплексной амплитуды света, прошедшего через диапозитив. Выразить частоты компонент света, прошедшего сквозь диапозитив, через f_1 и f_2 , а их амплитуды — через функцию $M(f)$ при соответствующих значениях аргумента.

7.12. Необходимо создать «расфокусированную» (некогерентную) систему пространственной фильтрации (ср. § 3) так, чтобы первый нуль ее передаточной функции совпал с частотой f_0 . Такая система показана на фиг. 93. Полагая, что данные, предназначенные для фильтрации, помещаются на расстоянии $2f$ перед круглой линзой диаметром L , найти зависимость «расстояния расфокусировки» Δ от f , L и f_0 . Чему равно расстояние Δ при $f_0 = 10$ период/см, $f = 10$ см и $L = 2$ см?



Ф и г. 93.

7.13. Для синтеза частотного фильтра применен метод Вандер Люфта. Как показано на фиг. 94 (вверху), «сигнал» в виде диапозитива с амплитудным коэффициентом пропускания $s(x, y)$ помещен



Ф и г. 94.

непосредственно перед собирающей линзой (а не в передней фокальной плоскости); интенсивность в задней фокальной плоскости записывается на фотопленку. Пленка обрабатывается так, что ее амплитудный коэффициент пропускания пропорционален экспозиции. Получившийся в результате диапозитив устанавливается в систему, показанную на фиг. 94 (внизу). Полагая, что в каждом случае исследуются соответствующие части выходной плоскости, найти, чему должно быть равно расстояние d между входной плоскостью и первой линзой, для того чтобы синтезировать

- а) фильтр с импульсным откликом $s(x, y)$;
- б) согласованный фильтр с импульсным откликом $s^*(-x, -y)$.

7.14. Радиолокатор посылает так называемые chirping-сигналы, которые можно аналитически записать в виде

$$s(t) = \text{rect}\left(\frac{t}{\tau}\right) \cos\left(2\pi f_r t + \frac{\Delta}{2r} t^2\right).$$

Эта волна отражается от неподвижной цели, находящейся на известном расстоянии от радиолокатора.

а) Каким образом можно использовать оптические методы для осуществления операции согласованной фильтрации сигнала, принятого в течение T (сек) после послышки импульса?

б) Каким образом, исследуя сигнал на выходе оптического оператора, можно найти расстояние до цели?

в) Каким образом можно отфильтровать одновременно с помощью одного оператора сигналы, полученные в течение ряда T -секундных разверток?

ЛИТЕРАТУРА

1. Optical Processing of Information, ed. D. K. Pollack, C. J. Koes-
ter, J. T. Tippett, Baltimore, 1963.
2. Optical and Electro-optical Information Processing, ed. J. T. Tip-
pett et al., Cambridge, Mass., 1965.
3. Porter A. B., Phil. Mag. (6), 11, 154 (1906).
О дифракционной теории изображения в микроскопе.
4. Zernike F., Zs. Techn. Phys., 16, 454 (1935).
Метод фазового контраста в микроскопии.
5. Armstrong E. H., Proc. IRE, 24, 689 (1936).
Метод уменьшения помех радиосвязи в системах с частотной
модуляцией.
6. Elias P., Grey D. S., Robinson D. Z., Journ. Opt. Soc.
Am., 42, 127 (1952).
Анализ Фурье в оптике.
7. Elias P., Journ. Opt. Soc. Am., 43, 229 (1953).
Оптика и теория связи.
8. O'Neill E. L., IRE Trans. Inform. Theory, IT-2, 56 (1956).
Пространственная фильтрация в оптике.
9. Communication and Information Theory Aspects of Modern
Optics, ed. E. L. O'Neill, Syracuse, 1962.
10. Cutrona L. J., et al., IRE Trans. Inform. Theory, IT-6, 386
(1960).
Оптическая обработка информации и системы фильтрации.
11. Cutrona L. J., et al., Proc. IEEE, 54, 1026 (1966).
О применении метода когерентной оптической обработки к ра-
диолокаторам синтетической апертуры.
12. Stroke G. W., Funkhouser A. T., Phys. Lett., 16, 272
(1965).
Фурье-спектроскопия с применением стационарных интерферо-
метров и голографического метода получения изображения без
использования вычислительных машин.
13. Jackson P. L., Appl. Opt., 4, 419 (1965).
Дифракционная обработка геофизических данных.
14. Mees C. E. K., The Theory of the Photographic Process, New
York, 1954.
15. Kodak Data Book P-9, Rochester, 1962.
Кодаковские пластинки и пленки для научно-технических целей.
16. Kozma A., Journ. Opt. Soc. Am., 56, 428 (1966).
Фотографическая запись пространственно-модулированного ко-
герентного света.
17. Leith E. N., Phot. Sci. Eng., 6, 75 (1962).
Фотопленка как элемент когерентной оптической системы.
18. Ingalls A. L., Phot. Sci. Eng., 4, 135 (1960).
Влияние колебаний толщины пленки на когерентный свет.
19. Goldberg E., U. S. Patent 1838389, Dec. 29, 1931.
Статистический прибор.

20. Kovasna y L. S. G., Arman A., Rev. Sci. Instr., 28, 793 (1957).
Оптическое измерение автокорреляции произвольных двумерных образов.
21. Trabka E. A., Roetting P. G., Journ. Opt. Soc. Am., 54, 1242 (1964).
Преобразование изображений для распознавания контуров в некогерентном свете с помощью биполярных масок.
22. Armitage J. D., Lohmann A. W., Appl. Opt., 4, 461 (1965).
Распознавание образов с помощью некогерентной пространственной фильтрации.
23. Vander Lugt A. V., Radar Lab., Rept. No. 4594-22-T, Institute of Science and Technology, The University of Michigan, Ann Arbor, 1963.
Обнаружение сигнала с помощью комплексной пространственной фильтрации.
24. Vander Lugt A. V., IEEE Trans. Inform. Theory, IT-10, 2 (1964).
Обнаружение сигнала с помощью комплексной пространственной фильтрации.
25. Kozma A., Kelly D. L., Appl. Opt., 4, 387 (1965).
Пространственная фильтрация для выделения сигнала из шума.
26. Tugin G. L., IRE Trans. Inform. Theory, IT-6, 311 (1960).
Введение в теорию согласованной фильтрации.
- 27*. Сороко Л. М., УФН, 90, вып. 1 (1966).
Голография и интерференционная обработка информации.
- 28*. Оптическая обработка информации, сб. статей, изд-во «Мир», 1966.
- 29*. Харкевич А., Борьба с помехами, М., 1965.

ГЛАВА 8

ПОЛУЧЕНИЕ ИЗОБРАЖЕНИЯ ВОССТАНОВЛЕНИЕМ ВОЛНОВОГО ФРОНТА, ИЛИ ГОЛОГРАФИЯ

В 1948 г. Дэнис Габор [1] предложил новый двухступенчатый процесс безлинзового получения изображений, который он назвал *восстановлением волнового фронта*. Габор установил, что в том случае, когда в пространстве одновременно присутствуют соответствующая когерентная опорная волна и волна, дифрагированная на объекте, то, несмотря на то что фотопластинка чувствительна только к интенсивности, на ней может быть записана информация как об амплитуде, так и о фазе дифрагированных волн. Он показал, что с помощью полученной таким образом интерферограммы (которую он назвал *голограммой*, что означает «полная запись») можно в конечном счете воссоздать изображение исходного объекта.

Если в период зарождения этого метода получения изображений, известного теперь под названием *голографии*, он не представлял существенного интереса, то последние достижения техники и усовершенствование экспериментальной методики позволили значительно расширить возможности его применений. В результате голография заняла важное место в современной теории построения изображений.

§ 1. ИСТОРИЧЕСКОЕ ВВЕДЕНИЕ

Изучение вопросов восстановления волнового фронта Габор начал под влиянием работ Брэгга по рентгеновской кристаллографии (см., например [2]), но основным его стимулом была надежда на то, что вновь разработанную методику удастся использовать в электронной микроскопии. Свою первоначальную идею Габор изложил в двух обширных статьях [3, 4], опубликованных в 1949 и 1951 гг. и касавшихся использования метода восстановления волнового фронта в микроскопии.

Правда, по техническим причинам этот метод не нашел еще сколько-нибудь важного применения, тем не менее развитие голографии в значительной мере определяется задачами микроскопии.

В 50-х годах многие авторы, в том числе Роджерс [5], Эль-Сум [6], Ломан [7], значительно развили теорию метода получения изображения восстановлением волнового фронта.

Однако только в начале 60-х годов в голографии началась настоящая революция. При этом опять большую роль сыграли работы, выполненные в Радиолокационной лаборатории Мичиганского университета, в особенности работы Лейта и Упатника [8]. Названные авторы обнаружили сходство между габоровским процессом восстановления волнового фронта и методом антенн синтетической апертуры и видоизменили первоначальную габоровскую схему, существенно усовершенствовав процесс. Вскоре исследователи Мичиганской группы с целью получения трехмерных фотографий предложили соединить новый метод с появившейся в то время лазерной техникой [9]. Высокое качество и натуральность трехмерных изображений, полученных методом голографии, вызвали всеобщий интерес к идее восстановления волнового фронта.

Вслед за разработкой новых, более совершенных схем голографии было выдвинуто и осуществлено множество идей практического использования этого метода. Однако, как мы увидим в § 9, вопреки широко распространенному мнению многие наиболее интересные и потенциально полезные свойства голограмм не связаны с возможностью получения трехмерных изображений.

§ 2. ПРОБЛЕМА ВОССТАНОВЛЕНИЯ ВОЛНОВОГО ФРОНТА

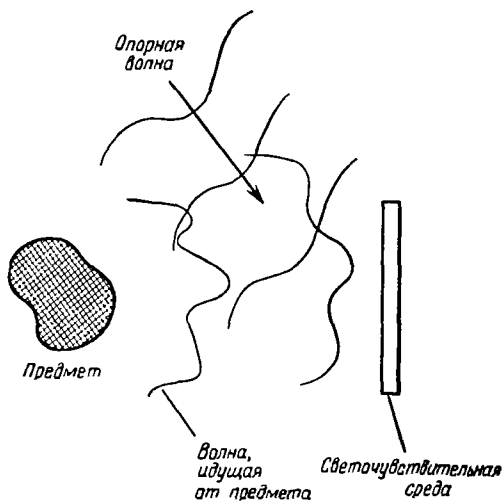
В этом параграфе наше внимание в основном будет сосредоточено на проблеме записи и последующего восстановления амплитуды и фазы оптической волны, проходящей от когерентно освещенного объекта. Эта проблема, конечно, является достаточно общей, так что ее интересно было бы рассмотреть не только для всего спектра электромагнитных волн, но и для акустических

волн, однако мы ограничимся оптической задачей. При изложении мы воспользуемся общим методом, аналогичным предложенному Колиером [10].

Запись амплитуды и фазы

Как уже говорилось, процесс восстановления волнового фронта должен состоять из двух стадий: запись или запоминание информации и операции восстановления. Рассмотрим сначала первую стадию.

Так как исследуемые волновые фронты когерентны, то необходимо сохранить информацию как об амплитуде, так и о фазе волн. Однако все используемые для записи среды чувствительны только к интенсивности. Поэтому необходимо, чтобы информация о фазе при записи каким-то образом превращалась в модуляцию интенсивности. Стандартный метод решения такой задачи — это *интерферометрия*. К неизвестному волновому фронту добавляется второй когерентный волновой фронт известной амплитудой и фазой (фиг. 95). Тогда их



Фиг. 95. Схема интерферометрической записи.

суммарная интенсивность будет зависеть как от амплитуды, так и от фазы исходного волнового фронта. Таким образом, если

$$a(x, y) = a(x, y) e^{-i\varphi(x, y)} \quad (8.1)$$

представляет волновой фронт, который необходимо записать, а

$$A(x, y) = A(x, y) e^{-i\psi(x, y)} \quad (8.2)$$

представляет «опорный» волновой фронт, с которым интерферирует волна a , то суммарная интенсивность будет

$$\mathcal{I}(x, y) = |A(x, y)|^2 + |a(x, y)|^2 + 2A(x, y)a(x, y)\cos[\psi(x, y) - \varphi(x, y)]. \quad (8.3)$$

В то время как первые два члена этого выражения зависят только от интенсивностей двух волн, третий член зависит от их разности фаз. Таким образом может быть записана информация как об амплитуде, так и о фазе волны a . Запись такой интерференционной картины от двух волновых фронтов можно рассматривать как голограмму.

До сих пор мы не говорили ни о каких особенностях опорной волны A . Свойства, которыми может или должна обладать опорная волна, будут выясняться по мере изложения.

Материалы для записи

Будем предполагать, что запись производится на фотопластинку или пленку. Несмотря на то что голограммы изготовлялись и на других материалах, в том числе на фотохромных стеклах и с помощью электронных детекторов, все же наиболее важным и наиболее широко используемым материалом для записи остается фотопленка. Как и ранее (см., например, гл. 7, § 5 и 7) будем считать, что после проявления фотоматериала его амплитудный коэффициент пропускания является линейной функцией интенсивности света, действовавшего в течение экспонирования. Таким образом, мы предполагаем, что диапазон экспозиций для выбранной нами пленки лежит в пределах линейного участка соответ-

ствующей t — E -кривой (эффекты, связанные с нелинейностью t — E -кривой, будут кратко рассмотрены в § 6). Кроме того, будем считать, что МПФ эмульсии простирается в область достаточно высоких частот, чтобы можно было записать всю пространственную структуру падающего света (эффекты, к которым приводит отказ от этого допущения, будут рассмотрены в § 5). Наконец, положим, что интенсивность $|A|^2$ опорной волны постоянна на всей поверхности пленки; в этом случае амплитудный коэффициент пропускания проявленной пленки можно записать в виде

$$t_f(x, y) = t_b + \beta' (|a|^2 + A^*a + Aa^*), \quad (8.4)$$

где t_b — постоянное «смещающее» пропускание, обусловленное только опорной волной, а β' — произведение наклона t — E -кривой β в рабочей точке на время экспонирования. Отметим, что так же, как и раньше, величина β' отрицательна для негатива и положительна для позитива.

Восстановление исходного волнового фронта

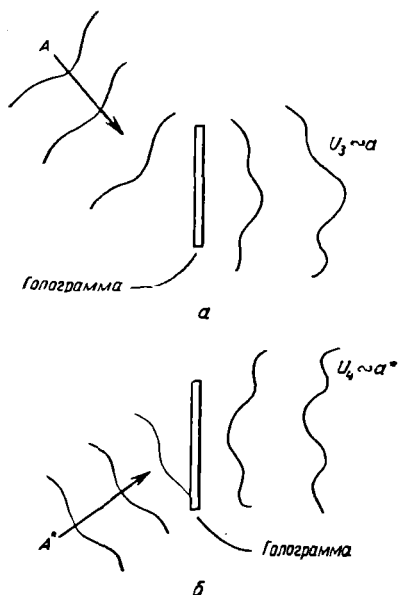
После того как записаны амплитуда и фаза волны объекта a , можно попытаться восстановить эту волну. Предположим, что полученный в результате проявления диапозитив освещается *при восстановлении* когерентной волной $B(x, y)$. Тогда, очевидно, амплитуда света, прошедшего через диапозитив, будет

$$B(x, y) t_f(x, y) = t_b B + \beta' a a^* B + \beta' A^* B a + \beta' A B a^* = \\ = U_1 + U_2 + U_3 + U_4. \quad (8.5)$$

Заметим теперь, что если волна B — точная копия первоначальной *опорной* волны A , то третий член этого соотношения будет иметь вид

$$U_3(x, y) = \beta' |A|^2 a(x, y). \quad (8.6)$$

Примем теперь во внимание наше допущение о том, что интенсивность $|A|^2$ опорной волны постоянна; тогда, очевидно, компонента U_3 с точностью до постоянного множителя будет копией исходного волнового фронта a (фиг. 96, a).



Фиг. 96. Восстановление волнового фронта: при освещении исходной опорной волной A (а) и при освещении комплексно сопряженной опорной волной A^* (б).

Аналогично, если выбрать волну $B(x, y)$ так, чтобы она была комплексно сопряжена опорной волне, т. е. равна $A^*(x, y)$, то для четвертого члена в восстановленной волне получим

$$U_4(x, y) = \beta' |A|^2 a^*, \quad (8.7)$$

т. е. он пропорционален величине, комплексно сопряженной исходному волновому фронту объекта. Этот случай показан на фиг. 96, б.

Отметим, что и в первом и во втором случаях интересующая нас компонента поля (т. е. U_3 при $B = A$ и U_4 при $B = A^*$) возникает одновременно с еще тремя дополнительными компонентами, каждая из которых дает интерференционные помехи. Поэтому для получения нужной нам точной копии волны a (или a^*), очевидно,

необходимо воспользоваться каким-то методом, позволяющим разделять различные компоненты прошедшего через диапозитив света.

Линейность процесса восстановления волнового фронта

Характерные свойства фотопленки, как видно из соотношения (8.4), проявляются в нелинейной связи между световым полем волны, падающей на пленку во время экспонирования, и полем волны, прошедшей через нее после ее проявления. Поэтому на первый взгляд может показаться, что теория линейных систем не имеет отношения к методу восстановления волнового фронта. Однако, к счастью, это не так. В то время как полный процесс отображения, осуществляемый пленкой, является нелинейным, отображение поля объекта a в компоненту U_3 света, прошедшего через пленку, представляет собой чисто линейный процесс, о чем свидетельствует и их прямая пропорциональность [см. (8.6)]. Аналогично связь между полем a^* и компонентой U_4 света, прошедшего через пленку, как видно из соотношения (8.7), также линейна. Следовательно, если поле a объекта рассматривать как входной сигнал, а компоненту поля U_3 (или U_4) прошедшего света — как выходной, то определенная таким образом система является линейной. Нелинейность пленки проявляется в том, что на выходе получается сразу несколько членов, но при этом интересующая нас компонента никак не искажается, конечно, при условии, что весь диапазон экспозиций лежит в пределах линейной области $t - E$ -кривой.

Получение изображений путем восстановления волнового фронта

До настоящего времени мы исследовали только вопрос о восстановлении волнового фронта, приходящего на поверхность детектора от когерентно освещенного предмета. Теперь, чтобы рассмотреть процесс восстановления волнового фронта как средство получения изображения, нам потребуется лишь слегка изменить нашу точку зрения.

Для этого заметим, что волна, соответствующая компоненте U_3 [см. (8.6)], является точной копией волны A исходного объекта. Поэтому наблюдателю она будет казаться распространяющейся от исходного предмета, хотя предмет на самом деле давно удален. Таким образом, если при восстановлении для освещения используется опорная волна A , то можно считать, что компонента U_3 прошедшей волны будет давать *мнимое изображение* исходного объекта. Этот случай показан на фиг. 97, *a* и *б* для объекта в виде точечного источника.

Аналогично, если для освещения при восстановлении используется волна A^* , комплексно сопряженная опорной, то компонента U_4 [см. (8.7)] тоже дает изображение, но на этот раз действительное, которое соответствует действительной фокусировке лучей в пространстве. Чтобы доказать это утверждение, обратимся к рассмотренным выше свойствам линейности в применении к объекту в виде простого точечного источника. Соответствующие результаты для объектов более сложного вида могут быть получены как линейная суперпозиция решений, найденных для объекта в виде точечного источника.

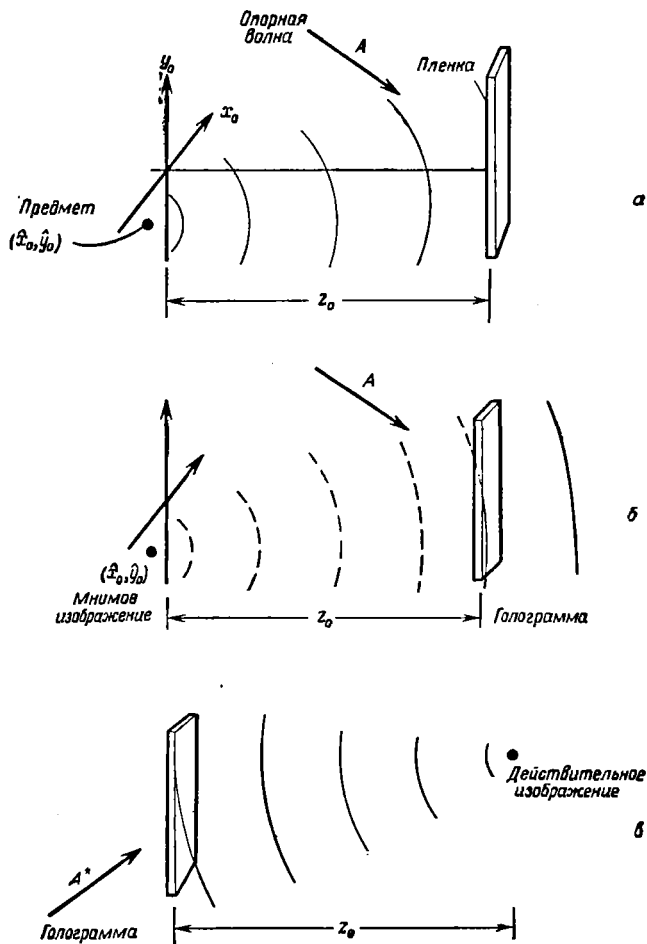
Падающий на пленку свет состоит из опорной волны A и простой сферической волны

$$a(x, y) = a_0 \exp \left[jk \sqrt{z_0^2 + (x - \hat{x}_0)^2 + (y - \hat{y}_0)^2} \right], \quad (8.8)$$

где (\hat{x}_0, \hat{y}_0) — соответственно (x, y) координаты точки объекта, а z_0 — расстояние по нормали до регистрирующей поверхности. Если осветить полученную голограмму волной A^* , то в прошедшем через нее свете появится компонента

$$U_4(x, y) = \beta' |A|^2 a^*(x, y) = \\ = \beta' |A|^2 a_0^* \exp \left[-jk \sqrt{z_0^2 + (x - \hat{x}_0)^2 + (y - \hat{y}_0)^2} \right], \quad (8.9)$$

которая представляет собой сферическую волну, *расходящуюся* от действительного центра, расположенного на расстоянии z_0 от голограммы (фиг. 97, *в*). Более сложный объект можно рассматривать как совокупность точечных источников с различными амплитудами и фазами. В соответствии со сказанным выше при условии



Фиг. 97. Получение изображения восстановлением волнового фронта.

а — запись голограммы предмета в виде точечного источника; б — получение мнимого изображения; в — получение действительного изображения.

линейности от каждого такого точечного источника получается свое действительное изображение. Таким образом, этим методом можно получить действительное изображение всего объекта.

Заметим, что амплитуда волны, описываемой соотношением (8.9), пропорциональна амплитуде a_0^* , комплексно сопряженной амплитуде исходного объекта в виде точечного источника. Аналогично для более сложного объекта действительное изображение, полученное с помощью голограммы, всегда комплексно сопряжено исходному объекту. Такое изменение фазы не влияет на интенсивность изображения, но может быть существенным для ряда приложений, в которых используется не только амплитуда, но и фаза изображения (см. § 9).

Следует еще раз подчеркнуть, что во всех этих случаях мы рассматривали только одну из четырех волн, прошедших через голограмму. Это возможно только в том случае, когда соответствующим выбором опорной волны можно устранить или отделить нежелательные компоненты от интересующего нас изображения. Если же это не так, то необходимо принимать во внимание интерференцию всех компонент прошедшего света.

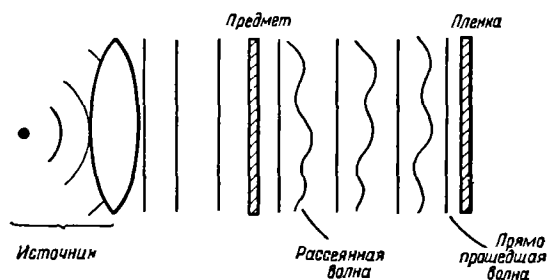
§ 3. ГОЛОГРАММЫ ГАБОРА

Учитывая вышесказанное, рассмотрим теперь процесс восстановления волнового фронта в том виде, как он первоначально был предложен и выполнен Габором. Затем (в § 4) обратимся к тем модификациям этого процесса, которые необходимы для получения изображения более высокого качества.

Источник опорной волны

Схема получения голограммы Габора представлена на фиг. 98. Пусть объект обладает большой прозрачностью, т. е. амплитудным коэффициентом пропускания

$$t(x_0, y_0) = t_0 + \Delta t(x_0, y_0), \quad (8.10)$$



Ф и г. 98. Запись голограммы Габора.

где t_0 характеризует высокое среднее пропускание, Δt — отклонение от этой средней величины и

$$|\Delta t| \ll |t_0|. \quad (8.11)$$

Если такой объект осветить когерентным параллельным пучком, как показано на фиг. 98, то прошедший через него свет будет состоять из двух компонент: 1) интенсивной однородной плоской волны, отвечающей наличию в пропускании члена t_0 , и 2) слабой рассеянной волны, отвечающей изменению пропускания $\Delta t(x_0, y_0)$. Распределение интенсивности света на фотопластинке, находящейся на расстоянии z_0 от объекта, можно записать в виде

$$\begin{aligned} \mathcal{I}(x, y) &= |A + a(x, y)|^2 = \\ &= A^2 + |a(x, y)|^2 + Aa(x, y) + Aa^*(x, y), \end{aligned} \quad (8.12)$$

где A — амплитуда плоской волны, а $a(x, y)$ — амплитуда рассеянного света.

Таким образом, благодаря высокому среднему пропусканию t_0 объект в некотором роде сам является источником опорной волны. В результате интерференции света, прямо прошедшего через объект и рассеянного им, получается распределение интенсивности, которое зависит как от амплитуды, так и от фазы рассеянной волны $a(x, y)$.

Сопряженные изображения

Предположим, что после обычной обработки фотопластинки амплитуда ее пропускания пропорциональна экспозиции. Тогда

$$t_f(x, y) = t_b + \beta' (|a|^2 + Aa + Aa^*). \quad (8.13)$$

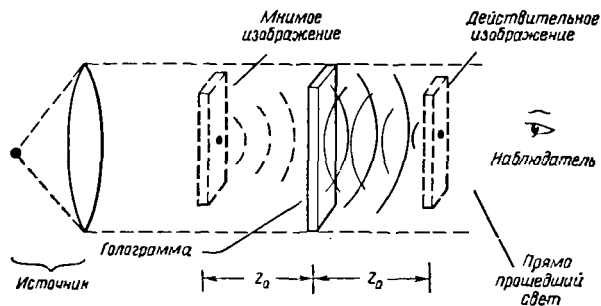
Если теперь осветить пластинку нормально падающей плоской волной с амплитудой B , то амплитуда светового поля за пластинкой описывается суммой четырех членов

$$Bt_f = Bt_b + \beta' B |a(x, y)|^2 + \beta' A B a(x, y) + \beta' A B a^*(x, y). \quad (8.14)$$

Первый член соответствует однородно ослабленной плоской волне, прошедшей без рассеяния через диапозитив. Вторым членом можно пренебречь из-за его малости в соответствии с допущением (8.11), из которого следует, что

$$|a(x, y)| \ll A. \quad (8.15)$$

Третий член описывает компоненту поля, амплитуда которой пропорциональна амплитуде исходной волны $a(x, y)$. Эта волна кажется исходящей из расположенного на расстоянии z_0 от голограммы мнимого изображения исходного предмета (фиг. 99). Совершенно аналогично присутствие в выражении (8.12) четвертого члена, пропорционального $a^*(x, y)$, приводит к тому, что на расстоянии z_0 от голограммы, но по другую сторону



Фиг. 99. Образование сопряженных изображений.

от нее, чем мнимое изображение, получается действительное изображение (см. фиг. 99).

Таким образом, в методе Габора на одной оси с голограммой получаются два изображения изменения пропускания предмета Δt — действительное и мнимое. Эти так называемые *сопряженные изображения* находятся на расстоянии $2z_0$ друг от друга и сопровождаются когерентным фоном Vt_b .

Из соотношения (8.14) видно, что пропускание может быть положительным и отрицательным; следовательно, амплитуды когерентного фона и волн, формирующих изображение, будут вычитаться или складываться в зависимости от знака пропускания (величина β' положительна при положительном пропускании и отрицательна при отрицательном пропускании). Так как в заключительной стадии процесса восстановления мы наблюдаем картину интерференции фона и волн, формирующих изображение, то голограммы с положительным пропусканием дают изображения с положительным контрастом, а голограммы с отрицательным пропусканием — изображения с отрицательным (т. е. обращенным) контрастом.

Недостатки голограммы Габора

Методу Габора присущи некоторые недостатки, которые ограничивают сферу его применения. Возможно, наиболее существенным из них является ограничение, связанное с допущением (8.11) о высокой прозрачности объекта, и вытекающее из него допущение (8.15). Если не выполняется указанное выше ограничение, то у волны, прошедшей через голограмму, уже нельзя пренебречь компонентой

$$U(x, y) = \beta' B |a(x, y)|^2. \quad (8.16)$$

Действительно, если средняя прозрачность невелика, то соответствующая второму члену в (8.14) волна может стать самой интенсивной и совершенно подавить менее интенсивные изображения. Таким образом, с помощью голограммы Габора удается восстановить изображения

объектов в виде, например, непрозрачных букв на прозрачном фоне, но не объектов в виде прозрачных букв на непрозрачном фоне. Это ограничение является существенным препятствием при использовании голограмм Габора во многих областях.

Другим серьезным ограничением является то, что при восстановлении получается вместо одного два сопряженных изображения. Недостатком является не само по себе наличие двух изображений, а то, что их невозможно разделить. Когда фокусируется действительное изображение, оно всегда сопровождается расфокусированным мнимым изображением. Аналогично наблюдатель, настроившийся на мнимое изображение, увидит одновременно и расфокусированное действительное изображение. Таким образом, даже для объектов с большим пропусканием качество изображений ухудшается из-за их раздвоенности. Для устранения или уменьшения этого недостатка предложен ряд методов (см., например, [7]), в том числе и способ, выдвинутый самим Габором [11]. Наиболее удачным из них является метод Лейта и Упатника [8]; на нем мы остановимся подробно в следующем параграфе.

§ 4. ГОЛОГРАММА ЛЕЙТА И УПАТНИКА

Наиболее широко используемый тип голограмм — это так называемые *голограммы Лейта — Упатника*, известные также под названием голограмм с *наклонным опорным пучком*. Основная их особенность состоит в том, что в процессе записи используется добавочная опорная волна, в отличие от метода Габора, в котором опорной служит волна, непосредственно прошедшая через объект. Более того, в методе Лейта — Упатника направление опорной волны не совпадает с направлением оси, на которой расположены объект и пленка, а составляет с ней конечный угол.

Впервые возможность практической реализации голограммы такого типа была доказана Лейтом и Упатником в 1962 г. [8]. С исторической точки зрения интересно отметить, что первые голограммы по этому методу были получены без использования лазерного источ-

ника. Однако пока метод не опирался на использование высококогерентного лазерного света, многие его возможности оставались не выявленными [9, 12].

Запись голограммы

Одна из возможных схем записи голограмм Лейта — Упатника показана на фиг. 100. Излучение от точечного источника собирается линзой L в параллельный пучок. Часть получившейся таким образом плоской волны освещает объект, которым обычно является диапозитив с общим коэффициентом пропускания по амплитуде $t_0(x_0, y_0)$. Другая часть плоской волны падает на призму P , расположенную над объектом, и отклоняется вниз под углом θ по отношению к оси пленки. Таким образом, на светочувствительной поверхности получается сумма двух когерентных волн, одна из которых представляет собой волну, пропущенную объектом, а вторая — наклонно падающую плоскую волну. Распределение амплитуды на пленке можно записать следующим образом:

$$U(x, y) = Ae^{-i2\pi\alpha y} + a(x, y), \quad (8.17)$$

где пространственная частота α опорной волны определяется выражением

$$\alpha = \frac{\sin \theta}{\lambda}. \quad (8.18)$$

Распределение интенсивности на пленке будет, очевидно, иметь вид

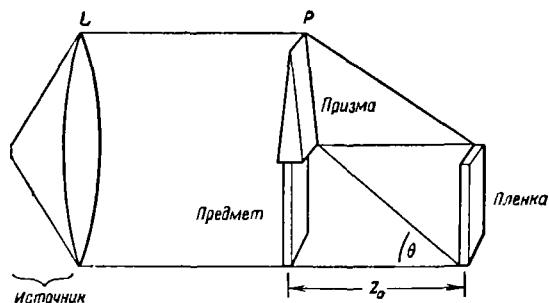
$$\mathcal{I}(x, y) = A^2 + |a(x, y)|^2 + 2Aa(x, y)e^{i2\pi\alpha y} + 2Aa^*(x, y)e^{-i2\pi\alpha y}. \quad (8.19)$$

Его можно записать иначе, выражая a через амплитуду и фазу

$$a(x, y) = a(x, y)e^{-i\varphi(x, y)} \quad (8.20)$$

и объединяя последние два члена в (8.19). Тогда получим

$$\mathcal{I}(x, y) = A^2 + a^2(x, y) + 2Aa(x, y) \cos [2\pi\alpha y - \varphi(x, y)]. \quad (8.21)$$



Фиг. 100. Запись голограммы Лейта и Упатника.

Это выражение показывает, что амплитуда и фаза света, прошедшего через предмет, записываются соответственно как амплитудная и фазовая модуляция несущей пространственной волны частоты α . Если несущая частота достаточно велика (сколь велика она должна быть, мы скоро увидим), то распределение амплитуды и фазы может быть однозначно восстановлено по интерференционной картине на пластинке.

Получение восстановленных изображений

Как обычно, фотопластинка после обработки представляет собой диапозитив с амплитудным коэффициентом пропускания, пропорциональным экспозиции. В этом случае пропускание пластинки можно записать как

$$t_f(x, y) = t_0 + \beta' (|a|^2 + Aae^{i2\pi\alpha y} + Aa^*e^{-i2\pi\alpha y}). \quad (8.22)$$

Для удобства введем для приведенных выше четырех слагаемых обозначения

$$\begin{aligned} t_1 &= t_0, & t_3 &= \beta' Aa(x, y)e^{i2\pi\alpha y}, \\ t_2 &= \beta' |a(x, y)|^2, & t_4 &= \beta' Aa^*(x, y)e^{-i2\pi\alpha y}. \end{aligned} \quad (8.23)$$

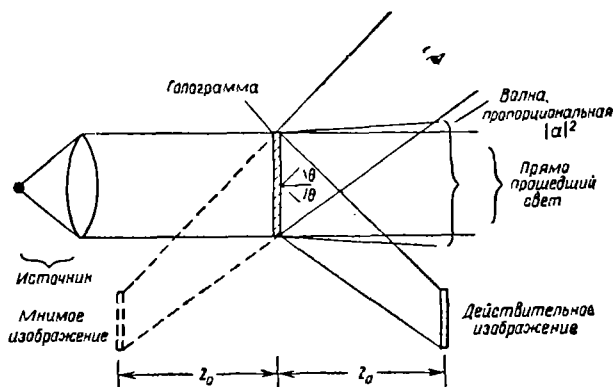
Предположим теперь, что голограмма освещается (фиг. 101) нормально падающей однородной плоской волной B . Поле волны, прошедшей через голограмму, может быть представлено в виде суммы четырех компонент, каждая из которых определяется соот-

ветствующим слагаемым пропускания [см. (8.23)]:

$$\begin{aligned} U_1 &= t_b B, & U_3 &= \beta' B A a(x, y) e^{j2\pi a y}, \\ U_2 &= \beta' B |a(x, y)|^2, & U_4 &= \beta' B A a^*(x, y) e^{-j2\pi a y}. \end{aligned} \quad (8.24)$$

Компонента поля U_1 — это просто плоская равномерно ослабленная падающая волна, которая распространяется по нормали к диапозитиву. Вторая компонента U_2 зависит от пространственных координат и поэтому состоит из набора плоских волн, распространяющихся под разными углами к оптической оси. Вскоре при более детальном рассмотрении мы убедимся, что если ширина спектра $a(x, y)$ мала по сравнению с несущей частотой α , то энергия во второй волне распространяется в направлении, достаточно близком к нормали к плоскости диапозитива, и в результате пространственно отделяется от интересующих нас изображений.

Компонента U_3 пропорциональна амплитуде волны a , идущей от исходного объекта, умноженной на экспоненциальный множитель. Пропорциональность амплитуде a означает, что этот член образует мнимое изображение объекта на расстоянии z_0 от диапозитива, в то время как экспоненциальный множитель $e^{j2\pi a y}$ показывает, что изображение отклонено от оси голограммы на угол θ (см. фиг. 101). Подобным образом, член U_4 пропорционален комплексно сопряженной амплитуде a^*

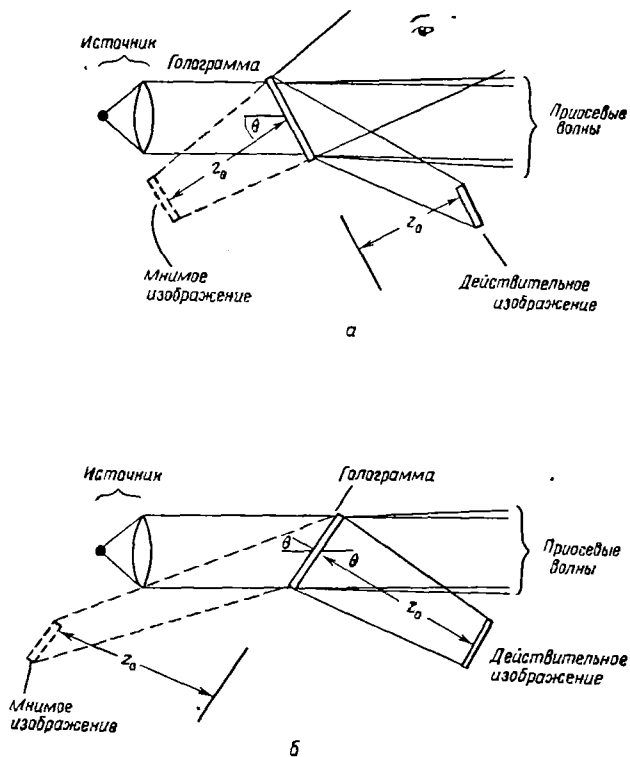


Фиг. 101. Восстановление изображений,

волны, идущей от объекта, поэтому образует действительное изображение, расположенное на расстоянии z_0 от диапозитива, но с противоположной стороны по отношению к мнимому изображению. Наличие экспоненциального множителя $\exp(-j2\pi u y)$ показывает, что действительное изображение получается в направлении, составляющем угол $-\theta$ с осью голограммы; это также показано на фиг. 101.

Самый существенный вывод из всего вышесказанного состоит в том, что, хотя в результате рассмотренного процесса появляются действительное и мнимое изображения, теперь они уже пространственно отделены друг от друга и от компонент U_1 и U_2 . Это разделение получается благодаря применению наклонного опорного пучка. Фактически для успешного разделения действительного и мнимого изображений необходимо, чтобы угол θ между опорным пучком и осью превышал некоторое минимальное значение $\theta_{\text{мин}}$ (скоро мы более подробно рассмотрим, что представляет собой минимальное значение опорного угла $\theta_{\text{мин}}$). Когда угол θ превышает $\theta_{\text{мин}}$, действительное и мнимое изображения не искажаются другими волновыми компонентами. Поскольку изображения могут наблюдаться независимо от когерентного фона, возникающего из-за наличия среднего пропускания, знак каждой из волновых компонент U_3 и U_4 [см. (8.24)] теперь несуществен. Таким образом, диапозитив может быть как позитивом, так и негативом; в любом случае получается «позитивное» изображение. Для практических целей вообще предпочтительнее пользоваться непосредственно негативами, избегая, таким образом, двухступенчатого процесса, необходимого для получения позитивных изображений.

Наконец, следует отметить, что, хотя мы выбрали для освещения голограммы нормально падающую плоскую волну, которая не являлась ни копией опорной волны, ни комплексно сопряженной ей волной, мы все же получили одновременно как действительное, так и мнимое изображения. Очевидно, введенное в § 2 ограничение вида восстанавливающей волны являлось слишком сильным. Однако если толщина регистрирующей



Фиг. 102. Возможные схемы восстановления изображений.

а — голограмма освещается копией исходной опорной волны; *б* — голограмма освещается волной, комплексно сопряженной исходной опорной волне.

эмульсии сравнима с масштабом поперечной структуры интерференционной картины, воспроизводимой на пластинке, то свойства восстанавливающего света становятся существенными. Как будет показано в § 7, в этом случае уже необходимо освещать голограмму в точности такой же волной, как опорная (для получения мнимого изображения), или комплексно сопряженной волной (для получения действительного изображения). Предоставляем читателю в качестве упражнения (см. задачу 8.1) убедиться, что соотношение между восстанавли-

вающей волной, голограммой и изображением для двух возможных случаев будет таким, как показано на фиг. 102.

Минимальный опорный угол

Вернемся к схеме восстановления, изображенной на фиг. 101. Чтобы можно было отделить сопряженные изображения друг от друга и от света, распространяющегося по нормали к диапозитиву, угол наклона опорной волны θ должен быть больше некоторого минимального значения $\theta_{\text{мин}}$. Чтобы определить это значение, достаточно найти минимальную величину несущей частоты α , при которой пространственно-частотные спектры компонент t_3 и t_4 (т. е. составляющих пропускания голограммы, ответственных за получение мнимого и действительного изображений) не перекрываются друг с другом и со спектрами компонент t_1 и t_2 . Когда такое перекрытие отсутствует, в принципе можно с помощью собирающих линз получить фурье-преобразование голограммы, устранить нежелательные компоненты с помощью соответствующих диафрагм, помещенных в фокальной плоскости, и затем осуществить второе фурье-преобразование с оставшейся частью прошедшего света, которое выделит одно из двух сопряженных изображений¹⁾.

Рассмотрим пространственно-частотные спектры различных слагаемых пропускания (8.23). Не учитывая конечность размеров диапозитива, непосредственно имеем

$$G_1(f_x, f_y) = \mathcal{F}\{t_1(x, y)\} = t_b \delta(f_x, f_y). \quad (8.25)$$

Используя теорему автокорреляции, можно написать также

$$G_2(f_x, f_y) = \mathcal{F}\{t_2(x, y)\} = \beta' G_a(f_x, f_y) * G_a(f_x, f_y). \quad (8.26)$$

¹⁾ На практике метод пространственной фильтрации редко применяется для разделения сопряженных изображений. Если опорный угол удовлетворяет указанным здесь требованиям, то изображения будут разделяться сами по себе из-за разницы в направлениях распространения соответствующих волновых компонент (ср. фиг. 101). Однако рассмотрение пространственной фильтрации позволяет сравнительно просто найти достаточные условия такого разделения.

где

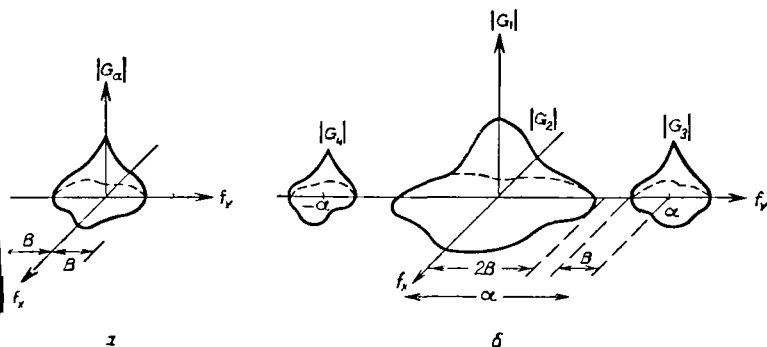
$$G_a(f_x, f_y) = \mathcal{F}\{a(x, y)\}, \quad (8.27)$$

а \star обозначает операцию автокорреляции. Наконец, имеем

$$G_3(f_x, f_y) = \mathcal{F}\{t_3(x, y)\} = \beta' AG_a(f_x, f_y - \alpha), \quad (8.28)$$

$$G_4(f_x, f_y) = \mathcal{F}\{t_4(x, y)\} = \beta' AG_a^*(-f_x, f_y - \alpha). \quad (8.29)$$

Заметим теперь, что G_a и исходный объект имеют одинаковые полосы частот, так как их спектры отличаются только передаточной функцией процесса распространения, которая является чисто фазовой функцией [см. (3.47)]. Предположим, что наибольшая пространственная частота для объекта равна B период/мм. Примерный спектр функции $|G_a|$ в этом случае приведен на фиг. 103, а. Соответствующий ему спектр пропускания голограммы изображен на фиг. 103, б. Член $|G_1|$ представляет собой просто δ -функцию в начале координат в плоскости $f_x f_y$. Член $|G_2|$, пропорциональный автокорреляционной функции $|G_a|$, простирается в область высоких частот вплоть до частоты $2B$. Наконец, член $|G_3|$ пропорционален функции $|G_a|$, но со смещенной в точку $(0, \alpha)$ центральной частотой, а член $|G_4|$ зеркально симметричен ему и пропорционален функции $|G_a|$, но с центральной частотой в точке $(0, -\alpha)$.



Фиг. 103. Спектры объекта (а) и голограммы (б).

Анализ фиг. 103, б показывает, что $|G_3|$ и $|G_4|$ можно отделить от $|G_2|$, если

$$\alpha \geq 3B \quad (8.30)$$

или если

$$\sin \theta \geq 3B\lambda. \quad (8.31)$$

Очевидно, минимально допустимый опорный угол определяется условием

$$\theta_{\min} = \arcsin 3B\lambda. \quad (8.32)$$

Когда опорная волна намного интенсивнее волны, идущей от объекта, это требование можно как-то ослабить. Физически член G_2 появляется из-за интерференции света, идущего от каждой точки объекта, со светом от всех других точек объекта, в то время как G_3 и G_4 появляются из-за интерференции опорной волны и волн от объекта. Когда волна от объекта намного слабее, чем опорная (т. е. когда $a \ll A$), член G_2 много меньше G_1 , G_3 и G_4 и им можно пренебречь. В этом случае минимальный опорный угол равен значению угла, при котором G_3 и G_4 еще разделяются, или

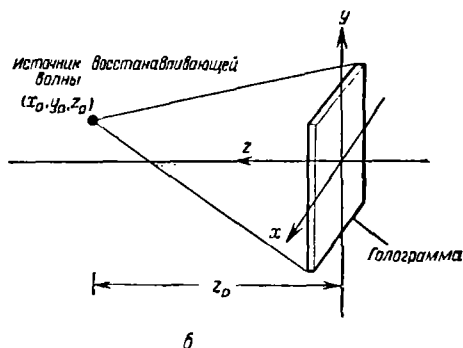
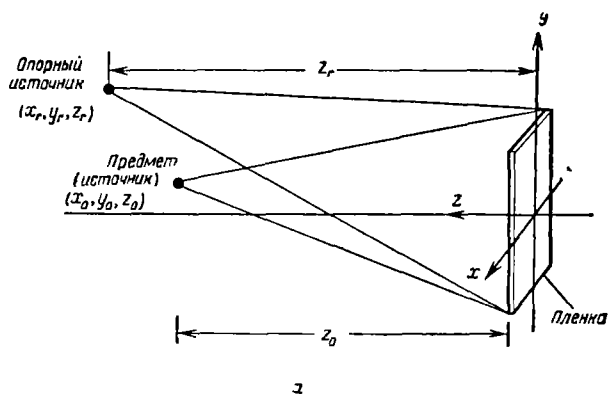
$$\theta_{\min} = \arcsin B\lambda. \quad (8.33)$$

Обобщение на случай неколлимированных опорной и восстанавливающей волн

Теперь естественно рассмотреть более общие схемы записи голограмм и восстановления изображений. Особый интерес представляют системы восстановления волнового фронта, в которых используются сферические опорные и сферические восстанавливающие волны¹⁾. Обращаясь к фиг. 104, а, будем полагать, что опорная волна излучается точечным источником с координатами (x_r, y_r, z_r) . Так как отображение амплитуды объекта остается линейным вне зависимости от характера опорной волны (см. § 2), достаточно²⁾ рассмотреть объект

¹⁾ Более полное обсуждение затронутого здесь вопроса, включая рассмотрение влияния аберраций, можно найти в работах [13, 14].

²⁾ При таком упрощении подразумевается, что угол наклона опорной волны достаточен, чтобы можно было разделить действительное и мнимое изображения.



Фиг. 104. Обобщенные схемы записи (а) и восстановления (б).

в виде точечного источника с координатами (x_0, y_0, z_0) . Как показано на фиг. 104, б, голограмма при восстановлении освещается сферической волной с центром в точке (x_p, y_p, z_p) . Для большей общности будем полагать, что длина волны λ_2 при восстановлении может отличаться от длины волны λ_1 , используемой в процессе записи.

Если для сферических волн воспользоваться квадратичным приближением, то суммарное поле падающее на фотопленку, можно записать как

$$U(x, y) = A \exp \left\{ j \frac{\pi}{\lambda_1 z_r} [(x - x_r)^2 + (y - y_r)^2] \right\} + a \exp \left\{ j \frac{\pi}{\lambda_1 z_0} [(x - x_0)^2 + (y - y_0)^2] \right\}, \quad (8.34)$$

где A и a — комплексные постоянные, представляющие относительные амплитуды и фазы двух сферических волн. Следовательно, соответствующее распределение интенсивности в интерференционной картине запишется следующим образом:

$$\begin{aligned} \mathcal{I}(x, y) = & |A|^2 + |a|^2 + \\ & + Aa^* \exp \left\{ j \frac{\pi}{\lambda_1 z_r} [(x - x_r)^2 + (y - y_r)^2] - \right. \\ & \left. - j \frac{\pi}{\lambda_1 z_0} [(x - x_0)^2 + (y - y_0)^2] \right\} + \\ & + A^* a \exp \left\{ -j \frac{\pi}{\lambda_1 z_r} [(x - x_r)^2 + (y - y_r)^2] + \right. \\ & \left. + j \frac{\pi}{\lambda_1 z_0} [(x - x_0)^2 + (y - y_0)^2] \right\}. \end{aligned} \quad (8.35)$$

Если амплитудный коэффициент пропускания полученного после проявления диапозитива пропорционален экспозиции, то интересующие нас члены в выражении для пропускания будут иметь вид

$$\begin{aligned} t_3 = \beta' A a^* \exp \left\{ j \frac{\pi}{\lambda_1 z_r} [(x - x_r)^2 + (y - y_r)^2] - \right. \\ \left. - j \frac{\pi}{\lambda_1 z_0} [(x - x_0)^2 + (y - y_0)^2] \right\}, \\ t_4 = \beta' A^* a \exp \left\{ -j \frac{\pi}{\lambda_1 z_r} [(x - x_r)^2 + (y - y_r)^2] + \right. \\ \left. + j \frac{\pi}{\lambda_1 z_0} [(x - x_0)^2 + (y - y_0)^2] \right\}. \end{aligned} \quad (8.36)$$

Голограмма освещается сферической волной

$$U_p(x, y) = B \exp \left\{ j \frac{\pi}{\lambda_2 z_p} [(x - x_p)^2 + (y - y_p)^2] \right\}. \quad (8.37)$$

Интересующие нас два волновых фронта за голограммой описываются выражениями, полученными перемножением членов (8.36) и (8.37):

$$\begin{aligned} U_3(x, y) = t_3 B \exp \left\{ j \frac{\pi}{\lambda_2 z_p} [(x - x_p)^2 + (y - y_p)^2] \right\}, \\ U_4(x, y) = t_4 B \exp \left\{ j \frac{\pi}{\lambda_2 z_p} [(x - x_p)^2 + (y - y_p)^2] \right\}. \end{aligned} \quad (8.38)$$

Чтобы установить природу этих полей, мы должны исследовать, каким образом они зависят от (x, y) . Так как мы учитываем только линейные и квадратичные члены относительно x и y , то выражения для U_3 и U_4 можно рассматривать в квадратичном приближении как сферические волны. Присутствие линейных членов указывает на то, что волны сходятся к точкам, которые не лежат на оси z (или расходятся из них). Теперь остается только найти точные положения (действительное и мнимое) этих точек фокусировки.

Заметим, что фазовые члены, зависящие от x^2 и y^2 , определяют z -координату фокуса, в то время как линейные члены определяют его смещения относительно оси. Если собрать все члены, содержащие $(x^2 + y^2)$, то экспоненциальные множители запишутся в виде

$$\exp \left[j\pi \left(\pm \frac{1}{\lambda_1 z_r} \mp \frac{1}{\lambda_1 z_0} + \frac{1}{\lambda_2 z_p} \right) (x^2 + y^2) \right], \quad (8.39)$$

где верхние знаки соответствуют одной волновой компоненте (U_3), а нижние — другой (U_4). Сравнивая это выражение с выражением для сферической волны, распространяющейся из точки, расположенной слева от голограммы на расстоянии z_i от нее (см. фиг. 104, б),

$$\exp \left[j \frac{\pi}{\lambda_2 z_i} (x^2 + y^2) \right], \quad (8.40)$$

получаем, что расстояние z_i от видимого нами источника до голограммы определится следующим образом:

$$z_i = \left(\frac{1}{z_p} \pm \frac{\lambda_2}{\lambda_1 z_r} \mp \frac{\lambda_2}{\lambda_1 z_0} \right)^{-1}, \quad (8.41)$$

где опять верхние знаки относятся к первой волне, а нижние — ко второй. Когда расстояние z_i положительно, получается мнимое изображение, расположенное слева от голограммы на фиг. 104, б; при отрицательном z_i изображение действительное и находится справа от голограммы.

Соотношение (8.41) определяет z -координаты изображений; теперь остается найти x - и y -координаты. Это можно сделать, если собрать все члены выражений (8.38), линейные относительно x и y , и сопоставить их

с выражением

$$\exp\left[j \frac{2\pi}{\lambda_2 z_t} (x_i x + y_i y)\right]. \quad (8.42)$$

Два интересующих нас члена имеют вид

$$\exp\left\{j2\pi\left[\left(\pm \frac{x_0}{\lambda_1 z_0} \mp \frac{x_r}{\lambda_1 z_r} - \frac{x_p}{\lambda_2 z_p}\right)x + \left(\pm \frac{y_0}{\lambda_1 z_0} \mp \frac{y_r}{\lambda_1 z_r} - \frac{y_p}{\lambda_2 z_p}\right)y\right]\right\}. \quad (8.43)$$

Сравнивая (8.42) и (8.43), получаем смещения изображений относительно оси:

$$\begin{aligned} x_i &= \pm \frac{\lambda_2 z_i}{\lambda_1 z_0} x_0 \mp \frac{\lambda_2 z_i}{\lambda_1 z_r} x_r - \frac{z_i}{z_p} x_p, \\ y_i &= \pm \frac{\lambda_2 z_i}{\lambda_1 z_0} y_0 \mp \frac{\lambda_2 z_i}{\lambda_1 z_r} y_r - \frac{z_i}{z_p} y_p. \end{aligned} \quad (8.44)$$

Отметим, в частности, что изменения координат объекта $(\Delta x_0, \Delta y_0)$ вызывают изменения координат изображения

$$\left(\pm \frac{\lambda_2 z_i}{\lambda_1 z_0} \Delta x_0, \pm \frac{\lambda_2}{\lambda_1} \frac{z_i}{z_0} \Delta y_0\right).$$

Таким образом, приходим к следующему выражению для увеличения, которое получается при использовании для восстановления описанной обобщенной схемы:

$$M = \left| \frac{\Delta x_i}{\Delta x_0} \right| = \left| \frac{\Delta y_i}{\Delta y_0} \right| = \left| \frac{\lambda_2 z_i}{\lambda_1 z_0} \right|. \quad (8.45)$$

Подставляя (8.41) в (8.45), находим для M равносильное выражение

$$M = \left| 1 - \frac{z_0}{z_r} \mp \frac{\lambda_1 z_0}{\lambda_2 z_p} \right|^{-1}. \quad (8.46)$$

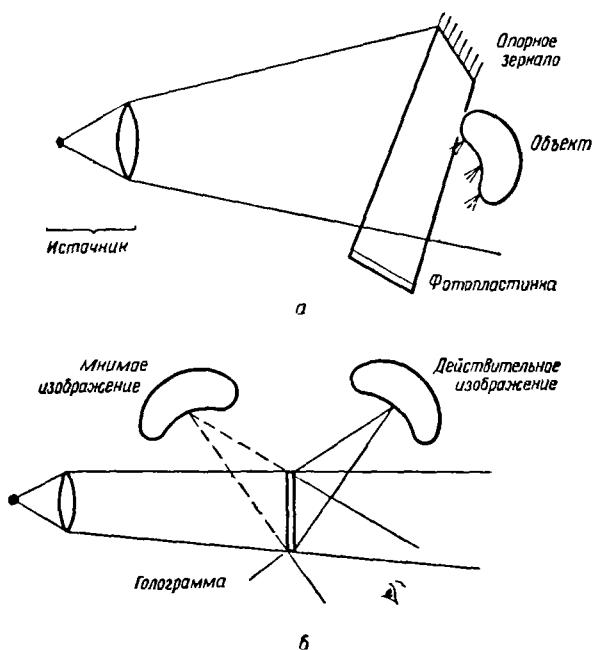
При использовании коллимированной опорной волны ($z_r = \infty$) и коллимированной восстанавливающей волны ($z_p = \infty$) увеличение, как можно видеть, равно единице вне зависимости от величины отношения λ_1/λ_2 ¹⁾. Единичное увеличение получается также для мнимого изображения, когда $\lambda_2 = \lambda_1$ и $z_p = z_r$, и для действительного изображения, когда $\lambda_2 = \lambda_1$ и $z_p = -z_r$ (см. задачу 8.3).

¹⁾ Однако расстояние z_i от плоскости диапозитива зависит от λ_1/λ_2 .

Приложение результатов к трехмерной фотографии

В 1964 г. Лейт и Упатник впервые опубликовали результаты своих успешных экспериментов по применению голографии для получения трехмерных изображений [9]. Успехи в этой области во многом были обязаны появлению гелий-неоновых лазеров с их высокой временной и пространственной когерентностью.

На фиг. 105, *а* приведена общая схема записи голограмм от трехмерных объектов. Интересующий нас объект освещается когерентным светом. Кроме того, часть света падает на «опорное» зеркало, которое располагается вблизи объекта. Отраженное зеркалом излучение попадает непосредственно на фотопластинку и



Фиг. 105. Получение голографического изображения трехмерного объекта.

а — запись голограммы; *б* — восстановление изображений.

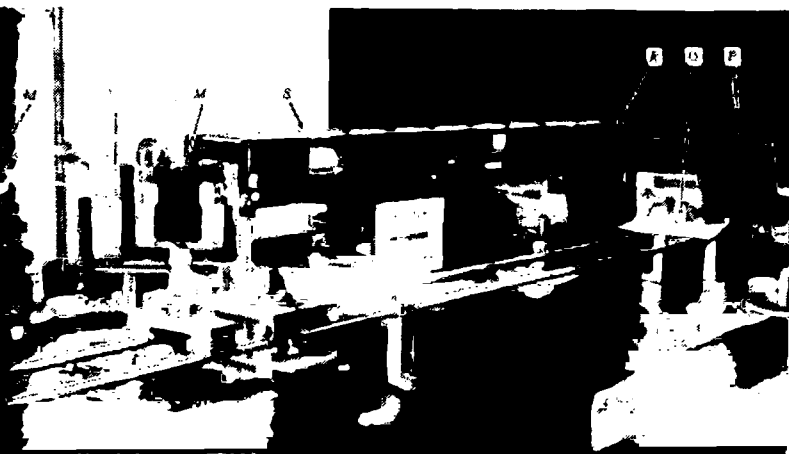
служит опорной волной, которая интерферирует со светом, отраженным от самого объекта. Таким образом, на фотопластинке записывается голограмма от трехмерного объекта.

Чтобы восстановить изображение, фотопластинку после проявления освещают восстанавливающей волной, как показано на фиг. 105, б. За пластинкой возникает трехмерное мнимое изображение. Так как волны, падающие на пластинку при записи, и волны, получающиеся при восстановлении, тождественны, изображение сохраняет все трехмерные свойства исходного объекта. В частности, хорошо заметен эффект параллакса: меняя точку наблюдения, мы можем как бы «заглянуть за объект», расположенный на переднем плане¹⁾.

Действительное изображение получается в пространстве между фотопластинкой и наблюдателем. В случае трехмерных объектов это изображение менее реалистично, чем мнимое изображение. Во-первых, точки объекта, расположенные ближе всего к фотопластинке (и, следовательно, ближе всего к наблюдателю, рассматривающему исходный объект), оказываются в восстановленном изображении опять ближе всего к фотопластинке, т. е. в данном случае являются *наиболее удаленными от наблюдателя* (ср. фиг. 105, б). Таким образом, для наблюдателя, рассматривающего действительное изображение, параллакс будет не таким, как у исходного объекта, и изображение будет казаться «вывернутым» в некотором особом смысле (чтобы полностью понять этот эффект, необходимо видеть его своими глазами). Такого рода изображения известны под названием *псевдоскопических*.

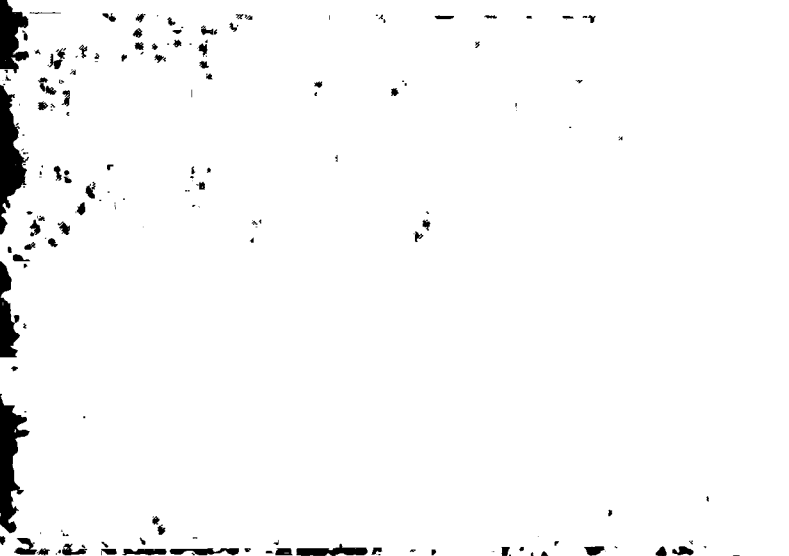
Вторым недостатком является малая глубина фокуса. Поэтому если попытаться непосредственно на пленке получить действительное изображение, то оказывается, что (для голограмм разумных размеров) невозможно изготовить сколько-нибудь разборчивый снимок. Эту проблему можно в какой-то мере решить, если освещать только

¹⁾ Параллакс ограничивается размерами голограммы как окошком, через которое мы смотрим на предметы, расположенные в трехмерном пространстве. — *Прим. ред.*



Ф и г. 106. Экспериментальная установка для записи голограмм трехмерных объектов.

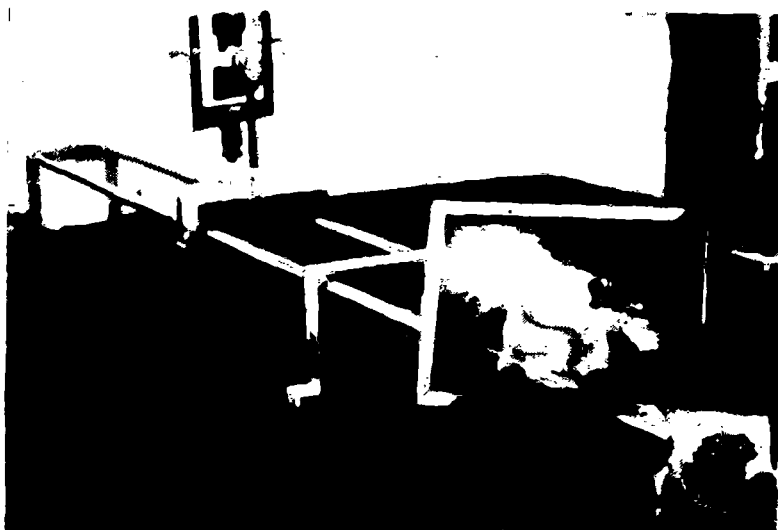
S — лазерный источник; *M*₁ — зеркало; *BS* — светоделитель; *M*₂ — зеркало; *O* — сцена; *R* — опорное зеркало; *P* — фотопластинка.



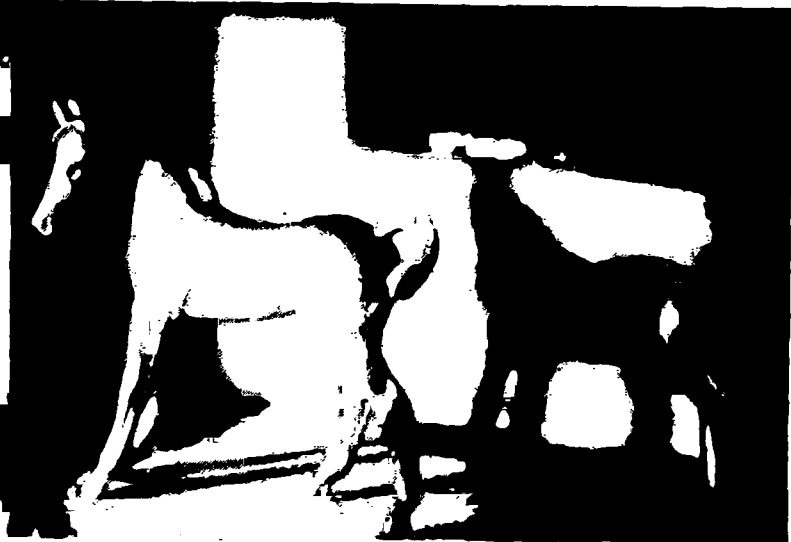
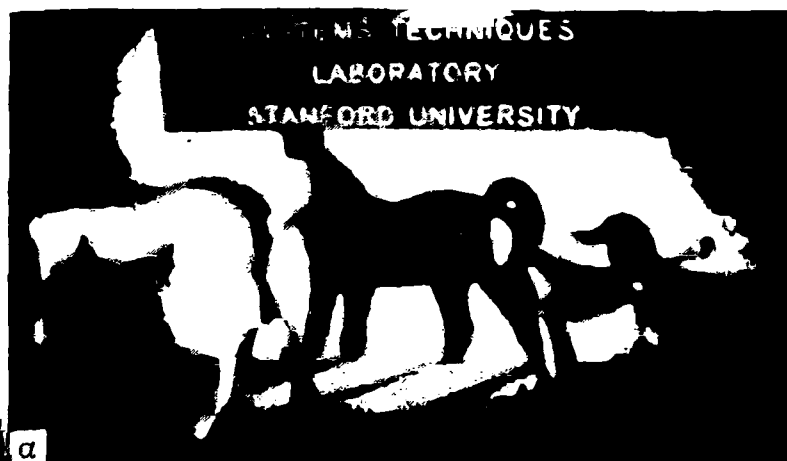
Ф и г. 107. Фотографический отпечаток самой голограммы.

небольшую часть голограммы. В этом случае глубина фокуса увеличивается, так что нужное двумерное изображение уже можно сфотографировать. Перемещая освещающий пучок по голограмме, мы будем получать двумерные изображения с различной перспективой. Таким образом, от каждого небольшого участка голограммы можно получить действительное изображение исходного объекта с различной перспективой!

На фиг. 106 показана установка, на которой записывались голограммы трехмерных объектов. Свет от лазера S с помощью зеркала M_1 направляется в светоделитель BS . Часть света, прошедшая через светоделитель, после отражения от зеркала M_2 попадает на трехмерную сценку O , которой в нашем случае служат две маленькие фигурки на фоне эмблемы. Другая часть света, отраженная светоделителем BS , попадает на опорное зеркало R , в котором здесь случайно видно отражение одной из фигур. Свет, отраженный опорным зеркалом, интерфери-



Фиг. 108. Экспериментальная установка для наблюдения многого изображения.



Ф и г. 109. Фотографии, иллюстрирующие трехмерный характер мнимого изображения.

рует на фотопластинке P со светом, приходящим от объекта.

На фиг. 107 приведена фотография проявленной голограммы. На ней незаметно никакой структуры, которую можно было бы отождествить с какой-либо частью объекта. В самом деле, можно сказать, что большая часть видимой структуры не имеет отношения к восстановлению, — она возникает за счет дефектов в оптической системе (например, из-за пылинок на зеркалах). Структура, позволяющая восстановить изображения, намного мельче той, которая видна на фотографии.

Фиг. 108 иллюстрирует метод получения мнимого изображения. В этом случае виден не сам лазер, а только зеркало M , отклоняющее его луч. За освещенной голограммой видно мнимое изображение фигурки лошади.

Для иллюстрации истинно трехмерного характера мнимого изображения обратимся к фиг. 109, на которой приведены две фотографии мнимого изображения. На фиг. 109, *a* камера сфокусирована на задний план мнимого изображения. Эмблема на заднем плане получилась резкой, в то время как изображение фигурки на переднем плане расфокусировано. Мы видим также, что хвост лошади закрывает тень от ее головы. Затем камера наводилась на передний план и сдвигалась для изменения перспективы. Полученное изображение показано на фиг. 109, *б*; теперь передний план сфокусирован, а задний расфокусирован. Хвост лошади уже не закрывает тень от головы, что является результатом изменения перспективы. Таким образом, нам удалось «заглянуть за» фигурку лошади, просто сместив фотоаппарат в сторону.

Экспериментальные трудности

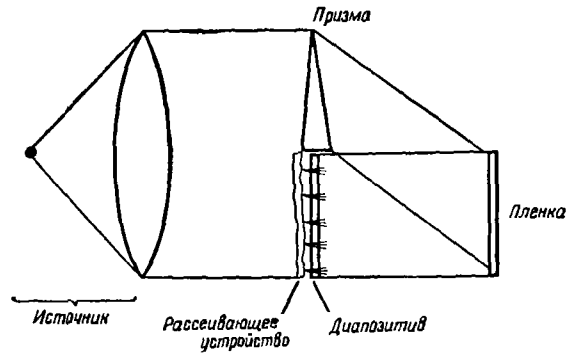
Несомненно, что с первых этапов развития голографии практически наиболее сложной была проблема источника с большой интенсивностью и высокой когерентностью. С появлением лазеров эта проблема существенно упростилась. Создание лазеров привело к тому, что голография нашла применение в новых и еще более сложных областях, например в описанной выше трехмерной фотографии. В большинстве случаев при приме-

нении голографии в новых областях исчерпываются все возможности сравнительно маломощных непрерывных лазеров, поэтому проблема источников света остается и в настоящее время главным препятствием для новых приложений голографии.

Процесс записи голограммы ненамного сложнее обычной интерференционной задачи. Как и в любом интерференционном эксперименте, если требуется записать ясные и четкие интерференционные полосы, то необходимо, чтобы положение оптического прибора было фиксировано в течение экспонирования с точностью до долей длины волны. Чем интенсивнее источник, тем меньше время экспонирования и тем менее жесткие требования предъявляются к стабильности. Необходимое в каждом случае время экспонирования зависит от множества факторов, в том числе от прозрачности или отражающей способности объекта, от применяемой схемы и расстояний и особенно от используемой пленки. На практике время экспонирования почти никогда не бывает меньше десятой доли секунды и больше нескольких часов. В некоторых случаях, когда в качестве источника можно использовать импульсный лазер, время экспонирования составляет несколько наносекунд.

Пожалуй, наиболее серьезные экспериментальные требования связаны с голографией трехмерных объектов. Для записи голограмм таких объектов необходима пленка с чрезвычайно большой разрешающей способностью (о требованиях к разрешающей способности пленки см. § 5). Самым употребительным фотоматериалом для такого рода задач являются спектроскопические пластинки типа Кодак 649F с разрешающей способностью более 2000 *линий/мм*. К сожалению, эмульсии, обладающие высоким разрешением, имеют чрезвычайно низкую чувствительность (в системе ASA чувствительность пластинок Кодак 649 F равна приблизительно 0,03, в то время как для пластинок Tg-X она равна 400). Из-за низкой чувствительности приходится увеличивать время экспонирования, что усиливает требования к стабильности.

В некоторых случаях важную роль играет еще одна практическая проблема — ограниченный динамический диапазон светочувствительного материала. Из фиг. 7С



Фиг. 110. Использование рассеянного света для записи голограмм.

видно, что линейное соотношение между амплитудным коэффициентом пропускания и величиной экспозиции выполняется только в ограниченном интервале. Конечно, желательно выбирать среднюю экспозицию так, чтобы она находилась вблизи средней точки линейного участка. Однако в том случае, когда объект представляет собой, например, диапозитив с довольно грубой структурой, на голограмме могут иметься заметные области, для которых экспозиция окажется много больше или много меньше значения, соответствующего линейному участку t — E -кривой. Из-за этой нелинейности можно ожидать некоторого искажения контраста изображения. Более подробно эффекты, связанные с нелинейностью, обсуждаются в § 6.

Проблему динамического диапазона можно в значительной мере решить с помощью метода, предложенного впервые Лейтом и Упатником [9]. Объект освещается через рассеивающее устройство, как показано на фиг. 110. Так как этот диффузор рассеивает свет в широком диапазоне направлений, то свет от каждого небольшого участка объекта попадает во все точки фотопластинки. Таким образом, в этом случае мы не будем иметь участков с очень большой и очень малой экспозициями, которые так или иначе получались бы при прямом освещении объекта. Нетрудно понять, что в этом случае мнимое изображение получается таким, как будто оно освещается сзади рассеянным светом.

§ 5. ВЛИЯНИЕ МПФ ПЛЕНКИ

До настоящего момента молчаливо предполагалось, что МПФ фотографической эмульсии, используемой для записи голограмм, одинакова для всех пространственных частот интерферограммы. В настоящем параграфе мы обсудим влияние частотного отклика пленки на качество восстановленных изображений. Мы рассмотрим две особенно важные схемы восстановления. Для дальнейшего изучения этого вопроса, включая анализ общих схем, читатель может обратиться к работам ван Литгена [15, 16].

Коллимированная опорная волна

Вначале рассмотрим, как влияет МПФ на преобразование амплитуд объекта в амплитуды действительного изображения для случая плоской опорной волны. Подобное же рассмотрение можно провести для любого действительного изображения, полученного из мнимого с помощью вспомогательной системы линз.

Как было отмечено раньше, если экспозиции лежат в пределах линейного участка $t - E$ -кривой, то преобразование амплитуд объекта в амплитуды изображения является линейным (см. § 2). Эта линейность не нарушается линейными процессами, которые ограничивают частотный отклик пленки. Поэтому достаточно найти влияние МПФ на одну частотную компоненту объекта; тогда общий результат получим просто как линейную суперпозицию.

Принимая это во внимание, положим, что опорная волна и волна от объекта имеют соответственно вид

$$\begin{aligned} U_r(x, y) &= A \exp(-j2\pi ay), \\ U_o(x, y) &= a \exp[-j2\pi(f_x x + f_y y)]. \end{aligned} \quad (8.47)$$

Таким образом, падающие на пленку волна от объекта и опорная волна являются плоскими и имеют различные направления распространения. Поэтому для распределения интенсивности при экспонировании получаем

$$\mathcal{I}(x, y) = A^2 + |a|^2 + 2A|a| \cos\{2\pi[f_x x + (f_y - \alpha)y] + \phi\}, \quad (8.48)$$

где $\varphi = \arg a$. Если представить частотный отклик эмульсии в виде $M(f_x, f_y)$, где

$$M(f_x, f_y) = M(f_x, f_y) \exp[-j\Omega(f_x, f_y)], \quad (8.49)$$

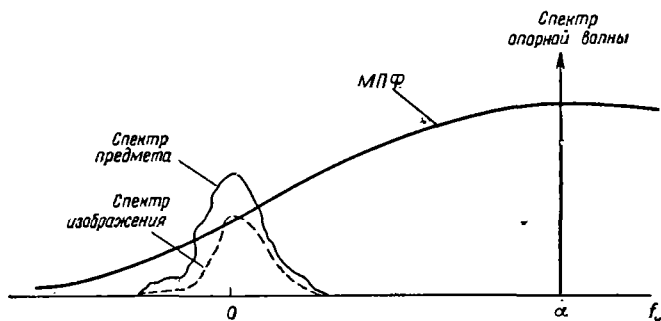
то в соответствии с (7.15) распределение *эффeктивной* интенсивности при экспозиции будет

$$\mathcal{U}_{\text{эфф}}(x, y) = A^2 + |a|^2 + 2A|a| M(f_x, f_y - \alpha) \times \\ \times \cos\{2\pi[f_x x + (f_y - \alpha)y] - \Omega(f_x, f_y - \alpha) + \varphi\}. \quad (8.50)$$

Отсюда заключаем, что соответствующая фурье-компонента объекта с пространственной частотой (f_x, f_y) *уменьшена* в $M(f_x, f_y - \alpha)$ раз и *сдвинута по фазе* на $\Omega(f_x, f_y - \alpha)$ по отношению к амплитуде и фазе, которые могли бы получиться в случае идеальной эмульсии (т. е. эмульсии, для которой $M = 1$ для всех частот).

Этот результат показан наглядно на фиг. 111. Пусть средняя частота спектра объекта лежит в точке $(f_x = 0, f_y = 0)$, а МПФ находится вблизи частоты $(f_x = 0, f_y = \alpha)$; тогда произведение спектра объекта и МПФ пленки дает спектр изображения.

Из приведенного выше рассмотрения ясно, что в том случае, когда используется коллимированная опорная волна, влияние частотного отклика пленки можно учесть просто умножением спектра объекта на $M(f_x, f_y - \alpha)$. Из этого результата непосредственно следует, что с про-



Фиг. 111. Влияние МПФ пленки для коллимированной опорной волны.

цессом получения изображения можно связать когерентную передаточную функцию

$$H(f_x, f_y) = M(f_x, f_y - \alpha).$$

Однако частотный отклик системы ограничен не только МПФ пленки, но и дифракционными эффектами, возникающими из-за ограниченности апертуры диапозитива. Таким образом, если $P(x, y)$ является функцией зрачка, т. е. если

$$P(x, y) = \begin{cases} 1 & (x, y) \text{ в пределах апертуры пленки,} \\ 0 & \text{вне пленки,} \end{cases}$$

то полную когерентную передаточную функцию можно записать в виде

$$H(f_x, f_y) = P(\lambda d_i f_x, \lambda d_i f_y) M(f_x, f_y - \alpha), \quad (8.51)$$

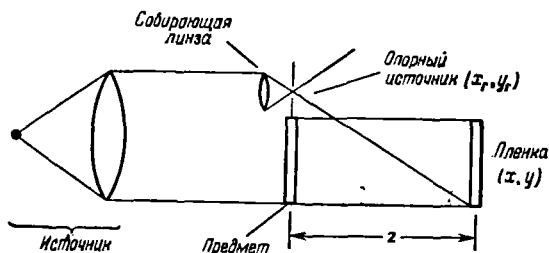
где первый множитель P введен непосредственно по аналогии с (6.19). Отметим, что влияние МПФ пленки можно считать эквивалентным наложению на входной зрачок системы получения изображения методом восстановления волнового фронта ослабляющей и сдвигающей фазу маски с амплитудным коэффициентом пропускания

$$t(x, y) = M\left(\frac{x}{\lambda d_i}, \frac{y}{\lambda d_i} - \alpha\right). \quad (8.52)$$

Безлинзовая фурье-голограмма

Рассмотрим еще одну возможную схему записи, в которой МПФ пленки сказывается существенно иначе, чем в предыдущем случае. Исследуемая здесь схема записи представлена на фиг. 112. Опорной волной теперь является сферическая волна, распространяющаяся из точки, компланарной объекту. По причинам, которые скоро станут понятны, голограмму, записанную по такой схеме, можно назвать *безлинзовой фурье-голограммой* [19].

Чтобы исследовать свойства голограмм этого типа, опять будем считать, что процесс получения изображений линейен, но на этот раз рассмотрим отклик формирующей изображение системы на объект не в виде плоской



Фиг. 112. Запись безлинзовой фурье-голограммы.

волны, а в виде одного точечного источника. Если (x_r, y_r) и (x_0, y_0) представляют собой соответственно координаты опорного точечного источника и точечного объекта, то распределение амплитуды на фотоэмульсии можно записать (во втором приближении) как

$$U_r(x, y) = A \exp \left\{ j \frac{\pi}{\lambda z} [(x - x_r)^2 + (y - y_r)^2] \right\},$$

$$U_0(x, y) = a \exp \left\{ j \frac{\pi}{\lambda z} [(x - x_0)^2 + (y - y_0)^2] \right\}.$$

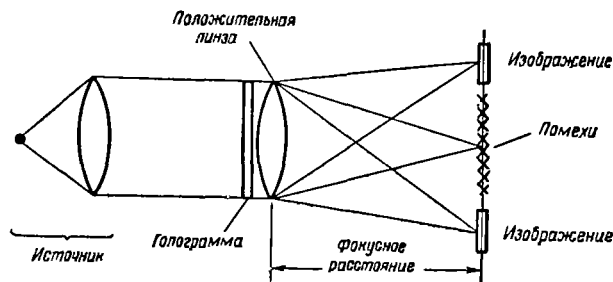
Таким образом, для интенсивности света при экспонировании имеем

$$\begin{aligned} \mathcal{I}(x, y) = & A^2 + |a|^2 + 2A|a| \cos \left[2\pi \frac{(x_0 - x_r)x}{\lambda z} + \right. \\ & \left. + 2\pi \frac{(y_0 - y_r)y}{\lambda z} + \theta(x_r, y_r; x_0, y_0) \right], \end{aligned} \quad (8.53)$$

где $\theta(x_r, y_r; x_0, y_0)$ — фазовый угол, зависящий от (x_r, y_r) и (x_0, y_0) , но не от координат пленки (x, y) .

Теперь становится понятным смысл названия *безлинзовая фурье-голограмма*. Волна, возникающая в точке объекта с координатами (x_0, y_0) , интерферируя с опорной волной, дает в результате распределение синусоидальных полос с пространственными частотами

$$\begin{aligned} f_x &= \frac{x_0 - x_r}{\lambda z}, \\ f_y &= \frac{y_0 - y_r}{\lambda z}. \end{aligned} \quad (8.54)$$



Фиг. 113. Получение изображений от безлинзовой фурье-голограммы.

Таким образом, при этой схеме записи существует однозначное соответствие между координатами объекта и пространственными частотами на голограмме. Такое соответствие характерно для операции фурье-преобразования, но получено оно без применения преобразующих линз.

Чтобы получить от такой голограммы изображение, ее освещают и устанавливают вслед за ней положительную линзу, как показано на фиг. 113. Из соотношения (8.41) при $z_p = \infty$ и $z_0 = z_r$ получаем, что оба сопряженных изображения, создаваемых самой голограммой, лежат на бесконечном расстоянии от диапозитива. Положительная линза переносит эти изображения в действительный фокус на расстоянии f от линзы¹⁾.

Чтобы определить, как влияет МПФ пленки на восстановленные изображения, объединим соотношения (7.15) и (8.53); в результате получим распределение эффективной интенсивности:

$$\begin{aligned}
 I_{\text{эфф}}(x, y) = & A^2 + |a|^2 + 2A|a| M \left(\frac{x_0 - x_r}{\lambda z}, \frac{y_0 - y_r}{\lambda z} \right) \times \\
 & \times \cos \left[2\pi \frac{x_0 - x_r}{\lambda z} x + 2\pi \frac{y_0 - y_r}{\lambda z} y + \right. \\
 & \left. + \theta(x_r, y_r; x_0, y_0) - \Omega \left(\frac{x_0 - x_r}{\lambda z}, \frac{y_0 - y_r}{\lambda z} \right) \right]. \quad (8.55)
 \end{aligned}$$

¹⁾ Восстановленные изображения можно получить с помощью точечного источника и одной линзы. Оба изображения будут находиться в плоскости, в которой линза изображает точечный источник. Положение голограммы относительно линзы при этом безразлично — от него зависят только размеры изображений. — Прим. ред.

Сдвиг фазы Ω , возникающий за счет МПФ пленки, влияет только на фазу света в точке изображения, соответствующей (x_0, y_0) , и поэтому мы можем его не учитывать. Однако множитель M соответствует ослаблению интенсивности света, дающего изображение точки (x_0, y_0) . Так как точки объекта, наиболее удаленные от опорной точки, дают на голограмме полосы с наивысшими пространственными частотами, интенсивность изображений таких точек ослабляется в наибольшей степени.

Таким образом, в схеме записи безлинзовой фурье-голограммы ограниченность частотного отклика пленки приводит к сужению поля зрения, но не влияет на разрешение в пределах этого поля зрения. Если, грубо говоря, максимальная пространственная частота, которая разрешается эмульсией, равна $f_{\text{макс}}$, то в изображении появляются только те точки объекта, координаты которых удовлетворяют соотношению

$$\sqrt{(x_0 - x_r)^2 + (y_0 - y_r)^2} \leq \lambda z f_{\text{макс}}. \quad (8.56)$$

Если для коллимированной опорной волны, как мы уже убедились, ограничение разрешающей способности пленки эквивалентно наложению маски на входной зрачок [см. (8.52)], то для второй схемы записи оно эквивалентно наложению маски на сам объект. В этом случае пропускание эквивалентной маски описывается соотношением

$$t(x_0, y_0) = M \left(\frac{x_0 - x_r}{\lambda z}, \frac{y_0 - y_r}{\lambda z} \right). \quad (8.57)$$

Так как амплитуда волны от каждой точки объекта умножается на коэффициент, который зависит от координат этой точки, рассмотренная схема не является пространственно-инвариантной и, строго говоря, для нее не существует когерентной передаточной функции. Однако с практической точки зрения можно разделить предметное пространство на так называемые *изопланатические области*, размеры которых достаточно малы, чтобы соотношение (8.57) было приблизительно верным в каждой такой области. Тогда каждой области можно поставить в соответствие свою когерентную передаточную функцию. В этом случае когерентную передаточную функцию мо-

жно записать в виде

$$H(f_x, f_y) \approx MP(\lambda d_{ifx}, \lambda d_{ify}), \quad (8.58)$$

где M определяется соотношением (8.57) для данной изопланатической области, а P опять есть функция зрачка, описывающая эффективный зрачок системы.

Более общие схемы записи

Полный анализ показанной на фиг. 104 более общей схемы записи был проведен ван Лигтенем [16], но он слишком громоздок, поэтому привести его здесь не представляется возможным. Анализ показывает, что влияние МПФ пленки во всех случаях эквивалентно действию маски, помещенной в некоторую плоскость между объектом и пленкой и имеющей амплитудный коэффициент пропускания, пропорциональный, как и ранее, множителю M . Положение эквивалентной маски зависит от схемы записи. Две схемы записи, рассмотренные подробно в настоящем параграфе, являются предельными случаями, в одном из которых эффективная маска совмещается с входным зрачком, а в другом — с объектом.

Для дальнейшего изучения эффектов, связанных с МПФ пленки, читатель может обратиться к цитированным работам; дополнительный материал приведен также в задаче 8.5.

§ 6. ЭФФЕКТЫ, СВЯЗАННЫЕ С НЕЛИНЕЙНОСТЬЮ ПЛЕНКИ

Во всех наших исследованиях голографического метода мы неоднократно делали предположение, что при экспонировании светочувствительного слоя рабочая точка лежит в пределах линейного участка кривой зависимости амплитудного коэффициента пропускания от экспозиции. Однако реальные фотопленки и пластинки никогда не бывают идеально линейными. Отклонение от линейности в значительной мере зависит от *диапазона экспозиций*, которым подвергается светочувствительный слой. В настоящем параграфе мы кратко рассмотрим, как сказывается нелинейность пленки на качестве восстановленного изображения. Следует подчеркнуть, что

в том случае, когда средняя экспозиция, обусловленная объектом, сравнима с экспозицией за счет опорной волны, возможно, именно нелинейные эффекты сильнее всего ограничивают качество изображения. Этот случай можно противопоставить случаю слабо освещенного объекта, где основным ограничивающим фактором является шум, связанный с зернистостью пленки [18].

Модель Косма

Косма предложил удобную модель для анализа влияния нелинейности пленки на голографически восстановленный волновой фронт [19]. Его подход к рассмотрению фотографической эмульсии во многом аналогичен подходу, используемому при анализе нелинейного элемента в теории связи [20, 21]. Принимается, что входным сигналом для такого элемента является распределение экспозиций, которым подвергается эмульсия, а выходным — распределение амплитудного коэффициента пропускания проявленного диапозитива.

Для простоты будем считать опорную волну плоской и запишем распределение ее поля и поля объекта соответственно в виде

$$\begin{aligned} U_r(x, y) &= A \exp(-j2\pi ay), \\ U_o(x, y) &= a(x, y) \exp[-j\varphi(x, y)]. \end{aligned} \quad (8.59)$$

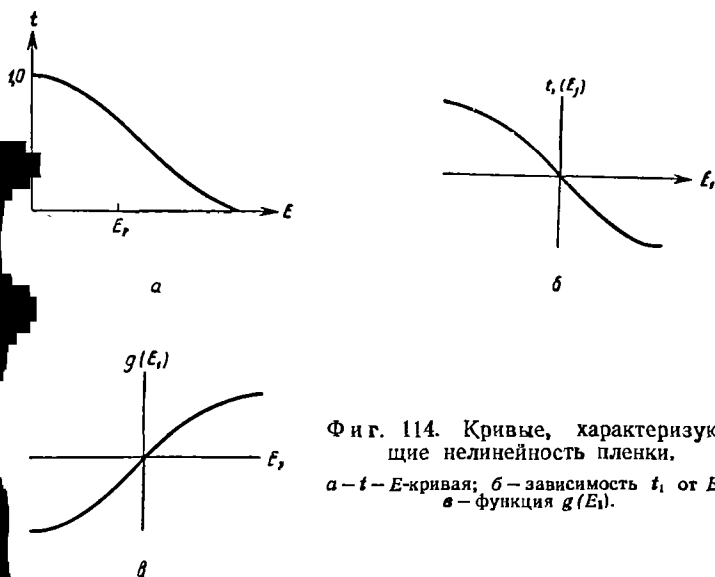
Тогда суммарную экспозицию можно записать следующим образом:

$$\begin{aligned} E(x, y) &= |U_r(x, y) + U_o(x, y)|^2 T = \\ &= A^2 T + a^2(x, y) T + 2Aa(x, y) T \cos[2\pi ay - \varphi(x, y)], \end{aligned} \quad (8.60)$$

где T — время экспонирования. Мы можем считать, что это распределение экспозиции равно сумме сравнительно постоянной экспозиции $E_r = A^2 T$, обусловленной опорной волной, плюс отклонения $E_1(x, y)$ от E_r , где

$$\begin{aligned} E_1(x, y) &= \\ &= a^2(x, y) T + 2Aa(x, y) T \cos[2\pi ay - \varphi(x, y)]. \end{aligned} \quad (8.61)$$

Амплитудный коэффициент пропускания $t(x, y)$ проявленного диапозитива можно представить в виде суммы



Фиг. 114. Кривые, характеризующие нелинейность пленки.

a — t — E -кривая; *б* — зависимость t_1 от E_1 ; *в* — функция $g(E_1)$.

сравнительно постоянного «главного» пропускания t_b , которое полностью определяется опорной волной, и изменения пропускания $t_1(x, y)$. Таким образом,

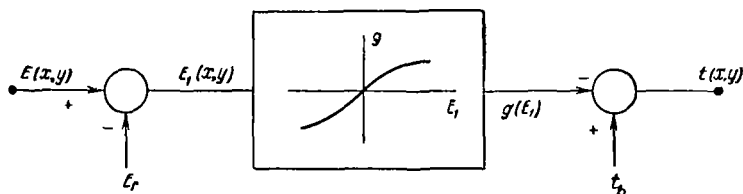
$$t(x, y) = t_b + t_1(x, y). \quad (8.62)$$

Именно зависимость t_1 от E_1 [т. е. функцией $t_1(E_1)$] и определяется линейность процесса. Для дальнейшего удобнее записывать эту функциональную зависимость как $g(E_1)$, где

$$g(E_1) = -t_1(E_1). \quad (8.63)$$

На фиг. 114 приведен типичный вид зависимостей t от E , затем t_1 от E_1 и окончательно g от E_1 .

Блок-схема модели Косма дана на фиг. 115. Постоянная экспозиция E_r вычитается из $E(x, y)$; оставшаяся экспозиция $E_1(x, y)$ подается на вход нелинейного элемента с характеристикой $g(E_1)$. Полное амплитудное пропускание $t(x, y)$ получается вычитанием выходного сигнала из



Ф и г. 115. Блок-схема модели Косма.

постоянного пропускания. Теперь остается, конечно, выяснить, как влияет нелинейное преобразование $g(E_1)$ на поле, проходящее через диапозитив.

Анализ методом преобразования

Один из возможных методов анализа модели Косма — это так называемый *метод преобразования*, используемый в теории связи [20, 21]. Этот метод основан на чисто математических операциях. Рассмотрим фурье-образ нелинейной характеристики $g(E_1)$:

$$G(f) = \int_{-\infty}^{+\infty} g(E_1) \exp(-j2\pi f E_1) dE_1. \quad (8.64)$$

Написанный фурье-образ не имеет реального физического смысла, так как он не представляет собой «спектр» временной или пространственной функции в обычном смысле слова. Однако в его полезности можно убедиться, выразив функцию $g(E_1)$ через ее фурье-образ $G(f)$:

$$\begin{aligned} g(E_1) &= \int_{-\infty}^{+\infty} G(f) \exp(j2\pi f E_1) df = \\ &= \int_{-\infty}^{+\infty} G(f) \exp[j2\pi f a^2(x, y) T] \times \\ &\quad \times \exp\{j4\pi f A a(x, y) T \cos[2\pi a y - \varphi(x, y)]\} df. \end{aligned}$$

Второй экспоненциальный множитель может быть раскрыт с помощью формулы Якоби — Ангера [22]

$$\exp(j\gamma \cos \theta) = \sum_{\nu=0}^{\infty} \epsilon_{\nu} j^{\nu} J_{\nu}(\gamma) \cos \nu \theta, \quad (8.65)$$

где ϵ_{ν} — множитель Неймана

$$\epsilon_{\nu} = \begin{cases} 1 & \nu = 0, \\ 2 & \nu > 0, \end{cases} \quad (8.66)$$

а J_{ν} — бесселева функция первого рода порядка ν . Таким образом

$$g(E_1) = \sum_{\nu=0}^{\infty} \epsilon_{\nu} j^{\nu} \left[\int_{-\infty}^{+\infty} G(f) \exp(j2\pi f a^2 T) J_{\nu}(4\pi T A a f) df \right] \times \\ \times \cos(2\pi \nu a y - \nu \varphi)$$

и окончательно

$$t(x, y) = t_0 - \sum_{\nu=0}^{\infty} H_{\nu}(a(x, y)) \cos[2\pi \nu a y - \nu \varphi(x, y)], \quad (8.67)$$

где

$$H_{\nu}(a(x, y)) = \epsilon_{\nu} j^{\nu} \int_{-\infty}^{+\infty} G(f) \exp(j2\pi f a^2 T) J_{\nu}(4\pi T A a f) df. \quad (8.68)$$

Запись результата в виде (8.67) очень наглядна. Член $\nu = 1$ представляет собой компоненту, благодаря которой получаются обычные действительное и мнимое изображения. Фазовая модуляция $\varphi(x, y)$ этой компоненты идентична фазовой модуляции распределения экспозиций, и, следовательно, не искажается нелинейностью. Однако амплитудная модуляция искажается в соответствии с нелинейностью функции $H_1(a(x, y))$, которую в принципе можно оценить для любой рассматриваемой нелинейной функции $g(E_1)$.

Кроме того, генерируются гармониками пространственной несущей частоты α . Несущая частота ν -й гармоники равна $\nu \alpha$. Отсюда следует, что волновые компоненты, соответствующие этим гармоникам, распространяются всегда под большими углами, чем недифрагированный

свет. Фазовая модуляция ν -й гармоники равна фазовой модуляции падающей волны, умноженной на ν , а амплитудная модуляция есть нелинейная функция $H_\nu(a(x, y))$ исходной амплитудной модуляции $a(x, y)$. Можно считать, что эти гармоники дают изображения «высших порядков» исходного объекта, хотя такие изображения будут, вообще говоря, сильно искажены из-за вносимого диапозитивом нарушения амплитудного и фазового распределений соответствующей волновой компоненты.

После того как получено общее выражение для пропускания нелинейного диапозитива, остаются еще два нерешенных вопроса, которые мы пока не затрагивали: 1) оценка функции $H_\nu(a)$ для некоторой разумной модели нелинейности пленки $g(E_1)$ и 2) определение влияния этих искажений не только на прошедшие волновые компоненты, но и на сами *изображения*. Мы не будем рассматривать здесь подробно ни одну из этих проблем. Однако некоторые частные случаи будут рассматриваться в дальнейшем при обсуждении свойств пленки с «резким контрастом». Кроме того, читатель может обратиться к работе Косма [19], в которой для нахождения функций H_ν используется так называемая модель *характеристики в виде функции ошибок*.

Прямой метод анализа. Резкий контраст

Анализ с помощью метода преобразования привлекателен благодаря своей общности, но в некоторых случаях и с помощью более прямого подхода удастся сравнительно легко получить простые результаты. Рассмотрим пример такого непосредственного подхода для случая *резкоконтрастной пленки*. Так мы будем называть пленку с зависимостью пропускания от экспозиции типа показанной на фиг. 116. Для всех экспозиций ниже некоторой пороговой E_0 амплитуда пропускания равна единице, а для всех экспозиций выше E_0 она равна нулю. Такая модель в лучшем случае является грубым приближением даже для пленки с очень большой контрастностью. Вообще говоря, она скорее соответствует нелинейной характеристике, с которой мы сталкиваемся в том случае,

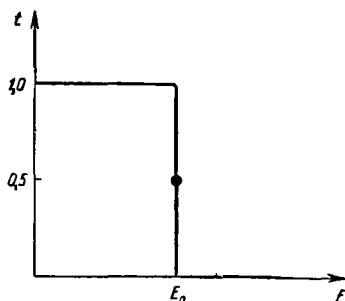
когда голограмма сканируется, а затем обрабатывается с помощью аналого-цифровой машины.

Чтобы проанализировать эту модель, сделаем два основных допущения. Во-первых, положим, что «несущая частота» α в соотношении (8.60) много больше ширины полос пространственных частот амплитуды $a(x, y)$ и фазы $\varphi(x, y)$. Если это так, то в любой области, малой по сравнению с обратными ширинами полос частот

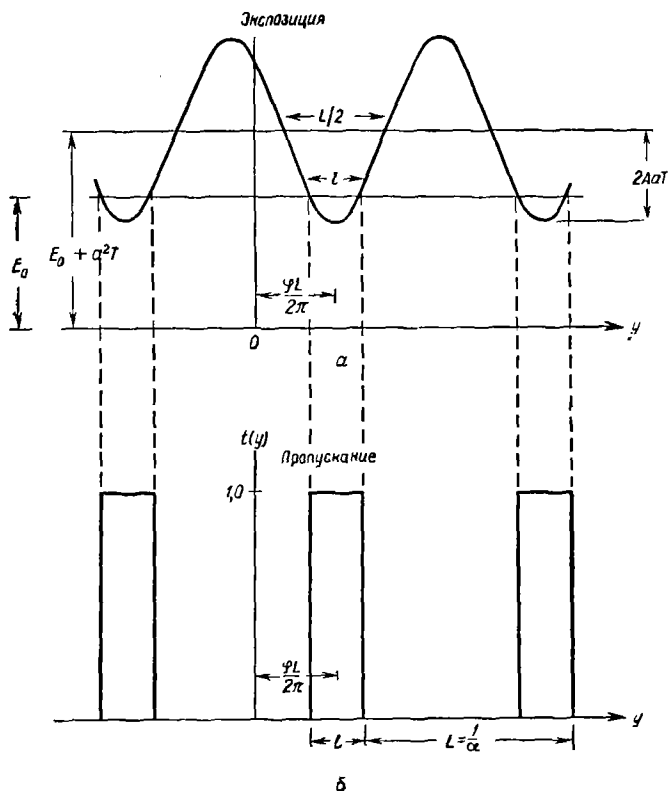
амплитуды $a(x, y)$ и фазы $\varphi(x, y)$, экспозицию можно рассматривать как сумму приблизительно постоянного смещения $E_r + a^2T$ и синусоидальных полос пространственной частоты α с почти постоянной амплитудой $2AaT$ и фазой φ . Во-вторых, допустим, что экспозиция опорной волны E_r выбрана так, чтобы она была в точности равна пороговой экспозиции E_0 . Тогда преобразование изменений экспозиции в изменения пропускания можно записать в виде

$$t = \begin{cases} 1 & \text{при } a^2T + 2AaT \cos(2\pi\alpha y - \varphi) < 0, \\ 1/2 & \text{при } a^2T + 2AaT \cos(2\pi\alpha y - \varphi) = 0, \\ 0 & \text{в остальных случаях.} \end{cases}$$

Фиг. 117 является наглядной иллюстрацией нелинейного действия пленки; с ее помощью можно определить важные величины, которыми мы будем оперировать в дальнейшем. Обозначим период функции пропускания $t(y)$ через $L = \alpha^{-1}$, а ширину импульсов пропускания с единичной амплитудой через l . Центр импульса смещен относительно начала координат по оси y на величину $\varphi L/2\pi$. Заметим, что информация об амплитудной модуляции $a(x, y)$ интерференционных полос падающего на пленку света не утрачивается при резком контрасте, так



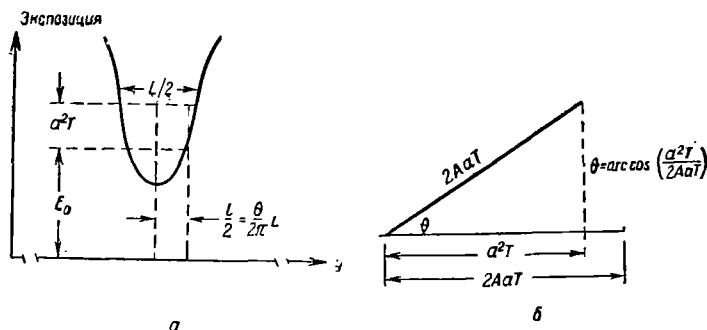
Фиг. 116. Кривая зависимости амплитудного коэффициента пропускания от экспозиции для идеальной резко-контрастной пленки.



Фиг. 117. Нелинейное преобразование изменений экспозиции (а) в изменения пропускания (б).

как от нее зависит ширина l импульсов пропускания. Таким образом, информация об амплитуде сохраняется благодаря модуляции ширины импульсов, но, конечно, в искаженном виде. Подобным образом фазовая модуляция $\varphi(x, y)$ влияет на положение импульсов пропускания относительно начала координат. Следовательно, информация о фазе сохраняется в виде модуляции положения импульсов.

Прежде чем рассматривать влияние нелинейности на различные волновые компоненты, прошедшие через голо-

Фиг. 118. Определение ширины импульса l .

грамму, мы найдем точное соотношение между шириной импульса l и амплитудой a волны, падающей на пленку от объекта. Это соотношение можно установить с помощью фиг. 118. Прежде всего заметим, что фазовый угол θ можно определить по фиг. 118, а; он равен

$$\theta = \frac{2\pi}{L} \frac{l}{2}. \quad (8.69)$$

Кроме того, из фазовой диаграммы на фиг. 118, б следует, что

$$\theta = \arccos\left(\frac{a^2\tau}{2Aa\tau}\right) = \arccos\left(\frac{a}{2A}\right). \quad (8.70)$$

Комбинируя (8.69) и (8.70), получаем

$$l = \frac{L}{\pi} \arccos\left(\frac{a}{2A}\right). \quad (8.71)$$

Этот результат иллюстрирует нелинейный характер преобразования амплитудной модуляции $a(x, y)$ в модуляцию ширины импульса $l(x, y)$.

Найдем теперь амплитудную и фазовую модуляции различных волновых компонент, прошедших через голограмму. Для этого воспользуемся предположением об «узкополосности» и применим анализ Фурье для исследования пропускания $t(x, y)$, как будто оно является чисто периодической функцией y и не зависит от x . Затем примем, что a и φ в решении представляют собой

медленно меняющиеся функции. Такое приближение является достаточно точным, если выполняется первое из сделанных выше допущений. Периодическую функцию $t(x, y) = t(y)$ можно разложить в ряд Фурье

$$t(y) = \sum_{\nu=-\infty}^{\infty} C_{\nu} \exp\left(j \frac{2\pi\nu y}{L}\right), \quad (8.72)$$

где

$$C_{\nu} = \frac{1}{L} \int_{-L/2}^{L/2} t(y) \exp\left(-j \frac{2\pi\nu y}{L}\right) dy. \quad (8.73)$$

После замены переменных $y' = y - (\varphi L/2\pi)$ середина импульса пропускания будет совпадать с началом системы координат y' ; тогда

$$C_{\nu} = \frac{\exp(-j\nu\varphi)}{L} \int_{-l/2}^{l/2} \exp\left(j \frac{2\pi\nu y'}{L}\right) dy'. \quad (8.74)$$

После интегрирования получим

$$C_{\nu} = \begin{cases} \frac{l}{L} & \nu = 0, \\ \frac{1}{\pi\nu} \sin\left(\pi\nu \frac{l}{L}\right) \exp(-j\nu\varphi) & \nu \neq 0. \end{cases} \quad (8.75)$$

Комбинируя этот результат с (8.71) и (8.72), находим¹⁾

$$t(x, y) = \frac{1}{\pi} \arccos\left(\frac{a}{2A}\right) + \sum_{\nu=1}^{\infty} \frac{2}{\pi\nu} \sin\left[\nu \arccos\left(\frac{a}{2A}\right)\right] \cos(2\pi\nu ay - \nu\varphi), \quad (8.76)$$

где теперь будем считать, что a и φ представляют собой медленно меняющиеся функции (x, y) . Если сравнить этот результат с выражением (8.67), полученным ранее методом преобразования, то совершенно ясно, что функ-

¹⁾ На простоту этого результата и прямой метод, с помощью которого можно его получить, автору впервые указал его ученик Гордон Найт.

ция H_ν в случае резкого контраста равна

$$H_\nu(a) = \begin{cases} \frac{1}{2} - \frac{1}{\pi} \arccos\left(\frac{a}{2A}\right) & \nu = 0, \\ -\frac{2}{\pi\nu} \sin\left[\nu \arccos\left(\frac{a}{2A}\right)\right] & \nu > 0. \end{cases} \quad (8.77)$$

Нас в первую очередь интересует фурье-компонента, ответственная за образование изображений первого порядка, т. е. компонента с $\nu = 1$. Поскольку

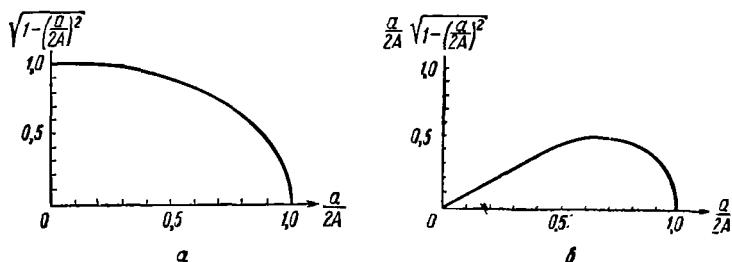
$$\sin\left[\arccos\left(\frac{a}{2A}\right)\right] = \sqrt{1 - \left(\frac{a}{2A}\right)^2}, \quad (8.78)$$

нетрудно найти, что соответствующая компонента пропускания есть

$$t_1(x, y) = \frac{2}{\pi} \sqrt{1 - \left[\frac{a(x, y)}{2A}\right]^2} \cos[2\pi a y - \varphi(x, y)]. \quad (8.79)$$

Таким образом, амплитудная модуляция несущей искажена, как это показано на фиг. 119, в то время как фазовая модуляция остается неискаженной. Заметим, что при $a \ll 2A$ амплитудная модуляция исчезает и пленка действует совершенно так же, как *ограничитель полосы пропускания* в теории связи (см. работу [20], стр. 288). Однако при значениях a , сравнимых с $2A$, изменение смещения a^2T , вызванное волной от объекта, приводит к искажению амплитудной модуляции пропускания t_1 .

В заключение интересно рассмотреть компоненту пропускания, ответственную за возникновение изобра-



Фиг. 119. Искажение амплитудной модуляции компоненты первого порядка (а) и компоненты второго порядка (б).

жений второго порядка:

$$t_2(x, y) = \frac{1}{\pi} \sin \left[2 \arccos \left(\frac{a}{2A} \right) \right] \cos [2\pi(2\alpha)y - 2\varphi]. \quad (8.80)$$

Используя соотношение

$$\begin{aligned} \sin \left[2 \arccos \left(\frac{a}{2A} \right) \right] &= 2 \sin \left[\arccos \left(\frac{a}{2A} \right) \right] \cos \left[\arccos \left(\frac{a}{2A} \right) \right] = \\ &= \frac{a}{A} \sqrt{1 - \left(\frac{a}{2A} \right)^2}, \end{aligned} \quad (8.81)$$

находим

$$t_2(x, y) = \frac{1}{\pi} \frac{a}{A} \sqrt{1 - \left(\frac{a}{2A} \right)^2} \cos [2\pi(2\alpha)y - 2\varphi]. \quad (8.82)$$

Искажение амплитудной модуляции этих волновых компонент показано на фиг. 119, б. Заметим, что для малых $(a/2A)$ амплитудная модуляция не искажена. Однако фазовая модуляция вдвое больше, чем это должно было бы быть; в результате можно ожидать появления искажений в изображении второго порядка. Для ознакомления с методом, позволяющим найти правильную частотную и фазовую модуляцию при $a \ll 2A$, читатель может обратиться к задаче 8.7.

Влияние нелинейности на восстановленные изображения

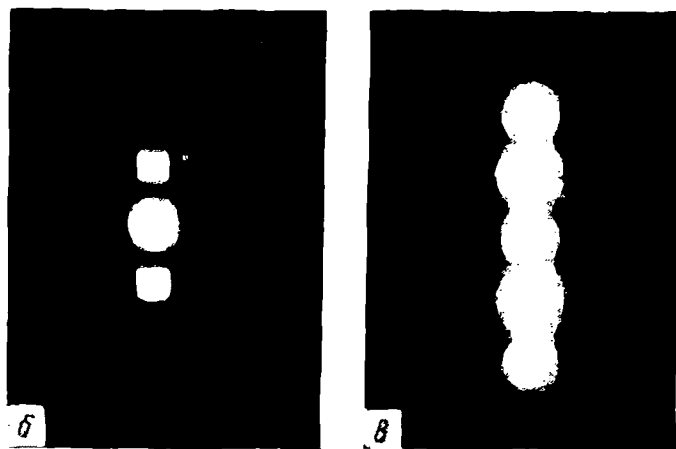
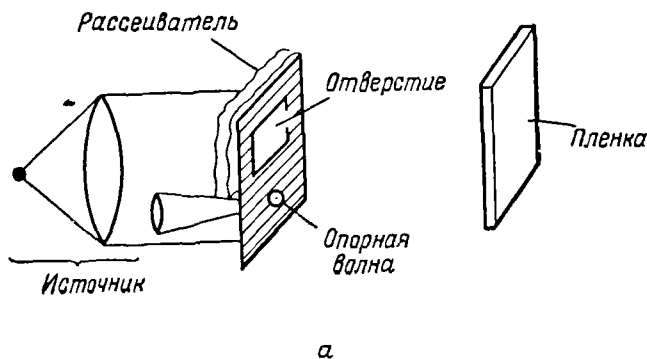
Определение влияния нелинейности на сами восстановленные изображения является, вообще говоря, трудной задачей, и мы не будем ее подробно рассматривать. Однако уместно сделать несколько простых замечаний.

Прежде всего, чтобы предсказать влияние нелинейности на изображение первого порядка для резко контрастной пленки, экспонированной так, что $a \ll 2A$, можно воспользоваться многими известными из теории связи свойствами ограничителя полосы пропускания [23]. Такой подход применили Фризем и Зеленка [24], которые установили как аналитически, так и экспериментально несколько важных закономерностей. Если объект состоит из двух точечных источников, причем амплитуда одного из них больше, чем другого, то должен наблюдаться (и это соответствует действительности) эффект подавления слабого сигнала. Иными словами, амплитуда

изображения более слабого точечного источника ослаблена по сравнению с амплитудой более интенсивного источника. Кроме того, из-за эффектов взаимной модуляции при взаимодействии двух точечных источников могут появиться ложные изображения, т. е. изображения точечных источников, которых на самом деле нет в объекте.

Влияние нелинейности пленки исследовалось также и для диффузного освещения объекта [25, 26]. В этом случае экспозицию лучше всего рассматривать как случайный процесс, используя методы анализа, которые выходят за рамки нашего рассмотрения. Со всеми деталями читатель может ознакомиться по цитированным работам.

В заключение приведем фотографии, иллюстрирующие природу искажений изображения, с которыми сталкиваются на практике. На фиг. 120, *a* приведена схема установки, используемой для записи интересующих нас сейчас голограмм. Объект представляет собой просто квадратное отверстие в непрозрачном экране, освещаемое сзади через рассеивающее устройство. Опорная волна излучается из точки, компланарной объекту; таким образом, здесь используется схема безлинзовой голографии Фурье. Голограмма записывается на спектроскопическую пластинку Кодак 649F. Отношение интенсивностей опорной волны и волны от объекта в плоскости записи во всех случаях равно 1,4:1. На фиг. 120, *b* мы видим два изображения первого порядка (плюс член нулевого порядка), полученные в том случае, когда средняя экспозиция была выбрана так, что постоянная компонента пропускания равна 0,5. Эта величина соответствует максимальному линейному динамическому диапазону на $t-E$ -кривой, поэтому нелинейные эффекты в данном случае очень малы. Изображения на фиг. 120, *в* получены с большим временем экспозиции, которое выбрано так, что постоянная компонента пропускания равна 0,24 при неизменных других факторах (включая отношение интенсивностей опорной волны и волны от объекта). В этом случае мы попадаем в область высокой нелинейности $t-E$ -кривой. Результаты наглядно видны. Прежде всего легко заметить присутствие членов второго



Фиг. 120. Влияние нелинейности пленки на восстановление изображения.

а — схема записи; б — изображения, полученные при $t_b = 0,5$, и в — изображения, полученные при $t_b = 0,21$.

порядка. Однако более существенно то, что четкие изображения первого порядка наложены теперь на размытый светлый фон. Этот фон есть непосредственное следствие вызванного нелинейностью пленки искажения за счет взаимной модуляции [26].

§ 7. ВЛИЯНИЕ ТОЛЩИНЫ ЭМУЛЬСИИ

Во всем предыдущем изложении предполагалось, что фотографическая эмульсия имеет пренебрежимо малую толщину и что запись голограммы можно рассматривать как чисто «поверхностное» явление. В действительности эмульсия всегда имеет какую-то отличную от нуля толщину, и поэтому важно знать, как эта толщина может влиять на процесс получения изображений.

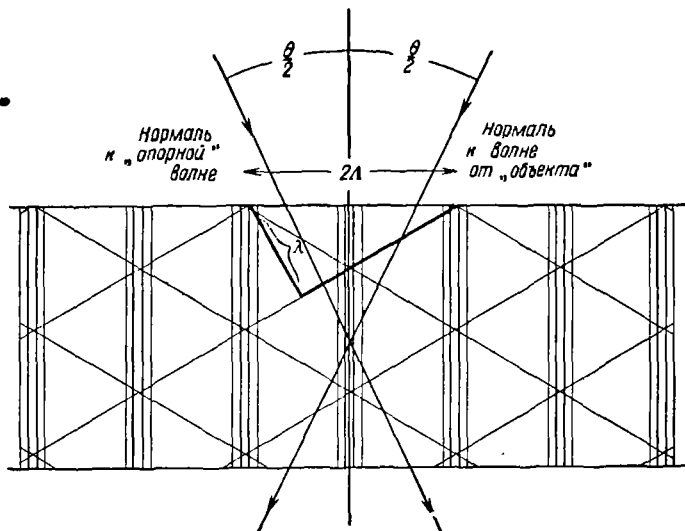
В качестве типичного примера укажем, что у широко используемых в голографии спектроскопических пластинок Кодак 649F толщина эмульсии составляет обычно около 16 мк. На первый взгляд может показаться, что это совершенно ничтожная величина; однако следует помнить, что период полос на поверхности эмульсии часто бывает сравним с длиной оптической волны, которая для обычно используемой красной линии гелий-неонового лазера равна приблизительно 0,6 мк. Таким образом, толщина эмульсии может иногда в 20 раз превышать период полос. В подобных случаях необходимо учитывать «объемный» характер процесса записи.

При рассмотрении этого вопроса мы будем в основном следовать работе Колиера [10], в которой читатель может найти дополнительные сведения.

Элементарная голограмма

Рассмотрим сначала очень простой случай: на эмульсию конечной толщины падают плоская опорная волна и плоская волна от объекта. Можно считать, что эти две простые волны дают элементарную голограмму.

Обращаясь к фиг. 121, положим для простоты, что волновые нормали этих двух волн (изображенные стрелками) составляют угол $\theta/2$ с нормалью к поверхности; тогда суммарный угол между ними будет θ . Волновые фронты, или последовательность линий нулевых фаз, показаны тонкими линиями. Расстояние между волновыми фронтами для каждой из волн равно длине оптической волны. Вдоль линий в эмульсии (т. е. в точках на нашем двумерном рисунке), где пересекаются волновые фронты двух волн, их амплитуды складываются в фазе, давая большую экспозицию. С течением времени волновые



Фиг. 121. Запись элементарной голограммы в толстой эмульсии.

фронты смещаются в направлениях своих нормалей, и линии конструктивной интерференции перемещаются по эмульсии, образуя *плоскости* больших экспозиций. Простое геометрическое построение показывает, что эти плоскости делят пополам угол θ между двумя волновыми нормальями и располагаются периодически в эмульсии.

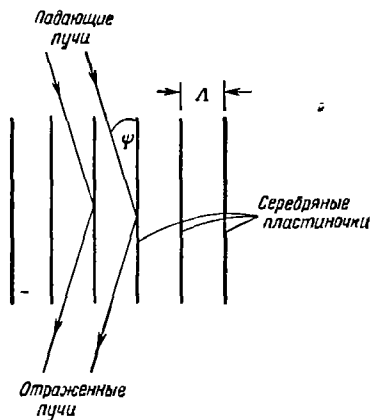
После проявления фотопластины атомы серебра будут образовывать *пластиночки* вдоль плоскостей большой экспозиции. Рассматривая прямоугольный треугольник, выделенный жирными линиями на фиг. 121, легко показать, что расстояние L между серебряными пластинками удовлетворяет соотношению

$$2L \sin \frac{\theta}{2} = \lambda. \quad (8.83)$$

Условие Брэгга

Положим, что мы пытаемся восстановить плоскую волну исходного *объекта*, освещая элементарную голограмму *восстанавливающей* плоской волной. Естественно

возникает вопрос: какой угол освещения надо выбрать, чтобы получить максимально интенсивную восстановленную волну? Чтобы ответить на этот вопрос, мы можем рассматривать каждую серебряную пластиночку как частично отражающее зеркало, которое отклоняет часть падающей плоской волны в соответствии с обычными законами отражения. Если плоская волна падает на пластинчатую структуру под углом ψ , как показано на фиг. 122, то отраженные плоские волны будут распро-



Фиг. 122. Схема восстановления.

страняться в указанном направлении. Однако для того, чтобы различные отраженные волны складывались *в фазе*, длины путей, проходимых волнами, отраженными от соседних пластинок, должны отличаться ровно на одну оптическую длину волны¹⁾. Обращаясь к фиг. 122, с помощью простого геометрического построения нетрудно показать, что это требование выполняется только в том случае, когда угол падения ψ удовлетворяет так называемому *условию Брэгга*

$$\sin \psi = \pm \frac{\lambda}{2L}. \quad (8.84)$$

Сравнение соотношений (8.83) и (8.84) показывает, что интенсивность будет максимальной при

$$\psi = \begin{cases} \pm \theta/2 \text{ или} \\ \pm (\pi - \theta/2). \end{cases} \quad (8.85)$$

Для получения максимальной интенсивности восстановленной волны необходимо применять для освещения

¹⁾ При разности хода в несколько *целых* длин волн также имеет место конструктивная (усиливающая) интерференция. Мы рассматриваем только условие первого порядка, соответствующее разности хода в *одну* длину волны.

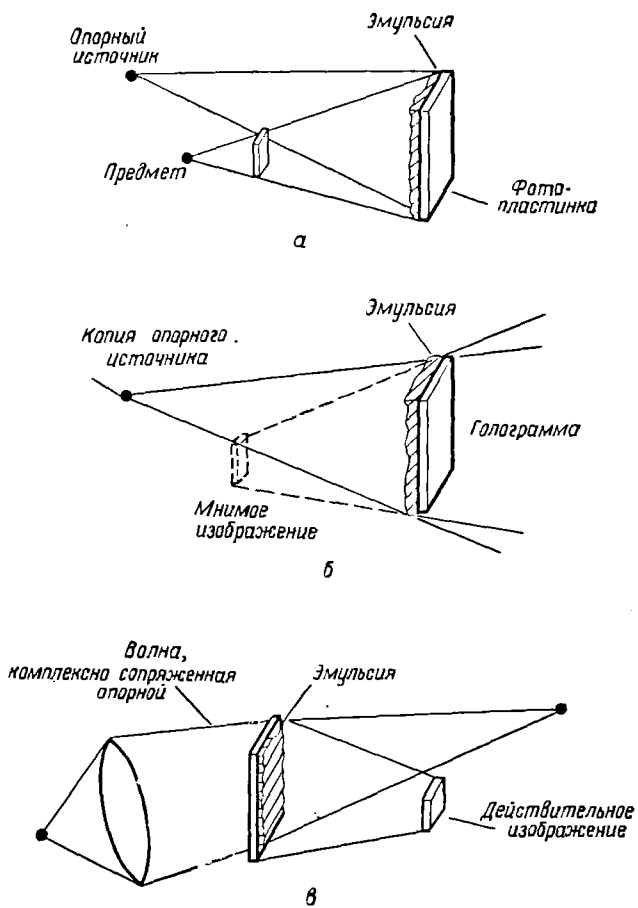
голограммы волну, являющуюся точной копией исходной опорной волны ($\psi = +\theta/2$). Эта восстановленная волна дает мнимое изображение объекта. Мы видим также, что при освещении голограммы волной исходного объекта получается копия опорной волны ($\psi = -\theta/2$). Кроме того, условие Брэгга удовлетворяется и для волн, распространяющихся в направлениях, противоположных направлениям исходных волн. Если осветить голограмму волной, распространяющейся навстречу опорной волне [$\psi = -(\pi - \theta/2)$], то возникающая при этом плоская волна объекта распространяется в направлении, противоположном тому, которое она имела при падении. Эта «комплексно сопряженная» волна дает *действительное* изображение исходного объекта. Подобным же образом действительное изображение источника опорной волны может быть получено при освещении голограммы волной, комплексно сопряженной волне объекта ($\psi = -\pi - \theta/2$).

Основной результат этого исследования состоит в том, что в случае «толстой» эмульсии максимальная яркость действительного и мнимого изображений получается только при соответствующем освещении голограммы¹⁾. Чтобы получить мнимое изображение, необходимо осветить голограмму копией опорной волны. Чтобы получить действительное изображение, голограмму необходимо осветить (с противоположной стороны) волной, комплексно сопряженной опорной. Эти результаты представлены на фиг. 123 для простого случая расходящейся сферической опорной волны.

Наконец, необходимо иметь в виду, что любая данная эмульсия может вести себя и как поверхностная дифракционная решетка (случай *тонкой* эмульсии) и как пространственный набор пластиночек (случай *толстой* эмульсии).

Весь вопрос заключается в том, преобладают ли в данном случае поверхностные или объемные эффекты, что зависит от толщины эмульсии и от углового спектра

¹⁾ Автор опускает очевидные рассуждения и переходит от элементарной голограммы к голограммам от сложных объектов. — *Прим. ред.*



Фиг. 123. Получение действительного и мнимого изображений от «толстой» голограммы.

а — запись голограммы; б — получение мнимого изображения; в — получение действительного изображения.

экспонирующих ее плоских волн. Как правило, эмульсию можно считать толстой, если ее толщина L много больше наименьшего периода Λ пластинчатой структуры; точно так же ее можно считать тонкой при $L \ll \Lambda$. Конечно, существует промежуточный случай, когда эффект Брэгга проявляется слабо.

§ 8. ОБОБЩЕНИЯ

В настоящем параграфе кратко рассматриваются некоторые обобщения обсуждавшихся нами ранее вопросов. Для более полного ознакомления с этим материалом интересующийся читатель может обратиться к цитируемым работам.

Голография движущихся объектов [27]

Во всех предшествовавших случаях мы полагали, что распределения полей объекта и опорной волны постоянны во времени, по крайней мере в течение экспонирования. В некоторых случаях объект может перемещаться, в результате чего интерференционная картина изменяется при экспонировании. Обобщим теперь нашу теорию так, чтобы включить в нее эффекты, связанные с движением объекта.

Положим, что применяется монохроматический опорный источник, и поэтому поле опорной волны, падающей на пленку, можно записать в комплексном представлении

$$U_r(x, y; t) = A(x, y) \exp(-j2\pi\nu_0 t). \quad (8.86)$$

Пусть поле объекта меняется во времени, тогда, используя комплексное представление, имеем

$$U_o(x, y; t) = a(x, y; t) \exp(-j2\pi\nu_0 t). \quad (8.87)$$

Таким образом, суммарная экспозиция

$$E(x, y) = E_r(x, y) + E_o(x, y) + A(x, y) \int_{-T/2}^{+T/2} a^*(x, y; t) dt + \\ + A^*(x, y) \int_{-T/2}^{+T/2} a(x, y; t) dt, \quad (8.88)$$

где E_r и E_0 — экспозиции, обусловленные соответственно опорной волной и волной объекта, а T — время экспонирования.

Ограничиваясь рассмотрением только одной компоненты экспозиции, ответственной в конечном счете за возникновение мнимого изображения, запишем

$$E'(x, y) = A^*(x, y) \int_{-T/2}^{+T/2} a(x, y; t) dt. \quad (8.89)$$

Это соотношение содержит всю информацию, необходимую для того, чтобы предсказать влияние движения объекта на изображение. Однако во многих случаях удобно записывать этот результат в пространстве частот. Конкретно мы можем написать, воспользовавшись табл. I (см. гл. 2),

$$\begin{aligned} E'(x, y) &= A^*(x, y) \int_{-\infty}^{+\infty} \text{rect}\left(\frac{t}{T}\right) a(x, y; t) dt = \\ &= A^*(x, y) T \int_{-\infty}^{+\infty} \text{sinc}(T\nu) G_a(x, y; \nu) d\nu, \end{aligned} \quad (8.90)$$

где $G_a(x, y; \nu)$ определяется выражением

$$G_a(x, y; \nu) = \int_{-\infty}^{+\infty} a(x, y; t) \exp(-j2\pi\nu t) dt, \quad (8.91)$$

а для получения последнего равенства использована обобщенная теорема Парсеваля [28].

Из соотношения (8.90) очевидно, что воздействие постоянной экспозиции в течение T секунд совершенно эквивалентно операции линейной фильтрации с (зависящей от времени) передаточной функцией

$$H(\nu) = \text{sinc}(T\nu), \quad (8.92)$$

действующей на $a(x, y; t)$. Этот результат можно интерпретировать следующим образом: если излучение, падающее на пленку от данной точки объекта (одинаковое на всей пленке), изменяется по отношению к опорному

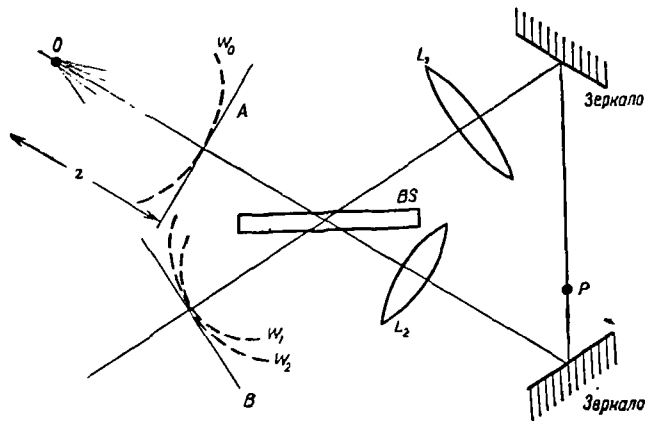
с частотой ν сек⁻¹, то интенсивность соответствующей точки изображения будет ослаблена в $\sin^2(T\nu)$ раз.

Полученный результат будет использован в § 9 при рассмотрении анализа вибраций. С другими применениями читатель может познакомиться по оригинальной работе [27].

Голография в пространственно-некогерентном свете

Если первоначально голографический метод получения изображения был разработан для случая когерентного освещения, то в настоящее время существуют способы, позволяющие получать голограмму и при некогерентном освещении. Обобщение голографической методики на случай некогерентного освещения впервые было предложено Мерцем и Юнгом [29]. Дальнейший вклад в теоретическое и экспериментальное изучение некогерентной голографии внесли Ломан [30], Строук и Рестрик [31] и Кохран [32].

Свет от какой-либо одной точки некогерентно освещенного объекта теперь уже не будет, конечно, интерферировать со светом от любой другой точки этого объекта. Тем не менее, если посредством некоторой опти-



Фиг. 124. Треугольный интерферометр для некогерентной голографии.

ческой манипуляции разделить свет, идущий от каждой точки объекта, на две части, то каждая пара волн, исходящих из одной точки, может интерферировать друг с другом и давать систему полос. Таким образом, информацию о каждой точке объекта можно зашифровать в виде некоторой системы полос. Если при этом шифровка осуществляется однозначно, т. е. так, что никакие две точки объекта не могут дать одинаковое распределение полос, то тогда в принципе можно получить и изображение объекта.

В настоящее время известно много оптических систем, пригодных для осуществления необходимого нам разделения волн, испускаемых объектом [30, 31]. Мы рассмотрим здесь только одну систему, предложенную Кохраном [32]. Как показано на фиг. 124, система состоит из треугольного интерферометра, в который помещены две линзы L_1 и L_2 с различными фокусными расстояниями f_1 и f_2 . Мы будем считать обе линзы положительными, хотя можно воспользоваться и комбинацией одной положительной и одной отрицательной линз. Линзы расположены на расстоянии $f_1 + f_2$ друг от друга, так что их фокусы совпадают в точке P . Плоскости A и B расположены соответственно на расстоянии f_1 от линзы L_1 и на расстоянии f_2 от линзы L_2 .

Свет от плоскости A может распространяться к плоскости B по любому из двух возможных направлений: через интерферометр по часовой стрелке или против часовой стрелки. Рассмотрим первый случай. Свет проходит расстояние f_1 от плоскости A до линзы L_1 , испытывая отражение в светоделителе BS . Затем свет проходит расстояние от L_1 до L_2 , равное $f_1 + f_2$, и расстояние f_2 от L_2 до плоскости B (опять испытывая отражение в BS). При нашем выборе отношения длин путей к фокусным расстояниям f_1 и f_2 плоскость A отображается на плоскость B , причем при показанной последовательности расположения линз L_1 и L_2 изображение получается с увеличением $M_1 = -f_2/f_1$.

При распространении против часовой стрелки свет в каждом случае проходит через светоделитель, а не отражается от него. Плоскость A опять отображается на плоскость B , но теперь свет проходит через систему линз

в обратном направлении, и мы получаем увеличение $M_2 = -f_1/f_2$.

Рассмотрим точку O (фиг. 124) некогерентного объекта, расположенную на расстоянии z от плоскости A . Так как свет от этой точки определяет распределение фаз на плоскости A , то сферическую волну, падающую на плоскость A (волновой фронт W_0 на фиг. 124), мы можем записать в виде комплексной функции

$$U_a(x, y) = U_0 \exp \left[j \frac{\pi}{\lambda z} (x^2 + y^2) \right]. \quad (8.93)$$

В плоскости B появятся две сферические волны (волновые фронты W_1 и W_2 на фиг. 124), одна с увеличением $M_1 = -f_2/f_1$, вторая с увеличением $M_2 = -f_1/f_2$. Таким образом, суммарная амплитуда будет

$$U_b(x, y) = U_1 \exp \left\{ j \frac{\pi}{\lambda z} \left[\left(\frac{x}{M_1} \right)^2 + \left(\frac{y}{M_1} \right)^2 \right] \right\} + \\ + U_2 \exp \left\{ j \frac{\pi}{\lambda z} \left[\left(\frac{x}{M_2} \right)^2 + \left(\frac{y}{M_2} \right)^2 \right] \right\}. \quad (8.94)$$

Соответствующее распределение интенсивности имеет вид

$$\mathcal{I}_b(x, y) = |U_1|^2 + |U_2|^2 + 2|U_1||U_2| \times \\ \times \cos \left[\frac{\pi}{\lambda z} \left(\frac{f_1^4 - f_2^4}{f_1^2 f_2^2} \right) (x^2 + y^2) + \psi \right], \quad (8.95)$$

где мы воспользовались соотношением

$$\frac{1}{M_1^2} - \frac{1}{M_2^2} = \frac{f_1^4 - f_2^4}{f_1^2 f_2^2}.$$

Если фотопластинка экспонируется светом с распределением интенсивности (8.95), а затем обрабатывается так, что получается диапозитив с амплитудным коэффициентом пропускания, пропорциональным экспозиции, то его пропускание можно записать в виде

$$t(x, y) = t_b + \beta' U_1 U_2^* \exp \left\{ j \left[\frac{\pi}{\lambda z} \left(\frac{f_1^4 - f_2^4}{f_1^2 f_2^2} \right) (x^2 + y^2) \right] \right\} + \\ + \beta' U_1^* U_2 \exp \left\{ -j \left[\frac{\pi}{\lambda z} \left(\frac{f_1^4 - f_2^4}{f_1^2 f_2^2} \right) (x^2 + y^2) \right] \right\}. \quad (8.96)$$

Второй и третий члены можно рассматривать как функции пропускания соответственно отрицательной и положительной линз [см. (5.10)], фокусное расстояние каждой из которых равно

$$f = \frac{f_1^2 f_2^2}{f_1^2 - f_2^2} z. \quad (8.97)$$

Таким образом, если диапозитив осветить волной от когерентного источника, то получается как мнимое, так и действительное изображения исходного точечного источника.

Переходя к объекту, состоящему из множества взаимно некогерентных точечных источников, получаем, что каждый такой источник дает на пленке свою собственную интерференционную картину. Поскольку различные источники некогерентны, полная интенсивность определяется просто как сумма интенсивностей всех получающихся таким образом распределений. Координаты (x, y) каждого точечного источника определяют центр отвечающей ему интерференционной картины и поэтому задают координаты (x, y) действительного и мнимого изображений, а z -координата точечного источника определяет соответствующее фокусное расстояние [см. (8.97)]; таким образом, получающееся изображение является трехмерным.

Хотя во многих приложениях было бы весьма желательно использовать вместо когерентного некогерентный свет, в настоящее время еще не решены многие серьезные практические проблемы, ограничивающие возможности некогерентной голографии. Трудности возникают из-за того, что в *некогерентной* голографии каждая элементарная интерференционная картина полос получается в результате интерференции двух чрезвычайно слабых пучков, составляющих ничтожную долю всего света, падающего на пленку, тогда как в *когерентной* голографии свет от каждой точки объекта интерферирует со всем светом опорной волны. Таким образом, при некогерентной голографии всегда приходится давать большие экспозиции смещения. Из-за этой основной проблемы некогерентная голография успешно применяется только для

объектов, состоящих из относительно малого числа элементарных частей. Дальнейшие возможности использования некогерентной голографии в огромной мере зависят от того, будет ли найден простой способ решения указанной проблемы.

Отражательная голограмма

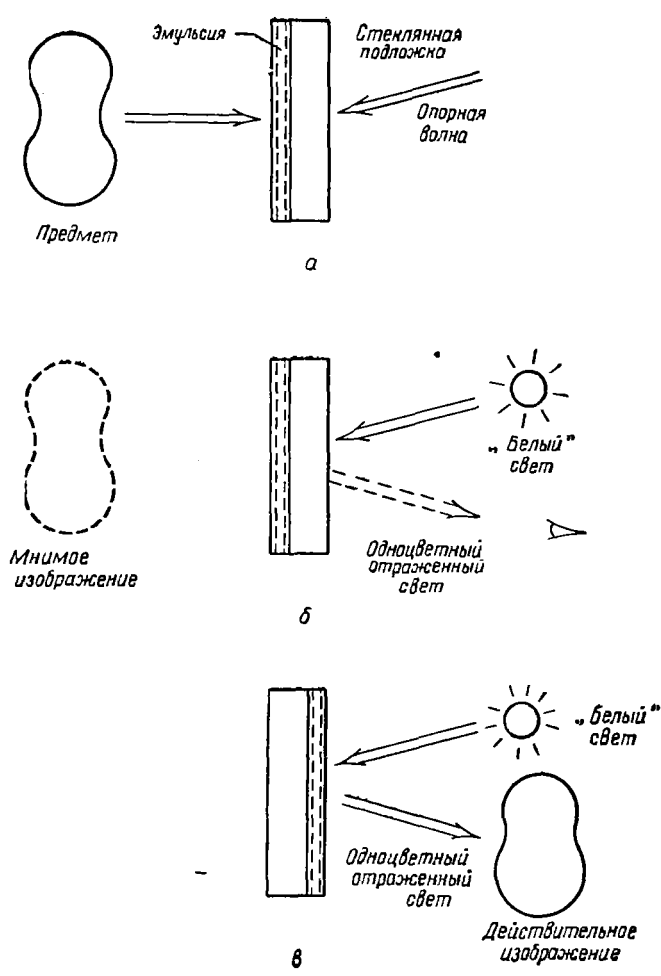
При соответствующем видоизменении техники когерентной голографии можно записать голограмму в (почти монохроматическом) когерентном свете, а изображения получить в «белом свете». В этом случае восстановленные волновые фронты получаются не при прохождении света через голограмму, а при отражении от нее. Этот метод основан на использовании трехмерных свойств фотографической эмульсии и тесно связан во многих отношениях с методикой цветной фотографии Липмана [33].

Основная идея такого метода голографии была впервые выдвинута в 1962 г. в Советском Союзе Денисюком [34]. Эта идея не была в должной мере оценена в Соединенных Штатах до опубликования работ Струка и Лабейри [35] и Лина и др. [36] в 1966 г. Довольно полный анализ отражательной голографии можно найти в работах Лейта и др. [37].

Чтобы записать отражательную голограмму, волна от когерентно освещенного объекта и опорная волна должны падать на фотографическую эмульсию с противоположных сторон (фиг. 125, а). Как уже было отмечено ранее (см. § 7), в результате интерференции этих двух волн пленка (после проявления) будет представлять собой диапозитив со структурой в виде слоев металлического серебра, действующих как отражающие плоскости. Так как угол θ близок к 180° , отражающие плоскости (делящие угол θ пополам) идут почти параллельно поверхности эмульсии. Из соотношения (8.83) следует, что плоскости располагаются на расстоянии

$$\Lambda \approx \frac{\lambda}{2}. \quad (8.98)$$

Чтобы получить мнимое изображение, мы должны осветить диапозитив копией опорной волны. Тогда мы



Ф и г. 125. Голография в отраженном свете.

а — запись голограммы; б — получение мнимого изображения; в — получение действительного изображения.



Фиг. 126. Фотография мнимого изображения, восстановленного от отражательной голограммы при освещении ее белым светом.

получим воспроизведение волнового фронта исходного объекта (фиг. 125, б), а это в данном случае и означает, что изображение получено в *отраженном* свете. Отметим также, что для данной геометрии освещения и наблюдения закон Брэгга при восстановлении будет выполняться только для одной длины волны. Таким образом, если освещение производится белым светом, то отражается только свет с длиной волны, удовлетворяющей закону Брэгга. весь же остальной проходит через эмульсию или частично поглощается.

Чтобы получить действительное изображение, нужно осветить диапозитив с противоположной стороны. Изображение, получающееся в отраженном свете, опять будет образовано волнами только одного цвета (см. фиг. 125, в).

В принципе цвет изображений должен быть почти таким же, как цвет света, использовавшегося при записи голограммы. На практике оказывается, что длина волны отраженного света меньше длины волны света, действовавшего при экспонировании. Это происходит из-за того, что при проявлении и фиксировании эмульсия сжимается. В результате расстояние между слоями серебра будет меньше ожидаемого значения. В тех случаях, когда нужно избежать изменения длины волны при восстановлении (это необходимо, например, в многоцветной голографии [36]), нужно принять какие-то меры против сжатия эмульсии. Если опустить процесс фиксирования, то сжатие в основном устраняется, но тогда проявленные пластинки постепенно теряют свои свойства. Возможно, наиболее приемлемым решением является какой-то метод расширения эмульсии после фиксирования. При достаточно аккуратном применении этого метода можно сделать так, что эмульсия будет отражать любую заданную длину волны в пределах широкого спектрального интервала.

На фиг. 126 приведена фотография мнимого изображения, полученного от отражательной голограммы. Голограмма была записана в свете гелий-неонового лазера с длиной волны 6328 Å. Без применения метода расширения эмульсии голограмма (отфиксированная) отражает зеленый свет.

§ 9. ПРИМЕНЕНИЯ ГОЛОГРАФИИ

В начале 60-х годов процесс восстановления волнового фронта был значительно модернизирован и усовершенствован, благодаря чему голография стала проникать во все новые и новые области. В настоящем параграфе мы рассмотрим несколько наиболее важных из этих приложений. Будем ограничиваться только голографией в *оптическом* диапазоне. Читатель должен иметь в виду, что этот процесс находит применение в СВЧ и в радиочастотном диапазоне спектра (см., например, [38]), а также и для получения изображения с помощью акустических волн [39].

Микроскопия

Исторически сложилась точка зрения, что наиболее вероятной областью применения голографии должна быть микроскопия: это было основным стимулом большинства ранних работ, посвященных процессу восстановления волнового фронта. Такое положение сохранялось и после первых работ Габора [1, 3, 4] и Эль-Сума [6].

На современном этапе развития найдено много новых приложений голографии, поэтому теперь уже нельзя сказать, как это было до сих пор, что развитие голографии стимулируется *исключительно* задачами микроскопии. К тому же, хотя сейчас уже в достаточной мере ясна полезность применения голограмм в микроскопии, совершенно очевидно, что новая методика не является серьезным конкурентом обычного микроскопа в традиционной микроскопии. Тем не менее существуют два аспекта, в которых использование голографии в микроскопии весьма перспективно. Сюда относятся прежде всего способ получения *объемных изображений высокого разрешения*, на чем мы остановимся более подробно в следующем параграфе. В обычной микроскопии высокое разрешение достигается за счет ограничения глубины фокуса. Открываемая применением голографии возможность получения трехмерных изображений потенциально должна привести к значительному увеличению глубины фокуса.

Другое применение в области микроскопии — использование голограмм для создания рентгеновского микроскопа — находится, надо сказать, еще в такой зачаточной стадии, что сейчас даже трудно оценить все его возможности. Основная идея состоит в том, что для записи голограммы используется рентгеновское излучение, а восстановление проводится в оптическом диапазоне. За счет различия длин волн можно достигнуть огромных увеличений [см. соотношение (8.46)]; однако более существенно то, что применение для освещения рентгеновских волн должно дать возможность получить разрешение порядка нескольких ангстрем ($1 \text{ \AA} = 10^{-10} \text{ м}$). Такое разрешение сравнимо с разрешением, достигаемым в элек-

тронном микроскопе. Однако от рентгеновских лучей образец нагревается меньше, чем от потока электронов; кроме того, отпадает необходимость в вакуумной системе, обязательной для электронного микроскопа. Успех или неудача голографического рентгеновского микроскопа, несомненно, зависит от дальнейшего прогресса в области создания источников и приемников излучений.

Интерферометрия

Голография дает возможность осуществить несколько довольно уникальных способов интерферометрии. Возможность выполнения интерферометрии с помощью процесса восстановления волнового фронта основана на том, что получающиеся в этом случае изображения *когерентны* и имеют хорошо известные распределения амплитуд и фаз. Любое использование голографии для получения суперпозиции двух когерентных изображений является потенциальной основой интерференционного метода.

Наиболее перспективный метод голографической интерферометрии основан на свойстве, отмеченном Габором и др. [40] и заключающемся в том, что посредством многократного экспонирования голограмм можно осуществить когерентное сложение комплексных амплитуд. Это свойство легко установить следующим образом. Пусть фотографическая эмульсия последовательно экспонируется N различными распределениями интенсивностей $\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_N$. Тогда полную экспозицию можно записать как

$$E = \sum_{k=1}^N T_k \mathcal{I}_k, \quad (8.99)$$

где T_1, T_2, \dots, T_N представляют собой N различных интервалов времени, в течение которых пленка подвергалась действию света. Теперь положим, что для каждого интервала засветки падающее излучение есть сумма фиксированной опорной волны $A(x, y)$ и волны от объекта $a_k(x, y)$, меняющейся от интервала к интервалу.

Тогда для суммарной экспозиции имеем

$$E = \sum_{k=1}^N T_k |A|^2 + \sum_{k=1}^N T_k |a_k|^2 + \sum_{k=1}^N T_k A^* a_k + \sum_{k=1}^N T_k A a_k^* \quad (8.100)$$

Если предположить, что мы работаем в линейной области t — E -кривой эмульсин, то получим компоненту пропускания

$$t_\alpha = \beta' \sum_{k=1}^N T_k A^* a_k \quad (8.101)$$

и компоненту

$$t_\beta = \beta' \sum_{k=1}^N T_k A a_k^* \quad (8.102)$$

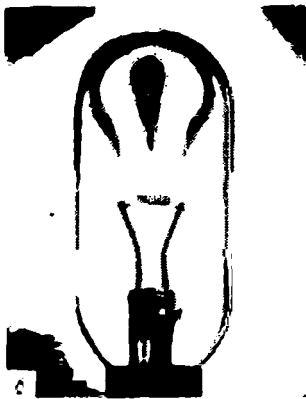
Из соотношения (8.101) ясно, что при освещении диапозитива волновым фронтом A в прошедшем поле появится компонента, пропорциональная сумме (с учетом веса) комплексных волновых фронтов a_1, a_2, \dots, a_N . В результате N когерентных мнимых изображений исходного объекта, которым соответствует совокупность волн a_i , будет испытывать линейную суперпозицию и, конечно, интерферировать.

Таким же образом при освещении диапозитива волновым фронтом A^* получается в соответствии с (8.102) N когерентных действительных изображений, которые интерферируют друг с другом.

Наиболее эффектные эксперименты, демонстрирующие возможности интерферометрии этого типа, были осуществлены Бруксом и др. [41], использовавшими рубиновый лазер с модулированной добротностью. На фиг. 127 приведены две фотографии, каждая из которых получена двойной экспозицией голограммы двумя лазерными импульсами. На фиг. 127, *a* первый импульс записывал только голограмму диффузного фона, а второй импульс давал наложенную на этот фон голограмму пули в полете. Возникающие при движении пули ударные волны вызывают изменение показателя преломления воздуха. В результате два изображения диффузного фона — одно записанное в отсутствие пули, а второе при

наличии неоднородностей показателя преломления — интерферируют, и мы получаем изображение ударных волн, возникающих при полете пули. Эти полосы кажутся фиксированными в трехмерном пространстве около пули.

На фиг. 127, б приведено полученное точно таким же способом изображение лампы накаливания. При первой экспозиции нить была выключена. При этом опять сначала записывалась голограмма диффузного фона, но на этот раз через стеклянную колбу лампы. Затем включалась нить накала и голограмма экспонировалась вторым лазерным импульсом. Некогерентный свет, излучаемый лампой, не интерферирует с лазерным светом, поэтому в конечном изображении нить накала кажется не светящейся. Однако нагрев газа в колбе вызывает изменение показателя преломления, что в свою очередь приводит к появлению в конечном изображении интерференционных полос, дающих представление о расширении газа. Следует подчеркнуть, что эти интерференционные полосы получены при наличии оптически несовершенной стеклянной колбы — возможность, которую не дает ни один другой «классический» метод интерферометрии.



Фиг. 127. Голографическая интерферометрия с помощью рубинового лазера с модулированной добротностью.

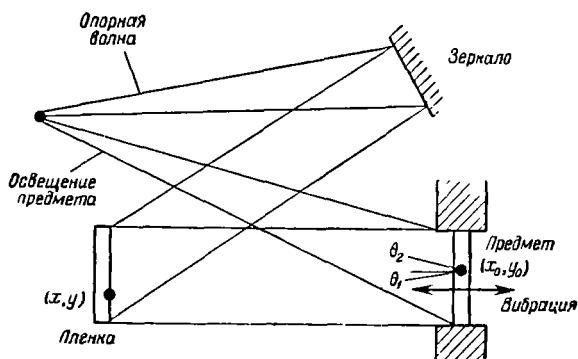
Анализ вибраций

Голографический способ анализа вибраций был впервые предложен и осуществлен Пауэлом и Стетсоном [42]. Этот способ можно рассматривать как обобщение метода интерферометрии с несколькими последовательными экспозициями на случай непрерывного экспонирования вибрирующего объекта.

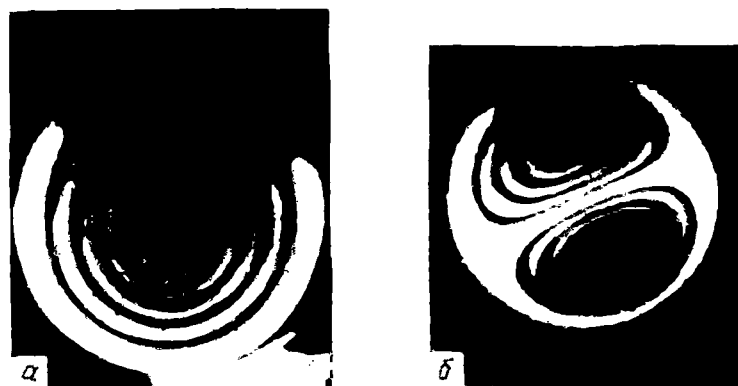
Эту схему можно проанализировать, вспоминая, что получение голограммы эквивалентно фильтрации, зависящей от времени (см. § 8). Обратимся к схеме на фиг. 128 и рассмотрим точку плоского объекта с координатами (x_0, y_0) , колеблющуюся по синусоидальному закону с угловой частотой Ω . Пусть максимальная амплитуда вибрации равна $m(x_0, y_0)$, а начальная фаза равна $\mu(x_0, y_0)$. Можно считать, что свет, приходящий из этой точки объекта в точку (x, y) на пленке, имеет меняющуюся во времени фазовую модуляцию

$$\varphi(x, y; t) = \frac{2\pi}{\lambda} (\cos \theta_1 + \cos \theta_2) m(x_0, y_0) \cos[\Omega t + \mu(x_0, y_0)], \quad (8.103)$$

где λ — оптическая длина волны источника света, θ_1 — угол между вектором смещения и линией, соединяющей точки (x_0, y_0) с (x, y) , а θ_2 — угол между вектором смещения и направлением распространения света, падаю-



Фиг. 128. Схема записи голограммы вибрирующего объекта.



Фиг. 129. Голографические изображения диафрагм, колеблющихся в двух различных модах.

щего в точку (x_0, y_0) . Так как эта фазовая модуляция имеет синусоидальный характер, частотный спектр $F(\nu)$ меняющегося во времени фазора для света, приходящего в точку (x, y) , можно, как обычно, выразить путем разложения волны с синусоидальной фазовой модуляцией по бесселевым функциям

$$F(\nu) = \sum_{k=-\infty}^{\infty} J_k \left[2\pi \frac{\cos \theta_1 + \cos \theta_2}{\lambda} m(x_0, y_0) \right] \delta \left(\nu - \frac{k\Omega}{2\pi} \right). \quad (8.104)$$

Когда время экспонирования много больше периода колебаний (т. е. $T \gg 2\pi/\Omega$), только «несущий» член с $k = 0$ попадает в полосу пропускания передаточной функции (8.92). Если изменения глубины модуляции, описываемые членом $\cos \theta_1$, почти не зависят от (x, y) [т. е. если угол, под которым видна пленка из точки (x_0, y_0) , мал], то уменьшение амплитуды изображения точки (x_0, y_0) будет определяться множителем

$$J_0 \left[\frac{2\pi}{\lambda} (\cos \theta_1 + \cos \theta_2) m(x_0, y_0) \right], \quad (8.105)$$

а уменьшение интенсивности — квадратом этого множителя. Таким образом, интенсивность в каждой точке

изображения зависит от величины $m(x_0, y_0)$ для соответствующей точки объекта.

На фиг. 129 приведены фотографии колеблющихся диафрагм, полученные экспериментально Пауэлом и Стетсоном. На фиг. 129, *а* колебания диафрагмы соответствуют моде наименьшего порядка с одним максимумом в центре диафрагмы. На фиг. 129, *б* колебания соответствуют моде более высокого порядка с двумя максимумами. Подсчитывая число полос от края диафрагмы до любой интересующей нас точки, можно с помощью выражения (8.105) определить амплитуду колебания в этой точке.

Объемное изображение высокого разрешения

Как уже упоминалось при обсуждении микроскопии, одно из наиболее важных свойств голографических изображений — это возможность получения большой глубины фокуса. Если какая-либо оптическая система (например, микроскоп) имеет высокое поперечное разрешение, то из самых общих соображений следует, что «в фокусе» одновременно может находиться только ограниченная часть пространства предмета. Конечно, можно исследовать большой объем *последовательно*, непрерывно меняя фокусировку, чтобы рассматривать новые области объекта. Однако в случае меняющегося или перемещающегося объекта такой подход часто оказывается неудовлетворительным.

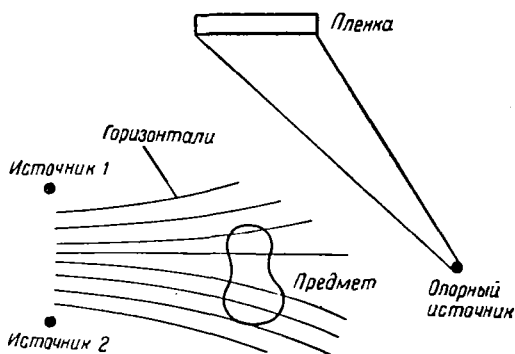
Эту проблему можно решить, если для записи голограммы воспользоваться импульсным лазером, позволяющим давать небольшие выдержки. Подвижный объект в этом случае как бы «замораживается» во времени, однако при записи сохраняется вся информация, необходимая для того, чтобы можно было исследовать весь объем объекта. Затем голограмма освещается обычным образом и получающиеся действительное и мнимое изображения исследуются с помощью вспомогательной оптической системы. Теперь уже возможно последовательное наблюдение объема изображения, так как объект (т. е. голографическое изображение) неподвижен.

Нокс [43] успешно использовал такой метод в трехмерной объемной микроскопии живых биологических объектов, а Томпсон, Вард и Цинки [44] применили его для измерения распределения частиц по размерам в аэрозолях. Для выяснения деталей читатель может обратиться к указанным работам.

Нанесение горизонталей на трехмерные объекты

Описанная выше интерференция нескольких когерентных изображений была использована при разработке способа получения трехмерных изображений с нанесенными на них линиями одинакового уровня (горизонталями). Этот способ применим для построения сечений и нанесения горизонталей. Гильдебранд и Хайнс [45] предложили для этого два совершенно различных метода. В первом из них объект освещается двумя взаимно когерентными, но пространственно разделенными точечными источниками. Голограмма может быть экспонирована либо один раз двумя источниками, либо два раза с различными положениями источников относительно объекта при каждом экспонировании.

При рассмотрении интерференционной картины, полученной при освещении объекта двумя источниками, нетрудно видеть, что она состоит из светлых и темных интерференционных полос в виде гипербол одинаковой разности хода (фиг. 130). Если объект освещать сбоку,



Фиг. 130. Получение горизонталей методом двух источников.



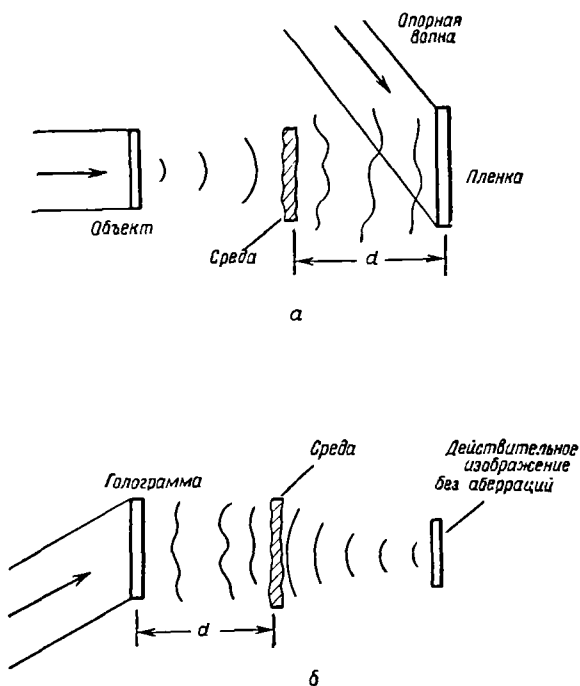
Ф и г. 131. Получение горизонталей двухчастотным методом.

а голограмму записывать сверху, то на восстановленном изображении легко заметить линии одинакового уровня. Результаты не зависят от того, используем ли мы одновременно два источника при одной экспозиции или каждый источник в отдельности при двух экспозициях; в любом случае происходит когерентное сложение освещенностей.

Нанесение горизонталей методом двух источников обладает тем недостатком, что направления освещения и наблюдения должны отличаться почти на 90° . Поэтому в случае рельефного объекта возникнут тени и некоторые части объекта просто не будут освещены. Этот недостаток можно устранить, если для получения горизонталей воспользоваться двухчастотным методом. В этом случае как волна, идущая от объекта, так и опорная волна содержат компоненты с двумя частотами. Если полученную таким образом голограмму освещать светом одной частоты, то возникнут два изображения, слегка отличающиеся положением в пространстве и увеличением. Эти два изображения будут интерферировать друг с другом. При некоторых условиях на изображениях видны горизонталы, позволяющие точно определить глубину. Мы не будем подробно останавливаться на анализе этого более сложного случая; интересующийся читатель может найти более подробные сведения в работе [45]. Фиг. 131 иллюстрирует результат нанесения горизонталей двухчастотным методом. На фиг. 131, а мы видим голографическое изображение монеты, полученное при использовании света одной частоты. Если для записи голограммы использовать свет двух частот, то получается изображение, представленное на фиг. 131, б. В указанном случае использовались две различные линии аргонового лазера. Длины волн этих линий отличаются на 65 \AA , поэтому расстояние между горизонталями на изображении равно $0,02 \text{ мм}$.

Передача изображения через искажающую среду

Во многих практически интересных случаях оптическая система должна давать изображение при наличии неконтролируемых аберраций. Аберрации могут появиться



Ф и г. 132. Использование исходной искажающей среды для компенсации aberrаций.

а — запись голограммы; *б* — восстановление изображения.

ся из-за дефектов в деталях самой оптической системы или могут быть результатом воздействия внешней среды, например атмосферы. Голографическая техника открывает совершенно исключительные возможности получения изображений при наличии таких aberrаций. Мы рассмотрим три различных голографических метода, позволяющих получить высокое разрешение при наличии сильных aberrаций.

Первый метод [46, 47] применим в том случае, когда искажающая среда неизменна во времени, но ее положение в пространстве можно изменять. Как показано на фиг. 132, голограмма искаженных aberrациями волн

объекта записывается при неискаженной опорной волне. Если теперь осветить голограмму таким образом, чтобы получилось действительное изображение объекта, то между голограммой и плоскостью изображения появится действительное изображение среды с оптическими дефектами. Пусть $U_0(\xi, \eta)$ представляет собой волну от исходного объекта, падающую на искажающую среду. Если амплитудный коэффициент пропускания искажающей среды равен $\exp[jW(\xi, \eta)]$, то при восстановлении действительного изображения среды мы получим комплексно сопряженную волну $U_0^*(\xi, \eta)\exp[-jW(\xi, \eta)]$. Если при восстановлении исходную искажающую среду расположить так, чтобы она точно совпала со своим комплексно сопряженным изображением, то прошедшая через среду волна будет

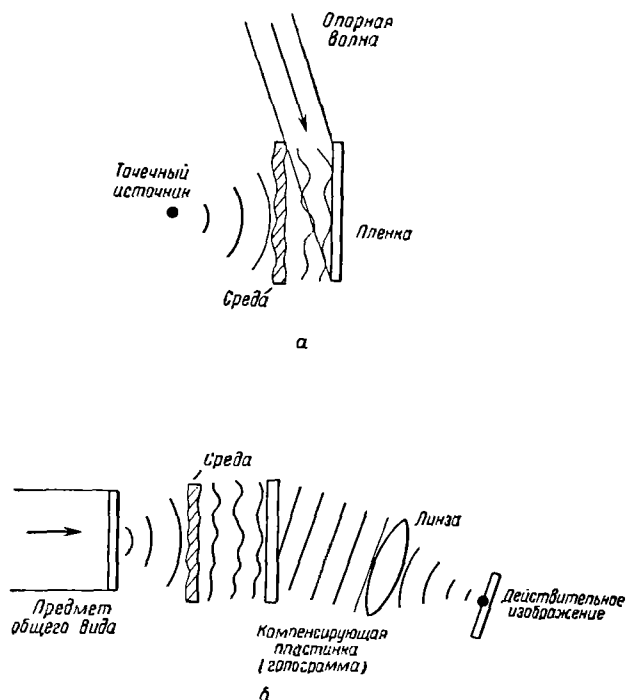
$$U_0^*(\xi, \eta)\exp[-jW(\xi, \eta)]\exp[jW(\xi, \eta)] = U_0^*(\xi, \eta).$$

Таким образом, аберрации точно компенсируются и мы получаем изображение без аберраций¹⁾.

Эта методика, несомненно, найдет широкое применение в криптографии или для надежной зашифровки донесений. Донесение, предназначенное для зашифровки, служит объектом. Рассеивающее устройство помещают между объектом и плоскостью пленки. При освещении голограммы обычным способом получается столь сильно искаженное изображение, что на нем невозможно что-либо прочитать. Однако какое-то лицо, у которого есть *точно такой же* диффузор, как и использовавшийся в процессе записи, может, поставив свою «декодирующую пластинку» соответствующим образом, получить неискаженное изображение.

Другой представляющий интерес метод иллюстрируется схемой на фиг. 133. В этом случае неподвижная искажающая среда не должна изменяться со временем (или по крайней мере меняться очень медленно). Чтобы при наличии аберраций получить изображения с хоро-

¹⁾ Требования к точности, с которой искажающая среда должна быть установлена при восстановлении, на практике могут быть очень высокими. Чтобы получить изображение через мелкозернистый диффузор, нельзя допустить ошибку даже в несколько микронов.



Ф и г. 133. Использование голографической компенсирующей пластинки.

а — запись компенсирующей пластинки; б — компенсация aberrаций.

шим разрешением, записывают голограмму искаженного абберациями волнового фронта от объекта в виде точечного источника, используя неискаженную опорную волну. Такую голограмму можно теперь использовать в качестве «компенсирующей пластинки», что позволит получить неискаженное изображение с помощью обычной оптической системы.

Пусть $\exp[iW(x, y)]$ представляет волну, падающую на фотопленку от одиночного точечного источника. Мы предполагаем при этом, что искажающая среда расположена достаточно близко к пленке, так что изменения интенсивности несущественны. Компонента пропускания

голограммы, дающая действительное изображение, пропорциональна $\exp[-jW(x, y)]$. Если заменить точечный источник более общим объектом и поместить голограмму в то же самое место, где она была первоначально записана, то оказывается, что искривление падающих на голограмму волн от объекта компенсируется при прохождении голограммы. При этом волны от различных точек объекта превращаются в плоские волны, распространяющиеся под разными углами. Тогда безабберационное изображение получается обычным образом с помощью линз.

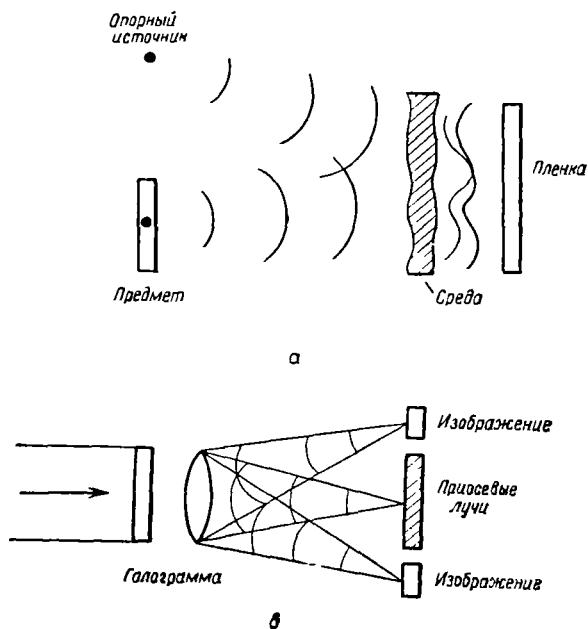
Этот метод будет давать достаточно хорошие результаты только в ограниченной части поля зрения. Действительно, если точка объекта находится слишком далеко от исходного точечного источника, использовавшегося при записи голограммы, то аберрации волны от объекта могут отличаться от аберраций, записанных на голограмме. Это ограничение наименее существенно, когда голограмма записана на очень близком расстоянии от искажающей среды. Упатник и др. [48] успешно применили этот метод для компенсации аберрации линз, для чего он особенно удобен.

Третий метод, позволяющий получать изображения через среду, которая может быть подвижной или неподвижной, меняющейся или не меняющейся во времени, состоит в том, что через искажающую среду посылаются обе волны — как опорная, так и волна от объекта. Как видно из фиг. 134, используется такая же схема, как и при записи *безлинзовой фурье-голограммы*. Для простоты положим, что искажающая среда расположена непосредственно перед плоскостью записи. Таким образом, достигающие пленки опорная волна и волна от объекта могут быть записаны в виде $A(x, y) \exp[jW(x, y)]$ и $a(x, y) \exp[jW(x, y)]$. При интерференции этих двух искаженных волн получается интерферограмма, на которой воздействие искажающей среды никак не сказывается,

$$\mathcal{I}(x, y) = |A(x, y) \exp[jW(x, y)] + a(x, y) \exp[jW(x, y)]|^2 =$$

$$= |A|^2 + |a|^2 + A^*a + Aa^*.$$

Следовательно, наличие искажающей среды никак не влияет на голограмму.



Фиг. 134. Получение неискаженного изображения для случая, когда опорная волна и волна от объекта имеют одинаковые аберрации.

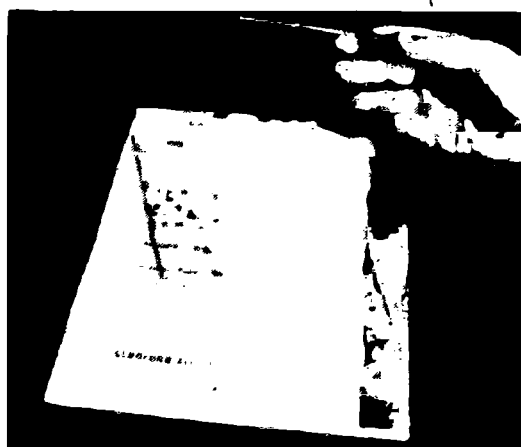
а — запись голограммы; б — получение изображения.

Как и раньше, метод дает хорошие результаты только для ограниченного поля объекта, так как точки, расположенные слишком далеко от опорного источника, могут давать волны с аберрациями, отличными от аберраций в опорной волне¹⁾. Поле зрения, в котором имеет место компенсация, максимально, когда аберрации вносятся вблизи плоскости записи. Для более детального изучения этого процесса читатель может обратиться к работе [49].

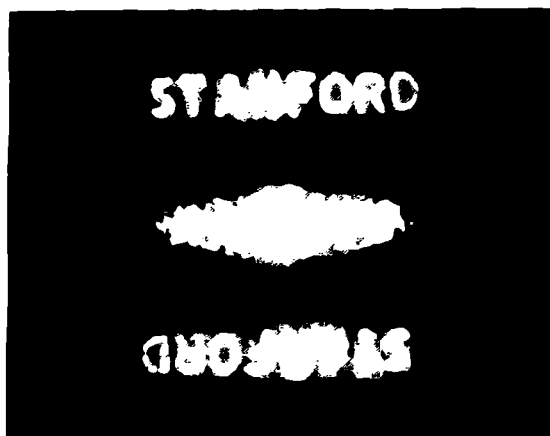
¹⁾ Точнее говоря, данная точка объекта будет изображаться без аберраций только в том случае, когда лучи из этой и из опорной точки приходят в какую-то точку пленки с одинаковыми задержками.



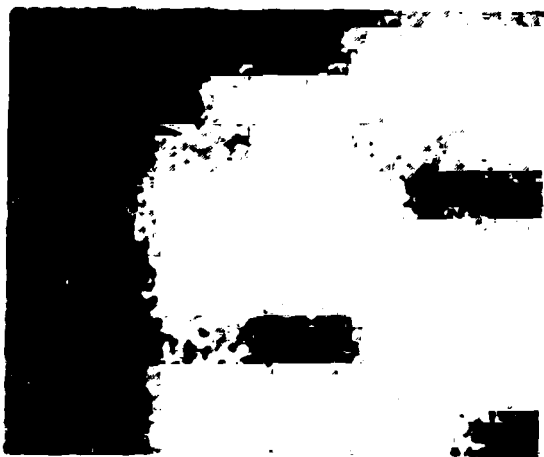
Ф и г. 135. Объект и опорный песточник.



Ф и г. 136. Искажающая среда.



Ф и г. 137. Изображения, полученные в присутствии искажающей среды методом восстановления волновых фронтов.



Ф и г. 138. Изображения, полученные обычным методом в присутствии искажающей среды.

В качестве иллюстраций полученных этим методом экспериментальных результатов рассмотрим фиг. 135—138. В этом случае объектом служил диапозитив, содержащий изображение эмблемы и яркую опорную точку, как показано на фиг. 135. Искажающей средой служило обычное волнистое стекло (фиг. 136). На фиг. 137 приведены восстановленные изображения, полученные при наличии искажающей среды. Для сравнения на фиг. 138 приведено изображение, полученное обычным способом: при этом пленка была заменена положительной линзой, непосредственно формировавшей изображение при наличии искажающей среды. Аберрации полностью уничтожают разрешение в изображении, полученном обычным способом, тогда как изображения, полученные восстановлением волнового фронта, искажены в очень малой степени.

ЗАДАЧИ

8.1. Голограмма плоского объекта записывается в плоскости, параллельной объекту. Показать, что получающиеся в результате изображения лежат в плоскостях, параллельных голограмме. (Для простоты считать опорную волну плоской.)

8.2. Для записи голограммы используется свет аргонового лазера (4880 Å), а для получения изображения — свет гелий-неонового лазера (6328 Å).

а) Полагая $z_p = \infty$, $z_r = \infty$ и $z_0 = 10$ см, найти расстояние до изображения z_i .

б) Полагая $z_p = \infty$, $z_r = 2z_0$ и $z_0 = 10$ см, найти z_i . Определить увеличение M .

8.3. Показать, что при $\lambda_2 = \lambda_1$ и $z_p = z_r$ получается мнимое изображение с единичным увеличением, а при $\lambda_2 = \lambda_1$ и $z_p = -z_r$ получается действительное изображение с единичным увеличением.

8.4. Голограмма записывается по схеме, приведенной на фиг. 104. Затем фотографический диапозитив увеличивается в m раз. Показать, что восстановленное изображение, полученное от увеличенной голограммы, находится от нее на расстоянии

$$z_i = \left[\frac{1}{z_p} \pm \frac{\lambda_2}{\lambda_1 z_r m^2} \mp \frac{\lambda_2}{\lambda_1 z_0 m^2} \right]^{-1}$$

и имеет увеличение

$$M = \left| \frac{\lambda_2 z_i}{m \lambda_1 z_0} \right|.$$

8.5. В нижеприведенной таблице указаны приблизительные значения предельных частот МПФ для пленок нескольких типов:

	линии/мм
Kodak Tri-X	50
Kodak High-Contrast Copy	60
Kodak SO-243	300
Agfa Agepan FF	600

Используется схема записи безлинзовой фурье-голограммы, причем длина волны света равна 6328 \AA , а опорный источник и объект находятся на расстоянии 10 см от пленки. Для каждой пленки оценить радиус круга (с центром в опорном источнике), обладающего тем свойством, что от точек объекта, расположенных вне этого круга, нельзя получить соответствующие точки изображения.

8.6. Нелинейная $t - E$ -кривая некоторой пленки в пределах рабочей области может быть описана функцией

$$t = t_b + \beta E_1^3,$$

где E_1 — отклонения экспозиции от опорной экспозиции.

а) Полагая, что на пленку падают опорная волна $A \exp(-j2\pi a y)$ и волна от объекта

$$a(x, y) \exp[-j\varphi(x, y)],$$

найти выражение для составляющей пропускания, которая ответственна за возникновение двух изображений первого порядка.

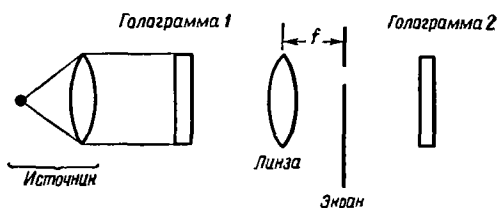
б) Какой вид принимает это выражение при $A \gg a$?

в) Сопоставить амплитудную и фазовую модуляции, полученные в случаях «а» и «б», с идеальной амплитудной и фазовой модуляциями для пленки с линейной $t - E$ -кривой.

8.7. Интенсивная опорная волна $A \exp(-j2\pi a y)$ и слабая волна от объекта $a(x, y) \exp[-j\varphi(x, y)]$ интерферируют на поверхности резкоконтрастной пленки (т. е. пленки с $t - E$ -кривой, показанной на фиг. 116). С полученной голограммы контактным способом изготавливается очень хорошая копия, так что в результате мы имеем две одинаковые голограммы.

Одна из голограмм освещается когерентным светом и с помощью оптической системы отображается на вторую идентичную ей голограмму, как показано на фиг. 139. В задней фокальной плоскости изображающей линзы помещена диафрагма, так что в плоскости изображения попадает только компонента, соответствующая одному из двух членов первого порядка пропускания голограммы 1. Показать, что при $a \ll A$ одна из компонент света, прошедшего через голограмму 2, имеет неискаженную амплитудную модуляцию $a(x, y)$ и неискаженную фазовую модуляцию $\varphi(x, y)$, несмотря на использование резкоконтрастной пленки.

8.8. Для записи голограммы объекта в виде квадратного диапозитива шириной L используется схема безлинзовой фурье-голограммы (см. фиг. 112). Амплитудный коэффициент пропускания объекта



Ф и г. 139.

равен $t_0(x_0, y_0)$, а расстояние от объекта до пленки равно z . Изображения получены с помощью положительной линзы с фокусным расстоянием f , как показано на фиг. 113.

а) Чему равно увеличение M двух изображений первого порядка?

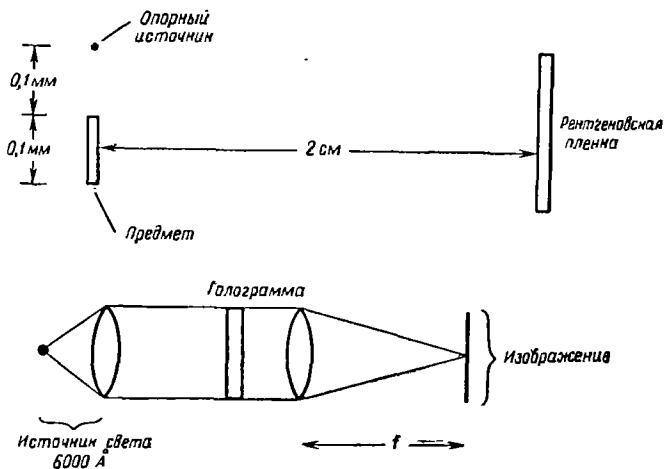
б) Показать, что амплитуда члена, соответствующего изображению нулевого порядка (т. е. световому полю между двумя изображениями первого порядка), пропорциональна автокорреляционной функции

$$t_0\left(\frac{x_0}{M}, \frac{y_0}{M}\right) \exp\left[j \frac{\pi}{\lambda z M^2} (x_0^2 + y_0^2)\right]$$

плюс амплитуда центрального дифракционного пятна.

в) На каком расстоянии от центра диапозитива, являющегося объектом, нужно поместить опорный источник, чтобы можно было гарантировать, что изображения первого порядка не перекрываются с нулевым порядком?

8.9. Для записи рентгеновской голограммы предполагается использовать излучение с длиной волны 1 \AA , а для восстановления



Ф и г. 140.

изображения применять свет с длиной волны 6000 Å. Выбрана схема безлинзовой фурье-голографии, показанная на фиг. 140 (вверху). Ширина объекта 0,1 мм. Минимальное расстояние между объектом и опорным источником выбрано равным 0,1 мм, чтобы сопряженные изображения не перекрывались. Рентгеновская пленка помещена на расстоянии 2 см от объекта.

а) Чему равна максимальная пространственная частота (линий/мм) в распределении интенсивности на пленке?

б) Пусть пленка имеет достаточное разрешение, чтобы можно было записать все изменения падающей на нее интенсивности. Предлагается для восстановления изображения воспользоваться обычным способом, показанным на фиг. 140 (внизу). Почему этот эксперимент потерпит неудачу?

ЛИТЕРАТУРА

1. Gabor D., Nature, **161**, 777 (1948).
Новый принцип микроскопии.
2. Gagg W. L., Nature, **149**, 470 (1942).
Рентгеновская микроскопия.
3. Gabor D., Proc. Roy. Soc., **A197**, 454 (1949).
Микроскопия с помощью восстановления волнового фронта.
4. Gabor D., Proc. Phys. Soc., **B64**, 449 (1951).
Микроскопия с помощью восстановления волнового фронта. II.
5. Rogers G. L., Nature, **166**, 237 (1950).
Габоровская дифракционная микроскопия. Голограмма как обобщенная зонная пластинка.
6. El-Sum H. M. A., Doctoral dissertation, Stanford University, 1952.
Микроскопия с помощью восстановления волнового фронта.
7. Lohmann A., Opt. Acta, **3**, 97 (1956).
Использование боковых полос первого порядка в габоровской микроскопии.
8. Leith E. N., Upatnieks J., Journ. Opt. Soc. Am., **52**, 1123 (1962).
Восстановление волновых фронтов и теория связи
9. Leith E. N., Upatnieks J., Journ. Opt. Soc. Am., **54**, 1295 (1964).
Восстановление волнового фронта при диффузном освещении трехмерных объектов.
10. Collier R. J., IEEE Spectrum, **7**, 67 (1966).
Современное состояние голографии.
11. Gabor D., Goss W. P., Journ. Opt. Soc. Am., **56**, 849 (1966).
Интерференционный микроскоп с полным восстановлением волнового фронта.
12. Leith E. N., Upatnieks J., Journ. Opt. Soc. Am., **53**, 1377 (1963).
Восстановление волнового фронта полутонных объектов.
13. Leith E. N., Upatnieks J., Haines K. A., Journ. Opt. Soc. Am., **55**, 981 (1965).
Микроскопия с помощью восстановления волнового фронта.
14. Meier R. W., Journ. Opt. Soc. Am., **55**, 987 (1963).
Увеличение и абберации третьего порядка в голографии.
15. Van Ligten R. F., Journ. Opt. Soc. Am., **56**, 1 (1966).
Влияние свойств фотопленки на восстановление волнового фронта. Плоские волновые фронты,

16. Van Ligtен R. F., Journ. Opt. Soc. Am., 56, 1009 (1966).
Влияние свойств фото пленки на восстановление волнового фронта. «Цилиндрические» волновые фронты.
17. Stroke G. W., An Introduction to Coherent Optics and Holography, New York, 1966. (Имеется перевод: Дж. Строук, Введение в когерентную оптику и голографию, изд-во «Мир», 1969.)
18. Goodman J. W., Journ. Opt. Soc. Am., 57, 493 (1967).
Влияние зернистости пленки на изображение, полученное методом восстановления волнового фронта.
19. Kozma A., Journ. Opt. Soc. Am., 56, 428 (1966).
Фотографическая запись пространственно модулированного когерентного света.
20. Davenport W. B., Jr., Root W. L., Random Signals and Noise, New York, 1958, Ch. 12, 13.
21. Middleton D., Statistical Communication Theory, New York, 1960, Ch. 5.
22. Magnus W., Oberhettinger F., Formulas and Theorems for the Functions of Mathematical Physics, New York, 1954, p. 18.
23. Davenport W. B. Jr., Journ. Appl. Phys., 24, 720 (1953).
Отношение сигнала к шуму в узкополосных фильтрах.
24. Friesem A. A., Zelenka J. S., Appl. Opt., 6, 1755 (1967).
Роль нелинейности пленки в голографии.
25. Goodman J. W., Journ. Opt. Soc. Am., 57, 560 (1967).
Влияние нелинейности пленки на изображения диффузных объектов, полученные методом восстановления волновых фронтов.
26. Knight G., Doctoral dissertation, Stanford University, 1967.
Роль нелинейности пленки в голографии.
27. Goodman J. W., Appl. Opt., 6, 857 (1967).
Свойства голограммы как временного фильтра.
28. Papoulis A., The Fourier Integral and Its Applications, New York, 1963, p. 27.
29. Mertz L., Young N. O., Proc. Conf. Optical Instruments and Techniques, ed. K. J. Habell, New York, 1963, p. 305.
Френелевское преобразование изображений.
30. Lohmann A. W., Journ. Opt. Soc. Am., 55, 1555 (1965).
Восстановление волновых фронтов некогерентных объектов.
31. Strowe G. W., Restrick R. C., Appl. Phys. Lett., 7, 229 (1965).
Голография в пространственно-некогерентном свете.
32. Cochrane G., Journ. Opt. Soc. Am., 56, 1513 (1966).
Новый метод осуществления преобразования Френеля в пространственно-некогерентном свете.
33. Lippmann G., Journ. Physique, 3, 97 (1894).
34. Денисюк Ю. Н., ДАН СССР, 144, № 6, 1275 (1962).
Об отображении оптических свойств объекта в волновом поле рассеянного им излучения.
35. Stroke G. W., Labeyrie A. E., Phys. Lett., 20, 368 (1966).
Восстановление голографических изображений в белом свете с помощью дифракционного эффекта Липпмана — Брэгга.
36. Lin L. H. et al., Bell System Techn. Journ., 45, 659 (1966).
Восстановление цветного голографического изображения при освещении белым светом,

37. Leith E. N. et al., Appl. Opt., 5, 1303 (1966).
Голографическая запись данных в трехмерной среде.
38. Tyler G. L., Journ. Geophys. Res., 71, 1559 (1966).
Двухлучевой метод непрерывной локализации для изучения поверхностей планет.
39. Kreuzer J. L., Proc. Symp. Modern Optics, New York (в печати).
Получение трехмерных изображений с помощью ультразвуковой голографической техники.
40. Gabor D. et al., Phys. Lett., 18, 116 (1965).
Синтез оптических изображений с помощью голографической техники фурье-преобразования (сложение и вычитание комплексных амплитуд).
41. Brooks R. E., Heflinger L. O., Wuerker R. F., IEEE Journ. Quantum Electron, QE-2, 275 (1966).
Получение голограмм с помощью импульсного лазера.
42. Powell R. L., Stetson K. A., Journ. Opt. Soc. Am., 55, 1593 (1965).
Интерферометрическое изучение вибраций с помощью восстановления волнового фронта.
43. Крех С., Science, 153, 989 (1966).
Голографическая микроскопия как методика записи подвижных микрообъектов.
44. Thompson B. J., Ward J. H., Zinky W. R., Appl. Opt., 6, 519 (1967).
Использование голографической техники для изучения размеров частиц.
45. Hildebrand V. P., Haines K. A., Journ. Opt. Soc. Am., 57, 155 (1967).
Применение многочастотной голографии и голографии с несколькими источниками для нанесения горизонталей.
46. Leith E. N., Upatnieks J., Journ. Opt. Soc. Am., 56, 523 (1966).
Получение голографических изображений через рассеивающую среду.
47. Kogelnik H., Bell Syst. Techn. Journ., 44, 2451 (1965).
Голографическое проектирование изображений через неоднородную среду.
48. Upatnieks J., Vander Lugt A., Leith E., Appl. Opt., 5, 589 (1966).
Применение голограмм для исправления аберраций линз.
49. Goodman J. W. et al., Appl. Phys. Lett., 8, 311 (1966).
Получение изображений восстановлением волнового фронта через неоднородную среду.
- 50*. Лейт Е., Упатниекс Дж., УФН, 87, вып. 3 (1965).
Фотографирование с помощью лазера.
- 51*. Налимов И. П., Лазерная голография — основные принципы и применения, Зарубежная радиоэлектроника (2) (1966).
- 52*. Смородинский Я. А., Сороко Л. М., Препринт ОИЯИ, Физика голографии, Р-4-4213, Дубна, 1968.
- 53*. Мериц Л., Интегральные преобразования в оптике, изд-во «Мир», 1969.

ПРИЛОЖЕНИЯ

I. ДЕЛЬТА-ФУНКЦИИ ДИРАКА

Широко используемую при анализе электрических цепей одномерную δ -функцию Дирака можно определить как предел последовательности импульсов убывающей ширины, увеличивающейся высоты и единичной площади. Существует, конечно, большое количество импульсов различной формы, удовлетворяющих такому определению; например, три определения

$$\delta(t) = \lim_{N \rightarrow \infty} N \exp(-N^2 \pi t^2), \quad (\text{П.1a})$$

$$\delta(t) = \lim_{N \rightarrow \infty} N \text{rect}(Nt), \quad (\text{П.1б})$$

$$\delta(t) = \lim_{N \rightarrow \infty} N \text{sinc}(Nt) \quad (\text{П.1в})$$

одинаково приемлемы.

В анализе цепей δ -функции используются для описания мощных импульсов тока или напряжения. Аналогичным понятием в оптике является точечный источник света, или *пространственный* импульс единичной площади. Определение δ -функции в двумерном пространстве есть просто обобщение определения одномерной δ -функции, хотя в этом случае существует даже еще большая свобода выбора функциональной формы импульсов. Возможны, например, следующие определения пространственной δ -функции:

$$\delta(x, y) = \lim_{N \rightarrow \infty} N^2 \exp[-N^2 \pi (x^2 + y^2)], \quad (\text{П.2a})$$

$$\delta(x, y) = \lim_{N \rightarrow \infty} N^2 \text{rect}(Nx) \text{rect}(Ny), \quad (\text{П.2б})$$

$$\delta(x, y) = \lim_{N \rightarrow \infty} N^2 \text{sinc}(Nx) \text{sinc}(Ny), \quad (\text{П.2в})$$

$$\delta(x, y) = \lim_{N \rightarrow \infty} \frac{N^2}{\pi} \text{circ}(N \sqrt{x^2 + y^2}), \quad (\text{П.2г})$$

$$\delta(x, y) = \lim_{N \rightarrow \infty} N \frac{J_1(2\pi N \sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}. \quad (\text{П.2д})$$

В определениях (П. 2а) — (П. 2в) переменные разделяются в прямоугольной системе координат, а в определениях (П. 2г) и (П. 2д) они входят симметрично. В некоторых приложениях одно из определений может оказаться более удобным, чем другие, и тогда для данной проблемы может быть выбрано наиболее подходящее из них.

Каждое из вышеприведенных определений имеет следующие основные свойства:

$$\delta(x, y) = \begin{cases} \infty & x = y = 0, \\ 0 & \text{в других точках;} \end{cases} \quad (\text{П. 3})$$

$$\int_{-e}^{+e} \int \delta(x, y) dx dy = 1, \quad e > 0; \quad (\text{П. 4})$$

$$\int_{-\infty}^{+\infty} \int g(\xi, \eta) \delta(x - \xi, y - \eta) d\xi d\eta = g(x, y) \quad (\text{П. 5})$$

во всех точках непрерывности функции g .

Свойство (П. 5) часто называют *фильтрующим* свойством δ -функции. С помощью любого из определений можно доказать еще одно важное свойство (см. задачу 2.10), а именно

$$\delta(ax, by) = \frac{1}{|ab|} \delta(x, y). \quad (\text{П. 6})$$

Нет никаких причин, по которым нельзя было бы определить δ -функцию в пространстве с более высоким числом измерений, однако их свойства совершенно схожи со свойствами их аналогов в пространстве меньшего числа измерений.

II. ДОКАЗАТЕЛЬСТВО ТЕОРЕМ О ФУРЬЕ-ПРЕОБРАЗОВАНИИ

Дадим краткое доказательство теорем о фурье-преобразовании. Более строгое доказательство приведено в работах [2.1, 2.2].

1. Теорема линейности. $\mathcal{F}\{\alpha g + \beta h\} = \alpha \mathcal{F}\{g\} + \beta \mathcal{F}\{h\}$.
Доказательство. Эта теорема непосредственно

следует из линейности интегралов, с помощью которых определяется фурье-преобразование.

2. Теорема подобия. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\mathcal{F}\{g(ax, by)\} = \frac{1}{|ab|} G\left(\frac{f_x}{a}, \frac{f_y}{b}\right).$$

Доказательство.

$$\begin{aligned} \mathcal{F}\{g(ax, by)\} &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(ax, by) \exp[-j2\pi(f_x x + f_y y)] dx dy = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(ax, by) \exp\left[-j2\pi\left(\frac{f_x}{a} ax + \frac{f_y}{b} by\right)\right] \frac{dax}{|a|} \frac{dby}{|b|} = \\ &= \frac{1}{|ab|} G\left(\frac{f_x}{a}, \frac{f_y}{b}\right). \end{aligned}$$

3. Теорема смещения. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\mathcal{F}\{g(x-a, y-b)\} = G(f_x, f_y) \exp[-j2\pi(f_x a + f_y b)].$$

Доказательство.

$$\begin{aligned} \mathcal{F}\{g(x-a, y-b)\} &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x-a, y-b) \exp[-j2\pi(f_x x + f_y y)] dx dy = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x', y') \exp\{-j2\pi[f_x(x'+a) + f_y(y'+b)]\} dx' dy' = \\ &= G(f_x, f_y) \exp[-j2\pi(f_x a + f_y b)]. \end{aligned}$$

4. Теорема Парсеваля. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)|^2 dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |G(f_x, f_y)|^2 df_x df_y.$$

Доказательство.

$$\begin{aligned}
 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |g(x, y)|^2 dx dy &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) g^*(x, y) dx dy = \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\xi d\eta G(\xi, \eta) \exp[j2\pi(x\xi + y\eta)] \right] \times \\
 &\times \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\alpha d\beta G^*(\alpha, \beta) \exp[-j2\pi(x\alpha + y\beta)] \right] = \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\xi d\eta G(\xi, \eta) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\alpha d\beta G^*(\alpha, \beta) \times \\
 &\times \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\{j2\pi[x(\xi - \alpha) + y(\eta - \beta)]\} dx dy \right] = \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\xi d\eta G(\xi, \eta) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\alpha d\beta G^*(\alpha, \beta) \delta(\xi - \alpha, \eta - \beta) = \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |G(\xi, \eta)|^2 d\xi d\eta.
 \end{aligned}$$

5. Теорема свертки. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$ и $\mathcal{F}\{h(x, y)\} = H(f_x, f_y)$, то

$$\mathcal{F} \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta \right\} = G(f_x, f_y) H(f_x, f_y).$$

Доказательство.

$$\begin{aligned}
 \mathcal{F} \left\{ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta \right\} &= \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\xi, \eta) \mathcal{F}\{h(x - \xi, y - \eta)\} d\xi d\eta = \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\xi, \eta) \exp[-j2\pi(f_x \xi + f_y \eta)] d\xi d\eta H(f_x, f_y) = \\
 &= G(f_x, f_y) H(f_x, f_y).
 \end{aligned}$$

6. Теорема автокорреляции. Если $\mathcal{F}\{g(x, y)\} = G(f_x, f_y)$, то

$$\mathcal{F}\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(\xi, \eta) g^*(\xi - x, \eta - y) d\xi d\eta\right\} = |G(f_x, f_y)|^2.$$

Доказательство.

$$\begin{aligned} \mathcal{F}\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(\xi, \eta) g^*(\xi - x, \eta - y) d\xi d\eta\right\} &= \\ &= \mathcal{F}\left\{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(\xi' + x, \eta' + y) g^*(\xi', \eta') d\xi' d\eta'\right\} = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\xi' d\eta' g^*(\xi', \eta') \mathcal{F}\{g(\xi' + x, \eta' + y)\} = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\xi' d\eta' g^*(\xi', \eta') \exp[j2\pi(f_x \xi' + f_y \eta')] G(f_x, f_y) = \\ &= G^*(f_x, f_y) G(f_x, f_y) = |G(f_x, f_y)|^2. \end{aligned}$$

7. Интегральная теорема Фурье. Во всех точках, где функция g непрерывна,

$$\mathcal{F}\mathcal{F}^{-1}\{g(x, y)\} = \mathcal{F}^{-1}\mathcal{F}\{g(x, y)\} = g(x, y).$$

В каждой точке, в которой функция g разрывна, с помощью двух последовательных преобразований можно получить среднее значение g в окрестности этой точки.

Доказательство. Пусть функция $g_R(x, y)$ определена как

$$g_R(x, y) = \iint_{A_R} G(f_x, f_y) \exp[j2\pi(f_x x + f_y y)] df_x df_y,$$

где A_R — круг радиусом R с центром в начале координат плоскости $f_x f_y$. Чтобы доказать теорему, достаточно продемонстрировать, что в точках, где функция g непрерывна,

$$\lim_{R \rightarrow \infty} g_R(x, y) = g(x, y)$$

и что в точках, где функция g разрывна,

$$\lim_{R \rightarrow \infty} g_R(x, y) = \frac{1}{2\pi} \int_0^{2\pi} g_0(\theta) d\theta,$$

где $g_0(\theta)$ — угловая зависимость g в малой окрестности вблизи исследуемой точки.

Проведем некоторые предварительные преобразования:

$$\begin{aligned} g_R(x, y) &= \iint_{A_R} \left\{ \iint_{-\infty}^{+\infty} d\xi d\eta g(\xi, \eta) \exp[-j2\pi(f_X \xi + f_Y \eta)] \right\} \times \\ &\quad \times \exp[j2\pi(f_X x + f_Y y)] df_X df_Y = \\ &= \iint_{-\infty}^{+\infty} d\xi d\eta g(\xi, \eta) \iint_{A_R} df_X df_Y \exp\{j2\pi[f_X(x - \xi) + f_Y(y - \eta)]\}. \end{aligned}$$

Замечая, что

$$\iint_{A_R} df_X df_Y \exp\{j2\pi[f_X(x - \xi) + f_Y(y - \eta)]\} = R \left[\frac{J_1(2\pi Rr)}{r} \right],$$

где $r = \sqrt{(x - \xi)^2 + (y - \eta)^2}$, имеем

$$g_R(x, y) = \iint_{-\infty}^{+\infty} d\xi d\eta g(\xi, \eta) R \left[\frac{J_1(2\pi Rr)}{r} \right].$$

Положим сначала, что функция g непрерывна в точке (x, y) . Тогда

$$\begin{aligned} \lim_{R \rightarrow \infty} g_R(x, y) &= \iint_{-\infty}^{+\infty} d\xi d\eta g(\xi, \eta) \lim_{R \rightarrow \infty} R \left[\frac{J_1(2\pi Rr)}{r} \right] = \\ &= \iint_{-\infty}^{+\infty} d\xi d\eta g(\xi, \eta) \delta(x - \xi, y - \eta) = g(x, y), \end{aligned}$$

где на второй стадии мы воспользовались соотношением (П. 2д). Таким образом, первая часть теоремы доказана.

Рассмотрим теперь точку, где функция g терпит разрыв. Без ограничения общности можно считать, что эта точка является началом координат.

Следовательно, можно записать, что

$$g_R(0, 0) \approx \int_{-\infty}^{+\infty} \int d\xi d\eta g(\xi, \eta) R \left[\frac{J_1(2\pi Rr)}{r} \right],$$

где $r = \sqrt{\xi^2 + \eta^2}$. Но для достаточно больших R выражение в скобках заметно отличается от нуля только в малой окрестности начала координат. Кроме того, в этой малой окрестности функция g зависит (приближенно) только от угла θ вблизи этой точки, поэтому

$$g_R(0, 0) \approx \int_0^{2\pi} g_0(\theta) d\theta \int_0^{\infty} rR \left[\frac{J_1(2\pi Rr)}{r} \right] dr,$$

где $g_0(\theta)$ дает зависимость g от θ вблизи начала координат. Наконец, замечая, что

$$\int_0^{\infty} rR \left[\frac{J_1(2\pi Rr)}{r} \right] dr = \frac{1}{2\pi},$$

находим

$$\lim_{R \rightarrow \infty} g_R(0, 0) = \frac{1}{2\pi} \int_0^{2\pi} g_0(\theta) d\theta,$$

и теорема полностью доказана.

ОГЛАВЛЕНИЕ

Предисловие редактора перевода	5
Предисловие	11
<i>Глава 1. ВВЕДЕНИЕ</i>	13
§ 1. Оптика и теория связи	13
§ 2. Об этой книге	14
<i>Глава 2. АНАЛИЗ ДВУМЕРНЫХ ЛИНЕЙНЫХ СИСТЕМ</i>	17
§ 1. Двумерный анализ Фурье	18
Определение и условия существования	18
Преобразование Фурье как разложение	21
Теоремы относительно преобразования Фурье	23
Функции с разделяющимися переменными	25
Функции, обладающие осевой симметрией. Преобразование Фурье — Бесселя	26
Некоторые часто используемые функции и полезные пары преобразований Фурье	29
§ 2. Линейные системы	33
Линейность и интеграл суперпозиции	34
Инвариантные линейные системы. Передаточные функции	36
§ 3. Двумерная теория выборки	39
Теорема Уиттекера — Шеннона	39
Задачи	44
Литература	48
<i>Глава 3. ОСНОВЫ СКАЛЯРНОЙ ТЕОРИИ ДИФРАКЦИИ</i>	50
§ 1. Исторический обзор	50
§ 2. Предварительные математические сведения	54
Уравнение Гельмгольца	54
Теорема Грина	56
Интегральная теорема Гельмгольца — Кирхгофа	56
§ 3. Формулировка задачи о дифракции на плоском экране, предложенная Кирхгофом	59
Применение интегральной теоремы	59
Граничные условия Кирхгофа	62
Формула дифракции Френеля — Кирхгофа и принцип Гюйгенса — Френеля	63
§ 4. Формулировка Релея — Зоммерфельда задачи о дифракции на плоском экране	65
Выбор альтернативных функций Грина	66

	Формула дифракции Релея — Зоммерфельда	68
	Дальнейшее обсуждение принципа Гюйгенса—Френеля	69
§ 5.	Обобщение теории Релея — Зоммерфельда на случай немонохроматических волн	70
§ 6.	Дифракция на границах	72
§ 7.	Угловой спектр плоских волн	74
	Угловой спектр и его физический смысл	74
	Распространение углового спектра	76
	Влияние ограничивающего отверстия на угловой спектр возмущения	78
	Линейный пространственный фильтр, эквивалентный распространению волн	79
	Задачи	81
	Литература	82
Глава 4. ДИФРАКЦИЯ ФРЕНЕЛЯ И ФРАУНГОФЕРА		84
§ 1.	Приближение, даваемое принципом Гюйгенса—Френеля	84
	Исходные приближения	84
	Приближения Френеля	86
	Приближение Фраунгофера	89
§ 2.	Примеры дифракционных картин Фраунгофера	90
	Прямоугольное отверстие	91
	Круглое отверстие	93
	Синусоидальная амплитудная решетка	95
	Синусоидальная фазовая решетка	99
§ 3.	Пример дифракционной картины Френеля	101
	Задачи	106
Глава 5. ЛИНЗЫ КАК ЭЛЕМЕНТЫ, ВЫПОЛНЯЮЩИЕ ПРЕ- ОБРАЗОВАНИЕ ФУРЬЕ И СОЗДАЮЩИЕ ИЗОБРА- ЖЕНИЕ		109
§ 1.	Тонкая линза как элемент, выполняющий фазовое пре- образование	109
	Функция толщины	110
	Параксиальное приближение	112
	Фазовое преобразование и его физический смысл	112
§ 2.	Линзы как элементы, выполняющие преобразование Фурье	116
	Предмет, расположенный вплотную к линзе	116
	Предмет, расположенный перед линзой	119
	Предмет, расположенный за линзой	122
§ 3.	Формирование изображения. Монохроматическое осве- щение	125
	Импульсный отклик положительной линзы	125
	Устранение квадратичных фазовых множителей. Фор- мула линзы	127
	Соотношение между предметом и изображением	130
	Задачи	133
	Литература	137

Глава 6. ЧАСТОТНЫЙ АНАЛИЗ ОПТИЧЕСКИХ СИСТЕМ . . .	138
§ 1. Обобщение представления о системах, образующих изображение	139
Обобщенная модель	139
Дифракционные эффекты	141
Немонохроматическое освещение. Когерентный и некогерентный случаи	144
§ 2. Частотный отклик дифракционно ограниченной когерентной оптической системы	150
Передаточная функция при когерентном освещении	150
Примеры передаточных функций при когерентном освещении для дифракционно ограниченных систем	152
§ 3. Частотный отклик дифракционно ограниченной некогерентной оптической системы	153
Оптическая передаточная функция и ее связь с когерентной передаточной функцией	154
Основные свойства ОПФ	156
ОПФ системы без aberrаций	158
Примеры ОПФ дифракционно ограниченных систем	159
§ 4. Aberrации и их влияние на частотный отклик	162
Обобщенная функция зрачка	163
Влияние aberrаций на когерентную передаточную функцию	164
Влияние aberrаций на ОПФ	165
Пример простой aberrации — ошибка фокусировки	166
§ 5. Сравнение формирования изображения при когерентном и некогерентном освещении	170
Частотный спектр интенсивности изображения	171
Разрешение двух точек	174
Другие эффекты	176
§ 6. Разрешающая способность, превышающая классический дифракционный предел	179
Основные математические положения	180
Пример схемы восстановления	180
Практические ограничения	182
Задачи	182
Литература	186
Глава 7. ПРОСТРАНСТВЕННАЯ ФИЛЬТРАЦИЯ И ОПТИЧЕСКАЯ ОБРАБОТКА ИНФОРМАЦИИ	188
§ 1. Исторические замечания	189
Эксперимент Аббе — Портера	189
Фазово-контрастный микроскоп Цернике	193
Улучшение качества фотографий. Метод Марешаля	195
Возникновение подхода, основанного на теории связи	197
Применение когерентной оптики к более общим задачам обработки информации	198
§ 2. Фотографическая пленка	199
Физические основы экспонирования	200
Определение основных величин	200

	Пленка в некогерентной оптической системе	203
	Пленка в когерентной оптической системе	205
	Модуляционная передаточная функция	209
§ 3.	Некогерентные системы обработки информации, основанные на геометрической оптике	212
	Передача изображения	213
	Свертка без перемещений	215
	Синтез импульсных откликов в расфокусированных системах	217
	Ограничения	218
§ 4.	Синтез в частотном пространстве	219
	Некоторые виды когерентных систем	221
	Ограничения	226
§ 5.	Фильтры Вандер Люгта	227
	Синтез маски в частотной плоскости	228
	Обработка данных, поступающих на вход	231
	Преимущества фильтра Вандер Люгта	234
§ 6.	Приложение к распознаванию образов	235
	Согласованный фильтр	236
	Проблема распознавания образов	238
	Оптический синтез устройства для распознавания образов	240
	Чувствительность к изменению масштаба и вращению	243
§ 7.	Обработка данных антенн синтетической апертуры	243
	Создание синтетической апертуры	243
	Прием и развертка сигналов	245
	Фокусирующие свойства пленки	248
	Получение конечного изображения	251
	Заключительные замечания	253
	Задачи	254
	Литература	259

Глава 8. ПОЛУЧЕНИЕ ИЗОБРАЖЕНИЯ ВОССТАНОВЛЕНИЕМ ВОЛНОВОГО ФРОНТА, ИЛИ ГОЛОГРАФИЯ 261

§ 1.	Историческое введение	261
§ 2.	Проблема восстановления волнового фронта	262
	Запись амплитуды и фазы	263
	Материалы для записи	264
	Восстановление исходного волнового фронта	265
	Линейность процесса восстановления волнового фронта	267
	Получение изображений путем восстановления волнового фронта	267
§ 3.	Голограммы Габора	270
	Источник опорной волны	270
	Сопряженные изображения	272
	Недостатки голограммы Габора	273
§ 4.	Голограмма Лейта и Упэтника	274
	Запись голограммы	275
	Получение восстановленных изображений	276
	Минимальный опорный угол	280

	Обобщение на случай неколлимированных опорной и восстанавливающей волн	282
	Приложение результатов к трехмерной фотографии	287
	Экспериментальные трудности	292
§ 5.	Влияние МПФ пленки	295
	Коллимированная опорная волна	295
	Безлинзовая фурье-голограмма	297
	Более общие схемы записи	301
§ 6.	Эффекты, связанные с нелинейностью пленки	301
	Модель Косма	302
	Анализ методом преобразования	304
	Прямой метод анализа. Резкий контраст	306
	Влияние нелинейности на восстановление изображения	312
§ 7.	Влияние толщины эмульсии	315
	Элементарная голограмма	315
	Условие Брэгга	316
§ 8.	Обобщения	320
	Голография движущихся объектов	320
	Голография в пространственно-некогерентном свете	322
	Отражательная голограмма	326
§ 9.	Применения голографии	329
	Микроскопия	330
	Интерферометрия	331
	Анализ вибраций	334
	Объемное изображение высокого разрешения	336
	Нанесение горизонталей на трехмерные объекты	337
	Передача изображения через искажающую среду	339
	Задачи	347
	Литература	350
	Приложения	353
	I. Дельта-функции Дирака	353
	II. Доказательство теорем о фурье-преобразовании	354

Дж. ГУДМЕН

Введение в фурье-оптику

Редактор *И. НАХИМСОН*

Переплет художника *С. Бычкова* Художественный редактор *П. Некундэ*
Технический редактор *Н. Новлева* Корректор *А. Рыбальченко*

Сдано в производство 22/IV 1970 г. Подписано к печати 13/XI 1970 г.
Бумага № 2 84×108¹/₁₆ = 5,69 бум. л. 19,11 печ. л. Уч.-над. л. 17,49. Изд. № 2/5306.
Цена I р. 39 к. Зак. 593.

ИЗДАТЕЛЬСТВО «МИР», Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени
Ленинградская типография № 2 имени Евгении Соколовой Главполиграфпрома
Комитета по печати при Совете Министров СССР, Измайловский проспект, 29.